Vorlesung 13

Erstklausur aus 2020/21

Besprechung der Lösungen

Bei den Aufgaben 1, 2, 4 und 6 können je 16 Punkte erreicht werden, bei den Aufgaben 3 und 5 können jeweils 18 Punkte erreicht werden.

Geben Sie kurze und treffende Begründungen für Ihre Ergebnisse.

1. a) Es sei T eine exponentialverteilte Zufallsvariable mit Parameter $1/2$. Für welche Zahl c gilt $\mathbf{P}(T > c) = 0.05$?

$$\boxed{\mathbb{P}(T > c) = e^{-\frac{1}{2}c} \stackrel{!}{=} 0.05}$$

$$\Updownarrow$$

$$\mathbb{P}(T \leq c) = 0.95$$

$$\ln\left(e^{-\frac{1}{2}c}\right) = \ln 0.05$$

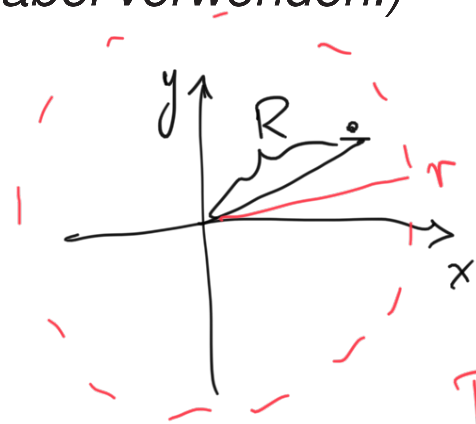
$$-\frac{1}{2}c = \ln \frac{1}{20} = -\ln 20$$

$$c = 2 \ln 20 = 5.99$$

b) Die Zufallsvariable (X, Y) sei standard-normalverteilt auf $\mathbb{R}^2$, d.h. (X, Y) ist eine $\mathbb{R}^2$-wertige Zufallsvariable mit Dichte $\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\,dx\,dy$. Der Abstand des Punktes (X, Y) vom Ursprung ist $R := \sqrt{X^2 + Y^2}$. Finden Sie dasjenige $r > 0$, für welches gilt:

$$\mathbf{P}(R \leq r) = 0.95.$$

(In der Vorlesung oder in den Übungen hergeleitete Tatsachen dürfen Sie dabei verwenden.)



$R = \sqrt{X^2 + Y^2}$

R^2 ist $\operatorname{Exp}(1/2)$-verteilt

$$\mathbb{P}(R \leq r) = \mathbb{P}(R^2 \leq r^2) = \mathbb{P}(T \leq r^2) \overset{!}{=} 0.95$$

siehe a), mit T $\operatorname{Exp}(1/2)$ verteilt

$$5.99 = c \overset{!}{=} r^2$$

$$\Longrightarrow \quad r = 2.45 \; (= \sqrt{5.99})$$

$X \cdot Y = 0$

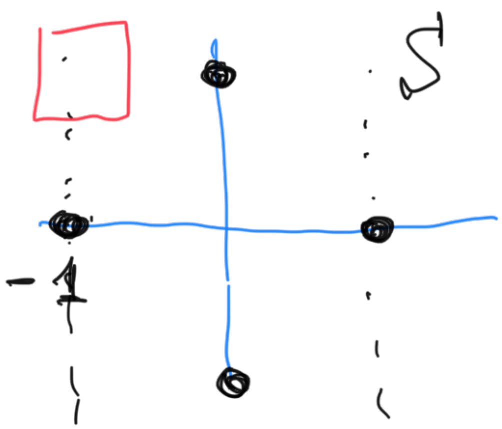
2. Wir betrachten die Menge $S := \{(1,0),(0,1),(-1,0),(0,-1)\}$; S ist also eine 4-elementige Teilmenge von $\mathbb{R}^2$. Die Zufallsvariable $Z := (X,Y)$ sei uniform verteilt auf S.

a) Sind X und Y

i) unabhängig?

