Test und Verlässlichkeit Foliensatz 3: Verteilungen

Prof. G. Kemnitz

June 10, 2020

Contents

Verteilungen

Verlässlichkeit wird durch eine Vielzahl von Kenngrößen beschrieben, die wir nicht exakt angeben können: Zählwerte für Fehler, FF, ... Überdeckungen, ... Mathematisch gesehen sind das Zufallsgrößen. Zweiwertige Zufallsgrößen, bei denen ein betrachtetes Ereignis eintreten oder nicht eintreten kann, werden durch die Eintrittswahscheinlichkeit charakterisiert.

Für Zufallsgrößen mit mehr als zwei möglichen Ergebnisse interessiert uns der wahrscheinliche Bereich. Dieser Foliensatz vermittelt ein auf die Vorlesung abgestimmten Werkzeugkasten aus der Stochastik, um solche Bereiche abzuschätzen.

Grundlage ist die Untersuchung und Abschätzung geeigneter Verteilungen. Ein Verteilung ordnet den möglichen Werten einer Zufallsgröße Wahrscheinlichkeiten zu ...

1 Grundlagen

1.1 Charakteristische Größen

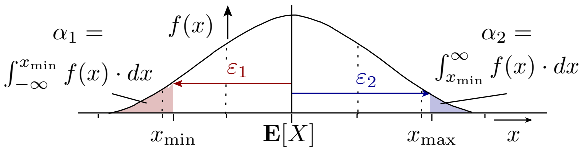
Charakteristische Größen einer Zufallsvariablen

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

Name	Definition
Verteilungsfunktion	$F_X(x) = \mathbb{P}[X \leq x]$
Dichtefunktion	$f_X(x) = \frac{dF_X(x)}{dx}$
Erwartungswert	$\mu = m_1 = \mathbb{E}[X]$
k-tes Moment	$m_k = \mathbb{E}\left[X^k\right]$
k-tes zentriertes Moment	$\mathbb{E}\left[(X - \mathbb{E}[X])^k\right]$
Varianz (2. zentr. M.)	$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$
Standardabweichung	$\mathrm{sd}[X] = \sqrt{\mathrm{Var}[X]}$
Vertrauensbereich*	$[x_{\min}, x_{\max}]$

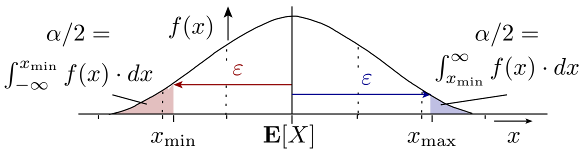
Wahrscheinlicher Bereich, $\mathbb{E}[X]$ und $\mathrm{sd}[X]$

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



$\alpha_1 = F(x_{\min})$	Irrtumswahrscheinlichkeit, dass Werte unterhalb des geschätzten Bereichs liegen.
$\alpha_2 = 1 - F(x_{\max})$	Irrtumswahrscheinlichkeit, dass Werte oberhalb des geschätzten Bereichs liegen.
$\varepsilon_{1/2}$	Intervallradius, Abstand der unteren / oberen Bereichsgrenze vom Erwartungswert.

Bei $\alpha_1 = 0$ / $\alpha_2 = 0$ wird nur eine Ober- / Untergrenze geschätzt.



Nach der tschebytscheffschen Ungleichung:

$$\alpha = \mathbb{P}[|x - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2} \tag{1}$$

ist die Wahrscheinlichkeit, das der Wert einer Zufallsgröße mehr als ein Intervallradius ε von seinem Erwartungswert abweicht, nicht größer als das Verhältnis der Varianz zum Quadrat des Intervallradius ε. Bei Zulassen einer Irrtumswahrscheinlichkeit α beträgt der Intervallradius maximal:

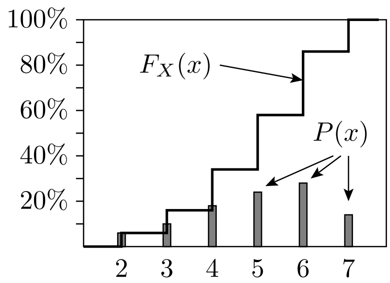
$$\varepsilon \leq \frac{\mathrm{sd}[X]}{\sqrt{\alpha}}$$

($\mathbb{E}[X]$ – Erwartungswert, $\mathrm{Var}[X]$ – Varianz; $\mathrm{sd}[X]$ – Standardabweichung der Zufassgröße X). Wenn mehr als ($\mathbb{E}[X]$ und $\mathrm{sd}[X]$ über X bekannt, ist der wahrscheinliche Bereich weiter einschränkbar.

Diskrete Verteilung

Zufallsgröße X kann nur (über-) abzählbare Werte x_i annehmen, z.B.:

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



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

Erwartungswert (mit den Auftrittswahrscheinlichkeiten gewichtete Mittelwert):

$$\mathbb{E}[X] = \sum_{i=1}^{m} p_i \cdot x_i \tag{2}$$

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

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

Varianz (2. zentriertes Moment):

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

Für das Beispiel:

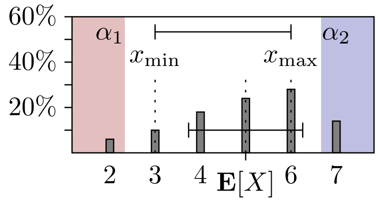
$$6\% \cdot (2-5)^2 + 10\% \cdot (3-5)^2 + \ldots + 14\% \cdot (7-5)^2 = 1{,}96$$

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

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

Für das Beispiel:

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



Bereich, in dem X mit 80% Wahrscheinlichkeit liegt

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

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

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

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

Verschiebungssatz

Die Varianz ist gleichfalls die Differenz aus dem Erwartungswert der Quadrate und dem Quadrat des Erwartungswertes[1]:

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \tag{3}$$

Herleitung:

$$\textstyle\sum_{i=1}^{m} p_i \cdot (x_i - \mathbb{E}[X])^2 = \sum_{i=1}^{m} p_i \cdot \left(x_i^2 - 2 \cdot x_i \cdot \mathbb{E}[X] + \mathbb{E}[X]^2\right)$$

$$\underbrace{\sum_{i=1}^{m} p_i \cdot x_i^2}_{\mathbb{E}[X^2]} + \mathbb{E}[X] \cdot \left(\mathbb{E}[X] \cdot \underbrace{\sum_{i=1}^{m} p_i}_{1} - 2 \cdot \underbrace{\sum_{i=1}^{m} p_i \cdot x_i}_{\mathbb{E}[X]} \right)$$

Für das Beispiel zuvor:

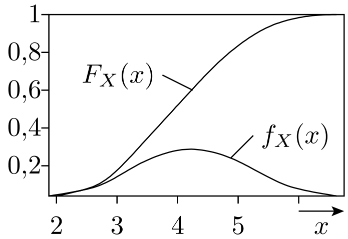
$$\mathrm{Var}[X] = 6\% \cdot 2^2 + 10\% \cdot 3^2 + 18\% \cdot 4^2 + 24\% \cdot 5^2 + 28\% \cdot 6^2 + 14\% \cdot 7^2 - 5^2 = 1{,}96$$

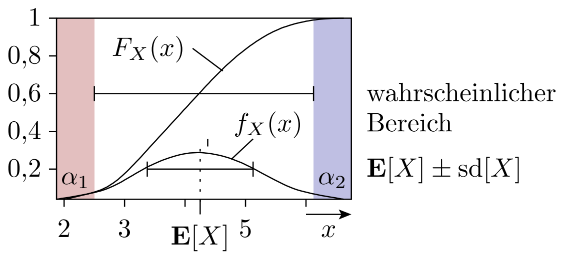
[1] Bei begrenzter Rechengenauigkeit u.U. numerisch problematisch.

