Wahrscheinlichkeitstheorie und Statistik

10. Woche

Dávid Tóth (BME SZIT)

4., 7. November 2024

Kovarianz

Definition. Seien X und Y auf demselben Wahrscheinlichkeitsraum definierte Zufallsvariablen. Dann ist *die Kovarianz* (kovariancia) von X und Y durch

$$\operatorname{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

definiert, wo wir annehmen, dass dieser Erwartungswert endlich ist.

Bemerkung.

- Die Kovarianz ist ein Maß für den Zusammenhang zwischen zwei Zufallsvariablen.
- Ihr Wert ist eine Aussage darüber, ob hohe Werte der einen Zufallsvariablen eher mit hohen oder eher mit niedrigen Werten der anderen Zufallsvariablen auftreten.

Kovarianz

Bemerkung.

$$\operatorname{cov}(X, X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{D}^2(X).$$

Behauptung.

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

D.h. falls der Ausdruck an einer Seite der Gleichung endlich ist, dann ist der Ausdruck an der andere Seite der Gleichung endlich, und sie sind gleich.

Kovarianz

Beweis.

$$\begin{aligned}
\operatorname{cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \\
&= \mathbb{E}(XY - \mathbb{E}(X) \cdot Y - \mathbb{E}(Y) \cdot X + \mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(\mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Kovarianz

Bemerkung. Wenn $\mathbb{D}^2(X)$ und $\mathbb{D}^2(Y)$ endlich sind, dann sind auch $\mathbb{E}(XY)$, $\mathbb{E}(X)$ und $\mathbb{E}(Y)$ endlich, also existiert die Kovarianz von X und Y wegen der Behauptung.

Erinnerung

Beispiel. Seien $\operatorname{ran} X = \{2, 3, 5\}$, $\operatorname{ran} Y = \{0, 1, 2\}$. Die Wahrscheinlichkeiten $\mathbb{P}(X = s,\ Y = t)$ sind in der folgenden Tabelle zusammengefasst:

Y \ X	2	3	5
0	0,05	0,15	0,1
1	0,1	0,2	0,1
2	0,05	0,2	0,05

Wir bestimmten früher die Werte $\mathbb{E}(X) = 3{,}3$, $\mathbb{E}(Y) = 1$ und $\mathbb{E}(XY) = 3{,}2$. Also gilt

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 3{,}2 - 3{,}3 \cdot 1 = -0{,}1.$$

Kovarianz

Behauptung. Seien X, Y und Z auf demselben Wahrscheinlichkeitsraum definierte Zufallsvariablen, deren Varianzen $\mathbb{D}^2(X)$, $\mathbb{D}^2(Y)$ und $\mathbb{D}^2(Z)$ endlich sind.

a) Die Kovarianz ist symmetrisch, d.h. $\mathrm{cov}\,(X,Y) = \mathrm{cov}\,(Y,X)$.

b) Die Kovarianz ist bilinear, also gelten

$$\mathrm{cov}\,(X, aY + bZ) = a \cdot \mathrm{cov}\,(X,Y) + b \cdot \mathrm{cov}\,(X,Z),$$
$$\mathrm{cov}\,(aX + bY, Z) = a \cdot \mathrm{cov}\,(X,Z) + b \cdot \mathrm{cov}\,(Y,Z)$$

für jede $a, b \in \mathbb{R}$.

c) Falls X oder Y konstant ist, dann gilt $\mathrm{cov}\,(X,Y) = 0$.

d) Falls X und Y unabhängig sind, dann gilt $\mathrm{cov}\,(X,Y) = 0$.

Kovarianz

Beweis. Jede in der Behautung gegebene Kovarianz existiert wegen der Endlichkeit von $\mathbb{D}^2(X)$, $\mathbb{D}^2(Y)$ und $\mathbb{D}^2(Z)$.

a)

$$\begin{aligned}\operatorname{cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(YX) - \mathbb{E}(Y)\mathbb{E}(X) = \operatorname{cov}(Y, X).\end{aligned}$$

Kovarianz

Beweis.

b) Wegen der Symmetrie ist es hinreichend, die erste Behauptung zu beweisen.

$$\operatorname{cov}(X, aY + bZ) =$$

$$= \mathbb{E}(X(aY + bZ)) - \mathbb{E}(X)\mathbb{E}(aY + bZ)$$

$$= \mathbb{E}(aXY + bXZ) - \mathbb{E}(X)(a\mathbb{E}(Y) + b\mathbb{E}(Z))$$

$$= a\mathbb{E}(XY) + b\mathbb{E}(XZ) - a\mathbb{E}(X)\mathbb{E}(Y) - b\mathbb{E}(X)\mathbb{E}(Z)$$

$$= a \cdot \operatorname{cov}(X, Y) + b \cdot \operatorname{cov}(X, Z).$$

Kovarianz

Beweis.

c) Wegen der Symmetrie ist es hinreichend, die Behauptung für die zweite Variable zu beweisen. Nehmen wir an, dass $Y = c$ konstant ist.

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(cX) - c\mathbb{E}(X) = 0.$$

Kovarianz

Beweis.

d) Nehmen wir an, dass X und Y unabhängig sind. Dann gilt $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, weil diese Werte existieren. Also

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0.$$

Kovarianz

Bemerkungen.

Teil c) Hier ist es genügend, dass X oder Y mit Wahrscheinlichkeit 1 konstant ist, also dass $\mathbb{P}(X = c) = 1$ oder $\mathbb{P}(Y = c) = 1$ für eine reelle Zalh c gilt. Auch dann gilt $\operatorname{cov}(X, Y) = 0$.

Teil d) Es gibt Zufallsvariablen X und Y mit $\operatorname{cov}(X, Y) = 0$, die nicht unabhängig sind. (Für ein Beispiel siehe Aufgabe 1. des 9. Übungsblatts.)

