$\boxed{1.a}$

$$\text{posterior} = f(\theta|X) = \frac{f(X|\theta)\, f(\theta)}{\int_0^{\frac{1}{\sqrt5}} f(X|\theta)\, f(\theta)\, d\theta} = \frac{\theta e^{-\theta X}\ (10\theta)}{\int_0^{\frac{1}{\sqrt5}} 10\theta^2 e^{-\theta X}\, d\theta}$$

$$= \frac{10\theta^2 e^{-\theta X}}{\left.\dfrac{-10(X^2\theta^2 + 2X\theta + 2)e^{-\theta X}}{X^3}\right|_{\theta=0}^{\theta=\frac{1}{\sqrt5}}} = \frac{\theta^2 e^{-\theta X}\, X^3}{2-\left(X^2\frac{1}{5} + 2X\frac{1}{\sqrt5} + 2\right)e^{-\frac{X}{\sqrt5}}} \qquad X \geq 0,\quad 0 \leq \theta \leq \frac{1}{\sqrt5}$$

$$\longrightarrow f(\theta|X=30) = \frac{27000\,\theta^2 e^{-30\theta}}{2-(180 + 12\sqrt5 + 2)e^{-6\sqrt5}}$$

$\boxed{1.b}$

$$\hat\theta = \underset{\theta}{\operatorname{argmax}}\ f(\theta|X=30) = \underset{\theta}{\operatorname{argmax}}\ \underbrace{\lg 27000}_{\text{const}} + \lg(\theta^2 e^{-30\theta}) - \underbrace{\lg\left(2-(180+12\sqrt5+2)e^{-6\sqrt5}\right)}_{\text{const}}$$

$$\xrightarrow{\frac{\partial}{\partial\theta}} \frac{2\theta\cancel{e^{-30\theta}} - 30\cancel{e^{-30\theta}}\theta^2}{\theta^2\cancel{e^{-30\theta}}} = \frac{2-30\theta}{\theta} = 0 \longrightarrow \theta = \frac{1}{15}$$

همان طور که ملاحظه می‌شود در MAP مخرج posterior مهم نیست زیرا ترم θ ندارد و در مرحله مشتق گیری 0 می‌شود

$\boxed{2.1}$

$$\mathbb{E}\left[\hat I_n(f)\right] = \mathbb{E}\left[\frac1n \sum_{i=1}^n \frac{f(x_i)}{g(x_i)}\right] = \frac1n \sum_{i=1}^n \underbrace{\mathbb{E}\left[\frac{f(x_i)}{g(x_i)}\right]}_{\int_a^b \frac{f(x_i)}{g(x_i)} g(x_i)\, dx_i} = \frac1n \times n \times \int_a^b f(x_i)\, dx_i = I(f)$$

$\boxed{2.2}$

باید نشان دهیم که تخمینگر $\hat I_n(f)$ consistent است. بنابراین باید نشان دهیم: $\lim_{n\to\infty} \mathbb{P}\left(\left|\hat I_n(f) - I(f)\right| \geq \varepsilon\right) = 0 \quad \forall\varepsilon>0$

طبق law of large numbers برای $\mathbb{E}$ می‌دانیم که اگر $Z_1 \dots Z_n$ به صورت iid باشد داریم: $\lim_{n\to\infty} \mathbb{P}\left(\left|\bar Z_n - \mathbb{E}[Z]\right| \geq \varepsilon\right) = 0 \quad \forall\varepsilon>0$

حال اگر قرار دهیم: $Z_i = \frac{f(x_i)}{g(x_i)}$ خواهیم داشت $\bar Z_n = \hat I_n(f)$ و $\mathbb{E}[Z] = I(f)$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat I_n(f) - I(f)\right| \geq \varepsilon\right) = 0 \quad \forall\varepsilon>0$$

بنابراین $\hat I_n(f)$ یک تخمینگر consistent است و در تعداد داده بالا به مقدار اصلی نزدیک می‌شود

