Tutorial 1: CS 215, Fall 2024

Prof. Sunita Sarawagi

12 August 2024

1. Let X be a random variable. If Var(X) = 0, then prove that (*Hint:* Use Chebyshev's Inequality)

$$P\{X = E[X]\} = 1$$

2. Let X and Y be two Random Variables.
   (a) If they are independent, show that they are uncorrelated (i.e. Corr(X, Y) = 0).
   (b) If they are uncorrelated (i.e. Corr(X, Y) = 0), does it imply they are independent? Prove or find a counterexample.

3. Let (X_i, Y_i), $i = 1, ..., n$ be a sequence of independent and identically distributed random vectors. That is, X_1, Y_1 is independent of, and has the same distribution as, X_2, Y_2, and so on. Although X_i and Y_i can be dependent, X_i and Y_j are independent when $i \neq j$. Let

$$\mu_x = E[X_i],\ \mu_y = E[Y_i]$$
$$\sigma_x^2 = Var(X_i),\ \sigma_y^2 = Var(Y_i)$$
$$\rho = Corr(X_i, Y_i)$$

   Find Corr($\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i$) (You might need to prove $Var(X + Y) = Var(X) + Var(Y)$ for independent R.V. X and Y)

4. A deck of 52 cards is shuffled and a bridge hand of 13 cards is dealt out. Let X and Y denote, respectively, the number of aces and the number of spades in the hand.
   (a) Show that X and Y are uncorrelated.
   (b) Are they independent?

5. (Gambler's fallacy) Mr. Jones has devised a gambling system for winning at roulette. When he bets, he bets on red and places a bet only when the 10 previous spins of the roulette have landed on a black number. He reasons that his chance of winning is quite large because the probability of 11 consecutive spins resulting in black is quite small. What do you think of this system?

6. Show that for any events E and F,

$$P(E|E \cup F) \geq P(E|F)$$

   In words, probability that E occurs given E or F occurs must be larger than if we just know that only F occurs. (Hint: Compute $P(E|E \cup F)$ by conditioning on whether F occurs)

7. Consider the following game played with an ordinary deck of 52 playing cards: The cards are shuffled and then turned over one at a time. At any time, the player can guess that the next card to be turned over will be the ace of spades; if it is, then the player wins. In addition, the player is said to win if the ace of spades has not yet appeared when only one card remains and no guess has yet been made. What is a good strategy? What is a bad strategy?

8. (a) A gambler has a fair coin and a two-headed coin in his pocket. He selects one of the coins at random; when he flips it, it shows heads. What is the probability that it is the fair coin?
   (b) Suppose that he flips the same coin a second time and, again, it shows heads. Now what is the probability that it is the fair coin?
   (c) Suppose that he flips the same coin a third time and it shows tails. Now what is the probability that it is the fair coin?

9. There are k + 1 coins in a box. When flipped, the i^{th} coin will turn up heads with probability i/k, i = 0, 1, … , k. A coin is randomly selected from the box and is then repeatedly flipped. If the first n flips all result in heads, what is the conditional probability that the (n + 1) flip will do likewise? Calculate for large k. (*Hint:* For large n, use approximation $\frac{1}{n}\sum_{i=0}^{n}\left(\frac{i}{n}\right)^m \approx \int_0^1 x^m dx$)

10. A person tried by a 3-judge panel is declared guilty if at least 2 judges cast votes of guilty. Suppose that when the defendant is in fact guilty, each judge will independently vote guilty with probability 0.7, whereas when the defendant is in fact innocent, this probability drops to 0.2. If 70 percent of defendants are guilty, compute the conditional probability that judge number 3 votes guilty given that

    (a) judges 1 and 2 vote guilty

    (b) judges 1 and 2 cast 1 guilty and 1 not guilty vote

    (c) judges 1 and 2 both cast not guilty votes

CS215 Fall, 2024: Tutorial 1

Atharva Tambat

September 15, 2024

Table of Contents

Question 1

Using Chebyshev's Inequality:

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Here, $\sigma^2 = 0$. Take $k = \frac{1}{n}$

$$P\left\{|X-\mu| \geq \frac{1}{n}\right\} \leq 0 \implies P\left\{|X-\mu| \geq \frac{1}{n}\right\} = 0$$

