

Midterm Solution

This is a 2 hour exam. The exam is closed book and closed notes except for one sheet (two sides) of notes. Begin each problem on a new page. Justify all your computations, partial credit will be given for answers that are well reasoned even if the final result is wrong.

1. (10 points) *Variance estimation.* Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i^4] < \infty$ and $\mathbb{E}[X_i] = 0$.

- a. Propose an unbiased estimator \hat{V}_n for $\text{Var}X_i$

Solution:

Since $\mathbb{E}[X_n] = 0$, then $\text{Var}X_n = \mathbb{E}[X_n^2]$, and we can use $\hat{V}_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ as the estimator. The estimator is unbiased if its expected value is equal to $\text{Var}X_n$, let's show that:

$$\mathbb{E}[\hat{V}_n] = \frac{1}{n} \mathbb{E} \left[\sum_{i=1}^n X_i^2 \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] = \text{Var}X_n$$

- b. Does your estimator converge to $\text{Var}X_i$ as $n \rightarrow \infty$? If yes, the in what sense does it converge? (It is enough to state just one)

Solution:

By the Law of Large Numbers as X_n^2 are i.i.d. and have a finite 4th moment, the sample mean converges to the expected value in probability.

- c. Provide an approximate estimate of large deviations probability, i.e.

$$\mathbb{P} \left\{ \left| \hat{V}_n - \text{Var}X_i \right| < \epsilon \right\} \text{ in terms of } Q\text{-function} \left(Q(x) = \int_x^{\infty} e^{-\frac{u^2}{2}} du \right)$$

Solution:

By the Central Limit theorem $Y_n = \sqrt{\frac{n}{\text{Var}(X_n^2)}} \left(\hat{V}_n - \text{Var}X_n \right)$ converges to

$Z_n \sim \mathcal{N}(0, 1)$ in distribution, so for large n we can approximate the cdf with a Gaussian cdf and get that:

$$\mathbb{P} \left\{ \left| \hat{V}_n - \text{Var}X_i \right| < \epsilon \right\} \approx \mathbb{P} \left\{ \sqrt{\frac{\text{Var}(X_n^2)}{n}} |Z| < \epsilon \right\} = 1 - 2Q \left(\sqrt{\frac{n}{\text{Var}(X_n^2)}} \epsilon \right)$$

2. (10 points) *Covariance matrices and Gaussians.*

- a. Which of the following can be a covariance matrix of a random vector of the corresponding size?

i. $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ **Solutions:** yes, for example if $X \sim \mathcal{N}(0, 1)$, $\mathbf{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} X$

ii. $\begin{bmatrix} 5 & 3 & 3 \\ 3 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$ **Solutions:** No, its enough to see that $\Sigma_{2,3} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \neq 0$, as

$\det \Sigma = -3$, but it is a covariance matrix for $\begin{bmatrix} X_2 \\ X_3 \end{bmatrix}$

- iii. $\begin{bmatrix} 3 & 2 \\ 3 & 4 \end{bmatrix}$ **Solutions:** No, the matrix is not symmetric
- b. X is a JGRV (Jointly Gaussian Random Variable), $X \in \mathbb{R}^n$. For which of the following transformations $Z = f(X)$ is also necessarily JGRV?
- i. $Z = MX$, where M is an $n \times n$ matrix, symmetric, not full-rank
Solutions: Yes, any linear transformation of a JGRV is a JGRV
- ii. $Z = MX$, where M is an $n \times n$ matrix, symmetric, full-rank, but not positive semi-definite
Solutions: Yes, any linear transformation of a JGRV is a JGRV
- iii. $Z = XX^T$
Solutions: No, for example $\mathbf{X} \sim \mathcal{N}(0, 1)$ than it is easy to see that the distribution is not symmetric and therefore cannot be a Gaussian distribution.
- iv. $Z = Y |X_1 - \mathbb{E}[X_1]|$, where $\mathbb{P}\{Y = 1\} = \mathbb{P}\{Y = -1\} = 1/2$
Solutions: Yes,

$$\begin{aligned} X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) &\Rightarrow X'_1 = X_1 - \mathbb{E}[X_1] \sim \mathcal{N}(0, \sigma_1^2) \\ &\Rightarrow f_{X'_1}(z) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{z^2}{2\sigma_1^2}} = f_{X'_1}(-z). \end{aligned}$$

