



# Test und Verlässlichkeit

## Foliensatz 3: Verteilungen

Prof. G. Kemnitz

Institut für Informatik, TU Clausthal (TV\_F3)  
November 6, 2022

## Inhalt Foliensatz TV\_F3: Verteilungen

### Grundlagen

- 1.1 Charakteristische Größen
- 1.2 Lineare Transformation
- 1.3 Verteilung von Zählwerten

### Näherungen für ZV

- 2.1 Binomialverteilung
- 2.2 Poisson-Verteilung
- 2.3 Bereichsschätzung, Poisson
- 2.4 Normalverteilung
- 2.5 Bereichsschätzung NVT

- 2.6 Varianzerhöhung
- 2.7 Bereichsschätz. Zählwerte  
Misch- und multimodale  
Verteilung  
Weitere Verteilungen
- 4.1 Pareto-Verteilung
- 4.2 Gammaverteilung
- 4.3 Exponentialverteilung
- 4.4 Verteilung der FF-Rate  
Test & Zuverlässigkeit  
Schaden durch FF

Vorlesung	1	2	3	4
bis Abschn.	2.?? (??)	4.?? (??)	4.?? (??)	5.?? (??)



## Zufall, Zufallsexperiment, Zufallsvariable

- Zufälliges Ereignis: Ereignis, das weder sicher noch unmöglich ist, sondern mit einer gewissen Wahrscheinlichkeit eintritt.
- Zufallsexperiment: Experiment mit mehreren möglichen Ergebnissen und zufälligem Ausgang.
- Zufallsvariable: Veränderliche, die ihre Werte in Abhängigkeit vom Zufall nach einer Wahrscheinlichkeitsverteilung annimmt.

## Bernoulli-Versuch

Das einfachste Zufallsexperiment. Zweipunktverteilung:

$$\mathbb{P}\{X = 0\} = 1 - p$$

$$\mathbb{P}\{X = 1\} = p$$

( $p$  – Eintrittswahrscheinlichkeit).

Die beiden mögliche Ergebnisse  $\{0, 1\}$  können auch  $\{\text{nein, ja}\}$ ,  $\{\text{falsch, wahr}\}$ , ... bedeuten.

Zufallsexperimente mit mehr als zwei möglichen Ergebnissen lassen sich in je einen Bernoulli-Versuch je Ergebnis aufspalten:

$$A_i = \begin{cases} 0 & \text{Ereignis nicht eingetreten} \\ 1 & \text{Ereignis eingetreten} \end{cases}$$



## Verteilungen

Eine Verteilung ordnet den möglichen Ergebnissen eines Zufallsexperiments Wahrscheinlichkeiten zu. Im weiteren interessierende Zufallsvariablen:

- Anzahl der FF (aufgetretene, erkannte, tolerierte, ...),
- Anzahl der Fehler (entstandene, beseitigte, vermiedene, ...),
- Lebensdauer,
- Schadenskosten, ...

---

Dieser Foliensatz vermittelt ein auf die Vorlesung abgestimmten Werkzeugkasten aus der Stochastik zur Abschätzung der

- Verteilungen,
- Erwartungswerte, Streuungen und
- wahrscheinlichen Bereiche

für Zählwerte, Überdeckungen, Schadenshöhen, Lebensdauern, ...



# Grundlagen



## Charakteristische Größen

## Charakteristische Größen einer Zufallsvariablen

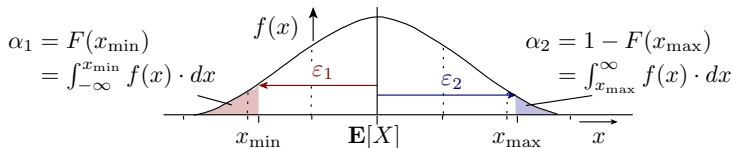
Wenn eine Zufallsvariable  $X$  mehr als 2 Werte annehmen kann, gibt es außer den Eintrittswahrscheinlichkeiten der einzelnen Werte weitere interessante Größen:

Name	Definition
Verteilungsfunktion	$F_X(x) = \mathbb{P}[X \leq x]$
Dichtefunktion	$f_X(x) = \frac{dF_X(x)}{dx}$
Erwartungswert	$\mathbb{E}[X] = \sum_{i=1}^m p_i \cdot x_i$ bzw. $\mathbb{E}[X] = \int_{-\infty}^{\infty} f_X(x) \cdot x \cdot dx$
k-tes Moment	$m_k = \mathbb{E}[X^k]$
k-tes zentriertes Moment	$\mathbb{E}[(X - \mathbb{E}[X])^k]$
Varianz (2. zentr. M.)	$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
Standardabweichung	$\text{sd}[X] = \sqrt{\text{Var}[X]}$
wahrscheinlicher Bereich	$[x_{\min}, x_{\max}]$



## Wahrscheinlicher Bereich, $\mathbb{E}[X]$ und $\text{sd}[X]$

Bereich  $[x_{\min}, x_{\max}]$ , die der Wert der Zufallsvariable  $X$  mit einer Wahrscheinlichkeit  $1 - (\alpha_1 + \alpha_2)$  annimmt:



$$\alpha_1 = F(x_{\min})$$

Irrtumswahrscheinlichkeit, dass Werte unterhalb des geschätzten Bereichs liegen.

$$\alpha_2 = 1 - F(x_{\max})$$

Irrtumswahrscheinlichkeit, dass Werte oberhalb des geschätzten Bereichs liegen.

$$\varepsilon_{1/2}$$

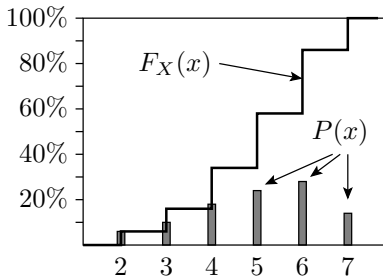
Intervallradius, Abstand der unteren / oberen Bereichsgrenze vom Erwartungswert.

Bei  $\alpha_1 = 0$  gibt es nur eine Ober- und bei  $\alpha_2 = 0$  nur eine Untergrenze.

## Diskrete Verteilung

Zufallsvariable  $X$  kann nur eine (über<sup>1</sup>) abzählbare Menge von diskreten Werten  $x_i$  annehmen, z.B.:

$x_i$	2	3	4	5	6	7
$P(x) = \mathbb{P}[X = x_i] = p_i$	6%	10%	18%	24%	28%	14%
$F_X(x) = \mathbb{P}[X \leq x_i]$	6%	16%	34%	58%	86%	100%



<sup>1</sup>falls die Menge nicht endlich ist.



## Erwartungswert und Varianz

$x_i$	2	3	4	5	6	7
$P(x) = \mathbb{P}[X = x_i] = p_i$	6%	10%	18%	24%	28%	14%
$F_X(x) = \mathbb{P}[X \leq x_i]$	6%	16%	34%	58%	86%	100%

**Erwartungswert** (mit den Auftrittswahrscheinlichkeiten gewichteter Mittelwert):

$$\mathbb{E}[X] = \sum_{i=1}^m p_i \cdot x_i \quad (1)$$

( $m$  – Anzahl der möglichen Werte). Für das Beispiel:

$$6\% \cdot 2 + 10\% \cdot 3 + 18\% \cdot 4 + 24\% \cdot 5 + 28\% \cdot 6 + 14\% \cdot 7 = 5$$

**Varianz** (2. zentriertes Moment):

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{i=1}^m p_i \cdot (x_i - \mathbb{E}[X])^2$$

Für das Beispiel:

$$6\% \cdot (2 - 5)^2 + 10\% \cdot (3 - 5)^2 + \dots + 14\% \cdot (7 - 5)^2 = 1,96$$

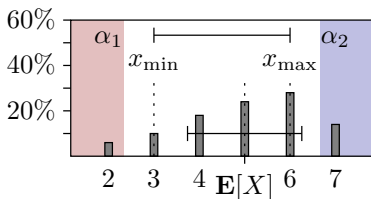
## Standardabweichung

Standardabweichung (Quadratwurzel aus der Varianz), Maß für die Abweichung vom Erwartungswert bzw. die Breite des wahrscheinlichen Bereichs von  $X$ :

$$\text{sd}[X] = \sqrt{\text{Var}[X]}$$

Für das Beispiel:

$$\text{sd}[X] = \sqrt{1,96} = 1,4$$



Bereich, in dem  $X$  mit 80% Wahrscheinlichkeit liegt

$$\mathbf{E}[X] \pm \text{sd}(X)$$

Irrtumswahrscheinlichkeiten für  $X$  außerhalb  $[x_{\min}, x_{\max}]$ :

$$\alpha_1 = \mathbb{P}[X < x_{\min}] = \sum_{x_i < x_{\min}} \mathbb{P}[X = x_i]$$

$$\alpha_2 = \mathbb{P}[X > x_{\max}] = \sum_{x_i > x_{\max}} \mathbb{P}[X = x_i]$$



## Verschiebungssatz

Die Varianz ist gleichfalls die Differenz aus dem Erwartungswert der Quadrate und dem Quadrat des Erwartungswertes<sup>2</sup>:

$$\text{Var} [X] = \mathbb{E} [X^2] - \mathbb{E} [X]^2 \quad (2)$$

Herleitung:

$$\begin{aligned} \sum_{i=1}^m p_i \cdot (x_i - \mathbb{E} [X])^2 &= \sum_{i=1}^m p_i \cdot (x_i^2 - 2 \cdot x_i \cdot \mathbb{E} [X] + \mathbb{E} [X]^2) \\ &= \underbrace{\sum_{i=1}^m p_i \cdot x_i^2}_{\mathbb{E}[X^2]} - 2 \cdot \mathbb{E} [X] \cdot \underbrace{\sum_{i=1}^m p_i \cdot x_i}_{\mathbb{E}[X]} + \mathbb{E} [X]^2 \cdot \underbrace{\sum_{i=1}^m p_i}_1 \end{aligned}$$

Für das Beispiel zuvor:

$$\text{Var} [X] = 6\% \cdot 2^2 + 10\% \cdot 3^2 + 18\% \cdot 4^2 + 24\% \cdot 5^2 + 28\% \cdot 6^2 + 14\% \cdot 7^2 - 5^2 = 1,96$$

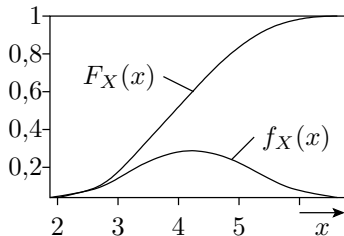
<sup>2</sup>Bei begrenzter Rechengenauigkeit u.U. numerisch problematisch wegen »kleiner Differenz großer Zahlen«.

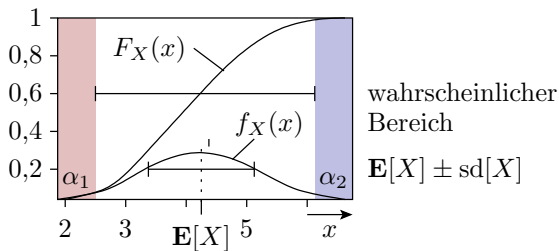
## Stetige Verteilungen

Zufallsvariable  $X$  ist stetig und hat im Intervall  $a \leq X \leq b$  unendlich viele Ausprägungen. Beschreibung durch die Dichte:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x) = \int_{-\infty}^x f(u) \cdot du \quad \text{für } X \in \mathfrak{R}$$



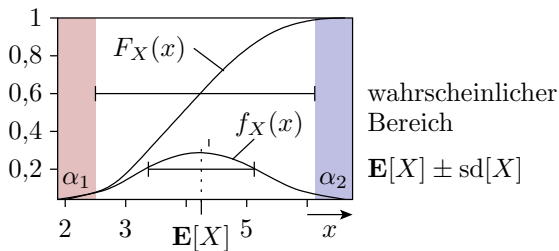


Erwartungswert:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} f_X(x) \cdot x \cdot dx$$

Varianz:

$$\begin{aligned} \text{Var}[X] &= \int_{-\infty}^{\infty} f_X(x) \cdot (x - \mathbb{E}[X])^2 \cdot dx \\ &= \int_{-\infty}^{\infty} f_X(x) \cdot x^2 \cdot dx - \mathbb{E}[X]^2 \end{aligned}$$



Wahrscheinlichkeit, dass  $X$  in einem Bereich  $[x_{\min}, x_{\max}]$  liegt:

$$\begin{aligned}
 \mathbb{P}[x_{\min} \leq x \leq x_{\max}] &= F_X(x_{\max}) - F_X(x_{\min}) \\
 &= \int_{x_{\min}}^{x_{\max}} f_X(x) \cdot dx
 \end{aligned}$$

Bereichsgrenzen:

$$x_{\min} = F^{-1}(\alpha_1)$$

$$x_{\max} = F^{-1}(1 - \alpha_2)$$



## Erwartungswert und Varianz einer Datenstichprobe

Für eine Datenstichprobe einer Zufallsvariable  $X$

$$w = (w_1, w_2, \dots, w_{\#w})$$

ist der im weiteren verwendete Schätzer für den Erwartungswert der Mittelwert:

$$\hat{\mathbb{E}}[X] = \bar{w} = \frac{1}{\#w} \cdot \sum_{i=1}^{\#w} w_i \quad (3)$$

Der Schätzer für die Varianz ist die mittlere quadratische Abweichung vom geschätzten Mittelwert:

$$\hat{\text{Var}}[X] = \frac{1}{\#w - 1} \cdot \sum_{i=1}^{\#w} (w_i - \hat{\mathbb{E}}[X])^2 \quad (4)$$

Der Quotient ist um eins kleiner als die Stichprobengröße  $\#w$ , d.h. die Abschätzung der Varianz erfordert mindestens Stichprobengröße  $\#w = 2$ .



## Lineare Transformation



## Lineare Transformation

Lineare Transformationen sind die Multiplikation und Addition einer Zufallsvariable mit reellen Zahlen. Der Erwartungswert vergrößert und verschiebt sich um dieselben Werte:

$$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b$$

Bei der Varianz entfällt die Verschiebung und der Skalierungsfaktor geht im Quadrat ein<sup>3</sup>:

$$\text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X] \quad (5)$$

Die Varianz ist insbesondere verschiebungsinvariant und bleibt bei einer Spiegelung der Verteilung gleich:

$$\text{Var}[-X] = (-1)^2 \cdot \text{Var}[X] = \text{Var}[X]$$

---

<sup>3</sup>Beweis durch »Nachrechnen« siehe HA bzw. Grosse Übung.

## Kontrolle am Beispiel



Realisierungen $x$ von $X$	1	2	3
Realisierungen $y$ von $Y = 5 - 2X$	3	1	-1
$\mathbf{P}[Y = y] = \mathbf{P}[X = x]$	0,3	0,5	0,2

$$\mathbb{E}[X] = 0,3 + 1 + 0,6 = 1,9$$

$$\text{Var}[X] = 0,3 + 2 + 1,8 - 1,9^2 = 0,49$$

$$\mathbb{E}[Y] = 0,9 + 0,5 - 0,2 = 1,2$$

$$\text{Var}[Y] = 2,7 + 0,5 + 0,2 - 1,2^2 = 1,96$$

$$\mathbb{E}[Y] = 5 - 2 \cdot \mathbb{E}[X]$$

$$\text{Var}[Y] = (-2)^2 \cdot \text{Var}[X]$$



## Summe von Zufallsvariablen

Die Verteilung der Summe von Zufallsvariablen ordnet jedem der möglichen Werte der Summe die Wahrscheinlichkeit zu, dass die Summe diesen Wert hat (Faltung):

$x$	1	3	4
$f_X(x)$	0,1	0,4	0,5
$y$	2	3	4
$f_Y(y)$	0,3	0,6	0,1

$$f_{X+Y} = f_X * f_Y :$$

$$\mathbb{P}[X + Y = 3] = \mathbb{P}[X = 1] \cdot \mathbb{P}[Y = 2]$$

$$\mathbb{P}[X + Y = 4] = \mathbb{P}[X = 1] \cdot \mathbb{P}[Y = 3]$$

$$\mathbb{P}[X + Y = 5] = \mathbb{P}[X = 1] \cdot \mathbb{P}[Y = 4] + \mathbb{P}[X = 3] \cdot \mathbb{P}[Y = 2]$$

$$\mathbb{P}[X + Y = 6] = \mathbb{P}[X = 3] \cdot \mathbb{P}[Y = 3] + \mathbb{P}[X = 4] \cdot \mathbb{P}[Y = 2]$$

$$\mathbb{P}[X + Y = 7] = \mathbb{P}[X = 3] \cdot \mathbb{P}[Y = 4] + \mathbb{P}[X = 4] \cdot \mathbb{P}[Y = 3]$$

$$\mathbb{P}[X + Y = 8] = \mathbb{P}[X = 4] \cdot \mathbb{P}[Y = 4]$$



Für die Summe von Zufallsvariablen ist der Erwartungswert gleich der Summe der Erwartungswerte:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Die Varianz ist die Summe der Varianzen plus doppelte Kovarianz:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{Cov}[X, Y] \quad (6)$$

mit der Kovarianz<sup>4</sup>:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])] \quad (7)$$

Für unabhängige Zufallsvariablen ist die Kovarianz null und die Varianz die Summe der Varianzen der Summanden:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

---

<sup>4</sup>Die Kontrollen durch »Nachrechnen« sind Übungsaufgaben.

