ו. תהא $X_1,\dots,X_n \overset{iid}{\sim} F$ כאשר F רציפה. בדומה ל-$F(x)$ פונקציית הסימן משאלה 2.

ראינו ב-$\hat F_n$ כאומד את F של ה-MISE (התנהגות שווה ל-$\frac{1}{6}$ (עבור F רציפה).

נניח כי מודל F שייך למשפחה פרמטרית $(F_\lambda)_{\lambda\in\mathbb{R}}$ כך ש-F_λ זה התפלגות $Exp(\lambda)$.

חשבו את הגבול של MISE במקרה הזה כאשר $n\to\infty$ עבור האומד $F_{\hat\lambda_n}$ כאשר $\hat\lambda_n$ MLE

פתרון:

כזכור, האומד של λ הוא $\hat\lambda_n=\frac{1}{\bar X_n}$. לכן

$$\mathbb{E}_\lambda\int_0^\infty n\left(F_{\hat\lambda_n}(x)-F(x)\right)^2 dF(x) = n\,\mathbb{E}_\lambda\int_0^\infty\left(e^{-\hat\lambda_n x}-e^{-\lambda x}\right)^2\lambda e^{-\lambda x}\,dx$$

$X\sim Exp(\lambda)$

$\cancel{\lambda\mathbb{E}_\lambda X=\mathbb{E}_1 X}$

$$\cancel{= n\,\mathbb{E}_1\int_0^\infty\left(e^{-\hat\lambda_n s}-e^{-s}\right)^2 e^{-s}\,ds}$$

מוה. לא במלה הגוש

$$= n\lambda\,\mathbb{E}_\lambda\int_0^\infty\left(e^{-2\hat\lambda_n x}-2e^{-\hat\lambda_n x-\lambda x}+e^{-2\lambda x}\right)e^{-\lambda x}\,dx$$

$$= n\lambda\,\mathbb{E}_\lambda\int_0^\infty e^{-x(\lambda+2\hat\lambda_n)}-2e^{-x(2\lambda+\hat\lambda_n)}+e^{-3\lambda x}\,dx$$

$$= n\lambda\,\mathbb{E}_\lambda\left(\frac{1}{\lambda+2\hat\lambda_n}-2\cdot\frac{1}{2\lambda+\hat\lambda_n}+\frac{1}{3\lambda}\right) = n\lambda\,\mathbb{E}_\lambda\left(\frac{2\lambda+\hat\lambda_n-2\lambda-4\hat\lambda_n}{(\lambda+2\hat\lambda_n)(2\lambda+\hat\lambda_n)}+\frac{1}{3\lambda}\right)$$

$$= n\lambda\,\mathbb{E}_\lambda\frac{-9\hat\lambda_n\lambda+2\lambda^2+\lambda\hat\lambda_n+4\lambda\hat\lambda_n+2\hat\lambda_n^2}{3\lambda(\lambda+2\hat\lambda_n)(2\lambda+\hat\lambda_n)} = n\cdot\frac{2}{3}\mathbb{E}_\lambda\frac{(\lambda-\hat\lambda_n)^2}{(\lambda+2\hat\lambda_n)(2\lambda+\hat\lambda_n)}$$

$$2\lambda^2-4\lambda\hat\lambda_n+2\hat\lambda_n^2=2(\hat\lambda_n-\lambda)^2$$

$$(\lambda-\hat\lambda_n)^2=\left(\lambda-\frac{1}{\bar X_n}\right)^2=\left(\frac{\lambda\bar X_n-1}{\bar X_n}\right)^2$$

$$\lambda+2\hat\lambda_n=\lambda+2\cdot\frac{1}{\bar X_n}=\frac{\lambda\bar X_n+2}{\bar X_n}$$

$$2\lambda+\hat\lambda_n=2\lambda+\frac{1}{\bar X_n}=\frac{2\lambda\bar X_n+1}{\bar X_n}$$

$$n\frac{2}{3}\mathbb{E}_\lambda\frac{(\lambda\bar X_n-1)^2}{(\lambda\bar X_n+2)(2\lambda\bar X_n+1)}$$

$$\cancel{\le n\cdot\frac{2}{3}\mathbb{E}_\lambda(\lambda\bar X_n-1)^2 = \frac{2}{3}\mathbb{E}_\lambda(\bar X_n-1/\lambda)^2\cdot\lambda^2=\frac{2}{3}n\lambda^2\cdot Var(\bar X_n)=\frac{2}{3}\cdot n\lambda^2\frac{1}{\lambda^2 n}=\frac{2}{3}}$$

