Summer School in Statistics for Astronomers & Physicists
June 5-10, 2005

Session on 'Statistical Inference for Astronomers'

Laws of Probability, Bayes' theorem, and the Central Limit Theorem

Donald Richards

Department of Statistics
Center for Astrostatistics
Penn State University

Deterministic experiment: One whose outcome is determined entirely by the conditions under which the experiment is performed.

Mathematical models for familiar deterministic experiments:

$s = ut + \frac{1}{2}at^2$: The distance traveled in time t by an object with initial velocity u and constant acceleration a:

Kepler's laws for planetary motion

$F = ma$: Newton's Second Law,

$V = kT$: Charles' gas law for V, the volume of a gas sample with t the gas temperature on the absolute scale

A *random experiment* is one which
(a) can be repeated indefinitely under essentially unchanged conditions, and
(b) whose outcome cannot be predicted with *complete certainty*, although all its possible outcomes can be described completely.

Toss a coin four times and observe the number of heads obtained.

Draw four cards from a fair deck and count the number of queens obtained.

The number of particles emitted by a radioactive substance in a one-second time period.

The annual number of radio flares observed on the Algol-type systems β Persei and δ Librae.

Sample space, S: The set of all possible outcomes of our random experiment.

Event: A subset of the sample space.

If A and B are events then:

$A \cup B$ is the event which occurs iff A or B occur.

$A \cap B$ is the event which occurs iff A and B occur.

$\bar{A}$ is the event which occurs iff A does not occur.

S: The event which always occurs

$\emptyset$: The event which never occurs

Mutually exclusive events: $A \cap B = \emptyset$

Let A and B be two events associated with $\mathcal{E}$, a random experiment. We repeat $\mathcal{E}$ n times. Let n_A and n_B be the number of times that A and B occur among all n repetitions.

The *relative frequency* of A is

$$f_A = \frac{n_A}{n},$$

the proportion of repetitions resulting in A.

$0 \leq f_A \leq 1$

$f_A = 1$ iff A occurred on each repetition.

$f_A = 0$ iff A never occurred on any repetition.

$f_{A \cup B} = f_A + f_B$ if A, B are mutually exclusive.

As $n \to \infty$, f_A "converges" to some number

This number is denoted by $P(A)$ and is called "the probability of A"

The Axioms of Probability

Let $\mathcal{E}$ be a random experiment with sample space S. With each subset A of S, we associate a number $P(A)$, the probability of A, such that:

(1) $0 \leq P(A) \leq 1$

(2) $P(S) = 1$

(3) If A, B are mutually exclusive then

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

(4) If $A_1, A_2, \ldots, A_k, \ldots$ are pairwise mutually exclusive events then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

Note that (4) $\Longrightarrow$ (3), but not *vice versa*.

Some simple properties of probability:

$P(\emptyset) = 0$

$P(\bar{A}) = 1 - P(A)$

If $A \subseteq B$ then $P(A) \leq P(B)$

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

The Inclusion-Exclusion Formula:

$$
\begin{aligned}
P(A \cup B \cup C) = \; & P(A) + P(B) + P(C) \\
& - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
& + P(A \cap B \cap C)
\end{aligned}
$$

Finite sample space: S consists of a finite number of "elementary" outcomes, $S = \{a_1, \ldots, a_m\}$

Equally likely outcomes: Each elementary outcome has the same probability of occurring

$$P(A) = \frac{\text{No. of ways in which } A \text{ can occur}}{\text{No. of possible outcomes}}$$

Toss a fair coin twice.

$P(\text{Two heads}) = \frac{1}{4}$

$P(\text{At least one head}) = 1 - P(\text{Two tails}) = \frac{3}{4}$

Reminder:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

is the number of ways in which k objects can be chosen from a set of n objects.

From a "fair" deck of 52 playing cards, we randomly draw:

1 card; P(We draw a queen) $=$

2 cards: P(We draw 2 queens) $=$

3 cards: P(We draw 3 queens) $=$

4 cards: P(We draw 4 queens) $=$

If someone draws 3 queens from a deck then it is implausible that the deck is fair.

Conditional Probability

A and B: subsets of S, the sample space

$P(A|B)$: "The probability of A given B"

The *conditional probability* of A, given that B has already occurred, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0$$

$P(A|B)$ is undefined if $P(B) = 0$

$P(A|B) \not\equiv P(A)$ because of the additional information that B has already occurred.