ii) unkorreliert?

i)



$$\mathbb{P}(X=-1, Y=+1) = 0$$

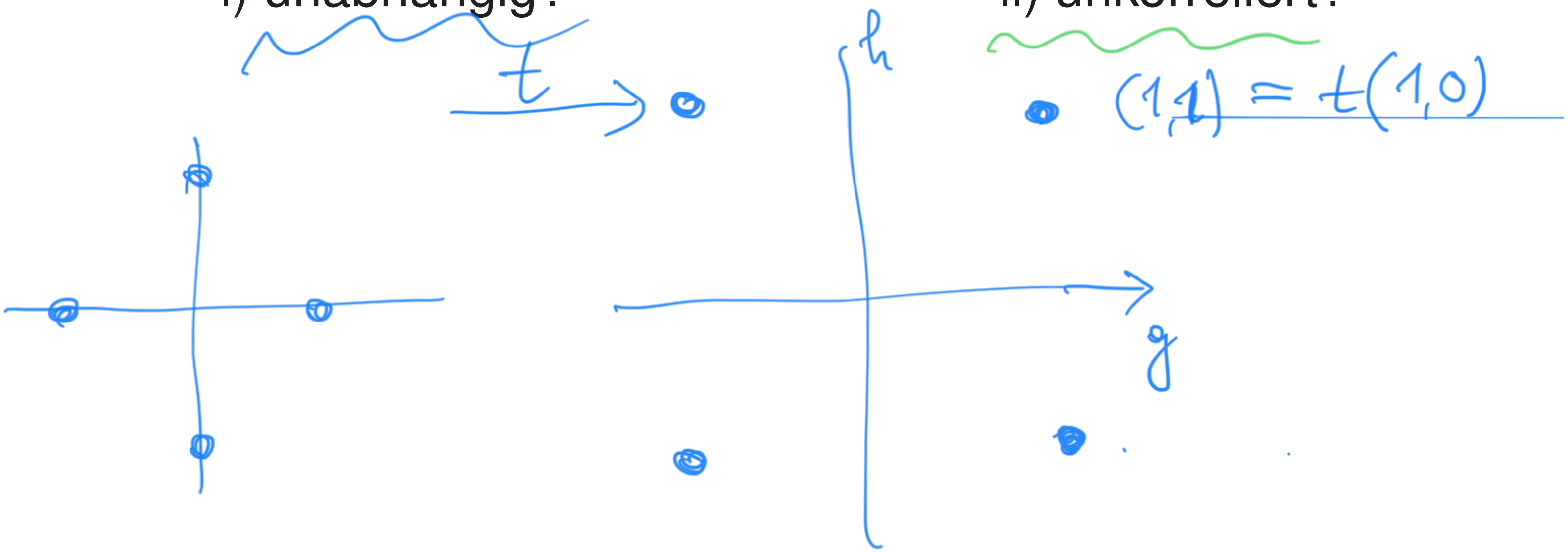
$$\underbrace{\mathbb{P}(X=-1)}_{1/4} \cdot \underbrace{\mathbb{P}(Y=+1)}_{1/4} = \frac{1}{16}$$

also: Produktformel verletzt! nicht erfüllt!

ii) $\mathrm{Cov}[X,Y] = \underbrace{\mathbb{E}[XY]}_{=0} - \underbrace{\mathbb{E}[X]}_{=0}\underbrace{\mathbb{E}[Y]}_{=0} = 0$ also unkorreliert!

b) Wir setzen $G := X + Y$, $H := X - Y$. Sind G und H

i) unabhängig? ii) unkorreliert?



$t(x,y) = (x+y,\ x-y)$

(G,H) ist uniform verteilt auf $\{-1,1\}^2$

also ein zweifacher ± 1 Münzwurf,

also unabh.

Alternativ: Prüfe die Produktformel

$\mathbb{P}(G=1)\mathbb{P}(H=1) = \frac{1}{2}\,\frac{1}{2} = \frac{1}{4} = \mathbb{P}(G=1,\ H=1) =$

(ii) Unabh $\Rightarrow$ Unkorreliert.