So, the distribution of X'_1 is symmetric. To prove that Z is Gaussian we can find its cdf:

$$\begin{aligned} \mathbb{P}\{Z \leq z\} &= \mathbb{P}\{YX'_1 \leq z\} = \mathbb{P}\{YX'_1 \leq z | Y = 1\} \mathbb{P}\{Y = 1\} \\ &+ \mathbb{P}\{YX'_1 \leq z | Y = -1\} \mathbb{P}\{Y = -1\} \\ &= \frac{1}{2} (\mathbb{P}\{X'_1 \leq z\} + \mathbb{P}\{X'_1 \geq -z\}) = \mathbb{P}\{X'_1 \leq z\}, \end{aligned}$$

where the last equality follows from the symmetry of the distribution of X'_1 and $Z \sim \mathcal{N}(0, \sigma_1^2)$

- v. $Z = [Z_1 \dots Z_n]^T$, where $Z_i = Y |X_i - \mathbb{E}[X_i]|$,
 $\mathbb{P}\{Y = 1\} = \mathbb{P}\{Y = -1\} = 1/2$
Solutions: No, because the all X_i must have the same signs. Note that $\mathbb{E}[X_i] = 0 \forall i$, if the distribution was JG, then $f_Z(z_1, z_2, \dots, z_n) = f_Z(-z_1, z_2, \dots, z_n)$ but if $\forall i z_i > 0$, the LHS is non-zero, while the RHS = 0
- vi. $Z = [Z_1 \dots Z_n]^T$, where $Z_i = Y_i |X_i - \mathbb{E}[X_i]|$,
 $\mathbb{P}\{Y_i = 1\} = \mathbb{P}\{Y_i = -1\} = 1/2$, X_i are independent for $i = 1, \dots, n$
Solutions: Yes, X_i are independent and, therefore, Z_i are independent as well, using the same proof as for the one dimensional case we get that each Z_i is Gaussian and their independence leads to vector Z being a JGRV.
- c. Let $\Sigma \in \mathbb{R}^{n \times n}$, such that its elements are $\sigma_{ij} = \min\{i, j\}$. Is $\Sigma \succeq 0$?
Solutions: Yes, we know that the covariance matrix is always positive semidefinite. Therefore, if we show that this matrix is a covariance matrix of some random vector than it has to be positive semidefinite. Choose

$X_i \sim \mathcal{N}(0, 1)$, X_i are i.i.d. If we choose

$$Y_n = \sum_{i=1}^n X_i, \text{ then}$$

$$\text{Cov}(Y_i, Y_j) = \sum_{k=1}^{\min\{i,j\}} \text{Var}(X_k) = \min\{i, j\}$$

3. (10 points) *Detection.* Alice and Bob are good students, but they don't like to attend lectures, especially in summer. One of the instructors, however, wants students to come to his lectures and he decided not to give credit to students who attended less than 80% of all the lectures. But the lecturer sometimes forgets to check attendance, especially when it's sunny outside and he's in a good mood. In this problem you will help Alice and Bob find an optimal strategy for coming to the lectures based on the weather conditions they see outside. Students have noticed that on a sunny day the probability of attendance check is 0.4, while on a cloudy day it's 0.6. The total probability of attendance being checked is 0.5.
- a. Bob wants to get a credit, so he needs a rule that would ensure 0.2 probability of being absent at a class, when the attendance is checked. He also wants to attend as few lectures, when the attendance is not checked, as possible.

