# Zuverlässigkeit

## 1. Teil

Prof. Erika Hausenblas

Montanuniversität Leoben, Österreich

10. Jänner 2017

# Inhalt

- Ein einzelner Bauteil
  - verschiedene Kenngrössen
  - Statistik der Kenngrössen
- Zuverlässigkeit grösserer Systeme
  - Komponenten in Serie geschaltet
  - Komponenten parallel geschaltet (Redundanz)
  - Kombinationen obiger Systeme
- Fehlerbaumanalyse
  - Qualitative Analyse von Fehlerbäumen
  - Minimal Cut's
- Maintainability
  - Systeme mit Reparatur
  - Modellierung von solcher Systemen mittels Markovketten
  - Monte–Carlo Simulation

# Inhalt

# Die Zuverlässigkeit eines Systems

$T_f$: Zeitpunkt, an dem das System ausfällt.

**Definition**

Die *Zuverlässigkeit* $R(t)$ eines Systems zu einem Zeitpunkt $t \geq 0$ ist die Wahrscheinlichkeit, dass im Zeitinterval $[0, t]$ kein Ausfall aufgetreten ist, mathematisch ausgedrückt,

$$\boxed{R(t) := \mathbb{P}\left(T_f > t\right), \quad t \geq 0.}$$

# Die Zuverlässigkeit eines Systems

$T_f$: Zeitpunkt, an dem das System ausfällt.

## Definition

Die *Zuverlässigkeit* $R(t)$ eines Systems zu einem Zeitpunkt $t \geq 0$ ist die Wahrscheinlichkeit, dass im Zeitinterval $[0, t]$ kein Ausfall aufgetreten ist, mathematisch ausgedrückt,

$$\boxed{R(t) := \mathbb{P}\left(T_f > t\right), \quad t \geq 0.}$$

## Definition

Die *Ausfallwahrscheinlichkeit* $F(t) = \mathbb{P}(T_f < t)$ in einem Intervall $[0, t]$ wird mit $F$ bezeichnet und kann mittels $R$ definiert werden, d.h.

$$\boxed{F(t) = 1 - R(t), \quad t \geq 0.}$$

# Einige Beispiele typischer Verteilungen

- **Die Exponential Verteilung mit Parameter $\lambda > 0$:** Die Dichtefunktion lautet
$$f(t) = \lambda \exp(-\lambda t), \quad t \geq 0.$$

# Einige Beispiele typischer Verteilungen

- **Die Exponential Verteilung mit Parameter $\lambda > 0$:** Die Dichtefunktion lautet
$$f(t) = \lambda \exp(-\lambda t), \quad t \geq 0.$$
- **Gamma Verteilung:** Die Dichtefunktion einer Gamma-verteilten Zustandsvariabler $X \sim \Gamma(k;\lambda)$ mit reellwertigen Parametern $k > 0$ und $\lambda > 0$ hat für $x > 0$ die Gestalt
$$f(t) = \frac{\lambda^k}{\Gamma(k)} \cdot t^{k-1} \cdot e^{-\lambda t}.$$

# Einige Beispiele typischer Verteilungen

- **Die Exponential Verteilung mit Parameter $\lambda > 0$:** Die Dichtefunktion lautet
$$f(t) = \lambda \exp(-\lambda t), \quad t \geq 0.$$
- **Gamma Verteilung:** Die Dichtefunktion einer Gamma-verteilten Zustandsvariabler $X \sim \Gamma(k;\lambda)$ mit reellwertigen Parametern $k > 0$ und $\lambda > 0$ hat für $x > 0$ die Gestalt
$$f(t) = \frac{\lambda^k}{\Gamma(k)} \cdot t^{k-1} \cdot e^{-\lambda t}.$$
- **Die Weibullverteilung:** Die Dichtefunktion der Weibull-Verteilung $\mathrm{Wei}(\alpha,\lambda)$, $\alpha, \beta > 0$ ist gegeben durch
$$f(t) = \begin{cases} \alpha\lambda^\alpha t^{\alpha-1} \mathrm{e}^{-(\lambda t)^\alpha} & \text{für } t > 0, \\ 0 & \text{sonst,} \end{cases}$$

# Einige Beispiele typischer Verteilungen

- **Die Exponential Verteilung mit Parameter $\lambda > 0$:** Die Dichtefunktion lautet
$$f(t) = \lambda \exp(-\lambda t), \quad t \geq 0.$$
- **Gamma Verteilung:** Die Dichtefunktion einer Gamma-verteilten Zustandsvariabler $X \sim \Gamma(k;\lambda)$ mit reellwertigen Parametern $k > 0$ und $\lambda > 0$ hat für $x > 0$ die Gestalt
$$f(t) = \frac{\lambda^k}{\Gamma(k)} \cdot t^{k-1} \cdot e^{-\lambda t}.$$
- **Die Weibullverteilung:** Die Dichtefunktion der Weibull-Verteilung $\mathrm{Wei}(\alpha,\lambda)$, $\alpha, \beta > 0$ ist gegeben durch
$$f(t) = \begin{cases} \alpha\lambda^\alpha t^{\alpha-1} \mathrm{e}^{-(\lambda t)^\alpha} & \text{für } t > 0, \\ 0 & \text{sonst,} \end{cases}$$
- **Die Lognormal Verteilung:** Ist $Y \sim \mathcal{N}(\mu,\sigma)$ dann ist $X$ gegeben durch $X = \exp(Y)$ Log Normal verteilt. Die Dichtefunktion lautet
$$f(t) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{t} e^{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2} & \text{für } t > 0 \\ 0 \text{ sonst.} \end{cases}$$

# Exkurs - Extremalverteilungen

Die Fréchet-verteilung
Eine Zufallsvariable $X$ ist *Fréchet verteilt* mit Parameter $a$, falls für die Verteilungsfunktion $F_X$ gilt

$$\Phi_a(x) := F_X(x) = \begin{cases} 0, & \text{falls } x \leq 0, \\ \exp(-x^{-a}), & \text{falls } x > 0. \end{cases}$$

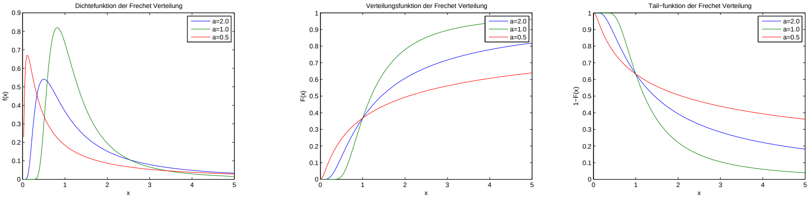



**Abbildung:** Die Dichte-, Verteilungs-, und Tailfunktion.

# Exkurs - Extremalverteilungen

## Die Weibull-verteilung

Eine Zufallsvariable $X$ ist Weibull verteilt mit Parameter $a$, falls für die Verteilungsfunktion $F_X$ gilt

$$\Psi_\alpha(x) := F_X(x) = \begin{cases} \exp(-(-x)^\alpha), & \text{falls } x \leq 0, \\ 1 & \text{falls } x > 0. \end{cases}$$

