Introduction to Probability

Probabilities are expressed in terms of ***events***

Examples of Events:

1. Six shows on a roll of a die
2. Jack of Spades is drawn from a deck of cards
3. Rains at 3:00 p.m. today in front of the Reitz Union
4. Have an automobile accident this year
5. Florida beat Florida State
6. Florida beat Tennessee
7. Florida beat Florida State and Tennessee
8. Yield of a randomly drawn citrus tree is greater than 8 boxes
9. Mean yield of 25 randomly drawn citrus trees is greater than 8 boxes

Events are denoted by capital letters, A, B, etc.

Probabilities of events are denoted P(A), or P(event occurs)

Terminology:

Terms in probability relate to terms in mathematical set theory:

Set theory	Probability
Universe	Universe
Subset of Universe	Event
Element in Universe	Outcome

Think of an outcome as the result of an experiment. The "universe" is the set of possible outcomes.
Example: Experiment="roll a die" → Universe={1,2,3,4,5,6}
Event A: "outcome" is even = {2,4,6} P(A)=3/6=.5

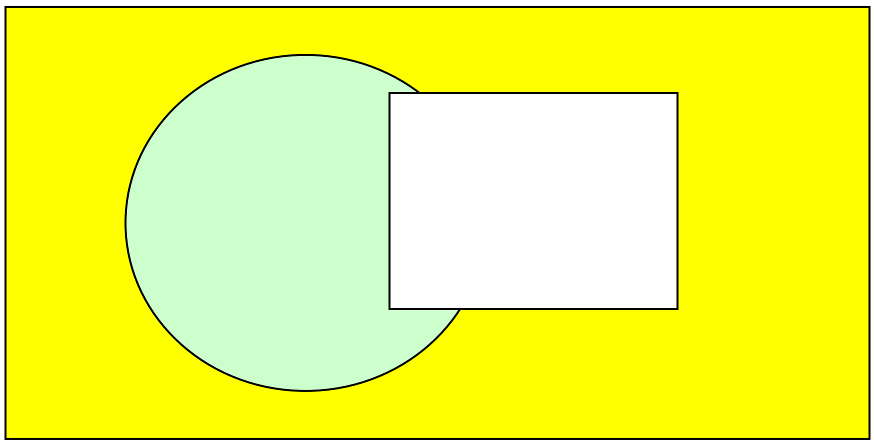
More on Terminology of Probability

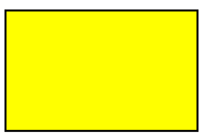
Think about the events "Florida beat FSU" and "Florida beat Tennessee." These are events pertaining to winning games in the football season. Let's try to determine a set of outcomes that would allow us to describe all possible such events. There may be several ways to do this. Perhaps the most basic is a list of all possible win-loss records for the regular season of 12 games. There are 2^{12}=4096 of these outcomes—that's the number of "elements" in the universe set. We could list them as follows:

Outcomes

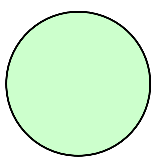
Game #									
1	W	W	W	W	…	W	…	W	L
2	W	W	W	W	…	W	…	L	L
3	W	W	W	W	…	W	…	L	L
4	W	W	W	W	…	W	…	L	L
5	W	W	W	W	…	L	…	L	L
6	W	W	W	W	…	W	…	L	L
7	W	W	W	W	…	L	…	L	L
8	W	W	W	W	…	W	…	L	L
9	W	W	W	W	…	W	…	L	L
10	W	W	W	L	…	W	…	L	L
11	W	W	L	W	…	W	…	L	L
12	W	L	W	W	…	W	…	L	L
# Wins	12	11	11	11	…	10	…	1	0
# Losses	0	1	1	1	…	2	…	11	12

Venn Diagrams



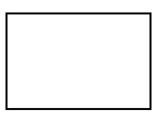


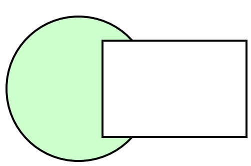
Universe



Event A

Event B





Event AUB

Determination of Probabilities

Probabilities are determined from:

1. Relative frequency computation
2. Subjective assessment

P(six on roll of die)=1/6, a relative frequency computation

P(Florida beat Florida State)=.8, a subjective assessment

P(Rain today at 3:00 at JWRU)=.4, combination of relative frequency and subjective assessment

Compound Events and Probabilities of Compound Events

Compound events are formed from combinations of events.

