Probabilities and Probability Distributions

George H Olson, PhD

Doctoral Program in Educational Leadership

Appalachian State University

May 2012

Contents

No course in statistics, especially *intermediate inferential* statistics would be complete without at least a casual stroll down the path of probability theory and probability distributions. After all, the field of statistics is all about computing the probability of events, as in what is the probability that the coin I toss will land on heads? Or, to take a more relevant example, what is the probability of a student getting seven of ten True-False item correct? Or, as yet another example, what is the probability that an observed difference in group means is just a chance occurrence?

In this week's lesson we will take that stroll, beginning with a primer on elementary probability theory followed by an examination of several probability distributions.

Basic Probability Theory

Probability*.* The probability of an event is defined as:

Probability of an event = the frequency of the event / number of possible events, or

$$\Pr\{\textbf{event}\} = \textbf{Freq}(\textbf{event}) / \textbf{N}(\textbf{events}).$$

For instance, suppose that at there are currently 2,214 graduate students enrolled Appalachian, and that, of these, 883 are male. Then the probability of selecting a male graduate student at random is given by

$$\begin{aligned}\textbf{Pr}\{\textbf{male}\} &= 883/2214 \\ &= .399.\end{aligned}$$

While we sometimes multiple the probability by 100 and talk about a nearly 40% chance of selecting a male from the Appstate graduate students, this is not actually strictly accurate. While probabilities, because they similar to proportions, can easily be converted to percentages, probabilities are always decimal values between zero (0) and one (1).

Now, suppose the 2214 graduate students are distributed as shown in the table below. What is the probability of selecting, at random, a male younger than 21? Or, a male younger than 31? Or a male > 30 and a female < 31 (assuming we are trying to arrange some match making)? These probabilities are more complicated. Before we answer these questions I need to introduce some additional ideas about probability.

Fictitious Distribution of Graduate Students at Appstate

Sex	Age Category			TOTAL
	<21	21-31	>31	
Female	210	575	546	1,331
Male	220	372	291	883
TOTAL	430	947	837	2,214

Independent vs. Dependent Events

Events are classified as belonging to one of two categories: independent events or dependent events.

Independent Events. Independent events are those events where the occurrence of one event has no effect on the occurrence of another event. The typical example is throwing a die and flipping a coin. The probability of throwing any face (1, 2, …, 6) on a die is 1/6, or .167. The probability of a head (or a tail) on a flip of a coin is .5. The probability of the outcome of a flip of a coin is *independent* of the outcome of a throw of a die.

Dependent Events. Events are *dependent* when the probability of the occurrence of one event is influenced (conditioned) by the probability of another event. Again, a typical example is drawing a card from a deck of 52 cards. Suppose, in a game of blackjack, the first card dealt is an Ace. What is the probability that the second card drawn is also an Ace?

The probability of an Ace on the first card dealt is 4/52, or .0769. The probability of an Ace on the second card dealt depends upon the fact that the first card dealt was an Ace—there are only three aces left in the deck. Hence, the probability that the second card dealt is an Ace is 3/51, or .0588. This is different from the probability of dealing an Ace second if the first card was anything other than an Ace: 4/51, or .0784.

Basic Rules of Probability

There are four basic rules of probability. These rules always hold—there are no exceptions.

Rule 1: . All probabilities lie in the inclusive interval, 0 – 1. An outcome of an event having probability 0 can never happen; an outcome of an event having probability 1 will always happen.

Rule 2: Compliments. If p = the probability of an outcome then $q = (1 - p)$ is the compliment of that probability. Hence, $p + q = 1$.

