## Common Probability Distributions

#### Nathaniel E. Helwig

Associate Professor of Psychology and Statistics University of Minnesota



August 28, 2020

Copyright  $\bigodot$  2020 by Nathaniel E. Helwig

- 1. Probability Distribution Background
- 2. Discrete Distributions
- 3. Continuous Distributions
- 4. Central Limit Theorem
- 5. Affine Transformations of Normal Variables

## Table of Contents

#### 1. Probability Distribution Background

- 2. Discrete Distributions
- 3. Continuous Distributions
- 4. Central Limit Theorem
- 5. Affine Transformations of Normal Variables

### Distribution, Mass, and Density Functions

A random variable X has a cumulative distribution function (CDF)  $F(\cdot)$ , which is a function from the sample space S to the interval [0, 1].

• 
$$F(x) = P(X \le x)$$
 for any given  $x \in S$ 

•  $0 \le F(x) \le 1$  for any  $x \in S$  and  $F(a) \le F(b)$  for all  $a \le b$ 

 $F(\cdot)$  has an associated function  $f(\cdot)$  that is referred to as a probability mass function (PMF) or probability distribution function (PDF).

• PMF (discrete): f(x) = P(X = x) for all  $x \in S$ 

• PDF (continuous): 
$$\int_a^b f(x)dx = F(b) - F(a) = P(a < X < b)$$

### Parameters and Statistics

A parameter  $\theta = t(F)$  refers to a some function of a probability distribution that is used to characterize the distribution.

• 
$$\mu = E(X)$$
 and  $\sigma^2 = E[(X - \mu)^2]$  are parameters

Given a sample of data  $\mathbf{x} = (x_1, \dots, x_n)^{\top}$ , a <u>statistic</u>  $T = s(\mathbf{x})$  is some function of the sample of data.

• 
$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
 and  $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$  are statistics

Parametric distributions have a finite number of parameters, which characterize the form of the CDF and PMF (or PDF).

• The parameters define a family of distributions

### Table of Contents

- 1. Probability Distribution Background
- 2. Discrete Distributions
- 3. Continuous Distributions
- 4. Central Limit Theorem
- 5. Affine Transformations of Normal Variables

# Bernoulli Distribution

A <u>Bernoulli trial</u> refers to a simple experiment that has two possible outcomes. The two outcomes are x = 0 (failure) and x = 1 (success), and the probability of success is denoted by p = P(X = 1).

- One parameter:  $p \in [0, 1]$
- Notation:  $X \sim \text{Bern}(p)$  or  $X \sim B(1, p)$

The Bernoulli distribution has properties:

• PMF: 
$$f(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$$
  
• CDF:  $F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \le x < 1 \\ 1 & \text{if } x = 1 \end{cases}$ 

• Mean: E(X) = p

• Variance: 
$$Var(X) = p(1-p)$$

# Bernoulli Distribution Example

Suppose we flip a coin and record the outcome as 0 (tails) or 1 (heads).

This experiment is an example of a Bernoulli trial, and the random variable X (the outcome of a coin flip) follows a Bernoulli distribution.



Figure 1: Bernoulli PMF with p = 1/2 (left) and p = 3/4 (right).

# **Binomial 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.

- Two parameters:  $n \in \{1, 2, 3, \ldots\}$  and  $p \in [0, 1]$
- Notation:  $X \sim B(n, p)$

The binomial distribution has the properties:

• PMF:  $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$  where  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ 

• CDF: 
$$F(x) = \sum_{i=0}^{\lfloor x \rfloor} {n \choose i} p^i (1-p)^{n-i}$$

• Mean: E(X) = np

• Variance: 
$$Var(X) = np(1-p)$$

# Binomial Distribution Example

Suppose that we flip a coin  $n \ge 1$  independent times, and assume that each flip  $Z_i$  has probability of success (i.e., heads)  $p \in [0, 1]$ .

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.



Figure 2: Binomial PMF with  $n \in \{5, 10\}$  and p = 1/2.

Nathaniel E. Helwig (Minnesota) Common Probability Distributions C A

# Limiting Distribution of Binomial

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.

Note that the binomial distribution (depicted on the previous slide) looks reasonably bell-shaped for n = 10.

## Discrete Uniform Distribution

Suppose that a simple random experiment has possible outcomes  $x \in \{a, a + 1, \dots, b - 1, b\}$  where  $a \leq b$  and m = 1 + b - a, and all m possible outcomes are equally likely, i.e., if P(X = x) = 1/m.

- Two parameters: the two endpoints a and b (with a < b)
- Notation:  $X \sim U\{a, b\}$

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:  $Var(X) = [(b a + 1)^2 1]/12$

# Discrete Uniform Distribution Example

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



Figure 3: Discrete uniform distribution PDF and CDF with a = 1 and b = 6.

