Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

Kapitel IV - Stichprobenfunktionen (Statistiken)

Induktive Statistik

Prof. Dr. W.-D. Heller
Hartwig Senska
Carlo Siebenschuh

Aufgabe der schließenden Statistik:

Festlegung einer Entscheidung auf der Grundlage eines Stichprobenergebnisses

Entscheidungsverfahren:

Zu jedem möglichen Stichprobenergebnis wird die zugehörige Entscheidung bestimmt.

$\Rightarrow$ Entscheidungsfunktion $\delta : \mathcal{X} \to \mathcal{A}$

$\mathcal{X}$ Stichprobenraum,

A Aktionenraum (Menge der Entscheidungsmöglichkeiten)

gesucht: “beste“ Entscheidungsfunktion

(benötigt wird dazu ein Bewertungskriterium s.u.)

Problem: $\mathcal{X} \subset \mathbb{R}^n$ ist i.a. sehr umfangreich, d.h. es gibt sehr viele, teilweise auch wenig sinnvolle Entscheidungsfunktionen.

Ansatz: Das Stichprobenergebnis enthält auch Informationen, die für die Entscheidung unwesentlich sind, z.B. die Reihenfolge der Messwerte bei einer Stichprobe mit Zurücklegen.

$\hookrightarrow$ Reduktion der Daten auf den für die Entscheidung wesentlichen Kern durch eine Auswertung des Stichprobenergebnisses
(z.B. mit Methoden der deskriptiven Statistik)

Beispiele:

1. Kontrolle einer Warenpartie:
   Stichprobenergebnis $(x_1, x_2, ..., x_n)$ mit

$$x_i = \begin{cases} 0 & i - \text{te Stichprobeneinheit ist gut} \\ 1 & i - \text{te Stichprobeneinheit ist schlecht} \end{cases}$$

   Auswertung:

$$\text{Anzahl schlechter Einheiten} = \sum_{i=1}^{n} x_i$$

   (Oder: Anzahl guter Teile, Stichprobenausschussanteil, ...)

2. Justierung einer Maschine:

   $x_1, x_2, ..., x_n$ Messwerte

$$\frac{1}{n}\sum_{i=1}^{n} x_i \text{ Stichprobenmittelwert}$$

   ist Kandidat für ein sinnvolles Auswertungsverfahren.

Stichprobenfunktion:

$\mathcal{X}$ Stichprobenraum einer Stichprobe $X_1, X_2, ..., X_n$

$T : \mathcal{X} \to \mathbb{R}^k$ heißt **Stichprobenfunktion**

$T(X_1, X_2, ..., X_n) = T(X)$ ist dann Zufallsvariable

Beispiele für Stichprobenfunktionen:

1. $T(x_1, x_2, ..., x_n) = x_3$ bzw. x_i für $i \in \{1, 2, ..., n\}$.
   T erhält in der Regel nicht die gesamte Information der Stichprobe, da nur eines der Stichprobenergebnisse benutzt wird.
2. $T(x_1, \dots, x_n) = (x_{(1)}, \dots, x_{(n)})$, d.h. die Stichprobenergebnisse werden ihrer Größe nach geordnet ($x_{(1)}, \dots, x_{(n)}$ ist die geordnete Urliste zu $x_1, \dots, x_n$). T heißt **Ordnungsstatistik** und erhält bei einer Stichprobe mit Zurücklegen alle relevanten Informationen, da in diesem Fall die Reihenfolge der Ergebnisse keine Bedeutung hat.
   Aber: Keine Datenreduktion.

3. $T(x_1,\ldots,x_n) = x_z$, x_z Zentralwert (Median) von $x_1, x_2, ..., x_n$, heißt **Stichprobenmedian(-zentralwert)**.

4. $T(x_1,\ldots,x_n)$ mit
$$T(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$
heißt **Stichprobenmittelwert**.

Beispiele 3 und 4 sind Kandidaten für Auswertungen bei Untersuchungen, die mit Lageparametern zu tun haben.

5. $T(x_1, \ldots, x_n)$ mit

$$T(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2, \ \bar{x} = \text{Stichprobenmittelwert},$$

heißt **Stichprobenvarianz**.

6. $T(x_1, \ldots, x_n)$

$$T(x_1, \ldots, x_n) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \ \bar{x} = \text{Stichprobenmittelwert},$$

heißt **korrigierte Stichprobenvarianz**.

