Vorlesung 23 | 29.1.2021 | 10:15–12:00 via Zoom

8 Anwendungen in der Statistik

(Kapitel 8 in Bovier Skript)

8.1 Einleitung

	W-theorie	Statistik
Gegeben	Modell (z.B. $\text{Ber}(1/2)$)	Stichprobe $(X_1, \ldots, X_n)$
Gesucht	Voraussage über das Verhalten einer Stichprobe $(X_1, \ldots, X_n)$, $X_k \sim \text{Ber}(1/2)$	Information über das zugrundeliegende Modell

Mit Modell annahmen: (z.B. Gaussverteilung der $X_1, \ldots, X_n$ mit unbekannte Mittelwert μ und Varianz v) $\rightarrow$ Parametrische Statistik (z.B. $(\mu, v) \in \mathbb{R} \times \mathbb{R}_+$ ist das Parameterraum) (endlich dimensionelle Probleme)

Ohne Annhame: Nicht parametrische Statistik unbekannte ist die ganze Verteilung $\mathbb{P}_X \in \Pi(\Omega)$ die Menge von alle W-Maße auf $(\Omega, \mathscr{F})$. (unendlich dimensionelle Probleme)

Beispiel. Neu Medikament.

Geuscht: Wirkungsgrad $\theta \in [0, 1] = \%$ der Leute, die von Medikament geheilt sind.

n Versuchpatienten $\xrightarrow{\text{Ergebnis}} (X_1, \ldots, X_n)$ mit

$$X_k = \begin{cases} 1, & \text{falls Patienten } k \text{ geheilt ist,} \\ 0, & \text{falls Patienten } k \text{ nicht geheilt ist.} \end{cases}$$

Modellannahme: X_k iid Bernoulli mit unbekannte Parameter θ.

Fragen:

a) Wie schätz man θ aus $(X_1, \ldots, X_n)$? Ein Schätzer T_n ist eine Funktion der Stichprobe, d.h.

$$T_n = f(X_1, \ldots, X_n).$$

b) Wie gross muss n sein, so dass der Messfehler $|T_n - \theta|$ klein genug ist. (Kosten v.s. Sicherheit)

c) Da θ nicht zufällig ist, aber T_n eine Z.V. ist. Gesucht werden sogenannte Konfidenzintervalle. $C_\theta = [T_n - \delta, T_n + \delta]$ s.d.

$$\mathbb{P}_\theta(\theta \in C_\theta) \geqslant 1 - \alpha,$$

mit α Irrtumsniveau ($\alpha \geqslant \mathbb{P}(\theta \notin C_\theta)$) Die Intervall ist unabhängig von θ! Die Modell $\mathbb{P}_\theta$ abhäng von θ. Wie wahlen wir T_n, δ?? (mit gegebenen Irrtumsniveau $\alpha \approx 0.1, 0.01, 0.05$)

d) Ein alte Medikament hat ein Wirkungsgrad von $\theta = 0.6$. Kann man aus $(X_1, \ldots, X_n)$ beschliessen ob den neuen Medikament besser/schlechter als den alten ist, oder reichen die Daten nicht aus? (Ja/Nein Antwort entscheiden).

8.2 Schätzer

Definition 1. *Ein Schätzer T_n ist eine (messbare) Funktion von $X_1, \ldots, X_n$. D.h. $T_n = f_n(X_1, \ldots, X_n)$.*

Dann, T_n ist einen Z.V.

Beispiel. Zwei Schäter für θ:

$$\begin{cases} T_n^{(1)} = \frac{1}{2} \\ T_n^{(2)} = \frac{1}{n}\sum_{k=1}^n X_k \\ T_n^{(3)} = X_1 \end{cases}$$

Alles sind Schätzer aber $T_n^{(1)}$ ist sinnlos, da nicht von den X_k abhängig ist!

Eine minimale Bedingung für ein „guten“ Schätzer ist Konsistenz.

Definition 2. *T_n ein Schätzer für θ heißt konsistent, falls*

$$\lim_{n\to\infty} T_n(\omega) = \theta \qquad \mathbb{P}_\theta - f.s.$$

für alle θ.

Aus GGZ, wissen wir, dass $T_n^{(2)}$ ist konsistent aber $T_n^{(1)}$ und $T_n^{(3)}$ sind es nicht.

