Normalverteilung

X heißt normalverteilt mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 \in (0, \infty)$, falls X die Dichte

$$\varphi_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right), \qquad x \in \mathbb{R}.$$

hat → **Gauß'sche Glockenkurve**.

Carl Friedrich Gauß (1777-1885): Berühmter Göttinger Mathematiker

Gauss

3D9
DEUTSCHE BUNDESBANK
Banknote
10
1777 1855 Carl Friedr. Gauß
ZEHN DEUTSCHE MARK
DN6833178D9


DEUTSCHE BUNDESBANK
Banknote

Eigenschaften

Verteilungsfunktion:

$$\Phi_{(\mu,\sigma^2)}(x) = \int_{-\infty}^{x} \varphi_{(\mu,\sigma^2)}(t)\, dt = p$$

Quantilfunktion (Umkehrfunktion):

$$x = \Phi^{-1}_{(\mu,\sigma^2)}(p)$$

Es gibt keine expliziten Formeln! $\rightarrow$ Computer / Tabellen
R: `p = pnorm(x)`, `x = qnorm(p)`.

Eigenschaften

Eigenschafen der Normalverteilung

1. $E(X) = \int_{-\infty}^{\infty} x \varphi_{(\mu,\sigma^2)}(x) dx = \mu.$
2. $\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \varphi_{(\mu,\sigma^2)}(x) dx = \sigma^2.$
3. Dichte hat Symmetriezentrum: μ
4. Dichte hat Wendepunkte bei $\mu \pm \sigma$

Rechnen mit normalverteilten ZVen

Rechenregeln

1. $X \sim N(\mu_1, \sigma_1^2)$ und $Y \sim N(\mu_2, \sigma_2^2)$ unabhängig, dann gilt: $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
2. Ist $X \sim N(\mu, \sigma^2)$ und sind $a, b \in \mathbb{R}$, dann gilt: $aX + b \sim N(a\mu + b, a^2\sigma^2)$.
3. Ist $X \sim N(\mu, \sigma^2)$, dann gilt:
$$X^* = \frac{X - \mu}{\sigma} \sim N(0, 1).$$
X^*: Standardisierte Version.
4. Ist $X^* \sim N(0, 1)$, dann gilt
$$\mu + \sigma \cdot X^* \sim N(\mu, \sigma^2).$$

Normalverteilung: Rechenregeln

Regel

Sind $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ unabhängig, dann ist das arithmetische Mittel normalverteilt mit Erwartungswert μ und Varianz σ^2/n:

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

und

$$\overline{X}^* = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - \mu}{\sigma} \sim N(0, 1).$$

Zufallsvektoren

Zufallsvektoren

Ω abzählbar, dann heißt jede Abbildung

$$\mathbf{X} : \Omega \to \mathbb{R}^n, \qquad \omega \mapsto \mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

in den n-dimensionalen Raum $\mathbb{R}^n$ **Zufallsvektor**.
Zusatz: Ist Ω überabzählbar, dann müssen alle X_i, $i = 1, \dots, n$, messbar sein.

Realisationen von $\mathbf{X} = (X_1, \dots, X_n)$ sind Vektoren $\mathbf{x}$ im $\mathbb{R}^n$:
$\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Verteilung eines Zufallsvektors (X, Y)

Verteilungsfunktion:

$$F(x, y) = P(X \leq x, Y \leq y), \qquad x, y \in \mathbb{R}$$

Es gilt: $F(-\infty, y) = P(\{X \leq -\infty\} \cap \{Y \leq y\}) = P(\emptyset \cap \{Y \leq y\}) = 0.$
Rand-Verteilungsfunktionen:

$$F_X(x) = P(X \leq x) = F(x, \infty) = \lim_{y \to \infty} F(x, y)$$

$$F_Y(y) = P(Y \leq y) = F(\infty, y) = \lim_{x \to \infty} F(x, y)$$

Für $(X_1, \ldots, X_n)$ entsprechend, z. B.

$$F_{(X,Y,Z)}(\infty, y, \infty) = P(Y \leq y), \qquad y \in \mathbb{R}.$$

Verteilung eines Zufallsvektors (X, Y)

Diskrete Zufallsvektoren: Verteilung durch Zähldichten gegeben

$$p(x, y) = P(X = x, Y = y), \qquad (x, y) \in \mathbb{R}^2$$

- Stäbe über der (x, y)-Ebene an denjenigen Stellen (x, y) mit $P(X = x, Y = x) > 0$, sonst 0.
- Entspricht der Kontingenztabelle der gemeinsamen Wahrscheinlichkeiten.

