richardwu.ca

STAT 333 Course Notes
Applied Probability

Steve Drekic • Winter 2018 • University of Waterloo

Last Revision: January 18, 2018

Table of Contents

Abstract

These notes are intended as a resource for myself; past, present, or future students of this course, and anyone interested in the material. The goal is to provide an end-to-end resource that covers all material discussed in the course displayed in an organized manner. If you spot any errors or would like to contribute, please contact me directly.

1 January 4, 2018

1.1 Example 1.1 Solution

What is the probability that we roll a number less than 4 given that we know it's odd?

Solution. Let $A = \{1, 2, 3\}$ (less than 4) and $B = \{1, 3, 5\}$ (odd). We want to find $P(A \mid B)$. Note that $A \cap B = \{1, 3\}$ and there are six elements in the sample space S thus

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

1.2 Example 1.2 Solution

Show that $Bin(n, p) \sim Pois(\lambda)$ when $\lambda = np$ for n large and p small.

Solution. Let $\lambda = np$. Note that $p = \frac{\lambda}{n} \quad n > 0$. From the pmf for $X \sim Bin(n, p)$

$$\begin{aligned} p(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{n(n-1)...(n-x+1)}{x!} (\frac{\lambda}{n})^x (1 - \frac{\lambda}{n})^{n-x} \\ &= \frac{n(n-1)...(n-x+1)}{n^x} \cdot \frac{\lambda^x}{x!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^x} \end{aligned}$$

Recall $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^n = e^{-\lambda}$ so

$$\lim_{n\to\infty} p(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

2 January 9, 2018

2.1 Example 1.3 Solution

Find the mgf of $Bin(n, p)$ and use that to find $E[X]$ and $Var(X)$.

Solution. Recall the binomial series is

$$(a + b)^m = \sum_{x=0}^{m} \binom{m}{x} a^x b^{m-x} \quad a, b \in \mathbb{R}, m \in \mathbb{N}$$

Let $x \sim Bin(n, p)$ and so

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \ldots, n$$

Taking the mgf $E[e^{tX}]$

$$\begin{aligned}\Phi_X(t) = E[e^{tX}] &= \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x(1-p)^{n-x}\end{aligned}$$

from the binomial series we have

$$\Phi_x(t) = (pe^t + 1 - p)^n \quad t \in \mathbb{R}$$

We can take the first and second derivatives for the first and second moment

$$\begin{aligned}\Phi'_X(t) &= n(pe^t + 1 - p)^{n-1}pe^t \\ \Phi''_X(t) &= np[(pe^t + 1 - p)^{n-1}e^t + e^t(n-1)(pe^t + 1 - p)^{n-2}pe^t]\end{aligned}$$

So $E[X] = \Phi_X(t) \mid_{t=0} = np$.
For the variance, we need the second moment

$$\begin{aligned}E[X^2] &= \Phi_X(t) \mid_{t=0} \\ &= np[1 + (n-1)p] \\ &= np + (np)^2 - np^2\end{aligned}$$

So

$$\begin{aligned}Var(X) &= E[X^2] - E[X]^2 \\ &= np + (np)^2 - np^2 - (np)^2 \\ &= np(1-p)\end{aligned}$$

2.2 Example 1.4 Solution

Show that $Cov(X, Y) = 0 \not\Rightarrow$ independence.

Solution. We show this using a counter example

	$p(x,y)$	y = 0	y = 1	$p_X(x)$
x	0	0.2	0	0.2
x	1	0	0.6	0.6
x	2	0.2	0	0.2
	$p_Y(y)$	0.4	0.6	1

Note that

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

where

$$E[XY] = \sum_{x=0}^{2}\sum_{y=0}^{1} xyp(x,y) = (1)(1)(0.6) = 0.6$$
$$E[X] = \sum_{x=0}^{2} xp_X(x) = (1)(0.6) + (2)(0.2) = 0.6 + 0.4 = 1$$
$$E[Y] = \sum_{y=0}^{1} yp_Y(y) = (1)(0.6) = 0.6$$

So $Cov(X,Y) = 0.6 - (1)(0.6) = 0$. However, $p(2,0) = 0.2 \neq p_X(2)p_Y(0) = (0.2)(0.4) = 0.08$, thus X and Y are not independent (they are dependent).