• נסמן:

$$Y_n=\frac{2}{3}\cdot\frac{\left(\sqrt{n}(\bar X_n-1/\lambda)\cdot\lambda\right)^2}{(\lambda\bar X_n+2)(2\lambda\bar X_n+1)}$$

עפ"י הלמה של סלוצקי ומשפט הגבול המרכזי, מתקיים ש

$$Y_n\xrightarrow[n\to\infty]{d}\frac{2}{27}\cdot Z^2\qquad ;Z\sim N(0,1)$$

נרצה להחליף סדר גבול ואינטגרל. לשם כך נראה ש- Y_n אינטגרבילית במ"ש
מבחן קריטריון de la Vallée-Poussin.

נשים לב:

$$\mathbb{E}_\lambda Y_n^2 \le \mathbb{E}_\lambda\left(\lambda^2\left(\sqrt{n}\left(\bar{X}_n-\frac{1}{\lambda}\right)\right)^2\right)^2 = \lambda^4\,\mathbb{E}_\lambda\, n^2\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\frac{1}{\lambda}\right)\right)^4$$

$$= \frac{\lambda^4}{n^2}\,\mathbb{E}_\lambda\left(\sum_{i=1}^{n}\left(X_i-\frac{1}{\lambda}\right)\right)^4$$

$\cancel{\Gamma(n,\lambda)\quad \sum_{i=1}^{n} X_i \quad \frac{n}{\lambda}}$

$\cancel{\mathbb{E}\Lambda^n = \theta^n\frac{\Gamma(n+k)}{\Gamma(k)}}$

$K=\alpha$

$\beta=\frac{1}{\theta}$

$\cancel{\mathbb{E}\Lambda^n = \left(\frac{1}{\beta}\right)^n\frac{\Gamma(n+\alpha)}{\Gamma(\alpha)}}$

לפי נוסחת מולטינום, סכום בינומים ולאחר שהתוחלות מתאפסות.

$$= \frac{\lambda^4}{n^2}\left[n\,\mathbb{E}_\lambda\left(X_1-\frac{1}{\lambda}\right)^4 + 3n(n-1)\underbrace{Var(X_1)}_{\frac{1}{\lambda^2}}{}^2\right]$$

$\frac{144}{\lambda^4}$ $\qquad !n = \left[\frac{n!}{e}\right]$

$$= \left[\frac{4!}{e}\right]\cdot\frac{1}{n} + \frac{3}{n}(n-1) = 3+\frac{8}{n}-\frac{3}{n} = 3+\frac{5}{n} < 10$$

$8 =$

de la Vallee Poussin:

(Y_n) אינטגרבילית במ"ש אם קיימת φ כך ש-

$$\lim_{x\to\infty}\frac{\varphi(x)}{x}\to\infty$$

כך ש- $\sup \mathbb{E}\varphi(|Y_n|)<\infty$.

לכן ניתן להחליף גבול ואינטגרל והגענו מהתרגיל ל- $1-\frac{2}{2\pi}$ כאשר $n\to\infty$.

תרגול #2 מודלים ב׳ - 15-16.3.22

שאלות רלוונטיות: 4,5,6 מפרק 1
1,2,3,4 מפרק 2

פרק 1 - אומד אי-פרמטרי ל-F:

שאלה 4:

בשאלה זו נחקור את הפיזור:

$$D_n = \sqrt{n}\sup_{x\in\mathbb{R}}|\hat{F}_n(x) - F(x)|$$

כאשר F בדידה עם תומך ב-$\{x_1,\dots,x_k\}$ והסתברויות $\{p_1,\dots,p_k\}$.