Kovarianz

Korollar. Seien X und Y auf demselben Wahrscheinlichkeitsraum definierte Zufallsvariablen, für die $\mathbb{D}^2(X)$ und $\mathbb{D}^2(Y)$ endlich sind. Dann ist auch $\mathbb{D}^2(X+Y)$ endlich, und

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2 \cdot \operatorname{cov}(X, Y).$$

Beweis.

$$\begin{aligned}
\mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2 \cdot \operatorname{cov}(X, Y) &= \\
&= \operatorname{cov}(X, X) + \operatorname{cov}(Y, X) + \operatorname{cov}(Y, Y) + \operatorname{cov}(X, Y) \\
&= \operatorname{cov}(X+Y, X) + \operatorname{cov}(X+Y, Y) \\
&= \operatorname{cov}(X+Y, X+Y) = \mathbb{D}^2(X+Y).
\end{aligned}$$

Kovarianz

Lemma. Sei X eine Zufallsvariable. Dann gilt

$$\mathbb{D}^2(X) = 0 \iff \mathbb{P}(X = c) = 1 \text{ für eine } c \in \mathbb{R}.$$

Beweis (Skizze). Man kann die Folgende beweisen:

$$\mathbb{P}(X = c) = 1 \Longrightarrow \mathbb{E}(X) = c.$$

Dann gilt

$$\begin{aligned}
\mathbb{P}(X = \mathbb{E}(X)) = 1 &\iff \mathbb{P}(X - \mathbb{E}(X) = 0) = 1 \\
&\iff \mathbb{P}((X - \mathbb{E}(X))^2 = 0) = 1 \\
&\Longrightarrow \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{D}^2(X) = 0.
\end{aligned}$$

Kovarianz

Lemma. Sei X eine Zufallsvariable. Dann gilt

$$\mathbb{D}^2(X) = 0 \Longleftrightarrow \mathbb{P}(X = c) = 1 \text{ für eine } c \in \mathbb{R}.$$

Beweis (Skizze).

$$0 = \mathbb{D}^2(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Man kann beweisen, dass der Erwartungswert einer nichtnegativen Zufallsvariablen nur dann 0 sein kann, falls die Variable mit Wahrscheinlichkeit 1 Null ist. Dann folgt

$$\mathbb{P}((X - \mathbb{E}(X))^2 = 0) = 1 \Longleftrightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1.$$

Kovarianz

Lemma. Seien X und Y Zufallsvariablen, für die $\mathbb{D}^2(X)$ und $\mathbb{D}^2(Y)$ endlich sind. Dann gilt

$$|\mathrm{cov}\,(X, Y)| \leq \frac{\mathbb{D}^2(X) + \mathbb{D}^2(Y)}{2}$$

Beweis. Die Varianz ist nichtnegativ, also

$$\begin{aligned} 0 \leq \mathbb{D}^2(X \pm Y) &= \mathbb{D}^2(X) + \mathbb{D}^2(\pm Y) + 2 \cdot \mathrm{cov}\,(X, \pm Y) \\ &= \mathbb{D}^2(X) + \mathbb{D}^2(Y) \pm 2 \cdot \mathrm{cov}\,(X, Y). \end{aligned}$$

Nach Umordnung ergibt sich die Behauptung.

Kovarianz

Korollar. Seien X und Y Zufallsvariablen, für die $\mathbb{D}^2(X)$ und $\mathbb{D}^2(Y)$ endlich sind. Dann gilt

$$|\mathrm{cov}\,(X, Y)| \leq \mathbb{D}(X)\mathbb{D}(Y)$$

Beweis. Falls $\mathbb{D}(X)$ oder $\mathbb{D}(Y)$ Null ist, dann ist X oder Y mit Wahrscheinlichkeit 1 konstant. In diesem Fall gilt auch $\mathrm{cov}\,(X, Y) = 0$.

Kovarianz

Korollar. Seien X und Y Zufallsvariablen, für die $\mathbb{D}^2(X)$ und $\mathbb{D}^2(Y)$ endlich sind. Dann gilt

$$|\operatorname{cov}(X, Y)| \leq \mathbb{D}(X)\mathbb{D}(Y)$$

Beweis. Sonst ist die Ungleichung äquivalent zu

$$\begin{aligned} 1 &\geq \frac{1}{\mathbb{D}(X)\mathbb{D}(Y)} |\operatorname{cov}(X, Y)| \\ &= \left| \operatorname{cov}\left(\frac{X}{\mathbb{D}(X)}, \frac{Y}{\mathbb{D}(Y)} \right) \right|. \end{aligned}$$

Kovarianz

Beweis. Sonst ist die Ungleichung äquivalent zu

$$1 \geq \frac{1}{\mathbb{D}(X)\mathbb{D}(Y)} \left|\operatorname{cov}(X, Y)\right|$$

$$= \left|\operatorname{cov}\left(\frac{X}{\mathbb{D}(X)}, \frac{Y}{\mathbb{D}(Y)}\right)\right|.$$

Wegen des zweiten Lemmas gilt aber

$$\left|\operatorname{cov}\left(\frac{X}{\mathbb{D}(X)}, \frac{Y}{\mathbb{D}(Y)}\right)\right| \leq \frac{\mathbb{D}^2\left(\frac{X}{\mathbb{D}(X)}\right) + \mathbb{D}^2\left(\frac{Y}{\mathbb{D}(Y)}\right)}{2} = 1.$$

Kovarianz

Definition. Seien X und Y Zufallsvariablen mit $0 < \mathbb{D}(X) < \infty$ und $0 < \mathbb{D}(Y) < \infty$, dann ist *der Korrelationskoeffizient* (korrelációs együttható) von X und Y durch

$$\operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}$$

definiert.

Bemerkung. Falls $\operatorname{corr}(X, Y) = 0$, dann sagen wir, dass X und Y *unkorreliert* (korrelálatlanok) sind. Zum Beispiel, falls X und Y unabhängig sind, dann sind sie unkorreliert.

Kovarianz

Korollar. Seien X und Y Zufallsvariablen mit $0 < \mathbb{D}(X) < \infty$ und $0 < \mathbb{D}(Y) < \infty$, dann gilt

$$-1 \leq \operatorname{corr}(X, Y) \leq 1.$$

Korrelationskoeffizient

Satz. Seien X und Y Zufallsvariablen mit $0 < \mathbb{D}(X) < \infty$ und $0 < \mathbb{D}(Y) < \infty$. Die Gleichung

$$\operatorname{corr}(X, Y) = \pm 1$$

gilt dann und nur dann, wenn es reelle Zahlen $\alpha, \beta \in \mathbb{R}$, $\beta \neq 0$ mit

$$\mathbb{P}(Y = \beta X + \alpha) = 1$$

gibt. In diesem Fall sind die Vorzeichen von β und $\operatorname{corr}(X, Y)$ gleich.

