MA 2621 Probability for Applications

Christopher Myers

June 13, 2020

Contents

1 Basic Concepts

1.1 Terminology

Random experiment A situation in which the outcomes are known, but which outcome actually happens is unknown until the event happens. For instance, tossing a coin is a random experiment because the outcomes (heads or tails) are known.

Sample space The set of all outcomes of a random experiment. Denoted by Ω. For example, in a coin toss, the sample space is $\{H, T\}$. If instead you toss a coin until a head is observed, $\Omega = \{H\}$. If you toss three coins $\Omega = \{HHH, HHT, HTH, \ldots TTT\}$

Event A subset of the sample space. If you know the sample space, take any subset of that and you have an event. For instance, $\{THT, HTT\}$ is an event. If $\Omega = \{1, 2, 3, 4, 5, 6\}$, the event of getting an odd number ($A = \{1, 3, 5\}$) is an event.

Events are usually denoted by capital letters, such as A, B, C, D, etc. Note that an event with no outcome is called a null event, and is denoted by Φ. This notation is borrowed from set notation.

Event Space The set of all events of the sample space Ω, otherwise known as the set of all subsets of Ω.

1.2 Properties of Events

Events, being subsets, have some set properties:

- $A \cup B$ is the set of all outcomes in either A or B. This is the union of A and B.
- $A \cap B$ is the set of outcomes in both A and B. This is the intersection of A and B.

There is also the axiom of probability. Let α be the event space of Ω. Then, probability assigns to every event $A \in \alpha$, a number $P(A)$, called the "probability of event A", satisfying the following axiom:

$$P : A \in \alpha \to P(A) \in [0, 1]$$

...where $P(A) \geq 0$ for any event $A \in \alpha$, if A and B are disjoint ($A \cap B = \Phi$) then $P(A \cup B) = P(A) + P(B)$, and $P(\Omega) = 1$. This means that the lowest probability is zero, but no probability can be greater than 1.

Example 1 Prove that $P(\Phi) = 0$.

For any event A, we know that $A \cap \Phi = \Phi$ — this is a set property. The probability function can be applied to both sides to form $P(A \cap \Phi) = P(\Phi)$. However, $A \cup \Phi = A$ and $P(A \cup \Phi) = P(A)$, so $P(A) + P(\Phi) = P(A)$, naturally meaning that $P(\Phi) = 0$.

Example 2 If $A_1, A_2, \ldots A_n$ are disjoint, then $P(A_1 \cup A_2 \cup \ldots A_n) = \sum\limits_{i=1}^{n} P(A_i)$ can be proven using the same logic as above.

Example 3 If the sample space consists of n equally divided outcomes, then $P(A) = \frac{n(A)}{n(\Omega)} = \frac{n(A)}{n}$, or the number of outcomes in A divided by the number of total outcomes.

Example 4 Consider an experiment of rolling a pair of four sided dice. The probability of each scenario is...

- The sum of the rolls is odd: $\Omega = \{(1, 1), (1, 2) \ldots (4, 4)\}$ for sixteen total outcomes. $P(A) = 1/2$.
- The sum of the rolls is even: Same as above, $P(B) = 1/2$.
- The first roll is equal to the second roll: $P(C) = 1/4$.

- The first roll is larger than the second roll: $P(D) = 3/8$?
- At least one roll is equal to four: $P(E) = 7$.

2 Properties of Probabilities

Let A, B, C be events...

- A^c or $(\bar{A})$ is the complement of event A, and $\bar{A} = \{x \in \Omega | X \notin A\}$. Furthermore, $P(\bar{A}) = 1 - P(\bar{A})$.
- $P(A) = P(A \cap B) + P(A \cap \bar{B})$, and $P(B) = P(B \cap \bar{A}) = P(B \cap \bar{A})$. This means that the probability of A is the same as the probability of the intersection of A and B, plus the intersection of A and B's complement.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Or, the union of A and B is the same as the probability of A and B added together, but minus their intersection.
- $P(A \cup B) \leq P(A) + P(B)$, since there may or ma not be an intersection between A and B.
- If $A \leq B$, then $P(A) \leq P(B)$. This one can b confusing — it means that if A is a subset of B, the probability of A is less than or equal to the probability of B.

3 Conditional Probability

Conditional probability is best explained using an example:

Example Consider rolling a six-sided die. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. Define event A as the case where an even number is rolled, and event B is where a 6 is rolled. That is to say, $A = \{2, 4, 6\}, B = \{6\}$. The probability of rolling a 6 is $\frac{1}{6}$, but suppose that we tack on the *condition* that the outcome is even. The notation used is something like "outcome is 6 | outcome is even", or "outcome is six *given that* outcome is even". Mathematically, this is $P(B|A)$. What is $P(B|A)$?

To answer this, we need to consider a new sample space. The new sample space is defined by the given condition A, which has only three elements. So, $P(B|A) = \frac{1}{3}$. Note that if the two events were disjoint, probability would be zero, e.g. $P(B|\neg A) = 0$.

From this example the definition of conditional probability is clear: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ where $P(B) > 0$.

Example 2 Using the previous example where $A = \{2, 4, 6\}, B = \{6\}, A \cap B = \{6\}$, find the same answer mathematically.

$$
\begin{aligned}
P(B|A) &= \\
P(A \cap B) &= \frac{1}{6} \\
P(A) &= \frac{1}{2} \\
&= \frac{\frac{1}{6}}{\frac{1}{2}} \\
&= \frac{1}{3}
\end{aligned}
$$

4 Independent Events

Two events A and B are independent if and only if the occurrence of A has no effect on B. For example, flipping a coin a few times results in multiple independent events, since the outcome of the coin flip doesn't depend on past coin flips. Drawing marbles out of a bag, however, is not independent if you don't replace the marbles.

Formally, two events are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$. This means that if A and B are disjoint events that have positive probabilities, then they are not independent ($P(A) \cdot P(B) \neq P(A \cap B) = 0$). The general result is: If $A_1, A_2, \dots A_n$ are said to be pairwise independent, the following is satisfied: $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ for all i, j in $1, 2, 3, \dots n$ where $i \neq j$. Also, if $A_1, A_2, A_3, \dots, A_n$ are independent, then $P(A_1 \cap A_2 \cap A_3 \cap \dots A_N) = P(A_1) \cdot \dots P(A_n)$. In more concise notation:

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

Example Let A = Alice and Betty have the same birthday, B = Betty and Carol have the same birthday, and C = Carol and Alice have the same birthday. Are A, B, and C pairwise independent? Are A, B, and C independent?

There are 365 possibilities for Alice's birthday. There are also 365 possibilities for Betty and Carol's individual birthdays. There are then 365^2 possibilities for Alice and Betty's birthdays. Since there are 365 different ways to have the same birthday, $P(A) = \frac{1}{365}$. The same applies for the other two events, so $P(A) = P(B) = P(C)$. We also need the probability that all three have the same birthday, so $P(A \cap B \cap C) = \frac{365}{365^3} = \frac{1}{365^2}$.

Now just verify that $P(X) \cdot P(Y)$ (for any events X and Y) are equal to the probability last calculated above. $P(A) \cdot P(B) = \frac{1}{365} \cdot \frac{1}{365} = \frac{1}{365^2}$ and the same holds for the other pairs, so the events are pairwise independent.

By the nature of this particular problem, $P(A \cap B \cap C) = P(A \cap B) = \frac{1}{365^2}$, but $P(A) \cdot P(B) \cdot P(C) = \frac{1}{365^3}$. Therefore, A, B, and C are not independent. This is a good demonstration of how **pairwise independence does not imply independence**.