- $P(X \in A, Y \in B) = \sum_{(x,y) \in A \times B} p(x, y)$
(Summe der Stäbe, die in $A \times B$ stehen.)

Stetige Zufallsvektoren: Verteilung gegeben durch Dichte $f(x, y)$

$$f(x, y) \geq 0, (x, y) \in \mathbb{R}^2, \qquad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dxdy = 1$$

- 'Gebirge' über der (x, y)-Ebene.
- $P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x, y)\, dydx$

Produktmodell / Produktverteilung

Produkt-Verteilungsfunktion

$$F(x,y) = F_X(x) \cdot F_Y(y), \qquad (x,y) \in \mathbb{R}^2$$

Produkt-Zähldichte

$$p(x,y) = p_X(x) \cdot p_Y(y), \qquad (x,y) \in \mathbb{R}^2$$

Produkt-Dichtefunktion

$$f(x,y) = f_X(x) \cdot f_Y(y), \qquad (x,y) \in \mathbb{R}^2$$

Dies entspricht der stochastischen Unabhängigkeit von X und Y.

Verteilung eines Zufallsvektors (X, Y)

Beispiel: $X \sim \text{Bin}(10, 0.4)$ und $Y \sim \text{Poi}(2)$ seien unabhängig. Dann gilt:

$$P(X = k, Y = l) = \binom{10}{k} 0.4^k 0.6^{10-k} \cdot \frac{2^l e^{-l}}{l!}, k = 0, \dots, n, k = 0, 1, 2, \dots$$

Beispiel: Sind $X, Y \sim N(0,1)$ unabhängig, dann hat (X, Y) die Dichte

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \cdot \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) \\ &= \frac{1}{2\pi} \exp(-(x^2 + y^2)/2) \end{aligned}$$

für $(x, y) \in \mathbb{R}^2$.

Beispiel

Beispiel: Gelte $(X, Y) \sim f(x, y)$ wobei

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)}, & x, y > 0 \\ 0, & \text{sonst} \end{cases}$$

Für alle $x, y > 0$ gilt:

$$f(x,y) = \lambda^2 e^{-\lambda(x+y)} = \lambda^2 e^{-\lambda x} e^{-\lambda y} = (\lambda e^{-\lambda x})(\lambda e^{-\lambda y})$$

$f(x, y)$ ist das Produkt von zwei Exp(λ)-Dichten (die 0 sind auf der nichtpositiven Halbachse). Also sind X und Y unabhängig und identisch Exp(λ)-verteilt.

Bedingte Verteilung, Unabhängigkeit

X, Y diskret mit Werten in $\mathcal{X} = \{x_1, x_2, \dots,\}$ bzw. $\mathcal{Y} = \{y_1, y_2, \dots,\}$.

1. Bedingte Wahrscheinlichkeit von $X = x_i$ gegeben $Y = y_j$:

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

   $p(x_i, y_j)$: gem. Zähldichte, $p_Y(y_j)$: Zähldichte von Y.

2. Definiert die **bedingte Zähldichte**

$$p(x|y) = p_{X|Y}(x|y) = P(X = x | Y = y) = \begin{cases} \frac{p(x,y)}{p_Y(y)}, & y \in \{y_1, y_2, \dots\}, \\ p_X(x), & y \notin \{y_1, y_2, \dots,\} \end{cases}$$

3. Endlicher Fall: $p(x_i, y_j)$: Tabelle (Kontingenzafel), $p_Y(y_j)$: Rand

4. Entsprechend definiert man die bedingte Wahrscheinlichkeit von $Y = y_j$ gegeben $X = x_i$.

Bedingte Zähldichten

Beispiel: X, Y seien Zufallsvariable und es gelte:

$$P(X = 1) = 0.1, \quad P(X = 2) = 0.5, \quad P(X = 3) = 0.4$$

$$P(Y = 0) = 0.2, \quad P(Y = 1) = 0.8$$

Stochastische Abhängigkeit bei Verletzung der Produktregel:

$Y \setminus X$	1	2	3	
0	0.02	0.08	0.1	0.2
1	0.08	0.42	0.3	0.8
	0.1	0.5	0.4	1

Bedingte Wahrscheinlichkeit für $X = 2$ gegeben $Y = 1$:

$$P(X = 2 | Y = 1) = \frac{p_{XY}(2, 1)}{p_Y(1)} = \frac{0.42}{0.8} = 0.525$$

Beispiel: Klassifikation im Machine Learning

Problem: Klassifiziere Objekte in 2 Klassen, 0: ok, 1: defekt
Klassifizierer (Algorithmus) berechnet $Y = g(X) \in \{0, 1\}$.
X: Zufallsvariable(n) (Input-Feature(s))
Z: wahres Label mit Werten in $\{0, 1\}$ (unbeobachtet).
Beispiele: Virus (defekt=infiziert), Produktion (Objekt=Produkt), ...
Entscheidung für Klasse 0: „0" $= \{Y = 0\}$,
Entscheidung für Klasse 1: „1" $= \{Y = 1\}$

Wahre Wahrscheinlichkeiten in der Population:

	wahres Label Z		
Klassifizierer Y	0	1	
„0"	p_{00}	p_{01}	$p_{00} + p_{01}$
„1"	p_{10}	p_{11}	$p_{10} + p_{11}$

Fehlklassifikationswahrscheinlichkeit: $p_{err} = p_{01} + p_{10}$.
Falsch-Positiv-Rate: $P(Z = 0|Y = 1) = \frac{p_{10}}{p_{10}+p_{11}}$

Beispiel: Klassifikation im Machine Learning

Falsch-Positiv-Rate: $P(Z=0|Y=1) = \frac{p_{10}}{p_{10}+p_{11}}$
Analyse des Zählers:

$$\begin{aligned} p_{10} &= P(Y=1|Z=0)P(Z=0) \\ &= (1 - P(Y=0|Z=0)) \cdot (1 - P(Z=1)) \end{aligned}$$

Hier ist $P(Y=0|Z=0)$ die **Spezifität** des Klassifiziers (richtige Entscheidung bei Nicht-Defekten) und $P(Z=1)$ der Defektanteil.
Falsch-Positiv-Rate kann sehr groß sein, wenn Spezifität nicht nahe 1 und/oder Defektanteil klein!
Sensitivität: $P(Y=1|Z=1)$ (Erfolgsrate unter Defekten)

- Algorithmen optimieren in der Regel $p_{err} = p_{01} + p_{11}$
- Sensitivität und Spezifität schätzbar aus Testfällen mit Z bekannt.
- Anwendung problematisch, wenn Falsch-Positiv-Rate hoch!

Simpson's Paradoxon

Situation: Verteilung von Y gegeben $X = x$ interessiert. Zusätzliche Variable Z

Beispiel: Fahrprüfungen: Y: best. / n. best., X: M/W, Z: Tag 1, 2

Tabellen nach Tagen (Werten von Z):

Tag $Z = 1$:

$Y \setminus X$	M	F
best.	1	9
n. best.	0	1
	1	10

Durchfallquoten: 0 %, 10 %

Tag $Z = 2$:

$Y \setminus X$	M	F
best.	3	1
n. best.	1	1
	4	2

Durchfallquoten: 25%, 50 %

Aggregiert über Z:

$Y \setminus X$	M	F
best.	4	10
n. best.	1	2
	5	12

Durchfallquoten: 20%, 16.67 %

Simpson's Paradoxon

Paradoxon: Das Ergebnis kehrt sich um, d.h. aggregriert anderes Ergebnis als aufgeschlüsselt!
Ursache: Verteilung von Variable Z unterschiedlich für Werte von X:
Verteilung von Z unter Männern: $1/5, 4/5$
Verteilung von Z unter Frauen: $10/12, 2/12$
(Männer wählen meist Tag 2, Frauen Tag 1)
Rechnerisch gilt: ($Y = 1$ für n. best., rel. Hf. als Wkeiten)

