Ideas for measuring the average distance of values of X around μ:

Suppose, X is discretely distributed with pmf $p(x_i)$, $i=1,\dots,n$

1st try:

$$\sum_{i=1}^{n} |x_i - \mu| \, p(x_i) = E[\,|X-\mu|\,]$$

2nd try:

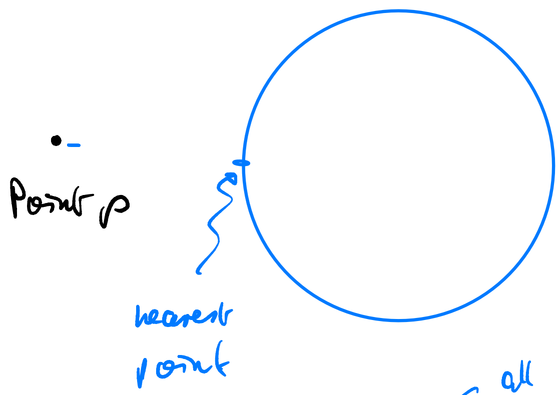
$$\max_{i=1}^{n} |x_i - \mu|$$

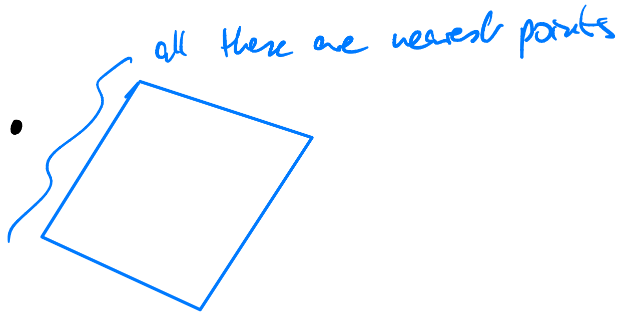
However, usually we have

$$\left(\sum_{i=1}^{n} (x_i - \mu)^2 \, p(x_i) \right)^{1/2} = E[\,(X-\mu)^2\,]^{1/2}$$

This the standard deviation of X

Round circles allow for unique approximation.





Eigenschaften der Kovarianz

$$\mathrm{Cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]$$
$$= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y$$
$$= E[XY] - \mu_X \mu_Y - \cancel{\mu_Y \mu_X} + \cancel{\mu_X \mu_Y}$$
$$= E[XY] - E[X] \cdot E[Y]$$

Cov. misst den Grad der Abh-

Beobachtung: X, Y unabh.

$$E[XY] = E[X] \cdot E[Y]$$
$$\Rightarrow \mathrm{Cov}(X, Y) = 0$$

Kovarianz und Addition

Satz $\mathrm{Cov}(X_1 + X_2, Y) \overset{?}{=} \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

$$\mathrm{Cov}(X_1 + X_2, Y) = E[(X_1 + X_2)\cdot Y] - E[X_1 + X_2] \cdot E[Y]$$
$$= E[X_1 \cdot Y + X_2 \cdot Y] - (E[X_1] + E[X_2]) \cdot E[Y]$$
$$= E[X_1 \cdot Y] + E[X_2 \cdot Y] - (E[X_1] \cdot E[Y] + E[X_2] \cdot E[Y])$$
$$= E[X_1 \cdot Y] - E[X_1] \cdot E[Y] + E[X_2 \cdot Y] - E[X_2] \cdot E[Y]$$
$$= \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$$

Theorem $\mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \mathrm{Cov}(X_i, Y_j)$

Varianz der Summe

$$Var\left(\sum_i X_i\right) = Cov\left(\sum_i X_i, \sum_i X_i\right)$$
$$= \sum_{i,j} Cov(X_i, X_j)$$
$$= \sum_i \left(\sum_{j \neq i} Cov(X_i, X_j) + Cov(X_i, X_i) \right)$$
$$= \sum_i \sum_{j \neq i} Cov(X_i, X_j) + \sum_i Var(X_i)$$
$$= \sum_i Var(X_i) + \underbrace{\sum_i \sum_{j \neq i} Cov(X_i, X_j)}_{=0 \text{, falls alle } X_i \text{ unabhängig voneinander}}$$

$$Var(X + Y) = Var(X) + Var(Y) + 2 \cdot Cov(X, Y)$$

$$(x+y)^2 = x^2 + y^2 + 2x \cdot y$$

Bedeutung der Kovarianz

$Cov(X, Y)$ ist

> 0 : $X - \mu_x$, $Y - \mu_y$ haben meist ähnliche Werte

< 0 : $X - \mu_x$, $Y - \mu_y$ haben meist entgegen-gesetzte Werte

≈ 0 : $X - \mu_x$, $Y - \mu_y$ haben meist nichts miteinander, sind gleich oft ähnlich wie entgegengesetzt

