Wahrscheinlichkeit und Statistik

Lösungen Serie 8

Lösung 8-1. (a) Setze $f(a,b) := E\left[(Y - aX - b)^2\right]$. Ausmultipliziert erhalten wir

$$f(a,b) = E[Y^2] + a^2 E[X^2] + b^2 - 2aE[XY] - 2bE[Y] + 2abE[X].$$

Um das minimum $(\widehat{a}, \widehat{b})$ zu finden, müssen wir den Gradienten ∇f von f gleich Null setzen und auflösen. Konkret bedeuted das:

$$\nabla f(\widehat{a}, \widehat{b}) = 0 \iff \quad \frac{\partial}{\partial a} f(\widehat{a}, \widehat{b}) = 0 \text{ und } \frac{\partial}{\partial b} f(\widehat{a}, \widehat{b}) = 0. \tag{1}$$

Die partiellen Ableitungen sind gegeben durch

$$\frac{\partial}{\partial a} f(\widehat{a}, \widehat{b}) = 2\widehat{a}E[X^2] - 2E[XY] + 2\widehat{b}E[X] \quad \text{und} \quad \frac{\partial}{\partial b} f(\widehat{a}, \widehat{b}) = 2\widehat{b} - 2E[Y] + 2\widehat{a}E[X]. \tag{2}$$

Kombinieren wir (2) mit der rechten Bedingung in (1) so müssen wir das System lösen

$$\begin{cases} \widehat{a}E[X^2] + \widehat{b}E[X] & = E[XY], \\ \widehat{a}E[X] + \widehat{b} & = E[Y]. \end{cases}$$

Elementare Umformungen ergeben, dass

$$\widehat{a} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}[X]} \quad \text{und} \quad \widehat{b} = E[Y] - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}[X]} E[X].$$

Der Prognosefehler ist gebeben durch

$$f(\widehat{a}, \widehat{b}) = \mathrm{Var}[Y] - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}[X]}.$$

(b) Für die gemeinsame Dichte $f_{X,Y}$ von X und Y muss gelten:

$$1 \stackrel{!}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c\mathbb{1}_{\{0, \leq y \leq 1, 0 \leq x \leq y^2\}} dx dy = \int_0^1 \int_0^{y^2} c dx dy \quad \Rightarrow \quad c = 3.$$

Zur Vorbereitung rechnen wir zunächst einige Momente aus. Allgemein gilt

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy,$$

wobei in unserem Fall $f_{X,Y}(x,y) = 3\mathbb{1}_{\{0, \leq y \leq 1, 0 \leq x \leq y^2\}}$ und somit

$$E[Y] = \int_0^1 \int_0^{y^2} 3y \; dx \; dy = \frac{3}{4} \qquad E[Y^2] = \int_0^1 \int_0^{y^2} 3y^2 \; dx \; dy = \frac{3}{5}$$

$$E[X] = \int_0^1 \int_0^{y^2} 3x \; dx \; dy = \frac{3}{10} \qquad E[X^2] = \int_0^1 \int_0^{y^2} 3x^2 \; dx \; dy = \frac{1}{7}$$

$$E[XY] = \int_0^1 \int_0^{y^2} 3xy \; dx \; dy = \frac{1}{4}.$$

Daraus folgt

$$\operatorname{Var}[Y] = E[Y^2] - E[Y]^2 = \frac{3}{80} \qquad \operatorname{Var}[X] = E[X^2] - E[X]^2 = \frac{37}{700}$$
$$\operatorname{Cov}[X,Y] = E[XY] - E[X]E[Y] = \frac{1}{40}.$$

Nach (a) gelten die Formeln:

$$\widehat{a} = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}[X]} = \frac{35}{74} \quad \text{and} \quad \widehat{b} = E[Y] - \widehat{a}E[X] = \frac{45}{74}.$$

Der erwartete quadratische Prognosefehler

$$E[(Y - \widehat{Y})^2] = \operatorname{Var}[Y] - \frac{\operatorname{Cov}[X,Y]^2}{\operatorname{Var}[X]} = \frac{19}{740} = 0.0256757.$$

Lösung 8-2. Sei X Pareto-verteilt mit den Parametern x_0 und $\alpha > 0$.