Rule 3 Union (also called the additive rule): The probability of the union of the outcomes of two or more, mutually exclusive events is the probability that **one** of the events will occur. This is represented using the union symbol, $\cup$. For example, let A represent one possible outcome, B, another possible outcome, and C, yet another possible outcome, then the probability of A *or* B *or* C is given by:

$$\Pr\{A \cup B \cup C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\}.$$

For instance, the probability of a 2, a 3, or a 4 on a single throw of a die is

$$\Pr\{2 \cup 3 \cup 4\} = \Pr\{2\} + \Pr\{3\} + \Pr\{4\}$$

$$= 1/6 + 1/6 + 1/6$$

$$= 3/6 = .5.$$

It should be obvious, from *Rule 1*, $\Pr\{A \cup B \cup C\} = 1$. Also, form *Rule 2*, if there is a fourth possible outcome, D, then $\Pr\{D\} = 1 - \Pr\{A \cup B \cup C\}$.

When the possible outcomes are NOT mutually exclusive, the probability of two outcomes, A *or* B, is given by

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}.$$

For three events the probability is

$$\Pr\{A \cup B \cup C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} - (\{\Pr\{A \cap B\} + \Pr\{A \cap C\} + \Pr\{B \cap C\} + \Pr\{A \cap B \cap C\}),$$

where the intersection (e.g., $\Pr\{A \cap B\}$) is defined in Rule 4, next.

Rule 4, intersection (also called the multiplicative rule): The probability of the intersection of the outcomes from two or more events is the probability that both (or all) the events will occur. The intersection is represented by the symbol, $\cap$. For example, if A, B, and C are three possible, *independent* outcomes. Then the probability of A *and* B *and* C is given by:

$$\Pr\{A \cap B \cap C\} = \Pr\{A\} \times \Pr\{B\} \times \Pr\{C\}.$$

For instance, the probability of "snake eyes" and a head on the throw of two dice and the flip of a coin is given by the product,

$$(1/6)(1/6)(.5) = .0139.$$

Conditional Probability

The conditional probability of an event is the probability of an event given the occurrence of some other event. For instance, for the distribution of Appstate graduate students, shown in the table, earlier, the probability of selecting, at random, a male graduate student (M) who is less than 21 years old (<21) is

$$\Pr\{\mathbf{MG} \mid (<21)\} = \frac{\Pr\{\mathbf{MG} \cap (<21)\}}{\Pr\{(<21)\}}.$$

(Note the expression, $\Pr\{\mathbf{MG} \mid \mathbf{A}_{<21}\}$, here, is read as, the probability of selecting a male graduate student from those who are less than 21 years old. $\Pr\{\mathbf{MG} \cap (<21)\}$ is the *joint* probability of $\mathbf{MG}$ *and* (<21), or

$$\begin{aligned}\left(\frac{883}{2214}\right)\left(\frac{430}{2214}\right) &= \left(\frac{883}{2214}\right)\left(\frac{430}{2214}\right)\\ &= (.3988)(.1942)\\ &= .0775.\end{aligned}$$

Similarly, the probability of female graduate student 21 years or older is given by (and this is just a little more complicated):

$$\begin{aligned}\Pr\{\mathbf{F} \mid (\geq 21)\} &= \frac{\Pr\{\mathbf{F} \cap (\geq 21)\}}{\Pr\{(\geq)\}}\\ &= \Pr\{\mathbf{F}\} \cap \Pr\{(21-31) \cup (>31)\}\\ &= \Pr\{\mathbf{F}\} \cap (\Pr\{(21-31) + \Pr\{(>31)\})\\ &= \Pr\{\mathbf{F}\} \times \left(\frac{947}{2214} + \frac{837}{2214}\right)\\ &= \left(\frac{1331}{2214}\right) \times \left(\frac{947}{2214} + \frac{837}{2214}\right)\\ &= (.6012) \times (.4277 + .3780)\\ &= .4844.\end{aligned}$$

Hence, we are much more likely to select a female over 30 than we are to select a male under 21.

You do not have to compute these, and other probabilities by hand. A Convenient Probability Calculator is available on the web.

Probability Distributions

In this section, we consider, in more or lesser depth, six probability distributions:

Poisson distribution,
Binomial distribution.
Normal distribution,
Chi square distribution,
t distribution, and
f distribution.

