# 1. Eigenschaften einer Zufallsstichprobe

## 1.1 Grundkonzepte

**Definition 1.1.1:** Man nennt die Zufallsvariablen $X_1, \ldots, X_n$ **zufällige Stichprobe** oder **Zufallsstichprobe** vom Umfang $n$ aus der Population (Grundgesamtheit) $f(x)$, wenn die $X_1, \ldots, X_n$ unabhängige Zufallsvariablen sind und die Randdichte oder Randwahrscheinlichkeitsfunktion jedes $X_i$ dieselbe Funktion $f(x)$ ist. Alternativ nennt man $X_1, \ldots, X_n$ **unabhängig und identisch verteilte** (independent and identically distributed, **iid**) Zufallsvariablen mit Dichte oder Wahrscheinlichkeitsfunktion $f(x)$.

**Bemerkungen:**

Die Beobachtungen werden derart genommen, dass der Wert einer Beobachtung keinen Effekt auf die übrigen Beobachtungen hat (Unabhängigkeit).

Die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion von $X_1, \ldots, X_n$ ist

$$f(x_1, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} f(x_i)\,.$$

Da die $X_i$ alle identisch verteilt sind, sind die marginalen $f(x_i)$ alle dieselben. Falls $f(x)$ ein Mitglied einer parametrischen Familie ist (Normal- ...) mit Dichte oder Wahrscheinlichkeitsfunktion $f(x|\theta)$, dann ist die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)\,.$$

Statistisches Denken: Wir nehmen an, dass die betrachtete Population ein Mitglied einer parametrischen Familie ist, aber der wahre Parameterwert unbekannt ist. Betrachtet man verschiedene Werte von $\theta$, so kann man untersuchen, wie sich eine Zufallsstichprobe für verschiedene Populationen verhält.

**Beispiel 1.1.1:** Sei $X_1, \ldots, X_n$ eine Zufallsstichprobe aus einer Exponential($\beta$)-Population, d.h. mit Dichtefunktion

$$f(x|\beta) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right), \quad 0 \leq x < \infty,\ 0 < \beta\,,$$

mit $\mathsf{E}(X) = \beta$ und $\mathsf{var}(X) = \beta^2$. Dazu korrespondiert die Verteilungsfunktion

$$F(x|\beta) = 1 - \exp\left(-\frac{x}{\beta}\right)\,.$$

Diese Zufallsstichprobe charakterisiere Ausfallszeiten von IC's in Jahren. Wir testen $n$ IC's und verwenden diese bis sie ausfallen.

Die gemeinsame Dichte der Zufallsstichprobe ist

$$f(x_1, \ldots, x_n|\beta) = \prod_{i=1}^{n} f(x_i|\beta) = \prod_{i=1}^{n} \frac{1}{\beta} \exp\left(-\frac{x_i}{\beta}\right) = \frac{1}{\beta^n} \exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} x_i\right) .$$

Damit ist z.B. die Wahrscheinlichkeit, dass alle IC's länger als 2 Jahre halten, gleich

$$\begin{aligned}
P(X_1 > 2, \ldots, X_n > 2|\beta) &= P(X_1 > 2|\beta) \cdots P(X_n > 2|\beta) \quad \text{Unabhängigkeit} \\
&= (P(X_1 > 2|\beta))^n \quad \text{identisch verteilt} \\
&= \left(e^{-2/\beta}\right)^n \quad \text{exponentialverteilt} \\
&= e^{-2n/\beta} .
\end{aligned}$$

Für große Werte von $\beta$ ist dies nahe 1.

**Bemerkungen:**

Diese Art Stichprobe nennt man *Stichprobe aus einer unendlichen Population*.

Gedanke: Erhalte die Werte von $X_1, \ldots, X_n$ sequentiell d.h.

1. Experiment: $X_1 = x_1$ wird beobachtet. Dann
2. Experiment: $X_2 = x_2$ wird beobachtet.

Unabhängigkeit bedeutet, dass die Verteilung von $X_2$ nicht von $X_1 = x_1$ abhängt.

Das Entfernen von $x_1$ aus der unendlich großen Population ändert diese nicht. Somit ist $X_2 = x_2$ noch immer zufällige Beobachtung aus derselben Population.

Bei einer *Stichprobe aus einer endlichen Population* muss zwischen mit/ohne Zurücklegen (replacement) unterschieden werden.

Population ist jetzt die endliche Menge von Zahlen $\{x_1, \ldots, x_N\}$. Sei $X_1, \ldots, X_n$ eine Stichprobe daraus.

(1) *mit Zurücklegen:* Wir nehmen an, dass die Wahrscheinlichkeit einen Wert aus der Population zu wählen gleich ist für alle $x_j$, also gleich $1/N$ ist (Urnenmodell, ziehe aus einem Hut).

Sei $X_1 = x_1$. Dann wird $x_1$ zurückgelegt und das Ziehen wiederholt. Wieder sind alle $N$ Werte gleichwahrscheinlich. Wir erhalten $X_2 = x_2$. Falls dieselbe Zahl gewählt wurde ist $x_2 = x_1$. Wir wiederholen das Experiment $n$ mal und erhalten die Stichprobe $X_1, \dots X_n$.

Die Bedingungen der Definition 1.1.1 sind bei dieser Art die Stichprobe zu ziehen erfüllt. Jedes $X_i$ ist eine diskrete Zufallsvariable, welche die Werte $x_1, \dots, x_N$ mit gleicher Wahrscheinlichkeit annimmt. Die Zufallsvariablen $X_1, \dots, X_n$ sind *unabhängig*, da der Auswahlprozess eines $X_i$ derselbe ist, unabhängig der Werte die zuvor gezogen wurden.

(2) *ohne Zurücklegen:* Annahmen wie zuvor jedoch wird jetzt nach $X_1 = x_1$ der Wert $x_1$ nicht mehr zurückgegeben. Der 2. Wert wird aus den verbliebenen $N-1$ gewählt. Jeder dieser $N-1$ Werte hat Wahrscheinlichkeit $1/(N-1)$ gezogen zu werden.

Wir erhalten $X_2 = x_2 \neq x_1$ u.s.w. Dieser Prozess liefert eine Stichprobe $X_1, \ldots, X_n$, jedoch ist ein gezogener Wert später nicht mehr verfügbar. Somit ist keine Unabhängigkeit in dieser Stichprobe und die Annahmen aus Definition 1.1.1 sind nicht erfüllt.

Betrachte die marginale Verteilung von $X_2$:

$$P(X_2 = x) = \sum_{i=1}^{N} P(X_2 = x | X_1 = x_i) P(X_1 = x_i) \quad \text{totale Wahrscheinlichkeit.}$$

Für **einen** Index $i$ ist $x = x_i$ und somit $P(X_2 = x | X_1 = x_i) = 0$. Für alle übrigen $N - 1$ Indizes $j \neq i$ gilt $P(X_2 = x | X_1 = x_j) = 1/(N-1)$ und es folgt

$$P(X_2 = x) = (N-1) \left( \frac{1}{N-1} \cdot \frac{1}{N} \right) = \frac{1}{N}\,.$$

Also sind die marginalen Verteilungen dieselben unter Ziehen der Stichprobe $mit$ und $ohne$ Zurücklegen. Jedoch gibt es keine Unabhängigkeit bei Ziehen ohne Zurücklegen. Annähernde Unabhängigkeit erreicht man für $N$ sehr groß.

Von nun an verwenden wir die Definition 1.1.1, um eine Zufallsstichprobe aus einer Population zu charakterisieren.