Korrelationskoeffizient

Bemerkung. Wenn das Verhältnis von X und Y nicht linear ist, dann kann der Korrelationskoeffizient eine beliebige Zahl in $(-1;1)$ sein. Zum Beilspiel, es gibt eine solche Zufallsvariable X, für die $\operatorname{corr}(X, X^2) = 0$ gilt (siehe Aufgabe 6. auf dem 10. Übungsblatt).

Korrelationskoeffizient

Beweis. Nehmen wir an, dass $\mathbb{P}(Y = \beta X + \alpha) = 1$ für einige reele Zahlen $\alpha, \beta \in \mathbb{R}$. Dann ist $Y - \beta X$ konstant, also gilt $\mathbb{D}^2(Y - \beta X) = 0$ (siehe das erste Lemma oben). D.h.

$$
\begin{aligned}
0 = \mathbb{D}^2(Y - \beta X) &= \operatorname{cov}(Y - \beta X, Y - \beta X) \\
&= \operatorname{cov}(Y, Y - \beta X) - \beta \cdot \operatorname{cov}(X, Y - \beta X) \\
&= \operatorname{cov}(Y, Y) - \beta \cdot \operatorname{cov}(Y, X) - \beta \cdot \operatorname{cov}(X, Y) + \beta^2 \cdot \operatorname{cov}(X, X) \\
&= \beta^2 \mathbb{D}^2(X) - 2\beta \cdot \operatorname{cov}(X, Y) + \mathbb{D}^2(Y).
\end{aligned}
$$

Dann ist β eine Wurzel des Polynoms

$$q(x) = \mathbb{D}^2(X) \cdot x^2 - 2\operatorname{cov}(X, Y) \cdot x + \mathbb{D}^2(Y).$$

Korrelationskoeffizient

Beweis. Das quadratische Polynom $q(x)$ hat eine reelle Wurzel, deshalb ist seine Diskriminante nichtnegativ:

$$\begin{aligned}4\operatorname{cov}^2(X,Y) - 4\mathbb{D}^2(X)\mathbb{D}^2(Y) &= \\ &= 4(\operatorname{cov}^2(X,Y) - \mathbb{D}^2(X)\mathbb{D}^2(Y)) \geq 0,\end{aligned}$$

d.h.

$$|\operatorname{cov}(X,Y)| \geq \mathbb{D}(X)\mathbb{D}(Y),$$

$$|\operatorname{corr}(X,Y)| = \frac{|\operatorname{cov}(X,Y)|}{\mathbb{D}(X)\mathbb{D}(Y)} \geq 1.$$

Aber $|\operatorname{corr}(X,Y)| \leq 1$ gilt auch, also folgt $\operatorname{corr}(X,Y) = \pm 1$.

Korrelationskoeffizient

Beweis.

$$0 = q(\beta) = \mathbb{D}^2(X) \cdot \beta^2 - 2\mathrm{cov}\,(X, Y) \cdot \beta + \mathbb{D}^2(Y),$$

also

$$2\mathrm{cov}\,(X, Y) \cdot \beta = \mathbb{D}^2(X) \cdot \beta^2 + \mathbb{D}^2(Y) > 0.$$

Es folgt, dass die Vorzeichen von β und $\mathrm{cov}\,(X, Y)$ gleich sind, aber die Vorzeichen von $\mathrm{cov}\,(X, Y)$ und $\mathrm{corr}\,(X, Y)$ sind auch gleich.

Korrelationskoeffizient

Beweis.
Nehmen wir an, dass $\operatorname{corr}(X, Y) = \pm 1$ gilt. Das ist äquivalent zu

$$\operatorname{cov}^2(X, Y) = \mathbb{D}^2(X)\mathbb{D}^2(Y),$$

und dann ist die Diskriminante des Polynoms $q(x)$ Null ist, und deshalb hat $q(x)$ eine eindeutige reelle Wurzel β.

Wir haben gesehen, dass

$$q(\beta) = 0 \Longleftrightarrow \mathbb{D}^2(Y - \beta X) = 0$$

gilt, also gilt auch

$$\mathbb{P}(Y = \beta X + \alpha) = 1$$

für eine $\alpha \in \mathbb{R}$. Wir haben noch $\beta \neq 0$, weil Y nicht mit Wahscheinlichkeit 1 konstant ist.

Korrelationskoeffizient

Beweis.
Wegen des ersten Teils des Beweises folgt endlich, dass die Vorzeichen von β und $\mathrm{corr}\,(X, Y)$ gleich sind.

Lineare Regression

- Im Folgenden betrachten wir zwei Zufallsvariablen X und Y mit $0 < \mathbb{D}(X), \mathbb{D}(Y) < \infty$. Wir möchten die Variable Y durch eine transformierte Variable $\beta X + \alpha$ approximieren.
- Wir suchen die "beste" Approximation. Es ist aber nicht eindeutig, was das Wort "best" bedeutet.
- Wir benutzen die Methode der kleinsten Quadrate. Also, wir suchen die Zahlen β und α, für die das Quadrat der Differenz $Y - (\beta X + \alpha)$ im Durchschnitt minimal ist.

Lineare Regression

Definition. Seien X und Y Zufallsvariablen. Die *lineare Regression von Y auf X* (az Y-nak az X-re vett lineáris regressziója) ist eine Zufallsvariable $\beta X + \alpha$ mit $\alpha, \beta \in \mathbb{R}$, für die der Erwartungswert

$$\mathbb{E}\left((Y - (\beta X + \alpha))^2\right)$$

minimal ist.

Lineare Regression

Satz. Seien X und Y Zufallsvariablen, für die $\mathbb{D}(X), \mathbb{D}(Y) < \infty$ und $\mathbb{D}(X) \neq 0$ gelten. Dann existiert die eindeutig bestimmte lineare Regression $\tilde{Y} := \beta X + \alpha$ von Y auf X.
Die Zahlen β und α sind durch

$$\beta = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}, \qquad \alpha = \mathbb{E}(Y) - \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}\mathbb{E}(X)$$

angegeben.

Lineare Regression

Wie kann man sich an die obige Formeln für α und β erinnern?

Die Zahlen β und α sind so gewählt, dass

$$\mathbb{E}(Y) = \mathbb{E}(\tilde{Y}) \quad \text{und} \quad \operatorname{cov}(X, Y) = \operatorname{cov}(X, \tilde{Y})$$

gelten.

