Intro to Mathematical Probability and Statistics

Dr. John Travis
Mississippi College

Terms:

- Experiment - any act that may be repeated under similar conditions resulting in a trial which yields an outcome.
- Sample - a part of the population analyzed to estimate the characteristics of the population; a collection of actual outcomes from a repeated experiment. Making a conjecture or observation about the population distribution based upon a sample is called a statistical inference.
- Random Experiment - an experiment for which the outcome cannot be predicted with certainty but for which the collection of outcomes can be described or listed. Performing the experiment in an unbiased way gives a random sample.
- Sample Space - a list of all the possible outcomes of a given random experiment.

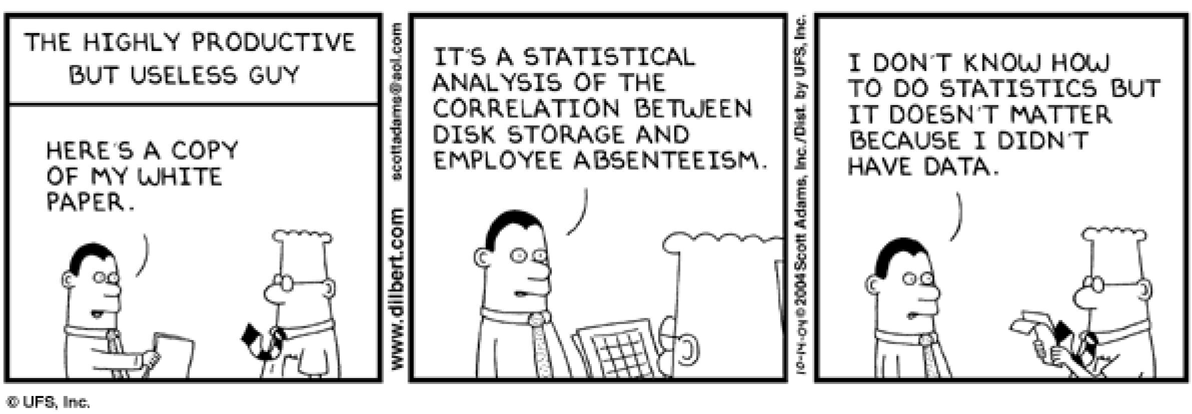
Defn: Given a random experiment with sample space S, a function X mapping each element of S to a unique real number is called a random variable. For each element s from the sample space S, denote this function by

$$X(s) = x$$

and call the range of X the space of X: $R=\{ x : X(s)=x, \text{ for some } s \text{ in } S\}$

Measurement Scales

- Nominal - Mutually Exclusive and Exhaustive categories for which the numerical value has only identification significance. Ex: Male = 1, Female = -1
- Ordinal - Discrete values ranked from lowest to highest or vice versa. Ex: Class grades for GPA.
- Interval - Ordinal data where distance between data values is of significance. Ex: Heights and Weights.
- Ratio - Interval data where ratios of observations have meaning. Ex: Percentile rankings



Goal: We will make various restrictions on the range of the random variable to fit different generalized problems. Then, we will be able to work on a problem (which may be inherently non-numerical) by using the random variable in subsequent calculations.

Ex: When dealing with only two outcomes, one might use

S = { success, failure}.

Choose X(success)=1, X(failure)=0. Then, R={0,1}.

Ex: When gambling with a pair of dice, one might use

S=ordered pairs of all possible rolls = {(a,b): a=die 1 outcome, b=die 2 outcome}.

Choose X((a,b)) = a+b. Then, R={2, 3, 4, 5, ..., 12}.

Ex: When rolling dice in a board game (like RISK), one might use

S={(a,b): a=die 1 outcome, b=die 2 outcome}

Choose X((a,b)) = max{a,b}. Then, R={1, 2, 3, 4, 5, 6}

Defn: R contains a countable number of points if either R is finite or there is a one to one correspondence between R and the positive integers. Such a set will be called discrete. We will see that often the set R is not countable. If R consists of an interval of points (or a union of intervals), then we call X a continuous random variable.

Remark: We would like to have a measure of how often we expect an outcome to occur. We will call this measure the probability of the outcome. However, we often do not have complete information. Therefore, we often repeat an experiment n times (n>1) generating several outcomes.