5 Random Variables

A random variable (sometimes abbreviated as rv) is a function that assigns a real number to each outcome in the sample space Ω of a random experiment.

Consider the tossing of a coin three times. Let X equal the number of heads out of three tosses, and $\Omega = HHH, HHT, HTH < THH, HTT, THT, TTH, TTT$. We can map a number onto each element of that set to get $X = 3, 2, 2, 2, 1, 1, 1, 0$, so X is a random variable.

Note that a random variable is called *discrete* if it can take only discrete numbers. Random variables that instead can take any value from an interval are called *continuous*, where that interval is any subset of the set of all real numbers.

Example Roll two dice and define X to be the sum of the two numbers. The possible values for X range from 2 to 12. If instead you roll a die until a 4 appears and define X as the number of rolls, $\Omega = 4, f4, \dots$ where f is a number that isn't a 4. This means the random variable X is a discrete, infinite random variable.

5.1 Probability Distributions

Let X be a discrete random variable and x be any real number. The probability distribution function, denoted by "P", is the probability that $\{X = x\}$. Other notations include $P(X = x)$, $P(x)$, or $P(\{X = x\})$, where $\{X = x\}$ is an event. Note that random variables are usually written using uppercase letters, while values are written using their corresponding lowercase variable (hence the X vs. x notation).

Example Consider the experiment of rolling two fair dice. Let the random variable X be the sum of the two numbers, for $x = 2, 3, 4, 5, \ldots 12$ (of course, many of these numbers can be formed in different ways, e.g. 6 can be formed from 1+5, 2+4, or 3+3). Find the probability of $x = 2$.

This can be written as $P(x = 2) = P(\{x = 2\}) = P(1, 1)$, which we know is $\frac{1}{36}$. If instead you calculate $P(x = 4)$, you must search for outcomes that result in a sum of four – in this case, there are three ways, so the probability is $\frac{3}{36} = \frac{1}{12}$. The same thing applies for every other number in X, a process that is not recorded here. Just know that the probability distribution of X can be summed up in different ways, such as a table or a graph, just the same as any other set of discrete data.

5.2 Geometric Distributions

Consider a sequence of independent and identical trials, such that each trial has only two outcomes, either success or failure (called **binary trials**) $P(success) = P$ on each trial. Define a random variable X as the number of trials you need to get a success. In this case, probability distribution is given by: $P(X = x) = (1-p)^{x-1}p, x = 1, 2, 3, \ldots \infty$. This is an example of a geometric distribution, as the odds of getting a success only after n trials decreases geometrically.

Example Suppose that a machine producing a product has a 3defective rate. What is the probability that the first defective item occurs at the fifth item inspected? What is the probability that the first defect occurs within the first five inspections?

Each product inspection can be regarded as a binary trial – either the product is defective (fail), or it's not. Let the random variable X be the number of trials to get the first defective item. $P(X = x) = (1-p)^{x-1}p$ in general, so $P(X - x) = 0.97^{x-1} \cdot 0.03$ for $x = 1, 2, 3, \ldots$. Just plug in the value 5 in for x to get a probability of 0.0256 for the fifth item being defective while the rest are not.

For the second part, we must sum the probabilities for the first five trials: $P(x \leq 5) \sum_{n=1}^{5} 0.97^n \cdot 0.03$. Doing some magical math yields a final answer of 0.14127.

Alternatively, you could just calculate the probability that the first defective item is *not* in the first 5 items, then subtract that from 1: $1 - P(x > 5)$. Unfortunately, doing it that way is much harder, as it involves summing an infinite geometric series. Not that it can't be done, just that it's not worth the added complexity.

Example In a baseball event, the first team to win four games wins the championship. What is the probability distribution for the number of games X?

Assume that each team has the same chance to win a game as any other team does (0.5). There are four games, so $x = 4$ (???). The first team to win four games wins, so the probability here is effectively equal to $\frac{1}{2}^2$. That's only for the first team, however, but it turns out that the probability of the second team winning all four is exactly the same, so $P(4) = P(x = 4) = \frac{1}{8}$ (???!).

Now suppose that five games are won, meaning that the second team wins at least once. There are four different ways for this to happen (the case where the other team wins the fifth game isn't included, there wouldn't be a fifth game in that scenario). The probability of the first team winning can be calculated as $(\frac{1}{2})^5 \cdot 4$.

5.3 Expected Value

The expected value is a measurement of the center of a random variable, otherwise known as the average or the mean. To find the average of a discrete random variable, $E(X) = \sum_x x * P(X = x)$. In other words, multiply the probability of each value by the value itself, and add them all together.

Example If you play roulette and bet \$1 on black, win \$1 with probability 18/38 and you lose \$1 with probability of 18/38, what is the expected probability?

Solving this is very simple. Add 18/38 to -20/38 to get -0.0526 (in other words, you lose just over 5 cents).

5.4 Expectation of a geometric distribution

We know what the probability distribution of a geometric distribution is defined by $P(x) = p(1-p)^{x-1}$. This means that the mean of a geometric distribution is defined as $\sum\limits_{x=1}^{\infty} xp(1-p)^{x-1} \ldots$

$$
\begin{aligned}
S_\infty &= 1 + 2(1-p) + 3(1-p)^2 + \ldots \\
S_\infty &= 1 + (1-p) + (1-p)^2 + \ldots \\
&= \frac{1}{1-(1-p)} \\
&= \frac{1}{p} \\
S_\infty &= \frac{1}{p^2} \\
E(X) &= \frac{1}{p}
\end{aligned}
$$

This doesn't quite make sense, but whatever. The end result is that **if X follows a geometric distribution with P(success)=p,** $E(X) = \frac{1}{p}$. Somehow this all works out.

There is also a general formula for the expectation of a function of a random variable. Let $g(x)$ be a function of random variable X. The expectation of $g(x)$ is then:

$$E(g(x)) = \sum_x g(x) \cdot p(x)$$

Example Consider the distribution of $p(x) = \frac{1}{16} : x = 0, \frac{6}{16} : x = 1, \frac{9}{16} : x = 2$. Find $E(x^2)$.

$$
\begin{aligned}
E(x^2) &= 0^2 \cdot \frac{1}{16} + 1^2 \cdot \frac{6}{16} + 2^2 \cdot \frac{9}{16} \\
&= \frac{21}{8}
\end{aligned}
$$

Expectation $E(x)$ is a linear function; you can use it for linear combinations, e.g. $E(ax+b) = aE(x)+b$, assuming a and b are constants. In order to maintain sanity, the proof is not shown here. If there are n random variables in the sequence $x_1, x_2, x_3, \cdots, x_n$, the expectation of the sum of all of them ($E(x_1+x_2+x_3+\cdots+x_n)$) is equal to $E(\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} E(x_i)$; i.e., you can swap the sum and the expectation function. This is true for any sequence of random variables.

