Chapter 11

Sums of Independent Random Variables and Limit Theorems

Outline

- 11.1 Moment-generating functions
- 11.2 Sums of independent random variables
- 11.3 Markov and Chebyshev inequalities
- 11.4 Laws of large numbers
- 11.5 Central limit theorem

Section 11.1

Moment-Generating Functions

Moments

- Moments are useful in practice as well as in developing theory.

$$\mu = E(X), \qquad \mu_X^{(r)} = E[(X-\mu)^r]$$

- The measure of **skewness** ($\mu_X^{(3)}/\sigma_X^3$)
  - Zero if symmetric
  - Negative if skewed to the left
  - Positive if skewed to the right
- The measure of **kurtosis** ($\mu_X^{(4)}/\sigma_X^4$)
  - Measures the relative flatness of the distribution function
  - For standard normal, it is 3
  - If larger than 3, it is more peaked than normal.
  - If smaller than 3, it is flatter than normal

Moment Generating Functions

- In this section, we study moment-generating functions
- They have a few properties

1) Moment-generating functions can be used to compute moments.
2) Moment-generating functions are unique. To compute the distribution function of a random variable, it is sufficient to compute its moment generating function.
3) Moment-generating functions are particularly useful to compute the distribution of the sum of independent random variables.

$$X = X_1 + X_2$$

$$\downarrow \quad \downarrow \quad \downarrow$$

$$MF \quad MF_1 \quad MF_2$$

Definition

- **Definition** For a random variable X, let
$$M_X(t) = E(e^{tX}).$$
If $M_X(t)$ is defined for all values of t in some interval $(-\delta, \delta)$, $\delta > 0$, then $M_X(t)$ is called the moment-generating function of X.
- If X is a discrete random variable with set of possible values A and probability mass function $p(x)$, then
$$M_X(t) = \sum_{x \in A} e^{tx} p(x).$$
- If X is a continuous random variable with probability density function $f(x)$, then
$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx.$$

Theorem 11.1 and Corollary

- **Theorem 11.1** Let X be a random variable with moment-generating function $M_X(t)$. Then

$$E(X^n) = M_X^{(n)}(0),$$

where $M_X^{(n)}(t)$ is the nth derivative of $M_X(t)$.

- **Corollary** The MacLaurin's series for $M_X(t)$ is given by

$$M_X(t) = \sum_{n=0}^{\infty} \frac{M_X^{(n)}(0)}{n!} t^n = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n.$$

Therefore, $E(X^n)$ is the coefficient of $t^n/n!$ in the MacLaurin's series representation of $M_X(t)$.

Proof of Theorem 11.1

- We prove the theorem for continuous random variables.
- For discrete random variables, the proof is similar.
- If X is continuous with probability density function f, then

$$M_X'(t) = \frac{d}{dt}\left(\int_{-\infty}^{\infty} e^{tx} f(x)\,dx\right) = \int_{-\infty}^{\infty} xe^{tx} f(x)\,dx,$$

$$M_X''(t) = \frac{d}{dt}\left(\int_{-\infty}^{\infty} xe^{tx} f(x)\,dx\right) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\,dx,$$

$$\vdots$$

$$M_X^{(n)}(t) = \int_{-\infty}^{\infty} x^n e^{tx} f(x)\,dx, \qquad (11.1)$$

- We have assumed that the derivatives of these integrals are equal to the integrals of the derivatives of their integrands.
- This property is valid for sufficiently smooth densities.
- Letting $t = 0$ in (11.1), we get

$$M_X^{(n)}(0) = \int_{-\infty}^{\infty} x^n f(x)\, dx = E(X^n).$$

Corollary of Theorem 11.1

- The MacLaurin's series for $M_X(t)$ is given by

$$M_X(t) = \sum_{n=0}^{\infty} \frac{M_X^{(n)}(0)}{n!} t^n = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n .$$

- Therefore, $E(X^n)$ is the coefficient of $t^n/n!$ in the MacLaurin's series representation of $M_X(t)$.
- If M_X is finite, then the moments of all orders of X must be finite.
    - The converse may not be true!