Lineare Regression

Also

$$\begin{aligned}\operatorname{cov}(X, Y) &= \operatorname{cov}(X, \beta X + \alpha)\\ &= \beta \cdot \operatorname{cov}(X, X) + \operatorname{cov}(X, \alpha)\\ &= \beta \cdot \mathbb{D}^2(X),\end{aligned}$$

d.h.

$$\beta = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}.$$

Weiter,

$$\mathbb{E}(Y) = \mathbb{E}(\beta X + \alpha) = \beta\mathbb{E}(X) + \alpha,$$

also

$$\alpha = \mathbb{E}(Y) - \beta\mathbb{E}(X) = \mathbb{E}(Y) - \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}\mathbb{E}(X).$$

Lineare Regression

Definition. Seien X und Y Zufallsvariablen mit $\mathbb{D}(X), \mathbb{D}(Y) < \infty$ und $\mathbb{D}(X) > 0$. Die Gerade

$$\{(x, y) \in \mathbb{R}^2 : \ y = \beta x + \alpha\}$$

heißt *die Regressionsgerade* von Y auf X (az Y-nak az X-re vett regressziós egyenese).

Bemerkung. Wenn die Werte der Vektorvariablen (X, Y) mit großer Wahrscheinlichkeit in der Nähe der Regressionsgerade liegen, dann ist die lineare Regression eine gute Approximation.

Lineare Regression

Beweis. Sei

$$\begin{aligned}
h(\alpha,\beta) &= \mathbb{E}\left((Y-(\beta X+\alpha))^2\right) \\
&= \mathbb{E}(Y^2+\beta^2X^2+2\alpha\beta X+\alpha^2-2\beta XY-2\alpha Y) \\
&= \mathbb{E}(Y^2)+\beta^2\mathbb{E}(X^2)+2\alpha\beta\mathbb{E}(X)+\alpha^2 \\
&\qquad -2\beta\mathbb{E}(XY)-2\alpha\mathbb{E}(Y).
\end{aligned}$$

$h(\alpha,\beta)$ ist eine nichtnegative Polynomfunktion in 2 Variablen vom totalen Grad 2.

Lineare Regression

Beweis.
Falls die Funktion h an einer Stelle (α_0, β_0) ein lokales Minimum hat, dann verschwinden ihre partiellen Ableitungen an der Stelle (α_0, β_0).

Falls die partiellen Ableitungen von h an einer Stelle (α_0, β_0) Null sind, und der Ausdurck

$$\frac{\partial^2 h}{\partial^2 \alpha}(\alpha_0, \beta_0)\frac{\partial^2 h}{\partial^2 \beta}(\alpha_0, \beta_0) - \left[\frac{\partial^2 h}{\partial \alpha\, \partial \beta}(\alpha_0, \beta_0)\right]^2$$

(also die Determinante der Hesse-Matrix von h), und auch

$$\frac{\partial^2 h}{\partial^2 \alpha}(\alpha_0, \beta_0)$$

positiv sind, dann hat h ein lokales Minimum an der Stelle (α_0, β_0).

Lineare Regression

Beweis.

$$\frac{\partial h}{\partial \alpha}(\alpha, \beta) = 2\alpha + 2\beta\mathbb{E}(X) - 2\mathbb{E}(Y),$$

$$\frac{\partial^2 h}{\partial^2 \alpha}(\alpha, \beta) = 2,$$

$$\frac{\partial h}{\partial \beta}(\alpha, \beta) = 2\beta\mathbb{E}(X^2) + 2\alpha\mathbb{E}(X) - 2\mathbb{E}(XY),$$

$$\frac{\partial^2 h}{\partial^2 \beta}(\alpha, \beta) = 2\mathbb{E}(X^2),$$

$$\frac{\partial^2 h}{\partial \alpha\, \partial \beta}(\alpha, \beta) = 2\mathbb{E}(X),$$

Lineare Regression

Beweis.
Also gelten

$$\frac{\partial^2 h}{\partial^2 \alpha}(\alpha, \beta) = 2 > 0,$$

und

$$\begin{aligned}\frac{\partial^2 h}{\partial^2 \alpha}(\alpha, \beta)\frac{\partial^2 h}{\partial^2 \beta}(\alpha, \beta) - \left[\frac{\partial^2 h}{\partial \alpha\, \partial \beta}(\alpha, \beta)\right]^2 &= 4\mathbb{E}(X^2) - 4\mathbb{E}(X)^2, \\ &= 4\mathbb{D}^2(X) > 0\end{aligned}$$

für jede (α, β).

Lineare Regression

Beweis.
Wir bestimmen α_0 und β_0, für die die partiellen Ableitungen Null sind. Es gilt

$$0 = \frac{\partial h}{\partial \alpha}(\alpha_0, \beta_0) = 2\alpha_0 + 2\beta_0 \mathbb{E}(X) - 2\mathbb{E}(Y),$$

$$0 = \frac{\partial h}{\partial \beta}(\alpha_0, \beta_0) = 2\beta_0 \mathbb{E}(X^2) + 2\alpha_0 \mathbb{E}(X) - 2\mathbb{E}(XY),$$

also sind α_0 und β_0 Lösungen des Systems

$$\alpha_0 + \mathbb{E}(X)\beta_0 = \mathbb{E}(Y),$$

$$\mathbb{E}(X)\alpha_0 + \mathbb{E}(X^2)\beta_0 = \mathbb{E}(XY).$$

Lineare Regression

Beweis.
Die erste Gleichung gibt

$$\alpha_0 = \mathbb{E}(Y) - \mathbb{E}(X)\beta_0,$$

und dann gibt die zweite Gleichung

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}(X)\alpha_0 + \mathbb{E}(X^2)\beta_0 \\ &= \mathbb{E}(X)(\mathbb{E}(Y) - \mathbb{E}(X)\beta_0) + \mathbb{E}(X^2)\beta_0 \\ &= \mathbb{E}(X)\mathbb{E}(Y) + \beta_0(\mathbb{E}(X^2) - \mathbb{E}(X)^2),\end{aligned}$$

also

$$\beta_0 = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\mathbb{E}(X^2) - \mathbb{E}(X)^2} = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}.$$

Lineare Regression

Beweis.
Wenn h ein globales Minimum hat, dann ist es auch ein lokales Mininmum, also muss h dieses Minimum an der Stelle (α_0, β_0) annehmen.