(a) Nach Definition des Erwartungswertes für stetige Verteilungen gilt $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$. Die Dichte f_X erhält man durch Ableiten der Verteilungsfunktion von X:

$$f_X(x) = F'_X(x) = \begin{cases} \alpha x_0^\alpha x^{-\alpha-1} & \text{für } x \geq x_0, \alpha > 0, \\ 0 & \text{sonst.} \end{cases}$$

(Bemerkung: $\alpha > 0$ muss gelten, weil sonst $f_X(x)$ keine Dichte ist.) Um $E[X]$ zu berechnen unterscheiden wir zwei Fälle. Für $\alpha \neq 1$ gilt

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_{x_0}^{\infty} x \alpha x_0^\alpha x^{-\alpha-1} dx \\ &= \alpha x_0^\alpha \int_{x_0}^{\infty} x^{-\alpha} dx = \alpha x_0^\alpha \left[\frac{1}{-\alpha+1} x^{-\alpha+1} \right]_{x_0}^{\infty} \\ &= \begin{cases} x_0 \frac{\alpha}{\alpha-1} < \infty & \text{falls } \alpha > 1, \\ \infty & \text{falls } \alpha < 1. \end{cases} \end{aligned}$$

Im Grenzfall $\alpha = 1$ gilt $E[X] = x_0 \int_{x_0}^{\infty} \frac{1}{x}\, dx = \infty$.

(b) Da $N \sim \mathcal{P}(\lambda)$ gilt $E[N] = \lambda$. Für die Nettoprämie P_{net} gilt dann $P_{net} = E[X]E[N] = x_0 \frac{\alpha}{\alpha-1} \lambda = 2 \cdot 10^6 \cdot 2 \cdot 3 = 12 \cdot 10^6$, also 12 Millionen CHF.

(c) Falls $\alpha \leq 1$, so ist $E[X] = +\infty$ (siehe **a)**). Eine Rückversicherung würde deshalb für keine noch so grosse Prämie Grossschäden versichern, die mit $\alpha \leq 1$ modelliert werden (müssen).

(d) Es gilt allgemein: Um Zufallsvariablen mit einer gewünschten Verteilungsfunktion F generieren zu können, muss F invertiert werden. Die Inverse F^{-1}, ausgewertet in den vorgegebenen Werten der uniformen Zufallsvariablen, ergibt die gewünschten Realisierungen. Hier wollen wir Realisierungen einer Pareto-verteilten Zufallsvariablen. Also muss als erstes F_X ($X \sim$ Pareto(x_0, α)) invertiert werden (nur für $x \geq x_0$!)

$$F_X(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha = y \implies x = \left(\frac{x_0^\alpha}{1-y}\right)^{1/\alpha} = \frac{x_0}{(1-y)^{1/\alpha}}.$$

Siehe nächstes Blatt!

Daraus folgt

$$F_X^{-1}(y) = \frac{x_0}{(1-y)^{1/\alpha}}\,,\ y \in (0,1)\,.$$

Einsetzen der Daten liefert

$$\begin{aligned} u &= 0.237 \implies x = F_X^{-1}(u) = \frac{2 \cdot 10^6}{(1-0.237)^{1/2}} = 2.29 \cdot 10^6\,, \\ u &= 0.733 \implies x = F_X^{-1}(u) = \frac{2 \cdot 10^6}{(1-0.733)^{1/2}} = 3.88 \cdot 10^6\,. \end{aligned}$$

Lösung 8-3. (a) Mit Chebyshev-Ungleichung:

$$P[|X-50| \geq 10] \leq \frac{\sigma^2}{10^2} = 0.25$$

Daher ist $P[|X-50| < 10] \geq 1 - 0.25 = 0.75$.

