Spring 2019 CX4240
Homework 1 Answer

Deadline: January 28, Monday, 11:59 pm

1 Expectation, Co-variance and Independence [25pts]

Suppose X, Y and Z are three different random variables.
Let X obeys Bernouli Distribution. The probability disbribution function is

$$p(x) = \begin{cases} 0.5 & x = 1 \\ 0.5 & x = -1. \end{cases}$$

Let Y obeys the standard Normal (Gaussian) distribution, which can be written as $Y \sim N(0, 1)$. X and Y are independent. Meanwhile, let $Z = XY$.

(a) What is the Expectation (mean value) of X? [3pts]

$$E[X] = \sum_x xp(x) = 0.5 \times 1 + 0.5 \times (-1) = 0$$

(b) Are Y and Z independent? (Just clarify, do not need to prove) [2pts]

No. Since $Z = XY$, if Y changes, Z will change correspondingly.

(c) Show that Z is also a standard Normal (Gaussian) distribution, which means $Z \sim N(0, 1)$. [10pts] (**hint**: Use the denominator part of the equation 1 on "other cases" at the page 16 of the slide 3.)

$$\begin{aligned} P(Z) &= P(X = 1)P(Z|X = 1) + P(X = -1)P(Z|X = -1) \\ &= P(X = 1)P(XY|X = 1) + P(X = -1)P(XY|X = -1) \\ &= P(X = 1)P(Y) + P(X = -1)P(-Y) \\ &= 0.5P(Y) + 0.5P(-Y) \end{aligned}$$

Due to $Y \sim N(0, 1)$, the PDF curve for standard normal distribution is symmetry based on vertical axis, so P(Y) = P(-Y).Thus,

$$P(Z) = 0.5P(Y) + 0.5P(Y) = P(Y).$$

Since $Y \sim N(0, 1)$, so $Z \sim N(0, 1)$.

(Derive CDF is also correct.)

(d) Are Y and Z uncorrelated(which means $Cov(Y, Z) = 0$)? (need to prove) [10pts]
(**hint[1]**: You may need this Theorem about Independence and Functions of Random Variables. Let

X and Y be independent random variables. Then, $U = g(X)$ and $V = h(Y)$ are also independent for any function g and h.)

$$\begin{aligned} Cov(Y,Z) &= E[YZ] - E[Y]E[Z] \\ &= E[YZ] - 0 \\ &= E[XY^2] \\ &= P(X=1)E[XY^2|X=1] + P(X=-1)E[XY^2|X=-1] \\ &= P(X=1)E[Y^2] + P(X=-1)E[-Y^2] \\ &= 0.5E[Y^2] + 0.5E[-Y^2] \end{aligned}$$

Due to the linearity of Expectation, where $E[aX] = aE[X]$, so $E[Y^2] = -E[Y^2]$, Then

$$Cov(Y,Z) = 0.5E[Y^2] - 0.5E[Y^2] = 0.$$

Method2:

Based on the hint, Y^2 is a function of Y, since X and Y be independent random variables, so X and Y^2 are also independent. Thus,

$$\begin{aligned} Cov(Y,Z) &= E[XY^2] \\ &= E[X]E[Y^2] \\ &= 0 \times E[Y^2] \\ &= 0 \end{aligned}$$

NOTE: From this problem, we find that $Independence \Rightarrow Cov = 0$, but $Cov = 0 \nRightarrow Independence$.

2 Maximum Likelihood [25pts]

2.1 Discrete Example [17pts]

Suppose you are playing two unfair coins. The probability of tossing a tail is θ for coin 1, and 2θ for coin 2. You toss each coin for several times, and you get the following results:

Coin No.	Result
1	head
2	head
1	tail
1	head
2	tail
2	tail

(a) What is the probability of tossing a head for coin 1 and for coin 2 [3pts]?

$$\begin{aligned} coin1 &: 1-\theta \\ coin2 &: 1-2\theta \end{aligned}$$

(b) What is the likelihood of the data given θ [7pts]?

$$\begin{aligned} L(\theta) &= (1-\theta)(1-2\theta)\theta(1-\theta)(2\theta)(2\theta) \\ &= (1-\theta)^2\theta(1-2\theta)(2\theta)^2 \end{aligned}$$

(c) What is maximum likelihood estimation for θ [7pts]?