## Gemessener Wert und Messfehler

In der Messtechnik gilt für jeden gemessenen Wert:

$$X_M = X + X_F$$

( $X$  – Messwert;  $X_F$  – Messfehler). Alle drei Größen haben einen Erwartungswert und eine Varianz. Mit dem Messwert und dem Messfehler als unabhängige Zufallsvariablen, gilt für diese:

$$\mathbb{E}[X_M] = \mathbb{E}[X] + \mathbb{E}[X_F]$$

$$\text{Var}[X_M] = \text{Var}[X] + \text{Var}[X_F]$$

- $\mathbb{E}[X_F]$  – Erwartungswert, systematischer Messfehler
- $\text{sd}[X_F] = \sqrt{\text{Var}[X_F]}$  – Standardabweichung, zufälliger Messfehler.

## Beispielaufgabe



Der gemessene Wert einer Widerstands-Charge ist im Mittel  $\mathbb{E}[R_M] = 1010 \Omega$  und hat eine Standardabweichung von  $\text{sd}[R_M] = 11,18 \Omega$ . Die Messung habe einen systematischen Fehler von  $\mathbb{E}[R_F] = 12 \Omega$  und eine Standardabweichung von  $\text{sd}[R_F] = 5 \Omega$ . Welchen Erwartungswert und welche Standardabweichung hat der (tatsächliche) Messwert?

---

$$\begin{aligned}\mathbb{E}[R] &= \mathbb{E}[R_M] - \mathbb{E}[R_F] = 1010 \Omega - 12 \Omega = 998 \Omega \\ \text{Var}[R] &= \text{Var}[R_M] - \text{Var}[R_F] = (11,18 \Omega)^2 - (5 \Omega)^2 = 100 \Omega^2 \\ \text{sd}[R] &= 10 \Omega\end{aligned}$$

Der (tatsächliche) Messwert hat eine kleinere Standardabweichung als der gemessene Wert.





## Verteilung von Zählwerten



## Verteilung von Zählwerten

Ein zufälliger Zählwert  $X$ , z.B. die Anzahl der korrekt ausgeführten oder fehlerhaft ausgeführten Service-Leistungen lässt sich als Summe

$$X = \sum_{i=1}^{\#X} X_i$$

»potentieller Zählwerte«  $X_i$  mit der Bernoulli-Verteilung:

$$\mathbb{P}[X_i = k] = \begin{cases} 1 - p_i & k = 0 \\ p_i & k = 1 \end{cases}$$

beschreiben.

Zählwert $X$	potentielle Zählwerte $X_i \in \{0, 1\}$
Fehlfunktionen	Service-Anforderungen
Fehler	potentielle Fehler
nachweisbare Fehler	vorhandene Fehler
...	



## Erwartungswert und Varianz

Der Erwartungswert der Einzelereignisse ist

$k$	0	1
$\mathbf{P}[X_i = k]$	$1 - p_i$	$p_i$

$$\mathbb{E}[X_i] = (1 - p_i) \cdot 0 + p_i \cdot 1 = p_i$$

Varianz nach Verschiebungssatz:

$$\text{Var}[X_i] = (1 - p_i) \cdot 0^2 + p_i \cdot 1^2 - p_i^2 = p_i \cdot (1 - p_i)$$

Der Erwartungswert der Summe ist die Summe der Erwartungswerte:

$$\mathbb{E}[X] = \sum_{i=1}^{\#X} p_i \quad (8)$$

Für die Varianz wird oft unterstellt, dass die zu zählenden Ereignisse, wie das Auftreten unterschiedlicher Fehlfunktion, nicht voneinander abhängen (Varianz der Summe gleich der Summe der Varianzen der Summanden, Kovarianz null):

$$\text{Var}[X] = \sum_{i=1}^{\#X} p_i \cdot (1 - p_i) \quad (9)$$



## Berechnung der Verteilung

Für die Verteilung gilt, dass bei Hinzunahme eines weiteren Experiments  $i$  sich mit Wahrscheinlichkeit  $p_i$  der Zählwert um eins erhöht und mit der Wahrscheinlichkeit  $1 - p_i$  gleich bleibt:

$$\mathbb{P}_i [X = k] = p_i \cdot \mathbb{P}_{i-1} [X = k - 1] + (1 - p_i) \cdot \mathbb{P}_{i-1} [X = k]$$

Berechnung der Verteilung:

$$\mathbb{P}_1 [X = 0] = 1 - p_1$$

$$\mathbb{P}_1 [X = 1] = p_1$$

	$p_i$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$\mathbf{P}[X_1 = k]$	30%	70%	30%			
$\mathbf{P}[X_1 + X_2 = k]$	50%	35%	50%	15%		
$\mathbf{P}[\dots + X_3 = k]$	40%	21%	44%	29%	6%	
$\mathbf{P}[\dots + X_4 = k]$	10%	18,9%	41,7%	30,5%	8,3%	0,6%

Wiederhole für  $i = 2$  bis  $N$

$$\mathbb{P}_i [X = 0] = \mathbb{P}_{i-1} [X = 0] \cdot (1 - p_i)$$

$$\mathbb{P}_i [X = i] = \mathbb{P}_{i-1} [X = i - 1] \cdot p_i$$

Wiederhole für  $k = 1$  bis  $i - 1$

$$\begin{aligned} \mathbb{P}_i [X = k] &= \mathbb{P}_{i-1} [X = k] \cdot (1 - p_i) \\ &+ \mathbb{P}_{i-1} [X = k - 1] \cdot p_i \end{aligned}$$

$\mathbb{P}_i [X_1 + \dots + X_i = k]$  – Wahrscheinlichkeit, dass die Summe der ersten  $i$  Zählwert  $k$  ist.

## Erwartungswert und Varianz für das Beispiel

$p_i$	30%	50%	40%	10%	
$k$	0	1	2	3	4
$P[X = k]$	18,9%	41,7%	30,5%	8,3%	0,6%

Erwartungswert der Summe aller  $N = 4$  Summanden:

$$\mathbb{E}[X] = 18,9\% \cdot 0 + 41,7\% \cdot 1 + 30,5\% \cdot 2 + 8,3\% \cdot 3 + 0,6\% \cdot 4 = 1,3$$

Als Summe aller  $p_i$  nach Gl. 8 ist die Berechnung kürzer:1

$$\mathbb{E}[X] = 30\% + 50\% + 40\% + 10\% = 1,3$$

Die Varianz beträgt nach dem Verschiebungssatz Gl. 2:

$$18,9\% \cdot 0^2 + 41,7\% \cdot 1^2 + 30,5\% \cdot 2^2 + 8,3\% \cdot 3^2 + 0,6\% \cdot 4^2 - 1,3^2 = 0,79$$

Die vereinfachte Berechnung nach Gl. 9:

$$\text{Var}[X] = 0,3 \cdot 0,7 + 0,5 \cdot 0,5 + 0,4 \cdot 0,6 + 0,1 \cdot 0,9 = 0,79$$

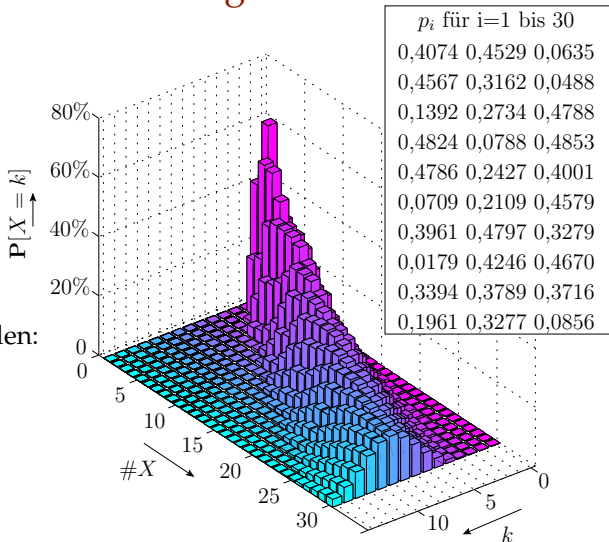
## Beispiel einer Zählverteilung

Mit Matlab schrittweise berechnete Zählverteilung. Die Eintrittswahrscheinlichkeiten der Zählereignisse siehe Kasten im Bild. Erwartungswert und Varianz für alle 30 Summanden:

$$\mathbb{E}[X] = 7,05,$$

$$\text{Var}[X] = 2,19$$

Wahrscheinlicher Bereich ca. 5 bis 15.





# Näherungen für ZV



# Binomialverteilung





# Binomialverteilung

$$\binom{4}{2} = \frac{4 \cdot 3}{2!}$$

Für den Sonderfall, dass gleichwahrscheinliche Ereignisse gezählt werden (alle  $p_i = p$ ), ist die Summe der gezählten Ereignisse binomialverteilt

$$X \sim \text{Bin}(n, p)$$

( $n$  – Anzahl der potentiellen Zählwerte;  $p$  – Wahrscheinlichkeit für Zählwert eins). Binomialverteilung:

$$\mathbb{P}[X = k] = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \quad (10)$$

Erwartungswert einer Binomialverteilung:1

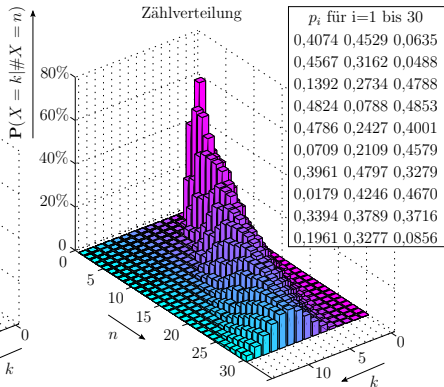
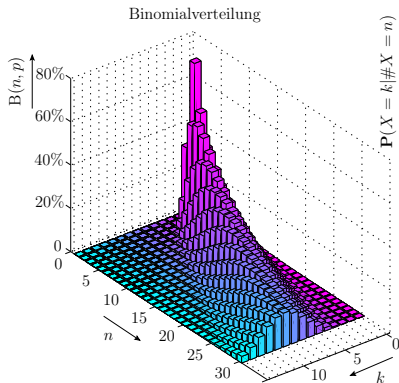
$$\mathbb{E}[X] = n \cdot p$$

Varianz und Standardabweichung einer Binomialverteilung:1

$$\text{Var}[X] = n \cdot p \cdot (1 - p) \quad (11)$$

$$\text{sd}[X] = \sqrt{n \cdot p \cdot (1 - p)} \quad (12)$$

## Binomialverteilung vs. allgemeine Zählverteilung



1

Eine Binomialverteilung mit  $p = \frac{1}{\#X} \cdot \sum_{i=1}^{\#X} p_i$  und  $n = \#X$  nähert eine Zählverteilung gut an und berechnet sich aus nur den zwei Parametern  $n$  und  $p$ .

## Bereichsschätzung für binomialverteilte #F



Die mittlere Nachweiswahrscheinlichkeit von 10 Fehlern sei 30%. Wie groß ist die Wahrscheinlichkeit, dass mindestens zwei von 10 Fehlern nachgewiesen werden?

---

$$\begin{aligned}\mathbb{P}[X \geq 2] &= 1 - \sum_{k=0}^1 \binom{10}{k} \cdot 0,3^k \cdot (1 - 0,3)^{10-k} \\ &= 1 - (0,7^{10} + 10 \cdot 0,3 \cdot 0,7^9) \\ &\approx 85\%\end{aligned}$$



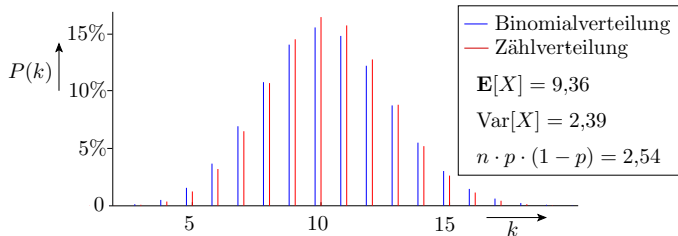
## Varianzobergrenze

### Satz

Bei gleicher Anzahl von unabhängigen Zählwerten  $n = \#X$  und  $p = \frac{1}{\#X} \cdot \sum_{i=1}^{\#X} p_i$  ist die Varianz der Binomialverteilung eine obere Schranke der Varianz einer Zählverteilung:

$$n \cdot p \cdot (1 - p) \geq \text{Var}[X] = \sum_{i=1}^n p_i \cdot (1 - p_i) \quad (13)$$

Für die beiden Verteilungen der Folie zuvor gilt für  $N = 30$ :





## Beweis

Ersatz der individuellen Auftrittswahrscheinlichkeiten der zu zählenden Ereignisse durch die mittlere Wahrscheinlichkeit und eine Differenz, die im Mittel null ist:

$$p_i = p + \delta_i \text{ mit } \sum_{i=1}^n \delta_i = 0$$

Varianz der Zählverteilung:

$$\begin{aligned} \text{Var}[X] &= \sum_{i=1}^n (p + \delta_i) \cdot (1 - p - \delta_i) \\ &= \sum_{i=1}^n (p - p^2 - p \cdot \delta_i + \delta_i - p \cdot \delta_i - \delta_i^2) \\ &= \underbrace{\sum_{i=1}^n (p - p^2)}_{n \cdot p \cdot (1-p)} + \underbrace{\sum_{i=1}^n (\delta_i - 2 \cdot p \cdot \delta_i)}_{(1-2p) \cdot \sum_{i=1}^n \delta_i = 0} - \sum_{i=1}^n \delta_i^2 \\ \text{Var}[X] &\leq n \cdot p \cdot (1-p) \leq \text{Varianz Binomialverteilung} \checkmark \end{aligned}$$



### Fakt 1

*Aus der mittleren Eintrittswahrscheinlichkeit*

$$p = \frac{1}{\#X} \cdot \sum_{i=1}^{\#X} p_i$$

*über eine Binomialverteilung berechnete wahrscheinliche Bereich für (unabhängige) Zählergebnisse (Fehler, Fehlfunktionen, ...) sind einfacher zu berechnen und bei gleichen Irrtumswahrscheinlichkeiten Obergrenzen bzw. bei gleicher Bereichsgröße sind die Irrtumswahrscheinlichkeiten Obergrenzen.*



# Poisson-Verteilung

## Poisson-Verteilung

Beim Zählen vieler seltener Ereignisse, z.B. der Fehlfunktionen bei Millionen von Service-Anforderungen, von denen nur wenige eintreten, streben die Eintrittswahrscheinlichkeit der Einzelereignisse und die Abweichung der Varianz vom Erwartungswert gegen null:

$$\begin{aligned} p_i &\rightarrow 0 \\ \text{Var}[X_i] - \mathbb{E}[X_i] = p_i \cdot (1 - p_i) - p_i = p_i^2 &\rightarrow 0 \end{aligned}$$

