روش بوت استرپ (خودگردان)

Bootstrap Method

به علت شباهت بسیار روش بوت استرپ و شبیه سازی مونت کارلو ابتدا این روش معرفی می‌گردد:

مقدمه: شبیه سازی مونت کارلو

Monte Carlo Simulation

فرض: $X_1, \dots, X_n \overset{iid}{\sim} F_X(\theta)$

یا $X_1, \dots, X_n \underset{SRSW}{\overset{iid}{\sim}} x_1, \dots, x_N$ $\qquad \theta = t(x_1, \dots, x_N)$: $\bar{x}_N$, σ^2, S_N^2

$\hat{\Theta} = t(X_1, \dots, X_n) \rightarrow$ برآوردگری برای پارامتر θ

$$\left.\begin{aligned} bias(\hat{\Theta}) &= E(\hat{\Theta}) - \theta \\ var(\hat{\Theta}) &= E[\hat{\Theta} - E(\hat{\Theta})]^2 \\ MSE(\hat{\Theta}) &= E(\hat{\Theta} - \theta)^2 \end{aligned}\right\}$$ اندازه‌گیری دقت $\hat{\Theta}$

$F_{\hat{\Theta}}(t) = P(\hat{\Theta} \le t)$ توزیع $\hat{\Theta}$

$P[h_1(\hat{\Theta}) < \theta < h_2(\hat{\Theta})] = 1 - \alpha$ بازه اطمینان برای θ

در مبانی نظریه آمار و استنباط آماری بعضی از
امید ریاضی (E) و احتمال (P) برای برخی از
برآوردگرها $\hat{\Theta}$ در توزیع‌های معلوم F و پارامترهای
θ جواب دارد و در اکثر جاها جواب ندارد
به علت پیچیده بودن $\hat{\Theta}$.

مثال ۱) $X_1, \dots, X_n \overset{iid}{\sim} F_X(\mu, \sigma^2)$

$\theta = \mu \qquad \hat{\Theta} = \bar{X}$

نااریب $\text{bias}(\bar{X}) = E(\bar{X}) - \mu = \mu - \mu = 0$

$\text{var}(\bar{X}) = \frac{\sigma^2}{n}$

توزیع $\bar{X}$ $\quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad F = N$ نرمال (مثلا)

$P(\bar{X} \mp z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$

تمرین ۱) $X_1, \dots, X_n \overset{iid}{\sim} \text{bernolli}(p)$

$\hat{\Theta} = \hat{p} = \frac{X}{n} = \frac{\sum X_i}{n}$

خواص نسبت (درصد) نمونه

مثال ۱) ادامه

$$\hat{\theta} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

واریانس نمونه

$$E(S^2)=\sigma^2$$ S^2 نااریب برای σ^2

$$Var(S^2) = ?$$

$F=N$ $$Var(S^2)=\frac{2\sigma^4}{n-1}$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$$

تمرین) فاصله اطمینان

تمرین) برای توزیع های غیر نرمال ؟؟؟

برآوردگر گری نمونه

تمرین) توزیع داده نمونه تغییرات R ، چندک p ام ، $\hat{Q}_p$ ، میانه ، نما (مد) ، ضریب تغییرات نمونه

اریبی ، واریانس ، فاصله اطمینان ؟

مثال دو متغیره)

$$(x_1, y_1), \ldots, (x_n, y_n) \overset{iid}{\sim} F_{xy}(\mu_x, \mu_y, \sigma_1^2, \sigma_2^2, \rho)$$

$\hat{\rho} = R$ ضریب همبستگی خطی پیرسون

برآوردگر اریب برای R

می‌تواند نااریب

$$\text{bias}(R_n) = E(R_n) - \rho \longrightarrow 0 \quad \text{as } n \to \infty$$

$$\text{var}(R) = ?$$

$R \sim ?$ توزیع R

در حالت n بزرگ Z فیشر

توزیع نرمال دو متغیره از توزیع t

دو نکته مهم) در شبیه سازی مونت کارلو

امید ریاضی E = میانگین دفعات مکرر آزمایش

احتمال P = فراوانی نسبی دفعات مکرر آزمایش

دو هدف مهم) در شبیه سازی مونت کارلو

1) تایید مبانی نظری اثبات شده در استنباط آماری به صورت شهودی و محاسباتی (شبیه سازی)

2) اثبات مبانی نظری لاینحل در استنباط آماری به صورت شهودی و محاسباتی (شبیه سازی)

دو فرض مهم) در شبیه سازی مونت کارلو

1) توزیع جامعه معلوم ^{F} / جامعه محدود $^{x_1 - x_N}$ با سرشماری

2) پارامترهای جامعه معلوم θ معلوم (مثلا: μ، $\bar{x}_N$ و ...)

