STAT 134: Concepts of Probability

—Midterm Review Guide—

Disclaimer: This study guide was created by the Statistics Undergraduate Student Association, which is in no way affiliated with the current course staff of Statistics 134. Additionally, this guide is designed merely to *supplement* to your class notes.

Last Updated: Fall 2020

Originally compiled by: Ethan Marzban

Contents

1
Basics of Probability

- A quick review of sets and set theory may be useful:
  - A **set** is a collection of unordered **elements**. Elements do not need to be numbers; for example, {Blue, Gold} is the set of official Berkeley colors (go bears!)
  - The **union** of two sets is the set containing all the elements of each set, and the **intersection** of two sets is the set containing elements common to both sets. For example, if $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$ then $A \cup B = \{1, 2, 3, 4\}$ and $A \cap B = \{2, 3\}$.
  - The **empty set** (denoted $\varnothing$) is the set containing no elements. Two sets are said to be **mutually exclusive** (or **disjoint**) if $A \cap B = \varnothing$.
  - A **subset** A of a set B is a set containing some (possibly all) of the elements in B. For example, $\{2, 4\} \subseteq \{1, 2, 3, 4\}$. Two sets A and B are said to be **equal** if $A \subseteq B$ and $B \subseteq A$.
  - Here is a summary of some set-related concepts:

Union	$A \cup B$	$:= \{x : x \in A \text{ or } x \in B\}$
Intersection:	$A \cap B$	$:= \{x : x \in A \text{ and } x \in B\}$
Difference:	$A \backslash B$	$:= \{x : x \in A \text{ and } x \notin B\}$
Subset:	$A \subseteq B$	$x \in A \implies x \in B$
Equality:	$A = B$	$A \subseteq B$ and $B \subseteq A$
Proper Subset:	$A \subset B$	$A \subseteq B$ and $A \neq B$

- The **outcome space** (denoted Ω) is the set containing all possible outcomes of a particular setup. **Events** are simply subsets of the outcome space.
  - If all events $A \subseteq \Omega$ are **equally likely**, we define the **probability of the event A** to be

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)}$$

    Here $\#(\cdot)$ denotes the number of elements in a set.
  - A set of pairwise disjoint events $\{B_1, \dots, B_n\}$ (that is, $B_i \cap B_j = \varnothing$ for any $i \neq j$) is said to **partition** the event B if

$$\bigcup_{i=1}^{n} B_i = B_1 \cup \cdots \cup B_n = B$$

- The **three axioms of probability** state

  (a) $\mathbb{P}(A) \geq 0$ for any $A \subseteq \Omega$

  (b) $\mathbb{P}(\Omega) = 1$

  (c) For mutually exclusive events A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

  - Define the **complement** of an event A to be the unique event $\overline{A}$ (sometimes notated $A^{\complement}$) such that $\{A, \overline{A}\}$ partitions the outcome space Ω. Then, by axioms (b) and (c), we have that

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$$

- The **inclusion-exclusion rule** provides a way to compute the probability of the union of events, even if the events are not mutually exclusive. For 2 events A and B, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

More generally, for n events $A_1, \ldots, A_n$ we have

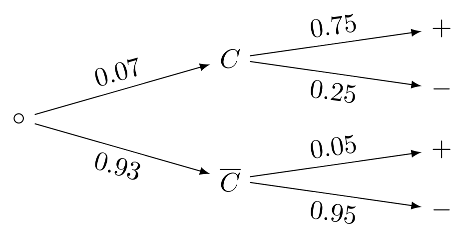
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

- **Conditional probabilities** are probabilistic quantities that reflect some change to the outcome space.

$$\mathbb{P}(A \mid B) = \frac{\#(A \cap B)}{\#(B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

The **multiplication rule** states that $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$

  - Two events A and B are said to be **independent** (notated $A \perp B$) if $\mathbb{P}(A \mid B) = \mathbb{P}(B)$. Alternatively, $A \perp B$ if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
  - **Probability Trees** can be useful in keeping track of conditional probabilities.
    * For example, suppose 7% of a population has a disease. Of those who have the disease, a test correctly identifies them as disease-positive 75% of the time. Of those who do not have the disease, the test correctly identifies them as disease-negative 95% of the time. The tree for this situation would be as follows:



Here, C denotes the event {person is actually a carrier}, $+$ denotes the event {the test tests positive}, and $-$ denotes the event {the test tests negative}.