(א) האם ההתפלגות של D_n תלויה רק ב-x_j-ים ולא ב-p_j-ים כאשר n קבוע?

פתרון:

במקרה הנ״ל, אם ל-$\hat{F}_n$ יש קפיצות ב-x_j-ים והגבהים המתאימים ל-F הם

$$\hat{p}_j(n) = \frac{1}{n}\sum_{m=1}^{n} 1\{X_m = x_j\} \quad ; \quad j=1,\dots,K$$

ולכן

יש sup על קבוצה סופית ולכן זה max.

$$D_n = \sqrt{n}\max_{1\le j\le k}|\hat{p}_j - p_j|$$

גם בשביל שעבר הוכחנו ש- $n\hat{p}\sim Mult(n,p)$. התפלגות זו לא תלויה ב-x_j-ים ~~בכלל~~ אלא בנוסף ב-p_j-ים, ~~[illegible]~~. אם כן, הערכים ש-D_n יכול לקבל ~~[illegible]~~ לא תלויים ב-x_j-ים, אבל כן תלויים בעיקרון ב-p_j-ים. לכן ההתפלגות של D_n תלויה ב-$\underline{p}$.
גם הפתרון דרך התלות.

הערה!

* למעשה, רק ההסתברויות שמתאימות להתפלגות של D_n, יכולות להיות תלויות ב-$\underline{p}$.

לדוגמה, עבור $k=2$,

$$D_n = \sqrt{n}\left|\frac{1}{n}\sum_{m=1}^{n} 1\{X_m = x_1\} - p_1\right|$$

כיוון ש- $D_n = \sqrt{n}\max\{|\hat{p}_1 - p_1|, |\hat{p}_2 - p_2|\}$ אז

$\hat{p}_2 = 1-\hat{p}_1$
$p_2 = 1-p_2$
ולכן יש את אותו ביטוי.

אבל לעומת ההסתברות שכל המידע שווה ל-x_1 היא:

$$P_F(D_n = \sqrt{n}(1-p_1)) = p_1^n$$

Scanned with CamScanner

ג) האם ההתפלגות של D_n מתכנסת כאשר $n\to\infty$? אם כן, האם ההתפלגות הגבולית תלויה

פתרון:

לפי CLT רב ממדי, מתקיים

מטריצת cov של התפלגות $Mult(n,p)$

$$\sqrt{n}(\hat{p}(n)-p) \xrightarrow[n\to\infty]{d} Y = N(0, diag(p)-pp^T) =$$

$$\max(f(x),g(x)) = \frac{1}{2}(f(x)+g(x)-|f(x)-g(x)|)$$

כעת שהמיפוי $x \longmapsto \max_{1\le j\le k} x_j$, $x\in\mathbb{R}^k$ רציף, אז

פונק' ליניארית. של פונקציות רציפות.

$$D_n \xrightarrow[n\to\infty]{d} \max_{j=1,\dots,k} |Y_j|$$

✓ מתכנסת.

ההתפלגות הגבולית יכולה להיות תלויה ב-p. למשל, עבור $k=2$,

$$D_n \xrightarrow[n\to\infty]{d} |Z| \quad , Z\sim N(0, p_1(1-p_1)) .$$

הערה: ההתפלגות הגבולית של $\hat{p}$ מאוד שונה מהמקרה הכללי שראינו, כאשר לא יצאה אי תלות ב-F רציפה/כללית.

שאלה 5: הוכיחו כי

$$(F(X_1),\dots,F(X_n)) \overset{d}{=} (V_1,\dots,V_n)$$

כאשר F רציפה ו- $V_1,\dots,V_n$ iid $U(0,1)$

פתרון: (סימון) $V := F(X)$. כמובן שזה מספיק כי הם iid.

