(الف)

$$P(x_1 \dots x_n|\theta) \overset{ind}{=} \prod_{i=1}^{n} P(x_i|\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$= \underbrace{\theta^n \left(\underbrace{\prod_{i=1}^{n} x_i}\right)^{\theta-1}}_{\varphi(u(x_i);\theta)} \times 1$$

پس $T = \prod_{i=1}^{n} x_i$ یک sufficient statistics است.

(ب)

$$P_\theta(x) = \theta a x^{a-1} e^{-\theta x^a} \Rightarrow P(x_1 \dots x_n|\theta) = \prod_{i=1}^{n} \theta a x_i^{a-1} \exp(-\theta x_i^a)$$

$$\theta^n a^n \left(\prod_{i=1}^{n} x_i\right)^{a-1} \exp\left(-\theta \sum_{i=1}^{n} x_i^a\right) = \underbrace{a^n \left(\prod_{i=1}^{n} x_i\right)^{a-1}} \underbrace{\theta^n \exp\left(-\theta \underbrace{\sum_{i=1}^{n} x_i^a}\right)},$$

بنابراین

$T = \sum_{i=1}^{n} x_i^a$ یک آماره کافی است.

(ج)

Known θ

unknown a

$$P(x|a) = \theta a^\theta x^{-(\theta+1)}$$

$$P(x_1 \dots x_n|a) = \prod_{i=1}^{n} P(x_i|a) = \theta^n \underbrace{a^{n\theta} \left(\underbrace{\prod_{i=1}^{n} x_i}\right)^{-(\theta+1)}}$$

$T = \prod_{i=1}^{n} x_i$ آماره کافی a

$T = \prod_{i=1}^{n} x_i$ آماره کافی θ

۲) اگر T یک آماره کافی باشد داریم که توزیع پسین فقط از طریق آماره $T(x)$ بستگی دارد

Bayes Formula: $$f_{w|x}(\theta|x)=\frac{f_x(x|\theta)\pi(\theta)}{\int_{\Theta} f_x(x|\psi)\pi(\psi)\nu(d\psi)}$$

با توجه به قضیه Neyman-Fisher Factorization داریم: $f_x(x|\theta)=h(x)g(\theta,T(x))$

$$f_{w|x}(\theta|x)=\frac{h(x)g(\theta,T(x))\pi(\theta)}{\int_{\Theta} h(x)g(\psi,T(x))\pi(\psi)\nu(d\psi)}$$

$$=\frac{g(\theta,T(x))\pi(\theta)}{\int_{\Theta} g(\psi,T(x))\pi(\psi)\nu(d\psi)}=f_{w|T(x)}(\theta|T(x))$$

✓

۲) الف

Bayesian Sufficiency $\Rightarrow$ Classical Sufficiency

$$\frac{f_x(x|\theta)\pi(\theta)}{f_x(x)}=f_{w|x}(\theta|x)=f_{w|T(x)}(\theta|T(x))$$

$$=\frac{f_{T(x)}(T(x)|\theta)\pi(\theta)}{f_{T(x)}(T(x))}$$

$$\Rightarrow f_x(x|\theta)=f_x(x)\frac{f_{T(x)}(T(x)|\theta)}{f_{T(x)}(T(x))}=h(x)g(\theta,T(x))$$

پس T یک آماره کافی است.

3 ابتدا تابع نرمال را بنویسیم.

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$L = \log f(x \mid \mu, \sigma^2) = -\tfrac{1}{2}\log(2\pi\sigma^2) - \tfrac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\frac{\partial L}{\partial \mu} = 0 + \frac{1}{2\sigma^2}(x-\mu) \times 2 = \frac{x-\mu}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial \mu^2} = \frac{-1}{\sigma^2} \Rightarrow \underbrace{I(\theta)}_{\text{fisher informati}} = -nE\left[\frac{\partial^2 L}{\partial L^2}\right] = \frac{n}{\sigma^2}$$

$$\text{var}(\hat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\sigma^2}{n}$$

با توجه به این که می‌دانیم $\bar{x}$ یک ~~MLE~~ sufficient statistic برای μ است پس $\bar{x}^3$ هم یک sufficient statistic برای μ^3 است.