Nehmen wir an, dass h kein globales Minimum hat. Dann gibt es ein Paar (α_1, β_1), für das $h(\alpha_1, \beta_1) < h(\alpha_0, \beta_0)$ gilt. Die Punkte der Gerade, die durch die Punkte (α_0, β_0) und (α_1, β_1) bestimmt ist, können in der Form

$$(t\alpha_0 + (1-t)\alpha_1, t\beta_0 + (1-t)\beta_1)$$

geschrieben werden. Wenn wir die Funktion h auf diese Gerade einschränken, dann bekommen wir ein Polynom

$$q(t) = h(t\alpha_0 + (1-t)\alpha_1, t\beta_0 + (1-t)\beta_1)$$

von Grad höchstens 2.

Lineare Regression

Beweis.

$$q(t) = h(t\alpha_0 + (1-t)\alpha_1, t\beta_0 + (1-t)\beta_1)$$

Das Polynom $q(t)$ hat ein lokales Minimum an der Stelle $t = 1$, also ist $q(t)$ quadratisch, und $t = 1$ ist ein globales Minimum von q, d.h.

$$h(\alpha_1, \beta_1) = q(0) > q(1) = h(\alpha_0, \beta_0).$$

Widerspruch.

Lineare Regression

Korollar. Seien X und Y Zufallsvariablen mit $\mathbb{D}^2(X), \mathbb{D}^2(Y) < \infty$ und $\mathbb{D}^2(X) > 0$. Sei noch $\beta X + \alpha$ die lineare Regression von Y auf X. Dann gilt

$$\begin{aligned}\mathbb{E}\left((Y - (\beta X + \alpha))^2\right) &= \mathbb{D}^2\left(Y - (\beta X + \alpha)\right) \\ &= \mathbb{D}^2(Y) - \frac{\operatorname{cov}(X, Y)^2}{\mathbb{D}^2(X)}.\end{aligned}$$

Falls auch $\mathbb{D}^2(Y) > 0$ gilt, dann ist dieser Ausdruck

$$\mathbb{D}^2(Y)(1 - \operatorname{corr}(X, Y)^2).$$

Also misst der Korrelationskoeffizient die lineare Abhängigkeit von Zufallsvariablen.

Lineare Regression

Beweis des Korollars. Für

$$\beta = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}, \qquad \alpha = \mathbb{E}(Y) - \frac{\operatorname{cov}(X, Y)}{\mathbb{D}^2(X)}\mathbb{E}(X)$$

haben wir

$$\mathbb{E}(Y - (\beta X + \alpha)) = 0,$$

also

$$\begin{aligned}\mathbb{D}^2(Y - (\beta X + \alpha)) &= \mathbb{E}\left((Y - (\beta X + \alpha))^2\right) - \mathbb{E}(Y - (\beta X + \alpha))^2 \\ &= \mathbb{E}\left((Y - (\beta X + \alpha))^2\right).\end{aligned}$$

Lineare Regression

Beweis des Korollars. Weiter,

$$\begin{aligned}
\mathbb{D}^2(Y - (\beta X + \alpha)) &= \mathbb{D}^2(Y - \beta X) \\
&= \mathbb{D}^2(Y) + \beta^2 \mathbb{D}^2(X) - 2\beta \cdot \operatorname{cov}(X, Y) \\
&= \mathbb{D}^2(Y) + \frac{\operatorname{cov}(X, Y)^2}{\mathbb{D}^2(X)} - 2\frac{\operatorname{cov}(X, Y)^2}{\mathbb{D}^2(X)} \\
&= \mathbb{D}^2(Y) - \frac{\operatorname{cov}(X, Y)^2}{\mathbb{D}^2(X)}.
\end{aligned}$$

Lineare Regression

Bemerkung. Nehmen wir an, dass $0 < \mathbb{D}(X), \mathbb{D}(Y) < \infty$ gelten. Es folgt dann aus dem Korollar, dass $|\operatorname{corr}(X, Y)| = 1$ dann und nur dann gilt, wenn

$$\mathbb{D}^2 (Y - (\beta X + \alpha)) = 0,$$

also wenn $\mathbb{P}(Y = \beta X + \alpha) = 1$ gilt, wo $\beta X + \alpha$ die lineare Regression von Y auf X ist.

In diesem Fall sind das Vorzeichen von β und das Vorzeichen von $\operatorname{cov}(X, Y)$ gleich. Aber das Vorzeichen von $\operatorname{cov}(X, Y)$ ist gleich das Vorzeichen von $\operatorname{corr}(X, Y)$, und wir haben den folgenden Satz bewiesen:

Lineare Regression

Satz. Seien X und Y Zufallsvariablen mit $0 < \mathbb{D}(X) < \infty$ und $0 < \mathbb{D}(Y) < \infty$. Die Gleichung

$$\operatorname{corr}(X, Y) = \pm 1.$$

gilt dann und nur dann, wenn es reelle Zahlen $\alpha, \beta \in \mathbb{R}$, $\beta \neq 0$ mit

$$\mathbb{P}(Y = \beta X + \alpha) = 1$$

gibt. In diesem Fall ist $\beta X + \alpha$ die lineare Regression von Y auf X, und die Vorzeichen von β und $\operatorname{corr}(X, Y)$ sind gleich.

Bedingter Erwartungswert

Erinnerung. Der Korrelationskoeffizient misst den linearen Zusammenhang zwischen zwei Zufallsvariablen, aber er zeigt andere Zusammenhänge nicht.

Zum Beispiel gibt es eine Zufallsvariable X, für die

$$\operatorname{corr}(X, X^2) = 0$$

gilt. Dann ist der Wert von X^2 bestimmt, wenn der Wert von X bekannt ist, trotzdem ist die (konstante) lineare Regression von X^2 auf X nicht eine gute Approximation von X^2.

Wie kann eine bessere Aproximation angegeben werden? Hier wird der bedigten Erwartungswert verwendet.

Bedingter Erwartungswert

Erinnerung. Wenn A ein Ereignis mit $\mathbb{P}(A) > 0$ ist, dann definiert die bedingte Wharscheilnichkeit

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

ein Wahrscheinlichkeitsmaß.

Definition. Sei Y eine Zufallsvariable, und sei A ein Ereignis mit $\mathbb{P}(A) > 0$. Dann wird *der Erwartungswert von Y unter der Bedingung A* (az Y-nak az A-ra vett feltétles várható értéke) als der Erwartungswert von Y bezüglich des bedingten Wahrscheinlichkeitsmaßes $\mathbb{P}(\cdot \mid A)$ definiert, falls dieser Wert existiert.