2.3 Example 1.5 Solution

Given $X_1, \ldots, X_n$ are independent r.v's where $\Phi_X(t)$ is the mgf of X_i, show that $T = \sum_{i=1}^n X_i$ has mgf $\Phi_T(t) = \prod_{i=1}^n \Phi_{X_i}(t)$.

Solution. We take the definition of the mgf of T

$$\begin{aligned}
\Phi_T(t) &= E[e^{tT}] \\
&= E[e^{t(X_1+\ldots+X_n)}] \\
&= E[e^{tX_1} \cdot \ldots \cdot e^{tX_n}] \\
&= E[e^{tX_1}] \cdot \ldots \cdot E[e^{tX_n}] && \text{independence} \\
&= \prod_{i=1}^{n} \Phi_{X_i}(t)
\end{aligned}$$

2.4 Exercise 1.3

If $X_i \sim Pois(\lambda_i)$ show that $T = \sum X_i \sim Pois(\sum \lambda_i)$.

Solution. Recall that $Pois(\lambda_i) \sim Bin(n_i, p)$ where $\lambda_i = n_i p$ and

$$\Phi_{X_i}(t) = (pe^t + 1 - p)^{n_i} \quad \forall t \in \mathbb{R}$$

where $X_i \sim Bin(n_i, p) \quad i = 1, \ldots, m$.
Therefore

$$\begin{aligned}
\Phi_T(t) &= \prod_{i=1}^{m} (pe^t + 1 - p)^{n_i} \\
&= (pe^t + 1 - p)^{n_1} \cdot \ldots \cdot (pe^t + 1 - p)^{n_m} \\
&= (pe^t + 1 - p)^{\sum n_i} \quad t \in \mathbb{R}
\end{aligned}$$

By the mgf uniqueness property, we have

$$T = \sum_{i=1}^{m} X_i \sim Bin(\sum_{i=1}^{m} n_i, p)$$

3 January 11, 2018

3.1 Theorem 2.1 - conditional variance

Theorem 3.1.

$$Var(X_1 \mid X_2 = x_2) = E[X_1^2 \mid X_2 = x_2] - E[X_1 \mid X_2 = x_2]^2$$

Proof.

$$\begin{aligned}
Var(X_1 \mid X_2 = x_2) &= E[(X_1 - E[X_1 \mid X_2 = x_2])^2 \mid X_2 = x_2] \\
&= E[(X_1^2 - 2E[X_1 \mid X_2 = x_2]X_1 + E[X_1 \mid X_2 = x_2]^2) \mid X_2 = x_2] \\
&= E[X_1^2 \mid X_2 = x_2] - 2E[X_1 \mid X_2 = x_2]E[X_1 \mid X_2 = x_2] + E[X_1 \mid X_2 = x_2]^2 \\
&= E[X_1^2 \mid X_2 = x_2] - E[X_1 \mid X_2 = x_2]^2
\end{aligned}$$

□

3.2 Example 2.1

Suppose that X and Y are discrete random variables having join pmf of the form

$$p(x, y) = \begin{cases} 1/5 & \text{, if x = 1 and y = 0,} \\ 2/15 & \text{, if x = 0 and y = 1,} \\ 1/15 & \text{, if x = 1 and y = 2,} \\ 1/5 & \text{, if x = 2 and y = 0,} \\ 2/5 & \text{, if x = 1 and y = 1,} \\ 0 & \text{, otherwise.} \end{cases}$$

Find the conditional probability of $X \mid (Y = 1)$. Also calculate $E[X \mid Y = 1]$ and $Var(X \mid Y = 1)$.