دو نکته مهم)

1) البته فرضیات بالا: F و θ معلوم فرض محال است

2) شبیه سازی در عالم واقع کاربرد ندارد برای محاسبات و شهودی در هدف بالاست !!!

الگوریتم مونت کارلو :

1) $X_1, \dots, X_n \overset{iid}{\sim} F_x(\theta)$ مجهول

یا $x_1, \dots, x_n \underset{SRSW}{\overset{iid}{\sim}} x_1, \dots, x_N$

2) $\hat{\Theta} = t(x_1, \dots, x_n)$

3) $\hat{\Theta}_1, \dots, \hat{\Theta}_m$ تکرار m بار مراحل ۱ و ۲ برای تشکیل

اندازه دقت

4) $bias(\hat{\Theta}) \simeq \underbrace{\frac{1}{m}\sum_{i=1}^{m} \hat{\Theta}_i}_{\bar{\hat{\Theta}}} - \theta$

$var(\hat{\Theta}) \simeq \frac{1}{m}\sum_{i=1}^{m} (\hat{\Theta}_i - \bar{\hat{\Theta}})^2$

$MSE(\hat{\Theta}) \simeq \frac{1}{m}\sum_{i=1}^{m} (\hat{\Theta}_i - \theta)^2$

توزیع $\hat{\Theta}$

$$F_{\hat{\Theta}}(t) = P(\hat{\Theta} \leq t) = \frac{\#(\hat{\Theta}_i \leq t \, ; \, i=1, \dots, m)}{m} = \frac{1}{m}\sum_{i=1}^{m} I(\hat{\Theta}_i \leq t)$$

آماره مرتب

بازه اطمینان

$$CI \left(\hat{\Theta}_{(\frac{\alpha}{2} \cdot m)} \, , \, \hat{\Theta}_{[(1-\frac{\alpha}{2})m]}\right)$$

روش بوت استرپ (خودگردان) ناپارامتری

Nonparametric Bootstrap Method

برادلی افرون در سال ۱۹۷۹ روش بوت استرپ را برای اندازه دقت برآورد برآوردگرها (مانند: اریبی، واریانس، خطای استاندارد و میانگین توان دوم خطا) و همچنین توزیع برآوردگرها و به علاوه بازه اطمینان و آزمون فرض برای پارامترها ارائه نمود.

$$X_1, \dots, X_n \overset{iid}{\sim} F_X(\theta)$$ مفروض ، نامعلوم

نمونه مشاهده شده $x_1, \dots, x_n$

برآوردگر $\hat{\Theta} = t(X_1, \dots, X_n)$

برآورد $\hat{\theta} = t(x_1, \dots, x_n)$

هدف اندازه گیری دقت برآوردگر $\hat{\Theta}$ مانند: اریبی، واریانس، خطای استاندارد و MSE و توزیع $\hat{\Theta}$ و همچنین بازه اطمینان و آزمون فرض برای θ در حالت نامعلوم بودن F و θ (ناپارامتری) است. برآورد

نکته: هدف اصلی شبیه سازی مونت کارلو وقتی F و θ نامعلوم هستند!

الگوریتم بوت استرپ ناپارامتری :

۱) نمونه بوت استرپ :

نمونه بوت استرپ $X_1^*, \ldots, X_n^*$ را به صورت یک نمونه تصادفی مستقل و هم توزیع (iid) از تابع توزیع تجربی $F_n(x)$ بدست می‌آوریم. در واقع نمونه بوت استرپ $X_1^*, \ldots, X_n^*$ یک نمونه تصادفی ساده با جایگذاری از نمونه مشاهده شده $x_1, \ldots, x_n$ است.

