Math Problem Set #1: Probability and Statistics

OSM Lab - University of Chicago
Geoffrey Kocks

Problem 1. Book exercises from Chapter 3 of Humpherys and Jarvis.

HJ 3.6

We first note that $A = \bigcup_{i \in I}(A \cap B_i)$, where I is a finite or countable set. The set B_i completely covers the probability space and each element in the set $\{B_i\}$ is disjoint since $B_i \cap B_j = \emptyset$ for all $i \neq j$. Therefore $(A \cap B_i) \cap (A \cap B_j) = \emptyset$ as well for all $i \neq j$. Then by applying the proposition that the probability of a union of disjoint events is the total sum of their probabilities, we have that:

$$P(A) = P(\bigcup_{i \in I}(A \cap B_i)) = \sum_{i \in I}(P(A \cap B_i)).$$

HJ 3.8

We use the proposition that the probability of the converse of an event E is equal to $1 - P(E)$. Then:

$$P(\bigcup_{k=1}^{n}(E_k)) = 1 - P((\bigcup_{k=1}^{n}(E_k))^C) = 1 - P(\bigcap_{k=1}^{n}(E_k^C))$$

by the elementary property that the complement of a union of sets is equivalent to the intersection of each of their complements. Because each E_i is independent, we also use the property that each E_i^C is therefore independent as well, so the probability of their intersections is the product of each E_i. Therefore:

$$P(\bigcup_{k=1}^{n}(E_k)) = 1 - \prod_{k=1}^{n}(E_k^C) = 1 - \prod_{k=1}^{n}(1 - P(E_K)).$$

HJ 3.11

Let C denote the event in which the individual of sample s committed the crime and let P denote the event in which the individual of sample s tested positive in the DNA database. Then if the individual of sample s tested positive, the probability that they are guilty of the crime is $P(C \mid P)$. By Bayes' Law, this probability is given by:

$$P(C \mid P) = \frac{P(P \mid C)P(C)}{P(P)} = \frac{P(P \mid C)P(C)}{P(P|C)P(C) + P(P|notC)P(notC)}.$$

Based on the information provided, we assume that the test will give a positive result with certainty if the individual committed the crime. Then:

$$P(C \mid P) = \frac{P(C)}{P(C) + P(P|notC)P(notC)}.$$

$$P(C \mid P) = \frac{\frac{1}{250{,}000{,}000}}{\frac{1}{250{,}000{,}000} + (\frac{1}{3{,}000{,}000})(\frac{249{,}999{,}999}{250{,}000{,}000})}. \approx .0119.$$

Therefore in the absence of any other evidence for or against the person's guilt, the probability that they are guilty is approximately 0.0119.

HJ 3.12

We first show that in the Monty Hall game with three doors, the probability of winning is better if the contestant chooses to change to the other unopened door. Without a loss of generality, assume that the contestant initially chooses Door 1, since the car is equally as likely to be behind any of the three doors. By sticking with the original choice, the probability of winning a car is therefore 1/3. Monty will then choose Door i, where $i = 2$ or $i = 3$. Let C denote the event that the contestant initially chose the car and let I denote the event that Monty picks Door i. By Bayes' formula:

$$P(C \mid I) = \frac{P(I \mid C)P(C)}{P(I)} = \frac{(1/2)(1/3)}{(1/2)(1/3) + (1/3)(0) + (1/3)(1)} = 1/3$$

In the denominator, 1/2 is the probability that Monty will select Door i if we initially selected the car, since there are two doors available that are chosen at random. 0 represents the probability that Monty will select Door i if the car is behind Door i, and 1 represents the probability that Monty will select Door i if the car is behind the door other than Door 1 and Door i. Then the probability of winning the car by switching doors is 2/3.

By the same logic if there are 10 doors and Monty opens 8 of the doors, our initial probability selecting the door with the car is 1/10. By switching our door selection when given the chance, our probability of winning the car increases to 9/10 because we will always select the car unless we initially chose the door that contained the car (which occurred with a probability of just 1/10).

