Overview of Statistical Concepts

Lecture 2

SIE 512 Spatial Analysis

Fall 2009

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 probably than others

A stochastic process is described by random variables and their probability

Events and sample space

The **sample space** is the set of all possible outcomes of a stochastic process.

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

Distinguish elementary and compound events

Random Variables

A random variable X, is a label attached to a 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. If a random variable can take only a finite number of distinct values, it must be discrete.

Random Variables and Probability Distributions

Probability is the chance of a random event happening.

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

Probability Concepts

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

Number of Symmetric Ways Definition

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

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 Concepts

- **Non-frequentist subjective view**

  Takes into account prior knowledge about events

  Bayes' theorem

- **Axiomatic approach – 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.

Discrete distributions are characterized by a probability mass function, *p* such that

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

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

Continuous probability distribution

Continuous distributions can be characterized by a probability density function

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(t)dt$$

Probability that a random variable takes a value in a range

If y is discrete

$$\sum_{y=a}^{b} f_Y(y)$$

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

If y is continuous

$$\int_a^b f_Y(y)dy$$

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

Cumulative Distribution Function

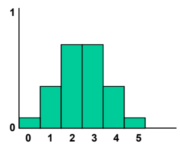
$$F_Y(y) = \begin{cases} \sum_{u=-\infty}^{y} f_Y(u) & \text{If Y is discrete} \\ \int_{-\infty}^{y} f_Y(u)\,\mathrm{d}u & \text{If Y is continuous} \end{cases}$$

Probability of Y taking any value less than or equal to y

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

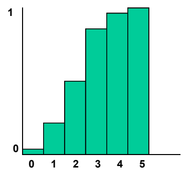
Discrete case : A random variable X has the following probability distribution p(x)

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

$$P(A \; and \; B) = P(A \cap B) = P(A)P(B)$$

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

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

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

$$P(A and\ B) = P(A)P(B)$$

And so conditional probability is:

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

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*].

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 probabilities of obtaining the amounts a_1, a_2, a_3,....a_n 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 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$

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

Expectation of a random variable

$$E(Y) = \begin{cases} \sum_{y=-\infty}^{\infty} y f_Y(y) & \text{discrete} \\ \int_{-\infty}^{\infty} y f_Y(y) dy & \text{continuous} \end{cases}$$

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.

$$Var(X) = E((X- E(X))^2$$

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

Covariance

Measures the correspondence or covariation of two random variables together

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

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

Skewness

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

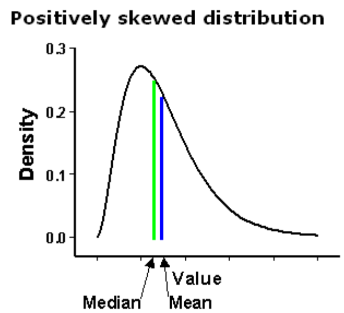
$$\gamma = \frac{E(x - E(x))^3}{\sigma^3} \qquad Skew = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \overline{x}}{\sigma}\right]^3$$

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



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.

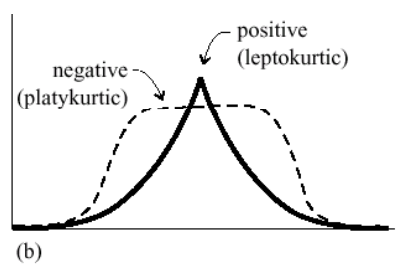
$$k = \frac{E(x - E(x))^4}{\sigma^4} - 3 \qquad k = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \overline{x}}{\sigma}\right]^4 - 3$$

Kurtosis = 0 … Standard Normal Distribution

Kurtosis > 0 … positive ("leptokurtic“)

Kurtosis < 0 … negative ("platykurtic“)

Kurtosis



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, takes value 1 with probability *p* and value 0 with probability *q* = 1 − *p*

One population parameter *p*

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

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

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

$$= p * 1^2 + (1-p) * 0^2 - p^2 = p(1-p)$$

$$= pq$$

Described by $X \sim Be(p)$

Binomial Distribution

Discrete probability distribution of the number of successes in a sequence of *n* independent yes/no experiments, each of which yields success with probability *p*.

