ECMA 31000: Solutions to Problem Set 1

Due October 7 by 11:59PM

1 Probability

Question 1 (Monty Hall) In the classical Monty Hall Problem, the contestant is faced with 3 doors, and must select one. Behind two of the doors is a goat and behind one door is a car. Monty, the host, knows which door contains the car. He behaves as follows: If the contestant selects the door with the car behind it, he picks one of the other doors each with probability 1/2. If the contestant selects a door with a goat behind it, Monty always selects the door with the other goat behind it. After Monty selects a door, he asks the contestant whether they would like to open the door they selected, or switch to the door neither of them selected. They win whatever is behind their final choice. Suppose the contestant always selects door 1 initially and that the car is equally likely to be behind any of the three doors. Independently of Monty's selection, the contestant switches their initial choice with probability p.

a) Write out the sample space, Ω, for the outcomes (Car position, Monty's Selection, Contestant Choice) and the corresponding probability mass function. e.g. $(1, 2, \text{Switch})$ represents the outcome that the car is behind door 1, Monty selects door 2, and the contestant switches. Outcomes that never occur may be omitted.

ANS:

$$\begin{aligned}\Omega = &\left\{(1,2,\text{Switch}),(1,3,\text{Switch}),(2,3,\text{Switch}),(3,2,\text{Switch})\right\}\\ &\cup\left\{(1,2,\text{Stay}),(1,3,\text{Stay}),(2,3,\text{Stay}),(3,2,\text{Stay})\right\}.\end{aligned}$$

We have

$$\begin{aligned}\mathrm{P}\left(\{(1,2,\text{Switch})\}\right) = \mathrm{P}\left(\{(1,3,\text{Switch})\}\right) &= \frac{1}{2}\times\frac{1}{3}\times p = \frac{p}{6},\\ \mathrm{P}\left(\{(2,3,\text{Switch})\}\right) = \mathrm{P}\left(\{(3,2,\text{Switch})\}\right) &= \frac{p}{3},\\ \mathrm{P}\left(\{(1,2,\text{Stay})\}\right) = \mathrm{P}\left(\{(1,3,\text{Stay})\}\right) &= \frac{1}{2}\times\frac{1}{3}\times(1-p) = \frac{1-p}{6},\\ \mathrm{P}\left(\{(2,3,\text{Stay})\}\right) = \mathrm{P}\left(\{(3,2,\text{Stay})\}\right) &= \frac{1-p}{3}.\end{aligned}$$

b) Write out the event corresponding to "Contestant wins the car" and compute its probability. If the contestant wants to maximize the probability of winning the car, what is the optimal choice of

p? Interpret your result.

ANS: $A = \{(2,3,\text{Switch}), (3,2,\text{Switch}), (1,2,\text{Stay}), (1,3,\text{Stay})\}$ with prob $\frac{2p}{3} + \frac{1-p}{3} = \frac{1+p}{3}$. Optimal choice sets $p = 1$. Since Monty always selects a door with a goat, switching reveals the car $2/3$ of the time, and not switching reveals the car $1/3$ of the time, so it is better to always switch.

c) (Malevolent Monty) Monty is upset because the contestant is using his door selection rule to get a better than fair chance of winning the car. Monty decides that in order to stop the contestant using his choice as information, he will ALWAYS pick door 3 when the car is behind door 1. Otherwise, he will pick the door containing the goat, as in the original game. This way, he thinks, his behaviour is the same as if the car were behind door 2, so that switching and not switching are now equally likely to produce a win for the contestant. How do your answers to a) and b) change? Did Monty succeed in reducing the probability the contestant wins the car? Why?

ANS:

$$\begin{aligned}\Omega = &\{(1,3,\text{Switch}), (2,3,\text{Switch}), (3,2,\text{Switch})\} \\ &\cup \{(1,3,\text{Stay}), (2,3,\text{Stay}), (3,2,\text{Stay})\}.\end{aligned}$$

and

$$\begin{aligned}\mathrm{P}(\{(1,3,\text{Switch})\}) = \mathrm{P}(\{(2,3,\text{Switch})\}) = \mathrm{P}(\{(3,2,\text{Switch})\}) &= \frac{p}{3}, \\ \mathrm{P}(\{(1,3,\text{Stay})\}) = \mathrm{P}(\{(2,3,\text{Stay})\}) = \mathrm{P}(\{(3,2,\text{Stay})\}) &= \frac{1-p}{3}.\end{aligned}$$

$$\begin{aligned}A &= \{(1,3,\text{Stay}), (2,3,\text{Switch}), (3,2,\text{Switch})\}, \\ \mathrm{P}(A) &= \frac{1+p}{3}.\end{aligned}$$

Monty failed because if he ever selects door 2 the car must be behind door 3. If Monty selects door 3, the contestant has a $\frac{1}{2}$ chance of winning by switching. So, a strategy of always switching is still optimal, and produces the same overall probability of winning the car.

