Einführung in die Empirische Wirtschaftsforschung

Übungsaufgaben 1

Bivariate Verteilungen - Unabhängigkeit - Kontingenztabellen

1. Seien (X, Y) unabhängige bivariat normalverteilte Zufallsvariablen mit Parametern $\mu_X = 0$, $\sigma_X^2 = 1$, $\mu_Y = -1$ und $\sigma_Y^2 = 4$. Berechnen Sie $P(X + Y > 0)$.

2. Betrachten Sie folgende bivariate diskrete Wahrscheinlichkeitsverteilung

$$p(x, y) = \frac{xy^2}{13} \mathbb{1}_S(x, y),$$

wobei $S = \{(1, 1), (1, 2), (2, 2)\}$ und

$$\mathbb{1}_S(x, y) = \begin{cases} 1 & \text{wenn } (x, y) \in S \\ 0 & \text{sonst.} \end{cases}$$

a) Stellen Sie die gemeinsame Wahrscheinlichkeitsverteilung in Tabellenform dar.

b) Sind die Bedingungen einer gemeinsamen Wahrscheinlichkeitsverteilung erfüllt?

c) Berechnen Sie $P(X + Y \leq 3)$.

d) Berechnen Sie die Randwahrscheinlichkeiten von X und Y.

e) Sind X und Y unabhängig?

3. Sei $X \sim Bernoulli(0.6)$ und

$$Y \sim \begin{cases} Bernoulli(0.2) & \text{wenn } X = 1 \\ Bernoulli(0.4) & \text{wenn } X = 0. \end{cases}$$

a) Stellen Sie die gemeinsame Wahrscheinlichkeitsverteilung in Tabellenform dar.

b) Berechnen Sie $E(X)$, $E(Y)$ und $E(X + Y)$.

c) Sind X und Y unabhängig?

4. Betrachten Sie die folgende bivariate Dichtefunktion

$$f(x, y) = \begin{cases} y(\frac{1}{2} - x) + x & \text{wenn } 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{sonst.} \end{cases}$$

Finden Sie die Randverteilung von X und Y.

5. Eine Versicherungsgesellschaft interessiert sich dafür, ob es einen Zusammenhang zwischen Rauchen und Autounfällen gibt. Es wird eine Zufallsstichprobe mit $N = 597$ erhoben. Die Daten sehen folgendermassen aus:

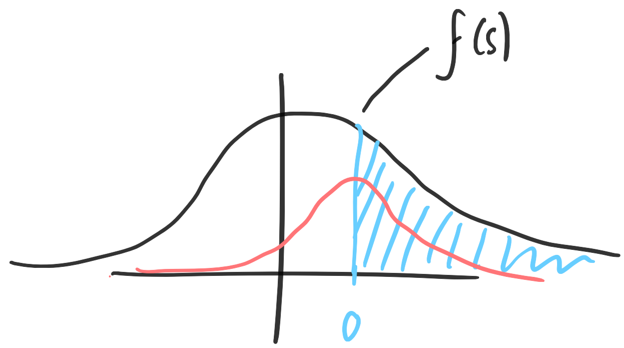
Anzahl Unfälle in den letzten 2 Jahren	Raucher	Nicht Raucher
0	35	170
1	79	190
2 oder mehr	57	66

Sollte die Versicherungsgesellschaft eine höhere Prämie bei Rauchern verlangen? Verwenden Sie zur Beantwortung der Frage einen geeigneten Hypothesentest mit Signifikanzniveau $\alpha = 5\%$.

1. Seien (X, Y) unabhängige bivariat normalverteilte Zufallsvariablen mit Parametern $\mu_X = 0$, $\sigma_X^2 = 1$, $\mu_Y = -1$ und $\sigma_Y^2 = 4$. Berechnen Sie $P(\underbrace{X+Y}_{:=S} > 0)$.

Satz: Die Summe zweier unabhängigen, normalverteilten Zufallsvar. X, Y ist normalverteilt mit

$$S \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$



Rep.: Standardisieren:

Falls $S \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$, dann

$$Z := \frac{S - (\mu_X + \mu_Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \sim N(0,1)$$

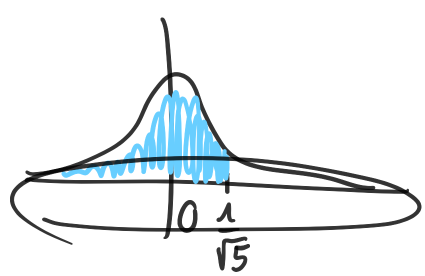
↳ Wir suchen $P(X+Y > 0) = P(S > 0) = P\left(\underbrace{\frac{S - (\mu_X + \mu_Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2}}}_{:=Z} > \frac{0 - (\mu_X + \mu_Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right)$

