Mathematische Konzepte aus der Risikoanalyse

Prof. Erika Hausenblas
Lehrstuhl für Angewandte Mathematik
Montanuniversität Leoben

Inhaltsverzeichnis

Kapitel 1

Eine kurze Einführung in die Wahrscheinlichkeitsrechnung

1 Ereignisse und Wahrscheinlichkeiten

Ein Wahrscheinlichkeitsraum ist ein Triplet(Ω,$\mathcal{F}$,$\mathbb{P}$), wobei Ω der Ergebnisraum ist und die Menge aller enthält. Die Definition haben Sie schon in der Vorlesung von Prof. Kirschenhofer kennengelernt. Wir wiederholen hier aber die notwendigsten Begriffe.

i Ereignisse als Mengen

Wir modellieren Ereignisse als Mengen. Dabei ist eine Menge eine Zusammenfassung von wohldefinierten und unterscheidbaren Dingen (Elemente) zu einem Ganzen. Die Menge Ω, die alle möglichen Ereignisse (Elementarereignisse) beinhaltet ist die Grundmenge, bzw. der Ergebnisraum. Weiters ist $\mathcal{F}$ die Menge aller möglichen Ereignisse und $\mathbb{P}$ eine Funktion auf $\mathcal{F}$ die jedem möglichen Ereignis einen Wert, die Wahrscheinlichkeit mit der dieses Ereignis eintritt, zuordnet.

Schreibweise:

- Ω Grundmenge, ω Element
- $\omega \in \Omega$: ω ist Element von Ω
- $\omega \notin \Omega$: ω ist nicht Element von Ω
- $A = \{a, b, c, d, \cdots\}$: Die Menge A besteht aus den Elementen $a, b, c, \ldots,$
- $A = \{\omega \in \Omega : \omega$ hat Eigenschaft $E\}$: A besteht aus denjenigen Elementen ω von Ω, die die Eigenschaft E haben.

Beispiel 1.1. $\Omega = \mathbb{N}$*;* $A = \{2; 4; 6; \ldots\} = \{n \in \mathbb{N} : n$ *ist durch* 2 *teilbar*$\}$.

Beispiel 1.2. der Münzwurf: $\Omega = \{K =' Kopf', Z =' Zahl'\}$*;* $A = \{K\} = \{$ *es wurde Kopf geworfen* $\}$.

Beispiel 1.3. der Würfel: $\Omega = \{1, 2, 3, 4, 5, 6\}$*;* $A = \{1, 2, 3\} = \{$ *es wurde eine Zahl kleiner als vier geworfen* $\}$.

Der Vergleich von Ereignissen erfolgt durch den Vergleich der Mengen, durch die die Ereignisse modelliert werden. Hierbei verwenden wir folgende Notation:

1. $A_1 \subset A_2$ bedeutet, A_1 ist Teilmenge von A_2, d.h., aus $\omega \in A_1$ folgt $\omega \in A_2$;
2. $A_1 \supset A_2$ bedeutet, A_2 ist Teilmenge von A_1, d.h., aus $\omega \in A_2$ folgt $\omega \in A_1$;
3. $A_1 = A_2$, falls $A_1 \subset A_2$ und $A_1 \supset A_2$;

Beispiel 1.4. *Betrachtet werden Ereignisse, die bei einem Zufallsexperiment (z.B. Münzwurf, Werfen eines Würfels, Roulette-Spiel, Erzeugen einer Pseudozufallszahl mit einem Zufallszahlengenerator) eintreten können. Dann ist*

- Ω = *Menge aller möglichen Versuchsergebnisse (Grundmenge, Grundgesamtheit, Merkmalraum, Stichprobenraum, Ereignisraum);* Ω *sicheres Ereignis (tritt immer ein)*
- $A_1, A_2 \subset \Omega$*: Ereignisse = Teilmengen von Versuchsergebnissen mit bestimmten Eigenschaften;*
- $\{\omega\} \in \Omega$*: Elementarereignis = ein (einzelnes) Versuchsergebnis;*
- *Angenommen: bei einem Versuch wird das Ergebnis* ω *erzielt. Dann sagen wir: Das Ereignis A tritt ein, falls* $\omega \in A$*.*
- *Für* $A_1 \subset A_2$ *gilt: Wenn* A_1 *eintritt, dann tritt auch* A_2 *ein.*

Beispiel 1.5. Der Würfel:*:* $\{\Omega = \{1, 2, 3, 4, 5, 6\}$*; Elementar–Ereignisse* $\{1\}, \{2\}, \cdots, \{6\}$*. Das Ereignis* $A_1 = \{2\}$ *tritt genau dann ein, wenn die Zahl 2 gewürfelt wird. Das Ereignis* $A_2 = \{2, 4, 6\}$ *tritt genau dann ein, wenn eine gerade Zahl gewürfelt wird. Also gilt:* $A_1 \subset A_2$*, d.h., wenn* A_1 *eintritt, dann tritt auch* A_2 *ein.*

Man hat also den Ereignisraum Ω. Diejenige Teilmenge von Ω, die kein Element enthält, heißt leere Menge und wird mit $\emptyset$ bezeichnet. Das Ereignis $\emptyset$ tritt niemals ein und wird deshalb unmögliches Ereignis genannt. Weiters haben wir die Menge $\mathcal{F}$ aller möglichen Ereignisse die eintreffen können aber nicht müssen, weiters liegt die leere Menge $\emptyset$ in $\mathcal{F}$ und der gesamte Raum Ω gehört zu $\mathcal{F}$. Zu jedem Ereignis $A \in \mathcal{F}$ kann man ein Gegenereignis $A^c = \Omega \setminus A$ bilden, dies gehört auch zu $\mathcal{F}$.

ii Ereignissysteme

Aus gegebenen Ereignissen $A_1, A_2, \ldots$ kann man durch deren *Verknüpfung* weitere Ereignisse bilden. Dies wird durch die folgenden Mengenoperationen modelliert.

Mengenoperationen und ihre probabilistische Bedeutung

1. *Vereinigungsmenge:* $A_1 \cup A_2$: Menge aller Elemente, die zu A_1 oder A_2 gehören.
   (Ereignis $A_1 \cup A_2$ = mindestens eines der Ereignisse A_1 oder A_2 tritt ein)
   $\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots =$ Menge aller Elemente, die zu mindestens einer der Mengen A_i gehören.
   (Ereignis $\cup_{i=1}^{\infty} A_i$ = mindestens eines der Ereignisse $A_1, A_2, \cdots$ tritt ein)
2. *Schnittmenge* $A_1 \cap A_2$: Menge aller Elemente, die zu A_1 und A_2 gehören
   (Ereignis $A_1 \cap A_2$ = beide Ereignisse A_1 und A_2 treten ein).
   Beachte: Zwei Ereignisse $A_1, A_2 \in \Omega$ mit $A_1 \cap A_2 = \emptyset$ heißen unvereinbar (disjunkt), d.h., sie können nicht gleichzeitig eintreten.
   $\cap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots =$ Menge aller Elemente, die zu jeder der Mengen A_i gehören. (Ereignis $\cap_{i=1}^{\infty} A_i$ = sämtliche Ereignisse $A_1, A_2, \ldots$ treten ein)
3. *Differenzmenge:* $A_1 \setminus A_2$: Menge aller Elemente von A_1, die nicht zu A_2 gehören. Spezialfall: $A^c = \Omega \setminus A$ (Komplement) (Ereignis A^c = Ereignis A tritt nicht ein)
4. *Symmetrische Mengendifferenz:* $A_1 \Delta A_2$: Menge aller Elemente, die zu A_1 oder A_2, jedoch *nicht* zu beiden gehören.

Beispiel 1.6. *Sei* $\Omega = \{a, b, c, d\}$*,* $A_1 = \{a, b, c\}$*;* $A_2 = \{b, d\}$*. Dann gilt* $A_1 \cup A_2 = \{a, b, c, d\}$*;* $A_1 \cap A_2 = \{b\}$*;* $A_1 \setminus A_2 = \{a, c\}$*;* $A_1^c = \{d\}$*;* $A_1 \Delta A_2 = \{a, c, d\}$*.*

Weitere Eigenschaften: Seien $A, B, C \subset \Omega$ beliebige Teilmengen. Dann gelten

- *Eindeutigkeitsgesetze:* $A \cup \emptyset = A$; $A \cap \emptyset = \emptyset$; $A \cup \Omega = \Omega$; $A \cap \Omega = A$
  (allgemein: falls $A \subset B$, dann gilt $A \cap B = A$; $A \cup B = B$);
- *de Morgansche Gesetze:* $(A \cup B)^c = A^c \cap B^c$; $(A \cap B)^c = A^c \cup B^c$;
- *Assoziativ Gesetze:* $A \cup (B \cup C) = (A \cup B) \cup C$; $A\ (B \cap C) = (A \cap B) \cap C$;
- *Distributiv Gesetze:* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Es ist oft nicht zweckmäßig, alle möglichen Teilmengen von Ω in die Modellierung einzubeziehen, sondern man betrachtet nur die Familie derjenigen Teilmengen von Ω, deren Wahrscheinlichkeiten tatsächlich von Interesse sind. Diese Mengenfamilie soll jedoch abgeschlossen sein bezüglich der Operationen $\cup, \cap, \setminus$, was durch die folgende Begriffsbildung erreicht wird.

Definition 1.1. *Eine nichtleere Familie $\mathcal{F}$ von Teilmengen von Ω heißt Algebra, falls*

A1: $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;

A2: $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$.

Beispiel 1.7. $\Omega = \{a, b, c, d\}$; $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c, d\}, \Omega\}$ *ist eine Algebra,* $\mathcal{F}_2 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ *ist dagegen keine Algebra.*

Definition 1.2. *Ein Algebra $\mathcal{F}$ heißt σ–Algebra, wenn zusätzlich gilt:*

A3: $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Bemerkung 1.1. *Die Unterscheidung zwischen Algebren und σ–Algebren ist nur wichtig, falls Ω nicht diskret ist (z.B. $\Omega = \mathbb{R}$). In diesem Fall sorgt A3 dafür, dass ein einzelner Punkt $\{a\}$ auch zur σ–Algebra gehören kann. Diese Unterscheidung ist aber eher mathematischer Natur, und hier nicht wichtig.*

Beachte:

- Das Paar $(\Omega, \mathcal{F})$ heißt Maßraum, falls $\mathcal{F}$ eine σ-Algebra ist.
- Für jedes Ω, Ω abzählbar, ist die Potenzmenge $\mathcal{P}(\Omega)$ die Familie aller Teilmengen von Ω, stets eine σ-Algebra.
- Wenn Ω endlich oder abzählbar unendlich ist, dann kann $\mathcal{F} = \mathcal{P}(\Omega)$ gewählt werden. Bei nicht abzählbarem Ω (z.B. $\Omega = \mathbb{R}$ oder $\Omega = [0, 1]$) muss eine kleinere σ-Algebra betrachtet werden (nicht $\mathcal{P}(\Omega)$), wir werden hier aber darauf nicht näher eingehen.

iii Definition und elementare Eigenschaften

Gegeben sei ein Maßraum $(\Omega, \mathcal{F})$. Betrachten wir eine Mengenfunktion, d.h. eine Abbildung $\mathbb{P} : \mathcal{F} \to [0, 1]$, die jeder Menge $A \in \mathcal{F}$ eine Zahl $\mathbb{P}(A) \in [0, 1]$ zuordnet. Dann heißt $\mathbb{P}(A)$ Wahrscheinlichkeit des Ereignisses $A \in \mathcal{F}$.

Beispiel 1.1. Der Münzwurf: *Sei $\Omega = \{K, Z\}$ und $\mathcal{F}$ sei die Menge aller möglichen Teilmengen von Ω, d.h. $\mathcal{F} = \{\{K, Z\}, \{K\}, \{Z\}, \emptyset\}$. Das Wahrscheinlichkeitsmaß $\mathbb{P} : \mathcal{F} \to [0, 1]$ definieren wir folgendermaßen: Sei $A \in \mathcal{F}$. Dann setzen wir $\mathbb{P}(A) = |A|/2$.*

Beispiel 1.2. Der Würfel: *Sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $\mathcal{F}$ sei die Menge aller möglichen Teilmengen von Ω, d.h. $\mathcal{F} = \{\Omega, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \{1, 2, 3, 5, 6\}, \ldots, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \emptyset\}$. Das Wahrscheinlichkeitsmaß definieren wir folgendermaßen: Sei $A \in \mathcal{F}$. Dann setzen wir $\mathbb{P}(A) = |A|/6$.*

Definition 1.3. *(Axiome von Kolmogorow) Die Mengenfunktion $\mathbb{P} : \mathcal{F} \to [0,1]$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{F}$, falls*

(P1) $\mathbb{P}(\Omega) = 1$ *(Normiertheit)*

(P2) $\mathbb{P}(\cup_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ *für paarweise disjunkte Mengen* $A_1, A_2, \ldots \in \mathcal{F}$ *(σ-Additivität)*

Wenn $(\Omega, \mathcal{F})$ ein Maßraum und $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$ ist, dann heißt das Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ Wahrscheinlichkeitsraum.

Satz 1.1. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, A_1, A_2, \ldots \in \mathcal{F}$. Dann gilt*

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $A_1 \subset A_2 \Rightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$;
- $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$;
- $\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$;

iv Definition von Zufallsvariablen

Betrachten wir einen beliebigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und ein beliebiges Element $\omega \in \Omega$, wobei wir so wie bisher $\{\omega\}$ als Elementarereignis bzw. Versuchsergebnis interpretieren. Beachte, häufig interessiert nicht ω selbst, sondern eine (quantitative oder qualitative) Kennzahl $X(\omega)$ von ω, d.h., wir betrachten die Abbildung $\omega \to X(\omega)$.

Der Münzwurf aus Beispiel 1.2 und der Würfel aus Beispiel 1.2 sind jeweils auf einen Wahrscheinlichkeitsraum Ω definiert. Man kann aber den Münzwurf mittels einer sogenannten Zufallsvariablen in die reellen Zahlen abbilden:

$$Y : \Omega \ni \omega \mapsto Y(\omega) = \begin{cases} 1 & \text{falls } \omega = K, \\ -1 & \text{falls } \omega = Z. \end{cases} \tag{1.1}$$

Beispiel 1.3. *Sei Ω die Menge aller möglichen Unfallverläufe eines Auffahrtsunfall, und $\mathcal{F}$ die Menge aller möglichen Teilmengen von Ω. Sei*

$$\begin{array}{rcl} X : \Omega & \to & \mathbb{R} \\ X(\omega) & \mapsto & \textit{Kosten die ein Unfall mit Unfallverlauf } \omega \textit{ verursacht.} \end{array}$$

Beispiel 1.4. *Sei Ω die Menge aller möglichen Wetterverläufe eines Tages und $\mathcal{F}$ die Menge aller möglichen Teilmengen von Ω. Sei*

$$\begin{array}{rcl} X : \Omega & \to & \mathbb{R} \\ X(\omega) & \mapsto & \textit{Wassermenge die sich in einen Gefäss am Schwammerlturm mit einen Quadratzentimeter} \\ & & \textit{Grundfläche bei Wetterverlauf } \omega \textit{ angesammelt hat.} \end{array}$$

Beispiel 1.5. *(Perioden zwischen zwei Erdbeben). Die Zeitintervalle in Tagen zwischen aufeinander folgenden schweren Erdbeben werden weltweit aufgezeichnet. Serious bedeutet eine Erdbebenstärke von mindestens* $7,5$ *auf der Richter-Skala oder, dass mehr als* 1000 *Menschen getötet wurden. Insgesamt* 63 *Erdbeben wurden aufgezeichnet, d.h.* 62 *Zwischenzeiten. Diese Datensatz umfasst den Zeitraum vom 16. Dezember 1902 bis 4. März 1977. Wartezeiten modelliert man mit exponentiell verteilte Zufallsvariable.*

Statt den Münzwurf über dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ zusammen mit der Zufallsvariablen X zu betrachten, ist es nun möglich gleich mit der von X induzierten Verteilung auf den reellen Zahlen

$\mathbb{R}$ zu arbeiten. Dazu definieren wir ein sogenanntes Maß P auf den reellen Zahlen. In der Wahrscheinlichkeitstheorie kann man zeigen, dass es genügt, so ein Maß auf der Menge aller halboffenen Intervalle zu definieren. Sei also $-\infty \leq a \leq b < \infty$. Dann sei

$$P((a,b]) = \begin{cases} 1 & \text{falls } 1 \in (a,b] \text{ und } -1 \in (a,b], \\ \frac{1}{2} & \text{falls entweder } 1 \in (a,b] \text{ oder } -1 \in (a,b], \\ 0 & \text{falls weder } 1 \in (a,b] \text{ noch } -1 \in (a,b]. \end{cases}$$

Definition 1.4. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein beliebiger Wahrscheinlichkeitsraum, und $X : \Omega \to \mathbb{R}$ sei eine beliebige Zufallsvariable. Die Verteilung der Zufallsvariablen X ist die Mengenfunktion $P_X : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ mit*

$$P_X(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}).$$

Man kann leicht sehen, dass folgende Beziehung gilt:

$$\mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in [a,b)\}\right) = P([a,b)), \quad -\infty < a \leq b < \infty.$$

In unserem Beispiel war der Wahrscheinlichkeitsraum sehr einfach. Es gibt aber auch weitaus komplexere Wahrscheinlichkeitsräume:

Beispiel 1.6. siehe Beispiel 1.3: *Sei Ω die Menge aller möglichen Unfallverläufe eines Auffahrtunfalls, und $\mathcal{F}$ die Menge aller möglichen Teilmengen von Ω. Sei*

$$\begin{array}{rcl} X : \Omega & \to & \mathbb{R} \\ X(\omega) & \mapsto & \textit{Kosten die ein Unfall mit Unfallverlauf } \omega \textit{ verursacht.} \end{array}$$

Für $\omega \in \Omega$ ist $\mathbb{P}(\omega)$ ist die Wahrscheinlichkeit, dass ein Unfall mit Unfallverlauf ω passiert. Möchte man aber die Kosten eines Unfalls wissen, so interessiert uns die Größe $\mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in (a,b)\}\right)$ für $0 \leq a \leq b$.

Sei $x > 0$. Die Wahrscheinlichkeit dass die Kosten des Unfalls mehr als x Euro betragen, kann durch die Verteilungsfunktion die durch $\mathbb{P}$ auf $\mathbb{R}$ induziert wird, berechnet werden. Damit ist die Verteilungsfunktion F_X[1] aus Beispiel 1.6 gegeben durch

$$F_X(x) := \mathbb{P}\left(X \leq x\right) = P((-\infty, x]), \qquad x \in \mathbb{R}.$$

Bemerkung 1.2. *Man kann zeigen, dass $F_X : \mathbb{R} \to \mathbb{R}$ folgende Eigenschaften besitzt:*

1. *$F_X : \mathbb{R} \to [0,1]$;*
2. *$\lim_{x \to -\infty} F_X(x) = 0$ und $\lim_{x \to \infty} F_X(x) = 1$;*
3. *$F_X(x) \leq F_X(y)$ für $x \leq y$ (F_X ist monoton steigend).*

Umgekehrt, erfüllt eine Funktion F die Punkte in Bemerkung 1.2, so besitzt diese Funktion eine Dichte. Dass heißt, es gibt eine nicht negative Funktion $f : \mathbb{R} \to \mathbb{R}$, sodass gilt

$$F(a) = \int_{-\infty}^{a} f(x)\, dx, \quad -\infty < a < \infty. \tag{1.2}$$

Umgekehrt, ist eine nichtnegative Funktion f die Ableitung von F, d.h.

$$\begin{array}{rcl} f(x) & = & \lim\limits_{\Delta x \to 0} \dfrac{F(x + \Delta x) - F(x)}{\Delta x} \\ & = & \dfrac{\mathrm{d}F(x)}{\mathrm{d}x} \end{array}$$

und $\int_{-\infty}^{\infty} f(x)\, dx = 1$, dann ist F definiert durch (1.2) eine Verteilungsfunktion.

Die Dichte in Beispiel 1.2 wird meistens mit Hilfe der Statistik eruiert, und in der Praxis nimmt man kurzerhand an, dass es diese Dichte gibt, sodass (1.2) erfüllt ist.

[1] Im Skriptum bezeichnen wir die Verteilungsfunktion einer gegebenen Zufallsvariable Y bei F_Y.

Bemerkung 1.3. *Mit Hilfe der Dichtefunktion, bzw. der Verteilungsfunktion lassen sich das erste Moment und das zweite Moment ganz leicht berechnen: Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und F_X bzw. f_X, die zugehörige Verteilungsfunktion, bzw. Dichte. Dann gilt für das* erste Moment *(auch Erwartungswert genannt)*

$$\mu_X \quad = \quad \mathbb{E}\, X = \int_{-\infty}^{\infty} x\, f_X(x)\mathrm{d}x = \int_{-\infty}^{\infty} (1 - F_X(x))\, \mathrm{d}x$$

und für das zweite Moment[2]

$$\mu_{X^2} \quad = \quad \mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2\, f_X(x)\mathrm{d}x = \int_{-\infty}^{\infty} x\, (1 - F_X(x))\, \mathrm{d}x.$$

Des erste Moment heißt auch Erwartungswert *und wird in der Literatur oft auch mit $\mathbb{E}X$ bezeichnet. Analog wird μ_{X^2} oft mit $\mathbb{E}X^2$ bezeichnet.*

Bemerkung 1.4. *Sei X wieder eine Zufallsvariable. Eine weitere wichtige Kenngrösse einer Zustandsvariablen ist deren* Varianz, *welche folgendermaßen berechnet wird.*

$$\begin{aligned}
\mathbf{Var}(X) = \sigma_X^2 &= \mathbb{E}\,(X - \mu_X)^2 \\
&= \int_{-\infty}^{\infty} (x - \mu_X)^2\, f_X(x)\, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^2\, f_X(x)\, \mathrm{d}x - 2\,\mu_X \int_{-\infty}^{\infty} x\, f_X(x)\mathrm{d}x + \mu_X^2 \int_{-\infty}^{\infty} f_X(x)\, \mathrm{d}x \\
&= \mu_{X^2} - 2\,\mu_X^2 + \mu_X = \mu_{X^2} - \mu_X^2
\end{aligned}$$

Ist $\mu_X = 0$, so gilt $\sigma_X = \sqrt{\mu_{X^2}}$

v Stochastische Unabhängigkeit und Faltung

Definition 1.5. *Zwei Zufallsvariable X und Y auf einem Wahrscheinlichkeitsraum $(\Omega; \mathcal{F}; \mathbb{P})$ heißen stochastisch unabhängig , wenn für beliebige Mengen A und $B \in \mathcal{F}$ gilt*

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A \text{ und } X(\omega) \in B\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \cdot \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}).$$

Beispiel 1.8. *Die Spielergebnisse beim Roulette werden als stochastisch unabhängige Zufallsvariable modelliert, da die Eintrittswahrscheinlichkeit des nächsten Ergebnisses gänzlich unabhängig von der gesamten Vorgeschichte ist. Entsprechend gilt: Die Wahrscheinlichkeit dafür, daß beim Roulette nach zehnmal* rot *nochmals* rot *kommt, ist* $18/37$.

In der Versicherungspraxis kommt die Faltung von Verteilungen in folgendem Zusammenhang vor: Für ein $i \in \mathbb{N}$ sei X_i die Schadensumme eines Geschäftsjahres aus dem i-ten Vertrag, modelliert als Zufallsvariable mit Verteilung $F_i = F_{X_i}$. Die n Verträge des Bestandes werden als stochastisch unabhängig angenommen. Gesucht ist die Verteilung von

$$S := X_1 + \ldots + X_n.$$

Deren Momente sind leicht zu berechnen:

$$\mathbb{E}S = \sum_{i=1}^{n} \mathbb{E}X_i.$$

[2]Das zweite, aber auch das erste Moment muss nicht immer existieren. So ist zum Beispiel das zweite Moment einer Cauchy-verteilten Zufallsvariablen unendlich.

$$\mathbf{Var}(S) = \sum_{i=1}^{n} \mathbf{Var}(X_i).$$

Die Gesamtschadensverteilung selbst, also die Verteilung von S; ist jedoch nicht so leicht zu berechnen. Hierzu benötigt man die Faltungen.

Definition 1.6. *Sind P und Q Verteilungen, so definiert man die Faltung $P * Q$ von P und Q als die Verteilung von $X + Y$, wobei X und Y stochastisch unabhängig sind mit $X \sim P$; $Y \sim Q$.*

Beispiel 1.9. *Sei $a \in \mathbb{R}$. Seien X und Y jeweils die Augensummen bei einmal Würfeln. Dann ist*

$$\begin{aligned}(P * Q)(\{a\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) + Y(\omega) = a\}) &= \sum_{i=1}^{6} \mathbb{P}(\{\omega \in \Omega : X(\omega) = i\})\mathbb{P}(\{\omega \in \Omega : Y(\omega) = a - i\}) \\ &= \sum_{i=1}^{6} P(\{i\}) \cdot Q(\{a - i\}).\end{aligned}$$

Dies liefert die allgemeine Faltungsformel für den diskreten Fall:

$$P \star Q(i) = \sum_{i \in \mathbb{Z}} P(\{i\}) \cdot Q(\{a - i\})$$

Im stetigen Fall (f Dichte von X, g Dichte von Y) hat $P \star Q$ die Dichte

$$h(x) = \int_{-\infty}^{\infty} f(x - y)g(y)\, dy.$$

Für die Faltung gilt das Assoziativgesetz, so daß wir definieren können:

$$P^{\star n} = P_1 \star P_2 \star \cdots \star P_n := P_1 \star (P_2 \star \cdots \star P_n) = (P_1 \star \cdots \star P_{n-1}) \star P_n.$$

Falls alle $P_i = P_j$, so führen wir eine weitere, abkürzende Schreibweise ein:

Man kann sich $P \star 0$ als die Verteilung der Konstanten *Null* vorstellen.

Beispiel 1.10. *Seien P_1 und P_2 zwei Zufallsvariablen mit Zähldichte*

$$P_i(\{1\}) = 1, P_i(\{0\}) = p; i = 1; 2 :$$

Dann hat $P_1 \star P_2$ Masse auf den Zahlen $0; 1; 2$ und die Zähldichten

$$\begin{aligned}P_1 \star P_2(\{0\}) &= \sum_b P_1\{0 - b\}P_2(\{b\}) = P_1\{0\}P_2\{0\} = (1 - p)^2. \\ P_1 \star P_2(\{1\}) &= \sum_b P_1\{1 - b\}P_2(\{b\}) = P_1(\{1\})P_2(\{0\}) + P_1(\{0\})P_2(\{1\}) = 2p(1 - p) \\ P_1 \star P_2(\{2\}) &= \sum_b P_1(\{2 - b\})P_2(\{b\}) = P_1(\{1\})P_2(\{1\}) = p^2.\end{aligned}$$

Allgemeiner gilt

$$P_1^{\star n}(\{k\}) = \binom{n}{k} \cdot p^k(1 - p)^{n-k}.$$

und man erkennt, daß $P_1^{\star n}$ die Binomialverteilung $b(n; p)$ ist.

Beispiel 1.11. *Seien $P_1 = Exp(a)$ und $P_2 = Exp(b)$ zwei Verteilungen mit $a \neq b$. P_1 hat daher die Dichte $a \exp(-ax)$ für $x > 0$, P_2 hat daher die Dichte $b \exp(-bx)$ für $x > 0$. Dann hat $P_1 \star P_2$ die Dichte*

$$h(x) = \int_0^{\infty} a \exp(-ay) b \exp(-b(x - y))\, dy = \frac{ab}{a - b} (\exp(-ax) - \exp(-bx)).$$

für $x \geq 0$. Dabei ist zu beachten, daß $y > 0$ und $x - y > 0$ gelten müssen.

vi Faltung und die Fouriertransformierte

Genauso wie man die Fouriertransformierte eines SIgnales oder einer Funktion berechnen kann, kann man die Fouriertransformierte einer Zufallsvariablen berechnen.

Definition 1.7.

- *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable über $(\Omega, \mathcal{F}, \mathbb{P})$. Dann ist*
$$\hat{X}(u) := \mathbb{E}\, e^{iXu}, \quad u \in \mathbb{R};$$
die Fouriertransformierte zu X.

- *Sei F eine Verteilungsfunktion und f deren Dichte. Dann lautet die Fouriertransformierte von F*
$$\hat{F}(u) := \int_{-\infty}^{\infty} e^{iux}\, dF(x) = \int_{-\infty}^{\infty} e^{iux}\, f(x)\, dx, \quad u \in \mathbb{R}.$$

Es gilt folgende wichtige Beziehung:

Satz 1. *Seine X und Y zwei Zufallsvariablen über $(\Omega, \mathcal{F}, \mathbb{P})$, X und Y sind unabhängig. Dann gilt*

$$\widehat{(X+Y)}(u) = \hat{X}(u) \cdot \hat{Y}(u), \quad u \in \mathbb{R}.$$

Beweis: Angenommen X und Y sind diskret. Dass heißt, es gibt Mengen $\{A_l^X : l = 1, \dots N_X\}$, $A_l^X \cap A_j^X = \emptyset$, $l \neq j$, und $\{A_l^Y : l = 1, \dots N_Y\}$, $A_l^Y \cap A_j^Y = \emptyset$, $l \neq j$, und Punkte $\{x_l^X : l = 1, \dots N_X\}$ und $\{y_l^Y : l = 1, \dots N_Y\}$ sodass gilt

$$X(\omega) = \sum_{l=1}^{N_X} x_l 1_{A_l^X}(\omega) \quad \text{und} \quad Y(\omega) = \sum_{l=1}^{N_Y} x_l 1_{A_l^Y}(\omega), \quad \omega \in \Omega.$$

Damit gilt

$$\begin{aligned}
\widehat{(X+Y)}(u) &= \mathbb{E}e^{i(X+Y)u} = \mathbb{E}e^{iXu}e^{iYu} \\
&= \sum_{l=1}^{N_X}\sum_{j=1}^{N_X} e^{ix_l u} e^{iy_j u} \mathbb{P}(A_l^X)\mathbb{P}(A_l^Y) \\
&= \sum_{l=1}^{N_X} e^{ix_l u}\mathbb{P}(A_l^X) \sum_{j=1}^{N_X} e^{iy_j u}\mathbb{P}(A_l^Y) = \hat{X}(u) \cdot \hat{Y}(u).
\end{aligned}$$

Da man beliebige Zufallsvariablen mit diskreten approximieren kann, folgt dies für alle Zufallsvariablen.

☺

Kapitel 2

Bayes'sche Statistik und Bayes'sche Filter

1 Bedingte Wahrscheinlichkeit und die Formel von Bayes

Häufig betrachtet man Ereignisse und Zufallsgrößen unter Bedingungen, die die Zufälligkeit einschränken oder zusätzliche Informationen berücksichtigen.

Beispiel 1.1. *1. Würfeln, wobei allerdings nur die gerade Augenzahl gezählt werden; Bedingung 'Augenzahl gerade' .*

2. Lebensdauerversuch unter der Bedingung, dass nur solche Bauteile berücksichtigt werden, die den ersten Einsatztag überstanden haben.

Nach diesem Beispiel verstehen wir die Formel für die bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B:

$$\mathbb{P}(A \mid B) \quad = \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{2.1}$$

Natürlich fordern wie $\mathbb{P}(B) > 0$. Tatsächlich arbeitet man auch mit bedingten Wahrscheinlichkeiten im Fall $\mathbb{P}(B) = 0$, aber dann lautet die Definition anders.

Analog gilt

$$\mathbb{P}(B \mid A) \quad = \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

für $\mathbb{P}(A) > 0$.

Beispiel 1.2. *1. Wie groß ist die Wahrscheinlichkeit dass man bei zweimaligen würfeln eine* 2 *(oder eine* 1*) erzielt unter der Bedingung dass der Wurf ein gerades Ergebnis lieferte ?*

2. Karton A enthält 8 *Glühbirnen von denen* 3 *defekt sind, Karton B enthält* 4 *Glühbirnen von denen* 2 *defekt sind. Jedem Karton wird zufällig eine Glühbirne entnommen. Wie groß ist die Wahrscheinlichkeit dass*

(a) beide Birnen nicht defekt sind ?

(b) eine defekt und eine nicht defekt ist ?

(c) die defekte aus Karton A stammt, wenn eine defekt ist und eine nicht defekt ist?

3. *Es sei X die Lebensdauer eines Bauteils und $A = \{X > T\}$, $B = \{X > t\}$, wobei $T > t > 0$ gilt. Dann ist $\mathbb{P}(A \mid B)$ die bedingte Wahrscheinlichkeit für eine Lebensdauer länger als T Zeiteinheiten unter der Bedingung dass ein Lebensdauer von mindestens t Zeiteinheiten erreicht wurde. Die Formel* (2.1) *liefert*

$$\mathbb{P}(A \mid B) \quad = \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(X > t \text{ und } X > T)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > T)}{\mathbb{P}(X > t)},$$

da natürlich $\{X > t \text{ und } X > T\} \subset \{X > T\}$.

Diese Formel entspricht unseren Vorstellungen: Wenn wir n Bauteile beobachten würden, dann würden wir die bedingte Wahrscheinlichkeit $\mathbb{P}(A \mid B)$ durch folgenden Bruch abschätzen:

$$\begin{aligned} \mathbb{P}(A \mid B) \quad \sim \quad & \frac{h_n(A)}{h_n(B)}, \\ & h_n(A) = \text{ Anzahl aller Teile, die länger als } T \text{ Zeiteinheiten funktionieren }, \\ & h_n(B) = \text{ Anzahl aller Teile, die länger als } t \text{ Zeiteinheiten funktionieren }. \end{aligned}$$

Man überlegt sich dass $\mathbb{P}(A \mid B)$ und $\mathbb{P}(A \cap B)$ etwas unterschiedliches bedeuten. Im allgemeinen wird

$$\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$$

gelten, vgl. die vorangegangenen Beispiele. Es kann aber auch vorkommen dass

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

gilt. Dann ändert das stellen der Bedingung B nichts an der Häufigkeit des Auftretens von A. Man sagt dass A und B voneinander unabhängig sind. Nach (2.1) gilt dann die Produktformel der Wahrscheinlichkeitsrechnung

Seien A und B zwei unabhängige Ereignisse. Dann gilt

$$\mathbb{P}(A \cap B) \quad = \quad \mathbb{P}(B) \cdot \mathbb{P}(A). \tag{2.2}$$

Wenn A und B unabhängig sind, dann sind auch die Paare A und B^C, A^C und B, und A^C und B^C unabhängig. Drei Ereignisse A, B und C heißen unabhängig, wenn

$$\begin{aligned} \mathbb{P}(A \cap B) &= P(A)P(B), \\ \mathbb{P}(A \cap C) &= P(A)P(C), \\ \mathbb{P}(B \cap C) &= P(B)P(C), \end{aligned}$$

sowie

$$\mathbb{P}(A \cap B \cap C) = P(A)P(B)P(C)$$

gilt. Man fordert also mehr als die Gültigkeit der letzten Gleichung.

Analog definiert man bei n Ereignisse $\{A_1, A_2, \ldots, A_n\}$, dass für jede Teilmenge $I \subset \{1, 2, \ldots, n-1, n\}$ gilt

$$\mathbb{P}(\cap_{i \in I}) = \prod_{i \in I} \mathbb{P}(A_i).$$

Der Nachweis der Unabhängigkeit von Ereignissen kann statistisch geführt werden;

i Der bedingte Erwartungswert

Definition 1.1. *Sei $\Omega = (\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $Y : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable. Das heißt es gibt eine Familie disjunkter Teilmengen $\{A_1^Y, \ldots, A_n^Y\}$ mit $\cup_{i=1}^n A_n^Y = \Omega$ und $\{y_1, \ldots, y_n\} \subset \mathbb{R}$ mit*

$$Y(\omega) = \sum_{i=1}^{n} 1_{A_1^Y}(\omega)\, y_i$$

Dann ist $\mathbb{E}[X \mid Y] : \Omega \to \mathbb{R}$ eine Y-messbare Zufallsvariable definiert durch

$$\mathbb{E}[X \mid Y](\omega) = \begin{cases} \mathbb{E}\left[1_{A_1^Y}(\omega)\, X(\omega)\right] / \mathbb{P}\left(A_1^Y\right), & \text{falls } \omega \in A_1^Y, \\ \mathbb{E}\left[1_{A_2^Y}(\omega)\, X(\omega)\right] / \mathbb{P}\left(A_2^Y\right), & \text{falls } \omega \in A_2^Y, \\ \vdots \quad \vdots & \vdots \\ \mathbb{E}\left[1_{A_n^Y}(\omega)\, X(\omega)\right] / \mathbb{P}\left(A_n^Y\right), & \text{falls } \omega \in A_n^Y. \end{cases}$$

Beispiel 1.3. *Nehmen wir einen Würfel mit $\Omega = (\Omega, \mathcal{F}, \mathbb{P})$ definiert in Beispiel 1.2. Sei $X : \Omega \to \mathbb{R}$ die zugehörige Zufallsvariable, i.e. $X(\omega) = \omega$, $\omega \in \Omega$. Angenommen ein Spieler würfelt und sagt Ihnen nur ob er eine gerade oder ungerade Zahl hat. Also*

$$\begin{array}{rcl} Y : \Omega & \to & \mathbb{R}, \\ \omega & \mapsto & \begin{cases} 1, & \omega = 1, 3, 5 \\ 0, & \omega = 2, 4, 6. \end{cases} \end{array}$$

Für eine Ereignis $A \in \mathcal{F}$ kann man jetzt die bedingte Wahrscheinlichkeit $\mathbb{P}(A \mid Y)$ und den bedingten Erwartungswert $\mathbb{E}[1_A \mid Y]$ betrachten, bzw. den bedingten Erwartungswert $\mathbb{E}[X \mid Y]$ betrachten.

ii Die Formel der totalen Wahrscheinlichkeit

Eine einfache, aber wichtige Formel ist *die Formel der totalen Wahrscheinlichkeit*

Seien B_i einander ausschließende (disjunkte) Ereignisse mit

$$B_1 \cup B_2 \cup \cdots = \Omega.$$

Dann gilt

$$\mathbb{P}(A) = \sum_i P(A \mid B_i)\, \mathbb{P}(B_i). \tag{2.3}$$

Mittels der Formel kann am Beispiele folgenden Typs berechnen:

Beispiel 1.4. *Drei Maschinen produzieren das gleiche Produkt, Maschine 1 20 %, Maschine 2 30 % und Maschine 3 50 % der Gesamtproduktion. Die Ausschussproduktion sind* 1, 2 *und* 1.5 *Prozent . Wie gross ist die Ausschussquote p der Gesamtproduktion ?*

Beispiel 1.5. *In diesem Beispiel betrachten wir eine alte Tankers der eine große Anzahl von Oberflächenrissen aufweist. Wir modellieren die Gesamtzahl der Risse durch ein Poisson verteilt Zufallsvariable N. Es gilt also*

$$\mathbb{P}(N = k) = \frac{m^k}{k!}\, e^{-m}, \quad m \in \mathbb{N}.$$

Der Parameter m ist die durchschnittliche Anzahl der Risse pro Quadratmeter, der zu schätzen ist. Aufgrund des Alters des Tankers, weiß man aus Erfahrungswerten, dass die Anzahl von Rissen pro Quadratmeter ca. $\lambda = 0,01$ beträgt. Die gesamte Oberfläche des Tankers beträgt $5000m^2$, also im Durchschnitt 50 Rißе pro Quadratmeter. Die Ortung der Rißе funktioniert vool automatisch. Durch Erfahrungswerte weiß man dass mit einer Wahrscheinlichkeit $p = 0.999$ ein Riß entdeckt wird. Wie groß ist die Wahrscheinlichkeit, dass Rißе übersehen wurden ?

Lösung: Definieren wir das Ereignis $B = \{$Alle Risse wurden repariert$\}$ und das Ereignissystem $A_i = \{$Es wurden i Risse auf der Oberfläche erkannt$\}$, d.h. $A_i = \{N = i\}$. Dann gilt

$$\mathbb{P}(B) = \sum_{i=0}^{\infty} \mathbb{P}(N = i)\, \mathbb{P}(B|N = i) = e^{-0,05} \sim 0,95.$$

Da die Oberfläche des Tankers groß ist, ist somit die Wahrscheinlichkeit dass einige Risse nicht erkannt wurden und nicht repariert wurden, nicht vernachlässigbar.

☺

Beispiel 1.6. *Vor Ihnen sehen drei Kisten. Sie wissen dass in einer Kiste nur Orangen sind, in einer Kiste nur Äpfel und in einer Kiste zur Hälfte Äpfel und zur anderen Hälfte Orangen sind. Vor Ihnen steht eine Kiste. Jemand macht diese auf, und ohne dass sie hereinschauen können reicht er Ihnen einen Apfel. Wie gross ist die Wahrscheinlichkeit, dass in diese Kiste nur Äpfel, Orangen oder Äpfel mit Orangen gemischt sind.*

Beispiel 1.7. *Sie nehmen an einer Spieleshow im Fernsehen teil, bei der sie eine von drei verschlossenen Türen auswählen sollen. Hinter einer Tür wartet der Preis, ein Auto, hinter den beiden anderer stehen Ziegen. Sie zeigen auf eine Tür, sagen wir Nummer eins. Sie bleibt vorerst geschlossen. Der Moderator weiß hinter welcher Tür sich das Auto befindet; mit den Worten 'ich zeige Ihnen mal was' öffnet er eine andere Tür zum Beispiel Nummer drei, und eine meckernde Ziege kommt heraus. Er fragt Sie: Bleiben Sie bei der Tür Nummer eins oder wechseln Sie zur Tür Nummer zwei ? Erhöhen Sie Ihren Gewinnchancen wenn Sie wechseln ?*

Beispiel 1.8. *Verschiedenen Bauteile werden von n verschiedenen Lieferanten $A_1, A_2, \ldots, A_n$ geliefert. Der Marktanteil der verschiedenen Lieferanten beträgt $p_1, p_2, \ldots p_n$ mit $\sum_{i=1}^{n} p_i = 1$. Die Wahrscheinlichkeit das die Festigkeit des Bauteils von Lieferanten A_i stärker ist als eine vorgegebene Schranke Y beträgt y_i. Wie gross ist die Wahrscheinlichkeit, dass ein gekauftes Bauteil eine Festigkeit größer als Y hat.*

2 Grundprinzipien der Bayes-Statistik

Im Mittelpunkt der Bayes Statistik steht das Bayes Theorem. Mit seiner Hilfe lassen sich unbekannte Parameter schätzen, Konfidenzintervalle für die unbekannten Parameter festlegen und die Prüfung von Hypothesen für die Parameter ableiten. In diesem Kapitel wird das Grundprinzip der Bayes-Statistik erläutert.

Für die Bayes Statistik wird eine Erweiterung des Begriffes der Wahrscheinlichkeit vorgenommen, indem die Wahrscheinlichkeit für Aussagen definiert wird. Im Gegensatz zur traditionellen Statistik bei der die unbekannten Parameter feste Größen repräsentieren, sind die Parameter in der Bayes Statistik Zufallsvariable. Das bedeutet aber nicht, dass die unbekannten Parameter nicht Konstanten repräsentieren dürfen, wie beispielsweise die Koordinaten eines festen Punktes. Mit dem Bayes Theorem erhalten die unbekannten Parameter Wahrscheinlichkeit-Verteilungen, aus denen die Wahrscheinlichkeit, also die Plausibilität von Werten der Parameter folgt. Die Parameter selbst können daher feste Größen darstellen und brauchen nicht aus Zufallsexperimenten zu resultieren.

Die Idee ist, dass man mit einer *a'priori* angenommenen Wahrscheinlichkeitsverteilung für den Parameter startet, diese dann mit den erworbenen Wissen (bzw. zusätzlicher gewonnener Information)

updatet. Das heißt mit der zusätzlichen Information kann man dann eine Wahrscheinlichkeitsverteilung ableiten (der *a'posteriori* Verteilung), aus der man dann den Schätzwert berechnet. Dazu aber als erstes zwei Beispiele.

Beispiel 2.1. *Eine Münze wird geworfen. Ist diese Münze fair, so gilt* $\mathbb{P}(Kopf) = \mathbb{P}(Zahl) = \frac{1}{2}$. *Sei* X *die Anzahl der geworfenen Köpfe, wenn man* n*-mal wirft. Es gilt ja*

$$\mathbb{P}(X = k) = b(k; n, \frac{1}{2}), \quad k \leq n.$$

Jetzt wird zehn mal geworfen, dabei kommt nur zwei mal Kopf heraus. Genauer, der fünfte und achte Wurf ist Kopf. Ist die Münze fair, oder gilt

$$\mathbb{P}(Kopf) = p,$$

wobei p *unbekannt ist.*

Beispiel 2.2. *Eine neue Maschine ist geliefert worden. Zwischenzeitlich kommt es immer wieder zu Ausfällen, die aber immer wieder schnell behoben werden können. Allerdings muss immer jemand anwesend sein. Sei* T *die Wartezeit zwischen zwei Ausfällen. Es hat sich gezeigt, dass Wartezeiten als exponential verteilt mit Parameter* λ *angenommen werden können, aber es gilt noch den Parameter* λ *zu schätzen. Da die Maschine schon einen Tag läuft, haben wir schon einige Vorinformation von einer Stichprobe (Sample)* T_1, T_2, *...,* T_n. *Genau genommen ist die Maschine vier mal ausgefallen, die Zwischenzeiten waren dabei eine halbe Stunde, 20 Minuten, 1 Stunde und 2 Stunden.*

In beiden Fällen haben wir das Geschehen schon beobachtet und haben eine Vorinformation. Wie kann man jetzt die Parameter schätzen und dabei die Vorinformation ausnützen.

Sei ϑ der Parameter und sei D die Menge aller möglichen Parameter. Sei $F_X(x;\vartheta)$ die Verteilungsfunktion von X die vom Parameter ϑ abhängt. Das heißt $\mathbb{P}(X \leq x \mid \vartheta) = F_X(x;\vartheta)$ und $\mathbb{P}(X = x \mid \vartheta) = f_X(x;\vartheta)$. Wir beobachten nun die Zufallsvariable $X_1, X_2, \ldots, X_n$, wobei wir annehmen dass diese unabhängig sind.

Die Idee ist, dass wir annehmen, dass es eine Zufallsvariable Θ gibt, die genau unserer Parameter ϑ ist. Dabei kann sein, dass wir keine Information über Θ besitzen. Egal, wir nehmen an dass Θ wie f^{priori} verteilt ist, wobei f^{prior} frei aber sinnvoll gewählt wird.

Was wir jetzt berechnen wollen ist $\mathbb{P}(\Theta = \vartheta \mid X_1, \ldots, X_n)$. Wir wissen ja dass aufgrund der angenommenen Unabhängigkeit von $X_1, X_2, \ldots, X_n$, gilt

$$\mathbb{P}(X_1, \ldots, X_n \mid \Theta = \vartheta) = f_X(X_1;\vartheta) \cdot f_X(X_2;\vartheta) \cdots f_X(X_n;\vartheta).$$

Damit gilt nach der Formel von Bayes

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)},$$

und somit für $A = \{\Theta = \vartheta\}$ und $B = \{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$

$$\begin{aligned}\mathbb{P}(\Theta = \vartheta \mid X_1 = x_1, \ldots, X_n = x_n) &= \frac{\mathbb{P}(X_1=x_1,\ldots,X_n=x_n|\Theta=\vartheta) f^{priori}(\vartheta)}{P(X_1=x_1,\ldots,X_n=x_n)} \\ &= \frac{f_X(x_1;\vartheta)\cdot f_X(x_2;\vartheta)\cdots f_X(x_n;\vartheta) f^{priori}(\vartheta)}{P(X_1=x_1,\ldots,X_n=x_n)}\end{aligned}$$

wobei gilt

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int_D f_X(x_1;\vartheta) \cdot f_X(x_2;\vartheta) \cdot \cdots \cdot f_X(x_n;\vartheta)\, f^{priori}(\theta)\, d\theta.$$

Die Funktion

$$D \ni \theta \mapsto \mathbb{P}(\Theta = \vartheta \mid X_1 = x_1, \ldots, X_n = x_n) \in [0, \infty)$$

heißt Liklihood Funktion. Man nimmt den Wert als Schätzer, wo $\mathbb{P}(\Theta = \vartheta \mid X_1 = x_1, \ldots, X_n = x_n)$ das Maximum besitzt, nämlich

$$\frac{d}{d\vartheta}\,\mathbb{P}(\Theta = \vartheta \mid X_1 = x_1, \ldots, X_n = x_n) = 0.$$

Beispiel 2.3. *Gehen wir zurück zum Beispiel 2.1. Hier nehmen wir als $f^{priori}(p) = 1$, d.h. wir nehmen an, dass der Parameter p gleichverteilt in $[0,1]$ ist. Hier hatten wir nur zwei mal Kopf bei zehn Würfen. Das heißt es gilt*

$$\mathbb{P}(P = p) = \frac{b(2;10,p)}{\int_0^1 b(2;10,p)\,dp} = \frac{\binom{10}{2} p^2(1-p)^8}{\frac{5\cdot 9}{495}}.$$

Da gilt

$$\frac{d}{dp}\mathbb{P}(P = p) = \frac{d}{dp}\left(\frac{495}{45}\binom{10}{2} p^2(1-p)^8\right) = 2(p-1)^7 p(5p-1),$$

ist $p = \frac{1}{5}$ unserer Schätzwert.

Bemerkung 2.1. *Es gibt mehrere Möglichkeiten den Schätzer auszuwählen. Eine anderer Schätzer wäre der Erwartungswert von ϑ, nämlich*

$$\hat{\vartheta} := \int \vartheta\,\mathbb{P}(\Theta = \vartheta \mid X_1 = x_1, \ldots, X_n = x_n)\,d\vartheta.$$

In diesem Fall nimmt man den Wert als Schätzer, der den quadratischen Fehler minimiert. Also genau den Wert $\hat{\vartheta}$, der den quadratischen Fehler

$$\int_0^1 \left(\vartheta - \hat{\vartheta}(X_1 = x_1, \ldots, X_n = x_n)\right)^2 \mathbb{P}(\Theta = \vartheta \mid X_1 = x_1, \ldots, X_n = x_n)\,d\vartheta,$$

minimiert. Berechnet man den Erwartungswert, erhält man $\frac{1}{4}$.

