2. Prinzipien der Datenreduktion

Man verwendet die Information in einer Stichprobe $X_1, \ldots, X_n$, um **statistische Inferenz** über einen unbekannten Parameter zu betreiben. Falls n groß ist, so ist die beobachtete Stichprobe $x_1, \ldots, x_n$ eine lange Liste von Zahlen die nur schwer zu interpretieren ist.

Man möchte die Information in dieser Stichprobe **zusammenfassen**, indem man einige besondere, zentrale Größen bestimmt (Statistiken, also Funktionen der Stichprobe). Beispielsweise könnte man empirisches Mittel und Varianz, Minimum und Maximum verwenden, um die Eigentümlichkeiten der Stichprobe damit zu charakterisieren.

Sei jetzt $\mathbf{X} = (X_1, \ldots, X_n)$ und $\mathbf{x} = (x_1, \ldots, x_n)$ die beobachteten Werte.

Jede **Statistik** $T(\mathbf{X})$ definiert eine Form der **Datenreduktion** oder Datenzusammenfassung.

Verwendet man den beobachteten Wert $T(\mathbf{x})$ anstelle der gesamten beobachteten Stichprobe, so wird man 2 Stichproben $\mathbf{x}$ und $\mathbf{y}$ als **identisch** behandeln, falls $T(\mathbf{x}) = T(\mathbf{y})$ gilt; obwohl sich die tatsächlichen Werte $x_1, \ldots, x_n$ von $y_1, \ldots, y_n$ unterscheiden.

Datenreduktion in Termen einer Statistik kann auch als **Partitionierung des Stichprobenraums** χ gesehen werden. Sei

$$\tau = \{t : t = T(\mathbf{x}) \;\text{ für beliebiges } \mathbf{x} \in \chi\}$$

das Bild von χ unter $T(\mathbf{x})$. Dann partitioniert $T(\mathbf{x})$ den Stichprobenraum in Mengen A_t, $t \in \tau$, mit

$$A_t = \{\mathbf{x} : T(\mathbf{x}) = t\}\,.$$

Die Statistik fasst nun alle Daten zusammen und zeigt nur an, dass

$$T(\mathbf{x}) = t \Longleftrightarrow \mathbf{x} \in A_t\,.$$

Wir interessieren uns für Reduktionsmethoden, die keine bedeutende Informationen über den unbekannten Parameter θ vernachlässigen.

2.1 Das Suffizienz-Prinzip

Eine **suffiziente Statistik** für den Parameter θ ist eine Statistik, die alle relevante Information der Stichprobe über θ beinhaltet. Jede zusätzliche Information in der Stichprobe wäre für θ irrelevant.

Suffizienz Prinzip: Falls $T(\mathbf{X})$ eine **suffiziente Statistik** für $\boldsymbol{\theta}$ ist, dann sollte jegliche Folgerung über $\boldsymbol{\theta}$ von der Stichprobe $\mathbf{X}$ nur von $T(\mathbf{X})$ abhängen. Dies hat zur Folge, dass für zwei Stichprobenpunkte $\mathbf{x}$ und $\mathbf{y}$ mit $T(\mathbf{x}) = T(\mathbf{y})$ die Folgerung über $\boldsymbol{\theta}$ dieselbe ist, unabhängig davon ob $\mathbf{X} = \mathbf{x}$ oder $\mathbf{X} = \mathbf{y}$ beobachtet wurde.

Definition 2.1.1: Eine Statistik $T(\mathbf{X})$ ist **suffizient** für θ, falls die bedingte Verteilung von $\mathbf{X}$, gegeben der Wert von $T(\mathbf{X})$, nicht von θ abhängt.

Direkte Konsequenz:

Satz 2.1.1: Bezeichne $p(\mathbf{x}|\theta)$ die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion der Stichprobe $\mathbf{X}$ und sei $q(t|\theta)$ die Dichte oder Wahrscheinlichkeitsfunktion von $T(\mathbf{X})$. Dann ist $T(\mathbf{X})$ eine suffiziente Statistik für den Parameter θ, falls für jedes $\mathbf{x}$ im Stichprobenraum der Quotient

$$\frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

konstant ist, als Funktion in θ betrachtet.

