STAT333 - Applied Probability

CLASSNOTES FOR WINTER 2017

by

Johnson Ng

BMath (Hons), Pure Mathematics major, Actuarial Science Minor

University of Waterloo

Table of Contents

List of Definitions

List of Theorems

Foreword

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.

1 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. $0 \leq P(A) \leq 1$
2. $P(S) = 1$ [1]
3. $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$, *where* $A_i \cap A_j = A_i A_j = \varnothing$ *for all* $i \neq j$ [2]

[1] This can also be stated as $P(\varnothing) = 0$, where $\varnothing$ is the null event.

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

Note

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*

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

Note

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

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

where the second last step is by Item 3, and the last step is by Definition 2.
$\square$

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

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

Note

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

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

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 $X \sim \text{NB}(k, p)$, where $k \in \mathbb{N}$ and $p \in [0,1]$, its pmf is

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

The Negative Binomial Distribution has a model that measures the probability that the kth success occurs.

Geometric Distribution Following the above distribution where $k = 1$, we have that X folows 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 $X \sim \text{HG}(N, rn)$, 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

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

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

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

Note

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 $X \sim \mathrm{Unif}(a, b)$, 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 $X \sim \mathrm{Gam}(n, \lambda)$, 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 $X \sim \mathrm{Exp}(\lambda)$, where its pdf is

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

1.3 *Moments*

Definition 3 (Expectation)

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

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*

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

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

Definition 6 (Standard Deviation)

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

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

We shall state the following properties without providing proof[3]:

[3] The proofs are very easy, but it serves as a strengthening exercise for the unfamiliar. Therefore,

Exercise 1.3.1
Proof both Proposition 3 and Proposition 4.

Proposition 3 (Linearity of the Expectation)

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

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

Proposition 4 (Linearity of the Variance)

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

$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).$$

Referring back to 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*

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

Note

1. *Observe that $\phi_X(0) = E\left[e^0\right] = 1$.*

2. *The reason such an expression is called a moment generating function is as follows: observe that*

$$\begin{aligned}\phi_X(t) = E\left[e^{tX}\right] &= E\left[\sum_{i=0}^{\infty} \frac{(tX)^i}{i!}\right] \\ &= E\left[1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \ldots + \frac{(tX)^n}{n!} + \ldots\right] \\ &= \frac{t^0}{0!}E[1] + \frac{t}{1!}E[X] + \frac{t^2}{2!}E\left[X^2\right] + \ldots + \frac{t^n}{n!}E\left[X^n\right] + \ldots\end{aligned}$$

by 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,

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

Index