Vorlesung 15b

Eine Klausur aus dem Vorjahr

mit Lösungen

1. T sei Exp(1) verteilt, V sei uniform auf der zweielementigen Menge $\{-1,+1\}$ verteilt und unabhängig von T.
Wir definieren $X := V\sqrt{2T}$.

a) Berechnen Sie
(i) $\mathbf{P}(X \geq a)$ (ii) $\mathbf{P}(X \leq -a)$ (iii) $\mathbf{P}(|X| \leq a)$
jeweils für $a \geq 0$.

b) Berechnen und skizzieren Sie die Dichte
(i) von X (ii) von $|X|$.

c) Welchen Wert hat das Integral $\int_0^\infty a^2 e^{-a^2/2} da$?
(Hinweis: Vergleichen Sie das Integral mit der Varianz der Standard-Normalverteilung.)

d) Berechnen Sie den Erwartungwert von $|X|$.

Lösung:
a) i) $\mathbf{P}(X \geq a) = \mathbf{P}(V\sqrt{2T} \geq a) = \mathbf{P}(V = 1; \sqrt{2T} \geq a)$
Weil V und T unabhängig sind, ist dies gleich

$$\mathbf{P}(V = 1)\mathbf{P}(\sqrt{2T} \geq a) = \frac{1}{2}\mathbf{P}(2T \geq a^2/2) = \frac{1}{2}e^{-\frac{a^2}{2}}.$$

ii) X ist symmetrisch um 0 verteilt. Also ist

$$\mathbf{P}(X \leq -a) = (X \geq a) = \frac{1}{2}e^{-\frac{a^2}{2}}.$$

iii) $\mathbf{P}(|X| \leq a) = 1 - (\mathbf{P}(X \geq a) + \mathbf{P}(X \leq -a)) = 1 - e^{-\frac{a^2}{2}}.$

b) Die Dichtefunktionen entstehen durch Ableiten der in a berechneten Verteilungsfunktionen.

i) Für $a \geq 0$ ist Dichtefunktion von X in a gleich $-\frac{d}{da}\frac{1}{2}e^{-\frac{a^2}{2}} = \frac{a}{2}e^{-\frac{a^2}{2}}$. Diese Funktion hat im Argument 0 den Wert 0 und die Steigung $1/2$. Erst wächst sie dank dem Faktor a, hat ihr Maximum bei $a = 1$ und schmiegt sich wegen des Faktors $e^{-\frac{a^2}{2}}$ für große a an die Abszisse.

Wegen der Symmetrie um 0 ist die Dichte von X gleich $\frac{|a|}{2}e^{-\frac{a^2}{2}}\,da, \quad a \in \mathbb{R}$.

ii) Die Dichtefunktion von $|X|$ ist $\frac{d}{da}(1 - e^{-\frac{a^2}{2}}) = ae^{-\frac{a^2}{2}}, \, a \geq 0$.

c) Für standard-normalverteiltes Z ist

$$1 = \mathbf{Var}[Z] = \mathbf{E}[Z^2] - (\mathbf{E}[Z])^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} a^2 e^{-a^2/2} da$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} a^2 e^{-a^2/2} da,$$

also hat das Integral $\int_0^{\infty} a^2 e^{-a^2/2} da$ den Wert $\sqrt{\frac{\pi}{2}}$.

d) Wegen b) ii) gilt

$$\mathbf{E}[|X|] = \int_{-\infty}^{\infty} a \cdot ae^{-a^2/2} da = 2 \int_0^{\infty} a^2 e^{-a^2/2} da = \sqrt{\frac{\pi}{2}}.$$

2. X und Y seien zwei reellwertige Zufallsvariable mit endlichen Standardabweichungen σ_X und $\sigma_Y := 4\sigma_X$.

Es sei bekannt, dass der Anstieg der Regressionsgeraden für Y auf der Basis von X den Wert 3 hat (anders gesagt: die beste affin lineare Vorhersage von Y auf der Basis von X ist von der Form $\widehat{Y} = 3X + \beta_0$).

Bestimmen Sie den Anstieg der Regressionsgeraden für

a) Y auf der Basis von $2X$

b) $2X$ auf der Basis von Y

c) $2X$ auf der Basis von X.

Lösung.

Für reellwertige Zufallsvariable G und H wollen wir hier den Regressionskoefffizienten für (die beste affin lineare Vorhersage von) H auf der Basis von G mit $\beta_{H|G}$ bezeichnen. Die zentrale Formel ist $\beta_{H|G} = \frac{\sigma_H}{\sigma_G}\kappa_{GH}$.

