Wahrscheinlichkeit & Statistik

Serie 9

1. In Aufgabe 3 von Serie 8 haben wir folgendes Modell betrachtet: Die Zufallsvariable X messe die Wasserhöhe in cm über der kritischen Marke von 140 cm über Normalniveau im Zürichsee. Zur Modellierung von X können wir eine sogenannte verallgemeinerte Pareto-Verteilung mit Dichte

$$f_X(x;\vartheta) = \begin{cases} \frac{1}{\vartheta}(1+x)^{-(1+\frac{1}{\vartheta})} & \text{falls } x > 0, \\ 0 & \text{falls } x \leq 0 \end{cases}$$

verwenden. Dabei ist $\vartheta > 0$ ein unbekannter Parameter, der auf Basis von Daten $x_1, \dots, x_n$ geschätzt werden soll; diese Daten werden wie üblich als Realisierungen von Zufallsvariablen $X_1, \dots, X_n$ aufgefasst, die für jede Wahl des Parameters ϑ unter P_ϑ i.i.d. sind mit Dichte $f_X(x;\vartheta)$. Zeigen Sie, dass der Schätzer

$$T^{(n)} = \sum_{i=1}^{n} \frac{\log(1+X_i)}{n}$$

der Maximum-Likelihood-Schätzer für ϑ ist.

2. Die Zufallsvariablen $X_1, \dots, X_n$ seien unabhängig und je $\mathcal{N}(\mu, \sigma^2)$-verteilt unter P_ϑ, wobei $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ ein unbekannter Parameter ist. Als Schätzer für σ^2 betrachte man

$$T^{(n)}(c) := c \sum_{i=1}^{n} (X_i - \overline{X}_n)^2,$$

wobei $\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$ und $c > 0$ ist.

a) Für welches $c > 0$ wird der mittlere quadratische Schätzfehler $E_\vartheta\left[\left(T^{(n)}(c) - \sigma^2\right)^2\right]$ minimiert?

Hinweis: Benutzen Sie, dass $\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$-verteilt ist unter P_ϑ.

b) Ist die resultierende Folge von Schätzern konsistent?

Siehe nächste Seite!

3. Eine Tankstelle veranschlagt für einen Ölwechsel mindestens $\vartheta_1 > 0$ Minuten. Die tatsächlich benötigte Zeit X variiert natürlich im Bereich $X \geq \vartheta_1$ und ist von Kunde zu Kunde verschieden. Man kann jedoch annehmen, dass diese Zeit durch eine exponentialverteilte Zufallsvariable gut modelliert wird. Die Zufallsvariable X besitze somit unter P_ϑ die Dichte

$$f(x;\vartheta) = \begin{cases} \vartheta_2 e^{\vartheta_1 \vartheta_2 - \vartheta_2 x} & \text{falls } x \geq \vartheta_1, \\ 0 & \text{sonst,} \end{cases}$$

d.h. $X = \vartheta_1 + Z$ mit $Z \sim Exp(\vartheta_2)$ unter P_ϑ. Um $\vartheta = (\vartheta_1, \vartheta_2)$ schätzen zu können, wurde von 10 zufällig ausgewählten Kunden die für den Ölwechsel benötigte Arbeitszeit in Minuten notiert:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
4.2	3.1	3.6	4.5	5.1	7.6	4.4	3.5	3.8	4.3

Bestimmen Sie den Maximum-Likelihood-Schätzer und den entsprechenden Schätzwert für den unbekannten Parameter ϑ.

4. Während der Vorbereitungszeit auf die Prüfungssession kommunizieren Sarah und Benno per E-mail, um auftretende Probleme zu bereinigen. Aufgrund der Tageszeiten, zu denen Benno Meldungen abschickt, hat Sarah den Verdacht, dass Benno mehr arbeitet als sie. Aus den Tageszeiten der letzten zehn Meldungen (umgerechnet in Stunden),

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
10.55	14.9	11.2	18.85	9.75	11.5	16.1	14.4	9.2	12.95

will Sarah Bennos Arbeitszeiten schätzen. Zu diesem Zweck nimmt sie an, dass Benno jeden Tag zur Zeit a anfängt zu lernen und zur Zeit b aufhört, und dass die Zeiten, zu denen die E-mails abgeschickt werden, unabhängig und alle uniform verteilt sind auf dem Intervall $[a, b]$. Bestimmen Sie für die Parameter a und b jeweils Schätzer und die entsprechenden Schätzwerte

a) mit der Maximum-Likelihood-Methode.
(**Vorsicht:** Die Likelihood-Funktion ist nicht differenzierbar. Das Maximum der Likelihood-Funktion muss daher auf eine andere Art bestimmt werden.)

b) mit der Momenten-Methode.

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 8

1. Sei Y eine χ^2_n-verteilte Zufallsvariable mit $n \in \mathbb{N}$, das heisst

$$Y = \sum_{i=1}^{n} X_i^2,$$

wobei $X_1, \dots, X_n$ unabhängig und $\mathcal{N}(0,1)$-verteilt sind.

a) Zeigen Sie, dass

$$E[Y] = n \quad \text{und} \quad \mathrm{Var}[Y] = 2n$$

gilt.

b) Geben Sie mit Hilfe der Chebyshev-Ungleichung eine untere Schranke für die Wahrscheinlichkeit

$$\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right].$$

Berechnen Sie die Schranke für $n = 12$.

c) Berechnen Sie für $n = 12$ eine Annäherung für die obige Wahrscheinlichkeit mit Hilfe des zentralen Grenzwertsatzes.

2. Seien $X_1, \dots, X_n$ unabhängige und auf $[\vartheta - 1, \vartheta]$ gleichverteilte Zufallsvariablen unter P_ϑ, wobei $\vartheta \in \mathbb{R}$ ein unbekannter Parameter ist. Für ϑ bieten sich

$$T^{(n)} = \frac{1}{n} \sum_{i=1}^{n} (X_i + 1/2) \quad \text{und} \quad \hat{T}^{(n)} = \max\{X_1, \dots, X_n\}$$

als Schätzer an. Wir untersuchen in dieser Aufgabe, welche Eigenschaften diese beiden Schätzer besitzen.

a) Untersuchen Sie, ob die Schätzer erwartungstreu sind.

b) Berechnen Sie $\mathrm{Var}_\vartheta[T^{(n)}]$ und $\mathrm{Var}_\vartheta[\hat{T}^{(n)}]$.

c) Sind die Schätzer konsistent?

Siehe nächste Seite!

