# Standard & Conditional Probability

ORIGIN := 0

## Probability as a Concept:

**"The probability of an event is the likelihood of that event expressed either by the relative frequency observed from a large number of data or by knowledge of the system under study" Zar 2010 p 60.**

**Statistics is typically based on a pair of quantities:**

**X <- observed sample values**
**P(X) <- probability of the sampled values under some model of probability.**

**In fact, associating these two quantities is not at all straightforward and is often a point of controversy as both a theoretical and practical matter.**

## Probability by knowledge of a system:

**Some systems are simple enough that concrete probabilities can be deduced directly by counting all possible outcomes. For instance, in flipping a "fair" coin there are only two possible outcomes with equal probabilities of occurring. Or, in rolling a single die, there are only six possible outcomes also with equal probabilities. Mendelian inheritance is another example in which probabilities of specific genotypes can be calculated using well-known ratios such as 9:3:3:1 for two dominant/recessive alleles on different chromosomes.**

**Counting all possible outcomes is greatly facilitated by considering *Permutations* and *Combinations.***

## Permutations:

**"A permutation is an arrangement of objects in a specific sequence." Zar 2010 p. 51**

**"The number of permutations of n things taken k at a time ... represents the number of ways of selecting k items out of n where the order of selection is important." Rosner 2006 Definition 4.8, p. 91.**

$n := 3$ **< n things...**

$k := 2$ **< taken k at a time**

$$nP_k := \frac{n!}{(n-k)!} \qquad nP_k = 6$$ **< Zar Eq. 5.6**

**meaning of factorials (!)**

$n! = 6 \qquad 3 \cdot 2 \cdot 1 = 6$

$k! = 2 \qquad 2 \cdot 1 = 2$

$(n-k)! = 1$

**^ number of permutations of n things taken k at at time**

**For example, let the n things be the letters: A, B, C. How many pairs of letters can we make where the order of letters is important?**

AB AC BC

BA CA CB

$$\frac{3!}{(3-2)!} = 6$$ **< fortunately the number n is relatively small, six!**

**What happens if:**

$n := 20$ **< n things...**

$k := 7$ **< taken k at a time**

$$nP_k := \frac{n!}{(n-k)!} \qquad nP_k = 3.907 \times 10^8$$

$n! = 2.433 \times 10^{18}$

$k! = 5.04 \times 10^3$

$(n-k)! = 6.227 \times 10^9$

**< Fortunately we have this formula, because listing all of the possilities and counting them up would take *a lot* of time...**

**^ number of Permutations of n things take k at at time**

## Degenerate Permutations:

**Zar 2010 gives additional formulas for permutations in which not all sequences are considered unique. These formulas are not of general importance, although for particular problems you may find them useful.**

**Circular Arrangements:**

$CnP_k := \dfrac{n!}{(n-k)! \cdot k}$ $\quad CnP_k = 5.581 \times 10^7$ $\quad (n-k)! \cdot k! = 3.138 \times 10^{13}$

**Indistinguishable objects:**

$i := 0..2$

$N := 20$ **< number of objects in three classes:** $n := \begin{pmatrix} 15 \\ 3 \\ 2 \end{pmatrix}$

**< 15 indistinguishable in class 1**

**< 3 indistinguishable in class 2**

**< 2 indistinguishtable in class 3**

$DnP_k := \dfrac{N!}{\prod_i n_i!}$ $\quad DnP_k = 1.55 \times 10^5$ $\quad \prod_i n_i! = 1.569 \times 10^{13}$ $\quad n_0! \cdot n_1! \cdot n_2! = 1.569 \times 10^{13}$

**^ Note in all degenerate cases, the denominator is larger than the non-degenerate case. Thus the number of permutations will always be smaller.**

## Combinations:

**"If groups... are important to us, but not the sequence of objects within groups, then we are speaking of combinations" Zar 2010 p. 55**

