Module 1

1 Basic Definitions and Results of Probability

We will begin the class by listing some of the basic rules, results, and definitions used of probability. Most of you have already seen what's in this module's notes, but it will help to review them. Many of the examples and illustrations are drawn from various sources.

Definition 1 The sample space of an experiment, S, is the set of all possible outcomes of that experiment.

Definition 2 An event is any subset of the sample space.

Result 1 If all events in S are equally likely, then the probability of an event $A \subset S$, $P(A)$, is calculated as $\frac{|A|}{|S|}$, where $|A|$ is the number of events in A.

Result 2 Associative Law: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, and $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$. I encourage you to draw Venn Diagrams to illustrate the associative law.

Result 3 $P(S) = 1$. The concept of this law is simple. The probability of anything happening is 1.

Result 4 If A_1 and A_2 are disjoint, $(A \cup A_2 = \emptyset)$ (i.e., if they don't overlap) then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. More generally, if $A_1, A_2, \ldots, A_n$ are mutually disjoint, i.e., none of them overlap, then

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_j).$$

Result 5 Let A^c denote the event that A does not happen. In this case, $P(A^c) = 1 - P(A)$.

Result 6 $P(\emptyset) = 0$. Mathematically, this states that the probability of the empty set (the probability that nothing happens) is 0.

Result 7 If $A \subset B$ (if A is a subset of B), then $P(A) \leq P(B)$.

Result 8 The probability of A or B happening, $P(A \cup B)$, is calculated as $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Definition 3 For any two events A and B, we define the conditional probability of A "given" B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

From this equation, it follows that the joint probability of A and B can be calculated as $P(A \cap B) = P(A|B) \times P(B)$.

Result 9 The Law of Total Probability: Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup B_i = S$ (i.e., they exhaust and encompass the entire sample space), and $B_i \cap B_j = \emptyset$ for any $i \neq j$ (i.e., none of them overlap). Then for any other event A,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n). \end{aligned}$$

Result 10 Bayes' Rule: We know from the law of total probability that if $B_1, B_2, \ldots, B_n$ are mutually exclusive and $\bigcup_{j=1}^{n} B_j = S$, then for any event A.

$$P(A) = \sum_{j=1}^{n} P(A \cap B_j) = \sum_{j=1}^{n} P(A|B_j)P(B_j).$$

With this information, we can calculate the probability of B_j given A. This is where Bayes' Rule kicks in.

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Definition 4 The events A and B are said to be "independent" if $P(A \cap B) = P(A)P(B)$, and the events $A_1, A_2, \ldots, A_n$ are said to be "mutually independent" if

$$P(A_1 \cap A_2 \cap A_3 \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n).$$

Intuitively, independence is a simple idea. If A and B are independent, then knowing whether A happened or not gives you no knowledge of whether B happened.

Below are some examples illustrating the concepts and definitions we have listed above.

Example 1 Driving to work, a commuter passes through three intersections with traffic lights. At each light, she either stops, "s", or continues, "c'.' The sample space, S, is the set of all possible outcomes and can be written as

$$S = \{sss, ssc, scs, css, ccs, csc, scc, ccc\}.$$

The first event in the sample space listed above is that she stops at all three lights, the second is that she stops at the first two lights and continues on the third, etc. Now let's consider the event A, where $A =$ the event that she stops once. A can be written as

$$A = \{ccs, csc, scc\}.$$

If all events are equally likely, then $P(A) = \frac{3}{8}$. Let B be the event that she continues at least twice. In this case,

$$B = \{ccc, ccs, csc, scc\}$$

and $P(B) = \frac{1}{2}$. The event that both A and B happen, written as $A \cap B$, is the event that she stops once and continues twice.

$$A \cap B = \{ccs, csc, scc\}.$$

Since there are only three events in this case,

$$P(A \cap B) = \frac{3}{8}.$$

Now consider the event $A \cup B$, the event that she stops once or continues twice. In this case,

$$A \cup B = \{ccs, ccc, csc, scc\}.$$

The probability of $A \cup B$ is $P(A \cup B) = \frac{4}{8}$.

