## STAT333 — Applied Probability

CLASSNOTES FOR WINTER 2017

by

Johnson Ng

BMath (Hons), Pure Mathematics major, Actuarial Science Minor University of Waterloo

# Table of Contents

Ta	ble o	f Contents	2	
Li	st of	Definitions	3	
Li	st of	Theorems	4	
Li	st of	Procedures	5	
1	Eler	nentary Probability Review	9	
	1.1	Introductions	9	
	1.2	Random Variables	11	
		1.2.1 Discrete Random Variables	11	
		1.2.2 Continuous Random Variables	13	
	1.3	Moments	14	
	1.4	Joint Distributions	17	
	1.5	Expectation of Joint Distributions	19	
	1.6	Independence of Random Variables	20	
2	Con	nditional Probabilities	23	
	2.1	Conditional Probability for Discrete Random Variables	23	
	2.2	Conditional Probability for Continuous Random Variables	31	
Bi	Bibliography			
In	Index			

# List of Definitions

1	■ Definition (Fundamental Definition of a Probability Function)	Ç
2	■ Definition (Conditional Probability)	ç
3	<b>■</b> Definition (Expectation)	14
4	■ Definition (Moment)	14
5	■ Definition (Variance)	14
6	■ Definition (Standard Deviation)	15
7	■ Definition (Moment Generating Function)	15
8	■ Definition (Joint CDF)	18
9	■ Definition (Marginal CDF)	18
10	■ Definition (Joint and Marginal Probability Mass Functions)	18
11	■ Definition (Joint and Marginal Probability Density Function)	18
12	■ Definition (Expectation of Joint Distributions)	19
13	■ Definition (Covariance of Joint Distributions)	20
14	■ Definition (Independence)	20
15	E Definition (Conditional Distribution of Discrete Random Variables)	23
16	■ Definition (Conditional Expectation for Discrete RVs)	
17	■ Definition (Conditional Distribution of Continuous Random Variables)	3:

# List of Theorems

1	■Theorem (Law of Total Probability)	10
2	Corollary (Bayes' Formula/Rule)	11
3	♦ Proposition (Linearity of the Expectation)	15
4	♦ Proposition (Linearity of the Variance)	15
5	♦ Proposition (Linearity of the Expectation over Addition)	20
6	♦ Proposition (Variance over a Linear Combination of RVs)	20
7	♦ Proposition (Expectation of Independent RVs)	21
8	Corollary (MGF of Independent RVs)	21
9	♦ Proposition (Conditional Variance of Discrete RVs)	24
10	♦ Proposition (Linearity of Conditional Expectation)	25
11	Corollary (General Linearity of Conditional Expectation)	26

# List of Procedures

## **F**oreword

I am transcribing this set of notes from my handwritten ones in Winter 2017, back at a time which I have yet to organize my notes by lecture. However, I will try my best to organize them by chapters and topics as presented in class.

I will try to be as rigourous as possible while transcribing my notes. However, given the nature of the course and the presentation, this will not always be possible, and I am mostly keeping these notes for "legacy purposes", and so I will not put too much effort into making the notes as complete as my newer ones.

For this course, you are expected to have basic knowledge of probability in order to be able to understand the material. You may want to have my STAT330 notes ready and/or reviewed.

## Elementary Probability Review

## 1.1 Introductions

## **■** Definition 1 (Fundamental Definition of a Probability Function)

For each event A of a sample space S, P(A) is defined as the "probability of the event A", satisfying these 3 conditions:

1. 
$$0 \le P(A) \le 1$$

2. 
$$P(S) = 1^{1}$$

3. 
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$
, where  $A_i \cap A_j = A_i A_j = \emptyset$  for all  $i \neq j^2$ 

<sup>1</sup> This can also be stated as  $P(\emptyset) = 0$ , where  $\emptyset$  is the null event.

 $^{2}$  We can also say that the sequence  $\{A_{i}\}_{i=1}^{n}$  has mutually exclusive elements

### 66 Note 1.1.1

By Item 2 and Item 3, we have

$$1 = P(S) = P(A \cup A^{C}) = P(A) + P(A^{C})$$

which implies that

$$P(A^C) = 1 - P(A).$$

## **■** Definition 2 (Conditional Probability)

Given events A and B in a sample space S, the conditional probability

of A given B is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) > 0. \tag{1.1}$$

## **66** Note 1.1.2

When B = S, Equation (1.1) becomes

$$P(A \mid S) = \frac{P(A \cap S)}{P(S)} = \frac{P(A)}{1} = P(A).$$

Also, we have, from Equation (1.1), that

$$P(A \cap B) = P(A \mid B) \cdot P(B).$$

## **■**Theorem 1 (Law of Total Probability)

Let S be a sample space. Let  $\{B_i\}_{i=1}^n$  be a sequence of mutually exclusive events such that

$$S = \bigcup_{i=1}^{n} B_i.$$

We say that the sequence  $\{B_i\}_{i=1}^n$  is a partition of S. Let  $A \subseteq S$  be an event. Then

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

#### Proof

Observe that

$$P(A) = P(A \cap S) = P\left(A \cap \left\{\bigcup_{i=1}^{n} B_i\right\}\right)$$
$$= P\left(\bigcup_{i=1}^{n} \left\{A \cap B_i\right\}\right) = \sum_{i=1}^{n} P(A \cap B_i)$$
$$= \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

Definition 2.