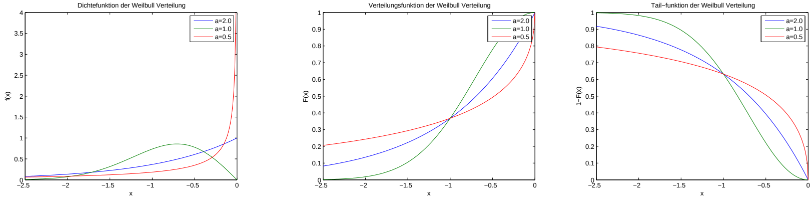



**Abbildung:** Die Dichte-, Verteilungs-, und Tailfunktion.

# Exkurs - Extremalverteilungen

Die Gumpel-verteilung Eine Zufallsvariable $X$ ist Gumpel verteilt, falls für die Verteilungsfunktion $F_X$ gilt

$$\Lambda(x) := F_X(x) = \exp(-e^{-x})$$

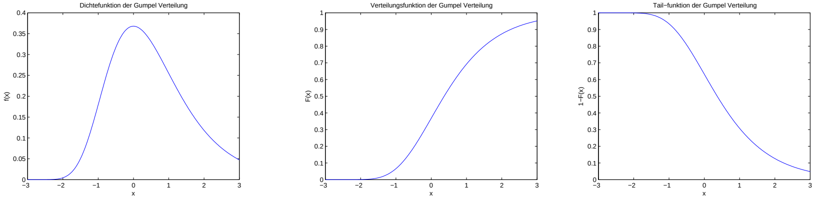



**Abbildung:** Die Dichte-, Verteilungs-, und Tailfunktion.

# Exkurs - Extremalverteilungen

## Theorem

*(Satz von Fisher-Liggett) Seien $\{X_n : n \in \mathbb{N}\}$ unabhängige identisch verteilte Zufallsvariabeln und sei*

$$M_n = \max\{X_1, X_2, \ldots, X_n\}, \quad n \in \mathbb{N}.$$

*Dann gibt es Konstanten $c_n > 0$ und $d_n$ sodass gilt*

$$c_n^{-1}(M_n - d_n) \to H, \quad n \to \infty$$

*für eine nicht degenerierte Zufallsvariable $H$. Dann genügt $H$ einer sogenannten Extremwertverteilung (d.h. Gumpel, Weilbull oder Fréchet Verteilung).*

# Exkurs - Extremalverteilungen

## Example

Seien $\{X_i : i \in \mathbb{N}\}$ unabhängig und exponential verteilt mit Parameter $\lambda$, d.h.

$$F(x) = 1 - e^{-\lambda x} \quad \text{für} \quad x > 0.$$

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 $\Lambda$.

Literatur: Embrechts, Klüppelberg,

# Weitere wichtige Kenngrössen

## Definition

Damit definieren wir die *momentane Fehlerrate* oder *momentane Ausfallrate* (im englischen *Hazardfunction*) als

$$\boxed{\lambda(t) = \frac{f(t)}{R(t)}.}$$

# Weitere wichtige Kenngrössen

## Definition

Damit definieren wir die *momentane Fehlerrate* oder *momentane Ausfallrate* (im englischen *Hazardfunction*) als

$$\boxed{\lambda(t) = \frac{f(t)}{R(t)}.}$$

## Definition

Erwartungswert der Ausfallzeit (englisch **MTTF**=mean time to failure)

$$\boxed{\mathbf{MTTF} = \mathbb{E}[T_f] = \int_0^\infty t f(t)\, dt.}$$

# Weitere wichtige Kenngrössen

(*mean residual life*):

$$\boxed{\mathbf{MRL}(t) = \mathbb{E}[T_f \mid \text{ Bauteil funktioniert zur Zeit } t], \quad t \geq 0.}$$

Die Formel von Bayes ergibt

$$\mathbf{MRL}(t) \;=\; \frac{\int_t^\infty (s-t)\, f(s)\, ds}{R(t)}.$$

# Weitere wichtige Kenngrössen

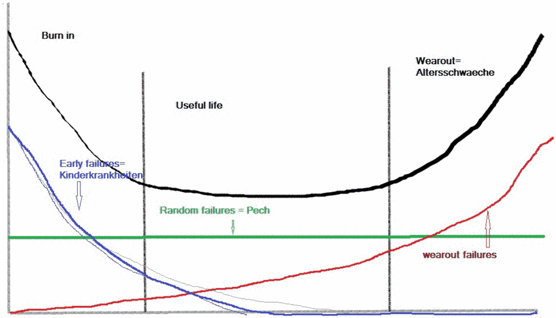
Gegeben sei eine Verteilungsfunktion $F$ mit Dichtefunktion $F$.
(*Mode einer Verteilung*):

$$\mathbf{t_{mode}} = \{t \in [0, \infty) : f(t) \geq f(s),\ s \geq 0\}.$$

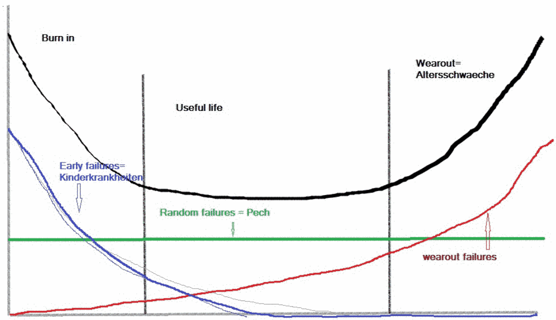
(*Median einer Verteilung*):

$$\mathbf{t_{median}} = \{t \in [0, \infty) : F(t) = 0.5\}.$$

# Die Badewannen Verteilung

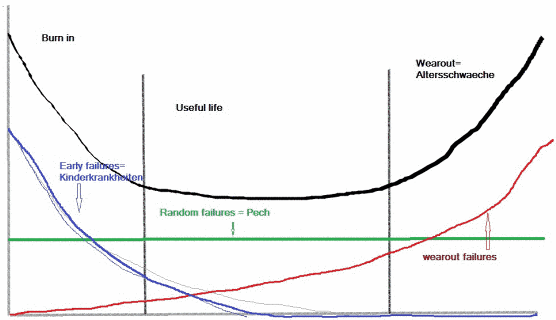


# Die Badewannen Verteilung



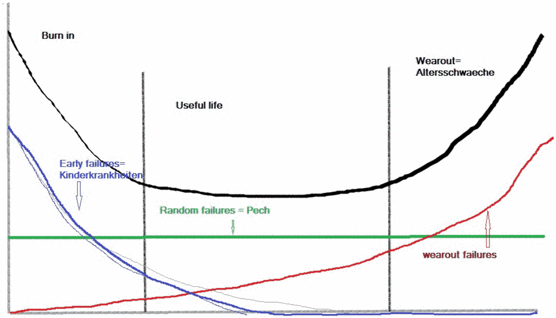

- **Burn in:** Fabrikationsfehler: Schlechte Schweißnähte, Sprünge, Riße, Ausschussteil, schlechte Qualitätskontrolle, Kontamination, . . . .

