Bedingte Erwartung

Bedingter Erwartungswert

Berechne EW mit bedingter Verteilung.

1. Sei (X, Y) nach der Zähldichte $p(x, y)$ verteilt. **Bedingte Erwartungswert** von X gegeben $Y = y$ gegeben durch

$$E(X|Y = y) = \sum_{x \in \mathcal{X}} x \cdot p_{X|Y}(x|y)$$

2. Sei $(X, Y) \sim f_{(X,Y)}(x, y)$ stetig. Bedingter Erwartungswert:

$$E(X|Y = y) = \int x \cdot f_{X|Y}(x|y)\, dx$$

3. $g(y) = E(X|Y = y)$ ist eine Funktion von y.
4. Einsetzen der Zufallsvariable Y liefert **bedingte Erwartung von X gegeben Y**. Notation: $E(X|Y) := g(Y)$.

Bedingte Dichten

Beispiel: Bedingte Dichte für $y > 0$:

$$f(x|y) = y \cdot e^{-y \cdot x} \mathbf{1}_{(0,\infty)}(x), \qquad x \in \mathbb{R}.$$

Bedingter Erwartungswert

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \cdot f(x|y)\, dx = \int_{0}^{\infty} xy \cdot e^{-y \cdot x}\, dx = \frac{1}{y}$$

(Wie beim partiellen Differenzieren behandle y als Konstante und integriere nach x).
Also zum Beispiel: Für $Y = 2$ erwarten wir

$$E(X|Y = 2) = \frac{1}{2}$$

Zufallszahlen

Quantil-Transformation:

Ist $U \sim U[0,1]$, dann ist die Zufallsvariable $F^{-1}(U)$ nach der Verteilungsfunktion F verteilt.

Beispiel: $-\ln(U)/\lambda$ ist Exp(λ)-verteilt.

Normalverteilung: Box-Muller-Methode

Sind U_1, U_2 unabhängig und identisch $U[0,1]$-verteilt, dann sind

$$
\begin{aligned}
Z_1 &= \sqrt{-2\ln U_1}\cos(2\pi U_2),\\
Z_2 &= \sqrt{-2\ln U_1}\sin(2\pi U_2),
\end{aligned}
$$

unabhängig und identisch $N(0,1)$-verteilt und

$$X_i = \mu + \sigma Z_i, \qquad i = 1, 2,$$

unabhängig und identisch $N(\mu, \sigma^2)$-verteilt.

Erwartungswertvektor

Erwartungswertvektor

$\mathbf{X} = (X_1, \ldots, X_n)'$ Zufallsvektor.
Gelte: $\mu_i = E(X_i)$, $i = 1, \ldots, n$, existieren.
Der (Spalten-) Vektor $\boldsymbol{\mu} = (E(X_1), \ldots, E(X_n))'$ heißt
Erwartungswertvektor von X.

Rechenregeln übertragen sich. Insbesondere gilt für zwei Zufallsvektoren $\mathbf{X}$ und $\mathbf{Y}$ sowie Skalare $a, b \in \mathbb{R}$:

$$E(a \cdot \mathbf{X} + b \cdot \mathbf{Y}) = a \cdot E(\mathbf{X}) + b \cdot E(\mathbf{Y}).$$

Rechenbeispiel

Beispiel: Der Zufallsvektor $\mathbf{X} = (X_1, X_2)^\top$ sei gegeben durch

X_1 : Anzahl der Sechsen bei 12 Würfen mit einem fairen Würfel,

$X_2 = 10 + 2 \cdot Z$

wobei $Z \sim N(0,1)$. Es gilt $X_1 \sim \text{bin}(12, 1/6)$ und somit

$$E(X_1) = 12 \cdot \frac{1}{6} = 2.$$

Nach den Rechenregeln für normalverteilte Zufallsvariablen ist $X_2 \sim N(10, 2^2)$ und somit $E(X_2) = 10$. Daher ist

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix} = \begin{pmatrix} 2 \\ 10 \end{pmatrix}.$$

Transformationsformel für Erwartungswerte

$E(Y) =?$ **für** $Y = g(\mathbf{X})$**:**

Ist $\mathbf{X}$ nach der diskreten Zähldichte $p_{\mathbf{X}}(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$, verteilt, dann gilt

$$E(Y) = E(g(\mathbf{X})) = \sum_{\mathbf{x}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}).$$

Ist $\mathbf{X}$ nach der stetigen Dichte $f_{\mathbf{X}}(\mathbf{x})$ verteilt, dann gilt

$$E(Y) = E(g(\mathbf{X})) = \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}.$$