- The **Rule of Average Conditional Probabilities** (also known as the **Law of Total Probability**) states that, for a partition $\{B_1, \ldots, B_n\}$ of the outcome space Ω,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) = \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \cdots + \mathbb{P}(A \mid B_n)\mathbb{P}(B_n)$$

That is, the probability of any event A can be computed as a weighted average of the probabilities of each event in a partition of Ω.

- **Bayes' Rule** provides another tool for evaluating conditional probabilities:

$$\begin{aligned}\mathbb{P}(B \mid A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}\end{aligned}$$

where $\{B_1, \ldots, B_n\}$ is a partition of Ω.

2 Random Variables and Distributions

- A **random variable** can be thought of as a measure of some random process. For example, if X denotes the number of heads in 2 tosses of a fair coin, then X is a random variable. The key idea is that X can take on different values, each with different probabilities.

- The **support** of a random variable is the set of all values the random variable is allowed to attain. For example, in the coin-tossing example above, X can be either 0, 1, or 2; it is impossible to toss 2 coins and observe more than 2 heads (or negative heads, for that matter).

- A **p.m.f.** (probability mass function) is an enumeration of the values of $\mathbb{P}(X = k)$ where X is a random variable and k is a value within the support of X. For instance, in the coin-tossing example:

$\boldsymbol{k}$	0	1	2
$\mathbb{P}(\boldsymbol{X = k})$	$(1/2)^2$	$(1/2)$	$(1/2)^2$

  The key to constructing tables (like the one above) is to translate each event into words. For example, $\{X = 2\}$ means "I toss two heads in two tosses of a fair coin." In this wording, it is clearer how to compute the associated probability.

  - The table above can be equivalently expressed as

$$\mathbb{P}(X = k) = \begin{cases} \binom{2}{k} (1/2)^k & \text{if } k = 0, 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

  - The **cumulative mass function** (CMF; notated $F_X(x)$) is defined to be $\mathbb{P}(X \leq x)$; the **survival** (sometimes notated $\overline{F_X}(x)$) is defined to be $\mathbb{P}(X > x)$.

- A **joint PMF** quantifies the probabilities associated with two related random variables, and is denoted $\mathbb{P}(X = x, Y = y)$.

  - Random variables X and Y are said to be **independent** (denoted $X \perp Y$) if $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.
  - A series of random variables $X_1, \dots, X_n$ are said to be **pairwise-independent** if $X_i \perp X_j$ for $i \neq j$. Note that pairwise independence does *not* imply independence, whereas independence *does* imply pairwise independence.
  - The **discrete convolution** provides a way of identifying the PMF of a sum of two random variables:

$$\mathbb{P}(X + Y = s) = \sum_{k=0}^{s} \mathbb{P}(X = k, Y = s - k)$$

- The **expected value** (or **expectation**) of a random variable is a measure of central tendency, and is defined to be

$$\mathbb{E}(X) := \sum_{k \in \text{ support}} k \cdot \mathbb{P}(X = k)$$

  The **variance** of a random variable is a measure of how "wide" a distribution is, and is defined to be

$$\text{Var}(X) := \mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

  The **standard deviation** is simply the square-root of variance: $\text{SD}(X) := \sqrt{\text{Var}(X)}$.

  - Expectation is linear: $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$. Variance is not: $\text{Var}(aX + b) = a^2\text{Var}(X)$.