Choose a day at random

$A = \{$It will snow on that day$\}$

$B = \{$That day's low temperature is $80^\circ F\}$

$P(A) \neq 0$, however $P(A|B) = 0$

Example: An urn contains 20 blue and 80 red balls. Mary and John each choose a ball at random from the urn. Let

$B = \{\text{Mary chooses a blue ball}\}, \quad P(B) = \frac{20}{100}$

$A = \{\text{John chooses a blue ball}\}, \quad P(A|B) = \frac{19}{99}$

The additional information that B has *already* occurred decreases the probability of A

$$P(A \cap B) = \frac{\text{No. of ways of choosing 2 blue balls}}{\text{No. of ways of choosing 2 balls}} = \frac{\binom{20}{2}}{\binom{100}{2}} = \frac{19}{495}$$

Note that

$$\frac{P(A \cap B)}{P(B)} = \frac{19}{495} \div \frac{20}{100} = \frac{19}{99} = P(A|B)$$

$P(A|B)$ satisfies the axioms of probability: If $P(B) \neq 0$ then

(1) $0 \leq P(A|B) \leq 1$

(2) $P(S|B) = 1$

(3) If A_1 and A_2 are mutually exclusive then

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$$

(4) If $A_1, A_2, \ldots, A_k, \ldots$ are pairwise mutually exclusive events then

$$P\Big(\bigcup_{i=1}^{\infty} A_i|B\Big) = \sum_{i=1}^{\infty} P(A_i|B).$$

Also,

$P(\emptyset|B) = 0$

$P(\bar{A}|B) = 1 - P(A|B)$

If $A_1 \subseteq A_2$ then $P(A_1|B) \leq P(A_2|B)$

$$P(A_1 \cup A_2|B)$$
$$= P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$$

The Multiplication Theorem: If $P(B) \neq 0$ then

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

Proof:

$$P(A|B)P(B) = \frac{P(A \cap B)}{P(B)}P(B) = P(A \cap B).$$

Repeat the argument: If $P(A_1 \cap A_2 \cap A_3) \neq 0$ then

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

Events $B_1, \ldots, B_k$ form a *partition* of the sample space S if

(1) $B_i \cap B_j = \emptyset$ for all $i \neq j$ (pairwise mutually exclusive)

(2) $B_1 \cup B_2 \cup \cdots \cup B_k = S$, the full sample space

(3) $P(B_i) \neq 0$ for all $i = 1, \ldots, k$

The Law of Total Probability: Let A be any event and $B_1, \ldots, B_k$ be a partition. Then

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_k)P(B_k)$$

By means of a simple Venn diagram,

$$P(A) = P(A \cap B_1) + \cdots + P(A \cap B_k)$$

By the Multiplication Theorem,

$$P(A \cap B_i) = P(A|B_i)P(B_i)$$

Example: An urn contains 20 blue and 80 red balls. Two balls are chosen at random and without replacement. Calculate $P(A)$ where

$$A = \{\text{The second ball chosen is blue}\}$$

Let $B = \{\text{The first ball chosen is blue}\}$

Then $\bar{B} = \{\text{The first ball chosen is red}\}$

$B, \bar{B}$ are a *partition* of the sample space

$B \cap \bar{B} = \emptyset$, $B \cup \bar{B} = S$, $P(B) = \frac{20}{100}$, $P(\bar{B}) = \frac{80}{100}$

By the Law of Total Probability,

$$\begin{aligned} P(A) =& P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}) \\ =& \frac{19}{99} \cdot \frac{20}{100} + \frac{20}{99} \cdot \frac{80}{100} \\ =& \frac{20}{100} \end{aligned}$$

Notice that $P(A) = P(B) = \frac{20}{100}$

Two cabinets, I and II, each have 2 drawers. I contains 2 silver coins in each drawer. II contains 2 silver coins in one drawer and 1 gold coin in the other drawer. Huygens rolls a fair die and chooses I if the die rolls 1, 2, 3, or 4; otherwise, he chooses II. Having chosen a cabinet, he chooses a drawer at random and opens it.