5.5 Variance

The k^{th} moment of a random variable X is defined as $E(x^k) = \sum\limits_x x^k \cdot p(x)$ (I have no idea why this is important). Variance is effectively just the spread of the probability distribution (?), and is defined by:

$$Var(x) = E((X - E(x))^2)$$

Using notation from earlier, $g(x) = (X - E(x))^2$. Calculating the variance probably will not be any easier than calculating the expected value, though, so here's a simpler formula:

$$Var(X) = E(x^2) - E(x)^2$$

This is much easier to compute than the above. Once again for reasons of sanity, the proof for this is not shown. Note that variance is *not* linear, so $Var(ax+b) = a^2 Var(x)$. Since this is important, here's a proof:

$$
\begin{aligned}
Var(ax+b) &= E((ax+b)^2) - E(ax+b)^2 \\
&= E(a^2x^2 + 2abx + b^2) - (aE(x)+b)^2 \\
&= a^2E(x^2) + 2abE(x) + b^2 - (a^2E(x)^2 + 2abE(x) + b^2) \\
&= a^2(E(x^2) - E(x)^2) \\
&= a^2Var(x)
\end{aligned}
$$

If you take the square root of the variance, you get the standard deviation.

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

Important: If X has a geometric distribution with P(success)$=p$ on each trial, the following is known:

- $P(x=n) - P(1-p)^{x-1}$
- $E(x) = \frac{1}{P}$
- $Var(x) = \frac{1-p}{p^2}$

Example Find the mean and the standard deviation of the following distribution: $P(x) = 0.02, 0.68, 0.07, 0.08, 0.10, 0.01, 0.0$ for sequence $0, 1, 2, 3, 4, 5, 8, 10$. $E(x) = 1.87$, by simple computation. To calculate the variance, first find $E(x^2)$, which is 9.06. The variance is just $E(x^2) - E(x)^2$, which is 5.5631. The standard deviation is just the square root of that, so $\sigma = 2.35$.

6 Combinatorial Probability

If Ω has finite equally likely outcomes, then for any event A, $P(A) = \frac{n(A)}{n(\Omega)}$. There are a few important rules for this (yes, these notes will be very disjointed).

6.1 Multiplication Rule

If a task consists of a sequence of choices in which there are n_1 selections for the first choice, n_2 selections for the second choice, etc..., then the task of making selections can be done in $n_1 \cdot n_2 \cdot n_3 \cdot \ldots$ different ways.

For example, if there are three locations A, B, C, with 3 different ways to go from A to B and 4 different ways to go from B to C, there are 12 different ways to go from A to B and then to C. For a more real-world example, calculating the number of possible US zip codes can be done by taking 10^5, since there are 5 digits to US zip codes and each digit has 10 different possibilities. If you impose a restriction that each digit must be unique, there are then $\frac{10!}{5!}$ possible combinations.

6.2 Permutations

Permutations are used to determine the number of different ways to pick k objects out of n different objects and arrange them on a line. The notation used is ${}^nP_k$, and is defined by ${}^nP_k = \frac{n!}{(n-k)!}$, where $k \leq n$. Note that if since there is only one way to pick zero items ($k = 0$), ${}^nP_0 = 1$. If instead you select all items from the group, ${}^nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$.

As a small example, if you select 9 players (for a batting order) from a 15 person team, there are ${}^{15}P_9 = \frac{15!}{6!}$ possible batting orders. My calculator says that 1,816,214,400 possible batting orders, which is probably a lot.

6.3 Combinations

Combinations are like permutations, except for where order doesn't matter. It's essentially the same notation as permutations, but with a C instead of a P. Combinations are defined as ${}^nC_k = \frac{n!}{(n-k)! \cdot k!}$. This also has the constraint that $k \leq n$.

Example From a group of 15 smokers and 21 nonsmokers, a researcher wants to randomly select 7 smokers and 6 nonsmokers for a study. In how many ways can the study group be selected?

Since order doesn't matter here, use combinations:

$$\begin{aligned} S &= (15C7)(21C6) \\ &= \frac{15!}{8!7!} \cdot \frac{21!}{15!6!} \\ &= 349188840 \end{aligned}$$

There are 349,188,840 ways to choose the study group.

6.4 Binomial Theorem

Binomial theorem is an application of numbers yielded by combinations or by Pascal's triangle (which are closely linked), and is given by the following summation:

$$(x+y)^n = \sum_{m=0}^{n} nC_m \cdot x^{n-m} \cdot y^m$$

This is a way of expanding a binomial raised to n into $n-1$ terms. For example, $(x+y)^2 = x^2+2xy+y^2$.

6.5 Partitions

Partitions are written on the disk using format defined by the partition table, often using either MBR (Master Boot Record) or GPT (GUID Partition Table). They are used to separa- wait, wrong class.

In probability, partitions are when you take a set of n elements and divide them into r disjoint subsets such that the i^{th} subset has n_i elements, where $i = 1, 2, \dots, r$ and $n_1 + n_2 + \cdots + n_r = n$. The total number of choices is given by both a very annoying string to LaTeX, but is much more concisely written as:

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots n_r!}$$

This is a generalized version of binomial theorem (?), and is called the multinomial coefficient (as opposed to nCr, which is called the binomial coefficient). There's a notation for it that looks similar to the binomial coefficient vertical-parentheses notation, but I don't know how to write that out.

Example A house has 12 rooms. How many ways can you paint 4 of those yellow, 3 purple, and 5 red.

Use the multinomial coefficient formula thingy:

$$\begin{aligned} S &= \frac{12!}{4! \cdot 3! \cdot 5!} \\ &= \frac{479001600}{17280} \\ &= 27720 \end{aligned}$$

Example 2 There are 39 students in a class. In how many ways can a professor give out 9 A's, 13 B's, 12 C's, and 5 F's?

$$\begin{aligned} S &= \frac{39!}{9! \cdot 13! \cdot 12! \cdot 5!} \\ &= 1.57 \cdot 10^{22} \end{aligned}$$

Example 3 How many different words can be obtained by rearranging the word "TATOO"?

If all letters are different, the answer would just be 5!, but it's not that simple. If you switch the O's, you don't get a different arrangement, so you must use the multinomial coefficient formula.

$$\begin{aligned} S &= \frac{5!}{2! \cdot 2!} \\ &= 30 \end{aligned}$$

Example 4 A class consisting of 4 graduate students and 12 undergraduate students randomly divides into four groups of 4. What is the probability that each group contains a graduate student?

The proper way to calculate this is by dividing the number of ways to have a grad in each group by the total number of choices. The denominator is fairly easy: $\frac{16!}{4! \cdot 4! \cdot 4! \cdot 4!}$.

If each group must have a graduate student, there are 4! ways to distribute them into 4 different groups. Now do the same for the undergraduates: $\frac{12!}{3! \cdot 3! \cdot 3! \cdot 3!}$. The number of ways to have a graduate student in each group is then $4! \cdot \frac{12!}{3! \cdot 3! \cdot 3! \cdot 3!}$.

Based on that, the probability that each group has a graduate student is then:

$$\frac{4! \cdot \frac{12!}{3! \cdot 3! \cdot 3! \cdot 3!}}{\frac{16!}{4! \cdot 4! \cdot 4! \cdot 4!}}$$

This is a moderately complex calculation, so it won't be finished here.

