MATH 503: Mathematical Statistics

Lecture 1: Prerequisite Concepts Review

Readings: Chapters 1-5

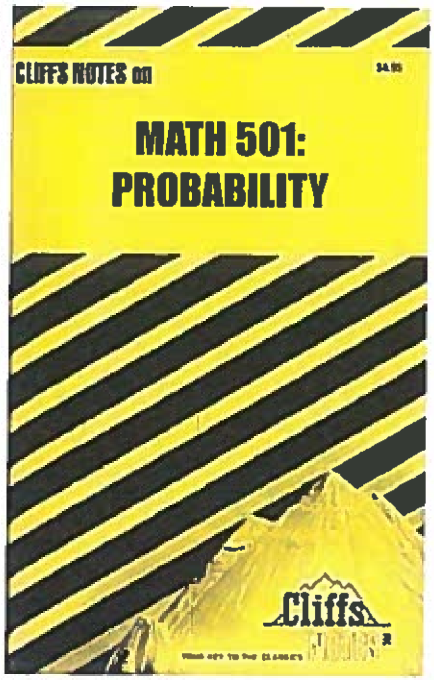
Kimberly F. Sellers

Department of Mathematics and Statistics

Georgetown University

Course Introduction

- What? MATH 503
- Who (teaches)? Dr. Kimberly Sellers
- Where? Depends
- When? Depends
- Why?
- Logistics:
  - Syllabus
- Textbook:
  - Casella and Berger (2002) *Statistical Inference, Second Edition.*



Today's Topics

- Probability, and Conditional Probability
- Random variables
- Basic distribution theory
- Transformations
- Expectation and Expected Value
- Moments, and Moment Generating Functions
- Distributions
- Inequalities
- Introduction to Statistics
- Convergences

Axioms of Probability

1. For any event A, $P(A) \geq 0$.
2. $P(S) = 1$.
3. If $A_1, A_2, \ldots, A_n$ is a finite collection of mutually exclusive events, then
$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$
4. If $A_1, A_2, A_3, \ldots$ is an infinite collection of mutually exclusive events, then
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1 \cup A_2 \cup \ldots) = \sum_{i=1}^{\infty} P(A_i)$$

Properties of Probability

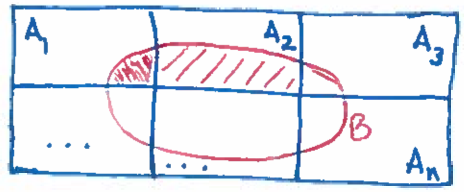
1. For any event, A, $P(A') = 1 - P(A)$.
2. If A and B are mutually exclusive, then $P(A \cap B) = 0$.
3. For two events A and B where $A \subset B$, $P(A) \leq P(B)$.
4. (General Addition Rule) For any two events A & B,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
5. (Boole's Inequality) Let $\{A_i\}$ be an arbitrary sequence of events. Then $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$

Conditional Probability

- The conditional probability of A given that the event B has occurred is
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$
where $P(B) > 0$.

- Note: by cross-multiplying both sides, we get "the multiplication rule":
$$P(A \cap B) = P(A|B) \cdot P(B)$$

The Law of Total Probability



$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \ldots + \mathbb{P}(B \cap A_n)$$
$$= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \ldots + \mathbb{P}(B|A_n)\mathbb{P}(A_n)$$

$$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)$$
$$= \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

A simpler case: $P(B) = P(B|A)P(A) + P(B|A')P(A')$

Bayes' Theorem/ Bayes' Law

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

← multiplication rule

← Law of Total Prob.

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)}$$
$$= \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

A simpler case: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$

Independence

- A, B are independent if and only if

$$P(A|B) = P(A).$$

In other words…

knowledge about B doesn't impact your probability about A occurring

- What is the multiplication rule for independent events?

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Exercise

If C_1 and C_2 are independent events, show that C_1 and C_2' are also independent.

