Prof. M. Maathuis | ETH Zürich | Sommer 2019

Prüfung
Stochastik (401-0603-00L)

Nachname: .... Vorname: .... Stud.-Nr: ....

Regeln zum Prüfungsablauf:

- Bitte legen Sie Ihre Legi gut sichtbar auf den Tisch.
- Es dürfen sich nur erlaubte **Hilfsmittel** auf dem Tisch befinden:
  10 hand- oder maschinengeschriebene A4-Seiten, Taschenrechner ohne Kommunikationsmöglichkeiten. Neutrales (nicht fachspezifisches) Deutsch-Englisch Wörterbuch.
- Die Benutzung von Mobiltelefonen ist nicht gestattet. Diese müssen ausgeschaltet sein und dürfen sich nicht auf dem Tisch befinden.
- Schreiben Sie Ihren Namen und Ihre Studierendennummer auf dieses Deckblatt und auf jede Seite, die Sie abgeben.
- Die für die Aufgaben benötigten Tabellen (Verwerfungsbereiche für den Wilcoxon-Test, Tabelle der Normal- und der t-Verteilung) wurden mit der Prüfung ausgeteilt.
- Die Prüfung besteht aus 4 Aufgaben auf 5 Seiten.
- Mit Bleistift, in Rot oder in Grün geschriebene Lösungen ergeben keine Punkte.
- Lösen Sie Aufgabe 1 durch Ankreuzen der korrekten Antwort auf der zusätzlich anghängten Seite. Schreiben Sie bei Aufgaben 2, 3 und 4 alle **Zwischenschritte** und **-rechnungen** sowie **Begründungen** auf.
- Sie haben 2 Stunden Bearbeitungszeit.

Viel Erfolg!

Korrektur:

Aufgabe	1	2	3	4
Punkte				
Kontrolle				

Punktetotal:	
Vollständigkeit:	

Prof. M. Maathuis | ETH Zürich | Sommer 2019

Stochastik

1. (9 Punkte) Beurteilen Sie die folgenden 18 Aussagen auf Ihre Richtigkeit. Die Aussagen sind jeweils in Dreiergruppen eingeteilt. **In einer Gruppe können alle, einige oder keine einzige Aussage richtig sein. Kreuzen Sie Ihre Antworten auf dem zusätzlich angehängten Antwortblatt an**. Pro korrekter Antwort gibt es einen halben Punkt. Es gibt keinen Punkteabzug für falsche Antworten.

a) Es seien A und B Ereignisse mit $\mathbb{P}[A] = 1/2$ und $\mathbb{P}[A \cap B] = 1/4$. Beurteilen Sie folgende Aussagen:

(i) Es gilt $\mathbb{P}[B] \geq 1/4$.
(ii) Es gilt $\mathbb{P}[B] = 1/2$.
(iii) Es gilt $\mathbb{P}[B] \leq 3/4$.

b) Sei $X \sim \text{Bernoulli}(p)$. Beurteilen Sie folgende Aussagen:

(i) $X^3 \sim \text{Bernoulli}(p)$.
(ii) $\text{Var}[X(1-X)] = 0$.
(iii) $2X \sim \text{Binomial}(2, p)$.

c) Sei $X \sim N(\mu, \sigma^2)$ mit $\mu = 2$ und $\sigma = 2$. Sei $\Phi(\cdot)$ die kumulative Verteilungsfunktion der Standardnormalverteilung. Beurteilen Sie folgende Aussagen:

(i) $\mathbb{P}[X = 2] = 0.5$.
(ii) $\mathbb{P}[X \leq c] = \Phi[(c-2)/2]$.
(iii) $\text{Var}(8 - 3X) = 18$.

d) Wir betrachten n i.i.d. Zufallsvariablen $X_1, ..., X_n$ mit $\mathbb{E}[X_i] = \mu$ und $\text{Var}(X_i) = \sigma^2$. Sei $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ und $S_n = \sum_{i=1}^{n} X_i$. Beurteilen Sie folgende Aussagen:

(i) Um die Standardabweichung von $\bar{X}_n$ zu halbieren, braucht man 4 mal so viele Beobachtungen.
(ii) Die Standardabweichung von S_n wächst linear in n.
(iii) $\bar{X}_n$ is eine stetige Zufallsvariable für alle $n = 1, 2, \ldots$.

