

Wahrscheinlichkeitstheorie und Statistik

12. Woche

Dávid Tóth (BME SZIT)

18. November 2024

Einführung in die Statistik

Grundbegriffe der Statistik

Das Ziel der Statistik

- In der Statistik stehen Daten im Vordergrund, aus denen mittels wahrscheinlichkeitstheoretischer Methoden Schlüsse gezogen werden sollen.
- Die deskriptive oder beschreibende Statistik beschäftigt sich mit der Darstellung von empirischen Daten (durch Tabellen, Kennzahlen, Grafiken, usw.).
- Unser Thema ist die induktive oder schließende Statistik, in der Daten benutzt werden, um von einer Teilgesamtheit auf eine Gesamtheit zu schließen, oder um eine Theorie zu überprüfen.

Grundbegriffe der Statistik

- Unser Ausgangspunkt ist immer eine Beobachtung, d.h. eine Messung oder eine Stichprobe.
- Z.B.: Körpergröße von Menschen, Wartezeit einiger Kunden, usw.
- Die zu beobachtenden Objekte bilden die *Grundgesamtheit*.
- Die Ergebnisse der Beobachtung können als zufällig betrachtet werden.
- Z.B.: man betrachtet die Körpergröße einer zufällig gewählten Person der Population, oder die Wartezeit eines Kunden ist von vielen Faktoren beeinflußt und kann als eine zufällige Quantität gehandelt werden, usw.

Grundbegriffe der Statistik

- Wir werden immer annehmen, dass eine zu messende Größe durch eine Zufallsvariable beschrieben wird, deren Verteilung im Prinzip die vollständige Information über diese Größe enthält.
- Falls diese Variable diskret ist, dann sagt man, dass sie ein *diskretes Merkmal* beschreibt. Falls sie absolut stetig ist und dann im Prinzip sämtliche Werte eines Intervalles von $\mathbb{R}$ annehmen kann, beschreibt sie ein *stetiges Merkmal*.
- Solche Zufallsvariablen (oder deren Verteilungen) können in Wirklichkeit nicht direkt beobachtet werden, also können wir nicht die vollständige Information bekommen.
- Unsere Stichproben werden aber als *Realisierungen* (also Werten) von solchen (identisch verteilten und unabhängigen) Zufallsvariablen betrachtet.

Grundbegriffe der Statistik

Das allgemeine Modell:

- Die Objekte der Grundgesamtheit sind Elemente eines Wahrscheinlichkeitsraumes.
- Die zu messende Größe wird durch eine Zufallsvariable Y beschrieben.
- Die Stichproben werden durch unabhängigen und identisch verteilten Zufallsvariablen modelliert, die die Verteilung von Y haben.
- Bei den konkreten Beispielen arbeiten wir mit Realisierungen dieser Variablen, also mit Datenvektoren $(x_1, \dots, x_n)$, deren Koordinaten als Werte dieser Variablen betrachtet werden.
- Aber im Allgemeinen arbeiten wir mit Variablen statt Datenvektoren, damit unsere Ergebnisse für jede Realisierung (also für jedes Beispiel) gültig sind.

Grundbegriffe der Statistik

Definition. Sei Y eine Zufallsvariable, die eine zu messende Quantität beschreibt. Eine *einfache Stichprobe* vom Umfang (oder von der Größe) n zu Y ist ein Vektor $\mathbf{X} = (X_1, \dots, X_n)$ von unabhängigen Zufallsvariablen, deren Verteilungen mit der Verteilung von Y übereinstimmen.

Oft wird die Variable Y nicht erwähnt und wir sagen, dass der Vektor $\mathbf{X}$ eine (einfache) Stichprobe vom Umfang n ist.

Grundbegriffe der Statistik

Zwar ist die Verteilung der Stichprobenvariablen (in den meisten Fällen) nicht bekannt, aber man kann oft annehmen, dass sie zu einer Familie von Verteilungen gehören (z.B. normalverteilt oder exponentialverteilt sind).

In vielen solchen Fällen ist die Verteilung durch eine oder mehrere Parameter $\vartheta \in \Theta$ bestimmt, wo $\Theta \subset \mathbb{R}^d$ die Menge der möglichen Werten der Parameter ist.

Also bestimmt jeder Vektor $\vartheta \in \Theta$ eine Verteilung eindeutig:

- die zugehörige Verteilungsfunktion wird durch F_{ϑ} bezeichnet,
- falls die Verteilung diskret bzw. absolut stetig ist, dann wird die zugehörige Wahrscheinlichkeitsfunktion bzw. Dichtefunktion durch f_{ϑ} bezeichnet.