Damit ist es möglich zu verifizieren, ob eine Statistik $T(\mathbf{X})$ suffizient ist für einen Parameter θ.

Beispiel 2.1.1: Seien $X_1, \ldots, X_n$ iid Bernoulli(θ)-Variablen, mit $0 < \theta < 1$.

Frage: ist $T(\mathbf{X}) = \sum_i X_i$ suffizient für θ?

Bekannterweise ist $T(\mathbf{X}) \sim$ Binomial(n, θ).

Sei $t = \sum_i x_i$, dann gilt

$$\begin{aligned}
\frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} &= \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{\theta^{\sum_i x_i}(1-\theta)^{\sum_i(1-x_i)}}{\binom{n}{\sum_i x_i}\theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}} \\
&= \frac{1}{\binom{n}{\sum_i x_i}},
\end{aligned}$$

was konstant ist in θ. Daher ist $T(\mathbf{X})$ eine suffiziente Statistik für θ.

Beispiel 2.1.2: Seien $X_1, \ldots, X_n$ iid Normal(μ, σ^2), mit $0 < \sigma^2$ bekannt.

Zeige, dass $T(\mathbf{X}) = \overline{X}$ suffizient ist für μ.

$$
\begin{aligned}
p(\mathbf{x}|\mu) &= f(\mathbf{x}|\mu) = \prod_{i=1}^{n} \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\sum_{i=1}^{n} \frac{1}{2\sigma^2}(x_i-\overline{x}+\overline{x}-\mu)^2\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\mu)^2\right]\right\},
\end{aligned}
$$

da $\sum_i (x_i - \overline{x})(\overline{x}-\mu) = (\overline{x}-\mu)\sum_i (x_i-\overline{x}) = 0$ gilt.

Weiters gilt bekanntlich $\overline{X} \sim$ Normal$(\mu, \sigma^2/n)$, also

$$q(\overline{x}|\mu) = (2\pi\sigma^2/n)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2/n}(\overline{x}-\mu)^2\right\}.$$

Wir erhalten somit für $\theta = \mu$

$$\frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} = n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\},$$

was wiederum konstant ist in μ.

Es ist ziemlich unhandlich, die Definition einer suffizienten Statistik unter einem Modell zum Auffinden einer solchen zu verwenden. Der nächste Satz (Halmos und Savage, 1949) erlaubt das Finden einer suffizienten Statistik, indem die gemeinsame Dichte- oder Wahrscheinlichkeitsfunktion der Stichprobe näher betrachtet wird.

Satz 2.1.2: (Faktorisierungssatz) Sei $f(\mathbf{x}|\theta)$ die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion einer Stichprobe $\mathbf{X}$. Eine Statistik $T(\mathbf{X})$ ist eine suffiziente Statistik für θ, genau dann wenn Funktionen $g(t|\theta)$ und $h(\mathbf{x})$ existieren, sodass für alle Stichprobenpunkte $\mathbf{x}$ und für alle Parameter θ gilt

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta) \cdot h(\mathbf{x})\,.$$

Verwende Faktorisierungssatz zur Konstruktion von suffizienten Statistiken:

Beispiel 2.1.2 (Fortsetzung): $X_i \overset{iid}{\sim}$ Normal(μ, σ^2), mit $0 < \sigma^2$ bekannt.

Gemeinsame Dichte

$$f(\mathbf{x}|\mu) = \underbrace{\left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right\}}_{h(\mathbf{x}) \text{ unabhängig von } \mu} \underbrace{\exp\left\{-\frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right\}}_{g(T(\mathbf{x})|\mu) \text{ mit } T(\mathbf{X}) = \overline{X}}\,.$$

Satz 2.1.2 erfordert, dass die Identität $f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$ für alle $\mathbf{x}$ und θ hält. Falls die Menge der $\mathbf{x}$, für die $f(\mathbf{x}|\theta)$ positiv ist, von θ abhängt, dann ist Vorsicht geboten. Die Funktionen $h(\cdot)$ und $g(\cdot)$ müssen derart definiert sein, dass ihr Produkt gerade dort Null ist wo auch $f(\cdot)$ Null ist.