Lemma 11.1

- Let X be a random variable with moment-generating function $M_X(t)$.
- For constants a and b, let $Y = aX + b$. Then the moment-generating function of Y is given by

$$M_Y(t) = e^{bt} M_X(at).$$

- **Proof.**
- By the definition of moment-generating function,

$$M_Y(t) = E(e^{tY}) = E\left[e^{t(aX+b)}\right] = e^{bt} E(e^{atX}) = e^{bt} M_X(at)$$

$$= E\left[e^{tb} \cdot e^{tax}\right]$$

A Remark

- It is important to know that if M_X is to be finite, then the moments of all orders of X must be finite.
- But the converse need not be true.
- That is, all the moments may be finite and yet there is no neighborhood of 0, $(-\delta, \delta), \delta > 0$, on which M_X is finite.

Example 11.2

- Let X be a binomial random variable with parameters (n, p). Find the moment-generating function of X, and use it to calculate $E(X)$ and $\mathrm{Var}(X)$.

- **Solution**
- The probability mass function of X, $p(x)$, is given by

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2 \ldots, n, q = 1 - p.$$

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$$

- Hence,

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + q)^n.$$

- To calculate the mean and the variance of X, note that

$$M_X'(t) = npe^t(pe^t + q)^{n-1}$$
$$M_X''(t) = npe^t(pe^t + q)^{n-1} + n(n-1)(pe^t)^2(pe^t + q)^{n-2}.$$

- Thus,

$$E(X) = M_X'(0) = np$$
$$E(X^2) = M_X''(0) = np + n(n-1)p^2.$$

- Therefore,

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = np + n(n-1)p^2 - n^2p^2 = npq.$$

Example 11.3

- Let X be an exponential random variable with parameter λ. Using moment-generating functions, calculate the mean and the variance of X.

- **Solution.**
- The probability density function of X is given by

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

- Thus,

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{(t-\lambda)x}\, dx.$$

- Since the last integral converges if $t < \lambda$, restricting the domain of $M_X(t)$ to $(-\infty, \lambda)$, we get

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

- Thus,

$$M_X'(t) = \frac{\lambda}{(\lambda - t)^2} \qquad M_X''(t) = \frac{2\lambda}{(\lambda - t)^3}.$$

- We obtain

$$E(X) = M_X'(0) = 1/\lambda, \qquad E(X^2) = M_X''(0) = 2/\lambda^2.$$

- Therefore,

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Example 11.5

- Let Z be a standard normal random variable.

a) Calculate the moment-generating function of Z.

b) Use part a) to find the moment-generating function of X, where X is a normal random variable with mean μ and σ^2.

c) Use part b) to calculate the mean and variance of X.

Solution of Example 11.5

- a) From the definition,

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz - z^2/2}\, dz.$$

- Since

$$tz - \frac{z^2}{2} = \frac{t^2}{2} - \frac{(z-t)^2}{2},$$

we obtain

$$\begin{aligned} M_Z(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[\frac{t^2}{2} - \frac{(z-t)^2}{2}\right] dz \\ &= \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} \exp\left[-\frac{(z-t)^2}{2}\right] dz. \end{aligned}$$

- Let $u = z - t$. Then $du = dz$ and

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{-u^2/2}\, du = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\, du = e^{t^2/2}.$$

This is an integral of a standard normal probability density function.

- b) Letting $Z = \frac{X-\mu}{\sigma}$, we have that Z is $N(0,1)$ and $X = \sigma Z + \mu$.

