CSE 312: Foundations of Computing II

Section 6: Continuous Random Variables

1. Review of Main Concepts

(a) **Multivariate: Discrete to Continuous:**

	Discrete	**Continuous**
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$
Joint range/support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x, s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Independence must have	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

(b) **Law of Total Probability (r.v. version):** If X is a discrete random variable, then

$$\mathbb{P}(A) = \sum_{x \in \Omega_X} \mathbb{P}(A|X = x) p_X(x) \qquad \text{discrete } X$$

(c) **Law of Total Expectation (Event Version)**: Let X be a discrete random variable, and let events $A_1, ..., A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \, \mathbb{P}(A_i)$$

(d) **Conditional Expectation**: See table bbelow. Note that linearity of expectation still applies to conditional expectation: $\mathbb{E}[X + Y \mid A] = \mathbb{E}[X \mid A] + \mathbb{E}[Y \mid A]$

(e) **Law of Total Expectation (r.v. Version)**: Suppose X and Y are random variables. Then,

$$\mathbb{E}[X] = \sum_{y} \mathbb{E}[X \mid Y = y] \, p_Y(y) \qquad \text{discrete version.}$$

(f) **Conditional distributions**

	Discrete	**Continuous**
Conditional PMF/PDF	$p_{X\|Y}(x\|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\|Y}(x\|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$\mathbb{E}[X \mid Y = y] = \sum_x x p_{X\|Y}(x\|y)$	$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X\|Y}(x\|y) dx$

(g)

- Law of Total Probability (continuous)

$$\mathbb{P}(A) = \int_{x \in \Omega_X} \mathbb{P}(A|X = x) f_X(x) dx$$

- Law of total expectation (continuous)

$$\mathbb{E}[X] = \int_{y \in \Omega_Y} \mathbb{E}[X \mid Y = y] \, f_Y(y) dy$$

Markov's Inequality: Let X be a non-negative random variable, and $\alpha > 0$. Then, $\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[-NoValue-][X]}{\alpha}$.

Chebyshev's Inequality: Suppose Y is a random variable with $\mathbb{E}[-NoValue-][Y] = \mu$ and $\text{Var}(Y) = \sigma^2$. Then, for any $\alpha > 0$, $\mathbb{P}(|Y - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$

Chernoff Bound (for the Binomial): Suppose $X \sim \text{Bin}(n, p)$ and $\mu = np$. Then, for any $0 < \delta < 1$

2. Zoo of Continuous Random Variables

(a) **Uniform**: $X \sim \text{Uniform}(a, b)$ iff X has the following probability density function:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$. This represents each real number from $[a, b]$ to be equally likely.

(b) **Exponential**: $X \sim \text{Exponential}(\lambda)$ iff X has the following probability density function:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$. The exponential random variable is the continuous analog of the geometric random variable: it represents the waiting time to the next event, where $\lambda > 0$ is the average number of events per unit time. Note that the exponential measures how much time passes until the next event (any real number, continuous), whereas the Poisson measures how many events occur in a unit of time (nonnegative integer, discrete). The exponential random variable X is memoryless:

$$\text{for any } s, t \geq 0,\ \mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

The geometric random variable also has this property.

3. Create the distribution

Suppose X is a continuous random variable that is uniform on $[0, 1]$ and uniform on $[1, 2]$, but

$$\mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X < 1).$$

Outside of $[0, 2]$ the density is 0. What is the PDF and CDF of X?