Beispiel 2.1.3: $X_1, \ldots, X_n$ iid aus der diskreten Gleichverteilung auf $\{1, 2, \ldots, \theta\}$. Der Parameter ist daher ein positiver Integer und die Wahrscheinlichkeitsfunktion eines X_i ist

$$f(x|\theta) = \begin{cases} 1/\theta & \text{falls } x = 1, \ldots, \theta \\ 0 & \text{sonst.} \end{cases}$$

Als gemeinsame Wahrscheinlichkeitsfunktion von $\mathbf{x} = (x_1, \ldots, x_n)$ ergibt sich

$$f(\mathbf{x}|\theta) = \begin{cases} \theta^{-n} & x_i \in \{1, \ldots, \theta\} \text{ für } i = 1, \ldots, n \\ 0 & \text{sonst.} \end{cases}$$

Die Einschränkung "$x_i \in \{1, \ldots, \theta\}$ für $i = 1, \ldots, n$" kann ersetzt werden durch "$x_i \in \{1, 2, \ldots\}$ für $i = 1, \ldots, n$, und $\max_i(x_i) \leq \theta$". Der Teil "$x_i \in \{1, 2, \ldots\}$ für $i = 1, \ldots, n$" ist darin unabhängig von θ.

Wir definieren $T(\mathbf{x}) = \mathrm{max}_i(x_i)$

$$
\begin{aligned}
h(\mathbf{x}) &= \begin{cases} 1 & x_i \in \{1, 2, \dots\} \text{ für } i = 1, \dots, n \\ 0 & \text{sonst} \end{cases} \\
g(t|\theta) &= \begin{cases} \theta^{-n} & t \leq \theta \\ 0 & \text{sonst} \end{cases}
\end{aligned}
$$

und die Faktorisierung

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

hält für alle $\mathbf{x}$ und θ.

Die größte Ordnungsstatistik $T(\mathbf{X}) = \mathrm{max}_i(X_i)$ ist somit suffizient für den Parameter θ.

Klarer wird dies noch mittels Indikatoren.

Bezeichne $\mathcal{N} = \{1, 2, \dots\}$ die Menge positiver Integer und sei $\mathcal{N}_\theta = \{1, 2, \dots, \theta\}$. Die gemeinsame Wahrscheinlichkeitsfunktion von $X_1, \dots, X_n$ ist damit

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} \theta^{-1} I_{\mathcal{N}_\theta}(x_i) = \theta^{-n} \prod_{i=1}^{n} I_{\mathcal{N}_\theta}(x_i)\,.$$

Mit $T(\mathbf{x}) = \max_i(x_i)$ gilt

$$\prod_{i=1}^{n} I_{\mathcal{N}_\theta}(x_i) = \left(\prod_{i=1}^{n} I_{\mathcal{N}}(x_i) \right) I_{\mathcal{N}_\theta}(T(\mathbf{x}))$$

und damit

$$f(\mathbf{x}|\theta) = \theta^{-n} I_{\mathcal{N}_\theta}(T(\mathbf{x})) \left(\prod_{i=1}^{n} I_{\mathcal{N}}(x_i) \right)\,.$$

Also ist $T(\mathbf{X}) = \max_i(X_i)$ suffizient für θ.

Manchmal kann aber auch die relevante Stichproben-Information nicht nur in einer Zahl zusammengefasst sein und es sind mehrere Zahlen dazu erforderlich. Dann ist die suffiziente Statistik ein Vektor $T(\mathbf{X}) = (T_1(\mathbf{X}), \ldots, T_r(\mathbf{X}))$. Dies ist oft gerade dann der Fall wenn $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)$ ein Parametervektor ist. Oft gilt dafür $r = s$; es sind jedoch auch verschiedene Dimensionen möglich.