$E[e^{tX}] = E[e^{t(\sigma Z + \mu)}] = e^{t\mu} E[e^{t\sigma Z}]$

$= e^{t\mu} \cdot M_Z(t\sigma)$

- Thus, by Lemma 11.1,

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} \exp\left(\frac{1}{2}\sigma^2 t^2\right) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

- c) Differentiating $M_X(t)$, we obtain

$$M_X'(t) = (\mu + \sigma^2 t)\exp\left(t\mu + \frac{1}{2}\sigma^2 t^2\right).$$

- Differentiating again gives

$$M_X''(t) = (\mu + \sigma^2 t)^2 \exp\left(t\mu + \frac{1}{2}\sigma^2 t^2\right) + \sigma^2 \exp\left(t\mu + \frac{1}{2}\sigma^2 t^2\right).$$

- Therefore,

$$E(X) = M_X'(0) = \mu, \qquad \text{and} \quad E(X^2) = M_X''(0) = \mu^2 + \sigma^2.$$

- Thus,

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2.$$

Theorem 11.2

- **Theorem 11.2** Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$. If for some $\delta > 0, M_X(t) = M_Y(t)$ for all values of t in $(-\delta, \delta)$, then X and Y have the same distribution.

- The textbook did not offer a proof.
- Theorem 11.2 says that moment-generating functions determines the distribution function uniquely.
- It serves as an important tool in the proof of many important theorems.

Section 11.2

Sum of Independent Random Variables

Convolution Theorem

- Section 8.4 introduced a theorem to compute the density function and the distribution function of the sum of two independent random variables.
- In this section, we extend this result from two random variables to more than two random variables.
    - We use moment generating functions to achieve this goal.

Theorem 8.9 – Convolution Theorem

- Let X and Y be continuous independent random variables with probability density functions f_1 and f_2 and distribution functions F_1 and F_2, respectively.
- Then g and G, the probability density and distribution functions of $X + Y$, respectively, are given by

$$g(t) = \int_{-\infty}^{\infty} f_1(x) f_2(t - x)\, dx,$$

$$G(t) = \int_{-\infty}^{\infty} f_1(x) F_2(t - x)\, dx,$$

$G(t)$
$= P(X+Y<t)$
$= \int_{-\infty}^{\infty} P(X+Y<t \mid X=x) f_1(x)dx$

$= \int_{-\infty}^{\infty} P(x+Y<t \mid X=x) f_1(x)dx = \int_{-\infty}^{\infty} P(Y<t-x \mid X=x) f_1(x)dx$

$= \int_{-\infty}^{\infty} P(Y<t-x) f_1(x)dx = \int_{-\infty}^{\infty} F_2(t-x) f_1(x)dx$

Theorem 11.3

- **Theorem 11.3** Let $X_1, X_2, \ldots, X_n$ be independent random variables with moment-generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$. The moment-generating function of $X_1 + X_2 + \cdots + X_n$ is given by

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t).$$

- **Proof.**
- Let $W = X_1 + X_2 + \cdots + X_n$.

$X_1, X_2, \ldots X_n$ indep

$\Rightarrow h_1(X_1), h_2(X_2), \ldots, h_n(X_n)$ indep

$X_1, X_2, \ldots, X_n$ indep

$\Rightarrow E[X_1 X_2 \ldots, X_n] = E[X_1] \ldots E[X_n]$

- By definition,

$$
\begin{aligned}
M_W(t) = E\left(e^{tW}\right) &= E\left(e^{tX_1+tX_2+\cdots+tX_n}\right) \\
&= E\left(e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right) \\
&= E\left(e^{tX_1}\right)E\left(e^{tX_2}\right)\cdots E\left(e^{tX_n}\right) \\
&= M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t).
\end{aligned}
$$