$$E[\bar{x}^2] = \text{var}(\bar{x}) + (E[\bar{x}])^2 = \frac{\sigma^2}{n} + \mu^2$$

حال امید ریاضی $\bar{x}^3$ را حساب می‌کنیم تا ببینیم unbiased هست یا خیر!

$$E[\bar{x}^3] = ? \xrightarrow{\text{می‌دانیم}} E[(\bar{x}-\mu)^3] = 0 \Rightarrow E[\bar{x}^3 - 3\bar{x}^2\mu + 3\bar{x}\mu^2 - \mu^3] = 0$$

$$E[\bar{x}^3] = \mu^3 + 3\left(\frac{\sigma^2}{n} + \mu^2\right)\mu - 3\mu^3 = \mu^3 + \frac{3\sigma^2}{n}\mu$$

$$E\left[\bar{x}^3 - \frac{3\sigma^2}{n}\bar{x}\right] = \mu^3 \rightarrow \text{bias} = 0$$

تخمین‌گر = unbiased

برای بدست آوردن min variance equivalent آن داریم:

تابعی که می‌خواهیم تخمین بزنیم

$$\varphi(T) = E[w \mid T] = E\left[\bar{x}^3 - \frac{3\sigma^2}{n}\bar{x} \,\middle|\, \bar{x}^3\right] =$$

$$\boxed{\bar{x}^3 - \frac{3\sigma^2}{n}\bar{x}}$$

ج.

الف

ج

$$L(\theta) = f(x_1, \dots x_n ; \theta) = \prod_{i=1}^{n} \theta^{-2} x_i e^{-\frac{x_i}{\theta}}$$

$$K = \ln L(\theta) = \sum_{i=1}^{n} \log(\theta^{-2} x_i e^{-x_i/\theta}) = -2n \log\theta + \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \frac{-x_i}{\theta}$$

$$\frac{dK}{d\theta} = \frac{-2n}{\theta} + 0 + \sum_{i=1}^{n} \frac{x_i}{\theta^2} = 0 \Rightarrow 2n\theta = \sum_{i=1}^{n} x_i \Rightarrow \boxed{\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{2n}}$$

آیا با افزایش تعداد نمونه درستنمایی افزایش می‌یابد؟

$$L(\theta) = \prod_{i=1}^{n} \theta^{-2} x_i e^{-x_i/\theta}$$

نه لزوماً، با رسم plot_{3D} آن متوجه شویم که به ازای $0 < x, \theta < 1$ ممکن است $L(\theta)$ به ازای هر نمونه بیشتر از یک باشد که ضرب آن‌ها قاعدتاً با افزایش n بیشتر خواهد شود ولی به ازای $\theta, x >$. اینطور نیست و $L(\theta) < 1$ است پس حتما با افزایش n مقدار $L(\theta)$ کاهشی خواهد بود.

ب

می‌خواهیم θ را تخمین بزنیم پس به گشتاورها نیاز داریم.

$$E[x] = \text{Sample mean} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$E[x^k] = \int_0^\infty \underbrace{\frac{x^k}{\theta^2} x}_{v} \underbrace{e^{-\frac{x}{\theta}} dx}_{du} = \frac{-\theta e^{-\frac{x}{\theta}} x^{k+1}}{\theta^2} \Big|_0^\infty - (k+1) \int \frac{-\theta e^{-\frac{x}{\theta}} x^k}{\theta^2} dx = (*)$$

$$du = e^{-\frac{x}{\theta}} dx \Rightarrow u = \int e^{-\frac{x}{\theta}} dx = -\theta e^{-\frac{x}{\theta}}$$

$$v = \frac{x^{k+1}}{\theta^2} \Rightarrow dv = \frac{(k+1) x^k}{\theta^2} dx$$

$$(*) = 0 + (k+1)\, k!\, \theta^k \Rightarrow E[x] = 2\theta$$

$$\hat{\theta}_{MM} = \frac{\sum_{i=1}^{n} x_i}{2n} \Rightarrow \boxed{\hat{\theta}_{MM} = \hat{\theta}_{MLE}}$$

برابر است !

d3b1

۴) الف) البته برای سوال ۴ قسمت اول باید بیشتر توضیح دهم. با توجه به اینکه در سوال گفته شده که $\log L(\theta)$ مدنظر است تغییرات زیر باید در نظر گرفته شود:

البته این تابع $\log L(\theta)$ همانند $L(\theta)$ رفتار یکنواختی ندارد و به صورت کلی می‌توان گفت که Likelihood با افزایش n رابطه خاصی ندارد. نمی‌توان گفت همیشه کم می‌شود که همیشه زیاد می‌شود

اما در اینجا اگر شرط $(x;\theta) < 1$ را لحاظ کنیم آنگاه $\log L(\theta) < 0$ همیشه منفی است و جمع یک سری عدد منفی نیز منفی خواهد بود که قطعاً تابعی هر بار کمتر می‌شود.

اما اگر شرط $(x,\theta) > 1$ را لحاظ کنیم $\log L(\theta) > 0$ می‌شود که هر بار بیشتر می‌شود

می‌توان این نتایج را با پلات کردن تابع (3D) $\log L(\theta)$ نسبت به θ و x دید

$$E\big[|x|\big] = \int_{-\infty}^{+\infty} |x|\, f_{x|\theta}(x|\theta)\, dx = \int_{-\infty}^{+\infty} |x| \frac{1}{2\theta} \exp\left(-\frac{|x|}{\theta}\right) dx$$

$$= \int_0^{\infty} \frac{x}{\theta} \exp\left(\frac{-x}{\theta}\right) dx \overset{}{\underset{\frac{x}{\theta}=v}{=\!=\!=}} \theta \int_0^{\infty} v \exp(-v)\, dv = \theta$$

$$E\left[|x|^2\right] = \int_{-\infty}^{+\infty} |x|^2 f_{x|\theta}(x|\theta)\, dx = \int_0^{\infty} x^2 \frac{1}{\theta^2} \exp\left(\frac{-x}{\theta}\right) dx =$$

$$= \theta^2 \int v^2 \exp(-v)\, dv = 2\theta^2$$

$$E(\hat{\theta}) = E\left(\frac{\sum_{i=1}^{n} |x_i|}{n}\right) = \frac{1}{n}\left(E\big[|x_1|\big] + \cdots + E\big[|x_n|\big]\right) = \frac{n\theta}{n} = \theta \rightarrow \text{bias} = 0$$

$$MSE(\hat{\theta}) = var(\hat{\theta}) + (\text{bias}(\hat{\theta}))^2 = \cancel{E(\theta - \hat{\theta})^2}$$

واریانس‌ها برابر

استقلال

$$var(\hat{\theta}) = var\left(\frac{|x_1| + \cdots |x_n|}{n}\right) = \frac{1}{n^2}\left(var(|x_1|) + \cdots + var(x_n)\right) =$$

$$\frac{var|x|}{n} = \frac{E(|x|^2) - (E(|x|))^2}{n} = \frac{2\theta^2 - \theta^2}{n} = \frac{\theta^2}{n}$$

$$\boxed{MSE(\hat{\theta}) = \frac{\theta^2}{n}}$$

$\cancel{P(x|\theta)}$

$$P(\theta|x) = \frac{P(\theta;a)\,P(x|\theta)}{P(x)} = \frac{P(\theta;a)\,P(x|\theta)}{\int P(x|\theta)\,P(\theta)\,d\theta}$$

می‌دانیم

$$P(\theta|x,a_i) = \frac{P(\theta;a_i)\,P(x|\theta;a_i)}{\int P(x|\theta,a_i)\,P(\theta;a_i)\,d\theta \longrightarrow P(x;a_i)}$$

حال

$$P(\theta|x) = \frac{\sum_i B_i\,P(x|\theta)\,P(\theta,a_i)}{\int \sum_i B_i\,P(x|\theta)\,P(\theta,a_i)\,d\theta} = \frac{\sum_i B_i\,P(\theta|x,a_i)\,P(x|a_i)}{\sum_i B_i\,P(x|a_i)}$$

$$= \sum_i \tilde{B}_i\,P_i(\theta|x), \quad \tilde{B}_i = \frac{B_i\,P(x|a_i)}{\sum_{i'} B_{i'}\,P(x|a_{i'})}$$

