حل تمارین سری سوم آمار

برآوردگرهای نقطه‌ای و بازه‌ای

برآوردگر نقطه‌ای

یک نقطه رو برای مقدار یا پارامتر نامشخص یک جامعه از روی یک sample از اون جامعه به ما میده.

$\bar{x}$ ← میانگین

S ← انحراف معیار

برآوردگر بازه‌ای

برخلاف برآوردگر نقطه‌ای این دسته یک بازه رو به ما میدن که کران‌های این بازه یک ضریب اطمینان هم به ما میدن

خطای اریب

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$MSE(\hat{\theta}) = E\left[(\hat{\theta}(x) - \theta)^2\right]$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (B(\hat{\theta}))^2$$

$$E\left[(\hat{\theta} - \theta)^2\right] = (E(\hat{\theta}) - \theta)^2 + Var(\theta)$$

بازه اطمینان

خطای بدست آمده براساس آزمایش‌ها

با توجه به بازه‌ی اطمینان از جدول خوانده می‌شود

$$\hat{e} \pm Z_{\alpha/2}\sqrt{\frac{e(1-e)}{n}}$$

تعداد آزمایش‌ها

پاسخ سوال ۱

a) CLT

برای استفاده از قضیه‌ی حد مرکزی مراحل زیر را طی میکنیم

۱. تعیین کردن شرایط استفاده از قضیه

۲. محاسبه‌ی $\sigma_{\bar{x}}$

۳. محاسبه‌ی بازه‌ی اطمینان ۹۰ درصدی

به علت اینکه سایز نمونه ۴۵ است و به اندازه‌ی کافی بزرگ هست و این نمونه‌ها از یکدیگر مستقل هستند حق استفاده از CLT را داریم.

برای محاسبه‌ی $\sigma_{\bar{x}}$ مانند رابطه‌ی زیر داریم

$$\sigma_{\bar{x}}^2 = \frac{s^2}{n} \Rightarrow \sigma_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{2}{45}} \approx 0.21$$

محاسبه‌ی بازه‌ی اطمینان به صورت زیر خواهد بود

$$\mathbb{P}\left(\mu - Z_{\alpha/2}\cdot\sigma_{\bar{x}} \le \bar{x} \le \mu + Z_{\alpha/2}\cdot\sigma_{\bar{x}}\right)$$

برای بازه‌ی اطمینان ۹۰ درصدی مقدار $Z_{\alpha/2}$ برابر با 1.645 است (با استفاده از جدول مقادیر بحرانی نرمال)

$$\mathbb{P}\left(\mu - (1.645\cdot 0.21) \le \bar{x} \le \mu + (1.645\cdot 0.21)\right)$$

$$\Rightarrow \mathbb{P}\left(\mu - 0.347 \le \bar{x} \le \mu + 0.347\right)$$

پس میانگین اصلی ما با احتمال 90% به مقدار 0.347 از میانگین 45 نمونه برداری فاصله دارد.

b) نامساوی چبیشف

$$\mathbb{P}(|X-\mu| \ge k\cdot\sigma) \le \frac{1}{k^2}$$

که در آن k تعداد انحراف معیارهایی است که تا μ داریم

مراحل

۱. محاسبه‌ی $\sigma_{\bar{x}}$ که در بخش CLT محاسبه شد

۲. پیدا کردن یک k که بازه‌ی اطمینان ۹۰ درصدی ما را نشان دهد

۳. محاسبه‌ی بازه‌ی اطمینان ۹۰ درصدی

$$\mathbb{P}(|\bar{x}-\mu| \ge k\cdot\sigma_{\bar{x}}) \le 0.1$$

$$0.1 = \frac{1}{k^2} \Rightarrow 0.1k^2 = 1 \to k^2 = 10 \to k \simeq 3.16$$

$$\mathbb{P}\left(\mu - k \cdot \sigma_{\bar{x}} \leq \bar{x} \leq \mu + k \cdot \sigma_{\bar{x}}\right) \Rightarrow \mathbb{P}\left(\mu - 3.16 \cdot 0.21 \leq \bar{x} \leq \mu + 3.16 \cdot 0.21\right)$$

$$\Rightarrow \mathbb{P}\left(\mu - 0.\overline{66} \leq \bar{x} \leq \mu + 0.\overline{66}\right)$$

پس با احتمال 90 درصد تخمین میانگین 45 نمونه با میانگین جامعه به اندازه $0.\overline{66}$ فاصله خواهد داشت.

