Matematikai statisztika
Témakör 3: Becslések nagymintás tulajdonságai

Elek Péter

Jelölések

- θ a becsülendő sokasági paraméter
- $\widehat{\theta}$ ennek becslőfüggvénye
- $\widehat{\theta}_n$ a jelölés, ha a mintaelemszámot hangsúlyozni kívánjuk

1. Aszimptotikus torzítatlanság

Aszimptotikus torzítatlanság

- $\widehat{\theta}_n$ a θ paraméter *aszimptotikusan torzítatlan* becslőfüggvénye, ha minden θ esetén $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$.
- (Ellenkező esetben $\widehat{\theta}_n$ aszimptotikusan torzított.)
- Példa: s^2 korrigálatlan tapasztalati variancia (mint a σ^2 sokasági variancia becslőfüggvénye)
  - $E\left(s^2\right) = \frac{n-1}{n}\sigma^2$, ezért s^2 torzított.
  - Azonban $\lim_{n\to\infty} E\left(s^2\right) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$, ezért s^2 aszimptotikusan torzítatlan.

2. Konzisztencia

2.1. Sztochasztikus konvergencia és konzisztencia

Konzisztencia

- $\widehat{\theta}_n$ a θ paraméter *konzisztens* becslőfüggvénye, ha *minden* $\varepsilon > 0$ esetén $\lim_{n\to\infty} \Pr\left(\left|\widehat{\theta}_n - \theta\right| < \varepsilon\right) = 1$.
  - Azaz $\widehat{\theta}_n$ mintavételi eloszlása egyre inkább "ráhúzódik" θ-ra.

- Ha ez legalább egy θ esetén nem teljesül, akkor $\widehat{\theta}_n$ inkonzisztens.

- Máshogyan, azt mondjuk, hogy $\widehat{\theta}_n$ *sztochasztikusan konvergens* és *sztochasztikus határértéke* θ.

- Jelölés: $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$ vagy $\widehat{\theta}_n \xrightarrow{p} \theta$

Sztochasztikus konvergencia (convergence in probability) (plim)

- A sztochasztikus konvergencia mindenféle v.v. esetén értelmezhető (és a határérték lehet nem konstans v.v. is).

- Legyen Z_n v.v.-k egy sorozata.

- Definíció: $\operatorname{plim}_{n\to\infty} Z_n = Z$ ha $\lim_{n\to\infty} \Pr\left(|Z_n - Z| < \varepsilon\right) = 1$ minden $\varepsilon > 0$ esetén.

- Megjegyzés:
  - $\lim$ definíciója az analízisben: $\lim_{n\to\infty} x_n = A$ ha minden $\varepsilon > 0$ esetén létezik olyan N_ε küszöb, hogy $n > N_\varepsilon$ esetén $|x_n - A| < \varepsilon$.
  - Valószínűségi változókra többféle konvergenciafogalom értelmezhető.

Konzisztencia elégséges feltétele

- Állítás: ha $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$ és $\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$, akkor $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$.
  - Azaz ha egy aszimptotikusan torzítatlan becslőfüggvény varianciája zérushoz tart, akkor az konzisztens.
  - Bizonyítás: Csebisev-egyenlőtlenség alapján

Példa: μ becslőfüggvényei

- A *mintaátlag* a μ várható érték konzisztens becslőfüggvénye.
  - Tudjuk, hogy $E\left(\overline{X}\right) = \mu$ and $Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$.
  - Ezért $\lim_{n\to\infty} E\left(\overline{X}\right) = \mu$ and $\lim_{n\to\infty} Var\left(\overline{X}\right) = 0$,
  - tehát $\overline{X}$ konzisztensen becsüli μ-t.

- X_1 első mintaelem a μ nem konzisztens becslőfüggvénye.
  - Mivel $E\left(X_1\right) = \mu$, a becslőfüggvény (aszimptotikusan) torzítatlan.
  - Azonban $n \to \infty$ esetén eloszlása változatlan, nem húzódik rá az ismeretlen paraméterre, tehát nem konzisztens.

