Exercise 09

Thursday, December 14, 2023 5:14 PM

Exercise 35 (4 points)

(a) (3 points) Let $\hat{\theta}$ be an estimator of θ. Show that $MSE(\hat{\theta}) = Var(\hat{\theta}) + B(\theta)^2$.

(b) (1 point) Let $X_1, \ldots, X_n$ be i.i.d. random variables with expectation μ and variance σ^2. Compute the mean squared error of $\frac{1}{n}\Sigma_{i=1}^n X_i$ as an estimator of μ in terms of σ^2.

(a) $Var(\hat{\theta})$ $\qquad E(\theta) = \theta$

$= E(\hat{\theta}^2) - (E(\hat{\theta}))^2$

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

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

$MSE(\hat{\theta}) = E((\hat{\theta}-\theta)^2)$

$= E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2)$

$\overset{\text{linearity}}{=} E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2$

$$Var(\hat{\theta}) + B(\hat{\theta})^2 = \underbrace{E(\hat{\theta}^2) - (E(\hat{\theta}))^2}_{Var} + \underbrace{(E(\hat{\theta}))^2 - 2E(\hat{\theta})\cdot\theta + \theta^2}_{B}$$

$= E(\hat{\theta}^2) - 2E(\hat{\theta})\cdot\theta + \theta^2$ $\qquad \theta$ is constant $\Rightarrow \theta = E\theta,\ \theta^2 = E(\theta^2)$

$= E(\hat{\theta}^2) - 2E(\hat{\theta})\cdot E(\theta) + E(\theta^2)$

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

(b)

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i \qquad E(\hat{\mu}) = E(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n EX_i = \frac{1}{n}\sum_{i=1}^n \mu = \mu$

so $B(\hat{\mu}) = E(\hat{\mu}) - \mu = 0$

$\sigma^2 = E((X_i-\mu)^2)$

$$MSE(\hat{\mu}) = E((\hat{\mu}-\mu)^2) = Var(\hat{\mu})$$

$= Cov(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{j=1}^n X_j)$

$= \underline{\frac{1}{n^2}(\sum_{i=1}^n\sum_{j=1}^n} Cov(X_i, X_j))$

$Cov(X_i, X_j) = 0$ if $i \neq j$

$= \frac{1}{n^2}\cdot n Var(X_i)$

$= \frac{1}{n}\sigma^2$

$MSE(\hat{\mu}) = E((\hat{\mu}-\mu)^2) = Var(\hat{\mu}) = E(\hat{\mu}^2) - (E(\hat{\mu}))^2$

$= E(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n X_iX_j) - \mu^2$

$= E(\frac{1}{n^2}(\sum_{i=1}^n\sum_{j=1}^n (X_iX_j - \mu^2)))$

$= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n (E(X_iX_j) - \mu^2)$

$= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n (E(X_iX_j) - \mu E(X_i) - \mu E(X_j) + \mu^2)$

$= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n (E[(X_i-\mu)(X_j-\mu)])$

$\overset{\text{i.i.d}}{=} \frac{1}{n^2}\sum_{i=1}^n E((X_i-\mu)^2) = \frac{1}{n}\sigma^2$

$E[(X_i-\mu)(X_j-\mu)] = 0$ if $i \neq j$

Exercise 36 (10 points)

Let $X_1, \ldots, X_k$ be i.i.d. $\mathbb{R}$-valued random variables with a normal distribution $\sim \mathcal{N}(\mu, \sigma^2)$. We know that their joint distribution has density function $\Pi_{i=1}^k \mathcal{N}(x_i|\mu, \sigma^2)$.
The maximum likelihood estimator for σ^2 is the random variable $\widehat{\sigma^2}_{ML} : \omega \mapsto \operatorname{argmax}_{\sigma^2} \Pi_{i=1}^k \mathcal{N}(X_i(\omega)|\mu, \sigma^2)$.
That is, for a sample ω giving rise to values $x_1 = X_1(\omega), \ldots, x_k = X_k(\omega)$ the value $\widehat{\sigma^2}_{ML}(\omega)$ is the σ^2 for which $\Pi_{i=1}^k \mathcal{N}(x_i|\mu, \sigma^2)$ assumes its maximum.

Show that $\widehat{\sigma^2}_{ML} = \frac{1}{k}\Sigma_{i=1}^k (X_i - \mu)^2$.

$$\hat{\sigma}^2_{ML} = \operatorname*{argmax}_{\sigma^2} \prod_{i=1}^k \mathcal{N}(X_i(\omega)|\mu, \sigma^2)$$

$$\prod_{i=1}^k \mathcal{N}(X_i(\omega)|\mu, \sigma^2) = \prod_{i=1}^k \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$$

Since $\log(x)$ is monotonically increasing.

$\operatorname*{argmax}_{\sigma^2} \prod_{i=1}^k \mathcal{N}(x_i(\omega)|\mu, \sigma^2)$

$= \operatorname*{argmax}_{\sigma^2} \log\left(\prod_{i=1}^k \mathcal{N}(X_i(\omega)|\mu, \sigma^2)\right)$

$= \operatorname*{argmax}_{\sigma^2} \left(-\sum_{i=1}^k \left(\frac{(X_i-\mu)^2}{2\sigma^2}\right) - \underbrace{\sum_{i=1}^k \log(\sqrt{2\pi})}_{\text{const}} - \sum_{i=1}^k \log(\sigma)\right)$

Let $a = \sigma^2$ $\quad = \operatorname*{argmin}_{a} \underbrace{\left(\sum_{i=1}^k \left(\frac{(X_i-\mu)^2}{2a}\right) + \sum_{i=1}^k (\log(\sqrt{a}))\right)}$

Q: Is ℓ convex ??