Solution. Note: for problems of this nature, construct a table.

		y			
	$p(x, y)$	0	1	2	$p_X(x)$
	0	0	2/15	0	2/15
x	1	1/5	2/5	1/15	2/3
	2	1/5	0	0	1/5
	$p_Y(y)$	2/5	8/15	1/15	1

Then we have

$$\begin{aligned}
p(0 \mid 1) = P(X = 0 \mid Y = 1) &= \frac{2/15}{8/15} = \frac{1}{4} \\
p(1 \mid 1) = P(X = 1 \mid Y = 1) &= \frac{2/5}{8/15} = \frac{3}{4} \\
p(2 \mid 1) = P(X = 2 \mid Y = 1) &= \frac{0}{8/15} = 0
\end{aligned}$$

The conditional pmf of $X \mid (Y = 1)$ can be represented as follows

x	0	1
$p(x \mid 1)$	$1/4$	$3/4$

We observe $X \mid (Y = 1) \sim Bern(3/4)$. We can take the known $E[X] = p$ and $Var(X)p(1-p)$ for $X \sim Bern(p)$, thus

$$\begin{aligned} E[X \mid (Y = 1)] &= 3/4 \\ Var(X \mid (Y = 1)) &= 3/4(1 - 3/4) = 3/16 \end{aligned}$$

3.3 Example 2.2

For $i = 1, 2$ suppose that $X_i \sim Bin(n_i, p)$ where X_1, X_2 are independent (but not identically distributed). Find conditional distribution of X_1 given $X_1 + X_2 = n$.

Solution. We want to find conditional pmf of $X \mid (X_1 + X_2 = n)$. Let this conditional pmf be denoted by

$$\begin{aligned} p(x_1 \mid n) &= P(X_1 = x_1 \mid X_1 + X_2 = n) \\ &= \frac{P(X_1 = x_1, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \end{aligned}$$

Recall: $X_1 + X_2 \sim Bin(n_1 + n_2, p)$ so

$$P(X_1 + X_2 = n) = \binom{n_1 + n_2}{n} p^n (1-p)^{n_1+n_2-n}$$

Next, consider

$$\begin{aligned} P(X_1 = x_1, X_1 + X_2 = n) &= P(X_1 = x_1, x_1 + X_2 = n) \\ &= P(X_1 = x_1, X_2 = n - x_1) \\ &= P(X_1 = x_1)P(X_2 = n - x_1) && \text{independence} \\ &= \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1 - x_1} \cdot \binom{n_2}{n - x_1} p^{n - x_1} (1-p)^{n_2 - (n - x_1)} \end{aligned}$$

provided that $0 \leq x_1 \leq n_1$ **and**

$$\begin{aligned} 0 \leq n - x_1 \leq n_2 \\ -n_2 \leq x_1 - n \leq 0 \\ n - n_2 \leq x_1 \leq n \end{aligned}$$

(from the binomial coefficients). Therefore our domain for x_1 is

$$x_1 = \max\{0, n - n_2\}, \ldots, \min\{n_1, n\}$$

Thus we have

$$\begin{aligned}
p(x_1 \mid n) &= \frac{P(X_1 = x, X_1 + x_2 = n)}{P(X_1 + X_2 = n)} \\
&= \frac{\binom{n_1}{x_1} p^{x_1}(1-p)^{n_1 - x_1} \cdot \binom{n_2}{n-x_1} p^{n-x_1}(1-p)^{n_2-(n-x_1)}}{\binom{n_1+n_2}{n} p^n (1-p)^{n_1+n_2-n}} \\
&= \frac{\binom{n_1}{x_1}\binom{n_2}{n-x_1}}{\binom{n_1+n_2}{n}}
\end{aligned}$$

for $x_1 = \max\{0, n-n_2\}, \ldots, \min\{n_1, n\}$.
Recall: A $HG(N, r, n)$ (hypergeometric) distribution has pmf

$$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{x}} \quad x = \max\{0, n-N+r\}, \ldots, \min\{n, r\}$$

So this is precisely $HG(n_1 + n_2, x_1, n)$.
If you think about it: we are choosing x_1 successes from n_1 trials from the first set X_1 and choosing the remaining $n - x_1$ successes from n_2 trials from X_2.