Beispiel:
Es gelte $\mathbf{X} = (X_1, X_2)' \sim f_{\mathbf{X}}$ mit

$$f_{(X_1,X_2)}(x_1, x_2) = \begin{cases} x_2^3 & \text{falls } x_1 \in [0,4] \text{ und } x_2 \in [0,1], \\ 0, & \text{sonst.} \end{cases}$$

Zu bestimmen sei $Eg(X_1, X_1)$ für die Funktion $g(x_1, x_2) = x_1 \cdot x_2,\ x_1, x_2 \in \mathbb{R}$. Wir erhalten

$$\begin{aligned} E(X_1 \cdot X_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 f_{(X_1,X_2)}(x_1, x_2)\, dx_1\, dx_2 \\ &= \int_0^1 \int_0^4 x_1 x_2 x_2^3\, dx_1\, dx_2 \\ &= \int_0^1 x_2^4 \left(\frac{x_1^2}{2} \Big|_{x_1=0}^{x_1=4} \right) dx_2 = \ldots = \frac{8}{5}. \end{aligned}$$

Kovarianz

Bekannt: X, Y unabhängig, dann gilt: $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.
Frage: Formel für den allgemeinen Fall?
Sei $\mu_X = E(X)$ und $\mu_Y = E(Y)$.
Ausquadrieren und Linearität des Erwartungswertes ausnutzen:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E\left((X + Y - (\mu_X + \mu_Y))^2 \right) \\
&= E\left(((X-\mu_X) + (Y-\mu_Y))^2 \right) \\
&= E\left((X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2 \right) \\
&= E((X-\mu_X)^2) + 2E((X-\mu_X)(Y-\mu_Y)) + E((Y-\mu_Y)^2) \\
&= \mathrm{Var}(X) + 2E(X-\mu_X)(Y-\mu_Y) + \mathrm{Var}(Y).
\end{aligned}$$

Sind X und Y stochastisch unabhängig, dann gilt für den mittleren Term

$$E((X-\mu_X)(Y-\mu_Y)) = E(X-\mu_X) \cdot E(Y-\mu_Y) = 0.$$

Kovarianz

Kovarianz, Kovarianzmatrix

1. Sind X und Y Zufallsvariablen mit existierenden Varianzen, dann heißt
$$\mathrm{Cov}(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)$$
Kovarianz von X und Y.
2. X, Y heißen **unkorreliert**, falls $\mathrm{Cov}(X,Y) = 0$.
3. Ist $\mathbf{X} = (X_1, \ldots, X_n)$ ein Zufallsvektor, dann heißt die symmetrische $(n \times n)$-Matrix $\mathrm{Var}(\mathbf{X}) = (\mathrm{Cov}(X_i, X_j))_{i,j}$ der n^2 Kovarianzen **Kovarianzmatrix von X.**
4. Alle X_i paarweise unkorreliert: $\mathrm{Var}(\mathbf{X})$ Diagonalmatrix.
5. Korrelation:
$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

Kovarianz: Rechenregeln

Rechenregeln

X, Y und Z Zufallsvariablen mit endlichen Varianzen. Dann gelten für alle $a, b \in \mathbb{R}$ die folgenden Rechenregeln:

1. $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y)$.
2. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.
3. $\mathrm{Cov}(X, Y) = 0$, wenn X und Y unabhängig sind.
4. $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$
5. Cauchy-Schwarz-Ungleichung: $\sigma_X^2 = \mathrm{Var}(X), \sigma_Y^2 = \mathrm{Var}(Y)$.

$$|\,\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)} = \sigma_X \sigma_Y.$$

(liefert: $\mathrm{Cor}(X, Y) \in [-1, 1]$.)

Kovarianz: Rechenbeispiel

Sei $Z \sim N(0,1)$ und $\mathbf{X} = (X_1, X_2)'$ gegeben durch

$$\begin{aligned} X_1 &= 1 + 2Z, \\ X_2 &= 3Z. \end{aligned}$$

Dann gilt

$$\operatorname{Var}(X_1) = 4, \qquad \operatorname{Var}(X_2) = 9$$

und die Kovarianz zwischen X_1 und X_2 berechnet sich zu

$$\begin{aligned} \operatorname{Cov}(X_1, X_2) &= \operatorname{Cov}(1 + 2Z, 3Z) \\ &= \operatorname{Cov}(2Z, 3Z) \\ &= 2 \cdot 3 \cdot \operatorname{Cov}(Z, Z) \\ &= 6\operatorname{Var}(Z) = 6. \end{aligned}$$

Somit erhalten wir für die Kovarianzmatrix

$$\operatorname{Cov}(\mathbf{X}) = \begin{pmatrix} \operatorname{Var}(X_1) & \operatorname{Cov}(X_1, X_2) \\ \operatorname{Cov}(X_1, X_2) & \operatorname{Var}(X_2) \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix}.$$