HJ 3.16

By definition, $Var[X] = E((X - \mu)^2)$. We expand the binomial to obtain:

$$Var[X] = E(X^2 - 2X\mu + \mu^2) = E(X^2) - 2E(X)\mu + \mu^2.$$

Since we define $\mu = E(X)$, this reduces to:

$$Var[X] = E(X^2) - 2(E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2.$$

We therefore have proved **Theorem 3.3.18** of Humpherys and Jarvis that:

$$\boxed{\mathrm{Var[X]} = \mathrm{E(X^2)} - (E(X))^2.}$$

HJ 3.33

We apply the Weak Law of Large Numbers and note that $E(\frac{B}{n}) = (1/n)E(B) = p$. We then directly apply the Law to determine that for any $\epsilon > 0$:

$$P(\mid \frac{B}{n} - p \mid \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}.$$

HJ 3.36

For each student admitted, there is a probability of 0.801 that they will accept the admissions offer. Therefore, each X_i has an expectation of 0.801 and a variance of $(0.801)(1-0.801)$. By the Central Limit Theorem:

$$P(\sum X_i \leq 5500) = P(\sum X_i - 6242(0.801) \leq 500).$$

$$P(\sum X_i \leq 5500) = P(\frac{\sum X_i - 6242(0.801)}{\sqrt{6242(0.801)(1-0.801)}} \leq \frac{500}{\sqrt{6242(0.801)(1-0.801)}}).$$

$$P(\sum X_i \leq 5500) \to \phi(15.85) \approx 1.$$

$\phi(x)$ denotes the cdf of the normal distribution. Therefore, the probability that the number of students enrolling will exceed 5500 is approximately 0 by the Central Limit Theorem.

Problem 2. Construct examples of events A, B, and C, each of probability strictly between 0 and 1, such that:

Part A. $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$, but $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

Solution. We roll a fair die two times. Define A as the event that we roll an odd number on the first roll. Define B as the event that we roll an even number on the second roll. Define C as the event that we roll the same number on both rolls. Then $P(A) = 1/2$, $P(B) = 1/2$, and $P(C) = 1/6$.

$P(A \cap C)$ refers to the probability that we roll the same odd number for both rolls and has probability $3/36 = 1/12$. $P(B \cap C)$ refers to the probability that we roll the same even number for both rolls and has probability $3/36 = 1/12$. $P(A \cap B)$ refers to the probability that we first roll an odd number and then roll and even number, which has probability $1/4$ because the rolls are independent.

However $P(A \cap B \cap C)$ refers to the probability that we first roll an odd number, and then roll an even number, but somehow roll the same number for both. This has a probability of 0, which is clearly not equal to the product of the three positive probabilities of the events.

Part B. $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(A \cap B \cap C) = P(A)P(B)P(C)$, but $P(B \cap C) \neq P(B)P(C)$.

Solution. We define 8 equally likely points $\{1, 2, 3, 4, 5, 6, 7, 8\}$. Define Event A as containing points $\{1, 2, 5, 6\}$, Event B as containing points $\{2, 3, 4, 6\}$, and Event C as containing points $\{5, 6, 7, 8\}$. Then $P(A) = P(B) = P(C) = 1/2$ and $P(A \cap B) =$

$P(A \cap C) = 1/4$. We also note that $P(A \cap B \cap C) = 1/8$. However $P(B \cap C) = 1/8 \neq 1/4$.

Problem 3. Prove that Benford's Law is, in fact, a well-defined discrete probability distribution.

Solution. We say that a distribution is a well-defined discrete probability distribution if: (i) the disjoint events that make up the distribution sum to 1 and (ii) each have a probability between 0 and 1 inclusive.

