CSE 312: Foundations of Computing II

Section 8: Variance, Important Discrete Distributions Solutions

0. What if we lose ?

Suppose 59 percent of voters favor Proposition 600. Use the Normal approximation to estimate the probability that a random sample of 100 voters will contain:

(a) at most 50 in favor. Mention any assumption that you make.

Solution:

We will make an assumption here. We will assume that the i^{th} person is in favor of the proposition with probability $\frac{59}{100}$. We define $X_i \sim$ Bernoulli$(\frac{59}{100})$ representing whether the i^{th} person is in favor or not. We define $X = \sum_{i=1}^{100} X_i$ representing the number of people who are in favor of the proposition. We can approximate X by $Y \sim N(100 \cdot 0.59, 100 \cdot 0.242)$. We need to find Pr$\left(\frac{Y-59}{\sqrt{(24.2)}} < \frac{50.5-59}{\sqrt{(24.2)}}\right)$(after continuity correction and standardization) which is equal to $\Phi(-1.729)$.

(b) more than 100 voters in favor or less than 0 voters in favor. Will the probability be non zero ?

Solution:

We will use our normal approximation Y from part(a). We are interested in Pr$(Y < 0.5)$+Pr$(Y > 100.5)$(after continuity correction) which is the same as

$$\text{Pr}\left(\frac{Y-59}{\sqrt{24.2}} < \frac{0.5-59}{\sqrt{24.2}}\right) + \text{Pr}\left(\frac{Y-59}{\sqrt{24.2}} > \frac{100.5-59}{\sqrt{24.2}}\right) = \Phi(-11.89) + 1 - \Phi(8.436)$$

. Yes, the probability will be non -zero because the density of the normal distribution is non-zero everywhere. Note that this result is acceptable because the normal distribution is an approximation.

1. By parts? !

Hopper loves exponentials, so he attempts to make a simple continuous distribution with an exponential. Hopper decides that the distribution will be of the form

$$\begin{cases} a \cdot e^x + b & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and that the distribution will have a mean equal to 0.5. Can hopper find such a distribution with $a \neq 0$?

Solution:

We know that $\int_0^1 a \cdot e^x + b \; dx = 1$. This implies that $\int_0^1 a \cdot e^x + b \; dx = a(e-1) + b = 1$. We also know that $\int_0^1 x \cdot (a \cdot e^x + b) dx = \frac{1}{2}$. Therefore, we have $\int_0^1 a \cdot x \cdot e^x + b \cdot x \; dx = \frac{b}{2} + a = \frac{1}{2}$. We have the equations $a(e-1) + b = 1$ and $\frac{b}{2} + a = \frac{1}{2}$ and we need to solve for a and b. By substituting for b, we get $-2a = a(1-e)$. So, $a = 0$ and $b = 1$. Unfortunately, Hopper cannot have the distribution of her choice.

2. Driving in Seattle

Yael's driving time to work is between 15 and 20 minutes if the day is sunny, and between 20 and 25 minutes if the day is rainy, with all times being equally likely in each case. Assume that a day is sunny with probability $\frac{1}{4}$ and rainy with probability $\frac{3}{4}$.

(a) What is the PDF of the driving time, viewed as a random variable X?

Solution:

Let the pdf in [15, 20] be a.Let the pdf in [20, 25] be b. Note that the pdf has to be zero everywhere else.
We know that $\int_{15}^{20} a \cdot dx = 5 \cdot a = \frac{1}{4}$ so $a = \frac{1}{20}$.
We know that $\int_{20}^{25} b \cdot dx = 5 \cdot b = \frac{3}{4}$ so $b = \frac{3}{20}$

(b) What is the CDF of the driving time?

Solution:

$$F_X(x) = \begin{cases} 0 & x < 15 \\ (x-15)\frac{1}{20} & 15 \le x \le 20 \\ (x-20)\frac{3}{20} + \frac{1}{4} & 20 \le x \le 25 \\ 1 & x > 25 \end{cases}$$

(c) What is $\mathbb{E}[X]$? You can leave your answer as an integral.