4 Tutorial 1

4.1 Exercise 1: MGF of Erlang

Find the mgf of $X \sim Erlang(\lambda)$ and use it to find $E[X], Var(X)$.
Note that the Erlang's pdf is for $n \in \mathbb{Z}^+$ and $\lambda > 0$

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad x > 0$$

Solution.

$$\begin{aligned}
\Phi_X(t) = E[e^{tX}] &= \int_0^\infty e^{tx} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx \\
&= \int_0^\infty \frac{\lambda^n x^{n-1} e^{-(\lambda - t)x}}{(n-1)!} dx
\end{aligned}$$

Note that the term in the integral is similar to the pdf of Erlang but for $\lambda = \lambda - t$. So we try to fix it so the integral is this pdf of Erlang

$$\begin{aligned}
\Phi_X(t) &= \int_0^\infty \frac{\lambda^n x^{n-1} e^{-(\lambda-t)x}}{(n-1)!} dx \\
&= (\frac{\lambda}{\lambda - t})^n \int_0^\infty \frac{(\lambda - t)^n x^{n-1} e^{-(\lambda-t)x}}{(n-1)!} dx \\
&= (\frac{\lambda}{\lambda - t})^n && t < \lambda
\end{aligned}$$

since the integral over the positive real line of the pdf of an $Erlang(n, \lambda - t)$ is 1 and $t < \lambda$ must hold so the rate parameter $\lambda - t$ is positive.

Differentiating,

$$\begin{aligned}
\Phi_X^{(1)}(t) &= \frac{d}{dt}\left(\frac{\lambda}{(\lambda-t)^n}\right) \\
&= \frac{n\lambda^n}{(\lambda-t)^{n+1}} \\
\Phi_X^{(2)}(t) &= \frac{d}{dt}\left(\frac{n\lambda^n}{(\lambda-t)^{n+1}}\right) \\
&= \frac{n(n+1)\lambda^n}{(\lambda-t)^{n+2}}
\end{aligned}$$

Thus we have

$$\begin{aligned}
E[X] = \Phi_X^{(1)}(0) &= \left.\frac{n\lambda^n}{(\lambda-t)^{n+1}}\right|_{t=0} = \frac{n}{\lambda} \\
E[X^2] = \Phi_X^{(2)}(0) &= \left.\frac{n(n+1)\lambda^n}{(\lambda-t)^{n+2}}\right|_{t=0} = \frac{n(n+1)}{\lambda^2} \\
Var(X) = E[X^2] - E[X]^2 &= \frac{n(n+1)}{\lambda^2} - \frac{n}{\lambda} = \frac{n}{\lambda^2}
\end{aligned}$$

Remark 4.1. To solve any of these mgfs, it is useful to see if one can reduce the integral into a pdf of a known distribution (possibly itself).

4.2 Exercise 2: MGF of Uniform

Find the mgf of the uniform distribution on $(0, 1)$ and find $E[X]$ and $Var(X)$.

Solution. Let $X \sim U(0, 1)$ so that $f(x) = 1 \quad 0 \leq x \leq 1$. We have

$$\begin{aligned}
\Phi_X(t) = E[e^{tX}] &= \int_0^1 e^{tx}(1)dx \\
&= \left.\frac{1}{t}e^{tx}\right|_{x=0}^{x=1} \\
&= t^{-1}(e^t - 1) \quad t \neq 0
\end{aligned}$$

Differentiating

$$\begin{aligned}
\Phi_X^{(1)}(t) &= \frac{d}{dt}(t^{-1}(e^t-1)) \\
&= t^{-1}e^t - t^{-2}(e^t - 1) \\
&= \frac{te^t - e^t + 1}{t^2} \\
\Phi_X^{(2)}(t) &= \frac{d}{dt}\left(\frac{te^t - e^t + 1}{t^2}\right) \\
&= \frac{t^2(te^t + e^t - e^t) - 2t(te^t - e^t + 1)}{t^4} \\
&= \frac{t^2e^t - 2te^t + 2e^t - 2}{t^3}
\end{aligned}$$

We may calculate the first two moments by applying **L'Hopital's rule** to calculate the limits

$$\begin{aligned} E[X] = \Phi_X^{(1)}(t)\Big|_{t=0} &= \lim_{t\to\infty} \frac{te^t - e^t + 1}{t^2} \\ &= \lim_{t\to\infty} \frac{te^t + e^t - e^t}{2t} \\ &= \lim_{t\to\infty} \frac{e^t}{2} = \frac{1}{2} \end{aligned}$$

Similarly

$$\begin{aligned} E[X^2] = \Phi_X^{(2)}(t)\Big|_{t=0} &= \lim_{t\to\infty} \frac{t^2e^t - 2te^t + 2e^t - 2}{t^3} \\ &= \lim_{t\to\infty} \frac{t^2e^t + 2te^t - 2te^t - 2e^t + 2e^t}{3t^2} \\ &= \lim_{t\to\infty} \frac{e^t}{3} = \frac{1}{3} \end{aligned}$$

So we have

$$Var(X) = E[X^2] - E[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

4.3 Exercise 3: Moments from PGF

Suppose X is a discrete r.v. on $\mathbb{N}$ with pmf $p(x)$. Show how to find the first two moments of X from its pgf.

