Die Varianz (Streuung)
Definition

Ang., die betrachteten Erwartungswerte existieren.

$$var(X) = \mathbf{E}(X - \mathbf{E}X)^2$$

heißt <u>Varianz</u> der Zufallsvariable X.

$$\sigma = \sqrt{Var(X)}$$

heißt <u>Standardabweichung</u> der Zufallsvariablen X.

Bez.: $var(X), Var(X), varX, \sigma^2, \sigma_X^2, \sigma, \sigma_X$.

Sei $\mu = \mathbf{E}X$.

Die Varianz

Stetige und diskrete Zufallsvariablen

Wenn X diskret, so gilt:

$$var(X) = \sum_{i=0}^{\infty} (x_i - \mu)^2 p_i$$

Wenn X stetig, so gilt:

$$var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx,$$

wobei f die Dichte von X ist.

$var(X)$: mittlere quadratische Abweichung von X und $\mathbf{E}X$.

Die Varianz

Eigenschaften der Varianz

$$
\begin{aligned}
var(X) &= \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X-\mu)^2 = \\
&= \mathbf{E}(X^2 - 2\mu X + \mu^2) = \\
&= \mathbf{E}X^2 - \mu^2.
\end{aligned}
$$

$$
\begin{aligned}
var(X) &= \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X-\mu)^2 = \\
&= \mathbf{E}(X^2 - 2\mu X + \mu^2) = \\
&= \mathbf{E}X^2 - \mu^2.
\end{aligned}
$$

$$
var(aX + b) = a^2 var(X), \quad a, b \in \mathbb{R}.
$$

Die Varianz
Eigenschaften der Varianz

$$
\begin{aligned}
var(X) &= \mathbf{E}(X-\mathbf{E}X)^2 = \mathbf{E}(X-\mu)^2 = \\
&= \mathbf{E}(X^2 - 2\mu X + \mu^2) = \\
&= \mathbf{E}X^2 - \mu^2.
\end{aligned}
$$

$$var(aX+b) = a^2 var(X), \quad a, b \in \mathbb{R}.$$

$$var(X) = 0 \Longleftrightarrow \exists c : \quad P(X=c) = 1.$$

Zwei Zufallsvariablen X und Y heißen unabhängig, falls

$$P(X \le x, Y \le y) = P(X \le x) \cdot P(Y \le y)$$

für alle $x, y \in \mathbf{R}$.

Zwei Zufallsvariablen X und Y heißen unabhängig, falls

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

für alle $x, y \in \mathbf{R}$.

Zwei Ereignisse A und B heißen unabhängig, falls

$$P(A, B) = P(A) \cdot P(B)$$

X und Y sind also unabhängig gdw. die Ereignisse $X \leq x$ und $Y \leq y$ unabhängig sind für alle $x, y \in \mathbf{R}$.

Zwei Zufallsvariablen X und Y heißen unabhängig, falls

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

für alle $x, y \in \mathbf{R}$.

Zwei Ereignisse A und B heißen unabhängig, falls

$$P(A, B) = P(A) \cdot P(B)$$

X und Y sind also unabhängig gdw. die Ereignisse $X \leq x$ und $Y \leq y$ unabhängig sind für alle $x, y \in \mathbf{R}$.

Seien X und Y unabhängig. Dann gilt

$$var(X + Y) = var(X) + var(Y).$$

Die Varianz

Poisson-Verteilung

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, 2, \ldots$$

$$\begin{aligned}
var(X) &= \mathbf{E}(X - \mathbf{E}X)^2 = \sum_{i=0}^{\infty} (i - \lambda)^2 p_i \\
&= \sum_{i=2}^{\infty} i \cdot (i-1) p_i + \sum_{i=0}^{\infty} i p_i - \\
&\quad 2\lambda \sum_{i=0}^{\infty} i p_i + \lambda^2 \sum_{i=0}^{\infty} p_i \\
&= e^{-\lambda} \lambda^2 \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} + \lambda - 2\lambda^2 + \lambda^2 = \lambda.
\end{aligned}$$

