Question 1

a) $(X \perp\!\!\!\perp Y, W \mid Z) \Rightarrow (X \perp Y \mid Z)$

True! Here's the proof:

$$P(X, Y, W \mid Z) = p(X \mid Z)\, p(Y, W \mid Z)$$ (definition of conditional probability)

Integrate w.r.t. to W:

$$\Rightarrow \int_W P(X, Y, W \mid Z)\, dW = p(X \mid Z) \int_W P(Y, W \mid Z)$$

$$\Leftrightarrow p(X, Y \mid Z) = p(X \mid Z)\, p(Y \mid Z)$$

That is $(X \perp\!\!\!\perp Y \mid Z)$ by definition. ■

b) $\underset{①}{(X \perp\!\!\!\perp Y \mid Z)}$ and $\underset{②}{(X, Y \perp\!\!\!\perp W \mid Z)} \Rightarrow (X \perp\!\!\!\perp W \mid Z)$

True! Here's a proof:

Start with $P(X, Y, W \mid Z) = p(X, Y \mid Z)\, p(W \mid Z)$ (from ②)

$$\Leftrightarrow p(X, Y, W \mid Z) = p(X \mid Z)\, p(Y \mid Z)\, p(W \mid Z)$$ (from ①)

Integrate w.r.t. y

$$\int_y P(x,y,w|z)\,dy = p(x|z)\,p(w|z)\underbrace{\int_y dy\, p(y|z)}_{=1}$$

$$\Leftrightarrow p(x,w|z) = p(x|z)\,p(w|z)$$

i.e. $(X \perp\!\!\!\perp W \mid Z)$ $\blacksquare$

c) $(X \perp\!\!\!\perp Y,W|Z)$ and $(Y \perp W|Z) \Rightarrow (X,W \perp Y|Z)$

I don't think this is true. Here's a counter example: Let

$Z = \{0,1\}$ be a binary variable,

$X = Z$

$$\left.\begin{aligned} Y &= Z \oplus \{0,1\}_y \\ W &= Z \oplus \{0,1\}_w \end{aligned}\right\}$$ here $\oplus$ is "XoR" and $\{0,1\}$ is another binary random variable. Note that $\{0,1\}_y$ and $\{0,1\}_w$ are independent

Since Z completely determines X, we have that $(X \perp\!\!\!\perp Y,W|Z)$.

Also, given Z, Y and W are independent of each other, so

$$(Y \perp\!\!\!\perp W \mid Z)$$

but $Y = Z \oplus \{0,1\} = X \oplus \{0,1\}$ $(X = Z)$

So it is not true that $(X \perp\!\!\!\perp Y \mid Z)$

So $(X, W \perp\!\!\!\perp Y \mid Z)$ is not true in general

d) $(X \perp\!\!\!\perp Y \mid Z)$ and $(X \perp\!\!\!\perp Y \mid W) \Rightarrow (X \perp\!\!\!\perp Y \mid Z, W)$

Not true!

Let $X, Y, Z = \{-1, 1\}$ be iid

$W = XYZ \quad \Rightarrow WZ = XY$

(multiplying or dividing by ± 1 is the same thing)

Since X, Y, Z are iid, clearly

$(X \perp\!\!\!\perp Y \mid Z)$ and $(X \perp\!\!\!\perp Y \mid W)$

but since knowing Z & W determine XY, it is not true that $(X \perp\!\!\!\perp Y \mid Z, W)$

Question 2 likelyhood

Let $\vec{X}_1, \ldots, \vec{X}_n \mid \vec{\pi} \overset{iid}{\sim} \text{Multinomial}(1, \vec{\pi})$ on k elements.

Values $\vec{x}_i$ that $\vec{X}_i$ can take is a one-hot vector.