Stetige Verteilungen

Zufallsvariable X ist stetig und hat in jedem Intervall a ≤ X ≤ b unendlich viele Ausprägungen. Beschreibung durch die Dichte:

$$f_X(x) = \frac{dF_X(x)}{dx}$$
$$F_X(x) = \int_{-\infty}^{x} f(u) \cdot du$$



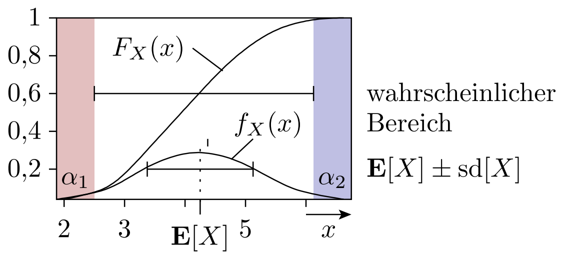


Erwartungswert:

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

Varianz:

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



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

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

Bereichsgrenzen:

$$x_{\min} = F^{-1}(\alpha_1)$$
$$x_{\max} = F^{-1}(1 - \alpha_2)$$

Erwartungswert und Varianz einer Datenstichprobe

Für eine Datenstichprobe einer Zufallsgröße X

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

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

$$\hat{\mathbb{E}}[X] = \bar{w} = \frac{1}{\#w} \cdot \sum_{i=1}^{\#w} w_i \tag{4}$$

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

$$\hat{\mathrm{Var}}[X] = \frac{1}{\#w - 1} \cdot \sum_{i=1}^{\#w} \left(w_i - \hat{\mathbb{E}}[X]\right)^2 \tag{5}$$

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

1.2 Lineare Transformationen, ...

Lineare Transformation

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

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

Bei der Varianz entfällt die Verschiebung und der Skalierungsfaktor geht im Quadrat ein[2]:

$$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X] \tag{6}$$

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

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

Kontrolle am Beispiel

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

$$\begin{aligned}
\mathbb{E}[X] &= 0{,}3 + 1 + 0{,}6 = 1{,}9 \\
\mathrm{Var}[X] &= 0{,}3 + 2 + 1{,}8 - 1{,}9^2 = 0{,}49 \\
\mathbb{E}[Y] &= 0{,}9 + 0{,}5 - 0{,}2 = 1{,}2 \\
\mathrm{Var}[Y] &= 2{,}7 + 0{,}5 + 0{,}2 - 1{,}2^2 = 1{,}96 \\
\mathbb{E}[Y] &= 5 - 2 \cdot \mathbb{E}[X] \\
\mathrm{Var}[Y] &= (-2)^2 \cdot \mathrm{Var}[X]
\end{aligned}$$

[2]Die Kontrolle der Gleichung ist eine Übungsaufgabe.

Summe von Zufallsgrößen

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

x	1	3	4
$f_X(x)$	0,1	0,4	0,5

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

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

$$\begin{aligned}
\mathbb{P}[X+Y=3] &= \mathbb{P}[X=1]\cdot\mathbb{P}[Y=2] \\
\mathbb{P}[X+Y=4] &= \mathbb{P}[X=1]\cdot\mathbb{P}[Y=3] \\
\mathbb{P}[X+Y=5] &= \mathbb{P}[X=1]\cdot\mathbb{P}[Y=4]+\mathbb{P}[X=3]\cdot\mathbb{P}[Y=2] \\
\mathbb{P}[X+Y=6] &= \mathbb{P}[X=3]\cdot\mathbb{P}[Y=3]+\mathbb{P}[X=4]\cdot\mathbb{P}[Y=2] \\
\mathbb{P}[X+Y=7] &= \mathbb{P}[X=3]\cdot\mathbb{P}[Y=4]+\mathbb{P}[X=4]\cdot\mathbb{P}[Y=3] \\
\mathbb{P}[X+Y=8] &= \mathbb{P}[X=4]\cdot\mathbb{P}[Y=4]
\end{aligned}$$

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

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

Die Varianz ist die Summe der Varianzen plus doppelte Kovarianz:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\cdot\mathrm{Cov}[X,Y] \tag{7}$$

mit der Kovarianz[3]:

$$\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\mathbb{E}[X])\cdot(Y-\mathbb{E}[Y])] \tag{8}$$

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

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

Gemessener Wert und Messfehler

In der Messtechnik gilt für jeden gemessenen Wert:

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

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

$$\begin{aligned}
\mathbb{E}[X_\mathrm{M}] &= \mathbb{E}[X_\mathrm{F}] + \mathbb{E}[X] \\
\mathrm{Var}[X_\mathrm{M}] &= \mathrm{Var}[X_\mathrm{F}] + \mathrm{Var}[X]
\end{aligned}$$

- $\mathbb{E}[X_\mathrm{F}]$ – Maß für den systematischen Messfehler
- $\mathrm{sd}[X_\mathrm{F}] = \sqrt{\mathrm{Var}[X_\mathrm{F}]}$ – Standardabweichung und Maß für den zufälligen Messfehler.

Beispielaufgabe

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

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

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

[3] Die Kontrolle der Gleichungen sind Übungsaufgaben.

1.3 Verteilung von Zählwerten

Verteilung von Zählwerten

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

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

»potentieller Zählwerte« X_i mit der Bernoulli-Verteilung:

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

beschreiben.

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

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

Der Erwartungswert der Einzelereignisse ist

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

Varianz nach Verschiebungssatz:

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

Der Erwartungswert der Summe ist die Summe der Erwartungswerte:

$$\mathbb{E}[X] = \sum_{i=1}^{\#X} p_i \tag{9}$$

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

$$\mathrm{Var}[X] = \sum_{i=1}^{\#X} p_i \cdot (1 - p_i) \tag{10}$$

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

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

Berechnung der Verteilung:

i	p_i	$X=0$	$X=1$	$X=2$	$X=3$	$X=4$
1	30%	70%	30%			
2	50%	35%	50%	15%		
3	40%	21%	44%	29%	6%	
4	10%	18,9%	41,7%	30,5%	8,3%	0,6%

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

`Wiederhole für` $i = 2$ `bis` N

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

`Wiederhole für` $k = 1$ `bis` $i - 1$

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

(i – Anzahl der berücksichtigten Summanden; k – Zählwert).

Erwartungswert und Varianz für das Beispiel

i	p_i	$X=0$	$X=1$	$X=2$	$X=3$	$X=4$
1	30%	70%	30%			
2	50%	35%	50%	15%		
3	40%	21%	44%	29%	6%	
4	10%	18,9%	41,7%	30,5%	8,3%	0,6%

Nach Gl. 2 beträgt der 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. 9 ist die Berechung kürzer:

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

Die Varianz beträgt nach dem Verschiebungssatz Gl. 3:

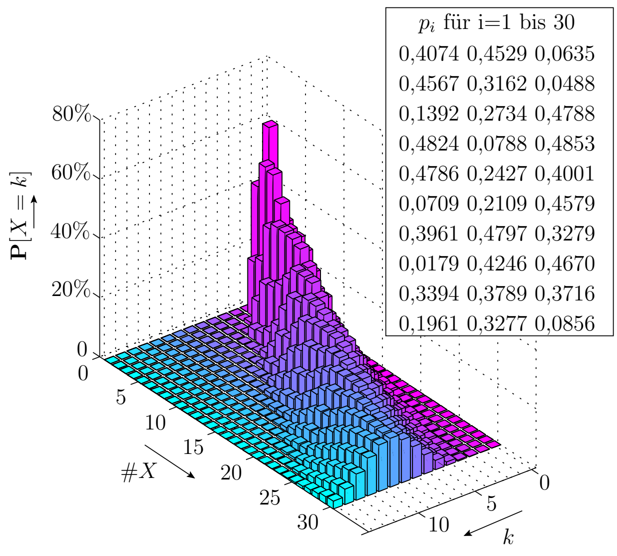
$$18{,}9\% \cdot 0^2 + 41{,}7\% \cdot 1^2 + 30{,}5\% \cdot 2^2 + 8{,}3\% \cdot 3^2 + 0{,}6\% \cdot 4^2 - 1{,}3^2 = 0{,}79$$

Die vereinfachte Berechnung nach Gl. 10:

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

Beispiel einer Zählverteilung

Das nachfolgende Säulendiagramm zeigt eine mit Matlab schrittweise berechnete Zählverteilung. Die Eintrittswahrscheinlichkeiten der Zählereignisse siehe Kasten im Bild. Erwartungswert und Varianz für alle 30 Summanden betragen $\mathbb{E}[X]$ =7,05, $\mathrm{Var}[X] = 2{,}19$:



2 Näherungen für ZV

2.1 Binomialverteilung

Binomialverteilung

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

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

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

$$\mathbb{P}[X = k] = \begin{pmatrix} n \\ k \end{pmatrix} \cdot p^k \cdot (1-p)^{n-k} \tag{11}$$

Erwartungswert einer Binomialverteilung:

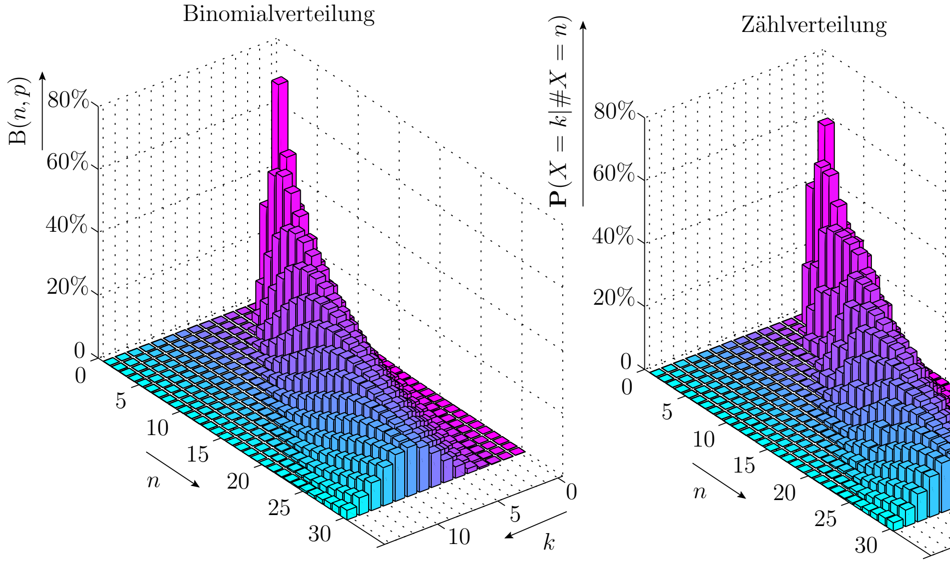
$$\mathbb{E}[X] = n \cdot p$$

Varianz und Standardabweichung einer Binomialverteilung:

$$\mathrm{Var}[X] = n \cdot p \cdot (1-p) \qquad (12)$$

$$\mathrm{sd}[X] = \sqrt{n \cdot p \cdot (1-p)} \qquad (13)$$

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

Beispielaufgabe

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

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

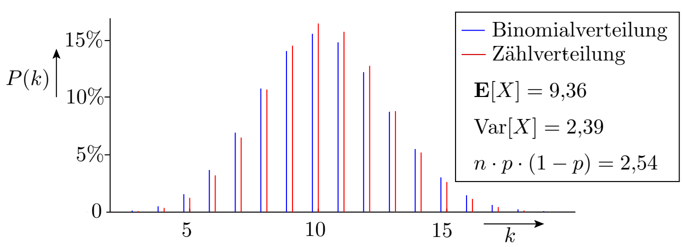
Varianzobergrenze

Satz

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

$$n \cdot p \cdot (1-p) \geq \mathrm{Var}[X] = \sum_{i=1}^{n} p_i \cdot (1-p_i) \qquad (14)$$

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



Beweis

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

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

Varianz der Zählverteilung:

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

Fakt 1. *Der über eine Binomialverteilung aus der mittleren Eintrittswahrscheinlichkeit*

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

berechnete wahrscheinliche Bereich für Zählergebnisse (Fehler, Fehlfunktionen, ...) ist einfacher zu berechnen, bei gleichen Irrtumswahrscheinlichkeiten garantiert größer bzw. bei gleicher Bereichsgröße sind die Irrtumswahrscheinlichkeiten kleiner.

2.2 Poisson-Verteilung

Poisson-Verteilung

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

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

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

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

Die Verteilung der Summe strebt gegen die Poisson-Verteilung:

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

Die Poisson-Verteilung

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

hat nur den Parameter λ, der die Summe der Eintrittswahrscheinlichkeiten, dass ein »potentieller Zählwert« eins, d.h. ein »echter Zählwert« und gleichzeitig Erwartungswert und Varianz ist:

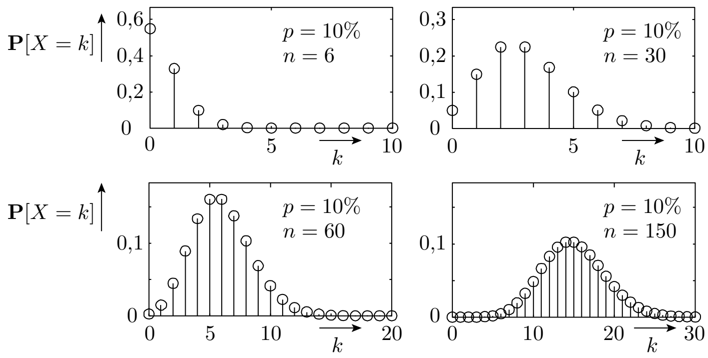
$$\mathbb{E}[X] = \mathrm{Var}[X] = \lambda = \sum_{i=1}^{\#X} p_i = n\cdot p$$

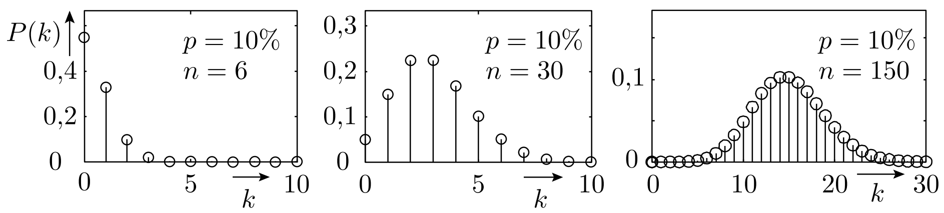
($n = \#X$ – Anzahl der potentiellen Zählwerte; p, p_i – mittlere bzw. individuelle Wahrscheinlickeit »Zählwert eins«).