**"The number of combinations of n things taken k at a time ... represents the number of ways of selecting k objects out of n *where the order of selection does not matter.*" Rosner 2006 Definition 4.11, p. 93.**

**Combinations are degenerate permutations in the same sense as above when membership is important but order is not.**

**For example with same n & k as the first one above:**

$n := 3$ **< n things...**

$k := 2$ **< taken k at a time**

$nC_k := \dfrac{n!}{k! \cdot (n-k)!}$ $\quad nC_k = 3$ $\quad n! = 6$, $k! = 2$, $(n-k)! = 1$

**< Note that this is half the number of Permutations with n=3, k=2.**

**^ number of Combination of n things take k at at time**

**What about the larger example above?**

$n := 20$ **< n things...**

$k := 7$ **< taken k at a time**

$nC_k := \dfrac{n!}{k! \cdot (n-k)!}$ $\quad nC_k = 7.752 \times 10^4$

**^ number of Combination of n things take k at at time**

$\dfrac{nP_k}{nC_k} = 5040$ **< a somewhat larger difference between Permutations and Combinations here!**

## Built-in Functions:

**Most software packages contain built-in functions for Permutations and Combinations:**

$n := 20$

$k := 7$

$\text{permut}(n,k) = 3.907 \times 10^{8}$ $\quad nP_k = 3.907 \times 10^{8}$ **< and match our calculations above**

$\text{combin}(n,k) = 7.752 \times 10^{4}$ $\quad (nC_k) = 7.752 \times 10^{4}$ **so serve as MathCad prototype...**

## Prototype in R:

```
# PERMUTATIONS AND COMBINATIONS
n=20
k=7
#PERMUTATIONS
# No built-in function found, so I'll calculate it directly
Permutations=factorial(n)/(factorial(n-k))
Permutations

# COMBINATIONS - calculated directly
Combinations=factorial(n)/(factorial(k)*factorial(n-k))
Combinations

# BUILT-IN FUNCTION FOR COMBINATIONS
choose(n,k)
```

## Important symmetry in calculation of Combinations:

$\text{combin}(n,k) = 7.752 \times 10^{4}$ $\quad \text{combin}[n,(n-k)] = 7.752 \times 10^{4}$

$\text{permut}(n,k) = 3.907 \times 10^{8}$ $\quad \text{permut}[n,(n-k)] = 4.827 \times 10^{14}$

**< k or (n-k) give the same result for combination but NOT permutation.**

## Probability by reference to large samples:

**There are two important perspectives:**

- **Frequentist (or Standard) Statistical Methods - mostly what we will do in this course.**
- **Bayesian Inference - increasingly prominent in several biological & biomedical fields.**

## Frequentist Method:

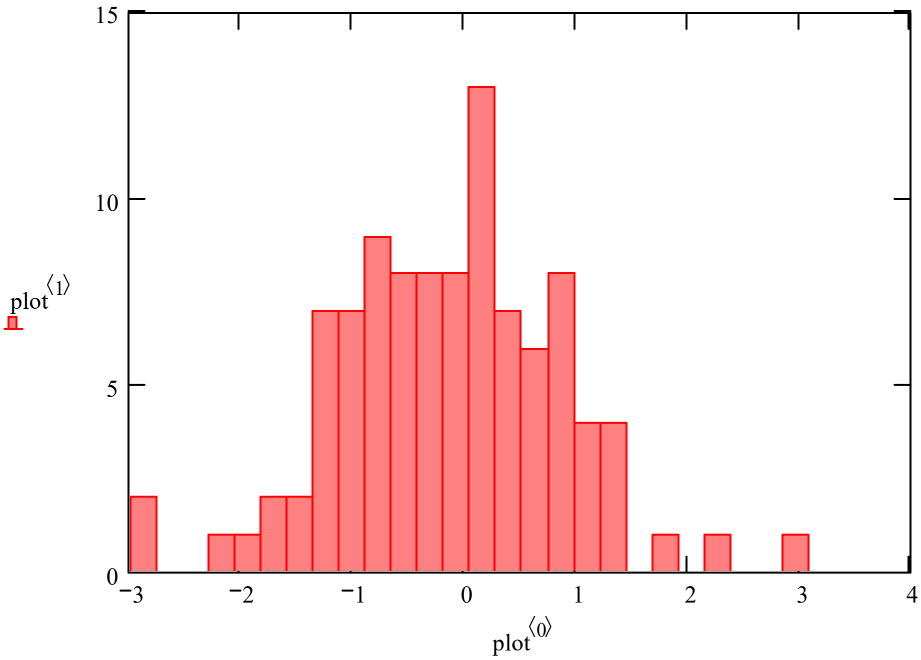
**"The probability of an event is the relative frequency of a set of outcomes over an indefinitely (or infinite) large number of trials." Rosner 206 p. 44 Definition 3.1**