Unter affin linearen Transformationen von X und Y ändert sich der Korrelationskoeffizient nicht. Insbesondere gilt $\kappa_{2X,Y} = \frac{\mathbf{Cov}[2X,Y]}{\sigma_{2X}\sigma_Y} = \kappa_{X,Y}$.

Also gilt: a) $\beta_{Y|2X} = \frac{\sigma_Y}{\sigma_{2X}}\kappa_{2X,Y} = \frac{1}{2}\frac{\sigma_Y}{\sigma_X}\kappa_{X,Y} = \frac{1}{2}\beta_{Y|X} = \frac{3}{2}$,

b) $\beta_{2X|Y} = \frac{\sigma_{2X}}{\sigma_Y}\kappa_{Y,2X} = 2\frac{\sigma_X}{\sigma_Y}\kappa_{X,Y} = 2\frac{1}{4}\kappa_{X,Y} = 2\frac{1}{4}\frac{1}{4}\beta_{Y|X} = \frac{3}{8}$,

c) $\beta_{2X|X} = \frac{\sigma_{2X}}{\sigma_X}\kappa_{2X,X} = 2\frac{\sigma_X}{\sigma_X}\kappa_{X,X} = 2$.

3. G sei uniform verteilt auf $S := \{1, 2, 3\}$.
Für jedes $a \in \{1, 2, 3\}$ gelte:
Gegeben $\{G = a\}$ ist H uniform verteilt auf $S \setminus \{1, a\}$.

(i) Bestimmen Sie die Matrix der Übergangsgewichte mit G als erster und H als zweiter Stufe.

(ii) Berechnen Sie die bedingte Verteilung von G gegeben $\{H = 2\}$.

Lösung

(i) Die Matrix der Übergangsgewichte ist $P = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 2 & 0 & 0 & 1 \\ 3 & 0 & 1 & 0 \end{array}$

(ii) Die Matrix der gemeinsamen Verteilungsgewichte von G und H entsteht durch Multiplikation der Einträge von P mit dn Gewichten der Startverteilung, hier $1/3$:

$$\begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & \frac{1}{6} & \frac{1}{6} \\ 2 & 0 & 0 & \frac{1}{3} \\ 3 & 0 & \frac{1}{3} & 0 \end{array} .$$

Die zweite Spalte dieser Matrix, normiert auf Gesamtgewicht 1, gibt die Antwort: Die Gewichte der bedingten Verteilung von G gegeben $\{H = 2\}$ sind $\frac{1}{3}, 0, \frac{2}{3}$.

4. Jedes von 10 Individuen ($i = 1, \ldots, 10$) bekommt rein zufällig einen von 100 möglichen Namen. Dabei stellen wir uns die an die Individuen $i = 1, \ldots, 10$ vergebenen Namen $X_1, \ldots, X_{10}$ als unabhängige, uniform auf $\{1, \ldots, 100\}$ verteilte Zufallsvariable vor.

a) Berechnen Sie die Wahrscheinlichkeit, dass keine zwei Individuen denselben Namen bekommen, mittels einer passenden, in der Vorlesung bereitgestellten Näherung.

b) Wir betrachten jetzt eine doppelt so große Anzahl von Individuen. Mit welchem Faktor muss man die Anzahl der Namen multiplizieren, damit die Wahrscheinlichkeit für die vollständige Identifizierbarkeit aller Individuen anhand ihrer Namen annähernd gleich bleibt?

Lösung

a) Es geht um die Kollisionswahrscheinlichkeit bei $n = 10$ Objekten und $g = 100$ Plätzen. Wegen $n \ll g$ verwenden wir die Näherung $e^{-\frac{1}{2}\frac{n(n-1)}{g}} \approx 0.64$..

b) Der gesuchte Faktor c ergibt sich aus der Gleichung

$$\frac{10 \cdot 9}{g} = \frac{20 \cdot 19}{cg}.$$

Also: $c = \dfrac{20 \cdot 19}{10 \cdot 9} \approx 4$.

5. In einer Population bestehend aus g Individuen ist der Mittelwert $\frac{1}{g}\sum_{j=1}^{g} w_j$ eines bestimmten individuellen Merkmals gleich μ. In einer ohne Zurücklegen gezogenen Stichprobe mit Umfang $n\,(\geq 20)$ ergibt sich der Stichprobenmittelwert 30 und die Stichprobenstandardabweichung 5.

Bestimmen Sie daraus ein approximatives 95%-Konfidenzintervall für den Populationsmittelwert μ, wenn

(i) g sehr groß gegenüber n ist

(ii) g doppelt so groß ist wie n.