3. Wir betrachten Pegelstände bei Hochwasser im Zürichsee. Hochwasser bedeute dabei, dass der Pegelstand die kritische Marke von 140 cm über Normalniveau überschreitet. Die Zufallsvariable X messe die Wasserhöhe in cm über der kritischen Marke. Zur Modellierung von X können wir eine sogenannte verallgemeinerte Pareto-Verteilung mit Dichte

$$f_X(x;\vartheta) = \begin{cases} \frac{1}{\vartheta}(1+x)^{-(1+\frac{1}{\vartheta})} & \text{falls } x > 0, \\ 0 & \text{falls } x \leq 0 \end{cases}$$

verwenden. Dabei ist $\vartheta > 0$ ein unbekannter Parameter, der auf Basis von Daten $x_1, \ldots, x_n$ geschätzt werden soll; diese Daten werden wie üblich als Realisierungen von Zufallsvariablen $X_1, \ldots, X_n$ aufgefasst, die für jede Wahl des Parameters ϑ unter P_ϑ i.i.d. sind mit Dichte $f_X(x;\vartheta)$. Als Schätzer für ϑ verwenden wir

$$T^{(n)} = \sum_{i=1}^{n} \frac{\log(1+X_i)}{n}.$$

a) Berechnen Sie den Erwartungswert und die Varianz von $T^{(n)}$ in jedem Modell P_ϑ. *Hinweis:* Benutzen Sie, dass $Y_i := \log(1+X_i) \sim Exp(\frac{1}{\vartheta})$ ist, d.h. Y_i hat unter P_ϑ die Dichte $f_Y(y) = \frac{1}{\vartheta}e^{-\frac{1}{\vartheta}y}I_{\{y\geq 0\}}$.

b) Ist $T^{(n)}$ ein erwartungstreuer Schätzer für ϑ?

c) Ist die Folge von Schätzern $T^{(n)}, n \in \mathbb{N}$, konsistent?

4. Wir betrachten eine stetige Verteilung mit Dichte

$$f(x) = \begin{cases} \frac{\vartheta}{x^{\vartheta+1}} & x \geq 1\,, \\ 0 & x < 1\,, \end{cases}$$

wobei $\vartheta > 0$ ein unbekannter Parameter ist. Wir wollen den Parameter ϑ mit Hilfe eines Datensatzes schätzen.

a) Sei $X_1, \ldots, X_n$ eine Stichprobe von unabhängigen Zufallsvariablen, welche alle die Dichte f besitzen. Bestimmen Sie die Likelihood- und log-Likelihood-Funktion.

b) Bestimmen Sie den zugehörigen Maximum-Likelihood-Schätzer für ϑ. Schreiben Sie zuerst die allgemeine Formel für n Beobachtungen hin und berechnen Sie den Schätzwert dann für die folgende konkrete Stichprobe:

x_1	x_2	x_3	x_4	x_5
12.0	4.0	6.9	27.9	15.4

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 10

1. Da jeder Computer nur endlich viele Zahlen darstellen kann, ist das Runden bei numerischen Auswertungen prinzipiell nicht zu vermeiden. Der Einfachheit halber werde jede reelle Zahl auf die nächstgelegene ganze Zahl gerundet, wobei der begangene Fehler durch eine Zufallsvariable $R \sim \mathcal{U}(-1/2, 1/2)$ beschrieben wird. Für verschiedene zu addierende Zahlen seien diese Fehler stochastisch unabhängig. Addiert man 1200 Zahlen, so könnten sich die Rundungsfehler $R_1, \dots, R_{1200}$ theoretisch zu ± 600 aufsummieren. Zeigen Sie: Es gilt approximativ

$$P\left[\left|\sum_{i=1}^{1200} R_i\right| \leq 20\right] \approx 0.9534.$$

2. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit $E[X_1^2] < \infty$ und $\sigma^2 := \operatorname{Var}[X_1]$. Die Stichprobenvarianz von $X_1, \dots, X_n$ ist definiert als

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

wobei $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$. Zeigen Sie, dass

$$\lim_{n\to\infty} S_n^2 = \sigma^2 \quad P\text{-fastsicher.}$$

3. Die Vereinigten Staaten wollen das derzeitige Niveau ihrer N strategischen Ölreserven ermitteln. Sie modellieren das Niveau einer einzelnen Ölreserve mit der Zufallsvariable X. Sie benutzen für X die Dichte

$$f_X(x; \vartheta_1, \vartheta_2) = c x^{-\vartheta_1} I_{\{x \geq \vartheta_2\}}\,,$$

wobei $\vartheta = (\vartheta_1, \vartheta_2)$ ein unbekannter Parameter ist mit $\vartheta_2 > 0$ und $\vartheta_1 > 1$.

a) Bestimmen Sie die Konstante c so, dass $f_X(x; \vartheta_1, \vartheta_2)$ zu einer Dichte wird.

Siehe nächste Seite!

b) Zeigen Sie, dass

$$E_\vartheta[X^s] = \frac{(\vartheta_1 - 1)\vartheta_2^s}{\vartheta_1 - 1 - s}$$

unter P_ϑ, wenn $\vartheta_1 > s+1$. Berechnen Sie dann $E_\vartheta[X]$ für $\vartheta_1 > 2$ und $\mathrm{Var}_\vartheta[X]$ für $\vartheta_1 > 3$ unter P_ϑ.

Nehmen Sie an, dass der Parameter ϑ_2 bekannt ist. Sie wollen den Parameter ϑ_1 aus einer Stichprobe mit n Beobachtungen schätzen.

c) Bestimmen Sie den Momentenschätzer und den Maximum-Likelihood-Schätzer für ϑ_1. Schreiben Sie die allgemeinen Formeln für n Beobachtungen hin.

Für $i = 1, \ldots, N$ sei X_i jene Zufallsvariable, welche das Niveau der i-ten Ölreserve repräsentiert. Sei $S_N = \sum_{i=1}^N X_i$ die Zufallsvariable, die das Gesamtniveau der N Ölreserven darstellt.

d) Wie gross ist etwa die Wahrscheinlichkeit, dass das Gesamtniveau der N strategischen Ölreserven eine Mindestgrenze δ unterschreitet? Treffen Sie dabei geeignete Annahmen, und beschreiben Sie diese. Drücken Sie diese Wahrscheinlichkeit mit ϑ_1, ϑ_2, N, δ und der kumulativen Verteilungsfunktion Φ der Standardnormalverteilung aus.

