WWZ
Wirtschaftswissenschaftliche Fakultät der
Universität Basel

Dr. Thomas Zehrt

Bitte in Druckbuchstaben ausfüllen:

Name	
Vorname	

Prüfung Statistik
Herbstsemester 2020

+ Lösungen

Hinweise:

- Die Prüfung umfasst 12 Aufgaben (1 bis 12).
- Die Bewertung der Teilaufgaben erfolgt gemäss den eingerahmten Punktzahlen.
- Provisorische Berechnungen sind auf den gekennzeichneten Blättern bzw. auf dem Umschlagblatt auszuführen. Diese Blätter sind ebenfalls abzugeben.
- Die definitive Lösung darf von jeder Aufgabe nur eine Version enthalten und hat direkt im Anschluss an diese Aufgabe (bzw. auf der Rückseite des entsprechenden Aufgabenblattes) zu erfolgen. Dabei sollten alle Rechenschritte klar ersichtlich sein.
- Die ausgeteilten Formelsammlungen dürfen nicht beschriftet werden und sind ebenfalls mit der Prüfung abzugeben.

1. (a) Ein Kioskbesitzer notiert 100 Tage lang die Zahl der verkauften Exemplare der Basler Zeitung.

Verkaufte Zeitungen	0	1	2	3	4
Anzahl der Tage	5	17	31	24	23

Bestimmen Sie die empirische Verteilungsfunktion F dieser Daten. $\boxed{2}$

(b) Es sei die folgende (empirische) Verteilungsfunktion F eines Merkmals X gegeben:

$$F(x) = \begin{cases} 0 & \text{für } x \in (-\infty, 2) \\ 0.20 & \text{für } x \in [2, 7) \\ 1 & \text{für } x \in [7, \infty) \end{cases}$$

Berechnen Sie (falls möglich) $F(-30)$, $F(1.999)$, $F(7)$ und $F(2020)$. $\boxed{2}$

(c) Skizzieren Sie den Graphen der Funktion $h(x) := \begin{cases} -1 & \text{für } x \in [-1, 1] \\ |x| & \text{sonst} \end{cases}$ $\boxed{2}$

Lösung:

(a)

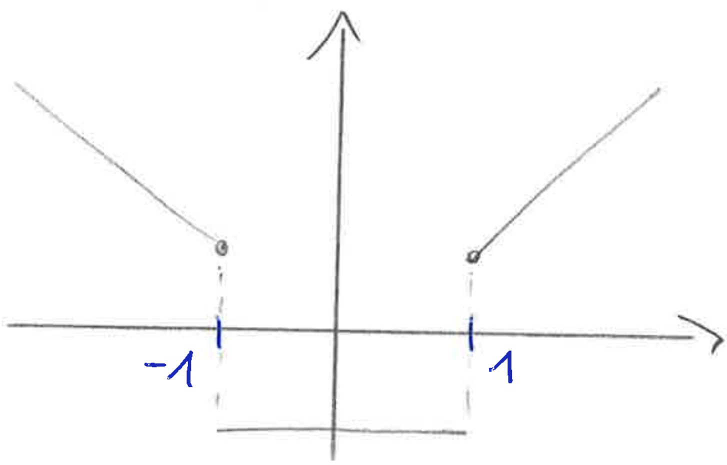
$$F(x) = \begin{cases} 0 & X < 0 \\ 5/100 & 0 \leq X < 1 \\ 22/100 & 1 \leq X < 2 \\ 53/100 & 2 \leq X < 3 \\ 77/100 & 3 \leq X < 4 \\ 1 & X \geq 4 \end{cases}$$

(b) $F(-30) = 0$ $\quad F(1.999) = 0$ $\quad F(7) = 1$

$F(2020) = 1$

(c)

