Prof. Dr. M. Schweizer — ETH Zürich — Sommer 2014

Wahrscheinlichkeit und Statistik
BSc D-INFK

Name:	
Vornahme:	
Stud. Nr.:	

Das Folgende bitte nicht ausfüllen!

Aufg.	Summe	Kontr.	Pkte.-Max.
1			10
2			10
3			10
4			10

Punktetotal:	
Vollständigkeit:	

Bitte wenden!

Hinweise zur Prüfung

Prüfungsdauer: 2 Stunden.

Hilfsmittel: 10 A4-Seiten resp. 5 Blätter Zusammenfassung. Kein Taschenrechner!

Bitte beachten Sie folgende Punkte:

- Legen Sie Ihre Legi offen auf den Tisch.
- Tragen Sie Ihre Daten in dieses Deckblatt ein und schreiben Sie auf jedes Blatt Ihren Namen.
- Beginnen Sie jede Aufgabe **auf einem neuen Blatt**.
- Schreiben Sie nicht mit Bleistift, rotem oder grünem Kugelschreiber.
- Um die volle Punktzahl zu erreichen, begründen Sie alle Resultate durch Zwischenschritte und -rechnungen (ausser Aufgabe 1) und vereinfachen Sie die Resultate so weit wie möglich.
- Lesen Sie alle Aufgaben durch, bevor Sie beginnen. Für eine genügende Note wird nicht erwartet, dass Sie alle Aufgaben in der Ihnen zur Verfügung stehenden Zeit lösen können.
- Es dürfen sich nur erlaubte Hilfsmittel auf dem Tisch befinden, d.h. 10 A4-Seiten resp. 5 Blätter Zusammenfassung. Kein Taschenrechner!

Siehe nächstes Blatt!

Aufgaben

---

1. (10 Punkte)

Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Für jede richtig beantwortete Frage erhält man 1 Punkt, für jede falsch beantwortete Frage erhält man $-1/2$ Punkte, und für jede nicht beantwortete Frage erhält man 0 Punkte. Ingesamt erhält man mindestens 0 Punkte für die gesamte Aufgabe. **Bitte benutzen Sie zur Beantwortung der Fragen das beiliegende Antwortblatt.**

a) Ein Hersteller produziert Computerchips in drei Fabriken A, B, C. Von den produzierten Chips ist eine gewisse Anzahl defekt. In einer bestimmten Woche ergeben sich folgende Zahlen:

	Fabrik A	Fabrik B	Fabrik C
Anzahl produzierter Chips	1000	500	1500
davon defekt	100	100	200

Wir nehmen an, dass ein aus allen produzierten Chips dieser Woche zufällig ausgewählter Chip tatsächlich defekt ist. Gegeben diese Information, wie gross ist die Wahrscheinlichkeit, dass dieser Chip in Fabrik A produziert wurde?

(i) $\frac{1}{30}$.
(ii) $\frac{1}{10}$.
(iii) $\frac{1}{4}$.

b) Wir wählen nun aus allen *defekten* Chips zufällig 100 Chips aus. Wie gross ist die Wahrscheinlichkeit, dass alle diese 100 Chips in Fabrik C produziert wurden?

(i) $\binom{200}{100}\left(\frac{1}{2}\right)^{200}$.
(ii) $\frac{\binom{200}{100}}{\binom{400}{100}}$.
(iii) $\frac{\binom{1500}{100}\binom{1500}{300}}{\binom{3000}{400}}$.

c) Wie oft muss man eine faire Münze mindestens werfen, um mit mindestens 90% Wahrscheinlichkeit mindestens einmal "Kopf" zu erhalten?

(i) 2 Mal.
(ii) 3 Mal.
(iii) 4 Mal.

d) X und Y seien zwei Zufallsvariablen mit $E[X] = E[Y] = 1$, $E[X^2] = E[Y^2] = 3$ und $\mathrm{Var}[1 + X + 2Y] = 14$. Berechnen Sie $\mathrm{Cov}(X, Y)$.