The *union* of events A and B occurs if *either* A or B occur

{Florida beat FSU}$\cup${Florida beat Tennessee}={Florida beat *either* FSU or Tennessee}

The *intersection* of events A and B occurs if *both* A and B occur

{Florida beat FSU}$\cap${Florida beat Tennessee}={Florida beat *both* FSU and Tennessee}

Calculating Probabilities of Compound Events

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

P{Florida beat *either* FSU or Tennessee}=P{Florida beat FSU}+P{Florida beat Tennessee}-P{Florida beat both FSU and Tennessee}

Two events, A and B, are *disjoint* if they are mutually exclusive; i.e., if $A \cap B = \Phi$, where $\Phi = \{\ \}$ is the "empty" set.

If A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$

Two events, A and B, are *independent* if $P(A \cap B) = P(A)P(B)$

Are the events {Florida beat FSU} and {Florida beat Tennessee} independent?

Are the events {Drives 4WD truck} and {Voted Republican} independent?

Are the events {Drives Prius} and {Voted Republican} independent?

Are the events {Lives in Florida} and {Voted Republican} independent?

Conditional Probability

The conditional probability of A *given B* **is the probability that A occurs, when it is known that B occurs.** It is calculated by the formula

$$P(A|B)=P(A\cap B)/P(B).$$

If A and B are independent, then P(A)=P(A|B)

Examples

P(J) = P(draw J from deck of cards) = $\frac{4}{52}=\frac{1}{13}$

P(C) = P(draw club from deck of cards) = $\frac{13}{52}=\frac{1}{4}$

Intersection: $P(J\cap C)=P(J \text{ and } C) = \frac{1}{52}$

Union:

$$P(J\cup C) = P(J \text{ or } C) = P(J) + P(C) - P(J\cap C)$$

$$=\frac{1}{13}+\frac{1}{4}-\frac{1}{52}=\frac{4+13-1}{52}=\frac{16}{52}$$

$P(J\cap Q) = 0$ $\quad$ $P(J\cup Q) = \frac{8}{52}=\frac{2}{13}$

Dependence:

$P(J\cap C)=P(J)P(C)$ "J" and "C" are independent

$P(J\cap Q) \neq P(J)P(Q)$ "J" and "Q" are not independent

Application

$$P(PC+) = \frac{50}{10000} = \frac{1}{200} = .005$$

$$P(PSA+) = \frac{398}{10000} = .0398$$

$$P(PC+ \cap PSA+) = \frac{48}{10000} = .0048$$

		Prostate Cancer		
		+	-	
PSA Test	+	48	350	398
	-	2	9600	9602
		50	9950	10000

Conditional Probability: P(A given B) = $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Sensitivity = P(PSA+ |PC+) = (48/10000)/(50/10000) = .96

Specificity = P(PSA- | PC-) = (9600/10000)/(9950/10000) = .9648

Predictive value = P(PC+ |PSA+) = P(PC+∩PSA+)/P(PSA+) = P(PSA+|PC+)P(PC+)/P(PSA+) = (.96x.005)/.0398 = .12

This is an example of the application of "Bayes' Rule." It allowed us to calculate the conditional probability P(PC+ |PSA+) using the other one, P(PSA+ |PC+).

More generally, suppose $A_1, A_2, \ldots A_n$ is a partitioning of the universe, i.e. U is the union of $A_1, A_2, \ldots A_n$, and the intersection of any two of the A_i sets is empty. You want to calculate $P(A_k|B)$, where know $P(B|A_1), \ldots, P(B|A_n)$, and you also know $P(A_1), \ldots, P(A_n)$. Bayes Rule says that

$$P(A_k|B) = P(B|A_k)P(A_k) / [P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)]$$

In the prostate cancer application, $A_1, A_2, \ldots A_n$ could represent the various stages of prostate cancer, and B is the event of a positive diagnosis. From epidemiological records you might know $P(B|A_i)$, the probability of positive diagnosis for each stage, and $P(A_i)$, the probabilities of various stages.