2.2. Tulajdonságok és tételek

Nagy számok (gyenge) törvénye (NSZT)

- *Nagy számok (gyenge) törvénye (NSZT)*: Ha $X_1, X_2, \ldots, X_n$ f.a.e. minta egy μ várható értékű eloszlásból, akkor $\operatorname{plim}_{n\to\infty} \overline{X} = E(X) = \mu$.
  - (Korábban ezt láttuk akkor, ha σ^2 véges.)
  - Azonban akkor is igaz, ha σ^2 nem véges.

plim **tulajdonságai**

- Ha c konstans, akkor $\operatorname{plim}_{n\to\infty} c = c$.
- Ha $x_1, x_2, x_3, \ldots$ nem véletlen sorozat, akkor $\operatorname{plim}_{n\to\infty} x_n = \lim_{n\to\infty} x_n$.
- Ha $\operatorname{plim}_{n\to\infty} \widehat{\theta}_{1n} = \alpha$ és $\operatorname{plim}_{n\to\infty} \widehat{\theta}_{2n} = \beta$, akkor
  - $\operatorname{plim}_{n\to\infty} \left(\widehat{\theta}_{1n} + \widehat{\theta}_{2n}\right) = \alpha + \beta$,
  - $\operatorname{plim}_{n\to\infty} \left(\widehat{\theta}_{1n}\widehat{\theta}_{2n}\right) = \alpha\beta$
  - továbbá$\beta \neq 0$ esetén $\operatorname{plim}_{n\to\infty} \left(\widehat{\theta}_{1n}/\widehat{\theta}_{2n}\right) = \alpha/\beta$.

Folytonos transzformáció és konzisztencia

- Tétel ("Continuous mapping theorem" stb.): ha $\operatorname{plim}_{n\to\infty} \widehat{\theta}_n = \theta$ és $g(\cdot)$ *folytonos* függvény a θ helyen, akkor $\operatorname{plim}_{n\to\infty} g\left(\widehat{\theta}_n\right) = g(\theta)$.
  - Bizonyítás: $\lim$ tulajdonságai alapján
- Tehát ha $\widehat{\theta}_n$ konzisztensen becsüli θ-t és g folytonos, akkor $g\left(\widehat{\theta}_n\right)$ konzisztensen becsüli $g(\theta)$-t.
- Hasonló állítás nem igaz a torzítatlanságra!

2.3. Alkalmazások

Konzisztencia bizonyításának szokásos módszerei

- Bizonyíthatjuk, hogy $\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta$ *és* $\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$.
- Vagy valamilyen átlagra alkalmazhatjuk a NSZT-t, majd kihasználhatjuk a plim operátor egyéb tulajdonságait (pl. viselkedését folytonos leképezések esetén).

σ^2 becslőfüggvényei

- s^2 (korrigálatlan tapasztalati variancia) konzisztensen becsüli σ^2-et.
  - Tudjuk, hogy $s^2 = \left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) - \left(\overline{X}-\mu\right)^2$.
  - Első tagra: $\operatorname{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) = E\left[(X_i-\mu)^2\right] = \sigma^2$.
  - A második tagra $\operatorname{plim}_{n\to\infty}\left[\left(\overline{X}-\mu\right)^2\right] = 0$, mert $\operatorname{plim}_{n\to\infty}\overline{X} = \mu$ és így $\operatorname{plim}_{n\to\infty}\left(\overline{X}-\mu\right) = 0$, majd alkalmazhatjuk a folytonos leképezésekre vonatkozó tulajdonságot.
- s^{*2} (korrigált tapasztalati variancia) is konzisztens.
  - $s^{*2} = \frac{n}{n-1}s^2$ és $\operatorname{plim}_{n\to\infty}\frac{n}{n-1} = 1$

σ becslőfüggvényei

- s és s^* (korrigálatlan és korrigált tapasztalati szórás) konzisztensen becsüli σ-t (a sokasági szórást).
  - $\operatorname{plim}_{n\to\infty}\sqrt{s_n^2} = \sqrt{\sigma^2}$, azaz $\operatorname{plim}_{n\to\infty}s_n = \sigma$.
- Hasonló állítás nem igaz a torzítatlanságra: s^* torzított, bár s^{*2} torzítatlan.

