Probability and Statistics for EECS: Homework #10 Solutions

Professor Ziyu Shao

Problem 1

Let $U_i \sim \text{Unif}(0, 1), i \geq 1$ be i.i.d. random variables. Define N as follows:

$$N = \max \left\{ n : \prod_{i=1}^{n} U_i \geq e^{-1} \right\}$$

(a) Estimate $\text{E}(N)$ by generating 5000 values of N and then use the sample mean.

(b) Estimate $\text{Var}(N)$.

(c) Estimate $P(N = i)$, for $i = 0, 1, 2, 3$.

(d) Can you find the exact distribution of N ?
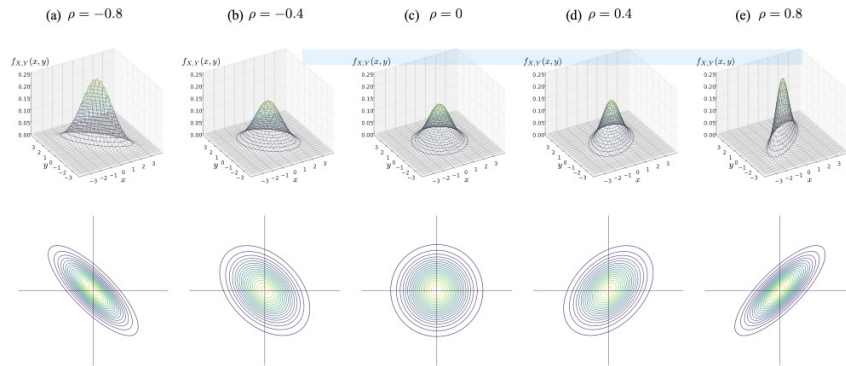
Solution
See the code file.

Problem 2

(a) Use the following transformation to generate samples from bivariate Normal distribution with correlation coefficient ρ:

$$X = Z$$
$$Y = \rho Z + \sqrt{1 - \rho^2} W,$$

where $-1 < \rho < 1$, Z and W are i.i.d. random variables following $\mathcal{N}(0, 1)$.

(b) The joint pdf function of bivariate Normal distribution with correlation coefficient ρ and the corresponding contour (or isocontour) are shown in the following figure for your reference:

Solution
See the code file.

Problem 3

Let $X_1 \sim \text{Expo}(\lambda_1)$, $X_2 \sim \text{Expo}(\lambda_2)$ and $X_3 \sim \text{Expo}(\lambda_3)$ be independent.

(a) Find $E(X_1 \mid X_1 > 2024)$.

(b) Find $E(X_1 \mid X_1 < 1997)$.

(c) Find $E(X_1 + X_2 + X_3 \mid X_1 > 1997, X_2 > 2014, X_3 > 2025)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.

Solution

(a) According to the memoryless property of exponential distribution, we have $E(X_1 - 2023 | X_1 > 2023) = E(X_1)$. Thus we can obtain the conditional expectation as follows:

$$E(X_1|X_1 > 2024) = 2024 + E(X_1 - 2024|X_1 > 2024) = 2024 + E(X_1) = 2024 + \frac{1}{\lambda_1}.$$

(b)

$$\begin{aligned}
E(X_1|X_1 < 1997) &= \frac{E(X_1) - E(X_1|X_1 > 1997)P(X_1 > 1997)}{P(X_1 < 1997)} \\
&= \frac{E(X_1) - (1997 + E(X_1))P(X_1 > 1997)}{P(X_1 < 1997)} \\
&= E(X_1) - \frac{1997 \times P(X_1 > 1997)}{P(X_1 < 1997)} \\
&= 1997 + E(X_1) - \frac{1997}{P(X_1 < 1997)} \\
&= 1997 + \lambda_1 - \frac{1997}{1 - e^{-\lambda_1 1997}}
\end{aligned}$$

(c) Since X_1, X_2, X_3 are independent, we have

$$\begin{aligned}
&E(X_1 + X_2 + X_3|X1 > 1997, X_2 > 2014, X_3 > 2025) \\
=\ & E(X_1|X_1 > 1997, X_2 > 2014, X_3 > 2025) \\
& + E(X_2|X1 > 1997, X_2 > 2014, X_3 > 2025) \\
& + E(X_3|X_1 > 1997, X_2 > 2014, X_3 > 2025) \\
=\ & E(X_1|X_1 > 1997) + E(X_2|X_2 > 2014) + E(X_3|X_3 > 2025) \\
=\ & E(X_1 - 1997|X_1 > 1997) + E(X_2 - 2014|X_2 > 2014) + E(X_3 - 2025|X_3 > 2025) + 6036 \\
=\ & E(X_1) + E(X_2) + E(X_3) + 6036 \\
=\ & \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + 6036
\end{aligned}$$

Problem 4

Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 6xy & \text{if } 0 \le x \le 1, 0 \le y \le \sqrt{x}, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the marginal distributions of X and Y. Are X and Y independent?