Example 2 Consider the population of people who attended the inauguration of President Donald Trump. Many of the people in attendance were supporters of Trump and came from out of town, while others were supports of Clinton and came to protest the inauguration. Many of those people also came from out of town. Assume the population can be split as such (and you should know that I'm totally making these numbers up):

	Trump	Clinton
DC Resident	.05	.36
Not DC Resident	.52	.07

The fractions within the table are joint probabilities. In other words, the probability that a randomly selected person at the inauguration was a supporter of Clinton and not a DC resident is 7%. With these joint probabilities, it is simple to calculate conditional probabilities. For example, to calculate the probability that a randomly selected attendant of the inauguration is a supporter for Trump given they are a DC resident is

$$P(\text{Trump}|\text{DC resident}) = \frac{P(\text{Trump } \cap \text{ DC Resident})}{P(\text{DC Resident})} = \frac{.05}{.41} = .12.$$

The probability that anyone at the inauguration voted for Trump (whether they were from the city or not) can be calculated as $P(\text{voted for Trump}) = .52 + .05 = .57$.

Example 3 An urn contains three red balls and one blue ball. What is the probability that a red ball is selected on the second draw?

In this case, let R_1 be the event that a red ball is drawn on the first draw, let B_1 be the event that a blue ball is selected on the first draw, and so on. The probability that a red ball is selected on the second draw can be written as $P(R_2)$ and calculated as

$$P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1) = \frac{2}{3} \times \frac{3}{4} + 1 \times \frac{1}{4} = \frac{3}{4}.$$

Example 4 Suppose that occupations are grouped into Upper (U), Middle (M), and Lower (L) levels such that

	U_2	M_2	L_2
U_1	0.45	0.48	0.07
M_1	0.05	0.70	0.25
L_1	0.01	0.50	0.49

This table should be read in the following way: $P(U_2|U_1) = .45$ implies that the probability that someone in generation 2 is in the upper class given that their parents were in the upper class is .45. $P(L_2|M_1) = .25$ implies that the probability that someone in generation 2 is in the lower class given that their parents were in the middle class is .25. With this table available, let's ask the following question: If 10% of Generation 1 is in U, 40% is in M, and 50% is in L, what is the probability that a randomly selected individual from the second generation will be in U? This probability can be written as $P(U_2)$ and calculated as

$$P(U_2) = P(U_2|L_1)P(L_1) + P(U_2|M_1)P(M_1) + P(U_2|U_1)P(U_1) = .01 \times .5 + .05 \times .4 + .45 \times .1 = .07.$$

Example 5 Polygraph Tests: Let + be the event that a polygraph test is "positive." Let L be the event that the subject really is lying. Let's also assume that the polygraph is designed so that $P(+|L) = 0.88$ (the probability that polygraph correctly identifies a subject as lying is 88%), and $P(-|L) = 0.12$ (the probability that the polygraph incorrectly identifies a subject as telling the truth is 12%). Let's also assume that $P(-|T) = 0.86$ (the probability that polygraph correctly identifies a subject as telling the truth is 86%. Given all of this information, and given that a vast majority of people don't lie, (P(T) = .99), what is the probability that someone is telling the truth given that the polygraph says they're telling the truth. This example is asking us to calculate $P(T|+)$, which can be calculated as

$$P(T|+) = \frac{P(+|T)P(T)}{P(+)} = \frac{P(+|T)P(T)}{P(+|T)P(T) + P(+|L)P(L)} = \frac{.14 \times .99}{.14 \times .99 + .88 \times .01} = .94.$$

Example 6 This example is taken from Rice's undergraduate text in mathematical statistics. Suppose that virus transmissions of AIDS in 500 acts of intercourse are mutually independent events. Also assume that the probability of transmissino in any one act is $\frac{1}{500}$. What is the probability of infection after 500 acts of intercourse? In this case, we take advantage of the fact that P(no infection) is easier to calculate than P(infection). We can calculate P(infection) as $1 - P$(no infection). Specifically,

$$P(\text{infection}) = 1 - P(\text{no infection}) = 1 - P\left(I_1^c \cap I_2^c \cap I_3^c \cap \cdots \cap I_{500}^c\right),$$

where I_j^c is the probability of not being infected on the j^{th} act of intercourse. The above probability can be calculated as

$$1 - P\left(I_1^c\right) P\left(I_2^c\right) \cdots P\left(I_{500}^C\right) = 1 - \left(1 - \frac{1}{500}\right)^{500} = .37.$$

2 Univariate Random Variables and Expectation

In this section, we introduce the concept of random variables, why they're used, how they're used, and how to calculate probabilities and expectations associated with them.