C_1, C_2 indpt $\Rightarrow \mathbb{P}(C_1 C_2) = \mathbb{P}(C_1)\mathbb{P}(C_2)$.

$$\begin{aligned}\mathbb{P}(C_1 C_2') &= \mathbb{P}(C_1) - \mathbb{P}(C_1 C_2) \\ &= \mathbb{P}(C_1) - \mathbb{P}(C_1)\mathbb{P}(C_2) \text{ because } C_1 \perp C_2 \\ &= \mathbb{P}(C_1)[1 - \mathbb{P}(C_2)] = \mathbb{P}(C_1)\mathbb{P}(C_2') \checkmark\end{aligned}$$

Alternatively, C_1, C_2 indpt $\Rightarrow \mathbb{P}(C_2 | C_1) = \mathbb{P}(C_2)$.

$$\mathbb{P}(C_2' | C_1) = 1 - \mathbb{P}(C_2 | C_1) = 1 - \mathbb{P}(C_2) \text{ [because } C_1 \perp C_2\text{]}$$
$$= \mathbb{P}(C_2') \checkmark$$

Definition: Random Variable

- For a given sample space, a random variable X is any rule that associates a number with each outcome in S.
- Example: define random variable, X, as

  $$X(\mathrm{S}) = 1 \qquad X(\mathrm{F}) = 0$$

  This is an example of a Bernoulli random variable.
- Random variables have probability distributions associated. These are referred to as probability mass functions [denoted $p(x)$] when X is a discrete random variable, or probability density functions (pdf) [denoted $f(x)$] where X is continuous.

What is a pmf? A pdf?

- A probability mass function (denoted pmf) is a function, $p(x) = P(X = x)$.

- A probability density function (denoted pdf; also called a density curve) is a function, $f(x)$, such that

  $$P(a \le X \le b) = \int_a^b f(x)\mathrm{d}x.$$

Properties Summary for Discrete & Continuous Random Variables

	Discrete r.v.	Continuous r.v.
Probabilities must be non-negative	$p(x) \geq 0$ for all x	$f(x) \geq 0$ for all x
Sum of all probabilities is 1	$\sum_x p(x) = 1$	$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$
Cumulative Distribution Function (CDF): $F(x) = P(X \leq x)$	$= \sum_{y: y \leq x} p(y)$	$= \int_{-\infty}^{x} f(y)\mathrm{d}y$

Cumulative Distribution Functions

- Denoted $F(x) = P(X \leq x)$
- $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ (non-negative)
- $F(\infty) = \lim_{x \to \infty} F(x) = 1$
- $F(x + \epsilon) \geq F(x)$, $\epsilon > 0$ (non-decreasing)
- $F(x^+) = \lim_{\epsilon \downarrow 0} F(x + \epsilon) = F(x)$ (right-continuous)

Example

A certain river floods every year. Suppose that the low-water mark is set at 1 and the high-water mark Y has distribution function $F_Y(y) = 1 - \frac{1}{y^2}, 1 \leq y < \infty$.

(a) Verify that $F_Y(y)$ is a cdf.

(b) Find $f_Y(y)$, the pdf of Y.

SEE SCRAP

Univariate Transformations

- For a random variable X with known distribution, we want to find the distribution of $Y = g(X)$. Let g be a one-to-one function.
  - For X discrete,

$$p_Y(y) = P[g(X) = y] = P[X = g^{-1}(y)] = p_X(g^{-1}(y))$$

  - For X continuous and g differentiable,

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|, \quad y \in S_Y$$

$$F_Y(y) = \begin{cases} 1 - 1/y^2, & 1 \le y < \infty \\ 0, & y < 1 \end{cases}$$

Ⓐ ① $F(-\infty) = \lim_{y\to-\infty} F_Y(y) = \lim_{y\to-\infty} 0 = 0$ ✓

② $F(\infty) = \lim_{y\to\infty} F_Y(y) = \lim_{y\to\infty} \left(1 - 1/y^2\right) = 1$ ✓

③ Show $F(y+\epsilon) \ge F(y)$:

For $\epsilon > 0$, $y+\epsilon \ge y \implies (y+\epsilon)^2 \ge y^2$

$$\frac{1}{(y+\epsilon)^2} \le 1/y^2$$

$$1 - \frac{1}{(y+\epsilon)^2} \ge 1 - 1/y^2$$

$$F_Y(y+\epsilon) = \begin{cases} 1 - \frac{1}{(y+\epsilon)^2}, & 1 \le y+\epsilon < \infty \quad (\therefore\ 1-\epsilon \le y < \infty) \\ 0 & y+\epsilon < 1 \quad (\therefore\ y < 1-\epsilon) \end{cases}$$

$\implies F(y) \le F(y+\epsilon)$

④ $F(y^+) = \lim_{\epsilon\downarrow 0} F(y+\epsilon) = \lim_{\epsilon\downarrow 0} \begin{cases} 1 - \frac{1}{(y+\epsilon)^2}, & 1 \le y+\epsilon < \infty \\ 0, & y+\epsilon < 1 \end{cases}$

$$= \begin{cases} 1 - \frac{1}{y^2}, & 1 \le y < \infty \\ 0, & y < 1 \end{cases} = F(y)$$

Ⓑ $f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\left(1 - 1/y^2\right) = \frac{d}{dy}\left(1 - y^{-2}\right) = 2y^{-3} = \frac{2}{y^3}, \quad 1 \le y < \infty$

Example (cont.)

If the low-water mark is reset at zero and we use a unit of measurement which is $\frac{1}{10}$ of that previously, the high-water mark becomes $Z = 10(Y-1)$. Find $F_Z(z)$.

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}(10(Y-1) \le z)$$
$$= \mathbb{P}\left(Y \le \frac{z}{10}+1\right) = F_Y\left(\frac{z}{10}+1\right)$$
$$= 1 - \frac{1}{\left(\frac{z}{10}+1\right)^2} = 1 - \frac{1}{\left(\frac{z+10}{10}\right)^2}$$
$$= 1 - \frac{100}{(z+10)^2} = \frac{z^2+20z+100-100}{(z+10)^2} = \frac{z^2+20z}{(z+10)^2}$$

where $y = \frac{z}{10}+1 \ge 1 \therefore z \ge 0$

Bivariate Transformations

- For bivariate random variables X_1, X_2 and transformations $Y_i = g_i(X_1, X_2),\ \ i = 1,2$, we have
  - for X_1, X_2 discrete,
$$p_{Y_1,Y_2}(y_1, y_2) = p_{X_1,X_2}[h_1(y_1, y_2), h_2(y_1, y_2)]$$
  - for X_1, X_2 continuous,
$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}[h_1(y_1, y_2), h_2(y_1, y_2)]|J|$$

    where
$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Exercise

Let X_1 and X_2 have the joint pdf
$h(x_1, x_2) = 2e^{-x_1-x_2}$, $0 < x_1 < x_2 < \infty$, zero o.w.
Find the joint pdf of $Y_1 = 2X_1$ and $Y_2 = X_2 - X_1$.

$\begin{cases} Y_1 = 2X_1 \\ Y_2 = X_2 - X_1 \end{cases} \Rightarrow \begin{cases} X_1 = Y_1/2 \\ X_2 = Y_2 + Y_1/2 \end{cases} \quad J = \begin{vmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{vmatrix}$

$\therefore |J| = (\tfrac{1}{2})(1) - (\tfrac{1}{2})(0) = \tfrac{1}{2}$

$f(y_1, y_2) = h(x_1 = y_1/2,\ x_2 = y_2 + y_1/2) \cdot (\tfrac{1}{2})$
$= \cancel{2} e^{-y_1/2 - (y_2 + y_1/2)} \cdot (\cancel{\tfrac{1}{2}}) = e^{-y_1 - y_2}$, $y_1 > 0$, $y_2 > 0$

Expected Value, Variance, and Standard Deviation (Discrete r.v.'s)