Beispiele 5 und 6 kommen bei Untersuchungen zu Streuungsparametern einer Verteilung in Betracht (aber auch die Stichprobenspannweite, die mittlere absolute Abweichung der Stichprobe, der Quartilsabstand der Stichprobe, ...).

Frage: Welche Auswertungen sind sinnvoll? Welche Stichprobenfunktionen dürfen verwendet werden, ohne gute Entscheidungsverfahren auszuschließen?

Antwort: T sollte alle entscheidungsrelevanten Informationen der Stichprobe über die Größe (Zustand) erhalten.

Wie kann dies überprüft werden?

1. Kontrolle einer Warenpartie:

   Intuition: Anzahl der schlechten Teile in der Stichprobe enthält alle Informationen über den Ausschussanteil, d.h. zwei Stichprobenergebnisse

   $$x = (x_1, \dots, x_n) \quad \text{und} \quad x' = (x'_1, \dots, x'_n)$$

   mit übereinstimmender Anzahl schlechter Teile unterscheiden sich nicht bezüglich ihrer Information über den Ausschussanteil.

   Ist dies richtig?

Dazu:

Wahrscheinlichkeitsverteilung der Stichprobe $X_1, \ldots, X_n$ auf Teilbereich
mit k schlechten Teilen: $T(X_1, \ldots, X_n) = k$

$\hookrightarrow$ Bedingte Wahrscheinlichkeitsverteilung unter der Bedingung:

$$T(X_1, \ldots, X_n) = k$$

$$
\begin{aligned}
P(X = x \mid T(X) = k) &= P\left(X_1 = x_1, \ldots, X_n = x_n \;\middle|\; \sum_{i=1}^{n} X_i = k\right) \\
&= \frac{P(X_1 = x_1, \ldots, X_n = x_n \text{ und } \sum\limits_{i=1}^{n} X_i = k)}{P(\sum\limits_{i=1}^{n} X_i = k)} \\
&= \begin{cases} 0 & \sum\limits_{i=1}^{n} x_i \neq k \\[2ex] \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{\binom{n}{k} p^k (1-p)^{n-k}} & \sum\limits_{i=1}^{n} x_i = k \end{cases}
\end{aligned}
$$

$X_1, X_2, \ldots, X_n$ unabhängig: Dann

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n)$$

Mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$

$$\prod_{i=1}^{n} P(X_i = x_i) = p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n - \sum_{i=1}^{n} x_i}$$

Damit:

$$P\left(X = x \;\middle|\; \sum_{i=1}^{n} X_i = k\right) = \begin{cases} 0 & \sum\limits_{i=1}^{n} x_i \neq k \\ \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} & \sum\limits_{i=1}^{n} x_i = k \end{cases}$$

$$= \begin{cases} 0 & \sum\limits_{i=1}^{n} x_i \neq k \\ \frac{1}{\binom{n}{k}} & \sum\limits_{i=1}^{n} x_i = k \end{cases}$$

Die bedingte Wahrscheinlichkeitsverteilung hängt also nicht von p ab

die Variation des Stichprobenergebnisses im Teilbereich mit $T(X_1, \ldots, X_n) = k$ ist also unbeeinflusst vom Ausschussanteil p

und damit auch ohne Information über p.

2. Fehleranzahlen

   Bei der Produktion von Teilen können ein oder mehrere Fehler entstehen (z.B. Lackierfehler an Automobilkarossen). Die Anzahl der Fehler an einer Produkteinheit entspreche einer Zufallsvariable Y.

   Y sei Poisson-verteilt mit Parameter λ,

   λ sei unbekannt.

   Durch eine Stichprobe soll Information über λ ermittelt werden.

