Overview of Statistical Concepts

Lecture 2

SIE 512 Spatial Analysis

Fall 2011

Basic Terminology for Statistical Models

Concerned with phenomena which are **stochastic**
- phenomena subject to uncertainty

A **stochastic process** is a process which is indeterminate in its future evolution – there are a number of possible outcomes – some more probable than others

A stochastic process is described by random variables and their probability distributions

Events and sample space

The set of all possible outcomes of a stochastic process is called the **sample space**, often denoted by Ω (omega)

An **event** is a subset of the sample space.

An elementary event consists of just one element of Ω

A compound event consists of more than 1

When a coin is tossed once the outcomes are {H, T}

When a coin is tossed twice Ω = {HH, HT, TH, TT}

A compound event for this sample space is that we get at least one head {HH, HT, TH}

Random Variables

A random variable X, is a label attached to a random event.

A random variable can be used to describe the process of rolling a fair die and the possible outcomes { 1, 2, 3, 4, 5, 6 }. Another random variable might describe the possible outcomes of picking a random person and measuring his or her height.

Unlike regular mathematical variables, a random variable cannot be assigned a value. It is a *function* that maps outcomes to numbers.

It does not describe the actual outcome of a particular experiment, but rather the possible, as-yet-undetermined outcomes.

Random variables are denoted by upper case letters (e.g.**Y, X, Z)**

Specific values or instances of random variables are expressed as lower case letters (e.g. x). In other words *x* is a possible value that random variable **X** can take.

Random Variables

Random variables can be discrete or continuous.

A discrete random variable is one which may take on only a countable number of distinct values such as 0, 1, 2, 3, 4, ...

Discrete random variables are usually (but not necessarily) counts. A discrete random variable can take only a finite or countably infinite number of distinct values.

Continuous random variables take on an infinite number of values - the range of real numbers, e. g. the level of ozone today, level of arsenic in groundwater, temperature.

Random Variables and Probability Distributions

Probability is the chance of a random event occuring.

Recording all the probabilities of values or ranges of a random variable *X* yields the probability distribution (discrete) or probability density of *X*.

Probability Concepts

Proportion of time we expect a result to come out in a particular way-before the experiment occurs

Proportion or Number of Symmetric Ways Definition

$$\text{Probability of an event} = \frac{\text{Number of outcomes leading to an event}}{\text{Number of outcomes possible}}$$

Probability of picking a queen from a well shuffled deck of cards:

$$P(\text{Queen}) = \frac{4}{52} = .07$$

Probability Concepts

Relative Frequency Definition

The proportion of time that events of the same kind will occur in the long run

m/n where m is the number of times an event occurs in n trials

Probability of a baby girl:

$$P(\text{Girl}) = \frac{\text{Number of girls born}}{\text{Number of births}}$$

Probability Concepts

- **Judgment or subjective view**

  Takes into account prior knowledge about events

  Bayes' theorem - probabilities are interpreted as representing beliefs, or knowledge

Probability that Obama will be reelected?
Probability of Red Sox winning the world series this year?

Bayesian probability interprets the concept of probability as "a measure of a state of knowledge". A Bayesian approach uses data to update prior probability estimates to generate an improved posterior probability estimate.

Probability Concepts

- **Probability Axioms – 3 necessary and sufficient axioms**

- The probability of an event is a non-negative real number
  $P(E) \geq 0, \ \forall E \subseteq \Omega$, where Ω is the sample space
- Probability of all events in sample space (Ω) sum to 1, $P(\Omega) = 1$
- $P(A \text{ or } B) = P(A) + P(B)$ for all mutually exclusive events A and B

Kolmogorov axioms

Discrete probability distribution

A probability distribution is discrete if there is a finite or countable set whose probability is 1.

The probability distribution of a discrete random variable X is a function which gives the probability $p(x_i)$ that the random variable X equals x_i

$$p(x_i) = P[X = x_i]$$

$f_X(x)$ probability that the random variable X takes the value x

Continuous probability distribution

Continuous distributions are characterized by a probability density function.