Beispiel 2.4. *Gehen wir zurück zum Beispiel 2.2. Dabei gilt $X_1 = \frac{1}{2}$, $X_2 = \frac{1}{3}$, $X_3 = 1$, und $X_4 = 2$. Als a'priori Abschätzung nehmen wir an Θ sei exponential verteilt mit Parameter 1, also*

$$f^{priori}(\vartheta) = \exp(-\vartheta).$$

Nun gilt

$$\mathbb{P}(X_1 = \frac{1}{2}, \ldots, X_4 = 2 \mid \Theta = \vartheta) = \vartheta^4 \exp(-(\frac{1}{2} + \frac{1}{3} + 1 + 2)\vartheta),$$

und somit

$$\mathbb{P}(\Theta = \vartheta \mid X_1 = \frac{1}{2}, \ldots, X_4 = 2) = \frac{\vartheta^4 \exp(-(\frac{1}{2} + \frac{1}{3} + 1 + 2)\vartheta)\exp(-\vartheta)}{\int_0^\infty y^4 \exp(-(\frac{1}{2} + \frac{1}{3} + 1 + 2)y)\exp(-y)dy}.$$

Ausrechnen ergibt

$$\int_0^\infty y^4 \exp(-(\frac{1}{2} + \frac{1}{3} + 1 + 2)y)\exp(-y)dy = \frac{24}{(\frac{1}{2} + \frac{1}{3} + 1 + 2 - 1)^4} = \frac{186624}{1419857} = 0.131439.$$

Berechnet man das Maximum erhält man $\vartheta = 4/(\frac{1}{2} + \frac{1}{3} + 1 + 2 - 1) = \frac{24}{17} = 1.41176$. Nimmt man den Erwartungswert erhält man $17/30 = 0.56666$.

Auch stellt sich hier die Frage, welchen Einfluss hat die a'priori Verteilung auf den Schätzwert. Wichtig ist, dass $f^{priori}(\vartheta) > 0$ ist, sobald $\vartheta > 0$ möglich ist. Weiß man z.B. dass ϑ sicher nicht größer

als 5 *ist, kann man* ϑ *gleichverteilt auf dem Intervall* $[0,5]$ *annehmen. Die gleiche Rechnung, nur mit* $f^{priori}(\vartheta) = \frac{1}{5}\,1_{[0,5]}(\vartheta)$ *führt zu folgender Rechnung:*

$$\mathbb{P}(\Theta = \vartheta \mid X_1 = \tfrac{1}{2}, \ldots, X_4 = 2) = \frac{\vartheta^4 \exp(-(\frac{1}{2}+\frac{1}{3}+1+2)\vartheta)\exp(-\vartheta)}{\frac{1}{5}\int_0^5 y^4 \exp(-(\frac{1}{2}+\frac{1}{3}+1+2)y)dy}$$

$$= 172.447\,\vartheta^4 \exp(-(\frac{1}{2}+\frac{1}{3}+1+2)\vartheta)\exp(-\vartheta).$$

Das Maximum liegt nun bei $\hat{\vartheta} = 1.04348$. *Nimmt man den Erwartungswert so kommt man auf* $\hat{\vartheta} = 1.30421$. *Im allgemeinen kann man sagen, dass das die Wahl des Maximums weniger sensibel bzgl der Anfangsverteilung ist, als die Wahl des Erwartungswertes.*

Beispiel 2.5. *Ein Draht wird benützt um schwere Gegenstände zu transportieren. Das Gewicht dieser Gegenstände schwankt von Tag zu Tag. Erfahrungsgemäß genügt das maximale Gewicht einer Gumbel Verteilung, d.h.*

$$P(X \le x) = \exp(-e^{-(x-b)/a}),$$

wobei X *das maximale Gewicht an einen Tag ist, und* $a = 156$ *und* $b = 910$ *ist.*

Kauft man so einen Draht, so ist die Qualität unterschiedlich. Der Verkäufer gibt nur die durchschnittliche Stärke und den Variationskoeffizient an. Aufgrund langjähriger Erfahrung weiß man dass die Stärke (oder maximale Traglast, benannt Y*) einer Weilbullverteilung genügt, d.h.*

$$P(Y = y) = \frac{c}{\alpha}\left(\frac{y}{\alpha}\right)^{c-1} e^{-\left(\frac{y}{\alpha}\right)^c}, \quad y \ge 0.$$

Die durchschnittliche maximale Traglast ist 1000 kg und der Variationskoeffizient ist 0.2. Damit gilt $c = 5.79$ *und* $\alpha = 1000/0.9259 \sim 1080$.

Der Seil hat jetzt ein Jahr lang durchgehalten. Soll man diesen jetzt im nächsten Jahr behalten oder einen besserer Qualität ($\mathbb{E}Y = 1200kg$ *und Variationskoeffizient 0.2) kaufen ?*

Lösung: Wir wissen dass die Verteilung der Tragkraft des Seils Y einer Weilbullverteilung mit Parametern $c = 5.79$ und $\alpha = 1000/0.9259 \sim 1080$ genügt. Damit gilt

$$f^{priori}(y) = \frac{c}{\alpha}\left(\frac{y}{\alpha}\right)^{c-1} e^{-\left(\frac{y}{\alpha}\right)^c}, \quad y \ge 0.$$

Sei nun B das folgende Ereignis:

$$B = \{\,\text{Das Seil ist nicht beschädigt im ersten Jahr}\,\}.$$

Da das Seil zerreißt sobald eine Traglast schwerer als Y ist, gilt $\mathbb{P}(B) = \mathbb{P}(X < Y)$. Angenommen die Tragkraft ist y. Dann gilt

$$\mathbb{P}(B \mid y) = P(X < y) = e^{-e^{-(y-b)/a}}.$$

(Die Tragkraft ist unabhängig von der Traglast). Nun gilt

$$\mathbb{P}(B) = \int_0^\infty \exp\left(-e^{-(y-90)/156}\right)\frac{5.79}{1080}\left(\frac{y}{1080}\right)^{4.79} e^{-(y/1080)^{5.79}}\,dy \sim 0.533.$$

Jetzt wollen wir aber die Information miteinbeziehen, dass das Seil das erste Jahr nicht gerißen ist. Also, $\mathbb{P}(y \mid B)$ berechnen. Es gilt ja

$$\begin{aligned} f^{posterior}(y) &= \frac{P(B \mid Y = y) f^{prior}(y)}{P(B)} \\ &= \frac{1}{0.533}\exp\left(-e^{-(y-910)/156}\right)\frac{5.79}{1080}\left(\frac{y}{1080}\right)^{4.79} e^{-(y/1080)^{5.79}}. \end{aligned}$$

Die Frage die wir uns jetzt stellen ist, wie gross ist die Wahrscheinlichkeit, dass dieses Seil das nächste Jahr durchhält. Sei

$$C = \{\text{ Das Seil hält das zweite Jahr unbeschädigt durch }\}.$$

Nun gilt, $P(C) = \int_0^\infty P(X < y)P(Y = y)\,dy$, d.h.

$$P(C) = \int_0^\infty P(C \mid y) f^{post}(y)\,dy = 0.705.$$

Angenommen, wir kaufen ein neues Seil, dann gilt

$$P(C) = \int_0^\infty P(C \mid y) f^{priori}(y)\,dy = 0.757.$$

Man muss nun den Preis eines neues Seil und die Wahrscheinlichkeit dass es nächstes Jahr nicht so schnell kaputt geht gegeneinander abwägen.

3 Bayes'sche Inferenz oder Rekursives Filtern

Oft bekommt man mit der Zeit immer mehr und mehr information über den Parameter. Dieses *updaten* des Parameters nennt man *Bayes'sche Inferenz*.

In diesem Kapitel wird Ihnen eine Strategie erklärt wie Sie einen Parameter Schätzer updaten können obwohl $F_X(x;\vartheta)$ eine komplizierte Funktion ist.

Fragestellung Eine biologische Kläranlage wurde installiert. Dabei arbeiten die verschiedenen biologischen Substanzen mit verschiedener Effizienz, die täglich durch alle möglichen Randbedingungen wechseln kann. Jeden Tag wird gemessen ob die Kläranlage den Standard entspricht. Wenn das Wasser in der Kläranlage den Standard entspricht, kann das Wasser ausgelassen werden. Wir schreiben $p = \mathbb{P}(B)$ wobei B das Ereignis $\{$*Die Kläranlage erfüllt den Standard* $\}$ beschreibt. Die Kläranlage arbeitet um so effizienter um so höher p liegt. Mit Hilfe der Konstante p kann man entscheiden ob eine neue Bakteriologischen Kultur das Wasser gut reinigt oder nicht. Man hat aber nur immer den Ausgang des Ereignisses B einmal in der Woche messen, da dazu das Becken entleert werden muss.

Das Ziel ist an einer Serie von Ausgängen von B die Wahrscheinlichkeit p zu schätzen. Um die Berechnung einfach zu halten, approximieren wir die Dichtefunktion durch eine stückweise stetige Funktion. Ander gesagt, wir unterscheiden dabei folgende Möglichkeiten für p:

$$A_1 = \{p = 0.1\}, \quad A_2 = \{p = 0.3\}, \quad A_3 = \{p = 0.5\}, \quad A_4 = \{p = 0.7\}, \quad A_5 = \{p = 0.9\}.$$

Wie wissen wir nun welches Ereignis $A_1, \ldots, A_5$ eintrifft. Wir können dies nur indem wir p abschätzen. Allerdings kennen wir $\mathbb{P}(B \mid A_i)$, nämlich $\mathbb{P}(B \mid A_1) = 0.1$, $\mathbb{P}(B \mid A_2) = 0.3$, $\mathbb{P}(B \mid A_3) = 0.5$, $\mathbb{P}(B \mid A_4) = 0.7$ und $\mathbb{P}(B \mid A_5) = 0.9$.

Dabei sei q_i^0 die Wahrscheinlichkeit dass A_i wahr ist. Da wir am Anfang eigentlich nichts wissen setzten wir $q_1^0 = q_2^0 = \cdots = q_5^0 = \frac{1}{5}$. Angenommen die $n$te Messung ergibt dass B wahr ist. Dann setzten wir

$$q_i^n = P(B|A_i)q_i^{n-1}, \quad i = 1, \ldots, 5.$$

Zeigt die $n$te Messung dass B falsch ist, setzten wir

$$q_i^n = (1 - P(B|A_i))q_i^{n-1}, \quad i = 1, \ldots, 5.$$

Angenommen wir erhalten jetzt folgende Folge von Ereignissen B, B^c, B, B, B. Berechnen wir nun q_i^5, $i = 1, \ldots, 5$ so zeigt sich dass q_4^5 den höchsten Wert hat. Wir werden also annehmen dass A_4 wahr ist.

Formale Beschreibung Gegeben sind Ereignisse $A_1, \ldots, A_k$ wobei wir abschätzen möchten, welches der Ereignisse A_i zutrifft. Dabei gilt $A_i \cap A_j = \emptyset$ und $\cup_{i=1,\ldots,k} A_i = \Omega$. Seien $q_i^{(0)}$ die Anfangsschätzer (*A priori* Schätzer) von A_i. Haben wir keine Information, setzten wir wieder $q_i^{(0)} = \frac{1}{k}$, $i = 1, \ldots, k$.

Seien $B_1, B_2, \ldots, B_n$ eine Folge von Ereignissen, die in zeitlicher Folge eintreten, unabhängig sind und die von den Ereignissen $A_1, \ldots, A_k$ abhängen. Das heißt man kann $\mathbb{P}(B_j \mid A_i)$, $j = 1, \ldots, n$, $i = 1, \ldots, k$ berechnen. Sei $q_i^{(n)}$ der *a posteriori* Schätzer von A_i. Die *Liklelihood* Funktion $L(A_i)$, bzw. $\mathbb{P}($ alle $B_1, \ldots, B_n$ treffen ein $\mid A_i)$ ist bekannt. Z.B, falls $n = 2$ ist, gilt (wegen der Unabhängigkeit von B_1 und B_2

$$\mathbb{P}(B_1 \cap B_2 \mid A_i) = \mathbb{P}(B_1 \mid A_i) \cdot \mathbb{P}(B_2 \mid A_i).$$

Dann gilt folgende Rekursion

$$q_i^{(n)} = \mathbb{P}(B_n \mid A_i)\ q_i^{(n-1)}, \quad n \in \mathbb{N}.$$

Gegeben B_n, berechnet man $q_1^{(n)}, \ldots, q_k^{(n)}$ und nimmt dieses A_i als wahr an, wo $q_i^{(n)}$ maximal ist.

Beispiel 3.1. *Gehen wir zurück zum Beispiel 2.2. Wir haben in Beispiel 2.4 einen Schätzer von ϑ berechnet und eine f^{post} verteilung von ϑ berechnet. Angenommen wir arbeiten an der Tankstelle. Jetzt besteht die Möglichkeiten die Wartezeiten weiter zu beobachten und in unsere Schätzung miteinzubeziehen. Da aber mit zunehmenden Wartezeiten die Funktion immer komplexer wird, diskretisieren wir den Wertebereich von ϑ und unterscheiden nur die folgende Werte:*

$$\vartheta_1 = 0.5, \vartheta_2 = 0.9, \vartheta_3 = 1.2, \vartheta_3 = 1.4, \vartheta_5 = 1.45, \vartheta_6 = 1.5, \vartheta_7 = 1.55, \vartheta_8 = 1.6, \vartheta_9 = 1.9, \vartheta_{10} = 2.4, \vartheta_{11} = 2.9.$$

Analog nennen wir $A_i = \{\vartheta = \vartheta_i\}$.

Sei $q_i^0 := f^{post}(\vartheta_i)$. Jetzt kommen langsam die Kunden herein, die Wartezeiten bezeichnen wir mit X_5, X_6, X_7, ... Bei jeden neuen Kunden können wir jetzt den Parameter updaten, i.,e. q_i neu berechnen:

$$q_i^k = \mathbb{P}(X_{k+4} \mid A_i) = \vartheta_i\, exp(-X_{k+4}\, \vartheta_i), \quad i = 1, \ldots 11.$$

Als Schätzer nehmen wir immer das ϑ_i, wo q_i^k am größten ist. Wollen wir jetzt nach einen Kunden eine kurze Pause einlegen, können wir uns ausrechnen wie lange die Pause dauern darf, sodass mit einer Wahrscheinlichkeit von 0.95 kein Kunde kommt:

$$\int_0^T \hat{\vartheta} \exp(-s\hat{\vartheta})\, ds \leq 0.05.$$

Eine kurze Rechnung ergibt:

$$(1 - e^{-T\hat{\vartheta}}) \leq 0.05 \quad \Rightarrow \frac{-\ln(0.95)}{\hat{\vartheta}} \geq T.$$

4 Bayes'sche Inferenz: Beispiele und Klassen:

i Preiskalkulation in der Autohaftpflichtversicherung - das Bonus Malus System

Ende der 60 Jahre wurde das Bonus Malus System für die Auto-Haftpflichtversicherung in der Schweiz eingeführt, die Mathematische Grundlage ist in Bichsel nachzulesen. Angenommen, Sie schließen eine Vertrag mit einer Autohaftpflichtversicherung ab. Nach einigen Jahren Unfallfreies Fahren werden Ihnen Bonus Punkte vergeben und Ihre Prämie sinkt. Falls Sie aber immer wieder selbst verursachte Unfälle haben, bleibt die Prämie gleich oder, steigt.

Allgemein gesehen, versuchen die Autoversicherung die Risiken in zwei Kategorien einzuteilen:

- den *objektiven Risiken*: z.B. die PS zahl eines Autos, der Hubraum, das Gewicht, etc .. ;

- und den *subjektiven Risiken* (nicht objektiv messbare Risiken): Risikobereitschaft, das Können, das Fahrverhalten des Fahrers, Temperament, die genaue Kilometeranzahl, etc ... .

Man hat zwei verschiedene Typen von Risiken - die einen sind objektiv erfassbar - bei jeder Autoversicherung muss man den Typenschein angeben - angeben ob man minderjährige Kinder hat, etc.., die anderen sind nicht erfassbar und sehr subjektiv. So sind auf Grund von Befragungen auf Österreichs Straßen nur sehr gute Autofahrer unterwegs - trotzdem passieren Unfälle. Damit hat jeder Autofahrer a sein individuelles Risikoprofil, das durch den Parameter ϑ_a beschrieben wird. Dieser Parameter kann Werte aus D_Θ annehmen, wobei D_Θ die Menge aller möglichen Werte für ϑ_a darstellt.

Für einen bestimmten Autofahrer ist der genaue Wert von ϑ_a zumeist unbekannt. Aus Statistiken kann man aber Rückschlüsse auf die Verteilung von ϑ_a machen, so sind die meisten Autofahrer vorsichtig, allerdings gibt es einige Ausreißer, die immer wieder Unfälle verursachen. Deswegen nimmt man eine A'priori Verteilung des Parameters ϑ_a an.

Die Autohaftpflichtversicherung ordnet a'priori Ihre Risikobereitschaft im Fahrverhalten bei Abschluss eines Vertrages hoch ein. Fahren Sie jetzt einige Zeit Unfallfrei, werden die Daten miteinbezogen und die Autohaftpflichtversicherung ordnet Ihre Risikobereitschaft geringer ein. Die Wahrscheinlichkeit im nächsten Jahr einen Unfall zu verursachen ist (statistisch gesehen) geringer und Ihre Prämie sinkt.

Statistischer Hintergrund

In diesem Abschnitt wird der Statistische Hintergrund zur Preiskalkulation bei Haftpflichtversicherung beschrieben. Sei $X_j, j = 1, \ldots, n$ die Forderung eines Versicherten im Jahr j. Mathematisch gesehen ist dies einen Zufallsvariable. Angenommen, diese Zufallsvariable hat Verteilung F_ϑ, wobei ϑ ein Parameter ist, der vom Können und der Risikobereitschaft des Fahrers abhängt.

Wir befinden uns jetzt in Jahre n. Mit Hilfe der Forderung von den letzten Jahren, d.h. $\mathbf{X} = (X_1, \ldots, X_n)'$ soll nun die Prämie für das nächste Jahr berechnet werden, d.h. die Verteilung von X_{n+1} soll geschätzt werden. Da unter der Bedingung dass ϑ bekannt ist, der Typ der Verteilung F_ϑ bekannt ist, heißt dies dass der Parameter ϑ geschätzt werden soll. Um ϑ zu schätzen, können wir nur auf die Daten $\mathbf{X}$ zurückgreifen.

Die Vorstellung dahinter ist, dass wir vor uns zwei Urnen haben. Zuerst wird der Parameter Θ aus einer Urne gezogen. Das heißt das Können und die Risikobereitschaft des Fahrers wird mit Hilfe einer Zufallsvariablen simuliert. Der Wert von Θ bestimmt die zweite Urne. Im zweiten Schritt werden die Forderungen $X_j, j = 1, 2, \ldots$ jedes Jahr aus der zweiten Urne gezogen. Diese Forderungen haben Verteilung F_Θ. Die Versicherung kennt aber Θ nicht, sondern nur die Forderungen aus den Jahren $1, \ldots, n$, i.e. $\mathbf{X} = (X_1, \ldots, X_n)'$. D.h. im n–ten Jahr, hat die Versicherung $\mathbf{X} = (X_1, \ldots, X_n)'$ gegeben.

Modellannahmen: Um das Modell mathematisch formulieren zu können werden folgende Annahmen gemacht:

- Ist $\Theta = \vartheta$ gegeben, dann sind $X_1, X_2, \ldots$ unabhängig, identisch verteilt und $X_1 \sim F_\vartheta$. *identische verteilt ist gleich zu setzten mit stationärität - damit kann man auf historische Daten zurückgreifen - Inflation und andere Einflussfaktoren müssen dabei herausgerechnet werden.*

- Θ ist eine Zufallsvariable mit Verteilungsfunktion U und Dichtefunktion u und Verteilungsfunktion U. Diese Verteilungsfunktion heißt *structural function of the collective*.

Die Versicherung steht jetzt vor den folgenden Problem. Gegeben $\mathbf{X} = (X_1, \ldots, X_n)'$, zu schätzen ist die Prämie für das nächste Jahr. Die Prämie hängt vom Parameter ϑ ab, aber dieser Parameter hängt von $(X_1, \ldots, X_n)'$. Damit ist T also ist eine Funktion

$$T : \mathbb{R}^n \longrightarrow \mathbb{R},$$

wobei D die Menge der möglichen Schätzewerte bezeichnet. Mittels $\hat{\vartheta} = T(X_1, \ldots, X_n)$ wird jetzt die Prämie errechnet.

Bemerkung 4.1. *Die Zufallsvariable X_{n+1} beschreibt die Forderungen die der Autofahrer im Jahre $n+1$ von der Autoversicherung einreichen wird und ist Ende des Jahres n noch nicht bekannt. Damit ist X_{n+1} eine Zufallsvariable dessen Verteilung wir noch nicht kennen. Trotzdem müssen wir die Art und Weise wie man die Prämie berechnet jetzt schon festlegen. Damit ist die Prämie eine Abbildung vom Raum aller möglichen Zufallsvariablen $X : \Omega \to \mathbb{R}^+$ in die reellen Zahlen, die jeder Zufallsvariablen X eine bestimmte Zahl zuordnet. Hier sind einige mögliche Beispiele gegeben:*

- $X \mapsto (1+\alpha)\mathbb{E}X, \quad \alpha > 0$ *(expectation principle)*
- $X \mapsto \mathbb{E}X + \beta\mathbf{Std}(X), \quad \beta > 0$ *(Standard Deviation Principle)*
- $X \mapsto \mathbb{E}X + \gamma\mathbf{Var}(X), \quad \gamma > 0$ *(Variance Principle)*
- $X \mapsto \frac{1}{\delta}\ln(\mathbb{E}e^{\delta X}), \quad \delta > 0$ *(Exponential Principle)*

In unseren Fall nehmen wir der Einfachheit halber das expectation principle. *Im allgemeinen wird ein Zuschlag in Abhängigkeit der Varianz bzw. der Standardabweichung miteinbezogen. Dies ist um die Volatilität mit einzubeziehen (Stärke der Fluktuationen). Fällt jetzt ein Jahr schlechter aus als erwartet, so hat die Versicherung genügend finanziellen Spielraum um nicht pleite zu gehen.*

In unseren Beispiel nehmen wir als Korrekte individuelle Prämie den Erwartungswert.

Definition 4.1. *Die Korrekte individuelle Prämie ist gegeben durch*

$$\mathcal{P}^{ind} = \mathbb{E}[X_{n+1} \mid \vartheta] = \int_0^\infty x\, dF_\vartheta(x) =: \mu(\vartheta).$$

Unsere Aufgabe ist jetzt aufgrund der Daten $\mathbf{X} = (X_1, \ldots, X_n)'$, den Wert $\mathbb{E}[X_{n+1} \mid \vartheta]$, bzw. $\mathbb{E}\left[\mu(\Theta) \mid \Theta\right| = \mu(\vartheta)$ zu Schätzen. Hier sei angemerkt, dass sobald der Wert von Θ bekannt ist, $\mu(\Theta)$ exakt berechnet werden kann. Im Unterschied zur individuellen Prämie kann man sich die Kollektive Prämie untersuchen.

Definition 4.2. *Die Kollektive ist gegeben durch*

$$\mathcal{P}^{coll} = \int_{D_\Theta} \mu(\vartheta)\, dU(\vartheta) =: \mu_0.$$

Hier bezeichnet D_Θ die Menge aller zulässigen (und möglichen) Parameter Θ, i.e. $D_\Theta := \{\vartheta \mid \vartheta$ ist als Parameter möglich $\}$.

Sei $L(\vartheta, T(\mathbf{x}))$ der *Verlust*, die sich ergibt, falls ϑ der richtige Parameter ist und $T(\mathbf{x})$ der aus der Beobachtung $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ geschätzte Wert von $\mu(\vartheta)$ ist. Eine mögliche Verlustfunktion wäre z.B. die quadratische Funktion

$$L(\vartheta, T(\mathbf{x})) = \left(\mu(\vartheta) - T(\mathbf{x})\right)^2.$$

Abweichungen in beide Richtungen führen zu einen Verlust. Setzt man die Prämie zu hoch an, gehen die Kunden zu einen anderen Anbieter über, und die Versicherung erleidet auch einen Verlust. Setzt man die Prämie zu niedrig an, verliert man Geld da man weniger einnimmt als ausgibt.

Sei nun $R_T(\vartheta, \mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ die Risikofunktion gegeben durch

$$R_T(\vartheta, \mathbf{x}) := \mathbb{E}_\vartheta\left[L(\vartheta, T(\mathbf{x}))\right] =$$

Definition 4.3. Baye'sche Riskoschätzer *von T mit Anfangsverteilung $U(\vartheta)$:*

$$R_T(\mathbf{x}) := \int_D R_T(\vartheta, \mathbf{x})\, dU(\vartheta).$$

Ziel ist es nun, einen Schätzer für $mu(\theta)$ zu finden, der die Risikofunktion $R_T(\mathbf{x})$ minimiert.

Im weiteren führen wir folgende Notation ein. Sei $g : \mathbb{R} \to \mathbb{R}$ eine (konvexe) Funktion. Das mathematische Zeichen $\min_{x\in\mathbb{R}} g(x)$ ist jedem (hoffentlich) hinreichend bekannt. Uns aber interessiert nicht der Wert des Minimums sondern an welchen Punkt $x \in \mathbb{R}$ das Minimum angenommen wird. Dieser Wert wird mit arg min bezeichnet. Genauer: Sei $D \subset \mathbb{R}$ und gilt $\tilde{x} \in D$ und $g(\tilde{x}) \leq g(x)$ für alle $x \in D$, dann schreiben wir $\tilde{x} = \arg\min_{x\in D} g(x)$.

Definition 4.4. *Der* Baye'sche Schätzer *von T ist gegeben durch*

$$\tilde{T}(\mathbf{x}) := arg\ min_{T:\mathbb{R}^n\to\mathbb{R}} R_T(\mathbf{x}),$$

wobei das Minimum über alle möglichen Abbildungen $T : \mathbb{R}^n \to \mathbb{R}$ genommen wird.

Mit dieser Definition ist $\tilde{T}(\mathbf{x})$ genau unserer Bayescher Schätzer den wir schon in unseren vorhergehenden Kapitel kennengelernt haben. Die Frage ist, welche Art von Abbildung $T : \mathbb{R}^n \to \mathbb{R}$ den Ausdruck $R_T(\mathbf{x})$ minimiert. Um das Minimum zu berechnen, ist dir Größe $R_T(\mathbf{x})$ nach T zu differenzieren und zu untersuchen für welches T die Ableitung gleich Null wird. Die Transformation T muss also so beschaffen sein, dass folgenden Ausdruck auf Null abgebildet wird:

$$(T(\mathbf{x}) - \mu(\vartheta))\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n)\, u(\vartheta).$$

Eine kurze Rechnung ergibt:

$$T(\mathbf{x}) = \frac{\mu(\vartheta)\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n)\, dU(\vartheta)}{\int_{D_\Theta} dF_\vartheta(x_1) \cdots dF_\vartheta(x_n)\, u(\vartheta)}. \tag{2.4}$$

Das heißt $T(\mathbf{x})$ wo T wie oben definiert ist, ist genau der Wert der $R_T(\mathbf{x})$ minimiert, und deswegen der Bayes'sche Schätzer der Prämie, also

$$\mathcal{P}^{bayes} = \frac{\int_{D_\Theta} \mu(\vartheta)\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n)\, dU(\vartheta)}{\int_D dF_\vartheta(x_1) \cdots dF_\vartheta(x_n)\, dU(\vartheta)}.$$

Damit hängt die genaue Gestalt von T noch von den Verteilungsfunktionen F_ϑ und U ab.

ii Die Preiskalkulation Der Versicherung

Im nächsten Schritt möchten wir die genauen Preiskalkulation der Prämien aufzeigen. Sei nun N_j die Anzahl der Forderungen und X_j die Gesamtforderung eines Versicherungsnehmer in Euro im j–ten Jahr.

Modellannahmen:

1. Gegeben sei das Risikoprofil ϑ des Autofahrers. Dann gilt

   $$\mathbb{E}[X_j \mid \Theta = \vartheta] = C \cdot \mathbb{E}[N_j \mid \Theta = \vartheta]$$

   wobei C nur von der PS-Zahl des Autos abhängt und $\mathbb{E}[N_j \mid \Theta = \vartheta]$ vom Fahrer (bzw. dessen Risikobereitschaft) anhängt.

2. Gegeben ist $\Theta = \vartheta$, dann sind die $\{N_j : j = 1, \ldots, n, n+1\}$ unabhängig und Poisson verteilt mit Parameter ϑ, i.e.

   $$f_\vartheta(N_j) = \mathbb{P}(N_j = k \mid \Theta = \vartheta) = e^{-\vartheta}\, \frac{\vartheta^k}{k!}.$$

3. Die Zufallsvariable Θ ist Gamma verteilt mit Parametern γ und β. Genauer, die Dichtefunktion lautet

   $$u(\vartheta) = \begin{cases} \frac{\beta^\gamma}{\Gamma(\gamma)}\, \vartheta^{\gamma-1} e^{-\beta\,\vartheta}, & \vartheta > 0, \\ 0 & \text{sonst}\ . \end{cases}$$

Bemerkung 4.2. *Es gilt*

$$\mathbb{E}[\Theta] = \frac{\gamma}{\beta} \quad \textit{und} \quad \mathbf{Var}[\Theta] = \frac{\gamma}{\beta^2}.$$

Durch eine kurze Rechnung lässt sich folgendes zeigen

$$\begin{aligned}\mathbb{E}[N_{n+1} \mid \Theta = \vartheta] &= \sum_{k=0}^{\infty} \mathbb{P}(N_{n+1} = k \mid \Theta = \vartheta) \cdot k = \sum_{k=0}^{\infty} e^{-\vartheta} \frac{\vartheta^k}{k!} \cdot k \\ &= \vartheta e^{-\vartheta} \sum_{k=1}^{\infty} \frac{\vartheta^{k-1}}{(k-1)!} = \vartheta e^{-\vartheta} e^{\vartheta} = \vartheta.\end{aligned}$$

Damit können wir folgendes Beweisen:

Proposition 4.1. *Unter den oben genannten Annahmen gilt*

$$\begin{aligned}\mathcal{P}^{ind} &= \mathbb{E}\left[N_{n+1} \mid \Theta\right] = \Theta, \\ \mathcal{P}^{coll} &= \mathbb{E}[\Theta] = \frac{\gamma}{\beta}, \\ \mathcal{P}^{bayes} &= \frac{\gamma + \sum_{j=1}^{n} N_j}{\beta + n} = \alpha \bar{\mathbf{N}} + (1-\alpha)\frac{\gamma}{\beta},\end{aligned}$$

wobei $\alpha = n/(n+\beta)$ *und* $\bar{\mathbf{N}} = \frac{1}{n}\sum_{j=1}^{n} N_j$. *Weiters gilt für die Abweichung der geschätzten Prämie zur eigentlichen Prämie*

$$\mathbb{E}\left[\left(\mathcal{P}^{Bayes} - \Theta\right)^2\right] = (1-\alpha)\,\mathbb{E}\left[\left(\mathcal{P}^{coll} - \Theta\right)^2\right] = \alpha \mathbb{E}\left[(\bar{N} - \Theta)^2\right].$$

Beweis: Die Punkte (b) und (c) folgen aus den Modellannahmen. Die Größen $\mathcal{P}^{ind}$ und $\mathcal{P}^{coll}$ kann man direkt durch das Einsetzten der Verteilungsfunktion F_ϑ und Bemerkung 4.2 berechnen. Der Bayes'sche Schätzer ist etwas komplizierter zu berechnen. Es gilt ja für den a'posteriori Dichtefunktion von Θ ($\mathbf{N} = (N_1, \ldots, N_n)$)

$$\begin{aligned}u(\vartheta \mid \mathbf{N}) &= \frac{u(\vartheta) \cdot \prod_{j=1}^{n} f_\vartheta(N_j)}{\int_0^\infty u(\vartheta) \cdot \prod_{j=1}^{n} f_\vartheta(N_j)\, d\vartheta} \\ &= \frac{\frac{\beta^\gamma}{\Gamma(\gamma)} \vartheta^{\gamma-1} e^{-\beta\,\vartheta} \cdot \prod_{j=1}^{n} e^{-\vartheta} \frac{\vartheta^{N_j}}{N_j!}}{\int_0^\infty \frac{\beta^\gamma}{\Gamma(\gamma)} \vartheta^{\gamma-1} e^{-\beta\,\vartheta} \cdot \prod_{j=1}^{n} e^{-\vartheta} \frac{\vartheta^{N_j}}{N_j!}\, d\vartheta} \\ &= \frac{\vartheta^{\gamma-1} e^{-\beta\,\vartheta} \cdot \prod_{j=1}^{n} e^{-\vartheta}\, \vartheta^{N_j}}{\int_0^\infty \vartheta^{\gamma-1} e^{-\beta\,\vartheta} \cdot \prod_{j=1}^{n} e^{-\vartheta}\, \vartheta^{N_j}\, d\vartheta}.\end{aligned}$$

Da gilt

$$\prod_{j=1}^{n} e^{-\vartheta}\, \vartheta^{N_j} = e^{-n\vartheta} \vartheta^{\sum_{j=1}^{n} N_j},$$

und

$$\int_0^\infty e^{-n\vartheta} \vartheta^{\sum_{j=1}^{n} N_j} \vartheta^{\gamma-1} e^{-\beta\,\vartheta} d\vartheta = (\beta+n)^{-\gamma-\sum_{j=1}^{n} N_j} \Gamma(\gamma + \sum_{j=1}^{n} N_j)$$

kommt man nach einiger Rechnung und unter Benützung von Gleichung (2.4) auf

$$T(\mathbf{N}) = \int_0^\infty \frac{\vartheta\, \vartheta^{\gamma+\sum_{j=1}^{n} N_j - 1} e^{-(\beta+n)\vartheta}}{(\beta+n)^{-\gamma-\sum_{j=1}^{n} N_j} \Gamma(\gamma + \sum_{j=1}^{n} N_j)}\, d\vartheta\,.$$

Man sieht leicht dass die Dichtefunktion innerhalb vom Integral, i.e. $\vartheta \mapsto \vartheta^{\gamma+\sum_{j=1}^{n} N_j - 1} e^{-(\beta+n)\vartheta}$, wieder einen Dichtefunktion der Gamma Verteilung ist, allerdings mit den neuen Parametern

$$\gamma' = \gamma + \sum_{j=1}^{n} N_j$$

und

$$\beta' = \beta + n.$$

Da der Erwartungswert einer Gamma verteilten Zufallsvariable der Quotient der beiden Parameter ist, folgt für den Schätzer

$$T(\mathbf{N}) = \frac{\gamma + \sum_{j=1}^{n} N_j}{\beta + n}.$$

☻

Einsetzten der Daten - Schätzen der Parameter γ und β: Wir haben jetzt die Struktur des Schätzers T abgeleitet, aber in Abhängigkeit der beiden Größen α und β. Diese ist ja die a'priori Abschätzung von Θ. Die Parameter werden aus allen Daten geschätzt, da man hier am meisten Daten zur Verfügung hat. Bichsel hatte folgende Daten aus dem Jahre 1961 zur Verfügung:

k	0	1	2	3	4	5	6	Total
$N(k) = \#$ Polizen mit k Forderungen	103 704	14 075	1766	255	45	6	2	119 853

Die Gesamtanzahl war also $I = 119853$. Man hat also Paare $\{(\Theta_i, N^{(i)}) : i = 1, 2, \ldots, I\}$, wobei Θ_i die Risikobereitschaft und $N^{(i)}$ die Anzahl der Unfälle im Jahre 1961 des i–ten Fahrer sind. Man kann nun α und β auf verschiedene Arten schätzen. Bichsel benützte die Momentenmethode. Dass heißt er schätze im ersten Schritt den Mittelwert und die Varianz der Grundgesamtheit des Samples:

$$\hat{\mu}(N) \quad = \quad \frac{1}{I} \sum_{i=1}^{I} N^{(i)} = 0.155$$

und

$$\hat{\mathbf{Std}}(N)^2 \quad = \quad \frac{1}{I-1} \sum_{i=1}^{I} (N^{(i)} - \hat{\mu}(N))^2 = 0.179.$$

Andererseits wissen wir auf Grund des Satzes der vollständigen Wahrscheinlichkeit

$$\mu_N = \mathbb{E}[N] = \mathbb{E}\left[\mathbb{E}[N \mid \Theta]\right].$$

Ist Θ gegeben, so ist N exponential verteilt mit Parameter Θ. Damit ist $\mathbb{E}[N \mid \Theta] = \Theta$ und somit gilt

$$\mu_N = \mathbb{E}\left[\Theta\right].$$

Nun wissen wir dass aufgrund der Modellannahme (c) Θ Gamma verteilt ist also

$$\mu_N = \frac{\gamma}{\beta}.$$

Ähnlich gilt für die Varianz

$$\begin{aligned}\mathbf{Std}^2(N) &= \mathbf{Var}[N] = \mathbb{E}[\mathbb{E}[N^2 \mid \Theta]] - \mathbb{E}[\mathbb{E}N \mid \Theta]]^2 = \mathbb{E}[\mathbb{E}[N^2 \mid \Theta]] - \mathbb{E}[\mathbb{E}[N \mid \Theta]^2] + \mathbb{E}[\mathbb{E}N \mid \Theta]]^2 \\ &= \mathbb{E}\left[\mathbb{E}[N^2 \mid \Theta\right] - (\mathbb{E}[N \mid \Theta])^2] + \mathbb{E}[(\mathbb{E}[N \mid \Theta])^2] - \mathbb{E}[N]].\end{aligned}$$

Der bedingte Erwartungswert ist auch eine Zufallsvariable, genau genommen gilt

$$\begin{aligned}\mathbb{E}[N \mid \Theta] : \Omega &\longrightarrow \mathbb{R} \\ \omega &\mapsto \mathbb{E}[N \mid \Theta(\omega)].\end{aligned}$$

Rechnet man die Varianz des bedingten Erwartungswert aus, kommt man auf

$$\mathbf{Var}(\mathbb{E}[N \mid \Theta]) = \mathbb{E}\left(\mathbb{E}[N \mid \Theta] - \mathbb{E}\mathbb{E}[N \mid \Theta]]\right)^2 = \mathbb{E}\left(\mathbb{E}[N \mid \Theta]\right)^2 - \mathbb{E}[\mathbb{E}[N \mid \Theta]]^2 = \mathbb{E}(\mathbb{E}[N \mid \Theta])^2 - \mathbb{E}[N]^2.$$

Auf der anderen Seite, ist die bedingte Varianz

$$\begin{aligned}\mathbf{Var}[N \mid \Theta(\omega)] : \Omega &\longrightarrow \mathbb{R} \\ \omega &\mapsto \mathbf{Var}[N \mid \Theta(\omega)],\end{aligned}$$

definiert durch

$$\mathbf{Var}[N \mid \Theta(\omega)] = \mathbb{E}\left[(N - (\mathbb{E}[N \mid \Theta])^2 \mid \theta\right] = \mathbb{E}\left[N^2 \mid \Theta\right] - (\mathbb{E}[N \mid \Theta])^2.$$

Beachte, die bedingte Varianz ist wie der bedingte Erwartungswert eine Zufallsvariable. Setzt man die bedingte Varianz und den bedingten Erwartungswert oben ein kommt man auf

$$\mathbf{Std}^2(N) = \mathbb{E}\left[\mathbf{Var}[N \mid \Theta]\right] + \mathbf{Var}[\mathbb{E}[N \mid \Theta]].$$

Mit Hilfe der Annahme (b) lässt sich $\mathbf{Var}[N \mid \Theta]$ und die Varianz von $\mathbb{E}[N \mid \Theta]$ berechnen. Genauer gesagt, is N exponential verteilt mit Parameter Θ ist der Erwartungswert und die Varianz gleich Θ. Es gilt daher

$$\mathbf{Var}[N] = \quad = \mathbb{E}\left[\Theta\right] + \mathbf{Var}[\Theta].$$

Aus Modellannahme (c) folgt, dass

$$\mathbf{Var}[N] = \quad = \frac{\gamma}{\beta}\left(1 + \frac{1}{\beta}\right).$$

In der Momentenmethode setzt man nun den geschätzten Wert gleich den theoretischen Wert, bzw. $\hat{\mu}(N)$ gleich $\mu(N)$ und $\hat{\mathbf{Std}}(N)$ gleich $\mathbf{Std}(N)$. Damit erhält man zwei Gleichungen mit zwei unbekannten (nämlich γ und β). Diese Rechnung führt zu (wir versehen hier die geschätzten Werte mit einen Hütchen)

$$\frac{\hat{\gamma}}{\hat{\beta}} = 0.1555, \qquad \frac{\hat{\gamma}}{\hat{\beta}}\left(1 + \frac{1}{\hat{\beta}}\right) = 0.179,$$

und somit

$$\hat{\gamma} = 1.001 \quad \text{und} \quad \hat{\beta} = 6.458.$$

☺

Verifikation des Modells Setzten wir die unbekannten Parameter β und γ für die Bayes Schätzer ein, so erhalten wir

$$\mathcal{P}^{bayes} = \frac{\hat{\gamma} + \sum_{j=1}^{n} N_j}{\hat{\beta} + n}.$$

Jetzt ist die Frage zu klären, ob die Modellannahmen auch passen. Dazu vergleichen wir zuerst die empirische Verteilung mit der theoretischen Verteilung. Dabei vergleichen wir das Modell mit Modellannahmen (a), (b) und (c) (im weiteren genannt Modell (A)) mit einem Modell ohne Risikoklassen, also folgenden Modellannahmen:

Modellannahmen des Modells B: Die Anzahl $N^{(i)}$ der Schadensmeldungen pro Jahr ist Poisson verteilt mit einem fixen Parameter λ.

Schätzt man λ mit dem Mittelwert ab, erhält man $\hat{\lambda} = 0.155$.

Für Modell (B) gilt also

$$p_k^B = \mathbb{P}\left(N^i = k\right) = e^{-\hat{\lambda}} \frac{\hat{\lambda}^k}{k!}$$

Multipliziert man p_k^B mit der Anzahl der Kunden, bekommt man die Werte in der Tabelle **??**.

In gleicher weise kann man sich die Werte für das Modell A ausrechnen. Zuerst berechnen wir p_k^A.

$$\begin{aligned}
p_k^A &= \mathbb{P}\left(N^{(i)} = k\right) = \int_0^\infty \mathbb{P}\left(N^{(i)} = k \mid \Theta = \vartheta\right)\, u(\vartheta)\, d\vartheta \\
&= \int_0^\infty \frac{\beta^\gamma}{\Gamma(\gamma)} \vartheta^{\gamma-1} e^{-\beta\vartheta}\, e^{-\vartheta} \frac{\vartheta^k}{k!}\, d\vartheta \\
&= \frac{\beta^\gamma}{\Gamma(\gamma)} \frac{1}{k!} \underbrace{\int_0^\infty \vartheta^{\gamma-1} e^{-(\beta+1)\vartheta}\, \vartheta^(\gamma + k - 1)\, d\vartheta}_{=\frac{\Gamma(\gamma+k)}{(\beta+1)^{\gamma+k}}} \\
&= \frac{\Gamma(\gamma+k)}{(\beta+1)^{\gamma+k} k!} \left(\frac{\beta}{(\beta+1)^\gamma}\right)^\gamma \cdot \left(\frac{1}{\beta+1}\right)^k \\
&= \binom{\gamma+k-1}{k} p^\gamma (1-p)^k, \quad \text{mit} p = \frac{\beta}{\beta+1}.
\end{aligned}$$

Das heißt es kommt eine negative Binomial Verteilung heraus. Multipliziert man wiederum p_k^A mit der Anzahl der Kunden, bekommt man die Werte in der Tabelle **??**.

k	beobachtete	Poisson ($\lambda = 0.1555$)	Negative Binomial ($\gamma = 1.001, \beta = 6.458$)
0	103 704	102 629	103 757
1	14 075	15 922	13 934
2	1 766	1 234	1 871
3	255	64	251
4	45	3	34
4	6	0	5
6	2	0	1

Man sieht sehr schön, das Modell A recht gut passt.

☺ ☺ ☺ ☺

Bichsel benützte dieses Modell und die obigen Daten als Basis um das Bonus Malus System für Autofahrer in der Schweiz zu konstruieren, die Prämie der Autofahrer die weniger als halb soviel Unfälle hatten als im Durchschnitt wurde halbiert. Mittlerweile wurde das Verfahren weiterentwickelt aber es sind ähnliche Überlegungen die zum heutigen Bonus Malus System führen. Nähere Information ist nachzulesen bei [**?**].

5 Konjugierte Klassen von Verteilungen

Wir haben im vorigen Beispiel gesehen, dass sich die Modellannahmen (b) und (C), i.e. N Poisson und Θ Gamma-Verteilt wunderbar ausgehen. Dies ist nicht immer der Fall. Es gibt Verteilungen die wunderbar

zusammenpassen, und Verteilungen, die analytisch sehr aufwendig sind. Passen Verteilungen sehr gut zusammen heißen diese Verteilungen konjugiert.

In diesem Kapitel werden wir kurz die Verteilungen die sehr gut zusammenpassen aufzählen und vorstellen.

i Der Binomial – Beta Fall

Ein Betrieb produziert Werkstücke . Diesen Werkstücke werden von verschiedenen Maschinen produziert jede Maschine hat eine bestimmte Ausfall Rate. Diese Ausfallrate kann bei verschiedenen Maschinen unterschiedlich sein. Allerdings hat jede Maschine eine konstante Ausfall Rate die nicht in der Zeit variiert. Die Aufgabe ist es diese Ausfallrate in Abhängigkeit von der Maschine zu schätzen.

Mathematisch gesehen, hat man verschiedene Gruppen von Werkstücken. Man ist an der Wahrscheinlichkeit des Auftretens eines Schadenfalles innerhalb einer bestimmten Gruppe interessiert. Hier nehmen wir an dass jedes Werkstück (Element) der Gruppe (oder Klasse) mit gleicher Wahrscheinlichkeit defekt wird und das dies unabhängig von den anderen Gruppenelementen passiert. Auch nehmen wir an dass ein Element in Falle eines Defektes oder Schadenfalls die Gruppe verlässt.

Wir führen nun folgende Zufallsvariable ein: Für ein fixes Jahr j und Maschiene sei

- $N = j$ die Anzahl der Schadensfälle innerhalb einer Gruppe;
- $V = j$ die funktionierenden Werkstücke am Anfang des Jahres j;
- und $X_j := \frac{N_j}{Vj}$ = die empirische Wahrscheinlichkeit dass ein Werkstück defekt wird.

Zur Zeit n sind alle Zufallsvariablen mit Index kleiner oder gleich n bekannt. Wir sind nun an

$$X_{n+1} = \frac{N_{n+1}}{V_{n+1}}$$

interessiert. V_{n+1} is bekannt, aber N_{n+1} ist unbekannt. Das heißt wir müssen eine Methode finden um N_{n+1} zu modellieren.

Modellannahmen:

- Unter der Bedingung dass der Parameter ϑ gegeben ist, sind die N_j, $j = 1, 2, \ldots$ unabhängig und binominal verteilt, i.e.

$$\mathbb{P}\left(N_j = k \mid P = \vartheta\right) = \binom{V_{n+1}}{k} \vartheta^k (1-\vartheta)^{V_{n+1}}.$$

- Der Parameter Θ ist Beta(a, b) verteilt mit $a, b > 0$, bzw.

$$u(\vartheta) = \frac{1}{B(a,b)}\, \vartheta^{a-1}(1-\vartheta)^{b-1}, \quad 0 \le \vartheta \le 1,$$

  wobei gilt $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{(a+b)}$.

Bemerkung 5.1.

- *Die ersten beide Momente der Beta Verteilung lauten*

$$\mathbb{E}\Theta = \frac{a}{a+b}, \quad \mathbf{Var}(\Theta) = \frac{a\,b}{(1+a+b)(a+b)^2}$$

- *Wir wollen den* disablement probability *schätzen;*
- *Die Familie der Beta Verteilungen die mit a und b beschrieben werden ist ziemlich gross.*

Proposition 5.1. *Unter den obigen Modellannahmen gilt*

$$\begin{aligned} F^{ind} &= \mathbb{E}[X_{n+1}|\Theta] = \Theta, \\ F^{coll} &= \mathbb{E}E[\Theta] = \frac{a}{a+b} \\ F^{Bayes} &= \frac{a+\sum_{j=1}^n N_j}{a+b+\sum_{j=1}^n V_j} = \alpha\bar{N} + (1-\alpha)\frac{a}{a+b} \\ & \quad \textit{wobei gilt} \quad \bar{N} = \frac{\sum_{j=1}^n N_j}{\sum_{j=1}^n V_j},\ \alpha = \frac{\sum_{j=1}^n V_j}{a+b+\sum_{j=1}^n V_j}. \end{aligned}$$

Der quadratische Verlust des Schätzers F^{Bayes} beträgt

$$\mathbb{E}(F^{Bayes} - \Theta)^2 = (1-\alpha)\mathbb{E}(F^{coll} - \Theta)^2 = \alpha\mathbb{E}(\bar{N} - \Theta)^2.$$

ii Der Normal Normal Fall

Angenommen eine Maschine produziert Bauteile einer bestimmten Länge L. Die Maschinen müssen adjustiert werden, dabei passiert es öfters das die Länge der Bauteile die erzeugt werden variieren, d.h. um einen Mittelwert schwanken. Wird so ein Bauteil gemessen, gibt es aber Messfehler, bzw. passieren bei der Produktion Fehler. Wir führen nun folgende Zufallsvariable ein: Für das $j$te Bauteil gilt sei

- L_j die Länge des j ten Bauteils
- die Maschine produziert Bauteile der Länge Θ

Modellannahmen:

- Unter der Bedingung dass der Parameter ϑ gegeben ist, sind die L_j, $j = 1, 2, \ldots$ unabhängig und normal verteilt mit Mittelwert ϑ und Varianz σ^2, i.e. $L \sim \mathcal{N}(\vartheta, 0.01)$.
- Der Parameter Θ ist Normalverteilt mit Mittelwert μ und Varianz τ^2.

Proposition 5.2. *Unter den obigen Modellannahmen gilt*

$$\begin{aligned} P^{ind} &= \mathbb{E}[X_{n+1}|\Theta] = \Theta, \\ P^{coll} &= \mathbb{E}[\Theta] = \mu, \\ P^{Bayes} &= \frac{\tau^2\mu + \sigma^2\left(\sum_{j=1}^n X_j\right)}{\tau^2 + n\sigma^2} = \alpha\bar{X} + (1-\alpha)\mu \\ & \quad \textit{wobei gilt} \quad \bar{X} = \frac{1}{n}\sum_{j=1}^n X_j,\ \alpha = \frac{n}{n^2 + \frac{\sigma^2}{\tau^2}}. \end{aligned}$$

Der quadratische Verlust des Schätzers F^{Bayes} beträgt

$$\mathbb{E}(F^{Bayes} - \Theta)^2 = (1-\alpha)\mathbb{E}(F^{coll} - \Theta)^2 = \alpha\mathbb{E}(\bar{X} - \Theta)^2.$$

iii Pareto Gamma Fall

Eine oft gemachte Annahme in der Versicherungsmathematik ist, dass die Beträge von verschiedenen Forderungen die größer als eine bestimmten Schwellenwert x_0 sind, Parteo verteilt mit einen bestimmten Parameter α sind. Dass dies eine sinnvolle Annahme ist, ist in Kapitel über seltene Ereignisse nachzulesen.