2.1 Random Variables

To motivate the concept of a random variable, we'll begin again with a random experiment. The classic random experiment that everybody likes to talk about is flipping a fair coin three times. In this case, the sample space is

$$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}.$$

And if we want to discuss the probability of events in the sample space, or if we we want to write it down, we would write

- Probability of getting two tails: $P(\text{getting two tails}) = P(\{TTH, THT, HTT\}) = \frac{3}{8}$.
- Probability of getting at least two tails: $P(\{TTH, THT, HTT, TTT\}) = \frac{4}{8}$.
- Probability of at least getting one head: $P(\{HTT, THT, TTH, HHT, HTH, THH, HHH\}) = \frac{7}{8}$.

But writing all of these events down (just as I have done) gets really annoying. And this is ultimately why random variables exist. Random variables ultimately make it easy to unambiguously define events in the sample space. To be more specific, a random variable is a map (or a function) from the sample space to the real line.

Let's use the concept of a random variable to ease the burden of writing down the probabilities listed above. Let X = the number of T in three tosses of a fair coin. The above probabilities can be written as

- Probability of getting two tails: $P(X = 2)$
- Probability of getting at least two tails: $P(X \geq 2)$.
- Probability of getting at least one head: $P(X \leq 2)$.

There are two types of random variables, discrete random variables and continuous random variables. Associated with every discrete random variable is a probability mass function, or pmf. The pmf of a discrete random variable X will be denoted in this set of notes as $p_X(x)$. The function $p_X(x)$ is the probability mass function of X evaluated at x and it is to be interpreted as $p_X(x) = P(X = x)$ To illustrate that X follows the pmf $p_X(x)$, we write $X \sim p_X(x)$. Most of the time it is understood in the context of the problem that $p_X(x)$ is the probability mass function of X, and in such cases $p_X(x)$ is simply written as $p(x)$. With the probability mass function of X, one can calculate the cumulative distribution function of X, $P(X \leq x)$, as $F_X(x) = P(X \leq x)$ as

$$P(X \leq x) = \sum_{x^*: x^* \leq x} p_X(x^*).$$

A continuous random variable can, theoretically, take any value in a given interval, and associated with it is a probability density function, $f_X(x)$. If X has probability density function $f_X(x)$, $X \sim f_X(x)$, then the probability that X is in between any two numbers a and b is calculated as t

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

The cumulative distribution function of X, $F_X(x)$, is calculated as

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt.$$

Properties of the cumulative distribution function include:

1. $\lim_{x\to\infty} F_X(x) = 1$. (conceptually, this is simple: $F_X(\infty) = P(X \leq \infty) = 1$).
2. If $x \leq y$, then $F_X(x) \leq F_X(y)$. (If y is bigger than x, then $P(X \leq x) \leq P(X \leq y)$).
3. $f_X(x) = \frac{d}{dx}F_X(x)$. This, of course, is a consequence of the Fundamental Theorem of Calculus, which states that $f(x) = \frac{d}{dx}\int_{-\infty}^x f(t)dt.$.

Here are some examples related to random variables

Example 1 An appliance store receives a shipment of 30 microwave ovens, 5 of which (unknown to the manager) are defective. Store manager selects four at random, without replacement, and tests to see if they are defective. Let $X =$ the number of defective microwave ovens found. What is the probability mass function of X?

To answer this question, let's first define the sample space.

$$S = \{G_1G_2G_3G_4,\ G_1G_2G_3B_4,\ \ldots, B_1B_2B_3B_4\},$$

where G_i means the i^{th} oven is a "good" oven and B_j means the j^{th} oven is a "bad" oven.