Solution:

The fact that X is uniform on each of the intervals means that its PDF is constant on each. So,

$$f_X(x) = \begin{cases} c & 0 < x \leq 1 \\ d & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Note that $F_X(1) - F_X(0) = c$ and $F_X(2) - F_X(1) = d$. The area under the PDF must be 1, so

$$1 = F_X(2) - F_X(0) = F_X(2) - F_X(1) + F_X(1) - F_X(0) = d + c$$

Additionally,

$$d = F_X(2) - F_X(1) = \mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X \leq 1) = 2 \cdot (F_X(1) - F_X(0)) = 2c$$

To solve for c and d in our PDF, we need only solve the system of two equations from above: $d+c=1$, $d=2c$. So, $d=\frac{2}{3}$ and $c=\frac{1}{3}$. Taking the integral of the PDF yields the CDF, which looks like

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{3}x & 0 < x \leq 1 \\ \frac{2}{3}x - \frac{1}{3} & 1 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

4. Max of uniforms

Let $U_1, U_2, \ldots, U_n$ be mutually independent Uniform random variables on $(0,1)$. Find the CDF and PDF for the random variable $Z = \max(U_1, \ldots, U_n)$.

Solution:

The key idea for solving this question is realizing that the max of n numbers $\max(a_1, ..., a_n)$ is less than some constant c, if and only if each individual number is less than that constant c (i.e. $a_i < c$ for all i). Using this idea, we get

$$\begin{aligned} F_Z(x) = \mathbb{P}(Z \leq x) &= \mathbb{P}(\max(U_1, ..., U_n) \leq x) \\ &= \mathbb{P}(U_1 \leq x, ..., U_n \leq x) \\ &= \mathbb{P}(U_1 \leq x) \cdot ... \cdot \mathbb{P}(U_n \leq x) && \text{[independence]} \\ &= F_{U_1}(x) \cdot ... \cdot F_{U_n}(x) \\ &= F_U(x)^n && \text{[where } U \sim \text{Unif}(0,1)] \end{aligned}$$

So the CDF of Z is

$$F_Z(x) = \begin{cases} 0 & x < 0 \\ x^n & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

To find the PDF, we take the derivative of each part of the CDF, which gives us the following

$$f_Z(x) = \begin{cases} n\, x^{n-1} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

5. Batteries and exponential distributions

Let X_1, X_2 be independent exponential random variables, where X_i has parameter λ_i, for $1 \leq i \leq 2$. Let $Y = \min(X_1, X_2)$.

(a) Show that Y is an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$. Hint: Start by computing $\mathbb{P}(Y > y)$. Two random variables with the same CDF have the same pdf. Why?

Solution:

We start with computing $\mathbb{P}(Y > y)$, by substituting in the definition of Y.

$$\mathbb{P}(Y > y) = \mathbb{P}(\min\{X_1, X_2\} > y)$$

The probability that the minimum of two values is above a value is the chance that both of them are above that value. From there, we can separate them further because X_1 and X_2 are independent.

$$\mathbb{P}(X_1 > y \cap X_2 > y) = \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = e^{-\lambda_1 y} e^{-\lambda_2 y}$$

$$= e^{-(\lambda_1+\lambda_2)y} = e^{-\lambda y}$$

So $F_Y(y) = 1 - \mathbb{P}(Y > y) = 1 - e^{-\lambda y}$ and $f_Y(y) = \lambda e^{-\lambda y}$ so $Y \sim \text{Exp}(\lambda)$, since this is the same CDF and PDF as an exponential distribution with parameter $\lambda = \lambda_1 + \lambda_2$.

(b) What is $Pr(X_1 < X_2)$? (Use the law of total probability.) The law of total probability hasn't been covered in class yet, but will be soon at which point it would be good to revisit this problem!

Solution:

By the law of total probability,

$$\mathbb{P}(X_1 < X_2) = \int_0^\infty \mathbb{P}(X_1 < X_2 | X_1 = x) f_{X_1}(x) dx = \int_0^\infty \mathbb{P}(X_2 > x)\lambda_1 e^{-\lambda_1 x} dx =$$

$$\int_0^\infty e^{-\lambda_2 x}\lambda_1 e^{-\lambda_1 x} dx = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

(c) You have a digital camera that requires two batteries to operate. You purchase n batteries, labelled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ, independently of all other batteries. Initially, you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process, you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

Solution:

Let T be the time until the end of the process. We are trying to find $\mathbb{E}[T]$. $T = Y_1 + ... + Y_{n-1}$ where Y_i is the time until we have to replace a battery from the ith pair. The reason it there are only $n-1$ RVs in the sum is because there are $n-1$ times where we have two batteries and wait for one to fail. By part (a), the time for one to fail is the min of exponentials, so $Y_i \sim Exp(2\lambda)$. Hence the expected time for the first battery to fail is $\frac{1}{2\lambda}$. By linearity and memorylessness, $\mathbb{E}[T] = \sum_{i=1}^{n-1} E[Y_1] = \frac{n-1}{2\lambda}$.