13 / 28

## Table of Contents

- 1. Probability Distribution Background
- 2. Discrete Distributions
- 3. Continuous Distributions
- 4. Central Limit Theorem
- 5. Affine Transformations of Normal Variables

# 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 (later defined).

- Two parameters: the mean  $\mu$  and the variance  $\sigma^2$
- Notation:  $X \sim N(\mu, \sigma^2)$

The <u>standard normal</u> distribution refers to a normal distribution where  $\mu = 0$  and  $\sigma^2 = 1$ , which is typically denoted by  $Z \sim N(0, 1)$ .

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  $\exp(x) = e^x$ 

• CDF: 
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$
 where  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} dz$ 

• Mean:  $E(X) = \mu$ 

• Variance: 
$$Var(X) = \sigma^2$$

# Normal Distribution Visualizations

The CDF has a elongated "S" shape, which is called an ogive.



Figure 4: Normal distribution PDFs and CDFs with different  $\mu$  and  $\sigma^2$ .

# Chi-Square 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.

- One parameter: the degrees of freedom  $\boldsymbol{k}$
- Notation:  $X \sim \chi_k^2$  or  $X \sim \chi^2(k)$

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: Var(X) = 2k

# Chi-Square Distribution Visualization

 $\chi^2(k)$  distribution approaches normal distribution as  $k \to \infty$ 



Figure 5: Chi-square distribution PDFs and CDFs with different DoF.

# F Distribution

A random variable X has an F distribution if the variable has the form

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

where  $U \sim \chi^2(m)$  and  $V \sim \chi^2(n)$  are independent chi-square variables.

- $\bullet\,$  Two parameters: the two degrees of freedom parameters m and n
- Notation:  $X \sim F_{m,n}$  or  $X \sim F(m,n)$

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  $B(u,v) = \int_0^1 t^{u-1}(1-t)^{v-1}dt$ • CDF:  $F(x) = I_{\frac{xm}{xm+n}}\left(\frac{m}{2},\frac{n}{2}\right)$  where  $I_x(u,v) = \frac{B(x;u,v)}{B(u,v)}$  and  $B(x;u,v) = \int_0^x t^{u-1}(1-t)^{v-1}dt$ • Mean:  $E(X) = \frac{n}{n-2}$  assuming that n > 2• Variance:  $Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$  assuming that n > 4

# F Distribution Visualization

Note that if  $X \sim F(m, n)$ , then  $X^{-1} \sim F(n, m)$ 



Figure 6: F distribution PDFs and CDFs with different DoF.

# Continuous Uniform Distribution

Suppose that a simple random experiment has an infinite number of possible outcomes  $x \in [a, b]$  where  $-\infty < a < b < \infty$ , and all possible outcomes are equally likely.

- Two parameters: the two endpoints a and b
- Notation:  $X \sim U[a, b]$  or  $X \sim U(a, b)$  (closed or open interval)

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:  $Var(X) = (b a)^2/12$

## Continuous Uniform Distribution Visualization



Figure 7: Continuous uniform PDF and CDF with a = 0 and b = 12.

### Table of Contents

- 1. Probability Distribution Background
- 2. Discrete Distributions
- 3. Continuous Distributions
- 4. Central Limit Theorem
- 5. Affine Transformations of Normal Variables

# 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  $\mu$  and variance  $\sigma^2 < \infty$ .

The central limit theorem reveals that as the sample size  $n \to \infty$ 

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

where  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$  is the sample mean.

For a large enough n, the sample mean  $\bar{x}_n$  is approximately normally distributed with mean  $\mu$  and variance  $\sigma^2/n$ , i.e.,  $\bar{x}_n \sim N(\mu, \frac{\sigma^2}{n})$ .

### Central Limit Theorem Visualization



Figure 8: 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.

## Table of Contents

- 1. Probability Distribution Background
- 2. Discrete Distributions
- 3. Continuous Distributions
- 4. Central Limit Theorem

#### 5. Affine Transformations of Normal Variables

## Definition of Affine Transformation

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.

Given a collection of variables  $X_1, \ldots, X_p$ , an <u>affine</u> 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.

• Linear transformations are a special case with a = 0

## Affine Transformations of Normals

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.,  $Y \sim N(\mu_Y, \sigma_Y^2)$  where  $\mu_Y = E(Y) = a + b\mu$  and  $\sigma_Y^2 = \operatorname{Var}(Y) = b^2 \sigma^2$ .

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

$$\mu_{Y} = E(Y) = a + \sum_{j=1}^{p} b_{j}\mu_{j}$$
$$\sigma_{Y}^{2} = \operatorname{Var}(Y) = \sum_{j=1}^{p} b_{j}^{2}\sigma_{j}^{2} + 2\sum_{j=2}^{p} \sum_{k=1}^{j-1} b_{j}b_{k}\sigma_{jk}$$

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