4. Sie haben einen Würfel mit sechs Seiten und wollen testen, ob der Würfel gezinkt ist und eher auf der Sechs landet. Also machen Sie ein Experiment, in dem Sie zehn Mal würfeln und jeweils die beobachtete Augenzahl notieren. Gehen Sie davon aus, dass alle Würfe unabhängig voneinander sind und die Wahrscheinlichkeit, eine 1, 2, 3, 4 oder 5 zu würfeln, gleich ist. Wir modellieren die Ausgänge der Würfe als eine Stichprobe $X_1, \ldots, X_{10}$, wobei $X_i = 1$ bedeutet, dass der i-te Wurf eine Sechs ist, und $X_i = 0$ sonst. Sie erhalten folgende Resultate:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	0	1	0	1	1	0	1	0	0

a) Bestimmen Sie ein geeignetes Modell $(P_\vartheta)_{\vartheta\in\Theta}$, d.h. einen Parameterraum und die Verteilungen von $X_1, \ldots, X_{10}$ unter jedem P_ϑ.

b) Bestimmen Sie eine geeignete Hypothese H_0 und Alternative H_A.

c) Sei $T = \sum_{i=1}^{10} X_i$ Ihre Teststatistik. Welcher Verteilung folgt T?

d) Sei $K = (4, 10]$ Ihr Verwerfungsbereich. Berechnen Sie die Wahrscheinlichkeit für einen Fehler 1. Art.

e) Beschreiben Sie Ihren Testentscheid.

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 11

1. Seien $X_1, \dots, X_{10}$ unabhängig und $\mathcal{P}(\vartheta)$-verteilt, wobei $\vartheta \in \Theta = (0, \infty)$ ein unbekannter Parameter ist. Geben Sie einen nichtrandomisierten Test (T, K) für die Hypothese $\Theta_0 = \{1/2\}$ und Alternative $\Theta_A = \{2\}$ zum Niveau $\alpha = 0.2$ an, welcher zudem folgende Eigenschaft erfüllt: Jeder andere Test (T', K') hat eine grössere Wahrscheinlichkeit für einen Fehler 1. Art oder 2. Art.

2. Sie erhalten den Auftrag, die Anzahl der Ausfälle eines Systems zu überprüfen. Nach Angabe des Herstellers sollen im Mittel 0.5 Ausfälle pro Stunde eintreten. Die Analyse nach 6 Betriebsstunden hat nun folgendes Ergebnis gebracht:

Betriebsstunde i	1	2	3	4	5	6
Anzahl Ausfälle x_i	1	0	1	1	2	1

Aufgrund der hohen Anzahl von Ausfällen haben wir den Verdacht, dass im Mittel 1 Ausfall pro Stunde passiert. Gehen Sie davon aus, dass die Anzahl der Ausfälle pro Betriebsstunde Poisson-verteilt ist und dass die Ausfälle in verschiedenen Betriebsstunden unabhängig voneinander sind. Prüfen Sie anhand eines einseitigen Tests auf dem Niveau 2.5%, ob tatsächlich im Mittel 0.5 Ausfälle pro Stunde angenommen werden können. Geben Sie

a) das Modell,

b) die Nullhypothese und die Alternative,

c) die Teststatistik,

d) die Verteilung der Teststatistik unter der Nullhypothese,

e) den Verwerfungsbereich,

f) den beobachteten Wert der Teststatistik, sowie

g) den Testentscheid an.

Siehe nächste Seite!

3. Seien $X_1, \ldots, X_{12}$ unabhängig und je $\mathcal{N}(\mu, \sigma^2)$-verteilt unter P_ϑ, wobei $\vartheta = (\mu, \sigma^2)$ ein unbekannter Parameter ist. Wir wissen, dass $\sigma = 0.0499$ gilt und haben folgende Daten für die Stichprobe gegeben:

x_1	x_2	x_3	x_4	x_5	x_6
1.00781	1.00646	1.00801	1.00833	1.00738	1.00687
x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1.00783	1.00936	1.00564	1.00543	1.00794	1.01060

Wir testen die Hypothese $H_0\colon \mu = \mu_0 = 1.0085$ gegen die Alternative $H_1 : \mu \neq \mu_0$.

a) Bestimmen sie a und b so, dass die Teststatistik $T := \frac{\sum_{i=1}^{12} X_i + b}{a} \sim \mathcal{N}(0,1)$-verteilt ist unter P_{μ_0}.

b) Wählen Sie $K := (-\infty, -c_{\neq}) \cup (c_{\neq}, \infty)$ als Ihren Verwerfungsbereich für ein zu bestimmendes $c_{\neq} \geq 0$. Testen Sie H_0 gegen H_A auf dem 5%-Niveau.

c) Berechnen Sie die Macht des oben durchgeführten Tests an der Stelle $\mu_A = 1.008$.

4. Ein Grossverteiler kauft bei einem regionalen Händler 2 Tonnen Galia-Melonen ein. Der Händler garantiert dem Grossverteiler, dass maximal 4% der Melonen faul seien. Zur Kontrolle entnimmt der Grossverteiler zufällig 50 Melonen, um zu untersuchen, ob die Aussage des Händlers stimmt.

a) Welche Verteilung eignet sich, um die Anzahl fauler Melonen unter den 50 untersuchten Melonen zu beschreiben? Welche Annahmen werden mit diesem Modell implizit gemacht?

b) Angenommen, unter den 50 Melonen befinden sich 4 faule. Hat der Händler über die Qualität seiner Melonen gemogelt? Formulieren Sie eine geeignete Nullhypothese und eine Alternative. Bestimmen Sie für das Signifikanzniveau von 5% den Verwerfungsbereich. Wie entscheidet der Test? Was ist die Interpretation des Ergebnisses?

c) Wie gross ist die Macht dieses Tests für eine Hypothese $\vartheta_0 = 0.04$ gegen eine Alternative $\vartheta_A = 0.1$? Welche Konsequenzen sind daraus zu ziehen?

Wahrscheinlichkeit & Statistik

Serie 13

1. Eine Klimaanlage schafft es, die Raumtemperatur bis auf eine Standardabweichung von einem halben Grad Celsius konstant zu halten. Die angestrebte Raumtemperatur beträgt 20.00 Grad Celsius. An zehn aufeinanderfolgenden Tagen wurden die folgenden Temperaturen gemessen:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
20.71	19.76	20.56	21.39	21.00	19.67	20.92	20.31	20.39	20.72

Wir nehmen an, dass die gemessenen Temperaturen unabhängig voneinander und identisch normalverteilt sind.

a) Führen Sie einen geeigneten Test auf dem 5%-Niveau durch, um zu beurteilen, ob die Klimaanlage wirklich auf den Sollwert von 20.00 Grad geeicht ist.

b) Berechnen Sie den realisierten p-Wert.