Take limit $n \to \infty$,

$$\lim_{n\to\infty} P\left\{|X-\mu| \geq \frac{1}{n}\right\} = P\left\{\lim_{n\to\infty}\left\{|X-\mu| \geq \frac{1}{n}\right\}\right\} = P\{X \neq \mu\} = 0$$

Table of Contents

Question 2 - Independence $\not\Longleftrightarrow$ Uncorrelated

1. Proof: Independence $\implies$ Uncorrelated
For Independent variables X and Y, E[XY] = E[X]E[Y]
Proof:

$$\begin{aligned} E[XY] = \mu_{xy} &= \sum_i \sum_j x_i y_j p(x_i, y_j) \\ &= \sum_i \sum_j x_i y_j p(x_i) p(y_j) \\ &= \left(\sum_i x_i p(x_i) \right) \left(\sum_j y_j p(y_j) \right) \\ &= E[X]E[Y] = \mu_x \mu_y \end{aligned}$$

Question 2 - Independence $\not\Leftrightarrow$ Uncorrelated

1. Proof: Uncorrelated $\not\Rightarrow$ Independence
   Proof: Counterexample: $X \in \{-1, 0, 1\}$

   $P(X = -1) = P(X = 0) = P(X = 1) = 1/3$

$$Y = \begin{cases} 1, & \text{if } X = 0 \\ 0, & \text{otherwise} \end{cases}$$

Table of Contents

Question 3 - Result 1

- First, prove that $Cov(X_1 + X_2, Y_1) = Cov(X_1, Y_1) + Cov(X_2, Y_1)$
  Proof:

$$Cov(X_1 + X_2, Y_1) = E[(X_1 + X_2)Y_1] - E[X_1 + X_2]E[Y]$$

$$= E[X_1 Y_1] + E[X_2 Y_1] - (E[X_1]E[Y_1] + E[X_2]E[Y_1])$$

$$= (E[X_1 Y_1] - E[X_1]E[Y_1]) + (E[X_2 Y_1] - E[X_2]E[Y_1])$$

$$= Cov(X_1, Y_1) + Cov(X_2, Y_1)$$

  - Since, $E[X + Y] = E[X] + E[Y]$ for R.V. X and Y

Question 3 - Result 2

- Next, prove that $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
  Proof:

$$\begin{aligned} Var(X+Y) &= E[(X+Y)^2] - E[X+Y]^2 \\ &= E[X^2 + Y^2 + 2XY] - (E[X] + E[Y])^2 \\ &= E[X^2 + Y^2 + 2XY] - (E[X]^2 + E[Y]^2 + E[X]E[Y]) \\ &= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2) + 2(E[XY] - E[X]E[Y]) \\ &= Var(X) + Var(Y) + 2Cov(X,Y) \end{aligned}$$

Question 3 - Final Step

$$Corr(\sum_i X_i, \sum_i Y_i) = \frac{Cov(\sum_i X_i, \sum_i Y_i)}{\sqrt{Var(\sum_i X_i) Var(\sum_i Y_i)}}$$

Since all the X_is are independent of each other (same for all Y_is), which means $Cov(X_i, X_j) = 0 \; \forall i \neq j$, Using Result 2

$$Corr(\sum_i X_i, \sum_i Y_i) = \frac{Cov(\sum_i X_i, \sum_i Y_i)}{\sqrt{\sum_i Var(X_i) \sum_i Var(Y_i)}}$$

Using Result 1,

$$= \frac{\sum_i \sum_j Cov(X_i, Y_j)}{\sqrt{\sum_i Var(X_i) \sum_i Var(Y_i)}}$$

$$= \frac{\sum_{i=j} Cov(X_i, Y_j) + \sum_{i \neq j} Cov(X_i, Y_j)}{\sqrt{(n\sigma_x^2)(n\sigma_y^2)}} = \frac{n\rho\sigma_x\sigma_y + 0}{\sqrt{(n\sigma_x^2)(n\sigma_y^2)}} = \rho$$

Table of Contents

Question 4

Define R.Vs (called indicator RVs - which indicate whether an event has happened or not) as follows

$$X_i = \begin{cases} 1, & \text{if you get the } i^{th} \text{ Ace in the dealt cards} \\ 0, & \text{otherwise} \end{cases}$$

$$Y_i = \begin{cases} 1, & \text{if you get the } i^{th} \text{ Spade in the dealt cards} \\ 0, & \text{otherwise} \end{cases}$$

So you have $X_1, X_2 ... X_4$ and $Y_1, Y_2 ... Y_{13}$ random variables. **Also notice two R.Vs (one from Xs and other from Ys), which denote Ace of Spades (WLOG, we'll call them X_1 and Y_1), will have the same value**. i.e $X_1 = Y_1$.