Solution:

First, let's write the problem in terms of indicator functions:

$X = 1\{\text{attendance is checked}\}$ and $Y = 1\{\text{weather is nice}\}$. Then we know that $\mathbb{P}\{X = 1\} = \mathbb{P}\{X = 0\} = 0.5$ and $p_{X|Y}(1|1) = 0.4 = p_{X|Y}(0|0)$, $p_{X|Y}(0|1) = 0.6 = p_{X|Y}(1|0)$. The optimal rule for Bob would be to use the Neyman-Pearson rule:

$$\Lambda(y) = \frac{p_{Y|X}(y|1)}{p_{Y|X}(y|0)} \leq \eta,$$

In our case of equally probable X :

$$\Lambda(y) = \frac{p_{X|Y}(1|y)}{p_{X|Y}(0|y)}$$

and $\hat{X} = 1\{\Lambda(y) > \eta\}$ $\Lambda(1) = 2/3$, $\Lambda(0) = 1.5$, so we should consider three intervals for η :

- $\eta < 2/3$: $\forall y \Lambda(y) > \eta$, so this rule implies $\hat{X}(y) = 1$ and $\mathbb{P}\{\hat{X} = 0|X = 1\} = 0$.
- $\eta \in (2/3, 1.5)$: $\Lambda(1) < \eta$, $\Lambda(0) > \eta$ so this rule implies $\hat{X}(1) = 0$, $\hat{X}(0) = 1$ and $\mathbb{P}\{\hat{X} = 0|X = 1\} = p_{Y|X}(1|1) = 0.4$.
- $\eta > 1.5$: $\forall y \Lambda(y) < \eta$, so this rule implies $\hat{X}(y) = 0$ and $\mathbb{P}\{\hat{X} = 0|X = 1\} = 1$.

As we can see, no rule achieves the desired value of $\mathbb{P}\{\hat{X} = 0|X = 1\} = 0.2$, so we have to use randomization to combine the first two rules described above with equal weights to get an optimal rule. This is done by setting $\eta = 2/3$ and tossing a coin if

$\Lambda(y) = \eta$. Basically, Bob should go to the lecture every cloudy day and toss a coin every sunny day: if it turns heads – he should stay home and if it lands tails, he should go to the lecture.

- b. Alice decides to come to as many as possible checked lectures and attend as few as possible lectures, when the attendance is not checked. This corresponds to MAP (maximum a posteriori probability) decision rule. Help her find the rule and calculate the probability of not attending a checked lecture. Whose rule is more relevant to the problem?

Solution:

MAP rule is to compare a posteriori probabilities ratio to 1:

$$\frac{p_{X|Y}(1|y)}{p_{X|Y}(0|y)} \leq 1$$

So, Alice's rule would be the same as Neyman-Pearson's rule for $\eta = 1$:

$\hat{X}(1) = 0, \hat{X}(0) = 1$. With this decision rule the probability of not attending a checked lecture is $\mathbb{P}\{\hat{X} = 0|X = 1\} = p_{Y|X}(1|1) = 0.4 > 0.2$, so Alice won't get a credit with high probability.

4. (10 points) *Estimation*. Suppose the random variable Y is a noisy measurement of the angular position X of an antenna, so $Y = X + Z$, where Z denotes the additive noise. Assume the noise is independent of the angular position, i.e., X and Z are independent random variables, with X uniformly distributed in the interval $[-1, 1]$ and Z uniformly distributed in the interval $[-2, 2]$.

Given that $Y = y$, we would like to determine the MMSE (Minimum Mean Squared Error) estimate $\hat{X}(y)$, the resulting mean square error: $\mathbb{E}\left[\left(\hat{X}(Y) - X\right)^2 \mid Y = y\right]$, and the overall mean square error averaged over all possible values y that the random variable Y can take.