- Expected value, $\mu = E(X) = \sum_x x p(x)$
- Variance of X, $\sigma_X^2 = V(X) = E[(X-\mu_X)^2]$
  $= E(X^2) - \mu_X^2$
  $= \left(\sum_x x^2 p(x)\right) - \left(\sum_x x p(x)\right)^2$
- Standard deviation of X, $\sigma_X = +\sqrt{\sigma_X^2}$

Expected Value, Variance, and Standard Deviation (Cont. r.v.'s)

- Expected value, $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$
- Variance of X, $\sigma_X^2 = E(X^2) - \mu^2$
$$= \left(\int_{-\infty}^{\infty} x^2 f(x)dx\right) - \left(\int_{-\infty}^{\infty} xf(x)dx\right)^2$$
- Standard deviation of X, $\sigma_X = +\sqrt{\sigma_X^2}$

Functions of Random Variables

- $E[h(X)] = \sum_x h(x)p(x)$ *for discrete X*
$= \int_x h(t)p(t)\mathrm{d}t$ *for continuous X*
- For some constants, a, b:
  - $E(aX + b) = aE(X) + b$
  - $V(aX + b) = a^2V(X)$
  - $\sigma_{aX+b} = |a|\sigma_X$
- Questions:

b is a constant, so its expected value is (naturally) itself.

$E(b) = b$. Interpret/explain.

$\sigma_b = 0$. Interpret/explain.

b is constant, so there is no variation associated with it.

Moment Generating Functions

- The moment generating function (mgf) of X is defined as $M_X(t) = E(e^{tX})$
- **Thm:** Let X and Y be random variables with mgfs, $M_X(t)$ and $M_Y(t)$, respectively, existing in open intervals about 0. Then $F_X(z) = FY(z)\ \forall\ z \in R$ iff. $M_X(t) = M_Y(t)$ $\forall t \in (-h, h)$ for some $h > 0$.

Moments

- For m (a positive integer), let $M^{(m)}(t)$ denote mth derivative (wrt. t) of mgf $M(t)$
- $E(X^m) = M^{(m)}(0)$ is the mth moment of X
- *Note:* mth central moment is $E\big((X-\mu)^m\big)$

Exercise

- Derive the mgf for a rv X with pdf

$$f(x|\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta},\quad 0\le x<\infty,\quad \alpha,\beta>0$$

- Use the mgf to derive $E(X)$.

$$M_X(t)=\mathbb{E}(e^{tX})=\int_0^\infty e^{tx}\cdot\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}\,dx$$

$$=\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-x(1/\beta-t)}\,dx$$

$$=\frac{1}{\cancel{\Gamma(\alpha)}\cancel{\beta^\alpha}}\cdot\cancel{\Gamma(\alpha)}\left(\frac{\cancel{\beta}}{1-\beta t}\right)^\alpha\cancelto{1}{\int_0^\infty\frac{1}{\Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha}x^{\alpha-1}e^{\frac{-x}{\beta/(1-\beta t)}}\,dx}$$

$$=\left(\frac{1}{1-\beta t}\right)^\alpha \text{ or } (1-\beta t)^{-\alpha}$$

$$M_X'(t)=\frac{\partial M(t)}{\partial t}=-\alpha(1-\beta t)^{-\alpha-1}(-\beta)$$

$$\therefore M_X'(t)\Big|_{t=0}=-\alpha(-\beta)=\alpha\beta=\mathbb{E}(X)$$

Conditional Distributions and Expectations

- The conditional pdf of X_1 given $X_2=x_2$ is

$$f_{X_1.X_2}(x_1|x_2)=\frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)},\quad f_{X_2}(x_2)>0$$

- Let (X_1,X_2) be a random vector s.t. the variance of X_2 is finite. Then,
  $E[E(X_2|X_1)] = E(X_2)$, and
  $\mathrm{Var}[E(X_2|X_1)]\le \mathrm{Var}(X_2)$

Correlation Coefficient

- The covariance of X and Y is
$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

- The correlation coefficient, denoted ρ, is
$$\rho = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2} = \frac{\mathrm{Cov}(X,Y)}{\sigma_1\sigma_2}$$

Independence

- Let the r.v.'s X_1 and X_2 have the joint pdf $f(x_1, x_2)$ and the marginal pdfs $f_1(x_1)$ and $f_2(x_2)$, respectively. The r.v.'s X_1 and X_2 are independent iff. $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. Analogously true for cdfs.