e) Beurteilen Sie folgende Aussagen zu einem statistischen Test:

(i) $P[\text{Fehler 2. Art}] \leq 1 - \alpha$.
(ii) Um den Verwerfungsbereich eines Tests zu bestimmen braucht man die Verteilung der Teststatistik unter der Nullhypothese und unter der Alternativhypothese.
(iii) Die Macht eines Tests kann man bestimmen aus dem Verwerfungsbereich und der Verteilung der Teststatistik unter der Alternativhypothese.

f) Seien $X_1, \ldots, X_{30}$ i.i.d. Beobachtungen von einer $N(\mu, \sigma^2)$ Verteilung, wobei μ und σ^2 unbekannt sind. Beurteilen Sie folgende Aussagen zu einem T-Test auf Niveau α, wobei $H_0 : \mu = 3$ versus $H_a : \mu \neq 3$ getestet wird:

(i) Der Test verwirft die Nullhypothese genau dann wenn das $(1-\alpha)100\%$ Vertrauensintervall für μ den Wert 3 enthält.

(ii) Je mehr Beobachtungen wir haben, desto kleiner wird die Wahrscheinlichkeit für einen Fehler 1. Art.

(iii) Sei t der realisierte Wert der Teststatistik. Der p-Wert ist gegeben durch $\mathbb{P}[T \geq t]$, wobei T eine t-Verteilung mit 29 Freiheitsgraden hat.

2. **(9 Punkte)** Eine Konditorei verwendet im Herstellungsprozess von Biskuitrollen 3 Maschinen: Maschine A, B und C. Diese führen nacheinander die einzelnen Produktionsschritte durch. Zuerst stellt Maschine A den Biskuitteig her. Im zweiten Produktionsschritt produziert Maschine B die Creme. Im letzten Produktionsschritt stellt Maschine C dann die Rolle inklusive Verzierung fertig.

Es kommt jedoch gelegentlich zu Fehlern im Produktionsprozess. Maschine A produziert mit Wahrscheinlichkeit 5% fehlerhaft, Maschine B mit Wahrscheinlichkeit 3%, Maschine C mit Wahrscheinlichkeit 15%. Dies passiert jeweils unabhängig voneinander.

Aufgrund Ihres guten Rufs kann die Konditorei nur fehlerfreie Biskuitrollen verkaufen.

a) *(1 Pkt)* Wir betrachten die Produktion von einer Biskuitrolle. Wie hoch ist die Wahrscheinlichkeit, dass die Biskuitrolle fehlerfrei produziert wird?

b) *(1 Pkt)* Nach der Produktion sehen wir, dass Maschine C einen Fehler gemacht hat. Wie hoch ist die bedingte Wahrscheinlichkeit, dass auch Fehler gemacht wurden von Maschinen A und/oder B?

Wir nehmen nun an, dass die Produktion einer Biskuitrolle mit 70% Wahrscheinlichkeit fehlerfrei ist.

c) *(1 Pkt)* Wir betrachten die Produktion von 3 Biskuitrollen. Sei X_3 die Anzahl der fehlerfrei produzierten Biskuitrollen. Bestimmen Sie die Verteilung inklusive Parameter von X_3 und geben Sie die Wahrscheinlichkeit an, dass mindestens 2 Biskuitrollen fehlerfrei sind.

d) *(2 Pkt)* Wir haben eine Anfrage zur Herstellung von 715 Biskuitrollen erhalten. Es ist aber fraglich, ob wir diese annehmen können. Innerhalb der Fertigstellungsfrist kann die Konditorei 1000 Biskuitrollen produzieren. Sei X_{1000} die Anzahl der entsprechend fehlerfrei produzierten Biskuitrollen. Bestimmen Sie die Verteilung von X_{1000} inklusive Parameter und berechnen Sie die erwartete Anzahl an fehlerfreien Biskuitrollen. Berechnen Sie mittels Normalapproximation die Wahrscheinlichkeit, dass mindestens 715 Biskuitrollen fehlerfrei produziert wurden.