Benford's Law states that the probability of d being the first digit of a number is given by $P(d) = \log_{10}(1 + \frac{1}{d}), d \in \{1, 2, 3, ..., 9\}$. We first show that the sum of the probabilities is 1, and note that the event of a number beginning with each individual digit is disjoint. The sum of each of these events that determine the probability space is given by:

$$\sum_{d=1}^{9}(\log_{10}(1 + \frac{1}{d})) = \sum_{d=1}^{9}(\log_{10}(\frac{d+1}{d})) = \sum_{d=1}^{9}((\log_{10}(d+1) - \log_{10}(d)))$$

This telescoping sum reduces to $\log_{10}(10) - \log_{10}(1) = 1$. We then show that each event of digit being the leading digit of a number has a probability between 0 and 1 inclusive. The function $\log_{10}(1 + \frac{1}{d})$ is strictly decreasing in the set of digits, and has a maximum when $d = 1$ and $\log_{10}(2) < 1$. The minimum occurs when $d = 9$ and $\log_{10}(10/9) > 0$. Therefore Benford's Law is a well-defined discrete probability distribution.

Problem 4. A person tosses a fair coin until a tail appears for the first time. If the tail appears on the nth flip, the person wins 2^n dollars. Let the random variable X denote the player's winnings.

Part A. (St. Petersburg paradox) Show that $E[X] = +\infty$.

Solution. $E[X] = \sum_{i=1}^{\infty}(x_i p_i)$ where p_i denotes the probability that a tail appears for the first time on flip i and x_i denotes the winnings from obtaining a tail for the first time on flip i. We note that the distribution for the first appearance of a tail is a geometric distribution with $p = 1/2$ so this probability equals $(1/2)^i$.

$$E[X] = \sum_{i=1}^{\infty}(2^i(1/2)^i) = \sum_{i=1}^{\infty}(1^i) = +\infty$$

.

Part B. Suppose the agent has log utility. Calculate $E[\ln X]$.

Solution. The expectation with log utility is given by:

$$E[\ln X] = \sum_{i=1}^{\infty}((\ln(2^i))(1/2)^i) = (\ln(2))\sum_{i=1}^{\infty}(i(1/2)^i) = (2)(\ln(2)) = \ln(4).$$

(with the infinite sum computed on Wolfram Alpha)

Problem 5. Suppose the exchange rate between USD and CHF is 1:1. Both a U.S. investor and a Swiss investor believe that a year from now the exchange rate will be either 1.25 : 1 or 1 : 1.25, with each scenario having a probability of 0.5. Both investors want to maximize their wealth in their respective home currency (a year from now) by investing in a risk-free asset; the risk-free interest rates in the U.S. and in Switzerland are the same. Where should the two investors invest?

Solution. This paradox suggests that each of the investors should invest in the currency of the other country. From the perspective of the U.S. investor, there is a 0.5 probability that CHF will be worth 1.25/1 USD in the future and a 0.5 probability that it will be worth 1/1.25 USD.Therefore the expected value of purchasing 1 CHF today is given by: $(0.5)\frac{1.25}{1} + (0.5)\frac{1}{1.25} = \frac{41}{40}$ USD.The U.S. investor should then purchase CHF today. The argument is exactly symmetric for the Swiss investor, who should purchase USD today, with 1 USD having an expected value of $\frac{41}{40}$ CHF in the future.

Problem 6. Consider a probability measure space with $\Omega = [0, 1]$.

Part A. Construct a random variable X such that $E[X] < \infty$ but $E[X^2] = \infty$.

Solution. Define random variable X as a continuous random variable with probability desntiy function:

$$f(x) = \begin{cases} 2/x^3, 1 \leq x \leq \infty, \\ 0, otherwise \end{cases}$$

Then

$$E[X] = \int_1^\infty (2/x^2)dx = 2 < \infty.$$

However

$$E[X^2] = \int_1^\infty (2/x)dx = \infty.$$

Part B. Construct random variables X and Y such that $P(X > Y) > \frac{1}{2}$ but $E[X] < E[Y]$.