(d) In the scenario of the previous part, what is the probability that battery i is the last remaining battery as a function of i? (You might want to use the memoryless property of the exponential distribution that has been discussed.)

Solution:

If there are two batteries i, j in the flashlight, by part (b), the probability each outlasts each other is $1/2$. Hence, the last battery n has probability $1/2$ of being the last one remaining. The second to last battery $n-1$ has to beat out the previous battery and the n^{th}, so the probability it lasts the longest is $(1/2)^2 = 1/4$. Work down inductively to get that the probability the i^{th} is the last remaining is $(1/2)^{n-i+1}$ for $i \geq 3$. Finally the first two batteries share the remaining probability as they start at the same time, with probability $(1/2)^{n-1}$ each.

6. Continuous joint density I

The joint probability density function of X and Y is given by

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) & 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Verify that this is indeed a joint density function.

(b) Compute the marginal density function of X.

(c) Find $Pr(X > Y)$. (Uses the continuous law of total probability which we have not covered in class as of 11/17.)

(d) Find $P(Y > \frac{1}{2}|X < \frac{1}{2})$.

(e) Find $E(X)$.

(f) Find $E(Y)$

Solution:

(a) A joint density function will integrate to 1 over all possible values. Thus, we integrate over the joint range range using Wolfram Alpha, and see that:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_0^2\int_0^1 \frac{6}{7}(x^2+\frac{xy}{2})dxdy = 1$$

We also need to check that the density is nonnegative, but that is easily seen to be true.

(b) We apply the definition of the marginal density function of X, using the fact that we only need to integrate over the values where the joint density is positive:

$$f_X(x) = \begin{cases} \int_0^2 \frac{6}{7}(x^2+\frac{xy}{2})dy = \frac{6}{7}x(2x+1) & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) First, we rearrange our initial probability. Then, by the continuous law of total probability:

$$\mathbb{P}(X > Y) = 1 - \mathbb{P}(X \le Y) = 1 - \int_{-\infty}^{\infty} \mathbb{P}(X \le Y|Y = y) f_Y(y)dy = 1 - \int_{-\infty}^{\infty} \mathbb{P}(X \le y) f_Y(y)dy$$

Once again, we can instead integrate over just the range of y, getting:

$$1 - \int_0^2 \mathbb{P}(X \le y) f_Y(y)dy$$

We have to remember that $f_X(x)$ is positive only when $0 < x < 1$. Thus, $F_X(x) = 1$ for $x \ge 1$, so we have:

$$1 - \int_0^1 \mathbb{P}(X \le y) f_Y(y)dy - \int_1^2 f_Y(y)dy$$

So, now we just need to find the CDF of X, and the marginal PDF of Y. For the former, for any $0 < x < 1$, we have

$$F_X(x) = \int_0^x \frac{6}{7}u(2u+1)du = \frac{1}{7}x^2(4x+3)$$

For the latter, for $0 < y < 2$, we have

$$f_Y(y) = \int_0^1 \frac{6}{7}(x^2+\frac{xy}{2})dx = \frac{1}{14}(3y+4)$$

Putting these together, we get that:

$$\mathbb{P}(X > Y) = 1 - \int_0^1 \frac{1}{7}y^2(4y+3)\frac{1}{14}(3y+4)dy - \int_1^2 \frac{1}{14}(3y+4)dy = 1 - \frac{253}{1960} - \frac{17}{28} = \frac{517}{1960} = 0.2638$$

