Test und Verlässlichkeit Foliensatz 4: Themenspezifische Einführung in Verteilungen

Prof. G. Kemnitz

Institut für Informatik, TU Clausthal (TV_F4)

10. November 2024

Inhalt Foliensatz 4

Vergangener Foliensatz »Wahrscheinlichkeit«:

- Definition und Abschätzung von Wahrscheinlichkeiten, Fehlerbäume, Markov-Ketten.
- Insbesondere Wahrscheinlichkeiten im Zusammenhang mit
  - Fehlernachweis,
  - Fehlerbeseitigung und
  - Fehlerentstehung.

Diesem Abschnitt »Verteilungen«:

- Zufallsgrößen mit mehr als zwei Werten,
- wahrscheinlicher Bereiche von Verlässlichkeitskenngrößen,
- insbesondere von geschätzten Zählwerten und Eintrittswahrscheinlichkeiten.

Verteilung

Grundbegriffe

- 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 Ergebnis eines Zufallsexperiments annimmt.
- Das einfachste Zufallsexperiment ist der Bernoulliversuch. Er hat nur die mögliche Ergebnisse 0 und 1, die Verteilung

$$\begin{aligned} \mathbb{P}[X=0] &= 1-p \\ \mathbb{P}[X=1] &= p \end{aligned} \tag{4.1}$$

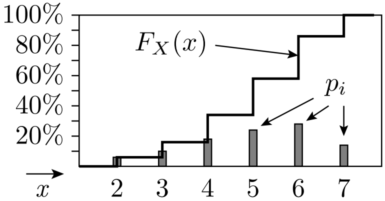
  und dient auch zur Definition der Wahrscheinlichkeit (Gl. 3.1).
- Zufallsexperimente mit (sehr) vielen möglichen Ergebnissen werden durch ihre Verteilungsfunktion beschrieben:

$$F_X(x) = \mathbb{P}[X \le x] \tag{4.2}$$

Diskrete Verteilung

Die Zufallsvariable X kann nur abzählbare Werten x_i annehmen, z.B.:

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



Anwendbar auf Zählwerte:

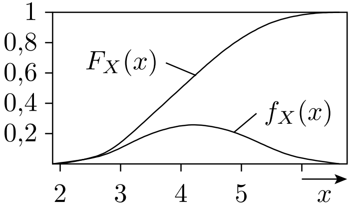
- Anzahl der Fehlfunktionen (aufgetretene, erkannte, ...),
- Anzahl der Fehler (entstandene, beseitigte, vermiedene, ...), ...

$\mathbb{P}[X = x_i]$ Verteilung der diskreten Zufallsvariablen X.
$F_X(x)$ Verteilungfunktion der Zufallsvariablen X.
p_i Eintrittswahrscheinlichkeit der Realisierung x_i.

Stetige Verteilungen

Zufallsvariable X ist stetig und hat im Intervall $a \leq X \leq b$ unendlich viele Ausprägungen. Beschreibbar auch durch ihre Dichtefunktion:

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} \tag{4.3}$$



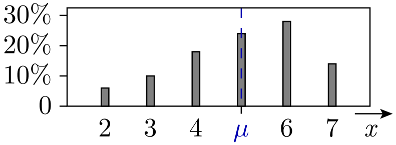
Auch als Näherung für diskrete Verteilungen mit sehr vielen möglichen Werten.

$F_X(x)$	Verteilungfunktion der Zufallsvariablen X.
$f_X(x)$	Dichtefunktion der Zufallsvariablen X.

Charakteristische Größen

Erwartungswert

Mittlerer Eintrittswert bei einer großen Versuchsanzahl:



- Diskrete Zufallsvariable:

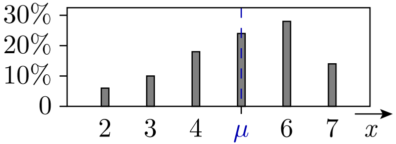
$$\mathbb{E}\left[X\right] = \mu = \sum_{i=1}^{\#x} p_i \cdot x_i \tag{4.4}$$

- Stetige Zufallsvariable:

$$\mathbb{E}\left[X\right] = \mu = \int_{x_{\min}}^{x_{\max}} f_X\left(x\right) \cdot x \cdot \mathrm{d}x \tag{4.5}$$

X	Zufallsvariable.
x_i	Realisierung (möglichen Wert) i der Zuvallsvariablen X.
$\#x$	Anzahl der möglichen Realisierungen der Zufallsvariablen X.
p_i	Eintrittswahrscheinlichkeit der Realisierung x_i.

Erwartungswert am Beispiel



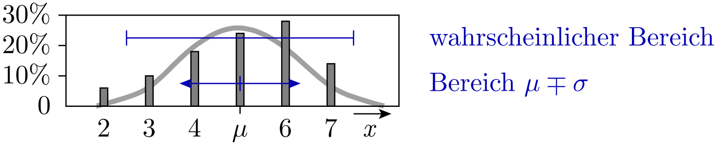
x_i	2	3	4	5	6	7
$\mathbb{P}[X = x_i] = p_i$	6%	10%	18%	24%	28%	14%
$F_X(x) = \mathbb{P}[X \leq x_i]$	6%	16%	34%	58%	86%	100%

(4.4) $$\mathbb{E}[X] = \mu = \sum_{i=1}^{\#x} p_i \cdot x_i$$

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

X Zufallsvariable.
x_i Realisierung (möglichen Wert) i der Zuvallsvariablen X.
$\#x$ Anzahl der möglichen Realisierungen der Zufallsvariablen X.
p_i Eintrittswahrscheinlichkeit der Realisierung x_i.
$\mathbb{E}[\ldots]$, μ Erwartungswert.

Varianz, Standardabweichung



Die Standardabweichung ist ein Maß für die Breite des wahrscheinlichen Bereichs. Die Varianz ist mittlere quadratische Abweichung vom Erwartungswert und das Quadrat der Varianz:

$$\mathrm{Var}\left[X\right] = \sigma^2 = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] \tag{4.6}$$

$$\mathrm{Var}\left[X\right] = \sigma^2 = \sum_{i=1}^{\#x} p_i \cdot \left(x_i - \mathbb{E}\left[X\right]\right)^2 \tag{4.7}$$

$$\mathrm{Var}\left[X\right] = \sigma^2 = \int_{x_{\min}}^{x_{\max}} f_X\left(x\right) \cdot \left(x - \mathbb{E}\left[X\right]\right)^2 \cdot \mathrm{d}x \tag{4.8}$$

$$\mathrm{sd}\left[X\right] = \sigma = \sqrt{\mathrm{Var}\left[X\right]} \tag{4.9}$$

$\mathrm{Var}\left[\ldots\right]$, σ^2	Varianz.
$\mathrm{sd}\left[\ldots\right]$, σ	Standardabweichung.

Steinerscher Verschiebungssatz

Die Varianz ist gleichfalls die Differenz aus dem Erwartungswert der Quadrate und dem Quadrat des Erwartungswertes:

$$\operatorname{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \tag{4.10}$$

Für diskrete Zufallsvariablen:

$$\operatorname{Var}[X] = \sum_{i=1}^{\#x} p_i \cdot (x_i)^2 - \mathbb{E}[X]^2 \tag{4.11}$$

Kontrolle durch Nachrechnen:

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

(Gl. 4.11) oft handlicher, aber bei kleinen Differenzen großer Werte numerische Probleme.

$\#x$	Anzahl der möglichen Realisierungen der Zufallsvariablen X.
x_i	Realisierung (möglichen Wert) i der Zuvallsvariablen X.
p_i	Eintrittswahrscheinlichkeit der Realisierung x_i.

Varianz am Beispiel

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

- Erwartungswert: $\mathbb{E}[X] = \mu = 5$ (siehe Gl. 4.4).
- Varianz nach

(4.7) $$\mathrm{Var}[X] = \sigma^2 = \sum_{i=1}^{\#x} p_i \cdot (x_i - \mathbb{E}[X])^2$$

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

- Varianz nach Verschiebesatz

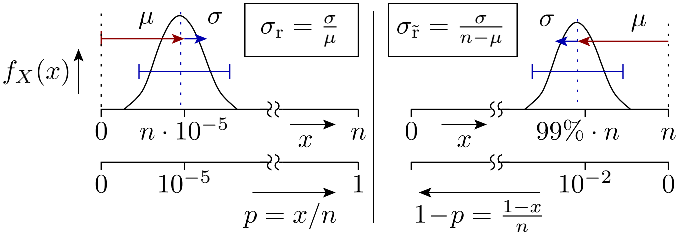
(4.11) $$\mathrm{Var}[X] = \sum_{i=1}^{\#x} p_i \cdot (x_i)^2 - \mathbb{E}[X]^2$$

$$\sigma^2 = 6\% \cdot 2^2 + 10\% \cdot 3^2 + 18\% \cdot 4^2 + 24\% \cdot 5^2 + 28\% \cdot 6^2 + 14\% \cdot 7^2 - 5^2 = 1{,}96\checkmark$$

- Standardabweichung:

$$\sigma = \sqrt{1{,}96} = 1{,}4$$

Varianzkoeffizient



Varianzkoeffizient für das Eintreten bzw. Nicht-Eintreten

$$\begin{aligned} \sigma_{\mathrm{r}} &= \frac{\sigma}{\mu} && \text{zweckmäßiges Mass für } \mu \le n/2 \\ \sigma_{\tilde{\mathrm{r}}} &= \frac{\sigma}{n-\mu} && \text{zweckmäßiges Mass für } \mu > n/2 \end{aligned} \tag{4.12}$$

Maße für relative Breite wahrscheinlicher Bereich im Verhältnis zum Abstand vom Minimum bzw. Maximum eines Zählwerts x und auch dessen Eintritts- bzw. Nicht-Eintrittswahrscheinlichkeit (Schätzgenauigkeit).

p	Eintrittswahrscheinlichkeit, Anzahl der Zählversuche.
σ_{r}, $\sigma_{\tilde{\mathrm{r}}}$	Varianzkoeffizient für das Eintreten bzw. Nichteintreten des Zählwerts.
μ, σ	Erwartungswert, Standardabweichung.

Erwartungswert einer Datenstichprobe

Wahrscheinlichkeitsmodelle dienen zur Vorhersage künftig zu erwartender Häufigkeiten von Datenmerkmalen. Umgekehrt dienen erhobene Daten zur Kontrolle der Wahrscheinlichkeitsmodelle. Wir werden im weiteren zwischen Modell und Daten vergleichen:

- Erwartungswert, Varianz und
- Verteilungsparameter.

Für eine Datenstichprobe einer Zufallsvariable X

$$v = (v_1, v_2, \ldots, v_{\#v})$$

ist der im weiteren verwendete Schätzer für den Erwartungswert:

$$\hat{\mathbb{E}}[X] = \hat{\mu} = \frac{1}{\#v} \cdot \sum_{i=1}^{\#v} v_i \tag{4.13}$$

$\#v$	Größe der Datenstichprobe.
v_i	Wert i der Datenstichprobe.
$.\hat{.}.$	Schätzwert.
$\mathbb{E}[\ldots]$, μ	Erwartungswert.

Varianz, Standardabweichung Datenstichprobe

Schätzer der Varianz ist die mittlere quadratische Abweichung vom geschätzten Erwartungswert:

$$\hat{\mathrm{Var}}[X] = \hat{\sigma}^2 = \frac{1}{\#v - 1} \cdot \sum_{i=1}^{\#v} (v_i - \hat{\mu})^2 \tag{4.14}$$

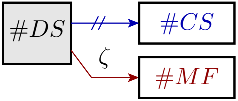
$$\hat{\mathrm{sd}}[X] = \hat{\sigma} = \sqrt{\hat{\mathrm{Var}}[X]} \tag{4.15}$$

Der Divisor ist um eins kleiner als die Anzahl der Datenwerte $\#v$, d.h Abschätzungen von Varianz und Standardabweichung erfordern mindestens eine Datenstichprobe $\#v \geq 2$.

$\#v$	Größe der Datenstichprobe.
v_i	Wert i der Datenstichprobe.
$\hat{\ldots}$	Schätzwert.
$\mathrm{Var}[\ldots]$, σ^2	Varianz.

Summen, lineare Transformationen

Summen von Zufallsvariablen



Zufallsgröße X:	$\#CS$
Anzahl Zählversuche:	$\#DS$
Summanden X_i:	jeder DS
Eintrittswahrsch. p_i:	$1-\zeta$

Komplexe Zufallsereignisse lassen sich oft als Summe linearer Transformationen einfacher zu untersuchender Ereignisse beschreiben.

Ein zufälliger Zählwert X, z.B. die Anzahl der korrekt ausgeführten Service-Leistungen, ist eine Summe elementarer Zählereignisse X_i:

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

Die Summanden haben die Bernoulli-Verteilungen (vergl. Gl. 4.1):

$$\begin{array}{lll} \mathbb{P}\left[X_i = 0\right] & = 1 - p_i & \text{Zählereignis nicht eingetreten} \\ \mathbb{P}\left[X_i = 1\right] & = p_i & \text{Zählereignis eingetreten} \end{array} \quad (4.16)$$

Lineare Transformation

Lineare Transformationen sind Multiplikationen und Additionen einer Zufallsvariablen mit reellen Zahlen:

$$Y = a \cdot X + b$$

Die (diskrete) Verteilung ordnet die Wahrscheinlichkeiten der Originalwerte den transformierten Werten zu:

$$\mathbb{P}\left[y = ax + b\right] = \mathbb{P}\left[x\right] \tag{4.17}$$

Der Erwartungswert wird genau wie jeder andere Wert transformiert:

$$\mathbb{E}\left[a \cdot X + b\right] = a \cdot \mathbb{E}\left[X\right] + b \tag{4.18}$$

Varianz und Standardabweichung sind verschiebungs- und spiegelungsinvariant, d.h. um Quadrat bzw. Betrag der Skalierung größer:

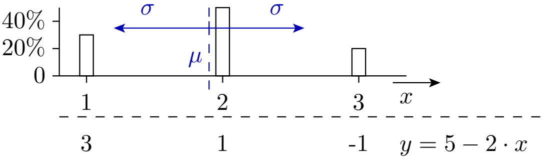
$$\mathrm{Var}\left[a \cdot X + b\right] = a^2 \cdot \mathrm{Var}\left[X\right] \tag{4.19}$$

$$\mathrm{sd}\left[a \cdot X + b\right] = \sqrt{\mathrm{Var}\left[a \cdot X + b\right]} = a \cdot \mathrm{sd}\left[X\right] \tag{4.20}$$

Kontrollen durch Nachrechnen folgen später als Übungsaufgaben.

$a,\ b$ Beliebige reele Zahlen.

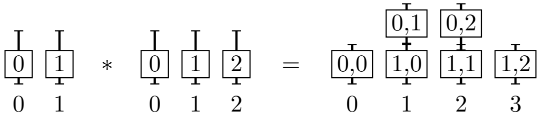
Veranschaulichung am Beispiel



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

$$
\begin{aligned}
\mathbb{E}[X] &= 0{,}3 \cdot 1 + 0{,}5 \cdot 2 + 0{,}3 \cdot 3 = 1{,}9 \\
\operatorname{Var}[X] &= 0{,}3 \cdot 1^2 + 0{,}5 \cdot 2^2 + 0{,}3 \cdot 3^2 - 1{,}9^2 = 0{,}49 \\
\mathbb{E}[Y] &= 0{,}3 \cdot 3 + 0{,}5 \cdot 1 + 0{,}2 \cdot (-1) = 1{,}2 \\
\operatorname{Var}[Y] &= 0{,}3 \cdot 3^2 + 0{,}5 \cdot 1^2 + 0{,}2 \cdot (-1)^2 - 1{,}2^2 = 1{,}96 \\
\mathbb{E}[Y] &= 5 - 2 \cdot \mathbb{E}[X] = 5 - 2 \cdot 1{,}9 = 1{,}2\surd \\
\operatorname{Var}[Y] &= (-2)^2 \cdot \operatorname{Var}[X] = 4 \cdot 0{,}49 = 4 \cdot 0{,}5 - 4 \cdot 0{,}1 = 1{,}96\surd
\end{aligned}
$$

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):

v	0	1	μ
$\mathbb{P}[X=v]$	0,4	0,6	0,6

v	0	1	2	μ
$\mathbb{P}[Y=v]$	0,3	0,6	0,1	0,8

$$\begin{aligned}
\mathbb{P}[X+Y=0] &= \mathbb{P}[X=0]\cdot\mathbb{P}[Y=0] = 0{,}4\cdot 0{,}3 = 0{,}12\\
\mathbb{P}[X+Y=1] &= \mathbb{P}[X=0]\cdot\mathbb{P}[Y=1] + \mathbb{P}[X=1]\cdot\mathbb{P}[Y=0] = 0.42\\
\mathbb{P}[X+Y=2] &= \mathbb{P}[X=0]\cdot\mathbb{P}[Y=2] + \mathbb{P}[X=1]\cdot\mathbb{P}[Y=1] = 0{,}4\\
\mathbb{P}[X+Y=3] &= \mathbb{P}[X=1]\cdot\mathbb{P}[Y=2] = 0{,}6\cdot 0{,}1 = 0{,}06
\end{aligned}$$

Erwartungswert und Varianz

Für Summen unabhängiger Zufallsvariablen sind Erwartungswert und Varianz gleich der Summe der Erwartungswerte bzw. Varianzen der Summanden:

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \tag{4.21}$$

$$\operatorname{Var}[X+Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] \tag{4.22}$$

v	0	1	2	3	μ
$\mathbb{P}[X=v]$	40%	60%			$40\% \cdot 0 + 60\% \cdot 1 = 0{,}6$
$\mathbb{P}[Y=v]$	30%	60%	10%		$60\% \cdot 1 + 10\% \cdot 2 = 0{,}8$
$\mathbb{P}[Y+Y=v]$	12%	42%	40%	6%	$42\% \cdot 1 + 40\% \cdot 2 + 6\% \cdot 3 = 1{,}4$

$$\begin{aligned}
\operatorname{Var}[X] &= 40\% \cdot 0 + 60\% \cdot 1^2 - 0{,}6^2 &&= 0{,}24\\
\operatorname{Var}[Y] &= 60\% \cdot 1 + 10\% \cdot 2^2 - 0{,}8^2 &&= 0{,}36\\
\operatorname{Var}[X+Y] &= 42\% \cdot 1 + 40\% \cdot 2^2 + 6\% \cdot 3 - 1{,}4^2 &&= 0{,}60
\end{aligned}$$

X, Y Zufallsvariablen.
$\mathbb{E}[\ldots]$ Erwartungswert von ...
$\operatorname{Var}[\ldots]$ Varianz von ...

Abhängigkeiten

Abhängigkeit bedeutet, dass die Eintrittswahrscheinlichkeit eines Summanden X vom zufälligen Ergebnis eines anderen Summanden Y abhängt. Für $X, Y \in \{0, 1\}$ als Zufallsvariablen für den Nachweis von zwei Haftfehlern Fehlern wurden (siehe Abschn. 2.1.5) z.B. folgende mögliche Abhängigkeiten behandelt:

- impliziter Fehlernachweis: wenn $Y = 1$ dann $X = 1$ und
- identischer Fehlernachweis $Y = X$.

Abhängigkeiten haben keinen Einfluss auf den Erwartungswert und erhöhen die Varianz um die doppelte Kovarianz*:

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y] \tag{4.23}$$

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])] \tag{4.24}$$

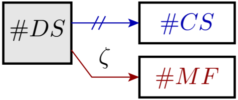
Wir werden Abhängigkeiten zwischen Zählwerten statt dessen durch eine aus Daten abschätzbare Varianzvergrößerung κ beschreiben (siehe später Abschn. 4.2.1).

$\mathrm{Cov}[X,Y]$ Kovarianz der beiden Zuvallsvariablen.

* Kontrolle durch Nachrechnen folgt später als Übungsaufgabe.

Verteilung von Zählwerten

Verteilung von Zählwerten



Zufallsgröße X:	$\#CS$
Anzahl Zählversuche:	$\#DS$
Summanden X_i:	jeder DS
Eintrittswahrsch. p_i:	$1-\zeta$

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

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

»potentieller Zählwerte« X_i mit den Bernoulli-Verteilungen

(4.16)
$$\begin{array}{lll} \mathbb{P}\left[X_i = 0\right] & = 1 - p_i & \text{ZÃ¤hlereignis nicht eingetreten} \\ \mathbb{P}\left[X_i = 1\right] & = p_i & \text{ZÃ¤hlereignis eingetreten} \end{array}$$

beschreiben.

---

p_i Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.
n Anzahl der Zählversuche, maximaler Zählwert.

Erwartungswert und Varianz der Summanden

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

Erwartungswerte der Einzelereignisse:

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

Varianzen nach dem Verschiebungssatz:

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

X_i	Potentielle Zählwerte, Zufallsvariablen mit Wertebereich $X_i \in \{0, 1\}$.
p_i	Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.

Der Erwartungswert der Summe ist die Summe der Erwartungswerte (Gl. 4.18):

$$\mathbb{E}\left[X\right] = \sum_{i=1}^{n} p_i \tag{4.25}$$

Für die Varianz wird oft unterstellt, das die zu zählenden Ereignisse, wie das Auftreten unterschiedlicher Fehlfunktionen, nicht voneinander abhängen (Varianz der Summe gleich der Summe der Varianzen der Summanden, Kovarianz null, Gl. 4.21):

$$\mathrm{Var}\left[X\right] = \sum_{i=1}^{n} p_i \cdot (1 - p_i) \tag{4.26}$$

Abhängigkeiten werden später durch die Varianzerhöhung κ modelliert (siehe Abschn. 4.2.1).

p_i	Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.
n	Anzahl der Zählversuche, maximaler Zählwert.

Berechnung der Verteilung

- Beginne mit der Verteilung von $S_1 = X_1$:

$$\mathbb{P}[S_1 = 0] = 1 - p_1$$
$$\mathbb{P}[S_1 = 1] = p_1$$

  - Wiederhole für $i = 1$ bis $n - 1$
    Berechne Verteilung von $S_{i+1} = X_{i+1} + S_i$:

$$\begin{array}{rlllll} (1-p_{i+1}) \cdot (& \mathbb{P}[S_i = 0], & \mathbb{P}[S_i = 1], & \ldots, \mathbb{P}[X_i = i-1], & 0 &) \\ + p_{i+1} \cdot (& 0, & \mathbb{P}[S_i = 0], & \ldots, \mathbb{P}[X_i = i-2], & \mathbb{P}[X_i = i-1] &) \end{array}$$

Beispiel:

	p_i	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$\mathbf{P}[S_1 = X_1 = k]$	30%	70%	30%			
$\mathbf{P}[S_2 = X_1 + X_2 = k]$	50%	35%	50%	15%		
$\mathbf{P}[S_3 = X_1 + X_2 + X_3 = k]$	40%	21%	44%	29%	6%	
$\mathbf{P}[S_4 = X_1 + X_2 + X_3 + X_4 = k]$	10%	18,9%	41,7%	30,5%	8,3%	0,6%

p_i Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.
$\mathbb{P}_i[S_i = k]$ Verteilung der Summe der ersten i Zählerwerte.

