علیرضا کریمی ۸۱۰۱۰۱۴۹۲

$$D = \{x_1, x_2, \dots, x_n\}$$

۱)

$$\hat{\theta}_{ML} = \arg\max_{\theta} P(D) = \arg\max_{\theta} \underbrace{\log P(D)}_{LL}$$

$$LL(\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta) \qquad \frac{dLL(\theta)}{d\theta} = 0 \longrightarrow \hat{\theta}_{ML}$$

$$\Longrightarrow \frac{d \sum_{i=1}^{n} \ln f(x_i|\theta)}{d\theta} = \sum_{i=1}^{n} \frac{d \ln \frac{\theta}{x_i^2}}{d\theta} = \sum_{i=1}^{n} \frac{1}{\theta} = \frac{n}{\theta}$$

از آنجایی که $\theta > 0 \Longrightarrow \frac{n}{\theta} > 0 \Longrightarrow \frac{dLL(\theta)}{d\theta} > 0 \Longrightarrow LL(\theta)$ همواره صعودی است

و هرچقدر θ بیشتر باشد، مقدار تابع $LL(\theta)$ نیز بیشتر است. از طرفی محدودیتی برای θ وجود دارد که

$$\theta \le x_i \Longrightarrow \hat{\theta}_{ML} = \min\{x_1, x_2, \dots, x_n\}$$

(۲ $m_1 = \mu = \int_{-\infty}^{+\infty} x f(x)\,dx = \frac{1}{2}\int_{-1}^{+1} (x+\theta x^2)\,dx = \frac{x^2}{4} + \frac{\theta x^3}{6}\Big|_{-1}^{1}$

$= \frac{\theta}{3} \Longrightarrow \theta = 3m_1 = 3\mu$ $\qquad \hat{m}_1 = \frac{1}{2}\sum_{i=1}^{n} x_i = \hat{\mu} \Longrightarrow$

$\hat{\theta} = 3\hat{m}_1 = 3\hat{\mu}$

بررسی bias بودن برآوردگر :

شرط اول : $\lim_{n\to\infty} E[\hat{\theta}] = \lim_{n\to\infty} E[3\hat{\mu}] = \lim_{n\to\infty} 3E[\hat{\mu}]$

$= 3E[\lim_{n\to\infty} \hat{\mu}] = 3E[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i] = 3E[\mu] = 3\mu = 3\times\frac{\theta}{3} = \theta$ ✓

شرط دوم : $\lim_{n\to\infty} Var(\hat{\theta}) = \lim_{n\to\infty} var(3\hat{\mu}) = 9\lim_{n\to\infty} var(\hat{\mu}) = 9\lim_{n\to\infty}\frac{\sigma^2}{n} = 0$ ✓

$\Leftarrow$ برآوردگر unbiased است.

۳) طبق جزوه استاد، داریم:

$$\hat{P} = \frac{X}{n} \sim N\left(P, \frac{P(1-P)}{n}\right)$$

ابتدا گزینه ب را بررسی می‌کنیم.

$0 < P \leq 1$

$$\Rightarrow \text{ب)}\quad P\left(\frac{X}{n} < P < 1\right) = P(\hat{P} < P < 1) = P(\hat{P} < P) =$$

$$P(\hat{P} - P < 0) = P\left(\frac{\hat{P} - P}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} < 0\right) = P(Z < 0) = \Phi(0) = 0{,}5$$

حال گزینه الف را بررسی می‌کنیم

چون P قطعاً مقداری بین صفر و یک دارد.

$$\Rightarrow \text{الف)}\quad P\left(\frac{X}{n} < P < \infty\right) = P\left(\frac{X}{n} < P < 1\right) + \cancel{P(1 < P < \infty)}$$