Solution:

$$\int_{15}^{20} x \cdot \frac{1}{20} dx + \int_{20}^{25} x \cdot \frac{3}{20} dx = \frac{170}{8}$$

(d) Find the variance of the driving time in two different ways. You can leave both answers as integrals.

Solution:

The first way :

$$\int_{15}^{20} x^2 \cdot \frac{1}{20} dx + \int_{20}^{25} x^2 \cdot \frac{3}{20} dx - \left(\frac{170}{8}\right)^2$$

The second way :

$$\int_{15}^{20} \left(x - \frac{170}{8}\right)^2 \cdot \frac{1}{20} dx + \int_{20}^{25} \left(x - \frac{170}{8}\right)^2 \cdot \frac{3}{20} dx$$

3. "Are we really twins ? " pmf asked pdf

Let $X \sim$ Uniform$(0,1)$ be a continuous uniform r.v.

(a) Find the CDF of X.

Solution:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

(b) Find the pdf of X^2. (Hint : sometimes, it is easier to calculate the CDF first.)

Solution:

Let $Y = X^2$. Let $f_Y(y)$ be the pdf of X^2 and let $F_Y(y)$ be the CDF of X^2. Note that $f(y) = \frac{d}{dy}F_Y(y)$. Now, $F_Y(y) = \Pr(X^2 \le y) = \Pr(X \le \sqrt{y}) = \sqrt{y}$ for $0 \le y \le 1$. So, $f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2\sqrt{y}}$ for $0 \le y \le 1$. Note that $f_Y(y) = 0$ for $y \notin [0,1]$.

(c) Was your answer in part(b) a uniform random variable ? If not, try to explain the result.

Solution:

Our solution was not a uniform random variable because our transformation was non-linear. The probability densities of X for x close to 0 got 'squeezed together' to give a high probability density of X^2 near zero and the probability densities of X for x close to 1 got 'stretched out' to give a low probability density of X^2 near 1. One way to analyze this is by noting that the derivative of $g(x) = x^2$ is equal to $2x$. The derivative implies that the larger the value of x, the same (change in x)= dx will correspond to a greater change in $g(x) = x^2$. In other words, the same densities corresponding to $[x, x + dx]$ for a larger x will be distributed across a larger range of X^2, reducing the density of X^2 over that range. This explains the downward sloping density function of X^2.

(d) Let $Y \sim$ Uniform(0, 1) which is independent of X. Use the analogies between continuous and discrete random variables to find $\Pr(X < Y)$. (Hint : The analogue of $\Pr(Y = k)$ is $p_Y(k) \cdot dy$.)

Solution:

We will use the law of total probability. The sums will turn into integrals and some probabilities will turn into probability densities.

$$\sum_y \Pr(X < Y \mid Y = y) \Pr(Y = y) < -- > \int_y \Pr(X < Y \mid Y = y)\, f_Y(y)dy$$

So, we have

$$\begin{aligned}\Pr(X < Y) &= \int_0^1 \Pr(X < Y \mid Y = y)\, f_Y(y)dy \\ &= \int_0^1 ydy \\ &= \frac{1}{2}\end{aligned}$$

4. I want views !!! :(

Sharpnel has three videos on youtube called A, B, and C. Video A receives 10 views per month on average, video B receives 15 views per month on average, and video C receives 25 views per month on average. Every view is received independently of any other view.

(a) What is the probability that Sharpnel receives at least a million views next month and becomes very rich ? You can leave your answer as a summation.

Solution:

The number of views received in a month can be modeled by $Y \sim$ Poisson$(15 + 10 + 25)$. So, $\Pr(Y \geq 1,000,000) = \sum_{k=1,000,000}^{\infty} e^{-50} \cdot \frac{(-50)^k}{k!}$

(b) What is the probability that Sharpnel only receive views on video A this month ?

Solution:

Let X_A, X_B, and X_C be random variables modeling the number of views received on websites A, B, and C respectively. We are interested in $\Pr((X_A > 0) \cap (X_B = 0) \cap (X_C = 0)) = \Pr(X_A > 0) \cdot \Pr(X_B = 0) \cdot \Pr(X_C = 0)$ (independence). Note that $X_A \sim$ Poisson(10), $X_B \sim$ Poisson(15), and $X_C \sim$ Poisson(25). So, the concerned probability is equal to

$$(1 - e^{-10}) \cdot (e^{-15}) \cdot (e^{-25})$$

.