– The expectation of a function of a random variable is given by the **Law of the Unconscious Statistician** (or LOTUS):

$$\mathbb{E}[g(X)] = \sum_{k \in \text{ support}} g(k)\mathbb{P}(X = k)$$

– For independent events $X_1, \ldots, X_n$, we have

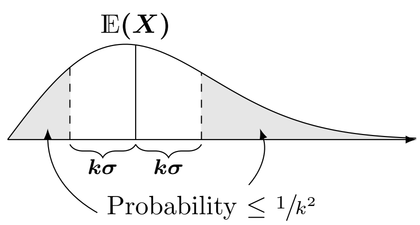
$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

If the events are not independent, the formula becomes a bit more complicated and requires material from chapter 6.

- There are two inequalities which can be used to identify an upper bound of probabilities *without* any knowledge of the underlying distribution:

  – **Markov's Inequality:** $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}(X)}{a}$ if $X \geq 0$, and if $a > 0$.

  – **Chebyshev's Inequality:** $\mathbb{P}\left[|X - \mathbb{E}(x)| \geq k \cdot \text{SD}(X)\right] \leq \dfrac{1}{k^2}$, for $k > 0$, and provided that the support of X contains only nonnegative numbers.



- An **indicator random variable** is a random variable defined as

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{If } A \text{ does not occur} \end{cases}$$

In this way, $\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A \text{ occurs})$.

  – Indicators are particularly useful in measuring counts. For example, let X denote the number of heads in 10 tosses of a $p-$coin. Then

$$X = \sum_{i=1}^{n} \mathbb{1}_{T_i} \quad \text{where} \quad \mathbb{1}_{T_k} = \begin{cases} 1 & \text{if } i\text{th toss lands heads} \\ 0 & \text{if } i\text{th toss lands tails} \end{cases}$$

  – More abstractly, say $X = \mathbb{1}_A + \mathbb{1}_B + \mathbb{1}_C + \mathbb{1}_D$. Further suppose that events A and C have occurred, whereas B and D have not. Then $\mathbb{1}_A = \mathbb{1}_C = 1$ and $\mathbb{1}_B = \mathbb{1}_D = 0$, so $X = 1 + 0 + 1 + 0 = 2$, which is precisely the number of events that have occurred.

3 Counting and Combinatorics

- Suppose we wish to pick k objects from a total of n objects. For illustrative purposes, say we wish to pick 3 letters from the set of $n = 5$ letters $\{a, b, c, d, e\}$.

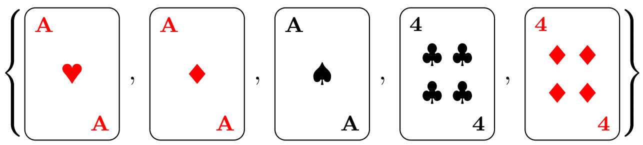
- If order matters (i.e. $\{a, b, c\}$ is not considered the same thing as $\{b, c, a\}$) then the number of ways to do this is

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

- If order does not matter (i.e. $\{a, b, c\}$ is considered the same thing as $\{b, c, a\}$), then the number of ways to do this is

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

- Always pick like objects together! It may be useful to demonstrate this through example. Given a poker hand of 5 cards drawn from a standard 52-card deck, we wish to compute the number of **full houses**. A full house is defined to be 3 cards of one rank, and 2 cards of another rank. For example,



We first find the number of ways to pick 3 cards from the first rank (in our example above this would be the number of ways to pick 3 aces from the deck): this number is $\binom{4}{3}$. Then we find the number of ways to pick 2 cards from the second rank (in our example above this would be the number of ways to pick 2 four's from the deck): this number is $\binom{4}{2}$.