Find the probability that the opened drawer contains 2 silver coins:

$A = \{$The opened drawer contains 2 silver coins$\}$

$B = \{$Chris chooses cabinet $I\}$
$\bar{B} = \{$Chris chooses cabinet $II\}$

$B, \bar{B}$ are a partition of the sample space:

$B \cap \bar{B} = \emptyset$, $B \cup \bar{B} = S$, $P(B) = \frac{4}{6}$, $P(\bar{B}) = \frac{2}{6}$

By the Law of Total Probability,

$$
\begin{aligned}
P(A) &= P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}) \\
&= 1 \cdot \frac{4}{6} + \frac{1}{2} \cdot \frac{2}{6} \\
&= \frac{5}{6}
\end{aligned}
$$

A more difficult problem

Given that the opened drawer contains 2 silver coins, find $P(\bar{B}|A)$, the probability that cabinet II was chosen.

Bayes' Theorem: Suppose that

1. $B_1, \ldots, B_k$ is a partition, and

2. A is an event with $P(A) \neq 0$.

Then

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + \cdots + P(A|B_k)P(B_k)}$$

1. Definition of conditional probability:

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(A \cap B_1)}{P(A)}$$

2. The multiplication theorem:

$$P(A \cap B_1) = P(A|B_1)P(B_1)$$

3. The Law of Total Probability:

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_k)P(B_k)$$

Return to the drawer problem

$P(A|B) = 1$, $P(B) = \frac{4}{6}$

$P(A|\bar{B}) = \frac{1}{2}$, $P(\bar{B}) = \frac{2}{6}$

By Bayes' theorem,

$$\begin{aligned} P(\bar{B}|A) &= \frac{P(A|\bar{B}) \cdot P(\bar{B})}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} \\ &= \frac{\frac{1}{2} \cdot \frac{2}{6}}{1 \cdot \frac{4}{6} + \frac{1}{2} \cdot \frac{2}{6}} \\ &= \frac{1}{5} \end{aligned}$$

Homework Assignment

Joseph Bertrand, *Calcul des Probabilités*, 1889

Bertrand's Box Paradox: A box contains three drawers. In one drawer there are two gold coins; in another drawer there are two silver coins; and in the third drawer there is one silver coin and one gold coin. A drawer is selected at random, and then a coin is selected at random from that drawer. Given that the selected coin is gold, what is the probability that the other coin is gold?

Three suspected burglars, Curly, Larry, and Moe, are held incommunicado in a distant country by a ruthless jailer. The three are told that one of them has been chosen at random to be imprisoned and that the other two will be freed, but they are not told whom. Curly asks the jailer who will be imprisoned, but the jailer declines to answer. Instead the jailer tells him that Larry will be freed. Given that Larry is to be freed, what is the probability that Curly will be imprisoned?

Hint: Search the Internet.

Independence: Events A and B are *independent* if $P(A \cap B) = P(A)P(B)$

Motivation: $P(A|B) = P(A)$, $P(B|A) = P(B)$

Theorem: If $\{A, B\}$ are independent then so are $\{A, \bar{B}\}$, $\{\bar{A}, B\}$, and $\{\bar{A}, \bar{B}\}$

Example: Roll two fair dice. Let
$A = \{$The first die shows an even number$\}$
$B = \{$The second die shows an odd number$\}$
$C = \{$The total rolled is an even number$\}$

$\{A, B\}$, $\{B, C\}$, $\{A, C\}$ are independent pairs

$P(A \cap B \cap C) = 0$; $P(A)P(B)P(C) = \frac{1}{8}$

Events A, B, C are *mutually independent* if:

$$P(A \cap B) = P(A)P(B),$$

$$P(A \cap C) = P(A)P(C),$$

$$P(B \cap C) = P(B)P(C),$$

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Pairwise independence does not imply mutual independence

Random variable: A numerical value calculated from the outcome of a random experiment

A function from S, the sample space, to $\mathbb{R}$, the real line

Discrete random variable: One whose possible values are a discrete set

The number of collisions daily between the ISS and orbital debris

Continuous random variable: One whose values form an interval

The length of time a "shooting star" is visible in the sky

Let X be the number of heads obtained among 2 tosses of a (fair) coin

The *possible values* of X are: $0, 1, 2$

The *probability distribution* of X is a listing of its possible values and the corresponding probabilities

Toss a coin twice:

$P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$, $P(X = 2) = \frac{1}{4}$

Notice that $\sum_{x=0}^{2} P(X = x) = 1$

Billingsley, *Probability and Measure*: The infinite coin toss

Bernoulli trial: A random experiment with only two possible outcomes, "success" or "failure"