Korrelation zwischen $\mathcal{X}$ und $\mathcal{Y}$

Idee: Normalisiere die Kovarianz auf Werte zwischen -1 und 1

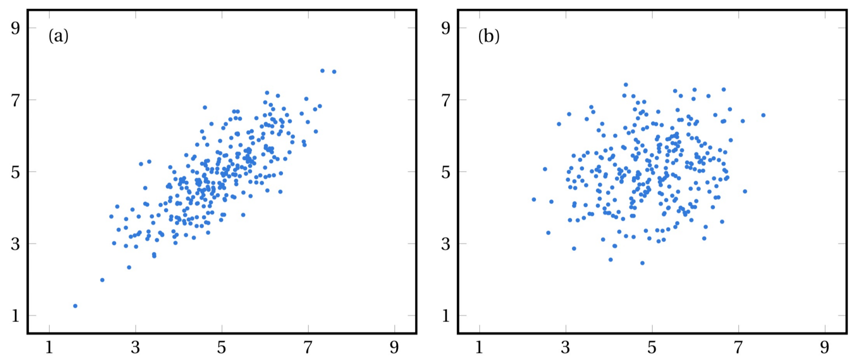
$$\mathcal{X}_0 := \frac{\mathcal{X}}{\sigma_{\mathcal{X}}} = \frac{\mathcal{X}}{\sqrt{Var(\mathcal{X})}}, \quad \mathcal{Y}_0 := \frac{\mathcal{Y}}{\sigma_{\mathcal{Y}}} = \frac{\mathcal{Y}}{\sqrt{Var(\mathcal{Y})}}$$

$$Var(\mathcal{X}_0) = Var\left(\frac{\mathcal{X}}{\sigma_{\mathcal{X}}}\right) = \frac{1}{\sigma_{\mathcal{X}}^2} Var(\mathcal{X}) = \frac{1}{Var(\mathcal{X})} \cdot Var(\mathcal{X}) = 1$$

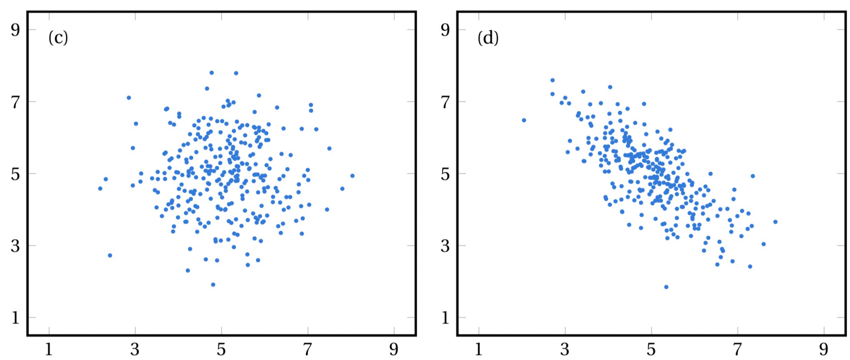
$$Corr(\mathcal{X}, \mathcal{Y}) := Cov(\mathcal{X}_0, \mathcal{Y}_0)$$

$$= Cov\left(\frac{1}{\sigma_{\mathcal{X}}} \mathcal{X}, \frac{1}{\sigma_{\mathcal{Y}}} \mathcal{Y}\right) = \frac{Cov(\mathcal{X}, \mathcal{Y})}{\sigma_{\mathcal{X}} \cdot \sigma_{\mathcal{Y}}} = \rho_{\mathcal{X},\mathcal{Y}}$$

Pearsonscher Korrelationskoeffizient (nach Karl Pearson)



$\rho_{\mathcal{X},\mathcal{Y}} = 0{,}75$ $\qquad$ $\rho_{\mathcal{X},\mathcal{Y}} = 0{,}2$



$\rho_{\mathcal{X},\mathcal{Y}} = 0$ $\qquad$ $\rho_{\mathcal{X},\mathcal{Y}} = -0{,}75$

Figure 9: Random variables $\mathscr{X}$ and $\mathscr{Y}$ with correlations (a) 0.75; (b) 0.2; (c) 0; and (d) -0.75.