6.6 Binomial and Multinomial Distributions

Consider n identical and independent trials such that each trial has only two outcomes, success and failure. The probability of success shall be equal to p, and let X be the number of successes out of n trials. X follows a binomial distribution with parameters n and p, with a probability distribution of:

$$P(X = x) =^n C_x p^x (1 - p)^{n-x}$$

This is also denoted as Binomial(n, p). If you sum all of the probabilities together, the result should be 1. It can also be shown that the expected value is just $E(X) = np$, and that the variance is given by $Var(x) = np(1 - p)$. For reasons of time and sanity, neither of these proofs are shown here. **Remember, p is only the probability of success for ONE trial.**

Example Suppose a die is rolled five times. What is the probability of getting exactly 2 fours? Of getting at least a four? What about getting at most 4 fours? Also, find the expected value and the variance of the number of fours out of five trials.

x is the number of 4's out of 5 trials. For the first problem:

$$\begin{aligned} n &= 5 \\ p &= \frac{1}{6} \\ P(X = 2) &= 5 \cdot \frac{1}{6^2} \cdot \frac{5^3}{6^3} \\ &= 0.0804 \end{aligned}$$

For the second problem, subtract from 1 the probability of getting no fours:

$$\begin{aligned} P &= 1 -^5 C_0 (\frac{5}{6})^5 \\ &= 0.665 \end{aligned}$$

For the third problem, subtract from one the probability of getting 5 fours:

$$
\begin{aligned}
P &= 1 -^5 C_5(\frac{1}{6})^5 \\
&= 0.999
\end{aligned}
$$

For the final part, the expected value is just $5 \cdot \frac{1}{6}$, and the variance is $5 \cdot \frac{1}{6} \cdot \frac{5}{6}$.

Example 2 A student takes a test with 16 multiple choice questions. Assuming each question has four choices and she chooses answers at random, what is the probability that she will get exactly 3 questions right?

Although the probability of success is not 0.5, this *is* a binary trial as there are only two outcomes possible per question: success or failure. Here, X is the number of correct answers, so use the binomial probability thingy for $n = 16$ and $p = 0.25$.

$$
\begin{aligned}
P(X = 3) &= \binom{16}{3}(\frac{1}{4})^3(\frac{3}{4})^{16-3} \\
&= 0.2079
\end{aligned}
$$

Now for multinomial distributions. Consider n identical and independent trials such that each trial has k (where $k > 2$) outcomes, with probabilities $p_1, p_2, p_3, \ldots, p_k$. The probability of getting n_i outcomes of type i with $n = n_1 + n_2 + \cdots + n_k$ is as follows:

$$P(n_1, n_2, \ldots, n_3) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots n_k!} P_1^{n1} P_2^{n2} P_3^{n3} \ldots P_k^{nk}$$

This is effectively just a more general form of the binomial case, which has only two outcomes. Just like the binomial case, the sums of all possible outcomes should be 1.

Example A baseball player gets a hit with a probability of 0.3, a walk with probability 0.1, and an out with probability of 0.6. If they bat four times during the game, what is the probability that they'll get 1 hit, 1 walk, and 2 outs?

We can immediately assign some values. $n = 4, k = 3, p_1 = 0.3, p_2 = 0.1, p_3 = 0.6, n_1 = 1, n_2 = 1, n_3 = 2$. Then just plug in the values:

$$\frac{4!}{1! \cdot 1! \cdot 2!} 0.3^1 \cdot 0.1^1 \cdot 0.6^2 = 0.1296$$

Example 2 The output of a machine is graded excellent 70% of the time, good 20% of the time, and defective 10% of the time. What is the probability that a sample size of 15 has 10 excellent, 3 good, and 2 defective items?

Use the same process as above, just plugging the values in:

$$\frac{15!}{10! \cdot 3! \cdot 2!} 0.7^{10} \cdot .2^3 \cdot 0.1^2$$

The final calculated value is not shown here, because reasons.

6.7 Poisson Approximation to the Binomial

Calculating binomial probabilities can be computationally expensive if you're doing it over a range. For that reason, there is an approximation that uses the Poisson distribution. A random variable X is said to have a Poisson distribution with parameter λ (or $X\ Poisson(\lambda)$ if):

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

...where $x = 0, 1, 2, 3, \ldots$. The Poisson distribution is used to find the number of successes when the average number of successes is given. λ is the number of successes, X is the actual number of successes (number of trials is not fixed). If you add all the probabilities together (from x to ∞), you get $\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}$, or the Taylor series of e^{λ}. Important note: it should be obvious that $E(X) = \lambda$ and the variance is also λ. The proof for variance is not shown here.

Example Suppose the average number of lions seen on a 1-day safari is 5. What is the probability that a tourist will see exactly 4 lions on the next day? What is the probability that the tourist will see fewer than 4 lions?

Here, seeing a lion is a "success". Define a random variable X as the number of lions seen. We already know that $\lambda = 5$. The probability distribution is then $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$. That means that $P(X = 4) = \frac{e^{-5}\lambda^4}{4!}$, which is 0.17547.

To calculate the probability of seeing fewer than 4 lions, you need to calculate $P(0)+P(1)+P(2)+P(3)$, which is $\frac{e^{-5}5^0}{0!} + \ldots$. That calculation isn't shown here because it's actually very simple, just very tedious.

6.8 Poisson Approximation to the Binomial, Round 2

(I have no idea why there's two of these sections, this one and the above section are somehow different)

Suppose Sn has a binomial distribution with parameters n and Pn. If $Pn \to 0$ and $n \cdot Pn \to \lambda$ as $n \to \infty$, then $P(Sn = x) \to \frac{e^{-\lambda}\lambda^x}{x!}$ For large n and small P, binomial probabilities can be approximated by the poisson distribution.

Example Suppose we roll two dice 12 times. Let D be the number of times a double six appears. Find exact and approximation values of $P(D = k)$ for $k = 0, 1, 2$.

We can start with some variables. $n = 12, p = P(6, 6) = \frac{1}{36}$. We already know how to calculate exact values for this using the binomial formula, but for the poisson approximation we'll need the parameter λ. Here, $\lambda = 12 \cdot \frac{1}{36} = \frac{1}{3}$.

Start by calculating the exact values:

$$\begin{aligned} P(X = 0) &= \binom{12}{0}(\frac{1}{36})^0(\frac{35}{36})^{12} \\ &= 0.7132 \end{aligned}$$

Now try doing it with the poisson approximation:

$$\begin{aligned} P(X = 0) &\approx \frac{e^{-\frac{1}{3}}}{0!} \\ &\approx 0.7165 \end{aligned}$$

Repeating this process for the remaining values of k yields exact/approximate value pairs of (0.2445, 0.2388) and (0.0384, 0.0398). All of these approximation values are fairly accurate, although there is some discrepancy.

Example 2 If you're in a group of 183 people, what is the probability that nobody else shares your birthday?

This is a binomial distribution problem, with 182 binary trials (the other people) with odds of success being $\frac{1}{365}$. The solution for this is simple: $\binom{182}{0}(\frac{1}{365})^0(\frac{364}{365})^{182} = 0.6069$.

If you were to take the approximate value instead, you would need to figure out the value of λ as $n \cdot p = \frac{182}{365}$. The formula used is then $P(X = 0) \approx \frac{e^{-\frac{182}{365}}}{0!} = 0.6073$.

7 Probabilities of Unions

Recall that for events A and B, the probability of their union $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (just make sure to know that for general events, $P(A \cap B) \neq P(A) \cdot P(B)$, that only applies if events A and B are known to be independent). Also, for events A, B, and C, $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$. This is a long formula that is not fun to prove (and the proof isn't going to be on the exam anyways...), so the proof is not shown here. The important bit from it, though, is that $P(A \cap (B \cup C))$ can be written as $P((A \cap B) \cup (A \cap C))$.

Example Suppose we roll 3 dice. What is the probability of getting at least one 6?

Let A_i be the probability of getting a 6 on the i^{th} roll, where $i = 1, 2, 3$. That means that $P(A_1) = \frac{1}{6} = P(A_2) = P(A_3)$. We can calculate $P(A_1 \cap A_2) = \frac{1}{36}$ as the events are all independent, and the same goes for $P(A_1 \cap A_2 \cap A_3 \cap)$, which is $\frac{1}{216}$. The final probability is then:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= \frac{3}{6} - \frac{3}{36} + \frac{1}{216} \\ &= \frac{91}{216} \\ &= 0.4213 \end{aligned}$$

I have no idea why you'd want to go to all of this work when you can just do $1 - (\frac{5}{6})^3 = 0.4213$.