Grundbegriffe der Statistik

Beispiele.

1. Auf der Straße finden wir eine Münze, von der nicht bekannt ist, ob sie regulär ist, also gibt sie einen Kopf mit Wahrscheinlichkeit $\vartheta \in \Theta = [0; 1] \subset \mathbb{R}$.

Der Parameter ϑ bestimmt die Verteilung der Würfe, denn

$$f_{\vartheta}(\text{Zahl}) = \mathbb{P}(\text{Zahl}) = 1 - \mathbb{P}(\text{Kopf}) = 1 - f_{\vartheta}(\text{Kopf}) = 1 - \vartheta$$

gilt.

Grundbegriffe der Statistik

Beispiele.

2. Von der Anzahl der Unfälle an Bahnübergängen in einem bestimmten Monat in Ungarn können wir annehmen, dass sie in guter Näherung Poisson-verteilt ist, da die Anzahl der Fahrer groß ist und jeder von ihnen einen solchen Unfall (mehr oder weniger) unabhängig von anderen Fahrern mit einer kleinen Wahrscheinlichkeit erleidet.

Diese Verteilung wird dann durch einen Parameter $\vartheta \in \Theta = (0, \infty) \subset \mathbb{R}$ bestimmt. Also gilt

$$f_{\vartheta}(k) = e^{-\vartheta} \frac{\vartheta^k}{k!}$$

für alle $k \in \mathbb{N}$.

Beispiele.

3. n Studentinnen in der BME werden uniform zufällig und unabhängig voneinander gewählt, sei X_i die Körpergröße der i -ten gewählten Studentin.

Dann sind die Stichprobenvariablen $X_1, \dots, X_n$ unabhängig und (approximativ) normalverteilt mit (unbekannten) Parametern $(\mu, \sigma^2) \in \mathbb{R} \times (0; \infty) \subset \mathbb{R}^2$. Die zugehörige Dichtefunktion ist

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

Grundbegriffe der Statistik

Bemerkung. In unserem Modell arbeiten wir auf demselben Wahrscheinlichkeitsraum Ω definierten Zufallsvariablen. In der Definition der Zufallsvariablen ist aber das Wahrscheinlichkeitsmaß nicht bedeutend, bei einer Zufallsvariablen X verlangen wir nur, dass die Menge $\{X \leq t\} \subset \Omega$ auch ein Ereignis ist. Das bedeutet, dass X für alle auf Ω definierten Wahrscheinlichkeitsmaßen $\mathbb{P}$ eine Zufallsvariable ist.

Grundbegriffe der Statistik

Das allgemeine Modell:

- Die Grundgesamtheit wird durch eine Ergebnismenge und eine σ -Algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$ (durch einen *messbaren Raum*) modelliert.
- Wir nehmen an, dass ein Wahrscheinlichkeitsmaß $\mathbb{P}_0$ auf Ω existiert, das in einer Menge

$$\{\mathbb{P}_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^d\}$$

von auf Ω definierten Wahrscheinlichkeitsmaßen enthalten ist. Unser Ziel ist: das Maß $\mathbb{P}_0$ zu finden, oder mindestens zu approximieren.

Das allgemeine Modell:

- Eine einfache Stichprobe vom Umfang n ist ein Vektor $\mathbf{X} = (X_1, \dots, X_n)$ von unabhängigen und identisch verteilten Zufallsvariablen mit Verteilungsfunktion $F_0(t) = \mathbb{P}_0(X \leq t)$.
- Dieser Variablen ordnet man aber eine Verteilungsfunktion $F_\vartheta(t) = \mathbb{P}_\vartheta(X_i \leq t)$ (also eine Verteilung) für alle Parameter $\vartheta \in \Theta$. Der entsprechende Erwartungswert bzw. die Varianz werden als $\mathbb{E}_\vartheta$ bzw. $\mathbb{D}_\vartheta^2$ bezeichnet, und es wird angenommen, dass die Variablen bezüglich aller Maßen $\mathbb{P}_\vartheta$ unabhängig sind.
- Wir (können und) werden annehmen, dass der Parameter ϑ die Funktion F_ϑ und auch das Maß $\mathbb{P}_\vartheta$ eindeutig bestimmt, also sind die Approximationen von F_0 und $\mathbb{P}_0$ gleichbedeutend.