- **Thm:** Let r.v.'s X_1 and X_2 have supports S_1 and S_2, respectively, and have the joint pdf $f(x_1, x_2)$. Then X_1 and X_2 are indpt. iff. $f(x_1, x_2)$ can be written as a product of a nonnegative function of x_1 and x_2, i.e.
$$f(x_1, x_2) = g(x_1)h(x_2),$$
where $g(x_1) > 0$, $x_1 \in S_1$, and $h(x_2) > 0$, $x_2 \in S_2$.

Independence (cont.)

- **Thm:** Suppose X_1 and X_2 are indpt and $E(u(X_1))$ and $E(v(X_2))$ exist. Then $E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)]$.

- **Thm:** Suppose the joint mgf,
$$M(t_1, t_2) = E(\exp(t_1 X_1 + t_2 X_2)),$$
exists for the r.v.'s X_1 and X_2. Then X_1 and X_2 are indpt. iff. $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$.

Common Distributions

- Discrete
  - Bernoulli
  - Binomial
  - Discrete Uniform
  - Geometric
  - Hypergeometric
  - Negative binomial
  - Poisson
- Continuous
  - Beta
  - Cauchy
  - Chi-squared
  - Exponential
  - F
  - Gamma
  - Normal
  - T
  - Uniform
  - Weibull

Refer to p. 621-627 regarding common distributions

Binomial Distribution

- Suppose we independently flip an unfair coin n times with $P(H)=p$.
- $X \sim \text{Bin}(n,p)$ and
$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},\ x=0,1,\ldots,n$$
is the pmf where $\binom{n}{x}=\frac{n!}{x!(n-x)!}$ is the number of ways of choosing x objects from n total objects, irrespective of order.
- n=1 ⇒ Bernoulli distribution

Binomial Distribution (cont.)

- For $X \sim \text{Bin}(n,p)$,
$$E(X)=np$$
$$V(X)=np(1-p)$$
$$\sigma_X=\sqrt{np(1-p)}$$
- Assumptions:
  - Two possible outcomes
  - Independence
  - One success probability value, p
- Distributions arise as a sum of Bernoulli trials
- Sums of indpt Bin(n_i,p) r.v.'s is Bin(Σn_i,p)

Geometric Distribution

- The pmf is $p(x) = p(1-p)^{x-1}$, $x=1,2,\ldots$; $0 \le p \le 1$
- $E(X) = 1/p$
- $Var(X) = (1-p) / p^2$
- $M_X(t) = \dfrac{pe^t}{1-(1-p)e^t}, \quad t < -\ln(1-p)$
- Can also be represented for $Y = X - 1$
- Special case of negative binomial distribution
- Has memoryless property, i.e.
$$P(X > s | X > t) = P(X > s - t)$$

The Poisson Distribution

- Simple distribution for counts
- Approximates the Binomial distribution as p gets small and n gets large ($\lambda = np$)
- X is a Poisson random variable [denoted $X \sim$ Poisson(λ)] with pmf
$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots.$$
- For X ~ Poisson(λ),
$$E(X) = \lambda$$
$$V(X) = \lambda$$
$$\sigma_X = \sqrt{\lambda}$$

Poisson Distribution Characteristics/Assumptions

- The experiment consists of counting the number of times a certain event occurs during a given unit of time or in a given area or volume (or any other unit of measurement)
- The probability that an event occurs in a given unit of time, area, volume (or any other unit of measurement) is the same for all the units
- The number of events that occur in one unit is independent of the number that occur in other units
- The expected number of events in each unit is denoted λ.