Prior on $\vec{\pi}$ is $\vec{\pi} \sim \text{Dirichlet}(\vec{\alpha})$, $\vec{\alpha} = (\alpha_1, \ldots, \alpha_k)$

$\alpha_j > 0 \; \forall j$

so $p(\vec{\pi}; \vec{\alpha}) = \dfrac{\Gamma\left(\sum_{j=1}^{k} \alpha_j\right)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \prod_{j=1}^{k} \pi_j^{\alpha_j - 1}$

prior

a) Since the $\vec{X}_i$ are iid given $\vec{\pi}$ we have that

$$(\vec{X}_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp \vec{X}_n \mid \vec{\pi}).$$

all mutually independent given $\vec{\pi}$!

b) I know that $\vec{X}_1, \ldots, \vec{X}_i \overset{iid}{\sim} \text{Multinomial}(1, \vec{\pi})$

with $\Omega_{X_i} = \{\vec{e}_1, \ldots, \vec{e}_k\}$ and $e_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ j^{th} position

Let $\vec{X} = \sum_{i=1}^{n} X_i \sim \text{Multinomial}(n, \vec{\pi})$

Now $\Omega_X = \{(n_1, \ldots, n_k) \mid n_j \in \mathbb{N}, \sum_{j=1}^{n} n_j = n\}$

Likelihood is $p(\vec{x} \mid \vec{\pi}) = \text{multinomial}(n, \vec{\pi})$

$$\Rightarrow p(\vec{x}, \vec{\pi}) = \underbrace{\binom{n}{(x)_1 \ldots (x)_k}}_{\text{multinomial coefficient}} \prod_{j=1}^{k} \pi_j^{(x)_j} \underset{\text{"proportional to"}}{\propto} \prod_{j=1}^{k} \pi_j^{(x)_j}$$

where $(x)_j = n_j$

Prior is $p(\vec{\pi} \mid \vec{\alpha}) = \text{Dirichlet}(\vec{\alpha})$

$$\Rightarrow p(\vec{\pi} \mid \vec{\alpha}) = \frac{\Gamma\left(\sum_{j=1}^{k} \alpha_j\right)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \prod_{j=1}^{k} \pi_j^{\alpha_j - 1} \propto \prod_{j=1}^{k} \pi_j^{\alpha_j - 1}$$

Can find the posterior using Baye's rule:

$$p(\vec{\pi} \mid \vec{x}) = \frac{p(\vec{\pi} \mid \vec{\alpha})\, p(\vec{x} \mid \vec{\pi})}{p(\vec{x})}$$

$$\propto p(\vec{\pi} \mid \vec{\alpha})\, p(\vec{x} \mid \vec{\pi})$$

$$\propto \left(\prod_{j=1}^{K} \pi_j^{\alpha_j - 1}\right)\left(\prod_{j=1}^{K} \pi_j^{(x)_j}\right)$$

$$\Rightarrow \boxed{p(\vec{\pi} \mid \vec{x}) \propto \prod_{j=1}^{K} \pi_j^{\alpha_j + (x)_j - 1}}$$

The posterior is a <u>Dirichlet</u> distribution with parameters $\underline{\vec{\alpha} + \vec{x}}$. In words, the posterior is the prior, but w/ parameters α_j "updated" by the number of counts in category j: $(x)_j$.

Can go a bit further to normalize this

$$\vec{\pi} \in \Delta_K = \left\{ \vec{\pi} \in \mathbb{R}^K \mid 0 \le \pi_j \le 1,\ \underline{\sum_{j=1}^{K} \pi_j = 1} \right\}$$

Same as

$$1 - \sum_{j=1}^{K-1} \pi_j = \pi_K$$

also let $u_j \equiv \alpha_j + (x)_j$

Then

$$A\int_{\Delta_K} d^K\vec{\pi}\, p(\vec{\pi}\,|\,\vec{x}) = A\int_{\Delta_K} d^K\vec{\pi} \prod_{j=1}^{K} \pi_j^{u_j-1} = 1$$

$$= A\int \cdots \int d\pi_1 \cdots d\pi_{\cancel{K}}\, \pi_1^{u_1-1} \cdots u_K^{u_K-1}$$

This is a special case of a Type I Dirichlet integral:

$$I = \int \cdots \int f(t_1 + \cdots + t_n)\, t_1^{a-1} \cdots t_n^{a_n-1}\, dt_1 \cdots dt_n$$

$$= \frac{\prod_{i=1}^{n} \Gamma(a_i)}{\Gamma\left(\sum_{i=1}^{n} a_n\right)} \int_0^1 f(\tau)\, \tau^{\left(\sum_{i=1}^{n} a_i - 1\right)} d\tau$$

