Prof. Dr. M. Maathuis ETH Zürich Sommer 2017

Stochastik
(BSc D-MAVT / BSc D-MATH / BSc D-MATL)

1. (10 Punkte) Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Kreise die richtige Antwort ein.

Beachte: Jede richtige Antwort gibt einen Punkt. Es gibt keinen Abzug für falsche Antworten. Wir empfehlen daher alle Fragen zu beantworten.

a) Sei $P(A) = 0.5$, $P(A \cap B) = 0.25$ und $P(A|B) = P(B|A)$. Welche Aussage ist richtig?

1. $P(B) = 0.25$.
2. $P(B) = 0.5$.
3. $P(B) = 0.75$.

b) Eine Mutter hat genau zwei Kinder. Mindestens eins von ihnen ist ein Junge. Wie gross ist die Wahrscheinlichkeit, dass beide Kinder Jungen sind?

1. 1/2.
2. 1/3.
3. 1/4.

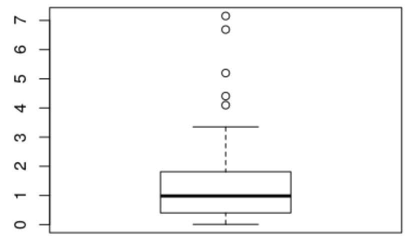
c) Welche Aussage ist richtig?

1. Wenn X_1 und X_2 Bernoulli-verteilt sind mit dem gleichen Parameter $p = 0.5$, dann ist $X_1 + X_2 \sim Bin(2, 0.5)$.
2. Wenn Y_1 und Y_2 unabhängig und identisch Poisson-verteilt sind mit Parameter λ, dann ist $Y_1 + Y_2$ Poisson-verteilt mit Parameter 2λ.
3. Wenn $Z_1 \sim Bin(n, 0.2)$ und $Z_2 \sim Bin(m, 0.3)$ unabhängig sind, dann ist $Z_1 + Z_2 \sim Bin(n + m, 0.5)$.

d) Es seien $X_1, ..., X_n$ i.i.d. F-verteilt, wobei F eine Verteilung ist mit Erwartungswert μ und Varianz σ^2. Wir schreiben $S_n := \sum_{i=1}^n X_i$ und $\bar{X}_n := S_n/n$. Welche Aussage ist richtig?

1. Falls F eine Normalverteilung ist, dann ist $S_n \sim N(n\mu, n\sigma^2)$.
2. Der zentrale Grenzwertsatz sagt aus, dass für eine beliebige Verteilung F und für grosses n, $\bar{X}_n$ ungefähr $N(\mu/n, \sigma^2/n)$-verteilt ist.
3. Aus dem Gesetz der Grossen Zahlen folgt, dass $P(|S_n - n\mu| > 0.1) \to 0$ für $n \to \infty$.

e) Betrachte den folgenden Boxplot. Welche Aussage ist richtig?



1. Diese Daten sind gut vereinbar mit einer Exponentialverteilung.
2. Ungefähr 75% der Daten sind kleiner als 3.5.
3. Das aritmetische Mittel der Daten wird durch die dicke schwarze Linie dargestellt.

f) Wir betrachten die Maximum Likelihood Methode. Welche Aussage ist richtig?

1. Für den geschätzten Parameter macht es keinen Unterschied, ob wir die Likelihood- oder die log-Likelihoodfunktion maximieren.
2. Die Maximum-Likelihood Methode kann konzeptionell nur für stetige Verteilung angewandt werden.
3. Bei stetigen Verteilungen ist die log-Likelihoodfunktion gegeben als der Logarithmus von der Summe der Dichtefunktionen.

g) Seien $X_1, ..., X_n$ i.i.d. Poisson-verteilt mit Parameter $\lambda > 0$. Welche Aussage ist richtig?

1. $\sum_{i=1}^n X_i^2/n$ ist ein Momentenschätzer für λ.
2. $\sum_{i=1}^n X_i/(n-1)$ ist ein Momentenschätzer für λ.
3. $\sum_{i=1}^n X_i/n$ ist ein Momentenschätzer für λ.

h) Die Unterrichtskommission möchte die Gesamtzufriedenheit der Studierenden in einer Vorlesung abschätzen. Welche Aussage ist richtig?