Poisson Process

- Assumptions: in an infinitesimal period of time δt
  - the probability of an event occurring is $\alpha \cdot \delta t$
  - the probability of more than one event is negligible
  - the probability of an event is independent of the number of events that previously occurred
- For a given time period t,

$$P(X = x) = \frac{e^{-\alpha t}(\alpha t)^x}{x!}, \quad x = 0, 1, 2, ...$$

i.e. X ~ Poisson(αt).

Uniform Distribution

- A Uniform(a,b) distribution has density curve,

$$f(x)=\frac{1}{b-a}, \quad a \le x \le b$$

- For uniform distribution,

$$\mu=\frac{a+b}{2}$$

$$\sigma^2=\frac{(b-a)^2}{12}$$

Gamma Distribution

- The $\Gamma(\alpha,\beta)$ distribution is

$$f(x|\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}, \quad 0 \le x < \infty, \quad \alpha,\beta>0$$

- $\Gamma(\alpha)=(\alpha-1)!$ and $\Gamma(½)=\sqrt{\pi}$
- Special cases:
  - Exponential: Exp(λ) = $\Gamma(\alpha=1, \beta=\lambda)$
  - Chi-squared, $X^2(\nu) = \Gamma(\alpha=\nu/2, \beta=2)$
- Sums of indpt $\Gamma(\alpha_i,\beta)$ distributions is $\Gamma(\Sigma\alpha_i,\beta)$

Exercise

Prove that, for $X_i \sim \Gamma(\alpha_i,\beta)$ indpt., $\Sigma X_i \sim \Gamma(\Sigma\alpha_i,\beta)$.

$$M_{X_i}(t) = \frac{1}{(1-\beta t)^{\alpha_i}}$$

$$M_{\Sigma X_i}(t) = \mathbb{E}\left(e^{t\Sigma X_i}\right) = \mathbb{E}\left(e^{tX_1+tX_2+\dots+tX_n}\right)$$

$$= \mathbb{E}\left(e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right) \cancel{\text{because } X_i \text{ indpt}}$$

$$= \prod_{i=1}^{n}\mathbb{E}\left(e^{tX_i}\right) \text{ because } X_i \text{ indpt}$$

$$= \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n}\frac{1}{(1-\beta t)^{\alpha_i}} = \frac{1}{(1-\beta t)^{\Sigma\alpha_i}},$$

which is the mgf of Gamma$(\Sigma\alpha_i,\beta)$ r.v. $\therefore \Sigma X_i \sim$ Gamma$(\Sigma\alpha_i,\beta)$

Beta Distribution

- The pdf of the B(α,β) distribution is

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \quad 0\le x\le 1, \quad \alpha>0, \quad \beta>0$$

- E(X) = $\dfrac{\alpha}{\alpha+\beta}$
- Var(X) = $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Normal Distribution

- X is normally distributed [ie $X \sim N(\mu, \sigma^2)$] implies that

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

is the pdf.

- The cdf is $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$
- No closed form for cdf (ie cannot solve integral)!
  - Normal probability tables
  - Computer software

Normal Distribution (cont.)

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

  $E(X) = \mu$

  $V(X) = \sigma^2$

  $\sigma_X = \sigma$
- If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, where $a \neq 0$, then $Y \sim N(a\mu + b, a^2\sigma^2)$
- The sum of two normal random variables is also normally distributed.
- Standard normal distribution: $Z \sim N(0,1)$ where

$$Z = \frac{X - \mu}{\sigma}$$

Important Inequalities

- **Markov's Inequality**
  - Let $u(X)$ be nonnegative function of r.v. X. If $E[u(X)]$ exists, then for every positive constant c,

$$P[u(X) \geq c] \leq \frac{E[u(X)]}{c}.$$

- **Chebyshev's Inequality**
  - Let r.v. X have distribution of probability about which we assume only that there is a finite variance σ^2. Then for every $k > 0$,

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

Important Inequalities (cont.)