Einerseits, haben Versicherungen zumeist schon eine Idee wir groß der Parameter sein wird, oder wie dieser verteilt ist. Andererseits, kann der Versicherer Information von vorhergehenden Forderungen

nutzen. Die Frage die hier auftritt, wie gut kann man beide Informationen kombinieren um ein möglichst guten Schätzer für α zu erhalten.

Sei $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ ein Vektor wobei im Vektor nur die Forderungen stehen die einen bestimmten Schwellenwert x_0 überschreiten (Rückversicherer). Wir nehmen aus oben genannten Grund an, dass x_j identisch und unabhängig Pareto verteilte Zufallsvariablen mit Parameter θ sind, also

$$x_j \sim \mathrm{Par}(\theta), \quad j = 1, \ldots, n.$$

mit Dichte

$$f_\theta(x) = \begin{cases} \dfrac{\theta}{x_{\min}} \left(\dfrac{x_0}{x}\right)^{-\theta+1} & x \geq x_{\min} \\ 0 & x < x_{\min} \end{cases}$$

und Verteilungsfunktion

$$F_\theta(x) = 1 - \left(\frac{x}{x_0}\right)^{-\theta}. \tag{2.5}$$

Eine kurze Rechnung ergibt für die Momente

$$\mu(\theta) = x_0 \frac{\theta}{\theta - 1}, \ \theta > 0, \quad \text{und} \quad \sigma^2(\theta) = x_0^2 \frac{\theta}{(\theta-1)^2(\theta-2)}, \ \theta > 2.$$

Um vorhandene Informationen miteinzubeziehen, nehmen wir an, dass der Parameter θ selber eine Zufallsvariable ist, mit Verteilungsfunktion $U(\theta)$. Hier nehmen wir an, dass $U(\theta)$ einer bestimmten Klasse von Verteilgungsfunktionen zugehört und zwar allen Verteilungsfunktionen der folgenden Familie von Verteilungsfunktionen:

$$\mathcal{U}' = \left\{ u_{\mathbf{x}} : u_{\mathbf{x}}(\theta) \sim \theta^n \exp\left(- \left(\sum_{j=1}^{n} \ln\left(\frac{x_j}{x_0}\right) \right) \theta \right) \right\}.$$

Die Familie $\mathcal{U}'$ besteht aus Γ–Verteilungen. Das heißt die natürliche Wahl als Ausgangsmenge ist die Famile aller Gamma Verteilungen, i.e.

$$\mathcal{U} = \left\{ \text{Verteilung mit Dichte } u(\theta) = \frac{\beta^r}{\Gamma(r)} \theta^{\nu-1} \exp(-\beta\theta) \right\}. \tag{2.6}$$

Da die Summe zweier Γ verteilten Zufallsvariablen wieder eine Γ verteilte Zufallsvariable ergibt, ist leicht zu sehen, dass die zu $\mathcal{U}$ konjugierte Klasse von Verteilungsfunktionen Pareto Verteilungen sind. Genauer,

$$\mathcal{F} = \{(x_0, \theta) : \theta, x_0 > 0\}$$

Hat man jetzt als Liklihoodfunktion (2.5) und a-priori Verteilung gegeben durch (2.6), erhält man als a'posteriori Verteilungsfunktion eine Verteilungsfunktion mit Dichte

$$u_{\mathbf{x}}(\theta) \sim \theta^{\gamma+n-1} \exp\left\{ - \left(\beta + \sum_{j=1}^{n} \ln\left(\frac{x_j}{x_0}\right) \right) \theta \right\}$$

Vergleicht man die Parameter mit den Parametern eine Γ–Verteilung so sieht man dass u_x wieder eine Γ–Verteilung ist mit Parametern

$$\gamma' = \gamma + n \quad \text{and} \quad \beta' = \beta + \sum_{j=1}^{n} \ln\left(\frac{x_j}{x_0}\right)$$

Der Bayes Schätzer für Θ ist gegeben durch

$$\tilde{\Theta} = E\left[\Theta|\mathbf{x}\right] = \frac{\gamma + n}{\beta + \sum_{j=1}^{n} \ln\left(\frac{x_j}{x_0}\right)} \tag{2.7}$$

Der Bayes Schätzer (2.7) kann leicht umgeschrieben werden. Dazu definieren wir zuerst den Maximum Liklihood Schätzer des Parameters einer Paretoverteilung, falls keine Vorinformation vorliegt. Dieser ist gegeben durch

$$\widehat{\Theta}_{MLE} = \frac{n}{\sum_{j=1}^{n} \ln\left(\frac{x_j}{x_0}\right)}$$

Mit oben kann man (2.7) umschreiben zu

$$\hat{\Theta} = \alpha\, \widehat{\Theta}_{MLE} + (1-\alpha)\frac{\gamma}{\beta}.$$

mit

$$\alpha = \left(\sum_{j=1}^{n}\left(\frac{x_j}{x_0}\right)\right)\left(\beta + \sum_{j=1}^{n}\ln\left(\frac{x_j}{x_0}\right)\right)^{-1}$$

Das heisst der Bayes Schätzer von Θ ist ein gewichteter Mittelwert vom Maximum Liklihood Schätzer der Paretoverteilung gegeben die Daten $\mathbf{x}$ und den Erwartungswerty der a'priori Verteilung.

iv Gamma - Gamma:

Wir beobachten Zufallsvariable X_j, $j = 1, 2, \ldots$ die bedingt Gamma verteilt sind mit Parameter Θ. Der Parameter Θ ist wiederum Gamma verteilt mit zu schätzenden Parameter.

v Geometrisch Beta:

Wir beobachten Zufallsvariable X_j, $j = 1, 2, \ldots$ die bedingt geometrisch verteilt sind mit Parameter Θ. Der Parameter Θ ist Beta verteilt mit zu schätzenden Parameter.

Kapitel 3

Risikotheorie

1 Mathematische Modellierung von Risiken

Betrachten wir ein Versicherungsunternehmen. Bei Eintreten des Versicherungsfalles entsteht dem Versicherungsunternehmen ein Schaden und es muss den vertraglich vereinbarten Betrag an den Versicherungsnehmer auszahlen. Jedoch kommt es nicht bei jedem Vertrag oder Police des Unternehmens zwangsläufig zur Auszahlung und die Höhe der Auszahlung ist in den meisten Fällen auch nicht vorher festgelegt. Der verursachte Schaden jedes Vertrages kann als ein Wert betrachtet werden, der dem Zufall unterworfen ist.

Eine Firma hat einen großen Maschinenpark. Es kommt immer wieder zum Ausfall einiger Maschinen, die Reparatur verursacht Kosten. Die Kosten sind aber je nach Schadensart verschieden und kann man im allgemeinen nicht vorhersagen. Dabei modelliert man die Kosten als Zufallsvariable, die nur negative Werte annnehmen kann. Relevante Fragen für die Firma sind z.B. wie groß die Rücklagen sein mindestens sein müssen, um mit einer 99.9 prozentiger Sicherheit den Schaden ohne Kredite aufzunehmen bezahlen kann.

Dazu wird im Folgenden als relevante Zufallsvariable meist der innerhalb der vorgegebenen Zeitperiode entstandene Schaden bzw. Verlust V betrachtet. Ein finanzieller Verlust bzw. eine Schadenshöhe wird dabei durch einen positiven Wert und ein finanzieller Gewinn durch einen negativen Wert von V beschrieben. Diese Konvention ist im Zusammenhang mit finanziellen Risiken zunächst etwas gewöhnungsbedürftig, besitzt aber auch dort einige Vorteile, z. B. kann man in Bilanzsimulationen positive Werte von V als erforderliches Sicherheitskapital interpretieren.

Alternativ ist es aber auch möglich, unmittelbar die Zufallsvariable V zu betrachten, also Gewinne mit einem positivem Vorzeichen zu versehen und Verluste mit einem negativen. In der Literatur kommen beide Vorzeichenkonventionen vor, je nachdem ob eher Schäden / Verluste oder mögliche Gewinne im Mittelpunkt der Betrachtung stehen.

Mathematisch gesehen definiert man ein Risiko (bzw. ein Portfolio an Risiken) als nicht negative Zufallsvariable.

Definition 1.1. *Eine nicht negative Zufallsvariable X heißt Risiko. Eine Menge $\{X_k : k = 1, \ldots\}$ von Risiken X_k heißt Portfolio.*

Im individuellen Modell eines Versicherungsunternehmens lässt sich ein Risiko X_k als der Schaden interpretieren, der sich aufgrund des k-ten Versicherungsvertrages (Police) in dem betrachteten Zeitraum, z.B. ein Jahr, ergibt. Offensichtlich ist dann die Gesamtsumme, die das Versicherungsunternehmen in einem Jahr auszahlen muss, gleich der Summe S_n der Risiken $X_1, \cdots, X_n$.

Wir werden in unserer Vorlesung auf folgende Punkte näher eingehen

Schadenshöheverteilungen für Einzelschäden Die Schadenhöhe pro eingetretenem Schadenfall für ein Einzelrisiko X_k, $k = 1, \ldots,$ wird durch eine nichtnegative Zufallsvariable X bzw. durch ihre zugehörige

Wahrscheinlichkeitsverteilung beschrieben. Meist wird sie zeitunabhängig modelliert. Die Dimension der Zufallsvariablen ist in der Regel eine Geldeinheit, etwa Euro. Das zu modellierende Einzelrisiko kann beispielsweise eine in der Produktion eingesetzte Maschine, ein versichertes Kraftfahrzeug oder etwas allgemeiner aufgefasst auch ein Kredit sein (wobei bei einem Kredit der *Schadenfalls* durch eine nicht ordnungsgemäße Bedienung der Zahlungsverpflichtungen durch den Kreditnehmer eintritt).

Schadensanzahlverteilungen Die Anzahl von Schäden gleichartiger Risiken (z.B. von gleichartigen Maschinen in einer Fabrik, Kfz in einem homogenen Versichertenkollektiv) innerhalb eines vorgegeben Zeitraums (z. B. ein Jahr) wird durch eine nicht negative ganzzahlige Zufallsvariable N beschrieben. Allgemeiner beschreibt $N(t)$ die Anzahl der Schäden im Zeitraum $[0, t]$. Etwas weiter gefasst tauchen Schadenanzahlverteilungen auch bei der Modellierung von Kreditausfällen in einem homogenen Bestand von Krediten auf.

Schadensanzahlprozesse Spielt die Zeit eine wichtige Rolle, modelliert man die die Schadensanzahl durch einen Schadensanzahlprozess, bzw. stochastischen Prozess

$$\begin{aligned} N : [0, \infty) \times \Omega &\to \mathbb{N}_0 \\ (t, \omega) &\mapsto N(t, \omega). \end{aligned}$$

Das bedeutet, dass gewisse Regeln für die mögliche Abfolge von Schadensfällen im Zeitverlauf aufgestellt werden; d.h.. die Schadensanzahl für zukünftige Zeiträume wird etwa in Abhängigkeit von den zuvor eingetretenen Schadensfällen modelliert (im Sinne bedingter Wahrscheinlichkeiten).

Gesamtschadenverteilungen Eine Gesamtschadenverteilung gibt die Wahrscheinlichkeitsverteilung der gesamten Schadenshöhe S bzw. S(t) einer bestimmten Gesamtheit (eines sogenannten Kollektivs) von Risiken im Zeitraum $[0, t]$ an, beispielsweise die Kfz-Schäden in einem Fuhrpark mit mehreren Fahrzeugen. Sie ergibt sich grob gesprochen aus der Schadenanzahl Verteilung und der Schadenshöhenverteilung pro Einzelschaden. Eine explizite Bestimmung der Gesamtschadenverteilung aus diesen beiden Grundbausteinen ist allerdings nur in einfachen Spezialfällen möglich, vor allem wenn allgemein auch Zusammenhänge (z.B. Korrelationskoeffizienten) der Einzelrisiken u. Ä. zu berücksichtigen sind.

Renditeverteilung Bildet ein Bauunternehmen Rücklagen und legt diese an, werden diese zu einen bestimmten Zinssatz verzinst. Ähnlich legt ein Versicherungsunternehmen die Prämien an, z.B. kauft es griechische Staatsanleihen. Das Problem ist, dass die Verzinsung, bzw. Rendite nicht fest sind sondern auch vom Zufall bestimmt ist. Eigentlich müsste man dies auch miteinbeziehen, aber hier in unseren Skript nehmen wir an, dass der Zinssatz fix ist.

Gesamtrisikoprozess Ähnlich wie Gesamtschadenverteilungen mehrerer Einzelrisiken kann man auch die Gesamtwertverteilung verschiedener unter Risiko stehender Wertobjekte berechnen, z. B. bezüglich des Gesamtwerts eines Wertpapierportfolios. Noch allgemeiner könnte man beispielsweise auch den zukünftigen Gesamtwert eines Unternehmens modellieren, indem alle relevanten Schadensprozesse, Preisprozesse u. Ä. in einer Gesamtbetrachtung zusammengeführt werden. Dies kann allerdings wegen der Vielzahl zu berücksichtigender Faktoren beliebig kompliziert werden und ist in der Regel nur unter starken Modellvereinfachungen durchführbar.

i Verteilungsmodelle für Einzelschäden

Im Folgenden werden verschiedene konkrete Verteilungsmodelle zur Modellierung der Schadenshöhe (synonym: Schadensumme) einzelner Schäden betrachtet. Für Schadenhöhenverteilungen kommen in erster Linie stetige Modelle infrage bei denen die Verteilung also durch eine sogenannte Dichtefunktion $f(x)$ beschrieben wird, welche in dem vorliegenden Kontext zudem nur für $x > 0$ von null verschiedene Werte besitzt.

Des Weiteren hängt der einzusetzende Verteilungstyp davon ab, ob es sich grundsätzlich um Schadenart handelt, bei der nur kleinere oder mittelgroße Schäden möglich sind, wie etwa der Kfz-Kasko-Versicherung, oder ob Großschäden möglich oder gar *wahrscheinlich* sind, beispielsweise in der Industrie-Haftpflichtversicherung (Haftpflichtversicherung für Unternehmen) oder der Rückversicherung (Versicherung für Versicherungsunternehmen) von Natur Katastrophen.

In der Praxis spricht man von Großschadenbeständen, wenn die Summe eines kleinen Teils der beobachteten Schäden einen großen Anteil in der Summe aller Schäden ausmacht. Sind $X_1, X_2, \dots X_n$ die beobachteten Schäden und $x(1), x(2), \dots x(n)$ die nach ihrer Größe geordneten Schäden, so spricht man beispielsweise von einem Großschadenbestand, wenn für ein $m \leq 0.2n$

$$\sum_{i=1}^{m} x(i) \geq 0.8 \sum_{i=1}^{n} X_i,$$

also wenn 20 Prozent aller Schäden mindestens 80 Prozent des Gesamtschadens ausmachen.

Kleinschadenverteilungen kommen z. B. auch bei Versicherungen mit Selbstbeteiligung vor. In diesem Fall ist besonders offensichtlich, dass Kleinschäden in der Regel durch eine feste Schadenobergrenze charakterisiert sind. Die zugehörige Wahrscheinlichkeitsverteilung hat also einen beschränkten Träger (Bereich mit Dichtefunktion $f(x) \neq 0$). Ferner bieten sich Verteilungen mit beschränktem Träger an, wenn Schadenquoten modelliert werden, d.h. wenn Schäden prozentual zu einer Bezugsgröße angegeben werden (üblich z.B.: Versicherungsleistungen bezogen auf Beitragseinnahmen).

Risiken, die typischerweise mit Schäden im kleinen und mittleren Bereich verbunden sind, zeichnen sich dadurch aus. dass die zugehörigen Schadenverteilungen einen schmalen Tail besitzen (falls nicht ohnehin eine feste Obergrenze vorliegt). Das heißt, im auslaufenden Ende der Wahrscheinlichkeilsverteilung befindet sich relativ wenig *Wahrscheinlichkeitsmasse*. Demgegenüber haben Großschadenverteilungen einen breiten Tail, sie besitzen also relativ viel Wahrscheinlichkeitsmasse im auslaufenden Ende der Verteilung; man spricht auch von Fat-Tail- oder Heavy-Tail-Verteilungen. Der Begriff der Heavy-Tail-Verteilung wird nicht einheitlich verwendet. In den meisten Fällen (auch hier) versteht man darunter eine Verteilung, deren Tail-Wahrscheinlichkeiten $P(X > x) = 1 - F(x)$ für $x \to \infty$ langsamer als jede exponentiell fallende Funktion gegen null konvergieren.

Definition 1.2. *Eine Verteilung Q auf $(\mathbb{R}^+; \mathcal{B}(\mathbb{R}^+))$ ist* heavy-tailed*, falls gilt:*

$$\int_{\mathbb{R}^+} \exp(sx) Q(dx) = \infty \quad \text{für alle } s > 0 :$$

Maßgeblich für die Risikomodellierung ist außer einer Annahme zur grundsätzlichen Form des Tails die Auswahl diverser charakteristischer Verteilungsparameter (Lage-, Form- und Skalenparameter), wie im Folgenden bei den einzelnen Verteilungsmodellen näher erläutert wird.

Wir werden später sehen, dass bei Wahrscheinlichkeitsverteilungen, die eine hohe Tail-Wahrscheinlichkeit besitzen, der Erwartungswert oder die Varianz sehr wenig aussagen. Dazu folgendes Beispiel.

Beispiel 1.1. *Sei X normalverteilt mit Varianz $\sigma = 0.092298493674638$ und Mittelwert $\nu = 1,1185$. Dann gilt $\mathbb{P}(X - \mu \leq 5\sigma) \leq 2.866 \cdot 10^{-7}$. Sei Z eine Log Normalverteilte Zufallsvariable mit Parameter $\nu = 7.64$ und $\sigma = 26.68$. Beide Zufallsvariable haben den gleichen Erwartungswert und die gleiche Varianz. Im zweiten Fall gilt aber $\mathbb{P}(Z - \mu \leq 5\sigma) \leq 0,158045336794519$. D.h. die Wahrscheinlichkeit dass ein Wert mehr als $5 \cdot \sigma$ vom Mittelwert abweicht ist nicht zu vernachlässigen !*

Hier eine kurze Klassifizierung von Verteilungen:

Beispiel 1.2.

- *Die Lognormalverteilung ist heavy-tailed.*
- *Die Normalverteilung ist nicht heavy-tailed.*
- *Die Gammaverteilung ist nicht heavy-tailed.*
- *Die Inverse-Gauss-Verteilung ist nicht heavy-tailed.*

- *Die Cauchy und Pareto Verteilungen sind heavy tailed.*

Bemerkung 1.1.
- *Oft modelliert man die Risikoverteilungen mittels der Gamma und Inverse-Gauss-Verteilung, obwohl diese nicht heavy-tailed sind. Dies kann trotzdem sinnvoll sein, wenn z.B. keine Schäden extremer Höhe zu erwarten sind oder wenn bei der Modellierung andere Aspekte im Vordergrund stehen. Die Lognormalverteilung ist heavy-tailed, modelliert also realistisch auch das Vorliegen von Großschäden.*

- *Für eine heavy-tailed Verteilung Q mit Verteilungsfunktion F gilt:*

$$\limsup_{x\to\infty} \exp(sx)(1-F(x)) = 1$$

für jedes $s > 0$.

Eine notwendige Bedingung für eine Verteilung, um heavy-tailed zu sein, gibt das nachfolgende Lemma an.

Lemma 1.1. *Es sei Q eine Verteilung auf $(\mathbb{R}^+; \mathcal{B}(\mathbb{R}^+))$ mit Verteilungsfunktion F. Falls*

$$\limsup_{x\to\infty} \frac{-\ln(1-F(x))}{x} = 0,$$

gilt, dann ist Q heavy-tailed.

Eine weitere Klasse von Verteilungen und mit denen Grosschäden modelliert werden, sind die subexponentiellen Verteilungen:

Definition 1.3. *Eine Schadenhöhenverteilung Q heißt* subexponentiell*, wenn für alle $x > 0$ die Beziehung $Q((x,\infty)) > 0$ gilt und wenn*

$$\lim_{x\to\infty} \frac{Q^{*2}((x,\infty))}{Q((x,\infty))} = 2$$

gilt.

Beispiele von Großschadensverteilungen, bzw. subexponentielle Verteilungen sind

- Die Pareto verteilung $Par(a, 1)$;
- Die Weilbull verteilung mit Parameter $\beta < 1$
- Die Lognormalverteilung mit Parametern μ und σ.
- Die Gumbel Verteilung ist subexponentiell;

Für nähere Information siehe Skriptum von Hipp (Risikotheorie) und das Buch von Klüppelberg, Mikosch und Embrechts (Modeling extremal events).

auch α stabile Verteilungen und domain of attraction !!!

Charakterisierung von Verteilungen: qq-plot - vergleich zweier Verteilungen (Heavy tail, exrtrem value

Definition 1.4. *Die* Mean excess function *(MeF) einer Zufallsvariablen ist gegeben durch*

$$e(u) \quad = \quad \mathbb{E}\left[X - u \mid X > u\right].$$

Exponentiell, Erlang und Gammaverteilung

Weilbull Verteilung

Logarithmische Normalverteilung

Log-Gamma Verteilung

Pareto Verteilung

Inverse Gaußverteilung

Die Aufgabe des Unternehmens ist die Einzelverteilungen der Kosten eines Schadens mit Hilfe der Statistik zu analysieren und den Verteilungstyp festzulegen. Hier ist es wichtig zu analysieren, welchen Typ die Verteilung hat, ob es eine heavy tailed, oder subexponentielle Verteilung ist. Im weiteren nehmen wir an, dass ein Portfolio $\{X_k : k \in \mathbb{N}\}$ existiert, wobei die Verteilung F_X der Einzelschäden bekannt ist, d.h. es gilt $X_k \sim F_X$, $k \in \mathbb{N}$.

ii Modellierung der Schadenszahl

Betrachtet man die Anzahl der Maschienenausfällen und daraus anfallenden Kosten innerhalb eines Jahres ist nicht nur die Verteilung der Kosten wichtig, sondern auch die Anzahl der Ausfälle. Dies wird durch eine ganzzahlige Zufallsvariable

$$\begin{array}{rcl} N : \Omega & \longrightarrow & \mathbb{N}_0 \\ \omega & \mapsto & N(\omega) \end{array}$$

modelliert.

Es gibt mehrere Möglichkeiten die Anzahl der Schäden in einen Zeitintervall gegebener Länge zu modellieren; bzw. hat man mehrere Verteilungen zur Verfügung die Anzahl der Schäden zu modellieren. Im den folgenden Kapiteln werden wir einigen Beispiele für solche Schadensanzahl Verteilungen vorstellen.

Bernoulliprozess und die Binominalverteilung

Beim Bernoulli prozess geht man davon aus, dass in einem (kleinen) Zeitraum der Länge 1 (z.B. ein Tag) höchstens ein Schaden auftreten kann, und zwar mit konstanter Wahrscheinlichkeit p;, d.h die Schadenseintrittwahrscheinlichkeit hängt nicht von der Vergangenheit ab.

Fixiere $T \in \mathbb{N}$. Für die Schadensanzahl $N(T)$ nach T Zeitperioden ergibt sich somit eine Binominalverteilung , also $N(T) \sim \mathbf{Bin}(T; p)$. Für die Wahrscheinlichkeit von n Schäden in einem Zeitintervall der Länge T gilt also

$$\mathbb{P}(N(T) = m) = p_n(T) = \binom{T}{n} \cdot p^n \cdot (1-p)^{T-n}.$$

Für zwei unabhängige Einzelverteilungen gilt

$$N_i \sim \mathbf{Bin}(T_i; p)\,(i = 1, 2) \quad \Rightarrow \quad N_1 + N_2 \sim \mathbf{Bin}(T_1 + T_2; p)$$

D.h. fasst man zwei Zeitperioden $[0, T_1]$ und $[T_1, T_1 + T_2]$ mit identischer Schadenseintrittswahrscheinlichkeit p zusammen, so erhält man unter der Annahme der Unabhängigkeit der Zeitperioden für den Gesamtzeitraum $[0, T_1 + T_2]$ wieder eine Binominalverteilung mit Parameter p. Es wird also vorausgesetzt dass die Schäden in $[0, T_1]$ nicht die Schäden in $[T_1, T_1 + T_2]$ beeinflussen, wie es etwa bei einen Ansteckungsprozess der Fall wäre.

Für den Erwartungswert und Varianz der Schadensanzahl ergibt sich

$$\begin{array}{rcl} \mathbb{E}N(T) & = & T \cdot p; \\ \mathbf{Var}N(T) & = & T \cdot p \cdot (1-p). \end{array}$$

Dieses Verteilungsmodell ist für kleine Schäden und homogene Bestände von Risiken geeignet und wird z.B. in der Lebensversicherung verwendet. Bei größeren Beständen arbeitet man in der Regel besser mit der Poisson Verteilung die sich formal aus der Binominalverteilung unter der Voraussetzung $\lambda = n \cdot p = Konstant$ für $T \to \infty$ ergibt.

Beispiel - ein Betriebes mit Maschinenpark Ein Betrieb setzt in seiner Produktion eine Maschine ein, die in einem einzelnen Monat mit Wahrscheinlichkeit $p = 0.1$ einen Schaden hat und repariert werden muss. Zur Vereinfachung sei angenommen, dass pro Monat lediglich ein Schaden auftreten kann und dass das Auftreten eines Schadens in den einzelnen Zeitperioden unabhängig voneinander ist.

Berechnet man die zu erwartende Schadenanzahl (Reparaturunfälle an der Maschine) nach sechs Monaten, so ergibt sich eine zu erwartende Schadensanzahl von $\mathbb{E}S_N = 6 \cdot 0.1 = 0.6$ bei einer Varianz von $\mathbf{Var}(S_N) = 6 \cdot 0.1 \cdot 0.9 = 0.54$ bzw. einer Standardabweichung von $\mathbf{Std}(S_N) \sim 0.7348$. Um das Modell hinsichtlich der Annahme, dass pro Zeitperiode nur ein Schaden auftreten kann, realistischer zu gestalten, könnte man als Zeitperiode beispielsweise auch nur einen Tag betrachten und die Wahrscheinlichkeit p_{Tag} des Ausfalls entsprechend kleiner wählen, i.e. $p_{Tag} = p/30$. Dies führt allerdings zu teilweise recht kleinen Zahlenwerten, mit insbesondere unzureichender grafischer Darstellungsmöglichkeit. Allerdings, ergibt sich im Grenzwert die Poisson Verteilung.

Beispiel - Unternehmensvorsorge Ein Unternehmen hat im Rahmen seiner betrieblichen Sozialleistungen jedem Angestellten eine Hinterbliebenenvorsorge zugesagt. Im Todesfall soll jeweils eine Summe von $e30.000$ an die Angehörigen ausgezahlt werden. Es gibt 400 Angestellte, deren einjährige Sterbewahrscheinlichkeit (d.h. die Wahrscheinlichkeit, dass die betreffende Person innerhalb des nächsten Jahres stirbt) zur Vereinfachung unabhängig von Alter, Geschlecht und aktuellem Gesundheitszustand mit $q = 0.005$ angesetzt wird. (Es ist in der Versicherungsmathematik üblich, Sterbewahrscheinlichkeiten mit q, und die Überlebenswahrscheinlichkeiten mit $p = 1 - q$ zu bezeichnen.) Aus diesen Informationen lassen sich Aussagen über die Wahrscheinlichkeitsverteilung (hier: Bernoulli-Verteilung unter der Voraussetzung der Unabhängigkeit der einzelnen Sterbefälle) ableiten; eine Modellierung als stochastischer Prozess ist nicht notwendig. Man berechnet beispielsweise, dass der Erwartungswert der Sterbefälle innerhalb eines Jahres $\mathbb{E}(\text{Sterbefälle}) = 400 \cdot 0.005 = 2$ beträgt. Pro Jahr ist also unter diesen Annahmen durchschnittlich mit zwei Sterbefällen zu rechnen, d.h. durchschnittlich mit einer Leistung des Unternehmens von $e60.000$ bzw. von $e150$ umgerechnet pro Kopf der Belegschaft. Aber es ist klar, dass es auch ganz anders kommen kann. Deshalb wird das Unternehmen sich möglicherweise durch den Abschluss einer Versicherung bei einem Versicherungsunternehmen absichern. Die errechnete durchschnittliche Leistung von $e60.000$ für die Gesamtbelegschaft bzw. 150 pro Person wäre dann ein Anhaltspunkt für die Prämie, die dem Versicherungsunternehmen für die Übernahme des Versicherungsschutzes für ein Jahr zu zahlen ist, wobei klar ist. dass die Versicherung auch noch Beratungs- und Verwaltungskoslen sowie einen Sicherheitszuschlag einkalkulieren muss. Die Notwendigkeit eines Sicherheitszuschlags ergibt sich zum einen aus unvorhergesehenen prinzipiellen Änderungen in der (Bevölkerungs-)Sterblichkeit, d.h. einer Veränderung des Parameters q (im Rahmen dieses Beispiels nicht thematisiert) und zum anderen aus den modellgemäßen Zufallsschwankungen in der Sterblichkeit. Ein Risikomaß für die Zufallsschwankungen in dem Beispiel wäre etwa die Varianz $\mathbf{Var}(\text{Sterbefälle}) = 400 \cdot 0.005 \cdot 0.995 = 1.99$ bzw. die Standardabweichung $\mathbf{Std}(\text{Sterbefälle}) \sim 1.41$, was grob gesprochen bedeutet, dass eine Abweichung von 1.41 (bezogen auf die Grundgesamtheit von 400), d.h. von etwas mehr als einem Todesfall mehr oder weniger, *nicht ungewöhnlich ist*. Risikokennzahlen werden später noch eingehender behandelt. Eine andere Möglichkeit des Unternehmens, sich auf seine finanziellen Verpflichtungen vorzubereiten, bestünde in der Bildung von entsprechenden Bilanzrückstellungen. Für deren notwendige Höhe kann der Erwartungswert ebenfalls ein Anhaltspunkt sein, wobei an dieser Stelle besonders deutlich wird, dass das Unternehmen auch mit Abweichungen rechnen (im wahrsten Sinne des Wortes) sollte. Das gerade erläuterte Beispiel ist insofern realitätsnah, als dass in der Tat in vielen Unternehmen Angestellte neben ihrem Gehalt auch betriebliche Sozialleistungen wie etwa Zusagen für eine spätere Betriebsrente oder für Leistungen an Hinterbliebene erhalten. Und zur Absicherung der damit eingegangenen finanziellen Verpflichtungen können Unternehmen tatsächlich Versicherungen abschließen oder auch selbst Bilanzrückstellungen bilden. Zu Einzelheiten

in diesem Zusammenhang gibt es allerdings umfangreiche gesetzliche Regelungen; zudem ist eine konkrete Prämien- bzw. Rückstellungsberechnung in der Realität nicht so einfach durchzuführen wie unter den hier nur zur Illustration der Bernoulli-Verteilung herangezogenen stark vereinfachten Annahmen.

Beispiel Zahlungsausfälle Bei der Vergabe von Krediten spielt die Einschätzung der Zahlungsfähigkeit der potentiellen Schuldner eine große Rolle. Sie erfolgt unmittelbar durch den Kreditgeber oder auch durch spezielle Rating-Agenturen. Insbesondere bei der Beurteilung der Kreditwürdigkeit von Unternehmen durch Rating-Agenturen ist die *Benotung* mit Hilfe von Rating-Klassen (bezeichnet z. B. mit AAA, AA, A, BBB, BB, B usw.) üblich. Einer vorgegebenen Rating-Klasse, beispielsweise B. wird eine bestimmte einjährige Ausfall Wahrscheinlichkeit w, beispielsweise w = 0.01 zugeordnet, die z. B. aus historischen Daten vergleichbarer Unternehmen geschätzt wird. Dieses Zahlenbeispiel ist fiktiv. Über Einzelheiten der Rating-Systeme von bekannten Rating-Agenturen wie Standard & Poor's oder Moody's kann man sich etwa auf deren Internet-Seiten informieren. Die einjährige Ausfallwahrscheinlichkeit w gibt die Wahrscheinlichkeit an, dass ein Schuldner aus der fraglichen Rating-Klasse innerhalb eines Jahres *ausfällt*, also seinen Zahlungsverpflichtungen nicht mehr genügen kann. Die Größe w kann als gewisses Analogon zur einjährigen Sterbewahrscheinlichkeit von vorher interpretiert werden. Hat also beispielsweise eine Bank T Kreditnehmer mit identischer Ausfallwahrscheinlichkeit w, so könnte sie in einem einfachen Modell die Anzahl N der Zahlungsausfälle innerhalb des nächsten Jahres in diesem Kreditportfolio als binomialverteilt $N \sim \mathbf{Bin}(T; w)$ annehmen. Es ist allerdings zu beachten, dass in diesem Beispiel die notwendige Annahme der Unabhängigkeit der Risiken oft unrealistisch sein dürfte.

Negative Binomialverteilung

Häufig wird als Verteilungsmodell für die Schadenanzahl bis zu einem festen Zeitpunkt t bzw. bezogen auf ein bestimmtes Zeitintervall auch die negative Binomialverteilung verwendet. Für $0 < p < 1$ und $n \in \mathbb{N}_0$ ist

$$\mathbb{P}(N = n) =: p_n = \binom{R+n-1}{n} \cdot p^r \cdot (1-p)^n.$$

Erwartungswert und Varianz sind dabei gegeben durch

$$\mathbb{E}(N) = \frac{r \cdot (1-p)}{p},$$

und

$$\mathbf{Var}(N) = \frac{r \cdot (1-p)}{p^2}.$$

Angenommen N_i beschreibt die Schadenszahl im Zeitintervall $[0, r_i]$ und es gilt

$$N_i \sim \mathbf{NB}(r_i; p),\ i = 1, 2.$$

Sind N_1 und N_2 unabhängig, so gilt

$$N_1 + N_2 \sim \mathbf{NB}(r_1 + r_2; p).$$

Poisson Verteilung

Die Binomialverteilung eignet sich zur Modellierung der Schadenzahlverteilung für kleine, homogene Bestände. Dagegen ist sie für große Bestände ungeeignet, da die Varianz dann sehr klein ausfällt. Desweiteren ist diese Verteilung wenig anpassungsfähig, da nur der Parameter p zur Modellierung dienen kann. Die Binomialverteilung kann für kleine Werte von p sehr gut durch die Poissonverteilung approximiert werden. Doch gerade kleine Werte von p, also Schäden, die nur mit einer sehr geringen Wahrscheinlichkeit auftreten, treten in der Versicherungsbranche häufig auf. Wir betrachteten im Folgenden die Poissonverteilung als Modellierung der Schadenzahlverteilung. Dieses Modell ist die am häufigste verwendete Verteilung der Schadenszahl im kollektiven Modell.

Definition 1.5. *Eine Zufallsvariable N auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ heißt poissonverteilt zum Parameter (Intensität) $\lambda > 0$, falls gilt:*

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!} \exp(-\lambda), \quad k \in \mathbb{N}_0.$$

Als Notation für die Poissonverteilung mit Parameter $\lambda > 0$ benutzen wir $\mathbf{Poi}(\lambda)$.

Sei N Poisson verteilt mit Parameter λ. Der Erwartungswert ist gleich der Varianz und ist gleich dem Parameter, i.e.

$$\mathbb{E}N = \mathbf{Var}N = \lambda.$$

Eine für die Schadensmodellierung wichtige Eigenschaft ist folgende. Seine $N_1 \sim \mathbf{Poi}(\lambda_1)$ und $N_2 \sim \mathbf{Poi}(\lambda_2)$. Dann gilt

$$N_1 + N_2 \sim \mathbf{Poi}(\lambda_1 + \lambda_2).$$

Beispiel - ein Betriebes mit Maschinenpark: Wie im Beispiel *ein Betriebes mit Maschinenpark* schon erwähnt, kann man die betrachteten Zeitperioden von einen Monat auf einen Tag kürzen, da die Annahme, dass in einem Monat nur eine Maschine ausfallen kann realitätsfern ist. In diesem Fall gilt $p_{Tag} = 0.1/31$ und sechs Monate haben 183 Tage. Unter der Annahme dass N Bernoulli verteilt ist, weiß mann, dass gilt

$$\mathbb{P}(N(183) = m) = p_n(T) = \binom{183}{m} \cdot p_{Tag}^m \cdot (1 - p_{Tag})^{183-m}.$$

Für den Erwartungswert und Varianz der Schadensanzahl ergibt sich

$$\mathbb{E}N(183) = 183 \cdot p_{Tag} = 0.6 \quad \text{und} \quad \mathbf{Var}N(183) = 183 \cdot p_{Tag} \cdot (1 - p_{Tag}) \sim 0.5980.$$

Nun hat man die Möglichkeit die einzelnen Stunden zu betrachten. Eine kurze Rechnung ergibt dass 6 Monate $183 \cdot 24$ Stunden sind, und $p_{stunde} = p_{Tag}/24$. Also

$$\mathbb{E}N(183 * 24) = 183 \cdot 24 \cdot p_{Stunde} = 0.6 \quad \text{und} \quad \mathbf{Var}N(183 * 24) = 183 \cdot 24 \cdot p_{stunde} \cdot (1 - p_{stunde}) \sim 0.5980.$$

Geht man auf immer kleiner Zeiteinheiten über, erhält man als Grenzverteilung die Poisson Verteilung mit Parameter $\lambda = 0.6$.

Nimmt man vom Anfang an die Poisson Verteilung, hat man den Vorteil, dass sich bestimmte Wahrscheinlichkeiten leichter ausrechnen lasse. Z.B. ist die Wahrscheinlichkeit dass mehr als 6 Schadensfälle in einem halben Jahr vorkommen

$$\mathbb{P}(N > 6) = 1 - \mathbb{P}(N \leq 6) = 1 - \lambda^0 e^{-\lambda} + \lambda^1 e^{-\lambda} + \cdots + \frac{\lambda^6}{6!} e^{-\lambda} \sim 3.29 \cdot 10^{-6}.$$

Beispiel - ein Betriebes mit Maschinenpark - mehrere Ausfallquoten: Wir haben wieder die Situation vom vorigen Beispiel. Angenommen die Wahrscheinlichkeit innerhalb der nächsten Zeiteinheit auszufallen sind verschieden. Genau genommen gibt es drei Gruppen von Maschinen, die einen sind neu und haben Ausfallwahrscheinlichkeit λ_{neu}, andere sind alt und haben eine Ausfallwahrscheinlichkeit λ_{alt}, der dritte Teil der Maschinen laufen schon einige Zeit und haben eine Ausfallwahrscheinlichkeit λ_0. Wir haben in jeder Teilgruppe gleich viele Maschinen. Es ergibt sich, dass N wieder Poisson verteilt ist mit Parameter $\lambda = \lambda_{alt} + \lambda_{neu} + \lambda_0$.

S 55, 56.

Panjer Verteilung

Eine weitere noch öfters beützte zur Modellierung der Schadenszahl N ist die Panjer Verteilung $N \sim \mathbf{Panjer}(a,b)$ mit $a+b \geq 0$. Die Wahrscheinlichkeitsverteilung genügt folgender Rekursion

$$\mathbb{P}(N=0) = p_1, \quad P(N=n) = p_n = p_{n-1} \cdot \left(a + \frac{b}{n}\right), \quad n \geq 1.$$

wobei a und b vorgegeben sind und den Startwert p_0 durch die Bedingung

$$\sum_{n=0}^{\infty} p_n = 1,$$

festlegen. Für den Erwartungswert und der Varianz des Panjer prozess gilt

$$\mathbb{E}(N) = \frac{a+b}{1-a}, \quad \mathbf{Var} = \frac{a+b}{(1-a)^2}.$$

Die Binomial, negative Binomial und Poisson Verteilung sind Spezialfälle der Panjer Verteilung. Genauer gilt $N \sin \mathbf{Bin}(t,p)$ mit

$$a = \frac{-p}{1-p}, \quad b = \frac{(t+1)p}{1-p}, \quad p_0 = (1-p)^t,$$

$N \sin \mathbf{NB}(r,p)$ falls

$$a = 1-p, \quad b = (r-1)(1-p), \quad p_0 = p^r,$$

und $N \sin \mathbf{Poi}(\lambda)$ falls

$$a = 0, \quad b = \lambda, \quad p_0 = e^{-\lambda}.$$

iii Modellierung der Schadensanzahl durch einen Schadensanzahlprozess

In den Beispielen vorher, war die Schadenszahl in einen festen Zeitintervall betrachtet. Oft ist man an der Fragestellung interessiert, wie die Schadenssumme zeitlich abhängt. Dies kann man auf die Fragestellung zurückführen, wie die Schadensanzahl zeitlich wächst.

Damit betrachtet man die Schadensfall als eine Funktion der Zeit. Da die Schadensanzahl auch von Ω abhängt, also eine ZV ist, ist, mathematisch gesehen, ist N ein stochastische Prozess

$$N : [0,\infty) \times \Omega \to \mathbb{N},$$

sodass gilt

$$N(t) = \#\{ \quad \text{Schäden, die im Zeitintervall } [0,t] \text{ aufgetreten sind} \quad \}.$$

Dabei kann man N mit verschiedenen Prozessen modellieren. Oft ist es besser nicht die Zeiten des Schadensanzahl zu modellieren sondern die Wartezeiten zwischen den Schadensmeldungen zu beschreiben.

Definition 1.6. *Es seien $\{\tau_i : i \in \mathbb{N}\}$ positive Zufallsvariablen und*

$$N(t) := \sum_{i=1}^{\infty} 1_{[T_i,\infty)}(t), \qquad \textit{wobei} \qquad T_k = \tau_1 + \cdots + \tau_k.$$

Dann heißt $(N(t) : t > 0)$ Schadenzahlprozess und die Zufallsvariablen $\{\tau_i : i \in \mathbb{N}\}$ Wartezeiten.

Bemerkung 1.2. *Der Prozess $(N(t) : t > 0)$ heißt auch Sprung- oder Zählprozess.*

Homogener Poisson Prozess

Bei Modellierung des Gesamtschadens in einem festen Zeitraum haben wir die Schadenzahl unter anderem durch eine poissonverteilte Zufallsvariable modelliert. In Analogie betrachten wir jetzt mit dem zusätzlichen zeitlichen Aspekt einen sogenannten Poissonprozess.

Dieser Poisson Prozess wird dadurch charakterisiert, dass die Wartezeiten zwischen den einzelnen Verlusten unabhängig und exponential verteilt sind mit einen bestimmten vorgegebenen Parameter.

Definition 1.7. *Es seien $\{\tau_i : i \in \mathbb{N}\}$ unabhängige Zufallsvariablen, die identisch exponentialverteilt mit Parameter $\lambda > 0$ sind. Definiert man*

$$N(t) := \sum_{i=1}^{\infty} 1_{[T_i,\infty)}(t), \qquad \text{wobei} \quad T_k = \tau_1 + \cdots + \tau_k.$$

Dann heißt $(N(t) : t > 0)$ (homogener) Poissonprozess mit Intensität λ.

Es gibt zahlreiche äquivalente Charakterisierung von Poisson Prozessen, von denen wir einige in dem folgenden Resultat aufführen.

Satz 1.1. *Es sei $N = (N(t) : t > 0)$ ein Zählprozess. Dann sind äquivalent:*

- *der Prozess N ist ein Poisson Prozess mit Intensität $\lambda > 0$;*
- *der Prozess N besitzt folgende Eigenschaften:*
  - *$N(t)$ ist für jedes $t > 0$ poissonverteilt mit Parameter $\lambda t > 0$;*
  - unabhängige Zuwächse: *für alle $0 \le t_0 < t_1 < \ldots < t_n$ und $n \in \mathbb{N}$ sind die Zuwächse*

    $$N(t_1) - N(t_0); N(t_2) - N(t_1); \cdots ; N(t_n) - N(t_{n-1})$$

    unabhängig;
  - stationäre Zuwächse: *für alle $0 \le t_0 < t_1 < \ldots < t_n$ und $n \in \mathbb{N}$ und für alle $h > 0$ hängen die Verteilungen von*

    $$N(t_1 + h) - N(t_0 + h); \cdots ; N(t_n + h) - N(t_{n-1} + h)$$

    nicht von h ab.
- *Der Prozess N besitzt unabhängige, stationäre Zuwächse und es gilt für alle $t > 0$:*
  - *$P(N(t + h) - N(t) = 1) = \lambda h + o(h)$ für $h \to 0$;*
  - *$P(N(t + h) - N(t) > 1) = o(h)$ für $h \to 0$.*

Bemerkung 1.3. *Besitzt $N = (N(t) : t > 0)$ unabhängige Zuwächse und ist für alle $t > 0$ die Verteilung von*

$$X(t + h) - X(t)$$

unabhängig von h, so besitzt $X(N)$ auch stationäre Zuwächse.

Beispiele

iv Die Gesamtschadenszahl

Bei der Behandlung des individuellen Modells steht die Modellierung des Erwartungswertes und der Varianz der Gesamtschadenverteilung im Vordergrund. Dazu setzten wir, bis auf verschiedene Versicherungssummen, ein homogenes Portfolio voraus. Jedoch reicht die Kenntnis bzw. Schätzung von Erwartungswert und Varianz nicht aus, um die Verteilung des Gesamtschadens ausreichend beschreiben zu können, z.B. zur Tarifkalkulation oder Berechnung von Rücklagen. Hier werden auch andere Größen, wie Quantilen oder Risikokennzahlen wichtig.

Die Schadensanzahl N gibt zu jedem Zeitpunkt die Anzahl der eingetretenen Schadensfälle eines bestimmten Risikos an.

Definition 1.8. *Der Gesamtschaden eines Portfolios $\{X_k : k \in \mathbb{N}\}$ (im kollektiven Modell) mit Schadenzahl S ist die Zufallsvariable*

$$S_N := \begin{cases} 0, & \text{falls } N = 0, \\ \sum_{k=1}^{N} X_i, & \text{falls } N > 0. \end{cases}$$

Bei der Behandlung des kollektiven Modells gehen wir wie teilweise zuvor von den folgenden Annahmen aus:

- Endlichkeit der Streuung $\mathbf{Var}[X_k]$; (Ist aber nicht immer gegeben !)
- alle Zufallsvariablen $N, X_1, X_2; \cdots$, sind auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert;
- $X_k > 0$ für alle $k \in \mathbb{N}$.

Im weiteren werden wir meistens auch folgende Voraussetzungen annehmen:

- $N, X_1, X_2, \ldots$ sind unabhängig; Die Unabhängigkeit der Risiken ist nicht immer gegeben, aber es ist eine Annahme die viele Rechnungen vereinfacht. Die Unabhängigkeit der Schadenzahl und der Schadenhöhen kann als realistisch betrachtet werden, aber auch hier kann eine genauere Betrachtung notwendig sein, z.B. Autohaftpflichtversicherung in einem Winter mit besonders viel vereisten Fahrbahnen: viele, jedoch kleine Schäden.
- $X_1, X_2, \ldots$ sind identisch verteilt; Zunächst scheint dies unserer Motivation für das kollektive Modell, der Vermeidung der Annahme von homogenen Portfolios, zu widersprechen. Jedoch werden im kollektiven Modell die Schadenhöhen nicht bestimmten Risiken zugeordnet, sondern es wird die Gesamtheit aller Schäden betrachtet. Deshalb kann sehr wohl eine identische Verteilung der Risiken angenommen werden, wenn man sich die Realisierung dieser Verteilung als ein zweistufiges Experiment vorstellt: zunächst wird zufällig eine bestimmte Verteilung (aus einer Klasse von ähnlichen, jedoch verschiedenen Verteilungen) bestimmt, und dann wird eine Realisation dieser zufällig bestimmten Klasse ausgewählt; Dies werden wir hier aber nicht abhandeln.

Zufallsvariablen der obigen Form werden zusammengesetzte Summenvariablen (compound random variable) genannt und entsprechend ihre Verteilung zusammengesetzte Summenverteilung. In den meisten Fällen ist es nicht möglich, die zusammengesetzte Summenverteilung in einer expliziten Form zu bestimmen. Jedoch kann man deren Laplace- Transformierte mittels den Transformierten der Schadenzahl und der Risiken angeben:

Satz 1.2. *Es seien $\{X_k : k \in \mathbb{N}\}$ ein Portfolio unabhängiger, identisch verteilter Risiken und N eine Schadenszahl.*

- *Falls die jeweiligen Momente existieren, dann gilt:*

$$\mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[X_1];$$

und

$$\mathbf{Var}[S_N] = (\mathbf{Var}[N])(\mathbb{E}[X_1])^2 + (\mathbb{E}[N])(\mathbf{Var}[X_1]).$$

- *Bezeichnet $\mathcal{G}(N)$ die erzeugende Funktion von N und $\mathcal{L}(S_N)$ und $\mathcal{L}(X_1)$ die jeweiligen Laplace-Transformierten von S_N und X_1, dann gilt:*

$$\mathcal{L}(S_N)(u) = \mathcal{G}(N)(\mathcal{L}(X_1)(u))$$

für alle $u > 0$:

Proof. Dieser Satz ist sehr einfach zu zeigen. Das erste Moment ist durch folgende Rechnung zu zeigen:

$$\mathbb{E}[S_N] = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\,\mathbb{E}[X_1+\cdots+X_k] = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\,k\,\mathbb{E}[X_1] = \mathbb{E}N \cdot \mathbb{E}[X_1].$$

Analog kann man das zweite Moment berechnen (Übungsaufgabe). Die Laplacetransofmierte kann man anhand folgender Rechnung berechnen.

$$\begin{aligned} L(S_N)(u) &= \mathbb{E}\left[e^{-uS_N}\right] = \sum_{k=0}^{\infty} \mathbb{P}(N=k) \cdot \mathbb{E}\left[e^{-u(X_1+\cdots+X_k)}\right] = \sum_{k=0}^{\infty} \mathbb{P}(N=k) \cdot \left(\mathbb{E}\left[e^{-uX_1}\right]^k\right) \\ &= \mathcal{G}(N)\left(\mathcal{L}(X_1)(u)\right). \end{aligned}$$

☺

□

Beispiel 1.3. *Angenommen die Schadensanzahl $N \sim \mathbf{NB}(1, 1-p)$ ist negative Binomial verteilt mit Parameter 1 und $1-p$ (also geometrisch verteilt, also $\mathbb{P}(N=k) = (1-p)\cdot p^k$) und das Portfolio ist exponentiell verteilt mit Parameter θ. Dann kann man zeigen, dass (Erlang Verteilung)*

$$\mathbb{P}(X_1 + X_2 + \ldots X_k = x) = \theta^{k-1} \cdot \frac{x^{k-1}}{(k-1)!} \cdot e^{-\theta x}.$$

Eingesetzt in die Summenformel, gilt

$$\begin{aligned} \mathbb{P}(S_N = x) &= \mathbb{P}(N=0)\cdot\delta_0(x) + \sum_{k=1}^{\infty} \mathbb{P}(N=k)\cdot\mathbb{P}(X_1+X_2+\ldots X_k = x) \\ &= (1-p)\delta_0(x) + \sum_{k=1}^{\infty}(1-p)\cdot p^{k-1}\cdot\theta^{k-1}\cdot\frac{x^{k-1}}{(k-1)!}\cdot e^{-\theta x} \\ &= (1-p)\delta_0(x) + (1-p)\cdot e^{-\theta x}\sum_{k=0}^{\infty}\frac{p^k\cdot\theta^k\cdot x^k}{k!} \\ &= (1-p)\delta_0(x) + (1-p)\cdot e^{-\theta x + p\theta x} = (1-p)\delta_0(x) + (1-p)\cdot e^{-\theta x(1-p)} \end{aligned}$$

☺

Ist die Schadenszahl exponentiell verteilt mit Parameter $\lambda > 0$ so gilt sogar folgendes:

$$\mathbf{Var}(S_N) = \mathbb{E}N \cdot \mathbb{E}X_1^2$$

und

$$\mathbb{E}[S_N - \mathbb{E}S_N]^3 = \mathbb{E}N \cdot \mathbb{E}X_1^3.$$

Als Verteilung des Gesamtschadens S eines Portfolios wo die Schadenszahl N Poisson verteilt mit Parameter $\lambda > 0$ erhält man

$$F_S = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\exp(-\lambda)F_{X_1}^{k\star}.$$

Die Verteilung von S heißt dann Poisson-Summenverteilung (compound Poisson distribution).