ادامه مثل قسمت ب

$$= P(\hat{P} < P) = 0{,}5$$

حال گزینه ج را بررسی می‌کنیم: طبق مفهوم بازه اطمینان برای نسبت برای بازه اطمینان $1-\alpha$ درصد داریم:

$$\Rightarrow \text{ج)}\quad P\left(\frac{X}{n} - 0{,}67\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}} < P < \frac{X}{n} + 0{,}67\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}}\right)$$

$$= P\left(\hat{P} - 0{,}67\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} < P < \hat{P} + 0{,}67\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right)$$

$$\Rightarrow Z_{1-\alpha/2} = 0{,}67 \xrightarrow{\text{طبق جدول}} 1 - \frac{\alpha}{2} = 0{,}7486 \Rightarrow \alpha = 0{,}5028$$

$$\Rightarrow 1 - \alpha = 0{,}4972$$

پس P به احتمال $1-\alpha$ یا همان $0{,}4972$ در بازه ج می‌افتد. در حالی که با احتمال $0{,}5$ در بازه‌های الف و ب می‌افتد

پس احتمال اینکه در بازه‌های "الف" و "ب" بیفتد بیشتر است.

۴)

$H_0: \mu = ۱۰٫۴$, $H_A: \mu \neq ۱۰٫۴ \longrightarrow \mu = ۱۰$

از میانگین به دست آمده از sample استفاده می‌کنیم

$\mu_0 = ۱۰٫۴$

$\bar{X} = ۱۰ \ , \ \sigma = ۲$

چون آزمون فرض دوطرفه است داریم:

$$\alpha = P(\bar{X} < \mu - \Delta \ \text{or} \ \bar{X} > \mu + \Delta \mid \mu = \mu_0)$$

$$P(\bar{X} < \mu - \Delta) = P(\bar{X} > \mu + \Delta)$$

$\mu = \mu_0 = ۱۰٫۴$

استاندارد کردن:

$$\Longrightarrow ۰٫۵ = ۲P(\bar{X} < ۱۰٫۴ - \Delta) = ۲P\left(\frac{\bar{X} - ۱۰٫۴}{\frac{\sigma}{\sqrt{n}}} < \frac{-\Delta}{\frac{\sigma}{\sqrt{n}}} \right)$$

$$\frac{\sigma}{\sqrt{n}} = \frac{۲}{\sqrt{۲۵}} = \frac{۲}{۵}$$

$$\Rightarrow ۲P\left(Z < -\frac{۵}{۲}\Delta\right) = ۲\Phi\left(-\frac{۵}{۲}\Delta\right) = ۰٫۰۵ \Rightarrow \Phi\left(-\frac{۵}{۲}\Delta\right) = ۰٫۰۲۵$$

$$\Longrightarrow \Phi\left(+\frac{۵}{۲}\Delta\right) = ۱ - ۰٫۰۲۵ = ۰٫۹۷۵ \Longrightarrow \frac{۵}{۲}\Delta = ۱٫۹۶ \Rightarrow \Delta = ۰٫۷۸۴$$

$$\beta = P(\text{Not}(\bar{X} < ۱۰٫۴ - \Delta \ \text{or} \ \bar{X} > ۱۰٫۴ + \Delta) \mid \text{Not} \ \mu = \mu_0)$$

$$\Rightarrow \beta = P(\overset{۱۰٫۴}{\mu_0} - \Delta < \bar{X} < \mu_0 + \Delta \mid \mu \neq \mu_0 = ۱۰٫۴)$$

$$\Longrightarrow \beta = P(\overset{۱۰٫۴}{\mu_0} - \overset{۰٫۷۸۴}{\Delta} < \bar{X} < \mu_0 + \Delta \mid \mu = ۱۰)$$

$$= P(۹٫۶۱۶ < \bar{X} < ۱۱٫۱۸۴ \mid \mu = ۱۰)$$

استاندارد کردن

$$= P\left(\frac{۹٫۶۱۶ - \overset{\mu}{۱۰}}{۲/\sqrt{۲۵}} < Z < \frac{۱۱٫۱۸۴ - ۱۰}{۲/\sqrt{۲۵}} \right) = P(-۰٫۹۶ < Z < ۲٫۹۶)$$

