4: Introduction to Probability

Introduction

Types of Probability

In statistics it is often necessary to make statements along the lines of "the probability of observing a given event is *p* percent." Dealing with probabilistic statements, however, is not often easy. Part of this difficulty stems from the innately abstract nature of probability, and part stems from fundamental questions about the nature of probability. Nevertheless, if we hope to use probability as a tool for inference, its must first be defined. Three definitions of probability are offered.

Definition 1 (Probability Based on Logic). This view presupposes we are dealing with a finite set of possibilities which are drawn at random. For example, in drawing from a deck of cards there are a finite set of possible hands in a deck of cards. Such problems are easily solved if we can recognize all possible outcomes and count the number of ways that a particular event may occur. Then, the probability of the event is the number of times it can occur divided by the total number of possibilities. To use a familiar example, we all recognize that there are 52 cards in a deck and 4 of these cards are Kings. Thus, the probability of drawing a King = 4 / 52 = .0769.

Definition 2 (Probability Based on Experience). This second view of probability assumes that *if* a process is repeated a large enough number of times *n*, and if event *A* occurs *x* of these times, then the probability of event *A* will *converge* on *x* / *n* as *n* becomes large. If we flip a coin many times we expect to see half of the flips turn up heads. Such estimates will become increasingly reliable as the number of replications (*n*) increases. For example, if a coin is flipped 10 times, there is no guarantee that exactly 5 heads will be observed **--** the proportion of heads will range from 0 to 1, although in most cases we would expect it to be closer to .50 than to 0 or 1. However, if the coin is flipped 100 times, chances are better that the proportions of heads will be close to .50. With 1000 flips, the proportion of heads will be an even better reflection of the true probability.

Definition 3 (Subjective Probability). In this view, probability is treated as a quantifiable level of belief ranging from 0 (complete disbelief) to 1 (complete belief). For instance, an experienced physician may say "this patient has a 50% chance of recovery." Presumably, this is based on an understanding of the relative frequency of might occur in similar cases. Although this view of probability is subjective, it permits a constructive way for dealing with uncertainty.

An appreciation of the various types of probability are not mutually exclusive. And fortunately, all obey the same mathematical laws, and their methods of calculation are similar. All probabilities are a type of relative frequency—the number of times something can occur divided by the total number of possibilities or occurrences. Thus,

$$\text{The probability of event A} = \frac{\text{no. of times event A can occur}}{\text{total no. of occurrences}}$$

Several Properties of Probabilities

At this point, many statistical texts would cover the laws and axioms of probability. We will take a less formal approach by introducing only selected properties of probabilities:

1. ***The range of possible probabilities.*** This may seem obvious, but keep in mind that probabilities can be no less than zero and no more than one. A statement that the probability is 110%, of course, is ridiculous.

2. ***Complements.*** We often speak of the complement of a event. The complement of an event is its "opposite," or event NOT happening. For example, if the event under consideration is being correct, the complement of the event is being incorrect. If we are denote an event with the symbol A, the complement may be denoted as the same symbol with a line overhead ($\bar{A}$). The sum of the probabilities of an event and its complement is always equal to one:

$$\Pr(A) + \Pr(\bar{A}) = 1$$

   Therefore, the probability of the complementary of an event is equal to 1 minus the probability of the event:

$$\Pr(\bar{A}) = 1 - \Pr(A)$$

   For example, if the probability of being correct is .95, the probability of being incorrect = 1 − .95 = .05. In contrast if the probability of being correct is .99, then the probability of being incorrect = 1 − .99 = .01.

3. ***Probability distributions.*** The usefulness of probability theory comes in understanding probability distributions (also called probability functions and probability densities or masses). Probability distributions list or describe probabilities for all possible occurrences of a random variable. There are two types of probability distributions:

   - Discrete distributions
   - Continuous distributions.

4. **Discrete probability distributions** describes a finite set of possible occurrences, for discrete "count data." For example, the number of successful treatments out of 2 patients is discrete, because the random variable represent the number of success can be only 0, 1, or 2. The probability of all possible occurrences—Pr(0 successes), Pr(1 success), Pr(2 successes)—constitutes the probability distribution for this discrete random variable.