Die Varianz der zu zählenden Ereignisse und die der Summe streben gegen den Erwartungswert

$$\begin{aligned} \text{Var}[X_i] &= \mathbb{E}[X_i] \\ \text{Var}[X] = \sum_{i=1}^{\#X} \text{Var}[X_i] &= \mathbb{E}[X] = \sum_{i=1}^{\#X} \mathbb{E}[X_i] = \lambda \end{aligned}$$

Die Verteilung der Summe strebt gegen die Poisson-Verteilung:

$$X \sim \text{Pois}(\lambda)$$





### Die Poisson-Verteilung

$$\mathbb{P}[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

hat nur den Parameter

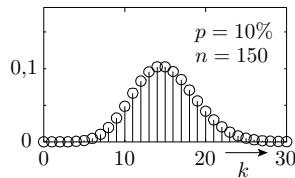
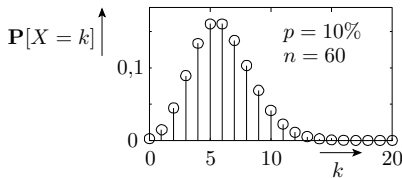
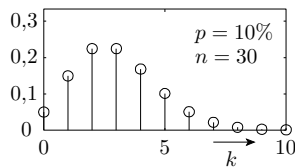
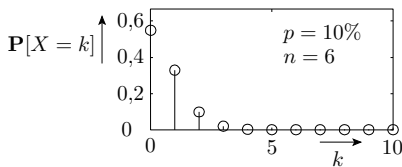
$$\lambda = \mathbb{E}[X] = \text{Var}[X] = \sum_{i=1}^{\#X} p_i$$

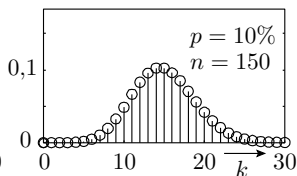
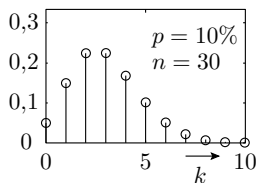
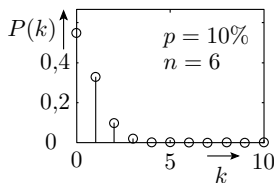
der die Summe der Eintrittswahrscheinlichkeiten, dass ein »potentieller Zählwert« eins, d.h. ein »echter Zählwert« und gleichzeitig Erwartungswert und Varianz ist ( $n = \#X$  – Anzahl der potentiellen Zählwerte;  $p_i$  – Wahrscheinlichkeit »Zählwert eins«,  $p$  – mittlere Wahrscheinlichkeit »Zählwert eins«).

Eine Poisson-Verteilung mit  $\lambda = n \cdot p$  nähert für  $p \ll 1$  eine Zählverteilung gut, an berechnet sich aus nur einem (zu schätzenden) Parameter. Geschätzter Bereich bei gleichen Irrtumswahrscheinlichkeiten garantiert größer als bei tatsächlicher Verteilung und Binomialverteilungsapproximation.

## Anzahl der Zählversuche und Verteilung

$$\mathbb{P}[X = k] = e^{-p \cdot n} \cdot \frac{(p \cdot n)^k}{k!}$$





Grobabschätzung der wahrscheinlichen Bereiche:

- Für  $\lambda = \mathbb{E}[X] = \text{Var}[X] p \cdot n < 3$  keine untere Schranke  $x_{\min} > 0$ .  
Ober Schranke:

$$k_{\max} > 3 \dots 5 \cdot \lambda$$

- Für  $\mathbb{E}[X] \approx 3 \dots 10$  zusätzlich unter Schranke:

$$k_{\max} < \frac{\lambda}{3 \dots 5}$$

- Für  $\mathbb{E}[X] > 10$  Abschätzung über Normalverteilung (siehe später):

$$[k_{\min}, k_{\max}] \approx \mathbb{E}[X] \cdot \left(1 \mp 2 \dots 4 \cdot \sqrt{\lambda}\right)$$



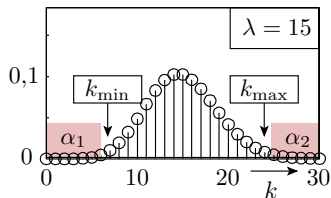
## Bereichsschätzung, Poisson

## Schätzen von $k_{\min}$

Vorgabe  $k_{\min}$  und  $\alpha_1$ . Numerische

Suche  $\lambda(k_{\min}, \alpha_1)$ , so dass

$$\sum_{k=0}^{k_{\min}-1} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \leq \alpha_1$$



$\alpha_1$	$k_{\min} = 1$	2	3	4	5	6
0,5%	5,298	7,430	9,273	10,978	12,593	14,150
1%	4,606	6,638	8,406	10,045	11,605	13,109
2%	3,912	5,834	7,516	9,084	10,580	12,027
10%	2,303	3,890	5,323	6,681	7,993	9,275
20%	1,609	2,995	4,279	5,514	6,721	7,906

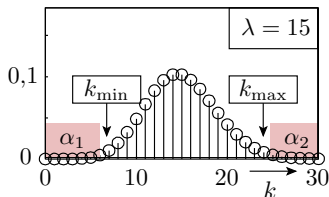
Beispielabschätzungen:

- $\lambda = 7$  und  $\alpha_1 \leq 1\% \Rightarrow k_{\min} = 2$
- $k_{\min} = 1$  und  $\alpha_1 = 2\% \Rightarrow \lambda \geq 3,912$

## Schätzen von $k_{\max}$

Vorgabe  $k_{\max}$  und  $\alpha_2$ . Numerische  
Suche  $\lambda(k_{\max}, \alpha_2)$ , so dass

$$\sum_{k=0}^{k_{\max}} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \geq 1 - \alpha_2$$



$\alpha_2$	$k_{\max} = 0$	1	2	3	4	5	6
0,5%	0,005	0,103	0,338	0,672	1,078	1,537	2,037
1%	0,01	0,148	0,436	0,823	1,279	1,785	2,330
2%	0,02	0,215	0,567	1,016	1,529	2,089	2,684
10%	0,105	0,532	1,102	1,744	2,432	3,152	3,894
20%	0,223	0,824	1,534	2,296	3,089	3,903	4,733

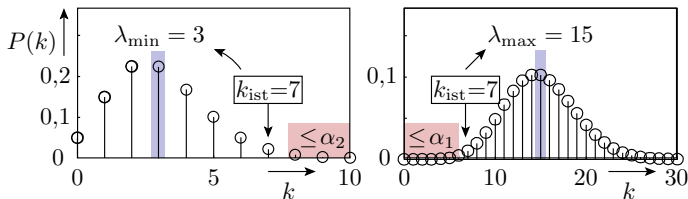
Beispielabschätzungen:

- $\lambda = 2$  und  $\alpha_2 \leq 1\% \Rightarrow k_{\max} = 6$
- $k_{\max} = 3$  und  $\alpha_2 = 2\% \Rightarrow \lambda \leq 1,016$



b

## Schätzen von $[\lambda_{\min}, \lambda_{\max}]$ aus $x_{\text{ist}}$



Aus den Tabellen der beiden Folien zuvor ist ablesbar:

$\alpha_1 = \alpha_2$	$k_{\text{ist}} = 1$	$k_{\text{ist}} = 2$	$k_{\text{ist}} = 3$
0,5%	[0,10, 5,30]	[0,34, 7,43]	[0,67, 9,27]
1%	[0,15, 4,60]	[0,44, 6,64]	[0,82, 8,41]
2%	[0,22, 3,91]	[0,57, 5,83]	[1,02, 7,52]
10%	[0,53, 2,30]	[1,10, 3,89]	[1,74, 5,32]
20%	[0,82, 1,61]	[1,53, 2,99]	[2,30, 4,28]





## 2. Näherungen für ZV 3. Bereichsschätzung, Poisson

$\alpha_1 = \alpha_2$	$k_{\text{ist}} = 4$	$k_{\text{ist}} = 5$	$k_{\text{ist}} = 6$
0,5%	[1,08, 11,0]	[1,54, 12,6]	[2,04, 14,2]
1%	[1,28, 10,0]	[1,79, 11,6]	[2,33, 13,1]
2%	[1,53, 9,08]	[2,09, 10,6]	[2,68, 12,0]
10%	[2,43, 6,68]	[3,15, 7,99]	[3,89, 9,28]
20%	[3,09, 5,51]	[3,90, 6,73]	[4,73, 7,91]

Für  $k_{\text{ist}} = 0$  ist  $\lambda_{\min} = 0$ . Für  $\lambda_{\max}$  gilt:

$$\sum_{k=0}^0 e^{-\lambda_{\max}} \cdot \frac{\lambda_{\max}^k}{k!} = e^{-\lambda_{\max}} = \alpha_1$$

$$\lambda_{\max} = -\ln(\alpha_1)$$

$\alpha_1$	0,5%	1%	2%	10%	20%
$\lambda_{\max}$	5,30	4,61	3,91	2,30	1,61%

Beispiele:

- Kein Schadensfall in 10 Jahren. Zu erwartende Anzahl von Schadensfällen in den nächsten 10 Jahren ( $\alpha_1 = 1\%$ ): 0 bis 4,61
- 5 FF pro Tag. Zu erwartende Anzahl der FF für die kommenden 10 Tage ( $\alpha_1 = 1\%$ ): 17,7 bis 116



## Abschätzungen einer FF-Rate



Mit  $n = 10^5$  Service-Anforderungen wurden drei Fehlfunktionen beobachtet. Auf welche Unter- und Obergrenze für die FF-Rate lässt sich mit Irrtumswahrscheinlichkeit  $\alpha_1 = \alpha_2 = 1\%$  schließen?

$\alpha_1 = \alpha_2 = 1\%$	$k_{\text{ist}} = 1$	$k_{\text{ist}} = 2$	$k_{\text{ist}} = 3$
$[\lambda_{\min}, \lambda_{\max}]$	$[0,15, 4,60]$	$[0,44, 6,64]$	$[0,82, 8,41]$

Abschätzbarer Bereich der FF-Rate:

$$\zeta_{\min} = \frac{\lambda_{\min}}{n} = 0,82 \cdot 10^{-5} \text{ FF/SL}$$

$$\zeta_{\max} = \frac{\lambda_{\max}}{n} = 8,41 \cdot 10^{-5} \text{ FF/SL}$$

Kleine Zählwerte erlauben nur grobe Abschätzungen. Genauere Abschätzungen verlangen größere Zählwerte.



## Schätzen der Maskierungswahrscheinlichkeit

Eine Überwachungseinheit hat von  $n = 10.000$  FF 5 FF nicht erkannt. In welchem Bereich liegt mit einer Irrtumswahrscheinlichkeit von  $\alpha = 1\%$  die Maskierungswahrscheinlichkeit?

$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = 0,5\%$	$k_{\text{ist}} = 4$	$k_{\text{ist}} = 5$	$k_{\text{ist}} = 6$
$[\lambda_{\min}, \lambda_{\max}]$	$[1,08, 11,0]$	$[1,54, 12,6]$	$[2,04, 14,2]$

Abschätzbarer Bereich der Maskierungswahrscheinlichkeit:

$$p_{F. \min} = \frac{\lambda_{\min}}{n} = 1,54 \cdot 10^{-4}$$

$$p_{F. \max} = \frac{\lambda_{\max}}{n} = 12,6 \cdot 10^{-4}$$



## Schätzen eines Zuverlässigkeitsbereichs

Beim Test eines Systems mit  $10^3$  Service-Leistungen wurden 6 Fehlfunktionen beobachtet. Auf welchen Bereich der Zuverlässigkeit kann nach diesem Versuchsergebnis mit den Irrtumswahrscheinlichkeiten  $\alpha_1 = \alpha_2 = 10\%$  geschlussfolgert werden?

$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = 10\%$	$k_{\text{ist}} = 4$	$k_{\text{ist}} = 5$	$k_{\text{ist}} = 6$
$[\lambda_{\min}, \lambda_{\max}]$	[2,43, 6,68]	[3,15, 7,99]	[3,89, 9,28]

- Abschätzbarer Bereich der FF-Rate:

$$\zeta_{\min} = 3,89 \cdot 10^{-3} \text{ FF/SL}$$

$$\zeta_{\max} = 9,28 \cdot 10^{-3} \text{ FF/SL}$$

- Abschätzbarer Bereich der Zuverlässigkeit:

$$Z_{\min} = \frac{1}{\zeta_{\max}} = 108 \text{ SL/FF}$$

$$Z_{\max} = \frac{1}{\zeta_{\min}} = 257 \text{ SL/FF}$$



# Normalverteilung

## Normalverteilung

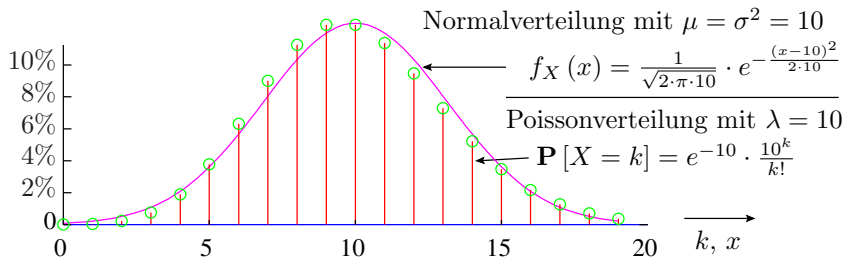
Die Summe sehr vieler unabhängiger Zufallsvariablen strebt unter sehr allgemeinen Bedingungen

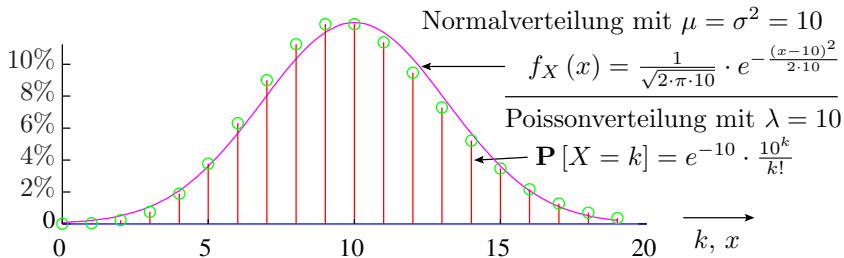
- kein Summand hat dominanten Einfluss, ...

gegen eine Normalverteilung:

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}} \quad \text{mit } \sigma = \text{sd}[X], \mu = \mathbb{E}[X]$$

Beispiel: Poisson- und Normalverteilung mit  $\mu = \sigma^2 = \lambda = 10$ :





Für unabhängige Zählwerte genügt die Annäherung der Zähl- durch eine Normalverteilung in der Regel bereits unter der Bedingung

$$10 \leq \mu \leq \#X - 10$$

( $\#X$  – Anzahl der Zählversuche;  $p_i$  – Eintrittswahrscheinlichkeiten;  
 $\mu = \sigma^2 = \sum_{i=1}^{\#X} p_i$  – Erwartungswert und Varianz der Zählwerte).

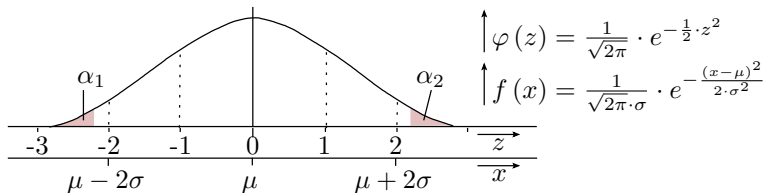
Die Annäherung durch eine Normalverteilung eignet sich gut für Abschätzung wahrscheinlicher Bereiche großer Zählwerte.