Г

Consequently, we have the following:

## Corollary 2 (Bayes' Formula/Rule)

Let  $\{B_i\}_{i=1}^n$  be a partition of a sample space S. Then for any event A, we have

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^n P(A \mid B_i) \cdot P(B_i)}.$$

## 1.2 Random Variables

#### 1.2.1 Discrete Random Variables

No formal definition of a discrete rv is given in class.

A discrete rv *X*:

- takes on either finite or countable number of possible values;
- has a probability mass function (pmf) expressed as

$$p(a) = P(X = a);$$

• has a cumulative distribution function (cdf) expressed as

$$F(a) = P(X \le a) = \sum_{x \le a}^{p(x)}$$

#### 66 Note 1.2.1

If  $X \in \{a_1, a_2, ...\}$  where  $a_1 < a_2 < ...$  such that  $p(a_i) > 0$  for all  $i \in \mathbb{N}$ , then

$$p(a_1) = F(a_1)$$
 and  $p(a_i) = F(a_i) - F(a_{i-1})$  for  $i = 2, 3, 4, ...$ 

THE FOLLOWING are some of the most common discrete distributions.

*Binomial Distribution* For an rv X that follows a Binomial Distribution, in which we denote as  $X \sim \text{Bin}(n, p)$ , where  $n \in \mathbb{N}$  and  $p \in [0, 1]$ , its pmf is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Bernoulli Distribution Following the above distribution where n = 1, we have that X follows what is called a Bernoulli Distribution, denoted as  $X \sim \text{Bernoulli}(p)$ .

*Negative Binomial Distribution* For an rv X that follows a Negative Binomial Distribution, in which we denote as  $X \sim NB(k, p)$ , where  $k \in \mathbb{N}$  and  $p \in [0, 1]$ , its pmf is

$$p(x) = {x-1 \choose k-1} p^k (1-p)^{x-k}$$

Geometric Distribution Following the above distribution where k = 1, we have that X follows what is called a Geometric Distribution, denoted as  $X \sim \text{Geo}(p)$ .

*Hypergeometric Distribution* For an rv X that follows a Hypergeometric Distribution, in which we denote as  $X \sim HG(N, rn)$ , where  $r, n \leq N \in \mathbb{N}$ , its pmf is

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

*Poisson Distribution* For an rv X that follows a Poisson Distribution, in which we denote as  $X \sim \text{Poi}(\lambda)$ , where  $\lambda > 0$ , its pmf is

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

The Negative Binomial Distribution has a model that measures the probability that the *k*th success occurs.

#### 1.2.2 Continuous Random Variables

No formal definition of a continuous rv is given in class.

A continuous rv *X*:

- takes on a continuum of possible values
- has a probability density function (pdf) expressed as

$$f(x) = \frac{d}{dx}F(x)$$

where F(x) is:

• (has a) cumulative distribution function (cdf) of

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y) \, dy$$

#### 66 Note 1.2.2

Note that our convention is that P(X = x) = 0 for a continuous rv(X).

THE FOLLOWING are some of the most common continuous distributions.

*Uniform Distribution* For an rv X that follows a Uniform Distribution, in which we denote as  $X \sim \text{Unif}(a,b)$ , where  $a,b \in \mathbb{R}$ , its pdf is

$$f(x) = \frac{1}{b-a}.$$

*Gamma Distribution* For an rv X that follows a Gamma Distribution, in which we denote as  $X \sim \text{Gam}(n, \lambda)$ , where  $n \in \mathbb{N}$  and  $\lambda > 0$ , its pdf is

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

Exponential Distribution Following the above distribution where n = 1, we have that X follows what is called an Exponential Distribution,

denoted as  $X \sim \text{Exp}(\lambda)$ , where its pdf is

$$f(x) = \lambda e^{-\lambda x}.$$

## 1.3 Moments

## **■** Definition 3 (Expectation)

Let X be an rv. Given a function g that is defined over X, the expectation of g(X) is given by

$$E[g(X)] = \begin{cases} \sum_{x} g(x)p(x) & \text{if } X \text{ is a discrete } rv \\ \int_{X} g(x)f(x) & \text{if } X \text{ is a continuous } rv \end{cases}.$$

Now if  $g(X) = X^k$  for some  $k \in \mathbb{N}$ , we have the following notion:

## **■** Definition 4 (Moment)

Let X be an rv. The kth moment of X is defined as  $E[X^k]$ .

Another notion that is commonly introduced after expectation is the variance.

## **■** Definition 5 (Variance)

Let X be an rv. The variance of X is given by

$$Var(X) = E[(X - E[X])^{2}] = E[X^{2}] - (E[X])^{2}$$

In relation to the variance, we have the standard deviation.