In my case $f(t_1 + \cdots + t_n) = 1$

$$\Rightarrow A \int \cdots \int d\pi_1 \cdots d\pi_K\, \pi_1^{u_1-1}\, \pi_K^{u_K-1}$$

$$= A \frac{\prod_{i=1} \Gamma(u_i)}{\Gamma\left(\sum_{i=1} u_i\right)} = 1$$

$$\Rightarrow \quad A = \frac{\Gamma\left(\sum_{i=1}^{k} u_i\right)}{\prod_{i=1}^{k} \Gamma(u_i)} = \frac{\Gamma\left(\sum_{i=1}^{k} (\alpha_j + (x)_j)\right)}{\prod_{i=1}^{k} \Gamma\left(\alpha_j + (x)_j\right)}$$

and $$p(\vec{\pi}|\vec{x}) = \frac{\Gamma\left(\sum_{j=1}^{k} (\alpha_j + (x)_j)\right)}{\prod_{j=1}^{k} \Gamma\left(\alpha_j + (x)_j\right)} \prod_{j=1}^{k} \pi_j^{\alpha_j + (x)_j - 1}$$

which is indeed a properly normalized Dirichlet distribution.

c) Compute $p(\vec{x})$. It is given by

$$p(\vec{x}|\vec{\alpha}) = \int d^k\vec{\pi}\, p(\vec{x}|\vec{\pi})\, p(\vec{\pi}|\vec{\alpha})$$

$$= \int d^k\vec{\pi} \left(\prod_{j=1}^{k} \pi_j^{(x)_j}\right)\left(\frac{\Gamma\left(\sum_{j=1}^{k} \alpha_j\right)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \prod_{j=1}^{k} \pi_j^{\alpha_j - 1}\right)$$

$$= \frac{\Gamma\left(\sum_{j=1}^{k} \alpha_j\right)}{\prod_{j=1}^{k} \Gamma(\alpha_k)} \int d^k\vec{\pi} \prod_{j=1}^{k} \pi_j^{\alpha_j + (x)_j - 1}$$

Type I Dirichlet integral again

$$= \frac{\prod_{j=1}^{K} \Gamma(\alpha_j + (x)_j)}{\Gamma\left(\sum_{j=1}^{K} (\alpha_j + (x)_j)\right)}$$

$$\Rightarrow \boxed{p(\vec{x} \mid \vec{\alpha}) = \frac{\Gamma\left(\sum_{j=1}^{K} \alpha_j\right)}{\Gamma\left(\sum_{j=1}^{K} (\alpha_j + (x)_j)\right)} \prod_{j=1}^{K} \frac{\Gamma(\alpha_j + (x)_j)}{\Gamma(\alpha_j)}}$$

is the marginal likelyhood

d) Derive the MAP estimate for $\vec{\pi}$, assuming $\alpha_j > 1 \ \forall j$.

computed in b)

$$\hat{\vec{\pi}}_{map}(\vec{x}) = \underset{\vec{\pi} \in \Delta_K}{\arg\max}\ p(\vec{\pi} \mid \vec{x})$$

$$= \underset{\vec{\pi} \in \Delta_K}{\arg\max} \frac{\Gamma\left(\sum_{j=1}^{K} (\alpha_j + (x)_j)\right)}{\prod_{j=1}^{K} \Gamma(\alpha_j + (x_j))} \prod_{j=1}^{K} \pi_j^{\alpha_j + (x_j) - 1}$$

does not depend
on K so drop.

$$= \underset{\vec{\pi} \in \Delta_K}{\arg\max} \prod_{j=1}^{K} \pi_j^{\alpha_j + (x)_j - 1}$$

Product is annoying so let me maximize the log of the posterior (log monotonically increasing so does not change the value of argmax)

$$\hat{\vec{\pi}}_{MAP} = \underset{\vec{\pi} \in \Delta_K}{\arg\max} \log \prod_{j=1}^{K} \pi_j^{\alpha_j + (x)_j - 1}$$

$$= \underset{\vec{\pi} \in \Delta_K}{\arg\max} \sum_{j=1}^{K} (\alpha_j + (x)_j - 1) \log \pi_j$$

optimize this with respect to π_j, under the constraint that $\sum_{j=1}^{K} \pi_j = 1$.

