Common Probability Distributions

Nathaniel E. Helwig

University of Minnesota

1 Overview

As a reminder, a random variable X has an associated probability distribution $F(\cdot)$, also know as a cumulative distribution function (CDF), which is a function from the sample space S to the interval $[0, 1]$, i.e., $F : S \rightarrow [0, 1]$. For any given $x \in S$, the CDF returns the probability $F(x) = P(X \leq x)$, which uniquely defines the distribution of X. In general, the CDF can take any form as long as it defines a valid probability statement, such that $0 \leq F(x) \leq 1$ for any $x \in S$ and $F(a) \leq F(b)$ for all $a \leq b$.

As another reminder, a probability distribution has an associated function $f(\cdot)$ that is referred to as a probability mass function (PMF) or probability distribution function (PDF). For discrete random variables, the PMF is a function from S to the interval $[0, 1]$ that associates a probability with each $x \in S$, i.e., $f(x) = P(X = x)$. For continuous random variables, the PDF is a function from S to $\mathbb{R}^+$ that associates a probability with each range of realizations of X, i.e., $\int_a^b f(x)dx = F(b) - F(a) = P(a < X < b)$.

Probability distributions that are commonly used for statistical theory or applications have special names. In this chapter, we will cover a few probability distributions (or families of distributions) that are frequently used for basic and applied statistical analyses. As we shall see, the families of common distributions are characterized by their parameters, which typically have a practical interpretation for statistical applications.

Definition. In statistics, a parameter $\theta = t(F)$ refers to a some function of a probability distribution that is used to characterize the distribution. For example, the expected value $\mu = E(X)$ and the variance $\sigma^2 = E((X - \mu)^2)$ are parameters that are commonly used to describe the location and spread of probability distributions.

2 Discrete Distributions

2.1 Bernoulli Distribution

Definition. In statistics, a Bernoulli trial refers to a simple experiment that has two possible outcomes, i.e., $|S| = 2$. The two outcomes are $x = 0$ (failure) and $x = 1$ (success), and the probability of success is denoted by $p = P(X = 1)$. It does not matter which of the two results we call the "success", given that the probability of the "failure" is simply $1 - p$.

The probability distribution associated with a Bernoulli trial is known as a Bernoulli distribution, which depends on the parameter p. The Bernoulli distribution has properties:

- PMF: $f(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$
- CDF: $F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1 \end{cases}$
- Mean: $E(X) = p$
- Variance: $\mathrm{Var}(X) = p(1-p)$

***Example* 1.** Suppose we flip a coin once and record the outcome as 0 (tails) or 1 (heads). This experiment is an example of a Bernoulli trial, and the random variable X (which denotes the outcome of a single coin flip) follows a Bernoulli distribution. A fair coin has equal probability of heads and tails, so $p = 1/2$ if the coin is fair. But note that the Bernoulli distribution applies to unfair coins as well, e.g., if the probability of heads is $p = 3/4$, the random variable X still follows a Bernoulli distribution.

***Example* 2.** Suppose we roll a fair dice and let $y \in \{1, \ldots, 6\}$ denote the number of dots. Furthermore, suppose we define $x = 0$ if $y \in \{1, 2, 3\}$ and $x = 1$ if $y \in \{4, 5, 6\}$. Then the random variable X follows a Bernoulli distribution with $p = 1/2$. Now suppose that we change the definition of X, such that $x = 0$ if $y < 6$ and $x = 1$ if $y = 6$; in this case, the random variable X follows a Bernoulli distribution with $p = 1/6$.

In both of these examples, note that there are two possible outcomes for X (0 and 1), and the distribution of these outcomes is determined by the probability of success p.