Poisson Probability Distribution

We begin with the Poisson distribution, named for the Paris statistician, Siméon-Denis Poisson (1781-1840). The Poisson probability distribution is particularly useful when we know the rate of occurrence of some event over a defined period of time and would like to know the likelihood of a greater (or lesser) rate of occurrence at some other period of time.

Suppose, over a 12 week period, 31 incidences of bullying were reported to have occurred in the playground. 31 incidences in 12 weeks averages out to 2.58 incidences per week. So, assuming a rate (r) of 2.58 incidences per week how likely are we to experience a week in which 0, 1, 2, 3, 5, … incidences of bullying are reported? Letting Δ (delta) = the number of incidences of bullying per week in which we are interested (e.g., Δ=0,1,2,3, 4, 5…), then using the Poisson probably distribution we compute

$$Pr\{\Delta\} = r^{\Delta} /(\Delta!)(e^{r}),$$

where Δ! Is delta factorial, and e is the constant, 2.178282. For the current example, with r = 2.58, we can compute the probability of Δ=3 as follows:

$$\begin{aligned} Pr\{\Delta=3\} &= 2.58^{3}/(3!)(e^{2.58}) \\ &= 17.174/(6e^{2.58}) \\ &= 17.174/(6\times 13.19714) \\ &= 17.174/79.18283 \\ &= .217. \end{aligned}$$

Hence, we can expect 3 incidences in about 2 out of every 10 weeks. Similarly, we can compute the probability of ., Δ=0,1,2,3, 4, 5, 6…, as shown in the following table. A quick inspection of the table reveals that there is a fairly low, but not zero, probability of a week with no reported bullying incidences. Weeks in which 1, 2, and 3 bullying incidences are reported are expected to occur around 20% of the time 4 incidences about 14% of the time. However, the probability of weeks in which 5 or more incidences of bullying are expected to occur fall off rapidly.

Predicted Probability of a Given Number Bullying Incidences in Any Given Week	
No. of Incidences	Probability
0	.076
1	.195
2	.252
3	.217
4	.140
5	.072
6	.031
7	.011
8	.004
9	.001
10	<.001

In the table you should note the following characteristics, which are characteristics of a Poisson probability distribution: The highest probability occurs close to the rate, and the probability distribution is skewed to the right (i.e., the probabilities bunch up around the rate of incidences and diminish rapidly as the predicted number of incidences increase).

Fortunately, you do not actually have to compute the Poisson probability distribution. There is an online Poisson Calculator available on the web.

Binomial Probability Distribution

Bernoulli trials. Before turning to the binomial distribution, we first need to introduce a new term, Bernoulli trials (or Bernoulli experiments). Basically, a Bernoulli trial is one of a series of random processes (or *experiments*) each of which consists of independent events with two possible, mutually exclusive outcomes with known probabilities of occurrence. Tossing coins, either several at one time, or several tosses in sequence, is an example of such a process. Each coin tossed has a .5 probably of landing on a head (or a tail).

The binomial distribution. The binomial distribution, sometimes inaccurately called the *Bernoulli* distribution, gives the probabilities of a number of successful outcomes in a series of Bernoulli trials. For instance, suppose you were playing red at a roulette wheel in a Las Vegas casino and had enough money to bet the same amount, $20, say, to play the wheel 10 times. Obviously, you would like to come out ahead. To come out even, you would need to win more than 5 times (actually six times, since a US wheel has two green slots, for 0 and double 0). Including the green slots, there are 38 possible outcomes of which 18 are red. Therefore, the probability of landing on red in a single spin is 18/38, or .4737. You want to know the probability of landing on red at least six times in 10 spins. This problem is solved using the binomial distribution.

The conditions that must be satisfied when using the binomial distribution to compute the probability of a random event, X, are that:

I. there is a fixed number of events (trials), *n*,
II. each event is independent of all other events,
III. the outcome of each event represents one of only two, possible, mutually exclusive outcomes, and
IV. The probability of the event, *p*, is the same on each outcome.