Sum of Binomial Random Variables

- **Theorem 11.4** Let $X_1, X_2, \ldots, X_r$ be independent binomial random variables with parameters $(n_1, p), (n_2, p), \ldots, (n_r, p)$, respectively. Then, $X_1 + X_2 + \cdots + X_r$ is a binomial random variable with parameters $n_1 + n_2 + \cdots + n_r$ and p.
- **Proof.** Let $q = 1 - p$.
- From Example 11.2, we know that

$$M_{X_i}(t) = \left(pe^t + q\right)^{n_i}, \quad i = 1, 2, \ldots, n.$$

- Let $W = X_1 + X_2 + \cdots + X_r$.
- By Theorem 11.3,

$$\begin{aligned} M_W(t) &= E\left(e^{tW}\right) = E\left(e^{tX_1+tX_2+\cdots+tX_r}\right) \\ &= \left(pe^t + q\right)^{n_1} \left(pe^t + q\right)^{n_2} \cdots \left(pe^t + q\right)^{n_r} \\ &= \left(pe^t + q\right)^{n_1+n_2+\cdots+n_r} \end{aligned}$$

- Note that $(pe^t + q)^{n_1+n_2+\cdots+n_r}$ is the moment-generating function of a binomial random variable with parameters $n_1 + n_2 + \cdots + n_r$ and p.
- The uniqueness property of moment-generating functions (Theorem 11.2) implies that W is binomial with parameters $(n_1 + n_2 + \cdots + n_r, p)$.

Sum of Poisson Random Variables

- **Theorem 11.5** Let $X_1, X_2, \ldots, X_n$ be independent Poisson random variables with means $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. Then, $X_1 + X_2 + \cdots + X_n$ is a Poisson random variable with mean $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.
- **Proof.** Let Y be a Poisson random variable with mean λ.
- Then

$$M_Y(t) = E\left(e^{tY}\right) = \sum_{y=0}^{\infty} e^{ty} \frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{(e^t \lambda)^y}{y!}$$

$$= e^{-\lambda} \exp\left(\lambda e^t\right) = \exp\left[\lambda(e^t - 1)\right].$$

- Let $W = X_1 + X_2 + \cdots + X_n$.
- By Theorem 11.3,

$$\begin{aligned} M_W(t) &= M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t) \\ &= \exp\left[\lambda_1(e^t - 1)\right] \exp\left[\lambda_2(e^t - 1)\right] \cdots \exp\left[\lambda_n(e^t - 1)\right] \\ &= \exp\left[(\lambda_1 + \lambda_2 + \cdots + \lambda_n)(e^t - 1)\right] \end{aligned}$$

- Now, since $\exp\left[(\lambda_1 + \lambda_2 + \cdots + \lambda_n)(e^t - 1)\right]$ is the moment-generating function of a Poisson random variable with mean $\lambda_1 + \lambda_2 + \cdots + \lambda_n$, the uniqueness property implies that $X_1 + X_2 + \cdots + X_n$ is Poisson with mean $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Sum of Normal Random Variables

- **Theorem 11.6** Let $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, ..., $X_n \sim N(\mu_n, \sigma_n^2)$ be independent random variables. Then

$$X_1 + X_2 + \cdots + X_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2).$$

- Prove by the same method.

Other Important Results

- Sums of independent geometric random variables are negative binomial.
- Sums of independent negative binomial random variables are negative binomial.
- Sums of independent exponential random variables are gamma.
- Sums of independent gamma random variables are gamma.

Sum of Weighted Normal Random Variables

- If X is a normal random variable with parameters (μ, σ^2), then for $\alpha \in \mathbf{R}$, we can prove that

$$M_{\alpha X}(t) = \exp\left[\alpha\mu t + (1/2)\alpha^2\sigma^2 t^2\right].$$

- This implies that $\alpha X \sim N(\alpha\mu, \alpha^2\sigma^2)$.
- It follows that linear combinations of sets of independent normal random variables are normal.
- **Theorem 11.7** Let $\{X_1, X_2, \ldots, X_n$ be a set of independent random variables and $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, \ldots, n$; then for constants $\alpha_1, \alpha_2, \ldots, \alpha_n$,

$$\sum_{i=1}^{n} \alpha_i X_i \sim N\left(\sum_{i=1}^{n} \alpha \mu_i, \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2\right).$$

- Let $X_1, X_2, \ldots, X_n$ be independent normal random variables with the same mean μ, and the same variance σ^2.
- Then, $S_n = X_1 + X_2 + \cdots + X_n$ is

$$N(n\mu, n\sigma^2).$$

- The sample mean

$$\bar{X} = \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is

$$N\left(\mu, \frac{\sigma^2}{n}\right).$$

Example 11.9

- Suppose that the distribution of students' grades in a probability test is normal, with mean 72 and variance 25.

a) What is the probability that the average of grade of such a probability class with 25 students is 75 or more?

b) If a professor teaches two different sections of this course, each containing 25 students, what is the probability that the each average of one class is at least three more than the average of the other class?