Anwendung: Multivariate Normalverteilung

Definition: Sind $X_1, \dots, X_n$ unabhängig und identisch $N(0,1)$-verteilte Zufallsvariablen, dann ist die gemeinsame Dichtefunktion des Zufallsvektors $\mathbf{X} = (X_1, \dots, X_n)'$ gegeben durch

$$\varphi(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2 \right), \qquad x_1, \dots, x_n \in \mathbb{R}.$$

$\mathbf{X}$ heißt **multivariat** oder n-dimensional **standardnormalverteilt**.
Notation: $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I})$

Erwartungswertvektor und Kovarianzmatrix von $\mathbf{X}$ sind gegeben durch

$$\boldsymbol{\mu} = E(\mathbf{X}) = \mathbf{0} = (0, \dots, 0)' \in \mathbb{R}^n, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Beispiel: Produktverteilung (Erinn: $\prod_{i=1}^{n} \varphi_i(x_i)$) aus Normalverteilungen

Sind $X_i \sim N(\mu_i, \sigma_i^2)$ unabhängig, $1 \le i \le n$, dann hat $\mathbf{X} = (X_1, \dots, X_n)'$ die gemeinsame Produktdichte

$$\varphi(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \right), \qquad x_1, \dots, x_n \in \mathbb{R}.$$

Erwartungswert: $E(\mathbf{X}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$
Kovarianzmatrix: $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2)$ (Diagonalmatrix).

Notation: $\mathbf{X} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$

Multivariate Normalverteilung

Satz: Ist $\mathbf{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$ ein Spaltenvektor und gilt $\mathbf{X} = (X_1, \ldots, X_n)' \sim N_n(\boldsymbol{\mu}, \mathbf{I})$, dann ist die Linearkombination $\mathbf{a}'\mathbf{X} = a_1 X_1 + \cdots + a_n X_n$ ebenfalls normalverteilt mit Erwartungswert

$$E(a_1 X_1 + \cdots + a_n X_n) = a_1 \mu_1 + \cdots a_n \mu_n = \mathbf{a}' \boldsymbol{\mu}$$

und Varianz

$$\mathrm{Var}(a_1 X_1 + \cdots + a_n X_n) = \mathrm{Var}(a_1 X_1) + \cdots + \mathrm{Var}(a_n X_n) = a_1^2 + \cdots + a_n^2 = \mathbf{a}'\mathbf{a}.$$

Kurz: Wenn $\mathbf{X} = (X_1, \ldots, X_n)' \sim N_n(\boldsymbol{\mu}, \mathbf{I})$ und $\mathbf{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$, dann folgt

$$\mathbf{a}'\mathbf{X} \sim N_n(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\mathbf{a}).$$

Anwendungsbeispiel: Künstliche Neuronale Netze (z.B. CNNs) berechnen für einen (zufälligen) d-dim. Input $\mathbf{X}$ lineare Transformationen

$$Z_j = b_j + \mathbf{a}_j'\mathbf{X}, \qquad 1 \leq j \leq m,$$

mit Gewichtungsvektoren $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^d$ und Intercept-Termen $b_1, \dots, b_m$, und dann Ausgaben

$$Y_j = \sigma(Z_j), \qquad 1 \leq j \leq m,$$

mit einer Aktivierungsfunktion $\sigma : \mathbb{R} \to \mathbb{R}$ (als Modell für ein Neuron mit m Dendriten).

Für normalverteilten Input $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ sind die Zwischenergebnisse (Projektionen! - vgl. LA) Z_j multivariat normalverteilt:

$$\mathbf{Z} = (Z_1, \dots, Z_m)' = \mathbf{b} + \mathbf{A}\mathbf{X} \sim N(\mathbf{b} + \mathbf{a}'\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

mit $\boldsymbol{\Sigma} = (\mathrm{Cov}(Z_i, Z_j))_{\substack{1 \leq i \leq m \\ 1 \leq j \leq m}} = (\mathbf{a}_i'\mathbf{a}_j)_{\substack{1 \leq i \leq m \\ 1 \leq j \leq m}}$, $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_m)'$, $\mathbf{b} = (b_1, \dots, b_m)'$.