$\boxed{2.3}$ $$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i)}{g(x_i)}\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[\frac{f(x_i)}{g(x_i)}\right] = \frac{1}{n}\left[\mathbb{E}\left[\left(\frac{f(x_i)}{g(x_i)}\right)^2\right] - I(f)^2\right]$$

$$= \frac{1}{n}\left[\int_a^b \frac{f(x_i)^2}{g(x_i)}\, dx_i - I(f)^2\right]$$

$\boxed{3}$

ابتدا حالتی را به‌دست می‌آوریم که σ^2 را می‌دانیم: متغیر تصادفی Z را به صورت زیر تعریف می‌کنیم:

$$Z = \frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1) \longrightarrow \mathbb{P}(|Z| \geq b) \leq 1-0.95 \longrightarrow b = -\Phi^{-1}\Big(\underbrace{\frac{1-0.95}{2}}_{0.025}\Big) = 1.96$$

$$\longrightarrow -1.96 \leq \frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} \leq 1.96 \longrightarrow \theta \in \left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right] \xrightarrow[\sigma=5]{\bar{X}=65} [61.9\ ,\ 68.1]$$

حال برای حالتی که σ^2 را ندانیم حساب می‌کنیم. متغیر تصادفی Z و Y را به صورت زیر تعریف می‌کنیم:

$$Z = \frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$$

$$Y = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{chi-squared}$$

طبق تعریف توزیع t-distribution می‌دانیم

$$\frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n)}{n}}} = T(n) \longrightarrow T = \frac{Z}{\sqrt{\frac{Y}{n-1}}} = \frac{\bar{X}-\theta}{s/\sqrt{n}} \sim T(n-1)$$

اگر CDF^{-1} توزیع $T(n-1)$ را t_{n-1} بنامیم مشابه قبل داریم:

$$\theta \in \left[\bar{X} - t_{n-1}\Big(\frac{1-0.95}{2}\Big)\frac{s}{\sqrt{n}}\ ,\ \bar{X} + t_{n-1}\Big(\frac{1-0.95}{2}\Big)\frac{s}{\sqrt{n}}\right]$$

$$\left.\begin{aligned} &\bar{X} = 65 \\ &s = 5.98 \\ &t_9(0.025) = 2.262 \end{aligned}\right\} \longrightarrow \theta \in [60.72\ ,\ 69.28]$$

⊛ همان طور که مشاهده می‌شود اگر واریانس را بدانیم بازه tight تری برای θ به‌دست می‌آوریم که منطقی هم هست زیرا اطلاعات بیشتری داشتیم.

⊛ به‌طور کلی وقتی واریانس را ندانیم روش‌های مختلفی وجود دارد مثلاً می‌توان برای σ یک تخمینگر بازه‌ای به‌دست آورد و با حد بالای آن به بازه‌ای برای θ رسید اما کاری که در اینجا کردیم جواب دقیق‌تری است.

$\boxed{4.a}$

با توجه به رابطه داده شده در رگرسیون خطی : $f(Y|X=x_i) \sim \mathcal{N}(\beta_0+\beta_1 x_i \,,\, \sigma^2)$

$$\hat\beta_0 , \hat\beta_1 = \underset{\beta_0 , \beta_1}{\arg\max} \prod_{i=1}^{n} \underbrace{f(Y=y_i | X=x_i)}_{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i-(\beta_0+\beta_1 x_i))^2}{2\sigma^2}\right)}$$

$$= \underset{\beta_0 , \beta_1}{\arg\max} \sum_{i=1}^{n} -\left(y_i - (\beta_0+\beta_1 x_i)\right)^2 = \underset{\beta_0 , \beta_1}{\arg\min} \sum_{i=1}^{n} \left(y_i - (\beta_0+\beta_1 x_i)\right)^2 \longrightarrow MSE$$

$\boxed{4.b}$

$$\frac{\partial MSE}{\partial \beta_0} = \sum_{i=1}^{n} -2(y_i - \beta_0 - \beta_1 x_i) = 0 \longrightarrow \hat\beta_0 = \frac{\sum (y_i - \hat\beta_1 x_i)}{n} = \boxed{\bar y - \hat\beta_1 \bar x}$$

$$\frac{\partial MSE}{\partial \beta_1} = \sum_{i=1}^{n} -2(y_i - \beta_0 - \beta_1 x_i) x_i = 0 \longrightarrow \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0$$

$$\longrightarrow \sum x_i y_i - \bar y \sum x_i + \underbrace{\beta_1 \bar x \sum x_i - \beta_1 \sum x_i^2}_{\beta_1 (n\bar x^2 - \sum x_i^2)} = 0$$

$$\longrightarrow \hat\beta_1 = \frac{\sum x_i y_i - n \bar y \bar x}{\sum x_i^2 - n \bar x^2} = \boxed{\frac{\sum (x_i - \bar x)(y_i - \bar y)}{\sum (x_i - \bar x)^2}}$$