## Bereichsschätzung NVT



## Bereichsschätzung mit Normalverteilung



- Transformation einer Zufallsvariablen  $X$  mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$  in Zufallsvariablen  $Z$  mit Erwartungswert null und Standardabweichung eins:

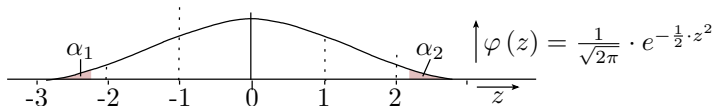
$$Z = \frac{X - \mu}{\sigma}$$

- Transformation der Werte  $x$  von  $X$  in  $z$  von  $Z$ , so dass  $F_X(x) = \Phi(z)$ :

$$z = \frac{x - \mu}{\sigma}$$

- Abschätzen der Irrtumswahrscheinlichkeiten mit einer Tabelle der standardisierten Normalverteilungsfunktion.

## Bestimmen der Irrtumswahrscheinlichkeiten



Standardisierte Normalverteilungsfkt. für  $z \geq 0$  in Schritten von 0,1:

$$\Phi(z) = \int_{-\infty}^z \varphi(z) \cdot dz$$

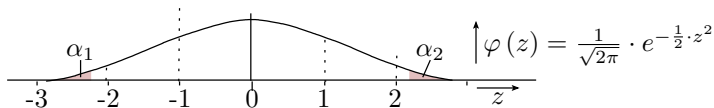
$z$	...,0	...,1	...,2	...,3	...,4	...,5	...,6	...,7	...,8	...,9
0,...	0,5000	0,5398	0,5793	0,6179	0,6554	0,6915	0,7257	0,7580	0,7881	0,8159
1,...	0,8413	0,8643	0,8849	0,9032	0,9192	0,9332	0,9452	0,9554	0,9641	0,9713
2,...	0,9772	0,9821	0,9861	0,9893	0,9918	0,9938	0,9953	0,9965	0,9974	0,9981
3,...	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000

Wegen Symmetrie für  $z < 0$ :  $\Phi(-z) = 1 - \Phi(z)$

Irrtumswahrscheinlichkeiten:  $\alpha_1 = \Phi(z_{\min}) = 1 - \Phi(-z_{\min})$

$\alpha_2 = 1 - \Phi(z_{\max})$

## Bestimmen wahrscheinlicher Bereiche



$z$	...,0	...,1	...,2	...,3	...,4	...,5	...,6	...,7	...,8	...,9
0,...	0,5000	0,5398	0,5793	0,6179	0,6554	0,6915	0,7257	0,7580	0,7881	0,8159
1,...	0,8413	0,8643	0,8849	0,9032	0,9192	0,9332	0,9452	0,9554	0,9641	0,9713
2,...	0,9772	0,9821	0,9861	0,9893	0,9918	0,9938	0,9953	0,9965	0,9974	0,9981
3,...	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000

Inverse standardisierte Normalverteilung zur Bereichsschätzung:

$\alpha$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\Phi^{-1}(1 - \alpha)$	2	3	4	2,05	2,33	2,57	2,88	3,10

Ablesen aus der Tabelle:  $z_{\min} = \Phi^{-1}(\alpha_1) = -\Phi^{-1}(1 - \alpha_1)$

$z_{\max} = \Phi^{-1}(1 - \alpha_2)$

Transformation:

$x_{\min} = \mu + \sigma \cdot z_{\min}$

$x_{\max} = \mu + \sigma \cdot z_{\max}$

## Beispielaufgaben

$z$	...,0	...,1	...,2	...,3	...,4	...,5	...,6	...,7	...,8	...,9
0,...	0,5000	0,5398	0,5793	0,6179	0,6554	0,6915	0,7257	0,7580	0,7881	0,8159
1,...	0,8413	0,8643	0,8849	0,9032	0,9192	0,9332	0,9452	0,9554	0,9641	0,9713
2,...	0,9772	0,9821	0,9861	0,9893	0,9918	0,9938	0,9953	0,9965	0,9974	0,9981
3,...	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000

$\alpha$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\Phi^{-1}(1 - \alpha)$	2	3	4	2,05	2,33	2,57	2,88	3,10

Zufallsvariable  $X$ ,  $\mu = 20$ ,  $\sigma = 5$ :

$$1 \quad \mathbb{P}[X \geq 30] \Rightarrow \mathbb{P}\left[Z \geq \frac{30-20}{5}\right] = 1 - \Phi(2) = 0,0227$$

$$2 \quad \mathbb{P}[X \leq 15] \Rightarrow \mathbb{P}\left[Z \leq \frac{15-20}{5}\right] = \Phi(-1) = 1 - \Phi(1) = 0,1587$$

$$3 \quad \alpha_2 \leq 1\% \Rightarrow z_{\max} = \Phi^{-1}(1 - 1\%) = 2,33$$

$$x_{\max} = 20 + 2,33 \cdot 5 = 31,65$$

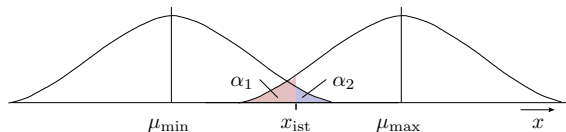
$$4 \quad \alpha_1 \leq 2\% \Rightarrow z_{\min} = \Phi^{-1}(2\%) = -\Phi^{-1}(1 - 2\%) = -2,05$$

$$x_{\min} = 20 - 2,05 \cdot 5 = 9,75$$

## Bereichsschätzung für den Erwartungswert

Der Erwartungswert zu einem beobachteten Ereignis ist

- mindestens so groß, dass die Wahrscheinlichkeit, dass ein größeres als das beobachtete Ergebnis eintritt, kleiner  $\alpha_2$ , und
- maximal so groß, dass ein kleineres als das beobachtete Ergebnis eintritt, kleiner  $\alpha_1$ , ist.



Untere und obere Bereichsgrenze des Erwartungswertes:

$$\mu_{\min} = x_{\text{ist}} - \sigma \cdot \Phi^{-1}(1 - \alpha_2)$$

$$\mu_{\max} = x_{\text{ist}} + \sigma \cdot \Phi^{-1}(1 - \alpha_1)$$

Beispiel:  $x_{\text{ist}} = 100$ ,  $\sigma = 10$ ,  $\alpha_1 = \alpha_2 = 1\%$ ,  $\Phi^{-1}(1 - 1\%) = 2,33$

$$[\mu_{\min}, \mu_{\max}] = 100 \mp 10 \cdot 2,33$$

## Bereichsschätzung unabhängiger Zählwerte

Für normalverteilte unabhängiger Zählwerte mit Erwartungswert  $\mu$  beträgt die Varianz max.:

$$\sigma^2 \leq \underbrace{\mu \cdot \left(1 - \frac{\mu}{\#X}\right)}_{\text{Var(Bin)}=n \cdot p \cdot (1-p)} < \underbrace{\mu}_{\text{Var(Pois)}}$$

- Abschätzung der Unter- und Obergrenze des wahrscheinlichen Bereichs:

$$k_{\min} \approx \mu - \sigma \cdot \Phi^{-1}(1 - \alpha_1)$$

$$k_{\max} \approx \mu + \sigma \cdot \Phi^{-1}(1 - \alpha_2)$$

- Abschätzung der oberen und unteren Schranke für den Erwartungswert zu einem Istwert:

$$\sigma^2 \approx x_{\text{ist}} \cdot \left(1 - \frac{x_{\text{ist}}}{\#X}\right) < x_{\text{ist}}$$

$$\mu_{\min} \approx x_{\text{ist}} - \sigma \cdot \Phi^{-1}(1 - \alpha_2)$$

$$\mu_{\max} \approx x_{\text{ist}} + \sigma \cdot \Phi^{-1}(1 - \alpha_1)$$

## Bereichsschätzung FF-Rate und Zuverlässigkeit

Bei der Abarbeitung von  $\#SL = 20.000$  SL wurden  $x_{\text{ist}} = 100$  FF beobachtet. In welchem Bereich liegt in 99% der Fälle die FF-Rate und die Zuverlässigkeit in einem künftigen Beobachtungszeitraum ( $\alpha_1 = \alpha_2 = 0,5\%$ , keine Abhängigkeiten)?

$\alpha$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\Phi^{-1}(1 - \alpha)$	2	3	4	2,05	2,33	2,57	2,88	3,10

$$\mu_{\min} \approx 100 - \sqrt{100} \cdot \Phi^{-1}(1 - 0,5\%) = 100 - 25,7 = 74,3$$

$$\mu_{\max} \approx 100 + \sqrt{100} \cdot \Phi^{-1}(1 - 0,5\%) = 100 + 25,7 = 125,7$$

Bereich der FF-Rate:

$$\zeta_{\min} = \frac{74,3}{20.000} = 0,37\%; \quad \zeta_{\max} = \frac{125,7}{20.000} = 0,64\%$$

Bereich der Zuverlässigkeit:

$$Z_{\min} = \frac{1}{\zeta_{\max}} = 159; \quad Z_{\max} = \frac{1}{\zeta_{\min}} = 269$$



# Varianzerhöhung





## Varianzerhöhung durch Abhängigkeiten

Abhängigkeiten erhöhen Varianz und Standardabweichung und die Breiten der wahrscheinlichen Bereiche.

Wenn z.B. zwei Zählereignisse immer paarweise gleichzeitig eintreten, ist das beschreibbar durch eine Summe von halb so vielen unabhängigen Zufallsvariablen mit den möglichen Werten 0 und 2:

$$X = \sum_{i=1}^{\#X/2} X_i$$

$$\mathbb{P}[X_i = k] = \begin{cases} 1 - p_i & k = 0 \\ p_i & k = 2 \end{cases}$$

Erwartungswert der Summanden:

$$\mathbb{E}[X_i] = 0 \cdot (1 - p_i) + 2 \cdot p_i = 2 \cdot p_i$$

Varianz der Summanden (nach Verschiebungssatz):

$$\begin{aligned} \text{Var}[X_i] &= (1 - p_i) \cdot 0^2 + p_i \cdot 2^2 - (2 \cdot p_i)^2 \\ &= 2^2 \cdot p_i \cdot (1 - p_i) \end{aligned}$$



Der gesamte Erwartungswert ist derselbe wie für  $\#X$  unabhängige Zählerereignisse mit paarweise gleichen Eintrittswahrscheinlichkeiten:

$$\mathbb{E}[X] = \sum_{i=1}^{\#X/2} 2 \cdot p_i$$

Die Varianz der Summe verdoppelt sich gegenüber der einer Summe unabhängige Zählerereignisse:

$$\text{Var}[X] = \sum_{i=1}^{\#X/2} 2^2 \cdot p_i \cdot (1 - p_i) = 2 \cdot \left( 2 \cdot \sum_{i=1}^{\#X/2} p_i \cdot (1 - p_i) \right)$$

und Standardabweichung und wahrscheinliche Bereiche vergrößern sich um  $\sqrt{2}$ .

Die Varianzerhöhung sei definiert als Verhältnis aus Varianz und Erwartungswert:

$$\kappa = \max \left( 1, \frac{\text{Var}[X]}{\mathbb{E}(X)} \right)$$

Für kleine  $p_i \ll 1$  ist sie im Beispiel  $\kappa = 2$ . Analog lässt sich zeigen, wenn immer  $n$  Zählereignisse gleichzeitig eintreten:

$$\kappa = n$$

## Schätzen der Varianzerhöhung

- Experimentelle Bestimmung von  $\#w \geq 2$  Zählwerten  $w_i$ .
- Schätzen des Erwartungswerts der Zählwertstichprobe:

$$\hat{\mathbb{E}}[X] = \frac{1}{\#w} \cdot \sum_{i=1}^{\#w} w_i$$

- Schätzen der Varianz der Zählwertstichprobe:

$$\hat{\text{Var}}[X] = \frac{1}{\#w - 1} \cdot \sum_{i=1}^{\#w} (w_i - \hat{\mathbb{E}}[X])^2$$

- Schätzwert der Varianzerhöhung:

$$\kappa = \max \left( 1, \frac{\hat{\text{Var}}[X]}{\hat{\mathbb{E}}(X)} \right)$$

## Beispielabschätzung der Varianzerhöhung

$n = 2.000$  Zählereignisse.  $\#w = 10$  Wiederholungen des Zählversuchs. Ergebnisse (Zählwerte):

Versuch $i$	1	2	3	4	5	6	7	8	9	10
Ergebnis $w_i$	44	87	58	62	59	57	65	57	75	67

- Erwartungswert der Zählwertstichprobe nach Gl. 3:

$$\hat{\mathbb{E}}[X] = \frac{1}{10} \cdot \sum_{i=1}^{10} w_i = 63,1$$

- Varianz der Zählwertstichprobe nach Gl. 4:

$$\hat{\text{Var}}[X] = \frac{1}{9} \cdot \sum_{i=1}^{10} (w_i - 63,1)^2 = 135$$

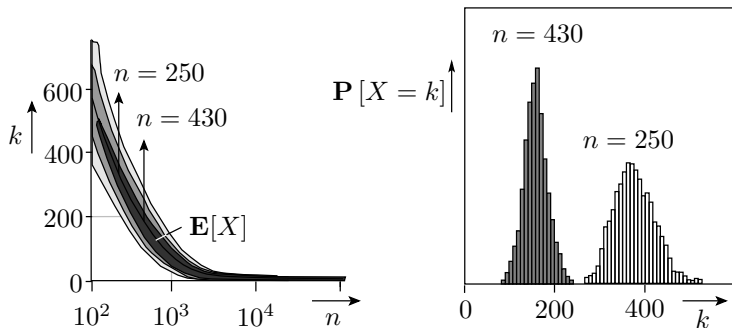
- Geschätzte Varianzerhöhung<sup>5</sup>:

$$\hat{\kappa} = \frac{135}{63,1} \approx 2,14$$

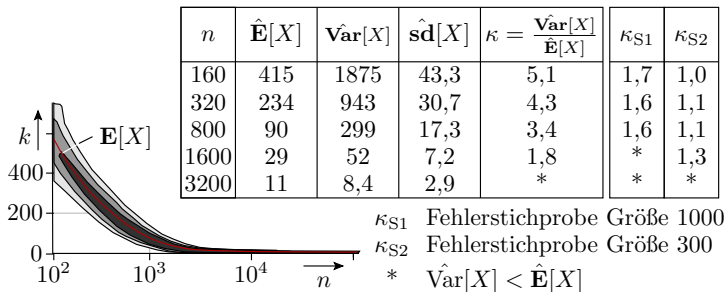
<sup>5</sup>Die Abhängigkeiten erhöhen die Varianz so, als ob mehr als 2 Zählereignisse fast immer gemeinsam eintreten.

## Experiment mit Haftfehlern

Kombinatorische Beispielschaltung (Benchmark c3540). 3606 simulierte, unterschiedlich nachweisbare Haftfehler. Zählwert  $X$  ist die Anzahl der nicht nachweisbaren Haftfehler. Abschätzung von  $\mathbb{P}[X = k]$  aus einer Stichprobe von  $\#w = 1000$  Zählwerten für verschiedene Zufallstestsätze der Länge  $n$ .



## Varianzerhöhung im Experiment



Zwischen den nicht nachweisbaren Fehlern gibt es offenbar Abhängigkeiten, die die Varianz so stark erhöhen, als ob 3...5 Modellfehler identisch Fehler nachweisbar wären. Identisch nachweisbare Fehler wurden jedoch nicht mitgezählt. Bleiben als Abhängigkeitsursache implizit nachweisbare Fehler sowie geteilte Steuer- und Beobachtungsbedingungen.

Bei weniger nicht nachweisbaren Fehlern oder einer Fehlerstichprobe statt der kompletten Modellfehlermenge ist  $\kappa$  deutlich kleiner.

## Fehlermodellierung und Vorhersagbarkeit

Die Breite des wahrscheinlichen Bereichs der Fehlerüberdeckung:

$$\text{Bereich } FC = \frac{\mu_{\max} - \mu_{\min}}{\#F_M} \sim \frac{\sqrt{\kappa \cdot FC \cdot (1 - FC) \cdot \#F_M}}{\#F_M} \sim \sqrt{\frac{\kappa}{\#F_M}}$$

$\mu_{\min}, \mu_{\max}$  – Schranken zu erwartenden Anzahl nachweisbare Fehler.