(d) By the definition of conditional probability:

$$\mathbb{P}\left(Y > \frac{1}{2}|X < \frac{1}{2}\right) = \frac{\mathbb{P}(Y > \frac{1}{2},\ X < \frac{1}{2})}{\mathbb{P}(X < \frac{1}{2})}$$

For the numerator, we have

$$\mathbb{P}(Y > \frac{1}{2}, X < \frac{1}{2}) = \int_{1/2}^{\infty}\int_{-\infty}^{1/2} f_{X,Y}(x,y)dxdy$$

$$= \int_{1/2}^{2}\int_{0}^{1/2} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)dxdy = \frac{69}{448}$$

For the denominator, we can integrate using the marginal distribution that we found before:

$$\int_0^{1/2} \frac{6}{7}x(2x+1)dx = \frac{5}{28}$$

Putting these together, we get:

$$\mathbb{P}(Y > \frac{1}{2}|X < \frac{1}{2}) = \frac{\frac{69}{448}}{\frac{5}{28}} = 0.8625$$

(e) By definition, and using $\Omega_X = (0,1)$:

$$\mathbb{E}[X] = \int_0^1 f_X(x)xdx = \int_0^1 \frac{6}{7}x(2x+1)xdx = \frac{5}{7}$$

(f) By definition, and using $\Omega_Y = (0,2)$:

$$\mathbb{E}[Y] = \int_0^2 f_Y(y)ydy = \int_0^2 \frac{1}{14}(3y+4)ydy = \frac{8}{7}$$

7. Continuous joint density II

The joint density of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of W and V is given by

$$f_{W,V}(w,v) = \begin{cases} 2 & 0 < w < v, 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent? Are W and V independent?

Solution:

For two random variables X, Y to be independent, we must have $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all $x \in \Omega_X,\ y \in \Omega_X$. Let's start with X and Y by finding their marginal PDFs. By definition, and using the fact that the joint PDF is 0 outside of $y > 0$, we get:

$$f_X(x) = \int_0^{\infty} xe^{-(x+y)}dy = e^{-x}x$$

We do the same to get the PDF of Y, again over the range $x > 0$:

$$f_Y(y) = \int_0^\infty xe^{-(x+y)}dx = e^{-y}$$

Since $e^{-x}x \cdot e^{-y} = xe^{-x-y} = xe^{-(x+y)}$ for all $x, y > 0$, X and Y are independent.

We can see that W and V are not independent simply by observing that $\Omega_W = (0, 1)$ and $\Omega_V = (0, 1)$, but $\Omega_{W,V}$ is not equal to their Cartesian product. Specifically, looking at their range of $f_{W,V}(w, v)$. Graphing it with w as the "x-axis" and v as the "y-axis", we see that The shaded area is where the joint pdf is strictly positive. Looking at it, we can see that it is not rectangular, and therefore it is not the case that $\Omega_{W,V} = \Omega_W \times \Omega_V$. Remember, the joint range being the Cartesian product of the marginal ranges is not sufficient for independence, but it is *necessary*. Therefore, this is enough to show that they are not independent.

8. Variance of the geometric distribution

Independent trials each resulting in a success with probability p are successively performed. Let N be the time of the first success. Find the variance of N.

Solution:

Let $Y = 1$ if the first trial results in a success and $Y = 0$ otherwise. Now

$$\text{Var}(N) = \mathbb{E}[N^2] - (\mathbb{E}[N])^2$$

To calculate $\mathbb{E}[N^2]$, we condition on Y as follows:

$$\mathbb{E}[N^2] = \mathbb{E}[\mathbb{E}[N^2|Y]]$$

However,

$$\mathbb{E}[N^2|Y = 1] = 1$$
$$\mathbb{E}[N^2|Y = 0] = \mathbb{E}[(1 + N)^2]$$

These two equations follow because, if the first trial results in a success, then clearly $N = 1$ and so $N^2 = 1$. On the other hand, if the first trial results in a failure, then the total number of trials necessary for the first success will have the same distribution as one (the first trial that results in failure) plus the necessary number of additional trials. Since the latter quantity has the same distribution as N, we obtain that $\mathbb{E}[N^2|Y = 0] = \mathbb{E}[(1 + N)^2]$. Hence we see that