- Frequency: The number of times any given outcome A occurs is called the frequency of outcome A and is denoted |A|.
- Relative Frequency: The ratio |A|/n is called the relative frequency or sample probability of outcome A. As we repeat the experiment more and more (ie. n gets large) the relative frequency generally will stabilize toward a value p. One may show as n approaches infinity lim |A|/n = p.
- Probability: We call p the probability of event A and denote it P(A). Note, P(A) is actually independent of the number of trials and will represent the fraction of times that the outcome of a random experiment results in the "desired" outcome A in a large number of trials of that experiment.

Experiment: Explain the Monty Hall game. What would you expect to be the probability of winning? Play the game a large number of times and compare this to your relative frequency of winning. Watch video clip for the Monty Hall game from the TV show NUMBERS, Season I, Episode 13. (An online version of the Monty Hall game is available at http://www.stat.uiuc.edu/courses/stat100/cuwu/http://www.stat.uiuc.edu/courses/stat100/cuwu/ . A good analysis of the game is available on this site.)

Techniques for Representing Data: (Most to be discussed throughout this course.)

1. Tabular Methods - based on the entire population yielding a global picture
    1. frequency distributions
    2. relative frequency distributions
    3. cummulative frequency distributions
    4. Stem-and-Leaf Displays
    5. Box-and-Whisker Diagrams
2. Summary Methods
    1. Measures of the center
        1. Mean
        2. Median
        3. Mode
    2. Measures of spread
        1. Range
        2. Variance and Standard Deviation
        3. Quantiles
    3. Measures of Skewness - indicates the level of symmetry of the data
        1. Pearson Coefficient
        2. Standard Skewness
        3. Bowley's Measure
    4. Measures of Kurtosis - indicates flatness or roundedness of the peak of the data
        1. Standard Kurtosis
        2. Coefficient of Kurtosis
    5. Measures of Association for Bivariate Data - indicates the likeliness of functional correlation of the data.
        1. Pearson Correlation Coefficient
        2. Spearman Rank Correlation Cooeficient
        3. Quantile-Quantile Plots
    6. Detection of Outliers - indicates whether abnormally large or small data distorts other techniques
        1. Z-scores
        2. Trimming
        3. Winsorizing
    7. Tests for Normality - indictes if the data is bell-shaped
        1. Standard Percentages relative to standard deviations from the mean
        2. Chi-square
        3. Kolmogorov-Smirnov
        4. Lilliefors
        5. Shapiro-Wilk
    8. Tests for Randomness - indicates whether the data has a non-systematic pattern
        1. Runs Test
        2. Mean-Square Successive Differences

Remark: Many of these measures above are relative and some are absolute.

© Scott Adams, Inc./Dist. by UFS, Inc.

Simpson's Paradox: See example on page 9. See if you can come up with another example with at least 3 different "seasons".

Remark: By considering n = a large number of experiments, we collect the corresponding relative frequencies f_k / n and notice they satisfy:

- f_k / n > 0.   (The relative frequency of an outcome that occurs is positive.)
- $\sum_{x \in R} f_k / n = 1$ (An outcome from the sample always occurs.)
- For A = a collection of outcomes, |A|/n = $\sum_{x_k \in A} f_k / n$.   (Distinct relative frequencies are additive.)

Histograms: A graphical way to express data with vertical bars. Many times it is easier to understand a collection of date using a picture such as a histogram instead of a table of numbers. Experiment with drawing histograms with software..

Remark: We will relate this empirical set of observations to the definition of the theoretical probability mass function. This probability mass function (pmf) of a discrete random variable should satisfy the following properties:

- f(x) > 0, for all x in R.   (The probability of an outcome that may occur is positive.)
- $\sum_{x \in R} f(x) = 1$ (An outcome from the sample space always occurs.)
- For $A \subseteq R$, P(A) = $\sum_{x \in A} f(x)$.   (Distinct Probabilities are additive.)
- f(x) = 0, for all $x \in R^c$.   (The probability of something that cannot occur is zero.)