Erwartungswert und Varianz für das Beispiel

i	0	1	2	3	4
p_i	30%	50%	40%	10%	
$\mathbf{P}[S_4 = i]$	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. 4.25 ist die Berechung kürzer:

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

Die Varianz beträgt nach dem Verschiebungssatz Gl. 4.10:

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

Vereinfachte Berechnung nach Gl. 4.26:

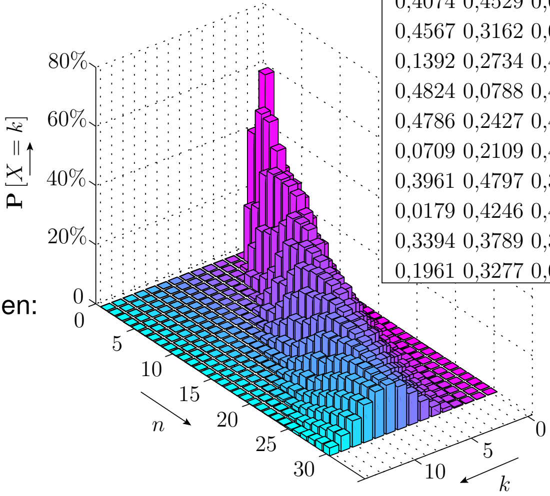
$$\mathrm{Var}[X] = 0{,}3 \cdot 0{,}7 + 0{,}5 \cdot 0{,}5 + 0{,}4 \cdot 0{,}6 + 0{,}1 \cdot 0{,}9 = 0{,}79$$

Beispiel einer Zählverteilung

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

$$\mathbb{E}[X] = 7{,}05$$
$$\sqrt{\mathrm{Var}[X]} = 2{,}19$$

Wahrscheinlicher Bereich ca. 5 bis 15.



p_i für i=1 bis 30		
0,4074	0,4529	0,0635
0,4567	0,3162	0,0488
0,1392	0,2734	0,4788
0,4824	0,0788	0,4853
0,4786	0,2427	0,4001
0,0709	0,2109	0,4579
0,3961	0,4797	0,3279
0,0179	0,4246	0,4670
0,3394	0,3789	0,3716
0,1961	0,3277	0,0856

n Anzahl der Zählversuche, maximaler Zählwert.

Messfehler

Gemessener Wert und Messfehler

Die Summenbeziehungen für Zufallsgrößen sind auch für andere Aufgabenstellung aus dem Bereich Test und Verlässlichkeit nützlich.

In der Messtechnik gilt für jeden gemessenen Wert:

$$X_{\mathrm{M}} = X + X_{\mathrm{F}} \tag{4.27}$$

Alle drei Größen haben einen Erwartungswert und eine Varianz. Mit Messwert und Messfehler als unabhängige Zufallsvariablen gilt:

$$\mathbb{E}\left[X_{\mathrm{M}}\right] = \mathbb{E}\left[X\right] + \mathbb{E}\left[X_{\mathrm{F}}\right] \tag{4.28}$$

$$\mathrm{Var}\left[X_{\mathrm{M}}\right] = \mathrm{Var}\left[X\right] + \mathrm{Var}\left[X_{\mathrm{F}}\right] \tag{4.29}$$

- $\mathbb{E}\left[X_{\mathrm{F}}\right]$ – Erwartungswert, systematischer Messfehler
- $\mathrm{sd}\left[X_{\mathrm{F}}\right] = \sqrt{\mathrm{Var}\left[X_{\mathrm{F}}\right]}$ – Standardabweichung, zufälliger Messfehler.

X_{M}	Zufallsvariable gemessener Wert.
X	Zufallsvariable zu messender Wert.
X_{F}	Zufallsvariable Messfehler.
$\mathbb{E}\left[\ldots\right]$	Erwartungswert von ...
$\mathrm{Var}\left[\ldots\right]$	Varianz von ...

Beispiel 4.1: Messfehler

Der gemessene Wert R_{M} einer Widerstands-Charge ist im Mittel $1010\,\Omega$ und hat eine Standardabweichung von $11{,}18\,\Omega$. Die Messung habe einen systematischen Fehler R_{F} von $12\,\Omega$ und eine Standardabweichung von $5\,\Omega$.

$\mathbb{E}\left[R_{\mathrm{M}}\right] = 1010\,\Omega$, $\mathrm{sd}\left[R_{\mathrm{M}}\right] = 11{,}18\,\Omega$, $\mathbb{E}\left[R_{\mathrm{F}}\right] = 12\,\Omega$ und $\mathrm{sd}\left[R_{\mathrm{F}}\right] = 5\,\Omega$.

Welchen Erwartungswert und welche Standardabweichung hat der zu messende Wert?

R_{M}	Gemessener Widerstandswert.
R_{F}	Messfehler des Widerstandswertes.
$\mathbb{E}\left[\ldots\right]$	Erwartungswert von ...
$\mathrm{Var}\left[\ldots\right]$	Varianz von ...
$\mathrm{sd}\left[\ldots\right]$	Standardabweichung von ...

$\mathbb{E}[R_\mathrm{M}] = 1010\,\Omega$, $\mathrm{sd}[R_\mathrm{M}] = 11{,}18\,\Omega$, $\mathbb{E}[R_\mathrm{F}] = 12\,\Omega$ und $\mathrm{sd}[R_\mathrm{F}] = 5\,\Omega$.

Welchen Erwartungswert und welche Standardabweichung hat der zu messende Wert?

(4.28) $$\mathbb{E}[X_\mathrm{M}] = \mathbb{E}[X] + \mathbb{E}[X_\mathrm{F}]$$

(4.29) $$\mathrm{Var}[X_\mathrm{M}] = \mathrm{Var}[X] + \mathrm{Var}[X_\mathrm{F}]$$

(4.9) $$\mathrm{sd}[X] = \sigma = \sqrt{\mathrm{Var}[X]}$$

Die Zufallsvariable in der Aufgabe ist R statt X und gesucht sind Erwartungswert und Standardabweichung des tatsächlichen Wertes:

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

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

R Zu messender Widerstandswert.

Zusammenfassung

Zufallsvariable und Verteilung

Die Verteilungsfunktion $F_X(x)$ gibt die Wahrscheinlichkeit an, dass eine Zufallsvariable X maximal den Wert x hat:

(4.2) $$F_X(x) = \mathbb{P}[X \le x]$$

Eine diskrete Zufallsgröße kann nur Werte x_i annehmen. Die Verteilung $\mathbb{P}[X = x_i]$ ordnet jedem der möglichen Werte eine Wahrscheinlichkeit p_i zu.

Eine stetige Verteilung hat im Intervall $x_{\min} \le X \le x_{\max}$ unendlich viele Ausprägungen und statt der Verteilung eine Dichtefunktion:

(4.3) $$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$$

Für die Modellierung von Zählwerten benötigen wir insbesondere die Bernoulli-Verteilung zur Beschreibung eines einzelnen Zählversuchs mit den möglichen Ergebnissen $x_i \in \{0, 1\}$ und der Verteilung:

(4.1) $$\begin{aligned} \mathbb{P}[X = 0] &= 1 - p \\ \mathbb{P}[X = 1] &= p \end{aligned}$$

Charakteristische Größen von Zufallsvariablen

Wir beschränken uns auf Erwartungswert:

(4.4) $$\mathbb{E}\left[X\right] = \mu = \sum_{i=1}^{\#x} p_i \cdot x_i$$

(4.5) $$\mathbb{E}\left[X\right] = \mu = \int_{x_{\min}}^{x_{\max}} f_X\left(x\right) \cdot x \cdot \mathrm{d}x$$

Varianz zur Abschätzung der Standardabweichung:

(4.6) $$\mathrm{Var}\left[X\right] = \sigma^2 = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right]$$

(4.7) $$\mathrm{Var}\left[X\right] = \sigma^2 = \sum_{i=1}^{\#x} p_i \cdot \left(x_i - \mathbb{E}\left[X\right]\right)^2$$

(4.8) $$\mathrm{Var}\left[X\right] = \sigma^2 = \int_{x_{\min}}^{x_{\max}} f_X\left(x\right) \cdot \left(x - \mathbb{E}\left[X\right]\right)^2 \cdot \mathrm{d}x$$

Standardabweichung als Mass für die Breite des wahrscheinlichen Bereichs:

(4.9) $$\mathrm{sd}\left[X\right] = \sigma = \sqrt{\mathrm{Var}\left[X\right]}$$

Berechnung der Varianz nach dem Verschiebesatz:

(4.10) $$\mathrm{Var}\left[X\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2$$

(4.11) $$\mathrm{Var}\left[X\right] = \sum_{i=1}^{\#x} p_i \cdot (x_i)^2 - \mathbb{E}\left[X\right]^2$$

Varianzkoeffizient für das Eintreten bzw. Nicht-Eintreten als Mass für die relative Bereichsbreite der Anzahl bzw. der Wahrscheinlichkeit für das Eintreten bzw. Nicht-Eintreten:

(4.12) $$\begin{aligned} \sigma_{\mathrm{r}} &= \tfrac{\sigma}{\mu} && \text{zweckmäßiges Mass für } \mu \leq n/2 \\ \sigma_{\tilde{\mathrm{r}}} &= \tfrac{\sigma}{n-\mu} && \text{zweckmäßiges Mass für } \mu > n/2 \end{aligned}$$

Datenstichprobe

Wahrscheinlichkeitsmodelle dienen zur Vorhersage künftig zu erwartender Häufigkeiten von Datenmerkmalen. Umgekehrt dienen erhobene Daten zur Kontrolle der Wahrscheinlichkeitsmodelle.

Schätzer für den Erwartungswert:

$$\hat{\mathbb{E}}[X] = \hat{\mu} = \frac{1}{\#v} \cdot \sum_{i=1}^{\#v} v_i \tag{4.13}$$

Schätzer für die Varianz:

$$\hat{\mathrm{Var}}[X] = \hat{\sigma}^2 = \frac{1}{\#v-1} \cdot \sum_{i=1}^{\#v} (v_i - \hat{\mu})^2 \tag{4.14}$$

Standardabweichung:

$$\hat{\mathrm{sd}}[X] = \hat{\sigma} = \sqrt{\hat{\mathrm{Var}}[X]} \tag{4.15}$$

Lineare Transformationen, Summe

- Verteilung, Erwartungswert und Varianz von $Y = a \cdot X + b$:

(4.17) $$\mathbb{P}[y = ax + b] = \mathbb{P}[x]$$

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

(4.19) $$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X]$$

(4.20) $$\mathrm{sd}[a \cdot X + b] = \sqrt{\mathrm{Var}[a \cdot X + b]} = a \cdot \mathrm{sd}[X]$$

- Erwartungswert und Varianz der Summe unabhängiger Zufallsvariablen:

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

(4.22) $$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

- Abhängigkeiten vergrößern die Varianz um die doppelte Kovarianz. (Wir brücksichtigen Abhängigkeiten später anders):

(4.23) $$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X, Y]$$

(4.24) $$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])]$$

Verteilung von Zählwerten

Erwartungswert und Varianz:

(4.25) $$\mathbb{E}[X] = \sum_{i=1}^{n} p_i$$

(4.26) $$\mathrm{Var}[X] = \sum_{i=1}^{n} p_i \cdot (1 - p_i)$$

Berechnung der Verteilung:

	p_i	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
$\mathbf{P}[S_1 = X_1 = k]$	30%	70%	30%			
$\mathbf{P}[S_2 = X_1 + X_2 = k]$	50%	35%	50%	15%		
$\mathbf{P}[S_3 = X_1 + X_2 + X_3 = k]$	40%	21%	44%	29%	6%	
$\mathbf{P}[S_4 = X_1 + X_2 + X_3 + X_4 = k]$	10%	18,9%	41,7%	30,5%	8,3%	0,6%

Messwerte und Messfehler

(4.27) $$X_{\mathrm{M}} = X + X_{\mathrm{F}}$$

(4.28) $$\mathbb{E}[X_{\mathrm{M}}] = \mathbb{E}[X] + \mathbb{E}[X_{\mathrm{F}}]$$

(4.29) $$\mathrm{Var}[X_{\mathrm{M}}] = \mathrm{Var}[X] + \mathrm{Var}[X_{\mathrm{F}}]$$

Näherungen

Binomialverteilung

Binomialverteilung

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

●	●	●	○	○	○
●	○	○	●	●	○
○	●	○	●	○	●
○	○	●	○	●	●

Für den Sonderfall, dass gleichwahrscheinliche Ereignisse gezählt werden, ist die Summe der gezählten Ereignisse binomialverteilt

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

Binomialverteilung:

$$\mathbb{P}\,[X = k] = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \tag{4.30}$$

Erwartungswert einer Binomialverteilung:

$$\mathbb{E}\,[X] = \mu = n \cdot p \tag{4.31}$$

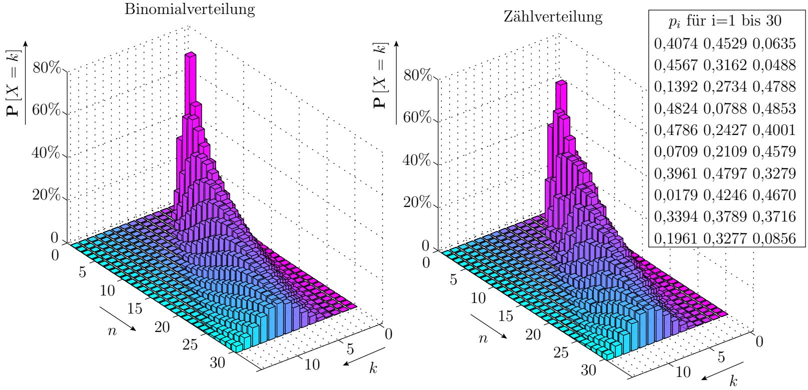
Varianz und Standardabweichung einer Binomialverteilung:

$$\mathrm{Var}\,[X] = \sigma^2 = n \cdot p \cdot (1 - p) \tag{4.32}$$

$$\mathrm{sd}\,[X] = \sigma = \sqrt{n \cdot p \cdot (1 - p)} \tag{4.33}$$

n Anzahl der Zählversuche, maximaler Zählwert.
p Eintrittswahrscheinlichkeit.

Binomialverteilung vs. allgemeine Zählverteilung



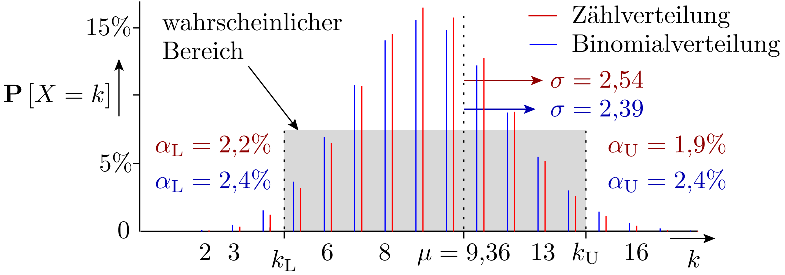
p_i für i=1 bis 30		
0,4074	0,4529	0,0635
0,4567	0,3162	0,0488
0,1392	0,2734	0,4788
0,4824	0,0788	0,4853
0,4786	0,2427	0,4001
0,0709	0,2109	0,4579
0,3961	0,4797	0,3279
0,0179	0,4246	0,4670
0,3394	0,3789	0,3716
0,1961	0,3277	0,0856

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

---

n Anzahl der Zählversuche, maximaler Zählwert.
p Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.
p_i Eintrittswahrscheinlichkeit Zählereignis i.

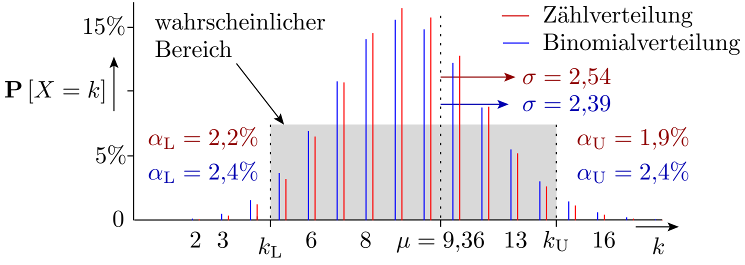
Bereichsschätzung für einen Zählwertbereich



- wahrscheinlicher Bereich $[k_{\mathrm{L}}, k_{\mathrm{U}}]$.
- Berechnung der Irrtumswahrscheinlichkeiten:

$$\mathbb{P}[X < k_{\mathrm{L}}] = \alpha_{\mathrm{L}} = \sum_{k=0}^{k_{\mathrm{L}}-1} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$\mathbb{P}[X > k_{\mathrm{U}}] = \alpha_{\mathrm{U}} = \sum_{k=k_{\mathrm{U}}+1}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$



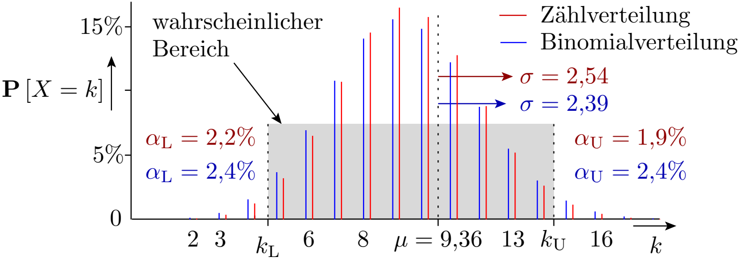
Für das Beispiel ($k_{\mathrm{L}} = 5$, $k_{\mathrm{U}} = 14$, $n = 30$, $p = 31{,}2\%$):

	$\mathbb{P}[X < 5]$	$\mathbb{P}[X < 14]$
Zählverteilung	2,2%	1,9%
Binomialverteilung	2,4%	2,4%

Bei gleicher Anzahl unabhängiger Zählversuche n liefert die Annäherung durch eine Binomialverteilung mit $p = \frac{1}{n} \cdot \sum_{i=1}^{n} p_i$ eine Worst-Case-Abschätzung:

- garantiert eingehaltene Irrtumswahrscheinlichkeiten bzw.
- einen mindestens garantierbaren Bereich.

Varianzobergrenze für unabhängige Zählwerte



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

$$\sigma^2 = \sum_{i=1}^{n} p_i \cdot (1 - p_i) \leq n \cdot p \cdot (1 - p) \tag{4.34}$$

Beweis durch Nachrechnen

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:

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^{n} (p+\delta_i)\cdot(1-p-\delta_i) \\
&= \sum_{i=1}^{n} \left(p - p^2 - p\cdot\delta_i + \delta_i - p\cdot\delta_i - \delta_i^2\right) \\
& \underbrace{\sum_{i=1}^{n} \left(p-p^2\right)}_{n\cdot p\cdot(1-p)} + \underbrace{\sum_{i=1}^{n} (\delta_i - 2\cdot p\cdot\delta_i)}_{(1-2p)\cdot\sum_{i=1}^{n}\delta_i=0} - \underbrace{\sum_{i=1}^{n} \delta_i^2}_{\geq 0} \\
\sigma^2 &\leq n\cdot p\cdot(1-p) \ \text{(Varianz Binomialverteilung)} \checkmark
\end{aligned}$$

Schätzer für μ, p und σ für Zählwerten

Die Ungleichheit in

(4.34) $\sigma^2 = \sum_{i=1}^{n} p_i \cdot (1 - p_i) \leq n \cdot p \cdot (1 - p)$

wird in einen Korrekturfaktor $\kappa \leq 1$ versteckt:

$$\sigma = \sqrt{\kappa \cdot n \cdot p \cdot (1 - p)} \tag{4.35}$$

Schätzer für Erwartungswert, Eintrittswahrscheinlichkeit und Standardabweichung für einen experimentell bestimmten Zählwert x_{AV}:

$$\hat{\mu} = x_{\mathrm{AV}} \tag{4.36}$$

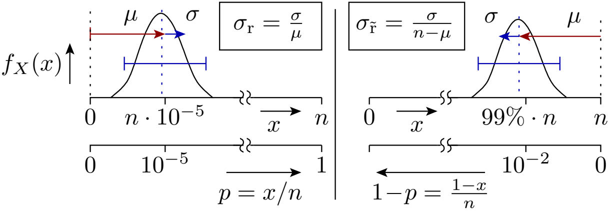
$$\hat{p} = \frac{\hat{\mu}}{n} = \frac{x_{\mathrm{AV}}}{n} \tag{4.37}$$

$$\hat{\sigma} = \sqrt{\kappa \cdot x_{\mathrm{AV}} \cdot \left(1 - \frac{x_{\mathrm{AV}}}{n}\right)} \tag{4.38}$$

Später wird gezeigt, dass größere Zählwerte oft normalverteilt und die Intervallradien 2 bis 3-mal so breit wie die Standardabweichung sind.

$\hat{\mu}$, $\hat{\sigma}$	Schätzwerte für Erwartungswert und Standardabweichung.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.
n, x_{AV}	Anzahl der Zählversuche, Experimentell bestimmter Ist-Zählwert.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \leq 1$.

Varianzkoeffizient als Genauigkeitsmaß



Der Varianzkoeffizient für das Eintreten bzw. Nicht-Eintreten ist ein Maße für die relative Breite des wahrscheinlichen Bereichs im Verhältnis zum möglichen Minimum bzw. Maximum und damit ein Maß für die relative Schätzgenauigkeit für:

- eintretende bzw. nicht eintretende Zählwerte und
- Eintritts- bzw. Nichteintrittswahrscheinlichkeiten.

(4.12)
$$\sigma_{\mathrm{r}} = \frac{\sigma}{\mu} \quad \text{zweckmäßiges Mass für } \mu \le n/2$$
$$\sigma_{\tilde{\mathrm{r}}} = \frac{\sigma}{n-\mu} \quad \text{zweckmäßiges Mass für } \mu > n/2$$

σ_{r}, $\sigma_{\tilde{\mathrm{r}}}$ Relative Standardabweichung zum erwarteten Eintritts- bzw. Nichteintrittszählwert.