# Die Badewannen Verteilung




- **Burn in:** Fabrikationsfehler: Schlechte Schweißnähte, Sprünge, Riße, Ausschussteil, schlechte Qualitätskontrolle, Kontamination, . . . .
- **Useful life:** Durch die Umgebung verursachte zufällige Belastung, menschliches Versagen, *Acts of God*, Pech, . . .

# Die Badewannen Verteilung




- **Burn in**: Fabrikationsfehler: Schlechte Schweißnähte, Sprünge, Riße, Ausschussteil, schlechte Qualitätskontrolle, Kontamination, . . . .
- **Useful life**: Durch die Umgebung verursachte zufällige Belastung, menschliches Versagen, *Acts of God*, Pech, . . .
- **Wear Out**: Ermüdungserscheinigungen, Korrision, Altern, Spannung, . . . .

# Die Badewannen Verteilung

**Example**

Eine vereinfachte *bathtub* Kurve kann man durch eine linearisierte *hazard rate* modellieren: In diesen Fall wird $\lambda$ folgendermaßen modelliert ($\lambda > 0$ sei fix vorgegeben):

$$\lambda(t) = \begin{cases} c_0 - c_1 t + \lambda & \text{für } 0 \leq t \leq \frac{c_0}{c_1} \\ \lambda & \text{für } \frac{c_0}{c_1} < t \leq t_0 \\ c_2(t - t_0) + \lambda & \text{für } t > t_0 \end{cases}$$

Es gilt somit

$$R(t) = \begin{cases} \exp\left(-\{(c_0 + \lambda)t - c_1(t^2/2)\}\right) & \text{für } 0 \leq t \leq \frac{c_0}{c_1} \\ \exp\left(-\left\{\lambda t + \frac{c_0^2}{2c_1}\right\}\right) & \text{für } \frac{c_0}{c_1} < t \leq t_0 \\ \exp\left(-\left\{\frac{c_2}{2}(t - t_0)^2 + \lambda t + \frac{c_0^2}{2c_1}\right\}\right) & \text{für } t > t_0 \end{cases}$$

# Exponential verteilung

# Exponential verteilung

Die Dichtefunktion lautet

$$f(t) = \lambda_0 \exp(-\lambda_0 t), \quad t \geq 0,$$

Die Zuverlässigkeitsfunktion lautet

# Exponential verteilung

Die Dichtefunktion lautet

$$f(t) = \lambda_0 \exp(-\lambda_0 t), \quad t \geq 0,$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = e^{-\lambda_0 t}, \quad t \geq 0,$$

und der Erwartungswert der Ausfallzeit

# Exponential verteilung

Die Dichtefunktion lautet

$$f(t) = \lambda_0 \exp(-\lambda_0 t), \quad t \geq 0,$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = e^{-\lambda_0 t}, \quad t \geq 0,$$

und der Erwartungswert der Ausfallzeit

$$\mathbb{E} T_f = \frac{1}{\lambda_0},$$

die zugehörige Varianz beträgt

# Exponential verteilung

Die Dichtefunktion lautet

$$f(t) = \lambda_0 \exp(-\lambda_0 t), \quad t \geq 0,$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = e^{-\lambda_0 t}, \quad t \geq 0,$$

und der Erwartungswert der Ausfallzeit

$$\mathbb{E} T_f = \frac{1}{\lambda_0},$$

die zugehörige Varianz beträgt

$$\mathbf{Var}(T_f) = \frac{1}{\lambda_0^2}.$$

# Exponential verteilung

Die Dichtefunktion lautet
$$f(t) = \lambda_0 \exp(-\lambda_0 t), \quad t \geq 0,$$
Die Zuverlässigkeitsfunktion lautet
$$R(t) = e^{-\lambda_0 t}, \quad t \geq 0,$$
und der Erwartungswert der Ausfallzeit
$$\mathbb{E} T_f = \frac{1}{\lambda_0},$$
die zugehörige Varianz beträgt
$$\mathbf{Var}(T_f) = \frac{1}{\lambda_0^2}.$$

- Die Ausfallrate ist konstant;

# Exponential verteilung

Die Dichtefunktion lautet

$$f(t) = \lambda_0 \exp(-\lambda_0 t), \quad t \geq 0,$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = e^{-\lambda_0 t}, \quad t \geq 0,$$

und der Erwartungswert der Ausfallzeit

$$\mathbb{E}\, T_f = \frac{1}{\lambda_0},$$

die zugehörige Varianz beträgt

$$\mathbf{Var}(T_f) = \frac{1}{\lambda_0^2}.$$

- Die Ausfallrate ist konstant;
- Die Exponential Verteilung ist gedächnislos

# Die Weibull Verteilung

$T_f \sim \mathrm{Wei}(\alpha, \lambda)$ mit Dichtefunktion

$$f(t) = \begin{cases} \alpha\lambda^\alpha t^{\alpha-1} \mathrm{e}^{-(\lambda t)^\alpha} & \text{für } t > 0, \\ 0 & \text{sonst,} \end{cases}$$

# Die Weibull Verteilung

$T_f \sim \mathrm{Wei}(\alpha, \lambda)$ mit Dichtefunktion

$$f(t) = \begin{cases} \alpha\lambda^\alpha t^{\alpha-1} \mathrm{e}^{-(\lambda t)^\alpha} & \text{für } t > 0, \\ 0 & \text{sonst,} \end{cases}$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = \int_t^\infty \lambda\alpha s^{\alpha-1} \exp(-(\lambda s)^\alpha)\, ds = \exp(-(\lambda t)^\alpha), \quad t \geq 0.$$

# Die Weibull Verteilung

$T_f \sim \mathrm{Wei}(\alpha, \lambda)$ mit Dichtefunktion

$$f(t) = \begin{cases} \alpha\lambda^\alpha t^{\alpha-1} \mathrm{e}^{-(\lambda t)^\alpha} & \text{für } t > 0, \\ 0 & \text{sonst,} \end{cases}$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = \int_t^\infty \lambda\alpha s^{\alpha-1} \exp(-(\lambda s)^\alpha)\, ds = \exp(-(\lambda t)^\alpha), \quad t \geq 0.$$

Die Ausfall Funktion lautet

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\lambda^\alpha \alpha t^{\alpha-1} \exp(-(\lambda t)^\alpha)}{\exp(-(\lambda t)^\alpha)} = \lambda^\alpha \alpha t^{\alpha-1}.$$

# Die Weibull Verteilung

$T_f \sim \mathrm{Wei}(\alpha, \lambda)$ mit Dichtefunktion

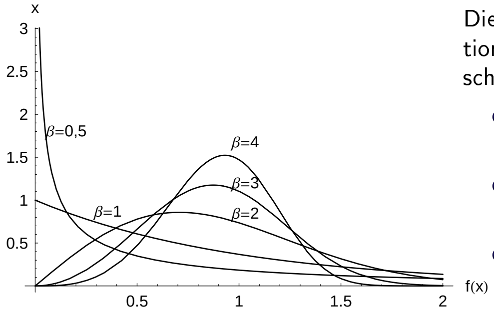
$$f(t) = \begin{cases} \alpha\lambda^\alpha t^{\alpha-1} \mathrm{e}^{-(\lambda t)^\alpha} & \text{für } t > 0, \\ 0 & \text{sonst}, \end{cases}$$