(i) $\mathrm{Cov}(X, Y) = 1$.
(ii) $\mathrm{Cov}(X, Y) = 2$.
(iii) Es liegen nicht genug Informationen vor, um $\mathrm{Cov}(X, Y)$ zu berechnen.

e) Sei $C \in [0, 1]$ und die Funktion $F : \mathbb{R} \to [0, 1]$ gegeben durch

$$F(x) = \begin{cases} 1 - Ce^{-x}, & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0. \end{cases}$$

Welche der folgenden Aussagen trifft zu?

(i) F ist für kein $C \in [0, 1]$ die Verteilungsfunktion einer Zufallsvariablen.
(ii) F ist für jedes $C \in [0, 1]$ die Verteilungsfunktion einer Zufallsvariablen.
(iii) F ist die Verteilungsfunktion einer Zufallsvariablen genau dann, wenn $C = 1$.

f) X, Y seien Zufallsvariablen mit gemeinsamer Dichte $f(x, y)$ und Randdichten $f_X(x)$ und $f_Y(y)$. Wie berechnet man die Wahrscheinlichkeit $P[X > Y]$?

(i) $P[X > Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f(x, y)\, dy\, dx$.
(ii) $P[X > Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f(x, y)\, dx\, dy$.
(iii) $P[X > Y] = 1 - \int_{0}^{Y} f_X(x)\, dx$.

g) Sei S_n die Anzahl der Sechsen beim n-maligen Werfen eines fairen Würfels. Welche der folgenden Aussagen trifft zu?

(i) S_n ist $Geom(\left(\frac{1}{6}\right)^n)$-verteilt.
(ii) $P[|S_n - \frac{n}{6}| \geq c] \geq \frac{n\frac{5}{36}}{c^2}$ für $n \in \mathbb{N}$ und $c > 0$.
(iii) $P[S_n \geq \frac{2}{3}n] \leq \left(\frac{e^3}{4^4}\right)^{\frac{n}{6}}$ für $n \in \mathbb{N}$.

h) Sei $X_1, X_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit Dichte $f_X(x)$ und endlichem Erwartungswert $\mu = E[X_i]$. Ferner sei $h : \mathbb{R} \to \mathbb{R}$ eine Funktion, so dass der Erwartungswert $E[h(X_1)]$ existiert. Welche der folgenden Aussagen trifft zu?

(i) $\frac{1}{n} \sum_{i=1}^{n} h(X_i) \underset{n\to\infty}{\longrightarrow} \mu$ P-fastsicher.
(ii) $\frac{1}{n} \sum_{i=1}^{n} h(X_i) \underset{n\to\infty}{\longrightarrow} h(\mu)$ P-fastsicher.
(iii) $\frac{1}{n} \sum_{i=1}^{n} h(X_i) \underset{n\to\infty}{\longrightarrow} \int_{-\infty}^{\infty} h(x) f_X(x)\, dx$ P-fastsicher.

i) Seien $X_1, \ldots, X_n$ i.i.d. $Geom(p)$-verteilt mit $p \in (0,1)$. Bestimmen Sie den Maximum-Likelihood-Schätzer T für p.

(i) $T = \frac{1}{\overline{X}_n}$.

(ii) $T = \frac{1}{2-\overline{X}_n}$.

(iii) $T = \frac{1}{1+\overline{X}_n}$.

j) Wir betrachten einen Test (T, K) zum Niveau 2.5% mit Teststatistik T und Verwerfungsbereich K, wobei Nullhypothese und Alternative beide einfach seien. Wir nehmen an, der beobachtete Wert der Teststatistik liege nicht in K. Welche der folgenden Aussagen trifft zu?

(i) Der Testentscheid des Tests (T, K) ist, die Nullhypothese nicht zu verwerfen, und die Wahrscheinlichkeit, dass dieser Entscheid falsch ist, ist höchstens 2.5%.

(ii) Ohne weitere Angaben kann man nicht sagen, mit welcher Wahrscheinlichkeit der Testentscheid falsch ist.

