Statistische Methoden

Dr. C.J. Luchsinger

2 Beweis CLT

Literatur Kapitel 2: Krengel: § 12

2.1 Einführung

Wir repetieren aus WTS, Kapitel 5: **Fragestellung:** Was geschieht mit dem Ausdruck

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}, \tag{2.1}$$

einfacherer Fall ($\mu = 0$, $\sigma = 1$)

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k,$$

falls $n \to \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen?

Theorem 2.1 [Zentraler Grenzwertsatz] *Sei $X_i, i \geq 1$, eine Folge von iid Zufallsgrössen mit $E[X_1] =: \mu$ und $0 < V[X_1] =: \sigma^2 < \infty$. Dann gilt für $a \in \mathbb{R}$ beliebig:*

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq a\right] \longrightarrow P[\mathcal{N}(0,1) \leq a]$$

wenn $n \to \infty$.

Mit "$P[\mathcal{N}(0,1) \leq a]$" meinen wir die Wahrscheinlichkeit, dass eine Standard-Normal-verteilte Zufallsgrösse Werte kleiner oder gleich a annimmt.
Die Verteilungsfunktion der zentrierten und normierten Summe (2.1) konvergiert also in jedem Punkt gegen die Verteilungsfunktion der Standard-Normal-Verteilung.

1. Auf Beiblättern zu Kapitel 5 in WTS finden Sie ein paar Plots, welche anhand des klassischen Falls der Binomialverteilung den CLT illustrieren. Auf CLT1 wird bei konstantem p das n erhöht; auf CLT2 wird bei $n = 50$ das p variiert.

2. Ausdruck (2.1) erinnert zu Recht an die Z-Transform. Betrachten wir dazu den Ausdruck

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

etwas genauer: mit

$$\sum_{k=1}^{n} X_k - n\mu$$

haben wir in einem ersten Schritt offenbar einfach *zentriert* (den Erwartungswert abgezogen - er ist jetzt 0), dann in einem zweiten Schritt mit $\sqrt{n}\sigma$ *normiert* und dabei

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

erhalten. Die Varianz ist jetzt 1. Zusammen nennt man diese beiden Schritte *standardisieren.*

And then a miracle occurs...,

denn das Ding hat am Schluss (bei $n = \infty$) eine Normalverteilung $\mathcal{N}(0, 1)$. Das *Überraschende* am CLT ist, dass mit wenigen Voraussetzungen (v.a. keine über die Verteilung der X_i's) immer eine Normalverteilung als Limesverteilung resultiert.

3. Ab welchem n "darf" man die Normalverteilung brauchen? Nichttriviale, allgemeine Fehler-Abschätzungen, welche nur von n abhängen und für alle Verteilungen gelten, existieren nicht. Wenn die X_i's aber Be(p)-Zufallsgrössen sind, so sagt eine Praktikerregel, dass $np(1-p) \geq 9$ gelten sollte, damit man mit gutem Gewissen die Normalverteilung benutzen darf.

4. Eine kleine Verbesserung der Approximation erreicht man im Fall von Summen diskreter Zufallsgrössen (also zum Beispiel Bernoulli/Binomial) mit der sogenannten Diskretisierungs-Korrektur (englisch: Continuity-Correction), siehe Cartoon-Guide Kapitel 5 oder Stahel 6.11 i.

5. "Nichttriviale, allgemeine Fehler-Abschätzungen, welche nur von n abhängen und für alle Verteilungen gelten, existieren nicht." Weshalb nicht?

Suchen Sie ein Gegenbeispiel für eine solche Schranke.

Nebenbemerkung: Es gilt der Satz von Berry-Essen: Seien $X_1, X_2, \ldots$ iid Zufallsgrössen. Sei $0 < \sigma^2 := V[X_1] < \infty$ und $\gamma := E[|X_1 - E[X_1]|^3] < \infty$. Dann gilt:

$$\sup_a \left| P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq a \right] - P[\mathcal{N}(0,1) \leq a] \right| \leq 0.8\gamma/(\sigma^3 \sqrt{n}).$$