**Sometimes, for theoretical reasons, aspects of the probablity distributions are known or are assumed. More commonly in practice, however, one takes a reasonably large empirical sample and compares it with known theoretical distributions, such as the Normal Distribution.**

$x := \text{rnorm}(100, 0, 1)$

**< For example data points drawn by this function gives following histogram...**

$\text{plot} := \text{histogram}(30, x)$

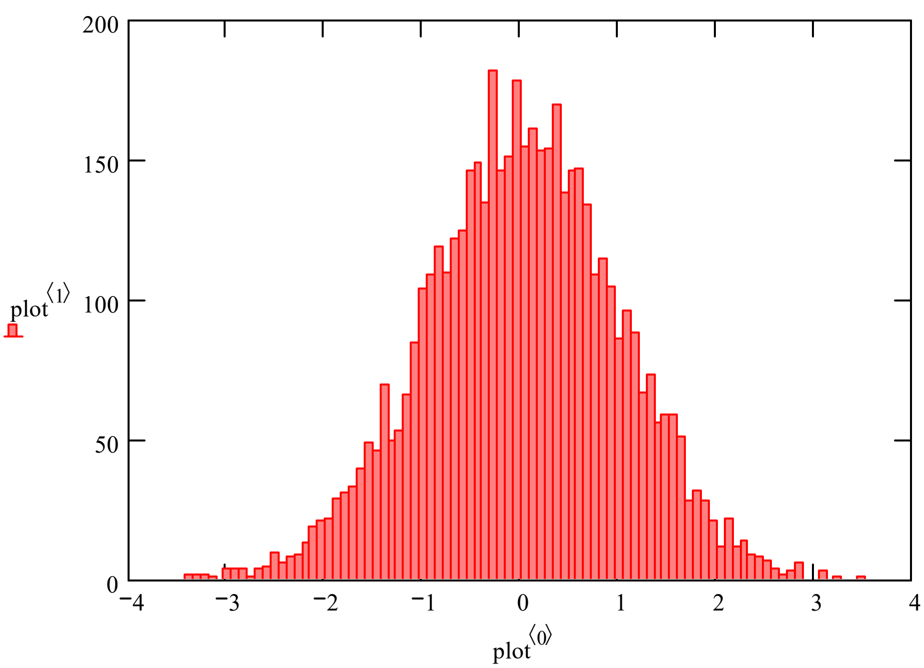



**^ From this limited sample, one might conclude that the population from which it was drawn has a Normal distribution...**

$x := \text{rnorm}(5000, 0, 1)$

**< But what if we draw a bigger sample, say 5000 values, and plot it with 100 bins instead of 30?**

$\text{plot} := \text{histogram}(100, x)$




**Conclusion: a bigger sample is usually better in guessing an underlying probability distribution... But other factors usually come into play including cost/time in conducting the study, and other biases of one sort or another.**