2.2 Binomial Distribution

The binomial distribution is related to the Bernoulli distribution. If $X = \sum_{i=1}^{n} Z_i$ where the Z_i are independent and identically distributed Bernoulli trials with probability of success p, then the random variable X follows a binomial distribution. Note that a binomial distribution has two parameters: $n \in \{1, 2, 3, \ldots\}$ and $p \in [0, 1]$. The number of Bernoulli trials n (sometimes called the "size" parameter) is known by the design of the experiment, whereas the probability of success may unknown.

The binomial distribution has the properties:

- PMF: $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient
- CDF: $F(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$
- Mean: $E(X) = np$
- Variance: $\mathrm{Var}(X) = np(1-p)$

***Example* 3.** Suppose that we flip a coin $n \geq 1$ independent times, and assume that each flip Z_i has probability of success $p \in [0, 1]$, where a result of heads is considered a "success". If we define X to be the total number of observed heads, then $X = \sum_{i=1}^{n} Z_i$ follows a binomial distribution with parameters n and p. See Figures 2 and 4 in the Random Variables notes for depictions of the PMF and CDF for the coin flipping example with $p = 1/2$.

If X follows a binomial distribution with parameters n and p, it is typical to write $X \sim B(n, p)$, where the $\sim$ symbol should be read as "is distributed as". Note that if $n = 1$, then the Binomial distribution is equivalent to the Bernoulli distribution, i.e., the Bernoulli distribution is a special case of the Binomial distribution when there is only one Bernoulli trial. As the number of independent trials $n \to \infty$, we have that

$$\frac{X - np}{\sqrt{np(1-p)}} \to Z \sim N(0, 1)$$

where $N(0, 1)$ denotes a standard normal distribution (later defined). In other words, the normal distribution is the limiting distribution of the binomial for large n.

2.3 Discrete Uniform Distribution

Suppose that a simple random experiment has possible outcomes $x \in \{a, a+1, \ldots, b-1, b\}$ where $a \leq b$ and $m = 1 + b - a$. If all of the m possible outcomes are equally likely, i.e., if $P(X = x) = 1/m$ for any $x \in \{a, \ldots, b\}$, the distribution is referred to as a discrete uniform distribution, which depends on two parameters: the two endpoints a and b.

The discrete uniform distribution has the properties:

- PMF: $f(x) = 1/m$
- CDF: $F(x) = (1 + \lfloor x \rfloor - a)/m$
- Mean: $E(X) = (a + b)/2$
- Variance: $\mathrm{Var}(X) = [(b - a + 1)^2 - 1]/12$

If X follows a discrete uniform distribution with parameters a and b, it is typical to write $X \sim U\{a, b\}$. Note that there also exists a continuous uniform distribution (later described), which has a similar notation. Thus, whenever you are using a uniform distribution, it is important to be clear whether or not you're assuming a discrete or continuous distribution.

***Example* 4.** Suppose we roll a fair dice and let $x \in \{1, \ldots, 6\}$ denote the number of dots. The random variable X follows a discrete uniform distribution with $a = 1$ and $b = 6$.