The probability density function is a function which can be integrated to obtain the probability that the random variable takes a value in a given interval

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(x)dx$$

Probability that a random variable takes a value in a range

If x is discrete

$$\sum_{x=a}^{b} f_X(x)$$

The probability X takes values in the range *(a,b)*

If x is continuous

$$\int_a^b f_X(x)dx$$

Probability density over the the range *(a,b)*

Cumulative Distribution Function

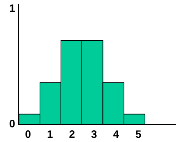
$$F_X(x) = \begin{cases} \sum_{u=-\infty}^{x} f_X(u) & \text{If X is discrete} \\ \int_{-\infty}^{x} f_X(u)\,du & \text{If X is continuous} \end{cases}$$

Probability of X taking any value less than or equal to x

Function giving the probability that the random variable X is less than or equal to x, for every value x.

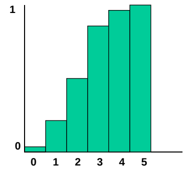
Discrete case : A random variable X has the following probability distribution p(x)

xi	0	1	2	3	4	5
p(xi)	1/32	5/32	10/32	10/32	5/32	1/32



The cumulative distribution function F(x) is then:

xi	0	1	2	3	4	5
F(xi)	1/32	6/32	16/32	26/32	31/32	32/32



Joint Probability

If two events *A* and *B* occur on a single performance of an experiment this is called the intersection or joint probability of *A* and *B*, denoted as

$$P(A \cap B)$$

If two events, *A* and *B* are independent then the joint probability is

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B)$$

For example, if two coins are flipped the chance of both being heads is $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$

Union of Events

If either event *A* or event *B* or both events occur on a single performance of an experiment this is called the union of the events *A* and *B* denoted as

$$P(A \cup B)$$

If two events are mutually exclusive then the probability of either occurring is

$$P(A or B) = P(A \cup B) = P(A) + P(B)$$

For example, the chance of rolling a 1 or 2 on a six-sided die is

$$P(1\, or\ 2) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

The Sum Rule

If the events are not mutually exclusive then

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

The probability that *A or B* will happen is the sum of the probabilities that *A* will happen and that *B* will happen, minus the probability that both *A and B* will happen

When drawing a single card at random from a regular deck of cards, the chance of getting a heart or a face card (J,Q,K) (or one that is both) is

$$\frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{11}{52}$$

Conditional Probability

The probability of an event given that another event has happened

The conditional probability of an event A given B is

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cup B)}{\mathrm{P}(B)}$$

read "the (conditional) probability of *A*, given *B*"

Independence

Probabilistic behavior of one random variable remains the same no matter what values the other variable takes

Joint probability distribution is the same as product of their individual probability distributions

For independent events

$$\mathrm{P}(A \cup B) = \mathrm{P}(A)\mathrm{P}(B)$$

And so conditional probability is:

$$\mathrm{P}(A|B) = \mathrm{P}(A)$$

Conditioning on independent events does not change the probability of an event

Expectation of a random variable

The probability distribution of random variable can usually be characterized by a small number of parameters.

It is often enough to know what its "average value" is. This is captured by the mathematical concept of expected value of a random variable, denoted E[*X*] – measures the center of a probability distribution

The **expected value or mathematical expectation of a** random variable is the sum of the probability of each possible outcome of the experiment multiplied by its payoff ("value"). Thus, it represents the average amount one "expects" as the outcome of the random trial when identical odds are repeated many times.

The value itself may not be expected in the general sense; it may be unlikely or even impossible.