- $P(X = 0) = P(\text{all good}) = P(G_1)P(G_2|G_1)P(G_3|G_2, G_1)P(G_4|G_1, G_2, G_3) = \frac{25}{30} \times \frac{24}{29} \times \frac{23}{28} \times \frac{22}{27} = .46$
- $P(X = 1) = P(B_1, G_2, G_3, G_4) + P(G_1, B_2, G_3, G_4) + P(G_1, G_2, B_3, G_4) + P(G_1, G_2, G_3, B_4) = \frac{5}{30}\frac{25}{29}\frac{24}{28}\frac{23}{27} + \frac{25}{30}\frac{5}{29}\frac{24}{28}\frac{23}{27} + \frac{25}{30}\frac{24}{29}\frac{5}{28}\frac{23}{27} + \frac{25}{30}\frac{24}{29}\frac{23}{28}\frac{5}{27} = .42$

Strategies similar to that above give us $P(X = 2) = .11$, $P(X = 3) = .01$, and $P(X = 4) = .00$. Verify these calculations on your own.

Example 2 A certain river floods every ear. The low water mark is set to 1 and the high water-mark Y has cumulative distribution function $F_Y(y) = 1 - \frac{1}{y^2}$. $0 \le y \le \infty$.

1. To verify that this is a cdf, we first observe that $\lim_{y\to\infty} F_Y(y) = 1$, and then observe that if $y_2 \ge y_1$, then $F_Y(y_2) = 1 - \frac{1}{y_2^2} \ge 1 - \frac{1}{y_1^2} = F_Y(y_1)$.
2. $f_Y(y) = \frac{d}{dy}\left\{1 - \frac{1}{y^2}\right\} = 2y^{-3}$.
3. What is $P(1 \le Y \le 3)$? $P(1 \le Y \le 3) = \displaystyle\int_1^3 2y^{-3}dy = \frac{8}{9}$.

2.2 Expectation

With probability mass functions and probability density functions, one can calculate the expected value of a random variable. The expected value is just at is sounds; it is the value of the random variable that you "expect." The most basic example of the expected value is for a fair coin. Assume you flip a fair coin once, and let

$$X = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T} \end{cases}$$

If you flip this coin one time, the expected value of X, $E(X)$ is 0.5. This expectation is calculated as $E(X) = 1 \times P(X = 1) + 0 \times P(X = 0) = 0.5$. In general, for a discrete random variable X with pmf $p_X(x)$,

$$E(X) = \sum_{\text{all } x} x p_X(x),$$

and for a continuous random variable Y with pdf $f_Y(y)$, the expected value is calculated as

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy.$$

It should be noted that the expected value of X is also denoted as μ_X, and it should also be noted that for any function of X, $h(X)$,

$$E[h(X)] = \sum_{\text{all } x} h(x)p_X(x) \text{ if } X \text{ is discrete,} \quad \text{and } E[h(X)] = \int_{-\infty}^{\infty} h(x)f_X(x)dx \text{ if} X \text{ is continuous.}$$

The variance of a random variable determines how much the value of X bounces around the mean. The variance of X, denoted as either Var(X) or σ_X^2, is calculated as

$$Var(X) = \sigma_X^2 = E\left[(X-\mu_X)^2\right] = \underbrace{\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx}_{\text{if } X \text{ continuous}} = \underbrace{\sum_{\text{all } x} (x-\mu_X)^2 p_X(x)}_{\text{if } X \text{ discrete}}.$$

A useful property of the variance is that if $Y = aX$, then $Var(Y) = a^2 Var(X)$.

Below is an example (and a trick) which shows how expected values and variances can be calculated.

Example 1 : Let X = the number of calls I get from Sirius XM in 1 week (they really want me to renew my subscription). Assume that $X \sim p_X(x)$, where

$$p_X(x) = \frac{e^{-1.2}1.2^x}{x!} \quad x = 0, 1, 2, 3, ...$$

What is the expected number of phone calls I get in one week?

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x p_X(x) = \sum_{x=0}^{\infty} x e^{-1.2}1.2^x/x! = 0 \times e^{-1.2}1.2^x/x! + \sum_{x=1}^{\infty} x e^{-1.2}1.2^x/x! \\ &= 1.2\left\{\sum_{x=1}^{\infty} e^{-1.2}1.2^x/(x-1)!\right\} = 1.2.\left\{\sum_{y=0}^{\infty} e^{-1.2}1.2^y/y!\right\} = 1.2 \end{aligned}$$

The last equality can be made if we let $Y = X - 1$. We know the last equality holds because the sum in brackets is just the sum of a probability mass function. This probability mass function can be thought of as the pmf of Y where $Y = X - 1$.