Example 2 You pick seven cards out of a deck of 52. What is the probability that you have at least one 3-of-a-kind?

Let A_i having 3 cards of type i, $i = 1, 2, 3, \ldots 13$ (where a number corresponds to a card type). This formula is going to be disgusting:

$$P(\cup_{i=1}^{13} A_i) = \sum_{i=1}^{13} P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

Note that there is no third term because a third term would require 9 cards, but we only have 7.

$$\begin{aligned} P(A_1) &= \frac{\binom{4}{3}\binom{48}{4}}{\binom{52}{7}} \\ P(A_1 \cap A_2) &= \frac{\binom{4}{3}\binom{4}{3}\binom{44}{31}}{\binom{52}{7}} \\ P(\cup_{i=1}^{13})A_i &= \binom{13}{1}\frac{\binom{4}{3}\binom{48}{4}}{\binom{52}{7}} - \frac{13}{2}\frac{\binom{4}{3}\binom{4}{3}\binom{44}{31}}{\binom{52}{7}} \end{aligned}$$

Example 3 Suppose you roll a dice 15 times. What is the probability that you don't see each of the 6 numbers at least once?

Let A_i be that you never see i where $i = 1, 2, 3, 4, 5, 6$. This is going to be another one of those disgusting formulas:

$$\begin{aligned} P(A_i) &= (\frac{5}{6})^{15} \\ P(A_i \cap A_j) &= (\frac{4}{6})^{15} \\ P(A_1 \cap A_j \cap A_k) &= (\frac{3}{6})^{15} \end{aligned}$$

...etc. Turns out there's a nice pattern here, so the final formula is:

$$P(\cup_{i=1}^{6} A_i) = \binom{6}{1}(\frac{5}{6})^{15} - \binom{6}{2}(\frac{4}{6})^{15} + \dots$$

No, I'm not going to finish that. It should be obvious, but it's very unpleasant to LaTeX.

8 Conditional Probability, Round 2

For any events A and B with $P(A) > 0$, the probability that B will occur given that A occurs is $P(B|A) = \frac{P(B\cap A)}{P(A)}$. Conditional probability follows some probability axioms:

- $P(\Omega|B) = 1$. This should be obvious.
- If A_1 and A_2 are disjoint (meaning they have no common outcomes), then $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$.

Example Suppose that five good fuses and 2 defective ones have been mixed up. To find the defective fuses, you test them one by one at random and without replacement. What is the probability that you find both of the defectives in the first two tests?

Let D_1 be the probability that the first draw is defective, and let D_2 be the probability that the second draw is defective. We know immediately that $P(D_1) = \frac{2}{7}$, but we can't directly tell what $P(D_2)$ is because it depends on the outcome of the first. However, $P(D_2|D_1) = \frac{1}{6}$, so $P(D_1 \cap D_2) = P(D_1) \cdot P(D_2|D_1)$. That ends up being $\frac{2}{7} \cdot \frac{1}{6} = \frac{1}{21}$.

8.1 Multiplication Rule

For any events $A_1, A_2, \dots, A_3$:

$$P(\cap_{i=1}^{n} A_i) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n| \cap_{i=1}^{n-1} A_i)$$

I guess this gets its own subsection now?

8.2 Two stage experiments

Time to introduce a new, confusingly named concept: partitions. Again.

Suppose you divide the sampling space Ω into multiple subsets, named $A_1, A_2, A_3, \dots A_n$. A collection of events is called a partition of Ω if $\cup_{n=1}^{n} A_i = \Omega$ and $A_i \cap A_j$ for all $i, j = 1, 2, 3 \dots n$ where $i \neq j$.

This is where total probability theorem becomes relevant. If you have a sampling space divided into partitions, plus some other event B that has any number of items from the different partitions, then $P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$.

Example You enter a chess tournament where you probability of winning a game is 0.3 against half of the players (type 1), 0.4 against a quarter of the players (type 2), and 0.5 against the remaining quarter of the players (type 3). You play a game against a randomly chosen opponent. What is the probability of winning?

Let A_i be the event of playing against a type i opponent ($i = 1, 2, 3$), and B be the probability of winning the game.

$$
\begin{aligned}
P(A_1) &= 0.5 \\
P(A_2) &= 0.25 \\
P(A_3) &= 0.25 \\
P(B|A_1) &= 0.3 \\
P(B|A_2) &= 0.4 \\
P(B|A_3) &= 0.5
\end{aligned}
$$

Since this is a partition problem, you can use the total probability theorem.

$$
\begin{aligned}
P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\
&= 0.375
\end{aligned}
$$

Example 2 You roll a four sided fair die. If the result is 1 or 2, you roll once more, but otherwise you stop. What is the probability that the sum of the rolls is at least 4?

Think of this as a random experiment. The outcome of the first roll can be anything from 1 to 4. Let A_i be the probability of getting i for the first roll, and B the probability that the sum of the roll is at least four. $P(A_i)$ is easily set at 1/4, so next to calculate is $P(B|A_1)$ and $P(B|A_2)$, which are 2/4 and 3/4, respectively. The other two probabilities are easy: $P(B|A_3) = 0, P(B|A_4) = 1$. By using the total probability theorem, the final result is 9/16.

8.3 Bayes' Theorem

Let $A_1, A_2, \ldots, A_n$ be a partition of Ω such that $P(A_i) > 0$ for all i. Then, for any event B, $P(A_i|B) = \frac{P(B|A_i)\cdot P(A_i)}{\sum_{j=1}^{n} P(B|A_i)\cdot P(A_j)}$ for $i = 1, 2, 3, \ldots n$. Using total probability theorem, that also means that it equals $\frac{P(B|A_i)\cdot P(A_i)}{P(B)}$.

Example Three factories make 20, 30 and 50% of the computer chips for a company. The probability of a defective chip is 0.04, 0.03, and 0.02 for the three factories, respectively. If a defective chip is found, what is the probability that it came from factory 1?

Let A_i be the chip coming from a factory ($i = 1, 2, 3$) and let B be the probability that the chip is defective.

$$
\begin{aligned}
P(A_1) &= 0.2 \\
P(A_2) &= 0.3 \\
P(A_3) &= 0.5 \\
P(B|A_1) &= 0.04 \\
P(B|A_2) &= 0.03 \\
P(B|A_3) &= 0.02 \\
P(A_1|B) &= \frac{P(B|A_1)\cdot P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\
&= \frac{8}{32} \\
&= 0.25
\end{aligned}
$$

8.4 Discrete Joint Distributions

Let X and Y be two random variables associated with the same random experiment. The probability distribution of X and Y is then given by:

$$P(X = x, Y = y) = P(X = x \cap Y = y)$$

This is called the *joint distribution* of X and Y. If you were to add all the possible probabilities ($\sum_{x,y} P(x, y)$), the result would be 1.

Example Roll two four-sided fair dice. Let X be the maximum of the two numbers, and Y be the sum. Find the joint distribution of X and Y, then find $P(x \leq 2, y \leq 3)$.

First, find the possible values of x and y: $1, 2, 3, 4$ and $2, 3, 4, 5, 6, 7, 8$, respectively. If you draw a table of the joint distribution, it turns out that there are 32 different outcomes, but most of them have a probability of zero. To answer the original question, the probability of $P(x \leq 2, y \leq 3)$ is just the sums of the only relevant non-zero points that satisfy the conditions, for $\frac{1}{16} + \frac{2}{16}$.

