3. Bedingte Verteilung, bedingte Erwartung

- Bedingte Verteilung
- Bedingter Erwartungswert

Erklärung und Prognose von Merkmalen

- ↪ Gegeben: Datensatz

Merkmal 1	...	Merkmal k	Merkmal $k+1$
x_{11}	...	x_{1k}	y_1
$\vdots$		$\vdots$	$\vdots$
x_{n1}	...	x_{nk}	y_n

,

  Modell n-fache (unabh.) Versuchswiederholung mit ZV $X_1, \ldots, X_k, Y$
- ↪ Aufgabe: Finde eine Funktion h mit $Y \approx h(X_1, \ldots, X_k)$, welche den Datensatz „gut“ beschreibt („Überwachtes Lernen“):
  - □ Erklärung von Y durch $X_1, \ldots, X_k$: Modell durch Festlegung von h an Daten anpassen.
  - □ Prognose von Y durch $X_1, \ldots, X_k$: Prognosefehler minimieren.
- ↪ Bedingte Sicht:
  - □ $X_1, \ldots, X_n$ sind Informationen, die zur Neubewertung der Modell-Informationen über Y führen
  - □ Aus der Wahrscheinlichkeitsrechnung bekanntes Konzept: Bedingte Wahrscheinlichkeiten von Ereignissen
  - □ Übertragung auf ZV: Bedingte Verteilung /bedingte Erwartung

3.1 Bedingte Verteilung

- $\hookrightarrow$ in WS-Rechnung:
  - □ WS von Ereignissen $\rightsquigarrow$ Verteilungen von ZV $P(X \in B) = P(X^{-1}(B))$
  - □ Bedingte WS von Ereignissen: $P(B|A) = P(B \cap A)/P(A)$ für $P(A) > 0$
- $\hookrightarrow$ Kombination: Bedingte WS von durch ZV X, Y induzierten Ereignissen
  - □ $P(Y \in B | X \in A) = \dfrac{P(Y^{-1}(B) \cap X^{-1}(A))}{P(X^{-1}(A))}$ für $P(X \in A) > 0$ $(*)$
- $\hookrightarrow$ Dabei ist z.B. in Erklärungsmodellen ein konkreter beobachteter Wert $X = x$ gegeben, d.h. $A = \{X = x\}$. Die WS für Y wird unter $X = x$ neu bewertet. (**Bedingte Verteilung** von Y unter $X = x$, Notation: $\mathcal{L}(Y|X = x)$)
  - □ Gerade in stetigen WS-Modellen gilt aber $P(X = x) = 0$,
  - □ d.h. $(*)$ kann dann direkt nicht umgesetzt werden.
- $\hookrightarrow$ Analog zur Bayes-Formel lässt sich für bivariate ZV (X, Y) mit gemeinsamer Dichte $f(x, y)$ „die" Dichte einer bedingten Verteilung $\mathcal{L}(Y|X = x)$ erklären.
- $\hookrightarrow$ Deren Kennzahlen lassen sich zur Prognose von Y gegeben $X = x$ einsetzen.

Übung: Es seien A, B Ereignisse mit $0 < P(A) < 1$ und $X = \mathbb{1}_A$, $Y = \mathbb{1}_B$. Berechnen Sie die bedingte Verteilung $\mathcal{L}(Y|X = x)$ für $x \in \{0, 1\}$.

Die bedingenden Ereignisse sind $\{X = 1\} = \{\omega : \mathbb{1}_B(\omega = 1\} = A$ und $\{X = 0\} = A^c$. Es müssen bedingte WS der Ereignisse $\{Y = 1\} = \{\omega : \mathbb{1}_B(\omega) = 1\} = B$ bzw. $\{Y = 0\} = B^c$ berechnet werden.