Solutions:

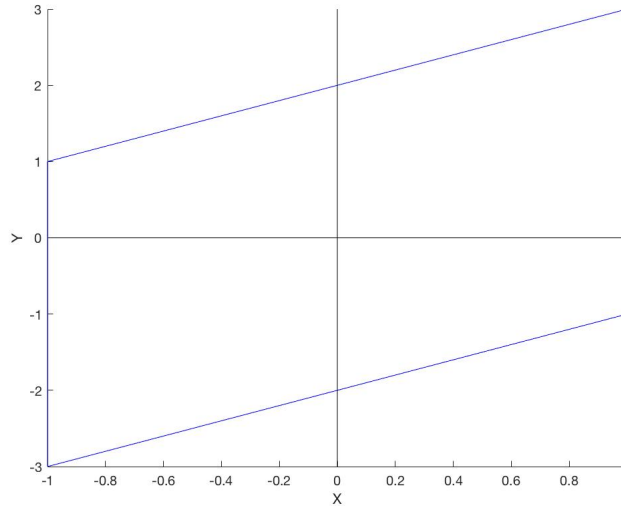
MMSE estimator: $\hat{X}(y) = E[X|Y = y]$

$$f_X(x) = \frac{1}{2} \quad \forall x \in [-1, 1]$$

$$f_{Y|X}(y|x) = \frac{1}{4} \quad \forall x \in [-1, 1], y \in [x - 2, x + 2]$$

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x) = \frac{1}{8} \quad \forall x \in [-1, 1] y \in [x - 2, x + 2]$$

The distribution is constant inside the parallelogram shown below.



As it was said in the lectures, conditioning on y is, taking the joint pdf $f_{X,Y}(x,y)$, fixing y and normalizing it over x . From the plot it's easy so see that:

$$\hat{X}(y) = E[X|Y = y] = \begin{cases} \frac{y+1}{2} & \text{if } y \in [-3, -1] \\ 0 & \text{if } y \in [-1, 1] \\ \frac{y-1}{2} & \text{if } y \in [1, 3] \end{cases}$$

Explanation for $y \in [-3, -1]$: because the distribution is uniform the expected value of $X|Y = y$ is the middle point between $x = -1$ and $x = y + 2$, the proof for the 2 other regions is the same.

MSE calculation:

$E[(X - \hat{X})^2|Y = y]$ is the variance of the distribution $f_{X|Y}(x|y)$.

$$E[(X - \hat{X})^2|Y = y] = \begin{cases} \frac{(3+y)^2}{12} & \text{if } y \in [-3, -1] \\ \frac{1}{3} & \text{if } y \in [-1, 1] \\ \frac{(3-y)^2}{12} & \text{if } y \in [1, 3] \end{cases}$$

Since $E[(X - \hat{X})^2] = E_Y[E[(X - \hat{X})^2|Y = y]]$ all we are left to do is to calculate $f_Y(y)$ and integrate.

Again, looking at the plot it's easy to see that

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{3+y} & \text{if } y \in [-3, -1] \text{ } x \in [-3, y+2] \\ \frac{1}{2} & \text{if } y \in [-1, 1] \text{ } x \in [-1, 1] \\ \frac{1}{3-y} & \text{if } y \in [1, 3] \text{ } x \in [y-2, 3] \end{cases}$$

$$f_Y(y) = \frac{f_{Y,X}(y,x)}{f_{X|Y}(x|y)} = \begin{cases} \frac{3+y}{8} & \text{if } y \in [-3, -1] \\ \frac{1}{4} & \text{if } y \in [-1, 1] \\ \frac{3-y}{8} & \text{if } y \in [1, 3] \end{cases}$$

$$\begin{aligned} E[(X - \hat{X})^2] &= \int_{-\infty}^{\infty} E[(X - \hat{X})^2 | Y = y] f_Y(y) dy \\ &= \int_{-3}^{-1} \frac{(3+y)^2 (3+y)}{12 \cdot 8} dy + \int_{-1}^1 \frac{1}{3} \frac{1}{4} dy + \int_1^3 \frac{(3-y)^2 (3-y)}{12 \cdot 8} dy = \frac{1}{4} \end{aligned}$$