Question 2 Fix a probability space $(\Omega, \mathcal{F}, \mathrm{P})$.

a) Show $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$ for any events $A, B \in \mathcal{F}$.

ANS: Note that $A \cup B = A \cup (B \cap A^c)$, where A and $B \cap A^c$ are disjoint events. Moreover, $B = (B \cap A) \cup (B \cap A^c)$, where $B \cap A$ and $B \cap A^c$ are disjoint events. Therefore, the definition of a probability measure allows us to write the following:

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B \cap A^c) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B),$$

where the second equality uses the fact that $\mathrm{P}(B) = \mathrm{P}(B \cap A) + \mathrm{P}(B \cap A^c)$.

b) Prove that if $A_1, A_2, \dots \in \mathcal{F}$ is any sequence of events, then $\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \lim_{k \to \infty} \mathrm{P}\left(\cup_{n=1}^{k} A_n\right)$.

Hint: Let $E_1 = A_1$, $E_2 = A_2 \backslash A_1$,... $E_k = A_k \backslash \left(\cup_{n=1}^{k-1} A_n \right)$. Show that the E_k are disjoint, and that $\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} E_n$ and $\cup_{n=1}^{k} A_n = \cup_{n=1}^{k} E_n$ for any k.

ANS: After showing the points in the hint, it follows that

$$\begin{aligned}
\mathrm{P}\left(\cup_{n=1}^{\infty} A_n\right) &= \mathrm{P}\left(\cup_{n=1}^{\infty} E_n\right) \\
\text{(Countable additivity)} &= \sum_{n=1}^{\infty} \mathrm{P}\left(E_n\right) \\
&= \lim_{k \to \infty} \sum_{n=1}^{k} \mathrm{P}\left(E_n\right) \\
&= \lim_{k \to \infty} \mathrm{P}\left(\cup_{n=1}^{k} E_n\right) \\
&= \lim_{k \to \infty} \mathrm{P}\left(\cup_{n=1}^{k} A_n\right).
\end{aligned}$$

c) Use your result to show that if $A_1, A_2, \ldots$ are independent events, $\mathrm{P}\left(\cap_{n=1}^{\infty} A_n\right) = \Pi_{n=1}^{\infty} \mathrm{P}\left(A_n\right)$.

Hint: De Morgan's laws provide: $\cap_{n=1}^{\infty} A_n = \left(\cup_{n=1}^{\infty} A_n^c\right)^c$. Now use part (b).

ANS: We get

$$\begin{aligned}
\mathrm{P}\left(\cap_{n=1}^{\infty} A_n\right) &= \mathrm{P}\left(\left(\cup_{n=1}^{\infty} A_n^c\right)^c\right) \\
&= 1 - \mathrm{P}\left(\cup_{n=1}^{\infty} A_n^c\right) \\
&= 1 - \lim_{k \to \infty} \mathrm{P}\left(\cup_{n=1}^{k} A_n^c\right) \\
&= 1 - \lim_{k \to \infty} \left[1 - \mathrm{P}\left(\cap_{n=1}^{k} A_n\right)\right] \\
&= \lim_{k \to \infty} \mathrm{P}\left(\cap_{n=1}^{k} A_n\right) \\
&= \lim_{k \to \infty} \Pi_{n=1}^{k} \mathrm{P}\left(A_n\right) \\
&= \Pi_{n=1}^{\infty} \mathrm{P}\left(A_n\right).
\end{aligned}$$

d) Your friend has an infinite amount of spare time. They plan to use it by throwing darts at a dartboard.. forever. The area of the board is 1, and the location of each of your friend's throws is uniformly distributed (denoted by P) across the board, which is labelled Ω. Each throw is independent of all others. They want to know if they can design a game with the following properties:

- For the $n-$th throw, there is a subset of the board A_n with $\mathrm{P}\left(A_n\right) < 1$ such that:
  - If the dart lands on a point $\omega \in A_n$, the game continues,
  - If the dart lands on a point $\omega \in \Omega \backslash A_n$, the game stops.
- The probability that the game never ends is neither 0 nor 1. (i.e. $\mathrm{P}\left(\cap_{n=1}^{\infty} A_n\right) \in (0, 1)$).

d(i) Show that if your friend chooses $A_n = A$ for all n, these properties are not satisfied.