דרך א':

$$\forall x\in(0,1)$$

$$\mathbb{P}_F(F(X)\le x) = \int_{\mathbb{R}} 1_{\{F(s)\le x\}}\, dF(s) \overset{\text{רציפות}}{=} \int_0^1 1_{\{u\le x\}}\, du = \int_0^x du = x$$

$$\mathbb{E}_F 1_{\{F(X)\le x\}}$$

☒

דרך ב':

$$\mathbb{P}_F(V\le x) = \mathbb{P}_F(F(X)\le x) = \int_{-\infty}^{\infty} 1_{\{F(s)\le x\}}\, dF(s) = \underbrace{1_{\{F(s)\le x\}} F(s)\Big|_{-\infty}^{\infty}}_{\xrightarrow[s\to\pm\infty]{} 0} - \int_{\mathbb{R}} F(s)\, d1_{\{F(s)\le x\}}$$

$s \longmapsto 1_{\{F(s)\le x\}}$ קופץ מ-1 ל-0 בנקודה $x^* = \inf\{s\in\mathbb{R} : F(s) > x\}$ ולכן

$$\mathbb{P}_F(V\le x) = -\int_{-\infty}^{\infty} F(s)\, d1_{\{F(s)\le x\}}$$

6:

לה כי נתקיים תכונות של D_n עבור התפלגויות רציפות.

(א) הראו ש

* נוסחה לחישוב בפועל.

$$D_n = \sqrt{n}\cdot\max_{1\le j\le n}\left(\left|\frac{j}{n}-F(X_{(j)})\right| \vee \left|\frac{j-1}{n}-F(X_{(j)})\right|\right)$$

כאשר $X_{(j)}$ הוא סטטיסטי הסדר ה-j $X_{(1)}\le\dots\le X_{(n)}$.

פתרון:

$$\frac{1}{\sqrt{n}}D_n = \sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\le x\}}-F(x)\right| = \sup_{u\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{F(X_i)\le u\}}-u\right|$$

$$\overset{\text{לא משנה מה נוכח סדר סוכמים.}}{=} \sup_{u\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{F(X_{(i)})\le u\}}-u\right|$$

נשים לב שהפונקציה בפנים היא ליניארית מקוטעים $[F(X_{(j-1)}),F(X_{(j)}))$ $(u\in)$ (כפונקציה של u).
ויתכנו קפיצות ב-$F(X_j)$-ים.

לכן הסופרמום יכול להתקבל רק בקצוות, כלומר בנקודות הקפיצה.
כיוון ש-F רציפה מימין עם גבולות משמאל, אז או שהוא מושג ב-$F(X_{(j)})$, (כאשר רציף מימין) או בגבול משמאל $F(X_{(j)}-)$. בכל אופן, יש רק מספר סופי של קצוות, לכן

$$\frac{1}{\sqrt{n}}D_n = \max_{1\le j\le n}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{F(X_{(i)})\le F(X_{(j)})\}}-F(X_{(j)})\right| \vee \left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{F(X_{(i)})<F(X_{(j)})\}}-F(X_{(j)})\right|$$

רציף מימין

גבול משמאל.

מהמונו- טוניות, וכיוון ש-F לא יורדת ורציפה

$$= \max_{1\le j\le n}\left(\left|\frac{j}{n}-F(X_{(j)})\right| \vee \left|\frac{j-1}{n}-F(X_{(j)})\right|\right)$$

תרגול #4 מודלים ב' 29-30.3.22

Bootstrap:

* שיטה לאמידת התפלגות של סטטיסטי או פונקציה של סטטיסטי. (למשל הטיה, סיכון, שונות וכו').
משמש די לבניית ר"ס ומבחנים סטטיסטיים.

* יהיו $X_1,\dots,X_n \overset{iid}{\sim} F$ ו- $\hat{F}_n$ פונ' ההתפלגות האמפירית. (נסתכל על

$$\mathbb{E}_F \Psi(F,\hat{F}_n,t)=0 \quad , t\in\mathbb{R}$$

ונרצה למצוא אומד $\hat{t}_n$ לפתרון המשוואה t. דרך פתרון תהיה Empirical Bootstrap:

← נחליף את F ב- $\hat{F}_n$ ואת $\hat{F}_n$ ב- $\hat{F}_n^*$, כאשר $\hat{F}_n^*$ היא פונ' ההתפלגות האמפירית של מדגם Bootstrap $X_1^*,\dots,X_n^* \overset{iid}{\sim} \hat{F}_n$

בפועל - דגימה בגודל אחיד עם החזרה מתוך $X_1,\dots,X_n$.