Need to use the method of lagrange multipliers: $\mathcal{L}(\vec{\pi}, \lambda) = \sum_{j=1}^{K} (\alpha_j + (x)_j - 1) \log \pi_j + \lambda g(\vec{\pi})$

Constraint $g(\vec{\pi})=0$

$$\sum_j \pi_j = 1 \Rightarrow 1-\sum_j \pi_j = 0$$

So $g(\vec{\pi}) = 1-\sum_j \pi_j$

Lagrangian is

$$\mathcal{L}(\vec{\pi},\lambda) = \sum_{j=1}^{K} (\alpha_j + (x)_j - 1)\log \pi_j + \lambda\left(1-\sum_{j=1}^{K}\pi_j\right)$$

Partial derivative w.r.t. π_i

$$\frac{\partial \mathcal{L}}{\partial \pi_i} = \frac{\alpha_j + (x)_j - 1}{\pi_j} - \lambda = 0$$

$$\Rightarrow \pi_j^* = \frac{\alpha_j + (x)_j - 1}{\lambda}$$

Use centraint to determine λ:

$$g(\vec{\pi}^*) = 0 \Rightarrow \sum_{j=1}^{K} \pi_j^* = 1$$

$$\sum_{j=1}^{K} \frac{\alpha_j + (x_j) - 1}{\lambda} = 1$$

$$\Rightarrow \quad \boxed{\lambda = \sum_{j=1}^{K} (\alpha_j + (x)_j - 1)}$$

So I get

$$\pi_j^* = \frac{\alpha_j + (x)_j - 1}{\sum_{j=1}^{K} (\alpha_j + (x)_j - 1)}$$

BTW, $\log \pi_j$ is a concave function, so $\nabla_{\pi_j} \log \pi_j = 0$ is a maximum

So

$$\boxed{\hat{\pi}_j^{MAP} = \frac{\alpha_j + (x)_j - 1}{\sum_{j=1}^{K} (\alpha_j + (x)_j - 1)}}$$

In class, we found the MLE estimator for the multinomial to be

$$\hat{\pi}_j^{MLE} = \frac{(x)_j}{\sum_{j=1}^{K} (x)_j}$$

We see that as the amount of data increases (so as the number of counts in each categories grows) the MAP estimate will collapse to the MLE estimate.

I.e.

$$\alpha_j + (x)_j - 1 \longrightarrow (x)_j \quad \text{as } N \to \infty$$

$$(x)_j \gg 1$$

$$\Rightarrow \hat{\vec{\pi}}^{MAP} \xrightarrow[N\to\infty]{} \hat{\vec{\pi}}^{MLE}$$

Question 3

a) $X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, $p(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$

Let $X = \sum_{i=1}^{n} X_i \Rightarrow p(x_1, \dots, x_n | \lambda) \overset{iid}{=} \prod_{i=1}^{n} p(x_i|\lambda)$

$$\Rightarrow p(x_1, \dots, x_n | \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

The MLE is

$$\hat{\lambda}_{MLE} = \underset{\lambda}{\text{argmax}}\ p(x_1, \dots, x_n | \lambda)$$

Maximize log likelyhood:

$$\hat{\lambda}_{MLE} = \underset{\lambda}{\operatorname{argmax}} \log(p(x_1,\dots,x_n \mid \lambda))$$

$$= \underset{\lambda}{\operatorname{argmax}} \log\left(\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right)$$

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \left(x_i \log\lambda - \underline{\log x_i!} + \log e^{-\lambda}\right)$$

does not depend on λ so drop.

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} (x_i \log\lambda - \lambda)$$

Maximize w.r.t. λ.

$$\frac{\partial}{\partial\lambda}\left(\sum_{i=1}^{n} (x_i \log\lambda - \lambda)\right)\Bigg|_{\lambda=\lambda^*} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{x_i}{\lambda^*} - \underbrace{\sum_{i=1}^{n} 1}_{=n} = 0$$

$$\Rightarrow \frac{1}{\lambda^*} \sum_{i=1}^{n} x_i = n$$

$$\Rightarrow \boxed{\hat{\lambda}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$ sample mean!