(b) Find $E[X \mid Y = y]$ and $\text{Var}[X \mid Y = y]$ for $0 \le y \le 1$.

(c) Find $E[X \mid Y]$ and $\text{Var}[X \mid Y]$.

Solution

(a) The supports of X and Y are both $[0,1]$. In this way, we have

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \\
&= \int_0^{\sqrt{x}} 6xy \, dy \\
&= 3xy^2 \Big|_{y=0}^{y=\sqrt{x}} \\
&= 3x^2,
\end{aligned}$$

and

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \\
&= \int_{y^2}^1 6xy \, dx \\
&= 3yx^2 \Big|_{x=y^2}^{x=1} \\
&= 3y - 3y^5.
\end{aligned}$$

Therefore,

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \begin{cases} 3y - 3y^5 & \text{if } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since $f_{X,Y}(x,y) \ne f_X(x) f_Y(y)$, X and Y are not independent.

(b) Since

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx,$$

to calculate $E[X|Y = y]$, we need to first calculate $f_{X|Y}(x|y)$.

If $y^2 \le x \le 1$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2x}{1 - y^4}.$$

In this way,

$$f_{X|Y}(x|y) = \begin{cases} \frac{2x}{1-y^4} & \text{if } y^2 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

5

Therefore,

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

$$= \int_{y^2}^{1} x \frac{2x}{1-y^4} dx$$

$$= \frac{2}{3(1-y^4)} x^3 \Big|_{x=y^2}^{x=1}$$

$$= \frac{2(1-y^6)}{3(1-y^4)}$$

$$= \frac{2}{3} \cdot \frac{1+y^2+y^4}{1+y^2}$$

Since

$$\mathrm{Var}[X|Y=y] = E[X^2|Y=y] - (E[X|Y=y])^2,$$

to calculate $\mathrm{Var}[X|Y=y]$, we need to first calculate $E[X^2|Y=y]$.

Since

$$E[X^2|Y=y] = \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y)dx$$

$$= \int_{y^2}^{1} x^2 \frac{2x}{1-y^4} dx$$

$$= \frac{1}{2(1-y^4)} x^4 \Big|_{x=y^2}^{x=1}$$

$$= \frac{1-y^8}{2(1-y^4)}$$

$$= \frac{1+y^4}{2},$$

we have,

$$\mathrm{Var}[X|Y=y] = E[X^2|Y=y] - (E[X|Y=y])^2$$

$$= \frac{1+y^4}{2} - \left(\frac{2(1-y^6)}{3(1-y^4)}\right)^2$$

$$= \frac{1+y^4}{2} - \frac{4}{9} \cdot \frac{(1+y^2+y^4)^2}{(1+y^2)^2}$$

(c) According to the result in question(b), we have

$$E[X|Y] = \frac{2}{3} \cdot \frac{1+Y^2+Y^4}{1+Y^2},$$

$$\mathrm{Var}[X|Y] = \frac{1+Y^4}{2} - \frac{4}{9} \cdot \frac{(1+Y^2+Y^4)^2}{(1+Y^2)^2}.$$

Problem 5

Let X be a discrete r.v. whose distinct possible values are $x_0, x_1, \ldots$, and let $p_k = P(X = x_k)$. The entropy of X is $H(X) = \sum_{k=0}^{\infty} p_k \log_2(1/p_k)$.

(a) Find $H(X)$ for $X \sim \text{Geom}(p)$.

(b) Let X and Y be i.i.d. discrete r.v.s. Show that $P(X = Y) \geq 2^{-H(X)}$. Hint: Jensen's Inequality.

Solution

(a) The PMF of X is $P(X = k) = p(1-p)^k$ since there is $X \sim \text{Geom}(p)$. Thus we have

$$H(X) = -\sum_{k=0}^{\infty} p(1-p)^k \log_2 \left(p(1-p)^k \right)$$

$$= -p \sum_{k=0}^{\infty} k(1-p)^k \log_2(1-p) - p \log_2 p \sum_{k=0}^{\infty} (1-p)^k$$

$$= -\log_2 p - \frac{1-p}{p} \log_2(1-p)$$

(b) Since X and Y are i.i.d random variables, via LOTP, we have

$$P(X = Y) = \sum_{k=0}^{\infty} P(X = Y | Y = k) \cdot P(Y = k)$$

$$= \sum_{k=0}^{\infty} P(X = k) \cdot P(Y = k) = \sum_{k=0}^{\infty} p_k^2$$

Denote Z as a new discrete random variable such that $P(Z = p_k) = p_k$, then we have:

$$E(Z) = \sum_{k=0}^{\infty} p_k \times p_k = P(X = Y)$$

Since $\log(\cdot)$ is a convex function, according to Jensen's inequality, we have $E(\log(Z)) \leq \log(E(Z))$, thus there is

$$\sum p_k \log_2 p_k \leq \log_2 \sum p_k^2$$
$$\Leftrightarrow -H(X) \leq \log_2 P(X = Y)$$
$$\Leftrightarrow P(X = Y) \geq 2^{-H(X)}.$$