Auch hier ist die Verwendung des Faktorisierungssatzes angebracht, um vektorwertige suffiziente Statistiken zu finden.

Beispiel 2.1.4: $X_1, \ldots, X_n$ iid Normal(μ, σ^2) mit beiden Parametern unbekannt. Definiere $\boldsymbol{\theta} = (\mu, \sigma^2)$. Bereits gezeigt, dass

$$f(\mathbf{x}|\boldsymbol{\theta}) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2\right]\right\}.$$

Alle Teile die von μ oder σ^2 abhängen müssen in der Funktion $g(\cdot)$ enthalten sein.

Die gemeinsame Dichte hängt nur über

$$T_1(\mathbf{x}) = \overline{x} \qquad \text{und} \qquad T_2(\mathbf{x}) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

von der Stichprobe ab. Daher definieren wir $h(\mathbf{x}) = 1$ und

$$g(\mathbf{t}|\boldsymbol{\theta}) = g(t_1, t_2|\mu, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)t_2 + n(t_1 - \mu)^2\right]\right\}.$$

Damit hält die Faktorisierung

$$f(\mathbf{x}|\mu, \sigma^2) = g(T_1(\mathbf{x}), T_2(\mathbf{x})|\mu, \sigma^2)h(\mathbf{x})$$

und $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X})) = (\overline{X}, S^2)$ ist eine suffiziente Statistik für (μ, σ^2).

Dieses Beispiel zeigt, dass es für ein Normalverteilungsmodell angebracht ist, nur das empirische Mittel und die empirische Varianz auszurechnen (sehr gebräuchliche Vorgehensweise).

Die Statistik $(\overline{X}, S^2)$ enthält sämtliche Information über (μ, σ^2) in der Stichprobe.

Jedoch ist die Definition einer suffizienten Statistik von der Modellannahme abhängig. Für ein **alternatives** Verteilungsmodell könnte $(\overline{X}, S^2)$ auch **nicht suffizient** für das Populationsmittel und die Populationsvarianz sein.

Satz 2.1.3: Seien $X_1, \ldots, X_n$ iid Beobachtungen aus einer Dichte- oder Wahrscheinlichkeitsfunktion $f(x|\boldsymbol{\theta})$, welche zur Exponentialfamilie gehört, also mit

$$f(x|\boldsymbol{\theta}) = h(x)\ c(\boldsymbol{\theta})\ \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right)\ ,$$

mit $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $d \leq k$. Dann ist

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

eine suffiziente Statistik für $\boldsymbol{\theta}$.

Beweis: Übung.

Zuvor fanden wir immer **eine** suffiziente Statistik zu jedem Modell. Tatsächlich gibt es aber für jedes Problem mehrere suffiziente Statistiken:

• Es gilt immer, dass die Stichprobe $\mathbf{X}$ selbst eine suffiziente Statistik ist, denn

$$f(\mathbf{x}|\theta) = f(T(\mathbf{x})|\theta)h(\mathbf{x}) \qquad \text{mit } T(\mathbf{x}) = \mathbf{x} \text{ und } h(\mathbf{x}) = 1.$$

• Jede invertierbare Funktion $r(\cdot)$ einer suffizienten Statistik ist auch suffizient. Angenommen $T(\mathbf{X})$ ist suffizient und sei $T^*(\mathbf{x}) = r(T(\mathbf{x}))$ für alle $\mathbf{x}$. Dann existieren wegen des Faktorisierungssatzes Funktionen $g(\cdot)$ und $h(\cdot)$ mit

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}) = g(r^{-1}(T^*(\mathbf{x}))|\theta)h(\mathbf{x})$$

Mit $g^*(t|\theta) = g(r^{-1}(t)|\theta)$ folgt $f(\mathbf{x}|\theta) = g^*(T^*(\mathbf{x})|\theta)h(\mathbf{x})$ und $T^*(\mathbf{X})$ ist somit suffiziente Statistik für θ.

