Prof. Dr. P. Embrechts | ETH Zürich | Winter 2009

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

Schreiben Sie für Aufgabe 2-4 stets alle Zwischenschritte und -rechnungen sowie Begründungen auf. Aufgabe 1 ist eine Multiple Choice Aufgabe (keine Begründungen notwendig).

Die für die Aufgaben benötigten Tabellen (Normalverteilung, Binomialverteilung) wurden mit der Prüfung ausgeteilt.

1. (10 Punkte) Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Es gibt pro richtig beantwortete Frage 1 Punkt, und pro falsche Antwort 1/2 Punkt Abzug. Minimal erhält man für die gesamte Aufgabe 0 Punkte. **Bitte benützen Sie das beiliegende Antwortblatt.**

a) Seien A und B Ereignisse mit $P[A] = 0.5$ und $P[B] = 1$. Dann gilt:

1) $P[B|A] = 0.5$
2) $P[B^c|A] = 0.5$
3) A und B sind unabhängig.

b) Sei X eine Zufallsvariable, welche nur Werte in $\{1, 2, 3, \ldots\}$ annehmen kann, und es gelte $P[X = k] = 1/2^k$ $(k = 1, 2, 3, \ldots)$. Für die Verteilungsfunktion F_X gilt:

1) $F_X(3) = 1/8$ 2) $F_X(3) = 7/8$ 3) F_X ist stetig.

c) Die nicht-negative Zufallsvariable X habe die Verteilungsfunktion F_X. Sei $Y := X^2$, dann gilt für die Verteilungsfunktion F_Y und $y \geq 0$, dass

1) $F_Y(y) = F(y)^2$ 2) $F_Y(y) = F(y^2)$ 3) $F_Y(y) = F(\sqrt{y})$.

d) Gegeben sind zwei Zufallsvariablen X und Y mit Verteilungsfunktionen F_X resp. F_Y. Dann gilt:

1) Die gemeinsame Verteilung von (X, Y) bestimmt F_X und F_Y eindeutig,
2) F_X und F_Y bestimmen (zusammen) eindeutig die gemeinsame Verteilung von (X, Y),
3) keine dieser zwei Aussagen ist richtig.

Bitte wenden!

e) Die Zufallsvariable X habe die Wahrscheinlichkeitsdichte f_X und erfülle $P[X \geq 0] = 1$. Dann kann der Erwartungswert $E[\sqrt{X}]$ berechnet werden durch die Formel

1) $\int_0^\infty \sqrt{f_X(x)}\,dx$ 2) $\int_0^\infty f_X(\sqrt{x})\,dx$ 3) $\int_0^\infty \sqrt{x} f_X(x)\,dx$.

f) Seien X und Y Zufallsvariablen mit $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. Dann folgt:

1) X und Y sind unabhängig,
2) X und Y sind unkorreliert,
3) X oder Y ist eine Konstante.

g) Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte (i.i.d.) Zufallsvariablen mit $E[X_i] = 2$ und $\mathrm{Var}(X_i) = 3$. Setze $S_n := \sum_{i=1}^n X_i$ für $n = 1, 2, \ldots$. Für grosse n gilt näherungsweise

1) $S_n \sim \mathcal{N}(2n, 3n)$ 2) $S_n \sim \mathcal{N}(2, 9/n)$ 3) $S_n \sim \mathcal{N}(2n, 3n^2)$.

h) Wir führen das gleiche physikalische Experiment mehrmals unabhängig aus und messen jedes mal einen Wert $x_i \in \mathbb{N}$. Wir stellen uns vor, dass dies Realisierungen von i.i.d. Zufallsvariablen sind. Wir stellen unsere Ergebnisse in einem Histogramm dar und stellen fest, dass ein Balken sehr hoch ist und die anderen sehr klein. Es folgt daraus, dass die Zufallsvariablen