   Stichprobenergebnis: $x_1, x_2, \ldots, x_n \quad (x_i$ ganze Zahl $\geq 0)$

Vermutung:
Wesentliche Information über den Parameter λ des Stichprobenergebnisses
ist die Gesamtzahl der Fehler (oder die durchschnittliche Fehlerzahl).
Zugehörige Stichprobenfunktion ist:

$$T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

Überprüfung der Vermutung:
Stichprobe mit Zurücklegen, d.h. $X_1, X_2, \ldots, X_n$ unabhängig, identisch verteilt wie Y

$$P\left(X = x_1, \ldots, X_n = x_n \;\middle|\; T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i = k\right) = ?$$

$$P\left(X = x_1, \ldots, X_n = x_n \;\middle|\; T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i = k\right)$$

$$= \begin{cases} 0 & \sum\limits_{i=1}^{n} x_i \neq k \\ \frac{P(X=x)}{P(\sum\limits_{i=1}^{n} X_i = k)} & \sum\limits_{i=1}^{n} x_i = k \end{cases}$$

mit

$$
\begin{aligned}
P(X = x) &= P(X_1 = x_1, \ldots, X_n = x_n) \\
&= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n) \\
&= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} \\
&= \prod_{i=1}^{n} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right)
\end{aligned}
$$

$\sum\limits_{i=1}^{n} X_i$ ist Poisson-verteilt mit Parameter $n\lambda$:

$$P(\sum_{i=1}^{n} X_i = k) = \frac{(n\lambda)^k}{k!} e^{-n\lambda}$$

Damit ist:

$$\frac{P(X{=}x)}{P(\sum\limits_{i=1}^{n} X_i{=}k)} = \frac{\prod\limits_{i=1}^{n} (\frac{\lambda^{x_i}}{x_i!} e^{-\lambda})}{\frac{(n\lambda)^k}{k!} e^{-n\lambda}} = \frac{(\prod\limits_{i=1}^{n} \frac{1}{x_i!}) \lambda^{\sum\limits_{i=1}^{n} x_i} e^{-n\lambda}}{\frac{(n\lambda)^k}{k!} e^{-n\lambda}}$$

$$= \frac{\prod\limits_{i=1}^{n} \frac{1}{x_i!}}{\frac{n^k}{k!}}$$

Die bedingte Wahrscheinlichkeitsverteilung

$$P\left(X_1 = x_1, \ldots, X_n = x_n \;\middle|\; T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i = k\right) =$$

$= \frac{k!}{n^k} \prod_{i=1}^{n} \frac{1}{x_i!}$ ist also unabhängig vom Parameter λ. Welche spezielle Stichprobe mit $(x_1, x_2, \ldots, x_n)$ mit Gesamtfehlerzahl k vorliegt, liefert demnach keine Information über λ.

Eine Stichprobenfunktion T enthält also die Information der Stichprobe über den Parameter der Verteilung, wenn die bedingte Wahrscheinlichkeit des Stichprobenvektors mit $T(X) = t$ als Bedingung nicht vom Parameter abhängt. T heißt dann **suffizient** bezüglich des Parameters.

Definition:

Sei Γ der Parameterraum zur Verteilungsannahme einer diskreten Zufallsvariablen Y. $X = (X_1, \ldots, X_n)$ sei eine Stichprobe zu Y mit Stichprobenraum $\mathcal{X}$. Eine Stichprobenfunktion $T : \mathcal{X} \to \mathbb{R}^k$ heißt suffizient bzgl. $\gamma \in \Gamma$, wenn

$$P_\gamma(X = x \mid T(X) = t)$$

für alle $x \in \mathcal{X}$ und $t \in \mathbb{R}^k$ mit $P_\gamma(T(X) = t) \neq 0$ von γ unabhängig ist,
also allein von x und t abhängt.

Folgerung:

Wegen

$$P_\gamma(X = x) = P(X = x \mid T(X) = t) P_\gamma(T(X) = t)$$

ist bei suffizientem T die Wahrscheinlichkeitsverteilung zerlegt in zwei Faktoren, von denen

$$P(X = x \mid T(X) = t)$$

nur von x (t ergibt sich aus $T(x) = t$) und

$P_\gamma(T(X) = t)$ von γ und t abhängt.

Es gilt demnach mit $P(X = x | T(X) = t) = h(x)$ und $P_\gamma(T(X) = t) = g(T(x), \gamma)$

$$(*)\ P_\gamma(X = x) = g(T(x), \gamma) h(x).$$

Aus (*) folgt aber :

$$\begin{aligned}
P_\gamma(X = x | T(X) = t) &= \frac{P_\gamma(X = x, T(X) = t)}{P_\gamma(T(X) = t)} = \frac{g(T(x), \gamma) h(x)}{\sum\limits_{\tilde{x} \in \mathcal{X}\ :\ T(\tilde{x}) = t} P_\gamma(X = \tilde{x})} \\
&= \frac{g(T(x), \gamma) h(x)}{\sum\limits_{\tilde{x} \in \mathcal{X}\ :\ T(\tilde{x}) = t} g(T(\tilde{x}), \gamma) h(\tilde{x})} = \frac{g(t, \gamma) h(x)}{\sum\limits_{\tilde{x} \in \mathcal{X}\ :\ T(\tilde{x}) = t} g(t, \gamma) h(\tilde{x})} \\
&= \frac{h(x)}{\sum\limits_{\tilde{x} \in \mathcal{X}\ :\ T(\tilde{x}) = t} h(\tilde{x})} \text{ hängt nicht von } \gamma \text{ ab}
\end{aligned}$$