Finally, we need to count the number of possible ranks we could have chosen for the three-of-a-kind: this is $\binom{13}{1}$. Then, from the remaining 12 ranks we pick one to be the rank of the two-of-a-kind: $\binom{12}{1}$. Putting everything together, the number of full houses is

$$\underbrace{\binom{13}{1}}_{\substack{\text{pick the rank}\\ \text{of the three-of-a-kind}}} \times \underbrace{\binom{4}{3}}_{\substack{\text{pick the cards}\\ \text{in the three-of-a-kind}}} \times \underbrace{\binom{12}{1}}_{\substack{\text{pick the rank}\\ \text{of the two-of-a-kind}}} \times \underbrace{\binom{4}{2}}_{\substack{\text{pick the cards}\\ \text{in the two-of-a-kind}}}$$

4 Approximations to the Binomial Distribution

- The **Standard Normal Distribution** is an example of a *continuous* distribution (continuous distributions will be discussed further after the midterm). The **standard normal distribution** (notated $\mathcal{N}(0, 1)$) has probability density function (the continuous analog of p.m.f's)

$$\phi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

and has cumulative density function (the continuous analog of c.m.f's)

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2} \, dx$$

The **normal distribution** (notated $\mathcal{N}(\mu, \sigma^2)$) is a nonstandardized version of the standard normal distribution with p.d.f.

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\left(\frac{X - \mu}{\sigma}\right) \sim \mathcal{N}(0, 1)$$

- Suppose $X \sim \text{Bin}(n, p)$. If p is not too small and if n is very large, then X is well approximated by the $\mathcal{N}(np, np(1-p))$ distribution.
- When using the normal approximation, it is advised to use the **continuity correction** to account for the fact that we are approximating a discrete random variable with a continuous one. Letting $X \sim \text{Bin}(n, p)$, we have

$$\mathbb{P}(X \le a) \approx \Phi\left(\frac{[a+0.5]-np}{\sqrt{np(1-p)}}\right)$$

$$\mathbb{P}(X \ge b) = 1 - \mathbb{P}[X \le (b-1)] \approx 1 - \Phi\left(\frac{[b-0.5]-np}{\sqrt{np(1-p)}}\right)$$

- Quantiles of the normal distribution cannot be obtained analytically; the use of a table (or computing software) is required.

- The **Poisson Distribution** (notated $\text{Pois}(\mu)$) is a discrete distribution with p.m.f.

$$\mathbb{P}(X = x) = e^{-\mu} \cdot \frac{\mu^x}{x!} \qquad x \in \{0, 1, 2, \dots\}$$

  - If $X \sim \text{Bin}(n, p)$ and p is very small or very large, then X is **not** well-approximated by a normal distribution and is better approximated by a $\text{Pois}(np)$ distribution.

- **Example:** Consider a coin that lands heads with probability $p = 0.4$. If I toss this coin 100 times and let X denote the number of heads in these 100 tosses, then X is approximately $\mathcal{N}(40, 24)$ and the probability of tossing 30 or less heads is approximately

$$\Phi\left(\frac{30.5-40}{\sqrt{24}}\right) \approx 0.02623975$$

  The exact answer, using the binomial distribution directly, is 0.02061342 so we see the error in approximation is quite small.

- **For the Mathematically Curious:** You might ask what we mean when we say that a distribution "approximates" another distribution. This is actually a deeper question that delves into topics relating to **notions of convergence**, and will be discussed further in Stat 135. If you're curious, you can look up the topics of **convergence in distribution** and **convergence in probability**.

5 Tips & Tricks

- When asked to compute the expectation of a quantity, there are three main tricks you can use:
  - **(i) The definition of expectation**. Though sometimes useful, this often leads to a lot of algebra (which in turn can lead to errors!).
  - **(ii) Indicators**. Again, if there's a count involved, see if you can use indicators.
  - **(iii) Relations.** If you're trying to find $\mathbb{E}(X)$, can you write X as the sum of other known random variables? For example, if $X \sim \text{Bin}(2, p)$ you can write $X = B_1 + B_2$ where $B_1, B_2 \overset{\text{i.i.d.}}{\sim} \text{Bern}(2, p)$ so $\mathbb{E}(X) = \mathbb{E}(B_1 + B_2) = \mathbb{E}(B_1) + \mathbb{E}(B_2) = 2p$. This is a lot easier than using the definition!

- **Maxes go with CMF's, Min's go with Survivals.** Consider random variables X_1, X_2, X_3. If the max of these RV's is less than k, it automatically follows that all three RV's must also be less than k. Similarly, if the minimum is greater than c, all three RV's must be greater than c.

