11. Statistik und Fehlerrechnung: Inhalt

11 Statistik und Fehlerrechnung

- Mittelwert, Varianz, Standardabweichung
- Fehlerfortpflanzung

Mittelwert, Varianz und Standardabweichung

Mittelwert

Der Erwartungswert $\langle x \rangle$ oder Mittelwert μ einer auf 1 normierten Wahrscheinlichkeitsdichteverteilung $p(x)$ ist definiert als

$$\langle x \rangle \; := \; \int_{-\infty}^{\infty} x \, p(x) \, \mathrm{d}x \; = \; \mu \qquad (182)$$

Varianz

Die Varianz ist der Erwartungswert von $(x-\mu)^2$:

$$V[x] \; = \; \left\langle (x-\mu)^2 \right\rangle \; = \; \int_{-\infty}^{\infty} (x-\mu)^2 p(x) \, \mathrm{d}x \; = \; \left\langle x^2 \right\rangle - \mu^2 \; = \; \sigma^2 \qquad (183)$$

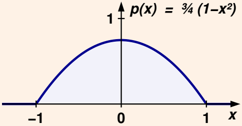
Die **Standardabweichung** σ ist die (positive) Wurzel der Varianz und ein Maß für die Breite der Verteilung.

Mittelwert, Varianz und Standardabweichung

Beispiel

Gegeben sei die Wahrscheinlichkeitsdichte

$$p(x) = \begin{cases} \frac{3}{4}(1-x^2), & -1 \le x \le 1 \\ 0, & \text{sonst} \end{cases}$$



Normierung:

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{-1}^{1} \frac{3}{4}\left(1-x^2\right)dx = \left[\frac{3}{4}x - \frac{1}{4}x^3\right]_{-1}^{1} = 2\left(\frac{3}{4} - \frac{1}{4}\right) = 1. \checkmark$$

Mittelwert:

$$\mu = \langle x \rangle = \int_{-\infty}^{\infty} x\,p(x)\,dx = \int_{-1}^{1} \frac{3}{4}x\left(1-x^2\right)dx = \left[\frac{3}{8}x^2 - \frac{3}{16}x^4\right]_{-1}^{1} = 0.$$

Varianz:

$$V = \int_{-\infty}^{\infty} (x-\mu)^2\,p(x)\,dx = \int_{-1}^{1} \frac{3}{4}x^2\left(1-x^2\right)dx = \left[\frac{1}{4}x^3 - \frac{3}{20}x^5\right]_{-1}^{1} = \frac{1}{5}.$$

$\Longrightarrow$ Standardabweichung $\sigma = \sqrt{\frac{1}{5}} \simeq 0{,}447.$

Die Gaußverteilung

Gaußverteilung

Die Gauß- oder Normalverteilung ist die Wahrscheinlichkeitsdichte

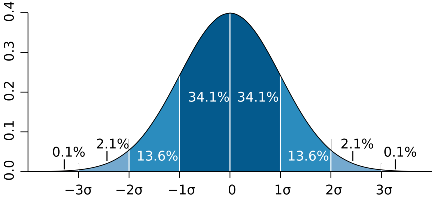
$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (184)$$

Mittelwert: $\langle x\rangle = \mu$

Varianz: $V[x] = \sigma^2$

Standardabweichung σ:

Innerhalb von $\pm 1\,\sigma$ befinden sich 68.2 % der Ereignisse, innerhalb von $\pm 2\,\sigma$ 95.5 % und innerhalb von $\pm 3\,\sigma$ 99.7 %.



Verwendung: Wiederholung von Messungen, Fehlerangaben u.v.a.m.

Mittelwert und Varianz von diskreten Verteilungen

Normalerweise hat man n Messwerte x_i, deren Wahrscheinlichkeitsdichteverteilung unbekannt ist, und muss aus den Messwerten Mittelwert und Varianz bestimmen.
Diese Größen ("Schätzer") gehen im Grenzfall $n \to \infty$ in die entsprechenden Größen der Wahrscheinlichkeitsdichteverteilung über.

Mittelwert von n Messwerten x_i

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad \lim_{n\to\infty} \overline{x} = \langle \overline{x} \rangle = \mu \tag{185}$$

Varianz von n Messwerten x_i (bei unbekanntem Mittelwert μ)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{n}{n-1} (\overline{x^2} - \overline{x}^2), \qquad \lim_{n\to\infty} s^2 = \sigma^2 \tag{186}$$

(Bei bekanntem Mittelwert μ verkleinert sich die Varianz um $\frac{n-1}{n}$.)

Geschätzer Mittelwert und Unsicherheit

Varianz von $\overline{x}$

$$V[\overline{x}] = \left\langle (\overline{x} - \langle \overline{x} \rangle)^2 \right\rangle = \frac{1}{n^2} \sum_{i=1}^{n} \left\langle (x_i - \mu)^2 \right\rangle = \frac{\sigma^2}{n} \tag{187}$$

Wird eine Messung n-mal wiederholt mit Messungen x_i, so wird $\overline{x}$ als Messwert und $\Delta x_{stat} = \frac{s}{\sqrt{n}}$ als Fehler angegeben.

Dieser Fehler ist statistischer Natur, weil er mit $\sqrt{n}$ abnimmt.

Der systematische Fehler (wenn nicht vernachlässigbar) muss noch abgeschätzt und quadratisch addiert werden. Dafür gibt es keine allgemeine Standardmethoden.