$\hookrightarrow$ $x = 1$:

- $y = 1$: $P(Y = 1|X = 1) = \frac{P(\{\mathbb{1}_A=1\}\cap\{\mathbb{1}_B=1\})}{P(\mathbb{1}_A=1)} = \frac{P(B\cap A)}{P(A)} = P(B|A)$
- $y = 0$: $P(Y = 0|X = 1) = \frac{P(\{\mathbb{1}_A=0\}\cap\{\mathbb{1}_B=1\})}{P(\mathbb{1}_A=1)} = \frac{P(B^c\cap A)}{P(A)} = P(B^c|A)$

$\hookrightarrow$ $x = 0$:

- $y = 1$: $P(Y = 1|X = 0) = \frac{P(\{\mathbb{1}_A=0\}\cap\{\mathbb{1}_B=1\})}{P(\mathbb{1}_A=0)} = \frac{P(B\cap A^c)}{P(A^c)} = P(B|A^c)$
- $y = 0$: $P(Y = 0|X = 0) = \frac{P(\{\mathbb{1}_A=0\}\cap\{\mathbb{1}_B=0\})}{P(\mathbb{1}_A=0)} = \frac{P(B^c\cap A^c)}{P(A^c)} = P(B^c|A^c)$

Die bedingte Verteilung wird also durch die verschiedenen bedingten Wahrscheinlichkeiten auf Grundlage von A, A^c bzw B, B^c gegeben.

Beschränkung auf bivariaten Fall: Gegeben sei ein Zufallsvektor (X, Y) mit

- diskreter/stetiger Dichte $f_{X,Y}(x, y)$
- Randdichten $f_X(x)$, $f_Y(y)$.

Bedingte Verteilung $\mathcal{L}(Y|X = x)$

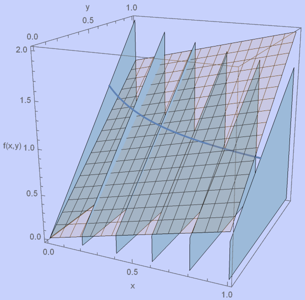
ist gegeben durch **bedingte Dichte**:

$$f_{Y|X=x}(y) = \begin{cases} f_{X,Y}(x, y)/f_X(x) & \text{falls Nenner} > 0 \\ f_Y(y) & \text{sonst} \end{cases} \qquad \text{(Bayes-Formel)}$$

„bedingte WS in y bilden bei festem x eine WS-Verteilung“.

Bed. Dichte zu stetigem Zufallsvektor (DuW)

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}x + \frac{3}{2}y & 0 \le x, y \le 1, \\ 0 & \textit{sonst} \end{cases}, f_X(x) = \frac{1}{2}x + \frac{3}{4}, f_Y(y) = \frac{3}{2}y + \frac{1}{4}$$



bed. Dichte für $0 < x, y < 1$:

$$f_{Y|X=x}(y) \quad = \quad \frac{f(x,y)}{f_X(x)} = \frac{\frac{1}{2}x + \frac{3}{2}y}{\frac{1}{2}x + \frac{3}{4}}$$

Übung: Berechne $\mathcal{L}(Y|X=x)$ zu $f(x,y) = \frac{\mathbb{1}_{]0;1[}(x)\mathbb{1}_{]0;x[}(y)}{x}$:

1. $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \dots$

$$\begin{aligned} \dots &= \int_{-\infty}^{\infty} \frac{\mathbb{1}_{]0;1[}(x)\mathbb{1}_{]0;x[}(y)}{x} dy = \frac{\mathbb{1}_{]0;1[}(x)}{x} \int_{-\infty}^{\infty} \mathbb{1}_{]0;x[}(y)dy \\ &= \frac{\mathbb{1}_{]0;1[}(x)}{x} \int_0^x 1dy = \frac{\mathbb{1}_{]0;1[}(x)}{x} \cdot x = \mathbb{1}_{]0;1[}(x) \end{aligned}$$

also ist $\mathcal{L}(X)$ eine Standard-Rechteck-Verteilung

2. $f_{Y|X=x}(y) = f(x,y)/f_X(x) = \dots$

$$\dots = \frac{\mathbb{1}_{]0;1[}(x)\mathbb{1}_{]0;x[}(y)}{x} \Big/ \mathbb{1}_{]0;1[}(x) = \frac{1}{x}\mathbb{1}_{]0;x[}(y)$$