Schätzer der W-Verteilung

Annahme: $X_1, \ldots, X_n$ iid Z.V. mit unbekannte Verteilung ν (eine unbekannte W-Maß auf $\mathbb{R}$). (nicht parametrische Stat.)
Gemessen: Frequenz der Ausgänge $X_k \in A \in \mathcal{B}(\mathbb{R})$.

$$\underbrace{\nu_n(A) := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_A(X_k).}_{\text{empirische Verteilung}}$$

$\nu_n\colon \Omega \to (\mathcal{B}(\mathbb{R}) \to [0,1])$ ist ein Zufällige Maß.

Lemma 3. *Für alle $A \in \mathcal{B}(\mathbb{R})$, $\nu_n(A)$ is ein konsistenten Schätzer von $\nu(A) = \mathbb{P}(X_1 \in A)$.*

Beweis. $(\mathbb{1}_A(X_k))_{k\geqslant 1}$ sind i.i.d. Z.V. mit $\mathbb{E}[\mathbb{1}_A(X_k)] = \mathbb{P}_\nu(X_k \in A) = \nu(A)$. Die Lemma folgt aus die GGZ, weil

$$\nu_n(A) = \frac{1}{n}\sum_{k=1}^n \mathbb{1}_A(X_k) \xrightarrow[n\to\infty]{\mathbb{P}_\nu - f.s.} \mathbb{E}[\mathbb{1}_A(X_1)] = \nu(A).$$

□

Wie gut ist die Approximation $\nu_n(A)$ von $\nu(A)$?

Lemma 4. *Seien $X_1, X_2, \ldots$ iid Z.V. mit Verteilung ν. Dann, für $A \in \mathcal{B}(\mathbb{R})$,*

$$\mathbb{P}_\nu\left(\underbrace{\left|\frac{\nu_n(A) - \nu(A)}{\nu(A)}\right|}_{\text{relative Fehler}} \geqslant \varepsilon\right) \leqslant \frac{1}{n\varepsilon^2 \nu(A)}.$$

Bemerkung. $\nu(A)$ klein $\Rightarrow$ $X_k \in A$ nicht so häufig ($\mathbb{E}(\#\{k$ s.d. $X_k \in A\}) = n\nu(A)$). braucht mehr Experimente um eine gute Approximation zu erreichen!

Beweis. $\nu(A) = \mathbb{P}_\nu(X_1 \in A)$

$$\nu_n(A) = \frac{1}{n}\sum_{k=1}^n \underbrace{\mathbb{1}_A(X_k)}_{\sim \mathrm{Ber}(\nu(A))}$$

$$\mathbb{E}[\nu_n(A)] = \nu(A)$$

Tchebichev

$$\mathbb{P}(|\nu_n(A)-\nu(A)|\geqslant t)\leqslant\frac{\mathrm{Var}(\nu_n(A))}{t^2}\underset{\text{unab.}}{=\!=}\frac{\mathrm{Var}(\mathbb{1}_A(X_k))}{nt^2}=\frac{\nu(A)(1-\nu(A))}{nt^2}$$

Nehmen $t=\varepsilon\nu(A)$. □

Jetz:

$$\mathbb{P}_\nu(|\nu_n(A)-\nu(A)|\leqslant\varepsilon\nu(A))=\mathbb{P}_\nu(\nu_n(A)-\varepsilon\nu(A)\leqslant\nu(A)\leqslant\nu_n(A)+\varepsilon\nu(A))$$

$$=\mathbb{P}_\nu(\nu(A)\in[\nu_n(A)-\varepsilon\nu(A),\nu_n(A)+\varepsilon\nu(A)])$$

aber $[\nu_n(A)-\varepsilon\nu(A),\nu_n(A)+\varepsilon\nu(A)]$ ist nicht ein Konfidenzintervall. Warum??? Weil das abhängt von $\nu(A)$ (unbekannt!!!)