$$
\begin{aligned}
P(Y=1|X=M) &= \underbrace{P(Y=1|X=M,Z=1)}_{\text{Quote aus Teilpopulation } Z=1} \cdot \underbrace{P(Z=1|X=M)}_{\text{Anteil } Z=1 \text{ unter } X=M} \\
&\quad + \underbrace{P(Y=1|X=M,Z=2)}_{\text{Quote aus Teilpopulation } Z=2} \cdot \underbrace{P(Z=2|X=M)}_{\text{Anteil } Z=2 \text{ unter } X=M} \\
&= \frac{0}{1} \cdot \frac{1}{5} + \frac{1}{4} \cdot \frac{4}{5} = \frac{1}{5} \\
P(Y=1|X=F) &= \frac{1}{10}\frac{10}{12} + \frac{1}{2}\frac{2}{12} = \frac{2}{12} = \frac{1}{6} = 0.1\bar{6}
\end{aligned}
$$

Simpson's Paradoxon: Klassisches Beispiel

Klassisches Beispiel: Y: Zulassung zum Studium, X: Geschlecht, Z: Studienfach.

Berkeley 1973: 35% der Frauen zugelassen, aber 44% der Männer.

Frage: Liegt eine Diskriminierung vor?

Analyse aufgeschlüsselt nach Studienfächern (Fakultäten) bestätigte den Verdacht nicht.

Ursache: Verteilung des Studienfachs Z bei Frauen anders als bei Männern. Frauen bewarben sich häufiger als Männer für Fächer mit niedrigen Zulassungsquoten für beide Geschlechter.

Simpson's Paradoxon und Big Data/ML/AI

• In der Regel ist die Detailanalyse nach den Werten Z informiver und liefert eine genauere Interpretation.
• Generell Vorsicht bei aggregierten Daten/Tabellen! Besser aufschlüsseln nach weiteren (potentiellen) Einflussvariablen.

Allgegenwärtiges, reales Problem bei Big Data, ML, AI:

Diskriminierung, Bias, Fairness-Begriff

Datenbestände können ungewollten gesellschaftlichen Bias und Diskriminierung abbilden. Schlechterstellung/Nachteil (sagen wir $Y = 1$) bei Vorliegen von bestimmten Ausprägungen ($X = x^*$) häufiger als bei anderen Ausprägungen von X. KI-Verfahren lernen dies aus den Daten. Was soll Fairness bedeuten? Gleiche Quoten marginal oder in Teilpopulationen ($Z = z$)?

Bedingte Dichten

Bedingte Dichte

X und Y stetig verteilt: $(X, Y) \sim f(x, y)$.
Bedingte Dichte von X gegeben $Y = y$ (y fest):

$$f(x|y) = f_{X|Y}(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)}, & f_Y(y) > 0, \\ f_X(x), & f_Y(y) = 0, \end{cases}$$

Dies ist als Funktion von x eine Dichte (für jedes $y \in \mathbb{R}$).
Notation: $X|Y = y \sim f_{X|Y}(x|y)$.

Beispiel: Bedingte Dichte für $y > 0$:

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

Ist $Y \sim U(1,3)$, dann ist die gemeinsame Dichte

$$f(x,y) = f(x|y) f_Y(y) = \begin{cases} \frac{1}{2} y \cdot e^{-y \cdot x}, & y \in [1,3], x > 0, \\ 0, & x \leq 0 \text{ oder } y < 1 \text{ oder } y > 3. \end{cases}$$

Kriterien für Unabhängigkeit

1. Diskreter Fall: X und Y sind genau dann stochastisch unabhängig, wenn für alle x und y gilt:

$$p_{X|Y}(x|y) = p_X(x) \qquad \text{bzw.} \qquad p_{Y|X}(y|x) = p_Y(y).$$

bzw. $p(x,y) = p_X(x)p_Y(y)$
(gem. Zähldichte = Produkt-Zähldichte)

2. Stetiger Fall: X und Y genau dann stochastisch unabhängig, wenn für alle x und y gilt:

$$f_{X|Y}(x) = f_X(x) \qquad \text{bzw.} \qquad f_{Y|X}(y) = f_Y(y).$$

bzw. $f(x,y) = f_X(x)f_Y(y)$ (gem. Dichte = Produktdichte).

3. X, Y ist genau dann stochastisch unabhängig, wenn für die gemeinsame Verteilungsfunktion $F_{(X,Y)}(x,y)$ für alle $x, y \in \mathbb{R}$ gilt: $F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y)$.