Grundbegriffe der Statistik

Bemerkung. In konkreten Fällen arbeiten wir mit konkreten Stichprobenwerten (also mit Datenvektoren). Diese werden als Werte von Stichprobenvariablen betrachtet und heißen Realisierungen von Stichproben.

Aus technischen Gründen betrachten wir nicht alle möglichen Datenvektoren aus der Wertemenge $\text{ran } \mathbf{X} = \text{ran } (X_1, \dots, X_n)$ einer Stichprobe als Realisierungen:

Definition. Nehmen wir an, dass eine Familie von diskreten oder absolut stetigen Verteilungen durch eine Parametermenge $\Theta \subset \mathbb{R}^d$ bestimmt ist. Dann ist die Menge der *wesentlichen Werte* der zum Parameter $\vartheta \in \Theta$ gehörenden Verteilung durch

$$S_{\vartheta} = \{t \in \mathbb{R} : f_{\vartheta}(t) > 0\}$$

angegeben.

Grundbegriffe der Statistik

Definition. Ein Vektor $(x_1, \dots, x_n) \in \mathbb{R}^n$ ist eine *Realisierung* einer Stichprobe $(X_1, \dots, X_n)$ für den Parameter ϑ , falls die Verteilung der (diskreten oder absolut stetigen) Stichprobenvariablen durch den Parameter ϑ bestimmt ist und $x_i \in S_{\vartheta}$ für alle $1 \leq i \leq n$ gilt.

Bemerkungen.

- Falls die durch ϑ bestimmte Verteilung absolut stetig ist, dann hängt die Menge der Realisierungen von der Wahl der Dichtefunktion f_{ϑ} ab.
- Die Wertemenge einer Stichprobenvariablen kann größer sein, als die zugehörige Menge der wesentlichen Werte, aber die Variable nimmt ihren Wert außer S_{ϑ} nur mit Wahrscheinlichkeit 0.

Grundbegriffe der Statistik

Beispiel. Wir werfen die Münze, die wir auf der Straße fanden, fünfmal. Ein Wurf gibt einen Kopf mit Wahrscheinlichkeit $\vartheta \in [0; 1]$, und dieser Parameter bestimmt die Verteilung des Wurfes. Dann ist der Vektor

(Kopf, Kopf, Zahl, Kopf, Zahl)

eine Realisierung der Stichprobe für alle $\vartheta \in (0; 1)$, aber für $\vartheta = 0$ oder $\vartheta = 1$ nicht!

Grundbegriffe der Statistik

Definition. Sei $(x_1, \dots, x_n) \in \mathbb{R}^n$ eine Realisierung einer Stichprobe, und sei $x \in \mathbb{R}$. Die *absolute Häufigkeit* von x ist

$$H(x) := |\{i : x_i = x\}|,$$

d.h., $H(x)$ gibt an, wie oft der Wert x im Vektor $(x_1, \dots, x_n)$ vorkommt. Die *relative Häufigkeit* von x ist

$$h(x) := \frac{H(x)}{n}.$$

Sei A eine Teilmenge der Werte $\{x_1, \dots, x_n\}$, dann ist die absolute Häufigkeit von A

$$H(A) = |\{i : x_i \in A\}|,$$

und die relative Häufigkeit von A ist

$$h(A) = \frac{H(A)}{n}.$$

Grundbegriffe der Statistik

Beispiel. Ein Würfel wird 15 Mal geworfen, die Ergebnisse sind

2, 5, 2, 1, 1, 6, 5, 4, 3, 5, 6, 2, 3, 5, 1.

Die absolute Häufigkeit von 5 ist 4, also ist ihre relative Häufigkeit $4/15$. Sei $A = \{2, 4, 6\}$, die relative Häufigkeit von A (also von den geraden Würfeln) ist $6/15 = 2/5$.

Diese Häufigkeiten sind einfacher zu bestimmen, falls die Liste der Ergebnisse geordnet ist:

1, 1, 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 5, 6, 6.

Grundbegriffe der Statistik

Definition. Sei $\mathbf{X} = (X_1, \dots, X_n)$ eine (einfache) Stichprobe mit Variablen $X_i : \Omega \rightarrow \mathbb{R}$. Für alle ω sei $(X_1^*(\omega), \dots, X_n^*(\omega))$ eine Auflistung der Elemente $X_1(\omega), \dots, X_n(\omega)$ so dass

$$X_1^*(\omega) \leq X_2^*(\omega) \leq \dots \leq X_n^*(\omega)$$

gilt. Dann sind $X_1^*, \dots, X_n^*$ auch Zufallsvariablen, die die entsprechende *geordnete Stichprobe* angeben.