- Für im Verhältnis zur Testobjektgröße kleiner  $\#F_M$  ist keine Varianzerhöhung zu erwarten ( $\kappa < 1$ ). Vorhersagbarkeit  $FC$  wächst mit der Modellfehleranzahl  $\#F_M$ .
- Für im Verhältnis zur Testobjektgröße großes  $\#F_M$  ist eine Varianzerhöhung  $\kappa \sim \#F_M$  zu erwarten. Keine Verbesserung der Vorhersagbarkeit von  $FC$  durch mehr Modellfehler.

Thesen:

- Es hat wenig Nutzen, zu viele Modellfehler im Verhältnis zur Testobjektgröße zur Abschätzung von  $FC$  zu verwenden.
- Insbesondere Fehlermodell, bei denen die Anzahl der Modellfehler überproportional mit der Testobjektgröße zunimmt, wie z.B. beim Pfadverzögerungsmodell, sind nicht zielführend.



## Bereichsschätz. Zählwerte



## Bereichsschätzung normalverteilter Zählwerte

$\alpha$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\Phi^{-1}(1 - \alpha)$	2	3	4	2,05	2,33	2,57	2,88	3,10

Wenn Abhängigkeiten zwischen Zählwerten bestehen (können), lassen sich diese durch eine (max. mögliche) Varianzerhöhung  $\kappa$  berücksichtigen. Der garantierbare wahrscheinliche Bereich verbreitert sich dann um  $\sqrt{\kappa}$ :

$$k_{\min} \approx \mu - \sqrt{\kappa} \cdot \sigma \cdot \Phi^{-1}(1 - \alpha_1)$$

$$k_{\max} \approx \mu + \sqrt{\kappa} \cdot \sigma \cdot \Phi^{-1}(1 - \alpha_2)$$

bzw. für  $\alpha = \alpha_1 + \alpha_2$ :

$$[k_{\min}, k_{\max}] \approx \mathbb{E}[X] \mp \sqrt{\kappa} \cdot \sigma \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

mit der oberen Schranke für die Varianz ohne Abhängigkeiten:

$$\sigma^2 \leq \underbrace{\mu \cdot \left(1 - \frac{\mu}{\#X}\right)}_{\text{Var(Bin)}=n \cdot p \cdot (1-p)} < \underbrace{\mu}_{\text{Var(Pois)}}$$

$\mu$  – zu erwartender Zählwert.



### Beispiel

$\alpha$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\Phi^{-1}(1 - \alpha)$	2	3	4	2,05	2,33	2,57	2,88	3,10

Der zu erwartende Zählwert für die Anzahl von Schadensfällen sei 100. Irrtumswahrscheinlichkeit  $\alpha_1 + \alpha_2 = 1\%$ , Varianzerhöhung  $\kappa = 2$ . In welchem Bereich wird bei künftigen Wiederholungen des Versuchs die Anzahl der Schadensfälle liegen?

$$\begin{aligned} [k_{\min}, k_{\max}] &\approx \mu \mp \sqrt{\kappa} \cdot \sigma \cdot \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \\ &= 100 \mp 10 \cdot \sqrt{2} \cdot 2,33 \end{aligned}$$

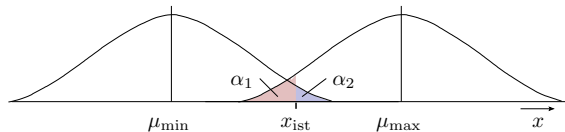
$$k_{\min} \leq 67$$

$$k_{\max} \geq 133$$

## Bereichsschätzung für den Erwartungswert

Der Erwartungswert zu einem beobachteten Ereignis ist

- mindestens so groß, dass die Wahrscheinlichkeit, dass ein größeres als das beobachtete Ergebnis eintritt, kleiner  $\alpha_2$ , und
- maximal so groß, dass ein kleineres als das beobachtete Ergebnis eintritt, kleiner  $\alpha_1$ , ist.



Untere und obere Bereichsgrenze des Erwartungswertes:

$$\mu_{\min} = x_{\text{ist}} - \sigma \cdot \Phi^{-1}(1 - \alpha_2)$$

$$\mu_{\max} = x_{\text{ist}} + \sigma \cdot \Phi^{-1}(1 - \alpha_1)$$

Abschätzung für die Standardabweichung für  $0 \ll x_{\text{ist}} \ll \#X$ :

$$\sigma \approx \sqrt{\kappa \cdot \mu} \approx \sqrt{\kappa \cdot x_{\text{ist}}}$$



## Beispiel

Von 1000 Modellfehlern wurden 32 nicht erkannt. Gesucht ist der Bereich für die zu erwartende Anzahl der nicht nachweisbaren Modellfehler mit  $\alpha_1 = \alpha_2 = 1\%$  und Varianzerhöhung  $\kappa = 2$ .

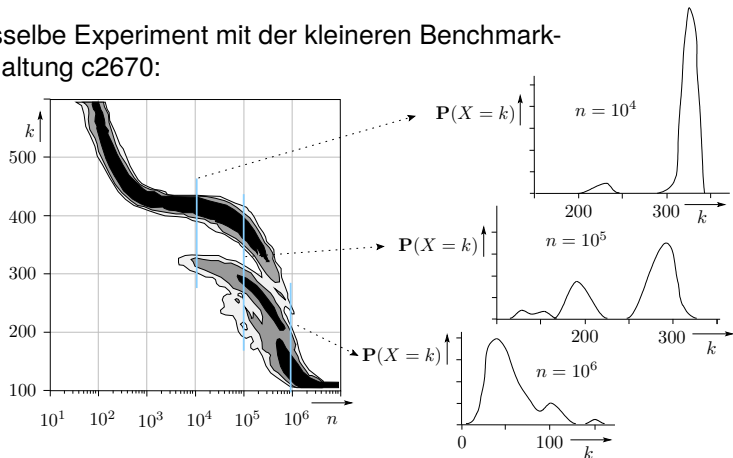
$\alpha$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\Phi^{-1}(1 - \alpha)$	2	3	4	2,05	2,33	2,57	2,88	3,10

$$\begin{aligned}
 [\mu_{\min}, \mu_{\max}] &\approx x_{\text{ist}} \mp \sqrt{\kappa \cdot x_{\text{ist}}} \cdot \Phi^{-1}(1 - \alpha_2) \\
 &= 32 \mp \sqrt{64} \cdot 2,33 = 32 \mp 18,64 \\
 \mu_{\min} &\approx 13,36 \\
 \mu_{\max} &\approx 50,64
 \end{aligned}$$

Die Abschätzung gilt nur für normalverteilte Zählwerte. Es ist allerdings nicht garantiert, dass untereinander abhängige Zählwerte normalverteilt sind.

## Nicht normalverteilte Zählwerte

Dasselbe Experiment mit der kleineren Benchmark-Schaltung c2670:



Im Bereich von  $n = 10^4$  bis  $10^6$  mehrere Gipfel. Keine näherungsweise Normalverteilung.

Wie ist das möglich?



# Misch- und multimodale Verteilung



### Mischverteilung

Aus einer Grundgesamtheit von  $\#X_{\text{ges}}$  Objekten, von den jeweils  $\#X_i$  Objekte eine Verteilung  $X_i \sim F_{X_i}$  haben wählt eine Zufallsvariable  $Y$  zufällig ein Objekt aus:

$$\mathbb{P}[Y = i] = \frac{\#X_i}{\#X_{\text{ges}}} = h_i$$

$$F_X(x) = \mathbb{P}[X \leq x] = \sum_{i=1}^{\#i} h_i \cdot F_{X_i}(x)$$

Für diskrete Verteilungen:

$$\mathbb{P}[X = x] = \sum_{i=1}^{\#i} h_i \cdot \mathbb{P}[X_i = x]$$

Für stetige Verteilungen:

$$f_X(x) = \frac{dF_X(x)}{dx} = \sum_{i=1}^{\#i} h_i \cdot f_{X_i}(x)$$



### Zufallsexperimente für Mischverteilungen

- Eigenschaft einer Schraube (z.B. Länge) bei zufälliger Auswahl auf einer Kiste mit Schrauben unterschiedlicher Hersteller.
- Fehleranzahl eines SW-Bausteins bei zufällige Auswahl aus Angeboten unterschiedlicher Programmierer mit unterschiedlichen Fehlerentstehungsraten.
- Schadenshöhe eines zufälligen Schadens auf einer Menge unterschiedlicher Schadensklassen mit unterschiedlicher Kostenverteilung.
- ...





### 3. Misch- und multimodale Verteilung

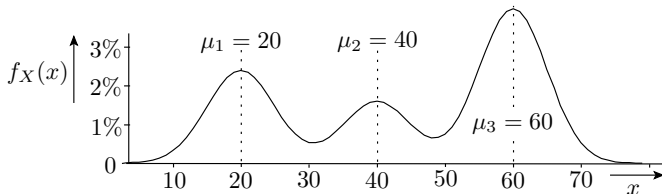
## Beispiel

Mischung von 3 normalverteilten Zufallsvariablen  $X_i$ :

$h_i$	0,3	0,2	0,5
$\mu_i$	20	40	60
$\sigma_i$	5	5	5

$$f_X(x) = \frac{dF_X(x)}{dx} = 0,3 \cdot \varphi\left(\frac{x-20}{5}\right) + 0,2 \cdot \varphi\left(\frac{x-40}{5}\right) + 0,5 \cdot \varphi\left(\frac{x-60}{5}\right)$$

$\varphi(z) = \varphi\left(\frac{x-\mu_i}{\sigma_i}\right)$  – Dichte der standardisierten Normalverteilung.





### 3. Misch- und multimodale Verteilung

## Varianzvergrößerung durch »Mischung«

Der Erwartungswert ist der gewichtete Mittelwert:

$$\mathbb{E}[X] = \mu = \sum_{i=1}^{\#i} h_i \cdot \mu_i$$

Varianz mit  $\mu = \mu_i + \delta_i$ :

$$\begin{aligned}\text{Var}[X] &= \sum_{i=1}^{\#i} h_i \cdot \mathbb{E}[(X_i - \mu_i - \delta_i)^2] \\ &= \sum_{i=1}^{\#i} h_i \left( \underbrace{\mathbb{E}[(X_i - \mu_i)^2]}_{\sigma_i^2} - \underbrace{\mathbb{E}[2 \cdot \delta_i \cdot (X_i - \mu_i)]}_{2 \cdot \delta_i \cdot \mathbb{E}[X_i - \mu_i] = 0} + \underbrace{\mathbb{E}[\delta_i^2]}_{\delta_i^2} \right) \\ \text{Var}[X] &= \sum_{i=1}^{\#i} h_i \cdot \sigma_i^2 + \sum_{i=1}^{\#i} h_i \cdot \delta_i^2\end{aligned}$$

Mittelwert der Einzelvarianzen plus mittlere quadratische Abweichung der Einzelerwartungswerte vom Gesamterwartungswert.



# 3. Misch- und multimodale Verteilung

Für das Beispiel  $\sigma_i = 5$  und

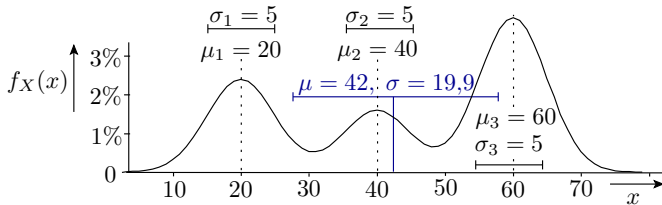
$h_i$	0,3	0,2	0,5
$\mu_i$	20	40	60

$$\mu = \sum_{i=1}^{\#i} h_i \cdot \mu_i = 0,3 \cdot 20 + 0,2 \cdot 40 + 0,5 \cdot 60 = 42$$

$$\sigma^2 = \sum_{i=1}^{\#i} h_i \cdot \sigma_i^2 + \sum_{i=1}^{\#i} h_i \cdot \delta_i^2$$

$$= 25 + 0,3 \cdot (20 - 42)^2 + 0,2 \cdot (40 - 42)^2 + 0,5 \cdot (60 - 42)^2 = 285$$

$$\sqrt{\sigma^2} = 16,9$$

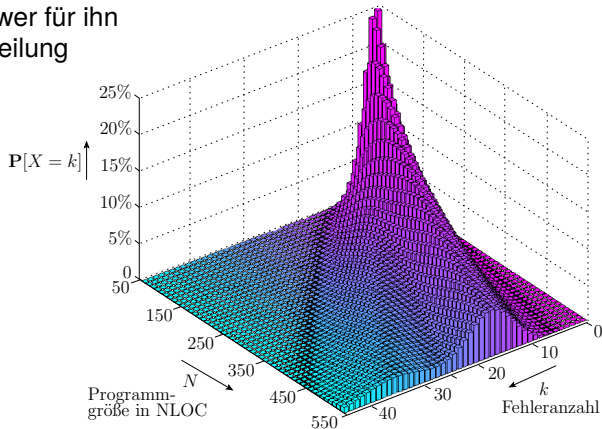




## Unterschiedlich gute Programmierer

Ein Anfänger und ein Profi entwickeln Software-Bausteine aus  $N$  NLOC (Netto Lines of Code), der Profi 66% mit ca. einem Fehler je 30 NLOC und der Anfänger 33% mit einem Fehler je 15 NLOC. Der Kunde weiß nicht, wer für ihn programmiert. Verteilung der Fehleranzahl:

$$\begin{aligned} \mathbb{P}[N, X = k] &= \underbrace{\frac{2}{3} \cdot e^{-\frac{N}{30}} \cdot \frac{\left(\frac{N}{30}\right)^k}{k!}}_{\text{Pois}\left(\frac{N}{30}\right)} \\ &+ \underbrace{\frac{1}{3} \cdot e^{-\frac{N}{15}} \cdot \frac{\left(\frac{N}{15}\right)^k}{k!}}_{\text{Pois}\left(\frac{N}{15}\right)} \end{aligned}$$

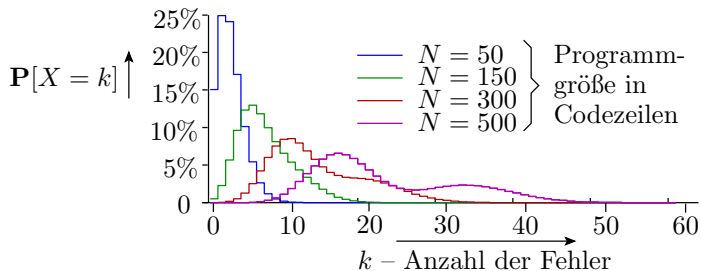




### 3. Misch- und multimodale Verteilung

Die Wahrscheinlichkeit, dass ein Modul genau  $k$  Fehler enthält, ist  $2/3$  mal die Wahrscheinlichkeit, dass es  $k$  Fehler enthält und vom Profi stammt plus  $1/3$  mal die Wahrscheinlichkeit, dass es vom Anfänger stammt:

$$\mathbb{P}[N, X = k] = \frac{2}{3} \cdot e^{-\frac{N}{30}} \cdot \frac{\left(\frac{N}{30}\right)^k}{k!} + \frac{1}{3} \cdot e^{-\frac{N}{15}} \cdot \frac{\left(\frac{N}{15}\right)^k}{k!} \quad (14)$$



Die Polarisierung nimmt mit der Größe der Software-Bausteine, die vom Profi und vom Anfänger getrennt entwickelt werden, zu.



### Beispiel: Identisch nachweisbare Fehler



In einer Modellfehlermenge aus  $N = 25$  Fehlern mit einer Nachweiswahrscheinlichkeit  $p = 60\%$  seien zehn Fehler identisch und die übrigen Fehler unabhängig voneinander nachweisbar. Gesucht:

- 1 Beschreibung als Mischverteilung von zueinander verschobenen Binomialverteilungen.
- 2 Erwartungswert, Varianz und Standardabweichung.
- 3 Varianzerhöhung  $\kappa$  gegenüber 25 mit  $p = 0,6$  unabhängig voneinander nachweisbaren Fehlern.