Die Zuverlässigkeitsfunktion lautet

$$R(t) = \int_t^\infty \lambda\alpha s^{\alpha-1} \exp(-(\lambda s)^\alpha)\, ds = \exp(-(\lambda t)^\alpha), \quad t \geq 0.$$

Die Ausfall Funktion lautet

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\lambda^\alpha \alpha t^{\alpha-1} \exp(-(\lambda t)^\alpha)}{\exp(-(\lambda t)^\alpha)} = \lambda^\alpha \alpha t^{\alpha-1}.$$




Die Grafik links zeigt die Dichtefunktionen der Weibull-Verteilung für verschiedene Werte von $\alpha = \beta$ .

- $0 < \alpha < 1$: $\lambda$ ist strikt monoton fallend.
- $\alpha = 1$: $\lambda(t) = \lambda$: Die Ausfallrate ist konstant.
- $\alpha > 1$: $\lambda(t)$ ist streng monoton steigend.

# Beispiel

Die Lebensdauer einer Ventilfeder einer Luftklappe ist Weilbull verteilt mit Parameter $\alpha = 2.25$ und $\lambda = 1.15 \cdot 10^{-4} h^{-1}$. Das Ventil wird 6 Monate ($t = 4380$ Stunden) durcharbeiten mit der Wahrscheinlichkeit

$$R(t) = e^{-(\lambda t)^{\alpha}} \sim 0.808.$$

# Beispiel

Die Lebensdauer einer Ventilfeder einer Luftklappe ist Weilbull verteilt mit Parameter $\alpha = 2.25$ und $\lambda = 1.15 \cdot 10^{-4} h^{-1}$. Das Ventil wird 6 Monate ($t = 4380$ Stunden) durcharbeiten mit der Wahrscheinlichkeit

$$R(t) = e^{-(\lambda t)^{\alpha}} \sim 0.808.$$

Dann gilt

$$\mathbf{MTTF} = \frac{\Gamma(1.44)}{1.15 \cdot 10^{-4}} h \sim 7706h.$$

und für den Median gilt $t_m \sim 7389h$.

# Beispiel

Die Lebensdauer einer Ventilfeder einer Luftklappe ist Weilbull verteilt mit Parameter $\alpha = 2.25$ und $\lambda = 1.15 \cdot 10^{-4} h^{-1}$. Das Ventil wird 6 Monate ($t = 4380$ Stunden) durcharbeiten mit der Wahrscheinlichkeit

$$R(t) = e^{-(\lambda t)^{\alpha}} \sim 0.808.$$

Dann gilt

$$\mathbf{MTTF} = \frac{\Gamma(1.44)}{1.15 \cdot 10^{-4}} h \sim 7706h.$$

und für den Median gilt $t_m \sim 7389h$.

Angenommen das Ventil arbeitet ohne Probleme die ersten 6 Monate. Es wird die nächsten 6 ohne Probleme durcharbeiten mit Wahrscheinlichkeit ($t_1 = t_2 = t = 4380$)

$$R(t_1 + t_2 \mid t_1) = \frac{R(t_1 + t_2)}{R(t_1)} = \frac{e^{-(\lambda(t_1+t_2))^{\alpha}}}{e^{-(\lambda t_1)^{\alpha}}} \sim 0.448.$$

Diese Wahrscheinlichkeit ist signifikant kleiner als die Wahrscheinlichkeit, dass das Ventil ohne Probleme die ersten 6 Monate durcharbeitet. Auch gilt

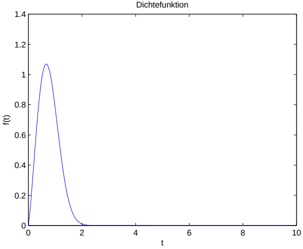
$$\mathbf{MRL}(t) = \frac{\int_t^{\infty} (s-t)\, f(s)\, ds}{R(t)} = \frac{1}{R(t)} \int_0^{\infty} R(t+x)\, dx \sim 4730h.$$

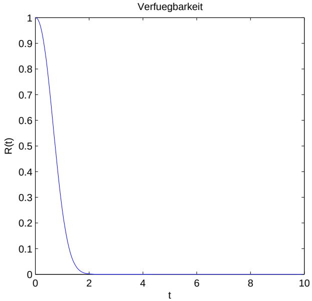
Beachte, $\mathbf{MRL}(t)$ kann nicht einfach ausgerechnet werden, und muss daher durch Näherungsrechnungen am Computer gefunden werden. Die Funktion

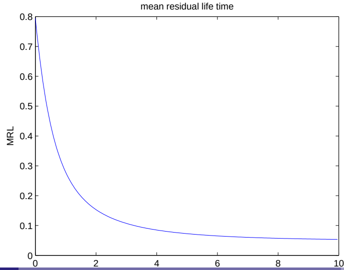
$$\mathbf{MRL}(t)$$

ist im nächsten Bild illustriert

# Zuverlässigkeit größerer Systeme



# Statistik

## Verschieden Modelle

- parameterfreie Statistik $\Rightarrow$ Man hat kein Vorwissen, man hat keine Vorinformation über den Typ der Verteilungsfunktion.

# Statistik

## Verschieden Modelle

- parameterfreie Statistik $\Rightarrow$ Man hat kein Vorwissen, man hat keine Vorinformation über den Typ der Verteilungsfunktion.
- parametrische Statistik $\Rightarrow$ Man modelliert die Verteilungsfunktion mit Hilfe einer Verteilungsfunktion bestimmten Typs, z.B. mit der Weibull Verteilung. Hier werden nur die Parameter geschätzt,

# Statistik

## Verschieden Modelle

- parameterfreie Statistik $\Rightarrow$ Man hat kein Vorwissen, man hat keine Vorinformation über den Typ der Verteilungsfunktion.
- parametrische Statistik $\Rightarrow$ Man modelliert die Verteilungsfunktion mit Hilfe einer Verteilungsfunktion bestimmten Typs, z.B. mit der Weibull Verteilung. Hier werden nur die Parameter geschätzt,

## Probleme:

Nicht vollständige Daten (incomplete Data):

# Statistik

## Verschieden Modelle

- parameterfreie Statistik $\Rightarrow$ Man hat kein Vorwissen, man hat keine Vorinformation über den Typ der Verteilungsfunktion.
- parametrische Statistik $\Rightarrow$ Man modelliert die Verteilungsfunktion mit Hilfe einer Verteilungsfunktion bestimmten Typs, z.B. mit der Weibull Verteilung. Hier werden nur die Parameter geschätzt,

## Probleme:

Nicht vollständige Daten (incomplete Data): $\Rightarrow$ Censoring - Fängt man zu messen an, weiß man zumeist nicht wann das Bauteil eingesetzt wurde, hört man auf, hat man keine Information wie lang das Bauteil noch funktioniert hat.