Bemerkung. Die Variablen $X_1^*, \dots, X_n^*$ sind im Allgemeinen nicht unabhängig und nicht identisch verteilt!

Grundbegriffe der Statistik

Im obigen Beispiel gibt die Realisierung

2, 5, 2, 1, 1, 6, 5, 4, 3, 5, 6, 2, 3, 5, 1

der Würfe die Realisierung

1, 1, 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 5, 6, 6

der entsprechenden geordneten Stichprobe.

Grundbegriffe der Statistik

Definition. Sei $n \in \mathbb{N}$ und $\mathbf{X} = (X_1, \dots, X_n)$ eine einfache Stichprobe. Die zugehörige *empirische Verteilungsfunktion* ist die Funktion

$$t \mapsto F_n^*(t) = \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}}}{n},$$

wo $t \in \mathbb{R}$ ist.

Bemerkung. Für alle t ist $F_n^*(t)$ eine Zufallsvariable, die eigentlich eine Funktion der Variablen X_i ist. Sie gibt das Verhältnis der Anzahl der Variablen in der Stichprobe, die kleiner oder gleich t sind.

Grundbegriffe der Statistik

Anhand der geordneten Stichprobe kann die empirische Verteilungsfunktion folgenderweise ausgedrückt werden:

$$F_n^*(t) = \begin{cases} 0, & \text{falls } t < X_1^*, \\ \frac{k}{n}, & \text{falls } X_k^* \leq t < X_{k+1}^*, \quad k = 1, \dots, n-1, \\ 1, & \text{falls } X_n^* \leq t. \end{cases}$$

Die obige Form zeigt, dass wenn man die Variablen X_i durch eine Realisierung von ihnen in $F_n^*(t)$ ersetzt, so erhält man wirklich eine Verteilungsfunktion.

Grundbegriffe der Statistik

Für eine Realisierung $(x_1, \dots, x_n)$ kann die empirische Verteilungsfunktion auch durch die (relative) Häufigkeiten der Werte x_i ausgedrückt werden. Nämlich, falls die in der Realisierung auftretenden Werten $a_1 < \dots < a_m$ mit relativen Häufigkeiten $h(a_1), \dots, h(a_m)$ sind, so gilt

$$F_n^*(t) = \begin{cases} 0, & \text{falls } t < a_1, \\ \sum_{i=1}^k h(a_i), & \text{falls } a_k \leq t < a_{k+1}, \quad k = 1, \dots, m-1, \\ 1, & \text{falls } a_m \leq t. \end{cases}$$

Anders formuliert, falls $A_k = \{a_1, \dots, a_k\}$, dann gilt

$$F_n^*(t) = \begin{cases} 0, & \text{falls } t < a_1, \\ h(A_k), & \text{falls } a_k \leq t < a_{k+1}, \quad k = 1, \dots, m-1, \\ 1, & \text{falls } a_m \leq t. \end{cases}$$

Grundbegriffe der Statistik

Im obigen Beispiel ist die Realisierung der geordneten Stichprobe

1, 1, 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 5, 6, 6,

die auftretenden Werte sind 1, 2, 3, 4, 5, 6, und die relative Häufigkeiten sind

$$h(1) = \frac{3}{15}, \quad h(2) = \frac{3}{15}, \quad h(3) = \frac{2}{15},$$

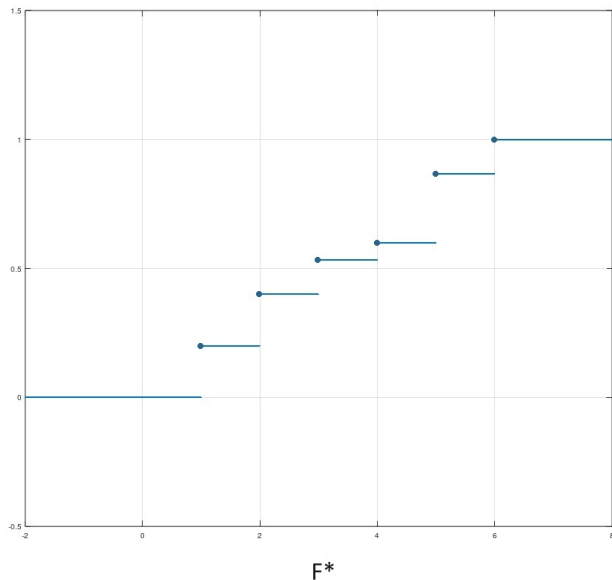
$$h(4) = \frac{1}{15}, \quad h(5) = \frac{4}{15}, \quad h(6) = \frac{2}{15}.$$