← מקבלים את משוואת ה- Bootstrap:

$$\mathbb{E}_{\hat{F}_n}\Psi(\hat{F}_n,\hat{F}_n^*,t)=0$$

בהרצאה עסקנו בדוגמה של אמידת הטיה:

$$\Psi(F,\hat{F}_n,t)=T(\hat{F}_n)-T(F)-t$$

ובמקרה הזה:

$$\hat{t}_n=\mathbb{E}_{\hat{F}_n}T(\hat{F}_n^*)-T(\hat{F}_n)$$

ונחשב את התוחלת כאשר $\underline{X}$ הן הנתונים.

← אומד מתוקן הטיה: $\hat{T}_n^B=T(\hat{F}_n)-\hat{t}_n$

היום: - אמידת שונות בעזרת Bootstrap

- אמידת התפלגות Bootstrap של המומנט האמפירי כאומד יחיד.

פרק 1.2 · אמידת שונות:

* נסתכל על בעיית אמידת השונות של $T(\hat{F}_n)$ (אומד הצבה ל-$T(F)$).

$$Var_F(T(\hat{F}_n)) = \mathbb{E}_F T(\hat{F}_n)^2 - \mathbb{E}_F^2 T(\hat{F}_n)$$

$\Leftarrow$ $F \to \hat{F}_n$, $\hat{F}_n \to \hat{F}_n^*$ ולכן

$$\hat{t}_n = \mathbb{E}_{\hat{F}_n} T(\hat{F}_n^*)^2 - \mathbb{E}_{\hat{F}_n}^2 T(\hat{F}_n^*)$$

תרגיל: (דוגמה 1.5 ברשומות)

~~מצאו~~ חשבו את אומד הבוטסטרפ לשונות עבור הפונקציונל

$$T(F) = \mu_F^2$$

פתרון:

תזכורת:

$$\mathbb{E}_{\hat{F}_n} X_i^* = \sum_{i=1}^{n} \frac{1}{n} \cdot X_i = \bar{X}_n \quad ; \quad Var_{\hat{F}_n}(X_i^*) = \mathbb{E}_{\hat{F}_n}(X_i^* - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

(סימון לצורך שימוש עתידי)

$$\hat{m}_p = \mathbb{E}_{\hat{F}_n}(X_i^* - \bar{X}_n)^p = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^p$$

$$\hat{t}_n = \mathbb{E}_{\hat{F}_n} \mu_{\hat{F}_n^*}^4 - \left(\mathbb{E}_{\hat{F}_n} \mu_{\hat{F}_n^*}^2\right)^2 = \underbrace{\mathbb{E}_{\hat{F}_n}(\bar{X}_n^*)^4}_{\text{I}} - \underbrace{\left[\mathbb{E}_{\hat{F}_n}(\bar{X}_n^*)^2\right]^2}_{\text{II}}$$

II $\quad \mathbb{E}_{\hat{F}_n}(\bar{X}_n^*)^2 = Var_{\hat{F}_n}(\bar{X}_n^*) + \left(\mathbb{E}_{\hat{F}_n}\bar{X}_n^*\right)^2 = \frac{Var_{\hat{F}_n}(X_1^*)}{n} + (\bar{X}_n)^2 = \underline{\frac{1}{n}\hat{m}_2 + (\bar{X}_n)^2}$

I $\quad \mathbb{E}_{\hat{F}_n}(\bar{X}_n^*)^4 = \mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n + \bar{X}_n)^4$