Falls Ihnen die Tabellen im Skript nicht ausreichen, können Sie diesen Rechner für die t-Verteilung und diese Tabelle für die Standardnormalverteilung benutzen.

Kennzahlen: $\overline{x}_{10} = 20.543$, $s^2 = 2.5808$.

2. Ein Weinhändler behauptet, dass die von ihm gefüllten Weinflaschen mindestens 70 Zentiliter enthalten. Ein skeptischer Konsument will diese Behauptung überprüfen. Deshalb kauft er 12 Weinflaschen und misst ihren Inhalt. Er findet die folgenden Werte (in Zentiliter):

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
71	69	67	68	73	72	71	71	68	72	69	72

Die Standardabweichung der Abfüllung sei im Voraus bekannt und betrage $\sigma = 1.5$ Zentiliter.

Siehe nächste Seite!

a) Führen Sie einen geeigneten Test durch. Spezifizieren Sie Ihre Annahmen und geben Sie die Hypothese, die Alternative, die Teststatistik sowie den Verwerfungsbereich an. Wie entscheidet der Test auf dem 5%-Niveau?

b) Berechnen Sie die Wahrscheinlichkeit eines Fehlers 2. Art, falls die Alternative $\mu_A = 69.5$ zutrifft (unter der Annahme von normalverteilten Beobachtungen).

Kennzahlen: $\overline{x}_{11} = 70.25$, $s^2 = 3.8409$.

3. In Aufgabe 2 von Serie 12 haben wir einen Test durchgeführt, um festzustellen, ob die mittlere Fahrzeit des Cisalpino auf der Strecke von Zürich nach Bellinzona von jener des Intercity abweicht. Die durchschnittliche Fahrzeit des Intercity-Zugs beträgt 146 Minuten. Mit dem Cisalpino wurden folgende Zeiten gemessen:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
152	145	141	137	145	146	139	147	138

Wir haben angenommen, dass diese Fahrzeiten unabhängige Realisierungen einer normalverteilten Zufallsvariablen X sind. Bei einem neuen Intercity-Zug misst man folgende Zeiten:

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
150	140	138	137	145	146	139	147	136

Man möchte diese Werte mit denen des Cisalpino vergleichen. Wir nehmen auch hier an, dass die Werte des neuen Zugs unabhängige Realisierungen einer normalverteilten Zufallsvariablen Y sind, welche dieselbe Varianz wie X besitzt. Sie möchten nun die Zeiten des neuen Zugs mit den Zeiten des Cisalpino vergleichen.

a) Handelt es sich um einen gepaarten oder einen ungepaarten Vergleich?

b) Führen Sie einen entsprechenden Test auf dem 5%-Niveau durch, um festzustellen, ob einer der beiden Züge (Cisalpino oder der neue) im Mittel schneller ist als der andere.

c) Bestimmen Sie den realisierten p-Wert und interpretieren Sie das Ergebnis.

Falls Ihnen die Tabellen im Skript nicht ausreichen, können Sie diesen Rechner für die t-Verteilung und diese Tabelle für die Standardnormalverteilung benutzen.

Kennzahlen: $\overline{x}_9 = 143.33$, $\overline{y}_9 = 142$, $s_X^2 = 24.25$, $s_Y^2 = 25.5$, $s^2 = 24.875$.

4. Eine Reifenfirma möchte ein neu entwickeltes Profil (A) mit einem bewährten Profil (B) vergleichen. Dazu werden n Fahrzeuge jeweils zuerst mit Reifen des Typs A und dann des Typs B bestückt und unter gleichen Bedingungen abgebremst. Es wird jeweils die Differenz der Bremswege (in Metern) gemessen. In einem konkreten Fall ergaben sich die Messwerte

Siehe nächste Seite!

Fahrzeug	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
Bremswegdifferenz $B - A$	0.4	-0.2	3.1	5.0	10.3	1.6	0.9	-1.4

Nehmen Sie an, dass die Bremswege vom Typ A und B jeweils normalverteilt sind. Wir möchten auf dem 2%-Niveau testen, ob sich die Bremszeit geändert hat.

a) Sind die Daten gepaart oder ungepaart? Welcher Test ist angebracht, um die gegebenen Informationen bestmöglich auszunützen?

b) Formulieren Sie die Hypothese und Alternative. Ist der Test ein- oder zweiseitig?

c) Führen Sie den Test durch. Wie entscheidet der Test?

Kennzahlen: $\overline{u}_8 = 2.4625$, $s^2 = 13.9598$.

Wahrscheinlichkeit & Statistik

Serie 12

1. Die Gamma-Verteilung mit Parametern $\alpha > 0$ und $\lambda > 0$ ist eine stetige Verteilung mit Dichte

$$f(z) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha z^{\alpha-1} e^{-\lambda z} I_{\{z \geq 0\}},$$

wobei

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$$

die Gammafunktion ist. Für $\alpha \in \mathbb{N}$ ist $\Gamma(\alpha) = (\alpha - 1)!$. Falls Z eine Zufallsvariable mit dieser Verteilung ist, schreiben wir $Z \sim Ga(\alpha, \lambda)$.

a) Seien X, Y unabhängig mit $X \sim Ga(\alpha, \lambda)$ und $Y \sim Ga(\beta, \lambda)$ für $\alpha > 0$, $\beta > 0$ und $\lambda > 0$. Zeigen Sie, dass $X + Y \sim Ga(\alpha + \beta, \lambda)$.

b) Seien $X_1, \ldots, X_n$ unabhängig und je $Exp(\lambda)$-verteilt mit $\lambda > 0$. Zeigen Sie, dass $X_1 + \cdots + X_n \sim Ga(n, \lambda)$.

2. Die durchschnittliche Fahrzeit von Zürich nach Bellinzona mit einem Intercity-Zug beträgt 146 Minuten. Mit dem Cisalpino werden die folgenden Zeiten gemessen:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
152	145	141	137	145	146	139	147	138

Wir nehmen an, dass diese Werte Realisierungen einer i.i.d. Stichprobe $X_1, \ldots, X_n$ sind mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ ein unbekannter Parameter und $\sigma^2 = 9$ ist. Führen Sie einen geeigneten Test durch, um auf dem 5%-Niveau festzustellen, ob die mittlere Fahrzeit des Cisalpino von jener des Intercity abweicht.