ANS:

$$\mathrm{P}\left(\cap_{n=1}^{\infty} A_n\right) = \lim_{N\to\infty} \Pi_{n=1}^{N} \mathrm{P}\left(A_n\right) = \lim_{N\to\infty} \left[\mathrm{P}\left(A\right)\right]^N = \begin{cases} 1 & \text{if } \mathrm{P}\left(A\right) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

d(ii) Find a sequence A_n with $\mathrm{P}\left(A_n\right) < 1$ for each n satisfying the desired properties.

ANS: The region of the board that allows our friend to continue must get larger with n. If this happens quickly enough, there is a positive probability the game will never end. For example, let A_n be a region of the board with area $1 - \frac{1}{n^2}$ for $n \geq 2$. Then

$$\begin{aligned} \ln\left[\mathrm{P}\left(\cap_{n=2}^{\infty} A_n\right)\right] &= \ln\left[\Pi_{n=2}^{\infty} \mathrm{P}\left(A_n\right)\right] \\ &= \sum_{n=2}^{\infty} \ln\left(\frac{n^2-1}{n^2}\right) \\ &= \lim_{k\to\infty} \sum_{n=2}^{k} \ln\left(\frac{n^2-1}{n^2}\right) \\ \text{(This is a telescoping series)} &= \lim_{k\to\infty} \sum_{n=2}^{k} \ln\left(\frac{n+1}{n}\right) - \ln\left(\frac{n}{n-1}\right) \\ &= \lim_{k\to\infty} \ln\left(\frac{k+1}{k}\right) - \ln 2 \\ &= -\ln 2. \end{aligned}$$

Thus, if $\mathrm{P}\left(A_1\right) = a > 0$, then

$$\mathrm{P}\left(\cap_{n=1}^{\infty} A_n\right) = \mathrm{P}\left(A_1\right) \mathrm{P}\left(\cap_{n=2}^{\infty} A_n\right) = \frac{a}{2} \in (0, 1).$$

Note that if the area doesn't increase quickly enough, the property fails and we get $\mathrm{P}\left(\cap_{n=1}^{\infty} A_n\right) = 0$. For example, if A_n has area $1 - \frac{1}{n} = \frac{n-1}{n}$, then

$$\begin{aligned} \Pi_{n=2}^{N} \mathrm{P}\left(A_n\right) &= \Pi_{n=2}^{N} \left(\frac{n-1}{n}\right) \\ &= \frac{1}{N} \to 0. \end{aligned}$$

2 Distributions

Question 3 Suppose $X^2 = c$, where X is a random variable and c is positive constant. Under what condition is $Var\left(X\right) > 0$? Justify your answer using Jensen's inequality.

ANS: Note that $Var(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$. Since both $\mathrm{E}\left(X\right)$ and $\mathrm{E}\left(X^2\right)$ exist, and $f(x) = x^2$ is a strictly convex function, Jensen's Inequality tells us that:

$$\mathrm{E}(X^2) > \mathrm{E}(X)^2 \iff X \text{ is not constant w.p.1.}$$

This means that $\mathrm{P}(X = \sqrt{c}) \in (0,1)$ is the required condition, i.e. X takes both values $\sqrt{c}$ and $-\sqrt{c}$ with positive probability.

Question 4 Show that for a random vector X, constant matrix A, and vector of constants b, $Var(AX+b) = AVar(X)A'$.