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 - 4ab^3 + b^4$$

$$= \underbrace{\mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n)^4}_{B} + 4\bar{X}_n \underbrace{\mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n)^3}_{A} + 6\bar{X}_n^2 \underbrace{\mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n)^2}_{= Var_{\hat{F}_n}(\bar{X}_n^*) = \frac{1}{n}\hat{m}_2} - 4\bar{X}_n^3 \underbrace{\mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n)}_{=0} + \underline{\bar{X}_n^4}$$

$$A: \quad \mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n)^3 = \mathbb{E}_{\hat{F}_n}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i^* - \bar{X}_n)\right]^3$$

$$= \frac{1}{n^3}\sum_{i=1}^{n} \mathbb{E}_{\hat{F}_n}(X_i^* - \bar{X}_n)^3 = \frac{1}{n^3} \cdot n \cdot \hat{m}_3 = \underline{\frac{1}{n^2}\hat{m}_3}$$

$$B: \quad \mathbb{E}_{\hat{F}_n}(\bar{X}_n^* - \bar{X}_n)^4 = \mathbb{E}_{\hat{F}_n}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i^* - \bar{X}_n)\right]^4 = \frac{1}{n^4} \cdot n \underbrace{\mathbb{E}_{\hat{F}_n}(X_i^* - \bar{X}_n)^4}_{\hat{m}_4} + \frac{1}{n^4}\cancel{\binom{4}{2}} 3\frac{n(n-1)}{2} \underbrace{\mathbb{E}_{\hat{F}_n}^2(X_i^* - \bar{X}_n)^2}_{\hat{m}_2^2}$$

$$= \frac{1}{n^3}\hat{m}_4 + \frac{3}{n^3}(n-1)\hat{m}_2^2$$

1:

נוכיח כי אומדי הבוטסטראפ לתוחלת μ_F המבוססים על המשוואות-

$$\mu_F - t = 0$$

$$\mathbb{E}_F \mu_{\hat{F}_n} - t = 0$$

שווים שניהם לממוצע האמפירי $\bar{X}_n$.

פתרון:

עבור המשוואה הראשונה, נציב $F \to \hat{F}_n$:

$$\mu_{\hat{F}_n} - t = 0 \quad \text{ולכן}$$

$$\hat{t}_n = \mu_{\hat{F}_n} = \bar{X}_n.$$

עבור המשוואה השנייה, $F \to \hat{F}_n$, $\hat{F}_n \to \hat{F}_n^*$:

$$\hat{t}_n = \mathbb{E}_{\hat{F}_n} \mu_{\hat{F}_n^*} = \mathbb{E}_{\hat{F}_n} \bar{X}_n^* = \mathbb{E}_{\hat{F}_n} X_1^* = \bar{X}_n.$$

■

שאלה 2: עבור מדגם בגודל $n=3$ ובהנחה שהתצפיות שונות זו מזו, מצאו את התפלגות הבוטסטראפ של הממוצע האמפירי.

פתרון:

נעניין למצוא את $\mathbb{P}_{\hat{F}_n}(\bar{X}^* = x)$ עבור כל x רלוונטי.
נסמן את המדגם (X_1, X_2, X_3) ואת מדגם הבוטסטראפ $X^* = (X_1^*, X_2^*, X_3^*)$
X^* מקבל ערכים מתוך אחיד על $\{X_1, X_2, X_3\}^3$. כלומר יש $3^3 = 27$ וקטורים אפשריים.
עכשיו נציין לעבור על כל הסדרים, ש-$\bar{X}^*$ יכול לקבל על פני הגדלים האפשריים, ולחשב את ההסתברות:

X^*	סידורים אפשריים כמה פעמים	Value x of $\bar{X}^*$	$\mathbb{P}_{\hat{F}_n}(\bar{X}^* = x)$
X_1, X_1, X_1	1	X_1	$1/27$
X_2, X_2, X_2	1	X_2	$1/27$
X_3, X_3, X_3	1	X_3	$1/27$
X_1, X_1, X_2	3	$\frac{2}{3}X_1 + \frac{1}{3}X_2$	$3/27 = 1/9$
X_1, X_1, X_3	3	$\frac{2}{3}X_1 + \frac{1}{3}X_3$	$1/9$
X_2, X_2, X_1	3	$\frac{2}{3}X_2 + \frac{1}{3}X_1$	$1/9$
X_2, X_2, X_3	3	$\frac{2}{3}X_2 + \frac{1}{3}X_3$	$1/9$
X_3, X_3, X_1	3	$\frac{2}{3}X_3 + \frac{1}{3}X_1$	$1/9$
X_3, X_3, X_2	3	$\frac{2}{3}X_3 + \frac{1}{3}X_2$	$1/9$
X_1, X_2, X_3	6	$\bar{X}_n$	$6/27 = 2/9$