The variance of X is

$$\begin{aligned} \sigma_X^2 &= E\left[(X-\mu_X)^2\right] = E\left[\left(X^2 - 2\mu_X X + \mu_X^2\right)\right] = E(X^2) - 2\mu_X E(X) + E\left(\mu_y\right)^2 \\ &= E\left(X^2\right) - 2\mu_X\mu_X + \mu_X^2 = E(X^2) - \mu_X^2 \end{aligned}$$

The equation above gives us an easy way to calculate σ_X^2; we could just calculate $E(X^2)$ and then subtract the squared mean. But that's not going to work in this case. Instead, we will calculate

$$\begin{aligned} E\left[X(X-1)\right] &= E(X^2) - E(X) = \sum_{x=0}^{\infty} x(x-1)e^{-1.2}1.2^x/x! = 0 + 0 + \sum_{x=2}^{\infty} e^{-1.2}1.2^x/(x-2)! \\ &= 1.2^2 \sum_{x=2}^{\infty} \frac{e^{-1.2}1.2^{x-2}}{(x-2)!} = 1.2^2\left\{\sum_{y=0}^{\infty} \frac{e^{-1.2}1.2^y}{y!}\right\} = 1.2^2 \end{aligned}$$

Again, the last equality holds because the sum in curly brackets is the sum of Y's pmf, where Y is now $X - 2$ ($Y = X - 2$).

$$\begin{aligned} E\left[X(X-1)\right] &= E\left(X^2\right) - E(X) = 1.2^2 \Longrightarrow E\left(X^2\right) = 1.2^2 + 1.2 \\ \Longrightarrow \sigma_X^2 &= E\left(X^2\right) - E(X)^2 = 1.2^2 + 1.2 - 1.2^2 = 1.2. \end{aligned}$$

2.3 Some Common Random Variables

Now let's talk about different types of probability distributions.

Bernoulli A Bernoulli random variable counts the number of successes in one success/no-success trial. The probability of success in one trial is p. In this case $X \sim p_X(x)$ where $p_X(x) = p^x(1-p)^{1-x}$ and $E(X) = p$.

Binomial X = the number of successes in n success/no-success trials with probability of a single success being p. In this case, $X \sim p_X(x)$, where $p_X(x) = P(x \text{ successes out of } n \text{ trials}) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x(1-p)^{n-x}$. For a Binomial random variable, $E(X) = np$ and $\text{Var}(X) = np(1-p)$.

Geometric X = the number of success/no-success trials until the first success. In this case, $X \sim p_X(x)$ where $p_X(x) = (1-p)^{x-1}p$. For a Geometric random variable, $E(X) = \frac{1}{p}$ and $\text{Var}(X) = \frac{1-p}{p^2}$.

Poisson A Poisson random variable is usually meant to count the number of occurrences of rare events. It has a rate parameter λ, and the pmf of X is $p_X(x)$, where $p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ and as we have shown in the above example, $E(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Uniform We say that $X \sim \text{Unif}(a,b)$ is X is uniformly distributed on the interval (a,b). The density of X is $f_X(x) = \frac{1}{b-a} \quad a \le x \le b$. The expected value of X in this case is $E(X) = \frac{b-a}{2}$.

Exponential If X has an exponential distribution with parameter λ, that is denoted as $X \sim \text{Exp}(\lambda)$. The pdf of X is $f_X(x) = \lambda \exp\{-\lambda x\}$, and the expected value is $E(X) = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$.

Gamma If X has a Gamma distribution with parameter λ and k ($X \sim \text{Gamma}(\lambda, k)$), then the density of X is $f_X(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} \exp\{-x\lambda\}$, The expected value of X is k/λ and the variance is $k/(\lambda^2)$.

Normal This is your standard bell-shaped curve. $X \sim N(\mu, \sigma^2)$ if $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_X^2}\right\}$, with $E(X) = \mu$ and $Var(X) = \sigma_X^2$.