$p \equiv P(\text{success})$, $q \equiv 1 - p = P(\text{failure})$

Perform n independent repetitions of a Bernoulli trial

X: No. of successes among all n repetitions

Possible values of X: $0, 1, 2, \ldots, n$

Probability distribution of X:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

The binomial distribution: $X \sim B(n, p)$

Every probability distribution satisfies:

1. $0 \le P(X = x) \le 1$

2. $\sum_{x=0}^{2} P(X = x) = 1$

A. Mészáros, "On the role of Bernoulli distribution in cosmology," *Astron. Astrophys.*, 328, 1-4 (1997).

n uniformly distributed points in a region of volume $V = 1$ unit

X: No. of points in a fixed region of volume p

X has a binomial distribution, $X \sim B(n, p)$

M. L. Fudge, T. D. Maclay, "Poisson validity for orbital debris ..." *Proc. SPIE*, 3116 (1997) 202-209, Small Spacecraft, Space Environments, and Instrumentation Technologies

ISS ... at risk from orbital debris and micrometeorite impact

fundamental assumption underlying risk modeling: orbital collision problem can be modeled using a Poisson distribution

assumption found to be appropriate based upon the Poisson ... as an approximation for the binomial distribution and ... that is it proper to physically model exposure to the orbital debris flux environment using the binomial.

The geometric distribution:

Perform independent repetitions of a Bernoulli trial

X: No. of trials needed to obtain one success

Possible values of X: $1, 2, 3, \ldots$

$P(X = k) = q^{k-1}p, \quad k = 1, 2, 3, \ldots$

Geometric series: $\sum\limits_{x=0}^{\infty} q^{k-1} = \frac{1}{1-q} = \frac{1}{p}$

The negative binomial distribution:

Perform independent repetitions of a Bernoulli trial

X: No. of trials needed to obtain r successes

Possible values of X: $r, r+1, r+2, \ldots$

$$P(X = k) = \binom{k-1}{r-1} q^{k-r} p^r$$

Neyman, Scott, and Shane (1953, ApJ): Counts of galaxies in clusters

ν: The number of galaxies in a randomly chosen cluster

Basic assumption: ν follows a negative binomial distribution

The Poisson distribution: $Y \sim Poisson(\lambda)$ if

$$P(Y = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

X: binomial random variable

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

If $n \to \infty$, $p \to 0$, and $np \to \lambda$ then

$$\lim P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Example: X is the number of typographical errors on a newspaper page

X has a binomial distribution, $X \sim B(n, p)$

n is very large, p is very small

The Poisson distribution is a good approximation to the distribution of X

Expected value of X: Average value of X based on many repetitions of the experiment,

$$\mu \equiv E(X) = \sum_{\text{all } k} k \cdot P(X = k)$$

If $X \sim B(n, p)$ then $E(X) = np$

If X is geometric then $E(X) = 1/p$

Standard deviation: A measure of average fluctuation of X around μ

Variance of X:

$$\text{Var}(X) = E(X - \mu)^2 = \sum_{\text{all } k} (k - \mu)^2 \cdot P(X = k)$$

$$S.D.(X) = \sqrt{\text{Var}(X)}$$

If $X \sim B(n, p)$ then $\text{Var}(X) = npq$

Every continuous random variable X has a probability density function f.

The three important properties of f:

$f(x) \geq 0$ for all x

$\int_{-\infty}^{\infty} f(x)\, dx = 1$

$P(X \leq t) = \int_{-\infty}^{t} \phi(x)\, dx$ for all t

Similar to discrete random variables,

$\mu \equiv E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$

$\mathrm{Var}(X) = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx$

The function

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty$$

is a probability density function:

1. $\phi(x) \geq 0$ for all x;

2. $\int_{-\infty}^{\infty} \phi(x)\, dx = 1$

Standard normal distribution: A continuous random variable Z having probability density function $\phi(x)$, i.e., for all t,

$$P(Z \leq t) = \int_{-\infty}^{t} \phi(x)\, dx$$

Homework: Show that $E(Z) = 0$, $\text{Var}(Z) = 1$

Normal distribution $\equiv$ Gaussian distribution

Normal Approximation to the Binomial Distn.:

DeMoivre (1733), $p = 1/2$
Laplace (1812), general p

If $X \sim B(n, p)$ then

$$\lim_{n\to\infty} P\left(\frac{X - np}{\sqrt{npq}} \le t\right) = P(Z \le t)$$

Proof: 1. Compute the Fourier transform of the density function of $(X - np)/\sqrt{npq}$