Im diskreten Fall ist eine Stichprobenfunktion genau dann suffizient bezüglich einem Parameter γ, wenn sich die Wahrscheinlichkeitsverteilung des Stichprobenvektors entsprechend (*) in Faktoren zerlegen lässt.
(Faktorisierungstheorem von Neyman).

Beispiele:

1. Bernoulliverteilung:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n) &= P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \\ &= p^{\sum\limits_{i=1}^{n} x_i} (1-p)^{n - \sum\limits_{i=1}^{n} x_i} \\ &= g(T(x), p) h(x) \text{ mit } h(x) = 1 \end{aligned}$$

2. Poissonverteilung:

$$
\begin{aligned}
P(X = x) &= P(X_1 = x_1, \ldots, X_n = x_n) \\
&= P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \\
&= \tfrac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \ldots \cdot \tfrac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \prod_{i=1}^{n} (\tfrac{\lambda^{x_i}}{x_i!} e^{-\lambda}) \\
&= \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \tfrac{1}{x_1! x_2! \cdot \ldots \cdot x_n!} = g(T(x), \lambda) h(x)
\end{aligned}
$$

mit $g(T(x), \lambda) = \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}$, $T(x) = \sum_{i=1}^{n} x_i$ und $h(x) = \frac{1}{x_1! x_2! \cdot \ldots \cdot x_n!}$

Im **stetigen** Fall sind die analogen Betrachtungen mit der **Dichtefunktion** durchzuführen.

Beispiel:

1. Exponentialverteilung
   Sei Y exponentialverteilt mit Parameter $\lambda > 0, X_1, X_2, \ldots, X_n$ eine einfache Stichprobe mit Zurücklegen zu Y. Dann gilt

$$
\begin{aligned}
f_{X,\lambda}(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\
&= \lambda^n e^{-\lambda \sum\limits_{i=1}^{n} x_i} = \ldots
\end{aligned}
$$

$$= g(\sum_{i=1}^{n} x_i, \lambda) \cdot h(x)$$

mit $g(T(x), \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$, $T(x) = \sum_{i=1}^{n} x_i$ und $h(x) = 1$.

Zur Berechnung der bedingten Dichte unter der Bedingung

$$T(X) = \sum_{i=1}^{n} X_i = t$$

wird die Verteilung von

$$\sum_{i=1}^{n} X_i$$

benötigt.

Man erhält (s. Übungsaufgabe aus der Wahrscheinlichkeitstheorie) die Erlang-Verteilung mit Stufenzahl n mit der Dichtefunktion

$$f(t) = \begin{cases} \frac{\lambda^n}{(n-1)!}\ t^{n-1} e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

... und damit die bedingte Dichte

$$
\begin{aligned}
f_{(X,\lambda|T(X)=t)}(x) &= \frac{\prod\limits_{i=1}^{n} \lambda e^{-\lambda x_i}}{\frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}} \\
&= \frac{\lambda^n e^{-\lambda \sum\limits_{i=1}^{n} x_i}}{\frac{\lambda^n}{(n-1)!} (\sum\limits_{i=1}^{n} x_i)^{n-1} e^{-\lambda \sum\limits_{i=1}^{n} x_i}} \\
&= (n-1)! \frac{1}{(\sum\limits_{i=1}^{n} x_i)^{n-1}}
\end{aligned}
$$

für $x = (x_1, \ldots, x_n)$ mit $\sum x_i = t$.

Die Dichte lässt sich also analog in zwei Faktoren zerlegen und die bedingte Dichte ist unabhängig vom Parameter. Die Stichprobenfunktion

$$T(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

ist also suffizient bezüglich λ.