Bias $\quad \text{Bias}(\hat{\lambda}) = \|\lambda - \mathbb{E}[\hat{\lambda}]\|_2$

$$\mathbb{E}[\hat{\lambda}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[x_i]$$ (expectation is linear)

use the fact that $E[X_i] = \lambda$ if $X_i \sim \text{Poisson}(\lambda)$

$$= \frac{1}{n}\sum_{i=1}^{n} \lambda = \frac{n\lambda}{n} = \lambda$$

So $\text{Bias}(\hat{\lambda}) = \|\lambda - \mathbb{E}[\hat{\lambda}]\|_2$

$$= \|\lambda - \lambda\|_2$$

$$\boxed{\text{Bias}(\hat{\lambda}) = 0}$$ unbiased

Variance $\quad \text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\text{Var}(x_i)}_{=\lambda \text{ for poisson}}$$ $(\text{Var}(nX) = n^2\text{Var}(X))$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \lambda$$

$$= \frac{n\lambda}{n^2}$$

$$\Rightarrow \boxed{\mathrm{Var}(\hat{\lambda}) = \frac{\lambda}{n}}$$

Is this consistent? Frequentist risk is

$$R(\lambda) = (\mathrm{Bias}\,\hat{\lambda})^2 + \mathrm{Var}(\hat{\lambda})$$

$$= 0 + \frac{\lambda}{n} = \frac{\lambda}{n}$$

So $\lim_{n\to\infty} R(\lambda) = \lim_{n\to\infty} \frac{\lambda}{n} \to 0$ Yes! Consistent!

b) $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Bernoulli}(p)$, $\hat{p} = \frac{1}{10}\sum_{i=1}^{10} X_i$ and $n > 10$. Pmf is $p(x|p) = p^x(1-p)^{1-x}$ with $\mathbb{E}[X] = p$ and $\mathrm{Var}(X) = p(1-p)$.

Bias

$$\mathrm{Bias}\,\hat{p} = \|p - \mathbb{E}[\hat{p}]\|_2$$

$$= \|p - \mathbb{E}[\tfrac{1}{10}\sum_{i=1}^{10} X_i]\|_2$$

$$= \|p - \frac{1}{10}\sum_{i=1}^{10} \underbrace{\mathbb{E}[X_i]}_{=p}\|_2$$

$$= \|p - \frac{1}{10}\sum_{i=1}^{10} p\|_2$$

$$= \| p - \frac{10p}{10} \|_2 = 0$$

$$\boxed{\text{Bias}\,\hat{p} = 0}$$ unbiased

<u>Variance</u>: $\text{Var}\,\hat{p} = \text{Var}\left(\frac{1}{10}\sum_{i=1}^{10} x_i\right)$

$$= \frac{1}{100}\sum_{i=1}^{10} \underbrace{\text{Var}\,x_i}_{= p(1-p)}$$

$$= \frac{1}{100}\sum_{i=1}^{10} p(1-p)$$

$$\boxed{\text{Var}(\hat{p}) = \frac{1}{10}p(1-p)}$$

Frequentist risk: $R(p) = (\text{Bias}\,\hat{p})^2 + \text{Var}\,\hat{p}$

$$= 0^2 + \frac{1}{10}p(1-p)$$

$$= \frac{1}{10}p(1-p)$$

So $\lim_{n\to\infty} R(p) = \lim_{n\to\infty} \frac{1}{10}p(1-p) = \frac{1}{10}p(1-p) \neq 0$

Not consistent!

c) $X_1, \dots, X_n \overset{iid}{\sim} \text{Uniform}(0, \theta)$,

I'm going to assume a continuous uniform distribution on $[0, \theta]$ w/ pdf

$$p(x_i|\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

Let $X = \sum_{i=1}^{n} X_i \Rightarrow p(x|\theta) \overset{iid}{=} \prod_{i=1}^{n} p(x_i|\theta)$

$$\Rightarrow p(x|\theta) = \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} \theta^{-n} & 0 \leq x \leq \theta \\ 0 & \text{elsewhere} \end{cases}$$