6. Rechnen wir zum CLT ein Beispiel (authentisch): Sie werden von Räubern im Wald zu einem Münzwurfspiel eingeladen. Wenn in 100 Würfen mehr "Kopf" als "Zahl" kommt, dürfen Sie von den Räubern ausgeraubt werden (Merke: Transparenz ist sehr wichtig für das funktionieren von Märkten). Das Resultat ist nach 100 Würfen 65 zu Ihren Ungunsten. Sie fragen sich jetzt (bisschen spät), ob die Münze fair war oder nicht. Dazu überlegen Sie sich, wie gross *bei einer fairen* Münze wohl die Wahrscheinlichkeit ist, 65 mal (oder mehr) "Kopf" zu werfen. Leider haben Sie keine Bin(100, 0.5)-Verteilungstabelle dabei. Rettung naht: Sie erinnern sich nämlich Gott sei Dank an den CLT:

Beachten Sie bitte: 1. Wir machen diese Rechnungen unter der Annahme $p = 0.5$. 2. Wir berechnen die Wahrscheinlichkeit, 65 mal *oder mehr* "Kopf" zu werfen.

2.2 Beweis

Wir folgen im Beweis 1 zu 1 dem Beweis aus dem Krengel (Beweis von Kersting); zum Teil ist er der mE besseren Lesbarkeit wegen ein wenig umgestellt und ergänzt.

Notation: Für Verteilungsfunktionen F und G verwenden wir die Metrik (HA: kontrollieren ob tatsächlich eine Metrik):

$$d(F,G) := \sup_x |F(x) - G(x)|.$$

Sind X, Y Zufallsvariablen mit Verteilungsfunktionen $F(x) := P[X \leq x]$ und $G(x) := P[Y \leq x]$, so schreiben wir manchmal statt $d(F,G)$ auch $d(X,Y)$ oder gar $d(X,G)$. Wir benötigen später das folgende

Lemma 2.2 *Ist $E[Y^2] \leq \eta$, so ist*

$$d(X+Y,\Phi) \leq d(X,\Phi) + 2\eta^{1/3}.$$

Dabei ist Φ die Verteilungsfunktion von $\mathcal{N}(0,1)$.

Beweis: Die maximale Steigung der Verteilungsfunktion Φ ist

$$\Phi'(0) = \varphi(0) = \frac{1}{\sqrt{2\pi}} \leq 1.$$

Also ist für alle x und alle $\delta > 0$

$$|\Phi(x \pm \delta) - \Phi(x)| \leq \delta.$$

Ist $X \leq x - \delta$, so ist $Y > \delta$ oder $X + Y \leq x$. Ist $X + Y \leq x$, so ist $X \leq x + \delta$ oder $Y < -\delta$. Daher gilt

$$P[X \leq x - \delta] - P[Y > \delta] \leq P[X + Y \leq x] \leq P[X \leq x + \delta] + P[Y < -\delta].$$

Daraus folgt

$$\begin{aligned} P[X \leq x - \delta] - \Phi(x - \delta) - \delta - P[Y > \delta] &\leq P[X + Y \leq x] - \Phi(x) \\ &\leq P[X \leq x + \delta] - \Phi(x + \delta) + \delta + P[Y < -\delta] \end{aligned}$$

und damit

$$d(X+Y,\Phi) \leq d(X,\Phi) + \delta + P[|Y| > \delta].$$

Setzt man $\delta = \eta^{1/3}$, so gilt wegen $\eta \geq E[Y^2] \geq \delta^2 P[|Y| > \delta]$ die Abschätzung $P[|Y| > \delta] \leq \eta^{1/3}$, und daraus folgt Lemma 2.2.

□

In einem nächsten Schritt beweisen wir vorerst den CLT unter Zusatzbedingungen:

Satz 2.3 [CLT-Light] *Sind $X_1, X_2, \ldots$ unabhängige Zufallsvariable mit endlich vielen Werten, die für alle m die Bedingungen*

$$E[X_m] = 0, \quad V[X_m] = 1$$

und

$$|X_m| \leq B < \infty$$

erfüllen, und ist $S_n := X_1 + \ldots + X_n$ und $S_n^* := S_n/\sqrt{n}$, so gilt

$$d(S_n^*, \Phi) \to 0.$$

Beweis: Sei Y_0 eine von den X_i unabhängige $\mathcal{N}(0,1)$-verteilte Zufallsgrösse (auf einem geeigneten Wahrscheinlichkeitsraum lassen sich Y_0 und die X_i auch definieren). Sei N zunächst eine fest gewählte natürliche Zahl. Wir setzen für $n \geq N$