Beidemale dürfen Sie mit der Normalapproximation rechnen. (Hinweis zu (ii) In einer Übungsaufgabe hatten wir gezeigt, dass der Korrekturfaktor für die Varianz beim Ziehen ohne Zurücklegen gleich $\frac{g-n}{g-1} \approx \frac{g-n}{g}$ ist.)

Lösung

Die geschätze Standardabweichung des Stichprobenmittelwertes ist $\frac{1}{\sqrt{n}}s$, mit $s = 5$. Die endliche Population-Korrektur bei Populationsgröße g (und bei Ziehen ohne Zurücklegen) ergibt den zusätzlichen Faktor $\sqrt{\frac{g-n}{g-1}}$. Ist g doppelt so groß ist wie n, dann ist dieser Faktor $\approx \frac{1}{\sqrt{2}}$.

Mig Verwendung der 2-sigma-Grenzen ist das Konfidenzintervall im Fall (i) gleich $30 \pm 2 \cdot \frac{5}{\sqrt{n}}$ und im Fall (ii) gleich $30 \pm 2 \cdot \frac{5}{\sqrt{2n}}$.

6. X sei eine Markovkette auf der Menge $S := \{a, b, c, d\}$ mit der durch die folgenden Eigenschaften festgelegten Übergangsmatrix P:

- Die Chancen, von a aus im nächsten Schritt a, b, c oder d zu erreichen, stehen im Verhältnis $0 : 3 : 2 : 1$.
- Die Chancen, von b aus im nächsten Schritt a, b, c oder d zu erreichen, stehen im Verhältnis $1 : 0 : 1 : 1$.
- $P(c, c) = P(d, d) = 1$.

Berechnen Sie:

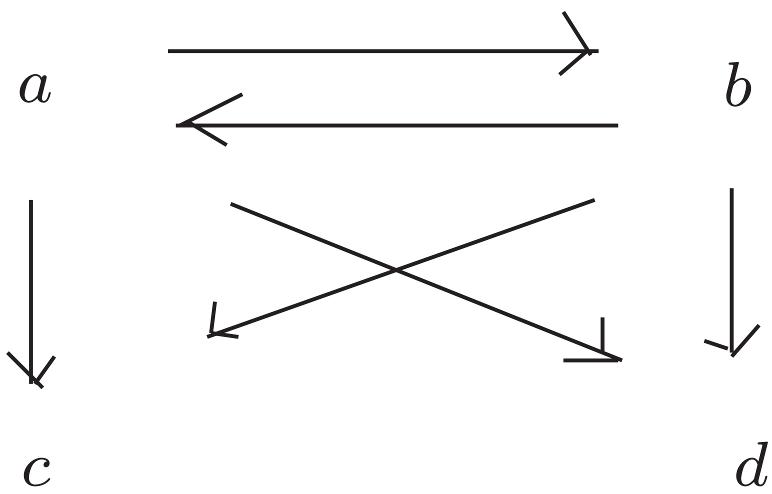
i) die Wahrscheinlichkeit, bei Start in a den Zustand c vor dem Zustand d zu erreichen,

ii) die erwartete Anzahl der Schritte bei Start in a bis zum erstmaligen Treffen der Menge $\{c, d\}$,

iii) drei verschiedene Gleichgewichtsverteilungen von X.

Lösung

Der Graph der Übergangswahrscheinlichkeiten ist



$$P(a,b) = \tfrac{1}{2},\ P(a,c) = \tfrac{1}{3},\ P(a,d) = \tfrac{1}{6}$$

$$P(b,a) = \tfrac{1}{3},\ P(b,c) = \tfrac{1}{3},\ P(b,d) = \tfrac{1}{3}.$$

(i) Für die Treffwahrscheinlichkeit "c vor d" in Abhängigkeit vom Startzustand ergibt sich das Gleichungssystem
$w(a) = \frac{1}{3} + \frac{1}{2} w(b)$
$w(b) = \frac{1}{3} + \frac{1}{3} w(a)$.
Einsetzen der zweiten in die erste Gleichung ergibt $w(a) = \frac{3}{5}$.

(ii) Für die erwartete Treffzeit von $\{c, d\}$ in Abhängigkeit vom Startzustand ergibt sich das Gleichungssystem
$e(a) = 1 + \frac{1}{2} e(b)$
$e(b) = 1 + \frac{1}{3} e(a)$.
Einsetzen der zweiten in die erste Gleichung ergibt $e(a) = \frac{9}{5}$.

iii) Drei verschiedene Gleichgewichtsverteilungen sind gegeben durch die Gewichte $(0, 0, 1, 0)$, $(0, 0, 0, 1)$, $(0, 0, \frac{1}{2}, \frac{1}{2})$.