- **Jensen's Inequality**
  - If g is convex on open interval I and X is r.v. whose support is contained in I and has finite expectation, then

$$g[E(X)] \leq E[g(X)]$$

- Example:

Let X be positive r.v., i.e. $P(X \leq 0) = 0$. Then,

$$E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}.$$

Types of Convergence

- Almost sure (a.s.)
- In probability (p)
- In distribution (d)

Almost Sure Convergence

A sequence of r.v.'s, X_1, X_2, …, <u>converges almost surely</u> to r.v. $X\left(X_n \xrightarrow{a.s.} X\right)$, if, for every ε>0,

$$P(\lim_{n\to\infty}|X_n-X|<\varepsilon)=1$$

- Similar to pointwise convergence of a sequence of functions
- Convergence doesn't need to occur on set with probability zero

Convergence in Probability

- Let $\{X_n\}$ be a sequence of r.v.'s and let X be a r.v. defined on a sample space. X_n <u>converges in probability</u> to X $\left(X_n \xrightarrow{p} X \right)$ if $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P[|X_n - X| \geq \varepsilon] = 0.$$

- Weaker than almost sure convergence
- Markov or Chebyshev Inequality can be used to show convergence in probability

Convergence in Probability

Thm1: Suppose $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Then

$$X_n + Y_n \xrightarrow{p} X + Y.$$

Thm2: Suppose $X_n \xrightarrow{p} X$ and a constant. Then $aX_n \xrightarrow{p} aX$.

Thm3: Suppose $X_n \xrightarrow{p} a$ and the real function g is continuous at a. Then $g(X_n) \xrightarrow{p} g(a)$.

Thm4: Suppose $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Then $X_n Y_n \xrightarrow{p} XY$.

Convergence in Distribution

Let $\{X_n\}$ be a sequence of r.v.s and let X be a r.v. Let F_{X_n} and F_X be, respectively, the cdfs of X_n and X. Let $C(F_X)$ denote the set of all points where F_X is continuous. X_n converges in distribution to X $\left(X_n \xrightarrow{d} X\right)$ if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \forall x \in C(F_X)$.

- CDFs converge… not the r.v.'s themselves.

MGF Technique

- Using definition of convergence in distribution directly is difficult
- Thm: Let $\{X_n\}$ be a sequence of r.v.'s with mgf $M_{X_n}(t)$ that exists for $-h<t<h$ for all n. Let X be a r.v. with mgf $M(t)$, which exists for $|t|\leq h_1 \leq h$. If $\lim_{n\to\infty} M_{X_n}(t) = M(t)$ for $|t|\leq h_1$, then $X_n \to X$ in distribution.

Example

Let Y_n have distribution Bin(n,p). What is the associated limiting distribution?

$M_{Y_n}(t) = (pe^t + q)^n = (pe^t + (1-p))^n$ where $\mu = np \therefore p = \mu/n$

$= \left(\frac{\mu e^t}{n} + 1 - \mu/n\right)^n = \left(1 + \frac{\mu(e^t - 1)}{n}\right)^n$

Recall: $\lim_{n\to\infty} (1 + x/n)^n = e^x$

$\therefore M_{Y_n}(t) = \left(1 + \frac{\mu(e^t-1)}{n}\right)^n \xrightarrow{n\to\infty} e^{\mu(e^t-1)}$, which is the mgf of Poisson(μ) r.v., thus the limiting distribution of $Y_n \sim$ Bin(n,p) is Poisson(μ) where $\mu = np$.

Convergence Results

- If $X_n \to X$ in probability, then $X_n \to X$ in distribution.
- If $X_n \to b$ in distribution, then $X_n \to b$ in probability.
- Suppose $X_n \to X$ in distribution and $Y_n \to 0$ in probability. Then $X_n + Y_n \to X$ in distribution.
- Suppose $X_n \to X$ in distribution and g is a continuous function on the support of X. Then $g(X_n) \to g(X)$ in distribution.