(iii) Die Macht des Tests (T, K) ist immer 97.5%.

2. (10 Punkte)

Das Restaurant *diwine* ist stolz auf sein hochwertiges Weinangebot. Sie sind besonders stolz darauf, dass es bei ihnen nur sehr selten vorkommt, dass ein Kunde einen Wein zurückschickt, nachdem er ihn probiert hat: Im Schnitt wird nur eine Flasche pro Monat von Kunden zurückgeschickt.

a) Sei Z die Anzahl der Flaschen, die in einem Monat zurückgeschickt werden. Welche Verteilung bietet sich sinnvollerweise für Z an? Geben Sie eine kurze Begründung. Bestimmen Sie unter dieser Verteilung für jedes $k \in \mathbb{N}_0$ die Wahrscheinlichkeit, dass k Flaschen im Monat abgelehnt werden.

Im vergangenen Monat wurden aussergewöhnlich viele Flaschen abgelehnt. Eine erste interne Qualitätskontrolle untersucht das Feedback von Kunden dieses Monats. Die Analyse ergibt, dass Weinsorte T mit Wahrscheinlichkeit $p = \frac{1}{6}$ ungeniessbar ist.

b) Sei A die Anzahl der Flaschen der Weinsorte T, die man öffnen muss, bis man den ersten ungeniessbaren Wein findet. Benennen Sie die Verteilung von A und berechnen Sie $P[A = 2]$.
Gehen Sie dabei davon aus, dass die Qualität des Weines in den einzelnen Flaschen unabhängig voneinander ist, und jede Flasche mit der selben Wahrscheinlichkeit von $p = \frac{1}{6}$ ungeniessbar ist.

Eine eingehende Untersuchung ergibt, dass die neueste Lieferung von 12 Flaschen der fraglichen Sorte T bei einem Zwischenhändler falsch gelagert wurde und der Wein mit einer Wahrscheinlichkeit von 80% ungeniessbar geworden ist. Man kann jedoch die neuen Flaschen äusserlich von den übrigen 60 Flaschen derselben Sorte nicht unterscheiden.

c) Berechnen Sie die Wahrscheinlichkeit, dass eine unmittelbar nach der Ankunft der neuen Lieferung zufällig ausgewählte Weinflasche der Sorte T ungeniessbar ist. Verwenden Sie dabei, dass bei allen Flaschen, die nicht aus der neuesten Lieferung stammen, der Wein nur mit Wahrscheinlichkeit 4% ungeniessbar ist.

d) Angenommen, diese zufällig ausgewählte Flasche wird einem Restaurantgast präsentiert und der Gast stellt fest, dass der Wein ungeniessbar ist. Gegeben diese Informationen, wie gross ist die Wahrscheinlichkeit, dass es sich um einen Wein aus der neuesten Lieferung handelt?
Hinweis: Falls Sie **c)** nicht gelöst haben, können Sie die Annahme $p = \frac{1}{6}$ für die Wahrscheinlichkeit, dass eine zufällig ausgewählte Flasche der Weinsorte T ungeniessbar ist, verwenden.

e) In der ersten Woche des vergangenen Monats wurden nur zwei der 72 vorhandenen Flaschen der fraglichen Sorte T den Restaurantgästen präsentiert. Sei X die Anzahl der ungeniessbaren Weine in dieser Stichprobe. Benennen Sie die Verteilung von X und berechnen Sie $P[X = 1]$.

Siehe nächstes Blatt!

3. (10 Punkte)

Die gemeinsame Verteilung von zwei Zufallsvariablen X und Y sei gegeben durch die gemeinsame Dichte

$$f_{X,Y}(x,y) = \begin{cases} c(x^2+y^2), & 0 \le x \le 1,\ 0 \le y \le x, \\ 0, & \text{sonst.} \end{cases}$$

für eine Konstante $c > 0$.

a) Berechnen Sie die Randdichten $f_X(x)$ sowie $f_Y(y)$ in Abhängigkeit von c. Bestimmen Sie zusätzlich die Konstante c, so dass $f_{X,Y}$ eine Dichte ist.

