Datenanalyse für Naturwissenschaftler*Innen
Statistische Methoden in Theorie und Praxis — **Uni Freiburg**
Fernandez Lebron, Andres (RZ: af231) — WiSe 2022/23

Blatt 07 — 9. Dezember 2022

Aufgabe 1 (Mittlerer quadratischer Fehler)

Zeigen Sie, dass für den Erwartungswert der quadratischen Abweichung des Schätzers $\hat{\theta}$ vom wahren Wert θ gilt:

$$E\left[\left(\hat{\theta}-\theta\right)^2\right] = V[\hat{\theta}] + b^2(\hat{\theta})$$

Lösung

z.z. $E\left[(\hat{\theta}-\theta)\right]^2 = V\left[\hat{\theta}\right] + b^2(\hat{\theta})$

Beweis:

Für x: $Var(x) = (E[x])^2 + E\left[x^2\right] \iff E\left[x^2\right] = Var(x) + E^2[x]$

$$\Rightarrow E\left[\left(\hat{\theta}-\theta\right)^2\right] = \underline{Var\left(\hat{\theta}-\theta\right)} + \overbrace{E^2\left[\left(\hat{\theta}-\theta\right)\right]}^{=b^2(\hat{\theta})}$$
$$= Var(\hat{\theta}) + b^2(\hat{\theta})$$

✓

$Var(\hat{\Theta}-\Theta)$
$= Var(\hat{\Theta})$
wegen Varianz translations invarianz

bzw.

$$\begin{aligned}
E\left[(\hat{\theta}-\theta)^2\right] &= E\left[\left(\hat{\theta}-E\left[\hat{\theta}\right]+E\left[\hat{\theta}\right]-\theta\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left[\hat{\theta}\right]\right)^2 + 2\cdot\left(\hat{\theta}-E\left[\hat{\theta}\right]\right)\cdot\left(E\left[\hat{\theta}\right]-\theta\right) + \left(E\left[\hat{\theta}\right]-\theta\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left[\hat{\theta}\right]\right)^2\right] + \underbrace{2\cdot E\left[\left(\hat{\theta}-E\left[\hat{\theta}\right]\right)\cdot\left(E\left[\hat{\theta}\right]-\theta\right)\right]}_{=0} + \left(E\left[\hat{\theta}\right]-\theta\right)^2 \\
&= E\left[\left(\hat{\theta}-E\left[\hat{\theta}\right]\right)^2\right] + \left(E\left[\hat{\theta}\right]-\theta\right)^2 \\
&= V\left[\hat{\theta}\right] + b^2(\hat{\theta})
\end{aligned}$$

2/2

✓

■

Aufgabe 2 (Schätzer für Parameter der Poissonverteilung)

Betrachten Sie n Messungen einer Messgröße k, die gemäß der Poisson-Wahrscheinlichkeitsdichtefunktion

$$f(k;\nu) = \frac{\nu^k}{k!}e^{-\nu}$$

mit festem ν verteilt sein soll.

(a) Berechnen Sie den „Maximum Likelihood"-Schätzer $\hat{\nu}$ für den Parameter ν (zeigen Sie auch, dass es wirklich ein Maximum ist).

(b) Ist der Schätzer erwartungstreu?

(c) Berechnen Sie die Effizienz des Schätzers.

Lösung

(a)

$$\mathcal{L} = \ln L = \sum_{i=1}^{n} \ln\left(\frac{v^{k_i}}{k_i!}e^{-v}\right) = \sum_{i=1}^{n}\left(\ln(e^{-v}) - \ln(k_i!) + \ln(v^{k_i})\right)$$

1,5/3

$$= \sum_{i=1}^{n}(v - \ln(k_i!) + k_i\ln(v))$$

$$= -\,nv + \sum_{i=1}^{n}\ln(k_i!) + \ln(v)\sum_{i=1}^{n}k_i$$

$$0 \stackrel{!}{=} \frac{\partial\mathcal{L}}{\partial v} = \frac{\partial}{\partial v}\left(nv + \sum_{i=1}^{n}\ln(k_i!) - \ln(v)\sum_{i=1}^{n}k_i\right) = n + 0 - \frac{1}{v}\sum_{i=1}^{n}k_i$$