Solution of Example 11.9

a) Let $X_1, X_2, \ldots, X_{25}$ denote the grades of the 25 students. Then $X_1, X_2, \ldots, X_{25}$ are independent random variables all being normal, with $\mu = 72$ and $\sigma^2 = 25$.

- The average of the grades of the class,

$$\bar{X} = (1/25) \sum_{i=1}^{25} X_i,$$

  is normal, with mean $\mu = 72$ and variance $\sigma^2/n = 25/25 = 1$.

- Hence,

$$P(\bar{X} \geq 75) = P\left(\frac{\bar{X} - 72}{1} \geq \frac{75 - 72}{1}\right)$$
$$= P(\bar{X} - 72 \geq 3) = 1 - \Phi(3) \approx 0.0013.$$

b) Let $\bar{X}$ and $\bar{Y}$ denote the means of the grades of the two classes.

- Then, as seen in part (a), $\bar{X}$ and $\bar{Y}$ are both $N(72, 1)$.
- Since a student does not take two sections of the same course, $\bar{X}$ and $\bar{Y}$ are independent random variables.
- By Theorem 11.7, $\bar{X} - \bar{Y}$ is $N(0, 2)$.

$\alpha_1 = 1 \qquad \alpha_2 = -1$

- By symmetry,

$$\begin{aligned}
P(|\bar{X}-\bar{Y}|>3) &= P(\bar{X}-\bar{Y}>3 \text{ or } \bar{Y}-\bar{X}>3) \\
&= 2P(\bar{X}-\bar{Y}>3) \\
&= 2P\left(\frac{\bar{X}-\bar{Y}-0}{\sqrt{2}} > \frac{3-0}{\sqrt{2}}\right) \\
&= 2P\left(\frac{\bar{X}-\bar{Y}}{\sqrt{2}} > 2.12\right) \\
&= 2[1-\Phi(2.12)] \approx 0.034.
\end{aligned}$$

Sum of Gamma Random Variables

- If $X_1, X_2, \cdots, X_n$ are n independent gamma random variables with parameters $(r_1, \lambda), (r_2, \lambda), \cdots, (r_n, \lambda)$, respectively, then

$$X_1 + X_2 + \cdots + X_n$$

is gamma with parameters

$$(r_1 + r_2 + \cdots + r_n, \lambda).$$

Example 11.11

- Let $X_1, X_2, \ldots, X_n$ be independent standard normal random variables.
- Then,

$$X = X_1^2 + X_2^2 + \cdots + X_n^2,$$

referred to as chi-squared random variable with n degrees of freedom, is gamma with parameters $(n/2, 1/2)$.

- An example of such a gamma random variable is the error of hitting a target in n-dimensional Euclidean space when the error of each coordinate is individually normally distributed.

- Since the sum of n independent gamma random variables, each with parameters $(1/2, 1/2)$, is gamma with parameters $(n/2, 1/2)$, it suffices to prove that for all i, $1 \leq i \leq n$, X_i^2 is gamma with parameters $(1/2, 1/2)$.
- To prove this assertion, note that

$$P(X_i^2 \leq t) = P(-\sqrt{t} \leq X_i \leq \sqrt{t}) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t}).$$

- Differentiating the preceding to obtain the probability density function f, ie

$$f(t) = \left(\frac{1}{2\sqrt{t}}\frac{1}{\sqrt{2\pi}}e^{-t/2}\right) - \left(-\frac{1}{2\sqrt{t}}\frac{1}{\sqrt{2\pi}}e^{-t/2}\right)$$
$$= \frac{1}{\sqrt{2\pi t}}e^{-t/2} = \frac{(1/2)e^{-t/2}(t/2)^{1/2-1}}{\sqrt{\pi}}.$$

- Since $\sqrt{\pi} = \Gamma(1/2)$, it follows that X_i^2 is gamma with parameters $(1/2, 1/2)$.