3. Sarah hat das neue Betriebssystem ‘Doors 2000’ auf ihrem Computer installiert und vermutet, dass sich die Zugriffszeit auf die Festplatte dadurch verändert hat. Mit dem neuen Betriebssystem erhält sie folgende Zugriffszeiten (in ms):

x_1	x_2	x_3	x_4	x_5	x_6
5.62	5.71	5.79	5.65	5.15	5.69

Siehe nächste Seite!

Sie geht davon aus, dass die Zeiten unabhängig voneinander und normalverteilt sind. Ohne 'Doors 2000' war die Zugriffszeit erfahrungsgemäss im Mittel 5.8 ms.

a) Welchen Test wählt Sarah? Mit welcher Hypothese und Alternative arbeitet sie? Führen Sie den Test zum 1%-Niveau durch.

b) Ändert sich etwas am Testergebnis, wenn man die fünfte Beobachtung (5.15) durch 5.65 ersetzt?

4. In einer Universitätsbibliothek wird normalerweise am Ende jedes Jahres eine Bestandsaufnahme durchgeführt, wo insbesondere überprüft wird, ob die Bücher richtig eingeordnet sind. Aufgrund einer neuen besseren Organisation ist der Bibliothekar der Meinung, dass die Bestandsaufnahme dieses Jahr verschoben werden könnte. Deshalb wählt er zufällig 100 Bücher und überprüft, ob diese richtig eingeordnet sind. Falls er dadurch statistisch belegen kann, dass der wirkliche Anteil an falsch eingeordneten Büchern kleiner als 3% ist, wird die Bestandsaufnahme verschoben.

a) Unter den 100 Büchern befindet sich nur eines, das falsch eingeordnet ist. Kann die Bestandsaufnahme daher verschoben werden? Führen Sie dazu einen statistischen Test durch, und gehen Sie dabei wie folgt vor:

i) Formulieren Sie eine geeignete Hypothese und Alternative, und geben Sie eine plausible Teststatistik an.

ii) Bestimmen Sie den Verwerfungsbereich des Tests auf dem 5%-Niveau.
Hinweis: Benutzen Sie $0.97^{100} = 0.0475$ und $100 \cdot 0.03 \cdot 0.97^{99} = 0.1407$.

iii) Wie entscheidet der Test?

b) Nehmen Sie nun an, dass der wahre Anteil an falsch eingeordneten Büchern 1% beträgt. Wie hoch ist die Wahrscheinlichkeit, dass die Bestandsaufnahme (unnötigerweise) verfrüht durchgeführt wird? Wie gross ist die Macht des Tests an der Stelle $\vartheta_A = 0.01$?
Hinweis: Benützen Sie $0.99^{100} = 0.3660$.

c) Um die Macht des Tests zu vergrössern, beschliesst der Bibliothekar, 1000 Bücher zu entnehmen. Nun findet er 10 Bücher, die falsch eingeordnet sind. Wie soll sich der Bibliothekar entscheiden? Führen Sie dazu wieder einen statistischen Test durch, und gehen Sie dabei wie folgt vor:

i) Formulieren Sie eine geeignete Hypothese und Alternative, und geben Sie eine plausible Teststatistik an.

ii) Berechnen Sie den Verwerfungsbereich des Tests auf dem 5%-Niveau.
Hinweis: Benützen Sie diesmal eine geeignete Approximation.

iii) Wie entscheidet der Test?
Hinweis: Benützen Sie $\sqrt{\frac{0.03 \cdot 0.97}{1000}} \approx 0.005$.

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 1

1. Wir analysieren einen Tennismatch von Roger Federer gegen Rafael Nadal. Der Match wird nach der Regel „best of 3“ gespielt; Sieger ist also, wer zuerst zwei Sätze gewinnt (es werden also maximal 3 Sätze gespielt). Wir nehmen an, dass Federer jeden einzelnen Satz – unabhängig von den anderen – mit Wahrscheinlichkeit $p = 1/3$ gewinnt. Mit A bezeichnen wir das Ereignis, dass Federer den ersten Satz gewinnt, und B bezeichne das Ereignis, dass Federer den Match (also zwei Sätze) gewinnt.

a) Drücken Sie $A \cup B$, $A^c \cap B$, $A \cap B^c$ und $A \setminus B$ in Worten aus. Berechnen Sie die bedingten Wahrscheinlichkeiten $P[B^c|A]$, $P[B|A]$ und $P[B|A^c]$.

b) Berechnen Sie die Wahrscheinlichkeit, dass Federer den Match gewinnt.

c) Berechnen Sie die bedingten Wahrscheinlichkeiten $P[A|B]$ und $P[A|B^c]$ mit Hilfe des Satzes von Bayes.

2. Bei einem Zufallsexperiment werden zwei Würfel gleichzeitig geworfen. Als Zufallsvariable S betrachten wir die Augensumme der beiden Würfel.

a) Welcher Grundraum Ω beschreibt dieses Zufallsexperiment?

b) Welche Werte kann S annehmen?

c) Berechnen Sie die Wahrscheinlichkeiten der Werte von S. Was nehmen Sie dabei an?

3. Sei X eine $\mathbb{N}$-wertige Zufallsvariable mit $0 < P[X = 1] < 1$ und der Eigenschaft

$$P[X = m + k | X > k] = P[X = m] \quad \text{für alle } k, m \in \mathbb{N}.$$

Zeigen Sie, dass $X \sim Geom(p)$ für ein $p \in (0, 1)$.

4. Sei $\Omega = \{1, 2, \ldots, n\}$ für ein $n \in \mathbb{N}$. Nehmen Sie an, dass die Menge Ω mit der Gleichverteilung P versehen ist, also

$$P[\{k\}] = \frac{1}{n} \quad \text{für alle } k \in \Omega.$$

Zeigen Sie, dass folgende zwei Aussagen äquivalent sind:

Siehe nächste Seite!

(i) Die Zahl n ist eine Primzahl.

(ii) Für alle unabhängigen Teilmengen $A, B \subseteq \Omega$ gilt, dass eine der beiden Mengen entweder leer oder die gesamte Grundmenge Ω ist.

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 3

1. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$. Zeigen Sie, dass A unabhängig von sich selbst ist genau dann, wenn $P[A] \in \{0, 1\}$.

2. Nehmen wir an, dass $-\infty < a < b < \infty$ und $c > 0$.

a) Sei $X \sim \mathcal{U}(0,1)$. Zeigen Sie, dass $a + (b-a)X \sim \mathcal{U}(a,b)$.

b) Sei $Y \sim Exp(\lambda)$ mit Parameter $\lambda > 0$. Welche Verteilung hat cY und was ist ihr Erwartungswert?