– ⚠ Be careful though! A common mistake is to write something like this:

$$\mathbb{P}(\max\{X_1, X_2, X_3\} \geq k) \implies \mathbb{P}(X_1 \geq k, X_2 \geq k, X_3 \geq k)$$

This is wrong!!! Suppose $X_1 = 2$, $X_3 = 5$, and $X_4 = 7$. Here, $\max\{X_1, X_2, X_3\} \geq 3$ however not all three RV's are greater than 3!

6 Exercises

Problem 1: Use what you know about distributions to evaluate each of the following sums:

(a) $\displaystyle\sum_{k=0}^{\infty} \frac{1}{k!}$

(b) $\displaystyle\sum_{k=0}^{n} \binom{n}{k}$

(c) $\displaystyle\sum_{k=r}^{\infty} \binom{k-1}{r-1}(1-p)^{k-r}$

(d) $\displaystyle\sum_{k=0}^{n} \left[\binom{n}{k}\right]^2$

Problem 2: At a carnival, 100 raffle tickets are divided equally among the 20 participants. Of these 100 tickets, 5 of them are winning tickets. Compute the probabilities of the following events:

(a) One participant receives all 5 winning tickets.

(b) There are exactly two winners (that is, only two people have winning tickets)

Problem 3: Suppose that 3% of the population has a certain disease. A test for the disease exists, however it is relatively imperfect: 20% of the time the test returns an inconclusive result, regardless of whether the person has or does not have the disease. Furthermore, 10% of the people who have the disease test negative, and 8% of people who are disease-free test positive. Given that a person tested positive, what is the probability that they have the disease?

Problem 4: If $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \text{Geom}(p)$ on $\{0, 1, 2, ...\}$, identify the distribution of $(X_1 + X_2)$.

Hint: Apply the discrete convolution formula to find $P(X_1 + X_2 = k)$, and recognize the resulting expression as the PMF of a known distribution.

Problem 5: Consider n independent events $A_1, A_2, A_3, ..., A_n$, where $P(A_i) = p_i$, for $i = 1, 2, ..., n$.

(a) Compute $\mathbb{P}(A_1 | A_2 \cup A_3)$.

(b) Find a simple expression for $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right)$ that does not involve a summation (that is, **don't** use the Inclusion-Exclusion Rule).

Problem 6: Alfred and Anne both (independently) roll a fair $k-$sided die. Let X denote the result of Alfred's roll and let Y denote the result of Anne's roll. Defining $Z := \max\{X, Y\}$, find $\mathbb{P}(Z = z)$.

7 Answers to Exercises

Problem 1: **(a)** e (Use the Poisson distribution)

(b) 2^n (Use the Binomial distribution)

(c) p^{-r} (Use the Negative Binomial distribution)

(d) $\binom{2n}{n}$ (Use the Hypergeometric distribution)

Problem 2: **(a)** $\frac{\binom{20}{1}\binom{5}{5}}{\binom{100}{5}} \approx 1.328 \times 10^{-6}$

(b) $\frac{\binom{20}{1}\binom{5}{1}\binom{19}{1}\binom{5}{4}+\binom{20}{1}\binom{5}{2}\binom{19}{1}\binom{5}{3}}{\binom{100}{5}} = (20)_2 \frac{\binom{5}{1}\binom{5}{4}+\binom{5}{2}\binom{5}{3}}{\binom{100}{5}}$

Problem 3: $\dfrac{(0.7)(0.03)}{(0.7)(0.03) + (0.08)(0.97)}$

Problem 4: $(X_1 + X_2) \sim \text{NegBin}(2, p)$

Problem 5: **(a)** $\dfrac{p_1 p_2 + p_1 p_3 - p_1 p_2 p_3}{p_2 + p_3 - p_2 p_3}$

(b) $1 - \prod_{i=1}^{n} p_i$

Problem 6: $\dfrac{(z+1)^2 - z^2}{4}$