üA



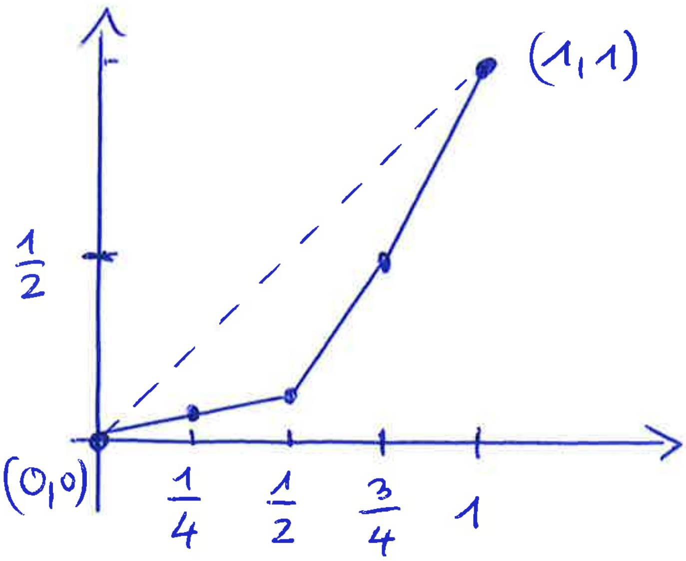
2. (a) Gegeben seien die folgenden vier Messungen eines Merkmals X: $1, 5, 3, 1$.
Berechnen und skizzieren Sie die Lorenz-Kurve. $\boxed{3}$

(b) Eine Messwertreihe $x_1, \ldots, x_n$ habe das arithmetische Mittel $\bar{x} = 14$ und die (empirische) Varianz $var(x) = 25$. Die Messwerte $y_1, \ldots, y_n$ errechnen sich, indem man jedes Element der x-Reihe mit der Konstanten $c = -3.5$ multipliziert und dann die Zahl $d = 5$ addiert. Wie gross sind $\bar{y}$ und $var(y)$? $\boxed{3}$

Lösung:

(a) ordnen(!): 1, 1, 3, 5 $\quad \Sigma = 10$

i	u_i	v_i
0	0	0
1	1/4	1/10
2	2/4	2/10
3	3/4	5/10
4	4/4	10/10



(b) (ÜA)

$$\bar{y} = (-3.5) \cdot 14 + 5 = -44$$

$$var(y) = (-3.5)^2 \cdot 25 = 306.25$$

3. (a) Gegeben seien die x-Werte $x_1 = 5$, $x_2 = 6$, $x_3 = 8$ und $x_4 = 11$. Bestimmen Sie **zwei** Reihen von y-Werten $y_1, \ldots, y_4$, so dass der Korrelationskoeffizient der Wertepaare $(x_1, y_1), \ldots, (x_4, y_4)$ gleich -1 ist. $\boxed{2}$

(b) Seien $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ zwei Datenreihen. Beweisen Sie **Schritt für Schritt und nachvollziehbar**, dass die folgende Gleichung gilt:

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}.$$

Dabei bezeichnen $\bar{x}$ bzw. $\bar{y}$ die jeweiligen arithmetischen Mittel. $\boxed{4}$

Lösung:

(a) linearer Zusammenhang ...

(b) $\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

$$= \sum_{i=1}^{n} (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\bar{y})$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{x} \cdot \underbrace{\sum_{i=1}^{n} y_i}_{= n \cdot \bar{y}} - \bar{y} \cdot \underbrace{\sum_{i=1}^{n} x_i}_{= n \cdot \bar{x}} + \sum_{i=1}^{n} \bar{x}\bar{y}$$

$$= \sum_{i=1}^{n} x_i y_i - n \cdot \bar{x}\bar{y} - n \cdot \bar{x} \cdot \bar{y} + n \cdot \bar{x}\bar{y}$$

$$= \sum_{i=1}^{n} x_i y_i - n \cdot \bar{x}\bar{y}$$

4. (a) Gegeben seien die folgenden Daten:

x_i	0	2	8
y_i	1	3	4

Bestimmen Sie die zugehörige Regressionsgerade. $\boxed{3}$

(b) Erläutern Sie kurz die Methode der kleinsten Quadrate (zur Konstruktion der Regressionsgeraden). $\boxed{3}$

Lösung:

(a) TR $\hat{y} = 0.3269 \cdot x + 1.5769$

oder

	x_i	y_i	x_i^2	$x_i y_i$
	0	1	0	0
	2	3	4	6
	8	4	64	32
$\sum$	10	8	68	38

Gleichungssystem
$$38 = \hat{a} \cdot 68 + \hat{b} \cdot 10$$
$$8 = \hat{a} \cdot 10 + \hat{b} \cdot 3$$

$$\rightarrow \quad \hat{a} = \frac{17}{52}$$
$$\hat{b} = \frac{41}{26}$$

(b) Siehe Skript

5. (a) In der Praxis wird oft ein additives Zeitreihenmodell $y_i = G_i + S_i + R_i$ verwendet.