Also ist $\mathcal{L}(Y|X=x)$ eine Rechteckverteilung $Re(0,x)$

Diskretes Beispiel: $Y_1, \ldots, Y_m$ u.i.v. $\sim Bin(1, p)$, $X = Y_1 + \cdots + Y_m$

Bestimme bedingte Verteilung von $(Y_1, \ldots, Y_m)$ unter $X = x \in \{0, \ldots, m\}$.
Betrachte hierzu (ausschließlich) $y_1, \ldots, y_m \in \{0, 1\}$ mit $y_1 + \cdots + y_m = x$:

$$\begin{aligned}
P(Y_1 = y_1, \ldots, Y_m = y_m | X = x) &= \frac{P(Y_1 = y_1, \ldots, Y_m = y_m, X = x)}{P(X = x)} \\
&= \frac{P(Y_1 = y_1, \ldots, Y_m = y_m)}{P(X = x)} \\
&= \frac{p^x (1-p)^{m-x}}{\binom{m}{x} p^x (1-p)^{m-x}} = 1/\binom{m}{x}
\end{aligned}$$

$\hookrightarrow$ $\mathcal{L}(Y_1, \ldots, Y_m | X = x)$ ist also eine Gleichverteilung auf den $\binom{m}{x}$ m-Tupeln $(y_1, \ldots, y_m) \in \{0, 1\}^m$ mit $y_1 + \cdots + y_m = x$.

$\hookrightarrow$ Die bedingte Verteilung ist unabhängig von p:

- Information über p ist schon vollständig in X enthalten.
- Schätzer, Tests, Konfidenzintervalle,... zu p dürfen Datenverdichtung X anstelle der Originaldaten $Y_1, \ldots, Y_m$ verwenden.
- Man sagt: $X = Y_1 + \cdots + Y_m$ ist **suffizient**.

3.2 Bedingter Erwartungswert

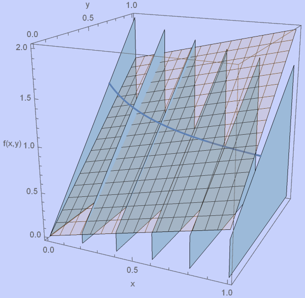
- ↪ Erwartungswert der bedingten Verteilung $\mathcal{L}(Y|X=x)$.
- ↪ Bei Vorliegen einer bedingten Dichte $f(y|x)$

$$h(x) := E(Y|X=x) = \begin{cases} \int\limits_{-\infty}^{\infty} y \cdot f(y|x)dy & \text{stetiger Fall} \\ \sum\limits_{y \in \mathcal{Y}} y \cdot f(y|x) & \text{diskreter Fall mit Träger } \mathcal{Y} \end{cases}$$

- ↪ Schreibweise: $E(Y|X) = h(X)$
- ↪ Der bed. EW löst das Problem $E((Y-h(X))^2) \stackrel{!}{=} \min$, d.h. gibt eine Funktion h an, welche den Ausdruck minimiert.
- ↪ Konzept bedingter Verteilungen/Erwartungswerte übertragbar auf Zufallsvektoren:
  $\mathcal{L}(Y_1, \dots, Y_m|X_1 = x_1, \dots, X_k = x_k)$ bzw. $E(Y|X_1 = x_1, \dots, X_n = x_n)$

Bed. Dichte zu stetigem Zufallsvektor (DuW)

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}x + \frac{3}{2}y & 0 \le x, y \le 1, \\ 0 & sonst \end{cases}, f_X(x) = \frac{1}{2}x + \frac{3}{4}, f_Y(y) = \frac{3}{2}y + \frac{1}{4}$$



bed. Dichte für $0 < x, y < 1$:

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{f(x,y)}{f_X(x)} = \frac{\frac{1}{2}x + \frac{3}{2}y}{\frac{1}{2}x + \frac{3}{4}} \\ E(Y|X=x) &= \int_0^1 \frac{\frac{1}{2}xy + \frac{3}{2}y^2}{\frac{1}{2}x + \frac{3}{4}} dy \\ &= \left[\frac{\frac{1}{4}xy^2 + \frac{1}{2}y^3}{\frac{1}{2}x + \frac{3}{4}} \right]_{y=0}^{y=1} \\ &= \frac{\frac{1}{4}x + \frac{1}{2}}{\frac{1}{2}x + \frac{3}{4}} \end{aligned}$$