Solution. By definition, the pgf of X is $\Psi_X(z) = E[z^X] = \sum_{x=0}^{\infty} z^x p(x)$.
If we let $z = 1$, then the sum equals 1. However, if we take its derivative with respect to z just once

$$\Psi_X^{(1)}(z) = \frac{d}{dz}\sum_{x=0}^{\infty} z^x p(x) = \sum_{x=1}^{\infty} xz^{x-1}p(x)$$

Letting $z = 1$ we can find the first moment

$$\begin{aligned} \Psi_X^{(1)}(1) &= \lim_{z\to 1}\sum_{x=1}^{\infty} xz^{x-1}p(x) \\ &= \sum_{x=1}^{\infty} xp(x) \\ &= \sum_{x=0}^{\infty} xp(x) && \text{when } x = 0 \text{ the term is 0 anyways} \\ &= E[X] \end{aligned}$$

For the second moment, we consider the second derivative

$$\begin{aligned} \Psi_X^{(1)}(z) &= \frac{d^2}{dz^2}\sum_{x=0}^{\infty} z^x p(x) \\ &= \sum_{x=2}^{\infty} x(x-1)z^{x-2}p(x) \end{aligned}$$

Letting $z = 1$

$$\begin{aligned}
\Psi_X^{(2)}(1) &= \lim_{z\to 1} \sum_{x=2}^{\infty} x(x-1)z^{x-2}p(x) \\
&= \sum_{x=2}^{\infty} x(x-1)p(x) \\
&= \sum_{x=0}^{\infty} x(x-1)p(x) \\
&= E[X(X-1)] \\
&= E[X^2] - E[X]
\end{aligned}$$

So we have $E[X^2] = \Psi_X^{(2)}(1) + \Psi_X^{(1)}(1)$. To find the variance

$$Var(X) = \Psi_X^{(2)}(1) + \Psi_X^{(1)}(1) - (\Psi_X^{(1)}(1))^2$$

4.4 Exercise 4: PGF of Poisson

Suppose $X \sim Pois(\lambda)$. Find the pgf of X and use it to find $E[X]$ and $Var(X)$. The pmf of $Pois(\lambda)$ for $\lambda > 0$

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

Solution.

$$\begin{aligned}
\Psi_X(z) = E[z^X] &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(z\lambda)^x}{x!} \\
&= e^{-\lambda} \cdot e^{z\lambda} \\
&= e^{\lambda(z-1)}
\end{aligned}$$

where the second equality holds since the summation is the Taylor expansion of $e^{z\lambda}$.
Differentiating

$$\begin{aligned}
\Psi_X^{(1)}(z) &= \frac{d}{dz} e^{\lambda(z-1)} \\
&= \lambda e^{\lambda(z-1)} \\
\Psi_X^{(2)}(z) &= \frac{d}{dz} \lambda e^{\lambda(z-1)} \\
&= \lambda^2 e^{\lambda(z-1)}
\end{aligned}$$

The moments are thus

$$\begin{aligned}
E[X] &= \Phi_X^{(1)}(1) = \lambda e^{\lambda(1-1)} = \lambda \\
E[X(X-1)] &= \Phi_X^{(2)}(1) = \lambda^2 e^{\lambda(1-1)} = \lambda^2 \\
E[X^2] &= E[X(X-1)] + E[X] = \lambda^2 + \lambda \\
Var(X) &= E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}$$

5 January 16, 2018

5.1 Example 2.3 Solution

Let $X_1, \ldots, X_m$ be independent r.v.'s where $X_i \sim Pois(\lambda_i)$. Define $Y = \sum_{i=1}^m X_i$. Find the conditional distribution $X_j \mid (Y = n)$.