További példák

- Példa 1: $\widehat{\mu}_1 = \frac{\sum_{i=1}^{n}X_i}{n-1}$. Konzisztens?
- Példa 2: $\widehat{\mu}_2 = wX_1 + (1-w)\frac{\sum_{i=2}^{n}X_i}{n-1}$. Konzisztens?
- Példa 3: Legyen $X_1,\ldots,X_n$ f.a.e. Bernoulli(p) eloszlású minta. Becsüljük $\frac{p}{1-p}$ értéket $\frac{\overline{X}}{1-\overline{X}}$ becslőfüggvénnyel. Torzítatlan? Konzisztens?

Megjegyzések a konzisztenciáról

- A fenti példák mutatják, hogy
  - konzisztens becslőfüggvények NEM feltétlenül torzítatlanok;
  - torzítatlan (illetve aszimptotikusan torzítatlan) becslőfüggvények NEM feltétlenül konzisztensek.

3. Aszimptotikus normalitás

3.1. Eloszlásbeli konvergencia és aszimptotikus normalitás

Eloszlásbeli konvergencia (convergence in distribution)

- A Z_n v. v.-k sorozata *eloszlásban konvergál* egy Z folytonos valószínűségi változóhoz, ha *minden* z valós szám esetén

$$\lim_{n\to\infty} \Pr\left(Z_n \leq z\right) = \Pr\left(Z \leq z\right).$$

- Intuíció: Z_n eloszlásának alakja egyre közelebb kerül Z eloszlásához
- Jelölés: $\xrightarrow{d}$
- A definíciót alkalmazhatjuk $\widehat{\theta}_n$ becslőfüggvényekre is.

Aszimptotikus normalitás

- Ha a határeloszlás normális, azaz $Z \sim N\left(\mu, \sigma^2\right)$, akkor Z_n-t *aszimptotikusan normális eloszlású*nak mondjuk.
  - Jelölés: $Z_n \overset{A}{\sim} N\left(\mu, \sigma^2\right)$.
- Ha $Z \sim N(0,1)$, akkor Z_n *aszimptotikusan standard normális eloszlású*. Ekkor nagy n esetén $\Pr\left(Z_n \leq z\right) \approx \Phi\left(z\right)$, ahol $\Phi(z)$ a standard normális eloszlásfüggvény.

3.2. Tulajdonságok és tételek

Eloszlásbeli konvergencia tulajdonságai

- *Folytonos leképezésekre vonatkozó tulajdonság* (continuous mapping theorem): Ha $Z_n \xrightarrow{d} Z$ és $g\left(\cdot\right)$ folytonos függvény, akkor $g\left(Z_n\right) \xrightarrow{d} g\left(Z\right)$.
- *Szluckij-lemma*: Ha $X_n \xrightarrow{d} X$ és $Y_n \xrightarrow{p} \alpha$ (ahol X egy eloszlás és α egy konstans), akkor
  - $X_n + Y_n \xrightarrow{d} X + \alpha$,
  - $X_n Y_n \xrightarrow{d} \alpha X$,
  - és ha $\alpha \neq 0$, akkor $X_n / Y_n \xrightarrow{d} X/\alpha$.
  - Fontos, hogy α *konstans* legyen.

Centrális határeloszlás-tétel (central limit theorem) (CHT, CLT)

- *Centrális határeloszlás-tétel*: Ha $X_1, \ldots, X_n$ f.a.e. minta egy olyan eloszlásból, amelyre $E(X_i) = \mu$ és $Var(X_i) = \sigma^2$, akkor

$$Z_n = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1).$$

- Korábbi eredmény: ha X_i *normális eloszlású*, akkor $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ pontosan teljesül minden n esetén. A CHT szerint *minden* véges szórású eloszlásra $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$.
- Tehát a normális eloszlás természetes határeloszlásként adódik.
- Megjegyzés: $\sqrt{n}$ a szükséges skálázó sorozat (mint általában az egyszerű statisztikai eljárások esetén).

Delta-módszer

- *Tétel*: Legyen $\sqrt{n}(Z_n - \theta) \xrightarrow{d} N\left(0, \sigma^2\right)$ és legyen $h()$ folytonosan differenciálható θ-ban úgy, hogy $h'(\theta) \neq 0$. Ekkor

$$\sqrt{n}\left(h(Z_n) - h(\theta)\right) \xrightarrow{d} N\left(0, \left[h'(\theta)\right]^2 \sigma^2\right).$$

- A becslőfüggvények aszimptotikus normalitásának belátása sokszor a delta-módszer és a CHT együttes alkalmazásával történhet.