2. Normalverteilung

Sei Y $N(\mu, \sigma^2)$-verteilt, also mit der Dichte

$$f_{Y,\mu,\sigma^2}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Dann ist für eine Stichprobe X mit Zurücklegen:

$$\begin{aligned}
f_{X,\mu,\sigma^2}(x_1,\ldots,x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\sum\limits_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{1}{2\sigma^2}(\sum\limits_{i=1}^{n} x_i^2 - 2\mu \sum\limits_{i=1}^{n} x_i + n\mu^2)} \\
&= \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n} x_i^2} e^{\frac{\mu}{\sigma^2}\sum\limits_{i=1}^{n} x_i} e^{-\frac{n\mu^2}{2\sigma^2}}
\end{aligned}$$

Dieser Ausdruck hängt außer von den Parametern μ und σ^2 nur von

$$\sum_{i=1}^{n} x_i^2 \text{ und } \sum_{i=1}^{n} x_i$$

ab. Die zweidimensionale Stichprobenfunktion

$$T(x_1, \ldots, x_n) = (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2) \text{ ist}$$

suffizient bezüglich (μ, σ^2).

In den Beispielen spielt die Exponentialfunktion eine wesentliche Rolle,

da das **Produkt** von Exponentialfunktionen mit der Exponentialfunktion

der **Summe** der Exponenten übereinstimmt:

$$\prod_{i=1}^{n} e^{z_i} = e^{\sum\limits_{i=1}^{n} z_i}$$

Folgerung:

Immer wenn wir die Wahrscheinlichkeiten bzw. die Dichte der Zufallsvariablen als Exponentialfunktion schreiben können, haben wir gute Aussichten, eine suffiziente Statistik zu erhalten.

Beispiele:

1. Poissonverteilung:

$$P_\lambda(Y = y) = \frac{\lambda^y}{y!} \cdot e^{-\lambda} = e^{y \ln \lambda} \frac{1}{y!} \cdot e^{-\lambda}$$

Daraus folgt für eine einfache Stichprobe $X = (X_1, \ldots, X_n)$ mit Zurücklegen

$$\begin{aligned} P_\lambda(X = x) &= \prod_{i=1}^{n} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = \prod_{i=1}^{n} \left(\frac{1}{x_i!} e^{x_i \ln \lambda} e^{-\lambda} \right) \\ &= \left[\prod_{i=1}^{n} \left(\frac{1}{x_i!}\right) \right] e^{\ln \lambda \sum\limits_{i=1}^{n} x_i} e^{-n\lambda} \end{aligned}$$

2. Normalverteilung:

$$f_{\mu,\sigma^2}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{y^2}{2\sigma^2}+\frac{\mu y}{\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} \cdot e^{\frac{-y^2}{2\sigma^2}} \cdot e^{\frac{2\mu y}{2\sigma^2}}$$

Daraus folgt für eine einfache Stichprobe $X = (X_1, \ldots, X_n)$ mit Zurücklegen

$$f_{\mu,\sigma^2}(x) = \prod_{i=1}^{n} f_{\mu,\sigma^2}(x_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} \right)^n \cdot e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} x_i^2} \cdot e^{\frac{\mu}{\sigma^2} \sum\limits_{i=1}^{n} x_i}$$

Nennen wir den Parameter γ, so ist in beiden Beispielen die Wahrscheinlichkeit
bzw. die Dichte von der Form

$$(+) \qquad a(\gamma)h(y)e^{\sum\limits_{k=1}^{r} b_k(\gamma)\tau_k(y)}$$

1. Poissonverteilung: $\quad r = 1 \quad e^{-\lambda}\frac{1}{y!}e^{yln\lambda} \quad a(\gamma)h(y)e^{b(\gamma)\tau(y)}$

mit

$$a(\gamma) = e^{-\lambda}, h(y) = \frac{1}{y!}, b(\gamma) = \ln\lambda, \tau(y) = y$$

2. Normalverteilung

$r = 2 \qquad a(\gamma)h(y)e^{b_1(\gamma)\tau(y)+b_2(\gamma)\tau_2(y)}$

mit

$$
\begin{array}{llll}
a(\mu,\sigma^2) = \frac{1}{\sqrt{\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} & , & h(y) = \frac{1}{\sqrt{2\pi}}, \\
\\
b_1(\mu,\sigma^2) = -\frac{1}{2\sigma^2} & , & \tau_1(y) = y^2, \\
b_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2} & , & \tau_2(y) = y.
\end{array}
$$