Solution. We set out to find

$$\begin{aligned}
p(x_j \mid n) = p(X_j = x_j \mid Y = n) &= \frac{P(X_j = x_j, Y = n)}{P(Y = n)} \\
&= \frac{P(X_j = x_j, \sum_{i=1}^m X_i = n)}{P(Y = n)} \\
&= \frac{P(X_j = x_j, X_j + \sum_{i=1, i\neq j}^m X_i = n)}{P(Y = n)} \\
&= \frac{P(X_j = x_j, \sum_{i=1, i\neq j}^m X_i = n - x_j)}{P(Y = n)} \\
&= \frac{P(X_j = x_j) P(\sum_{i=1, i\neq j}^m X_i = n - x_j)}{P(Y = n)} && \text{independence of } X_i
\end{aligned}$$

Remember that if $X_i \sim Pois(\lambda_i)$, then

$$Y = \sum_{i=1}^m X_i \sim Pois(\sum_{i=1}^m \lambda_i)$$

which can be derived from mgfs (Exercise 1.3). Therefore

$$\sum_{i=1, i\neq j}^m X_i \sim Pois(\sum_{i=1, i\neq j}^m \lambda_i)$$

Expanding out $p(x_j \mid n)$ with the pdfs

$$p(x_j \mid n) = \frac{\frac{e^{-\lambda_j}\lambda_j^{x_j}}{x_j!} \cdot \frac{e^{-\sum_{i=1,i\neq j}\lambda_i}(\sum_{i=1,i\neq j}\lambda_i)^{n-x_j}}{(n-x_j)!}}{\frac{e^{-\sum_{i=1}^m \lambda_i} \cdot (\sum_{i=1}^m \lambda_i)^n}{n!}}$$

where $x_j \geq 0$ and $n - x_j \geq 0 \Rightarrow 0 \leq x_j \leq n$ (from the factorials).

Cancelling out the e^λ terms and let $\lambda_Y = \sum_{i=1}^m \lambda_i$

$$\begin{aligned} p(x_j \mid n) &= \frac{n!}{(n-x_j)!x_j!} \frac{\lambda_j^{x_j}}{\lambda_Y^{x_j}} \frac{(\lambda_Y - \lambda_j)^{n-x_j}}{\lambda_Y^{n-x_j}} \\ &= \binom{n}{x_j} (\frac{\lambda_j}{\lambda_Y})^{x_j} (1 - \frac{\lambda_j}{\lambda_Y})^{n-x_j} \end{aligned}$$

This is the binomial distribution, so we have

$$X_j \mid Y = n \sim Bin(n, \frac{\lambda_i}{\lambda_Y})$$

5.2 Example 2.4 Solution

Suppose $X \sim Pois(\lambda)$ and $Y \mid (X = x) \sim Bin(x, p)$. Find the conditional distribution $X \mid Y = y$.
(Note: range of y depends on x (that is $y \leq x$). Graphically, we have integral points on and below the $y = x$ line starting from 0 for both x and y).

Solution. We wish to find the conditional pmf given by $X \mid Y = y$ or

$$p(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Note that also

$$\begin{aligned} P(Y = y \mid X = x) &= \frac{P(Y = y, X = x)}{P(X = x)} \\ \Rightarrow P(X = x, Y = y) &= P(X = x)P(Y = y \mid X = x) \\ &= \frac{e^{-\lambda}\lambda^x}{x!} \cdot \binom{x}{y} p^y (1-p)^{x-y} \end{aligned}$$

for $x = 0, 1, 2, \ldots$ **and** $y = 0, 1, 2, \ldots, x$ (range of y depends on x).
To find the marginal marginal pmf of Y, we use

$$p_Y(y) = \sum_x p(x, y)$$

To find the support for x, note that from the graphical region, we realize that $x = 0, 1, 2, \ldots$ **and** $y = 0, 1, 2, \ldots, x$ is equivalent to $y = 0, 1, 2, \ldots$ **and** $x = y, y + 1, y + 2, \ldots$.