Mit dem Schätzer für die Standardabweichung

$$\hat{\sigma} = \sqrt{\kappa \cdot x_{\mathrm{AV}} \cdot \left(1 - \frac{x_{\mathrm{AV}}}{n}\right)} \tag{4.38}$$

ergeben sich als Schätzer für die Varianzkoeffizienten für die später behandelten Bereichsschätzungen unter Annäherung der Zählwertverteilung durch eine Normalverteilung:

$$\begin{aligned} \hat{\sigma}_{\mathrm{r}} &= \frac{\hat{\sigma}}{x_{\mathrm{AV}}} = \sqrt{\kappa \cdot \left(\frac{1}{x_{\mathrm{AV}}} - \frac{1}{n}\right)} && \text{für } \hat{p} \le 50\% \\ \hat{\sigma}_{\tilde{\mathrm{r}}} &= \frac{\hat{\sigma}}{n - x_{\mathrm{AV}}} = \sqrt{\kappa \cdot \left(\frac{1}{n - x_{\mathrm{AV}}} - \frac{1}{n}\right)} && \text{für } \hat{p} > 50\% \end{aligned} \tag{4.39}$$

σ_{r}, $\sigma_{\tilde{\mathrm{r}}}$	Relative Standardabweichung zum erwarteten Eintritts- bzw. Nichteintrittszählwert.
$\hat{\sigma}$	Geschätzte Standardabweichung des Zählwerts.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
n	Anzahl der Zählversuche, maximaler Zählwert.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.

Poisson-Verteilung

Poisson-Verteilung

Beim Zählen vieler seltener Ereignisse, z.B. der Fehlfunktionen bei Millionen von Service-Anforderungen, von denen nur wenige eintreten

$$n \to \infty$$

$$p_i \to 0$$

$$\mathrm{Var}\left[X_i\right] = p_i \cdot (1 - p_i) \to p_i$$

strebt die Varianz der zu zählenden Ereignisse und deren Summe gegen den Erwartungswert:

$$\mathrm{Var}\left[X\right] = \mathbb{E}\left[X\right] = \sum_{i=1}^{n} p_i = \lambda$$

Die Verteilung der Summe strebt gegen die Poisson-Verteilung:

$$X \sim \mathrm{Pois}\left(\lambda\right)$$

---

n	Anzahl der Zählversuche, maximaler Zählwert.
p_i	Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.
λ	Parameter der Poisson-Verteilung (Erwartungswert und gleichzeitig Varianz).

Poisson-Verteilung:

$$\mathbb{P}[X = k] = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!} \tag{4.40}$$

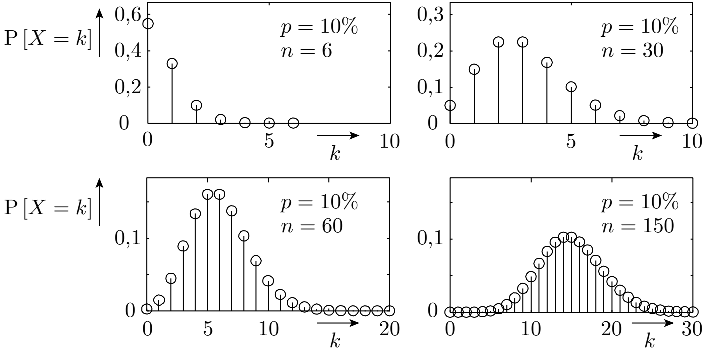
$$= \mathrm{e}^{-p \cdot n} \cdot \frac{(p \cdot n)^k}{k!} \tag{4.41}$$

Die Poisson-Verteilung hat nur λ als Parameter, der gleichzeitig Erwartungswert und Varianz ist.

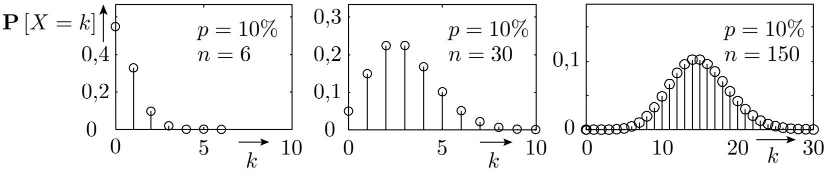
λ	Parameter der Poisson-Verteilung (Erwartungswert und gleichzeitig Varianz).
n	Anzahl der Zählversuche, maximaler Zählwert.
p	Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.

Anzahl der Zählversuche und Verteilung

(4.41)
$$= \mathrm{e}^{-p\cdot n}\cdot\frac{(p\cdot n)^k}{k!}$$



$\mathbb{P}[X=k]$	Verteilung der diskreten Zufallsvariablen X.
n	Anzahl der Zählversuche, maximaler Zählwert.
p	Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.



Grobabschätzung der wahrscheinlichen Bereiche:

- Für $\lambda = p \cdot n < 3$ keine untere Schranke $k_{\mathrm{L}} > 0$. Ober Schranke:

$$k_{\mathrm{U}} \approx 3 \ldots 5 \cdot \lambda$$

- Für $\lambda = p \cdot n \approx 3 \ldots 10$ zusätzlich untere Schranke:

$$k_{\mathrm{L}} \approx \frac{\lambda}{3 \ldots 5}$$

- Für $\lambda = p \cdot n > 10$ Abschätzung über Normalverteilung (Gl. 4.61):

$$\mathrm{sr} = [k_{\mathrm{U}}, k_{\mathrm{L}}] \approx \lambda \cdot \left(1 \mp 3 \ldots 5 \cdot \sqrt{\frac{1}{\lambda}}\right)$$

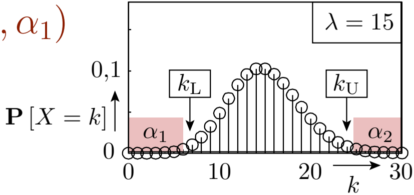
$k_{\mathrm{L}}, k_{\mathrm{U}}$	Untere (lower) und obere (upper) Bereichsgrenze.
sr	Symmetrischer Bereich der wahrscheinlichen Werte.
n	Anzahl der Zählversuche, maximaler Zählwert.
p	Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.

Bereichschätzung Pois

Schätzen $k_\mathrm{L}(\lambda, \alpha_1)$, $\lambda(k_\mathrm{L}, \alpha_1)$

Vorgabe k_L und α_1. Numerische Suche $\lambda(k_\mathrm{L}, \alpha_1)$, so dass

$$\sum_{k=0}^{k_\mathrm{L}-1} \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!} \le \alpha_1$$



α_1	$k_\mathrm{L} = 1$	2	3	4	5	6
0,5%	$\lambda =$ 5,298	7,430	9,273	10,978	12,593	14,150
1%	$\lambda =$ 4,606	6,638	8,406	10,045	11,605	13,109
2%	$\lambda =$ 3,912	5,834	7,516	9,084	10,580	12,027
10%	$\lambda =$ 2,303	3,890	5,323	6,681	7,993	9,275
20%	$\lambda =$ 1,609	2,995	4,279	5,514	6,721	7,906

Beispielabschätzungen:

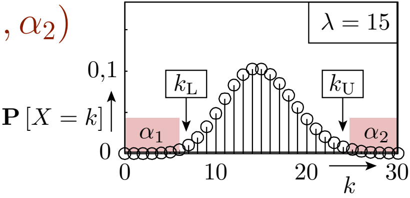
- $\lambda = 7$ und $\alpha_1 \le 1\% \Rightarrow k_\mathrm{L} = 2$
- $k_\mathrm{L} = 1$ und $\alpha_1 = 2\% \Rightarrow \lambda \ge 3{,}912$

α_1, α_2 Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
$k_\mathrm{L}, k_\mathrm{U}$ Untere (lower) und obere (upper) Bereichsgrenze.

Schätzen $k_U(\lambda, \alpha_2)$, $\lambda(k_U, \alpha_2)$

Vorgabe k_U und α_2. Numerische Suche $\lambda(k_U, \alpha_2)$, so dass

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



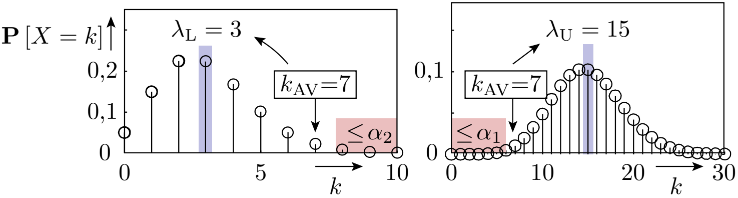
α_2	$k_U = 0$	1	2	3	4	5	6
0,5%	$\lambda =$ 0,005	0,103	0,338	0,672	1,078	1,537	2,037
1%	$\lambda =$ 0,01	0,148	0,436	0,823	1,279	1,785	2,330
2%	$\lambda =$ 0,02	0,215	0,567	1,016	1,529	2,089	2,684
10%	$\lambda =$ 0,105	0,532	1,102	1,744	2,432	3,152	3,894
20%	$\lambda =$ 0,223	0,824	1,534	2,296	3,089	3,903	4,733

Beispielabschätzungen:

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

α_2 Irrtumswahrscheinlichkeit, Werte oberhalb des geschätzten Bereichs.
k_U Obere Bereichsgrenze.

Schätzen von $[\lambda_L, \lambda_U]$ aus einem Ist-Zählwert k_{AV}



Von den Folien zuvor übernommene Werte $[\lambda_L, \lambda_U] = f(\alpha, k_{AV})$:

$\alpha_1 = \alpha_2$	$k_{AV} = 1$	$k_{AV} = 2$	$k_{AV} = 3$
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]

k_{AV} Ist-Zählwert (Actual count value).
α_1, α_2 Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
λ_L, λ_U Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.

$[\lambda_\mathrm{L},\ \lambda_\mathrm{U}] = f(\alpha, k_\mathrm{AV})$ für die Ist-Zählwerte $k_\mathrm{AV} = 4$ bis 6:

$\alpha_1 = \alpha_2$	$k_\mathrm{AV} = 4$	$k_\mathrm{AV} = 5$	$k_\mathrm{AV} = 6$
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]

$\lambda_\mathrm{U} = f(\alpha_1, k_\mathrm{AV})$ für den Ist-Zählwerte $k_\mathrm{AV} = 0$ ($\lambda_\mathrm{L} = 0$):

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

$$\lambda_\mathrm{U} = -\ln(\alpha_1)$$

α_1	0,5%	1%	2%	10%	20%
λ_U für $k_\mathrm{AV} = 0$	5,30	4,61	3,91	2,30	1,61%

k_AV Ist-Zählwert (Actual count value).
$\alpha_1,\ \alpha_2$ Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
$\lambda_\mathrm{L},\ \lambda_\mathrm{U}$ Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.

Beispiel 4.2: Poissonverteilte Anzahl Schadenfälle

a) *In den vergangen 10 Jahren ist kein Schaden eingetreten. Wie groß ist die zu erwartende Anzahl der Schadensfälle in den nächsten 10 Jahren mit einer Irrtumswahrschieinlichkeit* $\alpha_1 \leq 1\%$*?*

b) *In einem Nutzungjahr sind 5 Schadensfälle eingetreten. Auf welchen Bereich der zu erwartenden Anzahl der Schadensfälle lässt sich aus dieser Angabe für die nächsten 10 Nutzungsjahre mit den Irrtumswahrscheinlichkeiten* $\alpha_1 = \alpha_2 = 1\%$ *schließen?*

α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
k_{AV}	Ist-Zählwert (Actual count value).
$\lambda_{\mathrm{L}}, \lambda_{\mathrm{U}}$	Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.

a) *In den vergangen 10 Jahren ist kein Schaden eingetreten. Wie groß ist die zu erwartende Anzahl der Schadensfälle in den nächsten 10 Jahren mit einer Irrtumswahrschieinlichkeit $\alpha_1 \leq 1\%$?*

α_1	0,5%	1%	2%	10%	20%
λ_{U} für $k_{\mathrm{AV}} = 0$	5,30	4,61	3,91	2,30	1,61%

Bereich der zu erwartenden Anzahl der Schadensfälle: $\lambda_{\mathrm{L}} = 0$ bis $\lambda_{\mathrm{U}} = 4{,}61$. Die Anzahl der tatsächlichen Schadensfälle kann natürlich die geschätzte Obergrenze des Erwartungswert noch übersteigen.

α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
k_{AV}	Ist-Zählwert (Actual count value).
λ_{L}, λ_{U}	Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.

b) *In einem Nutzungjahr sind 5 Schadensfälle eingetreten. Auf welchen Bereich der zu erwartenden Anzahl der Schadensfälle lässt sich aus dieser Angabe für die nächsten 10 Nutzungsjahre mit den Irrtumswahrscheinlichkeiten* $\alpha_1 = \alpha_2 = 1\%$ *schließen?*

$\alpha_1 = \alpha_2$	$k_{\mathrm{AV}} = 4$	$k_{\mathrm{AV}} = 5$	$k_{\mathrm{AF}} = 6$
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]

Mindestens $10 \cdot \lambda_{\mathrm{L}} = 17{,}9$ und maximal $10 \cdot \lambda_{\mathrm{U}} = 116$ zu erwartende Schadensfälle.

α_1, α_2 Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
k_{AV} Ist-Zählwert (Actual count value).
λ_{L}, λ_{U} Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.

Beispiel 4.3: Fehlfunktionsrate

Bei $\#DS = 10^5$ Service-Leistungen wurden $k_{\mathrm{AV}} = 3$ Fehlfunktionen beobachtet.

Auf welche Unter- und Obergrenze für die Fehlfunktionsrate lässt sich mit den Irrtumswahrscheinlichkeit $\alpha_1 = \alpha_2 = 1\%$ unter Annahme einer Poissonverteilung schließen?

$\#DS$	Anzahl der erbrachten Service-Leistungen.
α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
k_{AV}	Ist-Zählwert (Actual count value).
$\left[\frac{\mathrm{MF}}{\mathrm{DS}}\right]$	Zählwertverhältnis in Fehlfunktionen je erbrachte Service-Leistung.

Bei $\#DS = 10^5$ Service-Leistungen wurden $k_{\mathrm{AV}} = 3$ Fehlfunktionen beobachtet.

Auf welche Unter- und Obergrenze für die Fehlfunktionsrate lässt sich mit den Irrtumswahrscheinlichkeit $\alpha_1 = \alpha_2 = 1\%$ unter Annahme einer Poissonverteilung schließen?

$\alpha_1 = \alpha_2 = 1\%$	$k_{\mathrm{AV}} = 1$	$k_{\mathrm{AV}} = 2$	$k_{\mathrm{AV}} = 3$
$[\lambda_{\mathrm{L}}, \lambda_{\mathrm{U}}]$	[0,15, 4,60]	[0,44, 6,64]	[0,82, 8,41]

Abschätzbarer Bereich der Fehlfunktionsrate:«

$$\zeta_{\mathrm{L}} = \frac{\lambda_{\mathrm{L}}}{\#DS} = 0{,}82 \cdot 10^{-5}\ [\mathrm{MF}/\mathrm{DS}]$$

$$\zeta_{\mathrm{U}} = \frac{\lambda_{\mathrm{U}}}{\#DS} = 8{,}41 \cdot 10^{-5}\ [\mathrm{MF}/\mathrm{DS}]$$

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

$\lambda_{\mathrm{L}}, \lambda_{\mathrm{U}}$ Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.

$\zeta_{\mathrm{L}}, \zeta_{\mathrm{U}}$ Untere und obere Bereichgrenze der geschätzten Fehlfunktionsrate.

Beispiel 4.4: Maskierungswahrscheinlichkeit

Eine Überwachungseinheit hat von $\#MF = 10.000$ Fehlfunktionen $k_{\mathrm{AV}} = 5$ Fehlfunktionen nicht erkannt.

Welchen Bereich der Maskierungswahrscheinlichkeit ergibt sich aus diesem Versuchsergebnis für eine Irrtumswahrscheinlichkeiten $\alpha_1 = \alpha_2 = 10\%$?

$\#MF$	Anzahl der Fehlfunktionen (Number of malfunctions).
k_{AV}	Ist-Zählwert (Actual count value).
α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.

Eine Überwachungseinheit hat von $\#MF = 10.000$ Fehlfunktionen $k_{\mathrm{AV}} = 5$ Fehlfunktionen nicht erkannt.

Welchen Bereich der Maskierungswahrscheinlichkeit ergibt sich aus diesem Versuchsergebnis für eine Irrtumswahrscheinlichkeiten $\alpha_1 = \alpha_2 = 10\%$?

$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = 10\%$	$k_{\mathrm{AV}} = 4$	$k_{\mathrm{AV}} = 5$	$k_{\mathrm{AV}} = 6$
$[\lambda_{\mathrm{L}},\ \lambda_{\mathrm{U}}]$	[2,43, 6,68]	[3,15, 7,99]	[3,89, 9,28]

Abschätzbarer Bereich der Maskierungswahrscheinlichkeit:

$$p_{\mathrm{ML}} = \frac{\lambda_{\mathrm{L}}}{\#MF} = 3{,}15 \cdot 10^{-4}$$

$$p_{\mathrm{MU}} = \frac{\lambda_{\mathrm{U}}}{\#MF} = 7{,}99 \cdot 10^{-4}$$

$\alpha_1,\ \alpha_2$ Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
$\lambda_{\mathrm{L}},\ \lambda_{\mathrm{U}}$ Minimaler bzw. maximaler Erwartungswert für einen poisson-verteilten Ist-Wert.
$p_{\mathrm{ML}},\ p_{\mathrm{MU}}$ Untere und obere Bereichsgrenze der geschätzten Maskierungswahrscheinlichkeit.

Beispiel 4.5: Zuverlässigkeitsbereich

Beim Test eines Systems mit $\#DS = 10^3$ Service-Leistungen wurden $k_{\mathrm{AV}} = 6$ Fehlfunktionen beobachtet.

Auf welchen Bereich der Zuverlässigkeit kann nach diesem Versuchsergebnis mit den Irrtumswahrscheinlichkeiten $\alpha_1 = \alpha_2 = 10\%$ geschlussfolgert werden?

$\#DS$	Anzahl der erbrachten Service-Leistungen.
k_{AV}	Ist-Zählwert (Actual count value).
α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.

Beim Test eines Systems mit $\#DS = 10^3$ Service-Leistungen wurden $k_{\mathrm{AV}} = 6$ Fehlfunktionen beobachtet.

Auf welchen Bereich der Zuverlässigkeit kann nach diesem Versuchsergebnis mit den Irrtumswahrscheinlichkeiten $\alpha_1 = \alpha_2 = 10\%$ geschlussfolgert werden?

$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = 10\%$	$k_{\mathrm{AV}} = 4$	$k_{\mathrm{AV}} = 5$	$k_{\mathrm{AV}} = 6$
$[\lambda_{\mathrm{L}},\ \lambda_{\mathrm{U}}]$	[2,43, 6,68]	[3,15, 7,99]	[3,89, 9,28]

- Abschätzbarer Bereich der MF-Rate:

$$\zeta_{\mathrm{L}} = 3{,}89 \cdot 10^{-3}\ [\mathrm{MF}/\mathrm{DS}]$$
$$\zeta_{\mathrm{U}} = 9{,}28 \cdot 10^{-3}\ [\mathrm{MF}/\mathrm{DS}]$$

- Daraus folgender Bereich der Zuverlässigkeit:

$$R_{\mathrm{L}} = \frac{1}{\zeta_{\mathrm{L}}} = 108\ [\mathrm{DS}/\mathrm{MF}]$$
$$R_{\mathrm{U}} = \frac{1}{\zeta_{\mathrm{U}}} = 257\ [\mathrm{DS}/\mathrm{MF}]$$

$\zeta_{\mathrm{L}}, \zeta_{\mathrm{U}}$ Untere und obere Bereichgrenze der geschätzten Fehlfunktionsrate.
$R_{\mathrm{L}}, R_{\mathrm{U}}$ Untere und obere Bereichgrenze der geschätzten Zuverlässigkeit.

Defektanteil

Zu erwartender Defektanteil und Fehleranzahl

Der zu erwartende Defektanteil ist die Wahrscheinlichkeit, dass ein Erzeugnis mindestens einen Fehler hat:

$$\mu_{\mathrm{DL}} = 1 - \mathbb{P}[X = 0]$$

Bei konstanter Fehlerentstehungsrate ist die zu erwartende Fehleranzahl für gleiche Produkte aus demselben Entstehungsprozess gleich und die Fehleranzahl ist poissonverteilt

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

mit $\lambda = \mu_{\mathrm{F}}$:

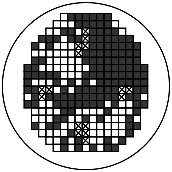
$$\mu_{\mathrm{DL}} = 1 - \mathrm{e}^{-\mu_{\mathrm{F}}} \qquad (4.42)$$

Für eine geringe Fehleranzahl $\mu_{\mathrm{F}} \ll 1$ sind der zu erwartende Fehleranteil und die zu erwartende Fehleranzahl gleich:

$$\mu_{\mathsf{DL}} = \mu_{\mathrm{F}} \qquad (4.43)$$

λ	Parameter der Poisson-Verteilung (Erwartungswert und gleichzeitig Varianz).
μ_{F}	Zu erwartende Fehleranzahl.
μ_{DL}	Zu erwartender Defektanteil.

Prozessgüteschwankungen und Fehlercluster



Fehlercluster auf einem Schaltkreiswafer [FSB87]

- □ funktionsfähige Schaltkreise
- ■ erkannte fehlerhafte Schaltkreise
- ⊠ Strukturen für Test und Diagnose

Örtlichen und zeilichen Schwankungen von Parametern, die die Fehlerentstehung und Beseitigung beeinflussen:

- Fehlerentstehungsraten,
- Fehlerabdeckungen, ...

verursachen Fehlercluster (lokale Fehlerhäufungen). Beispiele:

- Cluster von Schreibfehler in Texten,
- die qualitativ niederwertigen »Montagsprodukte«, ...

Fehlercluster liefern Hinweise auf Prozessverbesserungsmöglichkeiten.

Schaltkreisausbeute und Fehleranzahl

Die zu erwartende Schaltkreisausbeute ist die Wahrscheinlichkeit, dass die Anzahl der nachweisbaren Fehler aus den Entstehungsprozessen null ist. Für eine poisson-verteilte Fehleranzahl nach Gl. 4.42 mit $\mu_{\mathrm{F}} = FC \cdot \mu_{\mathrm{CF}}$:

$$\mu_{\mathrm{Y}} = \mathrm{e}^{-FC \cdot \mu_{\mathrm{CF}}} \tag{4.44}$$

Für jeden als fehlerfrei befundenen Schaltkreis müssen im Mittel

$$\frac{1}{\mu_{\mathrm{Y}}} = \mathrm{e}^{FC \cdot \mu_{\mathrm{CF}}}$$

Schaltkreise gefertigt werden.

μ_{Y}	Zu erwartende Ausbeute.
FC	Fehlerabdeckung (fault coverage), Anteil der nachweisbaren Fehler.
μ_{CF}	Zu erwartende Anzahl der Fehler aus den Entstehungsprozessen.