Maximum bestimmen

$$\Rightarrow \frac{1}{v}\sum_{i=1}^{n}k_i = n \Rightarrow \hat{v} = \frac{1}{n}\sum_{i=1}^{n}k_i$$

o.k.

es fehlt $\frac{\partial^2 \ln(L)}{\partial^2 \nu}$

∎

(b) Poisson: $v = E[k]$

$$E[\hat{v}] = E\left[\frac{1}{n}\sum_{i=1}^{n}k_i\right] = \frac{1}{n}\sum_{i=1}^{n}E[k_i] = \frac{1}{n}\sum_{i=1}^{n}v = \frac{n}{n}v = v$$

$\Rightarrow$ ist erwartungstreu ✓

$\Rightarrow b(\hat{\nu}) = E(\hat{\nu}) - \nu = \nu - \nu = 0$

∎

1/1

(c)

$$\frac{\partial^2\mathcal{L}}{\partial v^2} = \frac{\partial}{\partial v}\left(n - \frac{1}{v}\sum_{i=1}^{n}k_i\right) = -0 + \frac{1}{v^2}\sum_{i=1}^{n}k_i = \frac{1}{v^2}\sum_{i=1}^{n}k_i = \frac{1}{\nu^2}n\cdot\hat{\nu}$$

$$E\left[\pm\frac{\partial^2\mathcal{L}}{\partial v^2}\right] = E\left[\pm\frac{1}{v}\sum_{i=1}^{n}k_i\right] = -\frac{v}{v^2} = -\frac{1}{v} \Rightarrow MVB = -v \quad \text{f}$$

$E\left[-\frac{\partial^2\mathcal{L}}{\partial\nu^2}\right] = \frac{n}{\nu^2}E[\hat{\nu}] \stackrel{b)}{=} \frac{\nu}{n}$

$$Var(\hat{\theta}) = Var\left(\frac{1}{n}\sum_{i=1}^{n}k_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n}k_i\right) - \frac{1}{n^2}\sum_{i=1}^{n}Var(k_i) = \quad \text{f}$$

bekannt:

$Var[\hat{\nu}] = \frac{Var[k]}{n}$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} = \frac{v^2}{n}$$

$= \frac{\nu}{n}$

$$\Rightarrow E(\hat{v}) = \frac{MVB}{Var(\hat{\theta})} = \frac{v}{\left(\frac{v^2}{n}\right)} = nv$$

0,5/2

Aufgabe 3 (Integration der Planck-Verteilung)

Wir betrachten die Erzeugung von Photonen, die dem Gesetz von der Strahlung des Schwarzen Körpers folgen. Die Verteilung der skalierten Frequenz $x = h\nu/kT$ ist gegeben durch die Plancksche Strahlungsformel:

$$f(x) = c\frac{x^3}{e^x - 1}$$

wobei c eine Normierungskonstante ist. Im folgenden soll besprochen werden, wie man diese Funktion möglichst effizient zwischen $x_0 = 0$ und einem bestimmten x_{max} integrieren könnte.

(a) Warum lässt sich die Transformationsmethode nicht anwenden?

(b) Skizzieren Sie den Verlauf der Funktion.

(c) Beschreiben Sie, wie man nach der Akzeptanz-Zurückweisungsmethode vorgehen würde, um die Funktion zu integrieren bzw. Zufallszahlen zu erzeugen. Wie ist die Effizienz gegeben? Warum wird dieses Verfahren für hohe x sehr ineffizient?

(d) Nun soll ein Ansatz mit einer stückweise definierten Majorante gemacht werden. Im unteren Bereich bis zu einem Punkt x_1 machen wir den Ansatz $g_I(x) = f_{\max}$, oberhalb von x_1 hingegen den folgenden Ansatz:

(i) Warum könnte der Ansatz

$$g_{II}(x) = K \cdot c \cdot x^{-\epsilon} \exp(-x^{1-\epsilon})$$

mit $0 < \epsilon < 1$ und $K > 1$ ein geeigneter Ansatz für eine Majorante sein?