Zwar ist die Poissonverteilung auch wenig anpassungsfähig, da sie nur von einem Parameter abhängt, jedoch ist sie leicht handzuhaben, da viele Rechnungen explizit ausgeführt werden können. Ein wesentlicher Vorteil der Poissonverteilung ist die folgende Möglichkeit der Aufteilung eines inhomogenen Portfolios in mehrere homogene Portfolios. In vielen Situationen kann ein inhomogenes Portfolio aufgeteilt werden in m verschiedene Portfolios, die jeweils aus homogenen Risiken und einer poissonverteilten Schadenzahl bestehen, z.B. in der PKW-Haftpflichtversicherung erfahrene und unerfahrene Fahrer. Den verschiedenen Portfolios können unterschiedliche Risikoverteilungen Q_l und unterschiedliche Schadenintensitäten λ_l der Schadenzahl N_l für $l = 1, 2, \cdots, m$ zugrunde liegen. Jeder Gesamtschaden S_l der verschiedenen Portfolios besitzt dann die Verteilung:

$$F_{S_l} = \frac{\lambda_l^k}{k!} \exp(-\lambda_l) Q_l^{k\star}.$$

Das Versicherungsunternehmen ist aber interessiert an der Summe der verschiedenen Gesamtschäden.

Satz 1.3. *Es seien $S_1, S_2, \ldots S_m$ unabhängige Zufallsvariablen mit den Verteilungen*

$$F_{S_l} = \frac{\lambda_l^k}{k!} \exp(-\lambda_l) Q_l^{k\star}.$$

wobei $\lambda_l > 0$ und Q_l Verteilungsfunktionen sind. Dann gilt für $S := S_1 + S_2 \ldots + + S_m$

$$S \sim \sum_{k=1}^{N} Y_k$$

wobei N eine poissonverteilte Zufallsvariable zu dem Parameter $\lambda = \lambda_1 + \cdots + \lambda_m$ ist und $\{Y_j : j \in \mathbb{N}\}$ unabhängige, identisch verteilte Zufallsvariablen sind mit der Verteilung:

$$Q = \frac{1}{\lambda} \sum_{k=1}^{m} \lambda_k Q_k$$

ist.

Satz 1.3 besagt, dass ein Portfolio unabhängiger, identisch verteilter Risiken und einer poissonverteilten Zufallsvariablen bezüglich den Verteilungen als die Zusammenfassung mehrerer Portfolios mit unterschiedlich verteilten Risiken und Schadenzahlen aufgefasst werden kann. Betrachtet man ein Portfolio über mehrere Jahre hinweg, so sind oft trendartige und oszillatorische Veränderungen der Schadenzahl zu beobachten. Trendartige Veränderungen sind z.B. verbesserte Schadenverhütungsmaßnahmen wie Einbau von Sprinkelanlagen. Oszillatorische Veränderungen sind Schwankungen in der mittleren Schadenzahl, wie z.B. regenarme Sommer führen zu einer Zunahme von Bränden. Die oszillatorischen Veränderungen können modelliert werden, indem man den Parameter einer poissonverteilten Schadenzahl als einen zufällig gewählten Wert gemäß einer spezifizierten Verteilung betrachtet. Diese Verteilung modelliert die oszillatorischen Änderungen. Diese verbale Beschreibung resultiert in dem mathematischen Begriff der Poissonmischung:

Definition 1.9. *Es sei μ eine Verteilung auf $(\mathbb{R}+; \mathcal{B}(\mathbb{R}+))$. Dann wird durch*

$$Q(\{k\}) = \int_{\mathbb{R}_+} \frac{\delta^k}{k!} e^{-\delta} \nu(d\delta)$$

ein Wahrscheinlichkeitsmaß Q auf $(\mathbb{N}_0; \mathcal{P}(\mathbb{N}_0))$ definiert. Das Maß Q heißt Poissonmischung bezüglich des Mischungsmaßes.

Approximationen von der Gesamtschadenszahl S Die genaue Verteilung lässt sich nur in wenigen Fällen ausrichten. Zumeist ist man auf eine Monte Carlo Simulation des Gesamtschadens angewiesen (darauf werden wir später nochmals zurückkommen). Hier wollen wir kurz anführen, wie man die Dichteverteilungen aber trotzdem approximieren können. Eine Methode ist die Normalapproximation, die aber von keinen praktischen Wert ist. Der Grund, mittels der Normalverteilung kann man Zufallsvariablen zumeist nur in ihren mittleren Bereich gut approximieren, in den Randbereich wird die Qualität der Approximation immer schlechter. In der Risikoanalyse interessiert uns aber der Randbereich wo x große Werte annimmt, also genau der Bereich der nicht gut approximiert wird.

Rekursive Berechnung der Summenverteilung von Panjer Angenommen die Schadenszahl genügt einer Panjer verteilung, d.h. $\mathbb{P}(N = n) = p_n$, wobei $\{p_n : n \in \mathbb{N}\}$ folgender Rekursion genügt:

$$p_{n+1} = \left(a + \frac{b}{n+1}\right) p_n, \quad n \in \mathbb{N}.$$

Weiters nehmen wir an, dass die Verteilungsfunktion Q von dem Portfolio $\{X_k : k \in \mathbb{N}\}$ diskret ist, mit

$$Q(\{h, 2h, 3h, \ldots\}) = 1.$$

Sei P die Verteilungsfunktion von S_N. Dann genügt P folgender Panjer–Rekursion

$$P(\{0\}) = p_0,$$

und

$$P(\{(k+1)h\}) = \sum_{j=1}^{k+1} \left(a + \frac{b \cdot j}{k+1}\right) Q(\{jh\}) P(\{(k+1-j)h\}).$$

Ist N Poisson verteilt mit Parameter λ, dann ist $a = 0$ und $b = \lambda$ und wir erhalten

$$P(\{(k+1)h\}) = \frac{\lambda}{k+1} \sum_{j=1}^{k+1} jQ(\{jh\}) P(\{(k+1-j)h\}).$$

Summenverteilung von Großschäden Hat man ein Portfolio $\{X_i : i \in \mathbb{N}\}$ bestehend aus Großschäden, so dominiert das einzelne Ereignis über den Rest.

Satz 1.4. *Sei das Portfolio $\{X_i : i \in \mathbb{N}\}$ gegeben mit Verteilung Q, wo Q subexponentiell ist, und hat N eine Schadensanzahlverteilung, sodass es eine positive Zahl $\epsilon > 0$ gibt mit*

$$\sum_{k=0}^{\infty} \mathbb{P}(N = k) \cdot (1+\epsilon)^k < \infty,$$

dann gilt für die Summenverteilung $\mathcal{P}$ von S_N die Beziehung

$$\mathcal{P}([t, \infty)) \sim \mathbb{E}N \cdot Q([t, \infty)).$$

Somit ist $\mathbb{E}NQ([t, \infty))$ eine Approximation der Tailwahrscheinlichkeit von $\mathcal{P}$, deren relativer Fehler für $t \to \infty$ gegen Null konvergiert. Den Beweis bringen wir hier nicht, er ist im Skriptum von Prof Hipp (Risikotheorie I) zu finden.

Beispiel 1.4. *Sei Q die Pareto Verteilung mit Parametern α und 1 und N sei Poisson verteilt mit Parameter λ. Eine kurze Rechnung ergibt*

$$\sum_{k=0}^{\infty} \mathbb{P}(N = k) \cdot (1+\epsilon)^k = \sum_{k=0}^{\infty} \frac{(\lambda \cdot (1+\epsilon))^k}{k!} e^{-\lambda} = e^{(1+\epsilon)\lambda - \lambda} < \infty,$$

und damit gilt für die Gesamtschadensverteilung $\mathcal{P}$ die Beziehung

$$\mathcal{P}(t, \infty) \sim \lambda t^{-\alpha}.$$

v Risikoprozesse oder Geamtschadenszahlprozesse

In den vorangegangenen Abschnitten modellierten wir den Gesamtschaden eines Portfolios in einer bestimmten Zeitperiode, z.B. einem Jahr. In diesem Abschnitt interessieren wir uns für die Wahrscheinlichkeit, dass der Gesamtschaden eines Versicherungsunternehmens die Einnahmen, z.B. monatlich gezahlte Prämien, übertrifft. Da diese Ruinsituation nicht nur am Ende einer Beobachtungsperiode geschehen kann, kann man wie vorher noch eine zeitliche Komponente t einführen, um die Anzahl der Schäden bis zum Zeitpunkt $t > 0$ modellieren zu können.

Bezieht man wie vorher die Zeit mit ein und modelliert die Schadensanzahl durch einen Zählprozess so kommt man zu einem Schadensprozess:

Definition 1.10. *Ist $\{X_k : k \in \mathbb{N}\}$ ein Portfolio und gibt $N(t)$ die Schadenzahl in diesem Portfolio zur Zeit t an, so wird der Gesamtschaden modelliert durch:*

$$S(t) := \begin{cases} 0, & \text{falls } N(t) = 0, \\ \sum_{k=1}^{N(t)} X_i, & \text{falls } N(t) > 0. \end{cases}$$

Wie zuvor nehmen wir an:

- Endlichkeit der Streuung $\mathbf{Var}[X_k]$; (Ist aber nicht immer gegeben !)
- alle Zufallsvariablen $X_1, X_2; \cdots$, und der stochastische Prozess $N = N(t) : 0 \leq t < \infty\}$ sind auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert;
- $X_k > 0$ für alle $k \in \mathbb{N}$;
- $N, X_1, X_2, \ldots$ sind unabhängig;
- $X_1, X_2, \ldots$ sind identisch verteilt;

Für festes $t > 0$ übertragen sich die Aussagen aus Kapitel 2, z.B. Satz 1.2 lautet:

Satz 1.5. *Es seien $\{X_k : k \in \mathbb{N}\}$ ein Portfolio unabhängiger, identisch verteilter Risiken und $N = \{N(t) : t \geq 0\}$ eine Schadensprozess.*

- *Falls die jeweiligen Momente existieren, dann gilt:*

$$\mathbb{E}[S(t)] = \mathbb{E}[N(t)]\mathbb{E}[X_1];$$

und

$$\mathbf{Var}[S(t)] = (\mathbf{Var}[N(t)])(\mathbb{E}[X_1])^2 + (\mathbb{E}[N(t)])(\mathbf{Var}[X_1]).$$

- *Bezeichnet $\mathcal{G}(N(t))$ die erzeugende Funktion von $N(t)$ and der Stelle $t \geq 0$ und $\mathcal{L}(N(t))$ und $\mathcal{L}(X_1)$ die jeweiligen Laplace-Transformierten von X_1 und N, dann gilt:*

$$\mathcal{L}(S(t))(u) = \mathcal{G}(N(t))\,(\mathcal{L}(X_1)(u))$$

für alle $u > 0$.

Definition 1.11. *(***Cramer Lundberg Modell***) Es seien $\{X_k : k \in \mathbb{N}\}$ ein Portfolio unabhängiger, identisch verteilter Risiken und N Poissonprozess mit Intensität $\lambda > 0$. Wird der Gesamtschaden zur Zeit $t > 0$ modelliert durch*

$$S(t) := \begin{cases} 0, & \text{falls } N(t) = 0, \\ \sum_{k=1}^{N(t)} X_i, & \text{falls } N(t) > 0. \end{cases}$$

so heißt dieses Modell Cramer-Lundberg-Modell.

In der Praxis sind Aussagen über das dynamische Verhalten des Prozesses $(S(t) : t > 0)$ von Bedeutung. Die Verteilung von $S(t)$ für jedes $t > 0$ eindeutig bestimmt durch die Intensität λ des Poissonprozesses und der Verteilung von X_1. Die zeitlich abhängige Schadenzahl eines Portfolios kann durch zahlreiche andere Prozesse als ein Poissonprozess modelliert werden, jedoch beschränken wir uns in diesem Abschnitt auf diesen.

Als Verteilung des Gesamtschadens eines Portfolios wo N Poissonprozess mit Intensität $\lambda > 0$ erhält man

$$F_{S(t)} = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \exp(-\lambda t) F_{X_1}^{k\star}.$$

Die Verteilung von $S(T)$ heißt dann Poisson-Summenverteilung (compound Poissondistribution).

Ruinwahrscheinlichkeiten im Cramer-Lundberg-Modell

Die Bilanz eines Versicherungsunternehmens zum Zeitpunkt t setzt sich zusammen aus dem Gesamtschaden $S_{N(t)}$ als Verlust und den bezahlten Versicherungsprämien $p(t)$ als Einnahmen. Es addiert sich noch ein Startkapital $u > 0$ dazu, mit dem das Unternehmen zur Zeit $t = 0$ beginnt. Dies resultiert in der folgenden Definition:

Definition 1.12. *Es seien $\{X_k : k \in \mathbb{N}\}$ ein Portfolio und $N = (N(t) : t > 0)$ ein Schadenzahlprozess. Definiert man für eine Konstante $u > 0$ und eine monoton wachsende Funktion $p : \mathbb{R}^+ \to \mathbb{R}^+$ mit $p(0) = 0$*

$$R(t) := u + p(t) - S_{N(t)}; \quad t > 0;$$

so heißt $R = (R(t) : t > 0)$ Risikoprozess mit Anfangsrisikoreserve u. Falls $p(t) = ct$ für eine Konstante $c > 0$ gilt, dann heißt R klassischer Risikoprozess.

Der Ruin des Versicherungsunternehmens tritt ein, falls die Liquiditäten aufgebracht sind, das heißt, falls $R(t) < 0$ für ein $t > 0$. Natürlich ist das Versicherungsunternehmen interessiert daran, die Wahrscheinlichkeit für das Eintreten dieses Ereignisses möglichst gering zu halten. Um dies zu realisieren, kann das Versicherungsunternehmen entweder die Prämien $p(t)$ erhöhen oder mit einem höheren Startkapital beginnen. Den Einfluß der Prämien auf die Wahrscheinlichkeit des Ruins betrachten wir in Kapitel 4. In diesem Abschnitt sind wir interessiert an dem Verlauf des Prozesses R in Abhängigkeit von dem Startkapital u.

Definition 1.13. *Es sei $(R(t) : t > 0)$ ein Risikoprozess mit Anfangsrisikoreserve $u := R(0) > 0$. Dann heißt*

- $\tau(u) := \inf_{t>0}\{R(t) < 0\}$ *Ruinzeit (des Risikoprozesses R);*
- $\psi(u) = \mathbb{P}(\tau(u) < \infty)$ *Ruinwahrscheinlichkeit (des Risikoprozesses R);*
- $\phi(u) = 1 - \psi(u)$ *Uberlebenswahrscheinlichkeit (des Risikoprozesses R);*

Die (zufällige) Ruinzeit $\tau(u)$ muss keine endlichen Werte annehmen. Aus Ergebnissen in der Theorie von stochastischen Prozessen folgt, dass $\tau(u)$ eine Zufallsvariable ist. Man spricht auch von Ruinwahrscheinlichkeit in unendlicher Zeit (infinite-horizon ruin), da gilt:

$$\{\tau(u) < \infty\} = \inf_{t\geq 0}\{R(t) < 0\} = \cup_{t\geq 0}\{R(t) < 0\} = \{ \text{ es gibt ein } t > 0 \text{ mit } R(t) < 0.\}.$$

und das letzte Ereignis sich als das des Eintretens des Ruins in einer beliebig langen Zeitspanne interpretieren lässt.

Lemma 1.2. *Im Cramér-Lundberg-Modell mit Prämie β gilt für die Ruinwahrscheinlichkeit des klassischen Risikoprozesses für alle $u > 0$:*

$$\psi(u) = \mathbb{P}(\sup_{n\in\mathbb{N}} Y_n) \quad \textit{mit } Y_n = \sum_{k=1}^{n}(X_k - \beta\tau_k)$$

Lemma 1.2 erlaubt eine einfache Herleitung einer notwendigen Bedingung dafür, dass der Ruin nicht mit Wahrscheinlichkeit 1 eintritt. Denn nach dem starken Gesetz der großen Zahlen gilt:

$$\lim_{n\to\infty} \frac{1}{n} Y_n = \mathbb{E}[X_1] - \beta\mathbb{E}[\tau_1]$$

$\mathbb{P}$-f.s.

Falls also $\mathbb{E}[X_1] - \beta\mathbb{E}[\tau_1] > 0$ gilt, dann konvergiert Y_n $\mathbb{P}$-f.s. gegen 1 und gemäß Lemma 1.2 tritt der Ruin $\mathbb{P}$-f.s. ein. Mittels Resultaten über *zufällige Irrfahrten* lässt sich auch im Fall $\mathbb{E}[X_1] - \beta\mathbb{E}[\tau_1] = 0$ nachweisen, dass der Ruin $\mathbb{P}$-f.s. eintritt. Für ein Versicherungsunternehmen besteht nur dann die Gelegenheit, die Ruinwahrscheinlichkeit zu verringern, falls $\mathbb{E}[X_1] - \beta\mathbb{E}[\tau_1] < 0$ gilt. Diese Bedingung hat in der Literatur einen eigenen Namen:

Definition 1.14. *Der klassische Risikoprozess im Cramer Lundberg-Modell erfüllt die Nettoprofitbedingung, falls gilt:*

$$\mathbb{E}[X_1] - \beta\mathbb{E}[\tau_1] < 0$$

Definition 1.14 Im Cramér-Lundberg-Modell mit einem klassischen Risikoprozess heißt der Wert

$$\rho = \frac{\beta\mathbb{E}[\tau_1]}{\mathbb{E}[X_1]} - 1$$

Sicherheitszuschlag (safety loading).

Bemerkung 1.4. *Die Nettoprofitbedingung ist genau dann erfüllt, wenn der Sicherheitszuschlag positiv ist.*

Im allgemeinen kann man aber wenig über die Ruinwahrscheinlichkeit aussagen. Die Berechnung der Ruinwahrscheinlichkeit geht in die Erneuerungstheorie herein, oft kann man nur die Laplacetransformierte berechnen. Diese dann umzuwandeln ist nur in ein paar Spezialfällen möglich. Deswegen werden wir auf diesen Punkt hier nicht näher eingehen.

Es besteht aber die Möglichkeit der Monte Carlo simulation. Da das Cramer Lundberg aus recht einfachen und zumeist leicht zu simulierbaren Bausteinen besteht, kann man mittels Monte Carlo Simulation näheres über die Verteilung herausfinden (siehe Kapitel 4)

2 Risikokennzahlen

i Allgemeiner Risikobegriff

Werden in wissenschaftlichen Diskussionen Betrachtungen über das Risiko angestellt, stehen in ihren Mittelpunkt die Unsicherheit über künftige Ereignisse und Entwicklungen. Man möchte gerne das Risiko quantitativ messen. Dies führt zu dem Begriff Risikomaß, also einer einzige Zahl, die aussagen soll, wie risikoreich der drohende Verlust sein kann.

Schon im vorigen haben wir uns damit auseinander gesetzt, wie Risiken durch stochastische Verteilungsmodelle beschrieben werden können. Neben Risikokennzahlen, die sich auf die Verteilung beziehen, sind auch Kennzahlen interessant, die den Grad einer - oft in funktionaler Form gegebenen - Abhängigkeit von Wertbeeinflussenden Parametern erlassen. Zur Unterscheidung bezeichnen wir Risikokennzahlen, die sich auf die Wahrscheinlichkeitsverteilung beziehen, als stochastische Risikokennzahlen und Risikokennzahlen, die die Sensitivität funktionaler Abhängigkeiten messen, als analytische Risikokennzahlen.

ii Stochastische Risikokennzahlen

Für den Einsatz in Gesamtunternehmensmodellen sowie auch zum besseren Verständnis der grundsätzlich analogen Vorgehensweise ist es sinnvoll, Schadenrisiken und finanzielle Risiken mit einem einheitlichen Ansatz, also auf Basis einer gemeinsamen Werteskala, zu bewerten. Dazu wird im Folgenden als relevante Zufallsvariable meist der innerhalb der vorgegebenen Zeitperiode entstandene Schaden bzw. Verlust V betrachtet. Ein finanzieller Verlust bzw. eine Schadenshöhe wird dabei durch einen positiven Wert und ein finanzieller Gewinn durch einen negativen Wert von V beschrieben. Diese Konvention ist im Zusammenhang mit finanziellen Risiken zunächst etwas gewöhnungsbedürftig, besitzt aber auch dort einige Vorteile, z. B. kann man in Bilanzsimulationen positive Werte von V als erforderliches Sicherheitskapital interpretieren. Alternativ ist es aber auch möglich, unmittelbar die Zufallsvariable $W = -V$ zu betrachten, also Gewinne mit einem positivem Vorzeichen zu versehen und Verluste mit einem negativen. In der Literatur kommen beide Vorzeichenkonventionen vor, je nachdem ob eher Schäden / Verluste oder mögliche Gewinne im Mittelpunkt der Betrachtung stehen.

Das Ziel ist es eine einfache Zahl zu finden, die das Risiko ausdrückt. So eine Zahl nennt man ein Risikomaß.

Definition 2.1. *Ein* Risikomaß *ist eine Abbildung*

$$V \mapsto \rho(V) = \rho(F_V)$$

die einem Risiko V mit zugehöriger Verteilungsfunktion F_V eine reelle Zahl zuweist.

Es ist zu Bemerken, dass zwei Risiken mit der gleichen Verteilung stets der gleiche Risikowert zugeordnet wird.

iii Maße die auf Momente basieren

Der Mittelwert als Risikomaß

Der Mittelwert ist in Allgemeinen unzureichend. Dazu vergleiche folgendes Beispiel.

Beispiel 2.1. *Ein Betrieb besitzt drei gleichartige Maschinen. Bei beiden können Bauteil A oder Bauteil B ausfallen. Nach einem unplanmäßigen Stromausfall müssen mit jeweils* 10% *Wahrscheinlichkeit Bauteil A und Bauteil B erneuert werden. Es besteht kein Zusammenhang zwischen den jeweiligen Defekten bzw. den zugehörigen Neubeschaffungskosten. Der Gesamtschaden im Falle des Stromausfalls ergibt sich als Summe*

$$S = \sum_{i=1}^{6} X_i$$

der zufallsbedingten Auswechslungskosten derb betroffenen Bauteile. Die Wahrscheinlichkeitsverteilung von S soll explizit berechnet werden

- *für den Fall, dass beide Bauteile* 30.000 *€kosten (alle Angaben in e), d. h. für* $i = 1, \dots 6$ *ist*

$$X_i \quad = \quad \begin{cases} 30.000 & €mit\ Wahrscheinlichkeit\ 0.1, \\ 0 & mit\ Wahrscheinlichkeit\ 0.9, \end{cases}$$

- *für den Fall, dass Bauteil A* 20.000 *€und Bauteil B e*40.000 *kostet, d.h. für für* $i = 1, \dots 6$ *ist*

$$X_1, X_2, X_3 \quad = \quad \begin{cases} 20.000 & mit\ Wahrscheinlichkeit\ 0.1, \\ 0 & mit\ Wahrscheinlichkeit\ 0.9, \end{cases}$$

$$X_4, X_5, X_6 \quad = \quad \begin{cases} 40.000 & mit\ Wahrscheinlichkeit\ 0.1, \\ 0 & mit\ Wahrscheinlichkeit\ 0.9, \end{cases}$$

- *Zudem soll jeweils die erwartete Schadensumme und die Varianz des Schadens bestimmt werden.*

In Fall a) liegt eine Binomialverteilung vor. Der Gesamtschaden identisch ist mit 30.000 *mal Anzahl der Schäden. Der Erwartungswert beträgt* $6 \cdot 30.000^2 \cdot 0.1 \cdot 0.9 = 18.000$*, die Varianz* $6 \cdot 30.000^2 \cdot 0.1 \cdot 0.9^2 = 486.000.000^2$*, die Standardabweichung also* $22045, 41$*.*

Im Fall b) muss man die verschiedenen Kombinationsmöglichkeiten betrachten, wobei der Binomialkoeffizicnt für die Anzahl der Möglichkeiten steht, dass bei $k = 0, 1, 2$ *oder* 3 *Maschinen Bauteil A bzw. B defekt ist. Der Erwartungswert beträgt (*$3 \cdot 20.000 + 3 \cdot 40.000) \cdot 0.1 = 18.000$*. die Varianz* $(3 \cdot 20.000^2 + 3 \cdot 40.0O0^2) \cdot 0.1 \cdot 0.9 = 540.000.000$*. die Standardabweichung also* $23237, 90$*.*

Man erkennt an diesem Beispiel u.a. dass der erwartete Schaden nur ein unzureichendes Indiz für das tatsächliche Risiko ist. So ist die Varianz in den beiden Varianten bei gleichem Erwartungswert recht unterschiedlich. Weitere Risikokennzahlen werden in Kapitel 3 behandelt.

Varianz und Standardabweichung als Risikomaß

Es werden auch oft allgemeine Streuungsmaße wie die Varianz oder Standardabeweichung als Risikokennzahlen benützt. Sie geben an in welcher Größenordnung sich die Abweichungen von den tatsächlichen erzielten Gewinn zum Erwartungswert bewegt, und geben den Grad der Unsicherheit an. Ein Nachteil. ist aber dass kleinere Verluste und größere Verluste gleich behandelt werden.

Weitere relevante Risikokennzahlen ergeben sich, wenn man eine mittlere Abweichung vom Median oder vom Modus berechnet. Hier sind je nach verwendetem Abstandsbegriff verschiedene Definitionen möglich. Von größerer Bedeutung ist vor allem die Größe $E(|V - M(V)|)$, also die erwartete absolute Abweichung vom Median. Beim Median als Bezugsgröße ist, anders als bei der Definition der Varianz, tatsächlich der Absolutbetrag in der Regel das geeignete Abstandsmaß,. Im wichtigen Fall der Normal Verteilung ist die Wahrscheinlichkeilsmasse innerhalb bzw. außerhalb von $\pm\mathbf{Std}(V)$ Standardabweichungen um den Erwartungswert immer die gleiche, unabhängig von der speziellen Wahl von und , man spricht auch von σ -Äquivalenzen.

Darüber Hinaus werden auch manchmal Schiefmaße als Risikokennzahlen eingesetzt. Derartige Schiefmaße sind so konstruiert dass Abweichungen zwischen den Mittelwerten, Median und Mode betrachtet werden und ggf. normiert werden. Es gibt dazu verschiedene Varianten:

$$\begin{aligned} \text{Schiefmaß nach Pearson} \quad & \frac{\mathbb{E}(V) - \mathbf{mod}(V)}{\mathbf{Std}(V)}, \\ \text{Schiefmaß nach Yule Pearson} \quad & \frac{3(\mathbb{E}(V) - \mathbf{M}(V)}{\mathbf{Std}(V)}, \\ \text{Schiefmaß gemäß dritten Moment} \quad & \frac{\mathbb{E}(V - \mathbb{E}(V))^3}{\mathbf{Std}(V)^3}. \end{aligned}$$

Shortfallmaße

Die Gefahr der überschreitung einer vorgegebenen Verlustschwelle a messen sogenannte Shortfallmaße. Das einfachste Shortfallmaß ist die Shortfall Wahrscheinlichkeit:

Definition 2.2. *Für eine gegebene Verlustverteilung L ist die* Shortfallverteilung *bzgl. des Schwellenwertes x definiert als*

$$\mathbf{SW}_x(L) := \mathbb{P}(L > x).$$

Dies ist also die Wahrscheinlichkeit ϵ, dass der Verlust L den Wert x überschreitet.

Auch betrachtet man oft für Verluste die oberen partiellen Momente (upper partial moments):

$$\mathbf{UPM}_{(h,a)}(X) = \begin{cases} \mathbb{E}\left(\max(0, X-a)^h\right) & \text{für } h > 0, \\ \mathbb{P}(X \geq a) & \text{für } h = 0. \end{cases}$$

Spezialfälle sind die Überschreitungswahrscheinlichkeit der kritischen Grenze a ($h = 0$), die mittlere Überschreitung ($h = 1$) und die Semivarianz ($h = 2$).

Allgemeine Kritikpunkte mit Momentbasierten Maßen

Ein Problem ist, dass diese Momentbasierten Maße finanziell schwer interpretierbar sind. Am ehesten lässt sich noch die Standardabweichung als *durchschnittlicher Abstand zum Erwartungswert* interpretieren. Jedoch ist der euklidische Abstand zwar ein gutes Entfernungsmaß aber eben kein natürliches ß für finanzielle Risiken.

Für viele in der Versicherungsindustrie angewendete Verteilungen existieren höhere Momente nicht. Bei der Modellierung operationeller Risiken mit Hilfe der Verallgemeinerten Parteo Verteilung ist in der Praxis vorkommenden Parameter mitunter nicht einmal der Erwartungswert definiert.

iv Valued at Risk (VaR)

Gerade diese Vorteile sind es, die den Value at Risk, als Risikomaß, vor allem im finanzwirtschaftlichen Sektor so populär gemacht haben. Besonders die Morgan Guaranty Trust Company, kurz J.P. Morgan, hat 1994 mit ihrem RiskMetrics TM System zur Verbreitung des Value at Risk beigetragen. Seinen Durchbruch erlangte der Value at Risk in der zweiten Hälfte der 90er Jahre, als er im Aufsichtsrecht der Banken durch Basel II als verbindliches Risikomaß vorgeschrieben wurde.

Der Value-at-Risk (VaR) , welcher sich sowohl im Bereich des Marktes- als auch des Kreditrisiko-Management als zentrale Kennzahl des Portfoliosrisikos durchgesetzt hat, ist definiert als der maximale Verlust eines Kreditportfolios, der mit einer bestimmten Wahrscheinlichkeit innerhalb eines festgelegten Zeitraumes T nicht überschritten wird. Die Abhängigkeit von T kann man in die Modellierung des Verlustes mit hereinnehmen. Z.b. falls $L = R(T)$ gilt, wo $R = \{R(t) : t \geq 0\}$ ein Risikoprozess darstellt).

Ist der Verlust eines Kreditportfolios durch die Zufallsvariable L beschrieben, ist der VaR formal wie folgt definiert

$$\mathbf{VaR}_\alpha(L) = \inf_{l \in \mathbb{R}} \{\mathbb{P}(L > l) \leq 1 - \alpha\}.$$

Er ist die kleinste Zahl l, die der Verlust L mit einer Wahrscheinlichkeit von $1 - \alpha$ nicht überschreitet.

Ist die Verteilungsfunktion des Verlustes $F_L(l) = \mathbb{P}(L \leq l)$, lässt sich der VaR auch als α–Quantil der Verteilungsfunktion interpretieren,

$$\mathrm{VaR}_\alpha(L) = F^{-1}(\alpha).$$

Die für die Berechnung des VaR ausschlaggebende Wahrscheinlichkeit wird mittels eines Konfidenzinterval α festgelegt. In der Praxis übliche Werte für α sind 0.05 oder 0.01. Allerdings, ein hohes Konfidenznivaeau und die damit verbundene geringe Wahrscheinlichkeit VaR-überschreitende Verluste erzeugt zwar ein Gefühl der Sicherheit, erschwert aber die exakte und verläßliche Berechnung des VaR sowie die Überprüfung der Güte VaR–Modellen anhand tatsächlichen eintretenden Verlusten.

Kritikpunkte:

- Der VaR misst nur einen Punkt der Verlustverteilung und vernachlässigt dadurch das Risiko oberhalb dieses Punktes (Tail Risk). Dies hat zu Folge dass das tatsächliche Ausmaß der Abweichung vom VaR sowohl im positiven als auch im negativen keinen Einfluss auf eine Anlageentscheidung hat. Z.B. sind Portfolios denkbar, die den selben VaR haben und einem Entscheider somit gleichwertig erscheinen, aber bei einer Überschreitung des VaR zu stark unterschiedlichen Ausfällen führen.

  Beispiel 2.2. *Angenommen L_1 ist Gamma Verteilt mit Parameter $k = 3$ (shape parameter) und $\lambda = 1$ (scale parameter), und L_2 ist Log Normal verteilt mit Parametern $r = 1; \sigma = 0.51$. Dann gilt für $\alpha = 0.05$, $\mathbf{VaR}_\alpha(L_1) = \mathbf{VaR}_\alpha(L_1) = 6.2958$, aber der durchschnittliche Verlust ist im Falle der Gamma Verteilung, i.e. falls $L_1 \geq \mathbf{VaR}_\alpha(L_1)$ gleich 0.0041 und im Falle der Log Normalverteilung, i.e. falls $L_2 \geq \mathbf{VaR}_\alpha(L_2)$, gleich 0.0110, also mehr als doppelt so groß.*

- Der VaR ist nicht subadditiv und damit kein kohärentes Risikomaß. Das Axiom der subadditivität spiegelt die Idee der Subdervisifikation wider, wonach die Aufteilung eines Anlagebetrags auf mehrere Einzelrisiken immer zu einer Reduzierung des Gesamtrisiko führt. Da dieses Axiom nicht erfüllt ist, gibt es Portfolios deren VaR bei steigender Diversifikation steigt. Dieses soll anhand folgendes Beispiel gezeigt werden.

  Beispiel 2.3. *Es stehen 50 verschiedene Anleihen mit einem Kurswert von 95 Euro und einer Ausfallwahrscheinlichkeit von 2 Prozent zur Verfügung. Portfolio A besteht aus 100 Stück einer dieser Anleihen, Portfolio B aus je zwei Stück aller 50 Anleihen.*

  Aus Portfolio A resultiert mit einer Wahrscheinlichkeit von 98% ein Gewinn von e$100 \cdot 5 = 500$ und mit einer Wahrscheinlichkeit von 2% ein Verlust von e9500. Der $VaR_{0.05}$ dieses Portfolios ist kleiner als 0, da der mögliche Verlust eine geringer Wahrscheinlichkeit als 5% hat.

  Die Verlustverteilung des Portfolios B ist binominalverteilt. Bei einem α–Wert von 0.05 hat das entsprechende α Quantil einen Wert von 3, dass heißt es fallen mit einer Wahrscheinlichkeit , die grosser ist als 95% drei der 50 Anleihen aus. Dadurch einsteht ein Verlust von $6 \cdot 95$, die restlichen 94 führen zu einem Gewinn von je 5 Anleihen. Der VaR des Portfolios entspräche in diesem Fall $6 \cdot 95 - 94 \cdot 5 = 100$ und ist somit größer als der des konzentrierten Portfolios A.

- Bei riskioorientierten Ertragsmessung wird die Performanz eines Portfolios häufig als Verhältnis von Gewinn und eingesetzten ökonomischen Kapital gemessen Wird das ökonomische Kapital hierbei wie üblich mit dem VaR bestimmt, führt eine Optimierung dieser Performanz aufgrund der fehlenden Subadditivität wiederum zu einer Konzentration im Portfolio. Weiterhin wird die Optimierung dadurch erschwert, dass der VaR als Funktion der Portfolio-Positionierungen nicht konvex ist und mehrere lokale Extremwerte aufweisen kann.

v Expected Shortfall

Risikomaße, die berücksichtigen dass nur die negative Abweichung von einem erwarteten zukünftigen Wert als Risiko betrachtet wird, werden als Downside oder **Shortfall-Risikomaß** bezeichtnet. Ist man sich der oben dargestellten stark rechtsschiefen Verteilung der Kreditverluste bewusst, ist ersichtlich, dass die Gefahren eines Kreditportfolios primär bei den hohen, aber seltenen auftretenden Verlusten am rechten Rand (Tail) der Verlustverteilung liegen. Es werden also Risikomaße benötigt, welche die Eigenschaften in den Tails möglichst adäquat beschreiben.

Das bekannteste Shortfall Maß ist der Expected Shortfall. Der Expected Shortfall, auch *Conditional VaR* oder *Average Var*, wurde 1997 von Artzner et al. als kohärentes Risikomaß eingeführt. Er ist definiert als der erwartete Verlust für den Fall, dass **VaR** tatsächlich überschritten wird. Somit ist er der wahrscheinlichkeitgewichtete Durchschnitt aller Verluste, die höher sind als der **VaR**. Sei X eine Zufallsvariable die den Verlust eines Portfolios beschreibt und $\mathbf{VaR}_\alpha(X)$ der VaR bei einem Konfidenzniveau

von $100(1-\alpha)$ Prozent. Dann ist der Expected Shortfall definiert durch

$$\mathrm{ES}_\alpha(X) = \mathbb{E}\left[X \mid l \geq \mathbf{VaR}_\alpha(X)\right].$$

Die bedingte Erwartung berechnet sich bei einer stetigen Verteilung F_X durch

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha}\int_{\mathbf{VaR}_{\alpha(X)}}^{\infty} x\, dF_X(x).$$

Auch ist in der Literatur folgende (gleichwertige) Definition zu finden

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathbf{VaR}_z(X)\, dz.$$

Der Expected Shortfall ist ein kohärentes Risikomaß. Zudem wird schon aus seiner Definition ersichtlich, dass auf die bei Kreditausfällen eher seltenen Erignisse mit hohen Verlusten fokussiert wird. Somit bestehen die vom **VaR** bekannten, aus der Vernachlässigung des Tail Risks und der fehlenden resultierenden Probleme nicht.

Ebenso, ist der Expected Shortfall ein konvexes Risikomaß. Durch die Konvexeigenschaft lassen sich Optimierungsalgorithmen entwerfen. So lässt sich der Expected Shortfall z.B. unter Anwendungen der Linearen Programmierung minimieren.

Konzeptionell kann der Expected Shortfall dem **VaR** als überlegen angesehen werden. Es gibt aber auch einige Nachteile.

Für die Stabilität der Ergebnisse ist die möglichst genaue Schätzung des Tails der Verlustverteilung ausschlaggebend. Diese Schätzung ist mit herkömmlichen Methoden sehr schwierig. Z.B. verändern sich unter normalen Bedingungen beobachtete Korrelationen in Extremsituationen. Daher ist es nicht möglich den Tail dieser Verteilung mit konventionellen Monte-Carlo Simulationen zu schätzen.

Das Backtesting eines geschätzten Expected Shortfalls ist ebenfalls schwieriger als es für den **VaR** ist.

Tail VaR

Dieses Shortfall-Maßbezieht sich auf den Schwellenwert $x = VaR$. Zu einem vorgegebenen Konfidenzniveau $\alpha \in (0,1)$ ist der *Tail Value at Risk* definiert als

$$\mathrm{TVaR}(V;\alpha) := \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathrm{VaR}(V;s)\, ds.$$

Er ergibt sich als über die Mittelung aller **VaR**(V;s)-Werte für $s \geq \alpha$; damit wird also im Gegensatz zum VaR auch die Höhen der extremen Verluste berücksichtigt Der Tail **VaR** ist in der Literatur auch unter den Namen expected Shortfall und Conditional Value at Risk zu finden.

Der TailVaR ist ein kohärentes Risikomaß.

Worst conditional expectation

Dieses Shortfall-Maßbezieht sich auf den Schwellenwert $x = VaR_\alpha$. Hier ist die Struktur des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{F}, \mathbb{P})$ wichtig, insbesondere die σ–Algebra $\mathcal{F}$. Das heißt, da die σ–Algebra mit der Information über die Zufallsvariable gleichzusetzten ist, hängt dieses Risikomaß von den Wissen über die Zufallsvariable L ab. Zu einem vorgegebenen Konfidenzniveau $\alpha \in (0,1)$ ist der *Worst conditional expectation* definiert als

$$\mathrm{WCE}(V;\alpha) := \inf\{\mathbb{E}[X \mid A] : A \in \mathcal{F}, \mathbb{P}(A) > \alpha\}.$$

WCE ist subadditive, $WCE_\alpha \geq RCE_\alpha$.

vi Spektralmaße

Der Expected Shortfall lässt sich direkt verallgemeinern, um die individuelle Risikoaversion zu berücksichtigen. Statt über alle $\mathbf{VaR}_z^T(X)$ mit $z \geq \alpha$ mit gleichen Gewicht zu mitteln, kann man eine Gewichtsfunktion ϕ verwenden.

Eine Funktion $\phi : [0,1] \to \mathbb{R}$ heißt Gewichtsfunktion, falls gilt

- $\phi(\alpha) \geq 0$ für alle $\alpha \in [0,1]$;
- $\int_0^1 \phi(\alpha)\, d\alpha = 1$.

Definition 2.3. *Sei ϕ eine Gewichtsfunktion. Dann heißt das Risikomaß*

$$M_\phi(X) = \int_0^1 \mathbf{VaR}_\alpha(X)\phi(\alpha)\, d\alpha.$$

Mit einem Spektralmaßwird das Risiko auch in Abhängigkeit von der Seltenheit, mit der ein Verlust eintreten kann gewichtet. Das Konzept des Spektralmaßes ermöglicht somit die Abbildung eines individuellen Profils der Risikoaversion.

Man sieht leicht, dass der exspected Shortfall $\mathrm{ES}_\alpha(X)$ ein Beispiel für ein Spektralmaß mit Gewichtsfunktion $\phi = 1$ ist. Der Value at Risk kann als Grenzfall von Spektralmaßen verstanden werden, wo als Gewichtsfunktion δ_α gewählt wurde.

vii Anforderungen an eine Risikomaß

Allgemein ist ein Risikomaß ρ eine Abbildung, die jedem Portfolio mit dem zukünftigen Wert X eine Zahl $\rho(X)$ zuweist. Hier werden einige Eigenschaften vorgestellt, die ein Risikomaßbesitzen sollte.

Folgende Eigenschaften wären wünschenswert:

- **Subadditivität:** Das Risiko eines aggregierten Portfolios darf nicht größer sein als die Summe seiner Komponenten.

  $$\rho(X + Y) \leq \rho(X) + \rho(Y).$$

  Das bedeutet, dass durch eine Zusammenfassung von zwei Risiken niemals zusätzliches Risko entstehen kann, sondern sich das Gesamtrisiko eher vermindert. Dies entspricht dem bekannten Prinzip der Risikodiversifikation, wird aber nicht von allen Risikomassen erfüllt. Der Value-at-Risk erfüllt diese Bedingung nur für bestimmte Verteilungstypen wie etwa der Normalverteilung. Im Allgemeinen ist der Value-at-Risk jedoch kein subadditives Risikomaß.

  Insbesondere wenn das Risikomaß unmittelbar als Kapitalbedarf interpretiert werden kann, die Subadditivität aus unternehmerischer Sicht eine sinnvolle Eigenschaft, da sie eine dezentrale Risikosteuerung ermöglicht. Angenommen ein Unternehmen mit Gesamtrisiko X besteht zwei Geschäftseinheiten mit jeweiligen Risiken X_1 und X_2 Die Unternehmensleitung strebt das Gesamtrisiko $\rho(V) < 10$ Mio. Euro an. Wenn das verwendete Risikomaß subadditiv ist, d; kann die Unternehmensleitung ihr Ziel erreichen, indem sie den einzelnen Geschäftsbereich vorgibt, dass $\rho(V) < 5$ Mio. Euro und $\rho(V) < 5$ Mio. Euro sein sollte.

  Aus regulatorischer Sicht ist die Forderung der Subadditivität ebenfalls sinnvoll, wie die folgende Überlegung zeigt. Angenommen ein Unternehmen mit Gesamtrisiko V besteht wieder aus zwei Geschäftsbereichen mit entsprechenden Risiken V_1 und V_2. Wenn die Subadditivität gelten würde, wäre $p(V) = p(V_1 + V_2) > p(V_1) + p(V_2)$. Eine virtuelle Aufspaltung des Unternehmens in zwei Teilunternehmen wäre damit aus Sicht der Unternehmensleitung sinnvoll, sich dadurch das benötigte Risikokapital reduzieren würde. An der tatsächlichen Risikosituation des Gesamtunternehmens hätte sich durch diesen *Trick* jedoch nichts geändert.

- **Motonie:** Aus $X_1 \geq X_2$ fast überall folgt, dass gilt $\rho(X_1) \geq \rho(X_2)$;

  Das bedeutet das ein Portfolio, das in jeder möglichen Situation höhere Verluste als ein anderes Portfolio aufweist, auch zu einem höheren Risikokapital führen muss. Ein Beispiel wären zwei identische Portfolios, wobei das zweite Portfolio für die Prämien einen schadensabhängigen nachträglich gewährten Rabatt aufweist. Damit hat das zweite Portfolio ein höheres Risiko.

- **Positive Homogenität:** Das Risiko steigt proportional zu einem positiven Faktor λ, d.h.

$$\rho(\lambda \cdot X) = \lambda \cdot \rho(X).$$

  Positive Homogenität ist eine oftmals verwendete mathematische Eigenschaft, jedoch aus ökonomischer Sicht manchmal weniger sinnvoll. Positive Homogenität bedeutet, dass das Risiko einer Position proportional mit ihrer Größe wächst.

  Man beachte, dass diese Eigenschaft folgender Asymmetrie zwischen Gewinnen und Verlusten nicht gerecht wird: Vergrössert sich die Position um einen Faktor λ, so kann das mit ihr assoziierte Risiko stärker als λ wachsen, sofern die Kosten zur Verlustkompensation schneller als die Verlustgröße ansteigen (je größer eine Position, desto teurer wird es typischerweise, sie zu liquidieren). Das heißt, für ein kleines Unternahmen, welches nicht so grosse Rücklagen hat, kann es bei größeren Verlusten schnell zu Liquiditätsproblemen kommen, und damit hat ein dreifacher Verlust, ein mehr als dreifaches Risiko. Ein Ausweg ist, dass man als Kriterium

$$\rho(\lambda \cdot X) = \lambda^{\gamma} \cdot \rho(X).$$

  für ein $\gamma \geq 1$ annimmt.

- **Translationsinvarianz** (risk free condition, cash invariance)**:** Wird zusätzlich zum betrachteten Portfolio ein Betrag m angelegt, so reduziert sich das Risiko um den Betrag m, d.h.

$$\rho(X + m) = \rho(X) - m.$$

  Cash- Invarianz formalisiert die Vorstellung, dass Risiken auf einer monetären Skala gemessen werden: Wird ein risikofreier Geldbetrag m zur Position X addiert, dann reduziert sich das Risiko um m. Verteilungsinvarianz bedeutet, dass zwei Positionen, die den selben Verteilungen genügen, durch das selbe Risiko charakterisiert sind. Aus der Translationsinvarianz folgt außerdem $\rho(X - \rho(X)) = 0$. Das Risikokapital ist also genau der Geldbetrag der gehalten werden muss um bezüglich des Risikomaßes das Risko vollkommen abzufedern.

- **Komotone Additivität:** Es gilt $\rho(X + Y) = \rho(X) + \rho(Y)$ für X und Y comonoton.

  Das bedeutet, dass für zwei Risiken, die sich vollständig gleichläufig verhalten, das Risikomaß Zahl des Gesamtrisikos exakt der Summe der Risikomaßzahlen der Einzelrisiken entspricht. Wenn ein Risiko X als Summe der beiden komonotonen Risiken $X_1 = \frac{1}{2}X$ und $X_2 = \frac{1}{2}X$ betrachtet wird, besagt die komonotone Additivität, dass $\rho(X) = \rho(\frac{1}{2}X + \frac{1}{2}X) = \rho(\frac{1}{2}X) + \rho(\frac{1}{2}X) = 2 \cdot \rho(\frac{1}{2}X)$ gilt. In diesem Fall wäre also $\rho(X) = \frac{1}{2}\rho(\frac{1}{2}X)$, was der positiven Homogenität (R2) entspricht.

- **Konvexität** formalisiert die Idee, dass Diversifikation risikoreduzierend wirkt. Sei X_1, X_2 zwei Verlustverteilungen und und $\alpha \in (0, 1)$. Die Position $\alpha X_1 + (1 - \alpha)X_2$ wird dann als diversifiziert bezeichnet. Das Risikomaß ist konvex, wenn das Risiko einer diversifizierten Position niemals größer als die gewichtete Summe der Teilrisiken ist, d.h., wenn für alle Verlustverteilungen X_1, X_2 und und $\alpha \in (0, 1)$ gilt

$$\rho(\alpha X_1 + (1 - \alpha)X_2) \leq \alpha\rho(X_1) + (1 - \alpha)\rho(X_2).$$

(Seite 102)

Definition 2.4. *Zwei Risiken V_1 und V_2 heißen* komonoton*, wenn es eine Zufallsvariable Z und eine monoton steigende Funktion f_1 und f_2 gibt, sodass gilt*

$$V_1 = f_1(Z) \quad \text{und } V_2 = f_2(Z).$$

Anschaulich bedeutet dies, dass die beiden Risiken vollständig von einem gemeinsamen Risikofaktor Z abhängen.

S 102.

Definition 2.5. *Ein Risikomaßheißt* **kohärentes Risikomaß***, falls es translationsinvariant, positiv homogen, monoton und subadditiv ist.*

Beispiele kohärenter Risikomaße:

- Der expected Shortfall ist ein koährentes Risikomaß.
- Ist ϕ eine monoton wachsende Gewichtsfunktion, so ist das zugehörige Spektralmaß kohärent. Ein Spektralmaß ist also kohärent, wenn die individuelle Risikoaversion höheren Verluste auch höhere Gewichte zuordnet.
- Das Risikomaß $\mathbf{VaR}_\alpha$ ist translationsinvariant, positiv homogen und monoton;
- Das Risikomaß $\mathbf{VaR}_\alpha$ ist kohärent falls es auf normalverteilte Risiken eingeschränkt wird.