Kovarianz

Definition

Für zwei Größen x und y ist die **Kovarianz**

$$\begin{aligned} V_{xy} &= \big\langle (x-\mu_x)(y-\mu_y) \big\rangle = \langle x\, y \rangle - \mu_x \mu_y \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, y\, f(x,y)\, \mathrm{d}x\, \mathrm{d}y - \mu_x \mu_y \end{aligned} \tag{188}$$

Die Kovarianz verschwindet für unabhängige Variablen, denn es gilt $f(x,y) = f(x) \cdot f(y)$.

Die Kovarianz zwischen x und y aus n Messungen (x_i, y_i) ist

$$\hat{V}_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \frac{n}{n-1} (\overline{xy} - \overline{x} \cdot \overline{y}) \tag{189}$$

Für n Größen x_i kann man eine $(n \times n)$-Matrix $V = (V_{ij})$ definieren, mit den Diagonalelementen $V_{ii} = \sigma_i^2$.

Fehlerfortpflanzung

Seien $\boldsymbol{x} = (x_1, \ldots, x_n)$ n Größen mit Mittelwerten $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ und $f(\boldsymbol{x})$ eine Funktion der n Variablen.

Man kann für $f(\boldsymbol{x})$ um $\boldsymbol{\mu}$ eine Taylorentwicklung durchführen:

$$f(\boldsymbol{x}) \simeq f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}=\boldsymbol{\mu}} \cdot (x_i - \mu_i) \tag{190}$$

Dann gilt

$$\langle f(\boldsymbol{x}) \rangle \simeq f(\boldsymbol{\mu}) \tag{191}$$

$$\langle f^2(\boldsymbol{x}) \rangle \simeq f^2(\boldsymbol{\mu}) + \sum_{i,j=1}^{n} \left.\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{\mu}} \cdot V_{ij} \tag{192}$$

$$\sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2 \simeq \sum_{i,j=1}^{n} \left.\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{\mu}} \cdot V_{ij} \tag{193}$$

Fehlerfortpflanzung

Beispiel: Geschwindigkeitsmessung

Eine Geschwindigkeit $v = \frac{s}{t}$ soll über die Messungen einer Strecke s und der dafür benötigten Zeit t bestimmt werden.

Messwerte (voneinander unabhängig):
$$\begin{aligned} s &= \langle s \rangle \pm \sigma_s = (100 \pm 1)\,\mathrm{m} \\ t &= \langle t \rangle \pm \sigma_t = (10 \pm 1)\,\mathrm{s} \end{aligned}$$

Erwartungswert der Geschwindigkeit: $\langle v \rangle = \frac{\langle s \rangle}{\langle t \rangle} = \frac{100\,\mathrm{m}}{10\,\mathrm{s}} = 10\,\frac{\mathrm{m}}{\mathrm{s}}$.

Partielle Ableitungen: $\left.\frac{\partial v}{\partial s}\right|_{\substack{s=\langle s \rangle \\ t=\langle t \rangle}} = \frac{1}{\langle t \rangle}$ und $\left.\frac{\partial v}{\partial t}\right|_{\substack{s=\langle s \rangle \\ t=\langle t \rangle}} = -\frac{\langle s \rangle}{\langle t \rangle^2}$.

Da die Messungen voneinander unabhängig sind, brauchen nur die Diagonalelemente $V_{11} = \sigma_s^2$ und $V_{22} = \sigma_t^2$ berücksichtigt zu werden.

Fehlerfortpflanzung

$$\Longrightarrow \quad \sigma_v^2 = \sum_{i,j=1}^{2} \frac{\partial v}{\partial x_i}\frac{\partial v}{\partial x_j}\bigg|_{\substack{s=\langle s\rangle \\ t=\langle t\rangle}} \cdot V_{ij} = \left(\frac{\partial v}{\partial s}\bigg|_{\substack{s=\langle s\rangle \\ t=\langle t\rangle}}\right)^2 \cdot \sigma_s^2 + \left(\frac{\partial v}{\partial t}\bigg|_{\substack{s=\langle s\rangle \\ t=\langle t\rangle}}\right)^2 \cdot \sigma_t^2$$

$$= \frac{1}{\langle t\rangle^2}\sigma_s^2 + \frac{\langle s\rangle^2}{\langle t\rangle^4}\sigma_t^2 = \langle v\rangle^2 \cdot \left[\left(\frac{\sigma_s}{\langle s\rangle}\right)^2 + \left(\frac{\sigma_t}{\langle t\rangle}\right)^2\right]$$

$$= \left(10\,\tfrac{\mathrm{m}}{\mathrm{s}}\right)^2 \cdot \left[\left(\frac{1\,\mathrm{m}}{100\,\mathrm{m}}\right)^2 + \left(\frac{1\,\mathrm{s}}{10\,\mathrm{s}}\right)^2\right]$$

$$= \left(10\,\tfrac{\mathrm{m}}{\mathrm{s}}\right)^2 \cdot \left[\tfrac{1}{10\,000} + \tfrac{1}{100}\right] \simeq \left(10\,\tfrac{\mathrm{m}}{\mathrm{s}}\right)^2 \cdot \tfrac{1}{100} = \left(1\,\tfrac{\mathrm{m}}{\mathrm{s}}\right)^2.$$

Damit erhält man schließlich für die gesuchte Geschwindigkeit:

$$\boldsymbol{v = \langle v\rangle \pm \sigma_v = (10 \pm 1)\,\tfrac{\mathrm{m}}{\mathrm{s}}}.$$