Problem 6

Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1 - \delta$ confidence interval for a parameter p is an interval $[\hat{p} - \epsilon, \hat{p} + \epsilon]$ such that $P(p \in [\hat{p} - \epsilon, \hat{p} + \epsilon]) \geq 1 - \delta$. Now we toss a coin with probability p landing heads and probability $1 - p$ landing tails. The parameter p is unknown and we need to estimate its value from experiment results. We toss such coin N times. Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using

$$\hat{p} = \frac{X_1 + \ldots + X_N}{N}.$$

Find the $1 - \delta$ confidence interval for p, then discuss the impacts of δ and N.

(a) Method 1: Adopt Chebyshev inequality to find the $1 - \delta$ confidence interval for p, then discuss the impacts of δ and N.

(b) Method 2: Adopt Hoeffding bound to find the $1 - \delta$ confidence interval for p, then discuss the impacts of δ and N.

(c) Discuss the pros and cons of the above two methods.

Solution
Since $X_i \sim \text{Bern}(p), X_i \in \{0, 1\}$, we have $\mathbb{E}[X_i] = p$ and $\mathbb{V}[X_i] = p(1 - p)$. Therefore, we have

$$\mathbb{E}[\hat{p}] = p, \mathbb{V}[\hat{p}] = \frac{p(1 - p)}{N}.$$

Besides, we know that
$$P(p \in [\hat{p} - \varepsilon, \hat{p} + \varepsilon]) \geq 1 - \delta \Leftrightarrow P(|\hat{p} - p| \geq \varepsilon) \leq \delta.$$

(a) Applying Chebyshev's inequality on random variable $\hat{p}$, we have

$$P(|\hat{p} - p| \geq \epsilon) \leq \frac{p(1 - p)}{N\epsilon^2} \Rightarrow \delta = \frac{p(1 - p)}{N\epsilon^2}, \epsilon = \sqrt{\frac{p(1 - p)}{N\delta}}$$

Therefore, we know that δ negatively correlates with ϵ, i.e., given a fixed number of samples N, there is natural trade-off between accuracy and confidence. Besides, 1) Fix the confidence interval parametrized by δ, reducing the estimation error ϵ requires increasing the number of samples N. 2) Fix the estimation error ϵ, narrowing the confidence interval requires increasing the number of samples N. That is, the impacts of N is on both the "estimation accuracy" and "estimation confidence".

(b) Applying Hoeffding's inequality on random variable $\hat{p}$, we have

$$P(|\hat{p} - p| \geq \epsilon) \leq 2e^{-2N\epsilon^2} \Rightarrow \delta = 2e^{-2N\epsilon^2}, \epsilon = \sqrt{\frac{\ln(2/\delta)}{2N}}$$

The effects of δ and N are similarly discussed as in (a).

(c) Chebyshev's inequality (see Cantelli's inequality for the one-side improvement):

- Pros: 1) sharp bound and cannot be improved in general (given no extra assumption). 2) can be improved with extra distributional information on polynomial moments.
- Cons: 1) requires the existence of moments until the second order. 2) quadratic convergence rate.

Hoeffding's inequality (see Theorem 2.8 and 2.9 of paper "old and new concentration inequalities" for the one-side improvement):

- Pros: 1) exponential convergence rate. 2) does not require assumption on moments.
- Cons: 1) works only for sub-Gaussian (e.g., bounded random variables). 2) in general not sharp when the variance is small (e.g., see popoviciu's inequality on variances and Bernstein's inequality).

Problem 7

We consider the progressive Monty Hall problem. This time we assume there are n identical doors, where n is an integer satisfying $n \geq 3$. One door conceals a car, the other $n-1$ doors conceal goats. You choose one of the doors at random but do not open it. Monty then opens a door he knows to conceal a goat, always choosing randomly among the available doors. At this point he gives you the option either of sticking with your original door or switching to one of the remaining doors. You make your decision. Monty now eliminates another goat-concealing door (at random) and once more gives you the choice either of sticking or switching. This process continues until only two doors remain in play. What strategy should you follow to maximize your chances of winning? We consider three strategies: (1) Select a door at random and stick with it throughout. (2) Select a door at random, then switch doors at every opportunity, choosing your door randomly at each step. (3) Select a door at random, stick with your first choice until only two doors remain, and then switch. **When $n = 4$ and $n = 100$, please use simulation to compare such three strategies.**

Solution

See the code file.