The log-likelihood is

$$\begin{aligned} LL(\theta) &= log[(1-\theta)^2\theta(1-2\theta)(2\theta)^2] \\ &= 2log(1-\theta) + log\theta + log(1-2\theta) + 2log(2\theta) \end{aligned}$$

Take the derivative,

$$\begin{aligned} \frac{dLL(\theta)}{d\theta} &= \frac{d(2log(1-\theta) + log\theta + log(1-2\theta) + 2log(2\theta))}{d\theta} \\ &= (-2) \times \frac{1}{1-\theta} + \frac{1}{\theta} + (-2) \times \frac{1}{1-2\theta} + 2 \times 2 \times \frac{1}{2\theta} \\ &= \frac{12\theta^2 - 13\theta + 3}{(1-\theta)(1-2\theta)\theta} \end{aligned}$$

We want to maximum the log-likelihood, when the derivative is 0, then we will get the maximum the log-likelihood.

$\frac{dLL(\theta)}{d\theta} = 0,$
$12\theta^2 - 13\theta + 3 = 0,$
$\theta_1 = \frac{1}{3}, \theta_2 = \frac{3}{4}.$

Since the probablity for tossing coin1 or coin2 for both head and tail are suppoed to between 0 and 1, so $\theta = \frac{1}{3}$.

2.2 Continues Example [8pts]

A uniform distribution in the range of $[0, \theta]$ is given by

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & otherwise. \end{cases}$$

What is maximum likelihood estimation for θ?
(**hint:** Think of two cases, where $\theta < max(x_1, x_2, ..., x_n)$ and $\theta \geq max(x_1, x_2, ..., x_n)$.)

The likelihood function

$$\varphi(\theta) = \prod_{i=1}^{n} f(X_i|\theta) \;=\; \frac{1}{\theta^n} I(X_1, \ldots, X_n \in [0,\theta])$$
$$\;=\; \frac{1}{\theta^n} I(\max(X_1, \ldots, X_n) \leq \theta).$$

Here the indicator function $I(A)$ equals to 1 if event A happens and 0 otherwise. What the indicator above means is that the likelihood will be equal to 0 if at least one of the factors is 0 and this will happen if at least one observation X_i will fall outside of the 'allowed' interval $[0, \theta]$. Another way to say it is that the maximum among observations will exceed θ, i.e.

$$\varphi(\theta) = 0 \text{ if } \theta < \max(X_1, \ldots, X_n),$$

and

$$\varphi(\theta) = \frac{1}{\theta^n} \text{ if } \theta \geq \max(X_1, \ldots, X_n).$$

