Probability and Statistics

Jane Bae

Stanford University

hjbae@stanford.edu

September 16, 2014

Overview

1. Probability Concepts
   - Probability Space
   - Combinatorics
   - Random Variables and Distributions
   - Independence and Conditional Probability

Outline

1. Probability Concepts
   - Probability Space
   - Combinatorics
   - Random Variables and Distributions
   - Independence and Conditional Probability

Probability Space

A probability space consists of three parts:

- A sample space, Ω,
- A set of events $\mathcal{F}$,
- A probability P.

Sample Space

A sample space is the set of all possible outcomes.

- Single coin flip:

$$\Omega = \{H, T\}.$$

- Double coin flip:

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}.$$

- Single dice roll:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

- Double dice roll:

$$\Omega = ?$$

Set of Events

A event is a subset of the sample space.

- Single coin flip: (Event that a coin lands head)

$$A = \{H\}.$$

- Double coin flip: (Event that first coin lands head)

$$A = \{(H, H), (H, T)\}.$$

- Single dice roll: (Event that the dice roll is less than 4)

$$A = \{1, 2, 3\}.$$

- Double dice roll: (Event that the dice roll add up to less than 4)

$$A = \{(1, 1), (1, 2), (2, 1)\}$$

- The collection of all such events we would like to consider are called the set of events, $\mathcal{F}$.
- $\mathcal{F}$ is a σ-algebra.
    - $\Omega \in \mathcal{F}$.
    - $\mathcal{F}$ is closed under complementation.
    - $\mathcal{F}$ is closed under countable unions.

Probability

A set function on the set of events such that

- $0 \leq P(A) \leq 1$ for all events A.
- $P(\Omega) = 1$.
- For each sequence of mutually disjoint events $A_1, A_2, A_3, \ldots$,

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

(countable additivity).

Probability

- Fair coin

$$P(\emptyset) = 0$$
$$P(\{H\}) = 1/2$$
$$P(\{T\}) = 1/2$$
$$P(\{H, T\}) = 1$$

- Fair die

$$P(\emptyset) = 0$$
$$P(\{1\}) = 1/6$$
$$P(\{2\}) = 1/6$$
$$P(\{1, 2, 3\}) = 1/2$$

Some Properties (Good Exercise Problems)

- $P(\emptyset) = 0$.
- $P(A^c) = 1 - P(A)$.
- If $A \subseteq B$, then $P(A) \leq P(B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- For each sequence of events $A_1, A_2, A_3, \ldots$ (not necessarily disjoint),

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

Continuous Probability Space

- Similar idea as discrete probability space.
- Example
  - Arrival time of trains passing with approxmiately equal intervals.
  - Chord length of a circle of radius R with chord randomly selected.

Probability Space

A probability space consists of three parts:

- A sample space, Ω,
- A set of events $\mathcal{F}$,
- A probability P.

Sample Space

A sample space is the set of all possible outcomes.

- Random number between 0 and 1:

$$\Omega = [0, 1).$$

- Height of 10-year-olds (in cm):

$$\Omega = [0, \infty).$$

Set of Events

A event is a subset of the sample space.

- Random number between 0 and 1: (Event that it is less than 0.5)

$$A = [0, 0.5)$$

- Height of 10-year-olds (in cm): (Event that a 10-year-old is taller than 150cm)

$$A = [150, \infty)$$

Outline

1. Probability Concepts
   - Probability Space
   - Combinatorics
   - Random Variables and Distributions
   - Independence and Conditional Probability

Permutations

Permutation relates to the act of rearranging members of a set into a particular sequence or order.

- Ex) Six permutations of the set $\{1,2,3\}$
- Ex) Ways to sit around a table

k-permutation of n are the different ordered arrangements of a k-element subset of an n-set, usually denoted by

$$ {}_nP_k = \frac{n!}{(n-k)!}. $$

Combinations

Combination is a way of selecting members from a grouping, such that the order of selection does not matter.