$\frac{\sigma}{\sqrt{n}} \leftarrow ۲/\sqrt{۲۵}$

$$= \underbrace{P(۲٫۹۶)}_{۰٫۹۹۸۵} - (۱ - \underbrace{P(۰٫۹۶)}_{۰٫۸۳۱۵}) = \boxed{۰٫۸۳}$$

۵)

$$f_X(x_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\longrightarrow L(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \xRightarrow{\log} LL(\sigma^2) = \sum_{i=1}^{n} \ln f(x_i|\mu,\sigma^2)$$

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

مشتق گرفته و برابر صفر قرار میدهیم

$$\Rightarrow \frac{dLL(\sigma^2)}{d(\sigma^2)} = \sum_{i=1}^{n}\left(\frac{2(x_i-\mu)^2}{4\sigma^4} - \frac{1}{2\sigma^2}\right) = 0$$

$$\Rightarrow \left(\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^4}\right) = \frac{n}{2\sigma^2} \Longrightarrow \sum_{i=1}^{n}(x_i-\mu)^2 = n\sigma^2$$

$$\Longrightarrow \frac{1}{n}\sum_{i=1}^{n} var(x_i) = \hat{\sigma}^2 \longrightarrow$$ که همان $\hat{\sigma}^2_{ML}$ است

$$E(\hat{\sigma}^2_{ML}) = E\left[\frac{1}{n}\sum_{i=1}^{n} var(x_i)\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[var(x_i)]}_{\sigma^2}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \sigma^2 \Longrightarrow E(\hat{\sigma}^2_{ML}) = \sigma^2 \longrightarrow$$ برآوردگر unbiased است

(برای بدست آوردن ماکسیمم برآوردگر σ^2، نسبت به آن مشتق گرفته و برابر صفر قرار دادیم که $\hat{\sigma}^2_{ML}$ را پیدا کنیم.)

$n<30$

۶) سعی می‌کنیم این سوال را با کمک بازه اطمینان و با استفاده از توزیع t حل کنیم (چون تعداد نمونه‌ها کم است، واریانس جامعه را نیز نداریم. خب در ابتدا میانگین و واریانس نمونه‌ها را پیدا می‌کنیم و بعد از آن از توزیع t استفاده می‌کنیم.

$$\bar{X} = \frac{6+8.2+6.4+4.8+8.6+8+6+7.5+8.1+7.2}{10} = 7.08$$

محاسبه واریانس به صورت unbiased : $s^2 = \frac{1}{n-1}\sum_{i=1}^{n=10}(x_i-\bar{x})^2 \simeq 1.52 \Rightarrow s \simeq 1.23$

درجه آزادی : $r = n-1 = 9$

از آنجایی که $0.01 = \alpha$ است، بازه اطمینان $0.99 = 1-\alpha$ (۹۹ درصدی) به کمک توزیع برابر است با:

$$\left(\bar{x} - t_{\alpha/2}(r)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}(r)\frac{s}{\sqrt{n}}\right)$$

$$= \left(7.08 - \underbrace{(3.25)}_{t_{0.005}(9)} \times \frac{1.23}{\sqrt{10}} < \mu < 7.08 + (3.25) \times \frac{1.23}{\sqrt{10}}\right)$$

$$= (5.816 < \mu < 8.344)$$

از آنجایی که $7.5 = \mu_0$ در بازه ۹۹ درصدی بالا قرار دارد، پس H_0، reject نمی‌شود و نمی‌توان گفت که این نمونه از نوزادان کمبود وزن دارند.

$H_0 : \mu = 7.5$

$H_A : \mu < 7.5$

$t_{\alpha/2}(r) = t_{0.005}(9) = 3.25$

طبق جدول توزیع t (که در صفحه بعد آمده است)

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587