1.3 Wiederholung der Konvergenzkonzepte

Wir erlauben nun, dass der Stichprobenumfang n unendlich groß wird und untersuchen das Verhalten von Stichprobengrößen für diesen Fall. Dies liefert uns nützliche Approximationen für den endlichen Fall. Wir betrachten jetzt 3 Typen von Konvergenz. Speziell wird $\overline{X}_n$, das Mittel von n Beobachtungen, untersucht.

1.3.1 Konvergenz in Wahrscheinlichkeit

Definition 1.3.1: Eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ **konvergiert in Wahrscheinlichkeit** gegen eine Zufallsvariable X, falls für jedes $\epsilon > 0$ gilt

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0\,,$$

was äquivalent ist mit

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1\,.$$

Wir schreiben dafür $X_n \stackrel{P}{\to} X$.

Hierbei sind die $X_1, X_2, \ldots$ nicht notwendigerweise iid Variablen.

Statistiker beschäftigen sich häufig mit Situationen in denen die Grenzvariable eine Konstante und die Zufallsvariablen in der Folge das Stichprobenmittel ist.

Satz 1.3.1: (Schwaches Gesetz der großen Zahlen – WLLN) Seien $X_1, X_2, \ldots$ iid Zufallsvariablen mit $\mathrm{E}(X_i) = \mu$ und $\mathrm{var}(X_i) = \sigma^2 < \infty$. Dann gilt für jedes $\epsilon > 0$

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \epsilon) = 1\,,$$

also konvergiert das Stichprobenmittel $\overline{X}_n$ in Wahrscheinlichkeit gegen das Populationsmittel μ, d.h. $\overline{X}_n \xrightarrow{P} \mu$.

Das WLLN sagt aus, dass sich unter recht allgemeinen Bedingungen das Stichprobenmittel dem Populationsmittel nähert falls $n \to \infty$.

Diese Eigenschaft, dass sich eine Folge der gleichen Stichprobengröße für $n \to \infty$ einer Konstanten nähert, nennt man auch **Konsistenz**.

Zentrale Bedeutung hat im Folgenden auch die **Chebychev'sche Ungleichung**. Zur Erinnerung gilt für jedes beliebige $\epsilon > 0$

$$P(|X - a| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathsf{E}(X - a)^2$$

und speziell, dass

$$P(|X - \mathsf{E}(X)| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathsf{var}(X)\,.$$

Damit kann recht einfach der Satz 1.3.1 bewiesen werden.

Beispiel 1.3.1: Sei $X_1, X_2, \ldots$ eine Folge von iid Variablen mit $\mathsf{E}(X_i) = \mu$ und $\mathsf{var}(X_i) = \sigma^2 < \infty$. Hält das WLLN auch für

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Chebychev sagt aus, dass

$$P(|S_n^2 - \sigma^2| \geq \epsilon) \leq \frac{1}{\epsilon^2} \mathsf{var}(S_n^2)\,.$$

Somit ist $\mathsf{var}(S_n^2) \to 0$ für $n \to \infty$ eine hinreichende Bedingung dafür, dass S_n^2 in Wahrscheinlichkeit gegen σ^2 konvergiert.

Eine natürliche Erweiterung von Definition 1.3.1 bezieht sich auf Funktionen von Zufallsvariablen. Wir wollen der Frage nachgehen, ob bei Konvergenz in Wahrscheinlichkeit der Folge $X_1, X_2, \ldots$ gegen X oder gegen a, auch die Folge $h(X_1), h(X_2), \ldots$ konvergiert.

Satz 1.3.2: Angenommen die Folge von Zufallsvariablen $X_1, X_2, \ldots$ konvergiert in Wahrscheinlichkeit gegen eine Zufallsvariable X, und sei h eine stetige Funktion. Dann konvergiert $h(X_1), h(X_2), \ldots$ in Wahrscheinlichkeit gegen $h(X)$.

Beispiel 1.3.1 Fortsetzung: Falls die Stichprobenvarianz S_n^2 konsistent ist für σ^2, dann ist somit auch die Stichprobenstandardabweichung

$$S_n = \sqrt{S_n^2} = h(S_n^2)$$

konsistent für σ.

Bemerke jedoch, dass S_n verzerrt ist für σ. Diese Verzerrung verschwindet aber asymptotisch.

1.3.2 Fast sichere Konvergenz

Dieses Konzept ist stärker als die Konvergenz in Wahrscheinlichkeit.

Manchmal bezeichnet man es auch als *Konvergenz mit Wahrscheinlichkeit 1*.

Es ist ähnlich zur punktweisen Konvergenz einer Funktionenfolge: überall bis auf einer Ausnahmemenge (mit Wahrscheinlichkeit 0, **a**lmost **s**ure, a.s.).