3. In unseren Übungen hatten wir eine Population bestehend aus 100 Bäumen betrachtet, von denen 40 Bäume die Höhe 1 hatten, 30 die Höhe 2 und 30 die Höhe 3.

a) Wie groß ist der Erwartungswert und die Varianz der Höhe eines rein zufällig aus den 100 gezogenen Baumes?

siehe Übungen

$\mu = 1.9$

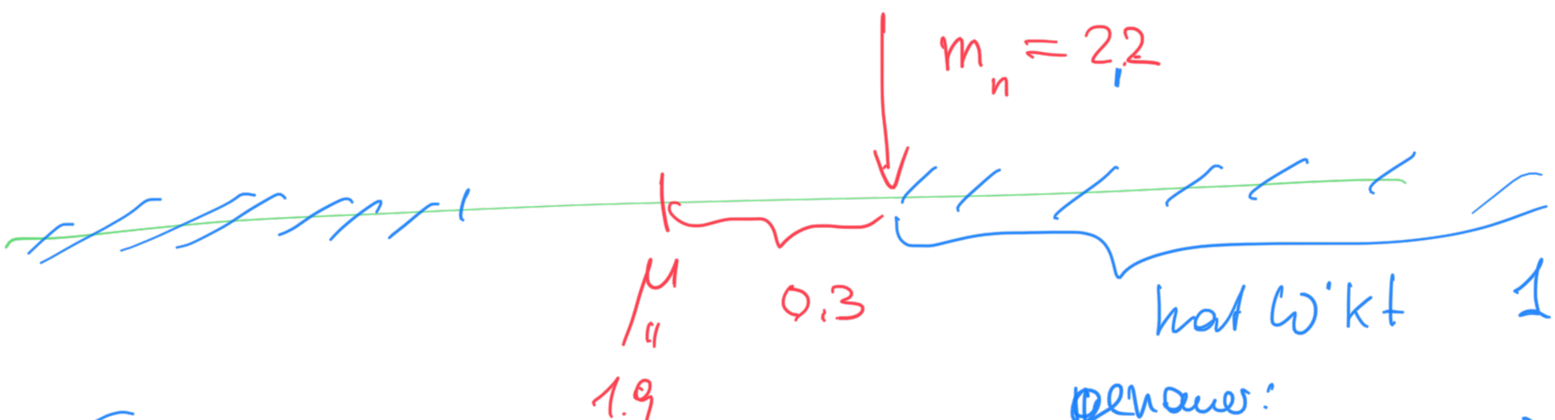
b) Wie groß ist die Standardabweichung des Stichprobenmittelwertes M_{25} einer Stichprobe vom Umfang 25, die rein zufällig und ohne Zurücklegen aus der Population entnommen wird?
(Zur Erinnerung: $M_n := \frac{1}{n}(X_1 + \cdots + X_n)$,
mit X_i:= Höhe des i-ten gezogenen Baumes.)

$n = 25$

$g = 100$

$$\sigma_{M_n} = \sqrt{\sigma^2_{M_n}} = \sqrt{\frac{1}{n}\sigma^2_{X_1}\frac{g-n}{g-1}} = 0{,}145$$

c) Aus der Population wurde nun eine Stichprobe von 25 Individuen ohne Zurücklegen entnommen, der Mittelwert in der Stichprobe war 2.2. Wir stellen die Hypothese auf den Prüfstand, dass die Auswahl der 25 Individuen rein zufällig erfolgte. Wie wahrscheinlich ist es, dass in der beschriebenen Situation der Mittelwert des Merkmals in einer rein zufälligen Stichprobe mindestens so weit vom Populationsmittelwert abweicht wie der in den Daten beobachtete? Verwenden Sie die Normalapproximation und drücken Sie Ihr Ergebnis durch die Standard-Normalverteilungsfunktion Φ aus.