Question 4

So, number of aces in the cards dealt (X) $= \sum_i X_i$.

Number of spades in the cards dealt (Y) $= \sum_i Y_i$

We'll show Cov(X, Y) $= 0$

Question 4

$$Cov(X, Y) = Cov(\sum_i X_i, \sum_j Y_j)$$

Using the result 1 from Question 3,

$$Cov(\sum_i X_i, \sum_j Y_j) = \sum_i \sum_j Cov(X_i, Y_j)$$

$$= Cov(X_1, Y_1) + \sum_{i \neq 1 \ or \ j \neq 1} Cov(X_i, Y_j)$$

Question 4

- First find $Cov(X_1, Y_1)$.

$$Cov(X_1, Y_1) = E[X_1 Y_1] - E[X_1]E[Y_1] = E[X_1^2] - E[X_1]^2$$

Since $X_1 = Y_1$

$$= Var(X_1)$$

- Notice two things
  - $E[X_1^2] = E[X_1]$ since X_1 is an indicator RV
  - $E[X_1] =$ **probability, p, of the event which is indicated by the indicator variable**. Since $E[X_1] = p.1 + (1-p).0 = p$
  - So, $Var(X_1) = \frac{1}{4}\left(1 - \frac{1}{4}\right) = \frac{3}{16}$

Question 4

Next, calculate $Cov(X_i, Y_j)$, where $i \neq 1$ *or* $j \neq 1$ - i.e. X_i and Y_j represent **different** cards. Note, there are **51 such terms**.

$$Cov(X_i, Y_j) = E[X_i Y_j] - E[X_i]E[Y_j]$$

- $$\begin{aligned} E[X_i Y_j] &= P(X_i = 1, Y_j = 1).1.1 + P(X_i = 1, Y_j = 0).1.0 \\ &+ P(X_i = 0, Y_j = 1).0.1 + P(X_i = 0, Y_j = 0).0.0 \\ &= \frac{50C11}{52C13} = 1/17 \end{aligned}$$

- $E[X_i] = E[Y_j] = 51C12/52C13 = 1/4$
- $Cov(X_i, Y_j) = -1/272$

Question 4 - The final nail in the coffin

$$Cov(X, Y) = \frac{3}{16} + 51(-\frac{1}{272}) = 0$$

Question 4 (ii)

These two events are not independent, even though they are uncorrelated - one more example of Question 2.

$$P(X = 4, Y = 13) = 0 \neq P(X = 4)P(Y = 13)$$

But for X and Y to be independent, $P(X, Y) = P(X).P(Y)$ should hold for all X and Y, therefore, contradiction.

Table of Contents

Question 5

We'll calculate the probability of Mr. Jones winning. Let W be the event of Mr. Jones winning at a game, and let T be the event that in the 10 previous spins, the ball has landed on a black number.

Since, Mr. Jones does not bet if in the 10 previous turns, the ball hasn't landed on a black number, we only care about $P(W|T)$

Since, W is independent of T, therefore the probability of him winning the match is

$$P(W|T) = P(W)$$

Which is equal to the probability of winning **any individual** game (not necessarily the one after getting 10 blacks)

Table of Contents

Question 6 - Law of Total Probability

Applying the law of total probability:

$$P(E|E \cup F) = P(E|E \cup F, F)P(F) + P(E|E \cup F, \bar{F})P(\bar{F})$$

1. $P(A|C, D)$ - Probability of event A, given events C **and** D have occurred
2. $P(E|E \cup F, F) = \frac{P(E \cap (E \cup F) \cap F)}{P((E \cup F) \cap F)} = P(E \cap F)/P(F) = P(E|F)$
3. $P(E|E \cup F, \bar{F}) = 1$. Why?