Notation: $\mathbb{E}(Y \mid A)$.

Bedingter Erwartungswert

Beispiel. Sei Y eine einfache Zufallsvariable, und sei A ein Ereignis mit $\mathbb{P}(A) > 0$, dann gilt

$$\begin{aligned}
\mathbb{E}(Y \mid A) &= \sum_{k \in \operatorname{ran} Y} k \cdot \mathbb{P}(Y = k \mid A) \\
&= \sum_{k \in \operatorname{ran} Y} k \cdot \frac{\mathbb{P}(\{Y = k\} \cap A)}{\mathbb{P}(A)} \\
&= \frac{1}{\mathbb{P}(A)} \sum_{k \in \operatorname{ran} Y} k \cdot \mathbb{P}(Y \cdot \mathbb{1}_A = k) \\
&= \frac{1}{\mathbb{P}(A)} \sum_{k \in \operatorname{ran} (Y \mathbb{1}_A)} k \cdot \mathbb{P}(Y \cdot \mathbb{1}_A = k) \\
&\qquad + \frac{1}{\mathbb{P}(A)} \sum_{k \in \operatorname{ran} Y \setminus \operatorname{ran} (Y \mathbb{1}_A)} k \cdot 0 = \frac{1}{\mathbb{P}(A)} \mathbb{E}(Y \mathbb{1}_A).
\end{aligned}$$

Bedingter Erwartungswert

Der obige Zusammenhang gilt auch im Allgemeinen:

Behauptung. Sei Y eine Zufallsvariable, deren Erwartungswert endlich ist. Sei noch A ein Ereignis mit $\mathbb{P}(A) > 0$, dann gilt

$$\mathbb{E}(Y \mid A) = \frac{1}{\mathbb{P}(A)} \mathbb{E}(Y \mathbb{1}_A).$$

Bedingter Erwartungswert

Wie kann dieser Begriff in einer Aufgabe auftreten?

Beispiel. Nehmen wir an, dass der Erwartungwert der Lebenszeit eines Apparats 3 Jahre ist, und der Erwartungwert der Lebenszeit eines anderen Apparats 4 Jahre ist. Wir wählen einen von den zwei Apparaten zufällig. Sei A das Ereignis, dass wir den ersten Apparat wählen, und sei B das Ereignis, dass wir den zweiten Apparat wählen. Sei Y die Lebenszeit des gewählten Apparats. Dann gilt

$$3 = \mathbb{E}(Y|A) = \frac{1}{\mathbb{P}(A)}\mathbb{E}(Y\mathbb{1}_A), \qquad 4 = \mathbb{E}(Y|B) = \frac{1}{\mathbb{P}(B)}\mathbb{E}(Y\mathbb{1}_B),$$

also

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(Y(\mathbb{1}_A + \mathbb{1}_B)) = \mathbb{E}(Y\mathbb{1}_A + Y\mathbb{1}_B)\\ &= \mathbb{E}(Y\mathbb{1}_A) + \mathbb{E}(Y\mathbb{1}_B) = 3\mathbb{P}(A) + 4\mathbb{P}(B).\end{aligned}$$

Bedingter Erwartungswert

Eine kleine Modifikation der obigen Argumentation gibt die folgende Behauptung:

Satz. (Satz vom totalen Erwartungswert für Ereignisse) Sei Y eine Zufallsvariable, und sei noch $A_1, \ldots, A_n$ ein vollständiges Ereignissystem mit $\mathbb{P}(A_i) > 0$ für alle $1 \leq i \leq n$. Falls $\mathbb{E}(Y)$ endlich ist, dann gilt

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(Y \mid A_i)\mathbb{P}(A_i).$$

Bedingter Erwartungswert

Beweis. Seien $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ die Indikatorvariablen der Ereignisse $A_1, \ldots, A_n$. Da diese Ereignisse ein vollständiges Ereignissystem bilden, folgt

$$\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n} = 1,$$

und deshalb gilt

$$\mathbb{E}(Y) = \mathbb{E}\left(Y(\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n})\right) = \sum_{i=1}^{n} \mathbb{E}(Y\mathbb{1}_{A_i}).$$

Der Satz folgt aus der letzten Behauptung, d.h. aus der Gleichung

$$\mathbb{E}(Y\mathbb{1}_{A_i}) = \mathbb{P}(A_i)\mathbb{E}(Y \mid A_i).$$

Bedingter Erwartungswert

Bemerkung. Seien X und Y Zufallsvariablen, wo $\mathbb{E}(Y)$ endlich ist. Als Bedingung kann ein Ereignis $\{X \leq s\}$ gewählt werden (oder $\{X = s\}$, wenn $\mathbb{P}(X = s) > 0$ gilt). D.h., $\mathbb{E}(Y \mid X \leq s)$ (bzw. $\mathbb{E}(Y \mid X = s)$) ist definiert.

Bedingter Erwartungswert

Beispiel. Zwei Würfel werden geworfen, seien Y und Z die Augenzahlen, und sei noch $X = Y + Z$. Wir berechnen $\mathbb{E}(Y \mid X \leq 6)$:

$$\mathbb{P}(Y = k \mid X \leq 6) = \frac{\mathbb{P}(Y = k, X \leq 6)}{\mathbb{P}(X \leq 6)} = \frac{6-k}{36} \Big/ \frac{15}{36} = \frac{6-k}{15}$$

für $k = 1, \ldots, 6$, also

$$\begin{aligned}\mathbb{E}(Y \mid X \leq 6) &= \sum_{k=1}^{6} k \cdot \mathbb{P}(Y = k \mid X \leq 6) \\ &= \sum_{k=1}^{6} k \cdot \frac{6-k}{15} = \frac{6 \cdot 21 - 91}{15} = \frac{7}{3}.\end{aligned}$$

Bedingter Erwartungswert

Beispiel. Weiter,

$$\mathbb{E}(Y \mid X = 5) = \sum_{k=1}^{6} k \cdot \mathbb{P}(Y = k \mid X = 5) = \sum_{k=1}^{4} k \cdot \frac{1}{4} = \frac{5}{2}.$$

Bedingter Erwartungswert

Bemerkung. Der Satz der totalen Wahrscheinlichkeit und der Satz vom totalen Erwartungswert bleiben auch dann richtig, falls $A_1, \ldots, A_n$ paarweise disjunkte Ereignisse sind und

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1$$

gilt. D.h.: die Vereinigung muss nicht die ganze Menge Ω sein, aber die Wahrscheinlichkeit der Differenz $\Omega \setminus (\cup_{i=1}^{n} A_i)$ muss Null sein. Hier können wir sogar abzählbar unendlich viele Ereignisse A_i betrachten, die obigen Sätze gelten auch dann.