هر دوی β_0 ، β_1 از ترکیب خطی و ضرب گاوسی ها به دست آمده اند پس گاوسی هستند.

$$\hat\beta_1 = \frac{\sum (x_i - \bar x)(\beta_1 (x_i - \bar x) + \varepsilon_i - \bar\varepsilon)}{\sum (x_i - \bar x)^2} = \beta_1 + \frac{\sum (x_i - \bar x)\varepsilon_i}{\sum (x_i - \bar x)^2} - \underbrace{\frac{\bar\varepsilon \overbrace{\sum (x_i - \bar x)}^{0}}{\sum (x_i - \bar x)^2}}_{0}$$

$$\mathbb{E}[\hat\beta_1] = \beta_1 + 0 = \beta_1 \longrightarrow \text{unbiased } \checkmark$$

$$Var[\hat\beta_1] = 0 + \frac{1}{(\sum (x_i - \bar x)^2)^2} \underbrace{var\left[\sum (x_i - \bar x)\varepsilon_i\right]}_{\sum (x_i - \bar x)^2 \sigma^2} = \frac{\sigma^2}{\sum (x_i - \bar x)^2}$$

$$\longrightarrow \hat\beta_1 \sim \mathcal{N}\left(\beta_1 \,,\, \frac{\sigma^2}{\sum (x_i - \bar x)^2}\right)$$

4.b ادامه

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon} - \hat{\beta}_1 \bar{x} = \beta_0 - \frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2}\bar{x} + \bar{\varepsilon}$$

$$\mathbb{E}[\hat{\beta}_0] = \beta_0 + 0 + 0 = \beta_0 \longrightarrow \text{unbiased } \checkmark$$

$$\operatorname{Var}[\hat{\beta}_0] = \frac{\bar{x}^2 \sigma^2}{\sum (x_i - \bar{x})^2} + \frac{\sigma^2}{n} = \frac{\sum x_i^2 - n\bar{x}^2 + n\bar{x}^2}{n\sum (x_i - \bar{x})^2}\sigma^2 = \frac{\sigma^2 \sum x_i^2}{n\sum (x_i - \bar{x})^2}$$

$$\longrightarrow \hat{\beta}_0 \sim \mathcal{N}\left(\beta_0 , \frac{\sigma^2 \sum x_i^2}{n\sum (x_i - \bar{x})^2}\right)$$

4.c

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} - \frac{\bar{y}\overbrace{\sum (x_i - \bar{x})}^{0}}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x}) x_i - \bar{x}\underbrace{\sum (x_i - \bar{x})}_{0}}$$

$$= \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x}) x_i}$$

$\checkmark \ \dfrac{\sum \gamma_i y_i}{\sum \gamma_i x_i} \longleftarrow \gamma_i = x_i - \bar{x}$ بنابراین اگر انتخاب کنیم

بنابراین عضو این خانواده است

$\checkmark \ \sum \gamma_i = \sum (x_i - \bar{x}) = 0$

4.d

$$\tilde{\beta}_1 = \frac{\sum \gamma_i y_i}{\sum \gamma_i x_i} = \frac{\sum \gamma_i (\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum \gamma_i x_i} = \frac{\beta_0 \overbrace{\sum \gamma_i}^{0} + \beta_1 \sum \gamma_i x_i + \sum \gamma_i \varepsilon_i}{\sum \gamma_i x_i} = \beta_1 + \frac{\sum \gamma_i \varepsilon_i}{\sum \gamma_i x_i}$$

$$\longrightarrow \mathbb{E}[\tilde{\beta}_1] = \beta_1 + 0 = \beta_1 \longrightarrow \text{unbiased}$$

$\boxed{4.e}$

$$\sum \gamma_i (x_i - \bar{x}) = \sum \gamma_i x_i - \underbrace{\bar{x} \sum \gamma_i}_{0} = \sum \gamma_i x_i \quad (*)$$