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

Anzahl der Zählversuche und Verteilung

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





Grobabschätzung der wahrscheinlichen Bereiche:

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

$$k_{\mathrm{max}}>3\ldots5\cdot\mathbb{E}\left[X\right]$$

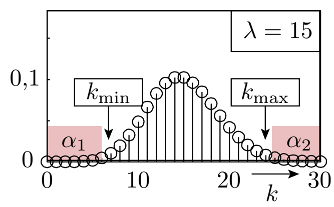
- Für $\mathbb{E}\left[X\right]\approx3\ldots10$ zusätzlich unter Schranke:

$$k_{\mathrm{max}}<\cdot\frac{\mathbb{E}\left[X\right]}{3\ldots5}$$

- Für $\mathbb{E}\left[X\right]>10$ (Normalverteilung günstiger):

$$\left[k_{\mathrm{min}},\,k_{\mathrm{max}}\right]\approx\mathbb{E}\left[X\right]\cdot\left(1\mp2\ldots4\right)$$

2.3 Bereichschätzung, Poisson



Schätzen von k_{min}

Vorgabe k_{min} und α_1. Numerische Suche $\lambda\left(k_{\mathrm{min}},\alpha_1\right)$, so dass

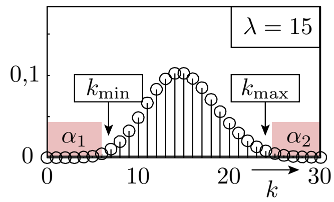
$\sum_{k=0}^{k_{\mathrm{min}}}e^{-\lambda}\cdot\frac{\lambda^k}{k!}\leq\alpha_1$

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

Beispielabschätzungen:

- $\lambda=7$ und $\alpha_1\leq1\%\Rightarrow k_{\mathrm{min}}=2$
- $k_{\mathrm{min}}=1$ und $\alpha_1=2\%\Rightarrow\lambda\geq3{,}912$

Schätzen von $k_{\max}$



Vorgabe $k_{\max}$ und α_2. Numerische Suche $\lambda(k_{\max}, \alpha_2)$, so dass

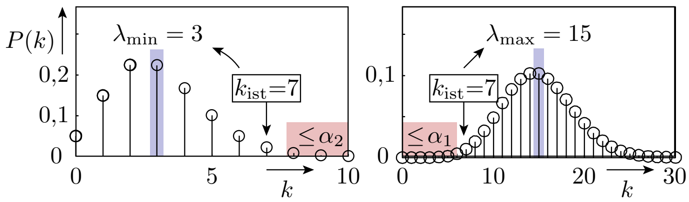
$\sum_{k=0}^{k_{\max}} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \geq 1 - \alpha_2$

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

Beispielabschätzungen:

- $\lambda = 2$ und $\alpha_2 \leq 1\% \Rightarrow k_{\max} = 5$
- $k_{\max} = 3$ und $\alpha_2 = 2\% \Rightarrow \lambda \leq 1{,}016$

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



Aus den Tabellen der beiden Folien zuvor ist ablesbar:

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

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

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

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

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

α_1	0,5%	1%	2%	10%	20%
$\lambda_{\max}$	5,30	4,61	3,91	2,30	1,61%

Abschätzungen einer FF-Rate

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

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

Abschätzbarer Bereich der FF-Rate:

$$\zeta_{\min} = \frac{\lambda_{\min}}{n} = 0{,}82 \cdot 10^{-5}$$
$$\zeta_{\max} = \frac{\lambda_{\max}}{n} = 8{,}41 \cdot 10^{-5}$$

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

Schätzen der Maskierungswahrscheinlichkeit

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

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

Abschätzbarer Bereich der Maskierungswahrscheinlichkeit:

$$p_{\text{F. min}} = \frac{\lambda_{\min}}{n} = 1{,}54 \cdot 10^{-4}$$
$$p_{\text{F. max}} = \frac{\lambda_{\max}}{n} = 12{,}6 \cdot 10^{-4}$$

Schätzen eines Zuverlässigkeitsbereichs

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

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

- Abschätzbarer Bereich der FF-Rate:

$$\zeta_{\min} = 3{,}89 \cdot 10^{-3}\,{}^{\text{FF}}/_{\text{SL}}$$
$$\zeta_{\max} = 9{,}28 \cdot 10^{-3}\,{}^{\text{FF}}/_{\text{SL}}$$

- Abschätzbarer Bereich der Zuverlässigkeit:

$$Z_{\min} = \frac{1}{\zeta_{\max}} = 108\,{}^{\text{SL}}/_{\text{FF}}$$
$$Z_{\max} = \frac{1}{\zeta_{\min}} = 257\,{}^{\text{SL}}/_{\text{FF}}$$

2.4 Normalverteilung

Normalverteilung

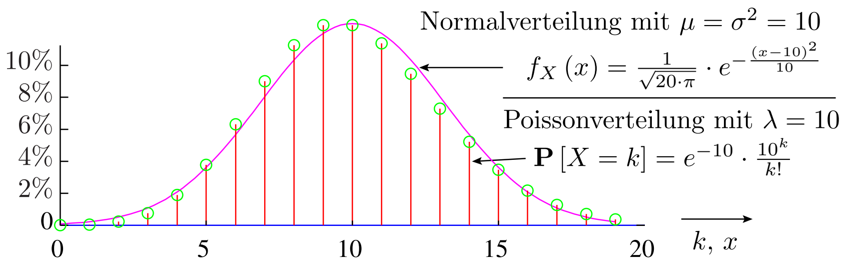
Die Summe sehr vieler unabhängiger Zufallsgrößen strebt unter sehr allgemeinen Bedingungen

- kein Summand hat dominanten Einfluss, ...

gegen eine Normalverteilung:

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

Beispiel: Poisson-Verteilung mit $\mathbb{E}[X] = \text{Var}[X] = \lambda = 10$:



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

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

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

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

2.5 Bereichschätzung NVT

Bereichsschätzung mit Normalverteilung

Die standardisierte Normalverteilung (Erwartungswert $\mu = 0$, Standardabweichung $\sigma = 1$). Verteilungsfunktion:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2} \cdot dz$$

Tabelliert für $z = 0$ bis 3,9 in Schritten von 0,1:

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 $z < 0$:

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

Transformation einer Zufallsgrößen X mit Erwartungswert μ und Standardabweichung σ in Zufallsgrößen Z mit Erwartungswert null und Standardabweichung eins:

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

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

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

Bestimmen der Irrtumswahrscheinlichkeiten:

- Transformation: $z_{\text{min}} = \frac{x_{\text{min}} - \mu}{\sigma}$ und $z_{\text{max}} = \frac{x_{\text{max}} - \mu}{\sigma}$
- Ablesen aus der Tabelle:

$$\alpha_1 = \Phi(z_{\text{min}}) = 1 - \Phi(-z_{\text{min}})$$
$$\alpha_2 = 1 - \Phi(z_{\text{max}})$$

Bestimmen des wahrscheinlichen Bereichs:

- Bestimme $z_{\text{min}} = -\Phi^{-1}(1 - \alpha_1)$ und $z_{\text{max}} = \Phi^{-1}(1 - \alpha_2)$.
- Transformation: $x_{\text{min}} = \sigma \cdot (\mu + z_{\text{min}})$, $x_{\text{max}} = \sigma \cdot (\mu + z_{\text{max}})$

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:

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

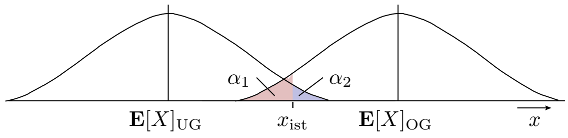
Beispieaufgaben: Zufallsgröße X, $\mu = 20$, $\sigma = 5$.