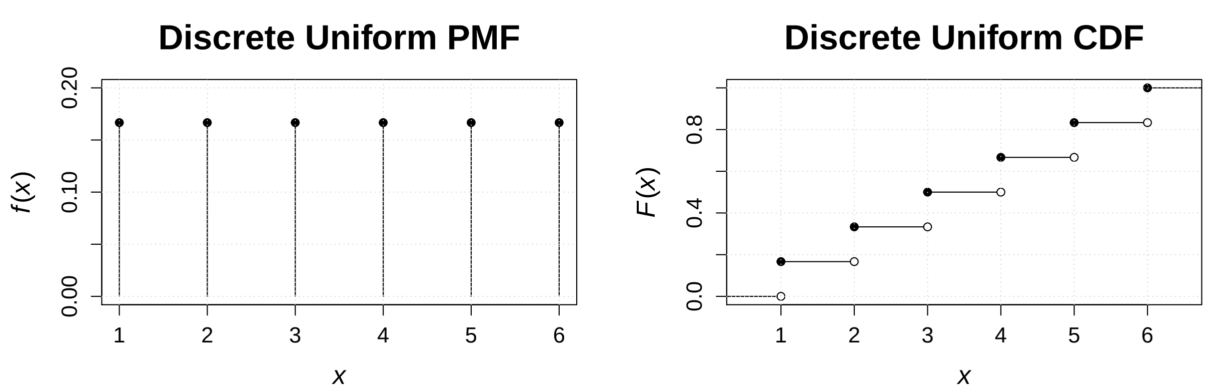

Figure 1: Discrete uniform distribution PDFs and CDFs with $m = 6$ possible outcomes.

3 Continuous Distributions

3.1 Normal Distribution

The normal (or Gaussian) distribution is the most well-known and commonly used probability distribution. The normal distribution is quite important because of the central limit theorem, which is discussed in the following section. The normal distribution is a family of probability distributions defined by two parameters: the mean μ and the variance σ^2.

The normal distribution has the properties:

- PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ where $\exp(x) = e^x$ is the exponential function
- CDF: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2} dz$ is the standard normal CDF
- Mean: $E(X) = \mu$
- Variance: $\mathrm{Var}(X) = \sigma^2$

To denote that X follows a normal distribution with mean μ and variance σ^2, it is typical to write $X \sim N(\mu, \sigma^2)$ where the $\sim$ symbol should be read as "is distributed as".

Definition. The standard normal distribution refers to a normal distribution where $\mu = 0$ and $\sigma^2 = 1$. Standard normal variables are typically denoted by $Z \sim N(0, 1)$.

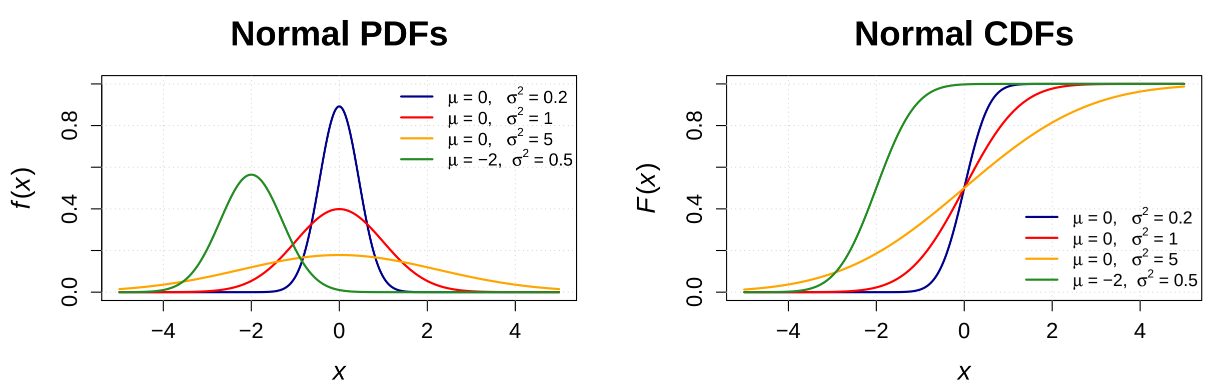

Figure 2: Normal distribution PDFs and CDFs with various different means and variances.

3.2 Chi-Square Distribution

The chi-square distribution is related to the normal distribution. If $X = \sum_{i=1}^{k} Z_i^2$ where the Z_i are independent standard normal distributions, then the random variable X follows a chi-square distribution with degrees of freedom k. The chi-square distribution is defined by a single parameter: the degrees of freedom k. Note that since the chi-square is the summation of squared standard normal variables, we have that $X > 0$.