i. Welche Eigenschaft hat die Funktion S? $\boxed{1}$

ii. Welcher Unterschied besteht zwischen G und $\hat{G}$? $\boxed{1}$

iii. Beschreiben Sie, wie man ein multiplikatives Modell $y_i = G_i \cdot S_i \cdot R_i$ in ein additives Modell umwandeln könnte. $\boxed{1}$

(b) Wir betrachten die folgende Zeitreihe.

t_i	1	2	3	4	5	6	7	8	9
y_i	−1.5	−2.5	−3.5	−4.5	−5.5	−6.5	−7.5	−8.5	−9.5

Schätzen Sie die Komponenten G, S (für $k = 3$) und R. Rechnung oder Begründung. $\boxed{3}$

Lösung:

(a) i. S ist periodisch

ii. $\hat{G}$ ist eine Schätzung der „echten" glatten Komponente G

iii. ln anwenden...

(b) Ansatz: $y_i = G_i + S_i + R_i$

Man sieht: Alle Punkte liegen auf der Geraden

$$y_i = -t_i - 0.5$$

$$\Rightarrow y_i = -t_i - 0.5 + \underbrace{0}_{\hat{S}_i} + \underbrace{0}_{\hat{R}_i}$$

6. (a) Bestimmen Sie die Wahrscheinlichkeit dafür, bei n Würfen mit einem Laplacewürfel mindestens einmal eine 1 zu erzielen. Für welche n ist die Wahrscheinlichkeit, mindestens eine 1 zu erzielen, grösser als 0.99? $\boxed{3}$

(b) Beschreiben Sie (kurz), wie man mit Hilfe eines Monte-Carlo Verfahrens ein bestimmtes Integral $\int_a^b f(x)dx$ schätzen könnte. $\boxed{3}$

Lösung:

(a) $E \mathrel{\hat{=}}$ „mind. eine 1 bei n Würfen"

$$P(E) = 1 - \left(\frac{5}{6}\right)^n \qquad \text{(üA)}$$

zu lösen: $1 - \left(\frac{5}{6}\right)^n > 0.99$

$$\Longleftrightarrow \quad 0.01 > \left(\frac{5}{6}\right)^n \quad \Big| \ln$$

$$\Longleftrightarrow \quad \ln(0.01) > \ln\left(\left(\tfrac{5}{6}\right)^n\right)$$

$$\Longleftrightarrow \quad \ln(0.01) > n \cdot \ln\left(\tfrac{5}{6}\right)$$

$$\Longleftrightarrow \quad \frac{\ln(0.01)}{\ln(5/6)} < n$$

$\uparrow$

$\ln(5/6) < 0$

$$\Longleftrightarrow \quad \underline{\underline{25.28 < n}}$$

(b) Siehe Skript

7. Aus zwei Urnen U_1 und U_2 wird zufällig eine Urne ausgewählt, wobei jede Urne dieselbe Wahrscheinlichkeit besitzt, in die Auswahl zu gelangen. Die zwei Urnen enthalten weisse und schwarze Kugeln, wobei sich in Urne

- U_1 : 8 weisse und 5 schwarze
- U_2 : 3 weisse und 8 schwarze

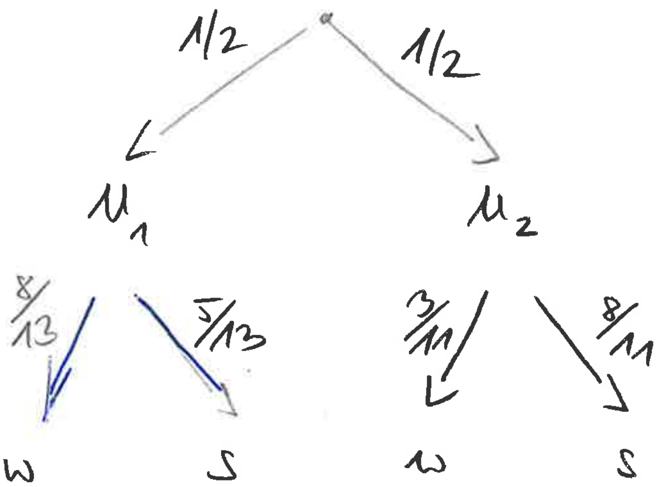
Kugeln befinden. Aus der zufällig gewählten Urne wird nun eine Kugel gezogen.