Ein Risikomaß das genau alles erfüllt heißt **kohärentes Risikomaß**. Zudem ist durch die positive Homogenität in Verbindung mit der Subadditivität sichergestellt, dass das Risikomaß konvex ist. Hierdurch ist eine Portfolio-Optimierung im Sinne der Minimierung des Risikos, auf der Grundlage dieses Risikomaßes stets lösbar.

Das in der finanzwirtschaftlichen Theorie und Praxis am meisten verbreitete Risikomaß ist die Varianz bzw. die Standardabweichung als ihre Quadratwurzel. Sie mißt die Streuung einer Verlustverteilung um den Erwartungswert und berücksichtigt dadurch gleichermaßen für den Kreditgeber positive und negative Abweichungen von diesem. Somit entspricht das durch die Standardabweichung beschriebene Risiko nicht den oben allgemein definierten Risikobegriff.

viii Diskussion

Ein Nachteil der Risikomaße $\mathbf{Var}(V)$, bzw. $\mathbf{Std}(W)$ ist. dass positive Abweichungen als genauso riskant eingestuft werden wie negative Abweichungen. Kennzahlen, die auf höheren Momenten basieren (wie Schiefe- und Wölbungsmaßzahlen), können Zusatzinformationen geben. Aber auch sie sind für eine Einschätzung darüber, wie wahrscheinlich unerwünschte oder sogar katastrophale Ereignisse sind, nur bedingt hilfreich.

seite 42 in wertorientiertes risikomangement;

Untehrnehmenssteuerung: wahrnehmen der Chancen nicht nur der Risiken -

ix Schätzen von Risikomaßen

Schätzen des $\mathbf{VaR}_\alpha$ - Schätzen von Quantilen

Die Quantile x_α zu einen Wert α einer Zufallsvariable X ist definiert durch

$$\mathbb{P}(X \leq x_\alpha) = 1 - \alpha.$$

Angenommen X steht für die Festigkeit eines Bauteils. Möchte man, dass die Bauteile nur mit einer Wahrscheinlichkeit von $(1-\alpha)$ brechen, so ist man an den Anteilen von Bauteilen interessiert mit Festigkeit kleiner als x_α. Der Wert α an den wir interessiert sind ist zumeist sehr nahe bei 1, z.B. $\alpha = 0.995$. Das Problem das jetzt entsteht ist, dass die Anzahl der Beobachtungen sehr gering sind, und das wenn man auch die Verteilung modellieren kann, die Unsicherheiten für grosse x sehr gross sind.

Um diese Probleme zu umgehen, gibt es spezielle Lösungen, eine ist die **P***eaks* **o***ver* **T** *Threshold* (POT) Methode.

Satz 2.1. *Under geeigneten Bedingung an die Zufallsvariablen X, welche zumeist erfüllt sind, und u_0 sehr gross ist, gilt*

$$\mathbb{P}\left(X > u_0 + h \mid X > u_0\right) \sim 1 - F(h; a, c),$$

wobei $F(h; a, c)$ eine Verallgemeinerte Pareto Verteilung ist, i.e.

$$F(h; a, c) = \begin{cases} 1 - (1 - ch/a)^{1/c}, & \textit{falls } c \neq 0, \\ 1 - \exp(-h/a) & \textit{falls } c = 0, \end{cases}$$

für $) < h < \infty$, falls $c \leq 0$ und für $0 < h < a/c$ falls $c > 0$ gilt.

Diskrete Schadensverteilung

Ein Betrieb besitzt drei gleichartigen Maschinen. Beide können Bauteil A oder Bauteil B ausfallen. Nach einem unplanmässigen Stromausfall Bauteile A und B wurden jeweils in verschiedenen Maschinen eingebaut. Insgesamt in sechs Maschinen, drei vom Typ A und drei vom Typ B.

Zunächst sei nur der Bestand aus den drei Maschinen vom Typ A betrachtet.

Kapitel 4

Monte-Carlo Simulation

1 Was ist Monte-Carlo Simulation

Monte-Carlo-Simulation oder Monte-Carlo-Studie, auch MC-Simulation, ist ein Verfahren aus der Stochastik, bei dem sehr häufig durchgeführte Zufallsexperimente die Basis darstellen. Es wird dabei versucht, mit Hilfe der Wahrscheinlichkeitstheorie analytisch nicht oder nur aufwändig lösbare Probleme numerisch zu lösen. Als Grundlage ist vor allem das Gesetz der großen Zahlen zu sehen. Die Zufallsexperimente werden in allgemeinen dabei am Computer mit Zufallszahlengeneratoren erzeugt.

Das Zufallsexperiment ist durch eine Zufallsvariable X beschrieben. Die kann z.B. der Risikoprozess zu einen fixen Zeitpunkt t sein. Hier in diesem Fall kann man die einzelnen Bausteine aus denen die Zufallsvariable zusammengesetzt ist, leicht simulieren. Allerdings ist es oft schwer möglich, die Verteilung von X analytisch zu beschreiben.

Im Allgemeinen möchte man folgendes Problem lösen: Gegeben ist die Zufallsvariable

$$X : \Omega \to \mathbb{R},$$

und eine Funktion $\phi : \mathbb{R} \to \mathbb{R}$. Berechne

$$\mathbb{E}\phi(X).$$

Diese Probleme können allerdings etwas variieren. Möchte man den Value at Risk für ein bestimmtes Signivikanzniveau $\alpha \in (0,1)$, d.h. $\mathbf{VaR}_\alpha$, eines Riskos X berechnen, so ist man an folgender Größe interessiert:

$$\inf_{x\in\mathbb{R}} \mathbb{E}\left[1_{(-\infty,x]}(X)\right] \geq 1-\alpha.$$

Bei der Monte Carlo Simulation macht man sich das starke Gesetz der großen Zahlen zunutze. Das heißt, der empirische Mittelwert einer Stichprobe nähert mit steigender Stichprobengröße sich den Mittelwert der Grundgesamtheit, aus der die Stichprobe gezogen wird, an. Mathematisch formuliert lautet dies so:

Satz 1.1. *Sei Y eine Zufallsvariable mit Erwartungswert a. Sei $\{Y_i : i = 1,\dots,n\}$ eine Stichprobe, d.h. $\{Y_i : i = 1,\dots,n\}$ ist Familie von unabhängigen Zufallsvariablen Y_i mit $Y_i \sim Y$. Sei*

$$D_n := \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

Dann gilt

$$\mathbb{P}\left(\lim_{n\to\infty} D_n = a\right) = 1.$$

Bei der Monte Carlo simulation berechnet man eine Stichprobe $\{Y_i : i = 1, \ldots, n\}$ um dann den Wert a abzuschätzen.

Als Beispiel nehmen wir den Risikoprozess $R = \{R(t) : 0 \leq t\}$ im Cramer Lundberg Modell. Angenommen wir interessieren uns für die Wahrscheinlichkeit, dass einen Firma am Ende eines Jahres rote Zahlen schreibt, d.h. wir interessieren und für den Wert ob das Ereignis $R(T) \geq 0$ zum Zeitpunkt $T = 1$ eintritt oder nicht. Dass heißt wir sind an der Zufallsvariablen $X = R(1)$, bzw. and der Grosse $\mathbb{P}(X > 0) = \mathbb{P}(R(T) > 0)$ interessiert. Setzt man $\phi(x) = 1_{(0,\infty)}(x)$ dann gilt $\mathbb{E}1_{(\infty,0)}(X) = \mathbb{P}(X > 0) = \mathbb{P}(R(T) > 0)$. Hier weiß man aus welchen Komponenten $X = R(T)$ aufgebaut ist, auch kennt man die einzelnen Verteilungen der Komponenten. Allerdings ist es oft analytisch sehr schwierig eine Verteilung der Zufallsvariablen $R(T)$ herzuleiten. Das heißt, es ist analytisch oft nicht möglich die Wahrscheinlichkeit, dass $R(T)$ grösser als Null ist, d.h. den Wert $c = \mathbb{P}(R(T) > 0)$ abzuleiten.

In der Monte Carlo Simulation erzeugt (berechnet) man eine Stichprobe $\{R_1, R_2, \cdots, R_n\}$ mit $R_i \sim R(T)$, $i = 1, \ldots, n$. Dies ist oft recht einfach, da man die Verteilungen der einzelnen Bauteile kennt. Um a zu berechnen, wertet man nur

$$a_n := \frac{1}{n} \sum_{i=1}^{n} 1_{[0,\infty)}(R_i)$$

aus. Aus den Gesetzt der großen Zahlen weiß man dass a_n mit $n \to \infty$ gegen a konvergiert. Wählt man nun n genug gross, ist a_n eine gute Näherung für a.

Die Geschwindigkeit der Konvergenz (in n) hängt natürlich von der Varianz von X ab. Hier gibt es auch Techniken die die Varianz reduzieren und damit die Konvergenzgeschwindigkeit erhöhen, wir werden aber hier nicht näher darauf eingehen.

Hier wird nur besprochen, wie man verschiedene Zufallsvariablen, bzw. einen Zählprozess generiert, das Gesamtrisiko simuliert und den **VaR**, bzw. expected Shortfall berechnet. Auch werden wir kurz eine Monte Carlo simulation des Cramer Lundberg Modell schildern.

Eine Einführung in Monte Carlo Algorithmen verweisen wir auf *Müller-Gronbach, Erich Novak, Klaus Ritter: Monte Carlo Algorithmen (2012)*, *Glassermann: Monte Carlo Methods in Financial engineering,(2004)*.

2 Erzeugung von Zufallszahlen

Die meisten Programmiersprachen enthalten zwei Arten von Zufallsgeneratoren. Einmal, einen Zufallsgenerator der eine auf einen Intervall $[0, 1]$ gleichverteilte Zufallsvariable X generiert, und einen Zufallsgenerator der eine ganzzahlige Zufallsvariable $Z : \Omega \to \{1, 2, \ldots, , n\}$ generiert. Da die Simulation der meisten anderen Verteilungen in der einen oder anderen Form auf diesen Generatoren beruht, ist es äußerst wichtig, dass gute Algorithmen implementiert wurden und die Implementierung valide Ergebnisse liefert. Hinsichtlich des letzten Punkts sind auch mögliche IT-spezifische Probleme zu beachten. Eine Einführung in die Problematik von Zufallszahlengeneratoren findet sich bei Lecuer.... Anzumerken ist, dass statistische Softwarepakete wie MatLab oder R enthalten zumeist Zufallszahlengeneratoren die Zufallsvariablen mit den gängigen Verteilungen erzeugen.

In Matlab stehen einige Zufallszahlengeneratoren zur Verfügung, der Befehl heißt *random* in MatLAb und z.B. ran in R. Die entsprechende Befehle sind in der Hilfe nachzulesen.

Trotzdem werden wir hier einige Verfahren vorstellen mit dessen Hilfe man aus gleichverteilten Zufallsvariablen Zufallsvariablen mit einer gegebener Verteilung gewinnt. Als erstes stellen wir zwei Verfahren zur Erzeugung von Zufallsvariablem mit stetiger Verteilungsfunktion vor, danach stellen wir auf die Erzeugung von Zufallsvariablen mit diskreter Verteilungsfunktion vor.

i Erzeugung stetiger Zufallszahlen - die Inversionsmethode

Gesucht wird eine Methode, um Zufallszahlen $\{X_i : i = 1, \ldots n\}$ gemäß einer gegebenen Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$ zu erzeugen. Die Inversionsmethode arbeitet mit der Quantilenfunktion. Dabei wird folgende wichtige Tatsache benützt:

Satz 2.1. *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion F und U eine $[0,1]$–gleichverteilte Zufallsvariable. Dann gilt für $F^{\leftarrow}(u) = \inf_{x\in\mathbb{R}}\{u \leq F(x)\}$*

$$F^{\leftarrow}(U) \quad \sim \quad X. \tag{4.1}$$

Ist F stetig und streng monoton steigend, so ist $F^{\leftarrow} = F^{-1}$. Hat F Sprünge oder ist F über ein Intervall konstant, so ist eigentlich F nicht invertierbar, aber $F^{\leftarrow}$ immer noch wohldefiniert und obige Beziehung (4.1) gilt trotzdem.

Satz 2.1 hat folgende Bedeutung für die Praxis: Wenn die inverse Verteilungsfunktion bzw. Quantilfunktion $F^{\leftarrow}$ bekannt ist, kann eine Realisierung der Zufallsvariablen X erzeugt werden, indem zunächst eine Zufallsvariable U die gleichverteilung ist, simuliert wird. Anschließend wird dieser Wert in die inverse Verteilungsfunktion eingesetzt, d.h. $X := F^{\leftarrow}(U)$. Satz 2.1 besagt nun, dass X tatsächlich eine Zufallsvariable der vorgegebenen Verteilung ist. Damit ergibt sich der folgende allgemeine Algorithmus:

Algorithmus zur Erzeugung einer Realisierung von X mit Verteilungsfunktion F gemäß der Inversionsmethode

1. Erzeuge U wobei U gleichverteilt auf $[0,1]$ ist.
2. Setze $X := F^{\leftarrow}(U)$.

Beispiel: Die Exponential-Verteilung Für die Exponential Verteilung $Exp(\lambda)$ mit Parameter $\lambda > 0$ ist die Verteilung gegeben durch $F(x) = 1 - \exp(-\lambda x)$. Da F streng monoton steigend ist, lässt sich $F^{\leftarrow}$ folgendermaßen berechnen

$$\begin{aligned} y &= 1 - \exp(-\lambda x) \\ 1 - u &= \exp(-\lambda x) \\ -\ln(1-u) &= \lambda x \\ -\frac{1}{\lambda}\ln(1-u) &= x. \end{aligned}$$

Da, falls U gleichverteilt auf $[0,1]$ ist, ist auch $1 - U$ gleichverteilt auf $[0,1]$ ist und es gilt

$$F^{\leftarrow}(u) = -\frac{1}{\lambda}\ln(u).$$

Damit lautet der Algorithmus zur Erzeugung einer exponentialverteilter Zufallsvariablen X:

1. Erzeuge U wobei U gleichverteilt auf $[0,1]$ ist.
2. Setze $X := -\frac{1}{\lambda}\ln(U)$.

ii Erzeugung stetiger Zufallszahlen - die Verwerfungsmethode

Bei der Inversionsmethode trifft das Problem auf, dass nicht alle Verteilungsfunktionen einfach zum invertieren sind. Ein Beispiel ist die Log Gamma Verteilung. ☹

Im Gegensatz, zur Inversionsmethode handelt es sich bei der Verwerfungsmethode (Acceptance-Rejection Method) um eine indirekte Methode zur Erzeugung von Zufallszahlen einer gewünschten Verteilung. Zunächst wird eine Zufallszahl gemäß einer leichter zu simulierenden Verteilung erzeugt und dann

gemäß einer bestimmten Entscheidungsregel angenommen oder verworfen. Dabei muss der Verwerfungsmechanismus so funktionieren, dass die schließlich akzeptierten Zufallszahlen der gewünschten Verteilung gehorchen. Das Ziel ist die Simulation einer stetigen Zufallsvariablen X die F als Verteilungsfunktion mit Dichte f besitzt.

Man sucht sich nun eine weitere Verteilung G mit Dichte g, die sich *gut* simulieren lässt und für die es eine Zahl $c > 0$ gibt, sodass gilt

$$f(x) < c \cdot g(x)$$

für alle Werte von $x \in \mathbb{R}$. Im ersten Schritt wird nun eine Zufallsvariable Y mit Verteilungsfunktion G erzeugt. Diese Zufallsvariable wird nun in einen zweiten Schritt mit Wahrscheinlichkeit $f(Y)/cg(Y)$ als Realisierung von X akzeptiert oder mit Wahrscheinlichkeit $1 - f(Y)/cg(Y)$ verworfen. Dies kann erreicht werden, indem zunächst eine gleichverteilte Zufallszahl U erzeugt wird. Ist die Bedingung $U \leq f(Y)/cg(Y)$ erfüllt, wird die Zahl Y akzeptiert, ansonsten wird sie verworfen. Als Algorithmus stellt sich die Methode wie folgt dar:

Algorithmus zur Erzeugung einer Realisierung von X mit Verteilungsfunktion f gemäß der Verwerfungsmethode:

1. Erzeuge Y mit Verteilungsfunktion G;
2. Erzeuge U, wobei U gleichverteilt auf $[0, 1]$ ist;
3. Wenn $U < f(Y)/cg(Y)$ ist, setze $X := Y$. Ansonsten gehe zu Schritt 1.

Die resultierende Zufallszahl besitzt tatsächlich die geforderte Verteilung. Man kann zeigen, dass die Anzahl der Iterationen, die notwendig ist, um eine Realisierung von X zu erzeugen, eine geometrisch verteilt Zufallsvariable mit Parameter c ist. Das heisst im Mittel werden c Simulationen benötigt, um eine Simulation von X zu erzeugen. Daher sollte c möglichst klein, d.h. nahe bei eins, gewählt werden.

Beispiel: Gamma Verteilung Angenommen wir möchten eine Zufallsvariable X, welche *Gamma* verteilt ist mit Parametern $\frac{3}{2}$ und 1 simulieren. Die Dichtefunktion ist gegeben durch

$$f(x) = \frac{1}{\Gamma(\frac{3}{2})} \cdot x^{\frac{1}{2}} \cdot e^{-x} = \frac{\sqrt{\pi}}{2} \cdot \sqrt{x} \cdot e^{-x}, \quad x > 0.$$

Sei nun g die Dichtefunktion der Exponentialverteilung mit Parameter $\frac{2}{3}$, d.h.

$$g(x) = \frac{2}{3} \cdot e^{-\frac{2}{3}x}, \quad x > 0.$$

Dann gilt ja

$$f(x) \leq \frac{3^{\frac{3}{2}}}{\sqrt{2\pi e}} \cdot g(x), \quad x > 0.$$

Damit lautet der Algorithmus:

Algorithmus zur Erzeugung einer Zufallsvariablen X die Gamma verteilt ist mit Parametern $\frac{3}{2}$ und 1 mittels der Verwerfungsmethode:

1. Erzeuge U_1 wobei U_1 gleichverteilt auf $[0,1]$ ist;
2. Setzte $Y = -\frac{3}{2}\ln(U_1)$;
3. Erzeuge U_2, wobei U gleichverteilt auf $[0,1]$ ist;
4. Wenn
$$U_2 < \frac{\sqrt{2e}}{\sqrt{3}}\sqrt{Y}e^{-\frac{1}{3}Y}$$
ist, setze $X := Y$. Ansonsten gehe zu Schritt 1.

iii Erzeugung Normalverteilter Zufallsvariablen

Prinzipiell ist es möglich, normalverteilte Zufallsvariable mit der Inversionsmethode zu erzeugen, dazu muss aber die Quantilenfunktion der Standardnormal-verteilung ausgewertet werden. Da es keine einfache exakte Formel für diese Funktion gibt, muss in der Praxis auf Approximationen zurückgegriffen werden, wodurch es zu Fehlern kommen kann. Eine andere Simulationsmethode ist die *Box-Müller-Transformation*, die durch folgenden Algorithmus aus zwei gleichverteilen Zufallsvariablen U_1 und U_2 zwei unabhängig standardnormalverteilte Zufallsvariablen X_1 und X_2 erzeugt.

Box-Müller-Algorithmus zur Erzeugung zweier unabhängiger standard normalverteilten Zufallsvariablen X_1 und X_2:

1. Erzeuge zwei unabhängige auf $[0,1]$ gleichverteilte Zufallsvariable U_1 und U_2;
2. Setze
$$\begin{aligned} X_1 &= \sqrt{(-2\ln(U_1)}\cdot\cos(2\pi\cdot U_2); \\ X_2 &= \sqrt{(-2\ln(U_1)}\cdot\sin(2\pi\cdot U_2); \end{aligned}$$

iv Erzeugung diskreter Zufallszahlen

In den vorangegangenen Abschnitten wurden Simulationsalgorithmen für stetige Verteilungen vorgestellt. Die Inversionsmethode funktioniert in etwas abgewandelter Form jedoch auch für diskrete Verteilungen.

Ziel hier ist es, eine Zufallsvariable X zu simulieren, die Werte $x_1 < x_2 < \cdots < x_n < \ldots$ mit den Wahrscheinlichkeiten $p_1, p_2, \ldots, p_n, \ldots$ annimmt, d.h.

$$X = \begin{cases} x_1 & \text{mit Wahrscheinlichkeit } p_1, \\ x_2 & \text{mit Wahrscheinlichkeit } p_2, \\ \ldots & \ldots \\ x_n & \text{mit Wahrscheinlichkeit } p_n, \\ \ldots & \ldots \end{cases}$$

Genauso wie bei der Inversionsmethode mit stetiger Verteilungsfunktion wird im ersten Schritt eine auf $[0,1]$ gleichverteilte Zufallszahl U erzeugt. Die gesuchte Zufallsvariable X wird dann definiert durch

$$X = \begin{cases} x_1 & \text{falls } U < p_1, \\ x_2 & \text{falls } p_1 < U < p_1 + p_2, \\ \dots & \dots \\ x_n & \text{falls } \sum_{i=1}^{n-1} p_i < U < \sum_{i=1}^{n} p_i, \\ \dots & \dots \end{cases}$$

Diese Definition liefert das richtige Ergebnis, denn es ist

$$\mathbb{P}(X = x_n) = \mathbb{P}\left(\sum_{i=1}^{n-1} p_i < U < \sum_{i=1}^{n} p_i\right) = \sum_{i=1}^{n} p_i - \sum_{i=1}^{n-1} p_i = p_n.$$

Beispiel: Poisson Verteilung Es soll eine Zufallsvariable mit Verteilung $Pois(\lambda)$ erzeugt werden. In diesem Fall ist die Menge $\{x_1, x_2, \cdots, x_n, \dots\}$ gleich $\{1, 2, \dots, n-1, \dots\}$ und

$$p_1 = e^{-\lambda}, \quad p_i = \frac{\lambda^{(i-1)}}{(i-1)!} = p_{i-1}\frac{\lambda}{i-1}.$$

Hier gilt also

$$p_i = e^{-\lambda}\frac{\lambda^{(i-1)}}{(i-1)!}.$$

In diesem Fall lassen sich die Wahrscheinlichkeiten p_i, $i = 1, \dots,$ durch eine Rekursion darstellen: $p_1 = e^{-\lambda}$, $p_i = p_{i-1}\frac{\lambda}{i-1}$.

Algorithmus zur Erzeugung von Poisson verteilten Zufallsvariablen mit Parameter λ:

1. Erzeuge eine auf $[0, 1]$ gleichverteilte Zufallsvariable U;
2. Setze $i := 0$, $p := e^{-\lambda}$, $F := p$;
3. Ist $U < F$, setze $x = i$ und stoppe;
4. Setze $p = \frac{\lambda}{i+1} \cdot p$, $F := F + p$, $i = i + 1$;
5. Erzeuge eine auf $[0, 1]$ gleichverteilte Zufallsvariable U;
6. Gehe zu 3.

3 Ruinwahrscheinlichkeiten im Cramer Lundberg Modell

Als wichtige Anwendung des Poisson Prozesses stellen wir hier das *Cramer Lundberg Modell* vor. Wir nehmen also an, dass $N = \{N(t) : 0 \leq t\}$ ein homogener Poisson Prozess mit intensität λ ist, und $\{X_i : i \in \mathbb{N}\}$ ein Portfolie unabhängig, identisch verteilter Zufallsvariablen mit Verteilungsfunktion F ist. Die Prämie ist mit a bezeichnet, dass Startkapital wird mit u bezeichnet. Die ZUufallsvariable $R(T)$ ist nun folgendermaßen aufgebaut:

$$R(T) := -u - at + \begin{cases} 0 & \text{falls } N(T) = 0, \\ \sum_{k=1}^{N(T)} x_k & \text{falls } N(T) > 0, \end{cases}$$

interessiert. Wir interessieren und für den Wert ob das Ereignis $R(T) \geq 0$ zum Zeitpunkt $T = 1$ eintritt oder nicht. Dass heißt wir sind an der Größe $c := \mathbb{P}(R(T) > 0) = \mathbb{E}1_{(\infty,0)}(R(T))$ interessiert.

Da $T = 1$ fix vorgegeben ist, genügt die Anzahl der Schadensfälle $N(T)$ bis zum Zeitpunkt T einer Poisson Verteilung mit Parameter λT. Wie man Anzahl $N(T)$ simuliert, wurde schon beschrieben. Das heißt, um $R(T)$ zu berechnen muss man zuerst $N(T)$ simulieren und dann $N(T)$ verschiedene Zufallsvariablen mit Verteilung F generieren.

In unseren Beispiel nehmen wir an, dass die $\{X_i : i \in \mathbb{N}\}$ einer Großschädenverteilung genügen, dass heißt Pareto verteilt sind mit Parametern $\alpha, \kappa > 0$. Das heißt,

$$F(x) = \begin{cases} 1 - (\kappa/(\kappa + x))^{\alpha}, & \text{falls} \quad x > 0, \\ 0, \quad \text{sonst.} \end{cases}$$

Eine kurze Rechnung zeigt, dass F leicht zu invertieren ist, und es gilt

$$F^{\leftarrow}(u) = \kappa - \frac{\kappa}{(1-u)^{\frac{1}{\alpha}}}, \quad u \in (0,1).$$

Algorithmus zur Erzeugung des Wertes $R(T)$:

1. setzte $R = -u$;
2. Erzeuge eine Poissonverteilte Zufallsvariable N;
3. Falls $N = 0$ gilt, setzt $R = R - aT$ und breche ab;
4. Falls $N > 0$ gilt, erzeuge N auf $[0,1]$ gleichverteilte Zufallsvariable $\{U_1, \ldots, U_N\}$;
5. Setzte $X_i = F^{\leftarrow}(U_i)$, $i = 1, \ldots, , N$;
6. Setzte $R = R + \sum_{i=1}^{N} X_i - aT$;

6

Um die Wahrscheinlichkeit zu berechnen ob $R(T)$ positiv ist, ist ein Sample $\{R_k : k = 1, \ldots, n\}$ zu berechnen. Es gilt dann

$$\mathbb{P}\left(R(T) > 0\right) \sim c_n := \sum_{k=1}^{n} 1_{(0,\infty)}(R_k).$$

Sind sie an $\mathbb{E}R(T)$ selber interessiert, müssen Sie folgendes beachten. Da die Pareto Verteilung langsam variierend mit Exponent κ ist, hängt die Konvergenzgeschwindikeit, bzw. das Konfidenzinterval von κ ab. Ist $\kappa > 2$, ist das zweite Moment endlich und das schwache Gesetzt der großen Zahlen kann angewendet werden um die Konfidenzintervalle zu berechnen.

Ist $\kappa < 2$, ist das zweite Moment nicht endlich ! Aus folgenden Überlegungen können Sie sich einen Richtwert für die Sample-Größe ableiten.

Seinen $\{X_1, \ldots, X_n\}$ Zufallsvariablen mit langsam variierender Tail Funktion mit Exponent κ. Sei $S_n = \sum_{i=1}^n X_i$ und wir sind an $c = \mathbb{E}X_1$ interessiert. Soll das Konfidenzintervall mit Sicherheit $1-\delta$ eingehalten sein, setzten Sie

$$d = \delta^{-\frac{1}{\kappa}} \quad \text{oder auch } d = F_\kappa^{\leftarrow}(\delta).$$

Nun gilt ja nach Satz 9 dass

$$\frac{1}{n^{\frac{1}{\kappa}}}(S_n - nc) \sim F_\kappa,$$

wobei F_κ eine κ stabile Verteilung inne hat. Damit gilt

$$\mathbb{P}\left(\left|\frac{1}{n^{\frac{1}{\kappa}}}(S_n - nc)\right| \leq d\right) \sim 1 - \delta.$$

Nach Normierung erhält man dass

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - c\right| \leq dn^{\frac{1}{\kappa}-1}\right) \sim 1 - \delta.$$

Dies ist aber eine sehr grobe Abschätzung. Im Falle einer Normalverteilung, würde ca. gelten

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - c\right| \leq xn^{-\frac{1}{2}}\right) \sim 1 - \delta.$$

wobei $x = \Phi^{\leftarrow}(1-\delta)$ ist.

Im vorigen Beispiel haben wir berechnet wie gross die Wahrscheinlichkeit ist, dass $R = \{R(t) : 0 \leq t < \infty\}$ zu einem festen Zeitpunkt $T = 1$ positiv ist. Möchte man wissen ob während der Zeit $[0, T]$ der Risikoprozess positiv wird, ist man an folgender Größe interessiert

$$c = \mathbb{P}\left(\min_{0 \leq t \leq T} R(t) < 0\right).$$

Das heißt man muss den gesamten Prozessverlauf $[0, T] \ni t \mapsto R(t)$ simulieren. Hier kann man aber auf folgende Konstruktion des Risikoprozesses zurückgreifen:

Sei a die Prämie pro Zeiteinheit, $\{\tau_i : i \in \mathbb{N}\}$ eine Familie von unabhängigen exponentialverteilten Zufallsvariabeln mit Parameter λ und sei $\{X_i : i \in \mathbb{N}\}$ eine Familie von unabhängiger Zufallsvariabeln, welche einer bestimmten Verteilung F genügen. Sei $T_i = \sum_{k=1}^i \tau_i$, $i \in \mathbb{N}$.

Der Risikoprozess $[0, T] \ni t \mapsto R(t)$ im Zeitintervall $[0, T]$ ist gegeben durch

$$R(t) = -u - a \cdot t + \begin{cases} 0, & \text{falls} \quad \tau_1 > t, \\ \sum_{i=1}^\infty 1_{[T_i, \infty)}(t)\, X_i, & \text{sonst.} \end{cases}$$

Sei nun $X = 1$ falls $\min_{0 \leq t \leq T} R(t) > 0$ und $X = 0$ falls $\min_{0 \leq t \leq T} R(t) \leq 0$.

Algorithmus zur Erzeugung des Wertes X:

1. Setzte $R = u$, $X = 0$ und $\tau_0 = 0$;
2. Erzeuge eine exponentialverteilte Zufallsvariable τ mit Parameter λ;
3. Falls $\tau + \tau_0 > T$ gilt, breche ab;
4. Falls $\tau + \tau_0 \leq T$ gilt, erzeuge eine F verteilte Zufallsvariable V;
5. Setzte $R = R - a\tau + V$;
6. Falls $R < 0$ gilt, setze $X = 1$ und breche ab;
7. Falls $R \geq 0$ gilt, setze $\tau_0 = \tau + \tau_0$ und gehe zu Punkt (2);

Kapitel 5

Ausgewählte Kapitel aus den Statistische Methoden

1 Extremwert Verteilungen und *worst case Szenarien*

Man spricht von einer 100 Jahr Flut, falls durchschnittlich jede hundert Jahre eine Flut mit mindestens dieser Höhe passiert. Allgemein, sei Y das Maximum der Flut in einem Jahr. Gegeben ein Level $x \geq 0$. Ein Ereignis $A = \{Y > x\}$ heißt $T(x)$ Jahres Ereignis falls es durchschnittlich einmal in $T(x)$ Jahren auftritt, d.h.

$$T(x) = \frac{1}{\mathbb{P}(Y > x)}.$$

Wird, z.B. ein Deich in den Niederlanden oder Belgien geplant, so werden die Deiche so ausgelegt, dass sie eine 10^4 -Jahres Flut aushalten. Man gibt sich also $T = T(x)$ vor und ist an x interessiert. Das heißt man ist an der Quantilen Funktion für Werte sehr nahe an 1 interessiert. Hier tritt aber ein Problem auf. Meistens passiert so eine Flut eben alle 10^4 Jahre, dass heißt solche Ereignisse sind extrem selten. Trotzdem möchte man die Wahrscheinlichkeit so eines Ereignisses abschätzen. Um dies zu tun ist die Extremwerttheorie innerhalb der Wahrscheinlichkeitsrechnung ein wichtiges Instrument. Die Extremwerttheorie beschäftigt sich mit folgender Fragestellung:

Sei $\{X_n : n \in \mathbb{N}\}$ eine Familie unabhängiger identisch verteilter Zufallsvariablen. Sei

$$M_n = \max_{1 \leq i \leq n} X_i$$

Wie ist M_n verteilt ?

Beispiel 1.1. *Man möchte die Wahrscheinlichkeit berechnen, dass innerhalb eines Jahre ein Zugseil reißt. Man weiß dass bei einer bestimmten Last, dieses Zugseil reißen kann. Die Last an einen Tag sei eine normalverteilte Zufallsvariable. Beobachtet man z.B. die Lastenverteilung eines Jahres für jeden Tag, so erhält man eine Familie von unabhängigen normalverteilten Zufallsvariable* $\{X_n : n = 1, \ldots, 365\}$. *Die Frage stellt sich jetzt, wie ist das Maximum*

$$M_{365} = \max_{1 \leq n \leq 365} X_n$$

verteilt. Im allgemeinen kann man M_{365} *mittels einer Extremwertverteilung approximieren.*

Solche Fragestellen tauchen in der Hydrologie (Abschätzen des höchsten Wasserstandes in den nächsten Jahren), Meterologie (Maximale Windgeschwindigkeit, Temperaturmaximum), Lastenberechnungen, Versicherungen (bzw. Rückversicherer), etc.... auf.

Hier in diesem Kapitel werden wir einige Grenzverteilungen, sogenannte Extremwertverteilungen, vorstellen und auch die Frage beantworten, wie $\{X_i : i = 1, \ldots, n\}$, verteilt sein müssen damit M_n für $n \to \infty$ gegen eine bestimmte Grenzverteilung zu konvergieren.

Angenommen die Famile $\{X_n : n \in \mathbb{N}\}$ von Zufallsvariablen hat Verteilungsfunktion F. Dann ist ein naiver Ansatz

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \cdots, X_n \leq x) = \mathbb{P}(X_1 \leq x) \cdots \mathbb{P}(X_n \leq x) = [F(x)]^n.$$

Wir werden uns für die Eigenschaften von M_n für große Werte von n interessieren. Im folgenden Satz berechnen wir den Wert, gegen den die Zufallsvariable M_n für $n \to \infty$ in Wahrscheinlichkeit konvergiert. Wir definieren den rechten Endpunkt der Verteilungsfunktion F durch $x_R = \sup\{t \in \mathbb{R} : F(t) < 1\}$. Hier, zwei Beispiele wo einmal der rechte Endpunkt endlich ist, d.h. 1 und einmal der rechte Endpunkt unendlich ist: Im ersten Bild bei Figure 5.5 ist die Verteilungsfunktion der β–Verteilung mit Parametern $p = 2$ und $q = 3$, i.e.

$$F(x) = \frac{1_{[0,1]}(x)}{B(p,q)} \int_0^x u^{p-1}(1-u)^{q-1}\, du$$

gegeben. Hier ist der rechte Endpunkt 1. Im zweiten Bild bei Figure 5.5 ist die Verteilungsfunktion einer Standard Normalverteilung gegeben. Hier ist der rechte Endpunkt $x_R = \infty$.

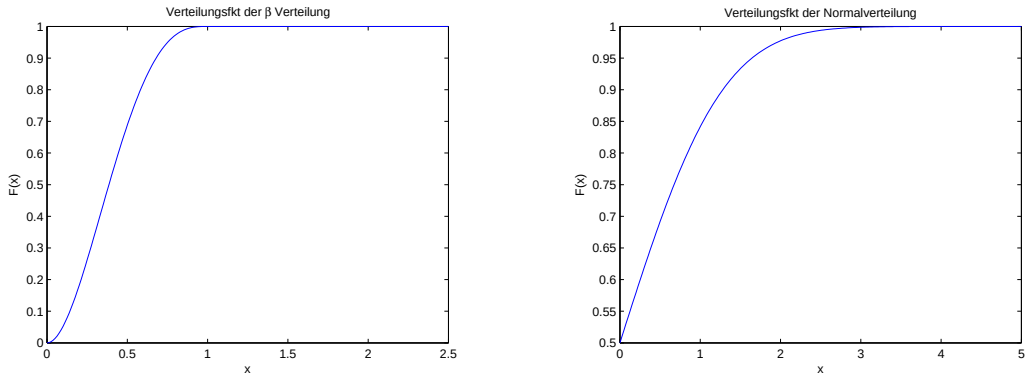


Abbildung 5.1: Die Beta und Normalverteilung.

Man kann nun zeigen dass folgendes gilt:

Satz 2. *Für $n \to \infty$ konvergiert die Zufallsvariable M_n in Wahrscheinlichkeit gegen den Wert x_R. Anders*

$$\mathbb{P}(M_n \to x_R) = 1.$$

Das Problem mit diesem Satz ist, dass die Konvergenz zu stark ist um welche Aussagen machen zu können. Erinnern wir uns an das Gesetz der großen Zahlen. Hier gibt es zwei Versionen, einmal das starke Gesetz der großen Zahlen, und das schwache Gesetz der großen Zahlen. Beim starken Gesetz der großen Zahlen betrachten wir den Grenzwert von

$$\frac{X_1 + X_2 + \ldots X_{n-1} + X_n}{n}$$

Sind X_i identisch und unabhängig verteilt mit Erwartungswert μ gilt

$$\frac{X_1 + X_2 + \ldots X_{n-1} + X_n}{n} \to \mu.$$

Beim schwachen Gesetz der großen Zahlen, skaliert man mit $\sqrt{n}$ und betrachtet den Grenzwert von

$$\frac{X_1 + X_2 + \ldots X_{n-1} + X_n}{\sqrt{n}}$$

Sind X_i identisch und unabhängig verteilt mit Erwartungswert μ und Varianz σ^2 gilt

$$\mathcal{L}aw\left(\frac{X_1 + X_2 + \ldots X_{n-1} + X_n - n\mu}{\sqrt{n}}\right) \to \mathcal{N}(0, \sigma).$$

Das heißt um vernünftige Aussagen zu treffen, sollte man M_n noch skalieren, d.h. die Verteilung von M_n/c_n für $n \to \infty$ untersuchen. Das ist der Inhalt des Fisher-Tapett Theorems, aber vorher werden noch die verschiedenen Verteilungstypen vorgestellt zu dem der skalierte Maximalwert streben kann. Dabei gibt es drei Typen von Verteilungen mit deren Hilfe M_n abgeschätzt werden kann:

- Eine Zufallsvariable X ist Fréchet verteilt mit Parameter a, falls für die Verteilungsfunktion F_X gilt

$$\Phi_a(x) := F_X(x) = \begin{cases} 0, & \text{falls } x \leq 0, \\ \exp(-x^{-a}), & \text{falls } x > 0. \end{cases} \tag{5.1}$$

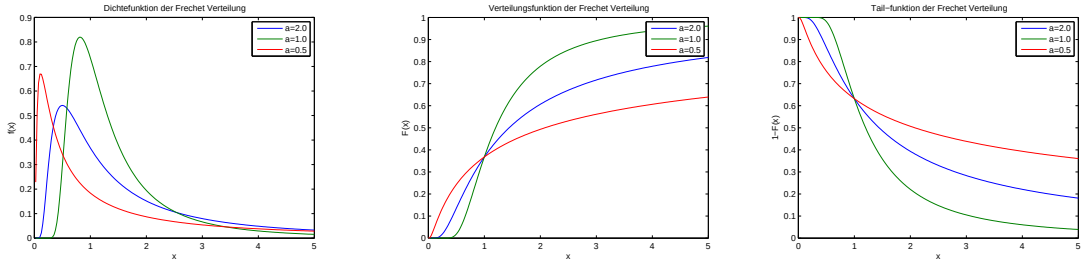


Abbildung 5.2: Die Dichte-, Verteilungs-, und Tailfunktion.

- Eine Zufallsvariable X ist Weibull verteilt mit Parameter a, falls für die Verteilungsfunktion F_X gilt

$$\Psi_\alpha(x) := F_X(x) = \begin{cases} \exp(-(-x)^\alpha), & \text{falls } x \leq 0, \\ 1 & \text{falls } x > 0. \end{cases} \tag{5.2}$$

- Eine Zufallsvariable X ist Gumpel verteilt, falls für die Verteilungsfunktion F_X gilt

$$\Lambda(x) := F_X(x) = \exp(-e^{-x}) \tag{5.3}$$

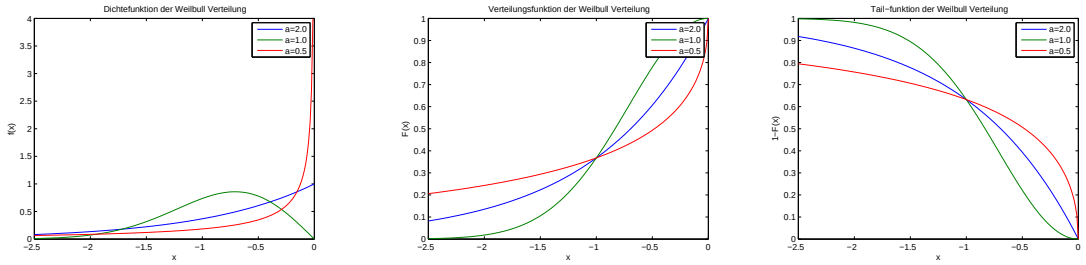


Abbildung 5.3: Die Dichte-, Verteilungs-, und Tailfunktion.

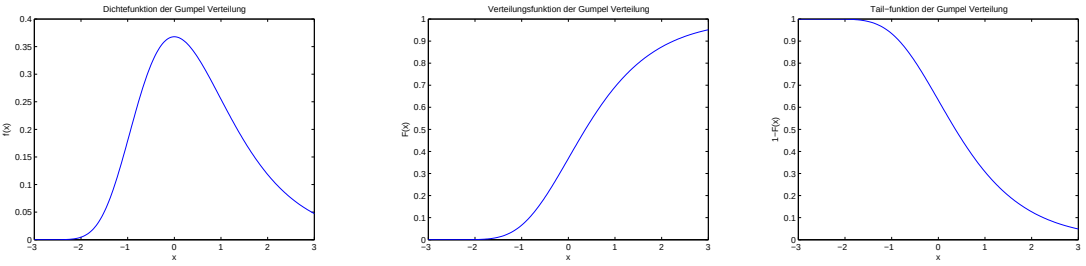


Abbildung 5.4: Die Dichte-, Verteilungs-, und Tailfunktion.

Das Fisher-Tapett Theorem besagt nun folgendes: [1]

Satz 3. *Seien $\{X_n : n \in \mathbb{N}\}$ unabhängige identisch verteilte Zufallsvariabeln sodass gilt: es gibt Konstanten $c_n > 0$ und d_n gibt mit*

$$c_n^{-1}(M_n - d_n) \to H, \quad n \to \infty$$

für eine nicht degenerierte Zufallsvariable H. Dann genügt H einer sogenannten Extremwertverteilung (d.h. Gumpel, Weilbull oder Fréchet Verteilung).

Beispiel 1.2. *Seien $\{X_i : i \in \mathbb{N}\}$ unabhängig und exponential verteilt mit Parameter λ, d.h.*

$$F(x) = 1 - e^{-\lambda x} \quad \text{für} \quad x > 0.$$

Man kann sich z.B. vorstellen, dass die X_i die Zerfallsdauer von radioaktiven Isotopen modellieren. Die Tail–Wahrscheinlichkeit $\bar{F}$ ist gegeben durch $\bar{F}(x) := 1 - F(x) = e^{-\lambda x}$. Damit gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n}{\lambda} - \frac{\log n}{\lambda} < x\right) = e^{-e^{-x}} = \Lambda(x), \quad x \in [0, \infty).$$

Mit anderen Worten konvergiert die Zufallsvariable $(M_n - \log n)/\lambda$ für $n \to \infty$ in Verteilung gegen Λ.

Beispiel 1.3. *Seien $\{X_i : i \in \mathbb{N}\}$ unabhängig und Pareto-verteilt mit Parameter $a > 0$, d.h.*

$$F(x) = \begin{cases} 1 - \frac{1}{x^a} & \text{für } x > 1, \\ 0 & \text{für } x \leq 1. \end{cases}$$

[1] Wir sprechen von nicht degenerierten Zufallsvariablen, falls H nicht ein fester Punkt ist.

Also, $\bar{F}(x) := x^{-a}$. Dann gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n}{n^{\frac{1}{a}}} < x\right) = \mathbb{P}\left(M_n < x n^{\frac{1}{a}}\right) = \begin{cases} e^{-x^{\frac{1}{a}}} & \text{für } x > 0, \\ 0 & \text{für } x \le 0. \end{cases}$$

Mit anderen Worten konvergiert die Zufallsvariable $n^{-\frac{1}{a}} M_n$ in Verteilung gegen Φ_a.

Man kann das Ergebnis wie folgt interpretieren: Für großes n nimmt das Maximum M_n sehr große Werte auf der Skala $n^{1/a}$ an. Reskaliert man M_n mit dem Faktor $n^{1/a}$, so erhält man approximativ Frechet-verteilte Werte.

Beispiel 1.4. *Die $\{X_i : i \in \mathbb{N}\}$ haben die Verteilungsfunktion*

$$F(x) = \begin{cases} 0 & \text{für } x \le 0, \\ 1 - (1-x)^{\alpha} & \text{für } 0 \le x < 1, \\ 1 & \text{für } x \ge 1. \end{cases}$$

wobei $\alpha > 0$ ein Parameter ist. Also, $\bar{F}(x) := (x_R - x)^{\alpha}$, $x_R = 1$, $x \in (0, x_R)$. Dann gilt

$$\lim_{n\to\infty} \mathbb{P}\left((M_n - 1)\, n^{\frac{1}{a}} < x\right) = \begin{cases} e^{-(-x)^{\alpha}} & \text{für } x \le 0, \\ 1 & \text{für } x \ge 0, \end{cases}$$

Mit anderen Worten konvergiert die Zufallsvariable $n^{\frac{1}{\alpha}}(M_n - 1)$ in Verteilung gegen Ψ_α.

Man kann das Ergebnis wie folgt interpretieren: Für großes n nähert sich das Maximum M_n dem Wert 1 von unten an. Die Differenz $M_n - 1$ nimmt sehr kleine negative Werte auf der Skala $n^{-1/\alpha}$ an. Reskaliert man $(M_n - 1)$ mit dem Faktor $n^{1/\alpha}$, so erhält man approximativ Weibull-verteilte Werte.

Bemerkung 1.1. *Eine Frechet-verteilte Zufallsvariable nimmt nur positive Werte an. Eine Weibull-verteilte Zufallsvariable nimmt nur negative Werte an. Eine Gumbel–verteilte Zufallsvariable kann sowohl positive als auch negative Werte annehmen.*

Bemerkung 1.2. *Es kommt oft vor dass man die Verteilung des Minimum eines Samples wissen möchte. Z.B. besteht eine Maschine aus n Bauteilen gleicher Art. Geht eines der Bauteile kaputt, so funktioniert die Maschine nicht mehr. Angenommen wir kennen die Lebensdauer der einzelnen Bauteile, d.h. wir kennen die Verteilung von $\{T_1, T_2, \ldots, T_n\}$ wo T_i die Lebensdauer der Maschine i bezeichnet. Die Lebensdauer der Maschine T is die Verteilung des Minimums, i.e. $T = \min(T_1, \ldots, T_n)$. Sie nun $\tilde{T}_i = -T_i$ und $\tilde{T} = \max(\tilde{T}_1, \ldots, \tilde{T}_n)$. Dann gilt $T = -\tilde{T}$. Das heißt, kennen wir die Verteilung des Maximums von $\{\tilde{T}_1, \ldots, \tilde{T}_n\}$, so kennen wir die Verteilung von T.*

2 Maximale Anziehungsbereiche

Neben seiner technischen Bedeutung beantwortet dieser Satz auch die Frage nach der Eindeutigkeit der Grenzverteilung normierter Maxima. Existiert für geeignet normierte Maxima $(M_n - d_n)/c_n$ von iid Zufallsvariablen eine nicht-entartete Grenzverteilung, so ist diese bis auf Typ-Gleichheit auch eindeutig bestimmt. Darüber hinaus charakterisiert dieser Satz sämtliche Konstantenfolgen $c_n > 0$ und d_n, für die $F^n(x/c_n + d_n)$ gegen eine nicht-entartete Verteilungsfunktion konvergiert in Abhängigkeit von einem schon gefundenen Paar (c_n), $c_n > 0$ und (d_n). Noch zu klären ist gegen welche Verteilungsfunktion die normierten Maxima konvergieren.

Seien $\{X_i : i \in \mathbb{N}\}$ unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion F. Wir sagen, dass F im Max-Anziehungsbereich einer Verteilungsfunktion H liegt, falls es reellwertige Folgen $c_n > 0$ und d_n gibt mit

$$c_n^{-1}(M_n - d_n) \to H, \quad n \to \infty$$

Im Folgenden werden wir die Maximalen-Anziehungsbereiche der Verteilungen Φ, Ψ und Λ beschreiben.

i Max-Anziehungsbereich der Frechet-Verteilung

Um die Konvergenz genauer zu untersuchen, also um c_n, d_n und die Verteilung von H näher zu spezifizieren, muss man das links auslaufende Ende der Verteilung charakertisieren. Dazu folgende Definition.

Definition 2.1. *Sei F eine Verteilung der Zufallsvariablen X, dann bezeichnen wir mit $\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$.*

Wir haben bereits gesehen, dass in Beispiel 1.2 die Verteilung mit der Tailfunktion $\bar{F}(x) = 1-F(x) = x^{-a}, x > 1$, im Max-Anziehungsbereich von Ψ_a liegt. Wir werden zeigen, dass der Anziehungsbereich von aus allen Verteilungen besteht, deren Tailfunktionen wo sich im gewissen Sinne wie x^{-a} verhalten. Der entsprechende Begriff ist die regulare Variation.

Definition 2.2. *Eine Funktion $L : [0, \infty) \to [0, \infty)$ heißt* langsam variierend *(slowly varying (at infinity)) falls für alle $a > 0$, gilt*

$$\lim_{x\to\infty} \frac{L(x)}{L(ax)} = 1.$$

Falls der Grenzwert

$$\lim_{x\to\infty} \frac{L(x)}{L(ax)}$$

endlich, aber ungleich eins ist, heißt die Funktion regulär variierend *(regularly varying function).*

Eine Funktion $L : [0, \infty) \to [0, \infty)$ heißt langsam variierend *(slowly varying (at infinity)) mit index α falls für alle $a > 0$, gilt*

$$\lim_{x\to\infty} \frac{L(x)}{L(ax)} = a^{\alpha}.$$

Beispiel 2.1.