Question 6 - Law of Total Probability

Thus,

$$\begin{aligned} P(E|E \cup F) &= P(E|F)P(F) + (1 - P(F)) \\ &\geq P(E|F)P(F) + P(E|F)(1 - P(F)) \quad \textbf{Why?} \end{aligned}$$

So,

$$P(E|E \cup F) \geq P(E|F)(P(F) + 1 - P(F)) = P(E|F)$$

Table of Contents

Optimal Strategy

Ans: **Any strategy** has a winning probability of 1/52!!

Proof: Let us prove a stronger point - in a game with n cards, the winning probability is always 1/n regardless of the strategy.

1. $p \rightarrow$ probability that the strategy chooses the first card
2. $G \rightarrow$ Event that the first card is guessed

Two cases of winning:

- 1^{st} card is ace of spades - happens with probability 1/n
- 1^{st} card is not ace of spades: What is the probability of win, **given** we skip the first chance?

Optimal Strategy

H = first card is not ace of spades

$$P(H).P(\{win\}|H)$$

But $P(\{win\}|H)$ = probability of winning with n-1 cards = $\frac{1}{n-1}$ by induction hypothesis

$$P(H).P(\{win\}|H) = \frac{n-1}{n}\frac{1}{n-1} = \frac{1}{n}$$

Using the Law of total probability:

$$\begin{aligned} P(\{win\}) &= P(\{win\}|G)P(G) + P(\{win\}|\bar{G})(1 - P(G)) \\ &= \frac{1}{n}p + \frac{1}{n}(1-p) \\ &= \frac{1}{n} \end{aligned}$$

Table of Contents

Coin Flip

- $C_i \rightarrow$ Event that coin i is chosen
- $F_n \rightarrow$ Event that first n tosses are heads
- $H \rightarrow$ Event that the $(n+1)^{th}$ is a head

$$P(H|F_n) = ?$$

$$P(H|F_n) = \sum_{i=0}^{k} P(H|F_n C_i) P(C_i|F_n)$$

Coin Flip

$P(H|F_nC_i)$ - means given i^{th} coin is selected and first n tosses are heads, what is the probability $(n+1)^{th}$ toss is a head.

$$P(H|F_nC_i) = P(H|C_i) = \frac{i}{k} \quad \textbf{Why?}$$

Also,

$$P(C_i|F_n) = \frac{P(C_iF_n)}{P(F_n)} = \frac{P(F_n|C_i)P(C_i)}{\sum_{j=0}^{k} P(F_n|C_j)P(C_j)}$$

Coin Flip

$$P(C_i|F_n) = \frac{(i/k)^n[1/(k+1)]}{\sum_{j=0}^{n}(j/k)^n[1/(k+1)]}$$

Thus,

$$P(H|F_n) = \sum_{i=0}^{k} \frac{(i/k)^{n+1}}{\sum_{j=0}^{k}(j/k)^n} = \frac{\sum_{i=0}^{k}(i/k)^{n+1}}{\sum_{j=0}^{k}(j/k)^n}$$

Now, use the approximation to simplify the numerator and denominator

Problems from TextBook: CS 215, Fall 2024

Prof. Sunita Sarawagi

29 August 2024

This is a sheet containing practice questions. Solve them and ask any doubts regarding them in tutorial classes. Refer to Ross Textbook Edition 3.

1 Ross Textbook Chapter 4

Try Questions: 6, 7, 8, 11, 25, 26, 28

2 Ross Textbook Chapter 5

Try Questions: 4, 5, 6

3 Questions

1. Prove or disprove the following:
   (a) If event B includes the event A, then $P(A) \leq P(B)$.
   (b) If $P(B \mid A) = P(B \mid \bar{A})$, then A and B are independent.
   (c) If P(A) = a and P(B) = b, then $P(A|B) \geq \frac{a+b-1}{b}$
2. Find the probability that a randomly written fraction will be irreducible. Assume both numerator and denominator belongs to set of natural numbers
3. Under what conditions does the following equality hold:

$$P(A) = P(A|B) + P(A|\bar{B})?$$

Practice Problems

CS215

September, 2024

Problem 1

Here is a nice application of basic probability. I hope you will like it! Suppose some n individuals have arrived at a lab for RTPCR testing. An RTPCR test involves extracting the nasal mucus of the individual and testing it within an RTPCR machine. Suppose that the probability that any individual will test positive is p, and let us assume that the test results across all n individuals are independent. Instead of individually testing each person, we follow the two-step procedure described below to save on the number of tests: (1) We divide the people into n/g groups, each of size g where we assume that g divides n. Small, equal-volume portions of the mucus samples of all individuals belonging to the same group are mixed together. This mixture is tested, thus leading to n/g independent tests, one per group. (2) If the mixture tests negative (non-infected), then all group members are declared negative. If the mixture tests positive (infected), then each member of the group is individually tested in a second round of tests.