Wir fragen uns daher, ob vielleicht eine suffiziente Statistik irgendwie besser als die anderen sind. Aus Gründen der bestmöglichen Datenreduktion definieren wir:

Definition 2.1.2: Eine suffiziente Statistik $T(\mathbf{X})$ nennt man **minimale suffiziente Statistik**, falls für jede beliebige andere suffiziente Statistik $T'(\mathbf{X})$ gilt, dass $T(\mathbf{X})$ eine Funktion von $T'(\mathbf{X})$ ist.

Bemerkung:

Zu sagen, $T(\mathbf{x})$ ist eine Funktion von $T'(\mathbf{x})$ bedeutet einfach: Gilt $T'(\mathbf{x}) = T'(\mathbf{y})$ dann ist auch $T(\mathbf{x}) = T(\mathbf{y})$.

Alternative Interpretation: Sei $\tau = \{t : t = T(\mathbf{x})\}$. Bezeichne $\{B_{t'} : t' \in \tau'\}$ und $\{A_t : t \in \tau\}$ die Partitionsmengen unter $T'(\mathbf{x})$ und unter $T(\mathbf{x})$, dann gilt unter Definition 2.1.2, dass jedes $B_{t'}$ eine Untermenge eines A_t ist.

Also ist die Partitionierung zu einer **minimalen suffizienten** Statistik die **gröbstmögliche** Partitionierung für eine suffiziente Statistik und die minimale suffiziente Statistik erzielt eine **größtmögliche** Datenreduktion unter allen suffizienten Statistiken.

Beispiel 2.1.5: Seien wiederum $X_1, \ldots, X_n$ iid Normal(μ, σ^2) mit σ^2 bekannt.

Bereits gezeigt, $T(\mathbf{X}) = \overline{X}$ ist suffizient für μ.

Wir könnten auch das Ergebnis aus Beispiel 2.1.4 verwenden (σ^2 bekannt hier) und schließen, dass $T'(\mathbf{X}) = (\overline{X}, S^2)$ eine suffiziente Statistik für μ ist.

Natürlich erzielt $T(\mathbf{X})$ eine größere Datenreduktion als $T'(\mathbf{X})$.

Wir können $T(\mathbf{x})$ als Funktion von $T'(\mathbf{x})$ schreiben. Sei dazu $r(a, b) = a$. Damit ist $T(\mathbf{x}) = \overline{x} = r(\overline{x}, s^2) = r(T'(\mathbf{x}))$.

Da $T(\mathbf{X})$ und $T'(\mathbf{X})$ beides suffiziente Statistiken sind, beinhalten sie dieselbe Information über μ. Die zusätzliche Information über den Wert von S^2 trägt nichts zu unserem Wissen über μ bei, da σ^2 als bekannt angenommen wurde.

Natürlich ist für den Fall von σ^2 **unbekannt** $T(\mathbf{X})$ **keine** suffiziente Statistik und $T'(\mathbf{X})$ enthält mehr Information über die Parameter (μ, σ^2) als $T(\mathbf{X})$.

Es ist wiederum sehr unpraktisch die Definition 2.1.2 zu verwenden um eine minimale suffiziente Statistik zu finden. Wir bräuchten wieder eine Idee einer minimalen suffiziente Statistik $T(\mathbf{X})$ und müssten dafür dann die Bedingungen in der Definition prüfen.

Das folgende Resultat von Lehmann und Scheffé (1950) liefert eine einfachere Methode um eine minimale suffiziente Statistik zu finden:

Satz 2.1.4: Sei $f(\mathbf{x}|\theta)$ die Dichte oder Wahrscheinlichkeitsfunktion einer Stichprobe $\mathbf{X}$. Wir nehmen an, dass eine Funktion $T(\mathbf{x})$ existiert, für die der Quotient $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ als Funktion in θ betrachtet konstant ist genau dann wenn $T(\mathbf{x}) = T(\mathbf{y})$ für zwei beliebige Stichprobenpunkte $\mathbf{x}$ und $\mathbf{y}$ gilt. Dann ist $T(\mathbf{X})$ eine minimale suffiziente Statistik für θ.