1. $\mathbb{P}[X \geq 30] \Rightarrow \mathbb{P}\left[Z \geq \frac{30-20}{5}\right] = 1 - \Phi(2) = 0{,}0228$
2. $\mathbb{P}[X \leq 15] \Rightarrow \mathbb{P}\left[Z \leq \frac{15-20}{5}\right] = \Phi(-1) = 1 - \Phi(1) = 0{,}1587$
3. $\alpha_2 \leq 1\% \Rightarrow z_{\text{max}} = \Phi^{-1}(1 - 1\%) = 2{,}33 \Rightarrow x_{\text{max}} = 20 + 2{,}33 \cdot 5 = 31{,}65$
4. $\alpha_1 \leq 2\% \Rightarrow z_{\text{min}} = \Phi^{-1}(2\%) = -\Phi^{-1}(1 - 2\%) = -2{,}05$ $x_{\text{min}} = 20 - 2{,}05 \cdot 5 = 9{,}75$

Bereichsschätzung für den Erwartungswert

Der Erwartungswert zu einem beobachteten Ereignis ist

- mindestens so groß, dass die Wahrscheinlichkeit, dass ein größeres als das beobachtete Ergebnis eintritt, kleiner α_2, und
- maximal so groß, dass ein kleineres als das beobachtete Ergebnis eintritt, kleiner α_1, ist.



Untere und obere Bereichsgrenze des Erwartungswertes:

$$\mathbb{E}[X]_{\text{UG}} = x_{\text{ist}} - \text{sd}[X] \cdot \Phi^{-1}(1 - \alpha_2)$$
$$\mathbb{E}[X]_{\text{OG}} = x_{\text{ist}} + \text{sd}[X] \cdot \Phi^{-1}(1 - \alpha_1)$$

Bereichsschätzung unabhängiger Zählwerte

Für die Summe unabhängiger Zählwerte

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

mit den Zweipunktverteilungen:

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

ist der Erwartungswert eine obere Schranke der Varianz:

$$\mathrm{Var}[X] \le \mathbb{E}[X] \cdot \left(1 - \frac{\mathbb{E}[X]}{\#X}\right) \le \mathbb{E}[X]$$

Unter- und Obergrenze des wahrscheinlichen Bereichs[4]:

$$k_{\min} \ge \mathbb{E}[X] - \sqrt{\mathbb{E}[X]} \cdot \Phi^{-1}(1 - \alpha_1)$$
$$k_{\max} \le \mathbb{E}[X] + \sqrt{\mathbb{E}[X]} \cdot \Phi^{-1}(1 - \alpha_2)$$

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

Für eine Irrtumswahrscheinlichkeit $\alpha = \alpha_1 + \alpha_2$ ist der Bereich $k_{\max} - k_{\min}$ für eine normalverteilte Zufallsgröße am kleinsten für $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ (Erwartungswert genau in der Bereichsmitte):

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

$k_{\min}$ und $k_{\max}$ vorheriger bzw. nächster ganzzahliger Wert:

$$k_{\min} \ge x_{\min}$$
$$k_{\max} \le x_{\max}$$

Bereichsschätzung der Anzahl der FF

Die zu erwartende Anzahl der Fehlfunktionen (FF) bei der Abarbeitung von $\#SL = 20.000\,\mathrm{SL}$ sei $\mu = 100\,\mathrm{FF}$. In welchem Bereich wird in 99% der Fälle die Anzahl der FF liegen ($\alpha_1 = \alpha_2 = 0{,}5\%$, keine Abhängigkeiten)?

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

$$100 - \sqrt{100} \cdot \Phi^{-1}(1 - 0{,}5\%) = 100 - 25{,}7 \le k_{\min} = 75$$
$$100 + \sqrt{100} \cdot \Phi^{-1}(1 - 0{,}5\%) = 100 + 25{,}7 \ge k_{\max} = 125$$

2.6 Varianzerhöhung

Varianzerhöhung durch Abhängigkeiten

Abhängigkeiten erhöhen die Varianz. Wenn z.B. zwei Zählereignisse immer paarweise gleichzeitig eintreten, werden halb so viele unabhängige Zufallsgrößen mit den möglichen Werten 0 und 2 aufsummiert:

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

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

Erwartungswert der Summanden:

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

Varianz der Summanden (nach Verschiebungssatz):

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

[4]Für seltene unabhängige Zählereignisse mit Erwartungswert $\mathbb{E}[X] \ge 10$.

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

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

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

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

Varianzerhöhung sei definiert als Verhältnis aus Varianz und Erwartungswert:

$$\kappa = \frac{\mathrm{Var}[X]}{\mathbb{E}(X)}$$

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

$$\kappa = n$$

Schätzen der Varianzerhöhung

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

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

- Schätzen der Varianz der Zählwertstichprobe:

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

- Varianzerhöhung:

$$\kappa = \frac{\hat{\mathrm{Var}}[X]}{\hat{\mathbb{E}}(X)}$$

Beispielabschätzung der Varianzerhöhung

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

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

- Erwartungswert der Zählwertstichprobe nach Gl. 4:

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

- Varianz der Zählwertstichprobe nach Gl. 5:

$$\hat{\mathrm{Var}}[X] = \frac{1}{9} \cdot \sum_{i=1}^{10} (w_i - 50{,}7)^2 = 248$$

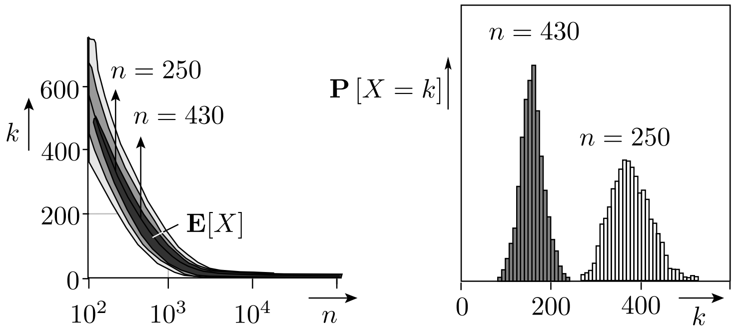
- Geschätzte Varianzerhöhung[5]:

$$\hat{\kappa} = \frac{248}{63{,}1} \approx 4$$

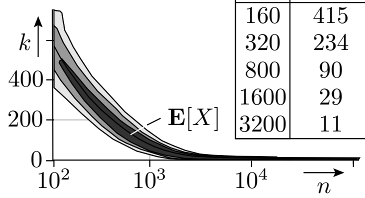
[5] Die Abhängigkeiten erhöhen die Varianz so, als ob 4 Zählereignisse fast immer gemeinsam eintreten.

Experiment mit Haftfehlern

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



Varianzerhöhung im Experiment



n	$\hat{\mathbf{E}}[X]$	$\hat{\mathbf{Var}}[X]$	$\hat{\mathbf{sd}}[X]$	$\kappa=\frac{\hat{\mathbf{Var}}[X]}{\hat{\mathbf{E}}[X]}$
160	415	1875	43,3	5,1
320	234	943	30,7	4,3
800	90	299	17,3	3,4
1600	29	52	7,2	1,8
3200	11	8,4	2,9	0,76

Zwischen den nicht nachweisbaren Fehlern gibt es offenbar Abhängigkeiten, die die Varianz so stark erhöhen, als ob 3...5 Modellfehler identisch Fehler nachweisbar wären. Identisch nachweisbare Fehler wurden jedoch nicht mitgezählt. Bleiben als Abhängigkeitsursache implizit nachweisbare Fehler sowie geteilte Steuer- und Beobachtungsbedingungen. Bei weniger nicht nachweisbaren Fehlern oder einer Fehlerstichprobe statt der kompletten Modellfehlermenge ist κ deutlich kleiner.