Falls **a)** nicht gelöst wurde, verwenden Sie im folgenden zum Rechnen die (falschen) Annahmen, dass X und Y die Randdichten

$$f_X(x) = 4(x - x^3), \quad 0 \le x \le 1, \quad \text{und} \quad f_Y(y) = \frac{4}{5}(3y^3 + y), \quad 0 \le y \le 1,$$

haben und ihre gemeinsame Dichte wie oben ist mit $c = 2$.

b) Berechnen Sie $E[X]$ und Var$[X]$.

c) Sind X und Y unabhängig? Begründen Sie ihre Antwort mit einem mathematischen Argument.

Alphons möchte für seine Hochzeit Apfelsaft in 2 Liter Beuteln direkt ab dem Bauernhof kaufen. Die Abfüllmaschine des Bauern ist leider etwas ungenau. Im Mittel enthält jeder Beutel zwar 2 Liter Apfelsaft, die Varianz des tatsächlichen Volumens ist jedoch 0.5 [Liter^2]. Wir nehmen an, dass die tatsächlichen Abfüllvolumina für verschiedene Beutel unabhängig sind und die gleiche Verteilung haben. Genauere Angaben über die Verteilung sind jedoch nicht bekannt.

d) Benützen Sie die Chebyshev-Ungleichung, um die Wahrscheinlichkeit von oben abzuschätzen, dass das totale Apfelsaftvolumen in n Beuteln mehr als 10% von seinem Erwartungswert abweicht. Wie viele Beutel muss Alphons mindestens kaufen, damit diese Wahrscheinlichkeit höchstens 10% beträgt?

e) Alphons kauft 50 dieser Beutel. Verwenden Sie den Zentralen Grenzwertsatz, um die Wahrscheinlichkeit, dass diese 50 Beutel mindestens 111 Liter Apfelsaft beinhalten, approximativ zu bestimmen.

4. (10 Punkte)

Ein russischer Milliardär zieht eine Investition in der Schweiz in Betracht. Dazu beobachtet er die Renditen von 25 Schweizer Hedgefonds über ein Quartal. Es wird angenommen, dass die Renditen $X_1, \ldots, X_{25}$ (in Prozent, d.h. $X_1 = 5$, falls die Rendite des 1. Hedgefonds im Quartal 5% beträgt) dieser 25 Hedgefonds unabhängig und normalverteilt seien mit unbekanntem Mittelwert μ und unbekannter Varianz σ^2. Der Investor möchte die Behauptung der Hedgefonds überprüfen, dass die mittlere Rendite pro Quartal exakt 7% sei.

a) Welchen Test eignet sich aufgrund der Modellannahmen, um diese Behauptung zu überprüfen? Ist der Test einseitig oder zweiseitig?

b) Stellen Sie den statistischen Test auf, um die Behauptung der Hedgefonds zu überprüfen. Geben Sie

(i) die Nullhypothese H_0 und die Alternative H_A,

(ii) die Teststatistik und deren Verteilung unter H_0,

(iii) den Verwerfungsbereich zum 10%-Niveau an.

c) Am Ende des Quartals beträgt das empirische Stichprobenmittel $\overline{x}_{25} = 9$ und die empirische Stichprobenvarianz $s^2 = 16$. Was ist der beobachtete Wert der Teststatistik, und was ist der Testentscheid?

d) Die Finanzberater des Milliardärs haben aufgrund der Renditen der Vergangenheit die Varianz der X_i zu $\sigma^2 = 25$ bestimmt. Welchen Test können Sie nun durchführen? Wie lauten die Teststatistik, der Verwerfungsbereich, der beobachtete Wert der Teststatistik und der Testentscheid?

e) Berechnen Sie auf Basis des empirischen Stichprobenmittels $\overline{x}_{25} = 9$ den P-Wert des Tests in **d)**, d.h. das kleinste Niveau, auf dem der Test die Nullhypothese gerade noch verwirft.

Viel Erfolg!