## Lösung

- 1 Verteilung ohne die 10 nur gemeinsam nachweisbaren Fehler:

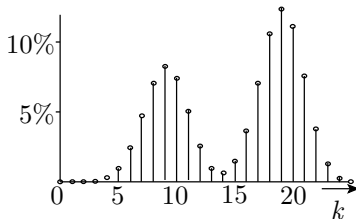
$$\mathbb{P}[X_0 = k] = \begin{cases} \binom{N-10}{k} \cdot p^k \cdot (1-p)^{N-10-k} & 0 \leq k \leq N-10 \\ 0 & \text{sonst} \end{cases}$$

mit den 10 nur gemeinsam nachweisbaren Fehlern:

$$\mathbb{P}[X_1 = k] = \begin{cases} \binom{N-10}{k-10} \cdot p^{k-10} \cdot (1-p)^{N-k} & 10 \leq k \leq N \\ 0 & \text{sonst} \end{cases}$$

Mischverteilung:

$$\mathbb{P}[X = k] = (1-p) \cdot \mathbb{P}[X_0 = k] + p \cdot \mathbb{P}[X_1 = k]$$





### 3. Misch- und multimodale Verteilung

$i$	$\mathbb{E}[X_i]$	$\text{Var}[X_i]$
0	$(N - 1) \cdot p = 15 \cdot 0,6 = 9$	$(N - 1) \cdot p \cdot (1 - p) = 3,6$
1	$\mathbb{E}[X_0] + 10 = 19$	$\text{Var}[X_0] = 3,6$

#### 2 Erwartungswert:

$$\mathbb{E}[X] = 25 \cdot p = 25 \cdot 60\% = 15$$

$$\stackrel{!}{=} (1 - p) \cdot \mathbb{E}[X_0] + p \cdot \mathbb{E}[X_1] = (1 - 0,6) \cdot 9 + 0,6 \cdot 19 = 15\checkmark$$

Varianz als Summe der Varianzen der Summanden:

$$\text{Var}[X] = 15 \cdot p \cdot (1 - p) + 10^2 \cdot p \cdot (1 - p) = 115 \cdot p \cdot (1 - p) = 27,6$$

$$\stackrel{!}{=} \underbrace{n \cdot p \cdot (1 - p)}_{3,6} + \underbrace{(1 - p) \cdot (\mathbb{E}[X_0] - \mathbb{E}[X])^2}_{+ 0,4 \cdot (9 - 15)^2} + \underbrace{p \cdot (\mathbb{E}[X_1] - \mathbb{E}[X])^2}_{+ 0,6 \cdot (19 - 15)^2} = 27,6\checkmark$$

$$\text{sd}[X] = \sqrt{\text{Var}[X]} = 5,25$$

#### 3 Varianz für 25 mit $p = 0,6$ unabhängig nachweisbare Fehler:

$$\text{Var}[X_{U25}] = 25 \cdot p \cdot (1 - p) = 25 \cdot 0,6 \cdot (1 - 0,6) = 6$$

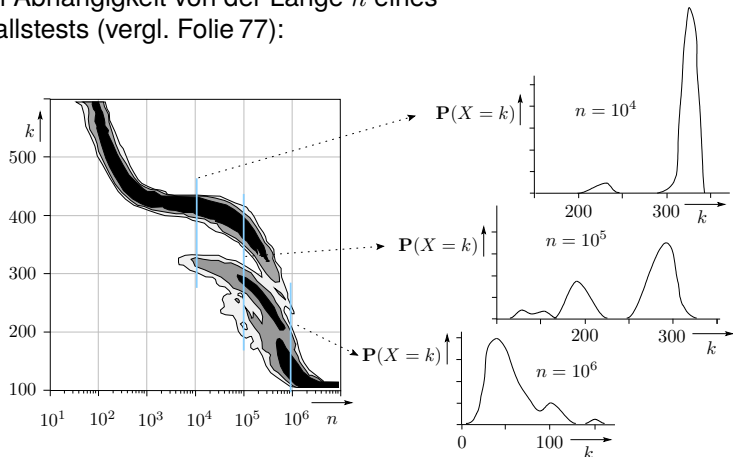
Varianzerhöhung:

$$\kappa = \frac{27,6}{6} = 4,6$$



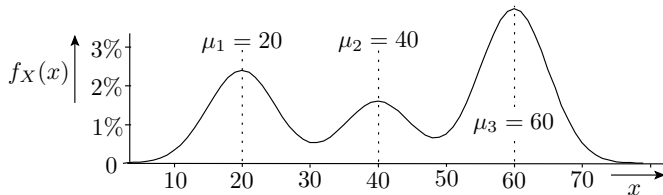
## Dichte nicht nadannchweisbare Fehler c2670

... in Abhängigkeit von der Länge  $n$  eines  
Zufallstests (vergl. Folie 77):



Im Bereich von  $n = 10^4$  bis  $10^6$  multimodale Verteilung. Offenbar ca. 80  
sehr ähnlich nachweisbare Fehler mit FF-Rate  $\zeta_i \approx 10^{-5} \text{ FF/SL}$ .

## Multimodalität und Fehlervermeidung

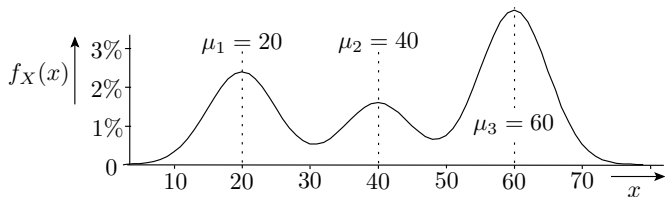


Wenn die Erwartungswerte deutlich auseinander liegen, entsteht eine multimodale (mehrgipflige) Verteilung. Die Multimodalität deutet auf Polarisierungen der Beobachtungswerte (Zugehörigkeit zu unterschiedlichen Verteilungen). Polarisierungen können wichtige Informationen über die Natur der untersuchten Variablen liefern:

- Abhängigkeiten bei der Fehlerentstehung, bei Ausfällen beim Fehlernachweis und beim Versagen von Service-Leistungen,
- Vorlieben oder Neigung befragter Experten, z.B. bei der Einschätzung von Gefährdungen und Risiken,
- Probleme eines Messverfahrens, ...



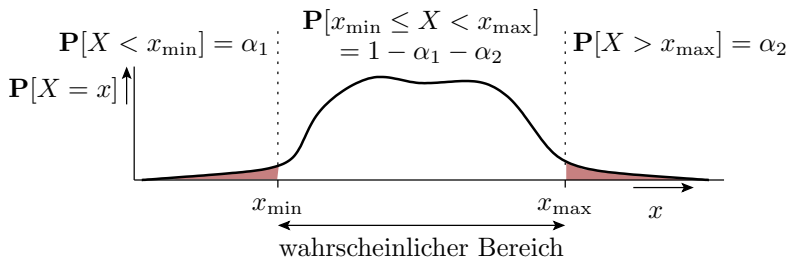
### 3. Misch- und multimodale Verteilung



Wenn die Zufallsvariable ein Gütemaß ist, hat man es offenbar mit einer zufälligen Mischung von besser und schlechter funktionierenden Prozessabläufen zu tun. Dann ist es natürlich interessant, warum der Entstehungsprozess mal besser und mal schlechter funktioniert, um das schlechtere Funktionierende zu eliminieren.

Auch bei normalen Parametern, die in einem Toleranzbereich liegen müssen, deutet Multimodalität in der Regel auf Prozessfehler, bei denen es sich lohnt, sie zu suchen und zu beseitigen.

## Bereichsschätzung, wenn Verteilung unbekannt



Die Bestimmung eines wahrscheinlichen Intervalls  $[x_{\min}, x_{\max}]$

- auch möglich, wenn Verteilung unbekannt, multimodal, ...
- Voraussetzung: eine hinreichend kleine Varianz.



## Das schwache Gesetz der großen Zahlen

Nach der tschebyscheffschen Ungleichung:

$$\mathbb{P}[|x - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

ist die Wahrscheinlichkeit, das der Wert einer Zufallsvariable mehr als ein Intervallradius  $\varepsilon$  von seinem Erwartungswert abweicht, nicht größer als das Verhältnis der Varianz zum Quadrat des Intervallradius  $\varepsilon$ .

Intervallradius für  $\alpha = \alpha_1 + \alpha_2$ :

$$\varepsilon \geq \sqrt{\frac{\text{Var}[X]}{\alpha}} = \frac{\text{sd}[X]}{\sqrt{\alpha}}$$

Wahrscheinlicher Bereich des Erwartungswerts bei einer bekannten Realisierung  $x_{\text{ist}}$ :

$$[\mathbb{E}[X]_{\min}, \mathbb{E}[X]_{\max}] = x_{\text{ist}} \mp \varepsilon$$

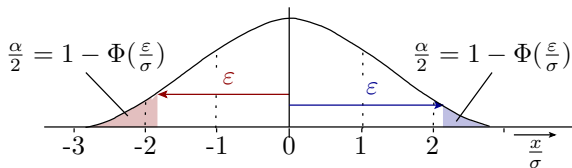
Wahrscheinlicher Bereich künftiger experimenteller Ergebnisse bei bekanntem Erwartungswert  $\mathbb{E}[X]$ :

$$[x_{\min}, x_{\max}] = \mathbb{E}[X] \mp \varepsilon$$



### 3. Misch- und multimodale Verteilung

## Zum Vergleich: Intervallradius bei Normalverteilung



Intervallradius für Normalverteilung und  $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ :

$$\epsilon = \sigma \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

Intervallradius für Verteilung unbekannt:

$$\epsilon = \sigma \cdot \alpha^{-0,5}$$

$\alpha$	4,55%	0,26%	0	4%	2%	1%	0,4%	0,2%
$\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$	2	3	4	2,05	2,33	2,57	2,88	3,10
$\alpha^{-0,5}$				5	7,07	10	15,8	22,4



### Beispielaufgabe



Gegeben sei eine Stichprobe gemessener Widerstandswerte in  $k\Omega$ :

$$R_i : 10,3, 10,5, 9,7, 8,9, 10,1, 11,0, 10,2, 9,5$$

Aus dieser Stichprobe soll

- 1 ohne weitere Vorkenntnisse über die Verteilung und
- 2 unter der Annahme, dass die Widerstandswerte normalverteilt sind,

auf den möglichen Bereich des Erwartungswertes geschlussfolgert werden. Zugelassene Irrtumswahrscheinlichkeit  $\alpha = 2\%$ .



### 3. Misch- und multimodale Verteilung

## Lösung

$\alpha$	4,55%	0,26%	0	4%	2%	1%	0,4%	0,2%
$\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$	2	3	4	2,05	2,33	2,57	2,88	3,10
$\frac{1}{\sqrt{\alpha}}$				5	7,07	10	15,8	22,4

Erwartungswert und Standardabweichung der Datenstichprobe:

$$\hat{\mathbb{E}}[R] = \frac{1}{8} (10,3 + \dots) \text{ k}\Omega = 10,025 \text{ k}\Omega$$

$$\hat{\text{s}}\text{d}[R] = \sqrt{\frac{1}{7} ((10,3 - 10,025)^2 + \dots)} \text{ k}\Omega^2 = 647 \Omega$$

Garantierbarer Bereich für den Erwartungswert:

**1** Ohne Kenntnis der Verteilung:

$$\hat{\mathbb{E}}[R] \mp \frac{\hat{\text{s}}\text{d}[R]}{\sqrt{\alpha}} = 10,025 \text{ k}\Omega \mp \frac{647 \Omega}{\sqrt{2\%}} = [5,3 \text{ k}\Omega, 14,8 \text{ k}\Omega]$$

**2** Für normalverteilte Widerstandswerte:

$$\hat{\mathbb{E}}[R] \mp \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \hat{\text{s}}\text{d}[R] = 0,025 \text{ k}\Omega \mp 2,33 \cdot 647 \Omega = [8,5 \text{ k}\Omega, 11,5 \text{ k}\Omega]$$





# Weitere Verteilungen



## 4. Weitere Verteilungen

Verteilungen für

- Nachweislängen,
- Schadenskosten,
- Lebensdauer.



# Pareto-Verteilung



# Das Pareto-Prinzip<sup>6</sup>

Statistisches Phänomen, dass ein kleiner Teil der Ursachen für den überwiegenden Teil der Wirkungen verantwortlich ist:

- Wenige Entstehungsursachen  $\Rightarrow$  Mehrheit der Fehler.
- Wenige Fehler  $\Rightarrow$  Mehrheit der FF.
- Wenige FF  $\Rightarrow$  Mehrheit der Schadenskosten.
- Wenige Zufallstests erkennen die Mehrheit der Fehler.

---

<sup>6</sup>Der italienische Ökonom Vilfredo Pareto untersuchte 1906 die Verteilung des Grundbesitzes in Italien und fand heraus, dass ca. 20 % der Bevölkerung ca. 80 % des Bodens besitzen. Das ist in den Sprachgebrauch als Pareto-20%-80%-Regel eingegangen.

## Pareto-Verteilung

$X \sim \text{Par}(k, x_{\min})$  ist eine stetige Wahrscheinlichkeitsverteilung auf einem rechtsseitig unendlichen Intervall  $[x_{\min}, \infty)$  und genügt dem Potenzgesetz:

$$F_X(x) = \mathbb{P}[X \leq x] = 1 - \left(\frac{x_{\min}}{x}\right)^k \quad \text{für } x \geq x_{\min} \text{ und } k > 0$$

( $k$  – Formfaktor;  $x_{\min}$  – Skalenparameter). Dichtefunktion für  $x \geq x_{\min}$ :

$$f_X(x) = \frac{k \cdot x_{\min}^k}{x^{k+1}}$$

Für kleine Exponenten gehört sie zu den endlastigen Verteilungen, bei denen ein erheblicher Teil der Wahrscheinlichkeitsmasse auf große  $x$  entfällt.



Ein Erwartungswert:

$$\mathbb{E}[X] = \int_{x_{\min}}^{\infty} \frac{k \cdot x_{\min}^k}{x^{k+1}} \cdot x \cdot dx = \frac{k \cdot x_{\min}^k}{1-k} \cdot \left( \lim_{x \rightarrow \infty} x^{1-k} - x_{\min}^{1-k} \right)$$

existiert nur für  $k > 1$ :

$$\mathbb{E}[X] = x_{\min} \cdot \frac{k}{k-1}$$

Eine Varianz existiert nur für  $k > 2$ :

$$\text{Var}[X] = x_{\min}^2 \cdot \frac{k}{(k-2)(1-k)^2}$$

Wir werden mit Parato-Verteilungen annähern:

- die Nachweislänge für reale Fehler mit zufälligen Eingaben und
- die Schadenskosten am Beispiel von Haftpflichtschäden für KFZ.

Die Verteilung der Nachweislänge wird wegen  $0 < k < 1$  weder einen Erwartungswert noch eine Varianz haben und die Schadenskosten werden wegen  $1 < k < 2$  keine Varianz haben, was die Versicherung der Schadensfälle schwierig macht.