Definition 1.3.2: Eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ **konvergiert fast sicher** gegen eine Zufallsvariable X, falls für jedes $\epsilon > 0$ gilt

$$P\left(\lim_{n\to\infty} |X_n - X| < \epsilon\right) = 1\,.$$

Wir schreiben dafür $X_n \stackrel{f.s.}{\to} X$.

Eine Zufallsvariable ist eine reellwertige Funktion definiert auf einem Stichprobenraum S. Falls die Elemente in S mit s bezeichnet sind, dann sind $X_n(s)$ und $X(s)$ Funktionen die auch auf S definiert sind.

$X_n(s)$ konvergiert fast sicher gegen $X(s)$, für alle $s \in S$, ausgenommen vielleicht für $s \in N$, mit $N \subset S$ und $P(N) = 0$.

Beispiel 1.3.2: Sei S das abgeschlossene Intervall $[0,1]$ mit der Gleichverteilung darauf definiert. Sei $X_n(s) = s + s^n$ und $X(s) = s$.

Für jedes $s \in [0,1)$ gilt $s^n \to 0$ für $n \to \infty$ und $X_n(s) \to s = X(s)$.

Jedoch ist $X_n(1) = 2$ für alle n, sodass $X_n(1) \not\to 1 = X(1)$.

Aber Konvergenz tritt ein auf dem Bereich $[0,1)$ und $P([0,1)) = 1$. Damit konvergiert X_n gegen X fast sicher.

Fast sichere Konvergenz impliziert die Konvergenz in Wahrscheinlichkeit.

Das Starke Gesetz der großen Zahlen (SLLN) beschreibt die fast sichere Konvergenz zu einer Konstanten.

Satz 1.3.3: (Starkes Gesetz der großen Zahlen – SLLN) Seien $X_1, X_2, \ldots$ iid Zufallsvariablen mit E$(X_i) = \mu$ und var$(X_i) = \sigma^2 < \infty$. Dann gilt für jedes $\epsilon > 0$

$$P\left(\lim_{n\to\infty} |\overline{X}_n - \mu| < \epsilon\right) = 1\,,$$

also konvergiert das Stichprobenmittel $\overline{X}_n$ fast sicher gegen das Populationsmittel μ, d.h. $\overline{X}_n \stackrel{f.s.}{\to} \mu$.

1.3.3 Konvergenz in Verteilung

Definition 1.3.3: Eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ **konvergiert in Verteilung** gegen eine Zufallsvariable X, falls

$$\lim_{n\to\infty} F_{X_n} = F_X$$

in allen Punkten x, in denen $F_X(x)$ stetig ist. Wir schreiben dafür $X_n \xrightarrow{D} X$.

Hierbei konvergieren eigentlich die Verteilungsfunktionen (und nicht die Zufallsvariablen)!

Was ist eigentlich die Grenzverteilung von $\overline{X}_n$?

Satz 1.3.4: Falls die Folge von Zufallsvariablen $X_1, X_2, \ldots$ in Wahrscheinlichkeit gegen die Zufallsvariable X konvergiert, dann konvergiert diese Folge auch in Verteilung gegen X.

Satz 1.3.5: Die Folge von Zufallsvariablen $X_1, X_2, \ldots,$ konvergiert in Wahrscheinlichkeit gegen eine Konstante μ, dann und nur dann, wenn diese Folge auch in Verteilung gegen μ konvergiert. Die Aussage

$$P(|X_n - \mu| > \epsilon) \to 0 \quad \text{für jedes } \epsilon > 0$$

ist daher äquivalent mit

$$P(X_n < x) \to \begin{cases} 0, & \text{falls } x < \mu\,, \\ 1, & \text{falls } x > \mu\,. \end{cases}$$

Satz 1.3.6: (Zentraler Grenzwertsatz – CLT) Seien $X_1, X_2, \ldots$ iid Zufallsvariablen mit $\mathrm{E}(X_i) = \mu$ und $\mathrm{var}(X_i) = \sigma^2 < \infty$. Bezeichne $G_n(x)$ die Verteilungsfunktion von $\sqrt{n}(\overline{X}_n - \mu)/\sigma$. Dann gilt für jedes $x \in \mathbb{R}$

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy\,.$$

Also hat $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ als Grenzverteilung die Standard-Normalverteilung.

Bemerkungen:

Wir beginnen eigentlich ohne Annahmen (bis auf Unabhängigkeit und endliche Varianzen) und enden in der Normalverteilung.

Diese Normalität stammt von den Summen von *kleinen* (endliche Varianzen) Störtermen.