Multivariate Normalverteilung

Seien allg. $\mathbf{a} = (a_1, \ldots, a_n)'$ und $\mathbf{b} = (b_1, \ldots, b_n)'$ Spaltenvektoren sowie

$$U = \mathbf{a}'\mathbf{X} = a_1 X_1 + \cdots + a_n X_n,$$
$$V = \mathbf{b}'\mathbf{X} = b_1 X_1 + \cdots + b_n X_n,$$

zwei Linearkombinationen der Zufallsvariablen $X_1, \ldots, X_n$.
Ist der Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_n)'$ nun $N_n(\mathbf{0}, \mathbf{I})$-verteilt, dann ist aufgrund der Unabhängigkeit der X_i

$$\begin{aligned} \mathrm{Cov}(U, V) &= \mathrm{Cov}(a_1 X_1 + \cdots + a_n X_n, b_1 X_1 + \cdots b_n X_n) \\ &= \mathrm{Cov}(a_1 X_1, b_1 X_1) + \cdots + \mathrm{Cov}(a_n X_n, b_n X_n) \\ &= a_1 b_1 + \cdots + a_n b_n = \mathbf{a}'\mathbf{b}. \end{aligned}$$

Somit sind die Zufallsvariablen U und V genau dann unkorreliert (also unabhängig), wenn $\mathbf{a}'\mathbf{b} = 0$.

Multivariate Normalverteilung

Ist $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I})$, $\mathbf{b} \in \mathbb{R}^n$ und $\mathbf{A}$ eine $m \times n$-Matrix, dann ist

$$\mathbf{Y} = \mathbf{b} + \mathbf{A}\mathbf{X} \sim N(\mathbf{b}, \boldsymbol{\Sigma})$$

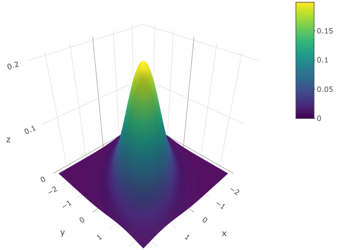
mit $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$.

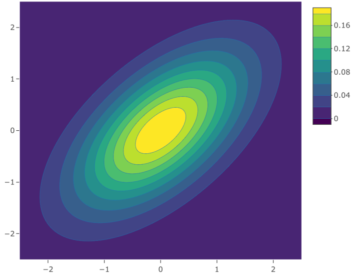
Erzeugung von $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-verteilten Vektoren:

- Bestimme Matrix $\mathbf{A}$ mit $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$ (geht über Singulärwertzerlegung)
- Erzeuge $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$ (Box-Muller)
- Berechne $\mathbf{Y} = \mathbf{b} + \mathbf{A}\mathbf{X}$

Multivariate Normalverteilung

Dichte einer bivariaten Normalverteilung (Korrelation 0.6, $\boldsymbol{\mu} = \mathbf{0}$).





Höhenlinien sind Ellipsen (Achsen d. Ellipsen gegeben durch Eigenvektoren $\mathbf{v}_i$ von $\boldsymbol{\Sigma}$, Länge der Halbachsen = Wurzel der Eigenwerte).
(Plots von Stichproben sind aufsteigende Punktewolken, die ellipsenförmig aussehen.)

Grenzwertsätze und Konvergenzbegriffe

Gegeben: Eine Folge von Zufallsvariablen:

$$X_1, X_2, X_3, \dots$$

Formaler:

$$X_n, \qquad n = 1, 2, \dots$$

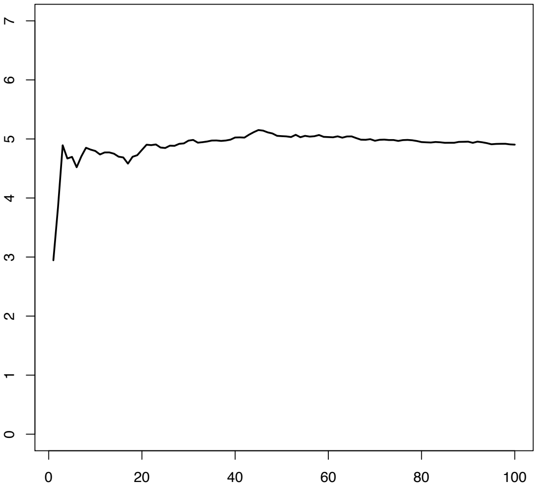
Betrachte die Folge der arithmetischen Mittelwerte:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad n = 1, 2, \dots$$

$$X_1, \quad \frac{X_1 + X_2}{2}, \quad \frac{X_1 + X_2 + X_3}{3}, \dots$$

Computerexperiment I: Folge der arithmetischen Mittel

Computerexperiment I:



Gesetz der großen Zahl

Beobachtung:

- Die Folge der $\overline{X}_n$ nähert sich einem festen Wert an.
- Welcher Wert ist das?
- Wie kann man diese 'Konvergenz' beschreiben?

Fundamentales Resultat 1: Gesetz der großen Zahl