$$Z_n := \sqrt{\frac{N}{n}} Y_0 + \frac{1}{\sqrt{n}}(S_n - S_N).$$

Sei $F_n(x) := P[Z_n \leq x]$. Es gilt

$$Z_{n+1} = \sqrt{\frac{n}{n+1}} Z_n + \frac{1}{\sqrt{n+1}} X_{n+1}.$$

Sind $x_1, x_2, \ldots, x_k$ die möglichen Werte von X_{n+1} und $p_1, \ldots, p_k$ die zugehörigen Wahrscheinlichkeiten, so ist - wegen der Unabhängigkeit der letzten beiden Summanden -

$$\begin{aligned} F_{n+1}(x) &= \sum_{i=1}^{k} p_i P\left[\sqrt{\frac{n}{n+1}} Z_n \leq x - \frac{1}{\sqrt{n+1}} x_i\right] \\ &= \sum_{i=1}^{k} p_i F_n\left[\sqrt{\frac{n+1}{n}} x - \frac{1}{\sqrt{n}} x_i\right] \\ &= E\left[F_n\left[\sqrt{\frac{n+1}{n}} x - \frac{1}{\sqrt{n}} X_{n+1}\right]\right]. \end{aligned}$$

Sei $\alpha := \sqrt{(n+1)/n}$, $\beta := \frac{1}{\sqrt{n}}$. Wir erhalten

$$|F_{n+1}(x) - \Phi(x)| \leq |E[F_n(\alpha x - \beta X_{n+1}) - \Phi(\alpha x - \beta X_{n+1})]| + |E[\Phi(\alpha x - \beta X_{n+1}) - \Phi(x)]|$$

und damit

$$d(F_{n+1}, \Phi) \leq d(F_n, \Phi) + \sup_x |E[\Phi(\alpha x - \beta X_{n+1}) - \Phi(x)]|. \tag{2.2}$$

Die Taylorentwicklung von Φ (zB Forster Analysis I, p 175 mit Lagrange-Form des Restgliedes) liefert

$$\Phi(\alpha x - \beta x_i) = \Phi(\alpha x) - \beta x_i \Phi'(\alpha x) + \frac{(\beta x_i)^2}{2} \Phi''(\alpha x) - \frac{(\beta x_i)^3}{6} \Phi'''(\alpha x - \theta_i(x) \beta x_i)$$

mit $|\theta_i(x)| \leq 1$. Der letzte Term ist durch ein $O(n^{-3/2})$ beschränkt, denn $\Phi''' = \varphi''$ ist beschränkt, die $|x_i|$ sind durch B beschränkt; zudem gilt $\beta = n^{-1/2}$.

Es ist (Taylor) $\alpha = 1 + 1/(2n) + O(n^{-2})$. Φ und alle Ableitungen von Φ sowie $x\Phi'(x)$ und $x^2\Phi''(x)$ sind beschränkt. Die Taylorentwicklung von $\Phi(\alpha x)$ liefert

$$\Phi(\alpha x) = \Phi(x) + \left(\frac{1}{2n} + O(1/n^2)\right) x\Phi'(x) + \left(\frac{1}{2n} + O(1/n^2)\right)^2 \frac{x^2}{2} \Phi''\left(x + \theta x\left(\frac{1}{2n} + O(1/n^2)\right)\right)$$

mit $0 \leq |\theta| \leq 1$. Damit erhalten wir

$$\Phi(\alpha x) = \Phi(x) + \frac{x}{2n}\Phi'(x) + O(1/n^2).$$

Analog zeigt man

$$\Phi''(\alpha x) = \Phi''(x) + \frac{x}{2n}\Phi'''(x) + O(1/n^2).$$

Bildet man nun $E[\Phi(\alpha x - \beta X_{n+1})] = \sum_{i=1}^{k} \Phi(\alpha x - \beta x_i)p_i$ und berücksichtigt, dass

$$\sum x_i p_i = 0, \qquad \sum x_i^2 p_i = 1,$$

so ergibt sich

$$E[\Phi(\alpha x - \beta X_{n+1})] = \Phi(x) + \frac{x}{2n}\Phi'(x) + \frac{1}{2n}\Phi''(x) + O\left(\frac{1}{n^{3/2}}\right). \tag{2.3}$$