پاسخ سوال 2

(a)

چون که داده از توزیع لاگ نرمال تبعیت میکنه پس
باید پارامترهای μ و σ^2 رو براش بدست بیاریم

$X \sim \log Normal(\mu, \sigma^2)$

مراحل زیر رو طی می‌کنیم

1 نوشتن MLE برای تابع pdf
2 محاسبه‌ی $\hat{\mu}$ و $\hat{\sigma}^2$ با استفاده از MLE
3 بررسی اریب بودن
4 محاسبه‌ی کارایی

$$X \sim LogNormal(\mu, \sigma^2) \longrightarrow pdf_X = f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,x\sigma} \exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}$$

$$\Rightarrow MLE_X(\mu,\sigma^2) = \prod_{i=1}^{n} f_X(x_i;\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,x_i\sigma} \exp\left\{-\frac{(\ln(x_i)-\mu)^2}{2\sigma^2}\right\}$$

$$\ln MLE_X(\mu,\sigma^2) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\,x_i\sigma} \exp\left\{-\frac{(\ln(x_i)-\mu)^2}{2\sigma^2}\right\}\right)$$

$$= -n\ln(\sigma) - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n}\ln(x_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2$$

محاسبه‌ی $\hat{\mu}$: $\frac{\partial \ln MLE}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu) = 0 \longrightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)$

محاسبه‌ی $\hat{\sigma}^2$: $\frac{\partial \ln MLE}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln(x_i)-\hat{\mu})^2$

چک کردن اریب بودن

$$E[\hat{\mu}] = E\left[\frac{1}{n}\sum_{i=1}^{n}\ln(x_i)\right] = \frac{1}{n}\sum_{i=1}^{n}E[\ln(x_i)] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

طبق خواص توزیع نرمال

$B(\hat{\mu}) = E[\hat{\mu}] - \mu = \mu - \mu = 0 \rightarrow$ اریب نیست

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(\ln(x_i)-\hat{\mu})^2\right] = \frac{1}{n}\sum_{i=1}^{n}E[\ln(x_i-\mu)^2]$$

$$= \frac{1}{n}\sum_{i=1}^{n}E[2\ln(x_i-\mu)] = \frac{1}{n}\sum_{i=1}^{n}2(E(\underset{\mu}{\cancel{\ln(x_i)}}) - E(\underset{\mu}{\cancel{\mu}})] = 0 \}$$

$\Rightarrow B(\hat{\sigma}^2) = 0 - \sigma^2 = -\sigma^2 \longrightarrow$ اریب است ...

محاسبه‌ی کارایی

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (bias(\hat{\theta}))^2$$

کارایی $\hat{\mu}$: چون که بایاس نیست و واریانس هم برابر با واریانس خودشه پس کاراست

کارایی $\hat{\sigma}^2$: چون که بایاس هست و برای واریانس هم مشکل داریم قطعاً کارا نیست (اما برای اینکه کاراش کنیم باید به جای $\frac{1}{n}$ از $\frac{1}{n-1}$ استفاده کنیم

$$\hat{\sigma}^2_{correct} = \frac{1}{n-1} \sum_{i=1}^{n} (\ln(x_i) - \hat{\mu})^2$$

البته باز هم از همین تخمینگر برای توزیع نرمال نشان داده شده که دیگه کارا (که همینطوریش بایاس میشه اما $\frac{2}{n-1}$ کمتر بایاس هست.

محاسبه‌ی کارایی رو نتونستم بنویسم چون از پس محاسباتش برنمیومدم و خیلی برام سنگین بود و نتیجه سرچ‌هایی با اینش برخورد کردم.

ط) در این جایی که سوال بخش a تخمین‌گرهای توزیع لاگ نرمال رو با استفاده از MLE به دست آوردیم در این بخش صورت مسئله رو بررسی می‌کنیم و در نهایت هم $\mathbb{P}(\mu \geq 7)$ رو به دست میاریم.