### **■** Definition 6 (Standard Deviation)

Note that this definition is actually the Law of the Unconscious Statistician

Let X be an rv. The standard deviation (sd) is given by

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

We shall state the following properties without providing proof<sup>3</sup>:

#### <sup>3</sup> The proofs are very easy, but it serves as a strengthening exercise for the unfamiliar. Therefore,

## ♦ Proposition 3 (Linearity of the Expectation)

Let X be an rv. Let  $a,b \in \mathbb{R}$ . We have that

$$E[aX + b] = aE[x] + b$$

Proof both 💧 Proposition 3 and **♦** Proposition 4.

Exercise 1.3.1

## • Proposition 4 (Linearity of the Variance)

Let X be an rv. Let  $a,b \in \mathbb{R}$ . We have that

$$Var(aX + b) = a^2 Var(X).$$

Referring back to  $\blacksquare$  Definition 3, if  $g(X) = e^{tX}$ , we have ourselves, what is called, the moment generating function.

## **■** Definition 7 (Moment Generating Function)

Let X be an rv. The moment generating function (mgf) of X is given by

$$\varphi_X(t) = E\left[e^{tX}\right].$$

#### 66 Note 1.3.1

- 1. Observe that  $\varphi_X(0) = E[e^0] = 1$ .
- 2. The reason such an expression is called a moment generating function

is as follows: observe that

$$\varphi_X(t) = E\left[e^{tX}\right] = E\left[\sum_{i=0}^{\infty} \frac{(tX)^i}{i!}\right] 
= E\left[1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \dots + \frac{(tX)^n}{n!} + \dots\right] 
= \frac{t^0}{0!}E[1] + \frac{t}{1!}E[X] + \frac{t^2}{2!}E\left[X^2\right] + \dots + \frac{t^n}{n!}E[X^n] + \dots$$

by lacktriangle Proposition 3. If we take the kth derivative wrt t and set t=0, we will obtain the kth moment of X. In other words,

$$E[X^k] = \varphi_X^{(k)}(0) = \frac{d^k}{dt^k} \varphi_X(t) \Big|_{t=0}.$$

It is important to note here that t = 0 can be interpreted in the sense of a limit, i.e.  $\lim_{t\to 0}$ .

#### Example 1.3.1

Suppose  $X \sim Bin(n, p)$ . Find mgf(X) and use it to find E[X] and Var(X).

#### Solution

First, observe the binomial formula:

$$(a+b)^m = \sum_{x=0}^m \binom{m}{x} a^x b^m.$$

Now

$$\varphi_X(t) = E\left[e^{tX}\right] = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \sum_{x=0}^n \binom{n}{x} \left(pe^t\right)^x (1-p)^{n-x}$$
$$= \left(pe^t + 1 - p\right)^n.$$

We observe that this works for all  $t \in \mathbb{R}$ .

To find the expectation and variance, we do

$$E[X] = \frac{d}{dt}\varphi_X(t)\Big|_{t=0} = n\left(pe^t + 1 - p\right)^{n-1} \cdot pe^t\Big|_{t=0} = np$$

and

$$\begin{split} E[X^2] &= \frac{d^2}{dt^2} \varphi_X(t) \Big|_{t=0} \\ &= npe^t \left( pe^t + 1 - p \right)^{n-1} + (n-1)np^2 e^{2t} \left( pe^t + 1 - p \right)^{n-2} \Big|_{t=0} \\ &= np(1 + np - p) \end{split}$$

and conclude that

$$Var(X) = np + np(n-1)p - n^2p^2 = np(1-p).$$

#### Exercise 1.3.2

*For*  $X \sim \text{Poi}(\lambda)$ *, show that* 

$$\operatorname{mgf}(X) = \varphi_X(t) = e^{\lambda(e^t - 1)}$$
, for  $t \in \mathbb{R}$ .

Find E[X] and Var(X) using the mgf.

#### Exercise 1.3.3

*For*  $X \sim \text{Exp}(\lambda)$ *, show that* 

$$\operatorname{mgf}(X) = \varphi_X(t) = \frac{\lambda}{\lambda - t}$$
, for  $t < \lambda$ .

Find E[X] and Var(X) using the mgf.

## 66 Note 1.3.2 (Important property of the MGF)

The mgf is important to us because it uniquely determines the distribution of an rv.

## 1.4 Joint Distributions

We shall only review bivariate distributions. One can easily extend the bivariate case to a multivariate situation.

## **■** Definition 8 (Joint CDF)

The joint cdf is defined by

$$F(a,b) = P(X \le a, Y \le b) = P(\{X \le a\} \cap \{Y \le b\})$$

for all  $a, b \in \mathbb{R}$ , where X, Y are rvs.

## **■** Definition 9 (Marginal CDF)

Given a joint cdf F, we define the marginal cdf as

$$F_X(a) = P(X \le a) := F(a, \infty) = \lim_{b \to \infty} F(a, b).$$

## **■** Definition 10 (Joint and Marginal Probability Mass Functions)

Suppose X,Y are discrete rvs. We define the joint probability mass function (pmf) of X and Y as

$$p(x,y) := P(X = x, Y = y).$$

The marginal pmf of X and Y are

$$p_X(x) = P(X = x) = \sum_y p(x, y)$$

and

$$p_Y(y) = P(Y = y) = \sum_{x} p(x, y),$$

respectively.

