

# MATH 318 Review Notes

Reese Critchlow

## Basic Concepts in Probability

Permutations: For some set  $S = \{a, b, c\}$ , the number of orderings of the set, or the number of *permutations* of the set is given by the factorial, that is, that the number of permutations for a set of distinct elements of size  $n$  is  $n!$ .

Redundancies in Permutations: It is important to note, that in a set of size  $n$  where there are  $m$  identical/redundant elements with count  $m_i$ , the number of permutations is given by:

$$\frac{n!}{\prod_{i=1}^m m_i!}$$

Choosing and the Binomial Coefficient: The number of ways to choose  $k$  unique objects from a set of size  $n$  is given by the binomial coefficient:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

Multinomial Coefficient: Let  $b_1, \dots, b_k$  be nonnegative integers, and let  $n = b_1 + b_2 + \cdots + b_k$ . The multinomial coefficient  $\binom{n}{b_1, b_2, \dots, b_k}$  describes:

- The number of ways to put  $n$  interchangeable objects into  $k$  boxes, such that box  $i$  has  $b_i$  objects in it, for  $1 \leq i \leq k$ .
- The number of ways to choose  $b_1$  interchangeable objects from  $n$  objects, then  $b_2$  from the remaining, until you choose  $b_{k-1}$  from what remains.
- The number of unique permutations of a word with  $n$  letters  $k$  distinct letters, such that the  $i$ -th letter occurs  $b_i$  times.

The multinomial coefficient can be given by:

$$\frac{n!}{b_1!b_2!\cdots b_k!}$$

Mathematical Definition of Probability: A *probability* is a function that assigns to each  $E$  contained in  $S$ , a number  $P(E)$  such that:

1.  $0 \leq P(E) \leq 1$ ,  $E \subseteq S$
2.  $P(S) = 1$
3.  $E_i \cap E_j = \emptyset, \forall i \neq j \implies P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n)$

Basic Terminology:

- **Sample Space**,  $S$ , describes the set of all possible outcomes of an experiment. It is either continuous or discrete, finite or infinite.
- **Event**,  $E$ , is a subset of the sample space  $E \subseteq S$ .
- A **Probability Space** is a triplet containing a sample space  $S$ , a set of events  $E$ , and a probability function  $P$ , that is  $(S, E, P)$ .
  - Often is the fact that  $S$  is finite, and every outcome is equally likely, such that
$$P(E) = \frac{\# \text{ of outcomes in } E}{\# \text{ of outcomes in } S}$$

Properties of Probability:

1.  $\forall E, P(E) + P(E^C) = P(S) = 1$

Inclusion-Exclusion Formula: For a union of non-disjoint events, the probability of the union is given:

$$P(E_1 \cup \dots \cup E_n) = \sum_{i=1}^n P(E_i) - \sum_{i < j} P(E_i \cap E_j) + \sum_{i < j < k} P(E_i \cap E_j \cap E_k) - \dots + (-1)^{n-1} \sum_{i < \dots} P(E_1 \cap \dots \cap E_n)$$

Conditional Probability: Two events  $E$  and  $F$  are said to be independent events if:

$$P(E \cap F) = P(E)P(F) \iff P(E|F) = P(E)$$

Hence, it is important to define the notation for a conditional probability:

$$P(E|F)$$

describes the probability of “ $E$  given  $F$ ”.

To generalize, events  $E_1, \dots, E_n$  are said to be independent if:

$$P(E_1 \cap \dots \cap E_n) = P(E_1) \dots P(E_n).$$

Finally we can give some formulae:

$$P(E \cap F) = P(E|F)P(F) \qquad P(E|F) = \frac{P(E \cap F)}{P(F)}$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Law of Total Probability: Let  $F_1, \dots, F_n$  be a partition of  $S$ . Then,

$$P(E) = \sum_{i=1}^n P(E|F_i)P(F_i)$$

Bayes Formula: Building off of the law of total probability:

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$$

## Random Variables

---

Definition: A random variable is a function that takes an event space  $S$  and produces an event.

$$X : S \rightarrow E$$

The event  $E$  is often a real number  $\mathbb{R}$ .

### Discrete Random Variables

---

Discrete random variables are random variables that take on values in a countable set.

Probability Mass Function: The probability mass function (p.m.f.)  $P : E \rightarrow \mathbb{R}$  associated with a discrete random variable defines the probability that the random variable takes certain values.

$$P(E) = P(X = E)$$

Note: it is required that  $\sum_i p(x_i) = 1$ .

