

7 Suffizienz und Vollständigkeit

7.1 Wiederholung

Bedingte Verteilungen

Sei (Ω, \mathcal{A}, P) Wahrscheinlichkeitsraum, $X : \Omega \rightarrow \mathbb{R}^k$, $Y : \Omega \rightarrow \mathbb{R}^s$ Zufallsvektoren.

Stochastik:

Es existiert Übergangswahrscheinlichkeit $P^{Y|X}$ mit

$$P^{(X,Y)} = P^X \otimes P^{Y|X} \quad (1)$$

$$P^{Y|X} : \begin{cases} \mathbb{R}^k \times \mathcal{B}^s \rightarrow \mathbb{R} \\ (x, B) \rightarrow P^{Y|X}(x, B) =: P^{Y|X=x}(B) \end{cases}$$

$$\begin{aligned} \text{mit } \forall x \in \mathbb{R}^k : & \quad P^{Y|X=x}(\cdot) \text{ Wahrscheinlichkeitsmaß auf } \mathcal{B}^s \\ \forall B \in \mathcal{B}^s : & \quad P^{Y|X=\cdot}(B) \text{ } \mathbb{R}^k \text{ -- messbar} \end{aligned}$$

$P^{Y|X}$ heißt (eine) bedingte Verteilung von Y bei gegebenem X.

$P^{Y|X=x}$ heißt (eine) bedingte Verteilung von Y bei gegebenem $X = x$.

Schreibweise:

$$P(Y \in B | X = x) := P^{Y|X=x}(B)$$

Dann (1) äquivalent zu

$$P^{(X,Y)}(A \times B) = P(X \in A, Y \in B) = \int_A P(Y \in B | X = x) P^X(dx)$$

$$\forall A \in \mathcal{B}^k, B \in \mathcal{B}^s$$

Insbesondere:

$$P(Y \in B) = \int P(Y \in B | X = x) P^X(dx)$$

Falls (X, Y) Dichte $f(x, y)$ bezüglich $\lambda \times \nu$ besitzt, so definiert man bedingte Dichte von Y gegeben $X = x$ durch

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

$$f_X(x) := \int f(x, y) \nu(dy) > 0$$

Damit:

$$P(Y \in B | X = x) = \int_B f_{Y|X}(y|x) \nu(dy)$$

$$\left[\begin{array}{l} P(X \in A, Y \in B) \stackrel{!}{=} \int_A [\int_B f_{Y|X}(y|x) \nu(dy)] f_X(x) \lambda(dx) \\ = \int_A \int_B f(x, y) d(\lambda \times \nu)(x, y) \end{array} \right]$$

7.2 Definition

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ Wahrscheinlichkeitsraum, $(\mathbb{R}^n, \mathcal{B}^n, \wp)$ statistischer Raum.

$X : \Omega \rightarrow \mathbb{R}^n$ Zufallsvektor, $T : \mathbb{R}^n \rightarrow \mathbb{R}^s$ Statistik.

T heißt **suffizient** für $\wp : \Leftrightarrow P^{X|T(X)}$ hängt nicht von $P \in \wp$ ab.

„Die bedingte Verteilung von X gegeben T ist bekannt.“

Falls $\wp = \{P_\vartheta : \vartheta \in \Theta\}$, so T **suffizient für** $\vartheta : \Leftrightarrow P^{X|T(X)}$ hängt nicht von ϑ ab.

7.3 Bemerkungen

(i) Wegen

$$\begin{aligned} \underbrace{P(X \in A, X \in B)}_{= \int_A P(X \in B | X=x) P^X(dx)} &= \int \mathbf{1}_{A \cap B}(x) P^X(dx) \\ &= \int_A \mathbf{1}_B(x) P^X(dx) \end{aligned}$$

$$\begin{aligned} &\text{gilt } P^{X|X=x}(B) = \mathbf{1}_B(x) \\ &\Rightarrow X \text{ suffizient für } \wp \end{aligned}$$

(ii) T suffizient für $\wp \Leftrightarrow \forall A \in \mathcal{B}^n : P(X \in A | T(X) = t)$ ist unabhängig von \wp für alle t (im Wertebereich von T)

(iii) Sei g bijektiv, g, g^{-1} messbar. Dann:

$$T \text{ suffizient} \Leftrightarrow g(T) \text{ suffizient}$$

7.4 Beispiel

$X = (X_1, \dots, X_n)$, $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bin}(1, \vartheta)$, $\vartheta \in (0, 1)$, $T(x) = \sum_{j=1}^n x_j$.
Sei $t \in \{0, 1, \dots, n\}$, $x \in \{0, 1\}^n$.