The chi-square distribution has the properties:

- PDF: $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$ where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the gamma function
- CDF: $F(x) = \frac{1}{\Gamma(k/2)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right)$ where $\gamma(u, v) = \int_0^v t^{u-1} e^{-t} dt$ is the lower incomplete gamma function
- Mean: $E(X) = k$
- Variance: $\mathrm{Var}(X) = 2k$

To denote that X follows a chi-square distribution with degrees of freedom k, it is typical to write $X \sim \chi_k^2$ or $X \sim \chi^2(k)$, where the symbol χ is the Greek letter "chi". Because the chi-square distribution is related to the normal distribution, the chi-square distribution is used (almost) as frequently as the normal distribution. In particular, the chi-square distribution is often used to assess the "goodness of fit" of a statistical model.

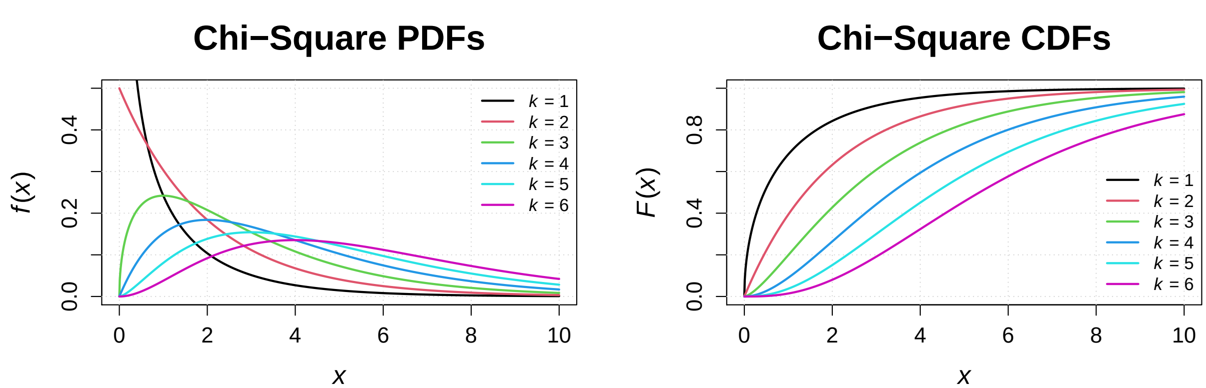

Figure 3: Chi-square distribution PDFs and CDFs with various different degrees of freedom.

3.3 F Distribution

The F distribution is related to the chi-square distribution. A random variable X has an F distribution if the variable can be written as

$$X = \frac{U/m}{V/n}$$

where $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are independent chi-square random variables. The F distribution depends on two parameters: the two degrees of freedom parameters m and n.

The F distribution has the properties:

- PDF: $f(x) = \frac{\sqrt{\frac{(xm)^m n^n}{(xm+n)^{m+n}}}}{xB(\frac{m}{2},\frac{n}{2})}$ where $B(u,v) = \int_0^1 t^{u-1}(1-t)^{v-1}dt$ is the beta function
- CDF: $F(x) = I_{\frac{xm}{xm+n}}\left(\frac{m}{2},\frac{n}{2}\right)$ where $I_x(u,v) = \frac{B(x;u,v)}{B(u,v)}$ is the regularized incomplete beta function and $B(x;u,v) = \int_0^x t^{u-1}(1-t)^{v-1}dt$ is the incomplete beta function
- Mean: $E(X) = \frac{n}{n-2}$ assuming that $n > 2$
- Variance: $\mathrm{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ assuming that $n > 4$

To denote that X follows an F distribution with degrees of freedom (m, n), it is typical to write $X \sim F_{m,n}$ or $X \sim F(m,n)$. Note that the F distribution was conceptualized by George Snedecor, who called it the F distribution in honor of R. A. Fisher.