(b) Mit Normalverteilung:

$$\begin{aligned} &P[|X-50| < 10] = \\ &\quad = P[40 < X < 60] = P[X < 60] - P[X < 40] = \\ &\quad = P[\frac{X-50}{5} < \frac{60-50}{5}] - P[\frac{X-50}{5} < \frac{X-40}{5}] = \\ &\quad = \Phi(2) - \Phi(-2) = \Phi(2) - (1-\Phi(2)) = 2\Phi(2) - 1 = \\ &\quad = 2 \cdot 0.977 - 1 = 0.954 \end{aligned}$$

Die Chebyshev-Ungleichung liefert also eine korrekte Abschätzung der Wahrscheinlichkeit (0.95 ist grösser als 0.75), allerdings ist sie nicht sehr genau (0.75 liegt nicht sehr nahe an 0.95). Dies war allerdings zu erwarten, weil die Chebyshev-Ungleichung für **jede** Zufallsvariable mit endlichem Erwartungswert gilt.

Lösung Challenge Serie 8. Es ist klar, dass wir zuerst ein Modell für dieses Problem formulieren müssen. Dazu seien $X_1, \ldots, X_n, \ldots$ i.i.d. uniform auf $[0,1]$ verteilte Zufallsvariablen. Die Anzahl Zahlen bis deren Summe grösser als 1 ist, ist dann

$$\tau := \inf\left\{ n > 0;\ \sum_{k=1}^{n} X_k > 1 \right\}.$$

Das heisst, gesucht ist die Grösse $E[\tau]$. Wegen Fromel **3)** aus Satz 2.2. im Skript, gilt

$$E\left[\tau\right] = \sum_{\ell=1}^{\infty} P\left[\tau \geq \ell\right]. \tag{3}$$

Per Konstruktion von τ gilt aber auch für $\ell \geq 1$, dass

$$P\left[\tau \geq \ell\right] = P\left[\sum_{k=1}^{\ell} X_k \leq 1\right] =: a_\ell.$$

Wir müssen nur noch einen Ausdruck für a_ℓ finden. Technisch gesehen könnte man probieren die Dichtefunktion der Zufallsvariable $\sum_{k=1}^{\ell} X_k$ zu berechen (mittels Faltung, siehe Skript Kapitel

Siehe nächste Seite!

4.5.). Die resultierende Verteilung heisst *Irwin-Hall Verteilung* und sieht ein bisschen zu kompliziert aus (man kann sie im Internet finden). Wir versuchen jedoch einen einfacheren Weg. Man kann kleine Werte für ℓ in a_ℓ einsetzten und versuchen eine Formel zu *raten*, die man dann mittels Induktion beweist. Dazu erweist es sich als geschickter, wenn wir

$$a_\ell(t) := P\left[\sum_{k=1}^{\ell} X_k \leq t\right] \quad \text{für } t \in [0,1]$$

betrachten. Für $\ell = 1$ gilt klar, dass $a_1(t) = P[X_1 \leq t] = t$. Für $\ell = 2$ erhalten wir

$$a_2(t) = \int_0^1 \int_0^1 \mathbb{1}_{\{x+y \leq t\}} \,\mathrm{d}x \,\mathrm{d}y = \int_0^t (t-y) \,\mathrm{d}y = \frac{t^2}{2}.$$

Eine analoge Rechnung ergibt $a_3(t) = \frac{t^3}{6}$. Wir behaupten damit, dass $a_\ell(t) = \frac{t^\ell}{\ell!}$.
Beweis: Mittels Induktion: Dazu nehmen wir an, dass $a_\ell(t) = \frac{t^\ell}{\ell!}$ korrekt ist und beweisen die Formel für $\ell + 1$:

$$a_{\ell+1}(t) = P\left[\sum_{k=1}^{\ell} X_k \leq t - X_{\ell+1}\right] = \int_0^t a_\ell(t-x) \,\mathrm{d}x = \frac{t^{\ell+1}}{(\ell+1)!}.$$

■

Zusammenfassend ergibt sich $a_\ell = a_\ell(1) = \frac{1}{\ell!}$ und damit erhalten wir mit (3), dass $E[\tau] = e$.

Bemerkung. Eine natürliche Frage, die sofort auftaucht ist: Was ist $E[S_\tau]$? Die Antwort dieser Frage benötigt ein sogenanntes *Martingalargument*. Dies ist jedoch nicht Gegenstand der Vorlesung. Die Antwort lautet $e/2$.