Bedingter Erwartungswert

Für eine diskrete Variable X definieren wir die Menge

$$S_X = \{s \in \mathbb{R} : \mathbb{P}(X = s) > 0\},$$

dann erfüllen die Ereignisse $\{X = s\}$ (wo s die Menge S_X durchläuft) die obigen Eigenschaften, also folgt der folgende

Satz. (Satz vom totalen Erwartungswert - diskreter Fall)
Seien X, Y Zufallsvariablen, wo $\mathbb{E}(Y)$ endlich ist. Falls X diskret ist, und

$$S_X = \{s \in \mathbb{R} : \mathbb{P}(X = s) > 0\},$$

dann gilt

$$\mathbb{E}(Y) = \sum_{s \in S_X} \mathbb{E}(Y \mid X = s)\mathbb{P}(X = s).$$

Bedingter Erwartungswert

Beispiel. Sei $Y \sim \mathrm{Geo}(p)$, wir berechenen $\mathbb{D}^2(Y)$ noch einmal. Wir können annehmen, dass Y die Anzalh von unabhängigen Experimenten bis zum ersten Erfolg ist, wo die Wahrscheinlichkeit des Erfolgs p ist.

Sei A das Ereignis, dass das erste Experiment erfolgreich ist. Dann gilt

$$\mathbb{E}(Y^2) = \mathbb{E}(Y^2 \mid A)\mathbb{P}(A) + \mathbb{E}(Y^2 \mid \overline{A})\mathbb{P}(\overline{A}).$$

Wir haben $\mathbb{E}(Y^2 \mid A) = 1$ und $\mathbb{P}(A) = p$.

Bedingter Erwartungswert

Weiter, wegen der Gedächtnislosigkeit gilt

$$\begin{aligned}
\mathbb{E}(Y^2 \mid \overline{A}) &= \sum_{k=2}^{\infty} k^2 \cdot \mathbb{P}(Y = k \mid Y > 1) \\
&= \sum_{k=1}^{\infty} (k+1)^2 \cdot \mathbb{P}(Y = k+1 \mid Y > 1) \\
&= \sum_{k=1}^{\infty} (k+1)^2 \cdot \mathbb{P}(Y = k) \\
&= \mathbb{E}((Y+1)^2) = \mathbb{E}(Y^2 + 2Y + 1)
\end{aligned}$$

Bedingter Erwartungswert

Also

$$\mathbb{E}(Y^2) = p + (1-p)(\mathbb{E}(Y^2) + 2\mathbb{E}(Y) + 1),$$

d.h.

$$p\mathbb{E}(Y^2) = 1 + \frac{2(1-p)}{p} = \frac{2-p}{p},$$

und dann

$$\mathbb{D}^2(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Randomisierter Quicksort

Anwendung: Durchschnittliche Laufzeit von Algorithmen

Definition. (Monte-Carlo- und Las-Vegas-Algorithmus)
Ein randomisierter Algorithmus, der (wenn er terminiert) immer eine richtige Lösung des Problems ausgibt, heißt *Las-Vegas-Algorithmus*.
Ein Algorithmus, der mit positiver Wahrscheinlichkeit ein falsches Ergebnis ausgibt, heißt *Monte-Carlo-Algorithmus*.

Randomisierter Quicksort

Der *randomisierte Quicksort* ist ein Beispiel für einen Las-Vegas-Algorithmus.

- Er ist ein elementarer Sortieralgorithmus.
- Er sortiert in einem rekursiven Verfahren die Eingabe, die aus einer Liste von Zahlen besteht.
- Bei jedem Schritt wird die Eingabe partitioniert, wo das Vergleichselement zufällig gewählt wird.

Randomisierter Quicksort

Randomisierter Quicksort:
Eingabe: Eine Sequenz $S = (x_1, \ldots, x_n)$ von paarweise verschiedenen Zahlen.

1. Wähle (gleichverteilt) zufällig ein Element Y von S.
2. Vergleiche jedes Element von $S \setminus \{Y\}$ mit Y, und gebe die Sequenzen $S_< := (y_1, \ldots, y_k)$ und $S_> := (z_1, \ldots, z_l)$ an, wo
   - $y_i < Y$ bzw. $z_j > Y$ für alle $1 \leq i \leq k$ bzw. $1 \leq j \leq l$ gelten,
   - die Reihenfolge der Elemente in $S_<$ und $S_>$ gleich wie in S ist,
   - und jedes Element von S in $S_<$ oder $S_>$ aufgelistet wird.
3. Ordne rekursiv die Sequeunzen $S_<$ und $S_>$ durch Wiederholen von 1. und 2.
4. Gebe die sortierte Sequenzen $S_<$ gefolgt von Y gefolgt von der sortierten Sequenz $S_>$ aus.

Ausgabe: Die Elemente von S aufsteigend sortiert.

Randomisierter Quicksort

Bemerkungen:

- Die Sequenzen $S_<$ und $S_>$ haben in jedem Fall eine kleinere Anzahl von Elementen als die ursprüngliche Sequenz S.
- Eine ein- oder nullelementige Menge ist offensichtlich geordnet, somit terminiert der Algorithmus nach endlich vielen Schritten.

Randomisierter Quicksort

Beispiel. Sei $S = (17, 6, 56, 23, 4, 19, 58, 28, 14, 43, 62, 3)$.

Wird $Y = 23$ als Vergleichselement ausgewählt, so erhält man

$$S_< = (17, 6, 4, 19, 14, 3) \quad \text{und} \quad S_> = (56, 58, 28, 43, 62).$$

Dann sortiert man die Sequnezen $S_<$ und $S_>$ rekursiv.

Randomisierter Quicksort

Laufzeit bei Quicksort.

Sei X_S die Anzahl der durchgeführten Paarvergleiche bei Eingabe einer Sequenz S.

