STA2112H: Probability I

Problem 1

Prove each of the following statements. Assume that any conditioning event has positive probability.

(a) If $P(B) = 1$, then $P(A|B) = P(A)$ for any A.

Solution

$$\begin{aligned}
P(B) = 1 &\implies P(B^c) = 1 - P(B) = 1 - 1 = 0 \\
P(B^c) = 0 &\implies P(B^c \cap A) = 0 \\
P(A) &= P(A \cap B) + P(A \cap B^c) = P(A \cap B) \\
P(A|B) &= \frac{P(A \cap B)}{P(B)} = P(A \cap B) = P(A)
\end{aligned}$$

■

(b) If $A \subset B$, then $P(B|A) = 1$ and $P(A|B) = P(A)/P(B)$.

Solution $A \subset B \iff x \in A \implies x \in B$. Therefore all elements of A are also in B. This implies that $A = A \cap B$. Then

$$\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1 \\
\text{and} \quad P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}.
\end{aligned}$$

■

(c) If A and B are mutually exclusive, then

$$P(A|A \cup B) = \frac{P(A)}{P(A) + P(B)}.$$

Solution

$$\begin{aligned}
A, B \text{ mutually exclusive} &\implies P(A \cap B) = 0 \\
& P(A \cup B) = P(A) + P(B)
\end{aligned}$$

$$\begin{aligned}
P(A|A \cup B) &= \frac{P(A \cap (A \cup B))}{P(A \cup B)} = \frac{P((A \cap A) \cup \overbrace{(A \cap B)}^{\text{the empty set}})}{P(A) + P(B)} \\
&= \frac{P(A)}{P(A) + P(B)}.
\end{aligned}$$

■

(e) $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$.

Solution

$P(A\cap B\cap C) = P(A\cap(B\cap C)) = P(A|(B\cap C))P(B\cap C) = P(A|(B\cap C))P(B|C)P(C)$.

$\blacksquare$

(d) If $P(A) = \frac{1}{2}$ and $P(B^c) = \frac{1}{4}$, can A and B be disjoint?

Solution No, because if A and B be disjoint then

$$P(A \cup B) = P(A) + P(B) = P(A) + (1 - P(B^c)) = \frac{1}{2} + \frac{3}{4} > 1.$$

More generally, if A and B are disjoint then $A \subset B^c$ so $P(A) \leq P(B^c)$. In this example, $P(A) > P(B^c)$.

Problem 2

Consider a clinical trial with two treatment groups and a binary outcome of success/failure. Assume the outcomes of different subjects are independent. In a 'play the winner' rule for assigning treatments, you randomly assign the first subject to either treatment group 1 or 2 with probability $\frac{1}{2}$. If that subject is a success, you assign the next subject to that same group, otherwise you assign the next subject to the other group. Suppose you continue assigning subjects in this manner indefinitely. Let $p_1 > 0$ be the probability of success in group 1 and $p_2 > 0$ be the probability of success in group 2.

(a) Suppose that the first subject is randomized to group 1. Let X be the random variable for number of subjects that are assigned to group 1 before switching over to group 2. What is the pmf of X?

Solution We may equivalently define X as the random variable for the number of trials until there is a failure in treatment group 1. The treatment for group 1 fails with probability $1 - p_1$. It then follows that $X \sim Geom(1 - p_1)$, which has pmf

$$p(X = k) = (p_1)^{k-1}(1 - p_1) \quad \text{for} \quad k = 1, 2...$$

(b) Suppose you are blinded to the first treatment group assignment, but you observe that the first n outcomes are successes. What is the probability that the first treatment assignment was group 1?

Solution Let A be the event that the first treatment assignment was group 1 and B be the event that the first n outcomes are successes. By Bayes' rule,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

From the problem statement we have, $P(B|A) = p_1^n$, $P(B|A^c) = p_2^n$, and $P(A) = 1/2$. It then follows that

$$P(A|B) = \frac{\frac{1}{2}(p_1^n)}{\frac{1}{2}(p_1^n + p_2^n)} = \frac{p_1^n}{p_1^n + p_2^n}$$

(c) Discuss a pro and a con of using the play the winner rule as the method of treatment allocation in a clinical trial as opposed to randomly choosing assignments.

Solution

Pro: More patients receive the more effective treatment.

Con: We can't start treatment for the next subject until we've observed the outcome of the previous subject which could be impractical if response times are slow.

Problem 3

Suppose you are making trail mix for a hike and have $n \geq 2$ unique ingredients to use. You decide to make the mix by picking an ingredient uniformly at random from each of the n ingredients n times. What is the most likely combination of ingredients you'll end up with?

Solution In this example, you are taking a uniform random sample of size n from the n ingredients with replacement. If we let X_i denote the random variable for the number of ingredients of type i for $i = 1, \ldots, n$, then $(X_1, \ldots, X_n)$ follows a Multinomial$(n, p_1, p_2, \ldots, p_n)$ where $p_i = 1/n$ for $i = 1, \ldots, n$. Using the definition of the multinomial distribution,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{n!}{x_1!x_2! \ldots x_n!} n^{-n}$$

This probability is maximized when $x_1 = x_2 = \cdots = x_n = 1$ so the most likely combination for your trail mix is the one that includes all of the ingredients.