Fertigungskosten und Fehleranzahl

Annahme, dass die zu erwartende Fehleranzahl und die reinen Fertigungskosten je Schaltkreis proportional mit der Transistoranzahl je Schaltkreis zunehmen:

$$\mu_{\mathrm{CF}} = \xi_{\mathrm{Tr}} \cdot \#Tr$$
$$C_{\mathrm{MIC}} = C_{\mathrm{Tr}} \cdot \#Tr$$

(vergl. Gl. 2.45). Kosten je nicht als defekt aussortierter Schaltkreis:

$$C_{\mathrm{IC}} = \frac{C_{\mathrm{MIC}}}{\mu_{\mathrm{Y}}} = C_{\mathrm{Tr}} \cdot \#Tr \cdot \mathrm{e}^{FC \cdot \xi_{\mathrm{Tr}} \cdot \#Tr} \tag{4.45}$$

μ_{CF}	Zu erwartende Anzahl der Fehler aus den Entstehungsprozessen.
ξ_{Tr}	Fehlerentstehungsrate in Fehlern je Transistor.
$\#Tr$	Anzahl der Transistoren.
C_{Tr}	Transistorkosten in Euro pro Transistor.
C_{MIC}	Zu erwartende Kosten je gefertigter Schaltkreis.
C_{IC}	Kosten je als gut befundener (verkaufbarer) Schaltkreis.
Aussprache:	ξ xi.

Beispiel 4.6: Schaltkreiskosten

Fehlerentstehungsrate $\xi_{\mathrm{Tr}} = 10^{-6}$ [Fehler je Transistor], Fertigungskosten $C_{\mathrm{Tr}} = 10^{-6}$ [Euro je Transistor], Schaltkreisgröße $\#Tr \in \left\{10^5, 10^6, 10^7\right\}$ [Transistoren], Fehlerabdeckung $FC = 1$.

Welche Kosten entfallen auf jeden als gut befundene (verkaufbaren) Schaltkreis?

ξ_{Tr}	Fehlerentstehungsrate in Fehlern je Transistor.
C_{Tr}	Transistorkosten in Euro pro Transistor.
$\#Tr$	Anzahl der Transistoren.
FC	Fehlerabdeckung (fault coverage), Anteil der nachweisbaren Fehler.

Fehlerentstehungsrate $\xi_{\mathrm{Tr}} = 10^{-6}$ [Fehler je Transistor], Fertigungskosten $C_{\mathrm{Tr}} = 10^{-6}$ [Euro je Transistor], Schaltkreisgröße $\#Tr \in \{10^5, 10^6, 10^7\}$ [Transistoren], Fehlerabdeckung $FC = 1$.

Welche Kosten entfallen auf jeden als gut befundene (verkaufbaren) Schaltkreis?

$\#Tr$	10^5	10^6	10^7
$\mu_{\mathrm{CF}} = \xi_{\mathrm{Tr}} \cdot \#Tr$ in Fehlern	0,1	1	10
$C_{\mathrm{MIC}} = C_{\mathrm{Tr}} \cdot \#Tr$ in Euro	0,1	1	10
$\mu_{\mathrm{Y}} = \mathrm{e}^{-FC \cdot \mu_{\mathrm{CF}}}$	$90{,}5$	$36{,}8\%$	$4{,}54 \cdot 10^{-5}$
$C_{\mathrm{IC}} = \frac{C_{\mathrm{MIC}}}{Y}$ in Euro	$0{,}11$	$2{,}72$	$2{,}2 \cdot 10^5$

Ab $\mu_{\mathrm{CF}} \geq 2 \ldots 3$ zu erwartenden Fehlern pro Schaltkreis benötigen Schaltkreise Reparaturmöglichkeiten:

- deaktivieren / Ersatz defekter Funktionsblöcke,
- Verkauf z.B. als Prozessoren mit weniger Cache, ...

μ_{CF} Zu erwartende Anzahl der Fehler je gefertigter Schaltkreis.
C_{MIC} Zu erwartende Kosten je gefertigter Schaltkreis.
μ_{Y} Zu erwartende Ausbeute.
C_{IC} Kosten je als gut befundener (verkaufbarer) Schaltkreis.

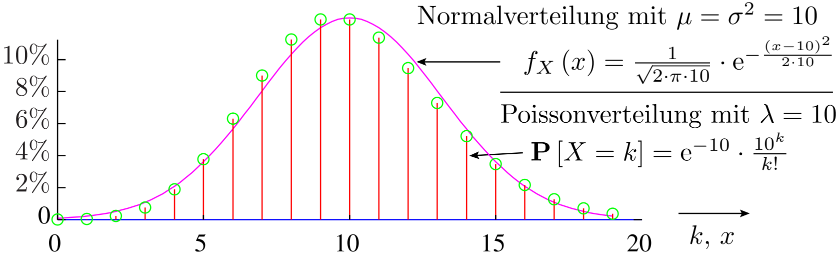
Normalverteilung

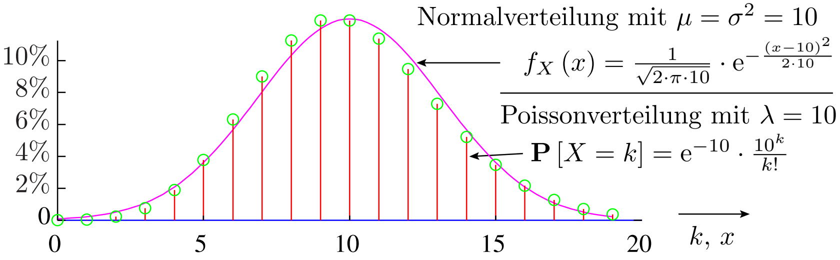
Normalverteilung

Die Summe sehr vieler unabhängiger Zufallsvariablen strebt unter sehr allgemeinen Bedingungen (keine dominanten Summanden, ...) gegen eine Normalverteilung:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}} \quad \text{mit } \sigma = \mathrm{sd}[X],\ \mu = \mathbb{E}[X] \tag{4.46}$$

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





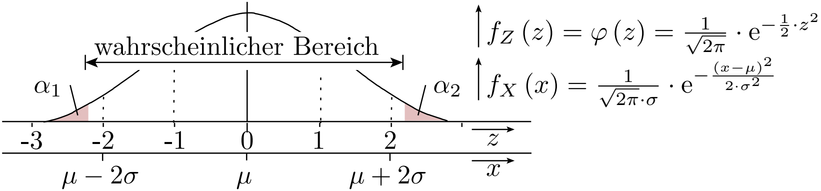
Für Zählwerte genügt die Annäherung der Zähl- durch eine Normalverteilung in der Regel bereits unter der Bedingung

$$10 \cdot \kappa \le \mu \le n - 10 \cdot \kappa \ \text{ mit } \mu = \sum_{i=1}^{n} p_i \tag{4.47}$$

n	Anzahl der Zählversuche, maximaler Zählwert.
p_i	Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.
μ	Erwartungswert.
σ	Standardabweichung.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \le 1$.

Bereichsschätzung NVT

Bereichsschätzung mit Normalverteilung



- Transformation einer Zufallsvariablen X mit Erwartungswert μ und Standardabweichung σ in eine Zufallsvariable Z mit Erwartungswert null und Standardabweichung eins:

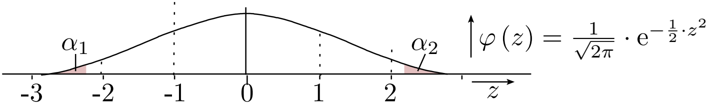
$$Z = \frac{X - \mu}{\sigma} \tag{4.48}$$

- Transformation der Werte x von X in z von Z:

$$z = \frac{x - \mu}{\sigma} \tag{4.49}$$

- Ablesen der Irrtumswahrscheinlichkeiten aus einer Tabelle der standardisierten Normalverteilungsfunktion $F_Z(z) = \Phi(z)$.

Standardisierte Normalverteilung



Verteilungsfunktion tabelliert für $z \geq 0$ in Schritten von 0,1:

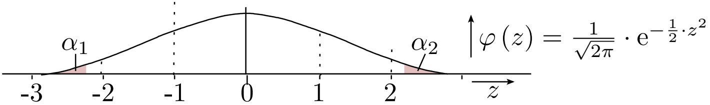
$$\Phi(z) = \int_{-\infty}^{z} \varphi(u) \cdot \mathrm{d}u$$

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 der Symmetrie gilt für $z < 0$:

$$\Phi(-z) = 1 - \Phi(z)$$

Inverse standardisierte Normalverteilung



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:

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

$$\Phi^{-1}(\alpha_1) = -\Phi^{-1}(1-\alpha_1)$$

X	Normalverteilte Zufallsvariable.
Z	Zufallsvariable mit Erwartungswert $\mu = 0$ und Standardabweichung $\sigma = 1$.
x, z	Werte der Zufallsvariablen X und Z.

Wahrscheinlichkeit der Bereichszugehörigkeit

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

Ausgehend von einer normalverteilten Zufallsvariablen X mit Erwartungswert μ und Standardabweichung σ und einem Bereich mit den Grenzen x_{L} und/oder x_{U}:

1. Transformation der Bereichsgrenze nach (Gl. 4.49)

$$z_{\mathrm{L}} = \frac{x_{\mathrm{L}} - \mu}{\sigma}$$

$$z_{\mathrm{U}} = \frac{x_{\mathrm{U}} - \mu}{\sigma}$$

2. Ablesen von $\Phi(z)$ bzw. für $z < 0$ von $\Phi(-z)$ aus der Tabelle.
3. Bestimmung der Irrtumswahrscheinlichkeiten:

$$\alpha_1 = \Phi(z_{\mathrm{L}}) = 1 - \Phi(-z_{\mathrm{L}}) = 1 - \Phi\left(\frac{\mu - x_{\mathrm{L}}}{\sigma}\right) \tag{4.50}$$

$$\alpha_2 = 1 - \Phi(z_{\mathrm{U}}) = 1 - \Phi\left(\frac{x_{\mathrm{U}} - \mu}{\sigma}\right) \tag{4.51}$$

Wahrscheinlicher Bereich

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

Ausgehend von einer einer normalverteilten Zufallsvariablen X mit Erwartungswert μ und Standardabweichung σ und zulässigen Irrtumswahrscheinlichkeiten α_1 oder/und α_2:

1. Ablesen aus der Tabelle:

$$z_{\mathrm{L}} = \Phi^{-1}\left(\alpha_1\right) = -\Phi^{-1}\left(1-\alpha_1\right)$$
$$z_{\mathrm{U}} = \Phi^{-1}\left(1-\alpha_2\right)$$

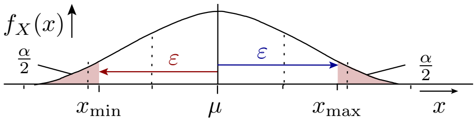
2. Transformation:

$$x_{\mathrm{L}} = \mu - \sigma \cdot z_{\mathrm{L}} = \mu - \sigma \cdot \Phi^{-1}\left(1-\alpha_1\right) \tag{4.52}$$
$$x_{\mathrm{U}} = \mu + \sigma \cdot z_{\mathrm{U}} = \mu + \sigma \cdot \Phi^{-1}\left(1-\alpha_2\right) \tag{4.53}$$

---

x_{L}, x_{U} Untere und oberen Schranke des wahrscheinlichen Bereichs von X.
z_{L}, z_{U} Transformierte unteren und obere Schranke.
μ, σ Erwartungswert, Standardabweichung.
α_1, α_2 Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.

Symmetrischer Bereich



α	4,54%	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

$$\varepsilon = \sigma\cdot\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right) \tag{4.54}$$

$$\mathrm{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \mu \mp \sigma\cdot\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right) \tag{4.55}$$

$$\mathrm{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \mu\cdot\left(1 \mp \sigma_{\mathrm{r}}\cdot\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right) \tag{4.56}$$

$f_X(x)$	Dichtefunktion der Zufallsvariablen X mit den möglichen Zählwerten x.
μ, x_{L}, x_{U}	Erwartungswert, untere und oberen Schranke des wahrscheinlichen Bereichs.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
ε, μ, σ	Intervallradius, Erwartungswert, Standardabweichung.
sr	Symmetrischer Bereich der wahrscheinlichen Werte.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
σ_{r}	Relative Standardabweichung zum erwarteten Eintrittszählwert.

Beispiel 4.7: Bereichschätzung Normalverteilung

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

a) *Mit welcher Wahrscheinlichkeit ist $X \geq 30$?*

b) *Mit welcher Wahrscheinlichkeit ist $X \leq 15$?*

c) *Welche obere Schranke x_{U} wird nur mit einer Irrtumswahrscheinlichkeit $\alpha_2 \leq 1\%$ überschritten?*

d) *Welche untere Schranke x_{L} wird nur mit einer Irrtumswahrscheinlichkeit $\alpha_1 \leq 2\%$ unterschritten?*

μ, σ	Erwartungswert, Standardabweichung.
α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
x_{L}, x_{U}	Untere und oberen Schranke des wahrscheinlichen Bereichs von X.

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

a) *Mit welcher Wahrscheinlichkeit ist $X \geq 30$?*

(4.51) $$\alpha_2 = 1 - \Phi\left(z_{\mathrm{U}}\right) = 1 - \Phi\left(\frac{x_{\mathrm{U}} - \mu}{\sigma}\right)$$

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

$$\alpha_2 = 1 - \Phi\left(\frac{30-20}{5}\right)$$
$$= 1 - \Phi\left(2\right)$$
$$\alpha_2 = 2{,}27\%$$

$\Phi(z)$ Verteilungsfunktion der standardisierten Normalverteilung.

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

b) *Mit welcher Wahrscheinlichkeit ist* $X \leq 15$*?*

(4.50) $$\alpha_1 = \Phi(z_L) = 1 - \Phi(-z_L) = 1 - \Phi\left(\frac{\mu - x_L}{\sigma}\right)$$

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

$$\alpha_1 = 1 - \Phi\left(\frac{20 - 15}{5}\right)$$
$$= 1 - \Phi(1)$$
$$\alpha_1 = 15{,}87\%$$

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

c) *Welche obere Schranke x_{U} wird nur mit einer Irrtumswahrscheinlichkeit $\alpha_2 \leq 1\%$ überschritten?*

(4.53) $$x_{\mathrm{U}} = \mu + \sigma \cdot z_{\mathrm{U}} = \mu + \sigma \cdot \Phi^{-1}\left(1 - \alpha_2\right)$$

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

$$\begin{aligned} x_{\mathrm{U}} &= 20 + 5 \cdot \Phi^{-1}\left(1 - 1\%\right) \\ &= 20 + 5 \cdot 2{,}33 \\ x_{\mathrm{U}} &= 31{,}65 \end{aligned}$$

$\Phi^{-1}(..)$ Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.

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

d) *Welche untere Schranke x_{L} wird nur mit einer Irrtumswahrscheinlichkeit $\alpha_1 \leq 2\%$ unterschritten?*

(4.52)
$$x_{\mathrm{L}} = \mu - \sigma \cdot z_{\mathrm{L}} = \mu - \sigma \cdot \Phi^{-1}\left(1 - \alpha_1\right)$$

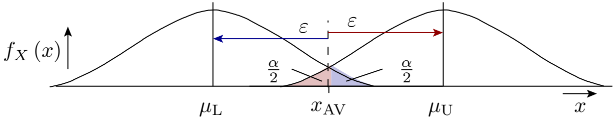
$\alpha_{1/2}$	2,27%	0,13%	0	2%	1%	0,5%	0,2%	0,1%
$\mp\Phi^{-1}\left(1-\alpha_{1/2}\right)$	2	3	4	2,05	2,33	2,57	2,88	3,10

$$\begin{aligned} x_{\mathrm{L}} &= 20 - 5 \cdot \Phi^{-1}\left(1 - 2\%\right) \\ &= 20 - 5 \cdot 2{,}05 \\ x_{\mathrm{L}} &= 9{,}75 \end{aligned}$$

Bereichsschätzung für den Erwartungswert

Der Erwartungswert μ zu einem beobachteten Zählwert x_{AV} ist

- mindestens so groß, dass $\mathbb{P}\left[X > x_{\mathrm{AV}}\right] < \frac{\alpha}{2}$ und
- maximal so groß, dass $\mathbb{P}\left[X < x_{\mathrm{AV}}\right] < \frac{\alpha}{2}$:

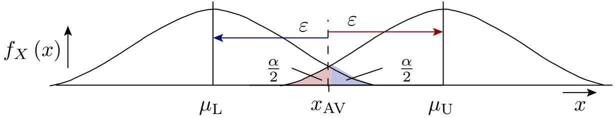


Intervallradius zwischen Erwartungswert und wahrscheinlichen Werten:

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

$f_X(x)$	Dichtefunktion der Zufallsvariablen X mit den möglichen Zählwerten x.
μ_{L}, μ_{U}	Untere und obere Schranke des wahrscheinlichen Bereichs des Erwartungswerts.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
ε	Intervallradius, Abstand zwischen Bereichsgrenzen und Erwartungswert.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
σ	Standardabweichung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.

Symmetrischer Bereich



Symmetrischer Bereich mit x_{AV} als Schätzwert:

$$\mathrm{sr}_\mu = [\mu_{\mathrm{L}}, \mu_{\mathrm{U}}] = x_{\mathrm{AV}} \mp \sigma \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{4.57}$$

$$= x_{\mathrm{AV}} \cdot \left(1 \mp \sigma_{\mathrm{r}} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \tag{4.58}$$

$f_X(x)$	Dichtefunktion der Zufallsvariablen X mit den möglichen Zählwerten x.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
$\mu_{\mathrm{L}}, \mu_{\mathrm{U}}$	Untere und obere Schranke des wahrscheinlichen Bereichs des Erwartungswerts.
sr_μ	Symmetrischer Bereich des wahrscheinlichen Erwartungswerts.
σ	Standardabweichung.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
σ_{r}	Relative Standardabweichung zum erwarteten Eintrittszählwert.

Beispiel 4.8: Bereichschätzung Erwartungswert

$x_{\mathrm{AV}} = 100$, $\sigma = 10$.

In welchem symmetrischen Bereich liegt der Erwartungswert mit Irrtumswahrscheinlichkeit $\alpha = 2\%$*.*

---

x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
σ	Standardabweichung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.

$x_{\mathrm{AV}} = 100$, $\sigma = 10$.

In welchem symmetrischen Bereich liegt der Erwartungswert mit Irrtumswahrscheinlichkeit $\alpha = 2\%$.

(4.57)
$$\mathrm{sr}_{\mu} = [\mu_{\mathrm{L}}, \mu_{\mathrm{U}}] = x_{\mathrm{AV}} \mp \sigma \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

α	4,54%	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

$$\Phi^{-1}(1 - 1\%) = 2{,}33$$
$$\mathrm{sr}_{\mu} = 100 \mp 10 \cdot 2{,}33 = 100 \mp 23{,}3$$

sr_{μ} Symmetrischer Bereich des wahrscheinlichen Erwartungswerts.
$\Phi^{-1}(..)$ Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.

Schätzen von Zählwerten

Schätzen von Zähl- und Erwartungswerten

Mit dem Erwartungswert und der Standardabweichung für Zählwerte:

(4.31) $$\mathbb{E}[X] = \mu = n \cdot p$$

(4.35) $$\sigma = \sqrt{\kappa \cdot n \cdot p \cdot (1-p)}$$

(siehe Binomialverteilungsnäherung) und

(4.52) $$x_{\mathrm{L}} = \mu - \sigma \cdot z_{\mathrm{L}} = \mu - \sigma \cdot \Phi^{-1}(1-\alpha_1)$$

(4.53) $$x_{\mathrm{U}} = \mu + \sigma \cdot z_{\mathrm{U}} = \mu + \sigma \cdot \Phi^{-1}(1-\alpha_2)$$

(4.55) $$\mathrm{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \mu \mp \sigma \cdot \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

ergeben sich folgende Bereichsgrenzen:

$$x_{\mathrm{L}} = p \cdot n - \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}(1-\alpha_1) \tag{4.59}$$

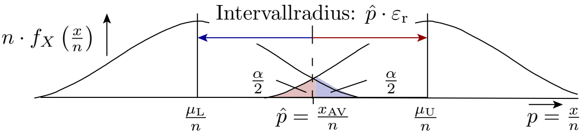
$$x_{\mathrm{U}} = p \cdot n + \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}(1-\alpha_2) \tag{4.60}$$

$$\mathrm{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = p \cdot n \mp \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \tag{4.61}$$

---

p Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.
n Anzahl der Zählversuche, maximaler Zählwert.
μ, x_{L}, x_{U} Erwartungswert, untere und oberen Schranke des wahrscheinlichen Bereichs.

Schätzen der Eintrittswahrscheinlichkeit



Schätzer der Eintrittswahrscheinlichkeit für n Zählversuche:

(4.37)
$$\hat{p}=\frac{\hat{\mu}}{n}=\frac{x_{\mathrm{AV}}}{n}$$

Varianzkoeffizient des Eintritts- bzw. Nichteintrittszählwerts:

(4.39)
$$\begin{aligned}\hat{\sigma}_{\mathrm{r}} &= \frac{\hat{\sigma}}{x_{\mathrm{AV}}}=\sqrt{\kappa\cdot\left(\frac{1}{x_{\mathrm{AV}}}-\frac{1}{n}\right)} && \text{für } \hat{p}\le 50\% \\ \hat{\sigma}_{\tilde{\mathrm{r}}} &= \frac{\hat{\sigma}}{n-x_{\mathrm{AV}}}=\sqrt{\kappa\cdot\left(\frac{1}{n-x_{\mathrm{AV}}}-\frac{1}{n}\right)} && \text{für } \hat{p}> 50\%\end{aligned}$$

$f_X(x)$	Dichtefunktion der Zufallsvariablen X.
μ_{L}, μ_{U}	Untere und obere Schranke des wahrscheinlichen Bereichs des Erwartungswerts.
n, x_{AV}	Anzahl der Zählversuche, Experimentell bestimmter Ist-Zählwert.
σ_{r}, $\sigma_{\tilde{\mathrm{r}}}$	Relative Standardabweichung zum erwarteten Eintritts- bzw. Nichteintrittszählwert.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa\le 1$.

Relative Invervallradien:

$$\varepsilon_{\mathrm{r}}=\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)}\text{ für }\hat{p}\le 50\% \tag{4.62}$$

$$\varepsilon_{\tilde{\mathrm{r}}}=\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{n-x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)}\text{ für }\hat{p}> 50\% \tag{4.63}$$

Symmetrischer Bereich:

$$\mathrm{sr}_{\mathrm{p}}=\left[p_{\mathrm{L}},p_{\mathrm{U}}\right]=\tfrac{x_{\mathrm{AV}}}{n}\cdot\left(1\mp\varepsilon_{\mathrm{r}}\right)\text{ für }\hat{p}\le 50\% \tag{4.64}$$

$$\mathrm{sr}_{\mathrm{p}}=\left[p_{\mathrm{L}},p_{\mathrm{U}}\right]=1-\left(1-\tfrac{x_{\mathrm{AV}}}{n}\right)\cdot\left(1\mp\varepsilon_{\tilde{\mathrm{r}}}\right)\text{ für }\hat{p}> 50\% \tag{4.65}$$

ε_{r}, $\varepsilon_{\tilde{\mathrm{r}}}$	Intervallradius realtiv zum erwarteten Eintritts- bzw. Nichteintritts-Zählwert.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa\le 1$.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
n, x_{AV}	Anzahl der Zählversuche, Experimentell bestimmter Ist-Zählwert.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.
sr_{p}	Geschätzter symmetrischer Bereich der Eintrittswahrscheinlichkeit.
$\left[-\frac{1}{n}\right]^{*}$	Term für kleine und große $\hat{p}$ und grobe Überschläge vernachlässigbar.