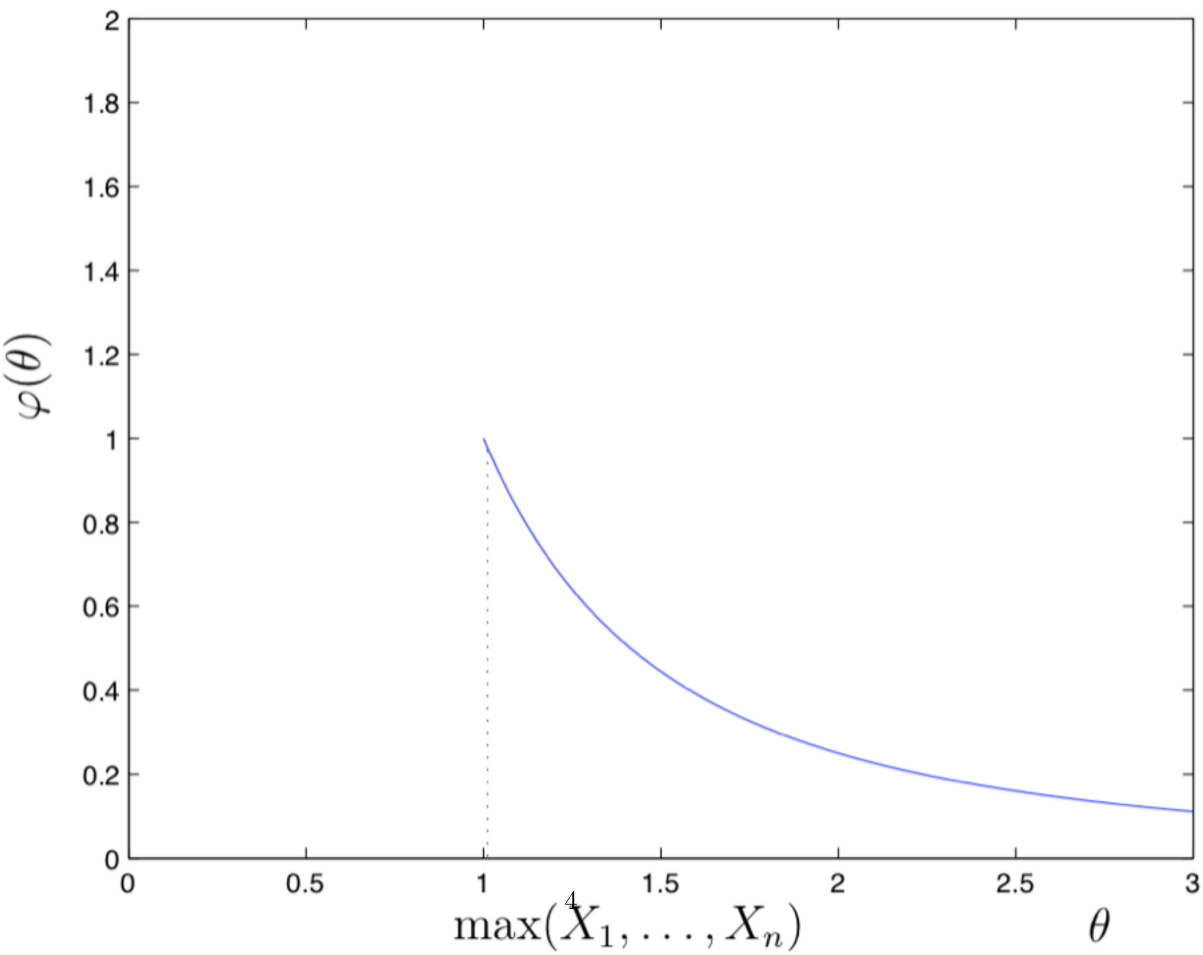

Figure 1: MLE for Uniform Distribution[2]

Thus, for $\theta < max(x_1, x_2, ..., x_n)$, no θ_{MLE}; for $\theta \geq max(x_1, x_2, ..., x_n))$, $\theta_{MLE} = max(x_1, x_2, ..., x_n)$.

3 Eigenvalues and Eigenvectors for Bivariate Gaussian Distribution [25pts]

Let two variables X_1 and X_2 are bivariately normally distributed with mean vector components μ_1 and μ_2 and co-variance matrix Σ shown below:

$$\Sigma = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

(a) What is the probability distribution function of joint Gaussian $P(X_1, X_2)$? (show it with μ and Σ) [5pts]

$p(X|\mu, \Sigma) = \dfrac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\dfrac{1}{2}(X-\mu)^T|\Sigma|^{-1}(X-\mu))$

Here, $\mu = \begin{bmatrix} \mu 1 \\ \mu 2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$

(b) What is the eigenvalues of co-variance matrix Σ? [3] [10pts]

Eigenvalues are obtained by solving the equation: $|\Sigma - \lambda I| = 0$

$\Sigma - \lambda I = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1-\lambda & r \\ r & 1-\lambda \end{bmatrix}$,

$|\Sigma - \lambda I| = (1-\lambda)^2 - r^2 = \lambda^2 - 2\lambda + 1 - r^2 = 0$,

$\lambda = \dfrac{2 \pm \sqrt{2^2 - 4(1-r^2)}}{2} = 1 \pm r.$

(c) Given the condition that the sum of squared values of each eigenvector are equal to 1, what is the eigenvectors of co-variance matrix Σ? (For example, if an eigenvector is $\mathrm{v}_1 = \begin{bmatrix} x1 \\ x2 \end{bmatrix}, then$ $\mathrm{x}_1^2 + x_2^2 = 1$)[10pts]

To obtain the corresponding eigenvectors, we must solve a system of equations below:

$(\Sigma - \lambda I)v_1 = 0$

This is translated for this specific problem in the expression below: $\begin{bmatrix} 1-\lambda & r \\ r & 1-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0.$

Yielding a system of two equations with two unknowns:

$(1-\lambda)x_1 + rx_2 = 0$
$rx_1 + (1-\lambda)x_2 = 0$

For $\lambda = 1 + r$, with $x_1^2 + x_2^2 = 1$, $\mathrm{v}_1 = \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

Similarly, for $\lambda = 1 - r$, with $x_1^2 + x_2^2 = 1$, $\mathrm{v}_2 = \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$

X/Y	1	2
0	$\frac{3}{10}$	0
1	$\frac{4}{10}$	$\frac{3}{10}$

4 Information Theory [25pts]

Suppose the joint probability distribution of two binary random variables X and Y are given as follows.