genauer: $\mathbb{P}(M_{25} \geq 2{,}2) \approx$ —"—

$$\frac{2.2 - 1.9}{\sigma_{M_n}} = \frac{0.3}{0.145} = \underline{\underline{2.07}}$$

$$\mathbb{P}(|M_{25} - \mu| \geq 0.3) \approx 2(1 - \Phi(2.07))$$
$$(= 2 \cdot \Phi(-2.07))$$

4. Die Zufallsvariable U sei uniform verteilt auf $[0,1]$. Berechnen Sie die Dichte der $\mathbb{R}_+$-wertigen Zufallvariablen

~~Die Wurzel aus einer Exp(1/2) Vert.~~ D.h. ist Standard Rayleigh-Vert.

$$X := \sqrt{-\ln(U^2)}.$$

oder ganz schnell

$$-\ln(U^2) = 2\underbrace{(-\ln U)}_{\text{ist Exp(1)}}$$

Exp(1/2) Vert.

Hinweis: Sie können entweder den Weg über die Bestimmung der Verteilungsfunktion von X gehen oder Bausteine aus der Lehreveranstaltung zusammenfügen. In jedem Fall sollten Sie sich daran erinnern, wie man $\ln(a^2)$ *umformen kann.*

$$F_X(b) := \mathbb{P}\left(\sqrt{-\ln(U^2)} \le b\right) = \mathbb{P}\left(-\ln(U^2) \le b^2\right)$$

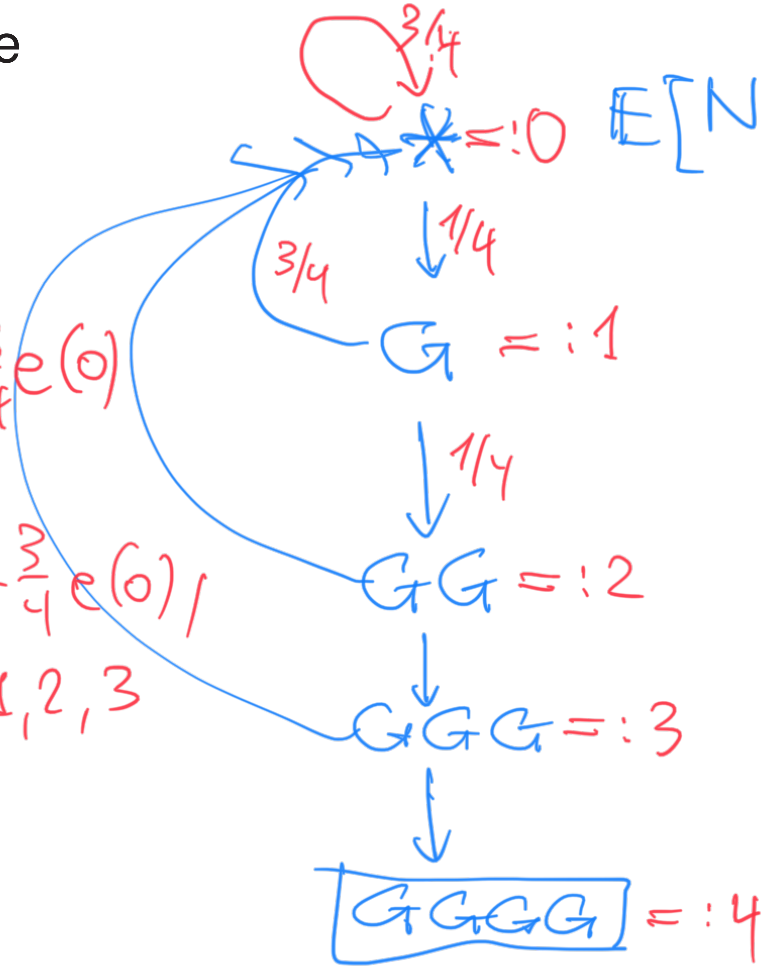
$$= \mathbb{P}\left(\ln U^2 \ge -b^2\right) = \mathbb{P}\left(U^2 \ge e^{-b^2}\right)$$

$$= \mathbb{P}\left(U \ge e^{-b^2/2}\right) = 1 - e^{-b^2/2}, \quad b \ge 0$$

$$f_X(b) = \frac{d}{db}\left(1 - e^{-b^2/2}\right) = b e^{-b^2/2}, \quad b \ge 0.$$