It is known that the mixture of different mucus samples, or using small portions of the sample, has no influence on the probability p or on test accuracy. This procedure is called Dorfman pooling and it was widely used during the COVID-19 pandemic.

(a) What is the expected total number of tests? Note that an individual test counts as one test, and the test of a mixture also counts as one test.

(b) Now suppose that exactly $k \ll n$ individuals are infected. In this scenario, what is the number of tests required in Dorfman's method in the worst case? For what value of g, expressed in terms of n and k, will this worst case number of tests be minimized? What is the number of tests in that case?

Solution

(a) The total number of tests is n/g in the first round. The probability of any one individual being infected is p. Hence the probability of any one individual being non-infected is $1-p$. The probability of obtaining a group of g individuals who are all non-infected is $(1-p)^g$, and hence the probability of obtaining a group of g individuals containing at least one infected individual is $1-(1-p)^g$. As the total number of groups is n/g, the expected number of groups with at least one infected member is $n/g \times [1-(1-p)^g]$. Hence the expected number of total tests is $n/g + g \times n/g \times [1-(1-p)^g] = n/g + n\,[1-(1-p)^g]$.

(b) In the worst case, each of the k infected people will lie in a different group. So each of these k groups will have to be tested in the second round, and each member of these k groups will be tested individually. This will give rise to $k \times g$ more tests. So the total number is $n/g + kg$ for the worst case number of tests. If the worst case number has to be optimal, we set the derivative of this number (w.r.t. g) to zero, giving rise to $-n/g^2 + k = 0$, that is, $g = \sqrt{n/k}$. The number of tests in this case will be $2\sqrt{nk}$.

Problem 2

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then express the CDF of $Y = aX + b$ in terms of the CDF of X. Also write down the PDF of Y. Here a, b are non-zero constants. If the PDF of X is $f_X(\cdot)$ and $Y = aX + b$ as before, write down an expression for the PDF of Y, i.e. $f_Y(\cdot)$ in terms of $f_X(\cdot)$.

Solution

In the case where $a > 0$, we have

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq (y-b)/a) = F_X((y-b)/a).$$

In the case that $a < 0$, we have

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \geq (y-b)/a) = 1 - F_X((y-b)/a).$$

This is the expression for the CDF of Y.
The PDF is obtained by taking the derivative w.r.t. x, giving $f_Y(y) = f_X((y-b)/a)/a$ when $a > 0$ and

$$f_Y(y) = -f_X((y-b)/a)/a \quad \text{when } a < 0.$$

This yields

$$f_Y(y) = f_X((y-b)/a)/|a|.$$

Now since

$$f_X(x) = \frac{e^{-(y-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}},$$

we have

$$f_Y(y) = \frac{e^{-(y-b-a\mu)^2/(2a^2\sigma^2)}}{|a|\sigma\sqrt{2\pi}} = \frac{e^{-(y-b-a\mu)^2/(2a^2\sigma^2)}}{|a|\sigma\sqrt{2\pi}}.$$

Problem 3

The entropy of a discrete random variable X is defined as

$$H(X) = -\sum_{i=1}^{K} p_i \log p_i$$

where $p_i = P(X = i)$ and $\sum_{i=1}^{K} p_i = 1$; $\forall i, 0 \leq p_i \leq 1$. In this definition, $0 \log 0$ is considered to be 0. For which PMF (i.e. for what values of $\{p_i\}_{i=1}^{K}$) will the entropy be maximum? What is this maximum value? Derive your answer by setting the first derivatives of the entropy to 0. Obtain the sign of the second derivatives, i.e. sign of $\frac{\partial^2 H}{\partial p_i^2}$. (Note: Given what you have learned so far, you will be able to find only a local maximum. But it turns out that the unexpectedness measure is a concave function, due to which a local maximum is also the global maximum. You are not expected to prove that it is a concave function.) For what PMF will the entropy be the least? Give an intuitive answer (it is not so easy to prove your answer for the minimum, in a quiz/exam). What is this minimum value?