1. In einer Vorlesung mit zweimal so vielen Studierenden braucht man eine viermal so grosse Stichprobe um die gleiche statistische Genauigkeit zu erreichen.
2. Wir schicken jedem Studierenden der Vorlesung eine Umfrage. Wenn mehr als 50% der Studierende die Umfrage ausfüllt, dann haben wir eine Zufallsstichprobe.
3. Eine kleinere zufällig gewählte Stichprobe ist statistisch gesehen oft besser als eine grössere nicht zufällig gewählte Stichprobe.

i) Wir führen einen z-Test zum Signifikanzniveau von 5% durch. Es gilt $n = 16$; $\sigma^2 = 1$; $H_0 : \mu = 0$; $H_A : \mu \neq 0$. Welche Aussage ist richtig?

1. Die Macht unter $H_A : \mu = -5$ ist kleiner als die Macht unter $H_A : \mu = -4$.
2. Wenn das beobachtete arithmetische Mittel $\bar{X}_n = -0.5$ ist, dann wird H_0 verworfen.
3. Wenn H_0 nicht verworfen wird, dann gilt $\mu = 0$.

j) Wir möchten testen, ob es bei den Schülern in der Schweiz einen Unterschied gibt zwischen der Durchschnittsnote in Musik und der Durchschnittsnote in Mathe. Dazu wählen wir zufällig 200 Schüler im Alter von 12-14 Jahren aus und betrachten ihre Noten in Mathe und Musik. Welcher statistischer Test ist am besten geeignet?

1. Einseitig, gepaart.
2. Einseitig, ungepaart.
3. Zweiseitig, gepaart.
4. Zweiseitig, ungepaart.

2. (8 Punkte) Eine Wetterstation versucht Sturmböen auf täglicher Basis vorherzusagen. Aus langjähriger Erfahrung wissen wir Folgendes: Sagt die Wetterstation für einen Tag Sturmböen voraus, gibt es sie an diesem Tag tatsächlich mit Wahrscheinlichkeit 0.7. Sagt die Wetterstation für einen Tag keine Sturmböen voraus, ist die Wahrscheinlichkeit 0.9, dass es an diesem Tag in der Tat keine Sturmböen gibt. Die Wahrscheinlichkeit, dass es an einem konkreten Tag Sturmböen gibt bezeichnen wir mit p.

a) Die Wahrscheinlichkeit, dass die Wetterstation für einen konkreten Tag Sturmböen vorhersagt, bezeichnen wir mit q. Zeige, dass $p = 0.1 + 0.6q$. Kann p den Wert 0.85 annehmen? Begründe deine Antwort.

Tipp: Schreibe V für das Ereignis, dass die Wetterstation Sturmböen vorhergesagt hat und B für das Ereignis, dass es Sturmböen gibt.

b) Jan behauptet, dass es an 85% der Tage Sturmböen gibt. Peter dagegen behauptet, dass es an 60% der Tage Sturmböen gibt. Wer hat sicher unrecht? Begründe deine Antwort.

c) Wir nehmen an, dass es an einem konkreten Tag Sturmböen gibt. Was ist die Wahrscheinlichkeit, dass die Vorhersage der Wetterstation für diesen Tag richtig war, wenn $p = 0.5$?

Die Wetterstation möchte p aus historischen Daten schätzen. In den vergangenen 50 Tagen gab es nur am 10. und am 42. Tag Sturmböen, an den restlichen 48 Tagen gab es keine. Wir beschreiben diese Daten mit $x_1, ..., x_{50}$, wobei $x_{10} = x_{42} = 1$ und $x_i = 0$ für $i \neq 10, 42$.

d) Betrachte $x_1, ..., x_{50}$ als Realisierungen von den unabhängigen und gleichverteilten Zufallsvariablen $X_1, ..., X_{50}$. Was ist die Verteilung von X_1?

e) Gebe die Likelihood Funktion zur gegebenen Stichprobe $x_1, ..., x_{50}$ an. Berechne daraus das MLE für p.

f) Nehmen wir an, dass $p = 0.05$. Was ist die Wahrscheinlichkeit, dass es in den kommenden 50 Tagen mindestens einmal Sturmböen gibt?

3. (11 Punkte) Ein Versicherer modelliert die jährlichen Schäden (in Millionen CHF) verursacht von Sturmböen mit der Zufallsvariable S und die jährlichen Schäden (in Millionen CHF) verursacht von Überflutungen mit der Zufallsvariable U. Er nimmt dabei die folgende gemeinsame Dichte an:

$$f_{S,U}(s,u) = \begin{cases} \frac{c}{s^4 u^4}, & \text{wenn } s \geq 1, u \geq 1, \\ 0, & \text{sonst.} \end{cases}$$

a) Bestimme die Konstante c.

b) Zeige, dass die Randdichte von S, f_S sich wie folgt angeben lässt:

$$f_S(s) = \begin{cases} \frac{3}{s^4}, & \text{wenn } s \geq 1, \\ 0, & \text{sonst.} \end{cases}$$

Was ist die Wahrscheinlichkeit, dass die von Sturmböen verursachten Schäden 3 Millionen CHF übersteigen?

c) Zeige, dass S und U unabhängig sind.