■

Scanned with CamScanner

המשך: לכן סה"כ

$$\hat{t}_n = \frac{1}{n^3}\hat{m}_4 + \frac{3}{n^3}(n-1)\hat{m}_2^2 + 6\bar{X}_n^2\cdot\frac{1}{n}\hat{m}_2 + \cancel{\bar{X}_n^4} - \left(\frac{1}{n}\hat{m}_2 + \bar{X}_n^2\right)^2 + \frac{1}{n^2}\hat{m}_3\cdot 4\cdot\bar{X}_n$$

$$\frac{3}{n^3}(n-1)\hat{m}_2^2 = \frac{3}{n^2}\hat{m}_2^2 - \frac{3}{n^3}\hat{m}_2^2$$

$$-\left(\frac{1}{n}\hat{m}_2 + \bar{X}_n^2\right)^2 = -\frac{1}{n^2}\hat{m}_2^2 - \frac{2}{n}\hat{m}_2\bar{X}_n^2 - \cancel{\bar{X}_n^4}$$

$$\cancel{= \frac{1}{n^3}\left[\hat{m}_4 + 3(n-1)\hat{m}_2^2\right]}$$

$$= \frac{1}{n^3}\left[\hat{m}_4 - 3\hat{m}_2^2\right] + \frac{1}{n^2}\left[2\hat{m}_2^2 + 4\hat{m}_3\bar{X}_n\right] + \frac{1}{n}\cdot 4\bar{X}_n^2\hat{m}_2$$

וזה האומד שלנו לשונות של $T(\hat{F}_n)$.

מצד שני, ממקודם:

$$Var_F(T(\hat{F}_n)) = Var_F(\bar{X}_n^2) = 4\mu^2_F\sigma^2_F\cdot\frac{1}{n} + O(n^{-3/2})$$

כלומר האומד שלנו מדויק לפחות עבור הביטוי בסדר גודל $O(\frac{1}{n})$. ∎

Scanned with CamScanner

3:

...או ביטוי לאומד הבוטסטרפ לשונות עבור אומד ההצבה לפונקציונל

$$T(F) = \int \psi dF$$

כאשר ψ פונקציה ידועה.

פתרון:

נסתכל על

$$Var_F(T(\hat{F}_n)) = \mathbb{E}_F T(\hat{F}_n)^2 - \mathbb{E}_F^2 T(\hat{F}_n)$$

ובמקרה הנ"ל,

$$\hat{t}_n = \mathbb{E}_{\hat{F}_n} T^2(\hat{F}_n^*) - \mathbb{E}^2_{\hat{F}_n} T(\hat{F}_n^*)$$

ומתקיים

$$T(\hat{F}_n^*) = \frac{1}{n}\sum_{i=1}^n \psi(X_i^*)$$

נחשב את התוחלות:

① $$\mathbb{E}_{\hat{F}_n} T(\hat{F}_n^*) = \mathbb{E}_{\hat{F}_n} \frac{1}{n}\sum_{i=1}^n \psi(X_i^*) = \mathbb{E}_{\hat{F}_n} \psi(X_1^*) = \frac{1}{n}\sum_{i=1}^n \psi(X_i) =: \bar{\psi}_n$$

② $$\mathbb{E}_{\hat{F}_n} T(\hat{F}_n^*)^2 = Var_{\hat{F}_n}(T(\hat{F}_n^*)) + \mathbb{E}^2_{\hat{F}_n} T(\hat{F}_n^*)$$

$$= \frac{1}{n} Var_{\hat{F}_n}(\psi(X_1^*)) + \bar{\psi}_n^2$$

$$Var_{\hat{F}_n}(\psi(X_1^*)) = \mathbb{E}_{\hat{F}_n}(\psi(X_1^*) - \bar{\psi}_n)^2 = \frac{1}{n}\sum_{i=1}^n (\psi(X_i) - \bar{\psi}_n)^2 = \underbrace{\frac{1}{n}\sum_{i=1}^n \psi^2(X_i)}_{\overline{\psi^2_n}} - \bar{\psi}_n^2$$