Example : Small Schools

Educational scientists found that among the schools that fare best in evaluations of teaching success, there are many more small schools than there are small schools among all schools.

(See statistics from North Carolina)

The Gates Foundation decided in the early 2000's to invest heavily in the establishment of small schools (e.g., by splitting larger schools into smaller ones)

Was that a good idea?

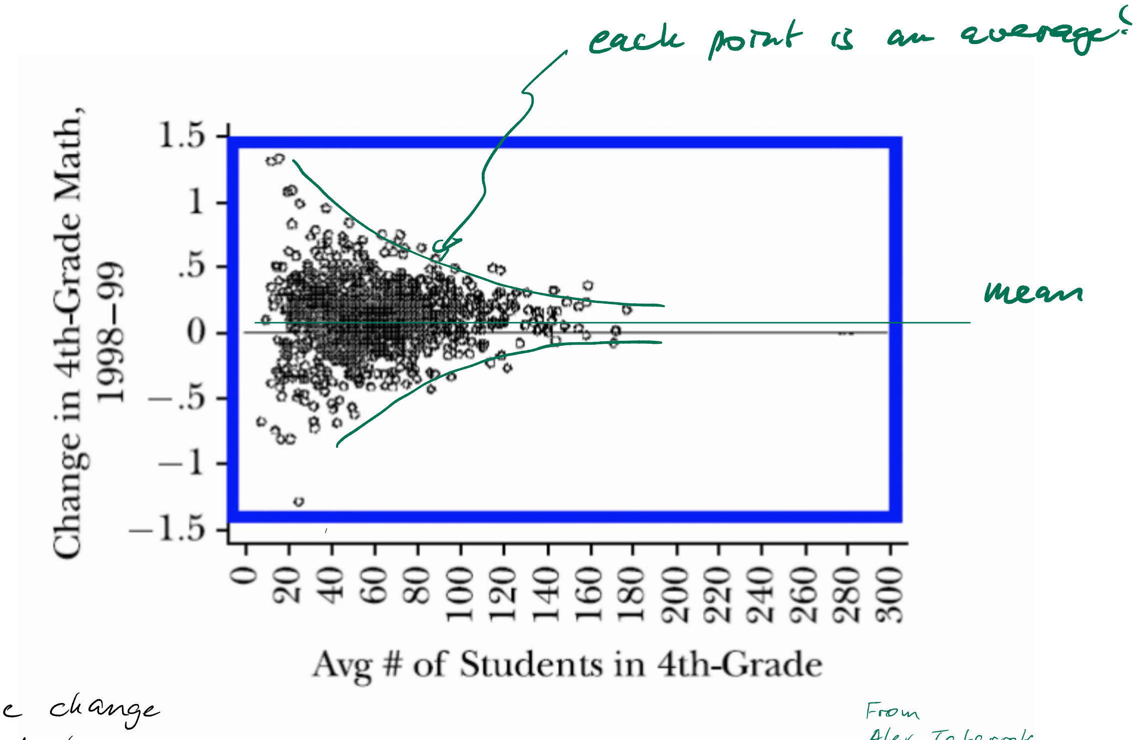
The story is from Daniel Kahnemann, "Thinking, Fast and Slow"

School Size	Percentage Ever "Top 25" 1997–2000
Smallest decile	27.7%
2nd	11.8
3rd	8.2
4th	3.6
5th	2.4
6th	3.6
7th	4.8
8th	7.1
9th	0
Largest decile	1.2
Total	7.0

Performance of small schools in North Carolina

From
Alex Tabarrok
"The Small Schools Myth",
September 2, 2010
https://marginalrevolution.com

Distribution of Performance wrt Student Numbers



average change of student performance

From Alex Tabarrok "The Small Schools Myth", September 2, 2010 https://marginalrevolution.com