So

$$\begin{aligned}
p_Y(y) &= \sum_{x=y}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} \frac{x!}{(x-y)!y!} p^y (1-p)^{x-y} \\
&= \frac{\lambda^y e^{-\lambda} p^y}{y!} \sum_{x=y}^{\infty} \frac{\lambda^{x-y}(1-p)^{x-y}}{(x-y)!} \\
&= \frac{e^{-\lambda}(\lambda p)^y}{y!} \sum_{x=y}^{\infty} \frac{[\lambda(1-p)]^{x-y}}{(x-y)!} \\
&= \frac{e^{-\lambda}(\lambda p)^y}{y!} e^{\lambda(1-p)} \\
&= \frac{e^{-\lambda p}(\lambda p)^y}{y!} \qquad\qquad y = 0, 1, 2, \ldots
\end{aligned}$$

Note that $p_Y(y) \sim Pois(\lambda p)$.
Thus

$$\begin{aligned}
p(x \mid y) &= \frac{P(X = x, Y = y)}{P(Y = y)} \\
&= \frac{\frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{x!}{(x-y)!y!} p^y (1-p)^{x-y}}{\frac{e^{-\lambda p}(\lambda p)^y}{y!}} \\
&= \frac{e^{-\lambda+\lambda p}[\lambda(1-p)]^{x-y}}{(x-y)!} \\
&= \frac{e^{-\lambda(1-p)}[\lambda(1-p)]^{x-y}}{(x-y)!} \qquad\qquad x = y, y+1, y+2, \ldots
\end{aligned}$$

This resembles the Poisson distribution with $\lambda = \lambda(1-p)$ but with a slightly modified domain.
So we see that

$$W \mid (Y = y) \sim W + y$$

where $W \sim Pois(\lambda(1-p))$. This is the **shifted Poisson pmf** y units to the right (note that W and y are random variables).
We can easily find the conditional expectations and variance e.g.

$$E[X \mid Y = y] = E[W + y] = E[W] + y$$

5.3 Example 2.5 Solution

Suppose the joint pdf of X and Y is

$$f(x, y) = \begin{cases} \frac{12}{5} x(2 - x - y) & , 0 < x < 1, 0 < y < 1, \\ 0 & , \text{ elsewhere} \end{cases}$$

Determine the conditional distribution of X given $Y = y$ where $0 < y < 1$. Also calculate the mean of $X \mid (Y = y)$. (Note: the graphical region is a unit square box where the bottom left corner is at $0, 0$: the inside of the box is the support).

Solution. Using our theory, we wish to find the conditional pdf of $X \mid (Y = y)$ given by

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

For $0 < y < 1$

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\
&= \int_0^1 \frac{12}{5}x(2-x-y)dx \\
&= \frac{12}{5}\int_0^1 (2x - x^2 - xy)dx \\
&= \frac{12}{5}(x^2 - \frac{x^3}{3} - \frac{x^2 y}{2})\Big|_0^1 \\
&= \frac{12}{5}(1 - \frac{1}{3} - \frac{y}{2}) \\
&= \frac{2}{5}(4-3y)
\end{aligned}$$

So we have

$$\begin{aligned}
f_{X|Y}(x \mid y) &= \frac{\frac{12}{5}x(2-x-y)}{\frac{2}{5}(4-3y)} \\
&= \frac{6x(2-x-y)}{4-3y}
\end{aligned}$$

Thus we have

$$\begin{aligned}
E[X \mid Y] &= \int_0^1 x \cdot f_{X|Y}(x \mid y)dx \\
&= \frac{5-4y}{2(4-3y)}
\end{aligned}$$

6 January 18, 2018

6.1 Example 2.6 Solution

Suppose the joint pdf of X and Y is

$$f(x,y) = \begin{cases} 5e^{-3x-y} & , 0 < 2x < y < \infty, \\ 0 & , \text{ otherwise} \end{cases}$$

Find the conditional distribution of $Y \mid (X = x)$ where $0 < x < \infty$.
Note the region of support is a "flag" (upright triangle with downawrd point) where the slanted part is the line $y = 2x$.

Solution. We wish to find

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}$$

For $0 < x < \infty$

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{2x}^{\infty} 5e^{-3x-y}dy \\
&= 5e^{-3x}\int_{2x}^{\infty} 5e^{-y}dy \\
&= 5e^{-3x}(-e^{-y})\Big|_{2x}^{\infty} \\
&= 5e^{-3x}e^{-2x}) \\
&= 5e^{-5x}
\end{aligned}$$

so we have $f_X(x) \sim Exp(5)$.