Defn: Suppose X is a discrete random variable with range R = {x_1, x_2, ..., x_n} and pmf f(x). In general:

The arithmetic mean is a measure of the middle

$$\mu = x_1 f(x_1) + \ldots + x_n f(x_n) = \sum_{k=1}^{n} x_k f(x_k) .$$

The mean is often called the centroid in the sense that if the x_k were locations of objects of weight $f(x_k)$, then the centroid would be the point where this system of n masses would balance.

Measures of spread:

- Average Deviation from the Mean – always zero for any distribution
- Average Absolute Deviation from the Mean – difficult to deal with algebra when absolute values are used
- Average Squared Deviation from the Mean – always non-negative and good with algebra and calculus

Defn: The *variance* is a measure of spread found by using the average squared deviation from the mean

$$\sigma^2 = \sum_{k=1}^{n} (x_k - \mu)^2 f(x_k)$$

if this value exists and is also denoted by Var(X). The positive square root of the variance is called the *standard deviation* and is denoted by σ.

Other Means:

- Geometric Mean = $(x_1\, x_2 \ldots x_n)^{1/n}$
- Harmonic Mean = $\frac{n}{\sum_{k=1}^{n} 1/x_k}$

Result: In general Harmonic Mean $\leq$ Geometric Mean $\leq$ Arithmetic Mean.

Written Assignment: Investigate the implications and applications of the geometric mean, harmonic mean and the arithmetic mean. Prepare a written report summarizing your findings. Be certain to list your collection of sources not all of which being web sites.

Result: (Chebyshev's Theorem) Given a random variable X with given mean μ and standard deviation σ, for k>1 at least $1 - \frac{1}{k^2}$ of the observations lie within k standard deviations from the mean.

HOMEWORK: page 10

Definitions:

- The *Universal Set*, denoted S, is the set of all objects under consideration at a given moment. The *population*. The sample space.
- An *Element* of a set is any member of the set. A *Subset* is any part of a set. An *Event* is a collection of possible outcomes and must be contained in S. The *Empty (Null) Set* is a notational idea representing the set (or subset) with no members.
- The basic set operations: *Union, Intersection, Complement*
- Two sets are *Mutually Exclusive* or *disjoint* if they have no elements in common (their intersection is disjoint).
- *Venn diagrams* are useful is graphically representing sets.
- A *Set Function* is a function whose domain is a collection of sets.

Defn: *Probability* – page 16

Result: P(A) = 1 - P(A'), or P(A') = 1 - P(A).

Pf: $S=A \cup A'$, $A \cap A'=0$.

Corollary: $P(\varnothing) = 0$.

Pf: Set $A=\varnothing$. Then, A' = S. Use above.

Result: If A is contained in B, $P(A) \leq P(B)$.

Pf: Notice, $A \cap (B \cap A') = (A \cap B) \cap A' = A \cap A' = \varnothing$, since A is contained in B.

So A and $B \cap A'$ are disjoint.

Hence, $B = B \cup A = (A \cup B) \cap S = (A \cup B) \cap (A \cup A') = A \cup (B \cap A')$.

So, $P(B) = P(A \cup (B \cap A')) = P(A) + P(B \cap A')$.

Corollary: $P(A) \leq 1$, for any event A.

Pf: Set B = S in above.

Result: For any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Pf: Look at Venn Diagram.

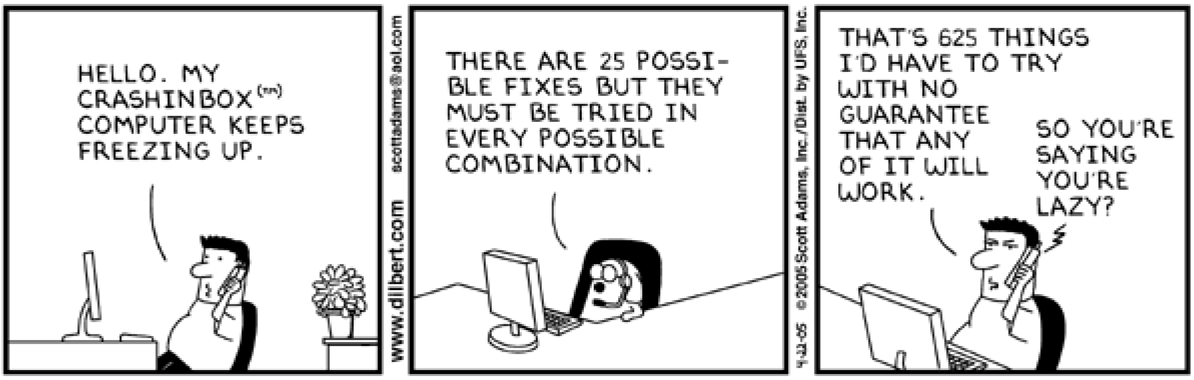
Defn: A simple event is an outcome which can not be broken down into a smaller outcome.