For the example just given, *n* is 10, *p* is .4737, the events (trials, or spins) are independent, and outcomes (red or not red) are mutually exclusive.

When these conditions are met, the binomial probability of *r* events occurring in *n* trials, where the individual probability of an event is *p* is given by,

$$\begin{aligned} P_r &= {}_n\mathbf{C}_r \times p^r \times (1-p)^{(n-r)} \\ &= \frac{n!}{r!(n-r)!} \times p^r \times (1-p)^{(n-r)} \\ &= \frac{n!}{r!(n-r)!} \times p^r q^{(n-r)}. \end{aligned}$$

Note: the notation, ${}_n\mathbf{C}_r$, is read as the "number of combinations of *n* things taken *r* at a time. For us, this is the combination of 6 reds in 10 trials. Hence, we have,

$$\begin{aligned} P_r &= {}_{10}\mathbf{C}_6 \times .4737^6 \times (1-.4737)^{(4)} \\ &= \frac{10!}{6!(10-6)!} \times .4737^6 \times (1-.4737)^{(4)} \\ &= \frac{10!}{6!(4)!} \times .4737^6 .5263^4 \\ &= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)(4 \cdot 3 \cdot 2 \cdot 1)} \times .4737^6 .5263^4 \\ &= \frac{10 \cdot 9 \cdot 8 \cdot 7}{(4 \cdot 3 \cdot 2 \cdot 1)} \times .4737^6 .5263^4 \\ &= \frac{5040}{24} \times .4737^6 .5263^4 \\ &= 210 \times .01130 \times .07672 \\ &= .1821. \end{aligned}$$

So, the probability of landing on red *exactly* six times in 10 spins is less than .2 (there is less than a 20% chance of landing on red *exactly* six times in 10 spins). But, landing on red six *or more* times is fine with us. What we want, then, is the probability of landing on red six or *more* times. To obtain this probability we need compute P_r for *r* = 7, 8, 9, and 10, and sum all of these along with the one we just computed. Fortunately, a convenient Binomial Calculator, provided by STAT TEC is available. Using the calculator, we find that the probability of landing on red six or more times in 10 spins is only .314—less than 30% of the time. We are better off saving our money for dinner and a show.

Some additional examples may help.

Example*.* If the probability of graduating from a particular college in four years is 0.4, then if 10 students are selected at random from that college, what is the probability that at least 3 of them will not graduate in 4 years? What we are given, then, is:

Number of trials, *n*, is 10,
Probability, *p*, of Graduating in 4 years is *p* = 0.4, and the
Probability, *q*, of NOT Graduating in 4 years is (1-*p*) or .6,

We are interested in computing the probability that at least 3 of the 10 students selected at random fail to graduate on time. This is equal to 1 minus the probability that, at *most*, 7 do graduate in four years. We can approach this problem in one of two ways. We can either calculate Pr{3 or more do NOT graduate} or 1-Pr{7 or more DO graduate}. Using the binomial calculator, and counting not graduating as a "success," we obtain

Pr{Not Graduating ≥ 3} = .988.

Calculating the other way, using the Binomial Calculator, we obtain

Pr{Graduating ≥ 7} = .012, and then compute

Pr{Not Graduating ≥ 3} = 1-.012 = .988.

So, either way, the probability of three of the ten students not graduating in four years is pretty high.

Another Example. Here is a more interesting example, originally offered by Ben P. Stein (2003). Stein wondered: assuming that the two teams meeting in the World Series are evenly matched, what are the chances that the World Series will go the full 7 games?" In arriving at his answer, Stein calculated the probabilities of a World Series ending at 4 games or going on to 5, 6, and 7 games. (Stein's article is fun to read, for baseball fans and statisticians alike.) How did Stein calculate these probabilities?