Geeignete Zählwertgrößen (ACR)

(4.62) $$\varepsilon_{\mathrm{r}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)} \text{ für } \hat{p}\le 50\%$$

(4.63) $$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{n-x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)} \text{ für } \hat{p}> 50\%$$

Geeignete Zählwertgrößen ergeben sich durch Auflösung nach der Anzahl der (nicht) eingetretenen Zählereignisse:

$$x_{\mathrm{AV}} \ge \frac{\kappa\cdot\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_{\mathrm{r}}^2}\cdot(1-\hat{p}) \text{ für } \hat{p}\le 50\% \tag{4.66}$$

$$n - x_{\mathrm{AV}} \ge \frac{\kappa\cdot\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_{\tilde{\mathrm{r}}}^2}\cdot\hat{p} \text{ für } \hat{p}> 50\% \tag{4.67}$$

Die Terme $(1-\hat{p})$ bzw. $\hat{p}$ sind mindestes 50% und max. eins.

ε_{r}, $\varepsilon_{\tilde{\mathrm{r}}}$	Intervallradius realtiv zum erwarteten Eintritts- bzw. Nichteintritts-Zählwert.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa\le 1$.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
n, x_{AV}	Anzahl der Zählversuche, Experimentell bestimmter Ist-Zählwert.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.
$\left[-\frac{1}{n}\right]^{*}$	Term für kleine und große $\hat{p}$ und grobe Überschläge vernachlässigbar.

Beispiel 4.9: Geeignete Zählwertgröße

Erforderliche Zählwerte für relative Intervallradien ε_{r} bzw. $\varepsilon_{\tilde{\mathrm{r}}}$ von 20% und 2% für zu schätzende Eintrittswahrscheinlichkeiten $\hat{p} \in \{10\%, 50\%, 90\%\}$. Irrtumswahrscheinlichkeit α so groß, dass $\Phi^{-1}\left(1-\frac{\alpha}{2}\right) = 2$, d.h. $\alpha = 4{,}52\%$, $\kappa = 1$.

ε_{r}	Intervallradius realtiv zum erwarteten Eintritts-Zählwert.
$\varepsilon_{\tilde{\mathrm{r}}}$	Intervallradius realtiv zum erwarteten Nichteintritts-Zählwerts.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \leq 1$.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
n	Anzahl der Zählversuche, maximaler Zählwert.

Erforderliche Zählwerte für relative Intervallradien ε_r bzw. $\varepsilon_{\tilde{\mathrm{r}}}$ von 20% und 2% für zu schätzende Eintrittswahrscheinlichkeiten $\hat{p} \in \{10\%, 50\%, 90\%\}$. Irrtumswahrscheinlichkeit α so groß, dass $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)=2$, d.h. $\alpha = 4{,}52\%$, $\kappa = 1$.

(4.66) $$x_\mathrm{AV} \geq \frac{\kappa \cdot \left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_\mathrm{r}^2} \cdot (1-\hat{p}) \text{ für } \hat{p} \leq 50\%$$

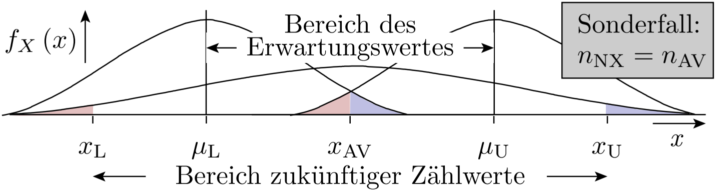
(4.67) $$n - x_\mathrm{AV} \geq \frac{\kappa \cdot \left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_{\tilde{\mathrm{r}}}^2} \cdot \hat{p} \text{ für } \hat{p} > 50\%$$

ε_r	$p=10\%$		$p=50\%$		$\varepsilon_{\tilde{\mathrm{r}}}$	$p=90\%$	
	$x_\mathrm{AV.min}$	n_min	$x_\mathrm{AV.min}$	n_min		$x_\mathrm{AV.max}$	n_min
20%	90	900	50	100	20%	810	900
2%	9.000	90.000	5.000	50.000	2%	81.000	90.000

Brauchbare Schätzungen verlangen eine große (Nicht-) Eintrittsanzahl $\gtrsim 100$ und insbesondere für sehr kleine (Nicht-) Eintrittswahrscheinlichkeiten eine sehr große Versuchsanzahl.

n_min Mindestanzahl der Zählversuche.
$x_\mathrm{AV.min}$ Minimal erforderliches Zählergebnis.
$x_\mathrm{AV.max}$ Maximal zulässiges Zählergebnis.

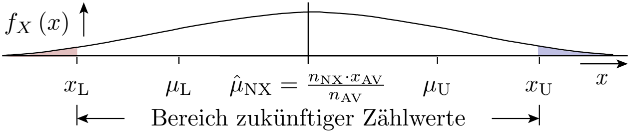
Schätzung künftiger aus bekannten Zählwerten



Für $n_{\mathrm{NX}} \neq n_{\mathrm{AV}}$ betragen die übereinstimmende Eintrittswahrscheinlichkeit und der Erwartungswert für künftige Zählwerte:

$$\hat{p} = \frac{x_{\mathrm{AV}}}{n_{\mathrm{AV}}} = \frac{\hat{\mu}_{\mathrm{NX}}}{n_{\mathrm{NX}}} \quad \Rightarrow \quad \hat{\mu}_{\mathrm{NX}} = \frac{n_{\mathrm{NX}} \cdot x_{\mathrm{AV}}}{n_{\mathrm{AV}}}$$

$f_X(x)$	Dichtefunktion der Zufallsvariablen X.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
μ_{L}, μ_{U}	Untere und obere Schranke des wahrscheinlichen Bereichs des Erwartungswerts.
x_{L}, x_{U}	Untere und oberen Schranke des wahrscheinlichen Bereichs von X.
n_{AV}	Anzahl der Zählversuche zur Schätzung der Eintrittswahrscheinlichkeit.
n_{NX}	Anzahl der Zählversuche zur Bestimmung eines künftigen Zählwerts.
$\hat{\mu}_{\mathrm{NX}}$	Erwartungswert des zu schätzenden Zählwerts.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.



	Erwartungswert	Varianz
X_{AV}	$p \cdot n_{AV}$	$\kappa \cdot n_{AV} \cdot p \cdot (1-p)$
X_{NX}	$p \cdot n_{NX}$	$\kappa \cdot n_{NX} \cdot p \cdot (1-p)$

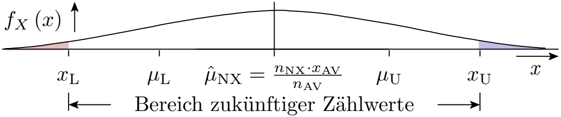
Die Abweichung der künftigen Werte von μ_{NX}:

$$X_\Delta = X_{NX} - \frac{n_{NX} \cdot X_{AV}}{n_{AV}}$$

Varianz der Differenz nach (Gl. 4.19):

$$\begin{aligned}\mathrm{Var}\left[X_\Delta\right] &= \mathrm{Var}\left[X_{NX} - \frac{n_{NX} \cdot X_{AV}}{n_{AV}}\right] = \mathrm{Var}\left[X_{NX}\right] + \left(\frac{n_{NX}}{n_{AV}}\right)^2 \cdot \mathrm{Var}\left[X_{AV}\right] \\ &= \kappa \cdot n_{NX} \cdot p \cdot (1-p) + \left(\frac{n_{NX}}{n_{AV}}\right)^2 \cdot \kappa \cdot n_{AV} \cdot p \cdot (1-p) \\ &= \kappa \cdot n_{NX} \cdot \left(1 + \frac{n_{NX}}{n_{AV}}\right) \cdot p \cdot (1-p)\end{aligned}$$

X_{AV} Zufallsvariable für die x_{AV} mit n_{AV} Zählversuchen bestimmt wird.
X_{NX} Zufallsvariable für die x_{NX} mit n_{NX} Zählversuchen bestimmt wird.



$\mathrm{Var}[X_\Delta]$ ist um den Faktor $\left(1 + \frac{n_{NX}}{n_{AV}}\right)$ größer, als die Varianz bei der Bereichsschätzung mit bekanntem Erwartungswert

(4.61) $\mathrm{sr} = [x_L, x_U] = p \cdot n \mp \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

Symmetrischer Bereich für die wahrscheinlichen künftigen Werte:

$$\mathrm{sr}_{NX} = [x_L, x_U] = \hat{\mu}_{NX} \mp \sqrt{\kappa \cdot n_{NX} \cdot \left(\frac{n_{NX}}{n_{AV}} + 1\right) \cdot \hat{p} \cdot (1 - \hat{p}) \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}$$
$$\text{mit} \quad \hat{p} = \frac{x_{AV}}{n_{AV}} \quad \text{und} \quad \hat{\mu}_{NX} = \frac{n_{NX} \cdot x_{AV}}{n_{AV}} \tag{4.68}$$

$f_X(x)$	Dichtefunktion der Zufallsvariablen X.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
μ_L, μ_U	Untere und obere Schranke des wahrscheinlichen Bereichs des Erwartungswerts.
x_L, x_U	Untere und oberen Schranke des wahrscheinlichen Bereichs von X.
n_{AV}	Anzahl der Zählversuche zur Schätzung der Eintrittswahrscheinlichkeit.
n_{NX}	Anzahl der Zählversuche zur Bestimmung eines künftigen Zählwerts.
sr_{NX}	Symmetrischer Bereich zukünftiger Zählergebnisse zu einem bekannten Ist-Zählwert.
$\hat{\mu}_{NX}$	Erwartungswert des zu schätzenden Zählwerts.
$\hat{p}$	Schätzwert der Eintrittswahrscheinlichkeit.

Beispiel 4.10: Bereich MF-Rate und künftige MF-Anzahl

Bei der Abarbeitung von 2.000 Service-Anforderungen wurden 100 Fehlfunktionen beobachtet. Keine Abhängigkeiten. Zugelassene Irrtumswahrscheinlichkeit 2%.

$n_{[\mathrm{AV}]} = 20.000\ [\mathrm{DS}]$, $x_{\mathrm{AV}} = 100\ [\mathrm{MF}]$, $\alpha = 2\%$, $\kappa = 1$.

a) *Schätzwert, Intervallradius, symmetrischer Bereich der MF-Rate?*

b) *Erforderliche Anzahl DS zur Verringerung des relativen Intervallradius auf* $\varepsilon_{\mathrm{r}} \leq 10\%$*?*

c) *Symmetrischer Bereich der Anzahl der Fehlfunktionen für* $n_{\mathrm{NX}} = 10.000\ [\mathrm{DS}]$ *mit der geschätzten MF-Rate aus Aufgabenteil a?*

n_{AV}	Anzahl der Zählversuche, mit denen x_{AV} bestimmt wurde.
$[\mathrm{DS}]$	Zählwert in erbrachten Service-Leistungen.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
$[\mathrm{MF}]$	Zählwert in Fehlfunktionen.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \leq 1$.
ε_{r}	Relativer Intervallradius der Fehlfunktionsrate.
n_{NX}	Anzahl der Zählversuche zur Bestimmung eines künftigen Zählwerts.

$n_{[\mathrm{AV}]} = 20.000$ [DS], $x_{\mathrm{AV}} = 100$ [MF], $\alpha = 2\%$, $\kappa = 1$.

a) *Schätzwert, Intervallradius, symmetrischer Bereich der MF-Rate?*

Die MF-Rate ist eine sehr kleine Eintrittswahrscheinlichkeit $\hat{\zeta} \ll 1$:

(4.62) $$\varepsilon_{\mathrm{r}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)} \text{ für } \hat{p} \le 50\%$$

(4.64) $$\mathrm{sr_p} = [p_{\mathrm{L}}, p_{\mathrm{U}}] = \frac{x_{\mathrm{AV}}}{n}\cdot(1 \mp \varepsilon_{\mathrm{r}}) \text{ für } \hat{p} \le 50\%$$

α	4,54%	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

$$\hat{\zeta} = \frac{100\,[\mathrm{MF}]}{20.000\,[\mathrm{DS}]} = 0{,}5\% \left[\frac{\mathrm{MF}}{\mathrm{DS}}\right]$$

$$\varepsilon_{\mathrm{r}} = \Phi^{-1}(1-1\%)\cdot\sqrt{\frac{1}{100}} = 0{,}233$$

$$\mathrm{sr}_{\zeta} = \hat{\zeta}\cdot(1 \mp \varepsilon_{\mathrm{r}}) = [0{,}38\%,\ 0{,}62\%] \left[\frac{\mathrm{MF}}{\mathrm{DS}}\right]$$

$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
$\hat{p}$, $\mathrm{sr_p}$	Geschätzte Wahrscheinlichkeit, symmetrischer Bereich der Wahrscheinlichkeit.
$\hat{\zeta}$, sr_{ζ}	Geschätzte Fehlfunktionsrate, symmetrischer Bereich der Fehlfunktionsrate.
ε_{r}	Intervallradius realtiv zum erwarteten Eintritts-Zählwert.

$n_{[\mathrm{AV}]} = 20.000\ [\mathrm{DS}],\ x_{\mathrm{AV}} = 100\ [\mathrm{MF}],\ \alpha = 2\%,\ \kappa = 1.$

b) *Erforderliche Anzahl DS zur Verringerung des relativen Intervallradius auf* $\varepsilon_{\mathrm{r}} \leq 10\%$*?*

(4.66) $$x_{\mathrm{AV}} \geq \frac{\kappa \cdot \left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_{\mathrm{r}}^2} \cdot (1-\hat{p}) \text{ für } \hat{p} \leq 50\%$$

Mindestanzahl der gezählten Fehlfunktionen:

$$x_{\mathrm{AV}} \geq \frac{2{,}33^2}{0{,}1^2} = 543\ [\mathrm{MF}]$$

Diese Zählergebnis wird erreicht nach etwa

$$n_{\mathrm{AV}} = \frac{x_{\mathrm{AV}}}{\zeta} = \frac{543\ [\mathrm{MF}]}{0{,}5\% \left[\frac{\mathrm{MF}}{\mathrm{DS}}\right]} = 108.000\ [\mathrm{DS}]$$

[MF]	Zählwert in Fehlfunktionen.
[DS]	Zählwert in erbrachten Service-Leistungen.

$n_{[\mathrm{AV}]} = 20.000\ [\mathrm{DS}],\ x_{\mathrm{AV}} = 100\ [\mathrm{MF}],\ \alpha = 2\%,\ \kappa = 1.$

c) *Symmetrischer Bereich der Anzahl der Fehlfunktionen für* $n_{\mathrm{NX}} = 10.000\ [\mathrm{DS}]$ *mit der geschätzten MF-Rate aus Aufgabenteil a?*

(4.68) $$\mathrm{sr}_{\mathrm{NX}} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \hat{\mu}_{\mathrm{NX}} \mp \sqrt{\kappa \cdot n_{\mathrm{NX}} \cdot \left(\frac{n_{\mathrm{NX}}}{n_{\mathrm{AV}}}+1\right) \cdot \hat{p} \cdot (1-\hat{p})} \cdot \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$
$$\text{mit} \quad \hat{p} = \frac{x_{\mathrm{AV}}}{n_{\mathrm{AV}}} \quad \text{und} \quad \hat{\mu}_{\mathrm{NX}} = \frac{n_{\mathrm{NX}} \cdot x_{\mathrm{AV}}}{n_{\mathrm{AV}}}$$

$$\hat{\mu}_{\mathrm{NX}} = 0{,}5\% \cdot 10.000 = 50\ [\mathrm{MF}]$$
$$\mathrm{sr}\left(x_{\mathrm{NX}}\right) = 50 \mp \sqrt{\left(\frac{10^4}{2\cdot 10^4} + 1\right) \cdot 50 \cdot (1 - 0{,}5\%)} \cdot 2{,}33\ [\mathrm{MF}]$$
$$= 50 \mp 20{,}1\ [\mathrm{MF}]$$

Relativer Intervallradius:

$$\varepsilon_{\mathrm{r}} = \frac{20{,}1\ [\mathrm{MF}]}{50\ [\mathrm{MF}]} \approx 40\%$$

$\hat{\mu}_{\mathrm{NX}}$ Erwartungswert des zu schätzenden Zählwerts.
$\mathrm{sr}_{\mathrm{NX}}$ Symmetrischer Bereich zukünftiger Zählergebnisse zu einem bekannten Ist-Zählwert.

Varianzerhöhung

Varianzerhöhung durch Abhängigkeiten

Abhängigkeiten erhöhen Varianz und Standardabweichung und den Intervallradius 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:

$$X = \sum_{i=1}^{\#X/2} X_i \quad \text{mit} \quad \mathbb{P}\left[X_i = k\right] = \begin{cases} 1 - p_i & k = 0 \\ p_i & k = 2 \end{cases}$$

Erwartungswert der Summanden:

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

Varianz der Summanden (nach Verschiebungssatz):

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

$\#X$ Anzahl der Zufallsvariablen.

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

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

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

$$\mathrm{Var}[X] = \sum_{i=1}^{\#X/2} 2^2 \cdot p_i \cdot (1 - p_i) = 2 \cdot \left(\underbrace{\sum_{i=1}^{\#X/2} 2 \cdot p_i \cdot (1 - p_i)}_{*} \right) = 2 \cdot \#X \cdot p \cdot (1 - p)$$

* Varianz von $\#X$ unabhängigen Zählwerten mit paarweise gleichem p_i. Standardabweichung und Intervallradius vergrößern sich um $\sqrt{2}$.

$\#X$	Anzahl der Zufallsvariablen.
p_i	Wahrscheinlichkeit, dass das Ergebnis von Zählversuch i eins ist.
p	Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.

Varianzerhöhung

Die Varianzerhöhung sei definiert als Verhältnis aus tatsächlicher Varianz und der Varianz einer Binomialverteilung mit derselben Versuchanzahl n und derselben mittleren Eintrittswahrscheinlichkeit p entsprechend Gl. 4.35:

$$\kappa = \frac{\mathrm{Var}[X]}{n\cdot p\cdot(1-p)} = \frac{\sigma^2}{\mathbb{E}[X]\cdot(1-p)} \tag{4.69}$$

Die Varianz der Binomialverteilung ist eine einfach abschätzbare Obergrenze für die Varianz unabhängiger Zählwerte. Im Beispiel »paarweise identisch nachweisbare Fehler« ist die Varianzerhöhung bei übereinstimmenden Eintrittswahrscheinlichkeiten p_i $\kappa = 2$ und bei stark abweichende p_{i} etwas kleiner.

Analog ergibt sich für »immer $\#IDC$ identische Zählereignisse« Varianerhöhung:

$$\kappa \le \#IDC$$

κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \le 1$.
$\mathrm{Var}\,[X]$	Varianze der Zufallsvariablen X.
$\mathbb{E}\,[X]$	Erwartungswert der Zufallsvariablen X.
$\#IDC$	Anzahl der identischen Zählwerte.

Schätzen der Varianzerhöhung

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

(4.13) $$\hat{\mathbb{E}}[X] = \hat{\mu} = \frac{1}{\#v} \cdot \sum_{i=1}^{\#v} v_i$$

- Schätzen der Varianz der Zählwertstichprobe:

(4.14) $$\hat{\mathrm{Var}}[X] = \hat{\sigma}^2 = \frac{1}{\#v-1} \cdot \sum_{i=1}^{\#v} (v_i - \hat{\mu})^2$$

- Schätzen der Eintrittswahrscheinlichkeit:

(4.37) $$\hat{p} = \frac{\hat{\mu}}{n} = \frac{x_{\mathrm{AV}}}{n}$$

- Geschätze Varianzerhöhung nach Gl. 4.69:

$$\hat{\kappa} = \frac{\hat{\mathrm{Var}}[X]}{\hat{\mathbb{E}}[X] \cdot (1-\hat{p})} = \frac{\hat{\sigma}^2}{\hat{\mu} \cdot (1-\hat{p})} \tag{4.70}$$

$\hat{\mathbb{E}}[X]$	Geschätzter Erwartungswert der Zufallsvariablen X.
$\#v$	Größe der Datenstichprobe.
v_i	Wert i der Datenstichprobe.
$\hat{\mathrm{Var}}[X]$	Geschätzte Varianze der Zufallsvariablen X.
$\hat{\kappa}$	Geschätzte Varianzerhöhung durch Abhängigkeiten.

Einführung

Beispiel 4.11: Varianzerhöhung

$\#v = 10$ Wiederholungen »Zählwertbestimmung für $n = 1.000$ Zählversuche:

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

Abschätzung der Varianzerhöhung $\hat{\kappa}$?

$\#v$	Größe der Datenstichprobe.
v_i	Wert i der Datenstichprobe.
$\hat{\kappa}$	Geschätzte Varianzerhöhung durch Abhängigkeiten.

Einführung

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

Abschätzung der Varianzerhöhung $\hat{\kappa}$?

(4.13) $$\hat{\mathbb{E}}[X] = \hat{\mu} = \frac{1}{\#v} \cdot \sum_{i=1}^{\#v} v_i$$

(4.14) $$\hat{\mathrm{Var}}[X] = \hat{\sigma}^2 = \frac{1}{\#v-1} \cdot \sum_{i=1}^{\#v} (v_i - \hat{\mu})^2$$

(4.70) $$\hat{\kappa} = \frac{\hat{\mathrm{Var}}[X]}{\hat{\mathbb{E}}[X]\cdot(1-\hat{p})} = \frac{\hat{\sigma}^2}{\hat{\mu}\cdot(1-\hat{p})}$$

$$\hat{\mu} = \frac{1}{10} \cdot \sum_{i=1}^{10} v_i = 63{,}1; \quad \hat{p} = \frac{\hat{\mu}}{n} = 6{,}3\%$$

$$\hat{\sigma}^2 = \frac{1}{9} \cdot \sum_{i=1}^{10} (v_i - 63{,}1)^2 = 135$$

$$\hat{\kappa} = \frac{135}{63{,}1 \cdot (1 - 6,3\%3)} = 2{,}28$$

Die Abhängigkeiten erhöhen die Varianz so, als ob mehr als 2 Zählereignisse fast immer gemeinsam eintreten.

Beispiel 4.12: Anzahl Schadensfälle mit Varianzerhöhung

Der zu erwartende Zählwert für die Anzahl von Schadensfällen sei 100. Irrtumswahrscheinlichkeit 2%, Varianzerhöhung 2.