Beispiel 2.1.6: Seien $X_1, \ldots, X_n$ iid Normal(μ, σ^2), und μ und σ^2 unbekannt. Seien $\mathbf{x}$ und $\mathbf{y}$ zwei Stichprobenpunkte und seien $(\overline{x}, s^2_{\mathbf{x}})$ und $(\overline{y}, s^2_{\mathbf{y}})$ ihre empirischen Mittel und Varianzen. Dann folgt mit Beispiel 2.1.4

$$\begin{aligned}
\frac{f(\mathbf{x}|\mu,\sigma^2)}{f(\mathbf{y}|\mu,\sigma^2)} &= \frac{(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[n(\overline{x}-\mu)^2+(n-1)s^2_{\mathbf{x}}\right]\right\}}{(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[n(\overline{y}-\mu)^2+(n-1)s^2_{\mathbf{y}}\right]\right\}} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left[n(\overline{x}^2-\overline{y}^2)-2n\mu(\overline{x}-\overline{y})+(n-1)(s^2_{\mathbf{x}}-s^2_{\mathbf{y}})\right]\right\}.
\end{aligned}$$

Dieser Term ist genau dann konstant in den Parametern μ und σ^2 wenn $\overline{x} = \overline{y}$ und $s^2_{\mathbf{x}} = s^2_{\mathbf{y}}$. Deshalb folgt mit Satz 2.1.4, dass $(\overline{X}, S^2)$ eine minimale suffiziente Statistik ist für die Parameter (μ, σ^2).

Falls der Bereich der $\mathbf{x}$ Werte, auf dem die Dichte- oder Wahrscheinlichkeitsfunktion positiv ist, vom Parameter θ abhängt, dann müssen Nenner und Zähler im Quotienten des Satzes 2.1.4 genau für diese Werte von θ positiv sein.

Beispiel 2.1.7: Seien $X_1, \ldots, X_n$ iid stetig gleichverteilt auf $(\theta, \theta+1)$, $\theta \in \mathbb{R}$. Die gemeinsame Dichte der Stichprobe ist

$$f(\mathbf{x}|\theta) = \begin{cases} 1 & \theta < x_i < \theta + 1,\ i = 1, \ldots, n \\ 0 & \text{sonst.} \end{cases}$$

Dies kann auch geschrieben werden als

$$f(\mathbf{x}|\theta) = \begin{cases} 1 & \max_i(x_i) - 1 < \theta < \min_i(x_i) \\ 0 & \text{sonst.} \end{cases}$$

Für 2 beliebige Punkte $\mathbf{x}$ und $\mathbf{y}$ ist Nenner und Zähler des Quotienten für denselben Wert von θ genau dann positiv, wenn $\min_i(x_i) = \min_i(y_i)$ und $\max_i(x_i) = \max_i(y_i)$. Falls Minima & Maxima gleich sind, ist der Quotient 1.

Sei $X_{(1)} = \min(X_i)$ und $X_{(n)} = \max(X_i)$, dann haben wir $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ als minimale suffiziente Statistik für θ. Hier ist die Dimension der minimalen suffizienten Statistik sogar größer als die Dimension der Parameter.

Eine minimale suffiziente Statistik ist **nicht eindeutig**. Jede invertierbare Funktion einer minimalen suffizienten Statistik ist selbst minimal suffizient. So ist für die Gleichverteilung auf $(\theta, \theta + 1)$ in Beispiel 2.1.7 auch beispielsweise die Statistik

$$T'(\mathbf{X}) = \big(X_{(n)} - X_{(1)}, \tfrac{1}{2}(X_{(1)} + X_{(n)})\big)$$

minimal suffizient. Für die Normal(μ, σ^2) Verteilung in Beispiel 2.1.6 ist auch

$$T''(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$$

minimal suffizient.