Random Variables:

Random variables are functions that describe the outcomes of an "experiment." They are usually denoted with a capital letter, such as *X* or *Y*. RVs can be either discrete of continuous. A discrete random variable has a "countable" number of values, meaning that they could be listed. A continuous random variable has an uncountable number of numeric values, meaning they cannot be listed. Here are some examples:

Discrete

- # spots on top face of die (1, 2, 3, 4, 5, 6)
- suit of drawn card (C, D, H, S)
- # aphids on leaf (0, 1, 2, 3, …)
- # defects in box of 1000 nails (0, 1, 2, …, 1000)
- # germinating seeds out of 50 (0, 1, 2, …, 50)

Continuous

- heights of people (0 -- ?)
- ph of soil (0 – 10)
- voltage in circuit (0 -- ?)

Discrete random variables have *probability (mass) function*s which we denote *p(x)*. The probability function gives probabilities of each individual value of the RV, e.g.

$$P(X = x) = p(x).$$

Continuous random variables have *probability density functions* which we denote *f(x)*. The probability density function can be used to give probabilities of ranges of values of the continuous RV. For example,

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$

There are two random variables that are especially important in statistics, the binomial and normal.

Binomial Random Variable

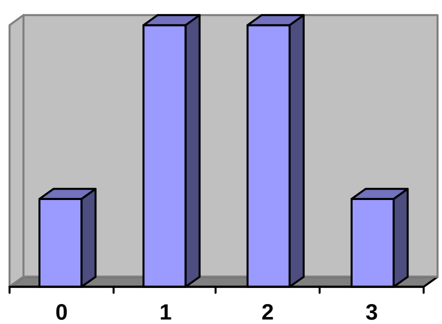
Let X = Number of successes out of n trials, in which the probability of success on each trial is a number π. Then X has a *binomial* distribution with parameters n and π. This is abbreviated $X \sim B(n, \pi)$.

Example: Consider flipping a coin, and declare a "success" if a head (H) appears. Then the probability of success is .5. That is, π=P(S)=.5. Suppose the coin is flipped n=3 times, and X = number of heads. Then $x \sim B(3, .5)$. The possible values of x are (0, 1, 2, 3). The probability of any event is $(1/2)^3 = 1/8$.

Events:	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
# Heads	3	2	2	2	1	1	1	0
P(Event)	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

P(3 H) = 1/8 P(2 H) = 3/8 P(1 H) = 3/8 P(0 H) = 1/8

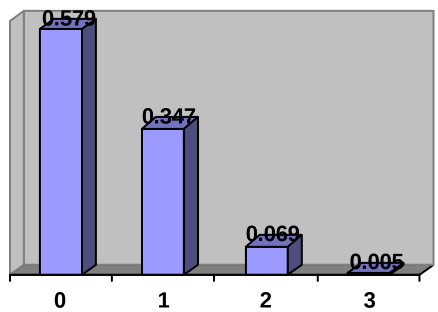
In general: $P(k\ H) = \frac{3!}{k!(3-k)!} \cdot \left(\frac{1}{2}\right)^3 = \frac{3!}{k!(3-k)!} \cdot \left(\frac{1}{8}\right)$



Binomial Random Variable (con't)

Example: X = number of 1's in 3 rolls of die (0, 1, 2, 3)

Events:	111	11X	1X1	X11	1XX	X1X	XX1	XXX
# 1's	3	2	2	2	1	1	1	0
P(Event)	$1^3/6^3$	$1^2 5^1/6^3$	$1^2 5^1/6^3$	$1^2 5^1/6^3$	$1^1 \cdot 5^2/6^3$	$1^1 \cdot 5^2/6^3$	$1^1 \cdot 5^2/6^3$	$5^3/6$



P(3 1's) =

$$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{6^3} = \frac{1}{216} = .0046$$

P(k 1's) = $\dfrac{3!}{k!(3-k)!} \left(\dfrac{1}{6}\right)^k \left(\dfrac{5}{6}\right)^{3-k}$