MLE is $\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, p(x|\theta)$

$$= \underset{\theta}{\operatorname{argmax}}\, \theta^{-n} \underbrace{\mathbb{1}_{\theta \geq \max(x_1, \dots, x_n)}}$$

this is the condition that $x_i \in [0, \theta]\ \forall i$

So need to maximize $L(\theta) = \theta^{-n} \mathbb{1}_{\theta \geq \max(x_1, \dots, x_n)}$

This is not differentiable on $\mathbb{R}$ ... because of the indicator function.

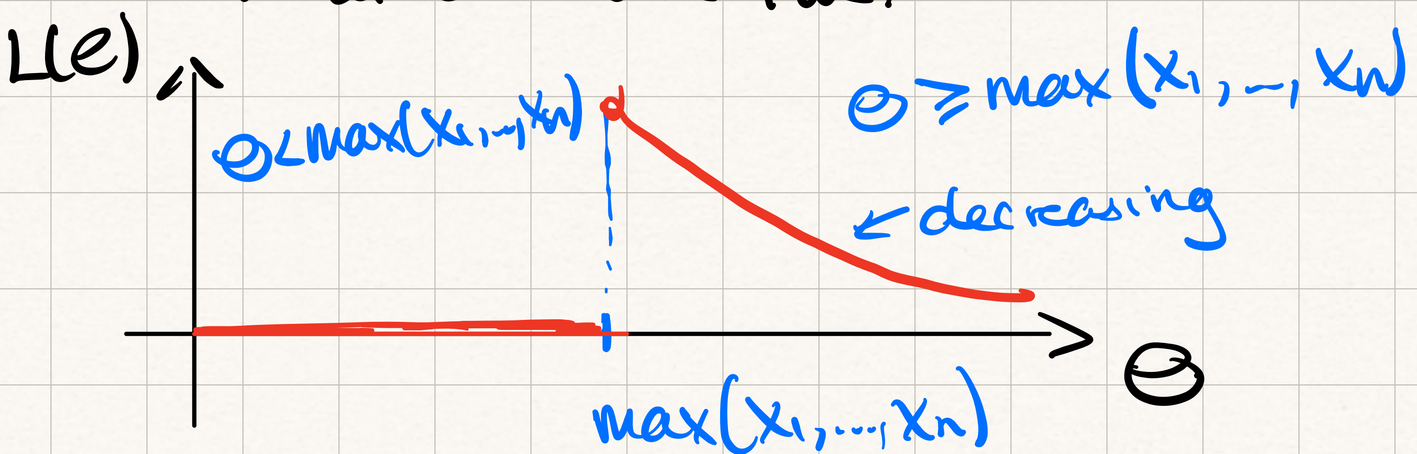
θ^{-n} is also a strictly decreasing function

... hard to maximize this.

but, $L(\theta) = \begin{cases} 0 & \text{if } \theta < \max(X_1, \dots, X_n) \\ \theta^{-n} & \text{if } \theta \geq \max(X_1, \dots, X_n) \end{cases}$

decreases.

which should look like this:



So I will say $\hat{\theta}^{MLE} = \max(X_1, \dots, X_n)$

<u>Bias</u> $\text{Bias}\,\hat{\theta} = \theta - \mathbb{E}[\hat{\theta}]$

Let $M = \max(X_1, \dots, X_n)$. The CDF of $\hat{\theta}_{MLE} = M$

is $P(M \leq c) = P(X_1 \leq c) \dots P(X_n \leq c)$

$$= \underbrace{\frac{c}{\theta} \cdots \frac{c}{\theta}}_{n \text{ times}} = \left(\frac{c}{\theta}\right)^n$$

Now, can get pdf of M by differentiating

$P(M \leq c)$ with respect to c:

$$p_M(c) = \frac{d}{dc}\left(\frac{c}{\theta}\right)^n = nc^{n-1}\,\theta^{-n} \quad (0 \leq c \leq \theta)$$