3.3. Alkalmazások

A binomiális eloszlás normális közelítése

- Legyen $\Pr(X = 1) = p$, $\Pr(X = 0) = 1 - p$, és ebből $X_1, X_2, \ldots, X_n$ f.a.e. minta.
- Ekkor a CHT szerint $\frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0,1)$, azaz (kissé pongyola jelöléssel) $\overline{X} \overset{A}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$.
- Ezért az $Y = \sum_{i=1}^{n} X_i = n\overline{X}$ valószínűségi változóra, amely Binom (n,p) eloszlású, igaz, hogy

$$\frac{Y - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

  azaz $Y \overset{A}{\sim} N(np, np(1-p))$.

A mintabeli t-érték aszimptotikus normalitása

- A CHT alapján $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$, valamint a NSZT és a folytonos leképezésekre vonatkozó tulajdonság miatt $\frac{\sigma}{s^*} \xrightarrow{p} 1$.
- Ezért az eloszlásbeli konvergencia tulajdonságai miatt $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\frac{\sigma}{s^*} = \frac{\overline{X}-\mu}{s^*/\sqrt{n}} \xrightarrow{d} N(0,1)$.
  - Tehát a mintabeli t-érték nagy n esetén közelítően standard normális (még ha az alapeloszlás nem is normális).
- Megjegyzés:
  - Tudjuk, hogy normális minta esetén bármilyen n-re $\frac{\overline{X}-\mu}{s^*/\sqrt{n}} \sim t_{n-1}$. Ebből is következik, hogy $n \to \infty$ esetén $t_n \xrightarrow{d} N(0,1)$. Ez utóbbi egyébként a NSZT alapján is következik a t-eloszlás definíciójából.

s^2 aszimptotikus normalitása

- Használjuk a következő jelölést: $H = \sqrt{Var\left[(X_i-\mu)^2\right]}$.
- Ekkor

$$\begin{aligned}
s^2 &= \frac{\sum_{i=1}^n (X_i-\mu)^2}{n} - \left(\overline{X}-\mu\right)^2 \\
s^2-\sigma^2 &= \frac{\sum_{i=1}^n (X_i-\mu)^2}{n} - \sigma^2 - \left(\overline{X}-\mu\right)^2 \\
\sqrt{n}\frac{s^2-\sigma^2}{H} &= \sqrt{n}\frac{\sum_{i=1}^n (X_i-\mu)^2/n-\sigma^2}{H} - \frac{1}{\sqrt{n}}\frac{\sigma^2}{H}\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2.
\end{aligned}$$

- Itt az első tag $\xrightarrow{d} N(0,1)$, a második tag pedig $\xrightarrow{p} 0$.
- Így a Szluckij-lemma miatt $\sqrt{n}\frac{s^2-\sigma^2}{H} \xrightarrow{d} N(0,1)$, azaz $s^2 \overset{A}{\sim} N\left(\sigma^2, \frac{H^2}{n}\right)$.

s^{*2} aszimptotikus normalitása

- $s^{*2} = \frac{n}{n-1}s^2$. Tehát:

$$\sqrt{n}\frac{s^{*2}-\sigma^2}{H} = \frac{n}{n-1}\sqrt{n}\frac{s^2-\sigma^2}{H} + \sqrt{\frac{n}{(n-1)^2}}\frac{\sigma^2}{H}.$$

- Itt $\frac{n}{n-1} \xrightarrow{p} 1$, $\sqrt{n}\frac{s^2-\sigma^2}{H} \xrightarrow{d} N(0,1)$, $\sqrt{\frac{n}{(n-1)^2}}\frac{\sigma^2}{H} \xrightarrow{p} 0$.
- Tehát $\sqrt{n}\frac{s^{*2}-\sigma^2}{H} \xrightarrow{d} N(0,1)$ és így s^{*2} és s^2 aszimptotikus eloszlása megegyezik.

Aszimptotikus eloszlások normális minta esetén

- Ha X_i normális eloszlású, akkor
$$H^2 = E\left[(X_i-\mu)^4\right] - E^2\left[(X_i-\mu)^2\right] = 2\sigma^4,$$
- ezért
$$\frac{s^2-\sigma^2}{\sqrt{\frac{2\sigma^4}{n}}} = \sqrt{\frac{n}{2}}\frac{s^2-\sigma^2}{\sigma^2} \xrightarrow{d} N(0,1),$$
- azaz
$$s^2 \overset{A}{\sim} N\left(\sigma^2, \frac{2\sigma^4}{n}\right).$$
- Ugyanez igaz s^{*2}-re is.