2.7 Bereichsschätzung Zählwerte

Bereichsschätzung normalverteilter Zählwerte

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

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

$$k_{\min}\geq\mathbb{E}[X]-\sqrt{\kappa\cdot\mathbb{E}[X]}\cdot\Phi^{-1}(1-\alpha_1)$$
$$k_{\max}\leq\mathbb{E}[X]+\sqrt{\kappa\cdot\mathbb{E}[X]}\cdot\Phi^{-1}(1-\alpha_2)$$

bzw. für $\alpha=\alpha_1+\alpha_2$:

$$k_{\min}\geq\mathbb{E}[X]-\sqrt{\kappa\cdot\mathbb{E}[X]}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$
$$k_{\max}\leq\mathbb{E}[X]+\sqrt{\kappa\cdot\mathbb{E}[X]}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

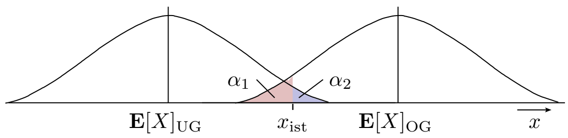
Annahme seltene Zählwerte mit Eintrittswahrscheinlichkeiten von im Mittel $p\ll 50\%$ und $\mathbb{E}[X]\geq\sqrt{\kappa}\cdot 10$. Für höhere Eintrittswahscheinlichkeit und $1\leq\frac{\mathbb{E}[X]}{\sqrt{\kappa}}\leq\#X-10$:

$$\ldots=\mathbb{E}[X]\mp\sqrt{\kappa\cdot\mathbb{E}[X]\cdot\left(1-\frac{\mathbb{E}[X]}{\#X}\right)}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

Bereichsschätzung für den Erwartungswert

Der Erwartungswert zu einem beobachteten Ereignis ist

- mindestens so groß, dass die Wahrscheinlichkeit, dass ein größeres als das beobachtete Ergebnis eintritt, kleiner α_2, und
- maximal so groß, dass ein kleineres als das beobachtete Ergebnis eintritt, kleiner α_1, ist.



Untere und obere Bereichsgrenze des Erwartungswertes:

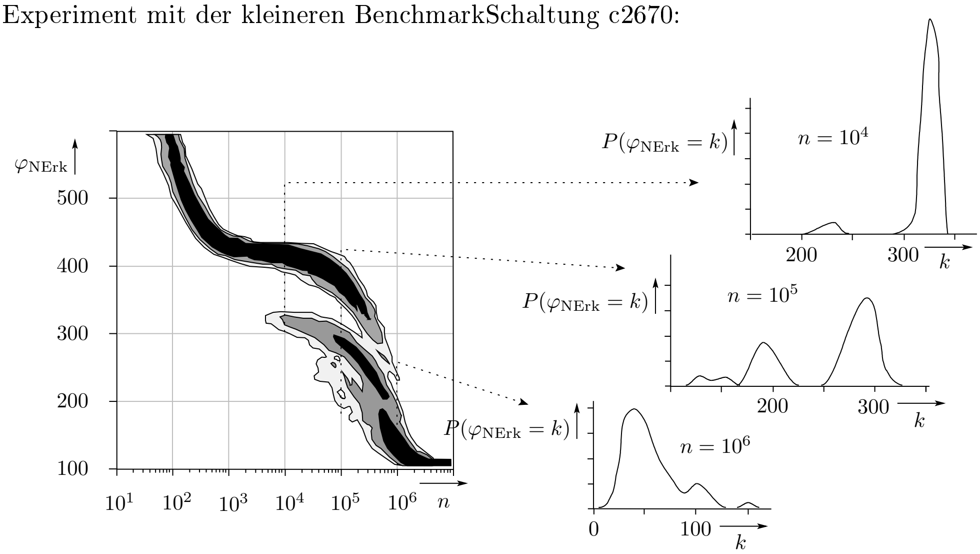
$$\begin{aligned}\mathbb{E}[X]_{\text{UG}} &= x_{\text{ist}} - \text{sd}[X] \cdot \Phi^{-1}(1-\alpha_2) \\ \mathbb{E}[X]_{\text{OG}} &= x_{\text{ist}} + \text{sd}[X] \cdot \Phi^{-1}(1-\alpha_1)\end{aligned}$$

mit

$$\text{sd}[X] = \sqrt{\kappa \cdot \mathbb{E}[X]} \approx \sqrt{\kappa \cdot x_{\text{ist}}}$$

Nicht normalverteilte Zählwerte

Dasselbe Experiment mit der kleineren BenchmarkSchaltung c2670:



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

3 Misch- und multimodale Verteilung

Mischverteilung

- Kiste mit Schrauben aus unterschiedlichen Herstellungsprozessen mit unterschiedlicher Verteilung der Länge X.
- Mehrere Programmierer schreiben SW-Bausteine. Jeder hat eine andere Fehlerentstehungsrate, ...
- Unterschiedliche Klassen von FF mit unterschiedlichen Verteilungen der Schadenshöhe. ...

Zufallsexperiment: Aus einer Grundgesamtheit von $\#X_{\text{ges}}$ Objekten, von den jeweils $\#X_i$ Objekte eine Verteilung $X_i \sim F_X^{(i)}$ haben:

$$F(x) = \mathbb{P}[X \geq x] = \frac{1}{\#X_{\text{ges}}} \cdot \sum_{\forall H_i} \#X_i \cdot F_X^{(i)}(x)$$

Mit der Auswahl als Zufallsgröße:

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

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

Dichtefunktion für stetige X:

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} = \sum_{i=1}^{\#i} h_i \cdot \frac{\mathrm{d}F^{(i)}(x)}{\mathrm{d}x} = \sum_{\forall h_i} h_i \cdot f_X^{(i)}(x)$$

Wahrscheinlichkeiten für diskrete X:

$$\mathbb{P}[X = k] = \sum_{i=1}^{\#i} h_i \cdot \mathbb{P}\left[X = k | X \sim F_X^{(i)}\right]$$

Erwartungswert und zweites Moment sind gleichfalls gewichtete Mittelwerte:

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

$$\mathbb{E}\left[X^2\right] = \sum_{i=1}^{\#i} h_i \cdot \mathbb{E}\left[X_i^2\right]$$

Varianz nach Anwendung Verschiebesatz $\sigma^2 = \mathbb{E}\left[X^2\right] - \mu^2$:

$$\sigma^2 = \sum_{i=1}^{\#i} h_i \cdot \left(\sigma_i^2 + \mu_i^2\right) - \mu^2$$

Varianzerhöhung durch unterschiedliche Erwartungswerte:

$$\begin{aligned}
\sum_{i=1}^{\#i} h_i \cdot \mu_i^2 - \mu^2 &= \sum_{i=1}^{\#i} h_i \cdot (\mu - \delta_i)^2 - \mu^2 \quad \text{mit } \sum_{i=1}^{\#i} h_i \cdot \delta_i^2 = 0 \\
&= \sum_{i=1}^{\#i} h_i \cdot \left(\mu^2 - 2\mu\delta_i + \delta_i^2\right) - \mu^2 \\
&= \underbrace{\sum_{i=1}^{\#i} h_i \cdot 2\mu\delta_i}_{0} + \underbrace{\sum_{i=1}^{\#i} h_i \cdot \delta_i^2}_{\geq 0} + \underbrace{\sum_{i=1}^{\#i} h_i \cdot \mu^2 - \mu^2}_{0}
\end{aligned}$$

Für Bereichsschätzungen ist eine multimodele Verteilung, auch bei Mischung normalverteilter Grundgesamtheiten mit deutlich voneinander abweichenden Erwartungswerten nicht mehr näherungsweise normalverteil. Die tschebytscheffsche Ungleichung:

$$\alpha = \mathbb{P}[|x - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\operatorname{Var}[X]}{\varepsilon^2} \tag{15}$$

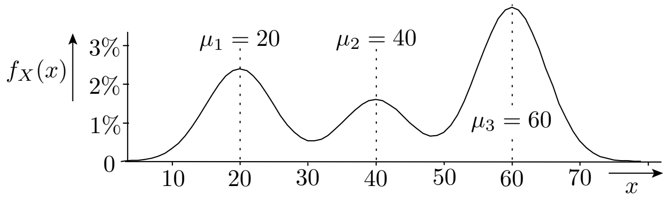
gilt immer, d.h. auch bei multimodaler Verteilung.