The probability mass function is given by the so-called Binomial Formula:

π = probability of success on single trial

P(x successes in n trials) = $\dfrac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x}$

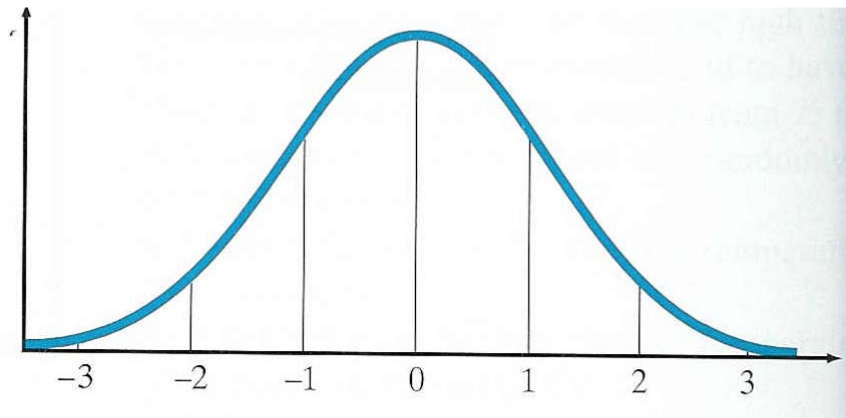
Mean of the binomial distribution: $n\pi$

Variance of the binomial distribution: $n\pi(1-\pi)$

Normal Distribution and normal random variable:

The notation $Y \sim N(\mu,\sigma^2)$ means "The random variable Y is distributed normally with mean μ and variance σ^2.

The *standard normal* distribution has mean μ=0 and variance σ^2=1. The letter Z is reserved to represent the standard normal random variable.



Computer programs and tables are available to obtain probabilities from the normal distribution. For example, you can discover that

- $P(-1 < Z < 1) = .68$
- $P(Z > 1) = .16$
- $P(-1.96 < Z < 1.96) = .95$
- $P(Z > 1.96) = .025$
- $P(Z > 1.42) = .0778$.

The probability density function of the normal random variable with mean μ and variance σ^2 is given by the formula

$$f(y) = (2\pi\sigma^2)^{-1/2} \exp(-(y-\mu)^2 / 2\sigma^2)$$

Using the Normal Distribution

Standardizing a Normal Distribution:

If $Y \sim N(\mu,\sigma^2)$, then $Z = (Y - \mu)/\sigma \sim N(0,1)$.

This result allows us to compute probabilities from *any* normal distribution using tables or a computer program for the standard normal distribution.

If you wanted to calculate the probability that a random variable *y* is greater that 1.42 standard deviations above its mean, you would compute:

$$P(Y > \mu + 1.42\sigma) = P(\frac{Y-\mu}{\sigma} > 1.42) = P(Z > 1.42) = .0778$$

As a more specific application, suppose you believe the egg weights to be normally distributed with mean 65.4 and standard deviation 5.17. You would calculate the probability that a randomly drawn egg is greater than 72 as:

$$P(Y > 72) = P((Y - 65.4)/5.2) > (72 - 65.4)/5.2) = P(Z > 1.27) = .1$$

Using the normal distribution—an application

Egg weights are normally distributed with mean $\mu = 65$ (g) and standard deviation $\sigma = 5.0$.

1. What is the probability one randomly drawn egg will exceed:
   a. 65 b. 66 c. 70 d. 75

   Let Y = egg weight. Then

   a. $P(Y > 65) = P\left(\frac{Y-65}{5} > \frac{65-65}{5}\right) = P(Z > 0) = ½ = .5$

   b. $P(Y > 66) = P\left(\frac{Y-65}{5} > \frac{66-65}{5}\right) = P(Z > .2) = .4207$

   c. $P(Y > 70) = P\left(\frac{Y-65}{5} > \frac{70-65}{5}\right) = P(Z > 1) = .1587$

   d. $P(Y > 75) = P\left(\frac{Y-65}{5} > \frac{75-65}{5}\right) = P(Z > 2) = .0228$

2. What is the probability one egg is between 66 and 70 g?
   $P(66 < Y < 70) = P(Y > 66) - P(Y > 70)$
   $= .4207 - .1587 = .262$

Probabilities of this type can be expressed in terms of the *cumulative distribution function*

$$P(Z < z) = F(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} \exp(-z^2/2)$$

The integral for the normal distribution is difficult to evaluate, so tables or computer programs are used to obtain actual values.