- Ex) Drawing marbles out of a box

k-combination of n are the different groupings of a k-element subset of an n-set, usually denoted by

$$_nC_k = \frac{n!}{k!(n-k)!}.$$

Outline

1. Probability Concepts
   - Probability Space
   - Combinatorics
   - Random Variables and Distributions
   - Independence and Conditional Probability

Random Variables

A random variable is a **function** from a sample space to a real number.

$$X : \Omega \to \mathbb{R}$$

Example:

- Single coin flip

$$X(H) = 1,$$
$$X(T) = -1.$$

- Double dice roll

$$X(i,j) = i$$
$$Y(i,j) = j$$
$$Z(i,j) = i + j.$$

Distributions of Discrete Random Variables

A discrete random variable X assumes values in discrete subset S of $\mathbb{R}$. The distribution of a discrete random variable is completely described by a **probability mass function** $p_X : \mathbb{R} \to [0,1]$ such that

$$P(X = x) = p_X(x).$$

Example:

- Bernoulli: $X \sim Ber(p)$ if $X \in \{0, 1\}$ and
$$p_X(1) = 1 - p_X(0) = p.$$
- Binomial: $X \sim Bin(n, p)$ if $X \in \{0, 1, ..., n\}$ and
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$
- Poisson: $X \sim Pois(\lambda)$ if $X \in \{0, 1, ...\}$ and
$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Distributions of Continuous Random Variables

A continuous random variable X assumes values in $\mathbb{R}$.
The distribution of continuous random variables is completely described by a **probability density function** $f_X : \mathbb{R} \to \mathbb{R}_+$ such that

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

Example:

- Uniform: $X \sim U(a, b)$, $a \leq b$ if

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}$$

- Chi-Square: $Q \sim \chi^2(k)$, $k \in \mathbb{Z}$ if

$$f_X(x) = \begin{cases} \frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

- Gaussian/Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ if

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability Distribution

Includes the probability mass function and probability density function mentioned above.
Also includes:

- Cumulative distribution function

$$F_X(x) = P(X \leq x), x \in \mathbb{R}.$$

- Characteristic function
- Any rule that defines a distribution

NOTATION: The right-hand sides of the previous displays are shorthand notation for the following:

$$P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$$

Note 1: Distribution can be identical even if the supporting probability space is different. Example:

$$X(H) = 1$$
$$X(T) = -1$$
$$Y(i) = \begin{cases} 1, & i \text{ is odd} \\ -1, & i \text{ is even} \end{cases}$$

Note 2: Distribution can be different even if the supporting probability space is identical.

Joint Distribution

Two random variables X and Y induce a probabillity $P_{X,Y}$ on $\mathbb{R}^2$:

$$P_{X,Y}((-\infty, x] \times (-\infty, y]) = P(X \leq x, Y \leq y).$$

A collection of random variables $X_1, X_2, ..., X_n$ induce a probabillity $P_{X_1,...,X_n}$ on $\mathbb{R}^n$:

$$P_{X_1,...,X_n}((\infty, x_1] \times \cdots \times (\infty, x_n]) = P(X_1 \leq x_1, \cdots, X_n \leq x_n).$$

Joint distribution of two discrete random variables X and Y assuming values in S_X and S_Y can be completely described by joint probability mass function $p_{X,Y} : \mathbb{R} \times \mathbb{R} \to [0,1]$ such that

$$P(X = x, Y = y) = p_{X,Y}(x, y).$$

Joint distribution of two continuous random variables X and Y can be completely described by joint probability density function $f_{X,Y} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ such that

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x, y) dy dx.$$

Outline

1. Probability Concepts
   - Probability Space
   - Combinatorics
   - Random Variables and Distributions
   - Independence and Conditional Probability

Independence

Two events A and B are independent if and only if their joint probability equals the product of their probabilities:

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

A finite set of events $\{A_i\}$ is pairwise independent iff every pair of events is independent. That is, if and only if for all distinct pairs of indices m, n

$$P(A_m \cap A_n) = P(A_m)P(A_n).$$

A finite set of events is mutually independent if and only if every event is independent of any intersection of the other events.That is, iff for every subset A_n

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

A set of random variables is pairwise independent iff every pair of random variables is independent.
That is either joint cumulative distribution function

$$F_{X,Y}(x, y) = F_X(x)F_Y(y),$$

or equivalently, joint density

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

A set of random variables is mutually independent iff for any finite subset $X_1, \ldots, X_n$ and any finite sequence of numbers $a_1, \ldots, a_n$, the events $\{X_1 \leq a_1\}, \ldots, \{X_n \leq a_n\}$ are mutually independent events.

Conditional Probability

The conditional probability of A given B is defined as

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

Conditional Probability Mass and Density

If X and Y are both discrete random variables with joint probability mass function $p_{X,Y}(x,y)$,

$$P(X = x|Y = y) = p_{X|Y}(x|y) := \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

If X and Y are both continuous random variables with joint density function $f_{X,Y}(x,y)$,

$$P(a \leq X \leq b|Y = y) = \int_b^a f_{X|Y}(x|y)dx,$$

where

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Bayes' theorem

The relationship between $P(A|B)$ and $P(B|A)$ is given by

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

That is, $P(A|B) = P(B|A)$ only if $P(B)/P(A) = 1$, or equivalently, $P(A) = P(B)$.

Exercise

- Consider a family with two children. Given that one of the children is a boy, what is the probability that both children are boys?
- Mr. Smith is the father of two. We meet him walking along the street with a young boy whom he proudly introduces as his son. What is the probability that Mr. Smiths other child is also a boy?

From 5 August 2011 New York Times article by John Allen Paulos:
"Assume that you are presented with three coins, two of them fair and the other a counterfeit that always lands heads. If you randomly pick one of the three coins, the probability that it's the counterfeit is 1 in 3. This is the prior probability of the hypothesis that the coin is counterfeit. Now after picking the coin, you flip it three times and observe that it lands heads each time. Seeing this new evidence that your chosen coin has landed heads three times in a row, you want to know the revised posterior probability that it is the counterfeit. The answer to this question, found using Bayes's theorem (calculation mercifully omitted), is 4 in 5. You thus revise your probability estimate of the coin's being counterfeit upward from 1 in 3 to 4 in 5."

Suppose a drug test is 99% sensitive. That is, the test will produce 99% true positive results for drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user?

Probability and Statistics

Jane Bae

Stanford University

hjbae@stanford.edu

September 15, 2014

Overview

1. Probability Concetps
   - Expectation and Moments
   - Conditional Expectation
   - Moment Generating Function, Characteristic Function
2. Limit Theorems
   - Modes of Convergence
   - Law of Large Numbers
   - Central Limit Theorem
3. Statistics
   - Estimators
   - Estimation Strategies

Outline

Expectation

For discrete random variable X, the expectation of X is

$$E[X] = \sum_{x \in S} x p_X(x).$$

For continuous random variable Y, the expectation of Y is

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy.$$

We can also compute expecation of $g(X)$ and $g(Y)$ as

$$E[g(X)] = \sum_{x \in S} g(x) p_X(x),$$

and

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y) f_Y(y) dy.$$

Basic Properties of Expectation

- Monotonicity
  If X and Y are random variables such that $X \leq Y$ "almost surely", then $E[X] \leq E[Y]$.
- Linearity
  $E[aX + bY + c] = aE[X] + bE[Y] + c$.

Moments

The expected values of the powers of X are called the **moments** of X.

- Mean: $E[X]$
- Variance: $E[(X - E[X])^2] = E[X^2] - E[X]^2$
  (Standard Deviation $= \sqrt{\text{Variance}}$)
- There are higher order moments of both X and $X - E[X]$.