Grundbegriffe der Statistik

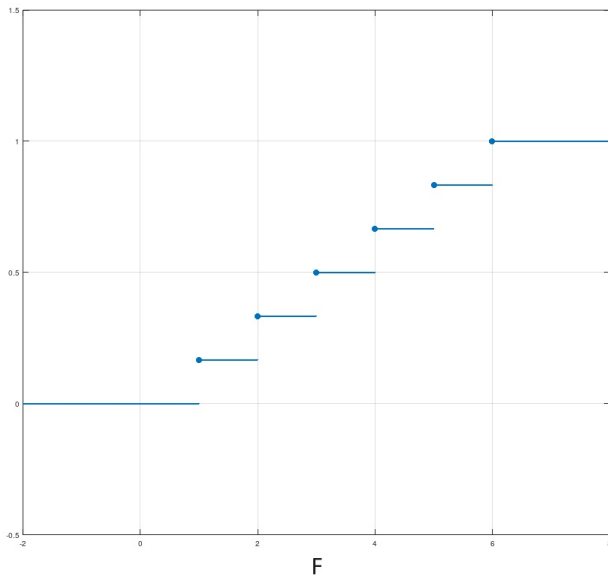
Also gilt

$$F_{15}^*(t) = \begin{cases} 0, & \text{falls } t < 1, \\ \frac{3}{15} = \frac{1}{5} & \text{falls } 1 \leq t < 2, \\ \frac{6}{15} = \frac{2}{5} & \text{falls } 2 \leq t < 3, \\ \frac{8}{15} & \text{falls } 3 \leq t < 4, \\ \frac{9}{15} & \text{falls } 4 \leq t < 5, \\ \frac{13}{15} & \text{falls } 5 \leq t < 6, \\ 1, & \text{falls } 6 \leq t. \end{cases}$$

Die empirische Verteilungsfunktion des Beispiels



Verteilungsfunktion des Wurfes mit einem regulären Würfel



Grundbegriffe der Statistik

Die folgende Behauptung zeigt, dass die empirische Verteilungsfunktion eine "gute" Schätzung der (echten) Verteilungsfunktion ist:

Behauptung. Sei $n \in \mathbb{N}$ und seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion F . Falls F_n^* die zugehörige empirische Verteilungsfunktion ist, dann

a) $\mathbb{E}(F_n^*(t)) = F(t),$

b) $\mathbb{D}^2(F_n^*(t)) = \frac{F(t)(1 - F(t))}{n},$ und

c) $\mathbb{P}(\lim_{n \rightarrow \infty} F_n^*(t) = F(t)) = 1.$

Grundbegriffe der Statistik

Beweis. a) Wegen der Linearität des Erwartungswertes gilt

$$\begin{aligned}\mathbb{E}(F_n^*(t)) &= \frac{\sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{X_i \leq t\}})}{n} \\ &= \frac{\sum_{i=1}^n \mathbb{P}(X_i \leq t)}{n} \\ &= \frac{n \cdot F(t)}{n} = F(t),\end{aligned}$$

weil die Variablen identisch verteilt sind.

Grundbegriffe der Statistik

Beweis. b) Aus der Unabhängigkeit der Zufallsvariablen $X_1, \dots, X_n$ folgt, dass die Ereignisse $\{X_i \leq t\}$ und folglich auch die Indikatorvariablen $\frac{\mathbb{1}_{\{X_i \leq t\}}}{n}$ unabhängig sind. Wegen der Eigenschaften der Varianz ergibt sich

$$\begin{aligned}\mathbb{D}^2(F_n^*(t)) &= \frac{\sum_{i=1}^n \mathbb{D}^2(\mathbb{1}_{\{X_i \leq t\}})}{n^2} \\ &= \frac{\sum_{i=1}^n \mathbb{P}(X_i \leq t)(1 - \mathbb{P}(X_i \leq t))}{n^2} \\ &= \frac{n \cdot F(t)(1 - F(t))}{n^2} = \frac{F(t)(1 - F(t))}{n}.\end{aligned}$$

Grundbegriffe der Statistik

Beweis. c) $\mathbb{E}(\mathbb{1}_{\{X_i \leq t\}}) = F(t)$, also gibt das starke Gesetz der großen Zahlen die Behauptung:

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}}}{n} = \mathbb{E}(\mathbb{1}_{\{X_i \leq t\}}) \right) = \mathbb{P}(\lim_{n \rightarrow \infty} F_n^*(t) = F(t)) = 1.$$