$$= \frac{1}{n}\overline{\psi^2_n} + \left(1 - \frac{1}{n}\right)\bar{\psi}_n^2$$

$$\Rightarrow \hat{t}_n = \frac{1}{n}\overline{\psi^2_n} + \left(1 - \frac{1}{n}\right)\bar{\psi}_n^2 - \bar{\psi}_n^2 = \frac{1}{n}\left(\overline{\psi^2_n} - \bar{\psi}_n^2\right)$$

תן ברמת $1-\alpha$ ביטחון ה-Bootstrap עבור $T(F)=F(x_0)$ עבור x_0 נקודה עניינית.

פתרון:

במקרה הנ״ל, נשאלת ה-Bootstrap הוא

$$\mathbb{P}_{\hat{F}_n}\left(\left|\hat{F}_n(x_0)-\hat{F}_n^*(x_0)\right|\le t\right)=1-\alpha$$

ונרצה למצוא את t.

$$\Rightarrow \mathbb{P}_{\hat{F}_n}\left(\left|\hat{F}_n(x_0)-\hat{F}_n^*(x_0)\right|\le t\right)=\mathbb{P}_{\hat{F}_n}\left(\hat{F}_n(x_0)-t\le \hat{F}_n^*(x)\le t+\hat{F}_n(x_0)\right)$$

כי ההתפלגות נמשך מ-

$$=\mathbb{P}_{\hat{F}_n}\left(\underbrace{n\hat{F}_n(x_0)-nt}_{I_1}\le \underbrace{\sum_{i=1}^{n}\mathbb{1}\{X_i^*\le x_0\}}_{\sim Bin(n,\hat{F}_n(x_0))}\le \underbrace{nt+n\hat{F}_n(x_0)}_{I_2}\right)$$

$$\mathbb{E}_{\hat{F}_n}\mathbb{1}\{X_i^*\le x_0\}=\sum_{j=1}^{n}\frac{1}{n}\cdot\mathbb{1}\{X_j\le x_0\}=\hat{F}_n(x_0)$$

$$=\sum_{\substack{k\in[I_1,I_2]\\ k:\left|\frac{k}{n}-\hat{F}_n(x_0)\right|\le t}}\binom{n}{k}\hat{F}_n(x_0)^k\left(1-\hat{F}_n(x_0)\right)^{n-k}$$

* לא ניתן למצוא את t, אבל אפשר למצוא t כך שההסתברות לו ~~[illegible]~~ שואפת ל- $1-\alpha$.

(2) הערכה את האינטרוול עבור $\alpha=0.2$, $n=10$, $\hat{F}_n(x_0)=\frac{1}{2}$

$$T(\hat{F}_n)\pm\hat{t}_{n,\alpha}$$

פתרון:

ונרצה למצוא t כך ש-

$$\left(\frac{1}{2}\right)^n\sum_{k:\left|\frac{k}{n}-\frac{1}{2}\right|\le t}\binom{n}{k}\ge 0.8$$

* $t=0 \Rightarrow k=\frac{n}{2}=\frac{10}{2}=5 \Rightarrow \left(\frac{1}{2}\right)^{10}\binom{10}{5}=0.2461$

* $t=\frac{1}{n}=0.1 \Rightarrow \left|\frac{k}{10}-\frac{1}{2}\right|\le 0.1 \Rightarrow 10(0.5-0.1)\le k\le(0.1+0.5)\cdot 10 \Rightarrow 4\le k\le 6$

$$\Rightarrow\left(\frac{1}{2}\right)^{10}\left[\binom{10}{6}+\binom{10}{5}+\binom{10}{4}\right]=0.65$$

נמשיך.

$t = \frac{2}{10} = 0.2 \Rightarrow 0.89 \ldots$

אם, רק ...

$\left[\frac{1}{2} \pm 0.2\right] = \left[\hat{F}(x_0) \pm \hat{\varepsilon}_{n,\alpha}\right]$