Expectation of a random variable

If we have a set A of values a_1, a_2, a_3,....a_n and their probabilities are p_1 , p_2 , p_3 ,… p_n

Then the mathematical expectation is

$$E(A) = a_1p_1 + a_2p_2 + \ldots + a_np_n$$

A weighted average – where weights are the probability for the value

Expectation of a random variable

Expectation for X, the number of heads obtained in three flips of a balanced coin

Probabilities for 0,1,2,3 heads are $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$

$$E(X) = 0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = \frac{12}{8} = 1.5$$

Expectation of a random variable

$$E(X) = \begin{cases} \sum_{y=-\infty}^{\infty} x f_X(x) & \text{discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{continuous} \end{cases}$$

Variance of a probability distribution

For a random variable X having a distribution P(X)with *known* population mean μ, the population variance var(X) , commonly written σ^2, is defined as

$$\sigma^2 \equiv \left\langle (X-\mu)^2 \right\rangle$$

where μ is the population mean and $\langle X \rangle$ denotes the expectation value of X. For a discrete distribution with N possible values of x_i, the population variance is

$$\sigma^2 = \sum_{i=1}^{N} P(x_i)(x_i - \mu)^2$$

whereas for a continuous distribution, it is given by

$$\sigma^2 = \int P(x)(x-\mu)dx$$

Variance of a probability distribution

The **variance of a** random variable (or equivalently a probability distribution) is a measure of its statistical dispersion, indicating how its possible values are spread around the expected value.

$$\text{Var}(X) = \sigma^2 = E((X- E(X))^2 = E(X^2)-E(X)^2$$

Standard deviation of a random variable is the square root of the Var(X)

Sample Variance

If the underlying distribution is not known, then the sample variance may be computed as

$$s_N^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

where $\bar{X}$ is the sample mean.

Covariance

Measures the correspondence or covariation of two random variables together

Cov(X,Y) = $E(X-E(X))(Y-E(Y))$

Correlation = $\dfrac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

Skewness

A measure of the asymmetry of the probability distribution of a real-valued random variable.

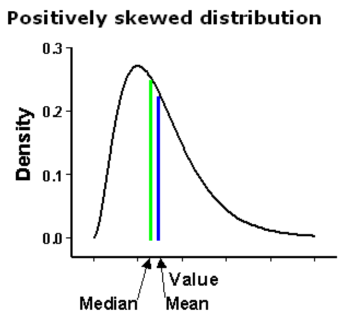
$$\gamma = \frac{E(x - E(x))^3}{\sigma^3} \qquad Skew = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^3$$

Skewness is computed as a third moment about the mean

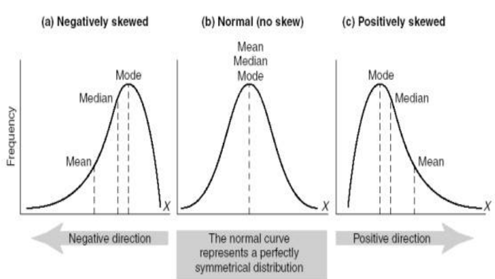
If the third moment is positive, the distribution is said to be positively skewed; if it is negative, the distribution is negatively skewed.

The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero.

Skewness



Skewness



Kurtosis

A measure of the "peakedness" of the probability distribution of a real-valued random variable. Higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations.

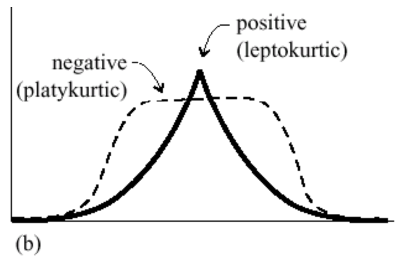
$$k = \frac{E(x - E(x))^4}{\sigma^4} - 3 \qquad k = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^4 - 3$$

Kurtosis = 0 … Standard Normal Distribution

Kurtosis > 0 … positive ("leptokurtic")

Kurtosis < 0 … negative ("platykurtic")

Kurtosis



Graphical representation: Histograms

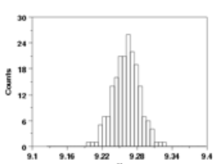
Histograms are one useful way of displaying the distribution of a data set.

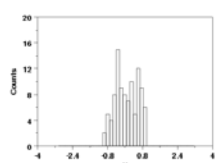
A histogram summarizes the data by placing observations into intervals (bins), counting the number of observations in each interval and dividing by sample size.