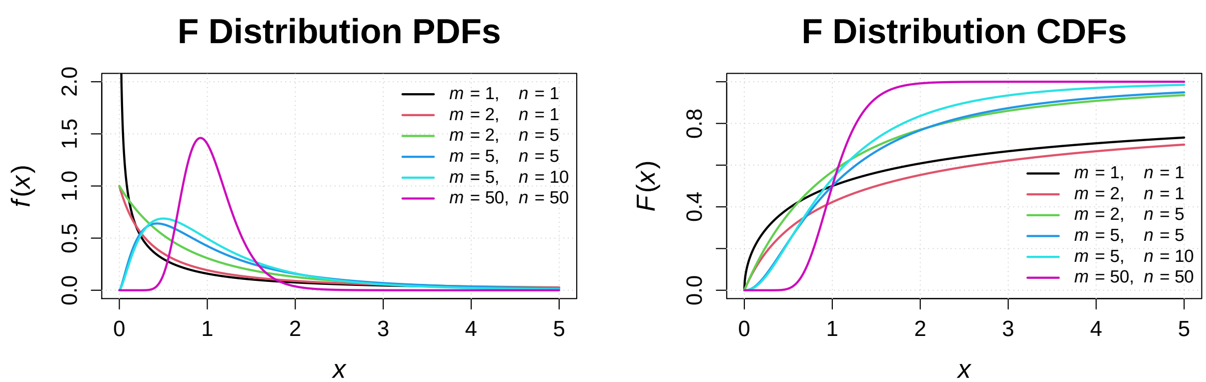

Figure 4: F distribution PDFs and CDFs with various different degrees of freedom.

3.4 Continuous Uniform Distribution

Suppose that a simple random experiment has an infinite number of possible outcomes $x \in [a, b]$ where $-\infty < a < b < \infty$. If all of the possible outcomes are equally likely, the distribution is referred to as a continuous uniform distribution, which depends on two parameters: the two endpoints a and b.

The continuous uniform distribution has the properties:

- PDF: $f(x) = \frac{1}{b-a}$ if $x \in [a, b]$ (note that $f(x) = 0$ otheriwse)
- CDF: $F(x) = \frac{x-a}{b-a}$ if $x \in [a, b]$ (note that $F(x) = 0$ if $x < a$ and $F(x) = 1$ if $x > b$)
- Mean: $E(X) = (a + b)/2$
- Variance: $\text{Var}(X) = (b - a)^2/12$

If X follows a continuous uniform distribution with parameters a and b, it is typical to write $X \sim U[a, b]$ or $X \sim U(a, b)$. Note that $U[a, b]$ assumes that the interval is closed, whereas the notation $U(a, b)$ assumes that the interval is open.

***Example* 5.** Suppose that we randomly spin the second hand around the face of a clock, and define X as the position where the second hand stops spinning. The random variable X follows a continuous uniform distribution with $a = 0$ and $b = 12$.

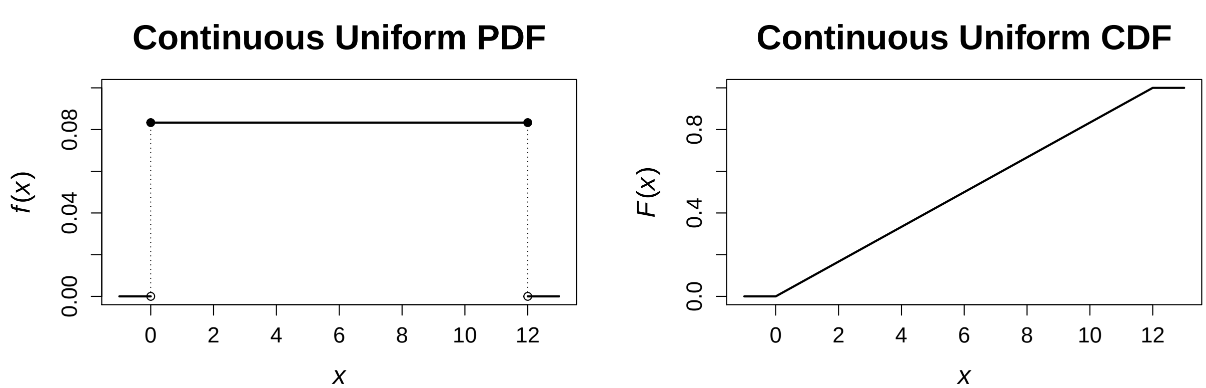

Figure 5: Continuous uniform distribution PDF and CDF with $a = 0$ and $b = 12$.