- *Die Funktion $f(x) = cx^a$, wobei $c > 0$, ist regulär variierend mit Index a.*
- *Die Funktion $f(x) = cx^a(\log x)\beta$, mit $c > 0$, $\beta \in \mathbb{R}$ ist regulär variierend mit Index a.*
- *Die Funktion $f(x) = (\log x)^{\beta}$ mit $\beta \in \mathbb{R}$ ist langsam variierend.*

Ist F streng monoton steigend, so ist F invertierbar und F^{-1} eindeutig definiert. Da aber F nicht unbedingt streng monoton steigen ist, aber trotzdem oft an den Inversen interessiert ist, definiert man sich das sogenannte Pseudoinverse durch

$$F^{\leftarrow}(x) = \inf_{x\in\mathbb{R}}\{F(x) \geq x\}.$$

Satz 4 (Konvergenz gegen die Fréchet Verteilung:)**.** *Eine Verteilungsfunktion F mit rechtem Endpunkt x_R liegt im Max-Anziehungsbereich der Frechet-Verteilung mit $a > 0$ genau dann, wenn folgende zwei Bedingungen erfüllt sind:*

- *$x_R = \infty$;*
- *Die Tailfunktion F ist regulär variierend mit Index $-a$, d.h.*

$$\lim_{x\to\infty} \frac{1 - F(\lambda x)}{1 - F(x)} = \lambda^{-a} \quad \text{für alle } \lambda > 0.$$

Gilt $(1 - F(x)) = x^{-\alpha}L(x)$ wobei L eine langsam variierende Funktion ist, dann gibt es eine Folge c_n mit

$$c_n^{-1} M_n \xrightarrow{d} \Phi_{\alpha}.$$

In der Praxis ist es zumeist noch wichtig c_n zu bestimmen. Hier kann man zeigen, dass

$$c_n = F^{\leftarrow}(1 - n^{-1}) = \inf_{x\in\mathbb{R}}\{F(x) \geq 1 - n^{-1}\}.$$

Beispiele

1. **Maxima einer Cauchy verteilten Zufallsvariablen:** Sei $\{X_n : n \in \mathbb{N}\}$ eine Familie von unabhängigen Cauchy verteilten Zufallsvariablen mit Dichtefunktion

$$f(x) = (\pi(1+x^2))^{-1}, \quad x \in \mathbb{R}.$$

Dann gilt $1 - F(x) \sim (\pi x)^{-1}$ und daher

$$\begin{aligned} \lim_{x\to\infty} \frac{1-F(x)}{(\pi x)^{-1}} &= \lim_{x\to\infty} \frac{f(x)}{\pi^{-1}x^{-2}} \\ \lim_{x\to\infty} \frac{\pi x^2}{\pi(1+x^2)} &= 1. \end{aligned}$$

Das heißt

$$\begin{aligned} \mathbb{P}\left(M_n \le \frac{nx}{\pi}\right) &= \left(1-(1-F(\frac{nx}{\pi}))\right)^n \\ = \left(1-\frac{1}{n}\left(\frac{1}{x}+o(1)\right)\right)^n &\to \exp(-x^{-1}). \end{aligned}$$

2. **Maxima der Pareto, Burr und stabile Verteilungen mit index $\alpha < 2$:** Genügt X einer der oben genannten Verteilungen, gilt $\bar{F}(x) = \mathbb{P}(X > x) \sim Kx^{-\alpha}$ mit $\alpha > 0$ und deswegen, $c_n = (Kn)^{\frac{1}{\alpha}}$.

3. **Maxima der Loggamma Verteilung:** Die Dichte der Loggamma verteilung ist gegeben durch

$$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha-1}, \quad x \ge 0, \quad \alpha,\beta > 0.$$

Eine kurze Rechnung zeigt, dass gilt

$$c_n = \left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{\frac{1}{\alpha}}.$$

ii Maximale Anziehungsbereich der Weilbull Verteilung:

Der Maximale Anziehungsbereich der Weibull-Verteilung hat eine ähnliche Charakterisierung wie der Maximale Anziehungsbereich der Frechet-Verteilung. Der Unterschied ist, dass im Fall der Weibull-Verteilung der rechte Endpunkt x_R endlich sein muss. Damit eine Verteilungsfunktion F im Maximale Anziehungsbereich von liegt, muss F in x_R im gewissen Sinne regulär variierend sein. Wir geben nun eine präzise Definition.

Definition 2.3. *Eine Funktion $L : (0,1] \to (0,\infty)$ heißt* langsam variierend in 0 *(slowly varying (at 0)) mit Index $\alpha > 0$ falls für alle $\lambda > 0$, gilt*

$$\lim_{x\to 0} \frac{L(x)}{L(\lambda x)} = \lambda^\alpha \quad \textit{für alle } \lambda > 0.$$

Beispiel 2.2. *Die Funktion $f(x) = x^a$ ist regulär variierend in 0 mit Index a.*

Satz 5 (Konvergenz gegen die Weibull Verteilung:)**.** *Eine Verteilungsfunktion F mit rechtem Endpunkt x_R liegt im Maximalen Anziehungsbereich der Weibull-Verteilung Ψ_α, dann und nur dann, wenn folgende zwei Bedingungen erfüllt sind:*

- $x_R < \infty$;

- *Die Funktion* $1 - F(x_R - x)$ *ist regulär variierend mit Index* α *in* 0*, d.h.*

$$\lim_{x\to\infty} \frac{1 - F(\lambda(x_R - x))}{1 - F(x_R - x)} = \lambda^\alpha \quad \textit{für alle } \lambda > 0.$$

Gilt $1 - F(x_R - x^{-1}) = x^{-\alpha}L(x)$ wobei L eine langsam variierende Funktion ist, dann gilt

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} \Phi_\alpha.$$

wobei $d_n = x_R$ und

$$c_n = x_R - \inf_{x\in\mathbb{R}}\{F(x)) \geq 1 - n^{-1}\}.$$

Beispiele

- **Gleichverteilung auf** $[0, 1]$: Hier gilt $c_n = n^{-1}$ und $d_n = 1$.
- **Power law behaviour at** x_R**:** Gilt $\bar{F}(x) = K(x_R - x)^\alpha$, für $x_R - K^{-\frac{1}{\alpha}} \leq x \leq x_R$, mit $K, \alpha > 0$, dann gilt $c_n = (Kn)^{-\frac{1}{\alpha}}$ und $d_n = x_R$.
- **Beta Verteilung:** Die Dichte der Beta Verteilung lautet

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}, \quad 0 \leq x \leq 1.$$

Hier gilt

$$c_n = \left(n\,\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(a+b)}\right)^{-\frac{1}{b}},$$

und $d_n = 1$.

iii Konvergenz gegen die Gumpel Verteilung:

Der Einzugsbereich der Gumpel Verteilung als Extremwertverteilung läuft von der Lognormalverteilung, der Exponential–Verteilung mit $\bar{F}(x) \sim e^{-x}$ bis zur Normalverteilung mit $\bar{F}(x) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/2}$ für $t \to \infty$ (Regel von l'Hospital).

Satz 6 (Konvergenz gegen die Gumpel Verteilung:)**.** *Eine Verteilungsfunktion* F *mit rechtem Endpunkt* x_R *liegt im Maximal Anziehungsbereich der Gumbel-Verteilung* $G(x) = e^{-e^{-x}}$ *genau dann, wenn es eine positive und messbare Funktion* $g(t)$ *gibt mit*

$$\lim_{x\to x_R} \frac{\bar{F}(x + x\,g(x))}{\bar{F}(x)} = e^{-x} \quad \textit{für alle } x \in \mathbb{R}.$$

Bemerkung 2.1. *x_R kann endlich oder unendlich sein.*

Bemerkung 2.2. *Die Standardnormalverteilung liegt im Anziehungsbereich der Gumpelverteilung. Man kann zeigen dass man Satz 6 anwenden kann mit* $g(t) = \frac{1}{t}$*. Genauer*

$$\begin{aligned}\lim_{t\to\infty} \frac{\bar{F}(t + xg(t))}{\bar{F}(t)} &= \lim_{t\to\infty} \frac{\frac{1}{t+xg(t)}e^{-(t+xg(t))^2/2}}{\frac{1}{t}e^{-t^2/2}} \\ = \lim_{t\to\infty} \frac{\frac{1}{t+x/t}e^{-t^2/2-x-x^2/(2t^2)}}{\frac{1}{t}e^{-t^2/2}} &= e^{-x}.\end{aligned}$$

Um die Verteilungen näher zu charakterisieren, bzw. c_n und d_n zu finden, ist es wichtig die Funktion a näher zu charakterisieren. Dazu führen wir als erstes die *von Mises Funktion* ein.

Definition 2.4. *Sei f eine Dichte Funktion mit rechten Endpunkt $x_R \leq \infty$. Angenommen, es existiert eine positive stetige Funktion $a : [0, \infty)$ und eine positive Konstante mit*

$$\bar{F}(x) = c \exp\left\{ - \int_z^x \frac{1}{a(t)}\, dt \right\}, \quad z < x < x_R$$

dann heißt F von Mises Funktion *und a die zu F gehörige Hilfsfunktion.*

Bemerkung 2.3. *In der Literatur wird auch verschiedentlich die* hazard Funktion *h durch*

$$h(x) = \frac{1 - F(x)}{f(x)} = \frac{\bar{F}(x)}{f(x)} \quad \text{für } x_L < x < x_R,$$

definiert. Dabei kann man sich folgenden Zusammenhang ableiten:

$$a(x) \sim h(x).$$

Nach Definition von a gilt

$$\int_x^\infty f(y)\, dy = \exp\left(- \int_\mu^x \frac{1}{a(y)}\, dy \right).$$

Differenzieren ergibt

$$-f(x) = \exp\left(- \int_\mu^x \frac{1}{a(y)}\, dy \right) \frac{-1}{a(x)} \bar{F}(x)\, \frac{-1}{a(x)}.$$

Es gilt somit

$$a(x) = \frac{\bar{F}(x)}{f(x)} \text{ für } x \to \infty.$$

Von Mises Funktionen gehören zu dem Einzugsbereich der Gumpel Verteilung. Folgenden Beispiele sind von Mises Funktionen:

- **Maxima von exponential verteilten Zufallsvariablen:** Sei $\{X_n : n \in \mathbb{N}\}$ eine Familie von unabhängigen exponential verteilten Zufallsvariablen. Hier gilt $\bar{F}(x) = e^{-\lambda x}$, $x \geq 0$, und $\lambda > 0$. Die Hilfsfunktion a lautet $a(x) = \lambda^{-1}$. Damit gilt

$$\begin{aligned} \mathbb{P}\left(M_n - \ln(n) \leq x\right) &= \left(\mathbb{P}\left(X \leq x + \ln(n)\right)\right)^n \\ = (1 - n^{-1} e^{-x})^n &\to \exp(-e^{-x}). \end{aligned}$$

- **Maxima von Weibull verteilten Zufallsvariablen:** (Beim Fisher Tippet Theorem wird die Weibullverteilung gespiegelt und das Maximum genommen, der rechte Endpunkt ist also 0. Hier wird dan Maximum der Weibullverteilung selbst genommen, der rechte Endpunkt ist also ∞.) Hier gilt

$$\bar{F}(x) \sim K x^\alpha \exp(-c x^r),$$

  mit $K, c, r > 0$ und $\alpha \in \mathbb{R}$. Somit gilt $\bar{F}(x) = \exp(-cx^r)$, $x \geq 0$, und $c, r > 0$. Die Hilfsfunktion a lautet $a(x) = c^{-1} r^{-1} x^{1-r}$. Oben eingesetzt ergibt

$$c_n = (cr)^{-1} \left(c^{-1} \ln n\right)^{\frac{1}{r} - 1},$$

  und

$$d_n = \left(c^{-1} \ln n\right)^{\frac{1}{r}} + \frac{1}{r} \left(c^{-1} \ln n\right)^{\frac{1}{r} - 1} \left\{ \frac{\alpha}{cr} \left(c^{-1} \ln n\right) + \frac{\ln(K)}{c} \right\}.$$

- **Maxima von Gamma verteilten Zufallsvariablen:** Für die Dichtefunktion gilt

$$f(x) = \frac{\lambda^k}{\Gamma(k)} \cdot x^{k-1} \cdot e^{-\lambda x}$$

Daraus ergibt sich nach kurzer Rechnung $c_n = \lambda^{-1}$, und

$$d_n = \lambda^{-1} \left(\ln n + (k-1) \ln \ln n - \ln \Gamma(k) \right).$$

- **Maxima von Erlang verteilten Zufallsvariablen:** Hier gilt

$$\bar{F}(x) = e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}, \quad x \geq 0, \text{ und } \lambda > 0,\, n \in \mathbb{N}.$$

Die Hilfsfunktion a lautet

$$a(x) = \sum_{k=0}^{n-1} \frac{(n-1)!}{(n-k-1)!} \lambda^{(k+1)} x^{-k}, \quad x \geq 0$$

.

Proposition 2.1. *(*Von Mises Funktionen und die Gumpel Verteilung) *Angenommen f ist eine von Mises Funktion. Dann gilt*

$$c_n^{-1}(M_n - d_n) \to \Gamma,$$

wobei

$$d_n = \inf_{x \in \mathbb{R}} \{F(x) \geq 1 - n^{-1}\}$$

und

$$c_n = a(d_n),$$

wobei a die Hilfsfunktion von f ist.

Es gehören nicht nur die von Mises Funktionen zum Einzugsbereich der Gumpel Verteilung sondern auch alle Verteilungen mit ähnlichen Verhalten von $\bar{F}(x)$ für $x \to \infty$, aber wir gehen hier nicht näher ein. Wir werden es aber im nächsten Beispiel benützen.

Beispiel 2.3. Die Normalverteilung:

Satz 7. *Seien $X_1, X_2, \ldots$ unabhängig und standard normalverteilt. Dann gilt für alle $x \in \mathbb{R}$:*

$$\lim_{n \to \infty} \mathbb{P}\left(\sqrt{2\log(n)} \left\{ M_n - \left(\sqrt{2\log(n)} - \frac{\log\log(n) + \log(4\pi)}{2\sqrt{2\log(n)}} \right) \right\} \leq x \right) = e^{-e^{-x}}.$$

Seien x_n normalverteilt mit Mittelwert μ und Varianz σ^2. Der erste Schritt ist die Hilfsfunktion zu finden. Falls die Normalverteilung eine Mises Funktion ist, muss folgende Gleichung gelten

$$\bar{F}(x) = \exp\left(- \int_1^x \frac{1}{a(y)}\, dy \right).$$

Die genaue Dichte eingesetzt gilt

$$\frac{1}{\sigma\sqrt{2\pi}} \int_x^\infty e^{-\frac{(y-\mu)^2}{2\sigma^2\pi}}\, dy = \exp\left(- \int_\mu^x \frac{1}{a(y)}\, dy \right).$$

Differenzieren ergibt

$$-f(x) = -\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2\pi}} = \exp\left(- \int_\mu^x \frac{1}{a(y)}\, dy \right) \frac{-1}{a(x)},$$

bzw.

$$-f(x) = -\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2\pi}} = \bar{F}(x)\frac{-1}{a(x)}.$$

Es gilt somit

$$a(x) = \frac{\bar{F}(x)}{f(x)} = \frac{\int_x^\infty e^{-\frac{(y-\mu)^2}{2\sigma^2\pi}}\, dy}{e^{-\frac{(x-\mu)^2}{2\sigma^2\pi}}}.$$

Weiters müssen wir das Verhalten von $\bar{F}$ für $x \to \infty$ charakterisieren. Die Regel von L'Hospital kann man zeigen dass gilt

$$\lim_{x\to\infty} \frac{\bar{F}(x)}{\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2\pi}}}{\frac{(x-\mu)}{\sigma^2\pi}}} = 1.$$

Angenommen wir haben eine Verteilungsfunktion mit

$$\bar{G}(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2\pi}}}{\frac{(x-\mu)}{\sigma^2\pi}},$$

für $x \to \infty$, dann haben F und G die gleichen Extremwertverteilung mit gleichen c_n und d_n. Damit reicht es für uns $\bar{G}$ zu invertieren (was leichter ist als $\bar{F}$). Wir lösen nun

$$\frac{e^{-\frac{(d_n-\mu)^2}{2\sigma^2\pi}}}{\frac{(d_n-\mu)}{\sigma^2\pi}} = 1 - \frac{1}{n},$$

nach d_n auf. Dies ergibt eine quadratische Gleichung für den Ausdruck $\frac{(d_n-\mu)}{\sigma^2\pi}$, nämlich

$$\left[\frac{(d_n-\mu)}{\sigma^2\pi}\right]^2 + \ln\left[\frac{(d_n-\mu)}{\sigma^2\pi}\right] + \ln(2) = \ln(n),$$

und somit

$$\left[\frac{d_n-\mu}{\sigma^2\pi}\right] = (2\ln n)^{\frac{1}{2}} - \frac{\ln\ln n + \ln 4\pi}{2(2\ln n)^{\frac{1}{2}}} + o\left((\ln n)^{-\frac{1}{2}}\right),$$

also mögliche Werte für d_n. Da gilt

$$a(x) = \frac{\bar{F}(x)}{f(x)} \sim \frac{\sigma^2\pi}{(x-\nu)}$$

erhalten wir

$$c_n = a(d_n) \sim \frac{\sigma^2\pi}{(d_n-\nu)} \sim (2\ln n)^{-\frac{1}{2}}.$$

Es gilt also

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} \Lambda(x).$$

mit obigen d_n und a_n.

Beispiel 2.4. *Angenommen ein Seil wird jeden Tag mit einer Last belastet. Die Last X_n ist normalverteilt mit Mittelwert 10 Tonnen und Varianz 0.25 Tonnen. Wie ist jetzt die Höchstlast in einem Jahr ? Wir wissen also $c_{365} \sim (2\ln 365)^{-\frac{1}{2}}$ und $d_{365} \sim \sigma^2\left((2\ln 365)^{\frac{1}{2}} - \frac{\ln\ln 365+\ln 4\pi}{2(2\ln 365)^{\frac{1}{2}}}\right) + \mu$. Es gilt also $M_n \sim G\, c_{365} + d_{365}$, wobei G Verteilungsfunktion Λ hat.*

Bemerkung 2.4. *Der Satz lässt sich wie folgt interpretieren: das Maximum M_n nimmt Werte an, die sehr nahe bei*

$$b_n = \sqrt{2\log(n)} - \frac{\log\log(n) + \log(4\pi)}{2\sqrt{2\log(n)}}$$

sind. Die Differenz zwischen M_n und b_n ist von der Größenordnung $a_n = 1/\sqrt{2\log n}$. Reskaliert man $M_n - b_n$ mit dem Faktor $\sqrt{2\log n}$, so erhält man eine approximativ Gumbel-verteilte Zufallsvariable.

3 Aussagen über die Konvergenzgeschwindigkeit

Konvergenzgeschwindigkeit normalisierter Extrema unabhängiger Zufallsvariablen Das Fisher-Tippett-Theorem gibt die möglichen Extremwertverteilungen fur richtig normierte Maxima von Folgen von iid Zufallsvariablen an. Um diesen Satz auch praktisch brauchbar zu machen, müssen wir wissen, wie schnell diese Konvergenz ist. d. h. wir müssen folgende Größe

$$|P(M_n = a_n x + b_n) - G(x)|$$

punktweise oder gleichmäßig in x abschätzen.

Satz 8. *Sei $c_n = F^{\leftarrow}(1 - \frac{1}{n})$, $n \geq 2$. Wir definieren die Funktion $r_n(x)$ wie folgt:*

- *Für F im Anziehungsbereich der Gumpel–Verteilung Λ*

$$r_n(x) = n\,(1 - F(c_n + xg(c_n))) - e^{-x}, \quad x \in \mathbb{R},$$

- *Für F im Anziehungsbereich der Frechet–Verteilung Φ_α*

$$r_n(x) = n\,(1 - F(c_n)) - x^{-\alpha}, \quad x > 0,$$

- *Für F im Anziehungsbereich der Weilbull–Verteilung Ψ_α*

$$r_n(x) = n\,(1 - F(x_R + x(X_R - c_n))) - (-x)^{\alpha}, \quad x < 0,$$

Dann gilt für hinreichendes grosses n

- *Für F im Anziehungsbereich der der Gumpel–Verteilung Λ*

$$\begin{aligned} |\mathbb{P}\,(M_n \leq a(c_n)\,x + c_n) - \Lambda(x)| &\leq \frac{0.6}{2n-1} r e^{-e^{-x}} \left| e^{-r_n(x)} - 1 \right| \\ &= \mathcal{O}(\max(\frac{1}{n}, |r_n(x)|), \quad x \in \mathbb{R}, n \to \infty. \end{aligned}$$

  Hier, ist die Funktion a identisch mit der Funktion aus Theorem 6

- *Für F im Anziehungsbereich der Frechet–Verteilung Φ_α*

$$\begin{aligned} |\mathbb{P}\,(M_n \leq x c_n) - \Theta_\alpha(x)| &\leq \frac{0.6}{2n-1} + e^{-x^{-\alpha}} \left| e^{-r_n(x)} - 1 \right| \\ &= \mathcal{O}(\max(\frac{1}{n}, |r_n(x)|), \quad x > 0, n \to \infty. \end{aligned}$$

- *Für F im Anziehungsbereich der Weilbull–Verteilung Ψ_α*

$$\begin{aligned} |\mathbb{P}\,(M_n \leq (x_R - c_n)x + x_R) - \Theta_\alpha(x)| &\leq \frac{0.6}{2n-1} + e^{-x^{-\alpha}} \left| e^{-r_n(x)} - 1 \right| \\ &= \mathcal{O}(\max(\frac{1}{n}, |r_n(x)|), \quad x > 0, n \to \infty. \end{aligned}$$

Als Vorüberlegung verwenden wir
Sei

$$h(x) = \frac{1 - F(x)}{f(x)}$$

für $x_R < x < x_R$. Exponential(1): $F(x) = 1 - exp(-x), f(x) = exp(-x)$ and so $h(x) = 1$ for all $x > 0$ (constant hazard).

Normal verteilung s 37,38 (pareto typ)
Sei

$$e_n = \sup_{x\in\mathbb{R}} |[F(a_n x + b_n)]^n - G(x)|.$$

dann gilt

$$O(e_n) = \max\{O(n^{-1}), h'(b_n) - \xi\}.$$

p. 50

Exponential(1) Previously we found that $h'(x) = 0$ und damit ist die Konvergenzrate and $O(n^{-1}$, also sehr schnell.

N(0,1) Previously we found that $h'(x) = x^2 + 3x^4 + ...\ \xi = 0$, d.h. konvergenzrate ist $O((logn)^{-1})$ also sehr langsam.

4 Statistische Methoden zum Schätzen von seltenen Ereignissen

There is always going to be an element of doubt, as one is extrapolating into areas one doesn't know about. But what Extrem Value Theory is doing is making the best use of whatever data you have about extreme phenomena - Richard Smith.

In der Statistik wird meistens das Hauptaugenmerk auf die Betrachtung von Mittelwerten gelegt. Für viele Bereiche ist das sinnvoll, doch z.B. bei der Investition in eine Aktie, ist es zwar interessant zu wissen wie viel im Durchschnitt an Gewinn oder Verlust zu erwarten ist, aber viel entscheidender ist die Frage nach dem *worst case*. Wie viel kann mit dieser Investition mit welcher Wahrscheinlichkeit im schlimmsten Fall verloren gehen? Auch in vielen anderen Bereichen, wie z.B. in Versicherungen oder bei der Planung von Deichen, ist es entscheidend, sich für den schlimmsten Fall abzusichern. Denn um einen Deich so hoch zu planen, dass im Normalfall keine Überschwemmungen zu befürchten sind, nützt es wenig, die durchschnittliche Wasserhöhe zu kennen. Stattdessen müssen die besonders hohen Wasserstände, oder im Fall der Versicherung die besonders hohen Versicherungsrisiken, abgeschätzt werden. Mit dieser Frage beschäftigt sich die Extremwerttheorie. Dabei werden, wie der Name schon sagt, hauptsächlich die extremen Werte, sogenannte *Ausreißer* betrachtet. Doch eine Abschätzung dieser Werte ist in vielen Fällen nicht einfach, da Aussagen über einen Bereich der Stichprobe gemacht werden sollen, in denen nur wenige Werte vorliegen

In vielen Schätzungen wird dazu eine Normalverteilung an die Daten angepasst und damit der Value at Risk berechnet. Das Problem dabei ist, dass die Normalverteilung bei der Art von Daten, die hier betrachtet werden sollen, oft keine sehr gute Anpassung an die extremen Werte der Daten liefert, da diese sehr selten auftreten. Da hier keine Mittelwerte betrachtet werden sollen, sondern grosse Verluste, ist aber gerade die Anpassung an diese extremen Werte wichtig. Eine bessere Möglichkeit die sogenannten Tails (Flanken) der unbekannten, zugrunde liegenden Verteilung zu schätzen, bietet die Peak over Threshold Methode (kurz: **POT**-Methode), bzw. die Block Methode. Mit diesen Methoden kann eine verallgemeinerte Paretoverteilung an die hohen Werte der Daten angepasst und schließlich der Value at Risk berechnet werden.

Beim Abschätzen solcher Ereignisse spielt die Verallgemeinerte Paretoverteilung eine Zentrale Rolle.

Definition 4.1 (Verallgemeinerte Pareto Verteilung)**.** *Sei $\xi \in \mathbb{R}$. Dann ist die Verallgemeinerte Pareto Verteilung mit Index $\xi > 0$ und Skalierungsparameter $\sigma > 0$ gegeben durch*

$$P_{\xi,\sigma}(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}, \quad x > 0.$$

Ist $\xi < 0$ dann ist der Definitionsbereich $x \in [0, -\sigma/\xi]$, und ist $\xi = 0$ dann ist die Verallgemeinerte Pareto Verteilung definiert durch

$$P_{0,\sigma}(x) = \lim_{\xi \to 0} \left(1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}\right) = 1 - e^{-x/\sigma}, \quad x > 0.$$

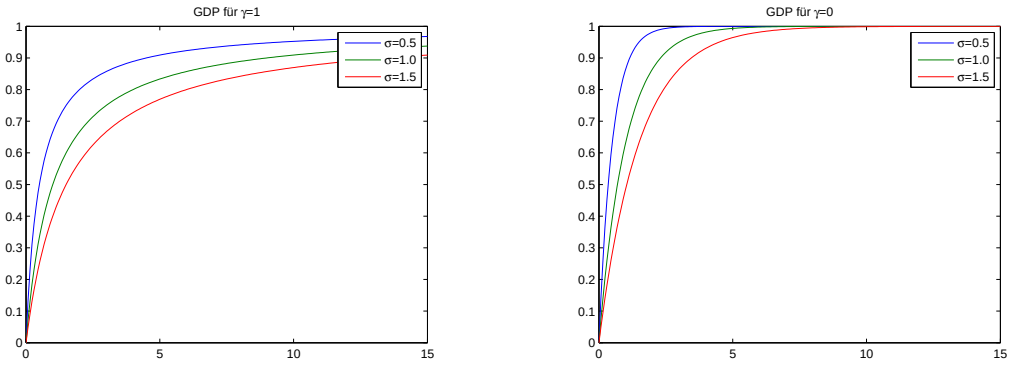


Abbildung 5.5: Verallgemeinerte Paretoverteilung mit verschiendenen Parametern.

i Die Blockmethode

Wie der Name dieser Methode schon suggeriert, teilen wir die Daten $\{X_1, X_2, \ldots, X_{km}\}$ (hier ist $n = km$) in m gleich große Blöcke der Länge k auf und bilden die Blockmaxima, das heißt, wir definieren $M_{k,j} = \max(X_{(j-1)k+1}, \ldots, Xjk)$ für $j = 1, \ldots, m$. Wenn die Daten $\{X_1, X_2, \ldots, X_{km}\}$ zum Beispiel die täglichen Verluste einer Feuerversicherung über m Monate sind (das heißt, X_l ist der Verlust am l-ten Tag), dann schätzen wir den maximalen Verlust in einem Monat. Die Blockgrößen sind ungefähr gleich groß, nämlich n ist zwischen 28 und 31 Tagen, und $M_{n,j}$ ist der maximale Verlust im j-ten Monat. Wir wissen aus den Satz von Fisher Tippett 3, dass die Verteilung von $M_{n,j}$ durch eine verallgemeinerte Extremwertverteilung approximiert werden kann, so es Konstanten $c_n > 0$ und d_n gibt mit

$$\mathbb{P}(M_n \leq u) \sim G_{\gamma,\theta,\vartheta}(u) \quad \text{für grosses } u.$$

wobei γ, θ, und ϑ Parameter sind, die geschätzt werden müssen, und die Konstanten a_n und b_n in θ und ϑ integriert sind. Zur Wiederholung, es gilt

$$G_{\gamma,\sigma,\mu}(u) = \begin{Bmatrix} e^{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}}, & \text{falls } \gamma \neq 0 \\ e^{-e^{-\frac{x-\mu}{\sigma}}}, & \text{falls } \gamma = 0 \end{Bmatrix} \text{ für } \begin{cases} 1 + \gamma\frac{x-\mu}{\sigma} > 0, & \text{falls } \gamma \neq 0, \\ x \in \mathbb{R} & \text{falls } \gamma = 0. \end{cases}$$

Es gibt verschiedene statistische Verfahren, um γ, θ, und ϑ zu schätzen, und wir bezeichnen mit $\hat{\gamma}, \hat{\theta}$, und $\hat{\vartheta}$ entsprechende Schätzwerte. Damit approximieren wir

$$\mathbb{P}(M_{m,j} \leq u) \sim G_{\hat{\gamma},\hat{\theta},\hat{\vartheta}}(u) \quad \text{für grosses } u.$$

ii Die PoT-Methode

Übersetzt heißt Peak Spitze oder Höchstpunkt und ein Threshold ist einfach ein Schwellenwert. Peak over Threshold bedeutet also so etwas wie Höchstpunkt oder Spitze über einem Schwellenwert (Threshold). Der Name beschreibt im Prinzip auch schon das Vorgehen der Methode. Es wird zunächst ein Schwellenwert im oberen Bereich der Daten gewählt. Dabei ist es entscheidend den Schwellenwert möglichst gut zu

wählen, damit es nicht zu viele, aber auch nicht zu wenige Schwellenwertüberschreitungen gibt. Schließlich sollen *nur* die extremen Werte betrachtet werden. Ist der Schwellenwert aber so hoch, dass zu wenige Überschreitungen übrig bleiben, ist die Anpassung evtl. die Schätzer für die Parameter zu ungenau. Ist der Schwellnewert zu niedrig, greift der Grenzübergang noch nicht und die Approximation der oberen Werte mittels einer verallgemeinerten Paretoverteilung ist nicht zulässig. Ist der Schwellenwert gewählt, kann schließlich eine *heavy tail* Verteilung an die Überschreitungen angepasst werden. Somit erhält man mit der Peak over Threshold Methode eine Verteilung, die die hohen Werte sehr gut beschreibt.

Problem: Gegeben ist eine Stichprobe $\{x_1, \ldots, x_n\}$ einer Grundgesamtheit die Verteilungsfunktion F hat. Zu finden ist die Wahrscheinlichkeit von $F(x)$ für x sehr gross (bzw. das q Quantil für q nahe 1, i.e. $y = F^{\leftarrow}(q)$).

Um mit der Peak over Threshold Methode eine geeignete Verteilung an die hohen Werte der Beobachtungen anzupassen, wird zunächst ein Schwellenwert (Threshold) u, der im oberen Teil des Wertebereichs der Daten liegen sollte, vorgegeben. Mit diesem Schwellenwert werden dann für $x_i > u$ die Exzesse über u bestimmt, i.e.

$$y_i := x_i - u, \quad i = 1, \ldots, n,$$

bestimmt. Die x_i, die $x_i > u$ erfüllen, heißen Exceedances.

Definition 4.2. *Sei X eine Zufallsvariable mit Verteilungsfunktion F. Die Exzess-Verteilungsfunktion von X bzw. F oberhalb des Schwellenwerts u wird definiert durch*

$$F_u(x) = \mathbb{P}\left(X - u \leq x \mid X > u\right) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

für $x \geq 0$ und festes $u < x_R$, wobei $x_R = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$ den rechten Endpunkt von F bezeichnet.

für $x > 0$ und festes $u < x_R$ definiert. Diese Verteilungsfunktion wird häufig auch Residual Life Verteilungsfunktion zum Alter u genannt. Dabei stellt $F_u(x)$ die Wahrscheinlichkeit dar, dass die verbleibende Lebenszeit kleiner als x ist, wenn das Alter u erreicht wurde.

Bemerkung 4.1. *Es gilt*

$$\bar{F}(u+y) = \bar{F}(u)\,\bar{F}_u(y).$$

Beispiel 4.1. *Sei X eine exponentialverteilte Zufallsvariable mit Parameter λ. Dann gilt*

$$\begin{aligned}
\mathbb{P}\left(X - u > t \mid X > u\right) &= \frac{\mathbb{P}\left(X - u > t, X > u\right)}{\mathbb{P}\left(X > u\right)} \quad \textit{Formel von Bayes} \\
&= \frac{\mathbb{P}\left(X > u + t\right)}{\mathbb{P}\left(X > u\right)} \quad \textit{Gedächnislosigkeit der Exponentialverteilung} \\
&= \frac{e^{-\lambda(u+t)}}{e^{-\lambda u}} = e^{-\lambda t}.
\end{aligned}$$

Beispiel 4.2. *Sei X eine Zufallsvariable aus den Anziehungsbereich der Frechetverteilung mit Index $-\alpha$, d.h.*

$$\lim_{x \to \infty} \frac{\bar{F}(\lambda x)}{\bar{F}(x)} = \lambda^{-\alpha}.$$

Dann gilt

$$\begin{aligned}
\mathbb{P}\left(\frac{X-u}{u} > t \mid X > u\right) &= \frac{\mathbb{P}\left(\frac{X-u}{u} > t, X > u\right)}{\mathbb{P}\left(X > u\right)} \quad \textit{Formel von Bayes} \\
&= \frac{\mathbb{P}\left(X > u + ut\right)}{\mathbb{P}\left(X > u\right)} \\
&= \frac{\bar{F}(u(t+1))}{\bar{F}(u)} \to (1+t)^{-\alpha}.
\end{aligned}$$

Beispiel 4.3. *Sei X eine standard normalverteilte Zufallsvariable. Dann gilt*

$$\begin{aligned}\mathbb{P}\left(u(X-u) > t \mid X > u\right) &= \frac{\mathbb{P}\left(u(X-u) > t, X > u\right)}{\mathbb{P}\left(X > u\right)} \quad \textit{Formel von Bayes} \\ &= \frac{\mathbb{P}\left(X > u + t/u\right)}{\mathbb{P}\left(X > u\right)} \\ &= \frac{e^{-u^2/2 - t - t^2/(2u^2)}}{e^{-u^2/2}} \to e^{-t}.\end{aligned}$$

Die Exzess-Verteilung repräsentiert also die Wahrscheinlichkeit, dass ein Verlust den Schwellenwert u mit höchstens einem Betrag x überschreitet, gegeben, dass der Schwellenwert überschritten wird. Mit Hilfe des Fisher Tippet Theorem, kann man Grenzsätze beweisen, gegen was F_u konvergiert, wenn u gegen den rechten Endpunkt der Verteilung F strebt.

Dazu spezifizieren wir aber als erstes die Typen von Verteilung die als Grenzwert in Frage kommen. Das wichtigste Verteilungsmodell für Exzesse über Schwellenwerts ist die verallgemeinerte Paretoverteilung (generalized pareto distribution, kurz: GPD).

Definition 4.3. *Die Verteilungsfunktion der verallgemeinerten Paretoverteilung wird definiert durch*

$$G_{\xi,\beta,\mu}(x) = \begin{cases} 1 - (1 + \xi \frac{(x-\mu)}{\beta})^{-\frac{1}{\xi}}, & \textit{falls } \xi \neq 0, \\ 1 - e^{-\frac{(x-\mu)}{\beta}}, & \textit{falls } \xi = 0. \end{cases}$$

wobei $\beta > 0$ und $x \geq 0$ im Falle $\xi \geq 0$ gilt und $0 \leq x \leq -\frac{\xi}{\beta}$ falls $\xi < 0$ gilt.

Bemerkung 4.2. *Für $\xi = 0$ ist dies die Gumpelverteilung.*

Der Parameter ξ ist der Form-Parameter (auch: Shape-Parameter) und β ein zusätzlicher Skalierungs-Parameter (auch: Scale-Parameter). Der Parameter μ ist der *Lokalisation Parameter* genannt. Für $\xi > 0$ ist $G_{\xi,\beta}$ eine reparametrisierte Version der gewöhnlichen Paretoverteilung, $\xi = 0$ entspricht einer Exponentialverteilung und $\xi < 0$ ist bekannt als eine Pareto Typ II Verteilung. Ist die Verteilungsfunktion F der Grundgesamtheit heavy tailed, dann ist vor allem der Fall $\xi > 0$ relevant.

Nun kann das Pickands-Balkema-de Haan Theorem formuliert werden. Es liefert ein wichtiges Resultat für die Schätzung von Exzessverteilungen oberhalb hoher Schwellenwerts

Satz 4.1. *Für alle $\xi \in \mathbb{R}$ liegt eine in x_R stetige Verteilungsfunktion F in $D(H)$ genau dann, wenn*

$$\lim_{u \to x_R} \sup_{0 < x \leq x_R - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0$$

für eine geeignete positive, messbare Funktion β gilt, wobei $x_R = \sup\{x \in \mathbb{R} : F(x) < 1\}$ den rechten Endpunkt von F bezeichnet.

Beispiel 4.4. *(siehe Beispiel 4.1) Sei X eine exponentialverteilte Zufallsvariable mit Parameter λ. Dann gilt*

$$\lim_{u \to \infty} \sup_{0 < x \leq \infty} \left| F_u(x) - G_{0,1/\lambda}(x) \right| = 0$$

Beispiel 4.5. *(siehe Beispiel 4.2) Sei X eine Zufallsvariable aus den Anziehungsbereich der Frechetverteilung mit Index $-\alpha$. Dann gilt*

$$\lim_{u \to \infty} \sup_{0 < x \leq \infty} \left| F_u(x) - G_{1/\alpha, u/\alpha}(x) \right| = 0$$

Beispiel 4.6. *(siehe Beispiel 4.3) Sei X eine standard normalverteilte Zufallsvariable. Dann gilt*

$$\lim_{u \to \infty} \sup_{0 < x \leq \infty} \left| F_u(x) - G_{0,\frac{1}{u}}(x) \right| = 0$$

Das Pickands-Balkema-de Haan Theorem besagt insbesondere, dass die verallgemeinerte Paretoverteilung für große u eine gute Approximation der Exzess- Funktion Fu darstellt, wenn F im Anziehungsbereich einer Extremwertverteilung liegt. Darüber hinaus ist der Shape-Parameter der GPD für die Exzesse derselbe wie der Shape-Parameter der GEV Verteilung für die Maxima.

Aus dem Pickands-Balkema-de Haan Theorem resultiert, dass die gesuchte Verteilung der Exzesse durch eine GPD approximiert werden kann, wenn der Schwellenwert groß genug ist. Dies legt nahe, einen relativ hohen Schwellenwert zu wählen, der sehr nah am rechten Endpunkt liegt. Andererseits bleiben bei einem sehr hohen Schwellenwert nur noch sehr wenig Exzesse übrig, mit denen das Modell geschätzt werden kann, was unter Umständen zu einer sehr hohen Varianz führt. Die Wahl eines Schwellenwerts ist somit grundsätzlich ein Kompromiss zwischen der Wahl eines ausreichend hohen Schwellenwerts, so dass das Theorem von Pickands-Balkema-de Haan als tatsächlich exakt betrachtet werden kann, und der Wahl eines ausreichend kleinen Schwellenwerts, so dass genügend Material zum Schätzen der Parameter zur Verfügung steht. Eine optimale, universale Lösung für die Schwellenwertwahl gibt es also nicht. Es gibt aber welche Hilfsmittel die wir hier vorstellen.

Feststellen des Verteilungstyp der Zufallsvariablen

Gegeben ist die Stichprobe $\{X_1, \ldots, X_n\}$. Die erste Aufgabe ist es nachzuprüfen in welchen Einzugsbereich die Verteilungsfunktion der Grundgesamtheit liegt. Dass heißt, man muss untersuchen ob sie in dem Einzugsbereich der Frechet Verteilung, oder Gumpelverteilung liegen.

Um den Tail einer Verteilung abzuschätzen gibt es verschiedene Methoden:

Hill Schätzer ($\xi > 0$: Sei $\{X_1, \ldots, X_n\}$ und $\{X_{(1)}, \ldots, X_{(n)}\}$ die geordnete Stichprobe, d.h. es gilt $X_{(i-1)} \leq X_{(i)}$ für alle $i = 2, \ldots, n$. Dann ist der Hill Schätzer von $\frac{1}{\xi} = \frac{1}{\alpha}$ der k ten Ordnung gegeben durch

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^{k} \log \left(\frac{X_{(i)}}{X_{(k+1)}} \right).$$

Es gilt

Satz 4.2. *Falls $n \to \infty$, $k \to \infty$ und $\frac{k}{n} \to 0$, dann gilt*

$$H_{k,n} \longrightarrow \frac{1}{\alpha}.$$

In der Praxis wird im allgemeinen ein Hill Plot von α gezeichnet, d.h. ein Plot der Punktepaare $\{(k, H_{k,n}^{-1}) : 1 \leq k \leq n\}$. Ist der Plot stabil, kann man α ablesen.

Pickands Schätzer $\xi \in \mathbb{R}$:

$$\hat{\xi}_{k,n}^{(Pickands)} := \left(\frac{1}{\log(2)} \right) \log \left(\frac{X_{(k)} - X_{(2k)}}{X_{(2k)} - X_{(4k)}} \right)$$

Deckers Einmal und de Haan Schätzer $\xi \in \mathbb{R}$:

$$\hat{\xi}_{k,n}^{DEdH} := \xi_{k,n}^{H(1)} + 1 - \frac{1}{2} \left[1 - \frac{\left(\xi_{(k)}^{H(1)} \right)^2}{\xi_{k,n}^{H(2)}} \right]^{-1}$$

mit

$$\xi_{k,n}^{H(r)} = \sum \frac{1}{k} \left[\log \left(\frac{X_{(i)}}{X_{(k+1)}} \right) \right]^r, \quad r = 1, 2, \ldots.$$

Falls $n \to \infty$, $k \to \infty$ und $\frac{k}{n} \to 0$, dann konvergieren alle Schätzer, weiters kann man die Konfidenzintervalle mittels der Normalverteilung ermitteln, genauer gilt

$$\begin{aligned} k^{\frac{1}{2}}\left(\xi_{k,n}^{(Pickand)} - \xi\right) &\longrightarrow \mathcal{N}\left(0, \xi^2 \frac{(2^{2\xi+1}+1)}{(2(2^{\xi}-1)\ln 2)^2}\right), \\ k^{\frac{1}{2}}\left(\xi_{k,n}^{DEdH} - \xi\right) &\longrightarrow \mathcal{N}\left(0, 1+\xi^2\right), \\ k^{\frac{1}{2}}\left(\xi_{k,n}^{H} - \xi\right) &\longrightarrow \mathcal{N}\left(0, \xi^2\right). \end{aligned}$$

QQ-Plots: eine weitere Möglichkeit ist einen qq plot der Verteilung mit einer heavy tail Verteilung zu vergleichen.

Wahl des Schwellenwertes (Threshold) u:

Im Folgenden stellt sich die Frage, wie ein geeigneter Threshold gewählt wird. Bei der Einführung der POT-Methode, wurde bereits erwähnt, dass der gewählte Threshold im oberen Teil des Wertebereichs der Daten liegen sollte, da sonst keine extremen Werte geschätzt werden können. Aus dem Pickands-Balkema-de Haan Theorem resultiert, dass die gesuchte Verteilung der Exzesse durch eine GPD approximiert werden kann, wenn der Threshold groß genug ist. Dies legt nahe, einen relativ hohen Threshold zu wählen, der sehr nah am rechten Endpunkt liegt. Andererseits bleiben bei einem sehr hohen Threshold nur noch sehr wenig Exzesse übrig, mit denen das Modell geschätzt werden kann, was unter Umständen zu einer sehr hohen Varianz führt. Die Wahl eines Thresholds ist somit grundsätzlich ein Kompromiss zwischen der Wahl eines ausreichend hohen Thresholds, so dass das Theorem von Pickands-Balkema-de Haan als tatsächlich exakt betrachtet werden kann, und der Wahl eines ausreichend kleinen Thresholds, so dass genügend Material zum Schätzen der Parameter zur Verfügung steht.

Kurz gesagt, ist u zu gross, so gibt es zu wenige Beobachtungen mit $X_i \geq u$, andererseits, ist u zu klein, ist die Approximation $\bar{F}_u(x) \sim \bar{G}_{\alpha,\beta(u)}(y)$ nicht sehr gut.

Eine optimale, universale Lösung für die Thresholdwahl gibt es also nicht. Es gibt aber zwei graphische Hilfsmittel mit denen geeignete Thresholds gefunden werden können. Eines davon ist der sogenannte Mean Residual Life Plot oder auch *mean excess Funktion* genannt.

Im Fall der Lebenszeit-Schätzung ist die im Alter u zu erwartende verbleibende Lebenszeit von Interesse. Diese wird durch die Mean Residual Life Funktion (kurz: mrl-Funktion) dargestellt:

Definition 4.4. *Die* Mean excess function *(MeF) einer Zufallsvariablen ist gegeben durch*

$$e(u) \quad = \quad \mathbb{E}\left[X - u \mid X > u\right].$$

Es gilt also, wenn Y GP-verteilt ist, mit den Parametern β und ξ, dann gilt für $\xi < 1$ $e(u) = \frac{\beta}{1-\xi}$. Für $\xi \geq 1$ ist der Erwartungswert unendlich. Sei nun u_0 der kleinste Threshold, für den das GP-Modell gilt, und $X_1, \ldots, X_n$ seien Zufallsgrößen mit der entsprechenden durch u festgelegten Verteilung. Vorausgesetzt $\xi < 1$, gilt dann für jedes X_i

$$\mathbb{E}(X_i - u_0 \mid X_i > u_0) = \frac{\beta_0}{1-\xi}.$$

wobei β_0 den Scale-Parameter, der sich für die Verteilung der Exzesse über den Threshold u_0 ergibt, darstellt. Aber wenn das GP-Modell gültig ist für Exzesse über den Threshold u_0, sollte es auch gültig sein für alle Thresholds $u > u_0$, allerdings mit einem anderen durch u gegebenen Scale-Parameter β_u. Zwischen β_u und β_0 besteht folgender Zusammenhang:

$$\beta_u = \beta_0 + \xi u.$$

Das folgt aus der Beweisskizze des Pickands-Balkema-de Haan Theorem, wenn anstatt der GPD-Verteilungsfunktion $G_{\xi,\nu,\beta}$, die GPD-Verteilungsfunktion G_{ξ,β_u}eingesetzt wird, d.h. die Verteilungsfunktion, die noch nicht

durch einen Loc- Parameter erweitert wurde. Für die mrl-Funktion ergibt sich daraus

$$\mathbb{E}(X_i - u \mid X_i > u) = \frac{\beta_u}{1-\xi} = \frac{\beta_{u_0} - \xi u}{1-\xi}.$$

Für $u > u_0$ ist die mrl-Funktion also linear in u.

Im folgenden Bild sind einige mean excess function von verschiedenen Verteilungen aufgezeichnet. Was wichtig ist, die mean excess function einer Pareto Verteilten Zufallsvariablen ist linear. Dies ist einfach nachzurechnen. Sei dazu X Parteo verteilt mit Parameter θ und a und $u \geq \theta$. Dann gilt

$$\begin{aligned}
\mathbb{E}\left[X - u \mid X > u\right] &= \frac{1}{\mathbb{P}\left(X > u\right)} \int_u^\infty f(x)\, x\, dx \\
&= \frac{1}{\mathbb{P}\left(X > u\right)} \int_u^\infty \bar{G}_{\xi,\beta,\nu}(x)\, dx \\
&= \frac{1}{\mathbb{P}\left(X > u\right)} \int_u^\infty \left(1 + \xi\frac{x}{\beta}\right)^{-\frac{1}{\xi}}\, dx \\
&= \frac{\frac{\beta}{\xi}}{\mathbb{P}\left(X > u\right)} \int_{1+\xi\frac{u}{\beta}}^\infty z^{-\frac{1}{\xi}}\, dz \\
&= \frac{1}{\left(1-\frac{1}{\xi}\right)} \frac{1}{\mathbb{P}\left(X > u\right)} \left(1 + \xi\frac{u}{\beta}\right)^{1-\frac{1}{\xi}} \frac{\beta}{\xi} \\
&= \frac{\xi}{\xi - 1} \left(1 + \xi\frac{u}{\beta}\right) \frac{\beta}{\xi} = \frac{\beta + \xi u}{1-\xi}\,.
\end{aligned}$$

Für die Wahl eines geeigneten Thresholds heißt das nun, dass die mrl-Funktion auf ihrem gesamten Wertebereich geplottet werden muss.

Hat man die Stichprobe $\{X_1, \ldots, X_n\}$ gegeben, so kann man eine empirische Mean Excess Funktion berechnen. Dazu sei

$$N_u = \#\{i : 1 \leq i \leq n : X_i > u\}$$

die Anzahl der Überschreitungen von u durch X_i, $i = 1 \ldots n$. Die empirische durchschnittliche Exzess Funktion ist gegeben durch

$$\hat{e}_n(u) = \frac{1}{N_u} \sum_{i=1}^{n} (X_i - u) 1_{\{X_i > u\}}.$$

Der Schwellenwert u_0 sollte so gross gewählt sein, dass der Plot $(e_n(X_i), X_i)$ für $u > u_0$ (annähernd) linear ist - die Steigung sollte circa $\frac{\xi}{(1-\xi)}$ betragen. Eine Anwendung des starken Gesetz der großen Zahlen ergibt, dass die empirische empirische mean excess-Funktion näherungsweise mit der tatsächlichen übereinstimmt. Somit liefert dieses Verfahren eine gute Hilfestellung für die Wahl des Schwellenwertes.

Beispiel 4.7.

- *Die Mean Excess Funktion der der Exponential Verteilung (Verallgemeinerten Paretoverteilung mit $\xi = 0$ und $\nu = 0$) lautet*

$$e(u) = 1, \quad u > 0$$

- *die Mean Excess Funktion der Pareto verteilung mit $\xi > 1$ und $\nu = 0$ lautet*

$$e(u) = \frac{u}{\xi - 1}, \quad u > 1$$

- *die Mean Excess Funktion der Beta Verteilung lautet*

$$e(u) = \frac{u}{\xi - 1}, \quad -1 \leq u \leq 0,$$

- *die Mean Excess Funktion Gumpel Verteilung*

$$e(u) = \frac{1+\xi u}{1-\xi}, \quad \begin{cases} 0 < u \text{ falls } 0 \le \xi < 1, \\ 0 < u < -1/\xi, \text{ falls } \xi < 0. \end{cases}$$

Dazu lässt man sich in R einen Mean Residual Life Plot (mrlplot). Auch kann man in R sich einen *tcplot* ausgeben lassen - siehe Hilfe in R (Internet).

iii Schätzen der Parameter in der GDP-Verteilung

Seien $\{Y_1, \ldots, Y_k\}$ die Exceedanten der Stichprobe $\{X_1, \ldots, X_n\}$, d.h. die Werte die u überschreiten, d.h. es gilt für alle $i = 1, \ldots, k$, $Y_k \ge u$. Achtung, in der Literatur werden oft die Wert $\{Z_1, \ldots, Z_k\}$ mit $Z_i = Y_i - u$ betrachtet.