Aus der Darstellung (+) für die Wahrscheinlichkeit bzw. Dichte einer Zufallsvariablen ergibt sich unmittelbar für die Wahrscheinlichkeit bzw. Dichte einer Stichprobe mit Zurücklegen

$$
\begin{aligned}
\prod_{i=1}^{n}\left(a(\gamma)h(x_i)e^{\sum\limits_{k=1}^{r} b_k(\gamma)\tau_k(x_i)}\right) &= a(\gamma)^n\left[\prod_{i=1}^{n} h(x_i)\right] e^{\sum\limits_{k=1}^{r} b_k(\gamma)\sum\limits_{i=1}^{n}\tau_k(x_i)} \\
&= g(T(x),\gamma)\tilde{h}(x) \\
\text{mit } T(x) &= \left(\sum_{i=1}^{n}\tau_1(x_i),\ldots,\sum_{i=1}^{n}\tau_r(x_i)\right), \\
g(T(x),\gamma) &= a(\gamma)^n e^{\sum\limits_{k=1}^{r} b_k(\gamma)\sum\limits_{i=1}^{n}\tau_k(x_i)}, \quad \tilde{h}(x) = \prod_{i=1}^{n} h(x_i)
\end{aligned}
$$

Falls bei der Verteilung der Zufallsvariable Y eine Darstellung der Form $(+)$ vorliegt, nennen wir die Verteilungsannahme der Zufallsvariable eine **Exponentialfamilie**.

Exponentialfamilie:

- Y diskret

$$P_\gamma(Y = y) = a(\gamma)h(y)e^{\sum_{j=1}^{r} b_j(\gamma)\tau_j(y)}$$

- Y stetig

$$f_{Y,\gamma}(y) = a(\gamma)h(y)e^{\sum_{j=1}^{r} b_j(\gamma)\tau_j(y)}$$

Folgerung:

$$T(x) = (\sum_{i=1}^{n} \tau_1(x_i), \ldots, \sum_{i=1}^{n} \tau_r(x_i))$$

ist suffiziente Stichprobenfunktion bezüglich γ.

Beispiel: Gammaverteilung

Dichte der Gammaverteilung ist

$$f_{\alpha,\lambda}(y) = \begin{cases} \frac{1}{\Gamma(\alpha)}\lambda^{\alpha} y^{\alpha-1} e^{-\lambda y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

$\Gamma(\alpha)$ ist dabei die Gammafunktion an der Stelle α;

für eine natürliche Zahl n gilt $\Gamma(n) = (n-1)!$

Hinweis: Für $\alpha = n$ erhalten wir also die Erlang-Verteilung mit Stufenzahl n.

Mit

$$
\begin{aligned}
&a(\alpha,\lambda) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}, \quad h(y) = \begin{cases} 1 & y \geq 0 \\ 0 & y < 0 \end{cases} \\
&b_1(\alpha,\lambda) = \alpha - 1, \quad \tau_1(y) = \ln(y), \\
&b_2(\alpha,\lambda) = -\lambda, \quad \tau_2(y) = y
\end{aligned}
$$

gilt

$$
f_{\alpha,\lambda}(y) = a(\alpha,\lambda)h(y)e^{b_1(\alpha,\lambda)\tau_1(y) + b_2(\alpha,\lambda)\tau_2(y)}
$$

Damit ist

$$T(x) = (\sum_{i=1}^{n} \ln(x_i), \sum_{i=1}^{n} x_i)$$

suffizient bezüglich α und λ.

Sei T eine suffiziente Stichprobenfunktion, so sollte die Entscheidung (Schlussfolgerung) aus dem Stichprobenergebnis nur vom Auswertungsergebnis bei T abhängen, d.h. sind

$$x = (x_1, \ldots, x_n) \text{ und } \tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n)$$

zwei Stichprobenergebnisse mit

$$T(x) = T(\tilde{x}),$$

so sollten die Entscheidungen bei x und $\tilde{x}$, übereinstimmen. Damit wird die Menge der möglichen Entscheidungsverfahren stark reduziert, falls eine geeignete suffiziente Statistik benutzt wird.

$$\begin{tikzcd}
x, x' \arrow[rrr] \arrow[ddd] & & & \delta(x) = \delta'(T(x)) = \delta'(T(x')) \\
& \mathcal{X} \arrow[r, "\delta"] \arrow[d, "T"'] & A & = \delta(x') \\
& \mathcal{X}^{T} \arrow[ur, "\delta'"'] & & \\
T(x) = T(x') \arrow[uuurrr, "\delta = \delta' \circ T"'] & & &
\end{tikzcd}$$

Reduktion auf die wesentliche Information zur

Entscheidungsfindung.