$1 - P(Z < \frac{1}{\sqrt{5}})$

$= P(Z > \frac{1}{\sqrt{5}}) \approx 0.3264$

↳ 1 - pnorm($q = \frac{1}{\sqrt{5}}$)

1 - pnorm(mean = -1, sd = $\sqrt{5}$, q = 0)



2. Betrachten Sie folgende bivariate diskrete Wahrscheinlichkeitsverteilung

$$p(x, y) = \frac{xy^2}{13} \underbrace{\mathbb{1}_S(x, y)},$$

wobei $S = \{(1,1), (1,2), (2,2)\}$ und

x, y

$$\mathbb{1}_S(x, y) = \begin{cases} 1 & \text{wenn } (x, y) \in S \\ 0 & \text{sonst.} \end{cases}$$

a) Stellen Sie die gemeinsame Wahrscheinlichkeitsverteilung in Tabellenform dar.

b) Sind die Bedingungen einer gemeinsamen Wahrscheinlichkeitsverteilung erfüllt?

c) Berechnen Sie $P(X + Y \leq 3)$.

d) Berechnen Sie die Randwahrscheinlichkeiten von X und Y.

e) Sind X und Y unabhängig?

a)

$y \backslash x$	$X=1$	$X=2$
$y=1$	$\frac{1}{13}$	0
$y=2$	$\frac{4}{13}$	$\frac{8}{13}$

b) Bedingungen:

(i) $0 \leq p(x,y) \leq 1$

(ii) $\sum_x \sum_y p(x,y) = 1$

$P(X=1, Y=1) + P(X=1, Y=2) + P(X=2 | Y=1)$

$+ P(X=2 | Y=2) = 1$

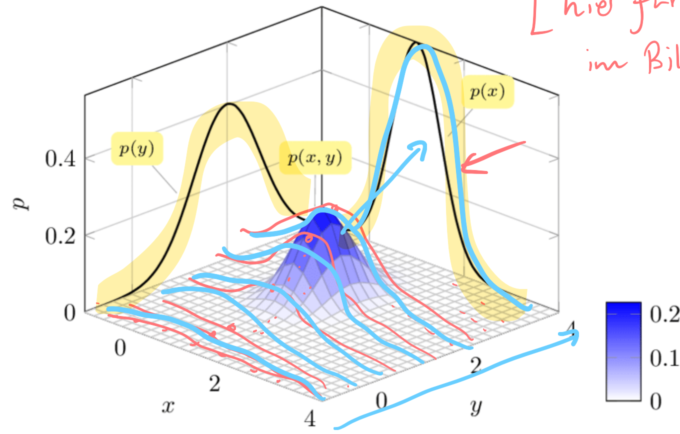
c) $P(X+Y \leq 3) = P(X=1, Y=1) + P(X=1, Y=2) + P(X=2, Y=1)$

$= 1/13 + 4/13 + 0 = 5/13$

Randverteilungen:

$$p(x) = \sum_y p(x,y)$$

$$p(y) = \sum_x p(x,y)$$



[hier für stetige ZV im Bild]

$$p(x) = \sum_y \frac{xy^2}{13} 1\!\!1_{S(x,y)}$$

$$1\!\!1_{S(x,y)} = \begin{cases} 1 & \text{falls } (x,y) \in S \\ 0 & \text{falls nicht} \end{cases}$$

- $p(x=1) = \underbrace{\frac{1 \cdot 1^2}{13}}_{y=1} + \underbrace{\frac{1 \cdot 2^2}{13}}_{y=2} = 5/13$

  $p(x=2) = 1 - P(x=1) = 8/13 \quad (\text{da } \sum_x p(x) = 1)$

- $p(y=1) = \frac{1 \cdot 1^2}{13} + \frac{2 \cdot 1^2}{13} \cdot 0 \quad , \text{da } (x=2, y=1) \notin S$

  $= \frac{1}{13}$

  $p(y=2) = 1 - P(y=1) = 12/13$

e) Unabhängigkeit:

Zwei ZV X, Y sind unabhängig, falls und nur falls: $(\Leftrightarrow)$

$$p(x,y) = p(x)\,p(y)$$

↳ Gegenbeispiel: $P(x=2, y=1) = 0$ aber $P(X=2)\,P(Y=1)$

$= 8/13 \cdot 1/13 \neq 0$

3. Sei $X \sim Bernoulli(0.6)$ und $\longrightarrow P(X=1)=0.6$

$$Y \sim \begin{cases} Bernoulli(0.2) & \text{wenn } X = 1 \\ Bernoulli(0.4) & \text{wenn } X = 0. \end{cases}$$

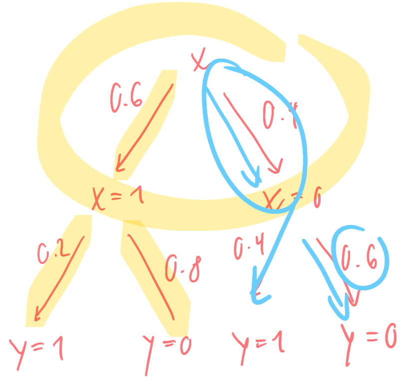
a) Stellen Sie die gemeinsame Wahrscheinlichkeitsverteilung in Tabellenform dar.

b) Berechnen Sie $E(X)$, $E(Y)$ und $E(X+Y)$.

c) Sind X und Y unabhängig?