Looking ahead: In a few weeks, we will learn about the bootstrap. The same calculations can be used to show that the most likely bootstrap resample is the original sample.

Problem 4

Suppose that $X_1, \dots, X_n$ are iid continuous random variables with cdf $F(x)$ and pdf $f(x)$. Define the random variables

$$U = min(X_1, \dots, X_n) \quad \text{and} \quad V = max(X_1, \dots, X_n)$$

.

a) Find the cdf of U.

Solution

$$\begin{aligned} P(U \leq u) &= P(min(X_1, \dots, X_n) \leq u) \\ &= 1 - P(min(X_1, \dots, X_n) \geq u) \\ &= 1 - P(X_1 \geq u, X_2 \geq u, \dots, X_n \geq u) \\ &= 1 - (P(X_1 \geq u)P(X_2 \geq u) \dots P(X_n \geq u)) \\ &= 1 - (1 - F(u))^n \end{aligned}$$

b) Find the cdf of V.

Solution

$$\begin{aligned} P(V \leq v) &= P(max(X_1, \dots, X_n) \leq v) \\ &= P(X_1 \leq v, \dots, X_n \leq v) \\ &= (P(X_1 \leq v)P(X_2 \leq v) \dots P(X_n \leq v)) \\ &= (F(v))^n \end{aligned}$$

c) Find the joint cdf of U and V.

Solution By definition $U \leq V$, so $f(u, v) = 0$ for $u > v$. When $u \leq v$, we have

$$\begin{aligned} P(U \leq u, V \leq v) &= P(V \leq v) - P(U > u, V \leq v) \\ &= P(V \leq v) - (P(u < X_1 \leq v)P(u < X_2 \leq v) \dots P(u < X_n \leq v)) \\ &= (F(v))^n - (F(v) - F(u))^n \end{aligned}$$

The first equality follows from the fact that

$$P(V \leq v) = P(V \leq v, U \leq u) + P(V \leq v, U > u)$$

.

Problem 5

Let X and Y be two random variables with mean 0 and variance 1. The correlation between X and Y is then given by

$$\rho = E(XY) = Cov(X,Y) = Cor(X,Y)$$

In the subsequent exercises, we will verify some of the properties of correlation and covariance we saw in class.

Part A

Find $E[(X-\rho Y)^2]$ and then use this result, together with the fact that $(X-\rho Y)^2$ is a nonnegative random variable, to show that $-1 \leq \rho \leq 1$.

Solution

$$\begin{aligned}
E[(X-\rho Y)^2] &= E(X^2 - 2XY\rho + Y^2\rho^2) \\
&= E(X^2) - 2\rho E(XY) + \rho^2 E(Y^2) \\
&= 1 - 2\rho^2 + \rho^2 = 1 - \rho^2
\end{aligned}$$

Since $(X-\rho Y)^2$ is nonnegative,

$$0 \leq E[(X-\rho Y)^2] = 1 - \rho^2$$

implying that

$$\rho^2 \leq 1 \rightarrow -1 \leq \rho \leq 1$$

Part B

Suppose that $Y = aX + b$ for some positive constants a and b. Find $Cor(X,Y)$.

Solution

$$\begin{aligned}
Cor(X,Y) &= \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \\
&= \frac{E[X(aX+b)] - E(X)E(aX+b)}{\sqrt{Var(X)Var(aX+b)}} \\
&= \frac{E(aX^2+bX) - E(X)E(aX+b)}{\sqrt{Var(X)Var(aX+b)}} \\
&= \frac{a}{\sqrt{a^2}} = 1
\end{aligned}$$

Part C

Using your result from Part A, show that X and Y are linearly dependent when $|\rho| = 1$.

Solution

From Part A, we have

$$E[(X - \rho Y)^2] = 1 - \rho^2$$

If $|\rho| = 1$,

$$E[(X - \rho Y)^2] = 0$$

Because $(X - \rho Y)^2$ is nonnegative, it follows that $(X - \rho Y)^2 = 0$ with probability 1 and therefore $X = \rho Y$. Since $|\rho| = 1$,

$$X = Y \quad \text{or} \quad X = -Y.$$

Part D

Let $U = X - Y$ and $V = X + Y$. Find $Cov(U, V)$ and $Cor(U, V)$.

Solution

$$\begin{aligned}
Cov(U, V) &= Cov(X - Y, X + Y) \\
&= E[(X - Y)(X + Y)] + E(X - Y)E(X + Y) \\
&= E[X^2 - Y^2] \\
&= E[X^2] - E[Y^2] = 0
\end{aligned}$$

Since $Cov(U, V) = 0$, $Cor(U, V) = 0$. This example shows that a correlation of zero between two random variables does not imply that they are independent. We also saw an example of this in class.