Remark: If several simple events are *equally likely*, then each has an equal chance of being selected and so should have probability 1/n. If a collection A of equally likely outcomes is being considered, then

$$P(A) = |A| / n .$$

HOMEWORK: page 20

Multiplication Principle: Consider successive *independent* experiments with n and m possible outcomes respectively. The number of possible pairs of outcomes then would be nm. This can be generalized to any number of successive *independent* outcomes.



Remark: We often need to count the number of ways something can happen when doing so could take a very long time. So, we would like to find ways to count without actually counting individual entries.

Terminology:

- Permutation - order of choosing items is important
- Combination - order of choosing items is not important
- With replacement/Without replacement
- Distinguishable/Not Distinguishable

Permutations when distinguishable and without replacement:

P(n,r) = $\frac{n!}{(n-r)!}$ = # of ways to get ordered samples of size r from a sample space of size n.

Also, P(n,n)=n!

Combinations when distinguishable and without replacement: (Binomial Coefficients)

C(n,r) = $\frac{n!}{r!(n-r)!} = \binom{n}{r}$ = # of ways to get unordered samples of size r from a sample space of size n.

Permutations when Not all items are distinguishable and without replacement: (Multinomial Coefficients)

If n items belong to s categories, n_1 in first, n_2 in second, ... , n_s in the last, the number of ways to pick all is

$$\frac{n!}{n_1!n_2!...n_s!}$$

Combinations when distinguishable and with replacement:

$\binom{n-1+r}{r}$ = Number of ways to get unordered samples of size r from n objects.

HOMEWORK: page 30

Defn: The *Conditional Probability* of A given that B has occured is

$$P(A \mid B) = P_B(A) = P(A \mid B) / P(B),$$

if P(B)>0.

Motivation: $\dfrac{|A\cap B| / |S|}{|B| / |S|} = \dfrac{|A \cap B|}{|B|}$. This is similar to restricting the problem to a smaller sample space.

Result: Conditional Probability is indeed a Probability Function satisfying the definition of probability.

Multiplication Rule: P(A | B) = P(A) P(B | A) = P(B) P(A | B).

HOMEWORK: page 40

Defn: Events A and B are *independent* (statistically independent, stochastically independent) if

$$P(A \mid B) = P(A)\, P(B).$$

That is, P(B | A)=P(B) and P(A | B) = P(A).

HOMEWORK: page 49

Defn: A partition of the sample space S is a collection of subsets which are exclusive (have no common elements) and exhaustive (all elements in S are in the union of the subsets)

Bayes Theorem: Suppose a partition of S is given by $S = B_1 \cup B_2 \cup \ldots B_n$. If A is any other event in S and P(A)>0, then

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}$$

Pf: $A = A \cap S = (B_1 \cap A) \cup (B_2 \cap A) \cup \ldots (B_n \cap A)$, a partition of the set A.

By the multiplication rule noting that the sets are mutually disjoint

$$P(A) = P(B_1 \cap A) + P(B_2 \cap A) + \ldots P(B_n \cap A) = \sum_{k=1}^{n} P(B_k \cap A) = \sum_{k=1}^{n} P(B_k)P(A \mid B_k).$$

From the definition of conditional probability and then replacing P(A) with the above, we get

$$P(B_k \mid A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k \cap A)}{\sum_{k=1}^{n} P(B_k)P(A \mid B_k)}.$$

Remark: This result gives us the opportunity to reverse the order of the events

Homework: Page 55