(a) Skizzieren Sie das zugehörige Baumdiagramm (Ereignisse, Wahrscheinlichkeiten und bedingte Wahrscheinlichkeiten). $\boxed{2}$

(b) Wie gross ist die Wahrscheinlichkeit, dass die gezogene Kugel weiss ist? $\boxed{2}$

(c) Die gezogene Kugel ist schwarz. Wie gross ist die Wahrscheinlichkeit, dass sie aus Urne U_2 stammt? $\boxed{2}$

<u>Lösung:</u>



$$P(w) = \frac{1}{2} \cdot \frac{8}{13} + \frac{1}{2} \cdot \frac{3}{11} = \underline{\underline{0.4440}}$$

$$P(U_2 \mid s) = \frac{P(s \mid U_2) \cdot P(U_2)}{P(s)} = \frac{8/11 \cdot 1/2}{1 - 0.444}$$

$$= \frac{104}{159} = \underline{\underline{0.6541}}$$

8. X und Y seien (unabhängige) Zufallsvariablen mit den folgenden Verteilungen:

k	$P(X=k)$
-1	$1/5$
1	$2/5$
2	$2/5$

k	$P(Y=k)$
-1	$1/3$
1	$2/3$

(a) Bestimmen Sie $E(X)$ und $E(Y)$. $\boxed{1}$

(b) Wie gross ist $Cov(X,Y)$? Berechnung oder Begründung! $\boxed{2}$

(c) Bestimmen Sie $E(X \cdot Y)$. $\boxed{1}$

(d) Bestimmen Sie die Verteilung der Zufallsvariablen $Z = X^2 + 1$. $\boxed{2}$

Lösung:

(a) $E(X) = (-1)\cdot\frac{1}{5} + 1\cdot\frac{2}{5} + 2\cdot\frac{2}{5} = \underline{\underline{1}}$

$E(Y) = \underline{\underline{\frac{1}{3}}}$

(b) $Cov(X,Y) = 0$, da X und Y nach Voraussetzung unabhängig sind.

(c) $Cov(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y)$

$Cov(X,Y) = 0$ Teil (b), $E(X) = 1$, $E(Y) = \frac{1}{3}$

$E(X\cdot Y) = \frac{1}{3}$ //

(d) $Z = X^2 + 1$

Z	$1^2+1=2$ $(-1)^2+1=2$	$2^2+1=5$
P	$3/5$	$2/5$

9. X sei eine binomialverteilte Zufallsgrösse mit $n = 10$ und $p = 0.75$.

(a) Welche Werte kann X annehmen? $\boxed{1}$

(b) Wie gross ist der Erwartungswert und die Varianz von X? $\boxed{1}$

(c) Bestimmen Sie die folgenden Wahrscheinlichkeiten: $\boxed{4}$

- $P(X = 2)$,
- $P(X \leq 2)$,
- $P(X = 4|X = 2)$ und
- $P(X \leq 2|X \geq 2)$.

Lösung: (üA)

(a) $0, 1, \dots, 10$

(b) $10 \cdot 0.75 = \underline{\underline{7.5}}$ $\quad 7.5 \cdot 0.25 = \underline{\underline{1.875}}$

(c) ≈ 0.0004

≈ 0.0004 $\quad \Sigma$

$= 0$

$\approx P(X=2)/P(X \geq 2) = \dfrac{0.0004}{0.9996}$

10. (a) Seien $X_1, X_2, \dots$ unabhängige und identisch (zweipunkt)verteilte Zufallsvariablen mit

$$X_i = \begin{cases} 0 & \text{mit } P(X_i = 0) = 1/4 \\ 3 & \text{mit } P(X_i = 3) = 3/4 \end{cases}$$

und sei $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$.