Beispiel: 10-mal unabhängig Würfeln

X_i Ergebnis des i-ten Wurfs

$$X := \sum_{i=1}^{10} X_i$$

$$Var(X) = Var\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

$$= \sum_{i=1}^{10} \frac{35}{12} = 10 \cdot \frac{35}{12}$$

Was ist σ_X?

$$\sigma_X = \sqrt{10 \cdot \frac{35}{12}} = \sqrt{10} \cdot \sqrt{\frac{35}{12}} = \sqrt{10} \cdot \sigma_{X_1}$$

- Var ist um den Faktor 10 gewachsen
- σ ist nur um den Faktor $\sqrt{10}$ gewachsen

2.8 Das schwache Gesetz der großen Zahlen

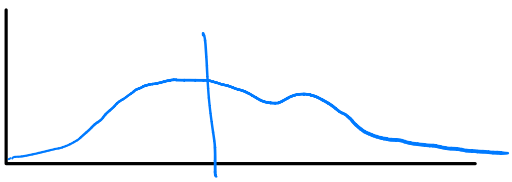
Weak Law of Large Numbers

2 Schritte: 2 Ungleichungen

- Markov
- Tschebyscheff

Markov Ungleichung

$$\boxed{X \geq 0, \quad X \sim f, \quad a > 0 \qquad P[X \geq a] \overset{?}{\leq} \frac{E[X]}{a}}$$



f wird irgendwann klein.
E[X] sagt uns, wann das spätestens passiert

$$E[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

$X \geq 0$

$$= \int_0^{\infty} x \, f(x) \, dx \geq \int_a^{\infty} \underbrace{x}_{\geq a} \cdot f(x) \, dx$$

$$\geq \int_a^{\infty} a \cdot f(x) \, dx = a \int_a^{\infty} f(x) \, dx$$

$$= a \, P[X \geq a]$$

Daher: $P[X \geq a] \leq \dfrac{E[X]}{a}$

Tschebyscheff-Ungl.

$$P[Y \geq k^2] \leq \frac{E[Y]}{k^2}$$

Wende Markov-Ungl. an auf

$$Y := (X-\mu)^2 \geq 0 \qquad a = k^2$$

Annahme: $Var(X) = \sigma^2 < \infty \Rightarrow E[Y] = Var(X) = \sigma^2$

$$P[\,|X-\mu| \geq k\,] \overset{\text{quadrieren}}{=} P[(X-\mu)^2 \geq k^2]$$

$$= P[Y \geq k^2] \leq \frac{E[Y]}{k^2} = \frac{\sigma^2}{k^2}$$

Sei X eine ZU mit MW. μ und Varianz σ, $k > 0$.
Dann gilt:

$$P[\,|X-\mu| \geq k\,] \leq \frac{\sigma^2}{k^2}$$

Szenario: Wir führen ein Experiment viele Male durch

Ergebnisse: $X_1, X_2, \dots, X_n$, so dass

1) alle X_i haben die gleiche Verteilung
2) alle X_i sind unabhängig.

Man sagt, die X unabhängig und identisch verteilt (uiv) (independent and identically distributed, iid)

$$\sigma^2 = Var(X_1) = Var(X_i)$$

$$\mu = E[X_i]$$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{Durchschnitt der } X_i$$

Erwartungswert und Varianz des Durchschnitts $\overline{X}_n$

$$E[\overline{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot n\cdot\mu = \mu$$

$$Var(\overline{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$
$$= \frac{1}{n^2}\cdot n\cdot \sigma^2 = \frac{1}{n}\sigma^2$$

Schwaches Ges. der gr. Zahlen

Tschebyscheff Ungleichung:

$$P[\,|X-\mu| \geq k\,] \leq \frac{\sigma^2}{k^2}$$

Anwendung:

$X \rightsquigarrow \overline{X}_n \qquad k \rightsquigarrow \varepsilon \qquad \sigma \rightsquigarrow \sqrt{Var(\overline{X}_n)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

ergibt

$$P[\,|\overline{X}_n - \mu| \geq \varepsilon\,] \leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{1}{n}\,\frac{\sigma^2}{\varepsilon} \longrightarrow 0$$

für $n \to \infty$

Wahrscheinlichkeit für Ausreißer (Outliers)

Whatever the threshold ε, the probability for outliers of the average $\overline{X}_n$ converges to 0 if $n \to \infty$.