Begin with the (unlikely) assumption that the two teams are equal, i.e., evenly matched. Define a successful event as a win. For four games (trials) the binominal probability is given by:

$$\begin{aligned} P_4 &= {}_4C_4 \times .5^4 \times .5^4 \\ &= \frac{4!}{4!(4-4)!} \times .5^4 \times .5^0 \\ &= \frac{4\times3\times2\times1}{(4\times3\times2\times1)\times0!} \times .625\times1 \\ &= .0625. \end{aligned}$$

Calculating the probability that the series ends after 5 games is a little more difficult. The probability is actually .25; but, how do we arrive at this probability? Well, we know that the only way the series can end after exactly 5 games is if one of the teams has won 3 of the first 4 games (if both teams have won two games, the series will go to at least 6 games).

The probability that one of the teams (Team A or Team B, say) wins exactly 3 of the first 4 games is

$$\begin{aligned} P_3 &= {}_4C_3 \times .5^3 \times .5^{(4-3)} \\ &= \frac{4!}{3!(4-3)!} \times .5^3 \times .5^1 \\ &= \frac{24}{6} \times .125 \times .5 \\ &= 4 \times .0625 \\ &= .25. \end{aligned}$$

Now, either team has an equal probability (.5) of winning the 5th game. The probability of Team A winning the fifth game is .25×.50=.125; the probability of Team B winning the fifth game is also .125. So, the probability of either Team A or Team B winning the series in 5 games is .125 + .125 = .25, the probability that the series ends in exactly 5 games.

For the series to end in exactly 6 games, one of the teams would have had to have won three of the previous 5 games. The probability of winning three of the previous five games is given by

$$\begin{aligned} P_3 &= {}_5C_3 \times .5^3 \times .5^{(5-3)} \\ &= \frac{5!}{3!(5-3)!} \times .5^3 \times .5^2 \\ &= 10 \times .125 \times .5 \\ &= .3125. \end{aligned}$$

Either team has a .3125 probability of winning three of the first five games and a .5 probability of winning the sixth game. Hence, the probability of Team A winning the series in six games is .3125 × .5 = .15625. Since both teams have the same probability of winning in 6 games the probability Team A **or** Team B winning in six games is .15625+.15625=.3125.

To complete the solution, compute the probability of three wins in the previous 6 games (.3125) and multiply this by the probability of a win in the 7th game (.5), yielding a .15625 probability one of the teams will win the series in seven games. But, since either team could win the series, the probability that one or the other team will win the series in the 7th game is .15625 + .15625, or .3125.

So the probabilities of winning the series in exactly 4, 5, 6, or 7 games, rounded to two decimal places, is:

Number of Games to Win	Probability
4	.06
5	.25
6	.31
7	.31

In his article, Stein points out that in 50 years' worth of World Series (1952-2002) the probabilities were actually quite different:

Probabilities of World Series Wins Over the Years, 1952 to 2002	
Number of Games to Win	**Probability**
4	.16
5	.16
6	.20
7	.48

For an explanation, see Stein's article.

Yet Another Example. We can use the binomial distribution to calculate number of items we need for a true-false test to yield meaningful results. By "meaningful results," here, I mean results that can tell us, reliably, whether an examinee is proficient (acquired knowledge in the subject area tested).

Suppose we want to administer a True-False test, but want a score of 70 percent items correct to be a reliable score. By reliable, here, I mean a score that is statistically, significantly greater than a chance score of 50 percent items correct. To be statistically significant, the score needs to be one that would occur by chance alone 5% or less of the time. Stated in probability terms, we want, $\Pr\{\mathbf{X}\} \leq .05$, or $\Pr\{70\% \text{ of K }\} \leq .05$, where K is the number of items in the test. In the following table, the probabilities of X equal to, or greater than, 70%, are shown for True/False tests of various lengths. You can see immediately that, for a 10-item test, to have a reliable score, X would have to be at least 9. But, a score of 9 out of 10 correct does not leave much room for differentiation. For a 15-item T/F test, a score of 11 is the lowest reliable score, but again, the differentiation occurs only for scores of 11 through 15, or 73%, 80%, 87% and 100%. A T/F score of 12 on a 16-item T/F gives, what would normally be considered a sufficient range of reliable scores.