Die Konditorei produziert auch Torten. In Schritt 1 wird der Teig hergestellt. Der Teig wird dann auf 2 Maschinen aufgeteilt, die in Schritt 2 jeweils eine Torte herstellen: eine vom Typ 1 und eine vom Typ 2. In Schritt 1 passieren zu 4% Fehler. In Schritt 2 passieren bei der Verarbeitung zu Typ 1 zu 6% Fehler. Bei der Verarbeitung zu Typ 2 passieren in Schritt 2 zu 3% Fehler. Die Fehler in den einzelnen Verarbeitungsschritten passieren unabhängig voneinander.

e) *(2 Pkt)* Wir betrachten einen Durchlauf der Produktionskette, also die Produktion von 2 Torten, einer vom Typ 1 und einer vom Typ 2. Für $i \in \{1, 2\}$ sei $Y_i = 0$, wenn die Torte vom Typ i fehlerhaft ist und $Y_i = 1$, wenn die Torte vom Typ i keine Fehler hat. Bestimmen Sie die gemeinsame Wahrscheinlichkeitsfunktion von Y_1 und Y_2, d. h. berechnen Sie die Werte

$$p_{1,1} = P[Y_1 = 1, Y_2 = 1], \qquad p_{1,0} = P[Y_1 = 1, Y_2 = 0],$$
$$p_{0,1} = P[Y_1 = 0, Y_2 = 1], \qquad p_{0,0} = P[Y_1 = 0, Y_2 = 0].$$

f) *(2 Pkt)* Um Folgefehler zu vermeiden, nehmen wir jetzt die folgende gemeinsame Wahrscheinlichkeitsfunktion von Y_1 und Y_2 an:

$$p_{1,1} = P[Y_1 = 1, Y_2 = 1] = 0.82, \qquad p_{1,0} = P[Y_1 = 1, Y_2 = 0] = 0.07,$$
$$p_{0,1} = P[Y_1 = 0, Y_2 = 1] = 0.03, \qquad p_{0,0} = P[Y_1 = 0, Y_2 = 0] = 0.08.$$

Bestimmen Sie die Kovarianz von Y_1 und Y_2.

3. (8 Punkte) Eine Verbraucherschutzorganisation hat die Lebensdauer von Mobiltelefonen eines Herstellers untersucht. Sie hat herausgefunden, dass die Lebensdauer in Jahren exponentialverteilt ist, wobei der Parameter vom Typ des Mobiltelefons abhängt.

Wir betrachten die Modelle R, S und T, mit den entsprechenden Parametern λ_R, λ_S und λ_T.

a) *(1 Pkt)* Geben Sie in Abhängigkeit von λ_R die Wahrscheinlichkeit an, dass ein Mobiltelefon vom Typ R mindestens 2 Jahre hält. Berechnen Sie den Wert, wenn $\lambda_R = 0.15$.

b) *(1.5 Pkt)* Berechnen Sie für $\lambda_R = 0.15$ den Median der Lebensdauer eines Mobiltelefons vom Typ R.

c) *(1 Pkt)* Aufgrund der Reputation des Herstellers können wir annehmen, dass die Mobiltelefone mindestens ein halbes Jahr halten. Wie hoch ist die darauf bedingte Wahrscheinlichkeit, dass ein Mobiltelefon vom Typ R innerhalb von 2 Jahren nicht mehr funktioniert? Benutzen sie auch hier den Parameter $\lambda_R = 0.15$.

d) *(3 Pkt)* Wir haben einen ausführlichen Testbericht zu einer mehrjährigen Studie über das Modell T gelesen. Es wurden bei unabhängiger Nutzung von 10 Mobiltelefonen des Typs T folgende Lebensdauern beobachtet:

i	1	2	3	4	5	6	7	8	9	10
x_i	9.42	4.31	0.61	0.50	4.38	8.51	10.28	2.07	5.61	1.84

mit Mittelwert $\bar{x} = 4.75$ und empirischer Varianz $s^2 = 13.17$. Leiten Sie für λ_T die folgenden zwei Momentenschätzer her: $\hat{\lambda}_T^{(1)}$ basierend auf dem ersten Moment und $\hat{\lambda}_T^{(2)}$ basierend auf dem zweiten Moment. Berechnen Sie die Werte für die beiden Schätzer.

e) *(1.5 Pkt)* Wir haben von den 3 Modellen je ein Telefon gekauft. Die Telefone wurden jeweils gleich und unabhängig genutzt. Sei U die Zeit (in Jahren), zu der das erste der 3 Telefone kaputt geht. Bestimmen Sie $P[U \leq u]$ für $u \geq 0$ in Abhängigkeit von λ_R, λ_S und λ_T. Benennen Sie die Verteilung von U inklusive Parameter.