Homework 13

- Section 11.1: 18, 22, 24
- Section 11.2: 7, 16, 20

- There is no need to submit solutions to this set of homework problems.
- Solution to this homework will be announced in the weekend.

Coverage of the final exam

- 7.2, 7.3, 7.4
- 8.1 – 8.3
- 9.1*
- 10.1 – 10. 4
- 11.1, 11.2

Section 11.3

Markov and Chebyshev inequalities

Theorem 11.8 (Markov's Inequality)

- **Theorem 11.8** Let X be a nonnegative random variable; then for any $t > 0$,
$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Proof of Markov's Inequality

- We prove the theorem for a discrete random variable X with probability mass function $p(x)$.
- For continuous random variables the proof is similar.
- Let A be the set of possible values of X and $B = \{x \in A : x \geq t\}$.
- Then

$$E(X) = \sum_{x \in A} x p(x) \geq \sum_{x \in B} x p(x) \geq t \sum_{x \in B} p(x) = t P(X \geq t).$$

- It follows that

$$P(X \geq t) \leq E(X)/t.$$

Theorem 11.9 (Chebyshev's Inequality)

- **Theorem 11.9** If X is a random variable with expected value μ and variance σ^2, then for any $t > 0$,

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

- **Proof.**
- Since $(X - \mu)^2 \geq 0$, by Markov's inequality

$$P\left((X - \mu)^2 \geq t^2\right) \leq \frac{E[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}.$$

- Chebyshev's inequality follows since $(X - \mu)^2 \geq t^2$ is equivalent to $|X - \mu| \geq t$.

A Remark on Chebyshev's Inequality

- Letting $t = k\sigma$ in Chebyshev's inequality, we get that

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

- That is, the probability that X deviates from its expected value at least k standard deviations is at most $1/k^2$.
- On the other hand,

$$P(|X - \mu| < k\sigma) \geq 1 - 1/k^2.$$

- The probability that X deviates from its mean less than k standard deviations is at least $1 - 1/k^2$.

Applications of Markov’s and Chebyshev’s Inequalities

- Markov’s inequality and Chebyshev’s inequality are useful to derive
  - estimates or
  - bounds

  for probabilities based on knowledge of moments only.

Section 11.4

Laws of Large Numbers

Definition of Convergence

- Recall that a sequence of real numbers $a_1, a_2, \cdots$ is said to converge to a if for any $\varepsilon > 0$, there exists an integer $N > 0$, such that for any $n > N$,

$$|a_n - a| < \epsilon$$

- Random variables are actually functions from the sample space to the real line.

$$X(\omega) : S \to R$$

- How do we define convergence of random variables, i.e. convergence of a sequence of functions?
- There are many ways!

Point-wise Convergence of Functions

- Point-wise convergence: A sequence of functions $\{f_n\}$ converges point-wise to function f, if for x in the domain

$$\lim_{n \to \infty} f_n(x) = f(x)$$

Converge Almost Surely

- Let V be the set of all points ω, in S, at which $X_n(\omega)$ converges to $X(\omega)$, i.e.

$$V = \{\omega \in S : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$$

- If $P(V) = 1$, we say that X_n converge to X almost surely.
  - Or, we say that X_n converge to X with probability 1.
- Almost surely convergence is similar to point-wise convergence except that it does not require the sequence of functions to converge to the limit at all points.
  - There can be an exceptional set of points.
  - The probability measure of this exceptional set is 0.

Converge Almost Surely

- $X_n(\omega)$ converges almost surely to $X(\omega)$, if

$$P\left(\left\{\omega \in S : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

- Almost surely convergence is also called convergence with probability 1.