1) einen grossen Erwartungswert haben,
2) eine grosse Varianz haben,
3) eine kleine Varianz haben.

i) Für den p-Wert eines beliebigen Tests gilt:

1) der p-Wert hängt vom Signifikanzniveau α ab,
2) der p-Wert gibt die Wahrscheinlichkeit an, dass die Nullhypothese verworfen wird,
3) weder 1) noch 2) ist richtig.

j) Wir führen das gleiche medizinische Experiment mehrmals unabhängig aus und messen jedes mal einen Wert $x_i \in \mathbb{R}$. Wir stellen uns vor, dass dies Realisierungen von i.i.d. normalverteilten Zufallsvariablen mit bekannter Varianz σ^2 sind. Mit dem z-Test möchten wir testen, ob der Erwartungswert $H_0 : \mu = 0$ ist, wobei $H_A : \mu \neq 0$. Indem wir das Experiment häufiger wiederholen (aber den Test-Typ nicht verändern), können wir

1) die Wahrscheinlichkeit eines Fehlers 1. Art verringern,
2) die Wahrscheinlichkeit eines Fehlers 2. Art verringern,
3) das Vertrauensintervall vergrössern.

Siehe nächstes Blatt!

2. (10 Punkte) Der Fernsehsender SkiTV möchte ein Skirennen mit 81 Skifahrern live übertragen. Der i-te Skifahrer ($i = 1, \ldots, 81$) erreicht nach einer Zeit T_i (in Minuten) das Ziel. Sobald ein Skifahrer im Ziel ist, kann der nächste ohne Zeitverzögerung direkt starten.
Die Zufallsvariablen T_i seien iid mit Dichte f_T, wobei

$$f_T(x) = \begin{cases} \lambda e^{-\lambda(x-1)}, & x \geq 1, \\ 0, & \text{sonst}, \end{cases}$$

und $\lambda = 3$.

a) Bestimme den Erwartungswert $E[T_i]$.
Hinweis: Falls $S \sim \mathrm{Exp}(\lambda)$, so ist $E[S] = 1/\lambda$ und $\mathrm{Var}(S) = 1/\lambda^2$.

b) Bestimme die Varianz $\mathrm{Var}(T_i)$.

c) Wie gross ist die Wahrscheinlichkeit, dass ein Skifahrer mehr als 70 Sekunden die Piste besetzt?

d) Aufgrund langjähriger Erfahrung hat der TV-Sender genauere Informationen, nämlich dass die erwartete Dauer bis ein Skifahrer im Ziel ist $\mu = 13/9$ beträgt, mit einer Varianz von $\sigma^2 = 1/9$. Da der TV-Sender das ganze Rennen live übertragen will, fragt er sich wie lange das Rennen dauern wird. Man möchte das Rennen mit einer Wahrscheinlichkeit von 97.7% vollständig übertragen können. Wie viele Stunden/Minuten muss der Sender in seinem Sendeplan für das Skirennen reservieren? Benutze den zentralen Grenzwertsatz um dieses Problem zu lösen.

Bitte wenden!

3. (10 Punkte) Eine Methode um Granit (G) und Basalt (B) zu unterscheiden, besteht darin einen Teil des infraroten Spektralbereiches der auf der Felsoberflaeche reflektierten Sonnenenergie zu untersuchen. Wir bezeichnen mit R_1, R_2 und R_3 die gemessenen Intensitäten drei verschiedener Wellenlängen. Typischerweise erhält man für Granit $R_1 < R_2 < R_3$ und für Basalt $R_3 < R_1 < R_2$. Wenn Messungen aus der Distanz (vom Flugzeug aus) gemacht werden können verschiedene Anordnungen der R_i's auftreten, egal ob der Fels aus Granit oder Basalt ist. Flüge über Regionen mit bekannter Felsbeschaffenheit haben die folgenden Werte ergeben:

	Granit (G)	**Basalt (B)**
$R_1 < R_2 < R_3$	60 %	10%
$R_1 < R_3 < R_2$	25 %	20%
$R_3 < R_1 < R_2$	15 %	70%

In der Region Paul-Valley weiss man, dass für einen zufällig ausgewählten Fels F gilt

$$P[F = G] = 0.25 \quad \text{und} \quad P[F = B] = 0.75.$$

a) Zeigen Sie, dass gilt

$$P\big[F = G\big|R_1 < R_2 < R_3\big] > P\big[F = B\big|R_1 < R_2 < R_3\big].$$

Falls Ihre Messungen $R_1 < R_2 < R_3$ ergeben, würden Sie den Fels also als Granit oder Basalt klassifizieren?

b) Analoge Rechnungen wie in **a)** liefern das folgende Klassifizierungsschema:

Beobachtung	Klassifizierung
$R_1 < R_2 < R_3$	Granit
$R_1 < R_3 < R_2$	Basalt
$R_3 < R_1 < R_2$	Basalt

Bestimmen Sie unter Verwendung dieses Schemas die Wahrscheinlichkeit, dass ein zufällig ausgewählter Fels in Paul-Valley falsch klassifiziert wird.

c) Angenommen $P[F = G] = p$ (statt 0.25) und $P[F = B] = 1 - p$ (statt 0.75). Bestimmen Sie alle Werte für p, so dass jeder Fels als Granit klassifiziert wird. *Hinweis:* Finden Sie alle $p \in [0, 1]$, so dass $P\big[F = G\big|R_1 < R_2 < R_3\big] > 0.5$, $P\big[F = G\big|R_1 < R_3 < R_2\big] > 0.5$ und $P\big[F = G\big|R_3 < R_1 < R_2\big] > 0.5$. Um Ihre Rechnungen übersichtlich zu halten können Sie zum Beispiel $P\big[F = G\big|R_1 < R_2 < R_3\big]$ mit $P_{G|123}$ usw. abkürzen.

Siehe nächstes Blatt!

4. (10 Punkte) Die Firma Pharma AG vertreibt Schmerzmittel in Pulverform und Tablettenform. Bei der Einnahme des Pulvers klagen 40% der Patienten über Nebenwirkungen. Sie möchten nun testen: Wenn Tabletten statt Pulver eingenommen werden, ändert sich dann der Anteil der Patienten mit Nebenwirkungen signifikant?

Dazu werden zufällig 50 Patienten ausgewählt und es sei X die Anzahl Patienten unter diesen 50, welche bei der Einnahme von Tabletten *keine* Nebenwirkungen haben. Wir nehmen an, dass $X \sim \text{Bin}(50, p)$, wobei p die Erfolgswahrscheinlichkeit im Einzelfall ist ("Erfolg" = keine Nebenwirkungen).

a) Obige Studie ergab das Folgende: Unter den 50 Patienten befanden sich 35, bei welchen durch die Tabletten keine Nebenwirkungen hervorgerufen wurden. Ist dies signifikant oder purer Zufall? Führen Sie dazu einen statistischen Test durch und gehen Sie dabei wie folgt vor:

i) Formulieren Sie eine geeignete Null- und Alternativ Hypothese.

ii) Ist der Test ein- oder zweiseitig durchzuführen? Begründen Sie kurz!

iii) Berechnen Sie den Verwerfungsbereich V des Tests für ein Signifikanzniveau von 5%.

iv) Wie enscheidet der Test?

b) Angenommen, der wahre Anteil der unter Nebenwirkungen leidenden Patienten bei der Einnahme der Tabletten sei 20%. Wie gross ist in diesem Fall die Wahrscheinlichkeit eines Fehlers 2. Art? *Falls Sie Teilaufgabe a) nicht lösen konnten, verwenden Sie einen einseitigen Test mit Verwerfungsbereich* $V = \{k \in \mathbb{N}_0 : k \geq 34\}$.