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:

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

Multinomial Coefficient: Let $b_1, \ldots, b_k$ be nonnegative integers, and let $n = b_1 + b_2 + \cdots b_k$. The multinomial coefficient $\binom{n}{b_1, b_2, \ldots, b_k}$ 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. $E_i \cap E_j = \emptyset, \forall i \neq j \implies P(E_1 \cup E_2 \cup \cdots E_n) = P(E_1) + P(E_2) + \cdots + P(E_n)$

Basic Terminiology:

- **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 face 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:

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

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 $E_1, \ldots, E_n$ are said to be independent if:

$$P(E_i \cap \cdots \cap E_n) = P(E_1) \cdots P(E_n).$$

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 $F_1, \cdots, F_n$ 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.

$$X : S \to E$$

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.) $P : E \to \mathbb{R}$ 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: $P(X > m+n \mid X > m) = P(x > n)$

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.	$p(i) = \binom{n}{i} p^i (1-p)^i$
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 λ, with the expectation of λ events in a given interval, it give the probability of k events occuring 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 emperically).

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	Memoryless: $P(x > 2\tau \mid x > \tau) = P(x > \tau) = \frac{1}{2}$ Half Life, etc: Exponential RVs are good for modelling exponential decay. We can then find the half life, such that $P(x > \tau) = \frac{1}{2}$, which is the time where it is equally likely for the material to have decayed than for it to have not.

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

Expectation Value

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

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

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:

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

Compositions of Random Variables

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

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

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:

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

In the continuous case, X and Y are jointly continuous with p.m.f. $f(x, y)$ if $P((X, Y) \in C) = \iint_C f(x, y)dxdy$. 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 Unconcious Statistician for 2 Random Variables: Let X, Y be R.V.s and g a function, then:

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

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:

$$p(x,y) = p_X(x)p_Y(y)$$
$$f(x,y) = f_X(x)f_Y(y)$$

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

<u>Compositions of Functions with 2 Random Variables:</u> Suppose that X, Y are independent RVs. Then,

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

<u>Covariance:</u> We can define the covariance of two RVs as:

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

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

<u>Correlation Coefficient:</u> The correlation coefficient is defined to be:

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

It can be interpreted as a "normalized" covariance.

<u>Theorem (Cauchy Schwartz Inequality):</u>

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

<u>Sum of Variance:</u> For two RVs X, Y, we can obtain the variance of their sum:

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

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

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

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

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

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

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

Midterm 1 Cutoff

Poisson Processes

Given some random variable N_t, which describes the number of events to occur by some time t, where each event happens sequentially, and has the the time to complete of $X_i \sim \text{Exp}(\lambda)$, this is called a Poisson Process:

- $\mathbb{E}N_t = \lambda t$.
- $P(N_t = m) = \frac{(\lambda t)^m e^{-\lambda t}}{m!}$.

Alternatively, we can also say that S_n is RV that represents the time it takes for the n-th event to occur. Hence, it follows that:

$$P(S_N > T) = P(N_T < N).$$

Moment Generating Functions and Characteristic Functions

We define the characteristic function of a random variable to be the following:

$$\phi(t) = \mathbb{E}e^{itx} = M(it) = \begin{cases} \sum_{n=-\infty}^{\infty} p(n)e^{itn} \\ \int_{-\infty}^{\infty} e^{itx} f(x)dx. \end{cases}$$

The terminology "moment generating functions" comes from the following property:

$$\left[\frac{d^n}{dt^n}\right]_{t=0} \phi(t) = i^n \mathbb{E}X^n.$$

We can also extract the pdf from a characteristic function:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt.$$

Properties of Characteristic Functions:

- $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$
- $P(X = c) = 1 \implies \phi_X(t) = e^{itc}$.

Continuity Theorem: Let $X_1, X_2, \ldots$ be rvs with CDFs $F_1, F_2, \ldots$ and characteristic functions $\phi_1, \phi_2, \ldots$, then:

(a) If $F_n \to F$ where F is the cdf of some rv X_i, with char fn. ϕ, then:

$$\lim_{n\to\infty} \phi_n(t) = \phi(t) \;\; \forall t \in \mathbb{R}.$$

(b) If $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ exists $\forall t \in \mathbb{R}$, then ϕ is the char. fn. of some RV X (with cdf F), and $F_n \to F$ and $X_n \to X$.