Struktur

$X \longrightarrow Y$



a)

$Y \backslash X$	$X=0$	$X=1$
$Y=0$	$0.4 \cdot 0.6$	$0.6 \cdot 0.8$
$Y=1$	$0.4 \cdot 0.6$	$0.6 \cdot 0.2$

b) $$E(X) = \sum_x x \, P(X=x) = 1 \cdot \underbrace{P(X=1)}_{=0.6} + 0 \cdot P(X=0) = 0.6$$

$$E(Y) = \sum_y y \, \underbrace{P(Y=y)}_{p(y)} = \sum_y y \underbrace{\sum_x \underbrace{p(y|x)p(x)}_{p(x,y)}}_{p(y)}$$

$$= 0 \cdot \big(p(y=0|x=0)\,P(x=0) + (p(y=0|x=1)\,p(x=1)\big)$$

$$+ 1 \cdot \big(p(y=1|x=0)\,P(x=0) + P(y=1|x=1)\,P(x=1)\big)$$

$$= 1 \cdot (0.4 \cdot 0.4 + 0.2 \cdot 0.6) = 0.28$$

$$E(X+Y) \overset{\text{immer}}{=} E[X] + E[Y] = 0.6 + 0.28$$

c) Gegenbeispiel:

$$p(x=0, y=0) = 0.4 \cdot 0.6 \quad \text{, wobei}$$

$$\neq$$

$$p(x=0)\,p(y=0) = 0.4 \cdot \underbrace{(0.4 \cdot 0.6 + 0.6 \cdot 0.8)}_{\sum_x p(y=0|x) \,=\, p(y=0|x=0) + p(y=0|x=1)}$$

4. Betrachten Sie die folgende bivariate Dichtefunktion

$$f(x,y) = \begin{cases} y(\frac{1}{2} - x) + x & \text{wenn } \underline{0 < x < 1}, \ \underline{0 < y < 2} \\ 0 & \text{sonst.} \end{cases}$$

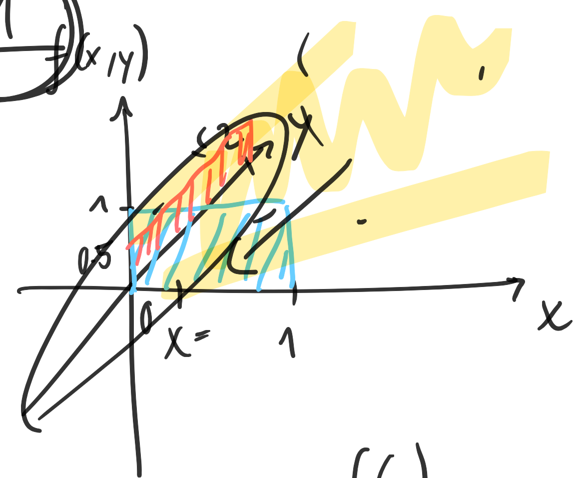
Finden Sie die Randverteilung von X und Y.

Recap: $p(x) = \sum_y p(x,y)$

$$\hookrightarrow f(x) = \int_0^2 f(x,y)\,dy = \int_0^2 \frac{1}{2}y - yx + x\,dy$$

$$= \left[\frac{1}{4}y^2 - \frac{1}{2}y^2 x + xy\right]_0^2 = \frac{1}{4}\cdot 2^2 - \cancel{\frac{1}{2}\cdot 2^2 x} + \cancel{2x}$$