3. Sei $r > 1$ und $f : \mathbb{R} \to \mathbb{R}$ gegeben durch

$$f(x) = \begin{cases} 0 & \text{für } x \leq 1, \\ cx^{-r} & \text{für } x > 1 \end{cases}$$

für ein $c > 0$.

a) Bestimmen Sie die Konstante c so, dass f zu einer Dichtefunktion wird.

b) Wir nehmen ab jetzt an, dass Sie die Konstante so bestimmt haben, dass f zu einer Dichtefunktion wird. Sei X eine Zufallsvariable mit Dichte $f_X = f$. Berechnen Sie die Verteilungsfunktion von X.

c) Nehmen Sie nun zusätzlich an, dass $r > 2$. Existiert der Erwartungswert von X? Falls ja, berechnen Sie ihn.

4. In dieser Aufgabe werden wir beweisen, dass die Exponentialverteilung die einzige stetige Verteilung ist, welche gedächtnislos ist. Aus der Vorlesung ist schon bekannt, dass die Exponentialverteilung diese Eigenschaft erfüllt. Zeigen Sie also: Wenn $X > 0$ eine Zufallsvariable ist, welche $P[X \leq 1] > 0$ und die Eigenschaft

$$P[X > t + s | X > s] = P[X > t] \quad \text{für alle } s, t > 0$$

erfüllt, dann ist $X \sim Exp(\lambda)$ für ein $\lambda > 0$.

Hinweis: Falls $f, g : (0, \infty) \to \mathbb{R}$ zwei Funktionen sind, welche rechtsstetig sind und auf allen positiven rationalen Zahlen übereinstimmen, dann ist $f = g$.

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 2

1. Betrachten wir den Münzwurf einer zufällig gezinkten Münze als mehrstufiges Zufallsexperiment: Wir nehmen an, dass wir zwei gezinkte Münzen M_1 und M_2 in einer Urne haben. Die Wahrscheinlichkeit, dass M_1, resp. M_2, auf Kopf landet, sei $p \in (0,1)$, resp. $q \in (0,1)$. Bei jedem Zufallsexperiment wird eine Münze aus der Urne gezogen, dann geworfen und schliesslich wieder in die Urne gelegt.

a) Zeichnen Sie den Ereignisbaum, welcher entsteht, wenn wir das obige Experiment zweimal hintereinander durchführen.

b) Was ist die Wahrscheinlichkeit, dass zweimal Kopf geworfen wird?

2. Drei Personen A, B und C wurden verdächtigt, mit einer ansteckenden Krankheit infiziert zu sein. Um dies zu überprüfen, wurde jeder Person eine Blutprobe entnommen. Das Ergebnis sollte vorerst nicht bekannt gegeben werden. A erfuhr jedoch, dass sich nur bei einer Person der Verdacht bestätigte, und bat den Arzt, ihm im Vertrauen den Namen einer der Personen B oder C zu nennen, die *gesund* ist. Der Arzt lehnte die Auskunft mit der Begründung ab, dass dadurch die Wahrscheinlichkeit, dass A infiziert sei, von $\frac{1}{3}$ auf $\frac{1}{2}$ ansteigen würde. Stimmt diese Aussage des Arztes?
Um dies zu beantworten, nehmen Sie an, dass der Arzt die Auskunft an A gäbe und dass der Arzt, falls A an der ansteckenden Krankheit leiden sollte, mit gleicher Wahrscheinlichkeit B oder C nennen würde.

a) Berechnen Sie die Wahrscheinlichkeit, dass der Arzt den Namen von B nennt. Berechnen Sie die Wahrscheinlichkeit, dass der Arzt den Namen von C nennt. Was nehmen Sie dabei an?

b) Berechnen Sie die bedingte Wahrscheinlichkeit, dass A infiziert ist nach der Auskunft des Arztes. Was kann man daraus schliessen?

3. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, das heisst, Ω ist eine nichtleere Menge, $\mathcal{F} \subseteq 2^\Omega$ ist eine σ-Algebra und $P : \mathcal{F} \to [0,1]$ ist ein Wahrscheinlichkeitsmass. Sind A und B beliebige Elemente in $\mathcal{F}$, so kennen wir schon die Additionsregel

$$P[A \cup B] = P[A] + P[B] - P[A \cap B].$$

Siehe nächste Seite!

Sei nun $n \geq 2$ eine beliebige natürliche Zahl. Zeigen Sie, dass für beliebige Elemente $A_1, \ldots, A_n$ in $\mathcal{F}$ die allgemeine Inklusions- und Exklusionsformel

$$P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\left[\bigcap_{j=1}^{k} A_{i_j}\right]$$

gilt.

4. An einer Party geben $n \in \mathbb{N}$ Personen ihren Hut an der Garderobe ab. Die Bedienung in der Garderobe mischt die Hüte und gibt sie in zufälliger Weise wieder zurück.

a) Definieren Sie einen passenden Grundraum und die Wahrscheinlichkeiten der Elementarereignisse.

b) Was ist die Wahrscheinlichkeit, dass am Ende des Abends niemand seinen Hut zurückbekommt? (*Hinweis:* Benutzen Sie die obige Inklusions- und Exklusionsformel.)

c) Was ist die Wahrscheinlichkeit, dass genau $k \in \{1, \ldots, n\}$ Personen ihren Hut zurückbekommen?

d) Berechnen Sie den Grenzwert dieser Wahrscheinlichkeiten für $n \to \infty$.

Wahrscheinlichkeit & Statistik

Serie 6

1. Sei $X_1, X_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit Verteilungsfunktion F. Die empirische Verteilungsfunktion $F_n : \mathbb{R} \times \Omega \to [0,1]$ ist gegeben durch

$$F_n(t,\omega) := \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i(\omega) \leq t\}}.$$

Der Wert $F_n(t,\omega)$ beschreibt also die relative Häufigkeit der $X_i(\omega)$ mit Werten $\leq t$ unter den ersten n. Damit ist $F_n(t) := F_n(t,\cdot)$ selbst eine Zufallsvariable für jedes $t \in \mathbb{R}$.

a) Sei $t \in \mathbb{R}$ beliebig und $Y_i := I_{\{X_i \leq t\}}$ für $i \in \mathbb{N}$. Zeigen Sie, dass $Y_1, Y_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen ist.