Solution. Define continuous random variable X with the following probability density function:

$$f(x) = \begin{cases} \frac{5}{59049}x^4, 0 \leq x \leq 9, \\ 0, otherwise \end{cases}$$

Define continuous random variable Y with the following probability density function:

$$f(y) = \begin{cases} 100, 7.5 \leq y \leq 7.51, \\ 0, otherwise \end{cases}$$

Then

$$P(X > Y) > P(X > 7.51) = \int_{7.51}^{9} (\frac{5}{59049}x^4)dx > 1/2.$$

However the expectations are given by:

$$E[X] = \int_{0}^{9} (\frac{5}{59049}x^5)dx = 7.5.$$

$$E[Y] = \frac{(7.5 + 7.51)}{2} = 7.505.$$

so $E[X] < E[Y]$.

Part C. Construct random variables X, Y, and Z such that $P(X > Y)P(Y > Z)P(X > Z) > 0$ and $E(X) = E(Y) = E(Z) = 0$.

Solution. Define random variable X with the following probability desntiy function:

$$f(x) = \begin{cases} 1/2, -1 \leq x \leq 1, \\ 0, otherwise \end{cases}$$

Define random variable Y with the following probability density function:

$$f(y) = \begin{cases} 1/4, -2 \leq y \leq 2, \\ 0, otherwise \end{cases}$$

Define discrete random variable Z with the following probability density function:

$$f(z) = \begin{cases} 1/6, -3 \leq z \leq 3, \\ 0, otherwise \end{cases}$$

Then $P(X > Y), P(Y > Z), P(X > Z)$ are all positive so their product is greater than 0, and each random variable has a mean of 0 because they are uniform distributions centered around 0.

Problem 7. Let the random variables X and Z be independent with $X \sim N(0, 1)$ and $P(Z = 1) = P(Z = -1) = \frac{1}{2}$. Define $Y = XZ$ as the product of X and Z. Prove or disprove each of the following statements.

Part A. $Y \sim N(0, 1)$.

Solution. True; The normal distribution is symmetric about its mean, so when there is a mean of 0, the distribution is still distributed as $N(0, 1)$ when multiplied by -1. In either case, Y is distributed as $N(0, 1)$ and the cases are independent.

Part B. $P(|X| = |Y|) = 1$.

Solution. True; It is always true that $X = Y$ or $X = -Y$. since Z always takes on a value of -1 or 1. In both of these cases, $(P(|X| = |Y|) = 1$.

Part C. X and Y are not independent.

Solution. True; If X and Y were independent, it would be the case that $P(Y) = P(Y|X)$. It is easy to come up with a counterexample where this does not hold. For example: $P(Y \leq -0.5) > 0$, while $P(Y \leq -0.5 | 0 \leq X \leq 0.25) = 0$. Therefore X and Y are not independent.

Part D. $Cov[X, Y] = 0$.

Solution. True; The covariance of X and Y is given by (since X and Z are independent):

$$Cov[XY] = E[XY] - E[X]E[Y] = E[X^2 Z] = E[X^2]E[Z] = 0.$$

Part E. If X and Y are normally distributed random variables with $Cov[X, Y] = 0$, then X and Y must be independent.

Solution. False; While it is true that independent random variables will always have a covariance of 0, the converse is not always true - as shown in Parts C and D.

Problem 8. Let the random variables X_i, $i = 1, 2, \ldots, n$, be i.i.d. having the uniform distribution on $[0, 1]$, denoted $X_i \sim U[0, 1]$. Consider the random variables $m = \min\{X_1, X_2, \ldots, X_n\}$ and $M = \max\{X_1, X_2, \ldots, X_n\}$. For both random variables m and M, derive their respective cumulative distribution (cdf), probability density function (pdf), and expected value.