Der Schlüssel zum Beweis ist jetzt die Identität

$$\Phi''(x) = -x\Phi'(x),$$

durch Differenzieren von $\Phi'(x)$ als kleine Hausaufgabe verifizieren. In (2.3) heben sich damit die Glieder mit $\Phi'(x)$ und $\Phi''(x)$ gegenseitig auf. Damit erhalten wir für ein geeignetes, von n und N unabhängiges, $K > 0$ aus (2.2)

$$|d(F_{n+1}, \Phi) - d(F_n, \Phi)| \leq Kn^{-3/2}.$$

Wegen $F_N = \Phi$ ist $d(F_N, \Phi) = 0$ und somit für alle $n \geq N$

$$d(Z_n, \Phi) = d(F_n, \Phi) \leq K \sum_{i=N}^{\infty} \frac{1}{i^{3/2}}.$$

Wegen der Konvergenz der Reihe auf der rechten Seite können wir bei vorgegebenem $\epsilon > 0$ N so gross wählen, dass $d(Z_n, \Phi) < \epsilon/2$ für alle $n \geq N$. Nun gilt aber bei festem N für $n \to \infty$

$$E[(S_n^* - Z_n)^2] = E\left[\left(\sqrt{\frac{N}{n}}Y_0 - \frac{1}{\sqrt{n}}S_N\right)^2\right] \to 0.$$

Nach Lemma 2.2 folgt damit für hinreichend grosses n

$$d(S_n^*, \Phi) < \epsilon.$$

□

Damit können wir jetzt durch Approximation den CLT in der Form von Theorem 2.1 beweisen:

Theorem 2.1 [Zentraler Grenzwertsatz] *Sei $X_i, i \geq 1$, eine Folge von iid Zufallsgrössen mit $E[X_1] =: \mu$ und $0 < V[X_1] =: \sigma^2 < \infty$. Dann gilt für $a \in \mathbb{R}$ beliebig:*

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq a\right] \longrightarrow P[\mathcal{N}(0,1) \leq a]$$

wenn $n \to \infty$.

Beweis: Wie im Krengel vorgeführt, definieren wir hier gleich

$$S_n^* := \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma};$$

statt obiger Aussage beweisen wir, dass dann

$$d(S_n^*, \Phi) \to 0.$$

Daraus folgt natürlich sofort Theorem 2.1.

Sei $f_m(x) = k/m$ für $k/m \leq x < (k+1)/m$ mit $-m^2 \leq k \leq m^2$, und $= 0$ sonst. Sei ferner $Y_{m,i} = f_m(X_i), \mu_m = E[Y_{m,i}]$ und $\sigma_m^2 = E[(Y_{m,i} - \mu_m)^2]$. Es ist nicht schwer zu zeigen, dass

$$\mu_m \to \mu = E[X_1], \quad \sigma_m^2 \to \sigma^2$$

und

$$E((Y_{m,i} - X_i)^2) \to 0.$$

Die Zufallsvariablen $X_{m,i} := (Y_{m,i} - \mu_m)/\sigma_m$ erfüllen für festes m die Voraussetzungen von Satz 2.3. Definiert man also

$$S_{m,n}^* := \frac{1}{\sqrt{n}}(X_{m,1} + \ldots + X_{m,n}),$$

so gilt $d(S_{m,n}^*, \Phi) \to 0$ wenn $n \to \infty$. Es ist

$$S_n^* - S_{m,n}^* = \big(\frac{\sigma_m}{\sigma} - 1\big)S_{m,n}^* + \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n [(X_i - \mu) - (Y_{m,i} - \mu_m)].$$

Bezeichnet man die beiden Summanden auf der rechten Seite mit $A_{m,n}$ und $B_{m,n}$, so ist

$$E[A^2_{m,n}] = (1 - \frac{\sigma_m}{\sigma})^2$$

und

$$E[B^2_{m,n}] = \frac{1}{\sigma^2} V(X_1 - Y_{m,1}) \leq \frac{1}{\sigma^2} E((X_1 - Y_{m,1})^2),$$

denn die Terme [...] sind iid und haben Erwartungswert 0. Für grosses m ist daher $E[(S^*_n - S^*_{m,n})^2]$ klein. Mit Lemma 2.2 folgt damit die Behauptung.

$\square$

Überlegen Sie sich, welche Voraussetzungen zum CLT dringend notwendig sind und welche eher aufgehoben werden können in Verallgemeinerungen. Sehen Sie dazu bitte auch das erste Übungsblatt.