Grundbegriffe der Statistik

Teil c) der obigen Behauptung zeigt, dass das Ereignis, dass $F_n^*(t)$ nicht gegen $F(t)$ konvergiert, für alle $t \in \mathbb{R}$ mit Wahrscheinlichkeit 0 eintritt. In der Tat gilt aber auch mehr:

Satz. (Glivenko-Cantelli) Sei $n \in \mathbb{N}$ und seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion F . Falls F_n^* die zugehörige empirische Verteilungsfunktion ist, dann gilt

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n^*(t) - F(t)| = 0 \right) = 1.$$

D.h.: die empirische Verteilungsfunktion $F_n^*(t)$ konvergiert fast immer gleichmäßig gegen $F(t)$.

Dieser Satz zeigt, dass wir die Verteilung der Variablen mit Wahrscheinlichkeit 1 beliebig gut annähern können, falls der Umfang der Stichprobe genügend groß ist.

Stichprobenfunktionen

Definition. Sei $T : \mathbb{R}^n \rightarrow \mathbb{R}$ eine symmetrische Funktion, d.h. eine solche Funktion, für die die Gleichung

$$T(x_1, \dots, x_n) = T(\pi(x_1), \dots, \pi(x_n))$$

für alle Vektoren $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ und alle Permutationen π der Koordinaten von $\underline{x}$ gilt. Sei noch $\mathbf{X} = (X_1, \dots, X_n)$ eine Stichprobe, und nehmen wir an, dass die Funktion $T(X_1, \dots, X_n)$ eine Zufallsvariable ist. Dann heißt T eine *Stichprobenfunktion* oder eine *Statistik* (statisztika).

Stichprobenfunktionen

Ein einfaches und wichtiges Beispiel ist das empirische Mittel:

Definition. Sei $\mathbf{X} = (X_1, \dots, X_n)$ eine Stichprobe, das *empirische Mittel* (mintaátlag) von ihr ist

$$\overline{X} = \frac{X_1 + \dots + X_n}{n}.$$

Wenn man den Umfang der Stichprobe betonen möchte, schreibt man $\overline{X}_n$. Wenn das empirische Mittel einer Realisierung berechnet wird, benutzt man typischerweise kleine Buchstaben:

$$\overline{x} = \frac{x_1 + \dots + x_n}{n}.$$

Wie oben, man schreibt $\overline{x}_n$, um den Umfang der Realisierung zu betonen.

Stichprobenfunktionen

Weitere wichtige Stichprobenfunktionen:

Definition. Sei $\mathbf{X} = (X_1, \dots, X_n)$ eine Stichprobe. Der *Stichprobenmedian* (oder *empirische Median*) von $\mathbf{X}$ (empirikus medián) ist definiert als der Wert in der Mitte der geordneten Liste. D.h., sei $X_1^*, \dots, X_n^*$ die geordnete Stichprobe, falls $n = 2k + 1$ ungerade ist, so ist der Median $m_{n,\mathbf{X}} = X_{k+1}^*$, und falls $n = 2k$ gerade ist, wird der Durchschnitt aus den beiden mittleren Werten gebildet, also $m_{n,\mathbf{X}} = \frac{X_k^* + X_{k+1}^*}{2}$.

Definition. Der *Modus* oder *Modalwert* (módusz) ist der häufigste Wert, der in der Stichprobe vorkommt. Er ist im Allgemeinen nicht eindeutig, alle häufigsten Werte werden als Modus betrachtet.

Stichprobenfunktionen

Bemerkung. Der Erwartungswert des empirischen Mittels ist

$$\mathbb{E}(\bar{X}) = \frac{\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}{n} = \frac{n \cdot \mathbb{E}(X_1)}{n} = \mathbb{E}(X_1),$$

also der Erwartungswert der Stichprobenvariablen.

Weiter, nach dem Gesetz der großen Zahlen konvergiert das empirische Mittel gegen $\mathbb{E}(X_1)$ mit Wahrscheinlichkeit 1.

D.h.: das empirische Mittel ist eine gute *Schätzer* für den Erwartungswert.

Punktschätzer

- Nähert man den Wert eines Parameters einer Verteilung an, so spricht man von Punktschätzung.
 - Eine Stichprobenfunktion, die bei der Schätzung benutzt wird, bezeichnet man im Rahmen der Schätztheorie als *Schätzfunktion* (oder einfach *Schätzer*).
- ⇒ Das empirische Mittel ist also eine Schätzfunktion für den Erwartungswert der Verteilung.