Nachdem ein geeigneter Schwellenwert gewählt wurde, können die Parameter der GPD-Verteilung geschätzt werden. Durch die Wahl des Schwellenwerts u ist der Lokation Parameter μ bereits festgelegt. Da die Wahrscheinlichkeit, den Schwellenwert zu überschreiten, bei einem Wert, der kleiner ist als u, gleich 0 ist, ist u der linke Endpunkt der Exzess-Verteilung. Dieser entspricht bei einer GPD-Verteilung dem Location Parameter μ. Nimmt man als Grundlage der Schätzung $\{Y_1, \ldots, Y_k\}$, so ist $\mu = u$, nimmt man als Grundlage der Schätzung $\{Z_1, \ldots, Z_k\}$, so ist $\mu = 0$.

Bleiben also noch der Shape- und der Scale-Parameter zu schätzen. Diese können durch verschiedene Methoden geschätzt werden. Die bekanntesten Methoden sind die Momentenmethode (method of moments, kurz: MoM), die Methode der wahrscheinlichkeitsgewichteten Momente (method of probability weighted moments, kurz: PWM) und die Maximum-Likelihood Methode (kurz: MLE).

Maximum-Likelihood-Schätzung Bei der Maximum-Likelihood-Schätzung versucht man, die unbekannten Parameter 9 einer Verteilung so zu bestimmen, dass die beobachtete Stichprobe möglichst wahrscheinlich ist. Dafür berechnet man erst die Likelihood-Funktion und maximiert diese dann fur die gesuchten Parameter (siehe Kapitel 2.1).

Die Dichte der Verallgemeinerten Pareto Verteilungsunktion (GPD) ist ja

$$f_{\xi,\beta,\mu}(x) = \frac{1}{\beta}\left(1 + \frac{\xi(x-\mu)}{\beta}\right)^{-\frac{1}{\xi}-1}, \quad x \ge \mu.$$

Daraus ergibt sich für die Log Likelihood Funktion (Schwellenwert = u)

$$l(\beta,\xi) = \log\left(\prod_{i=1}^{k} f_{\xi,\beta,u}(Y_i)\right) = \sum \log\left(f_{\xi,\beta,u}(Y_i)\right) = -k\log\beta - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{k}\log\left(1+\frac{\xi(Y_i-u)}{\beta}\right),$$

unter der Voraussetzung dass gilt $1 + \frac{\xi(Y_i-u)}{\beta} > 0$ für $i = 1, \ldots, k$. Für $\xi = 0$ gilt für die Dichte

$$f_{0,\beta,\mu}(x) = \beta^{-1} e^{-\frac{x-\mu}{\beta}}, \quad x \ge \mu.$$

Daraus erhält man als Log Likelihood Funktion

$$l(\beta) = -k\log\beta - \frac{1}{\beta}\sum_{i=1}^{k}(Y_i - u).$$

Die Maximum Likelihood Schätzer für ξ und $\beta(u)$ sind diejenigen Werte an denen $l(\beta,\xi)$ ein lokales Maximum annimmt. Die exakten Werte $\hat{\xi}$ und $\hat{\beta}$ wo $l(\beta,\xi)$ maximal wird, kann man mit Hilfe von numerischen Optimisierungsalgorithmen lösen. Wichtig ist die Instabilität um $\xi = 0$ zu beachten und gegebenenfalls $\xi = 0$ setzen.

Die Momenten Methode: Bei der Momenten-Methode werden die gesuchten Parameter in Abhängigkeit von den Momenten der Verteilung ausgedrückt (das r-te Moment von einer Zufallsvariable X ist der Erwartungswert der r-ten Potenz von X). Anschließend werden diese theoretischen Momente jeweils durch den empirisches Schätzwert ersetzt. Auf diese Weise erhält man dann die Momentenschätzer.

Angenommen X ist GPD verteilt mit Parameter ξ. Das r-te Moment der GPD existiert für $\xi < \frac{1}{r}$ Für den Fall, dass sie existieren, gilt:

$$\begin{aligned} \mathbb{E}X &= \frac{\beta}{1-\xi}, \\ Var(X) &= \frac{\beta^2}{(1-\xi)^2(1-2\xi)}. \end{aligned}$$

Nun schätzt man $\mathbb{E}X$ mit $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ab und $Var(X)^2$ durch $\hat{S} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ab. Auflösen der obigen Beziehung nach ξ und β ergibt

$$\begin{aligned} \hat{\xi} &= - \frac{1}{2}\left(1 - \frac{\bar{X}}{\hat{S}^2}\right), \\ \hat{\beta} &= \frac{\bar{X}}{2}\left(1 + \frac{\bar{X}^2}{\hat{S}^2}\right). \end{aligned}$$

Wahrscheinlichkeitsgewichtete Momenten-Methode Für eine Zufallsvariable X mit Verteilungsfunktion F sind die wahrscheinlichkeitsgewichteten Momente (PWM) definiert als:

$$M_{p,r,s} = \mathbb{E}X^p[F(X)]^r\,[1 - F(X)]^s.$$

$M_{p,0,0}$ ergibt den normalen Momentenschätzer. Aber oftmals ist die PWM-Methode jedoch günstiger, da sie robuster gegenüber Ausreißern ist und besonders für kleine Stichproben zuverlässigere Schätzungen liefert. Der Grund dafür ist, dass extreme Werte herabgewichtet werden. Für die Schätzung der Parameter β und ξ der verallgemeinerten Paretoverteilung sind $p = 1$ und $r = 0$ gute Werte. Also

$$M_{1,0,s} = \frac{\beta}{(s+1)(s+1-\xi)}.$$

Welche Methode man nehmen sollte hängt mitunter von den Skale Parameter ξ ab (Hosking und Wallis und 1995 von Rootz'en und Tajvidi).

- falls $\xi > 0.2$ dann ist der MLE schätzer stabil.
- falls ξ nahe 0 ist und $\xi > 0$, dann sollte man die Momentenmethode wählen;
- ist $\xi < -0.2$ sollte man die PWM Methode wählen

Smith (1985) gibt folgenden Vorschlag:

- Für $\xi > -0.5$ ist die MLE regulär, auch gilt die Normalapproximation zur Berechnung der Konfidenzintervalle;
- Für $-1 < \xi < -0.5$ ist die MLE anwendbar (bzw. lösbar), die Normalapproximation zur Berechnung der Konfidenzintervalle gilt i.A. nicht;
- Für $\xi < -1$ ist die MLE i.A. nicht lösbar;

Die Verteilungsfunktion der gesamten Verteilung für $x \geq u$ ist nun gegeben durch

$$F(x) = \mathbb{P}\,(x \geq u)\; G_{\hat{\xi},\hat{\beta},u}(x-u) \sim \frac{k}{n}\, G_{\hat{\xi},\hat{\beta},u}(x-u).$$

iv Überprüfung der Anpassung

Beim Finden des richtigen Wertes ist folgende Beobachtung hilfreich. Wir wissen ja dass $F_u(x) \sim G_{\xi,\beta(u),u}$ gilt. Ist $v > u$ so gilt $F_v(x) \sim G_{\xi,\bar{\beta},v}$ mit $\bar{\beta} = \beta(u) + \xi(v-u)$.

Verschiebt man also u, dann sollte beim Schätzen des Parameters β die gleiche lineare Beziehung wiederspiegeln. (siehe auch tcplot in R)

Weiters, vergleichen mit den Daten mittels einen Histogramm, QQ-Plot. Plot.

5 Schätzen des $\mathbf{VaR}_\alpha$ - Schätzen von Quantilen, Berechnung des Expected Shortfalls

Die Quantile x_α zu einen Wert α einer Zufallsvariable X ist definiert durch

$$\mathbb{P}(X \leq x_\alpha) = 1 - \alpha.$$

Angenommen X steht für die Festigkeit eines Bauteils. Möchte man, dass die Bauteile nur mit einer Wahrscheinlichkeit von $(1-\alpha)$ brechen, so ist man an den Anteilen von Bauteilen interessiert mit Festigkeit kleiner als x_α. Der Wert α an den wir interessiert sind ist zumeist sehr nahe bei 1, z.B. $\alpha = 0.995$. Das Problem das jetzt entsteht ist, dass die Anzahl der Beobachtungen sehr gering sind, und das wenn man auch die Verteilung modellieren kann, die Unsicherheiten für grosse x sehr gross sind.

Der erste Schritt ist mittels der *Peaks over Threshold* Methode die entsprechende GPD-Verteilung zu finden. Angenommen, wir haben $\hat{\xi}$ und $\hat{\beta}$ der gefunden. Aus

$$\bar{F}(x) = \bar{F}(u)\,\bar{F}_u(x-u),$$

und $\bar{F}(u) \sim \frac{k}{n}$ folgt

$$(\bar{F})_n(x) = \frac{k}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}.$$

Gibt man sich nun die α Quantile vor, so kommt man auf

$$\hat{\mathbf{VaR}}(\alpha) = F^{\leftarrow}(\alpha) = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{(1-\alpha)n}{k}\right)^{-\hat{\xi}} - 1\right).$$

Als nächstes möchten wir eine Formel für den *Expected Shortfall* ableiten. Wir wissen ja dass

$$\mathbf{es}_\alpha = \mathbb{E}\left[X \mid X > \mathbf{VaR}_\alpha\right] = \mathbf{VaR}_\alpha + \mathbb{E}\left[X - \mathbf{VaR}_\alpha \mid X > \mathbf{VaR}_\alpha\right]$$

Der zweite Term ist gerade der Erwartungswert der Verteilung $F_{\mathbf{VaR}_\alpha}(x)$ über dem Schwellenwert $\mathbf{VaR}_\alpha$. Man kann also die Verteilung $F_{\mathbf{VaR}_\alpha}(x)$ mittels der GDP mit schon geschätzten Parameter $\hat{\xi}$ und parameter $\beta + \hat{\beta} + \hat{\xi}(\mathbf{VaR}_\alpha - u)$ anpassen. Daraus folgt nach kurzer Rechnung ($\xi < 1$)

$$\mathbb{E}\left[X - \mathbf{VaR}_\alpha \mid X > \mathbf{VaR}_\alpha\right] = \frac{\hat{\beta} + \hat{\beta} + \hat{\xi}(\mathbf{VaR}_\alpha - u)}{(1-\xi)}.$$

Einsetzten der Dichtefunktion und Ausrechnen ergibt

$$\mathbb{E}\left[X - \mathbf{VaR}_\alpha \mid X > \mathbf{VaR}_\alpha\right] = \alpha \cdot \int_{\mathbf{Var}_\alpha}^{\infty} f_{\hat{\xi},\hat{\beta},\mathbf{Var}_\alpha}(x)\,x\,dx = \alpha\frac{\hat{\beta}}{1-\hat{\xi}},$$

und damit folgt

$$\mathbf{es}_\alpha = \mathbf{VaR}_\alpha + \alpha\frac{\hat{\beta}}{1-\hat{\xi}}.$$

6 Schätzen von Rückkehrzeiten oder der Parameter im Poisson Modell

Sei $N = \{N(t) : t \geq 0\}$ ein homogener Poisson Prozess. Gegeben ist eine Zeitreihe $X = \{x_i : i = 1 \dots n\}$. In diesem Kapitel behandeln wir das Problem, wie kann man aus der Zeitreihe X, den Poisson Parameter λ schätzen.

Anhang A

Verteilungen und Verteilungstypen

In diesem Kapitel werden die gängigen Verteilungen vorgestellt. Zu Begin werden zwei Klassen von Verteilungen und Konvergenz sätze erläutert, α–stabile Verteilungen und Extremwertverteilungen. Der Grund ist, dass diese Arten von Verteilungen relativ oft in der Risikoanalyse vorkommen.

1 Klassen von Verteilungen und Grenzsätze

i Stabile Verteilungen und das schwache Gesetzt der großen Zahlen

Seien $\{X_i : i \in \mathbb{N}\}$ identisch verteilte Zufallsgrössen mit Verteilung F gegeben und es sei

$$S_n := \sum_{i=1}^{n} X_n$$

die n-te Partialsumme der Folge $\{X_i : i \in \mathbb{N}\}$. Dann besagt dass starke Gesetzt der großen Zahlen, dass

$$\hat{\mu}_n := \frac{1}{n} S_n \to \mu := \mathbb{E}X_1,$$

falls $\mathbb{E}|X_1| < \infty$. Statistisch gesehen bedeutet dies, dass der empirische Mittwelwert einer Stichprobe gegen den Mittelwert ν der Grundgesamtheit strebt. Möchte man die Konfidenzintervalle des Schätzers $\hat{\nu}_n$ berechnen, verwendet man das schwache gesetzt der großen Zahlen. Das besagt, dass

$$\frac{1}{\sqrt{n}} S_n \to \mathcal{N}(\mu, \sigma),$$

wobei σ die Standardabweichung der Grundgesamtheit ist. Das diese Konvergenz hält ist aber wichtig, dass die Standardabweichung der Grundgesamtheit endlich ist und existiert.Genuegen aber $\{X_i : i \in \mathbb{N}\}$ einer Cauchy Verteilung, so ist σ nicht endlich, es existieren sogar keine ersten Momente ! In diesem Fall bleibt aber die Konvergenz erhalten, nur die Skalierung muss abgeändert werden und die Grenzverteilung ist nicht mehr die Normalverteilung sondern eine sogenannte stabile Verteilung.

Definition 1.1. *Eine Verteilung F heißt stabil, falls es für jedes n eine Verteilung F_n gibt und ein $\alpha \in (0, 2]$ sodass gilt*

$$F \stackrel{d}{=} n^{-\frac{1}{\alpha}} F_n^{(n)*}.$$

Besser kann man sich die Stabilität einer Zufallsvariaben vorstellen:

Definition 1.2. *Eine Zufallsvariable besitze eine stabile Verteilung, falls es für jedes $n \in \mathbb{N}$ eine Familie von n identische und unabhängig verteilter Zufallsvariablen $\{X_n^i : i = 1, \ldots, n\}$, eine reelle Zahl b_n und ein $\alpha \in (0, 2]$ gibt, sodass gilt*

$$X = \frac{1}{n^{\frac{1}{\alpha}}} \left(\sum_{i=1}^{n} X_n^i \right) - b_n.$$

Solche α–stabile Verteilung treten in der Versicherungsmathematik relativ oft auf. Z.b. ist die Cauchy Verteilung 1–stabil.

Bemerkung 1.1. *Positive stabile Verteilungen sind leptokurtotic und heavy-tailed.*

Hat man jetzt z.B. eine Familie $\{X_i : i \in \mathbb{N}\}$ von identisch verteilte Zufallsgroßen die Parteo verteilt sind, d.h. $F \sim Par$, dann ist das zweite Moment nicht endlich. Hier stellt sich nun die Frage, wie man die Summe S_n skalieren muss und gegen welche Verteilung diese Skalierte Summe strebt. Um das zu analysieren muss man die Tails der Verteilung abschätzen. Dazu führen wir folgende Definition ein:

Definition 1.3. *Eine Funktion $L : [0, \infty) \to [0, \infty)$ heißt* langsam variierend *(slowly varying (at infinity)) falls für alle $a > 0$, gilt*

$$\lim_{x \to \infty} \frac{L(x)}{L(ax)} = 1.$$

Falls der Grenzwert

$$\lim_{x \to \infty} \frac{L(x)}{L(ax)}$$

endlich, aber ungleich eins, ist, heißt die Funktion regulär variierend *(regularly varying function).*

Beispiele einer langsam variierender Funktion ist $[0,\infty) \ni x \mapsto \ln(x)$ und $[0,\infty) \ni x \mapsto \ln(x^2)$.

Definition 1.4.

- *Die durch* $\bar{F}(x) := 1 - F(x)$, $x \in \mathbb{R}$ *definierte Funktion heißt Tailfunktion von* F.
- $\bar{F}$ *heißt regulär variierend mit Index* α *in* ∞, *wenn*

$$\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{\alpha}, \quad x > 0$$

gilt.

- *Wir nennen* $x_R := \sup\{x : F(x) < 1\}$ *den rechten Endpunkt der Verteilungsfunktion* F *und* $x_L := \inf\{x : F(x) > 0\}$ *den linken Endpunkt der Verteilungsfunktion* F.

Satz 9.

- *Die Partialsummen* S_n *skaliert mit* $\frac{1}{\sqrt{n}}$ *konvergiert gegen die Normalverteilung, falls die Funktion*

$$x \mapsto \int_{|y|\le x} y^2 dF(y)$$

langsam variierend ist.

- *Die Partialsumme* S_n *skalierte mit* $\frac{1}{n^{\frac{1}{\alpha}}}$ *konvergiert gegen eine* α *– stabile Verteilung, falls es eine langsam variierende Funktion* L *und zwei Konstanten mit* $c_1 + c_2 > 0$ *gibt, sodass gilt*

$$F(-x) = \frac{c_1 + K}{x^{\alpha}} L(x) \text{ und } 1 - F(x) = \frac{c_2 + K_2}{x^{\alpha}} L(x) \text{ für } x \to \infty.$$

Durch den Skalierungsfaktor $\frac{1}{\alpha}$ stimmt die übliche Berechnung der Konfidenzintervalle für den Mittelwertschätzer nicht mehr. Diese Konfidenzintervalle werden unter Annahme der Normalverteilung berechnet !

2 Diskrete Verteilungen

i Die Gleichverteilung

Vorhanden sind n nummerierte Kugeln. Man zieht zufällig eine Kugel. Sei der Wert der Zufallsvariable X die Nummer auf der Kugel. Wie gross ist die Wahrscheinlichkeit dass diese Kugel genau den Wert 1 hat $(X = 1)$?

$$\mathbb{P}(X = k) = \frac{1}{n}, \quad k = 1, \ldots, n. \tag{A.1}$$

Der Erwartungswert lautet

$$\mathbb{E}X = (n+1)/2,$$

die Varianz lautet

$$\mathrm{Var}[X] = (n+1)(2n+1)/6.$$

ii Binominal Verteilung

Betrachten wir ein Experiment das nur zwei Mögliche Ausgänge hat, Erfolg oder Misserfolg. Die Wahrscheinlichkeit von Erfolg ist p, die Wahrscheinlichkeit von Misserfolg ist $q = 1 - p$. Dieses Experiment wird n mal wiederholt. Dabei nehme man an, dass die Ausgänge von den Experimenten unabhängig sind.

Sei X eine Zufallsvariable in der die genaue Anzahl der Erfolge steht. Dann ist X Binominal-verteilt, d.h.

$$\mathbb{P}(X = k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^k, \quad k = 0, 1, \ldots, n. \tag{A.2}$$

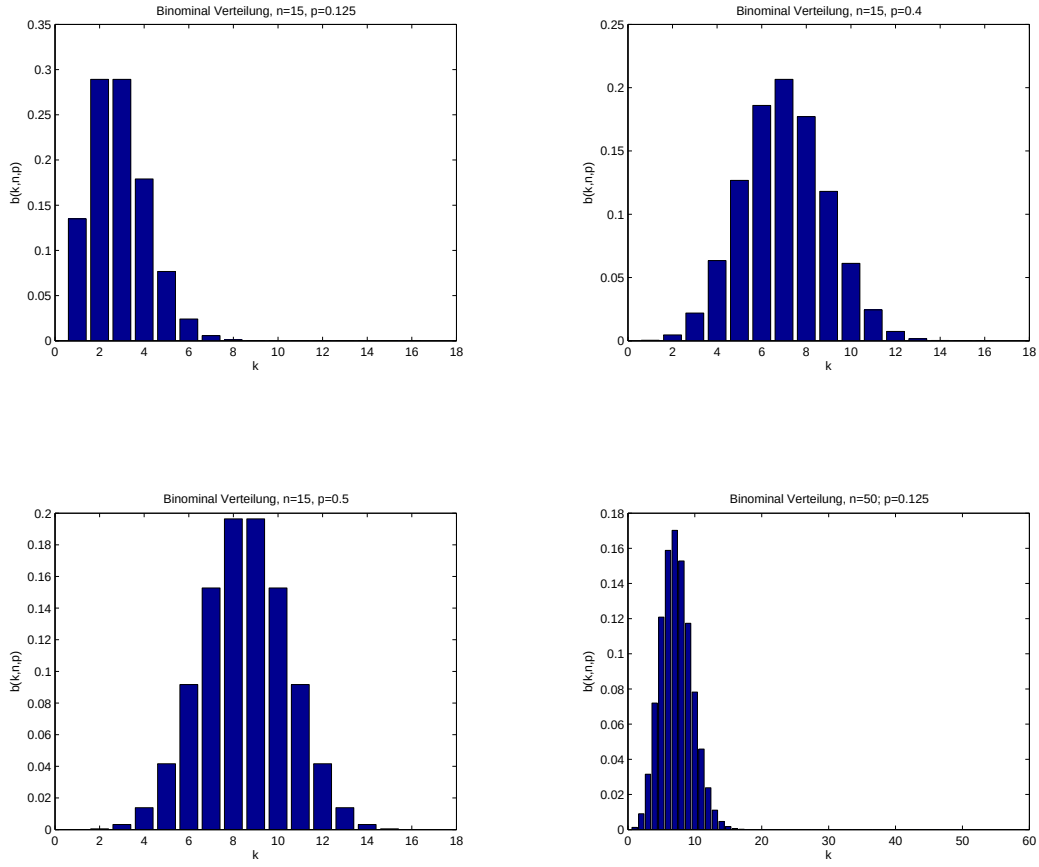
Der Erwartungswert lautet

$$\mathbb{E}X = np,$$

die Varianz lautet

$$\mathrm{Var}[X] = np(1-p).$$

Verteilungen für verschiedene Parameter sind in Bild A.1 illustriert.



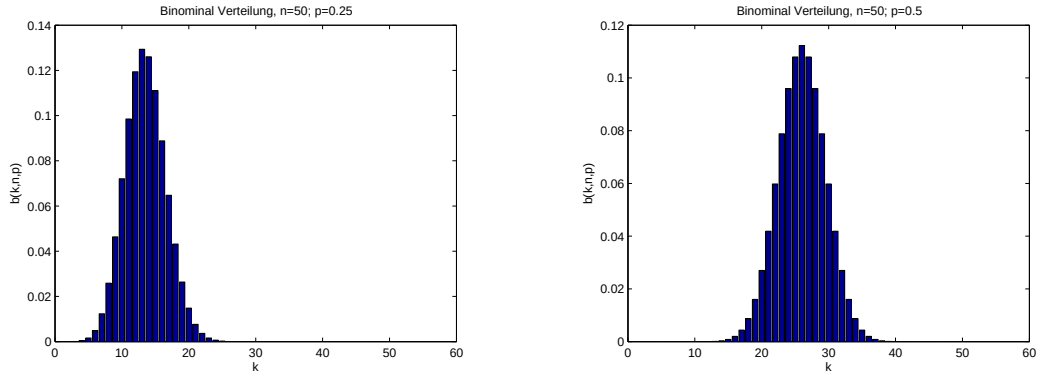


Abbildung A.1: Die Binominal-Verteilung.

iii Die Geometrische Verteilung

Betrachten wir wieder das Experiment von vorher, d.h. das Experiment beschrieben in dem ersten Absatz der Binominal-Verteilung. Jetzt aber fragen wir uns: nach dem wievielten Versuch ist das Experiment erfolgreich? Sei X die Anzahl der Versuche die bis zu dem Zeitpunkt stattgefunden haben wo das erste mal Erfolg eintritt.

Die Zufallsvariable X genügt einer geometrischen Verteilung, wobei gilt

$$\mathbb{P}(X = k) = g(k; p) = (1-p)^{k-1}p, \quad k = 1, \ldots, \ . \tag{A.3}$$

Der Erwartungswert lautet

$$\mathbb{E}X = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p[1 + 2(1-p) + 3(1-p)^2 + \cdots] = \frac{1}{p}.$$

Die Größe $\frac{1}{p}$ heißt oft Rückkehrzeit. Die Varianz lautet

$$\mathrm{Var}[X] = \frac{(1-p)}{p^2}.$$

Verteilungen für verschiedene Parameter sind in Bild A.2 illustriert.

Beispiel 2.1. *Jemand darf beim Würfelspiel erst aufhören, falls er zweimal die Eins geworfen hat. Wie groß ist die Wahrscheinlichkeit dass jemand mindestens sieben Mal keine eins geworfen hat, bevor er aufhören darf.*

iv Hypergeometrische Verteilung

Man hat eine Stückzahl von N Bauteilen. Genau D Bauteile sind nicht funktionstüchtig und $N - D$ Bauteile sind funktionstüchtig. Man greift zufällig n Bauteile heraus. Die Zufallsvariable X beschreibt die Anzahl der nicht funktionstüchtigen Bauteile. Dann ist X Hypergeometrisch-verteilt, d.h.

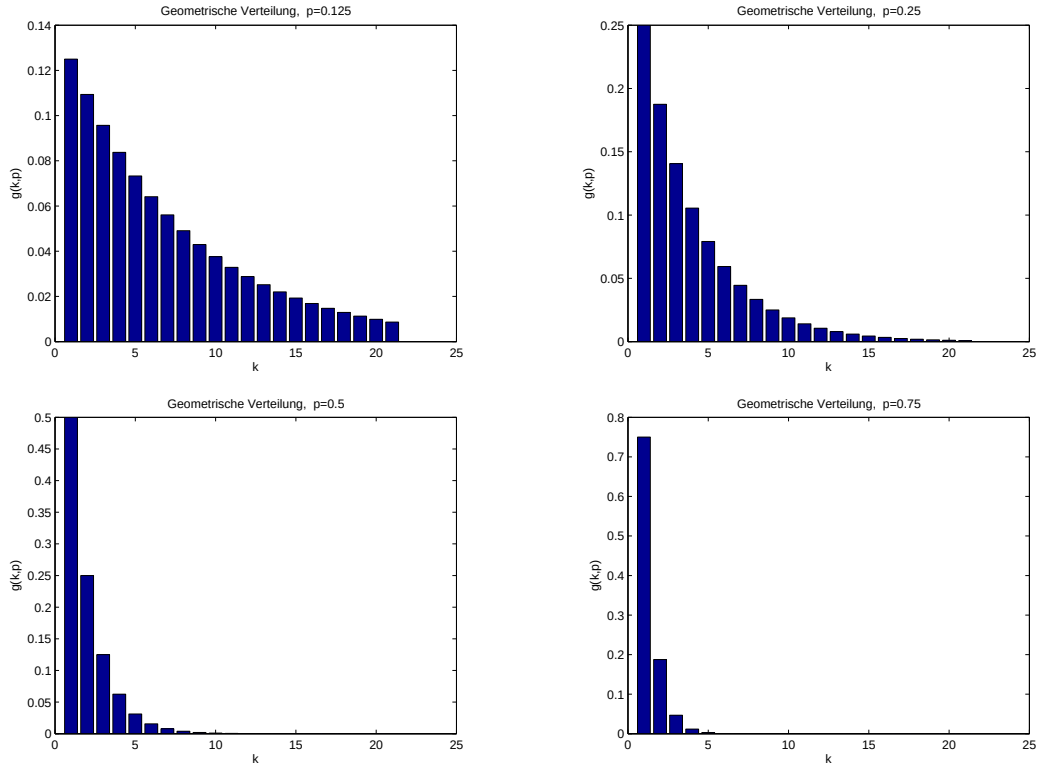


Abbildung A.2: Die Geometrische Verteilung.

$$\mathbb{P}(X = k) = h(k; N, D) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, \min(N, D). \tag{A.4}$$

Die Wahrscheinlichkeit $\mathbb{P}(X = k)$ sagt an wie groß die Wahrscheinlichkeit ist dass genau k Bauteile nicht funktionstüchtig sind. Der Erwartungswert lautet

$$\mathbb{E}X = np.$$

Die Varianz lautet

$$\mathrm{Var}[X] = npq\left(\frac{N-n}{N-1}\right), \quad p = \frac{D}{N},\ q = \frac{N-D}{N}.$$

Beispiel 2.2. *In einer Region studiert man die Population der Wölfe. 13 Wölfe fing man ein, bezeichnete sie und lies sie wieder frei. Nach einer längeren Zeit haben sich die Wölfe mit den nicht gezeichneten wieder vermischt und man fing 10 Wölfe, zwei von Ihnen waren gezeichnet. Berechnen Sie den Maximum Likleihood Schätzer der Wolfpopulation.*

Antwort: Sei N die Populationsgrösse. $D = 13$, $n = 10$. Die Anzahl der gezeichneten Wölfe beim zweiten Fang ist Hypergeometrisch verteilt. Also suchen wir den Parameter N, wo $h(2; N, 13)$ am größten

ist, also

$$h(2;N,13) = \frac{\binom{13}{2}\binom{N-13}{8}}{\binom{N}{10}}.$$

Da dies schwer zu Berechnen ist, schaut man sich den Quotienten $h(2;N,13)/h(2;N-1,13)$ an, und setzt diesen gleich 1. Die Lösung ist 65. [1]

v Poisson Verteilung

Betrachten Sie nun die Anzahl eines Ereignis das über einen Zeitraum geschehen kann. Zum Beispiel die Anzahl der Erdbeben in einem Jahr in einer bestimmten Region, die Anzahl aller Besucher einer Tankstelle in in der Mittagszeit $12.00 - 14.00$, die Ausfälle einer Maschine innerhalb eines Tages).

Die durchschnittliche Häufigkeit ($=\lambda$) dieses Ereignisses sei konstant und die Ereignisse selber seien als unabhängig angenommen.

Bezeichnet man die Anzahl solcher Ereignisse als K dann genügt K einer Poisson Verteilung, d.h.

$$\mathbb{P}\left(K = k \text{ im Zeitintervall } (0,t)\ \right) = p(k;(0,t),\lambda) = \frac{(\lambda t)^k}{k!} e^{-k\lambda t}, \quad k = 1,2,\ldots . \tag{A.5}$$

Der Erwartungswert lautet

$$\mathbb{E}X = \lambda t.$$

Die Varianz lautet

$$\mathrm{Var}[X] = \lambda t.$$

Verteilungen für verschiedene Parameter sind in Bild A.3 illustriert.

[1] Ist der Quotient größer eins, steigt $N \mapsto h(2;N,13)$, ist dieser Quotient kleiner 1 sinkt $N \mapsto h(2;N,13)$.

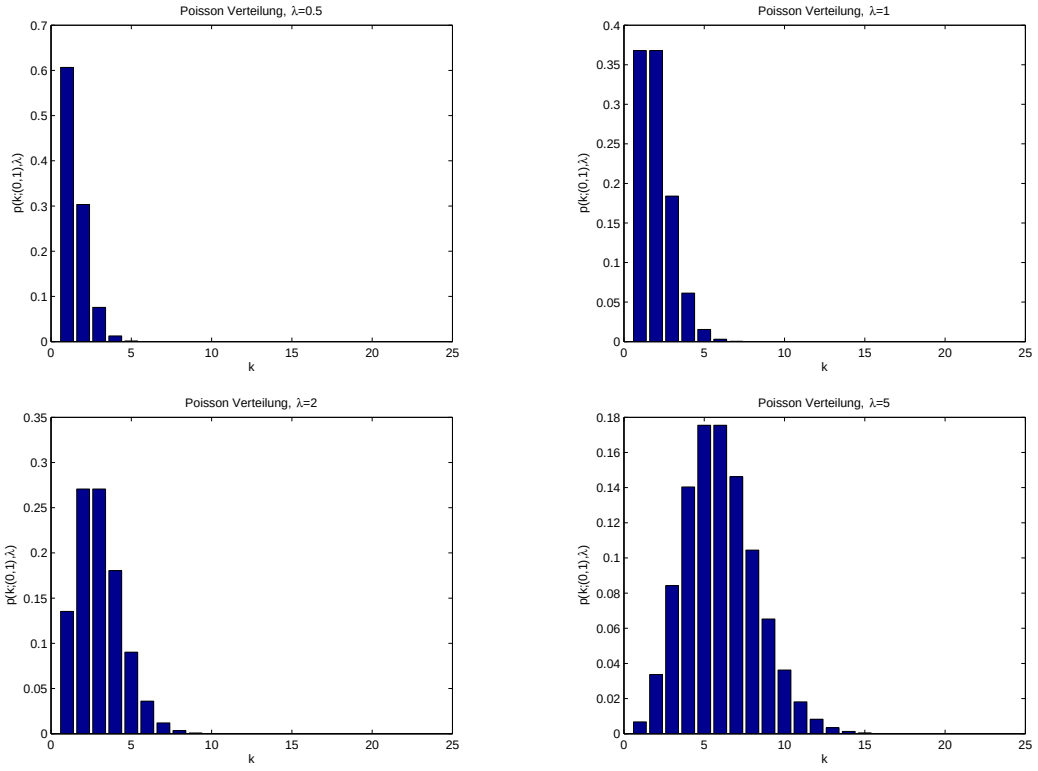


Abbildung A.3: Die Poisson-Verteilung.

Beispiel

Im Durchschnitt kommt es in einen Land jede fünf Jahre zu einen Erdbeben. Nehmen Sie an, dass die Anzahl der Erdbeben durch einen Poisson Prozess beschrieben werden kann. Berechnen Sie die Wahrscheinlichkeit folgender Ereignisse:

1. Das Erdbeben passierte genau einmal in drei Jahren;

2. Während der nächsten drei Jahr findet keine Erdbeben statt;

3. Es gibt in den nächsten zwei Jahren höchstens zwei Erdbeben

4. In den nächsten fünf Jahren gibt es höchstens ein Erdbeben.

vi Negative Binominal Verteilung

In seiner einfachsten Form (wenn r eine ganze Zahl ist), modelliert die negative Binomialverteilung die Anzahl der Ausfälle x in einer bestimmten Anzahl von Erfolgen. Hier nimmt man an das die einzelnen Experimente unabhängig und identisch verteilt sind. Die Parameter der Negativen Binominal Verteilung sind die Wahrscheinlichkeit des Erfolgs in einem einzigen Versuch, p, und die Anzahl der Erfolge r. Die geometrischen Verteilung ist ein Spezialfall der negativen Binomialverteilung für $r = 1$.

Eine Zufallsvariable ist negativ Binominalverteilt mit Parametern r und p, $0 < p < 1$, $X \sim \mathbf{NB}(r; p)$ falls gilt

$$\mathbb{P}(X = n) = \binom{r+n-1}{n} \cdot p^r \cdot (1-p)^n, \quad n = 0, 1, 2, \ldots . \tag{A.6}$$

Verteilungen für verschiedene Parameter sind in Bild A.4 illustriert.

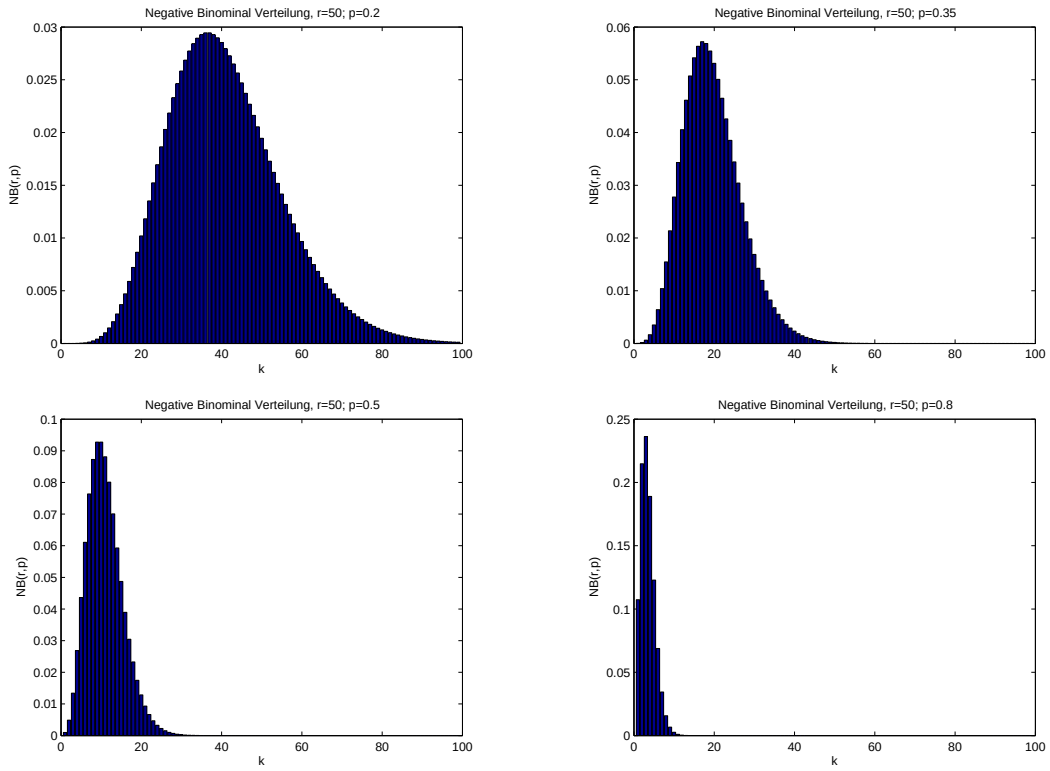


Abbildung A.4: Die Negative Binominal-Verteilung.

Ganz allgemein kann r auch nicht-ganzzahligen Werte annehmen. Die negativen Binomialverteilung kann nicht im Hinblick auf wiederholte Versuche interpretiert werden, ist aber wie die Poisson-Verteilung sinnvoll bei der Modellierung von Zähldaten. Auch ist die negative Binomialverteilung allgemeiner als die Poisson-Verteilung, da ihre Varianz größer als der Mittelwert sein kann. Im Grenzfall, falls r ins Unendliche wächst, nähert sich die negative Binomialverteilung der Poisson-Verteilung.
Angenommen, Sie sammeln Daten über die Anzahl der Autounfälle an einer viel befahrenen Autobahn, und möchten die Anzahl der Unfälle pro Tag modellieren. Da die Anzahl der Autos die vorbeifahren sehr groß ist und die Wahrscheinlichkeit eines Unfalls sehr gering ist, könnte man meinen, dass die Poisson-Verteilung paßt. Allerdings variiert die Wahrscheinlichkeit eines Unfalls sowie der Verkehr von Tag zu Tag, und somit sind die Annahmen für die Poisson-Verteilung nicht erfüllt. Insbesondere ist oft die Varianz von solchen Zähldaten größer als der Mittelwert. An den meisten Tagen passieren nur wenige oder gar keine Unfälle, und an ein paar wenigen Tagen passieren viele Unfälle.

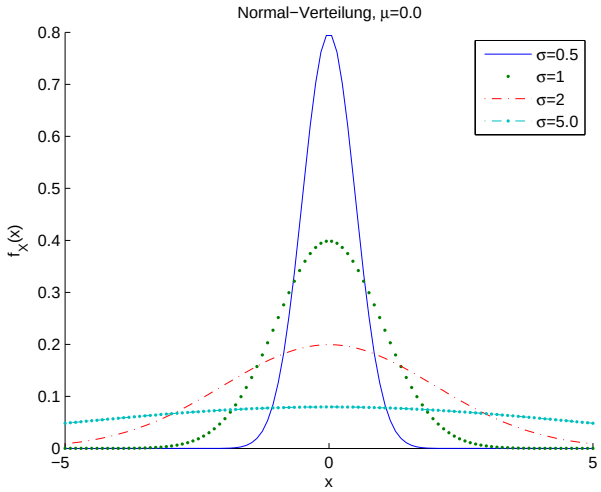


Abbildung A.5:

3 Stetige Verteilungen

i Normalverteilung

Sei X eine Zufallsvariable. Dann ist X normalverteilt[2], falls die Dichtefunktion f_X von X folgende Form hat:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\,\pi}}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}, \quad x \in \mathbb{R}. \tag{A.7}$$

In Bild A.5 ist die Dichte der Normalverteilung mit Mittelwert 0 und Varianz 1 abgebildet.

Die Parameter $\sigma \geq 0$ und $\mu \in \mathbb{R}$ in (A.7) kann man frei wählen. Man schreibt auch kurz, X ist $\mathcal{N}(\mu, \sigma)$ verteilt. Der Parameter μ heißt Mittelwert und es gilt

$$\int_{-\infty}^{\infty} x\, f_X(x)\, \mathrm{d}x = \mu.$$

Der Parameter σ heißt auch Standardabweichung und es gilt

$$\int_{-\infty}^{\infty} (x-\mu)^2\, f_X(x)\, \mathrm{d}x = \sigma^2.$$

Ist X eine normalverteilte Zufallsvariable, bzw. $X \sim N(0, \sigma)$, so kann man an der Standardabweichung sehen, wieviel im Durchschnitt für ein ω der Wert $X(\omega)$ vom Mittelwert 0 abweicht. Will man sich insbesondere die Wahrscheinlichkeit ausrechnen, dass X in dem Interval $[-\sigma, \sigma]$ liegt, so muss man den Wert des Integrals von $-\sigma$ nach σ über f_X berechnen, d.h.

$$\mathbb{P}\left(\{\omega \in \Omega : -\sigma(\omega) \leq X \leq \sigma\}\right) = \frac{1}{\sigma\sqrt{2\,\pi}} \int_{-\sigma}^{\sigma} e^{-\frac{x^2}{2\,\sigma^2}}\, \mathrm{d}x. \tag{A.8}$$

[2]Siehe Skriptum Kirschenhofer, Statistik B.2.

Analog, kann man die Wahrscheinlichkeiten berechnen, dass die Werte von X im Interval $[-2\sigma, 2\sigma]$, $[-3\sigma, 3\sigma]$, ..., liegt. Hier berechnet man den Wert des Integrals von -2σ nach 2σ (beziehungsweise von -3σ nach 3σ, ...) über f, das heißt

$$\mathbb{P}\left(\{\omega \in \Omega : -2\sigma \leq X \leq 2\sigma\}\right) = \frac{1}{\sigma\sqrt{2\,\pi}} \int_{-\sigma}^{\sigma} e^{-\frac{x^2}{2\,\sigma^2}}\,\mathrm{d}x. \tag{A.9}$$

In folgender Tabelle sind einige wichtige Werte eingetragen:

n	$\mathbb{P}(-n\,\sigma \leq x \leq n\,\sigma)$	$\mathbb{P}(\lvert x\rvert > n\,\sigma)$
1	68.3%	31.7%
2	95.4%	4.6%
3	99.7%	0.3%

ii Exponentielle Verteilung

Angenommen eine Maschine hat eine konstante Fehlerrate von λ. Angenommen in der Früh (genannt Zeitpunkt 0) wird diese Maschine in Betrieb gesetzt und läuft bis Mittag (also bis $t = 4$). Unterteilen wir diese Zeitlänge von vier Stunden in Teilintervalle der Länge $t/n(= \Delta t)$, $n \geq 10$. In jeden Teilintervall passiert entweder mit der Wahrscheinlichkeit $1 - \lambda\Delta t$ kein Fehler oder die Maschine stoppt mit wegen einen Fehler mit der Wahrscheinlichkeit $\lambda\Delta t$.

Fasst man alle diese Zeitintervalle zusammen, kann man jeden Zeitintervall ein Experiment zuordnen und die Wahrscheinlichkeit dass bei der Maschine kein Fehler auftritt als Bernoulli Experiment berechnen, d.h.

$$\begin{aligned} & \mathbb{P}\left(\text{ es tritt keine Fehler auf }\right) \\ = \quad & b(0; n, \lambda\Delta t) = \binom{n}{0} (\lambda\Delta t)^0 (1 - \lambda\Delta t)^{n-0} = (1 - \lambda\Delta t)^n. \end{aligned} \tag{A.10}$$

Betrachtet man nun einen stochastischen Prozess X der zu Beginn 0 ist, und zum Zeitpunkt T des ersten Fehlers auf eins springt, so kann man einfach durch einen Grenzübergang in (A.10) mit $n \to \infty$ die Wahrscheinlichkeit $\mathbb{P}(T > t) = \mathbb{P}(X(s) = 0,\, 0 \leq s < T)$ berechnen,

$$P(T > t) \quad = \quad \lim_{\substack{n\to\infty \\ \Delta t \to 0}} (1 - \lambda\Delta t)^n = \lim_{n\to\infty} (1 - \frac{\lambda}{n})^n = e^{-\lambda t}.$$

Die Exponentialverteilung tritt (näherungsweise) in Natur und Technik recht häufig auf. Typische Beispiele sind Lebensdauern von technischen Bauteilen, zeitliche und räumliche Abstände zwischen Autos auf wenig befahrenen Autobahnen, Zeitabstände zwischen Impulsen beim Radioaktiven Zerfall, Sehnenlängen zu unregelmässige verteilten Strukuren in ebenen Schliffen, Abstände zwischen Anrufen in einer Telefonzentrale, etc. .
Die Exponentialverteilung hängt stark mit den Poisson Prozess zusammen, hier sind die Wartezeiten zwischen den einzelnen Ereignissen eines Poissonprozess exponential verteilt.

Eine wichtige Eigenschaft der exponentialverteilung ist die *Gedächnislosigkeit*. Dazu stellen wir uns einen Lebensdauernversuch vor, der zur Zeit $t = 0$ beginnt und bei einem Ausfall eines Bauteils endet. Die Lebensdauer X eines Bauteils sei exponentialverteilt. Es gilt dann also

$$\mathbb{P}\left(X > t\right) = e^{-\lambda t}; \quad t \geq 0.$$

Nun möge die Beobachtung so verlaufen, dass das Bauteil bis zur Zeit t_0 nicht ausfällt. Welche Prognose können wir dann für das Ausfallen nach der Zeit t_0 machen ? Genauer, wie groß ist die Wahrscheinlichkeit dafür, dass das Bauteil t weitere Zeiteinheiten aushält ?

Diese Frage kann mit Hilfe bedingter Wahrscheinlichkeiten beantwortet werden. Es sei X_{t_0} die Restlebensdauer unseres Bauteils nach der Zeit t_0. Wir suchen $\mathbb{P}(X_{t_0} > t)$. Es gilt

$$\mathbb{P}(X_{t_0} > t) = \mathbb{P}(X > t + t_0 \mid X > x).$$

Nach der Formel für die bedingte Wahrscheinlichkeit erhalten wir

$$\begin{aligned}\mathbb{P}(X > t + t_0 \mid X > x) &= \frac{\mathbb{P}(X > t + t_0, X > x)}{\mathbb{P}(X > x)} \leq \frac{\mathbb{P}(X > t + t_0)}{\mathbb{P}(X > x)} \\ &= \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t}} = e^{-\lambda t_0}.\end{aligned}$$

Das zweite Gleichheitszeichen ergibt sich daraus, dass Ereignis $\{X > t_0 + t\}$ das Ereignis $\{X > t_0\}$ nach sich zieht: Wenn $X > t_0 + t$, dann ist selbstverständlich auch $X > t_0$. Unsere Rechnung hat also ergeben dass

$$\mathbb{P}(X_{t_0} > t) = \mathbb{P}(X > t) = e^{-\lambda t}.$$

Das Ergebnis besagt, dass unser Bauteil, wenn es einmal das Alter t_0 erreicht hat, in Zukunft die gleichen Chancen auszufallen hat wie bereits zur Zeit $t = 0$. Die verfloßenen Lebensdauer t_0 spielt keine Rolle.

Sicher ist, die exponentialverteilte Lebensdauer werden wir nicht beobachten können, wenn Verschleißerscheinungen eine Rolle spielen. Wir können Sie aber näherungsweise erwarten, wenn zufällige seltene Ereignisse die Ausfälle verursachen, z.B. Überlastung oder Fehlbedienung. Ansonsten sollte man die Weilbullverteilung oder Rayleigh Verteilung modellieren.

iii Die Rayleighverteilung

Seien $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige normalverteilte Zufallsvariable mit Mittelwert 0 und Varianz σ. Mittels X und Y kann man eine neue Zufallsvariable r definieren, die durch den Betrag des Vektors (X, Y) gegeben ist. Mathematisch ausgedrückt, ist r definiert durch

$$\begin{aligned} r : \Omega &\longrightarrow \mathbb{R}, \\ \omega &\mapsto r(\omega) := \sqrt{X^2(\omega) + Y^2(\omega)}. \end{aligned}$$

Die Dichtefunktion von r kann man leicht mittels Substitution ableiten: Sei $0 \leq a < b < \infty$. Dann gilt

$$\begin{aligned}\mathbb{P}(r \in [a, b]) &= \mathbb{P}(1_{[a,b]}(r)) = \mathbb{P}(1_{[a,b]}(\sqrt{X^2 + Y^2})) \\ &= \int_0^\infty \int_0^\infty f_X(x)\, f_Y(y)\, 1_{[a,b]}(\sqrt{x^2 + y^2}) \mathrm{d}x \mathrm{d}y \\ &= \frac{1}{\sigma^2\, 2\,\pi} \int_0^\infty \int_0^\infty e^{-\frac{x^2}{2\,\sigma^2}} e^{-\frac{y^2}{2\,\sigma^2}}\, 1_{[a,b]}(\sqrt{x^2 + y^2}) \mathrm{d}x \mathrm{d}y \\ &= \frac{1}{\sigma^2\, 2\,\pi} \int_0^\infty \int_0^{2\pi} e^{-\frac{r^2}{\sigma^2}}\, 1_{[a,b]}(r) \mathrm{d}\theta \mathrm{d}r.\end{aligned}$$

☺

Sei $-\infty < a < b \leq 0$. Dann ist $\mathbb{P}(r \in [a, b]) = 0$, da der Betrag eines Vektors nicht negativ sein kann. Damit lautet die Dichtefunktion dieser Zufallsvariablen r, die man Rayleigh-Verteilung nennt (kurz: $r \sim \text{Rayleigh}(\sigma)$)

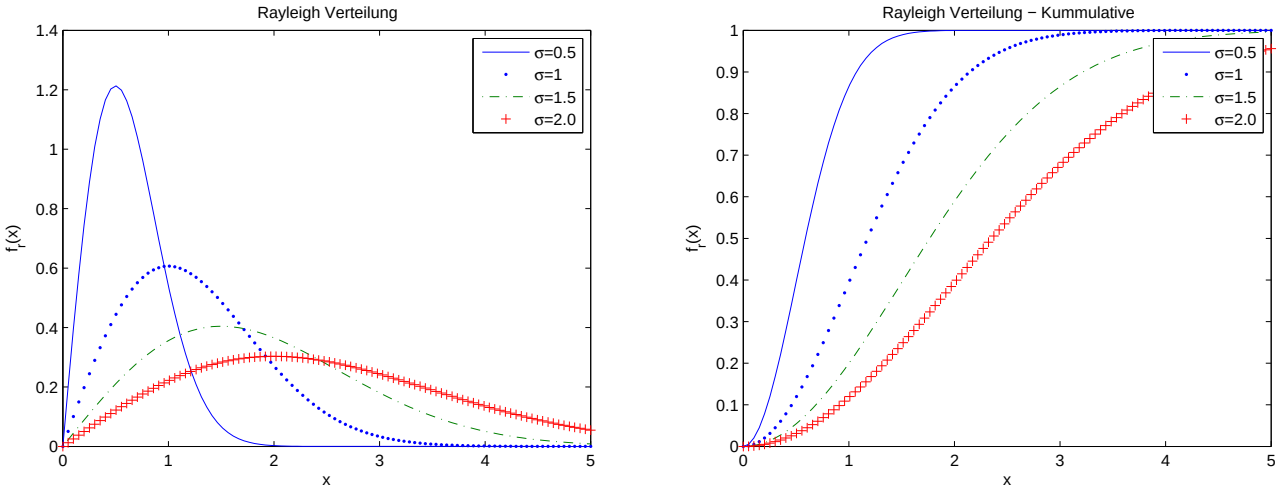


Abbildung A.6:

$$f_r(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

die Verteilungsfunktion lautet

$$F_r(x) = \begin{cases} 1 - e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0 \\ 0 & x < 0. \end{cases}$$

In Bild A.6 ist auf der linken Seite die Dichtefunktion für die Raighley Verteilung und auf der rechten Seite die Verteilungsfunktion der Raighley Verteilung gezeichnet. Die Varianz ist jeweils 1 gesetzt.