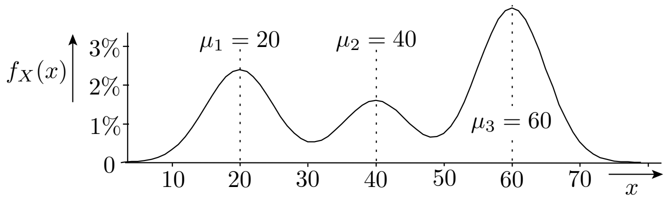
Mischung normalverteilter Grundgesamtheiten

Grundgesamtheit mit 3 verschieden normalverteilten Zufallsgrößen $\varphi\left(\frac{x-\mu_i}{\sigma=5}\right)$:

h_i	0,3	0,2	0,5
μ_i	20	40	60

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





- Erwartungswert der Mischverteilung:

$$\mu = 0{,}3\cdot\mu_1 + 0{,}2\cdot\mu_2 + 0{,}5\cdot\mu_3 = 44$$

- Varianz und Standardabweichung:

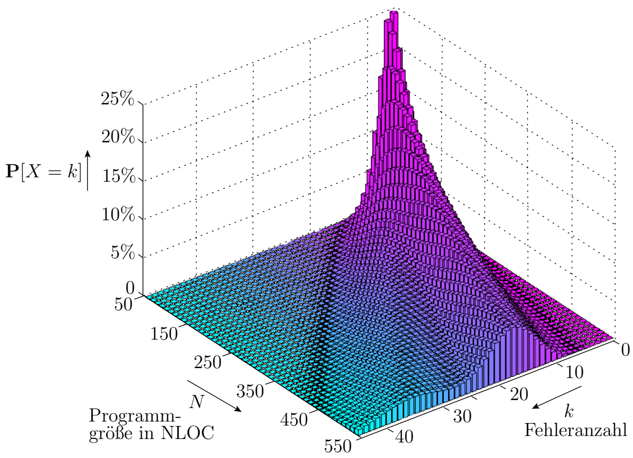
$$\sigma^2 = 0{,}3\cdot\left(5^2+\mu_1^2\right) + 0{,}2\cdot\left(5^2+\mu_2^2\right) + 0{,}5\cdot\left(5^2+\mu_3^2\right) - \mu^2 = 329$$
$$\sigma = 18{,}1$$

- Wahrscheinlicher Bereich ca. $\mu \mp 30$, Irrtumswahrscheinlichkeit nach Gl. 1 max.:

$$\mathbb{P}\left[|x-\mathbb{E}[X]| \geq \varepsilon\right] \leq \frac{\operatorname{Var}[X]}{\varepsilon^2} = \frac{18{,}1}{30^2} = 2\%$$

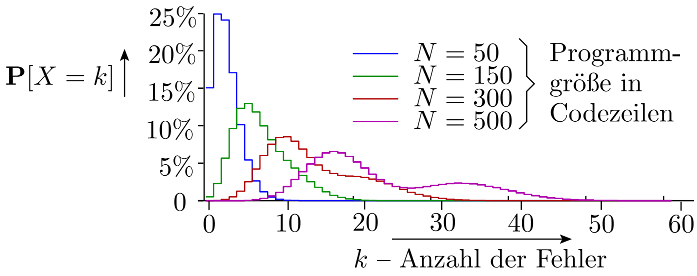
Unterschiedlich gute Programmierer

Beispiel sei ein Software-Team, in dem ein Anfänger und ein Profi gemeinsam Software-Bausteine aus N Code-Zeilen entwickeln, der Profi 66% der Bausteine mit ca. einem Fehler je 30 Codezeilen und der Anfänger 33% der Bausteine mit einem Fehler je 15 Codezeilen:



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 vom Anfänger stammt:

$$\mathbb{P}\left[N, X=k\right] = \frac{2}{3} \cdot e^{-\frac{N}{30}} \cdot \frac{\left(\frac{N}{30}\right)^k}{k!} + \frac{1}{3} \cdot e^{-\frac{N}{15}} \cdot \frac{\left(\frac{N}{15}\right)^k}{k!} \tag{16}$$



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

Beispiel: Identisch nachweisbare Fehler

In einer Modellfehlermenge aus $N = 25$ Fehlern mit einer Nachweiswahrscheinlichkeit $p = 40\%$ seien zehn Fehler identisch und die übrigen Fehler unabhängig voneinander nachweisbar. Gesucht:

1. Beschreibung als Mischverteilung von zueinander verschobenen Binomialverteilungen.
2. Erwartungswert und Standardabweichung.
3. Vergleich mit Erwartungswert und Standardabweichung für 25 mit $p = 40\%$ unabhängig voneinander nachweisbare Fehler.

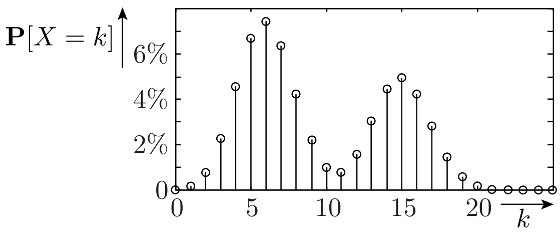
1. Beschreibung als Mischverteilung

   Verteilung der 15 unabhängig voneinander nachweisbaren Fehler:

   $$\mathrm{B}\left(15,\ k\right) = \begin{cases} \left(\begin{array}{c} 15 \\ k \end{array}\right) \cdot 0{,}6^k \cdot 0{,}4^{15-k} & 0 \le k \le 15 \\ 0 & \text{sonst} \end{cases}$$

   Mit mit $1-p$ sind nur unabhänge Fehlern nachweisbar, sonst zusätzlich die 10 identischen Fehlern (Verteilung $\mathrm{Bin}\left(k-10\right)$):

   $$\mathbb{P}\left[X=k\right] = 0{,}6 \cdot \mathrm{B}\left(15,\ k\right) + 0{,}4 \cdot \mathrm{B}\left(15,\ k-10\right)$$



Lösung Aufgabenteil 2 und 3

k	0	1	$E(X)_i$	$D^2(X)_i$
$P(X_i = k)$	$1-p$	p	p	$p\cdot(1-p)$

k	0	10	$E(X)_{16}$	$D^2(X)_{16}$
$P(X_{16} = k)$	$1-p$	p	$10\cdot p$	$10^2\cdot p\cdot(1-p)$