Bestimmen Sie die Verteilung der Zufallsvariablen Z_3, sowie den Erwartungswert $E(Z_3)$ und die Varianz $Var(Z_3)$. $\boxed{4}$

(b) Sei

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

Beweisen/Begründen Sie, dass $\Phi(-x) = 1 - \Phi(x)$ für alle $x \in \mathbb{R}$ gilt. $\boxed{2}$

Lösung:

(a) $Z_3 = \frac{1}{3}(X_1 + X_2 + X_3)$

0	1	2	3
$\left(\frac{1}{4}\right)^3$	$3 \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)$	$3\left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^2$	$\left(\frac{3}{4}\right)^3$

$E(X_i) = \frac{9}{4}$

$Var(X_i) = \left(0 - \frac{9}{4}\right)^2 \cdot \frac{1}{4} + \left(3 - \frac{9}{4}\right)^2 \cdot \left(\frac{3}{4}\right) = \frac{27}{16}$

$E(Z_3) = \frac{1}{3} \cdot 3 \cdot \frac{9}{4} = \underline{\underline{\frac{9}{4}}}$

$Var(Z_3) = \frac{1}{9} \cdot 3 \cdot \frac{27}{16} = \frac{1}{3} \cdot \frac{27}{16} = \underline{\underline{\frac{9}{16}}}$

(b) Beachten Sie den Unterschied zwischen φ (klein) und $\underline{\Phi}$ (gross)!

11. Für eine Zufallsvariable gelte $X \sim Po(\lambda)$.

(a) Welche Art von Zufallsexperiment führt typischerweise zu poissonverteilten Zufallsvariablen? Welche Bedeutung hat der Parameter λ? $\boxed{2}$

(b) Bestimmen Sie einen ML-Schätzer für λ bezüglich der konkreten Stichprobe $(3, 2, 5)$. Der Weg zur Herleitung **muss nachvollziehbar angegeben** werden. $\boxed{4}$

Lösung:

(a) Siehe z.B. Formelsammlung (Seite 14)

(b) $L(\lambda; 3, 2, 5)$

$$= \frac{\lambda^3}{3!} e^{-\lambda} \cdot \frac{\lambda^2}{2!} e^{-\lambda} \cdot \frac{\lambda^5}{5!} e^{-\lambda}$$

$$= \underbrace{\frac{1}{3!\,2!\,5!}}_{=:A} \cdot \lambda^{10} \cdot e^{-3\lambda} =: L$$

$$\ln(L) = \ln(A^{-1}) + \ln(\lambda^{10}) + \ln(e^{-3\lambda})$$

$$= \ln(A^{-1}) + 10 \cdot \ln(\lambda) - 3\lambda$$

$$\frac{\partial}{\partial \lambda} \ln(L) = \frac{10}{\lambda} - 3 \overset{!}{=} 0$$

$$\rightarrow \lambda = \frac{10}{3}$$

12. Ein neues Medikament soll nur dann eingesetzt werden, wenn die Wahrscheinlichkeit p für eine Heilung mindestens 0.75 ist. Die Wirksamkeit des Medikamentes müsste also * aktiv (statistisch) bewiesen werden.

Bei einer Studie wird das Medikament $n = 30$ Patienten verabreicht, wobei 25 Patienten geheilt werden. Wir wollen durch einen Test entscheiden, ob die beobachtete Anzahl der Heilerfolge für den Einsatz des Medikamentes spricht.

(a) Formulieren Sie H_0 und H_1. $\boxed{1}$

(b) Welche Fehler können bei diesem Test auftreten? Welchen Fehler sollte man möglichst vermeiden? $\boxed{2}$

(c) Bestimmen Sie den Verwerfungsbereich zum Signifikanzniveau $\alpha = 0.1$. Welche Entscheidung treffen Sie? $\boxed{3}$

Hinweis: $\sum_{k=0}^{25} \binom{30}{k} 0.75^k \, 0.25^{30-k} = 0.9021$ und rechnen Sie **ohne** Approximation.

Lösung:

(a) $H_0: p \leq 0.75$

$H_1: p > 0.75$

das muss ein rechtsseitiger Test sein!
(siehe *)

(b) Siehe Formelsammlung (Seite 18)

(c) $R = \{x, x+1, \dots, 30\}$

kleinste Zahl
s.d. $\sum_{k=0}^{x-1} \binom{30}{k} 0.75^k \cdot 0.25^{30-k} \geq 0.9$

$\xrightarrow{\text{Hinweis}}$ $x - 1 = 25$

$x = 26$