## Bayesian Inference:

**Here two kinds of probability are distinguished:**

**"The *prior probability* of an event is the best guess by the observer of an event's probability in the absence of data. This prior probability may be a single number, or it may be a range of likely values for the probability, perhaps with weights attached to each possible value." Rosner 2006 p. 63, Definition 3.16.**

**"The *posterior probability* of an event is the probability of an event after collecting some empirical data. It is obtained by integrating information from the prior probability with additional data related to the event in question." Rosner 2006 p. 64, Definition 3.17.**

**We'll look at aspects of Bayesian Inference shortly...**

## The general logic of probability:

**Under either of the above views, probability (both as a concept and a property) obeys fundamental logical (or mathematical) rules. These rules are very important to all aspects of statistical inference and in direct prediciton of outcomes.**

## Terminology:

| | |
|---|---|
| ***sample space*** | **= the set of all possible outcomes** |
| ***an event*** | **= any specific outcome** |
| **P(X)** | **= probability of event X, where** $0 \le P(X) \le 1$ |
| **complement of X** | **= (1-P(X)) = P(~X). Complement is the probability of X not happening.** |

## Mutually exclusive events:

**Two or more events (A,B,C...) are mutually exclusive if they can not both happen simultaneously.**

### Intersection of events is the empty set:

$P(A \wedge B) = \phi$ **< for two events (the smallest number where intersection has a meaning)**

$P(A_1 \wedge A_2 \wedge ....A_i) = \phi$ **< for two or more events *i***

### Union of events - The Law of Addition of probabilities applies:

$P(A \vee B) = P(A) + P(B)$ **< Probability of either A or B happening are their separate probabilities added together...**

$$P(A_1 \vee A_2 \vee A_3 \vee ....A_i) = (P(A_1) + P(A_2) + P(A_3) + .... + P(A_i))$$

**^ for multiple exclusive events, add them all.**

### Examples:

**Flipping a coin: H=event 'heads', T=event 'tails'**

**For a 'fair' coin, P(H) = 0.5 & P(T) = 0.5**
**Therefore:** $P(H \wedge T) = \phi$ **&** $P(H \vee T) = P(H) + P(T) = 0.5 + 0.5 = 1.0$

**Rolling a die: Here there are 6 mutally exclusive events: 1, 2, 3, 4, 5, 6**
**For a 'fair' die, P(i) = 1/6 for all 6 events**

**Therefore:** $P(2 \wedge 6) = \phi$ **&** $P(2 \vee 6) = P(2) + P(6) = 1/6 + 1/6 = 1/3$

## Potentially co-ocurring events:

**Two or more events (A,B,C...) may occur simultaneously. There are two kinds:**

### 1. Independent events:

**The probilities of events P(A), P(B), P(C) etc. have no bearing on each other.**

**Intersection of events - the Law of Multiplied probabilities applies:**

$P(A \wedge B) = P(A)\,P(B)$ **< multiply the separate probabilities to find the probability of both events occuring simultaneously.**

$$P(A_1 \wedge A_2 \wedge A_3 \wedge ....A_i) = P(A_1) \cdot P(A_2) \cdot P(A_3) .... P(A_i)$$

**^ multiply all of them for multiple events**

**Union of events - Expanded Law of Addition applies:**

$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

**< The probability of A or B happening is the separate probabilities added together minus the probability that both A & B occur together**

**Alternate equivalent forms for Independent events only:**

$P(A \vee B) = P(A) + P(B) \cdot (1 - P(A))$

$P(A \vee B) = P(B) + P(A) \cdot (1 - P(B))$

**< The probability of A or B happening is the probability of one plus the simultaneous occurrence of the other but not the first!**