Übung: Berechne $E(Y|X=x)$ zu $f(x,y) = \frac{\mathbb{1}_{]0;1[}(x)\mathbb{1}_{]0;x[}(y)}{x}$. Bereits gerechnet:

1. $f_X(x) = \mathbb{1}_{]0;1[}(x)$, d.h. $\mathcal{L}(X)$ ist Standard-Rechteck-Verteilung
2. $f_{Y|X=x}(y) = f(x,y)/f_X(x) = \frac{1}{x}\mathbb{1}_{]0;x[}(y)$, d.h. $\mathcal{L}(Y|X=x) = Re(0,x)$

Daraus jetzt den bedingten Erwartungswert:

Erwartungswert dieser Rechteckverteilung ist $(0+x)/2 = x/2$, zu Fuß:

$$E(Y|X=x) \quad = \quad \int_0^x \frac{1}{x} y dy = [\frac{y^2}{2x}]_0^x = \frac{x^2}{2x} = \frac{x}{2}$$

Regeln für bedingte Erwartungswerte

- $\hookrightarrow$ Linearität: $E(a + bY_1 + cY_2|X) = a + bE(Y_1|X) + cE(Y_2|X)$
- $\hookrightarrow$ Totale Wahrscheinlichkeit: $E(Y) = E(E(Y|X))$
- $\hookrightarrow$ Faktorisierung: $E(Y \cdot g(X)|X) = g(X) \cdot E(Y|X)$
- $\hookrightarrow$ Substituieren/Eliminieren (SE): Wenn X, Y st.u. sind, dann gilt:

(SE1) $P(h(X, Y) \in B|X = x)) = P(h(x, Y) \in B)$

(SE2) $E(h(X, Y)|X = x) = E(h(x, Y))$

Aussagen jeweils „fast sicher“ und unter Annahme existierender Erwartungswerte

Übung: Fortsetzung des Beispiels mit $f(x,y) = \mathbb{1}_{]0;1[}(x)\mathbb{1}_{]0;x[}(y)/x$.
Bereits berechnet wurden: $\mathcal{L}(X) = Re(0,1)$, $E(Y|X=x) = x/2$.

1. Berechne $E(Y)$ mit dem Satz von der totalen WS $E(Y) = E(E(Y|X))$.

$E(Y) = E(E(Y|X)) = E(X/2) = \frac{1}{2}E(X) = \frac{1}{4}$

2. Vergleiche mit dem direkten Rechenweg, d.h.
   - ☐ $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \dots$ für $y \in [0;1]$
   - ☐ $E(Y) = \int_{-\infty}^{\infty} y \cdot f_Y(y)dy = \dots$

☐ $\dots = \int_{-\infty}^{\infty} \mathbb{1}_{]0;1[}(x)\mathbb{1}_{]0;x[}(y)/x dx = \int_0^1 \mathbb{1}_{]0;x[}(y)/x dx$

Nun ist $\mathbb{1}_{[0;x[}(y) = 1 \Leftrightarrow y < x \Leftrightarrow \mathbb{1}_{[y;\infty]}(x) = 1$, also

$\mathbb{1}_{[0;x[}(y) = \mathbb{1}_{[y;\infty]}(x)$ für $y \in [0;1]$, daher für $y \in [0;1]$

$$f_Y(y) = \int_0^1 \frac{\mathbb{1}_{]0;x[}(y)}{x}dx = \int_0^1 \frac{\mathbb{1}_{[y;\infty]}(x)}{x}dx = \int_y^1 \frac{1}{x}dx = -\ln(y)$$

☐ $\dots = \int_0^1 -y\ln(y)dy = [-\frac{1}{2}y^2\ln(y)]_0^1 + \int_0^1 \frac{y^2}{2}\cdot\frac{1}{y}dy = [y^2/4]_0^1 = \frac{1}{4}$

Dieser Rechenweg liefert das gleiche Ergebnis (bestätigt die t.W.-Formel), ist aber wesentlich aufwändiger.