We are not equipped with the knowledge to formally define a **probability density function**, and so we shall only introduce it in a roundabout way.

## **■** Definition 11 (Joint and Marginal Probability Density Function)

Suppose X, Y are continuous rvs. We define joint probability density

$$f: \mathbb{R}^2 \to \mathbb{R}$$
.

The marginal pdf of X and Y are

$$f_X(x) \coloneqq \int_{\forall y} f(x,y) \, dy$$

and

$$f_Y(y) := \int_{\forall x} f(x,y) \, dx,$$

respectively.

## 66 Note 1.4.1

*Note that* 

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y).$$

## 1.5 Expectation of Joint Distributions

## **■** Definition 12 (Expectation of Joint Distributions)

Let X and Y be ros, and g(X,Y) a function. We define the expectation of a joint distribution of X and Y as

$$E[g(X,Y)] = \begin{cases} \sum_{x} \sum_{y} g(x,y) p(x,y) & X,Y \text{ are jointly discrete} \\ \int_{\forall x} \int_{\forall y} g(x,y) f(x,y) & X,Y \text{ are jointly continuous} \end{cases}$$

The following is a special case of an expectation, given for when

$$g(X,Y) = (X - E[X])(Y - E[Y]).$$

## **■** Definition 13 (Covariance of Joint Distributions)

#### ₩ Warning

It is important to note that

$$f(x,y) \neq P(X=x,Y=x),$$

and so we cannot put ■ Definition 10 and ■ Definition 11 as one definition.

We define the covariance of a joint distribution of X and Y as

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

## **66** Note 1.5.1

Observe that

$$Cov(X, X) = Var(X).$$

## ♦ Proposition 5 (Linearity of the Expectation over Addition)

Exercise 1.5.1

Prove Proposition 5.

Let X and Y be rvs. Then for  $a, b \in \mathbb{R}$ ,

$$E[aX + bY] = aE[X] + bE[Y].$$

## ♦ Proposition 6 (Variance over a Linear Combination of RVs)

Exercise 1.5.2

Prove 🌢 Proposition 6.

Let X and Y be rvs. Then for  $a, b \in \mathbb{R}$ ,

$$Var(aX + bY) = a^{2} Var(X) + b^{2} Var(Y) + 2ab Cov(X, Y).$$

## 1.6 Independence of Random Variables

## **■** Definition 14 (Independence)

Let X and Y be rvs, and F their joint cdf. We say that X and Y are independent if

$$F(x,y) = F_X(x) \cdot F_Y(y),$$

where  $F_X$  and  $F_Y$  are the marginal cdfs of X and Y, respectively. We shall denote this relationship between rvs as  $X \perp Y$ .

## 66 Note 1.6.1 (Equivalent definition of Independence)

One may also define independence of X and Y by

$$p(x,y) = p_X(x) \cdot p_Y(y)$$

if X and Y are discrete, and

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

if X and Y are continuous.

## ♦ Proposition 7 (Expectation of Independent RVs)

Let X and Y be independent rvs, and g a function of X and h a function of Y. Then

$$E(g(X)h(Y)) = E[g(X)] \cdot E[h(Y)].$$

#### Corollary 8 (MGF of Independent RVs)

Suppose  $X_1, X_2, ..., X_n$  are independent rvs, then consider  $T = \sum_{i=1}^n X_i$ . The mgf of T is then given by

$$\varphi_T(t) = \prod_{i=1}^n \varphi_{X_i}(t).$$

## Proof

Suppose  $X_1, X_2, ..., X_n$  are independent rvs, then consider  $T = \sum_{i=1}^{n} X_i$ . Then the mgf of T is

$$\varphi_T(t) = E\left[e^{tT}\right] = E\left[\exp\{t(X_1 + X_2 + \dots + X_n)\}\right]$$

$$= E\left[e^{tX_1} \cdot \dots \cdot e^{tX_n}\right]$$

$$= E\left[e^{tX_1}\right] \dots E\left[e^{tX_n}\right] \quad \therefore \quad \blacktriangle \quad Proposition \ 7$$

$$= arphi_{X_1}(t) \cdot \ldots \cdot arphi_{X_n}(t) = \prod_{i=1}^n arphi_{X_i}(t),$$

which is what we want.