Exercises

- A number is chosen at random from the set $S = \{-1, 0, 1\}$. Let X be the number chosen. Find the expected value, variance, and standard deviation of X.
- Let X and Y be independent random variables with uniform density functions on $[0, 1]$. Find
  (a) $E(|X - Y|)$.
  (b) $E(\max(X, Y))$.
  (c) $E(\min(X, Y))$.

Covariance

The covariance between two jointly distributed real-valued random variables X and Y with finite second moments is defined as

$$\sigma(X, Y) = E[(X - E[X])(Y - E[Y])].$$

By using the linearity property of expectations

$$\sigma(X, Y) = E[XY] - E[X]E[Y].$$

If $\sigma(X, Y) = 0$, the two random variables are **uncorrelated**.

- Exercise: Show that independent random variables are uncorrelated.

Outline

1. Probability Concetps
   - Expectation and Moments
   - Conditional Expectation
   - Moment Generating Function, Characteristic Function
2. Limit Theorems
3. Statistics

Conditional Expectation

For discrete random variables X and Y, the conditional expectation of X given Y = y is

$$E[X|Y=y] = \sum_{x \in S} x p_{X|Y}(x|y) dx = \sum_{x \in S} x \frac{P(X=x, Y=y)}{P(Y=y)}.$$

For continuous random variables X and Y, the conditional expectation of X given Y = y is

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

- Linearity and monotonicity also holds for conditional expectation.

- Chebyshev's Inequality:
  Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

- Markov's Inequality:
  If X is any nonnegative integrable random variable and $a > 0$, then

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

- Jensens Inequality:
  If X is a random variable and φ is a convex function, then
  $$\varphi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left[\varphi(X)\right].$$
- Holder's Inequality:
  If X and Y are random variables on Ω, and $p, q > 1$ with $1/p + 1/q = 1$,
  $$\mathbb{E}\left[|XY|\right] \leq \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \left(\mathbb{E}\left[|Y|^q\right]\right)^{1/q}.$$

Conditional Expectation Inequalities

- Markov's Inequality:
  If X and Y are any nonnegative integrable random variables and $a > 0$, then
  $$P(X \geq a|Y) \leq \frac{\mathbb{E}[X|Y]}{a}.$$
- Jensens Inequality:
  If X and Y are random variables and φ is a convex function, then
  $$\varphi\left(\mathbb{E}[X|Y]\right) \leq \mathbb{E}\left[\varphi(X)|Y\right].$$

- Holder's Inequality:
  If X Y and Z are random variables on Ω, and $p, q > 1$ with $1/p + 1/q = 1$,

$$\mathbb{E}\big[|XY||Z\big] \leq \big(\mathbb{E}\big[|X|^p|Z\big]\big)^{1/p}\,\big(\mathbb{E}\big[|Y|^q|Z\big]\big)^{1/q}.$$

Tower Property

Tower Property (Law of Iterated Expectation, Law of Total Expectation)

$$E[X] = E[E[X|Y]]$$

i.e.,

$$E[X] = \sum_{x \in S} E[X|Y = y]P(Y = y)$$

- If $Y \sim Unif(0, 1)$ and $X \sim Unif(Y, 1)$, what is $E[X]$?

Additional Properties of Conditional Expectation

- $E[Xg(Y)|Y] = g(Y)E[X|Y]$
- $E[E[X|Y, Z]|Y] = E[X|Y]$
- $E[X|Y] = X$, if $X = g(Y)$ for some g
- $E[h(X, Y)|Y = y] = E[h(X, y)]$
- $E[X|Y] = E[X]$, if X and Y are independent

Outline

Moment Generating Function and Characteristic Function

Moment generating function and characteristic function chracterizes the distribution of the random variable.

- Moment Generating Function

$$M_X(\theta) = E[exp(\theta X)]$$

- Characteristic Function

$$\Phi_X(\theta) = E[exp(i\theta X)]$$

Outline

1. Probability Concepts
2. Limit Theorems
   - Modes of Convergence
   - Law of Large Numbers
   - Central Limit Theorem
3. Statistics

Almost Sure Convergence

Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that X_n converges almost surely to X_∞ as $n \to \infty$ if

$$P(X_n \to X_\infty \text{ as } n \to \infty) = 1$$

We use the notation $X_n \stackrel{a.s.}{\to} X_\infty$ to denote almost sure convergence, or convergence with probability 1.

