Chapter 8 Limit Theorems

8.1 Introduction

- The most important theoretical results in probability theory are limit theorems.
- **Laws of large numbers**: The average of a sequence of random variables converges to the expected average.
- **Central limit theorems**: The sum of a large number of random variables has a probability distribution that is approximately normal.

8.2 Chebyshev's inequality and the weak law of large numbers

Proposition 2.1 Markov's inequality: If X is a random variable that takes only non-negative values, then for any value $a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

- For $a > 0$,

$$I = \begin{cases} 1 & \text{if } X \geq a, \\ 0 & \text{otherwise} \end{cases}$$

- $I \leq \frac{X}{a}$
- $E[I] = P\{X \geq a\} \leq \frac{E[X]}{a}$

Proposition 2.2 Chebyshev's inequality: If X is a random variable with finite mean μ and variance σ^2, then for any value $k > 0$,

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

- $P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$
- $P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$

- The importance of **Markov's** and **Chebyshev's** inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known.

- If the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds.

Example 8.2a. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

(a)

- By **Markov's inequality**

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b)

- By **Chebyshev's inequality**

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

- Hence

$$P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

Example 8.2b. If X is uniformly distributed over the interval $(0, 10)$, then, as $E[X] = 5$, $\text{Var}(X) = \frac{25}{3}$.

- It follows from Chebyshev's inequality that

$$P\{|X - 5| > 4\} \leq \frac{25}{3(16)} \approx .52$$

whereas the exact result is

$$P\{|X - 5| > 4\} = .20$$

- Thus, although Chebyshev's inequality is correct, the upper bound that it provides is not particularly close to the actual probability.
- Similarly, if X is a normal random variable with mean μ and variance σ^2.
- **Chebyshev's inequality** states that

$$P\{|X-\mu| > 2\sigma\} \leq \frac{1}{4}$$

whereas the actual probability is given by

$$P\{|X-\mu| > 2\sigma\} = P\left\{\left|\frac{X-\mu}{\sigma}\right| > 2\right\} = 2[1-\Phi(2)] \approx .0456$$

Proposition 2.3: If $\mathrm{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

- $P\{|X-\mu| > 1/n\} = 0$ for any $n \geq 1$.
- $0 = \lim_{n\to\infty} P\{|X-\mu| > 1/n\} = P\{\lim_{n\to\infty} |X-\mu| > 1/n\} = P\{X \neq \mu\}$.

Theorem 2.1 The weak law of large numbers: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \text{as } n \to \infty$$

- Assume the additional assumption that the random variables have variance σ^2.
- $E\left[\frac{X_1+\cdots+X_n}{n}\right] = \mu$
- $\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{\sigma^2}{n}$
- $P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$
- The weak law of large number was originally proved by James Bernoulli for the special case where the $X_i \sim$ Bernoulli random variables.
- The general form of the weak law of large

numbers presented in Theorem 2.1 was proved by the Russian mathematician Khintchine.

8.3 The central limit theorem

- The central limit theorem is one of the most remarkable results in probability theory.
- It not only provides a simple method for computing approximate probabilities for sums of independent random variables, but it also helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit bell-shaped curves.

Theorem 3.1 The central limit theorem: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2. Then the distribution of

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

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx \quad \text{as } n \to \infty$$

Lemma 3.1: Let $Z_1, Z_2, \ldots$ be a sequence of random variables having distribution function F_{Z_n} and moment generating functions M_{Z_n}, $n \geq 1$; and let Z be a random variable having distribution function F_Z and moment generating function M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is continuous.

Proof of the Central Limit Theorem

- Assume that $\mu = 0$ and $\sigma^2 = 1$.
- $M_{X_i}(t) = E[e^{tX_i}] = M(t)$
- $M_{X_i/\sqrt{n}}(t) = M\left(\frac{t}{\sqrt{n}}\right)$
- $M_{\Sigma_{i=1}^n X_i/\sqrt{n}}(t) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$
- $L(t) = \log M(t)$
  - $L(0) = L'(0) = 0$ and $L''(0) = 1$.
- $\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{1/n} = \frac{t^2}{2}$

Remark.

- It can be shown that the convergence of Theorem 3.1 is uniform in a.
- The first version of the central limit theorem was proved by DeMoivre around 1733 for the special case where X_i are Bernoulli random variables with $p = 1/2$.
- This was subsequently extended by Laplace to the case of arbitrary p.