$$\begin{aligned}
\mathbb{E}[N^2] &= \mathbb{E}[N^2|Y = 1]\mathbb{P}(Y = 1) + \mathbb{E}[N^2|Y = 0]\mathbb{P}(Y = 0) \\
&= p + (1 - p)\mathbb{E}[(1 + N)^2] \\
&= 1 + (1 - p)\mathbb{E}[2N + N^2]
\end{aligned}$$

Since we know that the expectation of a geometric random variable is given as $\mathbb{E}[N] = \frac{1}{p}$, by the Linearity of Expectation, we then have that

$$\begin{aligned}
\mathbb{E}[N^2] &= 1 + 2(1 - p)\mathbb{E}[N] + (1 - p)\mathbb{E}[N^2] \\
&= 1 + \frac{2(1 - p)}{p} + (1 - p)\mathbb{E}[N^2] \\
\mathbb{E}[N^2] - (1 - p)\mathbb{E}[N^2] &= \frac{2 - p}{p} \\
\mathbb{E}[N^2] &= \frac{2 - p}{p^2}
\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Var}(N) &= \mathbb{E}[N^2] - (\mathbb{E}[N])^2 \\ &= \frac{2-p}{p^2} - \frac{1}{p^2} \\ &= \frac{1-p}{p^2}\end{aligned}$$

9. 3 points on a line

(This problem uses the continuous law of total probability which has not yet be covered in class.) Three points X_1, X_2, X_3 are selected at random on a line L (continuous independent uniform distributions). What is the probability that X_2 lies between X_1 and X_3?

Solution:

Let $X_1, X_2, X_3 \sim Unif(0,1)$.

$$\begin{aligned}\mathbb{P}(X_1 < X_2 < X_3) &= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < X_2 < X_3 \mid X_2 = x)\, f_{X_2}(x)\, dx && \text{Continuous LoTP} \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < x, X_3 > x)\, f_{X_2}(x)\, dx && \text{Independence of } X_1, X_2, X_3 \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X_1 < x)\, \mathbb{P}(x < X_3)\, f_{X_2}(x)\, dx && \text{Independence of } X_1, X_3 \\ &= \int_{-\infty}^{\infty} F_{X_1}(x)\, (1 - F_{X_3}(x))\, f_{X_2}(x)\, dx \\ &= \int_0^1 x\, (1-x)\, 1\, dx \\ &= \frac{x^2}{2} - \frac{x^3}{3}\Big|_0^1 = \frac{1}{6}\end{aligned}$$

10. In between

(Covers ideas that have not been covered in class.) Suppose that X_1 and X_2 are discrete uniform random variables in $\{1, \ldots, 2n\}$ (i.e., X_1 and X_2 are equally likely to take any of the values $1, \ldots, 2n$) and let $Y = \min(X_1, X_2)$. What is the conditional pmf $p_{Y|X_1}(y \mid x_1)$ and conditional CDF $F_{Y|X_1}(y \mid x_1)$. What is $E[Y \mid X_1 = x_1]$? (For the definitions of conditional pmf, conditional CDF, see the review at the top of this worksheet.)

Solution:

The conditional pmf is

$$p_{Y|X_1}(y \mid x_1) = \mathbb{P}(\min(X_1, X_2) = y \mid X_1 = x_1) \quad = \frac{\mathbb{P}(\min(X_1, X_2) = y, X_1 = x_1)}{\mathbb{P}(X_1 = x_1)} = \begin{cases} 0 & \text{if } y > x_1 \\ 1 - \frac{x_1 - 1}{2n} & \text{if } y = x_1 \\ \frac{1}{2n} & \text{if } 1 \le y < x_1 \end{cases}$$

Explanation:

- Since $\min(X_1, X_2) \le X_1$, if $y > x_1$, then we have

$$\mathbb{P}(\min(X_1, X_2) = y, X_1 = x_1) = 0$$

- When $y = x_1$, we have

$$\mathbb{P}(\min(X_1, X_2) = x_1 \mid X_1 = x_1) = \mathbb{P}(X_2 \ge x_1 \mid X_1 = x_1) = \mathbb{P}(X_2 \ge x_1) = \frac{2n - x_1 + 1}{2n} = 1 - \frac{x_1 - 1}{2n}$$

  The second to last equality holds because X_1 and X_2 are independent.