Sei $\mathcal{X}^{\mathcal{T}}$ die Menge der möglichen Auswertungsergebnisse

bei der Stichprobenfunktion T, also

$\mathcal{X}^{T} = \{T(x) | x \in \mathcal{X}\}$

so ist $\delta^{T} : \mathcal{X}^{\mathcal{T}} \to \mathcal{A}$ eine Entscheidungsfunktion mit Benutzung von T,

d.h. zum Stichprobenergebnis x wird zunächst die Auswertung

$T(x)$ berechnet und abhängig von diesem Auswertungs-

ergebnis $T(x)$ die Entscheidung $\delta^{T}(T(x))$ getroffen.

Behauptung:

Wird eine Stichprobenfunktion T aus einer Darstellung der Wahrscheinlichkeits -

verteilung von Y als Exponentialfamilie gewonnen, so sind

zwei Entscheidungsfunktionen mit Benutzung von T durch

ihren Erwartungswert nahezu vollständig festgelegt.

Beispiel-Poisson-Verteilung:

$$T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

ist suffizient bezüglich dem Parameter $\lambda > 0$.
Seien δ_1^T und δ_2^T zwei Entscheidungsfunktionen mit Benutzung von T.
Erwartungswert von δ_i^T ist

$$\begin{aligned}
E_\lambda(\delta_i^T(T(X))) &= \sum_{k=0}^{\infty} \delta_i^T(k) P_\lambda(T(X) = k) = \sum_{k=0}^{\infty} \delta_i^T(k) P_\lambda(\sum_{j=1}^{n} X_j = k) \\
&= \sum_{k=0}^{\infty} \delta_i^T(k) \frac{(n\lambda)^k}{k!} e^{-n\lambda}
\end{aligned}$$

Stimmen die Erwartungswerte von δ_1 und δ_2 überein für alle $\lambda > 0$,
so gilt

$$0 = E_\lambda(\delta_1^T(T(X))) - E_\lambda(\delta_2^T(T(X))) = \sum_{k=0}^{\infty}(\delta_1^T(k) - \delta_2^T(k))\frac{(n\lambda)^k}{k!}e^{-n\lambda}$$
$$= e^{-n\lambda}\sum_{k=0}^{\infty}(\delta_1^T(k) - \delta_2^T(k))\frac{(n\lambda)^k}{k!}$$

Da $e^{-n\lambda} \neq 0$ ist, ist dies gleichwertig mit

$$\sum_{k=0}^{\infty}(\delta_1^T(k) - \delta_2^T(k))\frac{(n\lambda)^k}{k!} = 0 \text{ für alle } \lambda > 0$$
$$\Leftrightarrow \delta_1^T(k) = \delta_2^T(k) \text{ für alle k.}$$

Eine Stichprobenfunktion T mit der Eigenschaft, dass je zwei Funktionen δ_1^T und δ_2^T mit übereinstimmendem Erwartungswert auch selbst übereinstimmen,

heißt **vollständig**.

Definition:
Sei X eine Stichprobe zu Y, $\mathcal{X}$ der Stichprobenraum.
Y habe eine parametrische Verteilungsannahme mit Parameter γ aus dem Parameterraum Γ.
Eine Stichprobenfunktion(Statistik) T heißt **vollständig** bezüglich γ, wenn für je zwei Funktionen δ_1^T und δ_2^T aus

$$E_\gamma(\delta_1^T(T(X))) = E_\gamma(\delta_2^T(T(X))) \text{ für alle } \gamma \in \Gamma$$

folgt, dass

$$P_\gamma(\delta_1^T(T(X)) = \delta_2^T(T(X))) = 1$$

gilt.