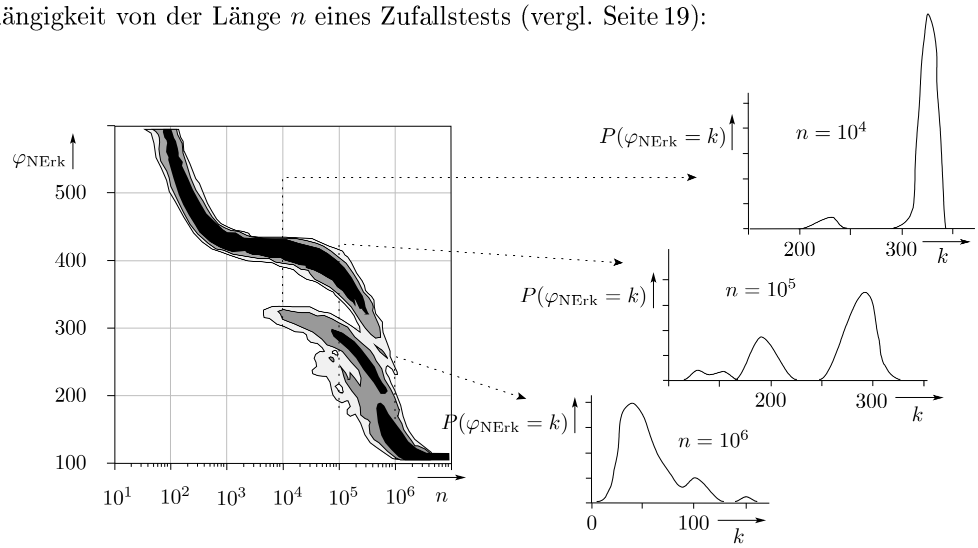
2.
$$\mathbb{E}[X] = 15\cdot p + 10\cdot p = 25\cdot p = 25\cdot 40\% = 10$$

3. Erwartungswert und Varianz als Summe der Varianzen der Summanden:
$$\mathrm{Var}[X] = 15\cdot p\cdot(1-p) + 10^2\cdot p\cdot(1-p) = 115\cdot p\cdot(1-p)$$

4. Gegenüber der Varianz der Summe von 25 unabhängigen Ereignissen mit Eintrittswahrscheinlichkeit p
$$\mathrm{Var}[X_{\text{unabh}}] = 25\cdot p\cdot(1-p)$$
Varianzerhöhung: $115/25 = 4{,}6$

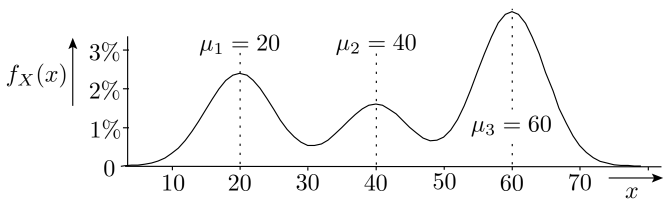
Dichte nicht nachweisbare Fehler Benchmark c2670

... in Abhängigkeit von der Länge n eines Zufallstests (vergl. Seite 19):



Im Bereich von $n = 10^4$ bis 10^6 multimodale Verteilung. Offenbar ca. 80 sehr ähnlich nachweisbare Fehler mit $p_{\text{Erk}} \approx 10^5$.

Multimodalität



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

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

4 Weitere Verteilungen

Verteilungen für

- Nachweislängen,
- Schadenskosten,
- Lebensdauer.

4.1 Pareto-Verteilung

Das Pareto-Prinzip[6]

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

- Wenige Entstehungsursachen ⇒ Mehrheit der Fehler.
- Wenige Fehler ⇒ Mehrheit der FF.
- Wenige FF ⇒ Mehrheit der Schadenskosten.
- Wenige Zufallstests erkennen die Mehrheit der Fehler.

Pareto-Verteilung

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

$$F_X(x) = \mathbb{P}\,[X \leq x]) = 1 - \left(\frac{x_{\min}}{x}\right)^k \qquad k > 0$$

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

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

Für kleine Exponenten gehört sie zu den endlastigen Verteilungen, bei denen ein erheblicher Teil der Wahrscheinlichkeitsmasse auf große x entfällt. Typisch für die Schadenskosten einer FF und die Nachweislänge[7] eines Fehlers durch Zufallstests.

Ein Erwartungswert:

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

existiert nur für $k > 1$:

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

Eine Varianz existiert nur für $k > 2$:

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

Auf Foliensatz 1 wurde für die Verteilung der Nachweislänge für den zufälligen Fehlernachweis de facto eine Pareto-Verteilung mit $0 < k < 1$ abgeschätzt. Diese hat dann keinen Erwartungswert.

Pareto-verteilte Schadenskosten mit $0 < k < 1$ haben auch keinen Erwartungswert und können entsprechend nur mit einer Haftungsbeschränkung auf einen maximal zu erstattenden Wert abgesichert werden.

[6] 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.

[7] Erforderliche Anzahl der zufällig ausgewählten Testbeispiele.

Pareto-Prinzip $\surd$

Der Anteil der Ursachen U mit der größten Wirkung:

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

haben mindestens die Wirkung:

$$w_{\min}=x_{\min}\cdot U^{-\frac{1}{k}}$$

die zu erwartende Gesamtwirkung (nur für $k>1$ angebbar):

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

Anteilige Gesamtwirkung ..

Der kleine Anteil der Ursachen U hat mindestens die Wirkung

$$w_{\min}=x_{\min}\cdot U^{-\frac{1}{k}}$$

und die anteilige Gesamtwirkung:

$$\begin{aligned}W&=\frac{\mathbb{E}\left[X|X>w_{\min}\right]}{\mathbb{E}\left[X\right]}=U^{\frac{k-1}{k}}\\k&=\frac{1}{\left(1-\frac{\ln(W)}{\ln(U)}\right)}\end{aligned}$$

Für das Pareto-20-80-Prinzip »$U=20\%$ der Bevölkerung besitzen $W=80\%$ des Bodens«:

$$\begin{aligned}k&=\frac{1}{\left(1-\frac{\ln(0{,}8)}{\ln(0{,}2)}\right)}=1{,}161\\w_{\min}&=x_{\min}\cdot 0{,}2^{-\frac{1}{k}}=x_{\min}\cdot 8{,}48\end{aligned}$$

4.2 Gammaverteilung

Gamma-Verteilung

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

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

Dichte:

$$f(x)=\begin{cases}e^{-\beta\cdot x}\cdot x^{\alpha-1}\cdot\frac{\beta^{\alpha}}{\Gamma(\alpha)} & x\geq 0\\0 & x<0\end{cases}$$

α – Formparameter; β – Skalenparameter; $\Gamma(\alpha)$ – Gamma-Funktion, Erweiterung der Fakultät auf reelle Zahlen. Für den Exponenten $0<\alpha\leq 1$ beträgt sie überschlagsweise $1/\alpha$ und für $\alpha>1$ gilt $\Gamma(\alpha+1)=\alpha\cdot\Gamma(\alpha)$.

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

Erwartungswert einer gamma-verteilten Zufallsgröße:

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

Varianz:

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

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

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

4.3 Exponentialverteilung

Exponentialverteilung

Die Exponentialverteilung $\operatorname{Exp}(\lambda) = \mathcal{G}(1, \lambda)$ ist eine Gamma-Verteilung mit Formparameter $\alpha = 1$. Beispieleanwendung: Lebensdauer von Bauteilen, wenn Alterungserscheinungen nicht betrachtet werden. Dichte:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

λ – Anzahl der zu erwartenden Ereignisse pro Zeitintervall. Verteilungsfunktion:

$$F(x) = \int_0^x f(x) \cdot \mathrm{d}x = \begin{cases} 1 - e^{-\lambda \cdot x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Erwartungswert:

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

Varianz:

$$\operatorname{Var}[X] = \frac{1}{\lambda^2}$$