Solution. We first consider the random variable $M = \max\{X_1, X_2, \ldots, X_n\}$. The cdf ($F(x)$) of M is given by:

$$F(x) = P(M \leq x) = P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x) = x^n.$$

since each X_i is independent and has a probability x of being less than or equal to x when distributed uniformly from 0 to 1. Then the pdf ($f(x)$) of M is given by:

$$f(x) = \frac{d}{dx}(x^n) = nx^{n-1}.$$

The expected value of M is given by:

$$E[M] = \int_0^1 (nx^n)dx = \frac{n}{n+1}x^{n+1}\Big|_{x=0}^1 = \frac{n}{n+1}.$$

Next we consider the random variable $m = \min\{X_1, X_2, \ldots, X_n\}$ The cdf ($G(x)$) of m is given by:

$$G(x) = P(m \leq x) = 1 - P(m \geq x) = 1 - (P(X_1 \geq x)P(X_2 \geq x)...P(X_n \geq x))$$

$$G(x) = 1 - (1 - x)^n.$$

Then the pdf ($g(x)$) of m is given by:

$$g(x) = \frac{d}{dx}(1 - (1 - x)^n) = n(1 - x)^{n-1}.$$

Finally, the expected value of m is given by:

$$E[m] = \int_0^1 (xn(1-x)^{n-1})dx = \frac{-(1-x)^n(nx+1)}{n+1}\Big|_{x=0}^{1} = \frac{1}{n+1}.$$

Problem 9. You want to simulate a dynamic economy (e.g., an OLG model) with two possible states in each period, a "good" state and a "bad" state. In each period, the probability of both shocks is $\frac{1}{2}$. Across periods the shocks are independent. Answer the following questions using the Central Limit Theorem and the Chebyshev Inequality.

Part A. What is the probability that the number of good states over 1000 periods differs from 500 by at most 2%?

Solution. If the number of good states over 1000 periods differs from 500 by at most 2% of 500, this means that the number of good states is between 490 and 510, inclusive. Let X_i be a random variable that equals 1 in a good state and 0 in a bad state. By the Central Limit Theorem:

$$P(\sum X_i \leq 510) = P(\sum X_i - 500 \leq 10).$$

$$P(\sum X_i \leq 510) = P(\frac{\sum X_i - 500}{\sqrt{1000(0.25)}} \leq \frac{500}{\sqrt{1000(0.25)}}).$$

$$P(\sum X_i \leq 510) \to \phi(0.632) \approx .7363.$$

Because the normal distribution is symmetric about its mean, $P(\sum X_i \leq 510) - 0.5 = P(500 \leq \sum X_i \leq 510) = P(490 \leq \sum X_i \leq 500)$. Therefore:

$$P(490 \leq \sum X_i \leq 510) = 2[\phi(0.632) - 0.5] \approx 0.4726.$$

Part B. Over how many periods do you need to simulate the economy to have a probability of at least 0.99 that the proportion of good states differs from $\frac{1}{2}$ by less than 1%?

Solution. The iid random variables X_i have a mean of $1/2$ and a variance of $1/4$. If the proportion of good states over 1000 periods differs from 0.5 by less than 1% of 0.5, then we are interested in being within 0.005 of the expected proportion of 0.5. By the Weak Law of Large Numbers (derived from the Chebyshev Inequality):

$$P(|\frac{\sum X_i}{n} - 0.5| \geq 0.005) \leq \frac{1}{4n(0.005^2)} \leq 0.01.$$

$$1 \leq 4n(0.005^2)(0.01)^{\cdot}$$

$$n \geq \frac{1}{4(0.01)(0.005^2)} = 1000000.$$

Therefore we would have to simulate the economy 1000000 times.

Problem 10. If $E[X] < 0$ and $\theta \neq 0$ is such that $E[e^{\theta X}] = 1$, prove that $\theta > 0$.

Solution. By Jensen's Inequality, if X is a random variable and f is a differentiable convex function, then $E[f(X)] \geq f(E[X])$. We note that $f(x) = e^{\theta X}$ is a differentiable convex function for all θ. Therefore:

$$E[e^{\theta X}] \geq e^{\theta E[X]}.$$

$$e^{\theta E[X]} \leq 1$$

$$\theta E[X] \leq 0$$

Because $E[X] < 0$, the final inequality will only hold if $\theta > 0$ since θ is nonzero.