# Schätzen der *Zuverlässigkeitsfunktion*

**Gegeben:**

$$\{t_i : i = 1, \ldots, n\}, \text{ mit } t_i < t_{i+1}, \; i = 1, \ldots, n-1.$$

Annahme: $t_i$ sind unabhängig und identisch verteilt.

Zugehörige geordnete Stichprobe:

$$\{t^{(i)} : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)}, \; i = 1, \ldots, n-1.$$

# Schätzen der *Zuverlässigkeitsfunktion*

**Gegeben:**

$$\{t_i : i = 1, \ldots, n\}, \text{ mit } t_i < t_{i+1}, \; i = 1, \ldots, n-1.$$

Annahme: $t_i$ sind unabhängig und identisch verteilt.

Zugehörige geordnete Stichprobe:

$$\{t^{(i)} : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)}, \; i = 1, \ldots, n-1.$$

**Möglicher Schätzer von $R(t)$:**

$$\hat{R}(t^{(i)}) = \frac{n-i}{n} = 1 - \frac{i}{n}.$$

# Schätzen der *Zuverlässigkeitsfunktion*

**Gegeben:**

$$\{t_i : i = 1, \ldots, n\}, \text{ mit } t_i < t_{i+1},\ i = 1, \ldots, n-1.$$

Annahme: $t_i$ sind unabhängig und identisch verteilt.

Zugehörige geordnete Stichprobe:

$$\{t^{(i)} : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \ldots, n-1.$$

**Möglicher Schätzer von $R(t)$:**

$$\hat{R}(t^{(i)}) = \frac{n-i}{n} = 1 - \frac{i}{n}.$$

Für große Werte von $t$ ist $R(t) = 0$.

# Schätzen der *Zuverlässigkeitsfunktion*

**Gegeben:**

$$\{t_i : i = 1, \ldots, n\}, \text{ mit } t_i < t_{i+1}, \; i = 1, \ldots, n-1.$$

Annahme: $t_i$ sind unabhängig und identisch verteilt.

Zugehörige geordnete Stichprobe:

$$\{t^{(i)} : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)}, \; i = 1, \ldots, n-1.$$

**Möglicher Schätzer von $R(t)$:**

$$\hat{R}(t^{(i)}) = \frac{n-i}{n} = 1 - \frac{i}{n}.$$

Für große Werte von $t$ ist $R(t) = 0$. $\Rightarrow$ Unterschätzen der Zuverlässigkeit für $t$ groß.

# Schätzen der *Zuverlässigkeitsfunktion*

**Gegeben:**

$$\{t_i : i = 1, \ldots, n\}, \text{ mit } t_i < t_{i+1},\ i = 1, \ldots, n-1.$$

Annahme: $t_i$ sind unabhängig und identisch verteilt.

Zugehörige geordnete Stichprobe:

$$\{t^{(i)} : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \ldots, n-1.$$

# Schätzen der *Zuverlässigkeitsfunktion*

**Gegeben:**

$$\{t_i : i = 1, \ldots, n\}, \text{ mit } t_i < t_{i+1},\ i = 1, \ldots, n-1.$$

Annahme: $t_i$ sind unabhängig und identisch verteilt.

Zugehörige geordnete Stichprobe:

$$\{t^{(i)} : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \ldots, n-1.$$

**Möglicher Schätzer der Zuverlässigkeitsfunktion $R(t)$:**

$$\hat{R}(t^{(i)}) = 1 - \frac{i}{n+1} = \frac{n+1-i}{n+1}.$$

# Schätzen der *Zuverlässigkeitsfunktion* - zensierte Daten (censored data)

**Gegeben:**

Stichrpobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$.
Annahmen: Die Zufallsvariablen $t_i$, $i = 1, \ldots, n$ sind unabhängig und identisch verteilt. Die Zeiten sind verschieden.

$$\{(t^{(i)}, \delta_i) : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \ldots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.

# Schätzen der *Zuverlässigkeitsfunktion* - zensierte Daten (censored data)

## Gegeben:

Stichrpobe $\{t_i : t_i$ ist eine Ausfallzeit$\}$.
Annahmen: Die Zufallsvariablen $t_i$, $i = 1, \ldots, n$ sind unabhängig und identisch verteilt. Die Zeiten sind verschieden.

$$\{(t^{(i)}, \delta_i) : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \ldots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.

## Der Kaplan Meier Schätzer:

$$\hat{R}(t) = \prod_{j \in J_t} \frac{n_j - 1}{n_j},$$

wobei (MATLAB ecdf)

- $J_t = \{j = 1, \ldots n : t^{(j)} \leq t\}$,
- $n_j = \#\{j \in J_t : t_j$ ist ein Ausfall$\}$,

# Schätzen der *Zuverlässigkeitsfunktion* - zensierte Daten (censored data)

**Gegeben:**

Stichrpobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$.
Annahmen: Die Zufallsvariablen $t_i$, $i = 1, \ldots, n$ sind unabhängig und identisch verteilt. Die Zeiten sind verschieden.

$$\{(t^{(i)}, \delta_i) : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)}, \ i = 1, \ldots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.

**Der Kaplan Meier Schätzer:**

$$\hat{R}(t) = \prod_{j \in J_t} \frac{n_j - 1}{n_j},$$

wobei (MATLAB ecdf)

- $J_t = \{j = 1, \ldots n : t^{(j)} \leq t\}$,
- $n_j = \#\{j \in J_t : t_j \text{ ist ein Ausfall}\}$,

*Der Kaplan Meier Schätzer ist der Maximum Liklihood Schätzer.*

# Eigenschaften des Kaplan Meier Schätzer

**Motivation:**

Sei $s_0 < s_1 < \cdots < s_n$ eine Partition des Zeitintervalls $[0, \infty)$. Sei dabei $s_i - s_{i-1}$ so klein, dass immer nur ein Ereignis im Intervall $[s_i, s_{i+1})$ stattfindet. Dann gilt

$$
\begin{aligned}
R(t) &= \mathbb{P}(T > t) \\
&= \mathbb{P}(T > s_0) \cdot \mathbb{P}(T > s_1 \mid T > s_0) \cdot \mathbb{P}(T > s_2 \mid T > s_1) \\
&\quad \cdots \mathbb{P}(T > s_n \mid T > s_{n-1}) \\
&= \mathbb{P}(T > s_1 \mid T > s_0) \cdot \mathbb{P}(T > s_2 \mid T > s_1) \cdots \mathbb{P}(T > s_n \mid T > s_{n-1}) \\
&= \prod_{j=0}^{n} p_j,
\end{aligned}
$$

wobei gilt
$p_j = \mathbb{P}(T > s_j \mid T > s_{j-1}) = \frac{n_j - 1}{n_j}$, $j = 1, \cdots n-1$, und $p_n = \mathbb{P}(T > t \mid T > s_n)$.