5. Wie in einer unserer Übungsaufgaben betrachten wir ein rein zufälliges $ACGT$-Würfeln. Es sei N_r die Anzahl der Würfe, bei der G erstmals viermal "in Reihe" aufgetreten ist, und N_s die Anzahl der Würfe, bei der G erstmals viermal "in Summe" aufgetreten ist. (Beispiel: Ist die realisierte Folge gleich $TGACGGGG$, dann hat N_r den Ausgang 8 und N_s den Ausgang 7.) Berechnen Sie

(i) Berechnen Sie $\mathbf{E}[N_r]$



$\mathbb{E}[N_r] = \mathbb{E}_*[T_{GGGG}]$
$= e(0)$

$e(0) = 1 + \frac{1}{4}e(1) + \frac{3}{4}e(0)$

$\vdots$

$e(k) = 1 + \frac{1}{4}e(k+1) + \frac{3}{4}e(0),$

$k = 0, 1, 2, 3$

$e(4) = 0$

$\boxed{e(0) = 340}$

(ii) Berechnen Sie $\mathbf{E}[N_s]$

Hinweis zu (ii): Wie ist die Verteilung der Anzahl der Würfe zwischen dem ersten und dem zweiten Auftreten von G verteilt?

$$N_s = W_1 + W_2 + W_3 + W_4.$$

Geom$(1/4)$-verteilt (und unabh.)

Linearität des EW liefert

$$\mathbb{E}[N_s] = 4 \cdot 4 = 16.$$

6. Z_1 und Z_2 seien unkorrelierte reellwertige Zufallsvariable mit Erwartungswert 0 und Varianz 1. Es sei

$\mu_X = 5$
$\mu_Y = -3$

$$X := 2Z_1 + 2Z_2 + 5$$
$$Y := 4Z_1 - Z_2 - 3$$

$\boxed{\begin{aligned}\sigma_X^2 &= 4+4 = 8\\ \sigma_Y^2 &= 16+1 = 17\end{aligned}}$

Bestimmen Sie

a) die Kovarianz von X und Y,

$$\mathrm{Cov}[2Z_1 + 2Z_2 + 5,\ 4Z_1 - Z_2 - 3]$$
$$= \underbrace{\mathrm{Cov}[2Z_1, 4Z_1]}_{2\cdot 4\,\underbrace{\mathrm{Var}[Z_1]}_{=1}} + \underbrace{\mathrm{Cov}[2Z_1, -Z_2]}_{=0} + \ldots$$
$$= 2\cdot 4 + 2\cdot(-1) = 6$$

b) den Korrelationskoeffizienten von X und Y

$$K_{XY} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y} = \frac{6}{\sqrt{8}\sqrt{17}}$$

c) die beste affin lineare Prognose von Y auf der Basis von X.

$$\boxed{\beta_1 = \frac{\mathrm{Cov}[X,Y]}{\sigma_X^2}} = \frac{6}{8} = \frac{3}{4}$$

$$y = \beta_1 x + \beta_0$$

$$-3 = \frac{3}{4} 5 + \beta_0$$

$$-\frac{12}{4} - \frac{15}{4} = \beta_0$$

$$-\frac{27}{4} = \beta_0$$

$$\boxed{\hat{Y} = \beta_1 X + \beta_0}$$

Alternativ: $\beta_1 = \frac{\sigma_Y}{\sigma_X} \kappa_{XY} = \frac{\sigma_Y \cdot \sigma_X \kappa_{XY}}{\sigma_X^2}$ $= \mathrm{Cov}[X,Y]$