So $$\mathbb{E}[\hat{\theta}] = \int_0^\theta c\,p_M(c)\,dc$$

$$= \int_0^\theta c \cdot nc^{n-1}\,\theta^{-n}\,dc$$

$$= n\theta^{-n}\int_0^\theta c^n\,dc = n\theta^{-n}\,\frac{c^{n+1}}{n+1}\Bigg|_0^\theta$$

$$= \frac{n\theta}{n+1}$$

So $$\text{Bias}\,\hat{\theta} = \theta - \frac{n\theta}{n+1}$$

$$= \frac{(\cancel{n}+1)\theta - \cancel{n\theta}}{n+1}$$

$$\boxed{\text{Bias}\,\hat{\theta} = \frac{\theta}{n+1}}$$

Variance $\text{Var}(\hat{\theta}) = \mathbb{E}[\hat{\theta}^2] - \underline{(\mathbb{E}[\hat{\theta}])^2}$

$$= \left(\frac{n\theta}{n+1}\right)^2$$

and $E[\hat{\theta}^2] = \int_0^\theta e^2 P_M(c)\,dc$

$$= \int_0^\theta e^2 n e^{n-1} \theta^{-n}\,de$$

$$= n\theta^{-n} \int_0^\theta e^{n+1}\,de$$

$$= n\theta^{-n} \left. \frac{e^{n+2}}{n+2} \right|_0^\theta$$

$$= \frac{n\theta^{-n}\,\theta^{n+2}}{n+2}$$

$$= \frac{n\theta^2}{n+2}$$

So $\mathrm{Var}(\hat{e}) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2$

$$\boxed{\mathrm{Var}(\hat{e}) = \theta^2\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)}$$

Frequentist risk is:

$$R(e) = (\mathrm{Bias}\,\hat{\theta})^2 + \mathrm{Var}\,\hat{\theta}$$

$$\lim_{n\to\infty} R(e) = \lim_{n\to\infty} \underbrace{\frac{\theta^2}{(n+1)^2}}_{\to 0 \text{ as } n\to\infty} + \underbrace{\theta^2\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)}$$

$$= 0 + \theta^2 \lim_{n\to\infty} \underbrace{\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)}$$

as $n \to \infty$, $n+2 \to n$

$n+1 \to n$

so $\frac{n}{n+2} - \frac{n^2}{n+1} \to \frac{n}{n} - \frac{n^2}{n^2} = 1-1 = 0$

$$= 0 + 0 = 0.$$ Yes! Consistent.

d) $X_1, \ldots, X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, $p(X_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i-\mu)^2}$

Let $X = \sum_{i=1}^{n} X_i$

Now, $p(X; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$

$\boxed{\text{MLE for } \mu:}$

$$\hat{\mu}_{MLE} = \underset{\mu}{\operatorname{argmax}}\, p(x; \mu, \sigma^2)$$

Maximize loglikelihood:

$$\hat{\mu}_{MLE} = \underset{\mu}{\operatorname{argmax}} \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$= \underset{\mu}{\operatorname{argmax}} \sum_{i=1}^{n} \Big(\underbrace{-\log\sqrt{2\pi\sigma^2}}_{\text{does not depend on } \mu\text{, so drop it}} + \log e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \Big)$$

$$= \underset{\mu}{\operatorname{argmax}} \underbrace{\sum_{i=1}^{n} -\frac{1}{2\sigma^2}(x_i-\mu)^2}_{\text{maximize this w.r.t. to } \mu\text{.}}$$

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

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0.$$

$$\Rightarrow \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mu = n\mu$$

So $\boxed{\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}}$

$\boxed{\text{MLE for } \sigma^2}$

$$\hat{\sigma}^2_{MLE} = \underset{\sigma^2}{\operatorname{argmax}} \log p(x;\mu,\sigma^2)$$

$$= \underset{\sigma^2}{\operatorname{argmax}} \sum_{i=1}^{n} \left(\underbrace{-\log\sqrt{2\pi}}_{\text{does not depend on } \sigma^2 \text{ so drop.}} - \log\sqrt{\sigma^2} + \log e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right)$$

$$= \underset{\sigma^2}{\operatorname{argmax}} \underbrace{\sum_{i=1}^{n} \left(-\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x_i-\mu)^2 \right)}_{\text{maximize this with respect to } \sigma^2.}$$

$$\frac{\partial}{\partial\sigma^2}\left(\sum_{i=1}^{n} -\cancel{\frac{1}{2}}\log\sigma^2 - \frac{1}{\cancel{2}\sigma^2}(x_i-\mu)^2 \right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\partial}{\partial\sigma^2}\left(\log\sigma^2 + \frac{1}{\sigma^2}(x_i-\mu)^2 \right) = 0.$$