4 Central Limit Theorem

Let $x_1, \ldots, x_n$ denote an independent and identically distributed (iid) sample of data from some probability distribution F with mean μ and variance $\sigma^2 < \infty$. The central limit theorem reveals that as the sample size $n \to \infty$, we have that

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

where $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ is the sample mean. Note that the symbol $\xrightarrow{d}$ should be read as "converges in distribution to". This theorem reveals that, for a large enough sample size n, the sample mean $\bar{x}_n$ is approximately normally distributed with mean μ and variance σ^2/n, i.e., $\bar{x}_n \dot{\sim} N(\mu, \frac{\sigma^2}{n})$ where the symbol $\dot{\sim}$ should be read as "is approximately distributed as".

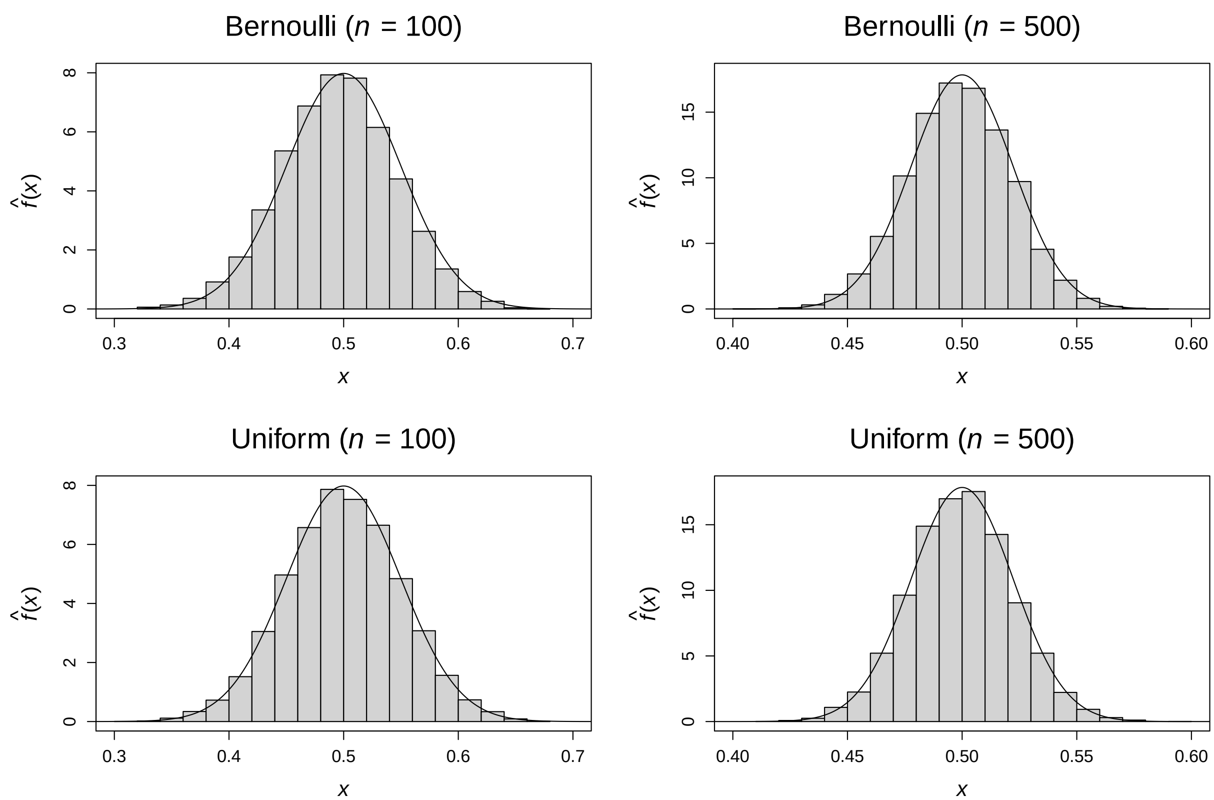

Figure 6: The central limit theorem. Top: $X \sim B(1, 1/2)$. Bottom: $X \sim U[\frac{2-\sqrt{12}}{4}, \frac{2+\sqrt{12}}{4}]$. Note that $\mu = 1/2$ and $\sigma^2 = 1/4$ for both distributions. The histograms depict the approximate sampling distribution of $\bar{x}_n$ and the lines denote the $N(\mu, \sigma^2/n)$ density.

5 Affine Transformations of Normal Variables

Another reason that the normal distribution is so popular for applied research is the fact that normally distributed variables are convenient to work with. This is because affine transformations of normal variables are also normal variables.

Definition. Given a collection of variables $X_1, \ldots, X_p$, an affine transformation has the form $Y = a + b_1 X_1 + \cdots b_p X_p$, where a is an offset (intercept) term and b_j is the weight/coefficient that is applied to the j-th variable.