Convergence in Probability

Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that X_n converges in probability to X_∞ if for each $\epsilon > 0$,

$$P(|X_n - X_\infty| > \epsilon) \to 0$$

as $n \to \infty$. We use the notation $X_n \xrightarrow{p} X_\infty$ to denote convergence in probability.

Weak Convergence

Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that X_n converges weakly to X_∞ if

$$P(X_n \leq x) \to P(X_\infty \leq x)$$

as $n \to \infty$ for each x at which $P(X_\infty \leq x)$ is continuous. We use the notation $X_n \Rightarrow X_\infty$ or $X_n \xrightarrow{\mathcal{D}} X_\infty$ to denote weak convergence or convergence in distribution.

Implecations

Almost Sure Convergence

$\downarrow$

Convergence in Probability

$\downarrow$

Weak Convergence

Outline

Weak Law of Large Numbers

Theorem (Weak Law of Large Numbers)
Suppose that $X_1, X_2, \ldots$is a sequence of i.i.d. r.v.s such that $E[X_1] < \infty$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E[X_1]$$

as $n \to \infty$

Theorem (Strong Law of Large Numbers)
Suppose that $X_1, X_2, ...$ is a sequence of i.i.d. r.v.s such that $E[X_1]$ exists. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{a.s.}{\to} E[X_1]$$

as $n \to \infty$

Outline

Central Limit Theorem

Theorem (Central Limit Theorem)

Suppose that $X_1, X_2, ..., X_n$ are i.i.d. r.v.s with common finite variance σ^2. Then, if $S_n = X_1 + \cdots + X_n$,

$$\frac{S_n - nE[X_1]}{\sqrt{n}} \Rightarrow \sigma\mathcal{N}(0,1)$$

as $n \to \infty$. From here, we can deduce the following approximation:

$$\frac{1}{n}S_n - E[X_1] \overset{\mathcal{D}}{\sim} \frac{1}{\sqrt{n}}\mathcal{N}(0,1)$$

What is statistics?

- Statistics is a mathematical body of science that pertains to the collection, analysis, interpretation or explanation, and presentation of data, or as a branch of mathematics.
- A statistic is random variable which is a function of the random sample, but not a function of unknown parameters.

Outline

Estimation

Estimation is making a best guess of an unknown parameter out of sample data.

Estimators

An **estimator** is a statistic that is used to infer the value of an unknown parameter in a statistical model. (rule of estimation)

Examples:

- Mean
- Standard deviation

Properties of Estimators

- For a given sample x, the **error** of the estimator $\widehat{\theta}$ is defined as

$$e(x) = \widehat{\theta}(x) - \theta,$$

- the **mean squared error** of $\widehat{\theta}$ is defined as

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta}(x) - \theta)^2].$$

- The **variance** of $\widehat{\theta}$ is defined as

$$\mathrm{var}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta} - \mathrm{E}(\widehat{\theta}))^2],$$

- the **bias** of $\widehat{\theta}$ is defined as

$$B(\widehat{\theta}) = \mathrm{E}(\widehat{\theta}) - \theta.$$

Properties of Estimators

- An estimator is **unbiased** if and only if $B(\widehat{\theta}) = 0$.
- The MSE, variance and bias are related

$$\text{MSE}(\widehat{\theta}) = \text{var}(\widehat{\theta}) + \left(B(\widehat{\theta})\right)^2$$

Confidence Interval

Consider the sample mean estimator $\widehat{\theta} = \frac{1}{n} S_n$. From CLT,

$$\frac{S_n - nE[X_1]}{\sqrt{n}} \Rightarrow \sigma \mathcal{N}(0,1).$$

Rearranging terms, (this is not a rigorous argument)

$$\frac{1}{n} S_n \overset{\mathcal{D}}{\sim} E[X_1] + \frac{1}{\sqrt{n}} \mathcal{N}(0,1)$$