Satz:

Sei X eine Stichprobe zu Y, deren Verteilung sich als Exponentialfamilie schreiben lässt, d.h. es gilt

$$\left.\begin{array}{ll} Y \text{diskret}: & P_\gamma(Y=y) \\ Y \text{stetig}: & f_{Y,\gamma}(y) \end{array}\right\} \quad = \quad a(\gamma)h(y)e^{\sum\limits_{j=1}^{r} b_j(\gamma)\tau_j(y)}$$

dann ist

$$T(x_1,\ldots,x_n) = \left(\sum_{i=1}^{n}\tau_1(x_i),\ldots,\sum_{i=1}^{n}\tau_r(x_i)\right)$$

eine suffiziente und vollständige Statistik bezüglich γ, wenn

$$B = \{(b_1(\gamma),\ldots,b_r(\gamma)) \mid \gamma \in \Gamma\} \subset \mathbb{R}^r$$

einen offenen Quader im $\mathbb{R}^r$ enthält.

Folgerungen:

Sei $X_1, X_2, \ldots, X_n$ eine einfache Stichprobe mit Zurücklegen zu Y. Dann

ist

$$T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

suffizient und vollständig, ...

... falls y

a) Bernoulli-verteilt mit Parameter $p \in [0, 1]$,
(Übungsaufgabe: Man stelle die Wahrscheinlichkeitsverteilung von Y als Exponentialfamilie dar.)

b) binomialverteilt mit Parameter $p \in [0, 1]$,

c) Poisson-verteilt mit Parameter $\lambda > 0$,

d) exponentialverteilt mit Parameter $\lambda > 0$.

Beispiel-Normalverteilung:

$$\begin{aligned} f_{\mu,\sigma^2}(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{y^2}{2\sigma^2}+\frac{\mu y}{\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} \cdot e^{\frac{-y^2}{2\sigma^2}} \cdot e^{\frac{2\mu y}{2\sigma^2}} \end{aligned}$$

1. μ und σ^2 unbekannt (und entscheidungsrelevant):

$$a(\mu, \sigma^2) = \frac{1}{\sqrt{\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} \quad , \quad h(y) = \frac{1}{\sqrt{2\pi}},$$

$$b_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2} \quad , \quad \tau_1(y) = y^2,$$

$$b_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2} \quad , \quad \tau_2(y) = y.$$

ergibt eine Darstellung als Exponentialfamilie.

$$T(x_1,\dots,x_n) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$$

ist eine suffiziente Statistik bezüglich (μ,σ^2). Sie ist auch vollständig bezüglich (μ,σ^2), da

$$\begin{aligned} B &= \{(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}) | \mu \in \mathbb{R}, \sigma^2 > 0\} \\ &= \{(\beta_1,\beta_2) | \beta_1 < 0, \beta_2 \in \mathbb{R}\} = \mathbb{R}_- \times \mathbb{R} \end{aligned}$$

einen offenen Quader im $\mathbb{R}^2$ enthält.

2. μ unbekannt (und entscheidungsrelevant), σ^2 fest:

Der Parameterraum ist jetzt $\Gamma = \mathbb{R}$. Mit

$$\begin{array}{ll} a(\mu) = \frac{1}{\sqrt{\sigma^2}} \cdot e^{-\frac{\mu^2}{2\sigma^2}} & , \quad h(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{y^2}{2\sigma^2}}, \\ b(\mu) = \frac{\mu}{\sigma^2} & , \quad \tau_2(y) = y \end{array}$$

erhalten wir eine Darstellung als Exponentialfamilie, so dass

$$T(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i$$

suffizient bezüglich μ ist. T ist auch vollständig, da

$$B = \{\frac{\mu}{\sigma^2} | \mu \in \mathbb{R}\} = \mathbb{R} \qquad \text{(enthält offenen Quader)}$$

3. σ^2 unbekannt (und entscheidungsrelevant), μ fest:

$$\begin{array}{lcl} a(\sigma^2) = \frac{1}{\sqrt{\sigma^2}} & , & h(y) = \frac{1}{\sqrt{2\pi}}, \\ b(\sigma^2) = -\frac{1}{2\sigma^2} & , & \tau(y) = (y-\mu)^2 \end{array}$$

ergibt jetzt eine Darstellung als Exponentialfamilie für den Parameter σ^2. Wegen

$$B = \{-\frac{1}{2\sigma^2} | \sigma^2 > 0\} = \{x \in \mathbb{R} | x < 0\}$$

ist damit

$$T(x_1, \ldots, x_n) = \sum_{i=1}^{n} (x_i - \mu)^2$$

suffizient und vollständig bezüglich σ^2.