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

1- p is probability of a failure

$$\Pr\{X = m\} = \frac{n!}{(n-m)!m!} p^m (1-p)^{n-m}$$

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

Binomial Distribution

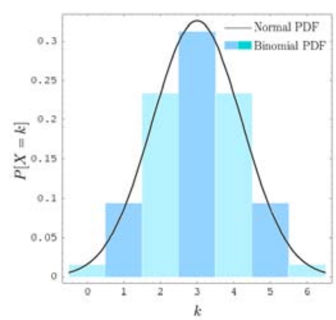
The expected value is:

$$E(X) = np$$

The variance is:

$$\mathrm{Var}(X) = np(1-p)$$

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



Binomial Distribution

Example 1
In the 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(8.02)

Poisson Distribution

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

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

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

Where

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

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

Poisson Distribution

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

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

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

Describedby X ~ Poisson(u)

Poisson Distribution

The Poisson distribution arises in connection with Poisson processes.

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.

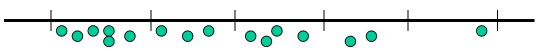
Examples of events that can be modeled as Poisson distributions include:

- 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 times a web server is accessed per minute
- The number of roadkill found per unit length of road.

Poisson Distribution

Find the probability of x successes during a time interval of length T

Divide an interval into n equal parts of length Δt so that T=n* Δt



Poisson Distribution

1. The probability of success in an interval is independent for all intervals.
2. The probability of success during a very small interval Δt is
   $\alpha * \Delta t$
3. The probability of more than one success in a sufficiently small interval Δt is essentially 0.

$b(x;n,p)$ with $n = \frac{T}{\Delta t}$ and $p = \alpha * \Delta t$

In the limit of the number of trials becoming large, the resulting distribution is a Poisson distribution

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$n \to \infty$ and $p \to 0$ while $np = \lambda$ remains constant

$$b(x;n,p) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x (1-\frac{\lambda}{n})^{n-x}$$

$$= \frac{n(n-1)(n-2)..(n-x+1)}{x!\,n^x}(\lambda)^x(1-\frac{\lambda}{n})^{n-x}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)..\left(1-\frac{x-1}{n}\right)}{x!}(\lambda)^x(1-\frac{\lambda}{n})^{n-x}$$

let $n \to \infty$

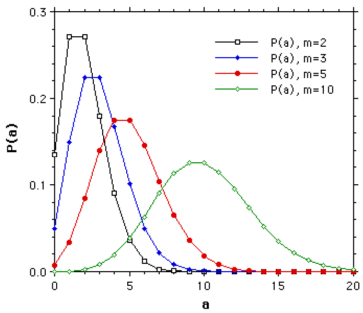
$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)*...*\left(1-\frac{x-1}{n}\right) \to 1$$

and

$$\left(1-\frac{\lambda}{n}\right)^{n-x} = \left[\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}}\right]^{\lambda}\left(1-\frac{\lambda}{n}\right)^{-x} \to e^{-\lambda}$$

Poisson Distribution $\frac{\lambda^x e^{-\lambda}}{x!}$ $for\ x = 0,1,2,...$

Poisson Distribution



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

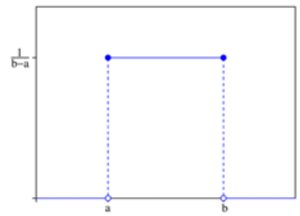
Uniform Distribution

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 $X \sim U(a,b)$

Uniform Distribution

The first two moments of the uniform distribution

$$m_1 = \frac{a+b}{2},$$

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

For a random variable following the uniform distribution:

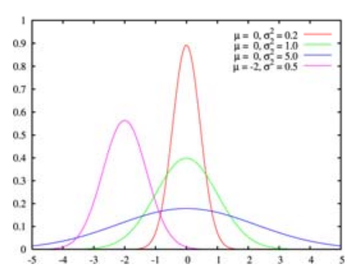
the expected value is m_1 = (*a* + *b*)/2
the variance is $m_2 - m_1^2 = (b - a)^2/12$.

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



Gamma Distribution

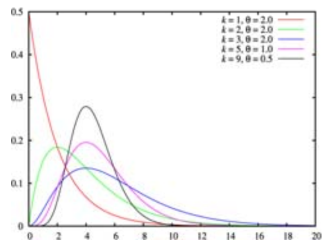
Describes the time until *n* consecutive rare random events occur in a process with no memory

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

where *k* > 0 is the shape parameter and θ > 0 is the scale parameter of the gamma distribution.

$E(X) = k\theta$

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



Described by X ~ Gamma(k, θ)

Gamma Distribution – special cases

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

Described by X ~ Exponential(θ)

The Chi Square distribution is a special case of gamma distribution where

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

Useful WebSites for Review

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

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

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

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