Solution

We have

$$H(X) = -\sum_{i=1}^{K-1} p_i \log p_i - p_K \log p_K = -\sum_{i=1}^{K-1} p_i \log p_i + (1 - \sum_{i=1}^{K-1} p_i) \log(1 - \sum_{i=1}^{K-1} p_i).$$

Hence, we have for $j \in \{1, \dots, K-1\}$ that

$$\frac{\partial H(X)}{\partial p_j} = -(1 + \log p_j) - \left(-1 - \log\left(1 - \sum_{i=1}^{K-1} p_i\right)\right) = -\log p_j + \log\left(1 - \sum_{i=1}^{K-1} p_j\right).$$

Setting this derivative to zero, we have $p_j = 1 - \sum_{i=1}^{K-1} p_j = p_K$. Thus, all the p_j values are equal to $1/K$, i.e. we have a discrete uniform PMD. The second derivative is given as

$$\frac{\partial^2 H(X)}{\partial p_j^2} = -1/p_j - 1/\left(1 - \sum_{j=1}^{K-1} p_j\right)$$

which is clearly less than 0, and thus the second derivative test is passed. Thus, a discrete uniform PMF maximizes the entropy. The least possible entropy occurs if $p_j = 1$ for some $j \in \{1, \dots, K\}$ with the other values being all 0. This is called a Kronecker delta function. The entropy is always non-negative and because $0 \leq p_i \leq 1$ and hence $-p_i \log p_i$ is always positive.

Problem 4

Let Y be a Gaussian random variable with mean μ and variance σ^2. Derive the CDF and PDF of the random variable $X = |Y|$. Also derive $E\left(X^2\right)$.

Solution

We have $F_X(x) = P(X \leq x) = P(|Y| \leq x) = P(-x \leq Y \leq x) = \int_{-x}^{+x} f_Y(y)dy = F_Y(x) - F_Y(-x) = \Phi((x-\mu)/\sigma) - \Phi((-x-\mu)/\sigma)$ where Φ stands for the CDF of a zero-mean Gaussian random variable with unit variance. The PDF of X is given by $f_X(x) = \frac{1}{\sigma}[\phi((x-\mu)/\sigma) + \phi((-x-\mu)/\sigma)] = \frac{1}{\sigma}[\phi((x-\ \mu)/\sigma) + \phi((x+\mu)/\sigma)]$ where ϕ stands for the PDF of a zero-mean Gaussian random variable with unit variance, and where we use the symmetry of ϕ. This yields $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\left[\exp\left(-(x-\mu)^2/2\sigma^2\right) + \exp\left(-(x+\mu)^2/2\sigma^2\right)\right]$.

Problem 5

In a certain town, there exist 100 rickshaws out of which 1 is red and 99 are blue. A person XYZ observes a serious accident caused by a rickshaw at night and remembers that the rickshaw was red in color. Hence, the police arrest the driver of the red rickshaw. The driver pleads innocence. Now, a lawyer decides to defend the hapless rickshaw driver in court. The lawyer ropes in an opthalmologist to test XYZ's ability to differentiate between the colors red and blue, under illumination conditions similar to those that existed that fateful night. The opthalmologist suggests that XYZ sees red objects as red 99% of the time and blue objects as red 2% of the time. What will be the main argument of the defense lawyer? (In other words, what is the probability that the rickshaw was really a red one, when XYZ observed it to be red?)

Solution

Let R_R, R_B be the events that the rickshaw was red, blue respectively. Let X_R, X_B be the events that XYZ perceived a rickshaw to be red, blue respectively. We have $P\left(X_R \mid R_R\right) = 0.99, R\left(X_R \mid R_B\right) = 0.02, P\left(R_R\right) = 0.01, P\left(R_B\right) = 0.99$. We need to evaluate $P\left(R_R \mid X_R\right) = P\left(X_R \mid R_R\right) P\left(R_R\right) / P\left(X_R\right) . P\left(X_R\right) = P\left(X_R \mid R_R\right) P\left(R_R\right) + P\left(X_R \mid R_B\right) P\left(R_B\right) = 0.99 \times 0.01 + 0.02 \times 0.99 = 0.99 \times 0.03$. Hence $P\left(R_R \mid X_R\right) = \frac{0.99 \times 0.01}{0.99 \times 0.03} = 1/3$. In other words, the probability that the rickshaw was red when XYZ observed it to be red is only 1/3. In other words, it is more probable that the rickshaw was a blue one, based on the available data!