## Pareto-Prinzip ✓

Der Anteil der Ursachen  $U$  mit der größten Wirkung:

$$U = \int_{w_{\min}}^{\infty} f(x) \cdot dx = \int_{w_{\min}}^{\infty} \frac{k \cdot x_{\min}^k}{x^{k+1}} \cdot dx = \left( \frac{x_{\min}}{w_{\min}} \right)^k$$

hat mindestens die Wirkung:  $w_{\min} = x_{\min} \cdot U^{-\frac{1}{k}}$ . Zu erwartende anteilige Wirkung und zu erwartende Gesamtwirkung:

$$\mathbb{E}[X|X \geq w_{\min}] = \int_{w_{\min}}^{\infty} \frac{k \cdot x_{\min}^k}{x^{k+1}} \cdot x \cdot dx = \frac{k}{k-1} \cdot x_{\min} \cdot \left( \frac{x_{\min}}{w_{\min}} \right)^{k-1}$$

$$\mathbb{E}[X] = \int_{x_{\min}}^{\infty} \frac{k \cdot x_{\min}^k}{x^{k+1}} \cdot x \cdot dx = x_{\min} \cdot \frac{k}{k-1}$$

(Voraussetzung  $k > 1$ ). Anteilige Gesamtwirkung:

$$W = \frac{\mathbb{E}[X|X \geq w_{\min}]}{\mathbb{E}[X]} = \left( \frac{x_{\min}}{w_{\min}} \right)^{k-1} = U^{\frac{k-1}{k}}$$

$U = 20\%^*$	$k = 1,1$	$k = 1,2$	$k = 1,15$	$k = 1,175$	$k = 1,16$
$W = U^{\frac{k-1}{k}}$	86,4%	76,5%	81,1%	78,7%	80,1*

\* Pareto-20%-80%-Regel



## Verteilung der Fehlenachweislänge

Die Nachweislänge  $N$  ist die Anzahl der SL bzw. die Zeit bis zum Nachweis eines Fehlers. Die Verteilungsfunktion als Wahrscheinlichkeit »Nachweislänge  $N$  nicht größer als Testsatzlänge  $n$ « ist gleich »zu erwartende Fehlerüberdeckung für Testsatzlänge  $n$ «:

$$F_N(n) = \mathbb{P}[N \leq n] = \mathbb{E}[FC(n)]$$

Bei einem Zufallstest verlangt eine Verringerung des Anteils der nicht nachweisbaren Fehler um eine Dekade in der Regel eine Erhöhung der Testsatzlänge um mehr als eine Dekade.

Annäherung durch das Potenzgesetz (siehe auf Foliensatz F1):

$$1 - \mathbb{E}[FC(n)] \approx \left(\frac{n}{n_0}\right)^{-k} \quad \text{mit } n \geq n_0 \text{ und } 0 < k < 1$$

$k$	1	0,5	0,33	0,25
$\frac{n}{n_0}$ für $1 - FC(n) = 0,1$	10	100	$10^3$	$10^4$



## Übergang zur Pareto-Verteilung

Zusammenfassung der Gleichungen der vorherigen Folie:

$$F_N(n) = \mathbb{P}[N \leq n] = \mathbb{E}[FC(n)]$$

$$1 - \mathbb{E}[FC(n)] [FC(n)] \approx \left(\frac{n}{n_0}\right)^{-k} \quad \text{mit } n \geq n_0 \text{ und } 0 < k < 1$$

$$F_N(n) = 1 - \left(\frac{n}{n_0}\right)^{-k}$$

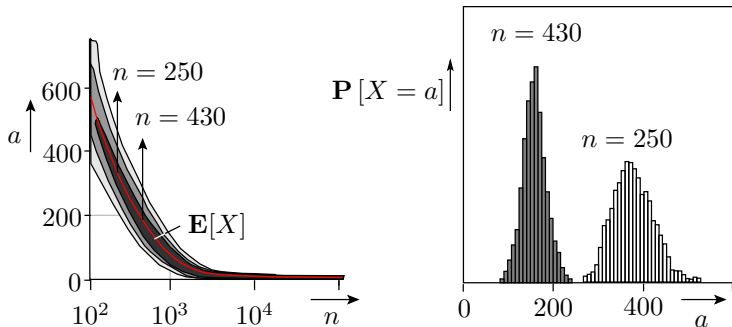
$n_0$  – Bezugstestsatzlänge für  $FC = 0$ ;  $n$  – Testsatzlänge incl.  $n_0$ .

Annäherung durch eine stetige Pareto-Verteilung durch Annäherung von  $n/n_0$  durch ein Verhältnis aus Testzeit und Bezugszeit  $t/t_0$ :

$$F_N(t) = \mathbb{P}[N \leq t] = 1 - \left(\frac{t}{t_0}\right)^{-k} \quad \text{mit } t \geq t_0 \text{ und } 0 < k < 1$$

Wegen  $k < 1$  existiert  $\mathbb{E}[N]$  nicht. Beim Betrieb von IT-System mit pareto-verteilter Nachweislänge und Beseitigung aller erkennbaren Fehler sind auch nach sehr langer Nutzungsdauer weitere Fehler nie ausschließbar.

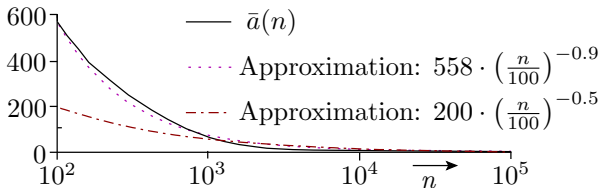
### Für das Haftfehlerexperiment



Kombinatorische Beispielschaltung (Benchmark c3540). 3606 simulierte, unterschiedlich nachweisbare Haftfehler. Zu erwartende Anzahl der nicht nachweisbaren Fehler:

$$\bar{a}(n) = 3606 \cdot (1 - \mathbb{E}[FC(n)])$$

## Pareto-Näherung



$$\bar{a}(n) = 3606 \cdot (1 - \mathbb{E}[FC(n)])$$

$$\mathbb{E}[FC(n)] = 1 - \frac{\bar{a}(n)}{3606} \approx F_N(t) = 1 - \left(\frac{n}{n_0}\right)^{-k}$$

$$\bar{a}(n) = \frac{3606}{n_0^k} \cdot \left(\frac{n}{n_0}\right)^{-k}$$

Nicht perfekt. Die Approximation mit  $k = 0,9$  nähert Bereich  $n < 1000$  und die mit  $k = 0,5$  Bereich  $n > 1000$  Testschritte besser an.

---

Gibt es bessere geeignete Verteilungen? Forschungsbedarf!



# Gammaverteilung



## Gamma-Verteilung

$\mathcal{G}(\alpha, \beta)$  ist eine stetige Wahrscheinlichkeitsverteilung auf einem rechtsseitig unendlichen Intervall  $[0, \infty)$  z.B. zur Modellierung

- von Bedien- und Reparaturzeiten (Warteschlangentheorie),
- kleiner und mittlerer Schäden (Versicherungsmathematik),
- der FF-Raten von Fehlern in IT-Systemen.

Dichte:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot e^{-\beta \cdot x} \cdot x^{\alpha-1} \quad \text{für } x \geq 0$$

$\alpha$  – Formparameter;  $\beta$  – Skalenparameter;  $\Gamma(\alpha)$  – Gamma-Funktion:

$$\Gamma(\alpha) = \int_0^\infty e^{-z} z^{\alpha-1} dz$$

$\alpha$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$\Gamma(\alpha)$	9,51	4,59	2,99	2,22	1,77	1,49	1,30	1,16	1,07

Für  $0 < \alpha \leq 1$  ist  $\Gamma(\alpha) \approx 1/\alpha$ . Für  $\alpha > 1$  gilt:

$$\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$$

## Eigenschaften der Gamma-Verteilung

Erwartungswert:

$$\mathbb{E}[X] = \frac{\alpha}{\beta}$$

Varianz:

$$\text{Var}[X] = \frac{\alpha}{\beta^2}$$

Die Summe gamma-verteilter Zufallsvariablen mit gleichem Skalenparameter  $X_1 \sim \mathcal{G}(\alpha_1, \beta)$  und  $X_2 \sim \mathcal{G}(\alpha_2, \beta)$  ist wieder gamma-verteilt:

$$X_1 + X_2 \sim \mathcal{G}(\alpha_1 + \alpha_2, \beta)$$

Beispiel: siehe später Verteilung der FF-Rate.



# Exponentialverteilung

## Exponentialverteilung

Die Exponentialverteilung  $\text{Exp}(\lambda) = \mathcal{G}(1, \lambda)$  ist eine Gamma-Verteilung mit Formparameter  $\alpha = 1$  und Skalenparameter  $\beta = \lambda$ . Dichte:

$$\begin{aligned} f_X(x) &= e^{-\beta \cdot x} \cdot x^{\alpha-1} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \quad \text{für } x \geq 0 \text{ mit } \alpha = 1 \text{ und } \beta = \lambda \\ &= \lambda \cdot e^{-\lambda \cdot x} \quad \text{für } x \geq 0 \end{aligned}$$

$\lambda$  – Anzahl der zu erwartenden Ereignisse pro Zeitintervall.  
Verteilungsfunktion:

$$F_X(x) = \mathbb{P}[X \leq x] = \int_0^x f(z) \cdot dz = 1 - e^{-\lambda \cdot x} \quad \text{für } x \geq 0$$

Erwartungswert:

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

Varianz<sup>7</sup>:

$$\text{Var}[X] = \frac{1}{\lambda^2} \quad \Rightarrow \quad \text{sd}[X] = \mathbb{E}[X]$$

---

<sup>7</sup>Keine Abnahme von  $\text{sd}[X]/\mathbb{E}[X]$  mit  $\mathbb{E}[X]$ , d.h. keine Abnahme des relativen Schätzfehlers wie bei Zählwerten. Schlecht für Bereichsabschätzungen.





## Beispiel: Verteilung Nachweislänge $\zeta_i$ bekannt

Für einen Fehler mit FF-Rate  $\zeta_i$  ist die Verteilungsfunktion der Nachweisdauer  $N_i$  die Wahrscheinlichkeit, dass Fehler  $i$  mit  $n$  Testschritten nachweisbar ist:

$$F_N(n) = \mathbb{P}[N_i \leq n] = p_i(n)$$

Für zufällige Testbeispiele:

$$p_i(n) = 1 - (1 - \zeta_i)^n = 1 - e^{\ln(1-\zeta_i) \cdot n}$$

Für  $\zeta \ll 1$  und der Taylor-Reihe  $\ln(1-x) = -\left(x + \frac{x^2}{2} + \frac{x^3}{3} + \dots\right)$

$$F_N(n) = p_i(n) = 1 - e^{-\zeta_i \cdot n}$$

Annäherung  $\zeta_i \cdot n$  durch  $\lambda_i \cdot t \Rightarrow$  exponentialverteilte Nachweislänge:

$$\mathbb{P}[N_i \leq t] = 1 - e^{-\lambda \cdot t} = \text{Exp}(t)$$



# Verteilung der FF-Rate



## Verteilung der FF-Rate

Bei exponential-verteilter Nachweislänge bei bekanntem  $\zeta_i$  (zufälliger Fehlernachweis):

$$\mathbb{P}[N_i \leq n] = p_i(n) = 1 - e^{-\zeta_i \cdot n}$$

entsteht die postulierte pareto-verteilte Nachweislänge für reale Fehler:

$$\mathbb{P}[N \leq n] = \mathbb{E}[FC(n)] = 1 - \left(\frac{n}{n_0}\right)^{-k} \quad \text{mit } n \geq n_0 \text{ und } 0 < k < 1$$

wenn die FF-Rate der Fehler gamma-verteilt ist

$$h(\zeta) = \frac{n_0^k}{\Gamma(k)} \cdot e^{-n_0 \cdot \zeta} \cdot \zeta^{k-1}$$

mit Formfaktor  $k$  und Skalenparameter  $n_0$ .



## Beweisgedanke

Wenn die Fehler mit einer von der Testsatzlänge abhängigen Wahrscheinlichkeit gefunden werden, ist die Fehlerüberdeckung die »nachweisbare Wahrscheinlichkeitsmasse«:

$$\mathbb{E}[FC(n)] = \int_0^1 h(\zeta) \cdot p(\zeta, n) \cdot d\zeta$$

Mit

$$p(\zeta, n) = 1 - e^{-\zeta \cdot (n - n_0)} \quad \text{mit } n \geq n_0$$

und

$$\mathbb{E}[FC(n)] = 1 - \left(\frac{n}{n_0}\right)^{-k} \quad \text{mit } n \geq n_0 \text{ und } 0 < k < 1$$

muss die Dichte der FF-Rate folgende Gleichung befriedigen:

$$1 - \left(\frac{n}{n_0}\right)^{-k} = \int_0^1 h(\zeta) \cdot \left(1 - e^{-\zeta \cdot (n - n_0)}\right) \cdot d\zeta$$

$$\left(\frac{n}{n_0}\right)^{-k} = \int_0^1 h(\zeta) \cdot e^{-\zeta \cdot (n - n_0)} \cdot d\zeta$$

Erwartete Lösung:

$$h(\zeta) = \frac{n_0^k}{\Gamma(k)} \cdot e^{-n_0 \cdot \zeta} \cdot \zeta^{k-1}$$

## Kontrolle durch Einsetzen

$$\left(\frac{n}{n_0}\right)^{-k} \stackrel{!}{=} \int_0^1 h(\zeta) \cdot e^{-\zeta \cdot (n-n_0)} \cdot d\zeta \quad \text{mit } h(\zeta) = \frac{n_0^k}{\Gamma(k)} \cdot e^{-n_0 \cdot \zeta} \cdot \zeta^{k-1}$$

$$\begin{aligned} \left(\frac{n}{n_0}\right)^{-k} &\stackrel{!}{=} \int_0^1 \left(\frac{n_0^k}{\Gamma(k)} \cdot e^{-n_0 \cdot \zeta} \cdot \zeta^{k-1}\right) \cdot \left(e^{-\zeta \cdot (n-n_0)}\right) \cdot d\zeta \\ &= \frac{n_0^k}{\Gamma(k)} \cdot \int_0^{\infty*} e^{-n \cdot \zeta} \cdot \zeta^{k-1} \cdot d\zeta \end{aligned}$$

\* Erweiterung der Integrationsgrenze von 1 auf  $\infty$  verlangt  $n \gg 1$  und ist dann zulässig, wegen:

$$\int_1^{\infty} e^{-n \cdot \zeta} \cdot \zeta^{k-1} \cdot d\zeta < \int_1^{\infty} e^{-n \cdot \zeta} \cdot d\zeta = \frac{1}{n} \cdot e^{-n} = 0 \text{ für } n \gg 1$$

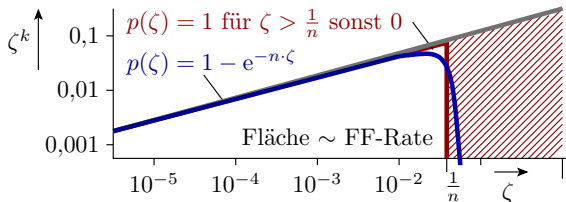
Substitution  $z = n \cdot \zeta$ ,  $d\zeta = \frac{dz}{n}$ :

$$\frac{n_0^k}{\Gamma(k)} \cdot n^{-k} \cdot \underbrace{\int_0^{\infty} e^{-z} z^{k-1} dz}_{\Gamma(k)} = \left(\frac{n}{n_0}\right)^{-k} \quad \checkmark$$



# Test & Zuverlässigkeit

## FF-Rate durch nicht beseitigten Fehler



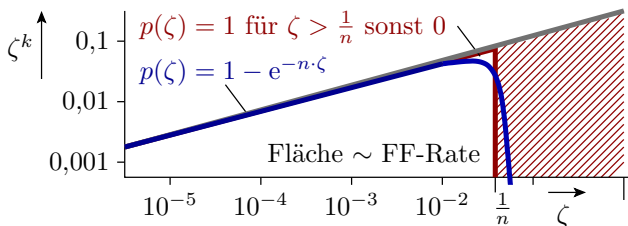
Zu erwartende Anzahl der nicht beseitigten Fehler zur Produktfreigabe:

$$\mathbb{E}[\#F(n)] = \mathbb{E}[\#F] \cdot (1 - \mathbb{E}[FC_{SF}]) \cdot \left(\frac{n}{n_0}\right)^{-k}$$

$\#F$  – Fehleranzahl aus den Entstehungs- und Fehlerbeseitigungsprozessen;  $FC_{SF}$  – Fehlerüberdeckung statische + fehlerorientiert gesuchte Tests;  $n_0$  – Anz. statische Tests;  $n$  – Anz. dynamische Tests incl.  $n_0$ .