5. **Continuous probability distributions** describe an "unbroken" continuum of possible occurrences. For example, the probability of a given birthweight can be anything from, say, half a pound to more than 12 pounds (or something like that). Thus, the random variable of birthweight is continuous, with an infinite number of possible points between any two values. (Think in terms of Xeno's Paradox.)

There are many families of discrete and continuous probability distributions. We will study only a couple, starting with the binomial distribution.

Binomial Distributions

Basics

We define a **Bernoulli trial**. as a random event that can take on only one of two possible outcomes, with the outcomes arbitrarily denoted as either a “success” or “failure.” For example, flipping a coin is an example of a Bernoulli trial since the outcome is either a head or a tail. (It doesn’t matter which of these outcomes we call a *success*, as long as we are consistent throughout the course of the problem.)

The total number of successes (X) observed in a series of n independent Bernoulli trials is a **binomial random variable,** and the listing of probabilities for all possible outcomes is a **binomial distribution**. For example, the number of heads occurring when tossing a fair coin twice is a binomial random variable with the following distribution:

Pr(0 heads) = 0.25
Pr(1 heads) = 0.50
Pr(2 heads) = 0.25

Although “the number of heads in tossing a fair coin twice” is a trivial example of a binomial distribution, binomial random variables are of central importance in biostatistics. Examples of relevant binomial random are:

- The number of patients out of n that respond to treatment.
- The number of people in a community of n people that have asthma.
- The number of people in a group of n intravenous drug users who are HIV positive.

The binomial distribution is actually a *family* of distributions with each family member identified by two binomial **parameters**:

n the number of independent trials
p the probability of success per trial

The **notation** for binomial random variables is $X \sim \mathrm{b}(n, p)$, meaning “X is a binomial random variable with parameters n and p.” For example, $X \sim \mathrm{b}(2, 0.5)$ means “X is distributed as a binomial random variable with $n = 2$ and $p = 0.5$.”

The **expected value** of a binomial random variable is its number of successes in the long run. The formula for the expectation of a binomial distribution is:

$$m = np \qquad \textbf{(4.1)}$$

For example, the expected value for X~b(2,.5) is $\mu = (2)(0.5) = 1$.

The **variance** (σ^2) of a binomial random variable quantifies the spread of the function. The variance of a binomial variable is:

$$s^{2} = npq \qquad \textbf{(4.2)}$$

where $q = 1 - p$. For example, the variance of the number of heads when tossing a fair coin twice is $\sigma^2 = (2)(0.5)(0.5) = 0.5$.

Calculation of Binomial Probabilities

To calculate binomial probabilities, we must first learn about the "choose function." The **choose function** quantifies the number of different ways to choose *i* objects of *n*. Let ${}_n\mathbf{C}_i$ denote the number of different ways to choose *i* items out of *n* without repetition. Then,

$$ {}_nC_i = \frac{n!}{i!(n-i)!} \quad \textbf{(4.3)} $$

where **"!"** represents the factorial function. The **factorial function** is the product of the series of integers from *n* to 1. Thus, $n! = (n)(n-1)\ (n-2)(n-3) \ldots (1)$. For example, 2! = (2)(1) = 2, 3! = (3)(2)(1) = 6, and 4! = (4)(3)(2)(1) = 24. By definition, 0! = 1

Using formula 4.3, ${}_3\mathrm{C}_2 = \frac{3!}{(2!)(3-2)!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1)(1)} = \frac{6}{2} = 3$. This says "3 choose 2 is equal to 3." In other words, there are three ways to choose 2 items out of 3: for items labeled A, B, and C, you may choose {A, B}, {A, C}, or {B, C}.

Let us now consider choosing 2 items out of 6. Using formula 4.3, ${}_4\mathrm{C}_2 = \frac{4!}{(2!)(4-2)!} = \frac{4 \cdot 3 \cdot \cancel{2!}}{(2 \cdot 1)(\cancel{2!})} = \frac{4 \cdot 3}{2} = 6$. In other words, there are 6 ways to choose 2 items out of 4: {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, or {C, D}.