$$\begin{aligned} P_\vartheta(X = x | T = t) &= \frac{P_\vartheta(X = x, T = t)}{P_\vartheta(T(x) = t)} \\ &= \begin{cases} 0 & , \sum_{j=1}^n x_j \neq t \\ \frac{P_\vartheta(X=x)}{P_\vartheta(T(x)=t)} = \frac{\prod_{j=1}^n \vartheta^{x_j} (1-\vartheta)^{1-x_j}}{\binom{n}{t} \vartheta^t (1-\vartheta)^{n-t}} = \frac{1}{\binom{n}{t}} & , \sum_{j=1}^n x_j = t \end{cases} \end{aligned}$$

Also:

$$P_\vartheta^{X|T(X)=t} = U(\{(s_1, \dots, s_n) : s_j \in \{0, 1\} \ \forall j, \sum_{j=1}^n s_j = t\})$$

Insbesondere ist T suffizient für ϑ .¹⁹

7.3(ii) \Rightarrow

$$P_\vartheta(X \in A) = \int \underbrace{P(X \in A | T = t)}_{\text{unabhängig von } \vartheta} P_\vartheta^T(dt)$$

Hier:

$$\begin{aligned} P_\vartheta(X = x) &= \sum_{t=0}^n P(X = x | T = t) P_\vartheta(T = t) \\ &= \sum_{t=0}^n \frac{1}{\binom{n}{t}} \mathbf{1}\{\sum_{j=1}^n x_j = t\} \cdot \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t} \\ &= \vartheta^{\sum x_j} (1-\vartheta)^{n-\sum x_j} \end{aligned}$$

„In Verteilung von T ist alle Information bezüglich ϑ enthalten.“

\Leftrightarrow Datenreduktion **ohne Informationsverlust**

7.5 Faktorisierungssatz

In der Situation von 7.2 existiere σ -endliches Maß μ auf \mathcal{B}^n mit $P \ll \mu$
 $\forall P \in \wp$. Dann sind äquivalent:

- (i) $T(X)$ ist suffizient für \wp .

¹⁹ $P_\vartheta^{X|T(X)=t}$ Gleichverteilung (auf Menge)

- (ii) $\exists h : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ messbar, $\forall P \in \wp$ existiert $g_P : \mathbb{R}^k \rightarrow \mathbb{R}_{\geq 0}$ messbar mit

$$\frac{dP}{d\mu}(x) = g_P(T(x)) \cdot h(x), \quad x \in \mathbb{R}$$

Ist speziell $\wp = \{P_\vartheta : \vartheta \in \Theta\}$, $f(x, \vartheta) := \frac{dP_\vartheta}{d\mu}(x)$, $g(T(x), \vartheta) = g_{P_\vartheta}(T(x))$, so gilt also:

$$T \text{ suffizient} \Leftrightarrow f(x, \vartheta) = g(T(x), \vartheta) \cdot h(x) \quad \forall x \in \mathbb{R}^n$$

Beweis:

z.B. Shao, Mathematische Statistik, S.104-106 oder Pruscha, S. 77-80

7.6 Beispiel (Ordnungsstatistik)

Sei $X = (X_1, \dots, X_n)$, X_1, \dots, X_n unabhängig identisch verteilt mit Verteilung $P \in \wp$, \wp die Familie aller Verteilungen auf \mathbb{R} mit Lebesgue-Dichte.

$$T(X_1, \dots, X_n) := (X_{(1)}, \dots, X_{(n)})$$

geordnete Stichprobe (Ordnungsstatistik).

$$\frac{dP^n}{d\lambda^n}(x) = \prod_{j=1}^n f(x_j) = \underbrace{\prod_{j=1}^n f(x_{(j)})}_{=g_P(T(x))} \cdot \underbrace{1}_{=h(x)}$$

$\xRightarrow{7.5} T$ suffizient für \wp .

Bemerkung:

Es gilt

$$P^{X|T(x)=(x_{(1)}, \dots, x_{(n)})} = U(\{(x_{\pi_1}, \dots, x_{\pi_n}) : (\pi_1, \dots, \pi_n) \in \mathcal{S}_n\})$$

7.7 Beispiel (Exponentialfamilien)

In der Situation von Satz 6.4 ist $T_{(n)}(X)$ suffizient für ϑ .
[Aufgabe 21(b)]

7.8 Satz von Rao-Blackwell

Sei $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^s\})$ statistischer Raum, $g : \Theta \rightarrow \mathbb{R}$, $X : \Omega \rightarrow \mathfrak{X}$,
 $U_g = \{S \mid S : \mathfrak{X} \rightarrow \mathbb{R} \text{ messbar, } E_\vartheta S = g(\vartheta) \forall \vartheta \in \Theta, E_\vartheta S^2 < \infty \forall \vartheta \in \Theta\}$.