نکته: در بخش قبل با استفاده از MLE تخمین‌گر $\hat{\sigma}^2$ کارا نیست و بایاس هست.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i)$$

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

$$\hat{\mu} = \frac{1}{n}(\ln(6.34) + \dots) \simeq \frac{1}{15}(25.01) \simeq 1.66$$

$$\hat{\sigma}^2 = \frac{1}{n}((\ln(6.34) - 1.66)^2 + \dots) \simeq \frac{1}{15}(0.534) \simeq 0.03$$

در ادامه چون که توزیع داده‌های ما لاگ نرمال است یعنی $Y \sim \ln(X)$, $Y \sim Normal(\mu, \sigma^2)$

و می‌دانیم که $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, $\bar{X} \sim Normal(E(\bar{X}), \frac{Var(X)}{n})$ که در آن $X \sim logNormal(\mu, \sigma^2)$

و $E[X] = \exp\{\mu + \frac{\sigma^2}{2}\}$, $Var[X] = (e^{\sigma^2} - 1)\exp\{2\mu + \frac{\sigma^2}{2}\}$

حال با استفاده از تخمینگرهایی که به دست آوردیم و اعمال CLT خواهیم داشت

$$\bar{X} \sim Normal\left(\exp\{\hat{\mu} + \frac{\hat{\sigma}^2}{2}\}, \frac{\hat{Var}(X)}{n}\right)$$

$$\Rightarrow \mathbb{P}(\bar{X} \geq 7) = 1 - \Phi\left(\frac{7 - E[\bar{X}]}{\sqrt{Var(\bar{X})}}\right) \quad s.t \begin{cases} E[\bar{X}] = 5.4 \\ Var(\bar{X}) = 0.07 \end{cases} \Rightarrow \mathbb{P}(\bar{X} \geq 7) \simeq 3.19 \times 10^{-9}$$

بعد از محاسبه‌هایی که بالا نوشتیم و جایگذاری $\hat{\mu}$ ، $\hat{\sigma}^2$ به دست آمدن

پاسخ سوال ۵

تقریب نرمال: $\hat{e} = \frac{11}{120}$ هستش چون ۱۲۰ بار تکرار کردیم میریم سراغ CLT

بازه‌های خطای ۰.۹ و ۰.۹۹ رو حساب میکنیم و خلاص

$$\hat{e} \pm z_{\alpha/2} \sqrt{\frac{\hat{e}(1-\hat{e})}{n}} \rightarrow \begin{cases} 90 & z_{\alpha/2} = 1.645 \rightarrow \hat{e} \pm 1.645\sqrt{\frac{\hat{e}(1-\hat{e})}{n}} = (0.04, 0.13) \\ 99 & z_{\alpha/2} = 2.576 \rightarrow \hat{e} \pm 2.576\sqrt{\frac{\hat{e}(1-\hat{e})}{n}} = (0.02, 0.15) \end{cases}$$

برای بازه‌های یک طرفه:

$$\hat{e} + z_{\alpha/2} \sqrt{\frac{\hat{e}(1-\hat{e})}{n}} \rightarrow \begin{cases} 90 & z_{\alpha/2} = 1.28 \rightarrow \hat{e} + 1.28\sqrt{\frac{\hat{e}(1-\hat{e})}{n}} = 0.125 \\ 99 & z_{\alpha/2} = 2.33 \rightarrow \hat{e} + 2.33\sqrt{\frac{\hat{e}(1-\hat{e})}{n}} = 0.153 \end{cases}$$

تفسیر بازه‌های دو طرفه:

با احتمال ۹۰ درصد خطای تقسیم کننده، در بازه‌ی ۴٪ و ۱۳٪ قرار داره

~ ~ ۹۹ ~ ~ ~ ~ ۲٪ و ۱۵٪ قرار داره ← ضریب اطمینان بیشتر = بازه‌ی بزرگتر

تفسیر بازه‌های یک طرفه:

با احتمال ۹۰ درصد خطای تقسیم کننده از ۱۲.۵٪ بیشتر نمیشه

~ ~ ۹۹ ~ ~ ~ ~ ~ ۱۵.۳٪ بیشتر نمیشه ← ضریب اطمینان بیشتر = بازه‌ی بزرگتر