Der Versicherer will seine Tätigkeit langfristig nur auf Sturmschaden, das heisst auf S konzentrieren. Er möchte dabei die Wahrscheinlichkeit approximieren, dass die durchschnittliche jährliche Sturmschaden 2 Millionen CHF übersteigen.

d) Berechne den Erwartungswert und die Varianz von S.

e) Wir nehmen an, dass die Verteilung von jährlichen Sturmschaden in diesen 30 Jahren gleich bleibt. Was ist (annähernd) die Wahrscheinlichkeit, dass in den nächsten 30 Jahren die Sturmschaden im Schnitt 2 Millionen CHF übersteigen. Triff dabei geeignete Annahmen und beschreibe sie.

Um einen Eindruck von den jährlichen Sturmschaden zu bekommen, setzt der Versicherer einen Zufallsgenerator für die uniforme Verteilung auf $[0, 1]$ ein. Die ersten zwei Werte, die vom Generator geliefert werden, sind 0.31 und 0.94.

f) Gib ein Verfahren an, das aus einer $Uni[0, 1]$-verteilten Zufallsvariable X eine Zufallsvariable Y generiert, die die gleiche Verteilung wie S hat. Wende dieses Verfahren auf die Zahlen 0.31 und 0.94 an.

4. **(10 Punkte)** Die Gesundheitsbehörde in Chile möchte das Arsengehalt des Trinkwassers in einer bestimmten Region den Landes testen. Die Weltgesundheitsorganisation (WHO) empfiehlt einen Höchstwert von $10\mu g$ Arsen pro Liter Trinkwasser. Ein Arsengehalt über diesem Grenzwert gilt als potenziell gesundheitsschädlich. Die Behörde möchte testen, ob dieser Grenzwert mit Signifikanzniveau 5% überschritten wird. Darum führt sie Messungen durch, die folgende Messdaten ergeben (in μg pro Liter):

Messung Nr.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Messwerte (x_i)	10.7	6.1	8.1	8.9	10.9	13.6	10.4	12.1	12.3	8.6

Die Daten (x_i) können als Realisierungen von unabhängig und identisch-verteilten Zufallsvariablen $X_1, ..., X_{10}$ aufgefasst werden. Man kann folgende Werte berechnen: $\bar{x} = 10.17$; $s_{10}^2 = 2.256$.

Das Unternehmen Bonaparte AG, ein Tochterunternehmen einer Arsen-Raffinerie, bietet sich an, die Messdaten auszuwerten. Bonaparte AG geht davon aus, dass die Daten einer nicht-symmetrischen Verteilung folgen.

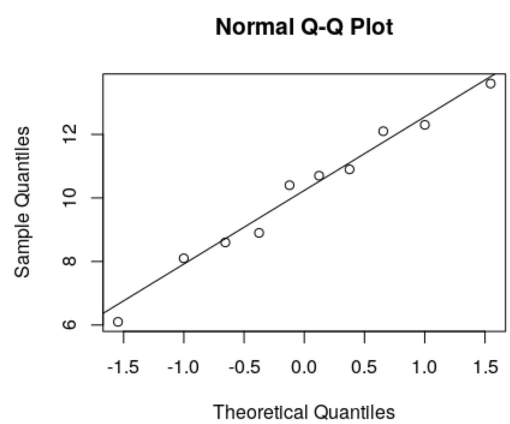
a) Welchen Test kann Bonaparte AG unter diesen Annahmen durchführen?

b) Was sind die Null und Alternativhypothesen für diesen Test?

c) Führe den in a) gewählten Test, mittels Berechnung des Verwerfungsbereichs, durch.

d) Berechne die Macht des Tests für die alternative Hypothese $P[X_1 > 10] = 0.75$.

Die Gesundheitsbehörde hat jedoch einen begründeten Verdacht auf Interessenkonflikt. Sie beauftragt deswegen ein unabhängiges Unternehmen, die Bovary AG, die Messdaten auszuwerten. Die Bovary AG stellt die Daten im folgenden Normal QQ-Plot dar.



e) Anhand des QQ-Plots, welchen Test soll Bovary AG durchführen?

f) Was sind nun die Null und Alternativhypothesen?

g) Führe den in e) gewählten Test mithilfe des p-Werts durch.