**Note: the complement here ^ = P(~A) or P(~B)**

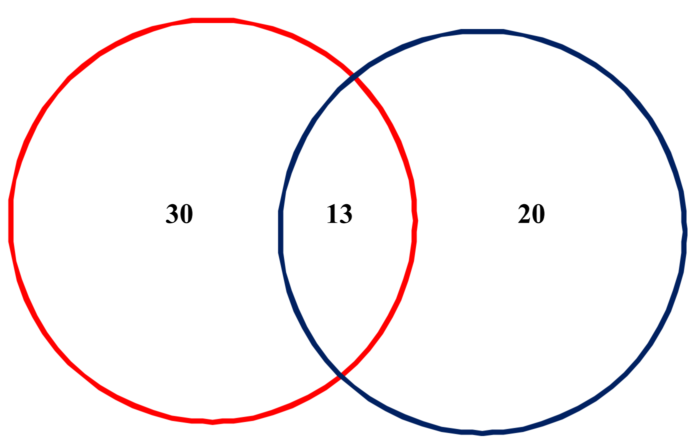
**Check Venn diagrams in the text to puzzle this out!**

**Note that Union for multiple independent events greater than two is not given...**

## Example:

**In a sample 50 birds, by chance 20 are male and 30 are female. Out of these, 7 males and 13 females test positive for a potentially dangerous virus. If a bird is chosen at random from this sample, what is the probability of obtaining a female or a bird testing positive for the virus - but not both occurring together?**

**P(female or pos) = P(female) + P(pos) - P(female and pos) = 30/50 + 20/50 - 13/50 = 37/50**



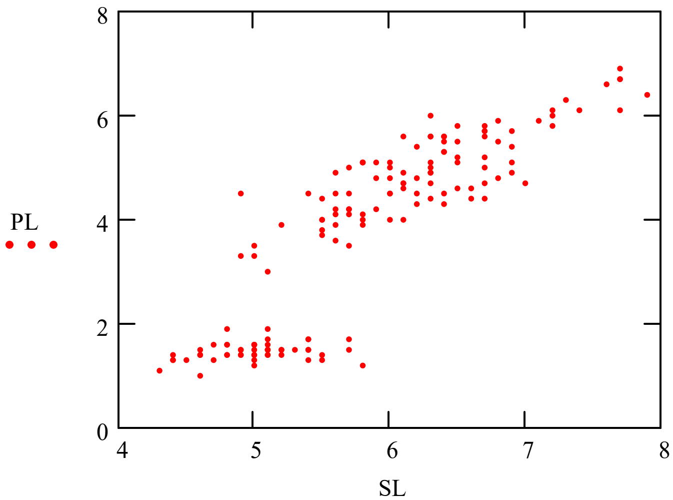
## 2. Dependent events:

**The probabilities of two events are related such that knowing the outcome of one event influences the probability of the other.**

iris := READPRN("c:/DATA/Biostatistics/iris.txt")

$SL := iris^{\langle 1\rangle}$ $SW := iris^{\langle 2\rangle}$ $PL := iris^{\langle 3\rangle}$ $PW := iris^{\langle 4\rangle}$

**< Reading the Famous Iris data again...**




**< A plot of Sepal Length (SL) and Petal Length (PL) shows dependence. Measuring one variable gives important information about more probable values of the other.**

$n := \text{length}(PL)$ $n = 150$

$i := 0..n-1$

$PL_{bar} := \text{mean}(PL)$ $PL_{bar} = 3.758$

$SL_{bar} := \text{mean}(SL)$ $SL_{bar} = 5.843$

$$\frac{n}{n-1}\cdot \text{cvar}(PL, SL) = 1.274315$$

### Covariance:

$$cov_{PLSL} := \left(\frac{1}{n-1}\right)\cdot\left[\sum_i \left(PL_i - PL_{bar}\right)\cdot\left(SL_i - SL_{bar}\right)\right] \qquad cov_{PLSL} = 1.274$$

### Intersection of events - the Law of Multiplied probabilities fails:

$P(A \wedge B) \neq P(A)\,P(B)$

**This is a more formal statement of what dependence actually means. In practical terms, one often assesses the separate probabilities for A and B, and then compares their product with a separately estimated probability of both events occuring simultaneously to see if they match.**