Aber wir können nehmen ($\nu(A)\leqslant 1$) dann

$$\mathbb{P}_\nu(\nu(A)\in[\nu_n(A)-\varepsilon,\nu_n(A)+\varepsilon])\geqslant\mathbb{P}_\nu(\nu(A)\in[\nu_n(A)-\varepsilon\nu(A),\nu_n(A)+\varepsilon\nu(A)])\geqslant 1-\frac{1}{n\varepsilon^2\nu(A)}.$$

Immer noch nicht gut… Wir können nicht n,ε wählen (unabhängig von $\nu(A)$) so dass

$$\mathbb{P}_\nu(\nu(A)\in[\nu_n(A)-\varepsilon,\nu_n(A)+\varepsilon])\geqslant 1-\alpha$$

mit gegebenen Irrtumniveau α.

9 Schätzung von Erwartungswert und Varianz

Seien $X_1,X_2,\dots$ iid Z.V. in $\mathscr{L}^2$. Nach GGZ, konvergiert das empirische Mittelwert

$$m_n:=\frac{1}{n}\sum_{k=1}^n X_k\xrightarrow[n\to\infty]{f.s.}\mu=\mathbb{E}[X_1].$$

So die empirische Mittelwert ist ein konsistent Schätzer von μ.

Dazu

$$\mathbb{E}[m_n]=\mu.$$

Definition 5. *Ein Schätzer T_n von θ heißt Erwartungstreu (unbiased estimator) falls $\mathbb{E}(T_n)=\theta$.*

Bemerkung. Erwartungstreu ist eine oft erwünschte Eigenschaft, aber nicht immer gefordert. Ist eine Schätzfunktion nicht erwartungstreu, spricht man davon, dass der Schätzer *verzerrt* ist. Das Ausmaß der Abweichung seines Erwartungswerts vom wahren Wert nennt man Verzerrung oder *Bias*. Die Verzerrung drückt den systematischen Fehler des Schätzers aus.

Wir „gut“ ist m_n als Schätzer von μ?

Lemma 6. *Seien $X_1,X_2,\dots,X_n$ iid Z.V. mit $\mu=\mathbb{E}[X_1]$ und $\sigma^2=\mathrm{Var}(X_1)<\infty$. Dann m_n ist ein erwartungstreuer Schätzer für μ und*

$$\mathbb{P}(|m_n-\mu|\geqslant\varepsilon\mu)\leqslant\frac{\sigma^2}{n\mu^2\varepsilon^2}\tag{1}$$

Beweis. Folgt aus Tchebichev Ungleichung. □

Ein Beispiel für Fragen in der Klausur:

Frage: Warum eine Z.V. in $\mathscr{L}^2$ ist auch in $\mathscr{L}^1$? Falls $\mathbb{E}[X^2] < \infty$, warum $\mathbb{E}[|X|] < \infty$?

Antwort 1: Jensen's ungleichung:

$$\mathbb{E}[|X|] = \mathbb{E}[(|X|^2)^{1/2}] \leqslant (\mathbb{E}[|X|^2])^{1/2}$$

Antwort 2: Cauchy-Schwartz:

$$\mathbb{E}[|X|] = \mathbb{E}[|X| 1] \leqslant [\mathbb{E}[X^2]\mathbb{E}[1^2]]^{1/2} \leqslant (\mathbb{E}[|X|^2])^{1/2}.$$

$$|\mathbb{E}[XY]|^2 \leqslant \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

Antwort 3:

$$\mathbb{1}_{|X| \geqslant 1} X^2 \geqslant \mathbb{1}_{|X| \geqslant 1} |X|, \qquad \mathbb{1}_{|X| < 1} |X| \leqslant \mathbb{1}_{|X| < 1}$$

dann

$$\mathbb{E}[|X|] = \mathbb{E}[\mathbb{1}_{|X| \geqslant 1}|X|] + \mathbb{E}[\mathbb{1}_{|X| < 1}|X|] \leqslant \mathbb{E}[\mathbb{1}_{|X| \geqslant 1} X^2] + \mathbb{E}[\mathbb{1}_{|X| < 1}] \leqslant \mathbb{E}[|X|^2] + \mathbb{P}(|X| \leqslant 1) \leqslant \mathbb{E}[|X|^2] + 1 < \infty.$$

Antwort 4: $X \in \mathscr{L}^2$ dann die Char Fkt ist C^2 dann ist auch C^1 und dann $X \in \mathscr{L}^1$. (das ist richtig aber zu kompliziert…)