Annahme: $U_g \neq \emptyset$

Sei $T : \mathfrak{X} \rightarrow \mathbb{R}^k$ suffizient für ϑ , $S \in U_g$.

Sei $\tilde{S}(X) := E[S(X)|T(X)]$.²⁰

Dann gilt:

$$\tilde{S} \in U_g \text{ und } \text{Var}_\vartheta \tilde{S}(X) \leq \text{Var}_\vartheta S(X) \quad \forall \vartheta \in \Theta$$

(Verbesserung erwartungstreuer Schätzer durch suffiziente Statistiken)

Beweis:

$$E_\vartheta \tilde{S}(X) = E_\vartheta E[S(X)|T(X)] = E_\vartheta S(X) = g(\vartheta) \quad \forall \vartheta \in \Theta$$

$$\begin{aligned} \text{Var}_\vartheta S(X) &= E_\vartheta [(S(X) - \tilde{S}(X) + \tilde{S}(X) - \underbrace{E_\vartheta S(X)}_{=g(\vartheta)})^2] \\ &= \underbrace{E_\vartheta (S(X) - \tilde{S}(X))^2}_{\geq 0} + \text{Var}_\vartheta \tilde{S}(X) \\ &\quad + 2E_\vartheta [\underbrace{E_\vartheta [(S(X) - \tilde{S}(X))(\tilde{S}(X) - g(\vartheta))|T(X)]}_{=(\tilde{S}(X) - g(\vartheta)) \cdot \underbrace{E_\vartheta [S(X) - \tilde{S}(X)|T(X)]}_{=\tilde{S}(X) - \tilde{S}(X)=0}}] \\ &\geq \text{Var}_\vartheta \tilde{S}(X) \end{aligned}$$

[Beachte: $E_\vartheta \tilde{S}(X) = E_\vartheta S(X) = g(\vartheta)$; Regeln bedingter Erwartungswert²¹]

²⁰Nicht von ϑ abhängig, da T suffizient. (Sonst wäre \tilde{S} kein Schätzer!)

²¹insbesondere einmal ohne Auswirkung Erwartungswert in Erwartungswert eines bedingten Erwartungswertes umgeschrieben

7.9 Beispiel

$X_1, \dots, X_n \stackrel{uiv}{\sim} U(0, \vartheta), \vartheta \in \Theta = (0, \infty), X = (X_1, \dots, X_n)$

$$S(X) = \frac{2}{n} \sum_{j=1}^n X_j$$

$$\Rightarrow E_{\vartheta} S(X) = 2E_{\vartheta} X_1 = \vartheta$$

d.h. S erwartungstreu für ϑ .

$$\text{Var}_{\vartheta} S(X) = \frac{4}{n} \text{Var}_{\vartheta} X_1 = \frac{4}{n} \cdot \frac{\vartheta^2}{12} = \frac{\vartheta^2}{3n}$$

$$T(X) := \max_{1 \leq j \leq n} X_j$$

Wegen

$$f(x, \vartheta) = \prod_{j=1}^n \frac{1}{\vartheta} \mathbf{1}_{(0, \vartheta)}(x_j) = \underbrace{\frac{1}{\vartheta^n} \cdot \mathbf{1}_{(0, \vartheta)}(\max x_j)}_{=g(T(x), \vartheta)} \cdot \underbrace{1}_{=h(x)}$$

ist $T(X)$ suffizient für ϑ .

Wegen

$$P^{X_1 | \max X_j} = \frac{1}{n} \delta_{\max X_j} + \frac{n-1}{n} U(0, \max X_j)$$

folgt

$$\begin{aligned} \tilde{S}(X) &= E[S(X) | \max_j X_j] \\ &= \frac{2}{n} \sum_{i=1}^n E[X_i | \max_j X_j] \\ &= 2E[X_1 | \max_j X_j] \\ &= 2\left(\frac{1}{n} \cdot \max_j X_j + \frac{n-1}{n} \frac{\max_j X_j}{2}\right) \\ &= \frac{n+1}{n} \max_j X_j \end{aligned}$$

$$\text{Var}_{\vartheta} \tilde{S}(X) = \dots = \frac{\vartheta^2}{n(n+2)} < \text{Var}_{\vartheta} S(X) \quad \text{für } n \geq 2$$

$$\text{Var}_{\vartheta} \tilde{S}(X) = \dots = \frac{\vartheta^2}{n(n+2)} = \text{Var}_{\vartheta} S(X) \quad \text{für } n = 1$$