Példa: mintaátlag aszimptotikus eloszlása

- Egy elemgyártó azt állítja, hogy az általa gyártott elemek várható élettartama $\mu = 54$ hónap, $\sigma = 30$ hónap szórással. Ezen állítás ellenőrzésére $n = 400$ elemű véletlen mintát veszünk: $X_1, X_2, \ldots, X_{400}$.
- Ha a gyártó igazat állít, akkor mi $\overline{X}$ mintavételi eloszlása? Mennyi annak a valószínűsége, hogy $\overline{X} \leq 50$ hónap?
- Megjegyzés: mit jelent a *"nagy mintaelemszám"*?
  - Ez függ a becslőfüggvénytől és a sokasági eloszlástól is.
  - Általában a sokasági várható érték becslésekor $n > 120$ (vagy akár $n > 60$) elég nagynak számít ahhoz, hogy a közelítések működjenek.

Példa: mintabeli arány aszimptotikus eloszlása

- Legyen X Bernoulli-eloszlású $p = 0.2$ paraméterrel.
- Ebből az eloszlásból $n = 400$ elemű f.a.e. mintánk van: $X_1, X_2, \ldots, X_{400}$.
- Mennyi a valószínűsége, hogy az $\overline{X}$ mintabeli arány 0.17 és 0.23 közé esik?
- Megjegyzés: szokásos gyakorlati feltétel a binomiális eloszlás nagymintás közelítésére: $np > 10$ és $n(1-p) > 10$.

4. Fontosabb becslőfüggvények összefoglalása

Mintaátlag: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

Jellemző	Normális minta	Általános minta
$E(\overline{X})$	μ	μ
$Var(\overline{X})$	σ^2/n	σ^2/n
$\operatorname{plim} \overline{X}$	μ	μ
$\overline{X}$ kismintás eloszlása	$\sim N\left(\mu, \frac{\sigma^2}{n}\right)$	általános
$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ kismintás eloszlása	$\sim N(0,1)$	általános
$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ aszimpt. eloszlása	$\sim N(0,1)$	$\xrightarrow{d} N(0,1)$
$\frac{\overline{X}-\mu}{s^*/\sqrt{n}}$ kismintás eloszlása	$\sim t_{n-1}$	általános
$\frac{\overline{X}-\mu}{s^*/\sqrt{n}}$ aszimpt. eloszlása	$\xrightarrow{d} N(0,1)$	$\xrightarrow{d} N(0,1)$

Sokasági variancia becslőfüggvényei normális eloszlású minta esetén

Jellemző	s^2	s^{*2}
E	$\frac{n-1}{n}\sigma^2$	σ^2
Var	$\frac{2\sigma^4(n-1)}{n^2}$	$\frac{2\sigma^4}{n-1}$
plim	σ^2	σ^2
kismintás eloszlás	$\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{(n-1)s^{*2}}{\sigma^2} \sim \chi^2_{n-1}$
aszimpt. eloszlás	$\frac{s^2-\sigma^2}{\sigma^2\sqrt{2/n}} \xrightarrow{d} N(0,1)$	$\frac{s^{*2}-\sigma^2}{\sigma^2\sqrt{2/n}} \xrightarrow{d} N(0,1)$

Sokasági variancia becslőfüggvényei általános eloszlású minta esetén

Jellemző	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$	$s^{*2} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$
E	$\frac{n-1}{n}\sigma^2$	σ^2
Var	általános	általános
plim	σ^2	σ^2
kismintás eloszlás	általános	általános
aszimpt. eloszlás	$\frac{s^2-\sigma^2}{H/\sqrt{n}} \xrightarrow{d} N(0,1)$	$\frac{s^{*2}-\sigma^2}{H/\sqrt{n}} \xrightarrow{d} N(0,1)$

Tananyag

- Jegyzet, feladatok és házi feladatok
- Wooldridge Appendix C.3
- Amemiya 6 (kivéve definition 6.1.3, Theorem 6.1.1, Theorem 6.2.3, Example 6.4.2) és section 7.2.6