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

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

So,

$$\boxed{\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = \frac{1}{n}\sum_{n=1}^{n}(x_i-\overline{x})^2}$$

Now, <u>bias</u> for $\hat{\sigma}^2$:

$$\text{Bias } \hat{\sigma}^2 = \sigma^2 - \mathbb{E}[\hat{\sigma}^2]$$

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

$$= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}x_i^2 - 2\hat{\mu}\underbrace{\sum_{i=1}^{n}x_i}_{\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i} + \underbrace{\sum_{i=1}^{n}\hat{\mu}^2}_{=n\hat{\mu}^2}\right]$$

$$\Rightarrow \sum_{i=1}^{n}x_i = n\hat{\mu}$$

$$= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}x_i^2 - 2n\hat{\mu}^2 + n\hat{\mu}^2\right]$$

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

$$\mathrm{Var}(x_i) = \mathbb{E}[x_i^2] - \mathbb{E}[x_i]^2$$

$$\Rightarrow \mathbb{E}[x_i^2] = \mathrm{Var}(x_i) + \mathbb{E}[x_i]^2 = \sigma^2 + \mu^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\mathrm{Var}(\hat{\mu}) + \underbrace{\mathbb{E}[\hat{\mu}]^2}_{=\mu^2}\right)\right)$$

$$= \frac{1}{n}\left(n\sigma^2 + \cancel{n\mu^2} - n\left(Var\left(\frac{1}{n}\sum_{i=1}^n x_i\right) - \cancel{\mu^2}\right)\right.$$

$$= \frac{1}{n}\left(n\sigma^2 - \frac{n}{n^2}\sum_{i=1}^n Var\, x_i\right)$$

$$= \frac{1}{n}\left(n\sigma^2 - \frac{1}{n}\sum_{i=1}^n \sigma^2\right)$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \mathbb{E}[\hat{\sigma}^2]$$

So $\text{Bias}\,\hat{\sigma}^2 = \sigma^2 - \left(\sigma^2 - \frac{\sigma^2}{n}\right)$

$$\Rightarrow \boxed{\text{Bias}\,\hat{\sigma}^2 = \frac{\sigma^2}{n}}$$

<u>Variance</u>

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

I want to use the trick that $Var\, \chi^2_{n-1} = 2(n-1)$

and $\chi^2_{n-1} = \frac{1}{\sigma^2}\sum_{n=1}^{1} (x_i - \hat{\mu})^2$

So $Var(\hat{\sigma}^2) = Var\left(\frac{\sigma^2}{n}\chi^2_{n-1}\right)$

$$= \frac{\sigma^4}{n^2} Var(\chi^2_{n-1})$$

$$\Rightarrow \boxed{Var(\hat{\sigma}^2) = \frac{2\sigma^4(n-1)}{n^2}}$$

check the consistency:

$$R(\sigma^2) = (Bias\,\hat{\sigma}^2)^2 + Var(\hat{\sigma}^2)$$

$$= \frac{\sigma^4}{n^4} + \frac{2\sigma^4(n-1)}{n^2}$$

$$\text{So } \lim_{n\to\infty} R(\sigma^2) = \lim_{n\to\infty}\left(\cancel{\frac{\sigma^4}{n^4}}^{\to 0} + \underbrace{\frac{2\sigma^4(n-1)}{n^2}}\right)$$

$$\frac{n-1}{n^2} = \frac{1}{n} - \frac{1}{n^2} \longrightarrow 0 \text{ as } n\to\infty$$

$$\lim_{n\to\infty} R(\sigma^2) = 0$$

Yes! Consistent!