Hinweis: Benutzen Sie, dass wie auch für diskrete Zufallsvariablen die Unabhängigkeit unter Transformationen erhalten bleibt.

b) Zeigen Sie, dass für jedes $t \in \mathbb{R}$ die empirische Verteilungsfunktion $F_n(t)$ fastsicher gegen $F(t)$ konvergiert für $n \to \infty$.

2. In dieser Aufgabe berechnen wir den Grenzwert

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

a) Sei $X \sim \mathcal{P}(\lambda)$ für ein $\lambda > 0$. Berechnen Sie $E[X]$ und $\mathrm{Var}[X]$.

b) Zeigen Sie mit dem zentralen Grenzwertsatz, dass der obige Grenzwert $1/2$ ist.

Hinweis: Für zwei unabhängige Zufallsvariablen X, Y mit $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$, und $Y \sim \mathcal{P}(\mu)$, $\mu > 0$, ist $X + Y \sim \mathcal{P}(\lambda + \mu)$.

3. Wir möchten in dieser Aufgabe verstehen, wie wir das Integral

$$A := \int_{-3}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

mit der Monte-Carlo-Methode berechnen können.

Siehe nächste Seite!

a) Wählen Sie eine geeignete Funktion f, sodass $A = E[f(U)]$ für $U \sim \mathcal{N}(0,1)$.

b) Seien $U_1, U_2, \ldots$ unabhängig und alle $\mathcal{N}(0,1)$-verteilt, und sei

$$S_n := \frac{1}{n} \sum_{i=1}^{n} f(U_i).$$

Welche Verteilung hat nS_n?

c) Berechnen Sie $E[S_n]$, und zeigen Sie, dass $\operatorname{Var}[S_n] = (A - A^2)/n$.

d) Zeigen Sie, dass

$$P[|S_n - A| \geq \varepsilon] \leq \frac{1}{n\varepsilon^2}$$

für jedes $\varepsilon > 0$ gilt. Was kann man nun über die Konvergenz von $S_1, S_2, \ldots$ aussagen? Konvergiert die Folge auch noch in einem anderen Sinn?

e) Welches Theorem können Sie anwenden, um auf schnellerem Weg die obige Konvergenz zu erhalten?

f) Erklären Sie kurz in Worten, wie man das Integral A numerisch approximieren kann.

4. Die Chebyshev-Ungleichung liefert eine Abschätzung von Wahrscheinlichkeiten, auch wenn die genaue Verteilungsfunktion nicht bekannt ist. Es müssen nur der Erwartungswert μ und die Varianz $\sigma^2 < \infty$ einer Zufallsvariablen X bekannt sein; dann gilt für jedes $k > 0$, dass

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}.$$

Die Anzahl der Turbinen, die pro Woche in einer Fabrik hergestellt werden, sei eine Zufallsvariable X mit $E[X] = 50$ und $\operatorname{Var}[X] = 25$.

a) Angenommen, es ist nichts weiter über die Zufallsvariable bekannt. Was kann man dann über die Wahrscheinlichkeit aussagen, dass in dieser Woche zwischen 40 und 60 Turbinen hergestellt werden?

b) Nun nehmen wir an, dass zusätzliche Informationen über die Verteilung von X erhältlich sind (dann kann die Wahrscheinlichkeit aus der letzten Teilaufgabe genauer berechnet werden und wir sehen, wie gut die Chebyshev-Abschätzung ist). Nehmen Sie an, dass $X \sim \mathcal{N}(\mu, \sigma^2)$. Wie gross ist dann die Wahrscheinlichkeit, dass in dieser Woche zwischen 40 und 60 Turbinen hergestellt werden?

Bemerkung: Die Chebyshev-Ungleichung ist bei vielen theoretischen Überlegungen zentral. Da sie aber unter so allgemeinen Bedingungen gültig ist, sind die angegebenen Schranken für die Wahrscheinlichkeit nicht sehr genau. In der Praxis ist sie daher nur dann sinnvoll, wenn wirklich keine Zusatzinformationen über die Verteilung erhältlich sind.

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 7

1. Die momenterzeugende Funktion einer Zufallsvariablen X ist

$$M_X(t) := E[e^{tX}] \quad \text{für } t \in \mathbb{R}.$$

a) Zeigen Sie, dass für $X_1, \ldots, X_n$ unabhängig gilt $M_{X_1+\cdots+X_n} = M_{X_1} \cdots M_{X_n}$, d.h.

$$M_{\sum_{i=1}^n X_i} = \Pi_{i=1}^n M_{X_i}.$$

b) Seien $X_1, \ldots, X_n$ unabhängig und $\mathcal{P}(\lambda)$-verteilt für ein $\lambda > 0$. Berechnen Sie die momenterzeugende Funktion von $X_1 + \cdots + X_n$.

2. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und $\mathcal{P}(\lambda)$-verteilten Zufallsvariablen mit $\lambda > 0$. Zeigen Sie, dass für $a > 0$ gilt

$$P\left[\frac{1}{n}\sum_{i=1}^n X_i \geq a\right] \leq e^{-nh(a;p)},$$

wobei $h(a;\lambda) = (\lambda - a) + a \log \frac{a}{\lambda}$.

3. Seien X_1, X_2, $\ldots$, X_n unabhängige und identisch verteilte Zufallsvariablen mit Dichte f und Verteilungsfunktion F. Seien

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\} \quad \text{und} \quad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}.$$

a) Bestimmen Sie die Dichtefunktionen $f_{(1)}$ von $X_{(1)}$ und $f_{(n)}$ von $X_{(n)}$.

b) Nehmen Sie an, dass $X_k \sim Exp(\lambda)$ für ein $\lambda > 0$. Seien

$$a_n := \frac{\log n}{\lambda} \quad \text{und} \quad b_n := \frac{1}{\lambda} \quad \text{für } n \in \mathbb{N}.$$

Zeigen Sie, dass

$$\lim_{n\to\infty} P\left[\frac{X_{(n)} - a_n}{b_n} \leq t\right] = e^{-e^{-t}} \quad \text{für } t > 0.$$

Bemerkung: Vergleichen Sie dieses Resultat mit dem zentralen Grenzwertsatz im Skript. Die Grenzverteilungsfunktion $e^{-e^{-t}} I_{\{t>0\}}$ heisst *Gumbel-Verteilung* und spielt eine fundamentale Rolle in der Extremwerttheorie.

Siehe nächste Seite!