2.4 Transformations

In this section we look at how to calculate the distributions and densities of random variables that have been transformed. For example, if X = the radius of a randomly generated sphere, and assuming it has density $f_X(x)$, then what is the density of the sphere's volume, Y, where $Y = \frac{4}{3}X^3$. We approach this problem generically and make use of the relationship between densities and distributions, i.e., that $f_Y(y) = \frac{d}{dy}F_Y(y)$ to derive the density of Y.

Here's how to do a transformation: Assume $X \sim f_X(x)$ and $Y = g(X)$ where g is some 1-to-1 function. To find the distribution of Y you must first recall that $f_Y(y) = \frac{d}{dy}F_Y(y)$. Let's do this calculation step-by-step.

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}P(Y \le y) = \frac{d}{dy}P\left(g(X) \le y\right) = \frac{d}{dy}P(X \le g^{-1}(y)) \\
&= \frac{d}{dy}F_X\left(g^{-1}(y)\right) = \frac{d}{dy}\int_{-\infty}^{g^{-1}(y)} f_X(x)dx \\
&= f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right|.
\end{aligned}
$$

Note that the reason g has to be 1-to-1 is so we can invert it and equate $P(g(X) \le y) = P(X \le g^{-1}(y))$.

Example 1 : Assume $Z \sim N(0,1)$. Let $X = Z^2$. What is the density of X? We begin by defining the density of Z. Recall that

$$
f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \quad -\infty < z < \infty.
$$

Now observe that $X = g(Z)$ where $g(Z) = Z^2 \Longrightarrow Z = g^{-1}(X) = \sqrt{X}$. Note that since g is not 1-to-1, I will redefine the density of Z so that it is. Since the density of Z is symmetric about 0 and $g(-z) = g(z)$, we could just as well say that Z is a half-normal random variable with density

$$f_Z(z) = \frac{2}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \quad 0 < z < \infty$$

and set $X = g(Z) = Z^2$, where $g^{-1}(X) = \sqrt{X}$. This is what we'll do. The necessary calculations to find $f_X(x)$ are then

$$f_X(x) = f_Z\left(x^{\frac{1}{2}}\right) \frac{1}{2} x^{-\frac{1}{2}} = \frac{x^{-\frac{1}{2}}}{\sqrt{2\pi}} \exp\left\{-\frac{x}{2}\right\}.$$

Example 2 : Assume that $X \sim f_X(x)$ and $Z = aX + b$. What is the density of Z? In this case, $g(X) = aX + b$ and $g^{-1}(Z) = \frac{(Z-b)}{a} = \frac{Z}{a} - \frac{b}{a}$. In this case,

$$f_Z(z) = f_X\left(g^{-1}(z)\right)\left(\frac{dg^{-1}(z)}{dz}\right) = f_X\left(\frac{z-b}{a}\right)\frac{1}{a}.$$

From this we can calculate the expected value of Z in terms of the expected value of X. and its variance, and we will use this result often.

$$\begin{aligned}
E(Z) &= \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} z f_X\left(\frac{z}{a} - \frac{b}{a}\right)\frac{1}{a} dz. \quad \text{Recall that } x = \frac{z}{a} - \frac{b}{a} \to dx = \frac{1}{a} dz \\
&= \int_{-\infty}^{\infty} (ax+b) f_X(x) \frac{1}{a} a dx = a\int_{-\infty}^{\infty} x f_X(x) dx + b\int_{-\infty}^{\infty} f_X(x) dx \\
&= aE(X) + b.
\end{aligned}$$

Prove to yourself that $\text{Var}(Z) = a^2 \text{Var}(X)$.

Example 3 : Here is another result that we will use extensively. Assume that $X \sim f_X(x)$, and let $Z = F_X(X)$. (So we're generating a value of X and putting it into it's own cumulative distribution function.) Then $Z \sim \text{Unif}(0,1)$. Here's a proof:

$$\begin{aligned}
f_Z(z) &= \frac{d}{dz} P(Z \leq z) = \frac{d}{dz} P\left(F_X(X) \leq z\right) = \frac{d}{dz} P\left(X \leq F_X^{-1}(z)\right) \\
&= \frac{d}{dz} F_X\left(F_X^{-1}(z)\right) = \frac{d}{dz} z = 1.
\end{aligned}$$

As I said before, this is a critical result that will eventually be used in generating random numbers from an arbitrary density.