4. (10 Punkte) Um den Kraftstoffverbrauch eines Fahrzeuges im Stadtverkehr zu messen, haben wir verschiedene Teststrecken zur Verfügung. Wir möchten damit die Herstellerangabe zum mittleren Verbrauch überprüfen.

In sieben unabhängigen Testversuchen mit dem selben Fahrzeugtyp erhalten wir den folgenden Kraftstoffverbrauch in Liter pro 100km:

Messung	1	2	3	4	5	6	7
Wert	6.61	6.02	6.79	7.23	7.28	7.15	7.14

Aus den Daten haben wir schon den Mittelwert $\bar{x} = 6.89$ und die Standardabweichung $s_X = 0.4558$ berechnet.

a) Sei μ_X der wahre mittlere Verbrauch dieses Fahrzeugtyps. Laut Angaben des Herstellers gilt $\mu_X = 6.49$. Wir möchten diese Aussage überprüfen. Wir können davon ausgehen, dass die Verteilung stetig und symmetrisch ist und wollen einen Wilcoxontest auf Signifikanzniveau 5% durchführen.

(i) *(1 Pkt)* Geben Sie einen Grund an, warum der Wilcoxontest in dieser Situation besser geeignet ist als der T-Test. Geben Sie auch einen Grund an, warum der Wilcoxontest in dieser Situation besser geeignet ist als der Vorzeichentest.

(ii) *(0.5 Pkt)* Formulieren Sie Null- und Alternativhypothese.

(iii) *(2 Pkt)* Berechnen Sie den realisierten Wert der Wilcoxon-Teststatistik.

(iv) *(1 Pkt)* Bestimmen Sie den Verwerfungsbereich. Können wir als Konsequenz aus dem Testergebnis die Herstellerangabe zum mittleren Verbrauch verwerfen?

b) Nach einem Softwareupdate der Motorsteuerung durch den Hersteller, führen wir neue Tests auf neuen Teststrecken durch. Diese liefern nun die folgenden Messwerte für den Kraftstoffverbrauch in Liter pro 100km:

Messung	1	2	3	4	5	6	7	8
Wert	7.56	5.85	7.33	6.99	7.15	7.07	7.45	7.21

mit Mittelwert $\bar{y} = 7.075$ und Standardabweichung $s_Y = 0.532$. Wir haben nun zusätzlich die Information erhalten, dass die Daten sowohl vor als auch nach dem Update normalverteilt mit Standardabweichung $\sigma = 0.8$ sind. Der mittlere Verbrauch nach dem Update sei μ_Y. Laut Hersteller hat sich der mittlere Verbrauch durch das Update nicht geändert. Wir bezweifeln das jedoch und wollen dies testen. Wir wollen einen ungepaarten Zweistichproben-Z-Test durchführen und auch hier ein Signifikanzniveau von 5% verwenden.

(i) *(0.5 Pkt)* Begründen Sie, warum wir einen ungepaarten und keinen gepaarten Test durchführen.

(ii) *(0.5 Pkt)* Formulieren Sie Null- und Alternativhypothese.

(iii) *(1.5 Pkt)* Formulieren Sie die Teststatistik und berechnen Sie den realisierten Wert.

(iv) *(1 Pkt)* Bestimmen Sie den Verwerfungsbereich. Können wir als Ergebnis des Tests die Herstelleraussage zum mittleren Verbrauch zurückweisen?

(v) *(2 Pkt)* Berechnen Sie die Macht des Tests, wenn der Mittelwert nach dem Update in Wahrheit um 0.5 höher ist, d. h. $\mu_Y = \mu_X + 0.5$.