Remark 6.1. The bounds on the integral are in terms of y: it is dependent on x in our $f(x,y)$ definition.

Now

$$\begin{aligned}
f_{Y|X}(y \mid x) &= \frac{5e^{-3x-y}}{5e^{-5x}} \\
&= e^{-y+2x} \quad y > 2x
\end{aligned}$$

Note: recognize the conditional pdf of $Y \mid (X = x)$ as that of a shifted exponential distribution ($2x$ units to the right). Specifically, we have

$$Y \mid (X = x) \sim W + 2x$$

where $W \sim Exp(1)$. Thus $E[Y \mid (X = x)] = E(W) + 2x$ and $Var[Y \mid (X = x)] = Var(W)$.

6.2 Example 2.7 Solution

Suppose $X \sim U(0,1)$ and $Y \mid (X = x) \sim Bern(x)$. Find the conditional distribution $X \mid (Y = y)$.
Note: X is continuous and $Y \mid (X = x)$ is discrete.

Solution. We wish to find

$$f_{X|Y}(x \mid y) = \frac{p(y \mid x)f_X(x)}{p_Y(y)}$$

From the given information, we have $f_X(x) = 1$ for $0 < x < 1$ Furthermore $p(y \mid x) = Bern(x) = x^y(1-x)^{1-y}$ for $y = 0,1$.
For $y = 0,1$ note that (from $\int f(x \mid y)dx = 1$)

$$\begin{aligned}
p_Y(y) &= \int_{-\infty}^{\infty} p(y \mid x)f_X(x)dx \\
p_Y(y) &= \int_0^1 x^y(1-x)^{1-y}dx
\end{aligned}$$

To compute this integral, let's check $p_Y(0)$ and $p_Y(1)$

$$\begin{aligned}
p_Y(0) &= \int_0^1 x^0(1-x)^{1-0}dx \\
&= \int_0^1 1 - xdx \\
&= x - \frac{x^2}{2}\Big|_0^1 \\
&= \frac{1}{2}
\end{aligned}$$

Similarly, take $y = 1$ where $p_Y(1) = \frac{1}{2}$.
In other words, we have that $p_Y(y) = \frac{1}{2}$ $y = 0, 1$ so

$$Y \sim Bern\left(\frac{1}{2}\right)$$

So

$$\begin{aligned}
f(x \mid y) &= \frac{p(y \mid x)f_X(x)}{p_Y(y)} \\
&= \frac{x^y(1-x)^{1-y} \cdot 1}{\frac{1}{2}} \\
&= 2x^y(1-x)^{1-y} \quad 0 < x < 1
\end{aligned}$$

6.3 Theorem 2.2 (law of total expectation)

Prove that for random variables X and Y, $E[X] = E[E[X \mid Y]]$.

Proof. WLOG assume X, Y are jointly continuous random variables. We note

$$\begin{aligned}
E[E[X \mid Y]] &= \int_{-\infty}^{\infty} E[X \mid Y = y]f_Y(y)dy \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} xf_{X|Y}(x \mid y)dx\right] f_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\frac{f(x,y)}{f_Y(y)} \cdot f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy \\
&= \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f(x,y)dy\right] dx \\
&= \int_{-\infty}^{\infty} xf_X(x)dx \\
&= E[X]
\end{aligned}$$

$\square$

6.4 Example 2.8 Solution

Suppose $X \sim Geo(p)$ with pmf $p_X(x) = (1-p)^{x-1}p$ where $x = 1, 2, 3, \ldots$. Calculate $E[X]$ and $Var(X)$ using the law of total expectation.

Solution. Recall $E[X] = \frac{1}{p}$ and $Var(X) = \frac{1-p}{p^2}$ where X models the number of (independent) trials necessary to obtain the first success.
Remember: we could manually solve $E[X] = \sum_{x=1}^{\infty}(1-p)^{x-1}p$ and similarly $Var(X) = E[X^2] - E[X]$, or take the derivatives of the mgf $\Phi_X(t) = E[e^{tX}]$. This is tedious in general.