Eigenschaften der Schätzfunktionen

Definition. Sei $\mathbf{X} = (X_1, \dots, X_n)$ eine einfache Stichprobe, wo die Verteilung der Variablen durch einen Parameter $\vartheta \in \Theta \subset \mathbb{R}^d$ bestimmt ist. Sei noch $\psi : \Theta \rightarrow \mathbb{R}$ eine Funktion, die jeder Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta$ eine zu schätzende Kennzahl (z.B. Erwartungswert, Varianz, usw.) zuordnet. Sei $T : \mathbb{R}^n \rightarrow \mathbb{R}$ ein Schätzer für $\psi(\vartheta)$.

a) Wir sagen, dass T *erwartungstreu* oder *unverzerrt* (torzíatlan becslés) ist, falls

$$\mathbb{E}_\vartheta(T(X_1, \dots, X_n)) = \psi(\vartheta)$$

für alle $\vartheta \in \Theta$ gilt.

b) Wir sagen, dass T *stark konsistent* (erősen konzisztens) ist, falls

$$\mathbb{P}_\vartheta \left(\lim_{n \rightarrow \infty} T(X_1, \dots, X_n) = \psi(\vartheta) \right) = 1$$

für alle $\vartheta \in \Theta$ gilt.

Eigenschaften der Schätzfunktionen

Beispiel. Sei $\psi(\vartheta) = \mathbb{E}_{\vartheta}(X_1)$, dann zeigt die Rechnung

$$\mathbb{E}_{\vartheta}(\bar{X}) = \frac{\mathbb{E}_{\vartheta}(X_1) + \cdots + \mathbb{E}_{\vartheta}(X_n)}{n} = \mathbb{E}_{\vartheta}(X_1),$$

dass das empirische Mittel eine erwartungstreue Schätzfunktion für den Erwartungswert ist.

Das starke Gesetz der großen Zahlen (bezüglich jedes Maßes $\mathbb{P}_{\vartheta}$) zeigt, dass das empirische Mittel auch stark konsistent ist.

Eigenschaften der Schätzfunktionen

Wir berechnen die Varianz des empirischen Mittels: wegen der Unabhängigkeit der Variablen und der Eigenschaften der Varianz gilt

$$\mathbb{D}_{\vartheta}^2(\bar{X}) = \mathbb{D}_{\vartheta}^2\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}_{\vartheta}^2(X_i) = \frac{\mathbb{D}_{\vartheta}^2(X_1)}{n}.$$

Eigenschaften der Schätzfunktionen

Was wäre ein guter Schätzer für die Varianz?

Eine naheliegende Möglichkeit ist die durchschnittliche quadratische Abweichung der Beobachtungswerte von ihrem Mittelwert:

$$S_n^2 := \frac{1}{n} \sum_{k=1}^n (X_i - \bar{X})^2.$$

Diese Stichprobenfunktion ist die *verzerrte empirische Varianz* (tapasztalati szórásnégyzet). Die Wurzel S_n von S_n^2 ist die *verzerrte empirische Standardabweichung*. Wie beim empirischen Mittel, man benutzt die Bezeichnung s_n^2 und s_n für die Realisierungen. Man kann den Umfang n von der Notation auch weglassen.

Eigenschaften der Schätzfunktionen

Die folgende Rechnung gibt eine andere nützliche Formel für S^2 an:

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i \cdot \bar{X} + \bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \cdot \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{n} \cdot n \cdot \bar{X}^2 \\ &= \overline{X^2} - 2\bar{X}^2 + \bar{X}^2 = \overline{X^2} - \bar{X}^2. \end{aligned}$$

Eigenschaften der Schätzfunktionen

Beispiel. Betrachten wir unser voriges Beispiel (15 Würfe eines Würfels) mit der Realisierung

2, 5, 2, 1, 1, 6, 5, 4, 3, 5, 6, 2, 3, 5, 1.