Die Varianz

Binomialverteilung, $X \sim B(n, p)$

$$P(X = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

$$var(X) = np(1-p).$$

(ohne Beweis, ÜA)

Die Varianz

Gleichverteilung auf (a, b)

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{sonst.} \end{cases}$$

$$\mathbf{E}X = \frac{a+b}{2}.$$

Die Varianz

Gleichverteilung auf (a, b)

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{sonst.} \end{cases}$$

$$\mathbf{E}X = \frac{a+b}{2}.$$

$$\begin{aligned} \mathbf{E}X^2 &= \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{3}x^3\Big|_a^b \cdot \frac{1}{b-a} \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}. \end{aligned}$$

Die Varianz

Gleichverteilung auf (a,b)

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{sonst.} \end{cases}$$

$$\mathbf{E}X = \frac{a+b}{2}.$$

$$\begin{aligned} \mathbf{E}X^2 &= \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{1}{3}x^3\Big|_a^b \cdot \frac{1}{b-a} \\ &= \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}. \\ var(X) &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\ &= \frac{1}{12}(4a^2+4ab+4b^2-3a^2 \\ &\quad -6ab-3b^2) \\ &= \frac{1}{12}(a^2-2ab+b^2) = \frac{(b-a)^2}{12}. \end{aligned}$$

Die Varianz
Exponentialverteilung

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & \text{falls } x \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

$$\begin{aligned} \mathbf{E}X &= \lambda. \\ \mathbf{E}X^2 &= \int_0^\infty x^2 \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\, dx = 2 \cdot \lambda^2 \quad \text{(ÜA)}. \\ var(X) &= \lambda^2. \end{aligned}$$

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx \\
\mathbf{E}(X-\mu)^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx \\
&= \sigma^2 \int_{-\infty}^{\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt \\
&= \sigma^2 \int_{-\infty}^{\infty} (-t)(-t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}})\, dt \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \Big(-te^{-t^2/2}\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1) e^{-\frac{t^2}{2}}\, dt \Big) \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt = \sigma^2.
\end{aligned}
$$

Normalverteilung

Besondere Eigenschaften

(schwaches) Gesetz der Großen Zahlen

Seien X_i unabhängig, identisch verteilt, $EX_i = \mu$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow_p \mathbf{E}X$$

Zentraler Grenzwertsatz

Seien X_i unabhängig, identisch verteilt,
$EX_i = \mu, varX_i = \sigma^2$.

$$Z_n := \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \rightarrow Z, \quad Z \sim N(0,1).$$

`Descr_Binomial_2.sas` `Descr_Exp.sas`

Normalverteilung

Fehlertheorie

Fehler sind unter folgenden Annahmen (asymptotisch) normalverteilt:

- Jeder Fehler ist Summe einer sehr großen Anzahl sehr kleiner, gleich großer Fehler, die verschiedene Ursachen haben.
- Die verschiedenen Fehlerkomponenten sind unabhängig.
- Jede Fehlerkomponente ist mit Wkt. 0.5 positiv und mit Wkt. 0.5 negativ.

bei gegebenen

Mittelwert μ und Varianz σ^2.

f: Wkt.dichte auf $(-\infty, \infty)$.

$\int xf(x)\,dx = \mu, \quad \int (x-\mu)^2 f(x)\,dx = \sigma^2$

Entropie:

$$H(f) := -\int f(x) \log f(x)\,dx$$

ist zu maximieren unter den obigen Bedingungen.

$\Longrightarrow \quad f =$Normaldichte.

Literatur: Rao: Lineare Statistische Methoden, 3.a.1.

Normalverteilung

Die Summe normalverteilter Zufallsvariablen

Die Summe normalverteilter Zufallsvariablen ist normalverteilt.

Seien $X_1 \sim N(\mu_1, \sigma_1^2)$ $X_2 \sim N(\mu_2, \sigma_2^2)$. Dann

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2).$$

(ρ: Korrelationskoeffizient zwischen X_1 und X_2, s.u.)

Beweis: über charakteristische Funktionen (Fouriertransformationen der Dichte) oder über die Faltungsformel (Stochastik-Vorlesung).