7.10 Definition

In der Situation von 7.2 heißt $T : \mathbb{R}^n \rightarrow \mathbb{R}^k$ vollständig für $P \in \wp$ (bzw. $\vartheta \in \Theta$, falls $\wp = \{P_\vartheta : \vartheta \in \Theta\}$), falls gilt:

Für jede messbare Funktion $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$ mit $E_P[\Psi(T)] = 0 \ \forall P \in \wp$ (bzw. $E_\vartheta[\Psi(T)] = 0 \ \forall \vartheta \in \Theta$) folgt $\Psi(T) = 0$ P-f.s. $\forall P \in \wp$ (bzw. P_ϑ -f.s. $\forall \vartheta \in \Theta$).

7.11 Beispiel (Fortsetzung von 7.9)

Behauptung:

$\overline{T(X)} := \max_j X_j$ vollständig

Beweis:

Sei $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ messbar.

$$E_\vartheta \Psi(T) = \int_0^\vartheta \Psi(t) \cdot \frac{n}{\vartheta} \left(\frac{t}{\vartheta}\right)^{n-1} dt = \frac{n}{\vartheta^n} \underbrace{\int_0^\vartheta \Psi(t) \cdot t^{n-1} dt}_{=: G(\vartheta)}$$

$$E_\vartheta \Psi(T) = 0 \ \forall \vartheta > 0 \Rightarrow G(\vartheta) = 0 \ \forall \vartheta > 0$$

$$\Rightarrow \Psi(\vartheta) \cdot \vartheta^{n-1} = 0 \quad \lambda^1|_{[0,\infty)}\text{-f.s.}$$

$$\Rightarrow \Psi(\vartheta) = 0 \quad \lambda^1|_{[0,\infty)}\text{-f.s.}$$

$$\Rightarrow P_\vartheta(\Psi(T) = 0) = 1$$

7.12 Beispiel

In einer strikt k-parametrischen Exponentialfamilie

$$f(x, \vartheta) = C(\vartheta) \cdot e^{\vartheta^T T(x)} h(x)$$

(mit natürlichem Parameterraum) ist die Statistik T vollständig.

(Beweis z.B. Shao, S.110 oder Pruscha, S.82)

Beispiel:

Sei $X_1, X_2 \overset{uiv}{\sim} \mathcal{N}(\vartheta, 1), \vartheta \in \mathbb{R}$.

$T(X_1, X_2) = X_1 + X_2$ ist vollständig nach 7.12.

$S(X_1, X_2) = X_1 - X_2$ dagegen nicht!

$$T \sim \mathcal{N}(2\vartheta, 2) = P_\vartheta^T$$

$$E_\vartheta \Psi(T) = \int_{\mathbb{R}} \Psi(t) \cdot \underbrace{\varphi_{2\vartheta, 2}(t)}_{\text{Dichte NV}} dt$$

$S \sim \mathcal{N}(0, 2) = P_\vartheta^S, \Psi(S) = S:$

$$E_\vartheta \Psi(S) = \vartheta - \vartheta = 0 \quad \forall \vartheta \in \Theta$$

$$\not\equiv \Psi(S) = X_1 - X_2 = 0 \quad P_\vartheta\text{-f.s.}$$

$\{P_\vartheta^T : \vartheta \in \mathbb{R}\} = \{\mathcal{N}(2\vartheta, 2) : \vartheta \in \mathbb{R}\}$ ist viel „reichhaltiger“ als $\{P_\vartheta^S : \vartheta \in \mathbb{R}\} = \{\mathcal{N}(0, 2)\}!$

7.13 Satz von Lehmann-Scheffé

In der Situation von 7.8 ($U_g \neq \emptyset$) sei die suffiziente Statistik T auch vollständig für ϑ . Dann existiert ein eindeutig bestimmter erwartungstreuer Schätzer für $g(\vartheta)$ der Gestalt

$$S^*(X) = h(T(X))$$

mit einer messbaren Funktion $h : \mathbb{R}^k \rightarrow \mathbb{R}$.

Dieser Schätzer ist UMVUE für $g(\vartheta)$.

Beweis:

Sei $S \in U_g$ und $\tilde{S}(X) := E[S(X)|T(X)]$.