One more example of the choose function: ${}_7\mathrm{C}_3 = \frac{7!}{(3!)(7-3)!} = \frac{7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{(3 \cdot 2 \cdot 1)(\cancel{4!})} = 35$. This says "7 choose 3 is equal to 35."

Binomial probabilities are calculated with the formula:

$$ \Pr(X = i) = {}_nC_i\, p^i q^{n-i} \quad \textbf{(4.4)} $$

where *i* is the observed number of successes
n is the number of trials
p is the probability of success for each trial
$q = 1 - p$

Illustrative Example. Suppose a treatment is successful 75% of the time (probability of success = .75). This treatment is used in 4 patients ($n = 4$). What is the probability of seeing 2 successes in these 4 patients? Let X represent the number of successful treatments. Thus, $\Pr(X = 2) = {}_4C_2\,(.75)^2\,(.25)^{4-2} = (6)(.5625)(.0625) = .2109$.

StaTable. StaTable is a public domain probability calculator available via http://www.cytel.com/statable/ . The program is available in a Java version (that runs through your browser) and a Window version (that must be downloaded and installed before use). The binomial function in the Windows version of the program is available by clicking the `Distribution` menu and then selecting `Discrete > Binomial`. Enter the number of successes in the field labeled *i* and binomial parameters n and p in their labeled fields. The binomial probability is listed in the field labeled Pr(i).

Binomial Probability Functions (Distributions)

The listing of probabilities for all possible outcomes of a binomial random variable is a *binomial distribution* or *binomial function*.

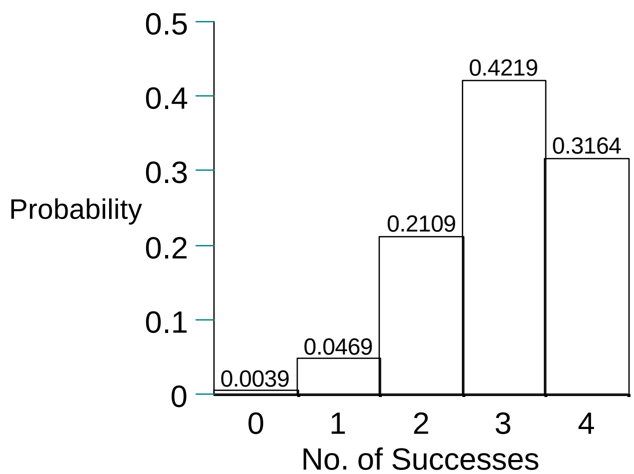
Binomial Illustrative Distribution 1. In tossing a fair coin twice, the number of "heads" is a binomial random variable with n = 2 and p = .5 with a probability distribution (function) of:

Probability of 0 heads	=	$\Pr(X = 0) = ({}_2C_0)(.5^0)(.5^{2-0})$	=	(1)(1)(.25)	=	.25
Probability of 1 head	=	$\Pr(X = 1) = ({}_2C_1)(.5^1)(.5^{2-1})$	=	(2)(.5)(.5)	=	.50
Probability of 2 heads	=	$\Pr(X = 2) = ({}_2C_2)(.5^2)(.5^{2-2})$	=	(1)(.25)(1)	=	.25

Binomial Illustrative Distribution 2. Suppose a treatment is successful 75% of the time. This treatment is used in 4 patients. The number of cures (X) is thus a binomial random variable with n = 4 and p = .75. The probability distribution for this random variable is:

The probability of 0 successes	=	Pr(X = 0)	=	${}_4C_0\,(.75)^0\,(.25)^{4-0}$	=	(1)(1)(.0039)	= .0039
The probability of 1 success	=	Pr(X = 1)	=	${}_4C_1\,(.75)^1\,(.25)^{4-1}$	=	(4)(.75)(.0156)	= .0469
The probability of 2 successes	=	Pr(X = 2)	=	${}_4C_2\,(.75)^2\,(.25)^{4-2}$	=	(6)(.5625)(.0625)	= .2109
The probability of 3 successes	=	Pr(X = 3)	=	${}_4C_3\,(.75)^3\,(.25)^{4-3}$	=	(4)(.4219)(.25)	= .4219
The probability of 4 successes	=	Pr(X = 4)	=	${}_4C_4\,(.75)^4\,(.25)^{4-4}$	=	(1)(.3164)(1)	= .3164

The complete probability function can be displayed as a **probability histogram**:



Notice that the **area contained in each bar** corresponds to its probability. For example, the bar for Pr(X = 0) has a height of .0039 and a width of 1 (from 0 to just before 1). Thus, its area = *height* × *width* = .0039 × 1 = .0039. Similarly, Pr(X = 1) = *height* × *width* = .0469 × 1 = .04169. (And so on.)

Cumulative Probability Functions

In addition to wanting to know the probability of a given number of successes, we often want to know the probability of observing *less than or equal* to a given number of successes. This is referred to as the **cumulative probability of the event**.

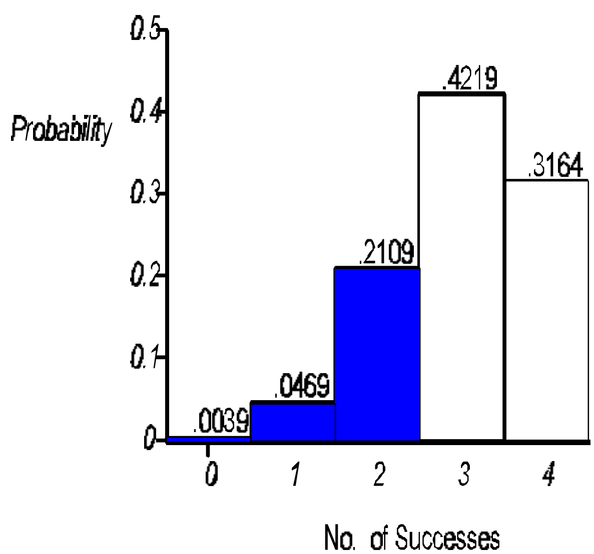
Cumulative Binomial Illustrative Distribution 1. The cumulative probability function for illustrative distribution 1 (the number of head out of 2 tosses) is:

Probability of at least 0 heads	=	$Pr(X \le 0) = Pr(X = 0)$		=	.25
Probability of at least 1 head	=	$Pr(X \le 1) = Pr(X \le 0) + Pr(X = 1)$	= .25 + .50	=	.75
Probability of at least 2 heads	=	$Pr(X \le 2) = Pr(X \le 1) + Pr(X = 2)$	= .75 + .25	=	1.00

Cumulative Binomial Illustrative Distribution 2. The cumulative probability distribution for illustrative distribution 2 (number of successful treatments out of 4 when p = .75) is:

The probability of at least 0 successes	$= Pr(X \le 0) = Pr(X = 0)$	=	.0039
The probability of at least 1 successes	$= Pr(X \le 1) = Pr(X \le 0) + Pr(X = 1) = .0039 + .0469$	=	.0508
The probability of at least 2 successes	$= Pr(X \le 2) = Pr(X \le 1) + Pr(X = 2) = .0508 + .2109$	=	.2617
The probability of at least 3 successes	$= Pr(X \le 3) = Pr(X \le 2) + Pr(X = 3) = .2617 + .4219$	=	.6836
The probability of at least 4 successes	$= Pr(X \le 4) = Pr(X \le 3) + Pr(X = 4) = .6836 + .3164$	=	1.0000

Cumulative probabilities corresponds to the **left tail** of a distribution. For example, the figure below shades the area corresponding to the cumulative probability of 2 successes i.e., $Pr(X \le 2)$:

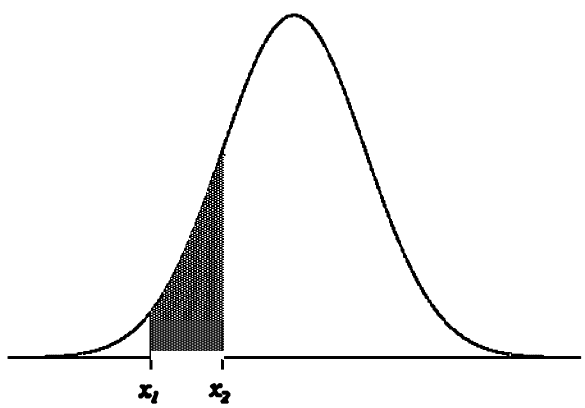


Normal Distributions

Continuous Probability Distributions

The previous considered the probability distribution of variables with only two kinds of possible outcomes: positives or negatives, people with or with an attribute, or successes or failures. The random number of successes of *n* made up a binomial distribution. The variable was an enumerations of the number of successes, and was thus *discrete*. We now turn to another kind of variable whose individuals are measured for a characteristic such as height or age. The variable flows without a break and is thus *continuous*, with no limit to the number of individuals with different measurements. Such measurements are distributed in any of a number of ways. We will consider the normal distribution as an example of a continuous distribution because of its utility and wide use in applied statistical methods.

Whereas discrete probability distributions (such as binomial distributions) are displayed with histograms, continuous probability distributions (such as normal distributions) are displayed with curves:



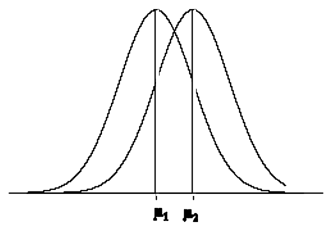
Probability curves demonstrate the following important properties:

1. The total area under the curve sums to 1
2. The probability of any exact value is equal to 0
3. The area under the curve (AUC) between any two points is the probability of values in that range (shaded area)

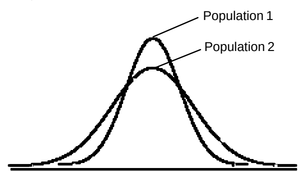
These properties (especially the first and third) allow use to determine the approximate probability of various potentialities using a variety of probability tables.

Characteristics of Normal Distributions

The **normal distribution** (also called the **Gaussian distribution**) is a family of distributions recognized as being symmetrical, unimodal, and bell-shaped. The normal distribution is characterized by two parameters: μ and σ. The **mean (μ)** determines the distribution's location. The figure below shows two normal distributions with different means:



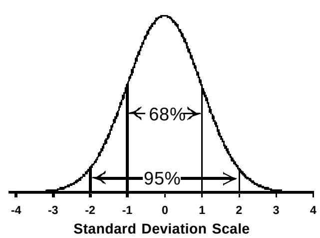
The **standard deviation (s)** of a particular normal distribution determines its spread. The figure below demonstrates two normal distributions with different spreads:



Notation: Let X~N(μ, σ) denote a particular normal random distribution with mean μ and standard deviation σ. For example, a normal random variable with a mean of 0 and standard deviation of 1 is denoted X~N(0,1). A normal random variable with a mean 100 and standard deviation of 15 is denoted X~N(100, 15).

For all normal distributions:

- 68% of the data lies within ±1 standard deviations of μ
- 95% of the data lies within±2 standard deviation of μ
- nearly all of the data lies with±3 standard deviation of μ



As an illustration, consider a normal random variable with a mean of 100 and standard deviation of 15. Sixty-eight percent (68%) of the areas under the curve for this distribution will lie in the range 100 ± 15 (between 85 and 115). In contrast, 95% of values lie in the range 100 ± 30 (between 70 and 130). Virtually all observations fall in the range 100 ± 45 (between 55 and 145).