## Example 1.6.1 (MGF of Independent Binomial Distributions with the Same Probability)

Let  $X_1, X_2, \ldots, X_m$  be an independent sequence of rvs where  $X_i \sim \text{Bin}(n_i, p)$ . Let  $T = \sum_{i=1}^m X_i$ . By  $\longleftarrow$  Corollary 8 mgf of T is

$$\varphi_T(t) = \prod_{i=1}^m \varphi_{X_i}(t) = \prod_{i=1}^m (pe^t + 1 - p)^{n_i} = (pelt + 1 - p)^{\sum_{i=1}^m n_i}.$$

We observe that

$$T \sim \operatorname{Bin}\left(\sum_{i=1}^{m} n_i, p\right).$$



## Conditional Probabilities

## 2.1 Conditional Probability for Discrete Random Variables

## ■ Definition 15 (Conditional Distribution of Discrete Random Variables)

Let  $X_1$ ,  $X_2$  be discrete rvs with joint pmf  $p(x_1, x_2)$  and marginal pmfs  $p_1(x_1)$  and  $p_1(x_2)$ , respectively. We call  $X_1 \mid X_2 = x_2$  the conditional distribution of  $X_1$  given  $X_2 = x_2$ .

The conditional pmf of  $X_1 \mid X_2 = x_2$  is defined as

$$p(x_1 \mid x_2) = P(X_1 = x_2 \mid X_2 = x_2) = \frac{P(X_1 = x_1 \land X_2 = x_2)}{P(X_2 = x_2)},$$

or more succinctly,

$$p(x_1 \mid x_2) = P(X_1 \mid X_2 = x_2) = \frac{p(x_1, x_2)}{p_2(x_2)},$$

where we require  $p_2(x_2) > 0$ .

## 66 Note 2.1.1

• If  $X_1$  and  $X_2$  are independent, then

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p_2(x_2)} = \frac{p_1(x_1)p_2(x_2)}{p_2(x_2)} = p_1(x_1).$$

• We may extend the above definition to multivariate cases.

### Example 2.1.1

Suppose  $X_1$ ,  $X_2$ ,  $X_3$  are discrete rvs, with  $p_3$  as the pmf of  $X_3$ . Then we define the conditional pmf of  $(X_1, X_2) \mid X_3 = x_3$  as

$$p((x_1,x_2) \mid x_3) = \frac{p(x_1,x_2,x_3)}{p_3(x_3)}.$$

## **■** Definition 16 (Conditional Expectation for Discrete RVs)

Given rvs  $X_1$  and  $X_2$ , and a function g on  $X_1$ , We define the conditional expectation of  $X_1$  given  $X_2 = x_2$  as

$$E[g(X_1) \mid X_2 = x_2] = \sum_{x_1} g(x_1) p(x_1 \mid x_2).$$

#### **66** Note 2.1.2

By lacktriangle Proposition 5, we have that given rvs  $X_1$  an  $X_2$ ,  $a,b \in \mathbb{R}$ , and functions g,h on  $X_1$ ,

$$E [ag(X_1) + bh(X_1) | X_2 = x_2]$$

$$= aE[g(X_1) | X_2 = x_2] + aE[h(X_1) | X_2 = x_2]$$

## ♦ Proposition 9 (Conditional Variance of Discrete RVs)

We have that

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

### Proof

Consider

$$g(x) = [X_1 - E[X_1 \mid X_2 = x_2]]^2,$$

which is the typical definition of a variance. Then

$$Var(X_1 \mid X_2 = x_2)$$

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

$$= E[X_1^2 - 2X_1E[X_1 \mid X_2 = x_2] + E[X_1 \mid X_2 = x_2]^2 \mid X_2 = x_2]$$

$$= E[X_1^2 \mid X_2 = x_2] - 2E[X_1 \mid X_2 = x_2]^2 + E[X_1 \mid X_2 = x_2]^2$$

$$= E[X_1^2 \mid X_2 = x_2] - E[X_1 \mid X_2 = x_2]^2.$$

Even better, we have the following proposition.

## • Proposition 10 (Linearity of Conditional Expectation)

Given rvs  $X_1$ ,  $X_2$  and  $X_3$ , we have

$$E[X_1 + X_2 \mid X_3 = x_3] = E[X_1 \mid X_3 = x_3] + E[X_2 \mid X_3 = x_3].$$

#### Proof

Let  $p_3$  be the pmf of  $X_3$ . Observe that

$$E[X_1 + X_2 \mid X_3 = x_3]$$

$$= \sum_{x_1} \sum_{x_2} (x_1 + x_2) p(x_1, x_2 \mid x_3)$$

$$= \sum_{x_1} \sum_{x_2} x_1 \frac{p(x_1, x_2, x_3)}{p_3(x_3)} + \sum_{x_1} \sum_{x_2} x_2 \frac{p(x_1, x_2, x_3)}{p_3(x_3)}$$

$$= \sum_{x_1} \frac{x_1}{p_3(x_3)} \sum_{x_2} p(x_1, x_2, x_3) + \sum_{x_2} \frac{x_2}{p_3(x_3)} \sum_{x_1} p(x_1, x_2, x_3)$$

$$= \sum_{x_1} \frac{x_1}{p_3(x_3)} p(x_1, x_3) + \sum_{x_2} \frac{x_2}{p_3(x_3)} p(x_2, x_3)$$

$$= \sum_{x_1} x_1 p(x_1 \mid x_3) + \sum_{x_2} x_2 p(x_2 \mid x_3)$$

$$= E[X_1 \mid X_3 = x_3] + E[X_2 \mid X_3 = x_3].$$

## Corollary 11 (General Linearity of Conditional Expectation)

Given an rv Y, a sequence of discrete rvs  $\{X_i\}_{i=1}^n$ , and a sequence of scalars  $\{a_i\}_{i=1}^n \subseteq \mathbb{R}$ , we have

$$E\left[\sum_{i=1}^{n} a_i X_i \mid Y = y\right] = \sum_{i=1}^{n} a_i E[X_i \mid Y = y].$$