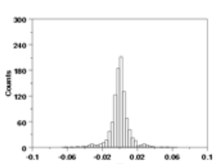
The y axis can indicate number of observations, percentage of total observations, fraction of total, density (frequency divided by bin width)

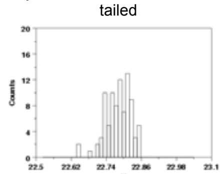
Histograms reveal location, spread, and shape of the distribution, presence of outliers in a data set.

Graphical representation: Histograms


normal


Symmetric, non-normal, short tailed


Symmetric, non-normal, long tailed


Non-symmetric, left skewed

Graphical Representation: Box Plots

A box plot (box and whisker) plots the quantiles of the data

A quantile is the number x_p such that a proportion *p* of the values are less than or equal to x_p

$x_{0.25}$ is the value such that 25% of all the values fall below this value.

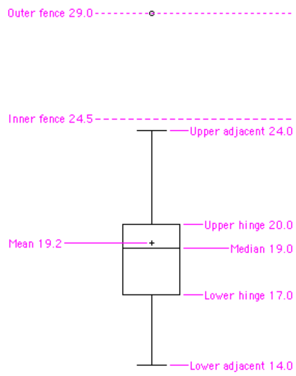
The edges of the box show the 25th and 75th percentiles and a line in the box indicates the median (50th percentile).

The whiskers can show the bounds of the data range (min and max) or the values that are 1.5 IQR

Graphical Representation: Box Plots

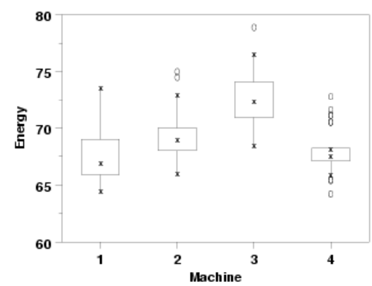
A boxplot allows one to quickly see characteristics of a data distribution

- The center of the data
- The spread of the data
- The presence of outliers
- The skewness of the data by looking at relative lengths of the box halves and whiskers



Graphical Representation: Box Plots

Allows comparison of several data distributions



Empirical probability distributions

An empirical estimate of a probability distribution is obtained from a sample of data

For large sample sizes the empirical estimates approach the population probability distribution

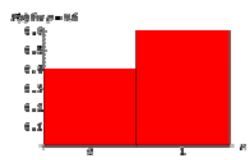
Theoretical probability distributions

Models of the probability distributions for observed data

A simple functional form determined by a small number of parameters θ

$$p(x) = f(x; \theta_1 \ldots, \theta_m)$$

Bernoulli Distribution



The Bernoulli distribution is a discrete distribution having two possible outcomes labeled by *n= 0* and *n=1* in which ("success") occurs with probability *p* and ("failure") occurs with probability q=1-p where 0<p<1 It therefore has probability density function

$$P(n) = \begin{cases} 1-p & \text{for } n=0 \\ p & \text{for } n=1, \end{cases}$$

Bernoulli Distribution

One population parameter *p*

$$P(X = x) = p^x (1- p)^{1-x}$$

$$E(X) = \sum_{x=0}^{1} x(p^x (1-p)^{1-x} = (0)(1-p) + (1)(p) = p$$

$$Var(X) = E(X^2) - (E(X))^2$$
$$= p * 1^2 + (1-p) * 0^2 - p^2 = p(1-p)$$
$$= pq$$

Described by $X \sim Be(p)$

Binomial Distribution

The binomial distribution gives the discrete probability distribution of obtaining exactly *n* successes out of *N* Bernoulli trials.