Standard Normal Distribution

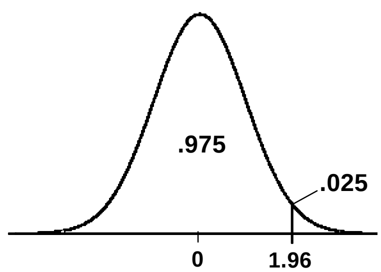
The **standard normal distribution** (also called the ***Z* distribution**) is a normal distribution with a mean of 0 and standard deviation of 1**.** The utility of the *Z* distribution is based on the availability of **standard normal tables** (Appendix 2).

Before using a standard normal table, data must be **standardized.** To standardize a value, subtract the distribution's mean (μ) and divide by its standard deviation (σ). A value, once standardized, is called a z-score.

$$z = \frac{x - m}{s} \qquad \textbf{(4.5)}$$

The z-score places the value above or below the mean in standard deviation units. For example, a *z* core of +1 shows the value to be one standard deviation *above* the population mean. In contrast, *z* score −1 shows the value to be 1 standard deviation *below* the population mean.

Notation: Let z_p denote a *z* score with a lower tail probability of *p*. For example, $z_{.975} = +1.96$:



This particular *z* score ($z_{.975}$) is now referred to as the 97.5th **percentile** of the standard normal distribution since it is greater than or equal to 97.5% of other scores on the distribution. You look up ***z* percentiles** on a standard normal table. To accomplish this it is helpful to draw the area under the curve. Several **key landmarks** on the standard normal curve are noted:

- $z_{.025} = -1.96$
- $z_{.500} = 0.00$
- $z_{.975} = +1.96$

Whenever working with z percentiles, it is helpful to DRAW THE CURVE!

Application of the Normal Distribution

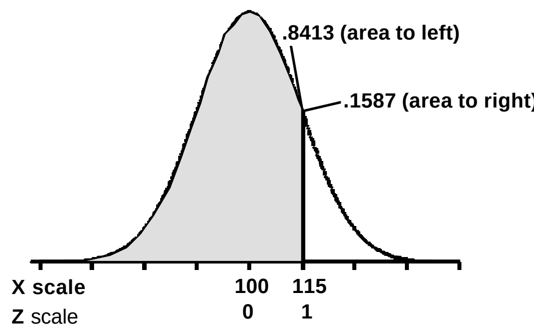
The normal distribution is helpful in solving many real-world problems. Many measurements are approximately distributed normally (e.g., IQ scores) and many can be transformed to something that is approximately normally distributed (e.g., the logarithm of cholesterol values).

Here is a method that seems to work for most students in solving normal probability problems:

1. Draw a diagram of a normal curve.
2. Mark the location of the mean on the curve.
3. Standardize the value you are considering.
4. Place the standardized value on the curve in relation to the mean or point of inflection (the points of inflection of the curve is $\pm 1\sigma$ from the mean.)
5. Shade the appropriate area under the curve that corresponds to the problem and then determine the area under the curve with your Z table.

Example: Suppose you have a normally distributed random variable with $\mu = 100$ and $\sigma = 15$ and you want to know what percentage of values fall below 115. Thus:

1. You draw a normal curve.
2. You mark the center of the curve at 100.
3. You standardize the value: $z = (115 - 100) / 15 = +1$
4. You place the z score 1 standard deviation above the mean
5. You look up a z score of +1 on the standard normal table and find that 0.8413 of the area under the curve is to its left. Thus, this value is greater than or equal to 84.13% of the values in the population.



Illustrative Example #2: Suppose scores on an exam are approximately normally distributed with a mean of 80 and standard deviation of 10. What percentage of scores are less than 65?

Solution:

1. A normal curve is drawn.
2. The middle of the curve is labeled with a value of 80.
3. The value of 65 is standardized: $z = (65 - 80) / 10 = -1.5$.
4. A score of 65 is placed on the X-axis 1.5 standard deviations to the *left* of the middle of the curve.
5. The area under the curve to the left of -1.5 on a standard normal table is equal to the area under the curve to the right of +1.5. The area to the left of +1.5 is .9332. Therefore, the area to the right of this point is $1 - .9332 = .0668$. Therefore, $\Pr(X < 65) = \Pr(Z < -1.5) = \Pr(Z > +1.5) = .0668$.