Notation: We say that a random variable  $X$  is distributed in a certain way with the notation  $X \sim \text{dist}(p)$ .

Geometric Random Variable	
Interpretation	The probability distribution of $i$ Bernoulli Trials (binary output) with probability $p$ of being successful before one success is obtained.
p.m.f.	$p(i) = (1 - p)^{i-1}p$
Symbol	$X \sim \text{geom}(p)$
Notes	Memoryless: $P(X > m + n   X > m) = P(x > n)$

Binomial Random Variable	
Interpretation	The binomial random variable describes the distribution of successful Bernoulli trials with probability of success $p$ over $n$ trials.
p.m.f.	$p(i) = \binom{n}{i} p^i (1 - p)^{n-i}$
Symbol	$X \sim \text{Bin}(n, p)$

Poisson Random Variable (with parameter $\lambda > 0$ )	
Interpretation	Arises as a simplification to the Binomial Random Variable. For some parameter $\lambda$ , with the expectation of $\lambda$ events in a given interval, it give the probability of $k$ events occurring in the same interval.
p.m.f.	$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$
Symbol	$X \sim \text{Bin}(n, p)$
Examples	Radioactive Decay: For $n$ atoms, each atom has probability $p$ of decaying in a 1 second interval. Thus $\lambda = np$ : the average number of decays in 1 second (determined empirically).

## Continuous Random Variables

Definition:  $X$  is a continuous random variable if

1. There exists a function  $f$  with  $f(x) \geq 0$ .
2.  $P(x \in B) = \int_B f(x)dx, \forall B \subseteq \mathbb{R}, x \in \mathbb{R}$ .

In this case  $f$  is called the probability density function for the random variable  $X$ , and  $f(a)$  indicates how likely it is for  $x$  to be near  $a$ , but importantly,  $f(a) \neq P(a)$ .

Cumulative Distribution Function: We can also define the Cumulative Distribution Function, or the c.d.f. for a random variable  $X$ :

$$F(a) = P(x \leq a) = P(x \in (-\infty, a]) = \int_{-\infty}^a f(x)dx.$$

Importantly, it follows that the relationship between the c.d.f. and the p.d.f. is that the p.d.f. is the derivative of the c.d.f., that is:

$$F'(x) = f(x).$$

Common Continuous Random Variables:

Uniform Random Variable	
Interpretation	All values in a range $[a, b]$ are equally likely.
p.m.f.	$p(x) = \frac{1}{b-a}$
Symbol	$X \sim \text{Unif}(a, b)$

Exponential Random Variable	
Interpretation	Time of occurrence for an unpredictable event.
p.m.f.	$p(x) = \lambda e^{-\lambda x}$
Symbol	$X \sim \text{Exp}(a, b)$
Notes	<p><u>Memoryless</u>: <math>P(x &gt; 2\tau \mid x &gt; \tau) = P(x &gt; \tau) = \frac{1}{2}</math></p> <p><u>Half Life, etc</u>: Exponential RVs are good for modelling exponential decay. We can then find the half life, such that <math>P(x &gt; \tau) = \frac{1}{2}</math>, which is the time where it is equally likely for the material to have decayed than for it to have not.</p>

Gaussian Random Variable	
Interpretation	Unsure
p.m.f.	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Symbol	$X \sim N(\mu, \sigma^2)$
Notes	<p><u>Scaling Property</u>: If <math>X</math> is normal (<math>X \sim N(\mu, \sigma^2)</math>), then <math>Y \sim N(0, 1)</math>, where <math>Y = \frac{X-\mu}{\sigma}</math>, and <math>X</math> is the value of the random variable. An example of this would be, if one is trying to find the probability that <math>X \in [a, b]</math>, with <math>X \sim N(\mu, \sigma^2)</math>, then the equivalent form would be to find the probability that <math>Y \in [\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}]</math>.</p>