$x_{\mathrm{AV}} = 100\ [\mathrm{D}]$, $n_{\mathrm{AV}} \gg x_{\mathrm{AV}}$, $\alpha = 2\%$, $\kappa = 2$.

In welchem symmetrischen Bereich wird bei künftigen Wiederholungen unter denselben Versuchsbedingungungen $(n_{\mathrm{NX}} = n_{\mathrm{AV}})$ *die Anzahl der Schadensfälle liegen?*

---

x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
n_{AV}	Anzahl der Zählversuche, mit denen x_{AV} bestimmt wurde.
n_{NX}	Anzahl der Zählversuche, mit denen x_{NX} bestimmt werden soll.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \leq 1$.
[D]	Zählwert in Schadensfällen.

$x_{\mathrm{AV}} = 100$ [D], $n_{\mathrm{AV}} \gg x_{\mathrm{AV}}$, $\alpha = 2\%$, $\kappa = 2$.

In welchem symmetrischen Bereich wird bei künftigen Wiederholungen unter denselben Versuchsbedingungungen $(n_{\mathrm{NX}} = n_{\mathrm{AV}})$ *die Anzahl der Schadensfälle liegen?*

$$\mathrm{sr}_{\mathrm{NX}} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \hat{\mu}_{\mathrm{NX}} \mp \sqrt{\kappa \cdot n_{\mathrm{NX}} \cdot \left(\frac{n_{\mathrm{NX}}}{n_{\mathrm{AV}}} + 1\right) \cdot \hat{p} \cdot (1-\hat{p})} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{4.68}$$

$$\text{mit} \quad \hat{p} = \frac{x_{\mathrm{AV}}}{n_{\mathrm{AV}}} \quad \text{und} \quad \hat{\mu}_{\mathrm{NX}} = \frac{n_{\mathrm{NX}} \cdot x_{\mathrm{AV}}}{n_{\mathrm{AV}}}$$

α	4,54%	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

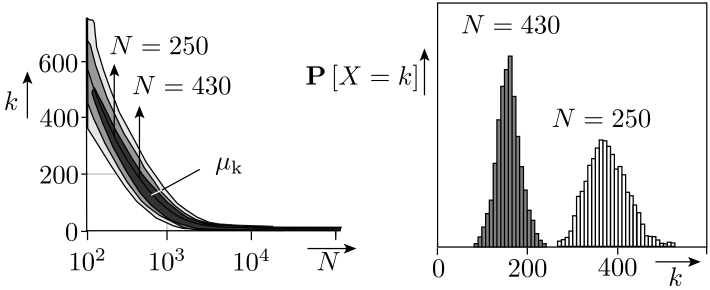
Für $n_{\mathrm{NX}} = n_{\mathrm{AV}}$ und $n_{\mathrm{AV}} \gg x_{\mathrm{AV}}$ ist $\mu_{\mathrm{NX}} = x_{\mathrm{AV}}$:

$$\mathrm{sr}_{\mathrm{NX}} = x_{\mathrm{AV}} \mp \sqrt{\kappa \cdot 2 \cdot x_{\mathrm{AV}}} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$
$$= 100 \mp 20 \cdot 2{,}33 = [53{,}4,\ 146{,}6]\ [\mathrm{D}]$$

$\mathrm{sr}_{\mathrm{NX}}$ Symmetrischer Bereich zukünftiger Zählergebnisse zu einem bekannten Ist-Zählwert.

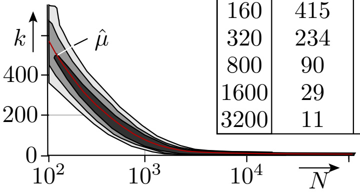
Experiment mit Haftfehlern

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



k	Anzahl der nicht nachweisbaren Modellfehler.
n	Anzahl der Modellfehler.
N	Anzahl der Tests.

Varianzerhöhung im Experiment



N	$\hat{\mu}$	$\hat{\sigma^2}$	$\hat{\sigma}$	$\hat{\kappa}_1 = \frac{\hat{\sigma^2}}{\hat{\mu}}$	$\hat{\kappa}_2$	$\hat{\kappa}_3$
160	415	1875	43,3	5,1	1,7	1,0
320	234	943	30,7	4,3	1,6	1,1
800	90	299	17,3	3,4	1,6	1,1
1600	29	52	7,2	1,8	*	1,3
3200	11	8,4	2,9	*	*	*

* $\hat{\sigma^2} < \hat{\mu}$

$\hat{\kappa}_1$: Fehlersimulation mit allen $n_1 = 3606$ Haftfehlern. Abhängigkeiten bis, als ob 3...5 Modellfehler identisch nachweisbar wären. Erkennbare Identitäten waren aber beseitigt. Bleiben als Abhängigkeiten impliziter Nachweis und geteilte Steuer- und Beobachtungsbedingungen.

$\hat{\kappa}_2$, $\hat{\kappa}_3$: Simulation mit Fehlerstichproben $n_2 = 1.000$ bzw. $n_3 = 300$. Abnahme der Abhängigkeiten mit der Verkleinerung der Fehlerstichprobe.

$\hat{\kappa}_{1/2/3}$ Geschätzte Varianzerhöhung für 1 – alle, $n_1 = 3606$, 2 – eine Stichprobe von $n_2 = 1000$ und 3 – eine Stichprobe von $n_3 = 300$ Modellfehlern.

Fehlermodellierung und Vorhersagbarkeit

Für den Fehlernachweis werden Weiswahrscheinlichkeit $p > 50\%$ angestrebt. Relativer Intervallradius:

(4.63) $$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{n-x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)}\text{ für }\hat{p} > 50\%$$

Das Experiment zuvor hat gezeigt, dass für eine große Anzahl von Modellfehlern bezogen auf die Testobjektgröße die Varianzerhöhung κ mit der Modellfehleranzahl n zunimmt. Damit lässt sich der relative Intervallradius $\varepsilon_{\tilde{\mathrm{r}}}$ als Maß der Schätzgenauigkeit nicht unbegrenzt durch mehr Modellfehler verringern.

$\varepsilon_{\tilde{\mathrm{r}}}$	Intervallradius realtiv zum erwarteten Nichteintritts-Zählwerts.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \leq 1$.
p	Mittlere Eintrittswahrscheinlichkeit der zu zählenden Ereignisse.
n	Anzahl der Zählversuche, maximaler Zählwert.
$\varepsilon_{\tilde{\mathrm{r}}}$	Relativer Intervallradius der Anzahl der nichtnachweisbaren Modellfehler.

Schlussfolgerungen

- Bei zufälliger Testauswahl hilft eine zu große Modellfehleranzahl im Verhältnis zur Testobjektgröße nicht, die Schätzgenauigkeit für die Fehlerabdeckung zu verbessern.
- Fehlermodelle, bei denen die Anzahl der Modellfehler überproportional mit der Testobjektgröße zunimmt, z.B. Kurzschlüsse und Pfadverzögerungsfehler, sind nicht zielführend (siehe später Abschn. 6.1.2 ff.).

Abschätzung mit mehreren Zufallstests

Alternative zur Verringerung des relativen Intervallradius

(4.63) $\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{n-x_{\mathrm{AV}}}\left[-\frac{1}{n}\right]^{*}\right)}$ für $\hat{p} > 50\%$

ist eine Bestimmung der Anzahl der nicht nachweisbaren Modellfehler x_{AV} für $\#v > 1$ unterschiedliche Zufallstestsätze wie im Experiment (siehe Folie Experiment mit Haftfehlern, Abschn. 4.2.8). Vergrößert die Anzahl der Zählversuche n und Zählwerte x_{AV} um $\#v$

$$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\kappa\cdot\left(\frac{1}{\#v\cdot(n-x_{\mathrm{AV}})}-\frac{1}{\#v\cdot n}\right)} \qquad (4.71)$$

ohne, dass sich die Abhängigkeiten zwischen den Zählversuche, beschrieben durch κ, in ähnlichem Maße mit erhöhen.

$\#v$	Anzahl der unterschiedlichen Zufallstestsätze.
$\varepsilon_{\tilde{\mathrm{r}}}$	Intervallradius realtiv zum erwarteten Nichteintritts-Zählwerts.
$\Phi^{-1}(..)$	Inverse Funktion zur Verteilungsfunktion der standardisierten Normalverteilung.
α, κ	Irrtumswahrscheinlichkeit, Varianzerhöhung durch Abhängigkeiten.
n, x_{AV}	Anzahl der Zählversuche, Experimentell bestimmter Ist-Zählwert.

Beispiel 4.13: Bereich der Modellfehlerabdeckung

Von 1000 Modellfehlern wurden 32 nicht erkannt. Varianzerhöhung maximal 2. Zulässige Irrtumswahrscheinlichkeit 2%.

$n = 1000$ [F], $x_{\mathrm{AV}} = 1000 - 32$ [F], $\kappa \leq 2$, $\alpha = 2\%$

a) *In welchem Bereich liegt die Modellfehlerabdeckung und wie groß ist der relative Intervallradius des Anteils der nicht nachweisbaren Fehler bei Abschätzung mit einem Zufallstestsatz* $(\#v = 1)$*?*

b) *Der relative Intervallradius soll max. 10% betragen. Für wie viele unterschiedliche Zufallstestsätz ist die Anzahl der nicht nachweisbaren Fehler dafür zu mitteln?*

n	Anzahl der Modellfehler.
[F]	Zählwert in Modellfehler.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.
κ	Varianzerhöhung durch Abhängigkeiten, für unabhängige Zählwerte $\kappa \leq 1$.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
$\#v$	Anzahl der unterschiedlichen Zufallstestsätze.

$n = 1000$ [F], $x_{\mathrm{AV}} = 1000 - 32$ [F], $\kappa \leq 2$, $\alpha = 2\%$

a) *In welchem Bereich liegt die Modellfehlerabdeckung und wie groß ist der relative Intervallradius des Anteils der nicht nachweisbaren Fehler bei Abschätzung mit einem Zufallstestsatz* $(\#v = 1)$*?*

(4.71) $$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\kappa \cdot \left(\frac{1}{\#v \cdot (n - x_{\mathrm{AV}})} - \frac{1}{\#v \cdot n}\right)}$$

(4.65) $$\mathrm{sr}_{\mathrm{p}} = [p_{\mathrm{L}}, p_{\mathrm{U}}] = 1 - \left(1 - \frac{x_{\mathrm{AV}}}{n}\right) \cdot (1 \mp \varepsilon_{\tilde{\mathrm{r}}}) \text{ für } \hat{p} > 50\%$$

α	4,54%	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

$$\varepsilon_{\tilde{\mathrm{r}}} = 2{,}33 \cdot \sqrt{2 \cdot \left(\frac{1}{32} - \frac{1}{1000}\right)} = 58\%$$

$$\mathrm{sr}_{\mathrm{p}} = 1 - \left(1 - \frac{1}{32}\right) \cdot (1 \mp 58\%) = [94{,}9\%,\ 98{,}7\%]$$

$\varepsilon_{\tilde{\mathrm{r}}}$ Relativer Intervallradius der Anzahl der nichtnachweisbaren Modellfehler.
sr_{p} Geschätzter symmetrischer Bereich der Fehlerüberdeckung.

$n = 1000$ [F], $x_{\mathrm{AV}} = 1000 - 32$ [F], $\kappa \leq 2$, $\alpha = 2\%$

b) *Der relative Intervallradius soll max. 10% betragen. Für wie viele unterschiedliche Zufallstestsätz ist die Anzahl der nicht nachweisbaren Fehler dafür zu mitteln?*

(4.71) $$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\kappa \cdot \left(\frac{1}{\#v \cdot (n - x_{\mathrm{AV}})} - \frac{1}{\#v \cdot n}\right)}$$

Umstellung nach $\#v$:

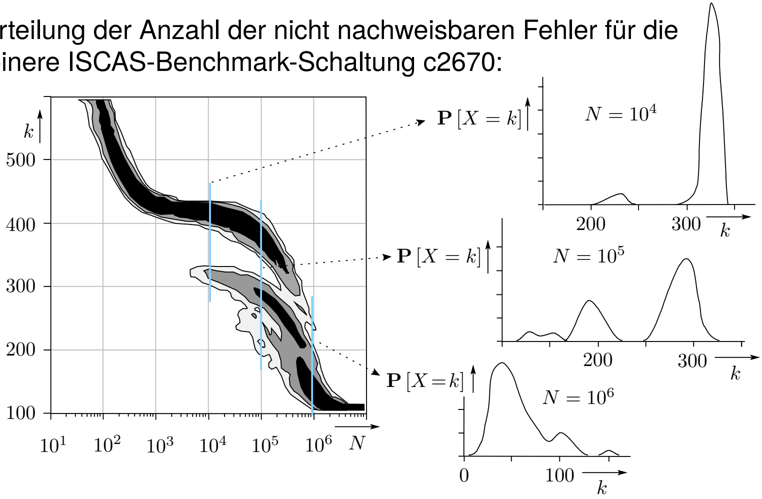
$$\#v \geq \kappa \cdot \left(\frac{1}{(n - x_{\mathrm{AV}})} - \frac{1}{n}\right) \cdot \left(\frac{\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)}{\varepsilon_{\tilde{\mathrm{r}}}}\right)^2$$

$$= 2 \cdot \left(\frac{1}{32} - \frac{1}{1000}\right) \cdot \left(\frac{2{,}33}{0{,}1}\right)^2 = 35$$

Die Modellfehlerabdeckung muss für mindestens 35 unterschiedliche Zufallstestsätze bestimmt und gemittelt werden.

Nicht normalverteilte Zählwerte

Verteilung der Anzahl der nicht nachweisbaren Fehler für die kleinere ISCAS-Benchmark-Schaltung c2670:



Im Bereich der Testsatzlänge $N = 10^4$ bis 10^6 mehrere Gipfel.

Wie ist das möglich? $\Rightarrow$ Mischverteilung, siehe nächster Abschnitt.

Zusammenfassung

Näherungen für Zählwertverteilungen

Wenn die Eintrittswahrscheinlichkeit p_i aller Zählwerte bekannt sind, gibt es einen Algorithmus für die Verteilungsberechung (siehe Abschnitt zuvor). Wenn weniger bekannt ist, z.B. nur die Anzahl der Zählversuche n und ein experimentell bestimmter Istwert x_{AV} gibt es Näherungen:

- Binomialverteilung mit demselben Erwartungswert und derselben Versuchsanzahl,
- für kleine Zählwerte und große Versuchsanzahl Poissonverteilung mit demselben Erwartungswert,
- bei mindestes etwa 10 eingetretenen und 10 nicht eingetretenen Zählwerten, Normalverteilung mit demselben Erwartungswert und einer aus dem Erwartungswert abschätzbaren Varianz.

Darauf aufsetzend praktischen Bereichsschätzungen:

- wahrscheinliche Bereiche*,
- Irrtumswahrscheinlichkeiten* und
- geeignete Zählwertgrößen für ausreichend genaue Schätzungen.

* Ausser für Zählwerte auch für allgemeine Zufallsgrößen, z.B. Messwerte.

Binomialverteilung

Für den Sonderfall, dass gleichwahrscheinliche Ereignisse gezählt werden, ist die Summe der gezählten Ereignisse binomialverteilt:

$$\mathbb{P}[X=k] = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \tag{4.30}$$

Erwartungswert:

$$\mathbb{E}[X] = \mu = n \cdot p \tag{4.31}$$

Varianz:

$$\mathrm{Var}[X] = \sigma^2 = n \cdot p \cdot (1-p) \tag{4.32}$$

Die Standardabweichung:

$$\mathrm{sd}[X] = \sigma = \sqrt{n \cdot p \cdot (1-p)} \tag{4.33}$$

Die Varianz einer Binomialverteilung ist bei gleichem Erwartungswert eine Obergrenze der Varianz der Verleilung unabhängige Zählwerte:

$$\sigma^2 = \sum_{i=1}^{n} p_i \cdot (1-p_i) \le n \cdot p \cdot (1-p) \tag{4.34}$$

Schätzer für Zählwerte

Erwartungswert, Standardabweichung, Eintrittswahrscheinlichkeit:

(4.36) $$\hat{\mu} = x_{\mathrm{AV}}$$

(4.38) $$\hat{\sigma} = \sqrt{\kappa \cdot x_{\mathrm{AV}} \cdot \left(1 - \frac{x_{\mathrm{AV}}}{n}\right)}$$

(4.37) $$\hat{p} = \frac{\hat{\mu}}{n} = \frac{x_{\mathrm{AV}}}{n}$$

Der Varianzkoeffizient für das Eintreten bzw. Nicht-Eintreten als Maße für die relative Breite des wahrscheinlichen Bereichs im Verhältnis zur (Nicht-) Eintrittsanzahl:

(4.39) $$\begin{aligned} \hat{\sigma}_{\mathrm{r}} &= \frac{\hat{\sigma}}{x_{\mathrm{AV}}} = \sqrt{\kappa \cdot \left(\frac{1}{x_{\mathrm{AV}}} - \frac{1}{n}\right)} && \text{für } \hat{p} \leq 50\% \\ \hat{\sigma}_{\tilde{\mathrm{r}}} &= \frac{\hat{\sigma}}{n - x_{\mathrm{AV}}} = \sqrt{\kappa \cdot \left(\frac{1}{n - x_{\mathrm{AV}}} - \frac{1}{n}\right)} && \text{für } \hat{p} > 50\% \end{aligned}$$

Benötigt für Bereichsschätzungen über Normalverteilung.

Poissonverteilung

Beim Zählen vieler seltener Ereignisse sind die Zählwerte näherungsweise poisson-verteilt:

(4.40) $$\mathbb{P}\left[X=k\right]=\mathrm{e}^{-\lambda}\cdot\frac{\lambda^k}{k!}$$

Der Verteilungsparameter λ, gleichzeitig Erwartungswert und Varianz, ist das Produkt aus der Anzahl der Zählversuche und der mittleren Eintrittswahrscheinlichkeit:

(4.41) $$=\mathrm{e}^{-p\cdot n}\cdot\frac{(p\cdot n)^k}{k!}$$

Ab $\lambda\geq 10$ können poissonverteilte Zählwerte näherungsweise als normalverteilt mit Varianz gleich Erwartungswert betrachtet werden.

Fehleranzahl, Defektanteil, Schaltkreiskosten

Neue Antworten auf alte Fragen:

- Zu erwartender Defektanteil in Abhängigkeit von der zu erwartenden Fehleranzahl bei konstanter Fehlerentstehungsrate:

(4.42) $$\mu_{\mathrm{DL}} = 1 - \mathrm{e}^{-\mu_{\mathrm{F}}}$$

- Abweichungen deuten auf Fehlercluster. Fehlercluster signalisieren abstellbare Prozessprobleme (Fehlervermeidung). In unserer idealisierten Fehlerkultur – Beseitigung aller erkennbarer Probleme – wird Fehlerzahl zumindest gegen Poissonverteilung tendieren.
- Zu erwartende Ausbeute für Schaltkreise und auch andere nicht reparierbare Systeme:

(4.44) $$\mu_{\mathrm{Y}} = \mathrm{e}^{-FC\cdot\mu_{\mathrm{CF}}}$$

- Kosten je verkaufbarer Schaltkreis und Entscheidungshilfe »Aussortieren oder Reparatur«:

(4.45) $$C_{\mathrm{IC}} = \frac{C_{\mathrm{MIC}}}{\mu_{\mathrm{Y}}} = C_{\mathrm{Tr}} \cdot \#Tr \cdot \mathrm{e}^{FC\cdot\xi_{\mathrm{Tr}}\cdot\#Tr}$$

Bereichsschätzung Poissonverteilung

Für die betrachteten kleinen Zählwerte von null bis <10 wurden Tabellen erstellt für

- $\lambda\left(k_{\mathrm{L}},\alpha_1\right)$: Mindesterwartungswert λ, um mit α_1 für einen Mindestzählwert k_{L} zu garantieren,
- $\lambda\left(k_{\mathrm{U}},\alpha_2\right)$: max. Erwartungswert λ, um mit α_2 für einen max Zählwert k_{U} zu garantieren,
- $\left[\lambda_{\mathrm{L}},\lambda_{\mathrm{U}}\right]\left(k_{\mathrm{AV}},\alpha\right)$: Min. und max. Erwartungswert zu einem tatsächlichen Zählwert.

Aus diesen Tabellen lassen sich

- für gegebene Irrtumswahrscheinlichkeiten und Erwartungswerte Minima und Maxima für die Zählwerte,
- für gegebene Erwartungswerte und Minima bzw. Maxima für die Zählwerte die Irrtumswahrscheinlichkeiten oder
- für gegebene Irrtumswahrscheinlichkeiten und Zählwerte Erwartungswertbereiche ablesen.

Vorhersagen künftiger seltener Ereignisse, z.B. Katastrophen, bei null oder wenigen bisher beobachteten Ereignissen sind sehr unscharf.