(ii) Ermitteln Sie die Transformationsvorschrift, die zu dieser Majorante gehört. *Hinweise:* $\int x^{-\epsilon} e^{-(x^{1-\epsilon})} = e^{-(x^{1-\epsilon})}/(\epsilon - 1)$. *Sie müssen nicht nach $x(r)$ auflösen.*

(iii) Wie wird mit Hilfe der Majorante entschieden, ob eine Zufallszahl $x < x_1$ oder $x > x_1$ genommen wird?

(iv) Wie groß ist der Effizienzgewinn durch diese Majorante im Vergleich zur einfachen Akzeptanz-Zurückweisungsmethode im Gebiet $x > x_1$ für $c = 1$, $K = 200$, $\epsilon = 0{,}1$, $x_{max} = 20$, $x_1 = 6$? Das Maximum der Verteilung $f(x)$ ist $f_{\max} \approx 1{,}42 \cdot c$.

Lösung

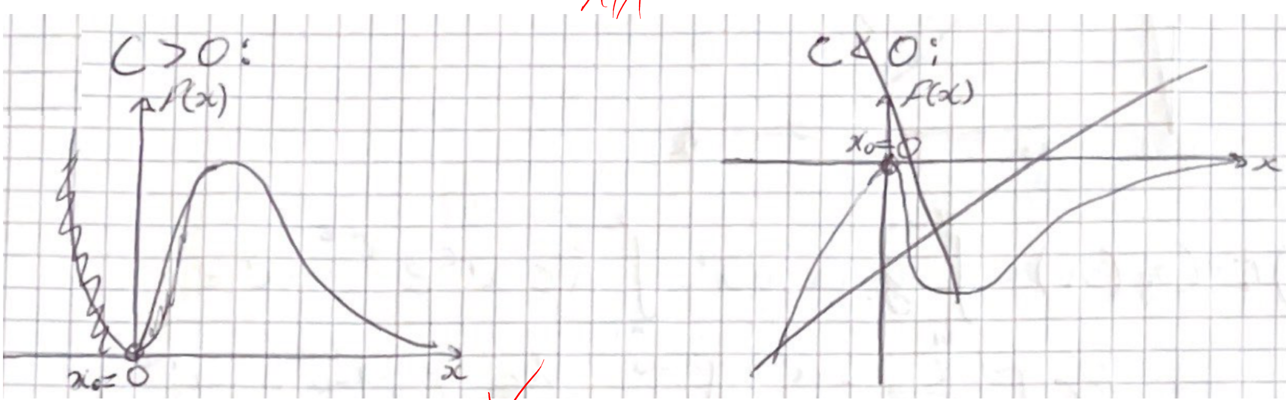
(a)

$f(x)$ ist nicht stetig und nicht stetig differenzierbar (undefiniert bei $x = 0$)

$\Rightarrow$ nicht invertierbar. ✓ 1/1

(b)

1/1



✓

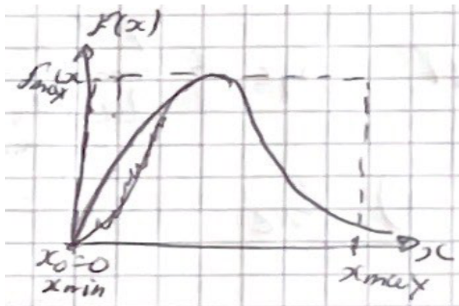
(c)

(1) WDF in einem Box

(2) Erzeuge ein Zufallzahl x im $[x_0, x_{max}]$:
$x = 0 + \Gamma_1(x_{max} - 0) = \Gamma_1 x_{max}$
mit $\Gamma_1$gleichmäßig im $[0, 1]$ ✓

(3) Erzeuge eine zweite unabhägige Zufallzahl w,
gleichmäßig verteilt im $[0, f_{max}]$: $w = \Gamma_2 f_{max}$ ✓

(4) Wenn $w < f(x)$, annehmen x.
Andernfalls, weise es zurück und wiederhole ✓



✓

x_0 anstatt 0 ist allgemeiner

$$Effizienz = \epsilon = \frac{\int_{\cancel{0}\,x_0}^{x_{max}} f(x')dx'}{f_{max}(x_{max} - \cancel{0}\,x_0)} \leq 1 \text{ or } \epsilon = \frac{\text{Anzahl der Annahmen}}{\text{Anzahl der Versuche}}$$