### Example 2.1.2

Suppose that *X* and *Y* are discrete rvs having joint pmf

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

Let us first find the conditional distribution of  $X \mid Y = 1$ . First, consider the following table:

$X \setminus Y$	0	1	2	$p_X$
0	0	2 15 2 5	0	$\frac{2}{15}$
1	$\frac{1}{5}$	<u>2</u> 5	$\frac{1}{15}$	$\frac{2}{3}$
2	1 5 1 5	0	0	2 15 2 3 1 5
$p_Y$	<u>2</u> 5	8 15	1 15	

Observe that

$$p(0,1) = \frac{\frac{2}{15}}{\frac{8}{15}} = \frac{1}{4} \text{ and } p(1,1) = \frac{\frac{2}{5}}{\frac{8}{15}} = \frac{3}{4}.$$

Thus the conditional distribution of  $X \mid Y = 1$  is given as in Table 2.2.

$$\begin{array}{c|cccc} X & 0 & 1 \\ \hline p(x \mid 1) & \frac{1}{4} & \frac{3}{4} \end{array}$$

Table 2.2: Conditional distribution of  $X \mid Y = 1$ 

With this we can calculate the conditional expectation and conditional variance of  $X \mid Y = 1$ . Note that

$$X \mid Y = 1 \sim \text{Bernoulli}\left(\frac{3}{4}\right).$$

Table 2.1: Tabulating values of p(x,y)

Thus

$$E[X \mid Y = 1] = \frac{3}{4} \text{ and } Var(X \mid Y = 1) = \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}.$$

#### Example 2.1.3

For i = 1, 2 suppose that  $X_i \sim \text{Bin}(n_i, p)$ , where  $X_1 \perp X_2$ . We want to find the conditional distribution of  $X_1$  given  $X_1 + X_2 = n$ , i.e. the conditional distribution of  $X_1 \mid X_1 + X_2 = n$ .

First, note that the sum of two **binomial distributions** is also a binomial distribution, where the number of trials is the sum of the number of trials from each distribution. In particular, we have that  $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$ .

Let the conditional pmf of  $X_1 \mid X_1 + X_2 = n$  be denoted as  $p(x_1 \mid n)$ . Then

$$p(x_1 \mid n) = P(X_1 = x_1 \mid X_1 + X_2 = n)$$

$$= \frac{P(X_1 = x_1, X_1 + X_2 = n)}{P(X_1 + X_2 = n)}$$

$$= \frac{P(X_1 = x_1, X_2 = n - x_1)}{P(X_1 + X_2 = n)}$$

$$= \frac{P(X_1 = x_1)P(X_2 = n - x_1)}{P(X_1 + X_2 = n)} :: X_1 \perp X_2$$

$$= \frac{\binom{n_1}{x_1}p^{x_1}(1-p)^{n_1-x_1}\binom{n_2}{n-x_1}p^{n-x_1}(1-p)^{n_2-n+x_1}}{\binom{n_1+n_2}{n}p^n(1-p)^{n_1+n_2-n}}$$

$$= \frac{\binom{n_1}{x_1}\binom{n_2}{n-x_1}}{\binom{n_1+n_2}{n}},$$

for  $0 \le x_1 \le n_1$  and  $0 \le n - x_1 \le n_2$ . We observe that  $X_1 \mid X_1 + X_2 = n$  has a **Hypergeometric Distribution**, i.e.

$$X_1 \mid X_1 + X_2 = n \sim HG(n_1 + n_2, n_1, n).$$

Recall that the intuition to understanding the hypergeometric distribution in this case is as follows: if we say that  $1, 2, ..., n_1$  are the indexed trials of  $X_1$  and  $1, 2, ..., n_2$  are those of  $X_2$ , then if we arrange the trials as

$$1, 2, \ldots, n_1, 1, 2, \ldots, n_2,$$

we may then think that we want to calculate the probability of getting  $x_1$  successes given that the rest of the  $n - x_1$  are failures.

Using the formulas for the expectation and variance of a hypergeometric distribution, we have that

$$E[X_1 \mid X_1 + X_2 = n] = \frac{n(n_1)}{n_1 + n_2}$$

and

$$Var(X_1 \mid X_1 + X_2 = n) = \frac{n(n_1)(n_1 + n_2 - n_1)(n_1 + n_2 - n)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$
$$= \frac{n(n_1)(n_2)(n_1 + n_2 - n)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

Let's deal with a more general case of the above, but this time with the individual rvs following  $Poi(\Lambda_i)$ .