# Eigenschaften des Kaplan Meier Schätzer

**Motivation:**

Sei $s_0 < s_1 < \cdots < s_n$ eine Partition des Zeitintervalls $[0, \infty)$. Sei dabei $s_i - s_{i-1}$ so klein, dass immer nur ein Ereignis im Intervall $[s_i, s_{i+1})$ stattfindet. Dann gilt

$$\begin{aligned}
R(t) &= \mathbb{P}(T > t) \\
&= \mathbb{P}(T > s_0) \cdot \mathbb{P}(T > s_1 \mid T > s_0) \cdot \mathbb{P}(T > s_2 \mid T > s_1) \\
&\quad \cdots \mathbb{P}(T > s_n \mid T > s_{n-1}) \\
&= \mathbb{P}(T > s_1 \mid T > s_0) \cdot \mathbb{P}(T > s_2 \mid T > s_1) \cdots \mathbb{P}(T > s_n \mid T > s_{n-1}) \\
&= \prod_{j=0}^{n} p_j,
\end{aligned}$$

wobei gilt
$p_j = \mathbb{P}(T > s_j \mid T > s_{j-1}) = \frac{n_j - 1}{n_j}$, $j = 1, \cdots n-1$, und $p_n = \mathbb{P}(T > t \mid T > s_n)$.
Ein Schätzer von $p_j$ ist gegeben durch

$$\hat{p}_j = \frac{n_j - 1}{n_j}.$$

# Schätzen der *Zuverlässigkeitsfunktion* - zensierte Daten (censored data)

**Gegeben:**

Stichrpobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$.
Annahmen: Die Zufallsvariablen $t_i$, $i = 1, \dots, n$ sind unabhängig und identisch verteilt. Die Zeiten sind verschieden.

$$\{(t^{(i)}, \delta_i) : i = 1, \dots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \dots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.

# Schätzen der *Zuverlässigkeitsfunktion* - zensierte Daten (censored data)

**Gegeben:**

Stichrpobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$.
Annahmen: Die Zufallsvariablen $t_i$, $i = 1, \ldots, n$ sind unabhängig und identisch verteilt. Die Zeiten sind verschieden.

$$\{(t^{(i)}, \delta_i) : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)}, \; i = 1, \ldots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.

**Varianz des Kaplan Meier Schätzer (Formel nach Greenwood):**

$$\hat{\mathbf{Var}}(\hat{R}(t)) = (\hat{R}(t)) \sum_{j \in J_t} \frac{1}{n_j(n_j - 1)},$$

# Nelson Schätzer für die kumulative Ausfallrate

**Die kumulative Ausfallrate:**

Die kumulative Ausfallrate ist gegeben durch

$$Z(t) \quad = \quad \int_0^t \lambda(s)\, ds = -\ln(R(t)).$$

Nimmt man den Kaplan Meier Schätzer für die Zuverlässigkeitsfunktion als Basis, erhält man

$$\hat{Z}(t) \quad \sim \quad -\ln(\hat{R}(t)) = \sum_{j=1}^{n} \ln(\hat{p}_j).$$

Da gilt

$$\ln \hat{p}_j = \ln\left(\frac{n_j - 1}{n_j}\right) = \ln\left(1 - \frac{1}{n_j}\right) = \frac{1}{n_j} + \frac{1}{2n_j^2} + \frac{1}{3n_j^3} + \cdots$$

nimmt man $\boxed{\hat{Z}(t) = \sum_{j \in J_t} \frac{1}{n_j}.}$

# Nelson Schätzer - Vollständige Daten

**Gegeben**

$\{t_i : i = 1, \ldots, n\}$ identisch verteilte unabhängige Lebensdauern. Sei $\{t^{(i)} : i = 1, \ldots, n\}$ die dazugehörige geordnete Menge.

# Nelson Schätzer - Vollständige Daten

**Gegeben**

$\{t_i : i = 1, \ldots, n\}$ identisch verteilte unabhängige Lebensdauern. Sei $\{t^{(i)} : i = 1, \ldots, n\}$ die dazugehörige geordnete Menge.

**Nelson Schätzer:**

$$\hat{Z}(t) = \begin{cases} 0 & \text{für } t < t_{(1)} \\ \sum_{j=1}^{r} \frac{1}{n-j+1} & \text{für } t_{(r)} \leq t < t_{(r+1)}. \end{cases}$$

# Nelson Schätzer - Zensurierte Daten

**Gegeben:**

Die Stichprobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$;
Annahmen: Die Elemente der Stichprobe sind unabhängig und identisch verteilt.
Zugehörgige geordnete Stichprobe:

$$\{(t^{(i)}, \delta_i) : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)}, \; i = 1, \ldots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.
Sei

- $J_t = \{j = 1, \ldots n : t_j \le t\}$,
- $n_j = \#\{j \in J_t : t_j \text{ ist ein Ausfall}\}$,

# Nelson Schätzer - Zensurierte Daten

**Gegeben:**

Die Stichprobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$;
Annahmen: Die Elemente der Stichprobe sind unabhängig und identisch verteilt.
Zugehörgige geordnete Stichprobe:

$$\{(t^{(i)}, \delta_i) : i = 1, \ldots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \ldots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.
Sei

- $J_t = \{j = 1, \ldots n : t_j \le t\}$,
- $n_j = \#\{j \in J_t : t_j \text{ ist ein Ausfall}\}$,

**Nelson Schätzer:**

$$\hat{Z}(t) = \sum_{j \in J_t} \frac{1}{n_j} = \sum_{\nu} \frac{1}{n - \nu + 1},$$

wobei $\nu$ durch alle Zahlen läuft mit $t_{(j)}$ ist eine Ausfallrate und $t_{(j)} < t$.

# Nelson Schätzer - Zensurierte Daten

## Gegeben:

Die Stichprobe $\{t_i : t_i \text{ ist eine Ausfallzeit}\}$;

Annahmen: Die Elemente der Stichprobe sind unabhängig und identisch verteilt.

Zugehörgige geordnete Stichprobe:

$$\{(t^{(i)}, \delta_i) : i = 1, \dots, n\}, \text{ mit } t^{(i)} < t^{(i+1)},\ i = 1, \dots, n-1.$$

Falls $t^{(i)}$ zensiert ist, ist $\delta_i = 1$, ansonsten $\delta_i = 0$.

Sei

- $J_t = \{j = 1, \dots n : t_j \le t\}$,
- $n_j = \#\{j \in J_t : t_j \text{ ist ein Ausfall}\}$,

## Nelson Schätzer:

$$\hat{Z}(t) = \sum_{j \in J_t} \frac{1}{n_j} = \sum_{\nu} \frac{1}{n - \nu + 1},$$

wobei $\nu$ durch alle Zahlen läuft mit $t_{(j)}$ ist eine Ausfallrate und $t_{(j)} < t$.

# Parametrische Statistik

Schätzen der Parameter einer Verteilungsfunktion

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind exponentiall verteilt oder Weibull verteilt mit unbekannten Parametern $\lambda$ oder $\alpha$ und $\lambda$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind exponentiall verteilt oder Weibull verteilt mit unbekannten Parametern $\lambda$ oder $\alpha$ und $\lambda$.