(a) Show the marginal distribution of X. [2pts]

$P(X = 0) = \frac{3}{10}$
$P(X = 1) = \frac{7}{10}$

(b) Find entropy $H(Y)$. [2pts]

$H(Y) = [P(Y = 1)\times log_2 P(Y = 1)+P(Y = 2)\times log2P(Y = 2)] = \frac{7}{10}\times log_2\frac{7}{10}+\frac{3}{10}\times log_2\frac{3}{10} = 0.88132$

(c) Find conditional entropy $H(X|Y)$ and $H(Y|X)$. [3pts]

$H(X|Y = 1)$
$= -[P(X = 0|Y = 1) \times log_2 P(X = 0|Y = 1) + Pr(X = 1|Y = 1) \times log_2 P(X = 1|Y = 1)]$
$= \frac{3}{7} \times log_2 \frac{3}{7} + \frac{4}{7} \times log_2 \frac{4}{7}] = 0.9852$
$H(X|Y = 2)$
$= -[P(X = 0|Y = 2) \times log_2 P(X = 0|Y = 2) + Pr(X = 1|Y = 2) \times log_2 P(X = 1|Y = 2)]$
$= \frac{0}{0.3} \times log_2 \frac{0}{0.3} + \frac{0.3}{0.3} \times log_2 \frac{0.3}{0.3}] = 0.9852$
$H(X|Y)$
$= P(Y = 1) \times H(X|Y = 1) + P(Y = 2) \times H(X|Y = 2)$
$= \frac{7}{10} \times 0.9852 + \frac{3}{10} \times 0$
$= 0.68964$

Similarly, $H(X|Y) = 0.68964$

(d) Find mutual information $I(X; Y)$. [3pts]

$I(X; Y) = H(X) - H(X|Y) = 0.88132 - 0.68964 = 0.19168$

(e) Find joint entropy $H(X, Y)$. [3pts]

$H(X, Y) = H(X) + H(Y|X) = 0.88132 + 0.68964 = 1.57096$

Method2:

$$\begin{aligned} H(X, Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x, y) log_2 p(x, y) \\ &= -(0.3 \times log_2 0.3 + 0 + 0.4 \times log_2 0.4 + 0.3 \times log_2 0.3) \\ &= 1.57096 \end{aligned}$$

hint: The following three proofs are not related to the given example. You need to prove it for any general case.

(f) Suppose X and Y are independent. Show that $H(X|Y) = H(X)$. [4pts]

If X and Y are independent, then I(X; Y) = 0. (Please refer slide 04 page 17 to the proof).

Thus,$I(X; Y) = H(X) - H(X|Y) = 0$
$H(X) = H(X|Y)$

(g) Suppose X and Y are independent. Show that $H(X, Y) = H(X) + H(Y)$. [4pts]

$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$.

Based on the above proof (f), $H(X) = H(X|Y)$, thus $H(X, Y) = H(X) + H(Y)$.

(h) Show that $I(X; X) = H(X)$. [4pts]

$I(X; X) = H(X) - H(X|X) = \sum_x P(x) \frac{1}{log_2 P(x)} - \sum_{x,x} P(x|x) \times \frac{1}{log_2 P(x|x)}$.

Given $X = x$, the probability of $P(X = x)$ is obviously 1, so $P(X = x|X = x) = 1$.
Then, $log_2 P(x|x) = log_2 1 = 0$

$I(X; X) = H(X) - H(X|X) = H(X) - 0 = H(X)$

5 Reference

You could check the reference for more details.

[1]https://imai.fas.harvard.edu/teaching/files/Expectation.pdf
[2]https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture2.pdf
[3]https://newonlinecourses.science.psu.edu/stat505/node/35/