Normalverteilung, Bereichsschätzungen

Die Normalverteilung ist eine stetige Verteilung mit der Dichtefunktion:

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

Zählwerte meist näherungsweise normalverteilt ab:

(4.47) $$10\cdot\kappa \le \mu \le n - 10\cdot\kappa \;\; \text{mit } \mu = \sum_{i=1}^{n} p_i$$

Für Bereichsschätzungen Transformation normalverteilter Zufallsvariablen und Bereichsgrenzen in die der standardisierten Normalverteilung:

(4.48) $$Z = \frac{X-\mu}{\sigma}$$

(4.49) $$z = \frac{x-\mu}{\sigma}$$

Ablesen Irrtumswahrscheinlichkeiten aus Tabelle $\Phi(z)$ (siehe Folie Standardisierte Normalverteilung, Abschn. 4.2.6):

- für »Wert kleiner als untere Bereichsgrenze«:

  (4.50) $$\alpha_1 = \Phi(z_{\mathrm{L}}) = 1 - \Phi(-z_{\mathrm{L}}) = 1 - \Phi\left(\frac{\mu - x_{\mathrm{L}}}{\sigma}\right)$$

- für »Wert größer als obere Bereichsgrenze«:

  (4.51) $$\alpha_2 = 1 - \Phi(z_{\mathrm{U}}) = 1 - \Phi\left(\frac{x_{\mathrm{U}} - \mu}{\sigma}\right)$$

Bereichsgrenzen für gegebene Irrtumswahrscheinlichkeiten:

(4.52)
$$x_{\mathrm{L}}=\mu-\sigma\cdot z_{\mathrm{L}}=\mu-\sigma\cdot\Phi^{-1}\left(1-\alpha_1\right)$$

(4.53)
$$x_{\mathrm{U}}=\mu+\sigma\cdot z_{\mathrm{U}}=\mu+\sigma\cdot\Phi^{-1}\left(1-\alpha_2\right)$$

Intervallradius und symmetrischer Bereich:

(4.54)
$$\varepsilon=\sigma\cdot\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)$$

(4.55)
$$\mathrm{sr}=\left[x_{\mathrm{L}},x_{\mathrm{U}}\right]=\mu\mp\sigma\cdot\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)$$

Symmetrischer Bereich des Erwartungswerts um einen experimentell bestimmten Istwert:

(4.57)
$$\mathrm{sr}_{\mu}=\left[\mu_{\mathrm{L}},\mu_{\mathrm{U}}\right]=x_{\mathrm{AV}}\mp\sigma\cdot\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)$$

Bereichsschätzung für Zählwerte

Ober Schranke, untere Schranke und symmetrischer Bereich:

(4.59) $$x_{\mathrm{L}} = p \cdot n - \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}\left(1-\alpha_1\right)$$

(4.60) $$x_{\mathrm{U}} = p \cdot n + \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}\left(1-\alpha_2\right)$$

(4.61) $$\mathrm{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = p \cdot n \mp \sqrt{\kappa \cdot n \cdot p \cdot (1-p)} \cdot \Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)$$

Eintrittswahrscheinlichkeit, Intervallradius, Varianzkoeffizient für Eintritt und Nichteintritt:

(4.37) $$\hat{p} = \tfrac{\hat{\mu}}{n} = \tfrac{x_{\mathrm{AV}}}{n}$$

(4.62) $$\varepsilon_{\mathrm{r}} = \Phi^{-1}\left(1-\tfrac{\alpha}{2}\right) \cdot \sqrt{\kappa \cdot \left(\tfrac{1}{x_{\mathrm{AV}}} \left[-\tfrac{1}{n}\right]^{*}\right)} \text{ für } \hat{p} \le 50\%$$

(4.63) $$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1-\tfrac{\alpha}{2}\right) \cdot \sqrt{\kappa \cdot \left(\tfrac{1}{n-x_{\mathrm{AV}}} \left[-\tfrac{1}{n}\right]^{*}\right)} \text{ für } \hat{p} > 50\%$$

(4.64) $$\mathrm{sr_p} = [p_{\mathrm{L}}, p_{\mathrm{U}}] = \tfrac{x_{\mathrm{AV}}}{n} \cdot (1 \mp \varepsilon_{\mathrm{r}}) \text{ für } \hat{p} \le 50\%$$

(4.65) $$\mathrm{sr_p} = [p_{\mathrm{L}}, p_{\mathrm{U}}] = 1 - \left(1 - \tfrac{x_{\mathrm{AV}}}{n}\right) \cdot (1 \mp \varepsilon_{\tilde{\mathrm{r}}}) \text{ für } \hat{p} > 50\%$$

Geeignete Zählwertgröße:

(4.66) $$x_{\mathrm{AV}} \geq \frac{\kappa\cdot\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_{\mathrm{r}}^2}\cdot(1-\hat{p})\ \text{für}\ \hat{p}\leq 50\%$$

(4.67) $$n - x_{\mathrm{AV}} \geq \frac{\kappa\cdot\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\varepsilon_{\tilde{\mathrm{r}}}^2}\cdot\hat{p}\ \text{für}\ \hat{p}> 50\%$$

Intervallradius künftiger Zählwerte um einen experimentell bestimmten Istwert:

(4.68) $$\mathrm{sr}_{\mathrm{NX}} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \hat{\mu}_{\mathrm{NX}} \mp \sqrt{\kappa\cdot n_{\mathrm{NX}}\cdot\left(\frac{n_{\mathrm{NX}}}{n_{\mathrm{AV}}}+1\right)\cdot\hat{p}\cdot(1-\hat{p})}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$
$$\text{mit}\quad \hat{p}=\frac{x_{\mathrm{AV}}}{n_{\mathrm{AV}}}\quad \text{und}\quad \hat{\mu}_{\mathrm{NX}}=\frac{n_{\mathrm{NX}}\cdot x_{\mathrm{AV}}}{n_{\mathrm{AV}}}$$

Varianzerhöhung

Varianzerhöhung und ihre Abschätzung:

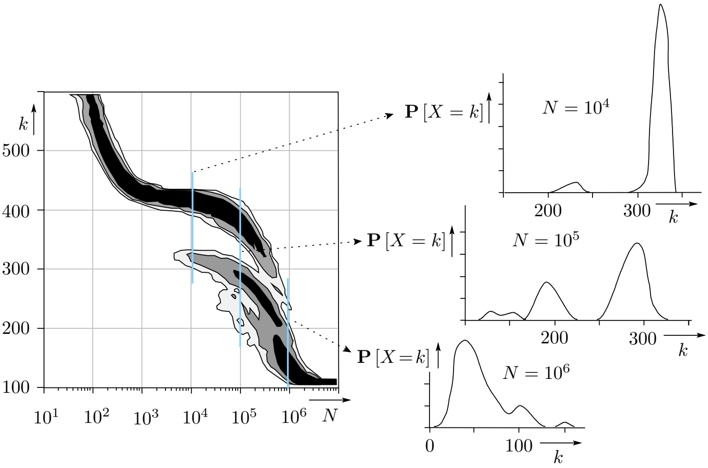
(4.69) $$\kappa = \frac{\mathrm{Var}[X]}{n \cdot p \cdot (1-p)} = \frac{\sigma^2}{\mathbb{E}[X] \cdot (1-p)}$$

(4.70) $$\hat{\kappa} = \frac{\hat{\mathrm{Var}}[X]}{\hat{\mathbb{E}}[X] \cdot (1-\hat{p})} = \frac{\hat{\sigma}^2}{\hat{\mu} \cdot (1-\hat{p})}$$

Zwischen Modellfehlern gibt es auch, wenn identisch nachweisbare Fehler zusammengefasst sind, erhebliche Varianzerhöhungen durch gemeinsame Anregungs- und Beobachtungsbedingungen.

Zu viele Modellfehler im Vergleich zur Objektgröße mindern ab einer gewissen Größe die Varianz nicht weiter und verbessern damit auch nicht weiter die Schätzgenauigkeit. Zur genaueren Abschätzung Mittelung der Fehlerabdeckungen mehrerer Zufallstestsätze:

(4.71) $$\varepsilon_{\tilde{\mathrm{r}}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\kappa \cdot \left(\frac{1}{\#v \cdot (n - x_{\mathrm{AV}})} - \frac{1}{\#v \cdot n}\right)}$$



Auf (siehe Folie Nicht normalverteilte Zählwerte, Abschn. 4.2.8) hat die experimentell bestimmte Verteilung der Anzahl der nicht nachweisbaren Haftfehler der ISCAS-Benchmark-Schaltung c2670 mehrere Gipfel. Wie ist das möglich?

Mischverteilung

Mischverteilung

Aus einer Grundgesamtheit gemischter Objekte mit unterschiedlichen Verteilungen wählt eine diskrete Zufallsvariable Y mit Verteilung

$$\mathbb{P}\left[Y = j\right] = h_j$$

zufällig ein Objekt X_i mit Verteilung F_{X_j} aus. Die resultierende Verteilungsfunktion, Verteilung bzw. Dichte ergeben sich durch gewichtete Mittelwertbildung:

$$F_X\left(x\right) = \mathbb{P}\left[X \leq x\right] = \sum_{j=1}^{\#Y} h_j \cdot F_{X_j}\left(x\right) \tag{4.72}$$

$$\mathbb{P}\left[X = x_i\right] = \sum_{j=1}^{\#Y} h_j \cdot \mathbb{P}\left(X_j = x_i\right) \tag{4.73}$$

$$f_X\left(x\right) = \frac{\mathrm{d}F_X\left(x\right)}{\mathrm{d}x} = \sum_{i=1}^{\#Y} h_j \cdot f_{X_j}\left(x\right) \tag{4.74}$$

---

h_j Wahrscheinlichkeit, dass ein Objekt mit Verteilungsfunktion j ausgewählt wird.

Zufallsvariablen mit einer Mischverteilung

- 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.
- ...

Eigenschaften

Varianzvergrößerung durch »Mischung«

Der Erwartungswert ist der gewichtete Mittelwert:

$$\mathbb{E}\left[X\right] = \mu = \sum_{j=1}^{\#Y} h_j \cdot \mu_j \tag{4.75}$$

Varianz bei abweichenden Erwartungswerten $\mu = \mu_j + \delta_j$:

$$\begin{aligned} \mathrm{Var}\left[X\right] = \sigma = \sum_{j=1}^{\#Y} h_j \cdot \mathbb{E}\left[(X_i - \mu_j + \delta_j)^2\right] \\ = \sum_{j=1}^{\#Y} h_j \left(\underbrace{\mathbb{E}\left[(X_j - \mu_j)^2\right]}_{\sigma_j^2} - 2 \cdot \delta_j \cdot \underbrace{\mathbb{E}\left[(X_j - \mu_j).\right]}_{0} + \underbrace{\mathbb{E}\left[\delta_j^2\right]}_{\delta_j^2} \right) \end{aligned}$$

$$\mathrm{Var}\left[X\right] = \sigma^2 = \sum_{j=1}^{\#Y} h_j \cdot \sigma_j^2 + \sum_{j=1}^{\#Y} h_j \cdot \delta_j^2 \tag{4.76}$$

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

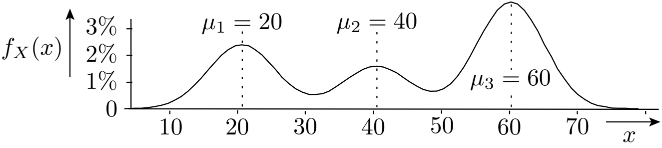
Multimodale Verteilung

Beim Mischen von Grundgesamtheiten mit deutlich abweichenden Erwartungswerten entstehen multimodale Verteilungen. Beispiel: Mischen von drei Normalverteilungen nach Gl. 4.74 mit:

h_i	0,3	0,2	0,5
μ_i	20	40	60
σ_i	5	5	5

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

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



(4.75) $$\mathbb{E}[X] = \mu = \sum_{j=1}^{\#Y} h_j \cdot \mu_j$$

(4.76) $$\operatorname{Var}[X] = \sigma^2 = \sum_{j=1}^{\#Y} h_j \cdot \sigma_j^2 + \sum_{j=1}^{\#Y} h_j \cdot \delta_j^2$$

h_i	0,3	0,2	0,5
μ_i	20	40	60
σ_i	5	5	5

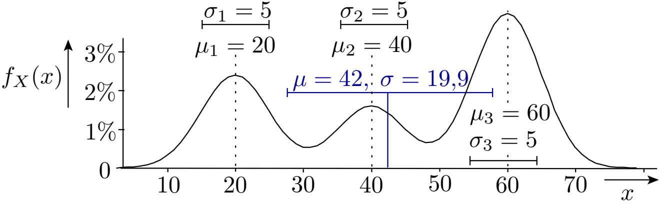
.

Erwartungswert:

$$\mu = 0{,}3 \cdot 20 + 0{,}2 \cdot 40 + 0{,}5 \cdot 60 = 42$$

Varianz, Standardabweichung:

$$\sigma^2 = 25 + 0{,}3 \cdot (20-42)^2 + 0{,}2 \cdot (40-42)^2 + 0{,}5 \cdot (60-42)^2 = 285$$
$$\sigma = 16{,}9$$



Anwendungen

Beispiel 4.14: Identisch nachweisbare Fehler

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

$n = 25$, davon 10 identisch nachweisbar, $p = 60\%$

a) *Mischverteilung als Zusammensetzung aus zueinander verschobenen Binomialverteilungen?*

b) *Erwartungswert und Varianz?*

c) *Standardabweichung und Varianzerhöhung?*

n	Anzahl der Zählversuche, maximaler Zählwert.
p	Eintrittswahrscheinlichkeit.

$n = 25$, davon 10 identisch nachweisbar, $p = 60\%$

a) *Mischverteilung als Zusammensetzung aus zueinander verschobenen Binomialverteilungen?*

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

(4.73) $$\mathbb{P}\left[X = x_i\right] = \sum_{j=1}^{\#Y} h_j \cdot \mathbb{P}\left(X_j = x_i\right)$$

Binomialverteilung ohne die 10 nur gemeinsam nachweisbaren Fehler:

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

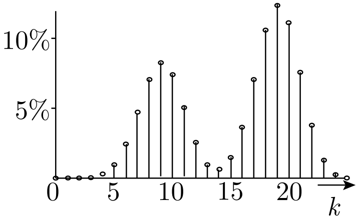
Mit den 10 nur gemeinsam nachweisbaren Fehlern verschieben sich alle Wahrscheinlichkeiten um 10 Realisierung:

$$\mathbb{P}\left[X_1 = k+10\right] = \begin{cases} \binom{n-10}{k} \cdot p^k \cdot (1-p)^{n-10-k} & 0 \le k \le n-10 \\ 0 & \text{sonst} \end{cases}$$

$n = 25$, davon 10 identisch nachweisbar, $p = 60\%$

Mischverteilung:

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



$n = 25$, davon 10 identisch nachweisbar, $p = 60\%$

b) *Erwartungswert und Varianz?*

(4.75) $$\mathbb{E}[X] = \mu = \sum_{j=1}^{\#Y} h_j \cdot \mu_j$$

(4.76) $$\operatorname{Var}[X] = \sigma^2 = \sum_{j=1}^{\#Y} h_j \cdot \sigma_j^2 + \sum_{j=1}^{\#Y} h_j \cdot \delta_j^2$$

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:

$$\operatorname{Var}[X] = 15 \cdot p \cdot (1-p) + 10^2 \cdot p \cdot (1-p) = 115 \cdot p \cdot (1-p) = 27{,}6$$
$$\stackrel{!}{=} \underbrace{n \cdot p \cdot (1-p)}_{3{,}6} + \underbrace{(1-p) \cdot (\mathbb{E}[X_0] - \mathbb{E}[X])^2}_{+\quad 0{,}4\cdot(9-15)^2} + \underbrace{p \cdot (\mathbb{E}[X_1] - \mathbb{E}[X])^2}_{+\quad 0{,}6\cdot(19-15)^2} = 27{,}6\checkmark$$

$n = 25$, davon 10 identisch nachweisbar, $p = 60\%$

c) *Standardabweichung und Varianzerhöhung?*

(4.9) $$\mathrm{sd}\left[X\right] = \sigma = \sqrt{\mathrm{Var}\left[X\right]}$$

(4.69) $$\kappa = \frac{\mathrm{Var}[X]}{n\cdot p\cdot(1-p)} = \frac{\sigma^2}{\mathbb{E}[X]\cdot(1-p)}$$

Standardabweichung:

$$\mathrm{sd}\left[X\right] = \sqrt{27{,}6} = 5{,}25$$

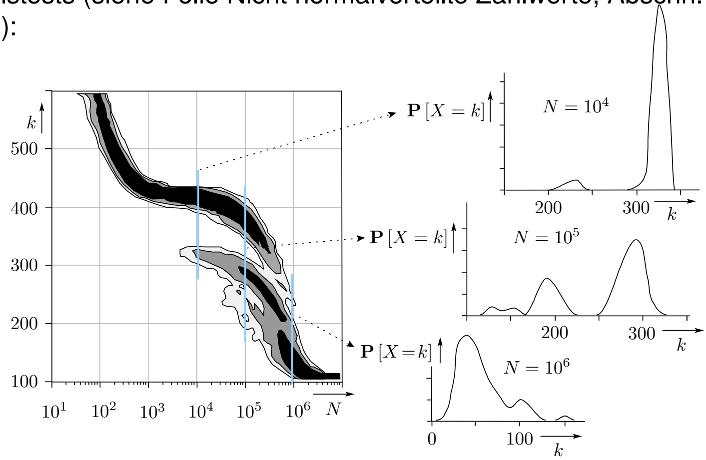
Varianzerhöhung: ?????

$$\kappa = \frac{27{,}6}{15} = 1{,}84$$

Etwas kleinere Varianzerhöhung als bei einer Modellfehlermenge aus paarweise identisch nachweisbaren Fehlern.

Dichte nicht nachweisbare Fehler c2670

... in Abhängigkeit von der Länge N eines Zufallstests (siehe Folie Nicht normalverteilte Zählwerte, Abschn. 4.2.8):

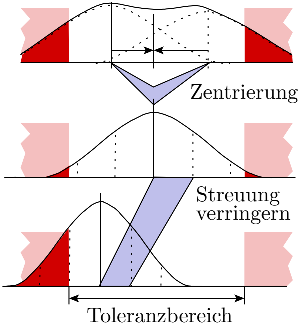


Im Bereich von $N = 10^4$ bis 10^6 multimodale Verteilung. Offenbar ca. 80 sehr ähnlich nachweisbare Fehler mit MF-Rate $\zeta_i \approx 10^{-5}\left[\frac{\mathrm{MF}}{\mathrm{DS}}\right]$.

Objekte aus unterschiedlichen Prozessen

Bei der mechanischen Fertigung haben die Zielparameter, z.B. bei einer Bohrung Durchmesser und Tiefe, eine Verteilung und einen Toleranzbereich. Entstehungshäufigkeit eines Parameterfehlers ist die Wahrscheinlichkeit, Parameter außerhalb Toleranzbereich. Bei erkennbarer Polarisierung der Messwerte eines Parameters:

- Lokalisierung der Prozesse, deren Ergebnisse gemischt werden.
- Prozesszentrierung: Verschiebung der Verteilung für jeden Einzelprozess mit Hilfe von Einstelloptionen in die Mitte des Toleranzbereichs.
- Prozessverbesserung: Verringerung der Streuung durch technologische Neuerungen, neue Maschinen, Verfahren, ...

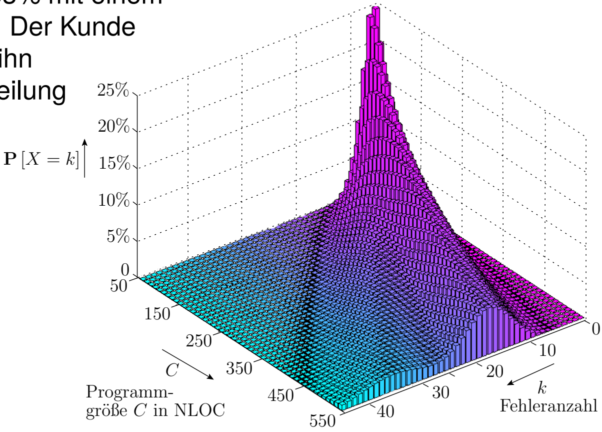


(siehe Folie Prozesszentrierung, Abschn. 2.3.2)

Unterschiedlich gute Programmierer

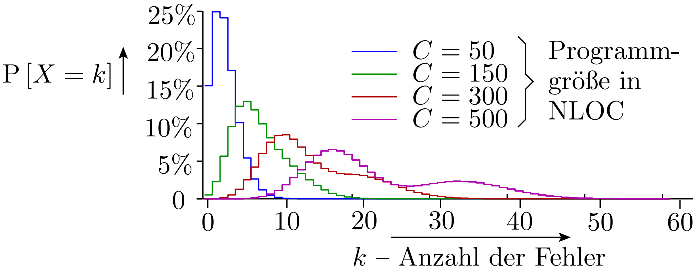
Ein Anfänger und ein Profi entwickeln Software-Bausteine aus C Netto Lines of Code (NLOC), der Profi 66% mit ca. einem Fehler je 30 NLOC und der Anfänger 33% mit einem Fehler je 15 NLOC. Der Kunde weiß nicht, wer für ihn programmiert. Verteilung der Fehleranzahl:

$$\mathbb{P}[X=k] = \underbrace{\frac{2}{3}\cdot \mathrm{e}^{-\frac{C}{30}}\cdot\frac{\left(\frac{C}{30}\right)^k}{k!}}_{\mathrm{Pois}\left(\lambda=\frac{C}{30}\right)} + \underbrace{\frac{1}{3}\cdot \mathrm{e}^{-\frac{C}{15}}\cdot\frac{\left(\frac{C}{15}\right)^k}{k!}}_{\mathrm{Pois}\left(\lambda=\frac{C}{15}\right)}$$



Die Wahrscheinlichkeit, dass ein Modul genau k Fehler enthält, ist 2/3 mal die Wahrscheinlichkeit, das es k Fehler enthält und vom Profi stammt plus 1/3 mal die Wahrscheinlichkeit, dass es k Fehler enthält und vom Anfänger stammt (vergl. Gl. 4.73):

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



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

C	Metrik für den Entstehungsaufwand, hier in NLOC (netto lines of code).
NLOC	Netto Lines of Code, Anzahl der Code-Zeilen ohne Kommentar und Leerzeilen.

Bei einem guten Entstehungsprozess streben überwachte Parameter oft gegen eine Normalverteilung. Polarisierungen (mehrere Gipfel) können wichtige Informationen über Schwachstellen und Ansatzmöglichkeiten für Verbesserungen liefern:

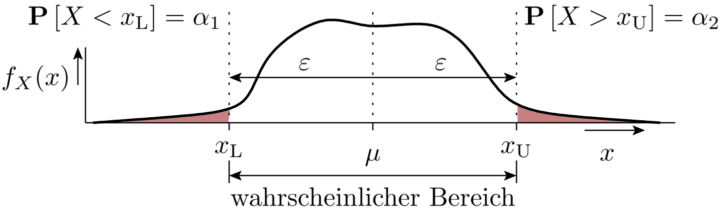
- Abhängigkeiten bei der Fehlerentstehung, bei Ausfällen beim Fehlernachweis und beim Versagen von Service-Leistungen,
- Vorliebe oder Neigung befragter Experten, z.B. bei der Einschätzung von Gefährdungen und Risiken,
- Probleme eines Messverfahrens, ...

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 schlechter Funktionierende zu eliminieren.

Spezielle Form des Lernen aus Fehlern.

Tschebytscheffsche Ungl.

Bereichsschätzung, wenn Verteilung unbekannt

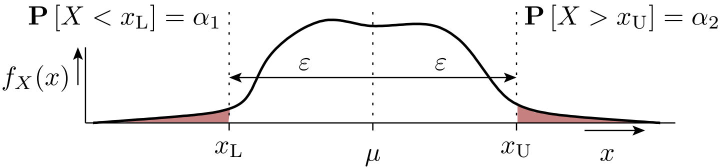


Die Bestimmung eines wahrscheinlichen Intervalls $[x_{\min}, x_{\max}]$

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

x_L, x_U	Untere und oberen Schranke des wahrscheinlichen Bereichs von X.
α_1, α_2	Irrtumswahrscheinlichkeit, Wert unterhalb bzw. oberhalb des geschätzten Bereichs.
$\mathbb{P}[\ldots]$	Eintrittswahrscheinlichkeit des Ereignisses ...
$f_X(x)$	Dichtefunktion der Zufallsvariablen X.
μ	Erwartungswert.
ε	Intervallradius, Abstand zwischen Bereichsgrenzen und Erwartungswert.