**Aufgabe:**

Schätzen der Parameter $\lambda$ oder $\alpha$ und $\lambda$. Man erhält Schätzer $\hat{\alpha}$ und $\hat{\lambda}$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind exponentiall verteilt oder Weibull verteilt mit unbekannten Parametern $\lambda$ oder $\alpha$ und $\lambda$.

**Aufgabe:**

Schätzen der Parameter $\lambda$ oder $\alpha$ und $\lambda$. Man erhält Schätzer $\hat{\alpha}$ und $\hat{\lambda}$.

**Berechnung der Zuverlässigkeitsfunktion und der Ausfallrate:**

Die Zuverlässigkeitsfunktion und die Ausfallrate wird mittels der exponential Verteilung mit Parameter $\hat{\lambda}$ oder der Weibull Verteilung Wei$(\hat{\alpha}, \hat{\beta})$ berechnet.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind exponentiall verteilt oder Weibull verteilt mit unbekannten Parametern $\lambda$ oder $\alpha$ und $\lambda$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind exponentiall verteilt oder Weibull verteilt mit unbekannten Parametern $\lambda$ oder $\alpha$ und $\lambda$.

**Maximum Liklihood Schätzer:**

Bei der Schätzung des unbekannten Parameters $\lambda$ oder $\alpha$ wählt man diesen so, dass die *beobachtete Stichprobe* für die Modellverteilung mit *diesem* Parameter *am wahrscheinlichsten ist*.

**Die kleinste Quadrate Methode:**

Die Parameter der Verteilung werden so an die Daten angepasst, dass das Kriterium der minimalen Fehlerquadratsumme erfüllt ist.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$
- zensierte Beobachtungen$\{t_1^+, \ldots, t_{n-m}^+\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- nicht zensiert;
- $t_i$ sind exponential verteilt mit unbekannten Parameter $\lambda$.

# Schätzen der Parameter einer Verteilungsfunktion

## Gegeben:

- Eine Stichprobe $\{t_i : i = 1, \dots, n\}$
- zensierte Beobachtungen$\{t_1^+, \dots, t_{n-m}^+\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- nicht zensiert;
- $t_i$ sind exponential verteilt mit unbekannten Parameter $\lambda$.

## Likelihood-Funktion

$$L(\lambda) = \prod_{i=1}^{m} \lambda \exp(-\lambda t_i) \prod_{i=1}^{n-m} \exp(-\lambda t_i^+).$$

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$
- zensierte Beobachtungen$\{t_1^+, \ldots, t_{n-m}^+\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- nicht zensiert;
- $t_i$ sind exponential verteilt mit unbekannten Parameter $\lambda$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$
- zensierte Beobachtungen$\{t_1^+, \ldots, t_{n-m}^+\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- nicht zensiert;
- $t_i$ sind exponential verteilt mit unbekannten Parameter $\lambda$.

**Schätzer des Erwartungswert (bzw. Parameters $\lambda$:**

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{1}{m}\left[\sum_{i=1}^{n} t_i + \sum_{i=1}^{n-m} t_i^+\right].$$

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- nicht zensiert;
- $t_i$ sind Weibul verteilt mit unbekannten Parameter $\lambda$ und $\alpha$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- nicht zensiert;
- $t_i$ sind Weibul verteilt mit unbekannten Parameter $\lambda$ und $\alpha$.

**Methode:**

Stammen die Beobachtungen tatsächlich aus einer Weibullverteilten (Parametern $\alpha$ und $\lambda$) Grundgesamtheit, müssen die Punkte der zugehörgigen Wertepaare auf der Geraden

$$y = \log(\lambda)\alpha + \alpha x = b + ax$$

liegen. Schätzen der Koeffizienten $\hat{a}$ und $\hat{b}$ führt zu Schätzern von $\alpha$ und $\lambda$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind Weibul verteilt mit unbekannten Parameter $\lambda$ und $\alpha$.

# Schätzen der Parameter einer Verteilungsfunktion

**Gegeben:**

- Eine Stichprobe $\{t_i : i = 1, \ldots, n\}$

Annahmen:

- $t_i$ sind unabhängig und identisch verteilt.
- $t_i$ sind Weibul verteilt mit unbekannten Parameter $\lambda$ und $\alpha$.

**Methode:**

Schätzen von $a$ und $b$ mittels einer Regressionsgeraden. Nach kurzer Rechnung erhält man:

$$\hat{\alpha} = a,$$

und

$$\hat{\lambda} = \exp\left(-\frac{\hat{b}}{\hat{a}}\right).$$

# Verteilungstest

**Gegeben:**

- Ein Stichprobe $\{x_1, x_2, \ldots, x_n\}$
- Eine Verteilungsfunktion $F$.

**Frage:**

Hat die Grundgesamtheit die Verteilungsfunktion $F$?

# Verteilungstest

**Gegeben:**

- Ein Stichprobe $\{x_1, x_2, \ldots, x_n\}$
- Eine Verteilungsfunktion $F$.

**Frage:**

Hat die Grundgesamtheit die Verteilungsfunktion $F$?

Nullhypothese und Alternativhypothese sind dabei:

- $H_0$: $X$ hat Verteilungsfunktion $F_X(x)$
- $H_1$: $X$ hat nicht Verteilungsfunktion $F_X(x)$

# Chi-Quadrat Test

## Berechnen der Testvariable:

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 $\Sigma$ ist die Testvariable, von der man weiß dass sie für grosse $n$ $\chi^2$ verteilt ist mit Freiheitsgraden $K-1$.
5. bestimme zu einem vorgegebenen Signifikanzniveau $\alpha$ einen kritischen Wert $c > 0$, der von der Testvariablen $\Sigma$ nicht überschritten werden darf.
6. Ist $\Sigma$ größer als $c$, verwerfe $H_0$ ansonsten nehme $H_0$ an.

# Kolmogorov Simirnov Test

Die Testvarible ist der betragsmässige größte Abstand der Werte empirischen Verteilungsfunktion $\hat{F}$ und der theoretischen Verteilungsfunktion:

$$D_n = \max_{x \in \mathbb{R}} |F_X(x) - \hat{F}_X(x)|.$$

# Kolmogorov Simirnov Test

Die Testvarible ist der betragsmässige größte Abstand der Werte empirischen Verteilungsfunktion $\hat{F}$ und der theoretischen Verteilungsfunktion:

$$D_n = \max_{x \in \mathbb{R}} |F_X(x) - \hat{F}_X(x)|.$$

Kolmogorov hat die asymptotische Verteilung von $\sqrt(n) D_n$ ermittelt und gezeigt dass für $\lambda > 0$ gilt

$$\lim_{n \to \infty} P(\sqrt{n} D_n \leq \lambda) = \sum_{k=-\infty}^{k=\infty} (-1)^k e^{-2k^2 \lambda^2}.$$

# Kolmogorov Simirnov Test

Die Testvarible ist der betragsmässige größte Abstand der Werte empirischen Verteilungsfunktion $\hat{F}$ und der theoretischen Verteilungsfunktion:

$$D_n = \max_{x \in \mathbb{R}} |F_X(x) - \hat{F}_X(x)|.$$

Kolmogorov hat die asymptotische Verteilung von $\sqrt(n)D_n$ ermittelt und gezeigt dass für $\lambda > 0$ gilt

$$\lim_{n \to \infty} P(\sqrt{n}D_n \leq \lambda) = \sum_{k=-\infty}^{k=\infty} (-1)^k e^{-2k^2\lambda^2}.$$

Den kritischen Vergleichswert $c$ wählt man zu einem Niveau $\alpha$ so, dass

$$P(D_n \leq c) = 1 - \alpha.$$

Der Reihenwert ist für $n > 50$ eine sehr gute Approximation. In Matlab können sie den Kolmogorov-Smirnov test mittels *kstest* aufrufen.