(c) What is the expected time Sharpnel has to wait for until he receives the next view?

Solution:

The waiting time for the next view can be modeled by an exponential random variable $Y \sim$ Exponential(50). So, $\mathbb{E}[Y] = \frac{1}{50} =$ 0.02 months $=$ 0.62 days.

5. Servers

Your company has 5000 servers. It costs a lot of money to run servers. From last year's data, you know $x_i =$ "money the i^{th} server cost you last year". Assume each server's cost is independent of any other server's cost.

(a) Using the law of large numbers, find an estimator for the expected cost per server.

Solution:

We assume that each server incurs a cost based on a common probability distribution each year. So, let $X_i =$ "cost incurred by the i^{th} computer in an year". So, $X = \frac{1}{5000}\sum_i^{5000} X_i$ represents the average cost per server. By the law of large numbers, we know that the Var(X) tends to zero as the number of servers increases. So, $\frac{1}{5000} \cdot (x_1 + x_2 + ... + x_{5000})$, will give us a good estimator of μ, the expected cost per server.

(b) Using the law of large numbers and the result from part (a), find an estimator for the variance per server. (Hint : Your first instinct is likely to be correct. Think about what you would do if you didn't know about the Law Of Large Numbers)

Solution:

Assume the expected cost per server from part(a) is μ. Let $V_i = (X_i - \mu)^2$ represent the variation of of the cost of the i^{th} server from the mean. Let $V = \frac{1}{5000} \cdot \sum_{i=1}^{1000} V_i$. Note that $E[V] = \sigma^2$, the variance of X_i. So, by the law of large numbers,

$$\frac{1}{5000} \cdot \left((x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + ... + (x_{5000} - \mu)^2\right)$$

will give us a good estimator of σ^2.

(c) Using a normal distribution and our results from part(a) and part(b), approximate the probability that the total cost will exceed C. Your answer will not be a number; it will be an expression in terms or ϕ.

Solution:

The total cost can be approximated by $N(5000 \cdot \mu, \sigma^2 \cdot 5000)$. We need Pr$(N > C)$. Note that Pr$(N > C) =$ Pr$\left(\frac{N - 5000\cdot\mu}{\sigma\cdot\sqrt{5000}} > \frac{C - 5000\cdot\mu}{\sigma\cdot\sqrt{5000}}\right)$. Hence, the probability we are look for is

$$1 - \Phi\left(\frac{C - 5000 \cdot \mu}{\sigma \cdot \sqrt{5000}}\right)$$

6. A square dartboard ?

You throw a dart at an $s \times s$ square dartboard. The goal of this game is to get the dart to land as close to the lower left corner of the dartboard as possible. However, your aim is such that the dart is equally likely to land at any point on the dartboard. Let random variable X be the length of the side of the smallest square B in the lower left corner of the dartboard that contains the point where the dart lands. That is, the lower left corner of B must be the same point as the lower left corner of the dartboard, and the dart lands somewhere along the upper or right edge of B. For X, find the CDF, PDF, $\mathbb{E}[X]$, and Var(X).

Solution:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ x^2/s^2, & \text{if } 0 \leq x \leq s \\ 1, & \text{if } x > s \end{cases}$$

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 2x/s^2, & \text{if } 0 \leq x \leq s \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \int_0^s x f_X(x) dx = \int_0^s \frac{2x^2}{s^2} dx = \frac{2}{s^2} \int_0^s x^2 dx = \frac{2}{3s^2} \left[x^3\right]_0^s = \frac{2}{3} s$$

$$\mathbb{E}\left[X^2\right] = \int_0^s x^2 f_X(x) dx = \int_0^s \frac{2x^3}{s^2} dx = \frac{2}{s^2} \int_0^s x^3 dx = \frac{1}{2s^2} \left[x^4\right]_0^s = \frac{1}{2} s^2$$

$$\mathsf{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = s^2 - \left(\frac{2}{3} s\right)^2 = \frac{1}{18} s^2$$