Introduction to Statistical Inference

- Parameter of interest, e.g. θ, unknown
- Information regarding θ determined from random sample $X_1,\ldots,X_n$.
  - Random variables $X_1, X_2, \ldots, X_n$ form a random sample of size n if X_i's are i.i.d.
- Statistic T is a function of the sample, $T=T(X_1,\ldots,X_n)$, used to estimate θ
  - Point estimator of θ
  - Examples:

$\bar{X}, S^2$ are point estimators of μ, σ^2, respectively

What is a point estimate?

- Just that – a single value (i.e. a point) that is considered a reasonable value for an unknown parameter, θ (i.e. it estimates θ).
- A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic (say T) is the point estimator of θ.

Standard Error

- The standard error of T is its standard deviation:

$$\sigma_T = \sqrt{V(T)}$$

If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into T yields the estimated standard error of T

$$\hat{\sigma}_T = s_T \doteq \text{estimated standard error}$$

Example: the std. error of $\overline{X}$ is $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$; $\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}}$

Properties of Estimators

- Let X be a r.v. with pdf $f(x;\theta)$, $\theta \in \Omega$. Let $X_1,\ldots,X_n$ be a random sample from distribution of X and let T denote statistic. T is an unbiased estimator of θ if $E(T)=\theta$, for all $\theta \in \Omega$.
- Let X be a r.v. with cdf $F(x;\theta)$, $\theta \in \Omega$. Let $X_1,\ldots,X_n$ be a random sample from distribution of X and let T_n denote statistic. T_n is a consistent estimator of θ if $T_n \xrightarrow{p} \theta$.

Expectations of Functions

1. Let $T = \sum_{i=1}^{n} a_i X_i$. Provided $E[|X_i|] < \infty$, for $i = 1,\dots,n$, $E(T) = \sum_{i=1}^{n} a_i E(X_i)$.
2. Let $T = \sum_{i=1}^{n} a_i X_i$ and let $W = \sum_{i=1}^{n} b_i Y_i$. If $E(X_i^2) < \infty$, and $E(Y_j^2) < \infty$ for $i = 1,\dots, n$ and $j = 1,\dots, m$, then
$$Cov(T,W) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j).$$
3. Let $T = \sum_{i=1}^{n} a_i X_i$. If $E(X_i^2) < \infty$, for $i = 1,\dots, n$, then
$$\begin{aligned} Var(T) &= Cov(T,T) \\ &= \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i<j} a_i a_j Cov(X_i, X_j). \end{aligned}$$

Laws of Large Numbers

- (Strong Law of Large Numbers) Let $\{X_n\}$ be an iid sequence of r.v.'s having common mean μ and variance $\sigma^2 < \infty$. Let $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. Then $\bar{X}_n \xrightarrow{a.s.} \mu$.

- (Weak Law of Large Numbers) Let $\{X_n\}$ be a sequence of r.v.'s having common mean μ and variance $\sigma^2 < \infty$. Let $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. Then $\bar{X}_n \xrightarrow{p} \mu$.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 > 0$. Let $G_n(x)$ denote the cdf of $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$. Then, for any x, $-\infty < x < \infty$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ has a limiting standard normal distribution.

Sampling Distributions

- A **statistic** is any value determined from sample data. Statistics have distributions associated because, prior to picking the sample and obtaining the data, there is uncertainty associated as to what the value is of the statistic. So, the statistic has a distribution associated with it, called a **sampling distribution**.

The Sample Mean and CLT

- The sample mean $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is used to estimate μ =E(X).
- CLT Interpretation: Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, if n is sufficiently large, $\overline{X}$ has approximately a normal distribution with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$. Similarly, if we define $T = \sum_{i=1}^{n} X_i$ then T has approximately a normal distribution with mean $\mu_T = n\mu$ and variance $\sigma_T^2 = n\sigma^2$
- Under the Central Limit Theorem,

$$\overline{X} \sim N\left(\mu_{\overline{X}} = \mu,\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\right)$$