2. Find the limit of this F.t. as $n \to \infty$

3. Check that the limiting F.t. is the F.t. of the density function of Z

Intuition: If $X \sim B(n, p)$ and n is large then

$$\frac{X - np}{\sqrt{npq}} \approx Z$$

Micrometeorite: piece of rock less than 1 mm in diameter (similar to a fine grain of sand); moving at up to 80 km/sec. (48 miles/sec.); difficult to detect; no practical way to avoid them

Claim: "The probability that the ISS will receive a micrometeorite impact during its (30-year) mission is nearly 100%"

n: No. of days in 30 years, $30 \times 365 = 10,950$

p: Probability that a micrometeorite hits the ISS on a randomly chosen day

X: No. of micrometeorite impacts on the ISS over its designed lifetime of 30 years

$X \sim B(n, p)$: a model for the distribution of the no. of impacts

Calculate $P(X \geq 1)$

Poisson approximation: n is large, p is small
$X \approx Poisson(\lambda)$, $\lambda = np$

$$P(X \geq 1) = 1 - P(X = 0) \simeq 1 - e^{-\lambda}$$

$n = 10,950$, $p = 10^{-2}$, $\lambda = 109.5$

Normal approximation:

$$\begin{aligned} P(X \geq 1) &= P\left(\frac{X - np}{\sqrt{npq}} \geq \frac{1 - np}{\sqrt{npq}}\right) \\ &\simeq P\left(Z \geq \frac{1 - 109.5}{\sqrt{108.405}}\right) \\ &= P(Z \geq -10.42) \\ &= 1 \end{aligned}$$

Important continuous random variables

Normal, chi-square, t-, and F-distributions

Z: standard normal distribution

Z^2: χ_1^2 (chi-square with 1 degree of freedom)

t-distribution: $\frac{N(0,1)}{\sqrt{\chi_p^2/p}}$

F-distribution: $\frac{\chi_p^2/p}{\chi_q^2/q}$

Las Vegas or Monte Carlo: $n = 10^6$ people each play a game (craps, trente-et-quarante, roulette, ...)

Each of the 10^6 repetitions is a Bernoulli trial

Casinos fear of too many (or too few) winners

X: the number of winners among all n people

p: probability of success on each repetition

Probability distribution: $X \sim B(n, p)$

Markov's Inequality: For any random variable X and $t > 0$,

$$P(X \geq t) \leq \frac{E(X)}{t}$$

We can compute the maximum probability of having too many winners

Craps game: $p = .492929292... = \frac{244}{495}$

$E(X) = np = 492,929.29$

Markov's inequality: $P(X \geq t) \leq \frac{492{,}929.29}{t}$

$t = 500,000$: $\quad P(X \geq 500,000) \leq 0.98585...$

Chebyshev's Inequality: For any $t > 0$,

$$P\left(\frac{|X - \mu|}{S.D.(X)} < t\right) \geq 1 - \frac{1}{t^2}$$

Craps game: $S.D.(X) = \sqrt{npq} = 499.95$

$$P\left(\frac{|X - 492,929.29|}{499.95} < t\right) \geq 1 - \frac{1}{t^2}$$

$t = 8$: $\quad P(488929 < X < 496929) \geq .984375$

Most of the distribution of values of X is concentrated around μ

The (weak) Law of Large Numbers

$X_1, X_2, \ldots$: independent, identically distributed random variables

μ: The mean of each X_i

$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$: The sample mean of the first n observations

For $t > 0$,

$$P\left(|\bar{X} - \mu| \geq t\right) \to 0 \qquad \text{as } n \to \infty$$

$X_1, \ldots, X_n$: independent, identically distributed random variables

μ: The mean (expected value) of each X_i

σ^2: The variance of each X_i

$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$: The average of the X_i's

The Central Limit Theorem: If n is large then

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx Z$$

where $Z \sim N(0, 1)$.

Proof: 1. Compute the Fourier transform of the density function of $(X_1 + \cdots + X_n - n\mu)/\sigma\sqrt{n}$

2. Find the limit of this F.t. as $n \to \infty$

3. Check that the limiting F.t. is the F.t. of the density function of Z

Historical note

The Central Limit Theorem was originally stated and proved by Laplace (Pierre Simon, the Marquis de Laplace).

Francis Galton, *Natural Inheritance*, 1889

"I know of scarcely anything so apt as to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error.' The Law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of unreason."