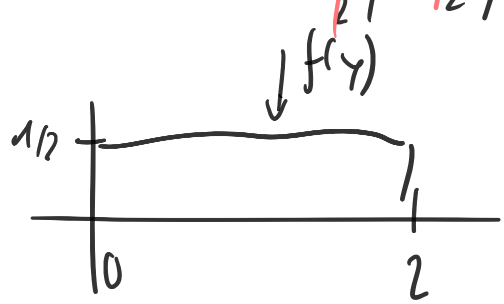
$$= 1$$

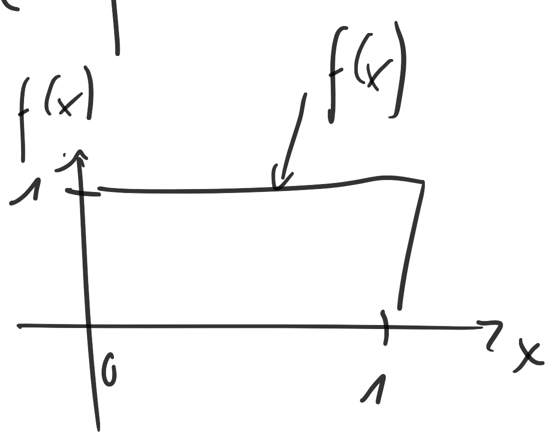


$$f(y) = \int_0^1 f(x,y)\,dx = \int_0^1 \frac{1}{2}y - yx + x\,dx$$

$$= \left[\frac{1}{2}xy - \frac{1}{2}yx^2 + \frac{1}{2}x^2\right]_0^1$$

$$= \cancel{\frac{1}{2}y} - \cancel{\frac{1}{2}y} + \frac{1}{2} = 1/2$$





5. Eine Versicherungsgesellschaft interessiert sich dafür, ob es einen Zusammenhang zwischen Rauchen und Autounfällen gibt. Es wird eine Zufallsstichprobe mit $N = 597$ erhoben. Die Daten sehen folgendermassen aus:

:= k

j =:

Anzahl Unfälle in den letzten 2 Jahren	Raucher	Nicht Raucher	
0	35	170	205
1	79	190	269
2 oder mehr	57	66	123
	171	426	597

Sollte die Versicherungsgesellschaft eine höhere Prämie bei Rauchern verlangen? Verwenden Sie zur Beantwortung der Frage einen geeigneten Hypothesentest mit Signifikanzniveau $\alpha = 5\%$.

h_{11}

Unabhängigkeit: $H_{jk} = P(X=j) \cdot P(Y=k) \cdot N$

$= P(\text{Anzahl Unfälle} = j) \, P(\text{Anzahl Raucher}) \cdot N$

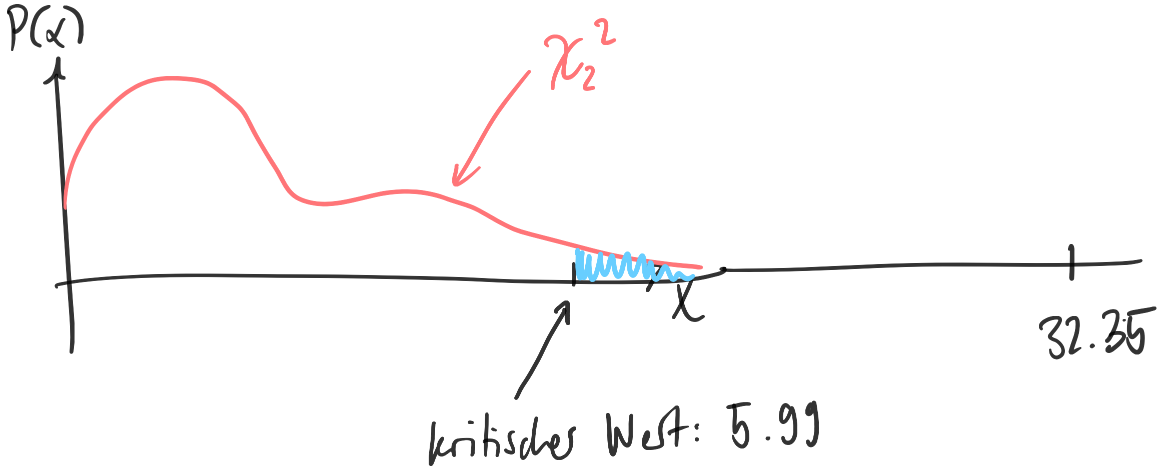
Unfälle	Raucher	Nicht-Raucher
0	$\frac{171}{597} \cdot \frac{205}{597} \cdot 597 = 58$	146
1	77	191
2 oder mehr	35	87

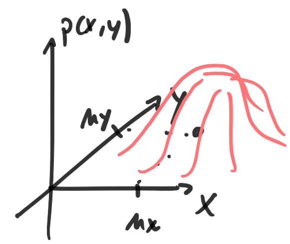
Test der Test-Statistik:

$$\boxed{\sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(h_{jk} - H_{jk})^2}{H_{jk}}} \sim \chi^2_{df}$$

$df := (\#\text{Spalten} - 1) \times (\#\text{Zeilen} - 1) = 2$

↳ hier: 32.35


P(α)
χ₂²
X
32.35
kritischer Wert: 5.99


p(x,y)
my
y
x
mx