ANS: Using the definition of variance, we obtain:

$$\begin{aligned}
Var(AX+b) &= \mathrm{E}\left[(AX+b-\mathrm{E}(AX+b))(AX+b-\mathrm{E}(AX+b))'\right] \\
&= \mathrm{E}\left[(AX-\mathrm{E}(AX))(AX-\mathrm{E}(AX))'\right] \\
&= \mathrm{E}\left[A(X-\mathrm{E}(X))(X-\mathrm{E}(X))'A'\right] \\
&= AVar(X)A'.
\end{aligned}$$

Question 5 Let $X = (X_1, X_2)'$ have bivariate CDF F_X, and X_1 have marginal CDF F_{X_1}. Show, using the result of Question 2b, that $\lim_{x_2\to\infty} F_X(x_1,x_2) = F_{X_1}(x_1)$.

ANS: Consider a sequence $\{x_{2,k}\}_k$ such that $\lim_{k\to\infty} x_{2,k} = +\infty$. Note that $\{X_1 \le x_1, X_2 \le x_{2,k}\}_{k\ge 1}$ forms a sequence of events such that

$$\bigcup_{k=1}^{\infty}\{X_1 \le x_1, X_2 \le x_{2,k}\} = \{X_1 \le x_1, X_2 \in \mathbb{R}\} = \{X_1 \le x_1\}.$$

Using the property derived in 2b,

$$\begin{aligned}
\lim_{x_2\to\infty} F_X(x_1,x_2) &= \lim_{n\to\infty} \mathrm{P}\left(\bigcup_{k=1}^{n}\{X_1 \le x_1, X_2 \le x_{2,k}\}\right) \\
&= \mathrm{P}\left(\bigcup_{k=1}^{\infty}\{X_1 \le x_1, X_2 \le x_{2,k}\}\right) \\
&= \mathrm{P}(X_1 \le x_1) = F_{X_1}(x_1).
\end{aligned}$$

Question 6 Show that $Var(Y) = \mathrm{E}(Var(Y|X)) + Var(E(Y|X))$. Explain intuitively why $Var(Y) \neq \mathrm{E}(Var(Y|X))$ using an example.

ANS: To prove the first claim, we write:

$$\begin{aligned}
\mathrm{E}(Var(Y|X)) + Var(E(Y|X)) &= \mathrm{E}\left(\mathrm{E}\left(Y^2|X\right) - [\mathrm{E}(Y|X)]^2\right) + Var(\mathrm{E}(Y|X)) \\
&= \mathrm{E}\left(Y^2\right) - \mathrm{E}\left([\mathrm{E}(Y|X)]^2\right) + \left[\mathrm{E}\left([\mathrm{E}(Y|X)]^2\right) - [\mathrm{E}(\mathrm{E}(Y|X))]^2\right] \\
&= \mathrm{E}\left(Y^2\right) - \mathrm{E}(Y)^2 \\
&= Var(Y),
\end{aligned}$$

where the first equality holds by definition of conditional variance, the second holds because $\mathrm{E}\left[\mathrm{E}\left(Y^2|X\right)\right] = \mathrm{E}\left(Y^2\right)$ (this is the law of iterated expectation), and the third holds by cancelling terms and noting that $\left[\mathrm{E}\left(\mathrm{E}\left(Y|X\right)\right)\right]^2 = \mathrm{E}\left(Y\right)^2$, again by the law of iterated expectation. Next, we consider the following example, which demonstrates that $Var\left(Y\right) \neq \mathrm{E}\left(Var\left(Y|X\right)\right)$. Suppose there are 2 countries, and that for a randomly drawn individual,

$$X = \begin{cases} 1 & \text{if individual from country A,} \\ 0 & \text{if individual from country B.} \end{cases}$$

where each country is equal in size, so $\mathrm{P}\left(X = 1\right) = \mathrm{P}\left(X = 0\right) = \frac{1}{2}$. Now suppose

$$Y|X = 0 \sim \mathcal{N}\left(175, 0.01\right); \qquad Y|X = 1 \sim \mathcal{N}\left(185, 0.05\right).$$

Both countries have vary little variation in height within country although the variation across countries is large. We find:

$$\mathrm{E}\left(Var\left(Y|X\right)\right) = \frac{1}{2} \times 0.01 + \frac{1}{2} \times 0.05 = 0.03,$$

which is small, while

$$\begin{aligned} Var\left(\mathrm{E}\left(Y|X\right)\right) &= \mathrm{E}\left[\left(\mathrm{E}\left(Y|X\right) - \mathrm{E}\left(Y\right)\right)^2\right] \\ &= \frac{1}{2} \times \left(185 - 180\right)^2 + \frac{1}{2} \times \left(175 - 180\right)^2 \\ &= 25, \end{aligned}$$

which is much larger. So, $Var\left(Y\right) = 25 + 0.03 = 25.03 \neq 0.03 = \mathrm{E}\left(Var\left(Y|X\right)\right)$.