2/2

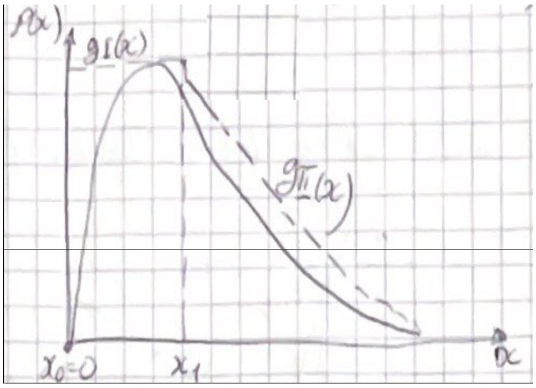
Ineffizient, weil für größere x oder $x \to \infty$ die Box wird nur bis x_{max} gewählt. ✓

(d)

$g_I(x) = f_{max}$

(i)

Weil $g_{II} > f(x)$ for $\forall x > x_1$: ✓



$\forall$ g_{II} hat invertierbare Stammfunktion
→ Transformationsmethode

vgl. $f(x)$ & $g_{II}(x)$ für große x

g_{II} nicht viel größer als $f(x)$ wg. Effizienz

1/2

(ii)

2.

$$\Gamma_2\ 2\ G_{II}(x) = \int_{x_1}^{x_{max}} g_{II}(x')dx = \int_{x_1}^{x_{max}} Kc\, x^{r^\epsilon} e^{-x^{M-\epsilon}} dx' =$$

$$= Kc \left[\frac{1}{6-1} e^{-x^{1-\epsilon}}\right]_{x_1}^{x_{max}} = \frac{Kc}{6-1}\left(e^{-x_{max}^{1-\epsilon}} - e^{-x_1^{1-\epsilon}}\right) = r$$

2/2

✓

(iii)

If $\Gamma_{2,1} \leq \frac{f(x_{2,1})}{g_{II/I}(x_{2,1})}$, dann annehmen $\Gamma_{2,1}$ foglich $x > x_1, x < x_1$

1/2

ausführlicher

$x < x_1$: wie in c)

$x > x_1$: 1) Zufallszahlen mit Transformationsvorschrift aus ii) generieren, $x_i = x(r_i)$

2) gleichverteilte Zufallszahlen $r_j \in [0,1]$ → $y_i = g_{II}(x_i)\, r_j$

akzeptieren wenn $y_i < f(x_i)$

(iv) $(x_{max} - x_1) f_{max} = 1{,}42(20-6)$

$$1)\ \epsilon_{AZ} = \frac{\int_{x_1}^{x_{max}=20} f(x')dx'}{f_{max}x_{max}} = \frac{\int_{x_1}^{20} f(x')dx'}{28.4}$$ f

$$2)\ \epsilon_{G_{II}} = \frac{\int_{x_1}^{x_{max}=20} f(x')dx'}{\frac{Kc}{\epsilon-1}\left(e^{-x_{max}^{1-\epsilon}} - e^{-x_1^{1-\epsilon}}\right)} = \frac{\int_{x_1}^{20} f(x')dx'}{\frac{200}{-0.9}\left(e^{-20^{0.9}} - e^{-6^{0.9}}\right)} = \frac{\int_{x_1}^{20} f(x')dx'}{1.47}$$

1,5/2

$$\Rightarrow \frac{\epsilon_{G_{II}}}{\epsilon_{AZ}} = \frac{28.4}{1.47} \approx 19.32$$ ff

$$\Rightarrow \epsilon_{G_{II}} = 19.32\ \epsilon_{AZ}$$

$g = \frac{\varepsilon_M}{\varepsilon_A} = \frac{(x_{max} - x_1) f_{max}}{\int_{x_1}^{x_{max}} g_{II}(x)\,dx}$

Aufgabe 4 (Monte Carlo Simulation)

Lösung

Siehe Jupyter Notebook Aufgabenblatt.