Limit Theorems

(Weak) Law of Large Numbers: Let X_i be i.i.d.. Assume $\mu = \mathbb{E}X_i < \infty$ and $Var(X_i) < \infty$. Let $S_n = X_1 + \cdots + X_n$. Then:

$$\frac{S_n}{n} \xrightarrow{\text{D}} \mu.$$

Central Limit Theorem: Let X_i be i.i.d. RVs with $\mathbb{E}X_i$ and $\text{Var}(X_i)$ finite. Let $S_n = X_1 + \cdots + X_n$. Then,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

We can interpret this by saying that $S_n - n\mu$ has fluctuations of $\sigma\sqrt{n}$.

Applying the Central Limit Theorem: Let X_i be some random variable with $\mathbb{E}X_i = \mu$ and $\text{Var}(X_i) = \sigma^2$, and $S_n = \sum_{i=1}^{n} X_i$. Then, it follows that:

$$P(S_n > a) \approx P\left(Z > \frac{a - n\mu}{\sigma\sqrt{n}}\right)$$

Statistical Estimation

Statistical Estimator: A statistical estimator is a function of the data:

- Sample Mean: $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$
- Sample Variance: $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{min}^{max}(X_i - \overline{X})^2$.

Unbiased Estimate: Given some data with parameter γ, if an estimate from the data for γ is $\hat{\gamma}$, then if $\mathbb{E}\hat{\gamma} = \gamma$, then $\hat{\gamma}$ is an unbiased estimate of γ.

Hypothesis Testing: Given some hypothesis of a mean value $\mu = b$, if μ is not in some confidence interval of probability p, then we reject the hypothesis that $\mu = b$.

Scenario 1: Given some i.i.d. data with known variance σ^2, but unknown mean μ, We can utilize the CLT to find the probability that the mean μ is in some range, with a given probability p (confidience interval).

$$P(|Z| < a) = p \implies \overline{X} \in \left[\mu - \frac{\sigma}{\sqrt{n}}a, \mu + \frac{\sigma}{\sqrt{n}}a\right] \iff \mu \in \left[\overline{X} - \frac{\sigma}{\sqrt{n}}a, \overline{X} + \frac{\sigma}{\sqrt{n}}\right]$$

Scenario 2: Given some i.i.d. data with both unkonwn mean μ and variance σ^2, then we utilize the CLT and the students-t distribution to determine the significance of the data.

$$P(|T| < a) = p \implies \mu \in \left[\overline{X} - a\frac{S}{\sqrt{n}}, \overline{X} + a\frac{S}{\sqrt{n}}\right].$$

Random Walks

Given some random walk in where each step by $\pm\vec{e}_i$ in $\mathbb{R}^n$ is equally likley, that is:

$$P(\vec{x}_i = \vec{e}_j) = \frac{1}{2d}$$

We say that the probability of returning to the origin is (where M is the number of visits to the origin):

- Recurrent: if $u = 1 \implies M = \infty$ (you always come back)
- Transient: if $u < 1 \implies M < \infty$

It is known that a random walk in $\mathbb{Z}^d$ is recurrent for $d = \{1, 2\}$ and transient for $d > 2$.

Gambler's Ruin

Assume that a "gambler" has k dollars and a "banker" has b dollars. They play games with probability p that the gambler wins, and play until one of the two goes broke. It follows that for k initial dollars for the gambler, and $N = k + b$ total dollars, then:

$$P(\text{gambler wins}) = \frac{\alpha^k - 1}{\alpha^N - 1} \qquad\qquad \alpha = \frac{1-p}{p}$$

Conditional Probability

Definition: The conditional pmf of X given that $Y = y$ is:

$$P_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)} = \frac{P(x, y)}{P_Y(y)}$$

Therefore, the conditional expectation can be described as:

$$\mathbb{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

Theorem: For X, Y discrete RVs:

$$\mathbb{E}X = \mathbb{E}\left[\mathbb{E}[X|Y]\right] = \sum \mathbb{E}\left[X|Y = y\right] P_Y(y).$$

For the continuous case:

Definition: The conditional pdf of X given that $Y = y$ is:

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

Therefore, it follows that:

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx.$$

As well as that:

$$\mathbb{E}X = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y) dy$$