Outline

1. Probability Concetps
2. Limit Theorems
3. Statistics
   - Estimators
   - Estimation Strategies

Maximum Likelihood Estimation

Finding most likely explanation:

$$\widehat{\theta_n} = \arg\max_{\theta} f(x_1, x_2, ..., x_n|\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdots f(x_n|\theta)$$

- Gold Standard: Gueranteed to be consistent ($\hat{\theta}_{\text{mle}} \xrightarrow{p} \theta_0$.)
- Often computationally challenging

Probability and Statistics

Jane Bae

Stanford University

hjbae@stanford.edu

September 16, 2014

Overview

1. Statistics
   - Hypothesis Testing
2. Markov Chains
   - Markov Chains
   - Absorbing Markov Chains
   - Ergodic and Regular Markov Chains
3. Random Walks
   - Random Walks
   - Gambler's Ruin
4. Simulation
   - Monte Carlo
   - Rare Event Simulation
5. Further Information
   - Classes at Stanford

Outline

Sample Size

Note that as the degree of confidence increases, the interval must become larger.

There is a way to improve both the degree of confidence and the precision of the interval: by increasing the **sample size**. However, in the real world, increasing the sample size costs time and money.

Hypothesis Testing

Oftentimes we want to determine whether a claim is true or false. Such a claim is called a hypothesis.

- Null Hypothesis: A specific hypothesis to be tested in an experiment.
- Alternative Hypoethesis: A specific hypothesis to be tested in an experiment.

1. Formulate the null hypothesis H_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H_a (commonly, that the observations show a real effect combined with a component of chance variation).
2. Identify a test statistic that can be used to assess the truth of the null hypothesis.
3. Compute the p-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the p-value, the stronger the evidence against the null hypothesis.
4. Compare the p-value to an acceptable significance value α (sometimes called an α value). If $p \leq \alpha$, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

Error in Hypothesis Testing

- Type 1 Error: Reject the null hypothesis when it is true
- Type 2 Error: Accept the null hypothesis when it is false

Example

Suppose that ordinary aspirin has been found effective against headaches 60 percent of the time, and that a drug company claims that its new aspirin with a special headache additive is more effective.

- Null hypothesis: $p = 0.6$
- Alternate hypothesis: $p > 0.6$,
  where p is the probability that the new aspirin is effective.

We give the aspirin to n people to take when they have a headache. We want to find a number m, called the critical value for our experiment, such that we reject the null hypothesis if at least m people are cured, and otherwise we accept it. How should we determine this critical value?

Outline

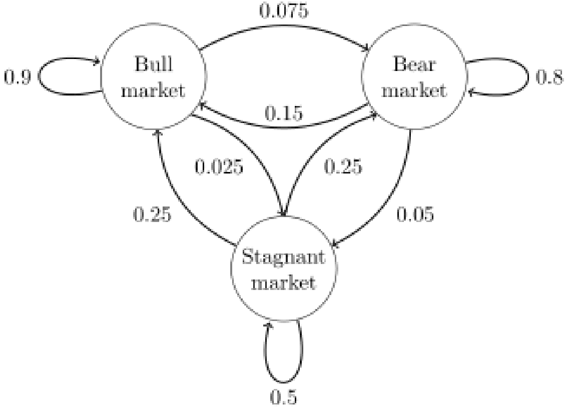
Markov Chains

A **Markov chain** is a sequence of random variables $X_1, X_2, X_3, \ldots$ with the Markov property, namely that, given the present state, the future and past states are independent.

$$P(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_{n+1} = x \mid X_n = x_n)$$

The possible values of X_i form a countable set S called the state space of the chain.