4. Der Airbus A380 bietet insgesamt 526 Fluggästen Platz. Da Kunden manchmal ihren Flug nicht antreten, lassen Fluggesellschaften zwecks optimaler Auslastung Überbuchungen zu. Es sollen möglichst viele Tickets verkauft werden, wobei jedoch die Wahrscheinlichkeit einer Überbuchung maximal 0.05 betragen soll. Wieviele Tickets dürfen dann maximal verkauft werden, wenn bekannt ist, dass ein Kunde mit Wahrscheinlichkeit 0.04 nicht zum Flug erscheint und vereinfachend angenommen wird, dass das Nichterscheinen für verschiedene Kunden unabhängig voneinander ist?

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 5

1. Seien X und Y zwei unabhängige und $\mathcal{U}(0,1)$-verteilte Zufallsvariablen. Bestimmen Sie die Dichte und die Verteilungsfunktion der Summe $X + Y$.

2. Seien X und Y zwei unabhängige und $Exp(\lambda)$-verteilte Zufallsvairablen mit Parameter $\lambda > 0$. Zeigen Sie, dass der Quotient $Z := X/Y$ die Verteilungsfunktion

$$F_Z(t) = \begin{cases} \frac{t}{1+t} & \text{für } t > 0, \\ 0 & \text{für } t \leq 0 \end{cases}$$

besitzt.

3. Aus der Vorlesung ist bekannt, dass aus der Unabhängigkeit von Zufallsvariablen ihre Unkorreliertheit folgt. In dieser Aufgabe zeigen wir, dass die Umkehrung dieser Aussage im Allgemeinen nicht gilt.

a) Sei $X \sim \mathcal{U}(-\pi, \pi)$. Zeigen Sie, dass $Y := \cos(X)$ und $Z := \sin(X)$ unkorreliert sind.

b) Begründen Sie mit einem mathematischen Argument, warum Y und Z nicht unabhängig sind.

4. Sei $X = (X_1, X_2)$ eine $\mathbb{R}^2$-wertige Zufallsvariable mit Dichte

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 \det(\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

für $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$, wobei

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2 \quad \text{und} \quad \Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

mit $\Sigma_{1,2} = \Sigma_{2,1}$ und $\det(\Sigma) > 0$.

Siehe nächste Seite!

a) Zeigen Sie, dass $\mathrm{Var}(X_1) = \Sigma_{1,1}$, $\mathrm{Var}(X_2) = \Sigma_{2,2}$ und $\mathrm{Cov}(X_1, X_2) = \Sigma_{1,2}$.

Hinweis: Benutzen Sie, dass $\int_{-\infty}^{\infty} \exp(-ax^2 - bx)dx = \sqrt{\frac{\pi}{a}} \exp(\frac{b^2}{4a})$ für $a > 0$ und $b \in \mathbb{R}$. Ferner gilt für $\tilde{X} \sim \mathcal{N}(\mu, \sigma^2)$, dass $E[\tilde{X}] = \mu$ und $\mathrm{Var}[\tilde{X}] = \sigma^2$.

b) Zeigen Sie, dass X_1 und X_2 genau dann unabhängig sind, wenn $\mathrm{Cov}(X_1, X_2) = 0$.

ETH Zürich FS 2021
D-INFK
Prof. Dr. Martin Schweizer

Koordinator
Marco Rodrigues

Wahrscheinlichkeit & Statistik

Serie 4

1. Wir betrachten eine Messsonde an einem Vulkankrater, welche den bevorstehenden Ausbruch beobachten soll. Ab Beginn der Messungen gehen wir davon aus, dass die Sonde innerhalb von einer Minute mit einer Wahrscheinlichkeit von $\frac{1}{20}$ wegen zu grosser Beschädigung ausfällt. Die Zufallsvariable Y bezeichne die Lebensdauer der Sonde in Minuten. Es gilt $Y \sim Exp(\lambda)$.

a) Bestimmen Sie λ.

Hinweis: Falls Sie a) nicht gelöst haben, so rechnen Sie für die weiteren Teilaufgaben mit $\lambda = -\ln(0.95)$.

b) Wie gross ist die Wahrscheinlichkeit, dass die Sonde mehr als 10 Minuten überlebt?

c) Wir wissen, dass die Sonde schon mehr als 20 Minuten überlebt hat. Wie gross ist dann die bedingte Wahrscheinlichkeit, dass die Sonde nochmals 10 Minuten überlebt?

d) Die Zufallsvariable Z bezeichne die Lebensdauer einer zweiten Sonde, und es gelte, dass $Z \sim Exp(\mu)$ und dass Y und Z unabhängig sind. Was ist die Wahrscheinlichkeit, dass die zweite Sonde eine längere Lebensdauer hat als die erste?

2. Seien U_1, U_2, U_3 unabhängige und $\mathcal{U}(0,1)$-verteilte Zufallsvariablen. Wir betrachten die reellen Zufallsvariablen $L := \min(U_1, U_2, U_3)$ und $M := \max(U_1, U_2, U_3)$.

a) Berechnen Sie die Dichte von M.

b) Berechnen Sie die gemeinsame Dichte von L und M.

c) Berechnen Sie die bedingte Dichte von L gegeben M.

3. Seien X_1 und X_2 zwei reelle Zufallsvariablen mit gemeinsamer Dichte

$$f_{X_1,X_2}(x_1, x_2) = \frac{\sqrt{2}}{\pi} \exp\left(-\frac{3}{2}x_1^2 - x_1 x_2 - \frac{3}{2}x_2^2\right) \quad \text{für } (x_1, x_2) \in \mathbb{R}^2.$$

Siehe nächste Seite!

a) Bestimmen Sie die Verteilungsfunktion der Randverteilung von X_1 und X_2. Wie sind X_1 und X_2 verteilt?

Hinweis: Benutzen Sie, dass $\int_{-\infty}^{\infty} \exp\left(-x_1 x_2 - \frac{3}{2} x_2^2\right) dx_2 = \sqrt{\frac{2\pi}{3}} \exp(\frac{x_1^2}{6})$.

b) Sind X_1 und X_2 unabhängig?

4. Sei $X = (X^1, X^2)$ eine $\mathbb{R}^2$-wertige Zufallsvariable.

a) Zeigen Sie, dass die Koordinaten X^1 und X^2 unabhängig sind, falls X uniform auf dem Quadrat $[-1/2, 1/2]^2$ verteilt ist.

b) Zeigen Sie, dass die Koordinaten X^1 und X^2 nicht unabhängig sind, falls X uniform auf dem Quadrat verteilt ist, welches entsteht, wenn man $[-1/2, 1/2]^2$ um den Winkel $\pi/4$ dreht.