- In the case where $y < x_1$, $\min(X_1, X_2) = X_2$. So for $y < x_1$,

$$\mathbb{P}(Y = y \mid X_1 = x_1) = \mathbb{P}(X_2 = y \mid X_1 = x_1) = \mathbb{P}(X_2 = y) = \frac{1}{2n}.$$

  Again, we used the independence of X_1 and X_2.

The conditional CDF can be computed as follows

$$F_{Y|X_1}(y \mid x_1) = \sum_{i=1}^{y} p_{Y|X_1}(i \mid x_1) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{y}{2n} & \text{if } y < x_1 \\ 1 & \text{if } y \geq x_1 \end{cases}$$

The expected value can be computed as follows:

$$E[Y \mid X_1 = x_1] = \sum_{y=1}^{2n} y\, p_{Y|X_1}(y \mid x_1) = \sum_{y=1}^{x_1-1} y\frac{1}{2n} + \sum_{y=x_1}^{x_1} y\left(1 - \frac{y-1}{2n}\right) = \frac{1}{2n}\sum_{y=1}^{x_1-1} y \; + \; x_1\left(1 - \frac{x_1-1}{2n}\right)$$

11. Tail bounds

Suppose $X \sim$ Binomial$(6, 0.4)$. We will bound $\mathbb{P}(X \geq 4)$ using the tail bounds we've learned, and compare this to the true result.

(a) Give an upper bound for this probability using Markov's inequality. Why can we use Markov's inequality?

(b) Give an upper bound for this probability using Chebyshev's inequality. You may have to rearrange algebraically and it may result in a weaker bound.

(c) Give an upper bound for this probability using the Chernoff bound.

(d) Give the exact probability.

Solution:

(a) We know that the expected value of a binomial distribution is np, so: $\mathbb{P}(X \geq 4) \leq \frac{\mathbb{E}[X]}{4} = \frac{2.4}{4} = 0.6$. We can use it since X is nonnegative.

$\mathbb{P}(X \geq 4) = \mathbb{P}(X - 2.4 \geq 1.6) \leq \mathbb{P}(|X - 2.4| \geq 1.6)$ we can add those absolute value signs because that only adds more possible values, so it is an upper bound on the probability of $X - 2.4 \geq 1.6$. Then, using Chebyshev's inequality we get:
$\mathbb{P}(|X - 2.4| \geq 1.6) \leq \frac{Var(X)}{1.6^2} = \frac{1.44}{1.6^2} = 0.5625$

(b) $\mathbb{P}(X \geq 4) = \mathbb{P}(X \geq (1 + \frac{2}{3})2.4) \leq e^{-(\frac{2}{3})^2\mathbb{E}[X]/4} = e^{-4\times 2.4/36} \approx 0.77$

(c) Since X is a binomial, we know it has a range from 0 to n (or in this case 0 to 6). Thus, the possible values to satisfy $X \geq 4$ are 4, 5, or 6. We plug in the PMF for each to get: $\mathbb{P}(X \geq 4) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) = \binom{6}{4}(0.4)^4(0.6)^2 + \binom{6}{5}(0.4)^5(0.6) + \binom{6}{6}0.4^6 \approx 0.1792$

(d) By definition, and using $\Omega_Y = (0, 2)$:

$$\mathbb{E}[Y] = \int_0^2 f_Y(y)y\,dy = \int_0^2 \frac{1}{14}(3y + 4)y\,dy = \frac{8}{7}$$