Markov chains are often described by a sequence of directed graphs, where the edges of graph n are labeled by the probabilities of going from one state at time n to the other states at time $n+1$, $P(X_{n+1} = x \mid X_n = x_n)$.



Specifying a Markov Chain

- State Space: $S = \{s_1, s_2, ...s_r\}$, the set of possible states.
- Transition Probability: $p_{ij} = P(X_{n+1} = s_j | X_n = s_i)$. (Transition Matrix)

Outline

Absorbing Markov Chains

A state s_i of a Markov chain is called **absorbing** if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step).

- probability of absorption
- time to absorption

Outline

Ergodic Markov Chains

- A Markov chain is **ergodic** if it is possible to go from every state to every state (Also known as **irreducible**).
- A Markov chain is **regular** if some power of the transition matrix has only positive elements.

$$\text{Regular} \Rightarrow \text{Ergodic}$$

The converse does not hold: Ergodic $\not\Rightarrow$ Regular
Ex) Let the transition matrix of a Markov chain be defined by

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Outline

Random Walks

Let $\{X_k\}_{k=1}^{\infty}$ be a sequence of independent, identically distributed discrete random variables. For each positive integer n, we let S_n denote the sum $X_1 + X_2 + ... + X_n$. The sequence $\{S_n\}_{n=1}^{\infty}$ is called a **random walk**.

Ex) If

$$X_i = \begin{cases} 1, & \text{with probability } 0.5 \\ -1, & \text{with probability } 0.5, \end{cases}$$

this is a symmetric random walk on a real line ($\mathbb{R}$) with equal probability of moving left or right.

Probability of First Return

In the symmetric random walk process in $\mathbb{R}$, what is the probability that the particle first returns to the origin after time $2m$?

Probability of Eventual Return

In the symmetric random walk process in $\mathbb{R}^m$, what is the probability that the particle eventually returns to the origin?

- For $m = 1, 2$ the probability of eventual return is 1. For other cases, it is strictly less than 1. (for $m = 3$, it is about 0.34)

Outline

Gambler's Ruin

Consider a nonsymmetric random walk on $\mathbb{R}$.

$$X_i = \begin{cases} 1, & \text{with probability } p \\ -1, & \text{with probability } q, \end{cases}$$

with $p + q = 1$.

A gambler starts with a stake of size s. He plays until his capital reaches the value M or the value 0.

Outline

Monte Carlo

Computational algorithms that rely on repeated random sampling to compute their results.

Theoretical Bases

- Law of Large Numbers guarantees the convergence

$$\frac{1}{n}\sum I_{(X_i \in A)} \to P(X_1 \in A)$$

- Central Limit Theorem

$$\frac{1}{n}\sum I_{(X_i \in A)} - P(X_1 \in A) \sim \frac{\sigma}{\sqrt{n}}\mathcal{N}(0,1)$$

Importance Sampling

We can express the expectation of a random variable as an expectation of another random variable.

eg.
Two continuous random variable X and Y have density f_X and f_Y such that $f_Y(s) = 0$ implies $f_X(s) = 0$. Then,

$$E[g(X)] = \int g(s) f_X(s) ds = \int g(s) \frac{f_X(s)}{f_Y(s)} f_X(s) ds = E[g(Y)L(Y)]$$

where $L(s) = \frac{f_X(s)}{f_Y(s)}$ is called a likelihood ratio.

Outline

Rare Events

Probability that a coin lands on its edge. How many flips do we need to see at least one occurrence?

Outline

Probability

- Basic Probability: STATS 116
- Stochastic Processes: STATS 215, 217, 218, 219
- Theory of Probability: STATS 310 ABC

Statistics

- Intro to Statistics: STATS 200
- Theory of Statistics: STATS 300 ABC

Applications

- Applied Statistics: STATS 191, 203, 208, 305, 315AB
- Stochastic Systems: MS&E 121, 321
- Stochastic Control: MS&E 322
- Stochastic Simulation: MS&E 223, 323, STATS 362
- Little bit of Everything: CME 308