(دقت کنید تابع توزیع تجربی $F_n(x)$ جرم $\frac{1}{n}$ را به هر x_i در نمونه مشاهده شده $x_1, \ldots, x_n$ نسبت می‌دهد)

نمونه بوت استرپ : $X_1^*, \ldots, X_n^* \overset{iid}{\sim} F_n(x)$

یا $X_1^*, \ldots, X_n^* \xleftarrow{SRSW} x_1, \ldots, x_n$

۲) برآوردگر بوت استرپ :

برآوردگر بوت استرپ را به صورت $\hat{\theta}^* = t(X_1^*, \ldots, X_n^*)$ مشابه برآوردگر $\hat{\theta} = t(X_1, \ldots, X_n)$ به دست می‌آوریم.

برآوردگر بوت استرپ : $\hat{\theta} = t(X_1^*, \ldots, X_n^*)$

3) برآوردگر نقطه‌ای بوت استرپ :

برآوردگر نقطه‌ای بوت استرپ اندازه‌گیری دقت (مانند : اریبی، واریانس، خطای استاندارد و MSE) و توزیع برآوردگر و همچنین، بازه اطمینان به صورت زیر محاسبه می‌گردد :

$$bias^*(\hat{\Theta}^*) = E^*(\hat{\theta}^*) - \hat{\theta}$$

$$var^*(\hat{\Theta}^*) = E^*\left[\hat{\Theta}^* - E^*(\hat{\Theta}^*)\right]^2$$

$$SE^*(\hat{\Theta}^*) = \sqrt{var^*(\hat{\Theta}^*)}$$

$$MSE^*(\hat{\Theta}^*) = E^*(\hat{\Theta}^* - \hat{\theta})^2$$

$$F^*_{\hat{\Theta}^*}(t) = P^*(\hat{\Theta}^* \leq t)$$

$$P\left[h_1^*(\hat{\Theta}^*) < \theta < h_2^*(\hat{\Theta}^*)\right] = 1-\alpha$$

نکته) کردشان E^* و P^* در واقع امید ریاضی شرطی و احتمال شرطی به شرط نمونه مشاهده شده $x_1, \dots, x_n$ هستند.

$$E^*(\cdot) = E^*(\cdot \mid x_1, \dots, x_n)$$

$$P^*(\cdot) = P^*(\cdot \mid x_1, \dots, x_n)$$

سؤال) فرض کنید $x_1, \dots, x_n \overset{iid}{\sim} F_X(\theta : \mu, \sigma^2)$

$\hat{\theta} = \bar{X}$

نمونه مشاهده شده $x_1, \dots, x_n$

$\hat{\theta} = \bar{x}$

$\downarrow$

نمونه بوت استرپ: $x_1^*, \dots, x_n^*$

$\downarrow$

برآوردگر بوت استرپ $\hat{\theta}^* = t(x_1^*, \dots, x_n^*) = \frac{1}{n}\sum_{i=1}^{n} x_i^* = \bar{x}^*$

حکم) $bias^*(\bar{x}^*) = E^*(\bar{x}^*) - \bar{x} = \bar{x} - \bar{x} = 0$

$\downarrow$

$$E^*(\bar{x}^*) = E^*\left(\frac{1}{n}\sum_{i=1}^{n} x_i^*\right) = \frac{1}{n}\sum_{i=1}^{n} E^*(x_i^*)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E^*(x_1^*) = E^*(x_1^*)$$

م.ت گسسته با مقادیر

x_i^*	$x_1 \dots x_n$
احتمال	$1/n \dots 1/n$

$$= \sum_{i=1}^{n} x_i \times \frac{1}{n} = \bar{x}$$

تمرین) 1) $var^*(\bar{x}^*) = \frac{\hat{\sigma}^2}{n}$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

2) $E(S^{2*}) = ?$ $\qquad var^*(S^{2*}) = ?$

4) برآوردگر تجربی بوت استرپ :

اگر برآوردگر نظری بوت استرپ در مرحله 3 جواب بسته‌ای نداشته باشد، آنگاه با استفاده از شبیه‌سازی مونت کارلو و تکرار B بار مراحل 1 و 2 برآوردگر تجربی بوت استرپ اندازه‌های دقت (اریبی، واریانس، خطای استاندارد و MSE) و توزیع برآوردگر و همچنین بازه اطمینان برای پارامتر θ به صورت زیر محاسبه می‌گردد :