Normal Approximation to the Binomial

You can use the normal distribution to approximate binomial probabilities. This often simplifies a computation. For example, suppose you are shooting free-throws in basketball. You know that you make 75% of your shots; that is, the probability of making any one shot is .75. You have entered a contest that awards a prize if you make at least 18 out of 20 shots. What is the probability that you will win a prize?

You need to calculate P($Y \geq 18$), where y is the number of shots you make out of 20. The exact probability is given by the binomial formula with $\pi = .75$ and n = 20:

$$P(Y \geq 18) = P(Y = 18) + P(Y = 19) + P(Y = 20)$$

$$= 20!/(18!2!).75^{18}.25^{2}$$

$$+20!/(19!1!).75^{19}.25^{1}$$

$$+20!/(20!0!).75^{20}.25^{0}$$

$$= .0069 + .0211 + .0032$$

$$= .0912$$

Normal Approximation to the Binomial

The calculation on the previous page would be tedious by hand, but many computer programs are available that can readily do it. However, even good computer programs may fail for computations involving extremely large n.

The normal approximation sets $\mu = n\pi$ and $\sigma^2 = n\pi(1-\pi)$, and assumes $Y \sim N(\mu,\sigma^2)$ to evaluate the probability. The approximation is improved by using a *continuity correction*, which means you compute $P(Y \geq 18-.5) = P(Y \geq 17.5)$.

The normal approximation is then computed as:

$\mu = n\pi = 20(.75) = 15$
$\sigma^2 = n\pi(1-\pi) = 20(.75)(.25) = 3.75$
$\sigma = 3.75^{1/2} = 1.94$

$$P(Y \geq 17.5) = P((y - 15)/1.94 \geq (17.5 - 15)/1.94)$$

$$= P(Z \geq 1.29) = 1 - .901 = .099.$$

This is a reasonable approximation to the exact binomial probability of $P(Y \geq 18) = .091$.

Means and Variances of Random Variables

Means of random variables are called *expected values*, denoted

$$\mu = E(X)$$

If X is a continuous RV, then

$$\mu_X = E(X) = \int xf(x)dx$$

If X is a discrete RV, then

$$\mu_X = E(X) = \sum_i x_i p(x_i)$$

Variances of random variables are also expected values.

If X is a continuous RV, then

$$\sigma_X^2 = E((X-\mu)^2) = \int (x-\mu)^2 f(x)dx$$

If X is a discrete RV, then

$$\sigma_X^2 = E((X-\mu)^2) = \sum_i (x_i-\mu)^2 p(x_i)$$

Means and Variances of Linear Functions of Random Variables

If X is an RV with mean μ_X and variance σ_X^2, and $Y=a+bX$, where a and b are constants, then:

$$\mu_Y = E(a+bX) = a + bE(X) = a + b\mu_X$$

and

$$\sigma_Y^2 = E((Y-\mu_Y)^2) = E(((a+bX)-(a+b\mu_X))^2)$$
$$= E(b^2(X-\mu_X)^2) = b^2\sigma_X^2$$

If $X_1, \ldots, X_k$ are RVs and $Y = X_1 + \ldots + X_k$ then

$$\mu_Y = E(X_1 + \ldots + X_k) = E(X_1) + \ldots + E(X_k) = \mu_1 + \ldots + \mu_k$$

If $X_1, \ldots, X_k$ are independent RVs, and $Y = X_1 + \ldots + X_k$ then

$$\sigma_Y^2 = E((X_1 + \ldots + X_k) - (\mu_1 + \ldots + \mu_k))^2$$
$$= E(X_1 - \mu_1)^2 + \ldots + E(X_k - \mu_k)^2 = \sigma_1^2 + \ldots + \sigma_k^2$$