**To proceed at this point, one needs a concept of conditional probability...**

$P(A \wedge B) = \text{"P(B|A)"}\cdot P(A)$

$P(B \wedge A) = \text{"P(A|B)"}\cdot P(B)$

**< intersection in terms of conditional probability.**
**Note that you can switch the roles of A & B depending on which is *prior probability* P(X) = known prior to collecting data for the study versus *conditional probability* "P(X|Y)" knowledge of Y influencing the probability of X**

*** See below for more than two events!**

## Conditional Probability:

**The concept of conditional probability can be applied to both the independent and dependent cases of potentially simultaneous events above, so I'll give both here..**

## Definition of Conditional Probability:

**Rearranging the Law of Multipied Probabilities for two dependent events to solve for one of the individual probabilities (i.e., P(B)), gives the definition for conditional probability:**

$$P\left(\frac{B}{A}\right) = \frac{P(A \wedge B)}{P(A)}$$

**< P(B) is also written as P(B|A) with no difference in meaning.**

**^ This is the conditional probability for B given prior knowledge of A...**

## Calculating Intersection with Conditional Probability:

### 1. Independent case:

$\mathbf{P(B|A) = P(B) = P(\sim A)}$

$P(A \wedge B) = P(A) \cdot P(B)$

**< Equalities here makes the Law of Multiplied probabilities a special case of the more general one below...**

### 2. More important Dependent case:

$$P(A \wedge B) = P\left(\frac{B}{A}\right) \cdot P(A)$$

**< For two events....**

$$P(A_1 \wedge A_2 \wedge A_3) = P(A_1) \cdot P\left[\frac{A_2}{(A_1)}\right] \cdot P\left[\frac{A_3}{(A_2 \wedge A_1)}\right]$$

**< For three events...**

$$P(A_1 \wedge A_2 \wedge A_3 \wedge ....A_i) = P(A_1) \cdot P\left[\frac{A_2}{(A_1)}\right] \cdot P\left[\frac{A_3}{(A_2 \wedge A_1)}\right] \cdot .. \cdot P\left(\frac{A_i}{A(i-1) \wedge A_3 \wedge A_2 \wedge A_1}\right)$$

**^ For more than three events...**

## Calculating Total Probability from Conditional Probability:

**This formulation is common to both the Independent and Dependent cases:**

$\mathbf{P(A) = P(A|B)\cdot P(B) + P(A|\sim B)\cdot P(\sim B)}$

$\mathbf{P(B) = P(B|A)\cdot P(A) + P(B|\sim A)\cdot P(\sim A)}$

**< For two possibilities A & B... Note that the roles of A & B are interchangeable**

$\mathbf{P(A) = \Sigma_\iota\ P(A|B_i)\cdot P(B_i)}$ **< For A given multiple prior probabilities $B_i$**

### 1. Independent case:

**The formulas simply reduces to multiplying P(B) or P($B_i$) depending on number of $B_i$'s**

### 2. Dependent case:

**The formula doesn't reduce and conditional probabilties must be used as stated above.**

## Bayes' Rule:

**The point of this procedure for two events A & B is to estimate one conditional probability P(B|A) from the other conditional probability P(A|B) and one total probablity P(B).**

$$P\left(\frac{B}{A}\right) = \frac{P\left(\frac{A}{B}\right) \cdot P(B)}{P\left(\frac{A}{B}\right) \cdot P(B) + P\left(\frac{A}{notB}\right) \cdot P(notB)}$$

**< Of course, as above, the defined roles of A & B here can be reversed.**

$$P\left(\frac{B_i}{A}\right) = \frac{P\left(\frac{A}{B_i}\right) \cdot P(B_i)}{\sum_i P\left(\frac{A}{B_i}\right) \cdot P(B_i)}$$

**< General form of Bayes' Rule giving multiple conditional probabilities for the B's given knowledge of multiple conditional probabilites P(A|$B_i$) and multiple total probabilites P($B_i$).**