As a reminder from our Introduction to Random Variables chapter, if X is a random variable with mean $\mu = E(X)$ and variance $\sigma^2 = E((X-\mu)^2)$, then the affine transformation $Y = a + bX$ has mean $\mu_Y = E(Y) = a + b\mu$ and variance $\sigma_Y^2 = \mathrm{Var}(Y) = b^2\sigma^2$. If the random variable X is normally distributed, i.e., if $X \sim N(\mu, \sigma^2)$, then the random variable $Y = a + bX$ is also normally distributed, i.e., $Y \sim N(\mu_Y, \sigma_Y^2)$.

As a reminder from our Introduction to Random Variables chapter, if X_1 and X_2 are random variables with means $\mu_j = E(X_j)$ and variances $\sigma_j^2 = E((X_j - \mu_j)^2)$ for $j \in \{1, 2\}$, then the affine transformation $Y = a + b_1 X_1 + b_2 X_2$ has mean and variance

$$
\begin{aligned}
\mu_Y &= E(Y) = a + b_1\mu_1 + b_2\mu_2 \\
\sigma_Y^2 &= \mathrm{Var}(Y) = b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + 2b_1 b_2 \sigma_{12}
\end{aligned}
$$

where $\sigma_{12} = E[(X_1 - \mu_1)(X_2 - \mu_2)]$ is the covariance between X_1 and X_2. Note that $\sigma_{12} = 0$ if X_1 and X_2 are independent of one another, but the converse is only true if X_1 and X_2 are normally distributed. If the random variables X_1 and X_2 are normally distributed, then the random variable $Y = a + b_1 X_1 + b_2 X_2$ is also normally distributed, i.e., $Y \sim N(\mu_Y, \sigma_Y^2)$.

More generally, if we define $Y = a + \sum_{j=1}^p b_j X_j$, then the mean and variance of Y are

$$
\begin{aligned}
\mu_Y &= E(Y) = a + \sum_{j=1}^p b_j \mu_j \\
\sigma_Y^2 &= \mathrm{Var}(Y) = \sum_{j=1}^p b_j^2 \sigma_j^2 + 2\sum_{j=2}^p \sum_{k=1}^{j-1} b_j b_k \sigma_{jk}
\end{aligned}
$$

where $\sigma_{jk} = E[(X_j - \mu_j)(X_k - \mu_k)]$. If the collection of variables $(X_1, \ldots, X_p)$ have a multivariate normal distribution, then $Y \sim N(\mu_Y, \sigma_Y^2)$.