Assumptions

- There are two possible outcomes for each trial
- The probability of a success is the same for each trial
- There are N trials, where N is constant
- The N trials are independent
- p is probability of success
- q=1- p is probability of a failure

$$Pr(n|N) = \binom{N}{n} p^n q^{N-n} \qquad Pr(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

Described by $X \sim Bin(N,p)$

Binomial Distribution

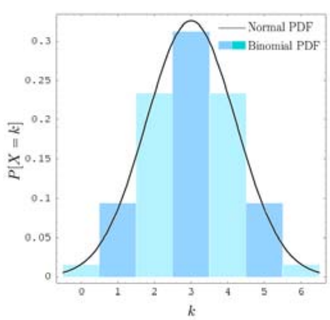
The expected value is:

$$E(X) = Np$$

The variance is:

$$Var(X) = Np(1-p)$$

For large *N*, the binomial distribution can be approximated by the normal distribution



Binomial Distribution

Example:
In an instant lottery with 20% winning tickets, if X is equal to the number of winning tickets among N = 8 that are purchased, the probability of purchasing 2 winning tickets

$$f(2) = P(X = 2) = \binom{8}{2} (0.2)^2 (0.8)^6 = .2936$$

The distribution of the random variable X is B(8,.20)

Poisson Experiment

A Poisson experiment is a statistical experiment with the following properties:

- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (μ) that occurs in a specified region is known.
- The probability that a success will occur is proportional to the size of the region.
- The probability that a success will occur in an extremely small region is virtually zero.

Poisson Distribution

Describes various discrete phenomena (those that may happen 0, 1, 2, 3, . times during a given period of time or in a given area) whenever the probability of the phenomenon happening is constant in time or space.

Examples of events that can be modeled as Poisson distributions include:

- The number of cars that pass through a certain point on a road during a given period of time.
- The number of phone calls at a call center per minute.
- The number of plants in a square meter
- The number of roadkill found per unit length of road.
- The number of diatoms per ml of water

Poisson Distribution

Expresses the probability of a number of events occurring in a fixed region or period of time if these events occur with a known average rate, and are independent of the last event.

The probability that there are exactly *m* occurrences (*m* being a non-negative integer, *m* = 0, 1, 2, ...) is

$$\Pr(X = m) = \frac{e^{-\mu}\mu^{m}}{m!}$$

Where

e is the base of the natural logarithm (*e* = 2.71828...),

μ is a positive real number, equal to the expected number of occurrences that occur during the given interval or region.

Poisson Distribution

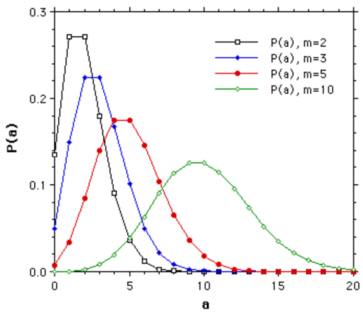
$$\Pr(X = m) = \frac{e^{-\lambda}\lambda^{m}}{m!}$$

The expected value of a Poisson distributed random variable is equal to (μ) and so is its variance.

$$E(X) = \lambda$$
$$Var(X) = \lambda$$

Described by $X \sim \text{Poisson}(\lambda)$

Poisson Distribution



As the mean increases the distribution moves to the right and broadens

Example: Cumulative Poisson probability

On average the number of moose seen while canoeing on the Allagash River in Maine per day in June is 5.

What is the probability that canoists will see fewer than four moose on a June day?

This is a Poisson experiment in which we know the following:

$\mu = 5$; since 5 moose are seen per day, on average.

x = 0, 1, 2, or 3; since we want to find the likelihood that canoeists will see fewer than 4 moose, e.g. we want the probability that they will see 0, 1, 2, or 3 moose.

e = 2.71828;

Example: Cumulative Poisson probability

Need to find the probability that canoeists will see 0, 1, 2, or 3 moose.

Thus, we need to calculate the sum of four probabilities:
P(0; 5) + P(1; 5) + P(2; 5) + P(3; 5)

$P(x \leq 3, 5) = P(0; 5) + P(1; 5) + P(2; 5) + P(3; 5)$

$P(x \leq 3, 5) = [\ (e^{-5})(5^0) / 0!\ ] + [\ (e^{-5})(5^1) / 1!\ ] + [\ (e^{-5})(5^2) / 2!\ ] + [\ (e^{-5})(5^3) / 3!\ ]$

$P(x \leq 3, 5) = [\ (0.006738)(1) / 1\ ] + [\ (0.006738)(5) / 1\ ] + [\ (0.006738)(25) / 2\ ] + [\ (0.006738)(125) / 6\ ]$