8.5 Marginal Distributions

Suppose $P(x, y)$ is the joint distribution of X and Y. The marginal distribution of X ($P(X = x)$) is then obtained by the summation $P(X = x) = \sum_{y} P(x, y)$. The marginal distribution of Y is also just $P(Y = y) = \sum_{x} P(x, y)$.

Example Suppose you draw 2 balls out of a container with 6 red balls, 5 blue balls, a 4 green balls. Define two random variables, X being the number of red balls and Y being the number of blue balls. Find the joint distribution of X and Y, then find the marginal distribution of X and Y.

The possible values of X and Y are both $0, 1, 2$. Since this is a more complicated distribution, the resulting table is more complex and involves some combinatorial math (unfortunately, since LaTeXtables are a pain, that table isn't shown here).

8.6 Functions of Multiple Random Variables

Let X and Y be two random variables, and let g be a function of X and Y. Then $Z = g(x, y)$ is also a random variable and the probability distribution is given by $P(Z = z) = \sum_{(x,y)|g(x,y)=z} P(x, y)$. In other words, sum for all cases where $g(x, y)$ exists. This means that $E(Z) = E(g(x, y)) = \sum_{x} \sum_{y} g(x, y)P(x, y)$. Also note that $E(ax + by + c) = aE(x) + bE(y) + c$ where a, b, and c are constants. This is true only for linear functions, but it can be used for any sequence of random variables:

For any sequence of random variables, $X_1, X_2, \ldots X_n$ and constants $a_1, a_2, \ldots a_n$, $E(a_1x_1 + a_2x_2 + \cdots + a_nx_n) = a_1E(x_1) + a_2E(x_2) + \cdots + a_nE(x_n)$.

8.7 Independence of Random Variables

For independent events, events A and B are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$. The concept is similar for random variables: two random variables X and Y are independent if an only if $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ for all X and Y. If you can find even one X, Y pair that does not follow by this rule, X and Y are not independent. Note that this also means that $E(g(X, Y)) = E(X) \cdot E(Y)$ if X and Y are independent. Also, still assuming that X and Y are independent, $Var(aX + bY + c) = a^2Var(X) + b^2Var(Y)$. Generalized, this means that $Var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2Var(X_1) + a_2^2Var(X_2) + \cdots + a_n^2Var(X_n)$.

There's an equation here for some reason, $Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$, but it only applies when the relevant random variables are all independent.

9 Distribution of Continuous Random Variables

So far, only discrete random variables have been covered, but what about continuous random variables? These types of random variables are continuous, so they can take *any* real number. Formalized: when a random variable can take any value in an interval it is called a continuous random variable. Furthermore, the probability distribution of a continuous random variable is called a probability density function, or "pdf" for short. This is opposed to the probability distribution of a discrete random variable, which is called a probability mass function, or "pms".

Anyways, a continuous random variable X is said to have a probability density function "f" if for all $a \leq b$, $P(a \leq x \leq b) = \int\limits_a^b f(x)dx$. The area under the curve has an actual meaning, specifically as the probability that a value will fall within some two values. In fact, the probability of a single point is 0 (i.e. $P(X = x) = 0$). That also means that there is effectively no difference between $P(a < x < b)$ and $P(a \leq x \leq b)$. Furthermore, for any PDF, $\int\limits_{-\infty}^{\infty} f(x)dx = 1$.

Example Let the pdf $f(x)$ be a constant c for $a < x < b$, but 0 otherwise. Calculate the value of c.

Since this is a PDF, the equation $\int\limits_{-\infty}^{\infty} f(x)dx$ should be satisfied. If you divide this into three parts, only one becomes non-zero: $\int\limits_a^b cdx = 1$. After a little bit of calculus and solving, $c = \frac{1}{b-a}$. This is called a uniform distribution, and X is called a uniform random variable.

Example 2 Let $f(x)$ be defined as $\frac{a}{\sqrt{x}}$ when $0 < x < 1$, and 0 otherwise. Find the value of a.

This is the same idea as the last problem, just a little more involved. Use the formula $\int x^k dx = \frac{x^{k+1}}{k+1}$, where $k \neq -1$. After some work, $a = 1/2$.

9.1 Expectations and Variance

The expected value of a continuous random variable X is given by:

$$E(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx$$

If g is a function of X, then the expectation of the function value is:

$$E(g(x)) = \int\limits_{-\infty}^{\infty} g(x)f(x)dx$$

...and the k^{th} moment is defined as:

$$E(x^k) = \int\limits_{-\infty}^{\infty} x^k f(x)dx$$

As for variance, the original formula still holds, more or less, just with a calc-y twist on it.

$$Var(x) = \int\limits_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$$

Note that the formula $Var(x) = E(X^2) - E(X)^2$ still holds, as does $E(aX + b) = aE(X) + b$ and $Var(aX + b) = a^2 Var(X)$.

Example Let X be a uniform random variable (a, b). Find the $E(X)$ and $Var(X)$.

For PDF, $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b\ldots$

$$\begin{aligned}
E(X) &= \int_a^b x\frac{1}{b-a}dx \\
&= \frac{1}{b-a}\int_a^b xdx \\
&= \frac{a+b}{2}
\end{aligned}$$

Now find $E(X^2)$:

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2\frac{1}{b-a}dx \\
&= \frac{b^3-a^3}{3(b-a)} \\
&= \frac{a^2+ab+b^2}{3}
\end{aligned}$$

When simplified, the final result for variance is:

$$\frac{(b-a)^2}{12}$$

If X is uniform, this formula, along with $E(X) = \frac{a+b}{2}$ may be used at any time without providing a proof.

9.2 Exponential distribution

Exponential distributions are often used to model time between two successive events. A random variable X follows an exponential distribution with parameter λ if the probability density function is given by the following where $x \geq 0$:

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

Here, x is the time between two events, and it can be shown that $\frac{1}{\lambda}$ is the average time between two events (i.e. the expected value). This should always be a valid probability density function, but the proof for this isn't shown here. The variance is $Var(X) = \frac{1}{\lambda^2}$, which is another complex calculus-y thing to prove. Also, $P(x \geq a) = e^{-\lambda a}$, where a is a constant greater than 0.

Example The time until a small meteorite lands anywhere in the Sahara desert is modeled as an exponential random variable with a mean of 10 days. The time is currently midnight. What is the probability that a meteorite first lands sometime between 6 AM and 6 PM of the first day.

If you take lambda to be days, $\lambda = \frac{1}{10}$. Let X be the time until the first landing in days. Here, it's equivalent to the time between the 0^{th} landing and the 1^{st} landing. Since the time range is from 6 AM to 6 PM (or 0600 to 1800), we need to find $P(\frac{1}{4} \leq x \leq \frac{3}{4})\ldots$

$$
\begin{aligned}
&= \int_{\frac{1}{4}}^{\frac{3}{4}} \frac{1}{10} e^{-\frac{1}{10}x} dx \\
P(X \geq a) &= e^{-\lambda a} \\
&= P(X \geq \frac{1}{4}) - P(x \geq \frac{3}{4} \\
&= e^{-\frac{1}{10}\cdot\frac{1}{4}} - e^{-\frac{1}{10}\cdot\frac{3}{4}} \\
&= 0.0476
\end{aligned}
$$

The final probability is 0.0476.