A Counter Example

- Consider an experiment in which we choose a random point from the interval $[0, 1]$.
- The sample space is $[0, 1]$.
- For all $\omega \in [0, 1]$, let $X_n(\omega) = \omega^n$ and $X(\omega) = 0$.
- Then, clearly

$$V = \left\{\omega \in [0, 1]\colon \lim_{n\to\infty} X_n(\omega) = W(\omega)\right\} = [0, 1)$$

- Since $P(V) = P([0, 1]) = 1$, X_n converges to X almost surely.
- But, this convergence is not point-wise.

Convergence in Probability

- **Definition.** Let $X_1, X_2, \cdots$ be a sequence of random variables defined on a sample space S. We say that X_n converges to a random variable X in probability if, for each $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0\,.$$

- Almost surely convergence implies convergence in probability.
- The converse is not true in general.

Convergence in Distribution

- Let $F(t)$ and $F_n(t)$ be the distribution function of X and X_n, respectively.
- We say that X_n converge in distribution to X (or F_n converges in distribution to F) if for all t at which F is continuous

$$\lim_{n\to\infty} F_n(t) = F(t).$$

- Convergence in probability implies convergence in distribution.
- The converse is in general not true.

Convergence in Quadric Moments

- We say that X_n converges to X in quadratic mean if

$$\lim_{n\to\infty} E(|X_n - X|^2) = 0$$

Theorem 11.10 (Weak Law of Large Numbers)

Theorem 11.10 (Weak Law of Large Numbers) *Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed random variables with $\mu = E(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i) < \infty$, $i = 1, 2, \ldots$. Then $\forall \varepsilon > 0$,*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

- The weak law of large number says that $(\frac{1}{n})(X_1 + X_2 + \cdots + X_n)$ **converges in probability** to the common mean μ.

Theorem 11.11
(Strong Law of Large Numbers)

Theorem 11.11 (Strong Law of Large Numbers) *Let* $X_1, X_2, \ldots$ *be an independent and identically distributed sequence of random variables with* $\mu = E(X_i)$, $i = 1, 2, \ldots$ *. Then*

$$\boldsymbol{P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right) = 1.}$$

- The strong law of large number says that $(\frac{1}{n})(X_1 + X_2 + \cdots + X_n)$ converges to μ **almost surely**.
- This book did not offer a proof of this theorem.

Section 11.5

- Central Limit Theorem

De Moivre-Laplace theorem

- The De Moivre-Laplace theorem is restated as follows:
- Let $X_1, X_2, X_3, \ldots$ be a sequence of independent Bernoulli random variables each with parameter p. Then, $\forall i$, $E(X_i) = p$, $\mathrm{Var}(X_i) = p(1-p)$, and

$$\lim_{n\to\infty} p\left(\frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

- The De Moivre-Laplace theorem motivates elegant generalizations.
- The first was given in 1887 by Chebyshev. But since its proof was not rigorous enough.
- Chebyshev's student Markov worked on the proof and made it both rigorous and simpler.
- In 1901, Lyapunov, another student of Chebyshev, weakened the conditions of Chebyshev and proved the following generalization of the De Moivre-Laplace theorem, which is now called the central limit theorem,

Theorem 11.12 (Central Limit Theorem)

- **Theorem 11.12 (Central Limit Theorem)** Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each with expectation μ and σ^2. Then the distribution of

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

converges to the distribution of a standard normal random variable. That is,

$$\lim_{n\to\infty} P(Z_n \le x) = \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le x\right)$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

Two Remarks

- **Remark 11.2** Note that Z_n is simply $X_1 + X_2 + \cdots + X_n$ standardized.
- **Remark 11.2** In Theorem 11.12, let $\bar{X}$ be the mean of the random variables $X_1, X_2, \ldots, X_n$, ie

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

the central limit theorem is equivalent to

$$\lim_{n \to \infty} p\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$