Question 7 Show that if F is a known discrete distribution, we can use draws of $Unif[0,1]$ random variables to generate random variables with distribution F.

ANS: Since F is discrete, there is a countable set of values $\{x_j\}_{j\geq 1}$ such that $\mathrm{P}_X\left(X \in \{x_j\}_{j\geq 1}\right) = 1$. For $j \geq 1$, denote $\mathrm{P}_X\left(x_j\right) = p_j$. The pmf is given by

$$p\left(x\right) = \begin{cases} p_j & x = x_j, \\ 0 & \text{otherwise.} \end{cases}$$

Now suppose $Y \sim U\left[0,1\right]$ and let $a_0 = 0$, $a_j = \sum_{i\leq j} p_i$ for $j \geq 1$. Consider

$$g\left(Y\right) = \sum_{j\geq 1} x_j \mathbf{1}\left(Y \in [a_{j-1}, a_j)\right).$$

Then $g\left(Y\right) = x_j$ iff $Y \in [a_{j-1}, a_j)$, which occurs with probability p_j, meaning $g\left(Y\right)$ has distribution F. This construction partitions $[0,1]$ into segments of size p_j and sets the random variable equal to x_j if the uniform draw lies inside the corresponding segment. Note that for scalar random variables the generalized inverse transform $F^{-1}\left(Y\right)$ always produces a random variable with distribution F, where

$$F^{-1}(a) := \inf \{x \in \mathbb{R} | F(x) \geq a\}.$$

This was discussed for the special case of continuous distributions in the Week 1 lecture slides but in fact holds for any distribution. This function is also known as the quantile function, and satisfies $F^{-1}(a) \leq t \iff a \leq F(t)$. To see this, note that if $F^{-1}(a) \leq t$, then $\inf \{x \in \mathbb{R} | F(x) \geq a\} \leq t$. Since $F(x)$ is weakly increasing, $F(t) \geq a$ must hold. Conversely, if $F(t) \geq a$, then $t \in \{x | F(x) \geq a\}$, so $t \geq \inf \{x \in \mathbb{R} | F(x) \geq a\}$, or $t \geq F^{-1}(a)$. Therefore, if $Y \sim U[0,1]$, $F^{-1}(Y) \leq t \iff Y \leq F(t)$ and so

$$\mathrm{P}\left(F^{-1}(Y) \leq t\right) = \mathrm{P}(Y \leq F(t)) = F(t).$$

Question 8 Prove that if X is represented by density f_X, and $Y = u(X)$, where u is strictly monotone with inverse u^{-1}, then

$$f_Y(t) = f_X\left(u^{-1}(t)\right) \cdot \left|\frac{d}{dt}\left(u^{-1}(t)\right)\right|.$$

You may use the fact that $\frac{d}{dt}F_X(t) = f_X(t)$ when X is represented by f_X.

ANS: If u is strictly increasing,

$$\begin{aligned} F_Y(t) = \mathrm{P}(u(X) \leq t) &= \mathrm{P}\left(X \leq u^{-1}(t)\right) \\ &= F_X\left(u^{-1}(t)\right). \end{aligned}$$

Now take derivatives:

$$F_Y'(t) = f_X\left(u^{-1}(t)\right) \cdot \frac{d}{dt}\left(u^{-1}(t)\right).$$

If u is strictly decreasing,

$$\begin{aligned} F_Y(t) &= \mathrm{P}(u(X) \leq t) \\ &= \mathrm{P}\left(X \geq u^{-1}(t)\right) \\ \text{(because } X \text{ is continuous)} &= 1 - \mathrm{P}\left(X \leq u^{-1}(t)\right) \\ &= 1 - F_X\left(u^{-1}(t)\right). \end{aligned}$$

Taking derivatives yields

$$\frac{d}{dt}F_Y(t) = -f_X\left(u^{-1}(t)\right) \cdot \frac{d}{dt}\left(u^{-1}(t)\right).$$

Since $\frac{d}{dt}u^{-1}(t) = \frac{1}{u'(u^{-1}(t))}$, it has the same sign as $u'\left(u^{-1}(t)\right)$, so the result follows.