9.3 Cumulative Distribution Functions

For a probability distribution function, the sum of all probabilities up to some point is a cumulative distribution function. These are denoted as $F(X)$.

More formally, the cumulative distribution function (or "cdf") of a discrete random variable X is given by:

$$F(X) = \sum_{k \leq x} P(k)$$

For continuous random variables, the cdf is defined as:

$$F(X) = \int_{-\infty}^{x} f(t)dt$$

Example Suppose $P(X)$ is given as $1/4$ when $x = 1, 2, 3, 4$, and is zero otherwise. Find the cumulative distribution function for $P(X)$.

This once is fairly simple. For values $1, 2, 3, 4$, the CDF is $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$. This is ultimately a step function; all cumulative distribution functions of discrete random variables are step functions.

CDFs of continuous random variables have a number of useful properties:

- $\lim_{x \to \infty} F(x) = 1$
- $\lim_{x \to -\infty} F(x) = 0$
- $F(x)$ is non-decreasing
- $F(x)$ is right-continuous, i.e. $\lim_{x \to a^+} F(x) = F(a)$

Example 2 Let X be a geometric distribution (p). Find the cdf of p.

Keep in mind, in geometric distributions, p is the probability of success. The probability mass function here is $P(x) = p(1-p)^{x-1}$ for $x = 1, 2, 3, \ldots$. Since this is a discrete random variable, a regular summation must be used here: $F(x) = \sum_{k=1}^{x} P(k) \ldots$

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

$$\sum_{k=1}^{x} p(1-p)^{k-1} = p + p(1-p) + p(1-p)^2 + \cdots + p(1-p)^{x-1}$$

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

$$= 1-(1-p)^x$$

Example 3 Let X follow an exponential distribution with parameter λ. Find the CDF of X.

These kinds of distributions follow the form of $f(x) = \lambda e^{-\lambda x}$ Since this is a continuous random variable, integrals must be used.

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$$= \int_{0}^{x} \lambda e^{-\lambda t} dt$$

$$= \lambda \cdot \frac{e^{-\lambda t}}{-\lambda}\Big|_0^x$$

$$F(x) = 1 - e^{-\lambda x}$$

Keep in mind, this only applies for $x \geq 0$; otherwise, $F(x) = 0$.

9.4 The Normal Distribution

The normal distribution an extremely useful continuous distribution that is widely used in statistics. It is defined as $N(\mu, \sigma^2)$ where μ is the mean and σ^2 is the variance. The actual formula is:

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

This is valid from $-\infty$ to ∞, as long as σ is a positive number. When you need to calculate probabilities using normal distributions, you must integrate for this thing (Note: $\int_{-\infty}^{\infty} N(\mu, \sigma^2) = 1$). Keep in mind, σ is the standard deviation, while σ^2 is the variance.

There are some useful properties of normal distributions. If you have a random variable X that follows a normal distribution, and you do $Y = aX + b$, it can be shown that Y is also normally distributed: $N(a\mu + b, a^2\sigma^2)$.

Remember the thing from earlier about integrating that ugly formula above? I lied. There is no way of computing the integral, you have to use some numerical method to find the values you're looking for. Usually this involves either a computer or a pre-computed table. These tables are usually based on the **standard normal distribution**, which is defined as $Z = N(0, 1)$. Here, more formulas:

- pdf of Z: $P(\zeta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta^2}{2}}$
- cdf of Z: $\Phi = P(Z \leq \zeta) = \int_{-\infty}^{\zeta} \frac{1}{\sqrt{2\pi}}$**INCOMPLETE**

Since most normal distributions won't be standard normal, you must standardize them before a standard normal distribution table can be applied. To do this, take $Z = \frac{x-\mu}{\sigma}$. This also means that $P(X > a) = P(\frac{x-\mu}{\sigma} > \frac{a-\mu}{\sigma})$.

Example An average lightbulb lasts 300 days with a standard deviation of 50 days Assuming that the bulb life is normally distributed, what is the probability that a lightbulb will last at most 365 days?

The bulbs follow the distribution $N(300, 50^2)$ (since in this notation, the second parameter is always the variance). First, standardize. $Z = \frac{365-300}{50} = 1.3$. Then just look up $P(Z < 1.3)$ (instead of the original $P(X < 365)$ and find it to be 0.9032.

Example 2 The annual snowfall for a particular location is modeled as a normal random variable with a mean of 60 inches and a standard deviation of 20 inches. What is the probability that this year's snowfall will be at least 8 inches?

Once again, standardize. $Z = \frac{80-60}{20} = 1$. Then just find $P(Z \geq 1)$. Since this is less convenient to calculate using a table, find $1 - P(Z < 1) = 1 - 0.8413 = 0.1587$.

9.5 Central Limit Theorem

The first thing to understand is the term *random sample.* A random sample is a sequence independent and identically distributed ("iid") random variables $x_1, x_2, x_3, \ldots x_n$. The random sample itself is a variable... somehow?

The sample total, produced by adding up everything inside the sample, is denoted as $S_n = x_1 + x_2 + x_3 + \cdots + x_n$. You can also take $E(S_n) = E(x_1 + x_2 + \cdots + x_n) = E(x_1) + E(x_2) + \cdots + E(x_n)$. The same ideas about variance from earlier apply here too: $Var(S_n) = Var(x_1) + Var(x_2) + \cdots + Var(x_n)$. This applies because all of these samples are by definition independent. The sample mean M_n is $\frac{x_1+x_2+\ldots x_n}{n} = \frac{S_n}{n}$, and $Var(M_n) = \frac{\sigma^2}{n}$ (no, I don't know why).

Note: if $x_1, x_2, \ldots, x_n = N(\mu, \sigma^2)$, then $S_n = N(n\mu, n\sigma^2)$. Also, $M_n = N(\mu, \sigma^2/n) = \frac{M_n-\mu}{\sqrt{\sigma^2/n}}$

Now, what if the distribution of the population from which samples are drawn is unknown? This is where central limit theorem becomes important. Let $x_1, x_2, \ldots x_n$ be a random sample with common mean μ and variance σ^2. Assume n (the sample size) is large. If S_n is the sample total $x_1 + x_2 + \cdots + x_n$, then then $\frac{S_n-n\mu}{\sqrt{n}\sigma}$ approximately follows a standard normal distribution $N(0, 1)$. The second part to this theorem is that if the sample mean is defined as $M_n = \frac{S_n}{n}$, then $\frac{M_n}{\sigma/\sqrt{n}}$ also approximately follows a standard normal distribution.

Example If you load on a plane 100 packages whose weights are independent random variables that are uniformly distributed between 5 and 50 pounds, what is the probability that the total weight exceeds 3,000 pounds?

Let x_i be the weight of the i^{th} package, where i is anything from 1 to 100. Then x_i follows a uniform distribution with parameters 5 and 50. Recall that $E(X) = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$. The expectation of a single variable is $\mu = E(x_i) = 27.5$. The variance of a single unit σ^2 is then $Var(x_i) = \frac{(50-2)^2}{12} = 168.75$. These values are true for all of the random variables (packages).

Now it's time to use central limit theorem. $E(S_{100}) = n\mu = 2750$, $Var(S_{100}) = n\sigma^2 = 16875$. Since the value actually needed is the standard deviation σ, the just take the square root the variance of the sample total, $\sqrt{16875}$. Now what you're looking for is $P(S_{100} \geq 3000)$. Standardize by subtracting the mean and dividing by the standard deviation to get $P(\frac{S_{100}-2750}{\sqrt{16875}} \geq \frac{3000-2750}{\sqrt{16875}}$. This works because we can assume this all follows a standard normal distribution.