- Die Zufallsvariable X_S misst die Laufzeit des randomisierten Quicksort.
- Bei einer unglücklichen Wahl der Vergleichselemente kann der Laufzeit so schlecht sein, wie möglich ist, d.h. es ist möglich, dass alle Paare der Glieder der Sequenz verglichen werden.
- Z.B.: Sei $S = (1, 2, \ldots, n)$, und nehmen wir an, dass das (zufällig gewählte) Vergleichselement in jedem Schritt des Algorithmus das letzte Element wird.
  Dann wird $S_< = (1, \ldots, n-k)$ und $S_> = \emptyset$ im k-ten Schritt des Algorithmus, also werden alle möglichen Paare verglichen, d.h.: die Anzahl der Paarvergleiche ist in diesem Fall

$$(n-1) + (n-2) + \cdots + 1 = \frac{n(n-1)}{2} = \binom{n}{2} = O(n^2).$$

Randomisierter Quicksort

Bemerkung.

- Wir nehmen immer an, dass die Eingabe eine (gleichverteilt) zufällige Sequenz von n paarweise verschiedenen Elementen ist.
- Der Wert von X_S hängt offensichtlich nur von der Anzahl der Elementen von S ab, weil nur die Relation der Elemente den Lauf des Algorithmus beeinflusst.
- Sei also X_n die Anzahl der durchgeführten Paarvergleiche, falls die Eingabe die Zahlen $1, 2, \ldots, n$ enthält. Dann werden die Variablen X_S und X_n für alle Sequenzen S mit $|S| = n$ identisch verteilt (hier und später bezeichnet $|S|$ die Länge von S).

Randomisierter Quicksort

Laufzeit bei Quicksort.

Zwar die Laufzeit kann in einigen Fällen schlecht sein, aber sie ist durch die Randomisierung jedoch in Erwartung sehr viel kürzer:

Behauptung. Im randomisierten Quicksort gilt

$$\mathbb{E}(X_n) < 2nH_n,$$

wobei $H_n = \sum_{k=1}^{n} \frac{1}{k}$ von der Größenordnung $\ln n$ ist:

$$H_n = \ln n + \gamma + o(1).$$

wo $\gamma = 0{,}55721\ldots$ die Eulersche Konstante ist.
Die durchschnittliche Laufzeit ist also bei einer zu sortierenden Liste der Länge n von der Ordnung $O(n \ln n)$.

Randomisierter Quicksort

Laufzeit bei Quicksort.

Wie oft wird der Algorithmus wesentlich langsamer als im Durchsnitt? D.h.: mit welcher Wahrscheinlichkeit hat die Laufzeit eine größere Ordnung als $O(n \ln n)$?

Zum Beispiel: wie groß kann die Wahrscheinlichkeit $\mathbb{P}(X_n \geq cn^\alpha)$ sein, wo $c > 0$ eine Konstante und $\alpha > 1$ eine beliebige reelle Zahl sind?

Aus der Markow-Ungleichung folgt

$$\mathbb{P}(X_n \geq cn^\alpha) \leq \frac{\mathbb{E}(X_n)}{cn^\alpha} \leq \frac{c' n \ln n}{cn^\alpha},$$

wo $c' > 0$ eine andere Konstante ist. Die rechte Seite konvergiert gegen 0, falls $n \to \infty$, also sind die "schlechten Fälle" sehr unwahrscheinlich (falls n groß ist).

Randomisierter Quicksort

Beweis der Behauptung. Wir geben eine rekursive Formel für die Zahl $C(n) := \mathbb{E}(X_n)$.

- Offensichtlich gilt $C(0) = C(1) = 0$, also können wir $n > 1$ annehmen.
- Sei Y der Wert des im ersten Schritt gewählten Vergleichselements.
- Wir vergleichen alle anderen Elemente mit Y, das gibt $n - 1$ Vergleiche.
- Falls $Y = k$ gilt, dann erhält man im ersten Schritt, dass $|S_<| = k - 1$ und $|S_>| = n - k$ gelten.
- Man überlegt leicht, dass $S_<$ und $S_>$ uniform zufällig geordnete Sequenzen von $k - 1$ bzw. $n - k$ paarweise verschiedenen Elementen sind.

Randomisierter Quicksort

Beweis der Behauptung. Es gilt immer

$$X_n = n - 1 + X_{S_<} + X_{S_>},$$

und unter der Bedingung $Y = k$ ergibt sich

$$\begin{aligned}
\mathbb{E}(X_n \mid Y = k) &= n - 1 + \mathbb{E}(X_{S_<} \mid Y = k) + \mathbb{E}(X_{S_>} \mid Y = k) \\
&= n - 1 + \mathbb{E}(X_{k-1}) + \mathbb{E}(X_{n-k}) \\
&= n - 1 + C(k - 1) + C(n - k)
\end{aligned}$$

wegen der Linearität des (bedingten) Erwartungswertes und der obigen Bemerkungen.

Randomisierter Quicksort

Beweis der Behauptung. Aus dem Satz vom totalen Erwartungswert folgt

$$\begin{aligned} C(n) = \mathbb{E}(X_n) &= \sum_{k=1}^{n} \mathbb{E}(X_n \mid Y = k)\mathbb{P}(Y = k) \\ &= \frac{1}{n}\left(n(n-1) + \sum_{k=1}^{n}[C(k-1) + C(n-k)]\right) \\ &= n - 1 + \frac{2}{n}\sum_{k=1}^{n-1} C(k). \end{aligned}$$

Diese Formel gilt auch für $n-1$:

$$C(n-1) = n - 2 + \frac{2}{n-1}\sum_{k=1}^{n-2} C(k).$$

Randomisierter Quicksort

Beweis der Behauptung. Also

$$\begin{aligned} nC(n) - (n-1)C(n-1) &= \\ &= n(n-1) - (n-1)(n-2) + 2C(n-1) \\ &= 2(n-1) + 2C(n-1), \end{aligned}$$

oder anders formuliert

$$\frac{C(n)}{n+1} = \frac{2(n-1)}{n(n+1)} + \frac{C(n-1)}{n} < \frac{2}{n} + \frac{C(n-1)}{n}.$$

Randomisierter Quicksort

Beweis der Behauptung. Nach $n-1$ Anwendungen dieser Ungleichung erhält man

$$\frac{C(n)}{n+1} < 2 \cdot \left(\frac{1}{n} + \cdots + \frac{1}{2}\right) = 2(H_n - 1),$$

wobei

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + O(1).$$

D.h.

$$C(n) < 2(n+1)(H_n - 1) = 2(nH_n + H_n - n - 1) < 2nH_n.$$