Das schwache Gesetz der großen Zahlen



Nach der tschebytscheffschen Ungleichung ist die Irrtumswahrscheinlichkeit α, das der Wert einer Zufallsvariablen mehr als ein Intervallradius ε von seinem Erwartungswert abweicht, nicht größer als das Verhältnis von Varianz und Intervallradius ε:

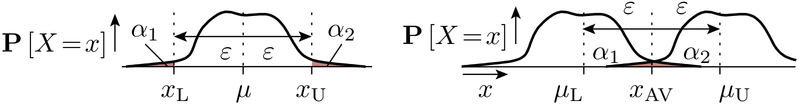
$$\mathbb{P}\left[|X-\mu| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2} \tag{4.77}$$

Intervallradius zur Irrtumswahrscheinlichkeit $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$:

$$\varepsilon \leq \sqrt{\frac{\sigma^2}{\alpha}} = \frac{\sigma}{\sqrt{\alpha}} \tag{4.78}$$

X	Zufallsvariable.
ε, μ, σ	Intervallradius, Erwartungswert, Standardabweichung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.

Bereichsschätzung



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

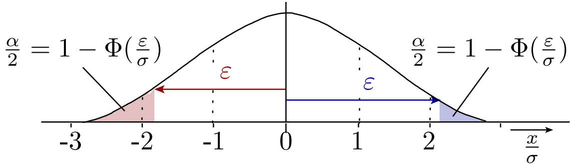
$$\mathrm{sr} = [x_L, x_U] = \mu \mp \frac{\sigma}{\sqrt{\alpha}} \tag{4.79}$$

Wahrscheinlicher Bereich des Erwartungswerts bei einer bekannten Realisierung x_{AV}:

$$\mathrm{sr}_\mu = [\mu_L, \mu_U] = x_{AV} \mp \frac{\sigma}{\sqrt{\alpha}} \tag{4.80}$$

ε, μ, σ	Intervallradius, Erwartungswert, Standardabweichung.
α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
μ_L, μ_U	Untere und obere Schranke des wahrscheinlichen Bereichs des Erwartungswerts.
x_L, x_U	Untere und oberen Schranke des wahrscheinlichen Bereichs von X.
sr	Symmetrischer Bereich der wahrscheinlichen Werte.
sr_μ	Symmetrischer Bereich des wahrscheinlichen Erwartungswerts.
x_{AV}	Experimentell bestimmter Ist-Zählwert, Schätzwert für den Erwartungswert.

Vergleich mit Intervallradius Normalverteilung



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

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

Intervallradius Tschebytscheffsche Ungleichung:

(4.78) $$\varepsilon \le \sqrt{\frac{\sigma^2}{\alpha}} = \frac{\sigma}{\sqrt{\alpha}}$$

$$\varepsilon \le \sigma \cdot \alpha^{-0,5}$$

α	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
$\alpha^{-0,5}$	4,68	19,60		5	7,07	10	15,8	22,4

Beispiel 4.15: Tschebytscheffschen Ungleichung

Aus eine Stichprobe gemessener Widerstandswerte in kΩ soll auf den möglichen Bereich des Erwartungswertes geschlussfolgert werden. Zugelassene Irrtumswahrscheinlichkeit 2%.

$\alpha = 2\%$, Wertestichprobe: $R_i:\ 10{,}3,\ 10{,}5,\ 9{,}7,\ 8{,}9,\ 10{,}1,\ 11{,}0,\ 10{,}2,\ 9{,}5$

a) *Ohne Kenntnisse der Verteilung über die Tschebytscheffsche Ungleichung?*

b) *Unter der Annahme, dass die Widerstandswerte normalverteilt sind?*

α	Irrtumswahrscheinlichkeit Werte außerhalb des geschätzten Bereichs.
R_i	Widerstandswerte in kΩ.

$\alpha = 2\%$, Wertestichprobe: R_i : $10{,}3,\ 10{,}5,\ 9{,}7,\ 8{,}9,\ 10{,}1,\ 11{,}0,\ 10{,}2,\ 9{,}5$

a) *Ohne Kenntnisse der Verteilung über die Tschebytscheffsche Ungleichung?*

(4.13) $\hat{\mathbb{E}}[X] = \hat{\mu} = \frac{1}{\#v} \cdot \sum_{i=1}^{\#v} v_i$

(4.14) $\hat{\operatorname{Var}}[X] = \hat{\sigma}^2 = \frac{1}{\#v-1} \cdot \sum_{i=1}^{\#v} (v_i - \hat{\mu})^2$

(4.79) $\operatorname{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \mu \mp \frac{\sigma}{\sqrt{\alpha}}$

$$\hat{\mu} = \frac{1}{8}(10{,}3 + \ldots)\ \mathrm{k\Omega} = 10{,}025\,\mathrm{k\Omega}$$

$$\hat{\sigma} = \sqrt{\frac{1}{7}\left((10{,}3 - 10{,}025)^2 + \ldots\right)\ \mathrm{k\Omega}^2} = 647\,\Omega$$

Bereich des Erwartungswerts:

$$\operatorname{sr}(R) = 10{,}025\,\mathrm{k\Omega} \mp \frac{647\,\Omega}{\sqrt{2\%}} = [5{,}3\,\mathrm{k\Omega},\ 14{,}8\,\mathrm{k\Omega}]$$

$\alpha = 2\%$, Wertestichprobe: R_i : $10{,}3,\ 10{,}5,\ 9{,}7,\ 8{,}9,\ 10{,}1,\ 11{,}0,\ 10{,}2,\ 9{,}5$

b) *Unter der Annahme, dass die Widerstandswerte normalverteilt sind?*

(4.55) $$\mathrm{sr} = [x_{\mathrm{L}}, x_{\mathrm{U}}] = \mu \mp \sigma \cdot \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)$$

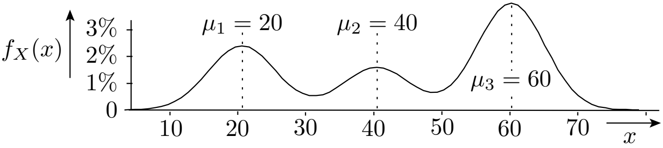
α	0,26%	0	4%	2%	1%	0,4%
$\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$	3	4	2,05	2,33	2,57	2,88

$$\begin{aligned} \mathrm{sr}\,(R) &= \hat{\mu} \mp \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \hat{\sigma} \\ &= 10{,}025\,\mathrm{k}\Omega \mp 2{,}33 \cdot 647\,\Omega \\ &= [8{,}5\,\mathrm{k}\Omega,\ 11{,}5\,\mathrm{k}\Omega] \end{aligned}$$

Weniger als halb so breiter Bereich im Vergleich zu $[5{,}3\,\mathrm{k}\Omega,\ 14{,}8\,\mathrm{k}\Omega]$ aus Aufgabenteil a »*ohne Kenntnis der Verteilung*«.

Zusammenfassung

Misch- und multimodale Verteilungen



Mischung diskreter Verteilungen:

(4.73) $$\mathbb{P}[X = x_i] = \sum_{j=1}^{\#Y} h_j \cdot \mathbb{P}(X_j = x_i)$$

Mischung stetiger Verteilungen:

(4.74) $$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \sum_{i=1}^{\#Y} h_j \cdot f_{X_j}(x)$$

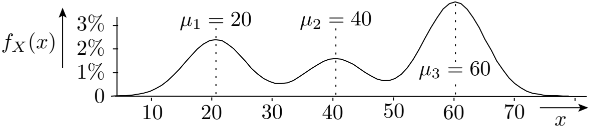
Erwartungswert:

(4.75) $$\mathbb{E}[X] = \mu = \sum_{j=1}^{\#Y} h_j \cdot \mu_j$$

Varianz:

(4.76) $$\mathrm{Var}[X] = \sigma^2 = \sum_{j=1}^{\#Y} h_j \cdot \sigma_j^2 + \sum_{j=1}^{\#Y} h_j \cdot \delta_j^2$$

Multimodalität



Bei im Verhältnis zur Standardabweichung großen Abweichungen der Erwartungswerte entsteht Multimodalität, z.B.:

- Für die Anzahl der nachweisbaren Fehler bei einem Zufallstest und großen Fehlerteilmengen, die (fast) gleich nachweisbar sind.
- Mischung von Objekten aus unterschiedlichen oder sich ändernden Entstehungsprozessen.

Die Maxima liefern Hinweise auf Verbesserungsmöglichkeiten

- für Entstehungsprozesse zur Fehlervermeidung,
- für die Bewertung erfasster Daten, z.B. Erkennen unerwartete Abhängigkeiten zwischen Zählwerten und Vorlieben von Experten bei Befragungen,
- für Messverfahren, ...

Bereichsschätzung

Bereichsschätzung für beliebige incl. multimodale Verteilungen:

- tschebytscheffschen Ungleichung:

(4.77) $$\mathbb{P}\left[|X-\mu| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2}$$

- garantierbarer Intervallradius:

(4.78) $$\varepsilon \leq \sqrt{\frac{\sigma^2}{\alpha}} = \frac{\sigma}{\sqrt{\alpha}}$$

- garantierbarer Bereiche für Realisierungen und Erwartungswerte:

(4.79) $$\mathrm{sr} = \left[x_{\mathrm{L}}, x_{\mathrm{U}}\right] = \mu \mp \frac{\sigma}{\sqrt{\alpha}}$$

(4.80) $$\mathrm{sr}_{\mu} = \left[\mu_{\mathrm{L}}, \mu_{\mathrm{U}}\right] = x_{\mathrm{AV}} \mp \frac{\sigma}{\sqrt{\alpha}}$$

Pareto-Verteilung

Pareto-Prinzip*

Ein kleiner Teil der Ursachen verursacht oft Mehrheit der Probleme:

- ein kleiner Teil der Fehler die Mehrheit der Fehlfunktionen,
- ein kleiner Teil der Fehlfunktionen den meisten Schaden,
- ein kleiner Teil der Tests weist die Mehrheit der Fehler nach.

Auch nach der Beseitigung der dominanten Ursachen gibt es in der Regel neue dominante Ursachen.

Ein bereits behandelten Beispiel ist die Verteilung der erforderlichen Testanzahl X für den Nachweis eines zufälligen Fehlers mit zufälligen Eingaben. Die Wahrscheinlichkeit, dass die erforderliche Nachweislänge nicht größer N ist, ist die zu erwartende Fehlerabdeckung mit N Tests nach (Gl. 4.87)

$$F_X(N) = \mathbb{P}[X \leq N] = \mu_{\mathrm{FC}}(N) = 1 - \left(\frac{N}{N_0}\right)^{-K}$$

N_0	Skalenparameter, hier Testanzahl, für die erkennbare Fehler bereits beseitigt sind.
$K > 0$	Formfaktor der Pareto-Verteilung.
*	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

Die Pareto-Verteilung

$$X \sim \operatorname{Par}(K, x_{\min})$$

ist die Verallgemeinerung auf eine stetige Zufallsgröße z.B. die Zeit bis zum erstmaligem Auftreten eines Problems oder die Schadenshöhe. Verteilungsfunktion:

$$F_X(x) = \mathbb{P}[X \le x] = \begin{cases} 0 & x \le x_{\min} \\ 1 - \left(\frac{x_{\min}}{x}\right)^K & \text{sonst} \end{cases} \quad (4.81)$$

Dichtefunktion für $x \ge x_{\min}$:

$$f_X(x) = \frac{K \cdot x_{\min}^K}{x^{K+1}} \quad (4.82)$$

$K > 0$ Formfaktor der Pareto-Verteilung.
$x_{\min} > 0$ Skalenparameter der Pareto-Verteilung.

Eigenschaften

Eigenschaften der Pareto-Verteilung

- Einen Erwartungswert

$$\mathbb{E}\left[X\right] = \int_{x_{\min}}^{\infty} \frac{K \cdot x_{\min}^{K}}{x^{K+1}} \cdot x \cdot \mathrm{d}x = \frac{K \cdot x_{\min}^{K}}{1-K} \cdot \left(\lim_{x\to\infty} x^{1-K} - x_{\min}^{1-K}\right)$$

hat eine pareto-verteilte Zufallsvariable nur für $K > 1$:

$$\mathbb{E}\left[X\right] = \mu = x_{\min} \cdot \frac{K}{K-1} \tag{4.83}$$

- Eine Varianz existiert nur für $K > 2$:

$$\mathrm{Var}\left[X\right] = \sigma^2 = x_{\min}^2 \cdot \frac{K}{(K-2)(1-K)^2} \tag{4.84}$$

Für kleine Exponenten, z.B. für die Fehlernachweislänge mit $0 < K < 1$, hat sie nicht einmal einen Erwartungswert.

Formfaktor für die Pareto-20%-80%-Regel

Der Anteil der Ursachen U mit der größten Wirkung:

$$U = \int_{w_{\min}}^{\infty} f(x) \cdot \mathrm{d}x = \int_{w_{\min}}^{\infty} \frac{K \cdot x_{\min}^K}{x^{K+1}} \cdot \mathrm{d}x = \left(\frac{x_{\min}}{w_{\min}}\right)^K \qquad (4.85)$$

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

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

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

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

$$W = \frac{\mathbb{E}[X|X \geq w_{\min}]}{\mathbb{E}[X]} = \left(\frac{x_{\min}}{w_{\min}}\right)^{K-1} = U^{\frac{K-1}{K}} \qquad (4.86)$$

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

Fehlernachweislänge

Verteilung der Fehlernachweislänge

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 (siehe Abschn. 2.2.2):

(3.11) $$\mu_{\mathrm{FC}} = 1 - \left(\frac{N_2}{N_1}\right)^{-K}$$

K	1	0,5	0,33	0,25
$\frac{N}{N_0}$ für $1 - \mu_{\mathrm{FC}}(N) = 0{,}1$	10	100	10^3	10^4

$\mu_{\mathrm{FC}}(N)$ Zu erwartende Fehlerabdeckung in Abhängigkeit von der Testanzahl.
N_0 Testanzahl, für die vorher alle Fehler beseitigt wurden, also für $FC = 0$.
N Anzahl der Tests, für die erkannten Fehler beseitigt werden, incl. N_0.

Daraus resultiert, dass die Nachweislänge X bei Abnahme des Anteils der nicht nachweisbaren Fehler nach Gl. 3.11 pareto-verteilt ist mit N_0 als Skalenparameter und K als Formfaktor:

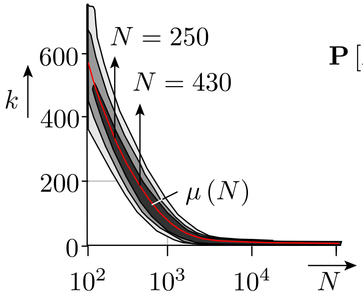
$$X \sim \mathrm{Par}\left(K, N_0\right)$$
$$F_X(N) = \mu_{\mathrm{FC}}(N) = 1 - \left(\frac{N_0}{N}\right)^K \quad \text{für } N \geq N_0 \tag{4.87}$$

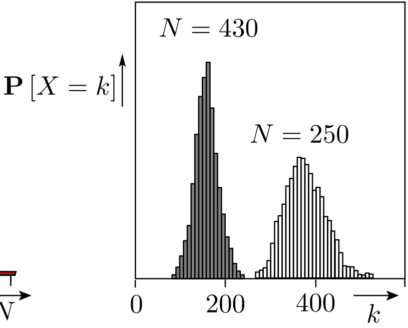
Wegen $K < 1$ hat die Nachweislänge keinen Erwartungswert. Beim Betrieb von IT-System mit pareto-verteilter Nachweislänge und Beseitigung aller erkennbaren Fehler sind auch nach sehr langer Nutzungsdauer weitere Fehler nicht ausschließbar.

Es folgt ein Experiment zu Untersuchung, wie gut eine Pareto-Verteilung die Verteilung der Nachweislänge von Fehlern in IT-Systemen annähert.

$F_X(N)$	Verteilungfunktion der Nachweislänge.
$\mu_{\mathrm{FC}}(N)$	Zu erwartende Fehlerabdeckung in Abhängigkeit von der Testanzahl.
N_0	Testanzahl, für die vorher alle Fehler beseitigt wurden, also für $FC = 0$.
N	Anzahl der Tests, für die erkannten Fehler beseitigt werden, incl. N_0.
K	Formfaktor der Dichte der Fehlfunktionsrate $(0 < K < 1)$.

Für das Haftfehlerexperiment



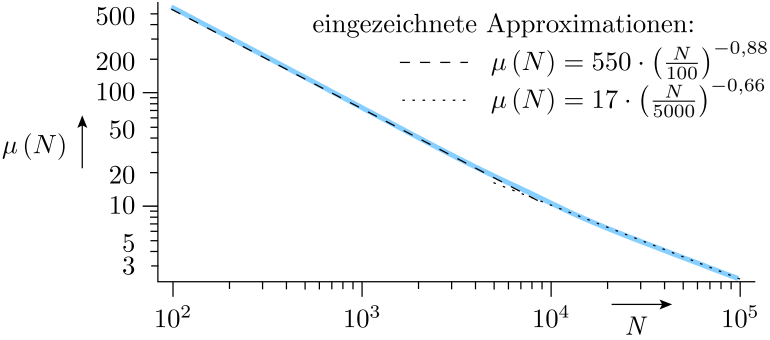


Kombinatorische Beispielschaltung (Benchmark c3540). 3606 simulierte, unterschiedlich nachweisbare Haftfehler. Zu erwartende Anzahl der nicht nachweisbaren Fehler:

$$\mu(N) = 3606 \cdot (1 - \mu_{\mathrm{FC}}(N))$$

$\mu(N)$	Zu erwartende Anzahl nicht nachweisbaren Modellfehler als Funktion der Testanzahl N.
k	Anzahl nicht nachweisbaren Modellfehler.

Pareto-Näherung



Formfaktor $K = 0{,}88$ nähert den Bereich $N \in \left[10^2, 10^4\right]$ und $K = 0{,}66$ den Bereich bis $N \in \left[10^4, 10^5\right]$ besser an. Aber Achtung, für $N \geq 10^4$ sind die Zählwerte sehr klein und der Schätzfehler für den Erwartungswert $\mu(N)$ und damit auch für den Formfaktor K groß.

$\mu(N)$	Zu erwartende Anzahl nicht nachweisbaren Modellfehler als Funktion der Testanzahl N.
$K > 0$	Formfaktor der Pareto-Verteilung.

Schaden durch MF

Schaden durch Fehlfunktionen

Die möglichen Schäden durch Fehlfunktionen von IT-Systemen sind vom Einsatz abhängig und reichen von »unerheblich« über sehr hoch (Verlust großer Datenmengen) bis zu unbezahlbaren Katastrophen (Krieg mit Atomwaffen (siehe Folie Der Preis fehlender Verlässlichkeit, Abschn. 1).

Anschaulich gilt das Pareto-Prinzip, dass ein kleiner Teil der MF den überwiegenden Teil des Schadens verursacht.

Mangels verfügbaren Schadensstatistik für IT-Fehlfunktionen betrachten wir die Verteilung von Haftpflichtschäden einer Autoversicherung.

Verteilung von Haftpflichtschäden

Haftpflichtschäden über 100.000 SF einer Schweizer Autoversicherung*:

```
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
```

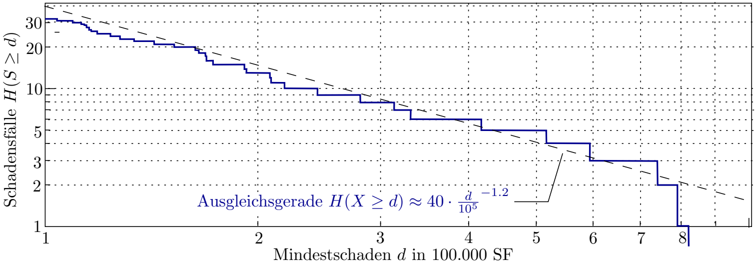
- Anzahl der Schadensfälle: 33
- Gesamtschadenssumme: 9.458.208 SF

SF Schadenskosten in Schweizer Franken.

* 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

Anzahl der Versicherungsfälle mit einem Schaden größer s:



Pareto-Verteilung der Schadenshöhe ab $d_{\min} = 10.000$ SF::

$$F_X(d) = \mathbb{P}[X \leq d] = 1 - \left(\frac{d}{d_{\min}}\right)^{-K} = 1 - \left(\frac{d}{10^5}\right)^{-1,2}$$

$H(X \geq d)$	Anzahl der Schadensfälle mit einem Schadenskosten größer d.
d	Schadenskosten in Schweizer Franken.
$d_{\min}$	Betrachteter Mindestschaden, Skalenparameter der Pareto-Verteilung.

Erwartungswert (Gl. 4.83):

$$\mu_{\mathrm{d}} = d_{\min} \cdot \frac{K}{K-1} = \frac{1{,}2}{1{,}2-1} \cdot d_{\min} = 600.000\,\mathrm{SF}$$

Eine Pareto-Verteilung hat erst für $K > 2$ eine Varianz. Alle im Kapitel behandelten Bereichsschätzungen incl. über die Tschebytscheffsche-Ungleichung nicht anwendbar. Abschätzung, wie viel Geld ein Versicherungsunternehmen als Rücklage haben muss, um jeden Schaden erstatten zu können, schwierig. Vermutlich haben Versicherungen deshalb eine max. Deckungssumme.

Der Autor geht davon aus, dass künftig Schäden durch IT z.B. in autonomen Fahrzeugen ähnlich wie heute Haftpflichschäden durch Personen versichert werden.

μ_{d}	Zu erwartende Schadenskosten in Schweizer Franken.
$d_{\min}$	Betrachteter Mindestschaden, Skalenparameter der Pareto-Verteilung.
$K > 0$	Formfaktor der Pareto-Verteilung.
SF	Schadenskosten in Schweizer Franken.

Zusammenfassung

Verteilungsfunktion und Dichte der Pareto-Verteilung:

(4.81) $$F_X(x) = \mathbb{P}[X \leq x] = \begin{cases} 0 & x \leq x_{\min} \\ 1 - \left(\frac{x_{\min}}{x}\right)^K & \text{sonst} \end{cases}$$

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

Erwartungswert für $k > 1$:

(4.83) $$\mathbb{E}[X] = \mu = x_{\min} \cdot \frac{K}{K-1}$$

Varianz für $k > 2$:

(4.84) $$\mathrm{Var}[X] = \sigma^2 = x_{\min}^2 \cdot \frac{K}{(K-2)(1-K)^2}$$

Die Pareto-20-80-Regel beschreibt den Sonderfall $K = 1{,}16$.

Für Überschläge kann die Fehlernachweislänge als pareto-verteilt mit $0 < K < 1$ betrachtet werden. Kein Erwartungswert:

(4.87) $$F_X(N) = \mu_{\mathrm{FC}}(N) = 1 - \left(\frac{N_0}{N}\right)^K \quad \text{für } N \geq N_0$$

Die Beispielsschadenszahlen ließen sich durch eine Pareto-Verteilung mit $K = 1{,}2$ annähern, d.h. Erwartungswert ja, Varianz nein.