$\boxed{4.d}$ طبق بخش

$$\longrightarrow Var\left[\tilde{\beta}_1\right] = \frac{\sigma^2 \sum \gamma_i^2}{\left(\sum \gamma_i x_i\right)^2}$$

$\boxed{4.b}$ طبق بخش

$$\longrightarrow Var\left[\hat{\beta}_1\right] = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

برای مقایسه این دو واریانس از نامساوی کوشی شوارتز استفاده می کنیم . به طور کلی داریم:

$$\underbrace{\left(\sum \gamma_i (x_i - \bar{x})\right)^2}_{\left(\sum \gamma_i x_i\right)^2 \; (*)} \leqslant \left(\sum \gamma_i^2\right)\left(\sum (x_i - \bar{x})^2\right)$$

$$\longrightarrow \frac{\sum \gamma_i^2}{\left(\sum \gamma_i x_i\right)^2} \geqslant \frac{1}{\sum (x_i - \bar{x})^2} \xrightarrow{\times \sigma^2} Var\left[\tilde{\beta}_1\right] \geqslant Var\left[\hat{\beta}_1\right]$$

$\boxed{5}$

$$\hat{\lambda} = \underset{\lambda}{\arg\max}\ \underbrace{f_\lambda(\lambda)}_{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}} \prod_{i=1}^{n} \underbrace{f(X_i|\lambda)}_{\lambda e^{-\lambda X_i}} = \underset{\lambda}{\arg\max}\ \alpha \lg\beta + (\alpha-1)\lg\lambda - \beta\lambda + \sum_{i=1}^{n} \lg\lambda - \lambda X_i$$

$$\xrightarrow{\frac{\partial}{\partial\lambda}} \frac{\alpha-1}{\lambda} - \beta + \sum_{i=1}^{n} \frac{1}{\lambda} - X_i = 0 \longrightarrow \frac{\alpha+n-1}{\lambda} = \beta + \sum_{i=1}^{n} X_i$$

$$\longrightarrow \hat{\lambda} = \frac{\alpha+n-1}{\beta + \sum_{i=1}^{n} X_i}$$

$\boxed{6}$

$$\hat{X} = \underset{X}{\arg\max}\ f(X|Y=3) = \underset{X}{\arg\max}\ P(Y=3|X)\ f(X)$$

آن فرم از توزیع geo را در نظر گرفتیم که ساپورتش شامل صفر نیست

$$= \underset{X}{\arg\max}\ (1-X)^2 X \ (3X^2) \qquad 0 \leqslant X \leqslant 1$$

$$\xrightarrow{\frac{\partial}{\partial X}} 9X^2(1-X)^2 - 6(1-X)X^3 = 0 \longrightarrow X = 0, \frac{3}{5}, 1$$

فقط $\frac{3}{5}$ ماکسیمم می کند $\longrightarrow \boxed{\hat{X} = \frac{3}{5}}$

$\boxed{7}$

$$f(Y|X;a) = f(aX+\varepsilon | X;a) \longrightarrow \text{گاوسی است}$$

$$\mathbb{E}[Y|X;a] = aX \quad , \quad \operatorname{Var}[Y|X;a] = \sigma^2$$

$$\hat{a} = \underset{a}{\arg\max} \prod_{i=1}^{n} \underbrace{f(Y_i|X_i;a)}_{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_i - aX_i)^2}{2\sigma^2}\right)} = \underset{a}{\arg\max} \sum_{i=1}^{n} \underbrace{-\lg(\sqrt{2\pi}\sigma)}_{\text{const}} - \frac{(Y_i - aX_i)^2}{2\sigma^2}$$

$$\xrightarrow{\frac{\partial}{\partial a}} \sum_{i=1}^{n} \frac{-1}{2\sigma^2} 2(Y_i - aX_i)(-X_i) = 0 \longrightarrow \frac{1}{\sigma^2} \sum_{i=1}^{n} Y_i X_i - aX_i^2 = 0$$

$$\longrightarrow \sum_{i=1}^{n} Y_i X_i = a \sum_{i=1}^{n} X_i^2 \longrightarrow a = \frac{\sum Y_i X_i}{\sum X_i^2}$$

- source 1