FF-Rate durch die nicht beseitigten Fehler:

$$\zeta_F = \mathbb{E}[\#F(n)] \cdot \bar{\zeta}$$



$$\zeta_{\text{F}} = \mathbb{E}[\#F(n)] \cdot \bar{\zeta}$$

Für eine gamma-verteilte FF-Rate:

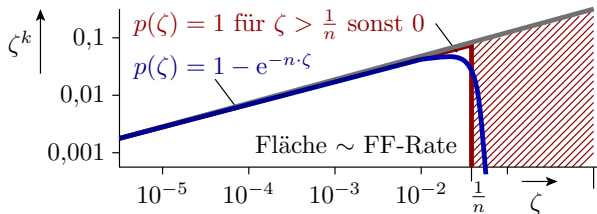
$$h(n, \zeta) = \frac{n^k}{\Gamma(k)} \cdot e^{-n \cdot \zeta} \cdot \zeta^{k-1} \quad \text{für } 0 \leq \zeta < \infty \text{ und } 0 < k < 1$$

$$\bar{\zeta} = \frac{k}{n} \quad (\text{Erwartungswert Gammaverteilung})$$

$$\begin{aligned} \zeta_{\text{F}} &= \mathbb{E}[\#F(n)] \cdot \frac{k}{n} \\ &= \mathbb{E}[\#F] \cdot (1 - \mathbb{E}[FC_{\text{SF}}]) \cdot \frac{k \cdot n_0^k}{n^{k+1}} \end{aligned}$$



## Mit der Vereinfachung von Foliensatz F1



$$h(\zeta) = k \cdot n^k \cdot \begin{cases} \zeta^{k-1} & 0 \leq \zeta < \frac{1}{n} \text{ und } 0 < k < 1 \\ 0 & \text{sonst} \end{cases}$$

$$\bar{\zeta} = \int_0^{\frac{1}{n}} \underbrace{\zeta \cdot k \cdot n^k \cdot \zeta^{k-1}}_{h(\zeta, n)} \cdot d\zeta = k \cdot n^k \cdot \int_0^{\frac{1}{n}} \zeta^k \cdot d\zeta = \frac{k}{(k+1)^* \cdot n}$$

$$\zeta_F = \mathbb{E}[\#F(n)] \cdot \frac{k}{(k+1)^* \cdot n}$$

\* Unterschied zur Annahme einer gamma-verteilten FF-Rate.



### Beispiel: Fehleranzahl und FF-Rate

Die FF-Rate eines Systems beträgt nach der effektiven Testlänge  $n_0 = 10^5$  Tests  $\zeta(n_0) = 10^{-4}$  FF/SL. Schätzen Sie für gamma-verteilte FF-Raten mit den Formfaktoren  $k \in \{0,3, 0,4, 0,5, 0,6, 0,7\}$ :

- 1 die zu erwartende Anzahl der nicht beseitigten Fehler für die aktuelle effektive Testlänge von  $n_0 = 10^5$ ,
- 2 die zu erwartende Anzahl der nicht beseitigten Fehler für die zehnfache effektive Testlänge  $n_1 = 10^6$ ,
- 3 die zu erwartende FF-Rate für die zehnfache effektive Testlänge  $n_1 = 10^6$ .

$$\zeta_F(n) = \mathbb{E}[\#F(n)] \#F(n) \cdot \frac{k}{n}$$
$$\mathbb{E}[\#F(n)] = \mathbb{E}[\#F(n_0)] \cdot \left(\frac{n}{n_0}\right)^{-k}$$
$$\zeta_F(n) = \zeta_F(n_0) \cdot \left(\frac{n}{n_0}\right)^{-(k+1)}$$



## Lösung

- 1 Für  $\zeta_F(n) = \mathbb{E}[\#F(n)] \cdot \frac{k}{n}$  sind gegeben  $n = 10^5$ ,  $\zeta_F(10^5) = 10^{-4}$  und  $k \in \{0, 3, \dots\}$ :

$$\mathbb{E}[\#F(n_0)] = \zeta_F(n_0) \cdot \frac{n_0}{k} = 10^4 \cdot \frac{10^5}{k} = \frac{10}{k}$$

- 2 Die zu erwartende Fehleranzahl nimmt mit der Erhöhung der effektiven Testlänge mit Exponent  $k$  ab:

$$\mathbb{E}[\#F(10^6)] = \mathbb{E}[\#F(10^5)] \cdot \left(\frac{10^6}{10^5}\right)^{-k} = \frac{10}{k} \cdot 10^{-k}$$

- 3 Die zu erwartende FF-Rate nimmt mit der Erhöhung der effektiven Testlänge mit Exponent  $-(k+1)$  ab:

$$\zeta_F(10^6) = \zeta_F(10^5) \cdot \left(\frac{10^6}{10^5}\right)^{-(k+1)} = 10^{-4} \cdot 10^{-(k+1)}$$



## 5. Test & Zuverlässigkeit

	$k=0,3$	$k=0,4$	$k=0,5$	$k=0,6$	$k=0,7$
$\zeta_F (10^5)$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{E} [\#F(10^5)] = \frac{10}{k}$	26,7	20	16	13,3	11,2
$\mathbf{E} [\#F(10^6)] = \frac{10}{k \cdot 10^k}$	13,4	7,96	5,06	3,45	2,28
$\zeta_F (10^6) = \frac{10^{-4}}{10^{k+1}}$	$5,01 \cdot 10^{-6}$	$3,98 \cdot 10^{-6}$	$3,16 \cdot 10^{-6}$	$2,51 \cdot 10^{-6}$	$2,00 \cdot 10^{-6}$

- Die FF-Rate eines Systems ist eine auch für den Anwender gut beobachtbare Größe.
- Mit einem Schätzwert für die bisherige effektive Testlänge lässt sich aus der FF-Rate auf die Anzahl der noch vorhandenen Fehler schließen, auch wenn über die Verteilung der FF-Rate wenig bekannt ist.
- Eine Verzehnfachung der effektiven Testlänge, z.B. durch Erhöhung der Reifedauer von 6 Monaten auf 5 Jahre reduziert die Fehlerzahl auf  $\frac{1}{2} \dots \frac{1}{5}$  und die FF-Raten auf  $\frac{1}{20} \dots \frac{1}{50}$ .



### Zuverlässigkeit und Sicherheit (korrigiert)

$$\mathbb{E}[\#F(n)] = \mathbb{E}[\#F] \cdot (1 - FC_{SF}) \cdot \left(\frac{n}{n_0}\right)$$

$$\zeta_F(n) = \frac{k \cdot \mathbb{E}[\#F] \cdot (1 - FC_{SF})}{n}$$

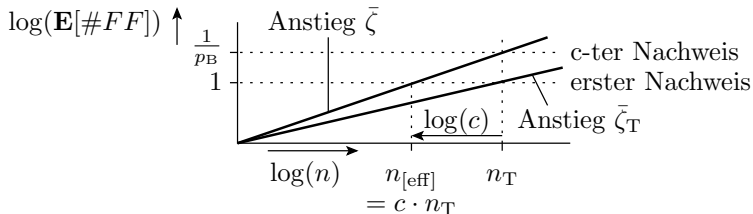
$$\mathbb{E}[Z_F(n)] = \frac{1}{\zeta_F(n)} = \frac{n}{k \cdot \mathbb{E}[\#F] \cdot (1 - FC_{SF})}$$

$$\mathbb{E}[S_F(n)] = \frac{\mathbb{E}[Z_F(n)]}{(1 - ROB) \cdot \eta_g}$$

$\zeta_F$ ,  $Z_F$ ,  $S_F$  – fehlerbezogene FF-Rate, Teilzuverlässigkeit, Sicherheit;  
 $\#F$  – Anzahl der Fehler aus den Entstehungs- und Fehlerbeseitigungsprozessen;  $FC_{SF}$  – Fehlerüberdeckung statische + fehlerorientiert gesuchte Tests;  $n_0$  – eff. Anzahl der gezielt gesuchten Tests;  $n$  – effektive Testanzahl incl.  $n_0$ ;  $k$  – Formfaktor der Verteilung der FF-Rate;  $ROB$  – Robustheit der FF-Behandlung;  $\eta_g$  – Anteil der die Sicherheit gefährdenden FF.

## Effektive Testsatzlänge (Wiederholung von F1)

Effektive Testsatzlänge  $n_{\text{[eff]}}$  ist die äquivalente Anzahl der Tests, für die alle erkennbaren Fehler beseitigt werden.



Wenn die mittlere FF-Rate je Fehler  $\bar{\zeta}_T$  in der Testsituation von  $\bar{\zeta}$  im Einsatz um einen Faktor  $c$  abweicht oder Fehler bei einer verursachten FF nur mit Wahrscheinlichkeit  $p_B = c$  beseitigt werden:

$$c = \frac{\bar{\zeta}_T}{\bar{\zeta}} \text{ bzw. } c = p_B \Rightarrow n_{\text{[eff]}} = c \cdot n_T$$

$n_{\text{[eff]}}$  – effektive Testsatzlänge;  $n_T$  – Anzahl der Test.

[Wenn Fehler in Wirklichkeit eine höhere FF-Rate haben als beim Test, dann ist die effektive Testsatzlänge kleiner als die Testsatzlänge, für die alle erkannten Fehler beseitigt werden]

## Abweichende effektive Testsatzlänge

- Beseitigung erkannter Fehler nur mit Wahrscheinlichkeit  $p_{Bes} < 1$ , z.B. Reifeprozess:

$$n_{[eff]} = p_{Bes} \cdot n_T$$

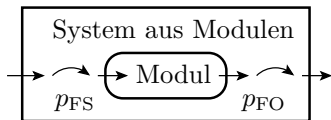
- Abweichende mittlere FF-Rate der Modellfehler  $\bar{\zeta}_{MF}$  von der der tatsächlichen Fehler  $\bar{\zeta}$  um einen Skalierungsfaktor:

$$c_{MF} = \frac{\bar{\zeta}_{MF}}{\bar{\zeta}}; \quad n_{[eff]} = c_{MF} \cdot n_T$$

- Modularer Test: FF-Rate im Betrieb:

$$\bar{\zeta} = p_{Anr} \cdot p_{Beob} \cdot n_{[eff]} \cdot \bar{\zeta}_{Modul}$$

$$n_{[eff]} = \frac{n_T}{p_{Anr} \cdot p_{Beob}}$$



$\bar{\zeta}_{Modul}$  – FF-Rate Modultest;  $p_{Anr}$  – relative Nutzungshäufigkeit des Moduls in der Systemumgebung;  $p_{Beob}$  – relative Häufigkeit, dass Verfälschungen am Modulausgang am Systemausgang beobachtbar sind.



## Fehlerüberdeckung und Testlängenskalierung

Fehlerüberdeckung für pareto-verteilter Nachweislänge:

$$\mathbb{P}[N \leq n] = \mathbb{E}[FC(n)] = 1 - \left(\frac{n}{n_0}\right)^{-k} \quad \text{mit } n \geq n_0 \text{ und } 0 < k < 1$$

verringert eine Testlängenskalierung  $c$  den zu erwartende Anteil der nicht nachweisbaren Fehler um:

$$C = \frac{\mathbb{E}[1 - FC(c \cdot n)]}{\mathbb{E}[1 - FC_C(n)]} = \frac{\left(\frac{c \cdot n}{n_0}\right)^{-k}}{\left(\frac{n}{n_0}\right)^{-k}} = c^{-k}$$

Fehleranteil- und Testlängenskalierung  $\Leftrightarrow$  Fehlerüberdeckung:

$$\begin{aligned} \mathbb{E}[FC(n)] &= 1 - C \cdot (1 - \mathbb{E}[FC_C(n)]) \\ &= 1 - c^{-k} \cdot (1 - \mathbb{E}[FC_C(n)]) \end{aligned}$$

$$\mathbb{E}[FC(c \cdot n)] = \mathbb{E}[FC_C(n)]$$

$k$	0,75	0,5	1/3	1/4
$C = c^{-k}$ für $c = 5$	0,3	0,48	0,58	0,67





## Beispiel: Fehler- und Haftfehlerüberdeckung

Foliensatz F2, Ende Abschnitt 3.2 Fehlerbeseitigung Ersatziteration:

*Typischen Literaturangaben: Absenkung des Fehleranteils durch Ersatziterationen von 50% auf 500 dpm mit  $FC_{SA} \approx 99\%$*

$$\Rightarrow C_{SA} \approx 0,1 \text{ und } c_{SA} = C_{SA}^{-\frac{1}{k}}:$$

$k$	0,75	0,5	$1/3$	$1/4$
$c_{SA} = C_{SA}^{-\frac{1}{k}}$ für $C_{SA} = 0,1$	22	100	1000	10.000

Die These  $C_{SA} \approx 0,1$  wäre glaubhaft, wenn:

$$\bar{\zeta} \approx c_{SA} \cdot \bar{\zeta}_{SA}; n_{[eff]} \approx c_{SA} \cdot n_T \text{ mit } c_{SA} = 20 \dots 10^5$$

gezeigt werden kann (siehe später Foliensatz F5, Abschnitt 2.4: Fehlermod. Nachweisbez.). Wenn nicht, würde gelten:

- Fehleranteil der Schaltkreise im Einsatz doch  $\gg 500$  dpm oder
- Typische Haftfehlerüberdeckungen doch  $\gg 99\%$ .



# Schaden durch FF



### Verteilung von Haftpflichtschäden

Mangels Zahlenmaterial über Schäden durch IT-Systeme hier die Haftpflichtschäden über 100.000 SF (SF – Schweizer Franken) einer Schweizer Autoversicherung<sup>8</sup>:

103.765, 109.168, 112.341, 113.800, 114.791, 115.731, 118.264,  
123.464, 127.611, 133.504, 142.821, 152.270, 163.491, 164.968,  
168.915, 169.346, 172.668, 191.954, 193.102, 208.522, 209.070,  
219.111, 243.910, 280.302, 313.898, 330.461, 418.074, 516.218,  
595.310, 742.198, 791.874, 822.787, 1.074.499

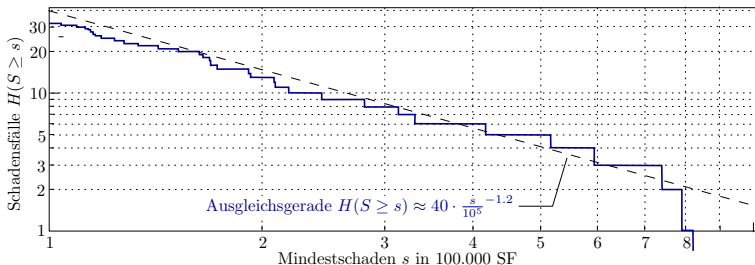
33 Schadensfälle, Gesamtschadenssumme: 9.458.208 SF

---

<sup>8</sup>Aus Klüppelberg, C. and Villasenor, J. A. (1993) Estimation of distribution tails – A semiparametric approach, Bl. Dtsch. Ges. Versicherungsmath. 21, No.2, 213-235.

## Annäherung durch eine Pareto-Verteilung

Schadenshäufigkeit in Abhängigkeit von der Schadenshöhe:



Verteilung der Schadenshöhe  $S$  für Schäden  $S > 10^5$  ist hier eine Pareto-Verteilung mit Formfaktor  $k = 1,2$  und Skalenparameter  $x_{\min} = 10^5$ :

$$F_S(s) = \mathbb{P}[S \leq s] = 1 - \left(\frac{s}{s_{\min}}\right)^{-k} = 1 - \left(\frac{s}{10^5}\right)^{-1,2}$$

Erwartungswert:

$$\mathbb{E}[S] = s_{\min} \cdot \frac{k}{k-1} = \frac{1,2}{1,2-1} \cdot s_{\min} = 600.000 \text{ SF}$$



### Abschlussbemerkungen

Eine Pareto-Verteilung hat erst für  $k > 2$  eine Varianz. Schwierig zu versichern. Vermutlich haben Versicherungen deshalb eine max. Deckungssumme.

---

Ich persönlich gehe davon aus, dass künftig Schäden durch IT z.B. in autonomen Fahrzeugen ähnlich wie heute Haftpflichtschäden durch Personen versichert werden müssen.

Der Abschnitt hier in der Vorlesung sollte zeigen, dass die mathematischen Grundlagen dafür aus der Versicherungsmathematik bekannt sind.