Faktorisierungssatz des bedingten Erwartungswerts \Rightarrow es existiert h messbar mit

$$\tilde{S}(X) = h(T(X))$$

Annahme: $\exists S_* \in U_g$ mit $S_*(X) = h_*(T(X))$ für ein h_*

$$\Rightarrow E_\vartheta[\underbrace{(h - h_*)}_{=: \Psi}(T)] = g(\vartheta) - g(\vartheta) = 0 \quad \forall \vartheta \in \Theta$$

$$\stackrel{(+)}{\Rightarrow} h = h_* \quad P_\vartheta\text{-f.s.} \quad \forall \vartheta \in \Theta$$

(+): Vollständigkeit von T ($\Psi = h - h_* = 0$)

$\tilde{S}(X)$ ist UMVUE für $g(\vartheta)$!

[Annahme: S_2 „besser“ als \tilde{S}

$$\Rightarrow \tilde{S}_2(X) = E[S_2(X)|T(X)]$$

„mindestens so gut“ wie S_2 (Rao-Blackwell); $\tilde{S}_2 = \tilde{S}$ wegen Eindeutigkeit]

7.14 Beispiel (Fortsetzung von 7.11)

$\frac{n+1}{n} \max_j X_j$ ist UMVUE für ϑ , falls $X_1, \dots, X_n \stackrel{uiv}{\sim} U(0, \vartheta)$, $\vartheta > 0$ unbekannt.

7.15 Beispiel (Anwendungen von Lehmann-Scheffé)

Sei T vollständig und suffizient für ϑ , $\vartheta \in \Theta$.

Finde h , so dass $E_{\vartheta}[h(T)] = g(\vartheta) \forall \vartheta \in \Theta$. (Lösen!, Raten!)

Falls $\text{Var}_{\vartheta}[h(T)] < \infty \Rightarrow h(T)$ UMVUE.

Sei $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$.

(i) Aufgabe 20:

$$\text{Var}_{\vartheta}((\bar{X}_n, S_n^2)^T) = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{bmatrix}$$

$$\frac{2\sigma^4}{n-1} > \text{CR-Schranke } \frac{2\sigma^4}{n}$$

$\Rightarrow (\bar{X}_n, S_n^2)$ nicht CR-effizient für ϑ

Aber:

$$T(X) = \left(\sum_i X_i, \sum_i X_i^2 \right)$$

suffizient und vollständig für $\bar{\vartheta} = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right)$ (nach 7.12).

$\Rightarrow T(X) = \left(\sum_i X_i, \sum_i X_i^2 \right)$ suffizient und vollständig für $\vartheta = (\mu, \sigma^2)$.

Sei $h(T(X)) = (\bar{X}_n, S_n^2)$.

$$\left. \begin{array}{l} E_{\vartheta}[h(T(X))] = \vartheta \forall \vartheta \\ \text{Var}_{\vartheta}[h(T(X))] \text{ existiert } \forall \vartheta \end{array} \right\} \Rightarrow (\bar{X}_n, S_n^2) \text{ ist UMVUE für } \vartheta = (\mu, \sigma^2)$$

Bemerkung:

Auch (\bar{X}_n, S_n^2) suffizient und vollständig für ϑ nach Bemerkung 7.3(ii) und analoge Aussage für Vollständigkeit.

(ii) Analog:

Der Schätzer aus Aufgabe 9 der Form $\sqrt{c_n S_n^2}$ ist UMVUE für σ .

(iii) Gesucht: UMVUE für $\frac{\mu}{\sigma}$

$$(T_1(X), T_2(X)) := \left(\sum_i X_i, \sum_i (X_i - \bar{X}_n)^2 \right)$$

$T_1(X), T_2(X)$ unabhängig, $\frac{T_1(X)}{\sigma^2} \sim \chi_{n-1}^2 = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$

$$\Rightarrow E_{\vartheta} T_2^{-\frac{1}{2}} = \frac{1}{\sigma} \cdot \frac{\Gamma\left(\frac{n}{2} - 1\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)} \quad (n \geq 3)$$

$$\text{Var}_{\vartheta} T_2^{-\frac{1}{2}} < \infty \text{ für } n \geq 4$$

$$\begin{aligned}
\Rightarrow E_{\vartheta}\left(\frac{T_1}{\sqrt{T_2}}\right) &= E_{\vartheta}T_1 \cdot E_{\vartheta}T_2^{-\frac{1}{2}} \\
&= \frac{\mu}{\sigma} \cdot \frac{n\Gamma(\frac{n}{2}-1)}{\sqrt{2}\Gamma(\frac{n-1}{2})} \\
&=: \frac{\mu}{\sigma}K_n
\end{aligned}$$

$(n \geq 3)$

$$\Rightarrow K_n^{-1} \cdot \frac{T_1}{\sqrt{T_2}}$$

ist UMVUE für $\frac{\mu}{\sigma}$ für $n \geq 4$.