Weitere Kenngrößen sind der Erwartungswert, die Momente und die Varianz. Der *Erwartungswert* lautet $\mu_r = \sigma\sqrt{\frac{\pi}{2}}$, was aus folgender Rechnung folgt:

$$\begin{aligned} \mu_r &= \int_0^\infty x\, f_r(x) \mathrm{d}x = \int_0^\infty \frac{x^2}{\sigma^2}\, e^{-\frac{x^2}{2\,\sigma^2}}\, \mathrm{d}x = \\ &= \int_0^\infty x \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}\, \mathrm{d}x = e^{-\frac{x^2}{2\sigma^2}}|_0^\infty + \int_0^\infty e^{-\frac{x^2}{2\sigma^2}}\, \mathrm{d}x = \\ &= \sigma\sqrt{2\,\pi} \cdot \frac{1}{\sigma\sqrt{2\,\pi}} \int_0^\infty e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}r = \sigma\sqrt{2\,\pi} \cdot \frac{1}{2} \cdot \frac{1}{\sigma\sqrt{2\,\pi}} \int_{-\infty}^\infty e^{-\frac{r^2}{2\sigma^2}} \mathrm{d}x = \\ &= \sigma\sqrt{2\,\pi} \cdot \frac{1}{2} = \sigma\sqrt{\frac{\pi}{2}}. \end{aligned}$$

☺

Der maximale Wert der Dichtefunktion wird in σ angenommen. Das zweite Moment lautet $\mu_{r^2} = 2\,\sigma^2$,

was aus folgender Rechnung folgt:

$$\begin{aligned}\mu_{r^2} &= \int_0^\infty x^2\, f_r(x)\,\mathrm{d}x = \int_0^\infty \frac{x^3}{\sigma^2}\, e^{-\frac{x^2}{2\,\sigma^2}}\,\mathrm{d}x \\ &= 2\,\sigma^2.\end{aligned}$$

☺

Allgemein gilt für das $k$te Moment:

$$\mu_{r^k} = \sigma^k 2^{k/2}\Gamma(1+k/2)^3, \qquad k \in \mathbb{N}.$$

Für die Varianz gilt: $Var(r) = \sigma_r^2 = \sigma^2(4-\pi)/2$.

$$\sigma_r^2 = \mu_{r^2} - (\mu_r)^2 = 2\,\sigma^2 - \sigma^2 \cdot \frac{\pi}{2} = \frac{4-\pi}{2} \cdot \sigma^2$$

☻

Eine gute Abschätzung für σ_r ist $\frac{2}{3}\,\sigma$.

Wie bei der Normalverteilung kann man sich fragen, wie gross die Wahrscheinlichkeit ist, dass eine Rayleighverteilte Zufallsvariable vom Mittelwert abweicht. Sei $n \in \mathbb{N}$. Berechnet man die Werte vom folgenden Integral

$$\mathbb{P}(r > n\sigma) = \int_{n\sigma}^\infty \frac{x}{\sigma^2}\, e^{-\frac{x^2}{2\,\sigma^2}}\,\mathrm{d}r$$

so kommt man auf folgende Tabelle:

n	$\mathbb{P}(r > n\sigma)$
0	100%
1	60.7%
2	13.5%
3	1.2%

iv Weibull Verteilung

Ein anschauliches Beispiel für die Anwendung der Weibull-Statistik ist die Ausfallwahrscheinlichkeit einer Kette. Das Versagen eines Glieds führt zum Festigkeitsverlust der ganzen Kette. Spröde Werkstoffe zeigen ein ähnliches Bruchverhalten. Es genügt ein Riss, der die kritische Risslänge überschreitet, um das Bauteil zu zerstören.

Die Dichtefunktion der Weibull-Verteilung Wei(α,β), $\alpha,\beta > 0$ ist gegeben durch

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1}\mathrm{e}^{-\alpha x^\beta} & \text{für } x > 0, \\ 0 & \text{sonst,} \end{cases}$$

und ihre Verteilungsfunktion lautet

$$F(x) = 1 - \mathrm{e}^{-\alpha x^\beta}, \quad x > 0.$$

Das Maximum der Dichtefunktion wird in $(\frac{\beta-1}{\alpha\beta})^\beta$ angenommen.

Verteilungen für verschiedene Parameter sind in Bild A.7 illustriert.

Die Weibullverteilung

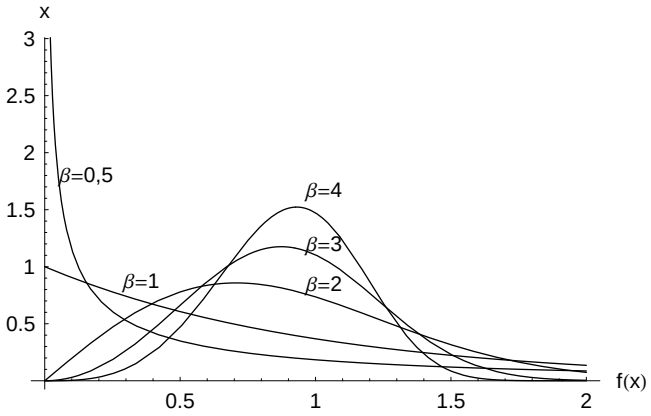


Abbildung A.7: Die Dichtefunktion ,für verschiedene Parameter

Die Grafik A.7 zeigt die Dichtefunktionen der Weibull-Verteilung für verschiedene Werte von α . Man sieht, dass der Fall $\alpha = 1$ die Exponentialverteilung ergibt. Für $\alpha \leq 1$ ist die Dicht streng monoton abfallend.

Für $\alpha = 3,4$ ergibt sich eine Verteilung ähnlich der Normalverteilung.

Den *Erwartungswert* ist gegeben durch

$$\mathbb{E}(X) = \alpha^{-1/\beta} \cdot \Gamma\left(\frac{1}{\beta} + 1\right).$$

Die Varianz ist gegeben durch

$$\sigma_X = \alpha^{-2/\beta} \cdot \left(\Gamma\left(\frac{2}{\beta} + 1\right) - \Gamma\left(\frac{1}{\beta} + 1\right)^2\right),$$

wobei Γ die Gammafunktion[4] bezeichnet. Weiters gilt

$$\text{Wei}\left(\frac{1}{2\sigma^2}, 2\right) = \text{Rayleigh}(\sigma^2)$$

Die Weibull-Verteilung ist eine statistische Verteilung, die beispielsweise zur Untersuchung von Lebensdauern in der Qualitätssicherung verwendet wird. Man verwendet sie vor allem bei Fragestellungen wie Materialermüdungen von spröden Werkstoffen oder Ausfällen von elektronischen Bauteilen, ebenso bei statistischen Untersuchungen von Windgeschwindigkeiten. Benannt ist sie nach dem Schweden Waloddi Weibull (1887-1979).
Die deutschen Ingenieure Rammler und Rosin benutzten sie schon in den 30 Jahren als Kerngrößenverteilung.
In der Schadensmodellierung wird die Weilbul Verteilung in der Regel nur für $k < 1$ verwendet, und zwar zur Abbildung von Großschäden, z.B. im Industriebereich, in der Kfz-Haftpflichtversicherung und Rückversicherung. Für $k > 1$ eignet sie sich nur für die Modellierung von Kleinschäden.

[4]Die Gammafunktion Γ ist eine spezielle Funktion in den mathematischen Teilgebieten der Analysis und der Funktionentheorie. Die Gammafunktion kann für positive reelle Zahlen x durch $\Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t} \mathrm{d}t$.

v Log-Normal Verteilung

Betrachten wir einen stochastischen Prozess der an einen Punkt anfängt und deren prozentualer Zuwachs normalverteilt ist. Genauer sei $\{r_i : i = 1, \dots n\}$ eine Familie von normalverteilten Zustandsvariablen mit Erwartungswert 0 und Varianz $1/n$. Sei $\{x_i : i = 1, \dots n\}$ ein diskreter stochastische Prozess mit

$$x_{i+1} = x_i + r_i x_i, \quad i = 0, \dots n-1.$$

Setzt man $\Delta x_i = \frac{x_{i+1} - x_i}{x_i}$ so gilt ja

$$r_1 + r_2 + \cdots + r_n = \frac{\Delta x_1}{x_1} + \frac{\Delta x_2}{x_2} + \cdots + \frac{\Delta x_n}{x_n}.$$

Läßt man n gegen unendlich gehen wissen wir aus den schwachen Gesetz der großen Zahlen dass gilt

$$\lim_{n\to\infty} (r_1 + r_2 + \cdots + r_n) \to \mathcal{N}(0, 1).$$

Auf der anderen Seite gilt

$$\begin{aligned} \lim_{n\to\infty} (r_1 + r_2 + \cdots + r_n) &= \lim_{n\to\infty} \left(\frac{\Delta x_1}{x_1} + \frac{\Delta x_2}{x_2} + \cdots + \frac{\Delta x_n}{x_n} \right) \\ &= \int_{x_0}^{x_n} \frac{du}{u} = \ln\left(\frac{x_n}{x_0}\right). \end{aligned}$$

Sei $Z = \lim_{n\to\infty}(r_1 + r_2 + \cdots + r_n)$ und $X = \lim_{n\to\infty} \frac{x_n}{x_0}$. Dann gilt $Z = \ln(X)$ und die Verteilung von X ist die Log-Normal Verteilung mit Parameter 0 und 1.

Sei f_Z und g_X die Dichte Funktion von Z und X. Mit Hilfe der Beziehung

$$f_Z \, dz = g_X \, dx$$

kann man zeigen dass gilt

$$g_X(x; \mu_Z, \sigma_Z) = \frac{1}{\sqrt{2\pi}\sigma_Z} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln(x) - \mu_Z}{\sigma_Z}\right)^2}, \quad x, \sigma_Z > 0, \tag{A.11}$$

wobei μ_Z und σ_Z der Erwartungswert und die Standardabweichung der Zufallsvariablen Z ist.

Der Erwartungswert lautet

$$\mathbb{E}X = e^{\mu_Z + \frac{\sigma_Z^2}{2}},$$

die Varianz lautet

$$\mathrm{Var}[X] = e^{2\mu_Z + \sigma_Z^2} \left(e^{\sigma_Z^2} - 1 \right).$$

Die Schiefe der Verteilung lautet

$$\gamma(X) = \sqrt{e^{\sigma_Z^2} - 1} \cdot \left(e^{\sigma_Z^2} + 2 \right).$$

Verteilungen für verschiedene Parameter sind in Bild A.8 illustriert.

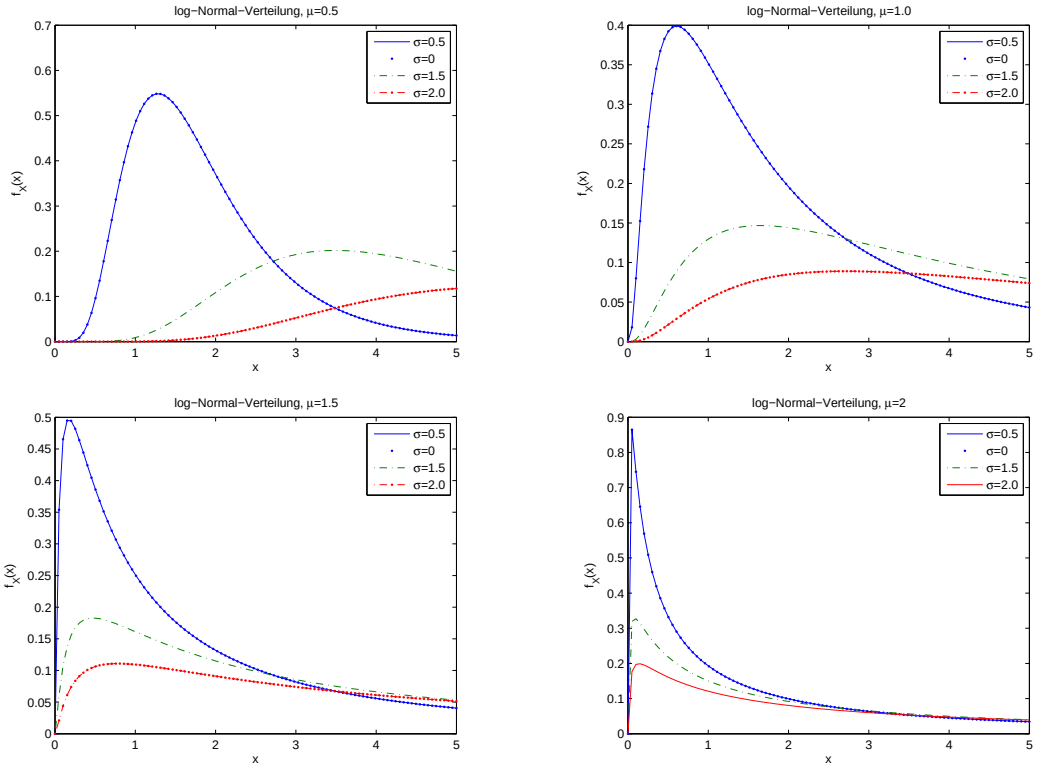


Abbildung A.8: Die Log-Normal-Verteilung.

Die Log Normalverteilung wird häufig angewendet. Da sie rechtsschief ist und die meisten empirischen Verteilungen diese Eigenschaften ebenfalls haben, benutzt man sie gern als Approximation, z.B. im Zusammenhang mit Lebensdauern, Belastungen oder Partikelparametern. Letztere Anwendungen ist sogar theoretisch begründet: Der berühmte russische Stochastiker Kolmogorov zeigte, dass fortlaufende Zerkleinerungsprozesse unter bestimmten Bedingungen asymptotisch zur Log Normalverteilung führen.
Da die Log-Normal Verteilung asymmetrisch und nach rechts schief verlagert ist, wird sie in der Versicherungsmathematik oft benutzt um die Unsicherheit der Ausfallraten von Maschinen zu modellieren. So werden Großschäden von Maschinen mit ihr modelliert.

vi Erlang und Gamma Verteilung

Die Exponentialverteilung und die Erlang Verteilung sind Spezialfälle der Gamma Verteilung. Die Dichtefunktion einer Gamma-verteilten Zustandsvariable $X \sim \Gamma(k;\lambda)$ mit reellwertigen Parametern $k > 0$ und $\lambda > 0$ hat für $x > 0$ die Gestalt

$$f(x) = \frac{\lambda^k}{\Gamma(k)} \cdot x^{k-1} \cdot e^{-\lambda x}. \tag{A.12}$$

Dabei ist

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \tag{A.13}$$

die sogenannte Gamma Funktion.

Der Erwartungswert lautet

$$\mathbb{E}X = \frac{k}{\lambda},$$

die Varianz lautet

$$\mathrm{Var}[X] = \frac{k}{\lambda^2}.$$

Die Schiefe der Verteilung lautet

$$\gamma(X) = \frac{2}{\sqrt{k}}.$$

Für ganzzahliges k wird die Gamma Verteilung auch als Erlang Verteilung bezeichnet. Für $k = 1$ ist es die exponential Verteilung.

Die Summe zweier unabhängig Γ-verteilter Zufallsvariablen $X_1 \sim \Gamma(k_1, \lambda)$, $X_2 \sim \Gamma(k_2, \lambda)$ $(k_1, k_2 > 0)$ mit demselben Verteilungsparameter $\lambda > 0$ ist selbst auch Γ verteilt mit Parameter λ und es gilt $X_1 + X_2 \sim \Gamma(k_1 + k_2, \lambda)$. Insbesondere lässt sich für k ganzzahlig eine Γ-verteilte Zufallsvariable als k–fache Summe von k unabhängigen exponentialverteilter Zufallsvariablen interpretieren.

Verteilungen für verschiedene Parameter sind in Bild A.9 illustriert.

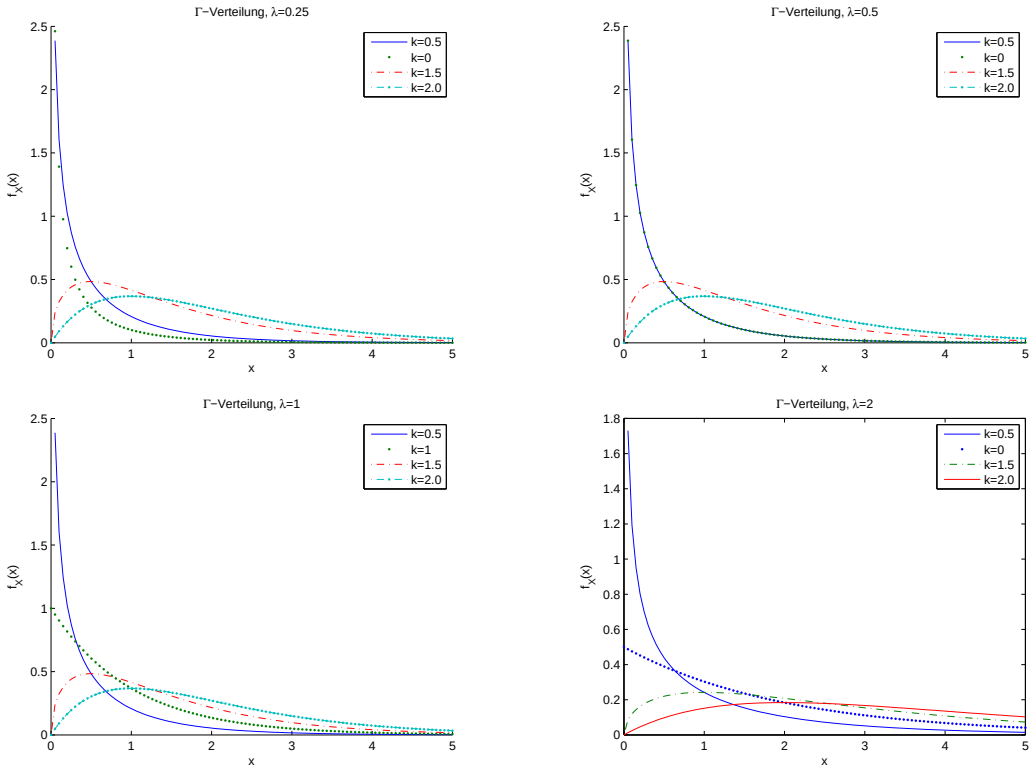


Abbildung A.9: Die allgemeine Gamma Verteilung.

Die Gamma Verteilung wird häufig zur Modellierung kleiner bis mittlerer Schäden verwendet, z.B. in der Hausrats- , Gewerbe-, Kfz-Kasko- und Kfz-Haftpflichtversicherung. Sie ist wegen ihrer zwei Parameter recht flexibel.

vii Die t-Verteilung

Die t-Verteilung spielt in der Statistik normalverteilter Zufallsvariablen eine wichtige Rolle und wird auch zur Modellierung von Risiken eingesetzt. Eine Zufallsvariable X ist t-verteilt mit $\nu > 0$ Freiheitsgraden (man schreibt $X \sim t_\nu$), wenn ihre Dichtefunktion f_X folgendermaßen gegeben ist

$$f_X(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} \cdot \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \tag{A.14}$$

Dabei bezeichnet Γ wieder die Γ funktion definiert in (**??**).

Wenn die Zufallsvariablen Y Normalverteilt, $Y \sim \mathcal{N}(0,1)$ und $Z\chi^2_\nu$ unabhängig sind, kann man zeigen das der Quotient

$$T = \frac{Y}{\sqrt{Z/\nu}}$$

die Verteilung t_ν besitzt.

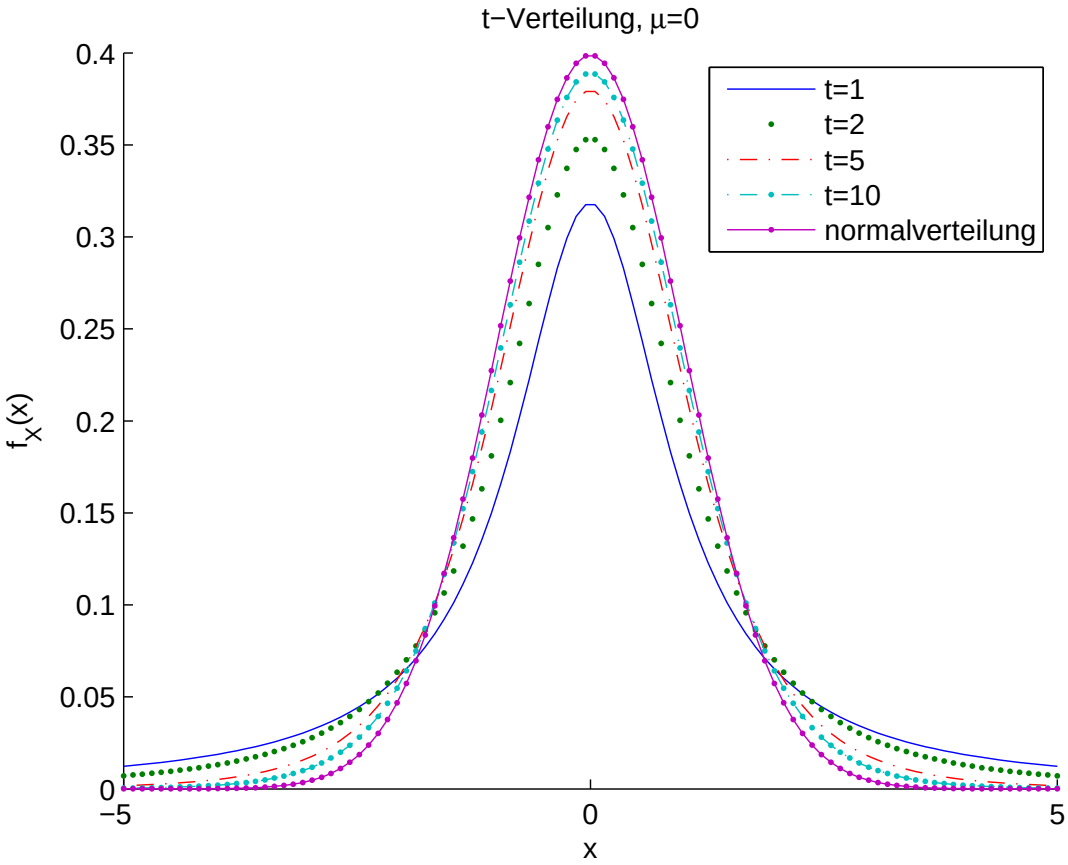


Abbildung A.10: Die t-Verteilung und Normalverteilung zum Vergleich, $\mu = 0$.

In der Abbildung **??** sind die Dichtefunktionen der t-Verteilung für verschiedene Freiheitsgrade sowie die Standardnormalverteilung dargestellt.

Offensichtlich fallen die Dichtefunktionen der t-Verteilung langsamer gegen null ab, als die der Normalverteilung. Die t-Verteilung besitzt also größere Tail-Wahrscheinlichkeiten als die Normalverteilung, somit lassen sich dadurch Risiken modellieren, bei denen sehr große Gewinne oder Verluste häufiger auftreten als unter der Normalverteilungsannahme. Dieser Effekt ist umso größer, je kleiner die Anzahl der Freiheitsgrade ν ist. Für große ν nähert sich die t-Verteilung immer mehr der Standardnormalverteilung an.
In der Statistik tritt die t Verteilung in natürlicher Weise für Mittelwertschätzer (unbekannte Varianz) auf.

viii Log-Gamma Verteilung

Ähnlich der Lognormalverteilung aus der Normalverteilung ergibt sich die Log-Gammaverteilung aus der Γ-Verteilung. Sie wird erzeugt durch $X = \exp(Y)$ mit $Y \sim \Gamma(k, \lambda)$. Die Dichtefunktion einer Log-gamma verteilten Zufallsvariable $X \sim \mathbf{LN}\Gamma(k; \lambda)$ mit den reellwertigen Parametern $k > 0$ und $\lambda > 0$ hat die Gestalt

$$f_X(x) = \begin{cases} \frac{\lambda^k}{\Gamma(k)} \cdot (\ln(x))^{k-1} \cdot x^{-\lambda-1} & \text{für } x > 1 \\ 0 \text{ sonst.} \end{cases} \tag{A.15}$$

Der Erwartungswert lautet

$$\mathbb{E}X = \left(1 + \frac{1}{\lambda}\right)^{-k},$$

die Varianz lautet

$$\mathrm{Var}[X] = \left(1 + \frac{2}{\lambda}\right)^{-k} - \left(1 + \frac{1}{\lambda}\right)^{-2k}.$$

Die Log-Gammaverteilung wird zur Modellierung von Großschäden benutzt. Mit $Y = X - 1$ erhält man einen Verteilung auf $(0, \infty)$; mit $Y = M \cdot X$ erhält man eine Verteilung auf (M, ∞), d.h. mit der Log Gammaverteilung lassen sich Schäden oberhalb eines vorgegebenen Schwellenwertes modellieren.

ix Pareto Verteilung

Die Pareto-Verteilung liefert nut oberhalb eines Schwellenwertes x_0 positive Wahrscheinlichkeiten. Die Dichtefunktion f_X und die Verteilungsfunktion F_X einer Pareto verteilten Zufallsvariablen $X \sim Pareto(x_0; a)$ mit den reelwertigen Parametern $x_0 > 0$, $\theta > 0$ und $a > 0$ lauten

$$f_X(x) = \begin{cases} a \cdot x_0^a \cdot (x - \theta)^{-a-1} & \text{für } x > \theta \\ 0 \text{ sonst.} \end{cases} \tag{A.16}$$

und

$$F_X(x) = \begin{cases} 1 - \left(\frac{x}{x_0}\right)^{-a} & \text{für } x > a \\ 0 \text{ sonst.} \end{cases} \tag{A.17}$$

Der Erwartungswert lautet

$$\mathbb{E}X = x_0 \cdot \frac{a}{a-1} \quad (\text{falls } a > 1),$$

die Varianz lautet

$$\mathrm{Var}[X] = x_0^2 \cdot \frac{a}{(a-1)^2(a-2)} \quad (\text{falls } a > 2).$$

Die Schiefe lautet

$$\gamma(X) = 2\frac{\sqrt{a-2}(a+1)}{\sqrt{a}(a-3)} \quad (\text{falls } a > 3).$$

x Die Gumpel Verteilung und andere Extremalverteilungen

Eine Zufallsvariable X ist Gumpel verteilt mit Parameter a und b, falls für die Verteilungsfunktion F_X gilt

$$F_X(x) = \exp(-e^{-(x-b)/a}) \tag{A.18}$$

Ist $a = 1$ und $b = 0$ schreiben wir $\Lambda = F_X$. Ist $b = 0$, $a = 1$ und U eine gleichverteilte Zufallsvariable, dann gilt $X = -\ln(-\ln(U))$. Weiters gilt für die Dichte Funktion

$$f_X(x) = \frac{1}{a}\, e^{-(x-b)/a}\, \exp(-e^{-(x-b)/a}) \tag{A.19}$$

Die Gumpelverteilung ist eine der drei für die Extremwerttheorie wichtigen Wahrscheinlichkeitsverteilungen. Sie wurde benannt nach Emil Julius Gumbel. Wie die Rossi-Verteilung und die Frechet-Verteilung gehört sie zu den Extremwertverteilungen, die den in einem Zeitraum T zu erwartenden höchsten Messwert berechnen. Sie wird u.ain folgenden Bereichen benutzt:

- Wasserwirtschaft für extreme Ereignisse wie Hochwasser und Trockenzeiten
- Verkehrsplanung
- Meteorologie (Wettervorhersage)
- Hydrologie.

Die Gumbel Verteilung ist eine typische Verteilungsfunktion für jährliche Serien. Sie kann nur auf Reihen angewendet werden, bei denen die Länge der Meßreihe mit dem Stichprobenumfang übereinstimmt. Ansonsten erhält man negative Logarithmen.

Man kann folgenden Sachverhalt zeigen: Seien X_i unabhängige identisch verteilte Zufallsvariabeln mit Verteilungsfunktion $F \sim \mathbf{Gump}(a, b)$. Sei

$$M_n = \max(X_1, X_2, \ldots, X_n).$$

Dann gilt für das Maximum

$$\mathbb{P}\,(M_n \leq x) = \exp(-e^{-(x-b-a\ln(n))/a}).$$

xi Die Cauchy Verteilung

Der Quotient zweier normalverteilter Zufallszahlen ist Cauchy-verteilt. Die Cauchy-Verteilung tritt bei der Brown'schen Molekularbewegung in Erscheinung (random walk). Der Mittelwert und die Varianz existieren nicht.

Weiters ist die Cauchy Verteilung ein Spezialfall der t-Verteilung ($t = 1$).

Bei der Cauchy-Verteilung ist die Wahrscheinlichkeit für extreme Ausprägungen sehr groß. Sind die 1X mindestens 2,58, beträgt bei einer Cauchy-verteilten Zufallsvariablen die entsprechende Untergrenze

ca. 31. Möchte man die Auswirkung von Ausreißern in Daten auf statistische Verfahren untersuchen, verwendet man häufig Cauchy-verteilte Zufallszahlen in Simulationen.

xii Die Beta-Verteilung

Die Beta-Verteilung beschreibt Zufallsvariablen die nur im Interval $[0, 1]$ Werte annehmen kann. Die Dichtefunktion einer Betaverteilten Zufallsvariablen X mit Parametern $a, b \in (0, 1)$, $X \sim \beta(a, b)$, lautet

$$f_X(x) = \frac{1}{B(a,b)} \begin{cases} x^{a-1}(1-x)^{b-1} & \text{für } 0 \le x \le 1 \\ 0 \text{ sonst.} \end{cases} \tag{A.20}$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Der Erwartungswert lautet

$$\mathbb{E}X = \frac{a}{a+b}$$

die Varianz lautet

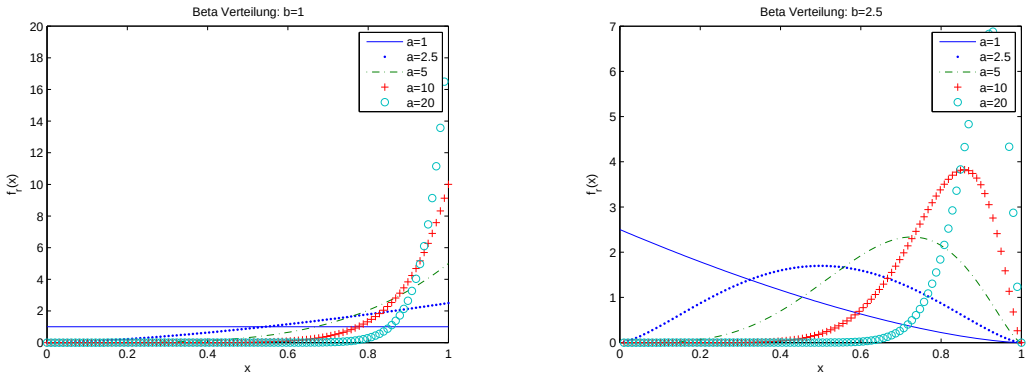
$$\mathrm{Var}[X] = \frac{ab}{(a+b)^2(a+b+1)}.$$

Ist Y eine t-Verteilte Zufallsvariable mit n Freiheitsgraden, dann ist die Zufallsvariable X mit

$$X = \frac{1}{2} + \frac{1}{2}\frac{Y}{\sqrt{\nu + Y^2}}$$

$\beta\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ verteilt.

Verteilungen für verschiedene Parameter sind in Bild A.11 illustriert.



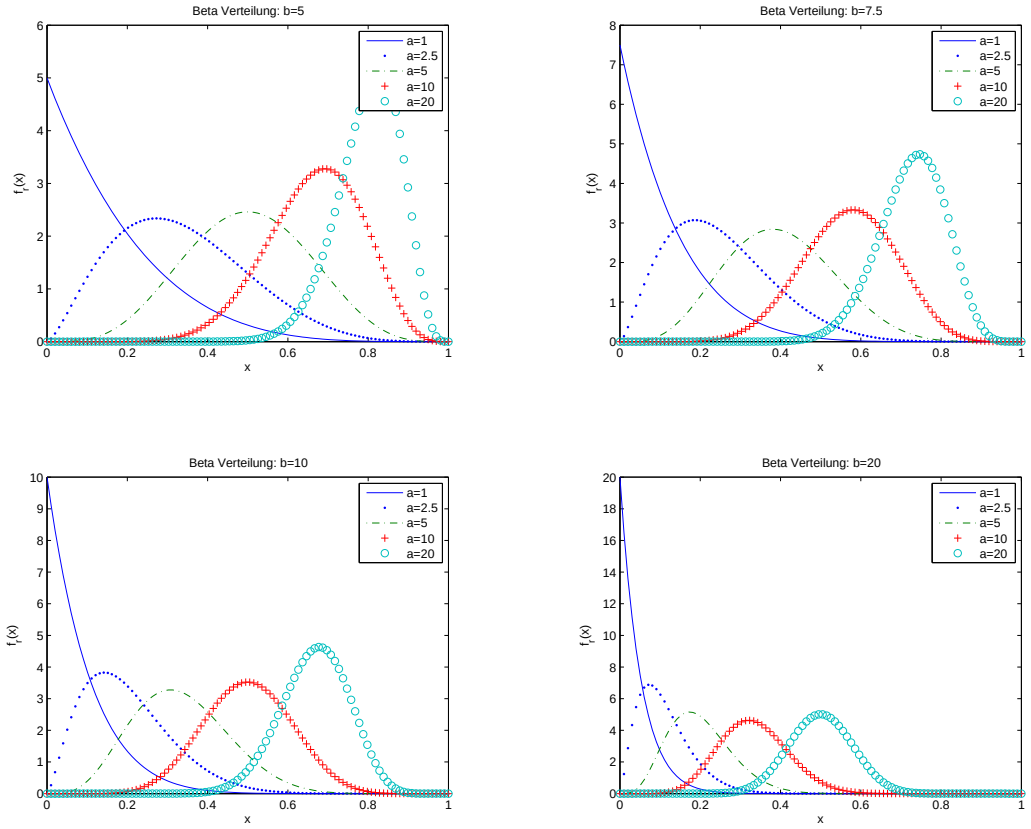


Abbildung A.11: Die Beta-Verteilung.

Wichtige Eigenschaften:

1. Gilt $X \sim \beta_{a,b}$, dann gilt $1/X \sim \beta_{b,a}$.
2. $\beta_{1,1}$ ist die Gleichverteilung.
3. Ist $X \sim \beta_{a,b}$, dann gilt $\sqrt{c+1}\frac{(x-\mu)}{\sqrt{\mu(1-\mu)}} \sim \mathcal{N}(0,1)$ für große $c = a + b$.

Mit der Beta Verteilung kann man Daten die eine untere und obere Schranke haben sehr gut beschreiben.

xiii Die χ-quadrat Verteilung

Die χ-quadrat Verteilung tritt in erster Linie in der Statistik auf. Sie ist ein Spezialfall der Γ Verteilung für $b = 2$. Eine Zufallsvariable X ist χ-quadrat verteilt mit Freiheitsgraden $n \in \mathbb{N}$, $X \sim \chi^2_n$, falls gilt

$$f_X(x) = \chi^2_{n,b}(x) = \begin{cases} \frac{1}{2^n \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{für } 0 \leq x \\ 0 \text{ sonst.} \end{cases} \tag{A.21}$$

Der Erwartungswert lautet

$$\mathbb{E}X = n,$$

die Varianz lautet

$$\mathrm{Var}[X] = 2n.$$

Die Verteilung ist in der Statistik wichtig. Hat mein eine Stichprobe mit n normalverteilten Zufallsvariablen, und s^2 deren empirische Varianz,

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

Weitere wichtige statistische Eigenschaften:

1. $\chi^2_n(x) = \gamma_{\frac{n}{2},2}(x)$. Speziell, falls $n = 2$ ist, erhält man die exponential Verteilung mit $\lambda = \frac{1}{2}$.
2. $X = \sum_{i=1}^n X_i^2$, wobei $X_i \sim \mathcal{N}(0,1)$ und unabhängig verteilt sind, dann gilt $X \sim \chi^2_n$.
3. $X = \sum_{i=1}^n (X_i \bar{X})^2$, wobei $X_i \sim \mathcal{N}(0,1)$ und unabhängig verteilt sind, dann gilt $X \sim \chi^2_{n-1}$.
4. $X = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \nu)^2$, wobei $X_i \sim \mathcal{N}(\mu,\sigma)$ und unabhängig verteilt sind, dann gilt $X \sim \chi^2_n$.
5. $X = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2$, wobei $X_i \sim \mathcal{N}(\mu,\sigma)$ und unabhängig verteilt sind, dann gilt $X \sim \chi^2_{n-1}$.
6. Gilt $X \sim \chi^2_n$, $Y \sim \chi^2_m$, dann gilt $X + Y \sim \chi^2_{n+m}$.

xiv Die F-Verteilung

Die F Verteilung kommt wiederum zumeist nur in der Statistik vor. Sie wurde von Ronald A. Fisher (1890-1962) eingeführt, um Quotienten von Varianzen zu studieren und ist nach ihm benannt. Der Quotienten zweier unabhängig Γ–verteilten Zufallsvariablen, jede durch ihre Varianz dividiert, ist F verteilt.

Die Dichtefunktion einer F-verteilten Zufallsgröße X lautet

$$f_X(x) = f_{n,m}(x) = \begin{cases} \frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \frac{x^{(n-2)/2}}{\left(1+\frac{n}{m}x^2\right)^{(n+m)/2}} & \text{für } 0 \le x \\ 0 \text{ sonst.} \end{cases} \quad \text{(A.22}$$

Der Erwartungswert lautet

$$\mathbb{E}X = \frac{m}{m-2} \quad (\text{falls } m > 2)\ ,$$

die Varianz lautet

$$\mathrm{Var}[X] = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)} \quad (\text{falls } m > 4)\ .$$

Weitere statistische Eigenschaften:

1. Ist $X_1 \sim \gamma_{a_1,p_1}, X_2 \sim \gamma_{a_2,p_2}$. Dann gilt $\frac{X_1/(a_1p_1)}{X_2/(a_2p_2)} \sim f_{2a_1,2a_2}$.
2. Ist $X \sim \beta_{a,b}$, dann gilt $\frac{X/a}{(1-X)/b} = \frac{X/\mu}{(1-X)/(1-\mu)} \sim f_{2a,2b}$.
3. Ist $X \sim f_{a,b}$, dann gilt $1/X \sim f_{b,a}$.
4. Ist $X \sim \chi^2_{n_1}, Y \sim \chi^2_{n_2}$, X und Y unabhängig, dann gilt $\frac{X/n_1}{Y/n_2} \sim f_{n_1,n_2}$.

5. Seinen $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ $n+m$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(\mu_1, \sigma_1)$ und $Y_i \sim \mathcal{N}(\mu_2, \sigma_2)$, dann gilt
$$\frac{\left(\sum_{i=1}^{n}(X_i - \mu_1)^2\right)/(n\sigma_1)}{\left(\sum_{i=1}^{n}(Y_i - \mu_2)^2\right)/(m\sigma_2)} \sim f_{n,m}$$

6. Seinen $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ $n+m$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(\mu_1, \sigma_1)$ und $Y_i \sim \mathcal{N}(\mu_2, \sigma_2)$, dann gilt
$$\frac{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)/((n-1)\sigma_1)}{\left(\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right)/((m-1)\sigma_2)} \sim f_{n-1,m-1}$$

xv Die Parteo-Verteilung

Die Pareto-Verteilung, benannt nach Vilfredo Pareto (1848–1923) ist eine stetige Wahrscheinlichkeitsverteilung auf einem rechtsseitig unendlichen Intervall $[x_{\min}, \infty)$. Sie ist skaleninvariant und genügt einem Potenzgesetz. Für kleine Exponenten gehört sie zu den *heavy tailed* Verteilungen. Die Verteilung wurde zunächst zur Beschreibung der Einkommensverteilung Italiens verwendet. Paretoverteilungen finden sich charakteristischerweise dann, wenn sich zufällige, positive Werte über mehrere Größenordnungen erstrecken und durch das Einwirken vieler unabhängiger Faktoren zustande kommen.

Eine stetige Zufallsvariable X heißt pareto-verteilt $\mathrm{Par}(k, x_{\min})$ mit den Parametern $k > 0$ und $x_{\min} > 0$, wenn sie die Wahrscheinlichkeitsdichte
$$f(x) = \begin{cases} \dfrac{k}{x_{\min}} \left(\dfrac{x_{\min}}{x}\right)^{k+1} & x \geq x_{\min} \\ 0 & x < x_{\min} \end{cases}$$
besitzt. Dabei ist $x_{\min}$ ein Parameter, der den Mindestwert der Verteilung beschreibt. Dieser ist auch gleichzeitig der Modus der Verteilung, also die Maximalstelle der Wahrscheinlichkeitsdichte. Mit steigendem Abstand zwischen x und $x_{\min}$ sinkt die Wahrscheinlichkeit, dass X den Wert x annimmt. Der Abstand zwischen den beiden Werten wird als Quotient, das heißt als Verhältnis zwischen beiden Größen, bestimmt. k ist ein Parameter, der das Größenverhältnis der Zufallswerte in Abhängigkeit von ihrer Häufigkeit beschreibt. Mit k wird der Quotient potenziert. Bei einem größeren k verläuft die Kurve deutlich steiler, das heißt, die Zufallsvariable X nimmt große Werte mit geringerer Wahrscheinlichkeit an.

Die Wahrscheinlichkeit, dass die Zufallsvariable X Werte größer x annimmt, ist gegeben mit:
$$\mathbb{P}\{X > x\} = 1 - \mathbb{P}\{X \leq x\} = \left(\frac{x_{\min}}{x}\right)^k, \quad \forall x \geq x_{\min}.$$

Der Erwartungswert ergibt sich zu:
$$\mathbb{E}(X) = \begin{cases} x_{\min}\dfrac{k}{k-1} & k > 1, \\ \infty & k \leq 1. \end{cases}$$

Varianz ist nicht für alle $k > 0$ definiert, nur für $k > 2$. Dann gilt
$$\mathbb{E}X^2 = k(2-k)x_{min}^2.$$

4 Verteilungstests

Hat man ein Sample $\{x_1, x_2, \ldots, x_n\}$ gegeben, so kann man mittels dem Histogramm des Samples versuchen mittels Vergleich die richtige Verteilungstyp herauszufinden. Die Parameter kann man mittels einen

Maximum Liklihood Schätzer berechnen. Für die meisten gängigen Verteilungstypen sind die Schätzer in MatLab mit einen Befehl zu berechnen. Nun stellt sich die Frage, ist dies aber wirklich die richtige Verteilung. Mit Hilfe eines Verteilungstest kann man klären ob ein gegebener Sample $\{x_1, x_2, \ldots, x_n\}$ dieser einer bestimmten Verteilung genügt oder nicht.

Die zentrale Idee beim Entwurf solcher Tests ist der Vergleich der empirischen Verteilungsfunktion $\tilde{F}_X$ mit der hypothetischen (zu testenden) Verteilungsfunktion.

Nullhypothese und Alternativhypothese sind dabei:

- H_0: X hat Verteilungsfunktion $F_X(x)$
- H_1: X hat nicht Verteilungsfunktion $F_X(x)$

Die bekanntesten dieser Tests sind der *Kolmogorov Smirnov Test* für stetige verteilte Merkmale und der *Chi Quadrat Test*. Im weiteren werden hier noch den Wilconschen Rangsummentest and QQ-Plot hier vorgestellt.

i Chi Quadrat Test

Mit Hilfe der Stichprobe $\{x_1, x_2, \ldots, x_n\}$ können wir die zugehörigen empirische Verteilung bestimmen und sie der hypothetischen Verteilung gegenüberstellen.

1. Teile den Wertebereich in K Klassen ein d.h., man erhält dann $y_0 < y_1 < \cdots < y_K$ mit $y_0 = \min\{x_1, x_2, \ldots, x_n\}$, und $y_K = \min\{x_1, x_2, \ldots, x_n\}$.

2. Zähle für jede Klasse wie viele Elemente in diese Klasse fallen und halte dies in einen Vektor $A = (a_1, \ldots, A_K)$ fest. Also
$$A_j = \#\{x_i \in [y_{j-1}, y_j)\}.$$

3. Berechne die theoretische Klassenhäufigkeit durch
$$B_j = (F_X(y_j) - F_X(y_{j-1})) \cdot n$$
wobei F_X die zu testende Verteilungsfunktion ist.

4. Berechne den quadratischen Fehler
$$\Sigma = \sum_{j=1}^{K} \frac{(B_j - A_j)^2}{B_j}.$$
Die Variable Σ ist die Testvariable, von der man weiß dass sie für grosse n χ^2 verteilt ist mit Freiheitsgraden $K - 1$.

5. bestimme zu einem vorgegebenen Signifikanzniveau α einen kritischen Wert $c > 0$, der von der Testvariablen Σ nicht überschritten werden darf.

6. Ist Σ größer als c, verwerfe H_0 ansonsten nehme H_0 an.

Man sollte darauf achten, dass in jeder Klasse mindestens 5 Elemente zu finden sind. Ist die Verteilung abhängig von Parametern, und wurden diese Parameter vorher geschätzt, so sollte man für jeden Parameter einen Freiheitsgrad abziehen.

ii Kolmogorov Simirnov Test

Der Kolmogorv Smirnov test ist eine Test für stetige Merkmale. Die Idee des Testes beruht auf den Vergleich der mit einer Stichprobe vom Umfang n ermittelten empirischen Verteilungsfunktion $\hat{F}_X(x)$ mit der hypothetischen Verteilungsfunktion $F_X(x)$. Das Vergleichskriterium ist der dabei betragsmässige größte Abstand der Werte dieser beiden Funktionen

$$D_n = \max_{x \in \mathbb{R}} |F_X(x) - \hat{F}_X(x)|.$$

Die Ermittelung des Maximums wird dabei durch die Tatsache vereinfacht, dass die Verteilung stetig sein soll. In diesem Fall kann das Maximum nur an einer Sprungstellen der empirischen Verteilungsfunktion auftauchen.

Anschaulich ist klar, dass die Nullhypothese, dass die Daten aus einer Grundgesamtheit mit Verteilungsfunktion $F_X(x)$ kommen, wohl dann abgelehnt werden muss, wenn D_n zu groß ist.

Den kritischen Vergleichswert c wählt man zu einem Niveau α so, dass

$$P(D_n \leq c) = 1 - \alpha.$$

Kolmogorov hat die asymptotische Verteilung von $\sqrt{(}n)D_n$ ermittelt und gezeigt dass für $\lambda > 0$ gilt

$$\lim_{n \to \infty} P(\sqrt{n} D_n \leq \lambda) = \sum_{k=-\infty}^{k=\infty} (-1)^k e^{-2k^2\lambda^2}.$$

Der Reihenwert ist für $n > 50$ eine sehr gute Approximation. In Matlab können sie den Kolmogorov-Smirnov test mittels *kstest* aufrufen.

iii Wilxconschen Radgsummen text

Gegeben sind hier zwei Stichproben $X = \{x_1, x_2, \ldots, x_n\}$ und $Y = \{y_1, y_2, \ldots, y_m\}$. Zu klären ist hier die Fragestellung ob beide Stichproben der gleichen Grundgesamtheit angehören, das heißt, ob beide die gleiche Verteilung haben. Man kann diesen Test als Verteilungtest für eine Stichprobe $\{x_1, x_2, \ldots, x_n\}$ benützen indem man die Stichprobe $\{y_1, y_2, \ldots, y_m\}$ mittels Monte Carlo Simulation gemäss einer bestimmten Verteilung F erzeugt. Mittels den Rangsummentest wird festgestellt ob die Stichprobe $\{x_1, x_2, \ldots, x_n\}$ die gleiche Verteilung wie die Stichprobe $\{y_1, y_2, \ldots, y_m\}$ besitzt, also auch F als Verteilungsfunktion hat.

Dazu werden die Stichproben als erstes nach Ihrer Größe nach geordnet und nummeriert. D.h., man erhält zwei Listen

$$\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$$

und

$$\{y_{(1)}, y_{(2)}, \ldots, y_{(m)}\}$$

mit $x_{(i)} \leq x_{(i+1)}$ für $1 \leq i \leq n-1$ und $y_{(i)} \leq y_{(i+1)}$ für $1 \leq i \leq m-1$.

Hierbei werden die Unterschiede zwischen den erreichten Punktzahlen durch Einführung einer Rangliste und einer entsprechenden Bewertung, die sich aus dieser Rangliste ergibt, quantitativ herausgearbeitet.

Der aus den Rangfolgen resultierende Test basiert nun auf folgender Idee. Wären die Ergebnisse der Gruppe X deutlich schlechter als die der Gruppe Y, so wären die zugehörigen Ränge und in Folge dessen die Rangsumme R klein. Die Rangsumme läge dann in der Nähe des Minimalwertes (die X-Werte nehmen die ersten n Ränge ein). Umgekehrt, wäre die Vorlesungsgruppe deutlich besser, so wären die Ränge groß und die Rangsumme R in der Nähe des Maximalwertes nin + (s. dazu Übung 89). Ist kein Unterschied zu erkennen (dies entspricht der Nullhypothese), so sind die Ränge, die den Werten von Äßugeordnet werden, *gut durchmischt* und es sollte eine mittlere Rangsumme errechnet werden. Verschiebt man den Wertebereich der Rangsumme durch Subtraktion des Minimalwertes, so erhalten wir die tatsächlich für den Wilcoxon'schen Rangsummentest verwendete Testvariable

QQ-Plot