Das empirische Mittel $\bar{x}$ ist

$$\frac{2 + 5 + 2 + 1 + 1 + 6 + 5 + 4 + 3 + 5 + 6 + 2 + 3 + 5 + 1}{15} = \frac{17}{5},$$

weiter

$$\overline{x^2} = \frac{3 \cdot 1^2 + 3 \cdot 2^2 + 2 \cdot 3^2 + 4^2 + 4 \cdot 5^2 + 2 \cdot 6^2}{15} = \frac{221}{15},$$

also

$$s^2 = \frac{238}{75}, \quad s \approx 1,7814.$$

Eigenschaften der Schätzfunktionen

Das Wort "verzerrt" kann jetzt folgenderweise erklärt werden:

$$\begin{aligned}\mathbb{E}_\vartheta(S_n^2) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\vartheta(X_i^2) - \mathbb{E}_\vartheta(\bar{X}^2) \\&= \frac{1}{n} \sum_{i=1}^n (\mathbb{D}_\vartheta^2(X_i) + \mathbb{E}_\vartheta(X_i)^2) - (\mathbb{D}_\vartheta^2(\bar{X}) + \mathbb{E}_\vartheta(\bar{X})^2) \\&= \frac{1}{n} \cdot n \cdot (\mathbb{D}_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2) - \left(\frac{1}{n} \mathbb{D}_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2 \right) \\&= \frac{n-1}{n} \mathbb{D}_\vartheta^2(X_1),\end{aligned}$$

also ist S_n^2 *nicht* eine erwartungstreue Schätzfunktion für die Varianz.

Eigenschaften der Schätzfunktionen

Definition. Sei $\mathbf{X} = (X_1, \dots, X_n)$ eine einfache Stichprobe, wo die Verteilung der Variablen durch einen Parameter $\vartheta \in \Theta \subset \mathbb{R}^d$ bestimmt ist. Sei $\psi : \Theta \rightarrow \mathbb{R}$ eine Funktion, die jeder Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta$ eine zu schätzende Kennzahl zuordnet. Sei $T : \mathbb{R}^n \rightarrow \mathbb{R}$ ein Schätzer für $\psi(\vartheta)$.

Wir sagen, dass T *asymptotisch erwartungstreu* (asymptotikusan torzítatlan) ist, falls

$$\lim_{n \rightarrow \infty} \mathbb{E}_\vartheta(T(X_1, \dots, X_n)) = \psi(\vartheta)$$

für alle $\vartheta \in \Theta$ gilt.

Da

$$\lim_{n \rightarrow \infty} \mathbb{E}_\vartheta(S_n^2) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \mathbb{D}_\vartheta^2(X_1) = \mathbb{D}_\vartheta^2(X_1),$$

gilt, ist die verzerrte empirische Varianz eine asymptotisch erwartungstreue Schätzfunktion für die Varianz.

Eigenschaften der Schätzfunktionen

Es ist aber leicht, diese Schätzfunktion zu verändern, damit man eine erwartungstreue Schätzfunktion bekommt:

Definition. Die Schätzfunktion

$$(S_n^*) := \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_i - \bar{X})^2$$

heißt die *empirische Varianz* (korrigált tapasztalati szórásnégyzet). Ihre Wurzel S_n^* ist die *empirische Standardabweichung*. Bei Realisierungen benutzt man die Bezeichnung $(s_n^*)^2$ und (s_n^*) , und man kann den Umfang n in der Notation weglassen.

Eigenschaften der Schätzfunktionen

Im obigen Beispiel gilt

$$s^2 = \frac{238}{75}; \quad (s^*)^2 = \frac{15}{14} \cdot \frac{238}{75} = \frac{17}{5} = 3,4; \quad s^* = \sqrt{3,4} \approx 1,8439$$

Eigenschaften der Schätzfunktionen

Behauptung. Die empirische Varianz ist eine erwartungstreue Schätzfunktion für die Varianz, nämlich

$$\mathbb{E}_{\vartheta}((S_n^*)^2) = \frac{n}{n-1} \mathbb{E}_{\vartheta}(S_n^2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \mathbb{D}_{\vartheta}^2(X_1) = \mathbb{D}_{\vartheta}^2(X_1).$$

Aufgabe. Zeigen wir, dass S_n^2 und $(S_n^*)^2$ stark konsistente Schätzfunktionen für die Varianz sind.

Eigenschaften der Schätzfunktionen

Bemerkung. Die empirische Standardabweichung ist *nicht* eine erwartungstreue Schätzfunktion für die Standardabweichung. Im Allgemeinen gilt $\mathbb{E}_\vartheta(S_n^*) \leq \mathbb{D}_\vartheta(X_1)$, also unterschätzt dieser Schätzer in den meisten Fällen die Standardabweichung der Grundgesamtheit. Falls $X_1, \dots, X_n$ unabhängig und normalverteilt sind, dann gilt

$$\mathbb{E}_\vartheta(S_n^*) = \sqrt{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \mathbb{D}_\vartheta(X_1),$$

und S_n^* ist in diesem Fall asymptotisch erwartungstreu.