A 4-choice, multiple-choice test would require considerably fewer items to yield a sufficient range of reliable scores. As shown in the next table, an 8-item MC test can yield a sufficient range of reliable scores (62% to 100%). And, with a 10-item MC test, a good range of reliable scores begins with a score of 6 items correct.

True-False Test

Number of Items (trials)	**Score (number correct) = X**	**Probability of X**	**Probability of X or Greater**
10	7 (=70%)	.111	.172
10	8 (=80%)	.044	.055
10	9 (=90%)	.001	.011
15	10 (=67%)	.092	.151
15	11 (=73%)	.042	.060
15	12 (=80%)	.014	.018
15	13 (=80%)	.003	.004
15	14 (=87%)	.000	.000
16	11 (=69%)	.067	.105
16	12 (=75%)	.028	.038
16	13 (=81%)	.009	.011
16	14 (=88%)	.002	.022
16	15 (=94%)	.000	.000

4-choice Multiple Choice Items

Number of Item (trials)	Score (number correct) = X	Probability of X	Probability of X or Greater
5	3 (=60%)	.088	.104
5	4 (=80%)	.015	.016
6	3 (=50%)	.132	.170
6	4 (=67%)	.033	.038
6	5 (=83%)	.004	.005
7	4 (=57%)	.058	.071
7	5 (=71%)	.011	.013
7	6 (=86%)	.001	.001
8	4 (=50%)	.086	.114
8	5 (=62%)	.023	.027

8	6 (=75%)	.003	.004
8	7 (=88%)	<.001	<.001
9	4 (=78%)	.117	.166
9	5 (=56%)	.039	.049
9	6 (=67%)	.008	.010
9	7 (=78%)	.001	.001
9	8 (=89%)	<.001	<.001
10	5 (=50%)	.058	.078
10	6 (=%)	.016	.020
10	7 (=%)	.003	.004
10	8 (=%)	<.001	<.001
10	9 (=%)	<.001	<.001

Mean and Variance of the Binomial Distribution

For an individual trial, where X=1 for a success and X=0 or a failure, the mean of X is equal to $\frac{\Sigma X}{n} = p$, and the variance is equal to *pq*. For a distribution with *n* trials the

Mean = *p*, and

Variance = *npq*.

Sample Proportions

If we know that the count X of "successes" in a group of n observations with success probability p has a binomial distribution with mean np and variance np(1-p), then we are able to derive information about the distribution of the sample proportion (the count of successes X divided by the number of observations n). By the multiplicative properties of the mean, the mean of the distribution of X/n is equal to the mean of X divided by n, or np/n = p. This proves that the sample proportion is an unbiased estimator of the population proportion p. The variance of X/n is equal to the variance of X divided by n^2, or $(np(1-p))/n^2 = (p(1-p))/n$. This formula indicates that as the size of the sample increases, the variance decreases.

Chi-Square and *t* Distributions.

Chi-square (χ^2) distribution. Point your browser to the Stat Trek tutorial on the Chi-square distribution. Read through the lesson and make sure you understand the problems (and the use of the Chi-Square Distribution Calculator) given at the end of the lesson.

***t* distribution**. Stat Trek also has an excellent tutorial on the *t* distribution. Go there and read the whole lesson. While, working through the lesson, follow the link to the central limit theorem. Be sure you understand the two problems given at the end of the tutorial.

Reference

Stein, B. P. (17 Oct, 2003). Are 7-Game World Series More Common Than Expected? *Inside Science News Service*. Downloaded from http://www.insidescience.org/current-affairs/are-7-game-world-series-more-common-than-expected on 6 May 2012.