Doesn't seem so

$$\text{Let } a=\sigma^2 \quad = \underset{a}{\operatorname{argmin}} \underbrace{\left(\sum_{i=1}^{k} \left(\frac{(x_i-\mu)^2}{2a} \right) + \sum_{i=1}^{k} \left(\log(\sqrt{a}) \right) \right)}_{:= l}$$

Q: Is l convex ??

Doesn't seem so.

To find the minimum,

$$\frac{\partial l}{\partial a} = -\sum_{i=1}^{k} \frac{(x_i-\mu)^2}{2a^2} + \sum_{i=1}^{k} \frac{1}{\sqrt{a}} \cdot \frac{1}{2\sqrt{a}} = 0 \quad , \quad a>0$$

$$= \frac{k}{2a} - \sum_{i=1}^{k} \frac{(x_i-\mu)^2}{2a^2} = 0$$

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

Exercise 38 (10 points)

In a pond there are yellow, silver and black fish. We catch $\boxed{n}$ fish (gently, by hand, nobody is hurt) and throw them back each time. Let X, Y, Z be the total numbers of yellow/silver/black fish that we caught. If p_X, p_Y, p_Z are the probabilities of catching a yellow/silver/black fish, then the joint distribution of X, Y, Z is given by $P(X = x, Y = y, Z = z) = \frac{n!}{x!y!z!} p_X^x p_Y^y p_Z^z$ (a multinomial distribution).

In this exercise you should calculate the maximum likelihood estimates for p_X, p_Y, p_Z, if we have caught x yellow fish, y silver fish and z black fish.

This means that you have to solve the constrained optimization problem of maximizing the likelihood function $l = l(x, y, z, p_X, p_Y, p_Z)$ with respect to p_X, p_Y, p_Z, and fixed given parameters x, y, z. You can do this however you want, but here is a sequence of hints that might make it easier:

(a) Calculate the partial derivatives of the likelihood function $l = l(x, y, z, p_X, p_Y, p_Z)$ with respect to p_X, p_Y, p_Z.

(b) Express each of these partial derivatives as multiples of the likelihood function; $\frac{\partial}{\partial p_i} l = c_i \cdot l$ for some c_i, $i \in \{X, Y, Z\}$

(c) Write down the equations with Lagrange multiplier λ for the constraint $p_X + p_Y + p_Z = 1$

(d) Show that $\lambda = n \cdot l$.

(e) Substitute this into the Lagrange multiplier equations to compute your estimates for p_X, p_Y, p_Z.

[Remark: If you follow these instructions: 2 points for each step, otherwise 10 altogether]

(a) $l = \frac{n!}{x!y!z!} p_X^x p_Y^y p_Z^z$

(b)

$$\frac{\partial l}{\partial p_X} = \frac{n!}{x!y!z!} \cdot x p_X^{x-1} p_Y^y p_Z^z = \frac{x}{p_X} l$$

Similarly $\frac{\partial l}{\partial p_Y} = \frac{y}{p_Y} l \quad , \quad \frac{\partial l}{\partial p_Z} = \frac{z}{p_Z} l$

(c) The Lagrangian can be written as

$$L(\lambda, p_X, p_Y, p_Z) = \frac{n!}{x!y!z!} p_X^x p_Y^y p_Z^z + \lambda(1-p_X-p_Y-p_Z)$$

(d) Gradient condition:

$$\begin{cases} \frac{\partial L}{\partial p_X} = \frac{x}{p_X} l - \lambda = 0 \\ \frac{\partial L}{\partial p_Y} = \frac{y}{p_Y} l - \lambda = 0 \\ \frac{\partial L}{\partial p_Z} = \frac{z}{p_Z} l - \lambda = 0 \end{cases} \Rightarrow \begin{cases} x l = p_X \lambda \\ y l = p_Y \lambda \\ z l = p_Z \lambda \end{cases} \Rightarrow (p_X+p_Y+p_Z)\lambda = \lambda$$

$$= (x+y+z) l$$

$$= n l$$

$$\Rightarrow \begin{cases} p_X = \frac{x}{n} \\ p_Y = \frac{y}{n} \\ p_Z = \frac{z}{n} \end{cases}$$

Convexity:

$$H_l(p_X, p_Y, p_Z) = \begin{pmatrix} \frac{x(x-1)}{p_X^2} & \frac{xy}{p_X p_Y} & \frac{xz}{p_X p_Z} \\ & \frac{y(y-1)}{p_Y^2} & \frac{yz}{p_Y p_Z} \\ \text{sym} & & \frac{z(z-1)}{p_Z^2} \end{pmatrix} \cdot l$$

All elements in $H_l \geq 0$, since $x, y, z \geq 0$, $p_X, p_Y, p_Z \geq 0$.

$\Rightarrow$ For $\forall p_X, p_Y, p_Z \geq 0$,

$p^T H_l p \geq 0 \Rightarrow l$ is convex. Since $p_X+p_Y+p_Z-1=0$ is affine

$\Rightarrow L$ is a convex optimization problem.

Exercise 39 (7 points)

The provided notebook shows you an estimator for the covariance matrix of a multvariate distribution that is given as an algorithm, instead of a formula. Look at the notebook and follow the instructions.