تکرار B بار مراحل 1 و 2 به صورت متعدد محاسبه

$\hat{\theta}_1 , \dots , \hat{\theta}_B$

$$\widehat{bias}^*(\hat{\theta}^*) = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}_i^* - \hat{\theta}$$

$$\widehat{var}^*(\hat{\theta}^*) = \frac{1}{B} \sum_{i=1}^{B} (\hat{\theta}_i^* - \bullet)^2$$

$$\widehat{SE}^*(\hat{\theta}^*) = \sqrt{\widehat{var}^*(\hat{\theta}^*)}$$

$$\widehat{MSE}^*(\hat{\theta}^*) = \frac{1}{B} \sum_{i=1}^{B} (\hat{\theta}^* - \hat{\theta})^2$$

$$\hat{F}_{\hat{\Theta}^*}(t) = \hat{P}^*(\hat{\Theta}^* \leq t) = \frac{1}{B}\sum_{i=1}^{B} I(\hat{\Theta}_i^* \leq t)$$

$$= \frac{\#(\hat{\Theta}_i^* \leq t \;;\; i=1,\dots,B)}{B}$$

یک فاصله اطمینان صدکی $100(1-\alpha)\%$ بوت استرپ به صورت زیر است :

$$\theta \in \left\{ \hat{\Theta}^*_{[(B.\alpha/2)]} \;,\; \hat{\Theta}^*_{[(B.(1-\alpha/2))]} \right\}$$

که در آن $\hat{\Theta}_{[(a)]}$ آماره مرتب شده a ام $\hat{\Theta}_1^*, \dots, \hat{\Theta}_B^*$ است.

Mouse Data

مثال (Pill , Efron)

مرجع) کتاب مقدمه‌ای بر بوت استرپ افرون و تیبشیرانی

Introduction to the Bootstrap

Efron & Tibshirani

1993

تعداد $n=16$ موش آزمایشگاهی به طور تصادفی انتخاب و به دو گروه $n_1=7$ تیمار و $n_2=9$ کنترل (شاهد) تقسیم شدند. در گروه تیمار به موشها داروی تزریق گردید که در طول عمر آنها موثر است (افزایش). تعداد روزهای زنده ماندن موشها ثبت گردید. نتایج در جدول زیر خلاصه شده است:

تیمار	94	197	16	38	99	141	23		
کنترل	52	104	146	10	50	31	40	27	46

تمرین ۱) I) فاصله اطمینان 0.95 برای $\frac{\sigma_1^2}{\sigma_2^2}$ و $\mu_1-\mu_2$

آزمون فرض در سطح $\alpha=0.05$

$$\begin{cases} H_0: \sigma_1^2=\sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases} \qquad \begin{cases} H_0: \mu_1=\mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$$

II) - آزمونهای ناپارامتری بخش I را انجام دهید.

- وضعیت نرمال بودن داده‌ها را بررسی کنید (نمودار - آزمون)

I)

تمرین ۲) روش بوت استرپ را برای برآورد اندازه دقت خطای استاندارد در برآوردگر میانگین نمونه $\bar{X}$ و میانه نمونه $\tilde{X}$ برای داده‌های گروه تیمار داده‌های موش اجرا کنید.

نمونه‌ای از دو برآورد بالا برای B‌های مختلف در جدول زیر ارائه شده است: ?

B	50	100	250	500	1000	∞
$\hat{SE}^*(\bar{x})$	19.72	23.63	22.32	23.79	23.02	23.36
$\hat{SE}^*(\tilde{x})$	32.21	36.35	34.46	36.72	36.48	37.83

II) برای داده‌های موش گروه تیمار با استفاده از آمار کلاسیک و بوت استرپ

$\hat{bias}$ و $\hat{bias}^*$: $\bar{x}$ و $\tilde{x}$ و S^2 و S

$\hat{SE}$ و $\hat{SE}^*$: " " " " "

توزیع ت : بوت استرپ کلاسیک " " " " " "

فاصله اطمینان : " " " " " "

III) تمرین ۱ را با استفاده از روش بوت استرپ جواب دهید.

Scanned by CamScanner