Einschränkung: Keine Kenntnis wie gut diese Approximation ist!

Beispiel 1.3.3: Sei $X_1, \ldots, X_n$ eine Zufallsstichprobe aus einer Negativ-Binomial Verteilung für die gilt

$$P(X_1 = x|r,p) = \binom{r+x-1}{x} p^r (1-p)^x\,, \qquad x = 0, 1, \ldots$$

mit

$$\mathsf{E}(X_1) = \frac{r(1-p)}{p}\,, \qquad \mathsf{var}(X_1) = \frac{r(1-p)}{p^2}\,.$$

Das CLT sagt aus, dass

$$\sqrt{n}\,\frac{\overline{X_n} - \frac{r(1-p)}{p}}{\sqrt{\frac{r(1-p)}{p^2}}}$$

asymptotisch $\mathsf{N}(0,1)$ verteilt ist.

Weiters gilt die Faltungseigenschaft $\sum_{i=1}^{n} X_i \sim$ Negativ-Binomial(nr, p).

Wir betrachten die Situation: $r = 10$, $p = 1/2$ und $n = 30$.

Exakt: Damit ergibt sich exakt

$$P(\overline{X}_{30} \leq 11) = P\left(\sum_{i=1}^{30} X_i \leq 330\right) = \sum_{x=0}^{330} \binom{300 + x - 1}{x} \left(\frac{1}{2}\right)^{300} \left(\frac{1}{2}\right)^{x} = 0.8916\,.$$

Diesem Ergebnis liegt eine ziemlich komplexe Berechnung zugrunde.

Approximativ: Das CLT liefert mit $\mathsf{E}(\overline{X}_{30}) = r = 10$ und $\mathsf{var}(\overline{X}_{30}) = 2r/n = 20/n$ die Approximation

$$P(\overline{X}_{30} \leq 11) = P\left(\sqrt{30}\frac{\overline{X}_{30} - 10}{\sqrt{20}} \leq \sqrt{30}\frac{11 - 10}{\sqrt{20}}\right) \approx P(Z \leq \sqrt{3/2}) = 0.88966\,,$$

wobei $Z \sim N(0, 1)$.

Ein Approximationswerkzeug, das häufig zusammen mit dem CLT verwendet wird, ist der folgende Satz:

Satz 1.3.7: (Satz von Slutsky) Falls $X_n \xrightarrow{D} X$ und $Y_n \xrightarrow{P} a$, mit einer Konstanten a, dann gilt

(a) $Y_n X_n \xrightarrow{D} aX$,

(b) $Y_n + X_n \xrightarrow{D} a + X$.

Beispiel 1.3.4: Angenommen

$$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{D} N(0,1)$$

aber σ^2 sei unbekannt. Aus Beispiel 1.3.1 wissen wir, dass $S_n^2 \xrightarrow{P} \sigma^2$ falls $\lim_{n\to\infty} \text{var}(S_n^2) = 0$.

Für stetige Transformationen $g(\cdot)$ liefert Satz 1.3.2: $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$.

Hier: $S_n^2 \xrightarrow{P} \sigma^2 \Rightarrow \sqrt{S_n^2} \xrightarrow{P} \sqrt{\sigma^2}$ und $S_n^2 \xrightarrow{P} \sigma^2 \Rightarrow 1/S_n \xrightarrow{P} 1/\sigma$.

Also gilt auch $\sigma/S_n \xrightarrow{P} 1$. Mittels Slutsky (Satz 1.3.7) folgt damit:

$$\underbrace{\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}}_{\xrightarrow{D} N(0,1)} \cdot \underbrace{\frac{\sigma}{S_n}}_{\xrightarrow{P} 1} = \sqrt{n}\frac{\overline{X}_n - \mu}{S_n} \xrightarrow{D} N(0,1)\,.$$

Satz 1.3.8: (Delta Methode) Sei Y_n eine Folge von Zufallsvariablen die

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$$

genügt. Sei g eine reellwertige Funktion mit $g'(\theta) \neq 0$. Dann gilt

$$\sqrt{n}\Big(g(Y_n) - g(\theta)\Big) \xrightarrow{D} N\Big(0, \sigma^2[g'(\theta)]^2\Big)\,.$$

Satz 1.3.9: (Delta Methode 2. Ordnung) Sei Y_n eine Folge von Zufallsvariablen die

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2)$$

genügt. Für eine Funktion g und für den Wert θ sei $g'(\theta) = 0$ und es existiere $g''(\theta) \neq 0$. Dann gilt

$$n\Big(g(Y_n) - g(\theta)\Big) \xrightarrow{D} \sigma^2 \frac{g''(\theta)}{2} \chi_1^2\,.$$