## Example 2.1.4

Let  $\{X_i\}_{i=1}^m$  be a sequence of independent rvs where  $X_i \sim \text{Poi}(\Lambda_i)$ , and i = 1, 2, ..., m. Let  $Y = \sum_{i=1}^m X_i$ . Let us try to deduce the conditional distribution of  $X_i \mid Y = n$ .

First, note that a sum of independent Poisson rvs is also a Poisson distribution, with each of its mean (which is also its parameter) summed up, i.e. given any  $Z_1 \sim \text{Poi}(\lambda_1), \ldots, Z_k \sim \text{Poi}(\lambda_k)$ , we have

$$\sum_{i=1}^k Z_i \sim \operatorname{Poi}\left(\sum_{i=1}^k \lambda_i\right).$$

Also a subtle point, note that

$$Z_j \perp \sum_{\substack{i=1\\i\neq j}}^k Z_i.$$

Observe that

$$P(X_j = x_j \mid Y = n)$$

$$= \frac{P(X_j = x_j, Y = n)}{P(Y = n)}$$

$$\begin{split} &= \frac{P\left(X_{j} = x_{j}, \sum\limits_{\substack{i=1\\i\neq j}}^{m} X_{i} = n - x_{j}\right)}{P(Y = n)} \\ &= \frac{P(X_{j} = x_{j})P\left(\sum\limits_{\substack{i=1\\i\neq j}}^{m} X_{i} = n - x_{j}\right)}{P(Y = n)} \\ &= \frac{\exp(-\lambda_{j})\lambda_{j}^{x_{j}}\left(\frac{1}{x_{j}!}\right) \cdot \exp\left(-\left(\sum\limits_{\substack{i=1\\i\neq j}}^{m} \lambda_{i}\right)\right)\left(\sum\limits_{\substack{i=1\\i\neq j}}^{m} \lambda_{i}\right)^{n - x_{j}}\left(\frac{1}{(n - x_{j})!}\right)}{\exp\left(-\left(\sum\limits_{i=1}^{m} \lambda_{i}\right)\right)\left(\sum\limits_{i=1}^{m} \lambda_{i}\right)^{m}\left(\frac{1}{n!}\right)} \\ &= \binom{n}{x_{j}}\frac{\lambda_{j}^{x_{j}}\left(\sum\limits_{i=1}^{m} \lambda_{i} - \lambda_{j}\right)^{n - x_{j}}}{\left(\sum\limits_{i=1}^{m} \lambda_{i}\right)^{x_{j}}\left(\sum\limits_{i=1}^{m} \lambda_{i}\right)^{n - x_{j}}} \\ &= \binom{n}{x_{j}}\left(\frac{\lambda_{j}}{\sum_{i=1}^{m} \lambda_{i}}\right)^{x_{j}}\left(1 - \frac{\lambda_{j}}{\sum_{i=1}^{m} \lambda_{i}}\right)^{n - x_{j}}, \end{split}$$

where  $x_i \in [0, n]$ . We see that

$$X_j \mid Y = n \sim \operatorname{Bin}\left(n, \frac{\lambda_j}{\sum_{i=1}^m \lambda_i}\right).$$

#### Exercise 2.1.1

It is a straightforward exercise to compute  $E[X_j \mid Y = n]$  and  $Var(X_j \mid Y = n]$ n).

## Example 2.1.5

Suppose  $X \sim \text{Poi}(\lambda)$  and  $Y \mid X = x \sim \text{Bin}(x, p)$ . Let us compute the conditional distribution of  $X \mid Y = y$ . We have that

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

$$= \frac{p(y \mid x) \cdot P(X = x)}{\sum_{x} p(y \mid x) P(X = x)}$$

$$= \frac{\exp(-\lambda)\lambda^{x} \left(\frac{1}{x!}\right) \cdot {\binom{x}{y}} p^{y} (1 - p)^{x - y}}{\sum_{x = y}^{\infty} \exp(-\lambda)\lambda^{x} \left(\frac{1}{x!}\right) \cdot {\binom{x}{y}} p^{y} (1 - p)^{x - y}}$$

$$= \frac{\exp(-\lambda)\lambda^{x} \cdot \frac{1}{y!(x-y)!} p^{y} (1-p)^{x-y}}{\sum_{x=y}^{\infty} \exp(-\lambda)\lambda^{x} \cdot \frac{1}{y!(x-y)!} p^{y} (1-p)^{x-y}}$$