v

$$P_1, \dots P_6 \sim \tfrac{1}{3}\, Dir(\alpha_{11}, \dots \alpha_{16}) + \tfrac{2}{3}\, Dir(\alpha_{21}, \dots \alpha_{26})$$

$$y \sim Cat(P_1, \dots P_6)$$

$$f(\Theta \mid D) = \text{Posterior} \propto f(\Theta, D) = f\left(P_1 \dots P_6 \,\middle|\, \begin{matrix} a_{11} \dots a_{16} \\ a_{21} \dots a_{26} \end{matrix}\right) \prod_{y_i \in D} f(y_i \mid P_1 \dots P_6)$$

$$\propto \frac{1}{3} \prod_{j=1}^{6} P_j^{\alpha_{1j}-1} \prod_{y_i \in D} \prod_{j=1}^{6} P_j^{I\{y_i=j\}} + \frac{2}{3} \prod_{j=1}^{6} P_j^{\alpha_{2j}-1} \prod_{y_i \in D} \prod_{j=1}^{6} P_j^{I\{y_i=j\}}$$

$$= \frac{1}{3} \prod_{j=1}^{6} P_j^{a_{1j}-1+\sum_{y_i \in D} I\{y_i=j\}} + \frac{2}{3} \prod_{j=1}^{6} P_j^{a_{2j}-1+\sum_{y_i \in D} I\{y_i=j\}}$$

$$= \frac{1}{3}\, Dir\left(a_{1j} + \textstyle\sum_{y_i \in D} I(y_i=j)\right) + \frac{2}{3}\, Dir\left(a_{2j} + \textstyle\sum_{y_i \in D} I(y_i=j)\right)$$

$$a_{1j}' = a_{1j} + \sum_{y_i \in D} I(y_i=j) \qquad a_{2j}' = a_{2j} + \sum_{y_i \in D} I(y_i=j)$$

Posterior Predictive

$$f(y=x \mid D) = \int f(y=x \mid \theta)\, f(\theta \mid D)\, d\theta = \int f(y=x \mid P_1 \dots P_6)\, f(P_1 \dots P_6)\, dS_K$$

$$= \frac{1}{3} \int P_x \frac{\Gamma\left(\sum_{j=1}^{K} \alpha_j'\right)}{\prod_{j=1}^{K} \Gamma(\alpha_j')} \prod_{j=1}^{K} P_j^{a_{ij}'-1}\, dS_K + \frac{2}{3} \int \frac{P_x \Gamma\left(\sum_{j=1}^{K} a_j'\right)}{\prod_{j=1}^{K} \Gamma(\alpha_j')} \prod P_j^{a_{2j}'-1}\, dS_K$$

$$= \frac{1}{3} \frac{\Gamma\left(\sum_{j=1}^{K} a_{ij}'\right)}{\prod_{j=1}^{K} \Gamma(\alpha_{ij}')} \int \prod_{j=1}^{K} P_j^{I(x=j)+a_{ij}'-1}\, dS_K + \frac{2}{3} \frac{\Gamma\left(\sum_{j=1}^{K} a_{2j}'\right)}{\prod_{j=1}^{K} \Gamma(a_{2j}')} \int \prod_{j=1}^{K} P_j^{I(x=j)+a_{2j}'-1}\, dS_K$$

ادامه ۷

$$= \frac{1}{3} \frac{\Gamma\left(\sum_{j=1}^{K} a_{ij}'\right) \prod_{j=1}^{K} \Gamma\left(I(x=j) + a_{ij}'\right)}{\prod_{j=1}^{K} \Gamma(\alpha_{ij}')\ \ \Gamma\left(1 + \sum_{j=1}^{K} a_{ij}'\right)} +$$

$$\frac{2}{3} \frac{\Gamma\left(\sum_{j=1}^{K} a_{2j}'\right) \prod_{j=1}^{K} \Gamma\left(I(x=j) + a_{2j}'\right)}{\prod_{j=1}^{K} \Gamma(\alpha_{2j}')\ \ \Gamma\left(1 + \sum_{j=1}^{K} a_{2j}'\right)}$$

$$= \frac{1}{3} \frac{a_x'}{\sum_{j=1}^{K} a_{ij}'} + \frac{2}{3} \frac{a_x'}{\sum_{j=1}^{K} a_{2j}'}$$