Problem 6

If X and Y are two independent continuous random variables with PDFs f_X and f_Y respectively, then prove that the PDF of $Z = XY$ is given by $f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z/x) \frac{1}{|x|} dx$.

Solution

The CDF of Z is given by:

$$\begin{aligned} F_Z(z) = P(Z \leq z) &= P(XY \leq z, X \geq 0) + P(XY \leq z, X \leq 0) \\ &= P(Y \leq z/X, X \geq 0) + P(Y \geq z/X, X \leq 0) \\ &= (*) \textstyle\int_0^{+\infty} f_X(x) \int_{-\infty}^{z/x} f_Y(y) dy dx + \int_{-\infty}^0 f_X(x) \int_{z/x}^{\infty} f_Y(y) dy dx \\ &= \textstyle\int_0^{+\infty} f_X(x) \left(F_Y(z/x) - 0\right) dx + \int_{-\infty}^0 f_X(x) \left(1 - F_Y(z/x) dx\right. \end{aligned}$$

where the step marked (*) follows due to independence. Taking derivatives w.r.t. z, we obtain the PDF of Z as follows:

$$\begin{aligned} f_Y(z) &= \int_0^{+\infty} f_X(x) f_Y(z/x)/x dx - \int_{-\infty}^0 f_X(x) f_Y(z/x)/x dx \\ &= \int_0^{+\infty} f_X(x) f_Y(z/x)/|x| dx + \int_{-\infty}^0 f_X(x) f_Y(z/x)/|x| dx \\ &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z/x)/|x| dx \end{aligned}$$

Practice on MLE (sec 7.2), Evaluating Point Estimators (sec 7.7), and Bayesian Estimator (sec 7.8)

1 Questions

1. Consider the task of performing line fitting to a set of points (x_i, y_i). For this question, model y_i as having the equation $mx_i + c$ (we do NOT know m and c) but with Gaussian Noise added to it in the form $\mathcal{N}(0, \sigma^2)$. Use the concept of MLE to find the estimate for m, c and σ.

We model our data as

$$y_i = mx_i + c + \epsilon$$

where ϵ is the noise sampled from a Gaussian.
Thus, $y_i \sim \mathcal{N}(mx_i + c, \sigma^2)$

$$P = \sum p(y_i|x_i, m, c) = \sum_{i=1}^{n} \frac{\exp\left(\frac{-(y_i - mx_i - c)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$$

Set partial derivative of $logP$ to 0 with respect to both m and c to obtain,

$$\begin{aligned} c' &= \sum_{i=1}^{n} \frac{y_i - m'x_i}{n} \\ m' &= \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sum_{i=1}^{n}(x_i - \mu_x)^2} \end{aligned}$$

where $\mu_x = (\sum x_i)/n$ and $\mu_y = (\sum y_i)/n$. use the MLE Gaussian estimation on $y_i - m'x_i - c'$ to get the variance,

$$\sigma'^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - m'x_i - c')^2$$

2. Find a MLE estimate for the Geometric Distribution probability p, where $P(x, p) = (1-p)^{x-1}p$.

Solution - $p' = \frac{n}{\sum x_i} = \frac{1}{\mu_x}$

3. Find a MLE estimator for θ for a sample size of n in the two sided exponential family with the pdf

$$f(x) = \frac{1}{2}e^{-|x-\theta|} \quad \forall\ x \in \mathcal{R}$$

Is this unbiased?

Solution- Use basic MLE derivation to obtain that x', the Median of the sample, is a MLE estimator of θ.
Reason using symmetry of the distribution and that we can take another sample mirrored around the true θ that the final bias will be 0.

4. Use MLE for normal distribution to estimate σ^2 while μ is known. What is the expected value of estimator?

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

2 Textbook Problems

Chapter 7, Problems 62, 63, 65 and Examples 7.8b, 7.8c, 7.8d