$P(x \leq 3, 5) = [\ 0.0067\ ] + [\ 0.03369\ ] + [\ 0.084224\ ] + [\ 0.140375\ ]$

$P(x \leq 3, 5) = 0.2650$

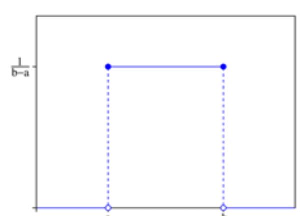
Uniform Distribution

A random variable is equally likely to take any value in an interval a,b

Pr{X=x}=1/(b-a)

E(X)=(a+b)/2

Var={(b-a)^2/12



Described by X ~ U(a,b)

Uniform Distribution

The first two moments of the uniform distribution

$$m_1 = \frac{a+b}{2} \qquad m_2 = \frac{a^2+ab+b^2}{3}$$

For a random variable following the uniform distribution:

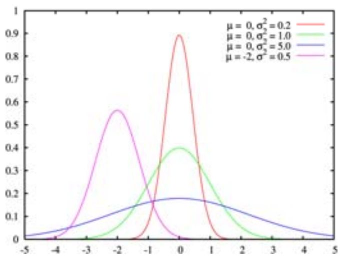
the expected value is $m_1 = (a + b)/2$

the variance is $m_2 - m_1^2 = (b - a)^2/12$.

Normal Distribution

The **normal distribution, also called Gaussian distribution (although** Gauss was not the first to work with it), is an extremely important probability distribution in many fields. It is a family of distributions of the same general form, differing in their *location* and *scale* parameters: the mean ("average") and standard deviation ("variability"), respectively.

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/(2\sigma^2)}$$



Described by $X \sim N(\mu, \sigma^2)$

Standard Normal Distribution

The standard normal distribution has zero mean and unit variance.

Described by $X \sim N(0,1)$

The area under the standard normal curve is sometimes referred to as the error function: $\Phi(x)$

Gamma Distribution

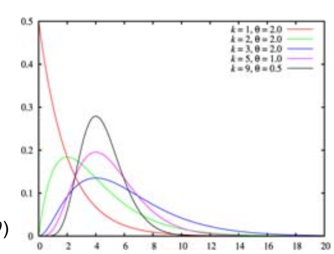
Used to model continuous variables that are always positive and have skewed distributions

$$f(x;k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \quad for\ x > 0$$

where $k > 0$ is the shape parameter and $\theta > 0$ is the scale parameter.

$$E(X) = k\theta$$

$$VAR(X) = k\theta^2$$



Described by $X \sim Gamma(k, \theta)$

Gamma Distribution – special cases

The Gamma distribution represents a family of shapes. The fundamental shapes are characterized by the values of k

When k is less than one, the Gamma distribution has an exponential shape.

When k is greater than one, the Gamma distribution assumes a mounded (unimodal), but skewed shape. The skewness reduces as the value of k increases.

The exponential distribution is a special case of the gamma distribution when k = 1

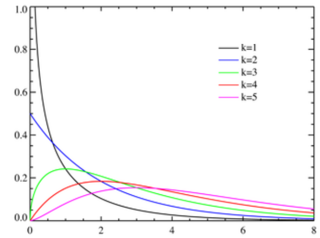
Described by $X \sim \text{Exponential}(\theta)$

Gamma Distribution – special cases

The Chi Square distribution is a special case of gamma distribution

Described by $X \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2}) = X \sim \chi_n^2$

The chi-square distribution has one parameter: *n* - a positive integer that specifies the number of degrees of freedom (i.e. the number of X_i's – independent random variables with mean zero and variance 1)



commonly used in goodness of fit tests for an observed distribution to a theoretical one

Useful WebSites on probability distributions

http://mathworld.wolfram.com/BinomialDistribution.html

http://mathworld.wolfram.com/PoissonDistribution.html

http://mathworld.wolfram.com/NormalDistribution.html

http://www.stats.gla.ac.uk/steps/glossary/probability_distributions.html#randvar

http://en.wikipedia.org/wiki/Probability_distribution