Now notice that we may work out the denominator to be

$$\begin{split} &\sum_{x=y}^{\infty} e^{-\lambda} \lambda^x \frac{1}{y!(x-y)!} p^y (1-p)^{x-y} \\ &= \frac{e^{-\lambda}}{y!} p^y \sum_{x=y}^{\infty} \frac{\lambda^x}{(x-y)!} (1-p)^{x-y} \\ &= \frac{e^{-\lambda} p^y \lambda^y}{y!} \sum_{x=y}^{\infty} \frac{\lambda^{x-y} (1-p)^{x-y}}{(x-y)!} \\ &= \frac{e^{-\lambda} p^y \lambda^y}{y!} \sum_{x-y=0}^{\infty} \frac{(\lambda (1-p))^{x-y}}{(x-y)!} \\ &= \frac{e^{-\lambda + \lambda (1-p)} p^y \lambda^y}{y!} \\ &= \frac{e^{-\lambda p} (\lambda p)^y}{y!}, \end{split}$$

where the second last equality follows since

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

by the Taylor Expansion of the exponential function. Thus we have

$$p(x \mid y) = \frac{\exp(-\lambda)\lambda^{x} \cdot \frac{1}{y!(x-y)!} p^{y} (1-p)^{x-y}}{\frac{(\lambda p)^{y} e^{\lambda p}}{y!}}$$
$$= \frac{e^{-\lambda(1-p)} (\lambda(1-p))^{x-y}}{(x-y)!},$$

for  $x \in [y, \infty)$ , which is a Poisson-like pmf.

We say that

$$X \mid Y = y \sim W + y$$

where

$$W \sim \text{Poi}(\lambda(1-p)).$$

This distribution, W + y, is called a **shifted Poisson** distribution.

Observe that it is relatively easy to compute the expectation and

variance due to their linearity properties: we have

$$E[X \mid Y = y] = E[W + y] = E[W] + y = \lambda(1 - p) + y$$

and

$$Var(X \mid Y = y) = Var(W + y) = Var(W) = \lambda(1 - p).$$

## 2.2 Conditional Probability for Continuous Random Variables

In the jointly discrete case, it was natural to define

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

due to Bayes' Rule, or simply by realizing that we are trying to determine the probability of X = x and Y = y given that we already know the probability of Y = y.

This, however, does not make sense in the continuous case, especially since  $f(x, y) \neq P(X = x, Y = y)$ , and  $f_Y(y) \neq P(Y = y)$ .

<sup>1</sup> We can use a similar definition of a single variable continuous rv and extend it for 2 rvs: we can consider

$$f(x,y) = \lim_{\substack{dx \to 0 \\ dy \to 0}} \frac{P(x \le X \le x + dx, y \le Y \le y + dy)}{dx \, dy}.$$
 (2.1)

Notice that for small dx and dy, we have

$$P(\underbrace{x \le X \le x + dx}_{A} \mid \underbrace{y \le Y \le y + dy}_{B}) = \frac{P(A \cap B)}{P(B)}$$

$$\approx \frac{f(x, y) \, dx \, dy}{f_{Y}(y) \, dy}$$

$$= \frac{f(x, y)}{f_{Y}(y)} \, dx.$$

## Definition 17 (Conditional Distribution of Continuous Random Variables)

Let X and Y be continuous rvs, with joint pdf f(x,y), and Y has the marginal pdf  $f_Y$ . We call  $X \mid Y = y$  the conditional distribution of X <sup>1</sup> I am actually not sure how this paragraph inspires our definition, cause some things don't seem to match up nicely.

given Y = y, whose pdf is defined as

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \lim_{\substack{dx \to 0 \\ dy \to 0}} \frac{P(x \le X \le x + dx, y \le Y \le y + dy)}{dx}.$$

## Example 2.2.1

Suppose the joint pdf of *X* and *Y* is given by

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

First, note that the **region of support** for X and Y is given as in Figure 2.1.

We compute the conditional pdf of  $X \mid Y = y$ :

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

$$= \frac{\frac{12}{5}x(2-x-y)}{\int_0^1 \frac{12}{5}x(2-x-y) dx}$$

$$= \frac{2x-x^2-xy}{x^2-\frac{1}{3}x^3-\frac{1}{2}x^2y\Big|_{x=0}^{x=1}}$$

$$= \frac{2x-x^2-xy}{1-\frac{1}{3}-\frac{1}{2}y}$$

$$= \frac{2x-xy-x^2}{\frac{2}{3}-\frac{1}{2}y}$$

$$= \frac{12x-6xy-6x^2}{4-3y}, \text{ for } 0 < x < 1.$$

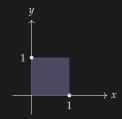


Figure 2.1: Region of support as a rectangle



## **Index**

Gamma Distribution, 13

Bayes' Formula, 11 Bayes' Rule, 11	Joint Probability Mass Function,		
Bernoulli Distribution, 12			
Binomial Distribution, 12	Law of Total Probability, 10		
Conditional Distribution, 23, 31 Conditional Expectation, 24 Conditional Probability, 9, 23 Covariance of Joint Distributions, 20 cumulative distribution function, 11, 13	Marginal CDF, 18  Marginal Probability Density Function, 18  Marginal Probability Mass Function, 18  Moment, 14  Moment Generating Function, 15		
Expectation, 14 Expectation of Joint Distributions, 19 Exponential Distribution, 13	Negative Binomial Distribution, 12 partition, 10 Poisson Distribution, 12		
Gamma Distribution 13	Poisson Distribution, 12		

Geometric Distribution, 12

Hypergeometric Distribution, 12

region of support, 32

Independence, 20

shifted Poisson, 30

Joint CDF, 18

Joint Probability Density Func-

probability density function, 13

tion, 18 Uniform Distribution, 13

Variance, 14