## Expectation Value

We can define the expectation value  $\mathbb{E}$  for a continuous random variable  $X$  as:

$$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx.$$

Similarly, for a discrete random variable, we have:

$$\mathbb{E}X = \sum_{\text{all } i} k_i p(k_i)$$

Lemma: Suppose  $X$  is a continuous RV with p.d.f.  $f$  and  $f(x) = 0 \forall x < 0$ , then:

$$\mathbb{E}X = \int_0^{\infty} P(X > x)dx.$$

Similarly, for a discrete RV:

$$\mathbb{E}X = \sum_{n=0}^{\infty} P(X > n).$$

## Compositions of Random Variables

Theorem (Law of the Unconscious Statistician): For a continuous RV  $X$  with p.d.f.  $f(x)$  and a function  $g: \mathbb{R} \rightarrow \mathbb{R}$ , then:

$$\mathbb{E}g(X) = \int g(x)f(x)dx$$

This can also be applied to discrete RVs:

$$\mathbb{E}g(X) = \sum_i g(x_i)p(x_i)$$

Linearity of Expectation: Expectation values are linear, that is:

$$\mathbb{E}(aX + b) = a\mathbb{E}X + b$$

## Moments of Random Variables and Related Values

---

We can define the n<sup>th</sup> moment of  $X$  to be:

$$\mathbb{E}X^n = \int_{-\infty}^{\infty} x^n f(x) dx$$

or, in the discrete case:

$$\sum_i x_i^n p(x_i).$$

Variance: We define variance as:

$$\begin{aligned}\text{Var}(X) &= \sigma^2 = \mathbb{E}[(X - \mathbb{E}X)^2] \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2.\end{aligned}$$

Standard Deviation: Is the square root of the variance:

$$\sigma = \sqrt{\text{Var}(x)}$$

## Joint Distributions

---

Discrete random variables  $X, Y$  have a joint p.m.f.

$$p(x, y) = P(\{X = x\} \cap \{Y = y\}).$$

They also have marginal p.m.f.s:

$$P_X(x) = \sum_y p(x, y) = P(X = x) \qquad P_Y(y) = \sum_x p(x, y) = P(Y = y).$$

In the continuous case,  $X$  and  $Y$  are jointly continuous with p.m.f.  $f(x, y)$  if  $P((X, Y) \in C) = \iint_C f(x, y) dx dy$ . Thus, we can also define the marginal p.m.f. for continuous distributions:

$$P(x \in A) = P(X \in A, y \in \mathbb{R}) = \int_{D_X} f(x, y) dy$$

where  $D_X$  is the domain of  $X$  for a given  $y$ . And thus, similarly,

$$P(y \in B) = \int_{D_Y} f(x, y) dx$$

Law of the Unconscious Statistician for 2 Random Variables: Let  $X, Y$  be R.V.s and  $g$  a function, then:

$$\mathbb{E}g(x, y) = \iint g(x, y) f(x, y) dx dy \qquad \mathbb{E}g(x, y) = \sum_x \sum_y g(x, y) f(x, y)$$

This implies that  $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$ .

Independence of Random Variables: Two independent random variables  $X, Y$  are independent if for any  $a, b \in \mathbb{R}$ ,

$$P(\{X \leq a\} \cap \{Y \leq b\}) = P(\{X \leq a\}) P(\{Y \leq b\})$$

It also follows that  $X, Y$  are independent iff:

$$\begin{aligned} p(x, y) &= p_X(x)p_Y(y) \\ f(x, y) &= f_X(x)f_Y(y) \end{aligned}$$

Finally, one can also say that  $X, Y$  are independent if the matrix of the values is full rank.

Compositions of Functions with 2 Random Variables: Suppose that  $X, Y$  are independent RVs. Then,

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}g(X) \cdot \mathbb{E}h(Y).$$

Covariance: We can define the covariance of two RVs as:

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y.$$

From this, it is implied that if  $X, Y$  are independent, then  $\text{Cov}(X, Y) = 0$ , but importantly, for only the forwards direction. Covariance describes the nature of random variables to grow or shrink together.

Correlation Coefficient: The correlation coefficient is defined to be:

$$\rho(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

It can be interpreted as a “normalized” covariance.

Theorem (Cauchy Schwartz Inequality):

$$|\mathbb{E}(XY)|^2 \leq \mathbb{E}X^2 \cdot \mathbb{E}Y^2.$$

Sum of Variance: For two RVs  $X, Y$ , we can obtain the variance of their sum:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y),$$

for which if  $X, Y$  are independent, then the covariance term is zero, so

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Sums of Random Variables: For continuous RVs, we can find the p.m.f. of their sum under the convolution:

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy.$$

Gamma Distribution: We define another distribution, the gamma distribution,  $X \sim \text{gamma}(n, \lambda)$  with p.m.f.:

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