# Datenanalyse - Nicht Parametrische Tests

**Gegeben:**

Ein Stichprobe $\{x_1, x_2, \ldots, x_n\}$

**Frage:**

Wie schaut die Verteilungsfunktion aus?

- Empirische Verteilungsfunktion (*ecdf*)
- Empirische Hazard Funktion (*ecdf mit cumulative hazard*)
- Konfidenz Intervalle;

# Datenanalyse - Parametrische Tests

**Gegeben:**

Ein Stichprobe $\{x_1, x_2, \ldots, x_n\}$

**Frage:**

Welcher Verteilungstyp passt ?

# Entscheidungskriterien:

1. Plotten des Histogramms: Anzahl der Klassen: Sturges Regel:
   $k = [1 + 3.3 \log_{10}(n)]$, $k$ Anzahl der Klassen.

# Entscheidungskriterien:

1. Plotten des Histogramms: Anzahl der Klassen: Sturges Regel:
   $k = [1 + 3.3 \log_{10}(n)]$, $k$ Anzahl der Klassen.
2. Berechnen vom Median und Mittelwert (mean):
   - Symmetrische Verteilung: mean $\sim$ median (Normal, Weibull mit shape parameter $\in (3, 4)$);
   - Rechtsschief: median $\ll$ mean (Exponential, Weibull oder Lognormal Verteilung);
     (Falls mean = Standardabweichung: Exponentailfunktion);
   - Linksschief: median $\gg$ mean.

# Entscheidungskriterien:

1. Plotten des Histogramms: Anzahl der Klassen: Sturges Regel:
   $k = [1 + 3.3 \log_{10}(n)]$, $k$ Anzahl der Klassen.
2. Berechnen vom Median und Mittelwert (mean):
   - Symmetrische Verteilung: mean $\sim$ median (Normal, Weibull mit shape parameter $\in (3, 4)$);
   - Rechtsschief: median $\ll$ mean (Exponential, Weibull oder Lognormal Verteilung);
     (Falls mean $=$ Standardabweichung: Exponentailfunktion);
   - Linksschief: median $\gg$ mean.
3. Zeichnen der Ausfallrate (Hazard rate)
   - Monoton sinkend: Weibull mit shape parameter $\alpha < 1$
   - Konstant: Exponential Verteilung
   - Monoton Steigend: Weibull (grosser shape parameter), Normal, Lognormal Verteilung.

# Entscheidungskriterien:

1. Plotten des Histogramms: Anzahl der Klassen: Sturges Regel: $k = [1 + 3.3\log_{10}(n)]$, $k$ Anzahl der Klassen.
2. Berechnen vom Median und Mittelwert (mean):
   - Symmetrische Verteilung: mean $\sim$ median (Normal, Weibull mit shape parameter $\in (3,4)$);
   - Rechtsschief: median $\ll$ mean (Exponential, Weibull oder Lognormal Verteilung);
     (Falls mean $=$ Standardabweichung: Exponentailfunktion);
   - Linksschief: median $\gg$ mean.
3. Zeichnen der Ausfallrate (Hazard rate)
   - Monoton sinkend: Weibull mit shape parameter $\alpha < 1$
   - Konstant: Exponential Verteilung
   - Monoton Steigend: Weibull (grosser shape parameter), Normal, Lognormal Verteilung.
4. Exponential, Weibull plots.

# Entscheidungskriterien:

1. Exponental Plot: Zeichne

$$\{(t_i; \ln(1/(1-\hat{F}(t_i)))), \quad i = 1, \ldots, n\}$$

# Entscheidungskriterien:

1. Exponental Plot: Zeichne

$$\{(t_i; \ln(1/(1-\hat{F}(t_i)))), \quad i = 1, \ldots, n\}$$

   Schätzen des Parameters durch Schätzung der Steigung:

$$\hat{\lambda} = \hat{b} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

   wobei $y_i = \ln(1/(1-F(t_i)))$ und $x_i = t_i$ gilt.

# Entscheidungskriterien:

1. Exponental Plot: Zeichne

$$\{(t_i; \ln(1/(1-\hat{F}(t_i)))), \quad i=1,\ldots,n\}$$

Schätzen des Parameters durch Schätzung der Steigung:

$$\hat{\lambda} = \hat{b} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

wobei $y_i = \ln(1/(1-F(t_i)))$ und $x_i = t_i$ gilt.

2. Weibull Plot: Zeichne

$$\{(\ln(t_i); \ln\left(\ln\left(1/(1-\hat{F}(t_i)\right)\right)), \quad i=1,\ldots,n\}$$

# Entscheidungskriterien:

1. Exponental Plot: Zeichne

$$\{(t_i; \ln(1/(1-\hat{F}(t_i)))), \quad i = 1, \dots, n\}$$

   Schätzen des Parameters durch Schätzung der Steigung:

$$\hat{\lambda} = \hat{b} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

   wobei $y_i = \ln(1/(1-F(t_i)))$ und $x_i = t_i$ gilt.

2. Weibull Plot: Zeichne

$$\{(\ln(t_i); \ln\left(\ln\left(1/(1-\hat{F}(t_i)\right)\right)), \quad i = 1, \dots, n\}$$

   Schätzer der Parameter sin Schätzer der Steigung und Nullstelle der Geraden.

# Schätzen der Parameter:

1. *wblfit(data),*
2. *lognfit(data)*
3. $phat = mle(data,'pdf',pdf,'start',start,'cdf',cdf)$

# Schätzen der Parameter:

1. *wblfit(data)*,
2. *lognfit(data)*
3. $phat = mle(data,'pdf',pdf,'start',start,'cdf',cdf)$

# Verteilungstest:

1. Chi$^2$ test *chi2gof(x)*
2. Kolmogorov Smirnov Test: *kstest*
3. $QQ$–plot:
4. Barlett's Test für die Exponentialverteilung; Mann's Test für die Weibulverteilung.

**Sonstige Möglichkeiten:**

empirsche Verteilungsfunktion; $\gamma$, $\beta$, Verteilung, Extremwertverteilung, Birnbaum Saudners, Burr Verteilung, Pearson, bimodale Verteilung.