This all turns out to be $P(Z \geq 1.92) = 0.02743$.

Example The income of college students is distributed with a mean income per year of \$12,000 and a standard deviation of \$6,000. If you randomly sample 50 college students, what is the expected average income of the sample? What is the varaince of the average income of the sample? What is the probability that the average income of the sample is less than \$10,000?

$\mu = 12000, \sigma^2 = 6000^2$. Let X_i be the income of the i^{th} student, where i is from 1 to 50. The expectation of the sample average is just μ, which is 12000 as stated. The variance $Var(M_{50})$ is $\frac{\sigma^2}{n}$ or 720,000. To answer the final question, calculate $P(M_{50} < 10000)$. Once standardized, this turns into $P(Z < -2.357)$, which is just 0.00921.

9.6 Functions of Random Variables

Functions of discrete random variables are simple enough, but things get more complicated with continuous random variables. There are two different methods: the cdf method, and the transformation method.

9.6.1 CDF Method

Let X be a continuous random variable and $y = g(x)$. First, find the cdf of y, $F_y(y)$, and try to find it in terms of the CDF of the original random variable ($F_x(x)$). Remember: $F_x(x) = \int\limits_{-\infty}^{x} f(t)dt$. Second, differentiate to obtain the pdf of y, $f_y(y)$.

Example Let X follow a uniform random distribution with parameters (0,1), and define $y = \sqrt{x}$. Find the pdf of y.

Since X is unifomr, the pdf is $\frac{1}{b-a}$ and the range is $a \leq x \leq b$. Here, that means that $f_x(x) = 1$, for $0 \leq x \leq 1$. Now find the pdf of y in terms of x.

$$\begin{aligned} F_y(y) &= P(Y \leq y) \\ &= P(\sqrt{x} \leq y) \\ &= P(x \leq y^2) \\ &= F_x(y^2) \end{aligned}$$

The cdf of the original variable has now been written in terms of the new variable. That's step one complete, so now differentiate with respect to y to get the pdf. This will need the chain rule ($f(g(x)) = f'(g(x)) \cdot g'(x)$)

$$\begin{aligned} f_y(y) &= f_x(y^2) \cdot \frac{d(y^2)}{dy} \\ &= 2y \end{aligned}$$

Now all that's left to find is the range. The range of the original pdf was $0 \leq x \leq 1$. To get the range, look at the graph for the cdf; in this case, it's the same range as the original.

Example 2 A man is driving from Boston to New York, a distance of 180 miles. His average speed is uniformly distributed between 30 and 60 miles per hour. What is the pdf of the duration of his trip?

Let X be the speed and Y be the duration. $x = \frac{180}{y}$, so $y = \frac{180}{x} = g(x)$. The pdf of X is just $\frac{1}{30}$. Then $F_y(y) = P(Y \leq y) = P(\frac{180}{x} \leq y) = P(\frac{x}{180} \geq \frac{1}{y}$. More math? $P(x \geq \frac{180}{y})$.

Now figure out how to write this in terms of the cdf of X, which is $F_x(X) = P(X \leq x)$. The inequality above is in the wrong direction, so invert and get $1 - F_x(\frac{180}{y})$. Doing the derivation stuff with respect to y gets you $\frac{6}{y^2}$. Doing a little graph magic somehow yields a range of $3 \leq y \leq 6$.

9.6.2 Transformation Method

Using this method is easier, but it has some restrictions; it cannot be used on all functions. Suppose X is a continuous random variable and g is a function of X such that g is monotonic (strictly increasing or strictly decreasing) over the range of X. Let $Y = g(X)$, and assume that $g(X)$ is differentiable over the range of X. The PDF of Y is then given by:

$$f_y(y) = f_x(g^{-1}(y)) \cdot |\frac{d}{dy}g^{-1}(y)|$$

The range of Y is given by taking the function g of the range of X. You may have to look at a graph to get this information.

Example Suppose X follows $N(\mu, \sigma^2)$ and $Y = aX + b$ where a and b are constants. Find the pdf of Y. From earlier it's already known that this comes out to $N(a\mu + b, a^2\sigma^2)$, but now provide it.

$$\begin{aligned}
f_x(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
y &= ax + b = g(x) \\
x &= \frac{y-b}{a} = g^{-1}(y) \\
\frac{d}{dy}\frac{y-b}{a} &= \frac{1}{a} \\
f_y(y) &= f_x(g^{-1}(y)) \cdot |\frac{d}{dy} g^{-1}(y)| \\
&= f_x(\frac{y-b}{a}) \cdot \frac{1}{|a|} \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} \cdot \frac{1}{|a|} \\
&= \frac{1}{\sqrt{2\pi}\sigma|a|} e^{\frac{-(y-(a\mu+b))^2}{2\sigma^2 a^2}} \\
&= N(a\mu + b, a^2\sigma^2)
\end{aligned}$$

Example 2 Let X be a random variable with pdf given by $f_x(X) = \frac{2x}{\pi^2}$ with range $0 \leq X \leq \pi$. Define $Y = \sin(X)$. Find the pdf of Y.

Since Y is not monotonic over the given range, the transformation method is not applicable here. Therefore, we must use the CDF method. First, find the cdf of y, $F_y(Y) = P(Y \leq y)$.

$$\begin{aligned}
F_y(Y) &= P(Y \leq y) \\
&= P(0 \leq x \leq \arcsin(y)) + P(\pi - \arcsin(y) \leq x \leq \pi) \\
&= P(x \leq \arcsin(y)) - P(x \leq 0) + P(x \leq \pi) - P(x \leq \pi - \arcsin(y)) \\
&= F_x(\arcsin(y)) - F_x(0) + F_x(\pi) - F_x(\pi - \arcsin(y)) \\
f_y(y) &= f_x(\arcsin(y)) \cdot \frac{d}{dy}(\arcsin(y)) - 0 + 0 - f_x(\pi - \arcsin(y)) \cdot (-1)\frac{d}{dy}(\arcsin(y)) \\
&= \frac{2\arcsin(y)}{\pi^2} \cdot \frac{1}{\sqrt{1-y^2}} + \frac{2(\pi - \arcsin(y))}{\pi^2} \cdot \frac{1}{\sqrt{1-y^2}} \\
&= \frac{2\pi}{\pi^2} \cdot \frac{1}{\sqrt{1-y^2}} \\
&= \frac{2}{\pi} \cdot \frac{1}{\sqrt{1-y^2}}
\end{aligned}$$

All that is left to find is the range of the above formula... using the above, $0 \leq y \leq 1$. Actually, wait, y cannot be 1, so it's $0 \leq y < 1$.

Example 3 Let X follow a uniform distribution $(0, 1)$ and $Y = X^2$. Find the pdf of Y.

Since this is monotonic, we can use the transformation method here. $f(x) = 1, 0 \leq x \leq 1$. Let $y = x^2 = g(x)$; this means that $g^{-1}(x) = \sqrt{y}$ (since the range of X is only positive, ignore the negative results of $\sqrt{y}$).

$$
\begin{aligned}
\frac{d}{dy}g^{-1}(y) &= \frac{1}{2\sqrt{y}} \\
f_y(y) &= f_x(g^{-1}(y) \cdot |\frac{d}{dy}g^{-1}(y)| \\
&= f_x(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\
&= \frac{1}{2\sqrt{y}}
\end{aligned}
$$

Y cannot be 0 here for obvious reasons, so the range is $0 < y \leq 1$.