- Laplace also discovered the more general form of the central limit theorem given in Theorem 3.1.
- A truly rigorous proof of the central limit theorem was first presented by the Russian mathematician Liapounoff in the period 1901-1902.
- This important theorem is illustrated by the central limit theorem module on the text diskette (Fig. 8.1).
- $p_k = P\{X_i = k\}$
- $p_0 = .25, p_1 = .15, p_2 = .1, p_3 = .2, p_4 = .3$
- $E\left[\sum_1^5 X_i\right] = 10.75, \mathrm{Var}\left(\sum_1^5 X_i\right) = 12.6375$
- $E\left[\sum_1^{10} X_i\right] = 10.75, \mathrm{Var}\left(\sum_1^{10} X_i\right) = 12.6375$
- $E\left[\sum_1^{25} X_i\right] = 53.75, \mathrm{Var}\left(\sum_1^{10} X_i\right) = 63.1875$
- $E\left[\sum_1^{100} X_i\right] = 215, \mathrm{Var}\left(\sum_1^{10} X_i\right) = 252.75$

Example 8.3a. An astronomer is interested in measuring, in light years, the distance from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance but merely an estimate. As a result the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated values of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variance of 4 (light years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within $\pm.5$ light year?

- Suppose that the astronomer decides to make n observations.

- If $X_1, \ldots, X_n$ are the n measurements, then

$$Z_n = \frac{\sum_{i=1}^n X_i - nd}{2\sqrt{n}} \approx N(0,1)$$

- $$\begin{aligned} P\left\{-.5 \leq \frac{\sum_{i=1}^n X_i}{n} - d \leq .5\right\} &= P\left\{-.5\frac{\sqrt{n}}{2} \leq Z_n \leq .5\frac{\sqrt{n}}{2}\right\} \\ &\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \end{aligned}$$

- $2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = .95$ or $\Phi\left(\frac{\sqrt{n^*}}{4}\right) = .975$
- $\frac{\sqrt{n^*}}{4} = 1.96$ or $n^* = (7.84)^2 \approx 61.47$
- $E\left[\sum_{i=1}^n \frac{X_i}{n}\right] = d \qquad \text{Var}\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \frac{4}{n}$
- Chebyshev's inequality yields that

$$P\left\{\left|\sum_{i=1}^n \frac{X_i}{n} - d\right| > .5\right\} \leq \frac{4}{n(.5)^2} = \frac{16}{n}$$

- If he makes $n = 16/.05 = 320$ observations, he can be 95 percent certain that his estimate will be accurate to within .5 light year.

Example 8.3b. The number of students that enroll in a psychology course is a Poisson random variable with mean 100. The professor in charge of the course has decided that of the number enrolling is 120 or more he will teach the course in two separate sections, whereas if fewer than 120 students enroll he will teach all of the students together in a single section. What is the probability that the professor will have to teach two sections?

- The exact solution $e^{-100}\sum_{i=120}^{\infty}(100)^i/i!$.
- A Poisson random variable with mean 100 is the sum of 100 independent Poisson random variables each with mean 1.
- Use the central limit theorem to obtain an approximate solution.
- X: The number of students that enroll in the course.

$$\begin{aligned} P\{X \geq 120\} &= P\left\{\frac{X-100}{\sqrt{100}} \geq \frac{120-100}{\sqrt{100}}\right\} \\ &\approx 1-\Phi(2) \\ &\approx .0228 \end{aligned}$$

Example 8.3c. If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40.

- X_i: The value of the ith die.
- $E[X_i] = 7/2$ and $\text{Var}(X_i) = 35/12$.

$$
\begin{aligned}
P\left\{30 \le \sum_{i=1}^{10} X_i \le 40\right\} &= P\left\{\frac{30-35}{\sqrt{\frac{350}{12}}} \le \frac{\Sigma_{i=1}^{10} X_i - 35}{\sqrt{\frac{350}{12}}} \le \frac{40-35}{\sqrt{\frac{350}{12}}}\right\} \\
&\approx 2\Phi\left(\sqrt{\frac{6}{7}}\right) - 1 \\
&\approx .65
\end{aligned}
$$

Example 8.3d. Let $X_i, i = 1, \ldots, 10$ be independent random variables, each uniformly distributed over (0,1). Calculate an approximation to $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$.

-

$$
P\left\{\sum_{1}^{10} X_i > 6\right\} = P\left\{\frac{\Sigma_1^{10} X_i - 5}{\sqrt{10(\frac{1}{12})}} > \frac{6-5}{\sqrt{10(\frac{1}{12})}}\right\}
$$

$$\approx 1 - \Phi(\sqrt{1.2})$$
$$\approx .16$$

- Hence only 16 percent of the time will $\sum_{i=1}^{10} X_i$ will be greater than 6.

Central limit theorems also exist when the X_i are independent but not necessarily identically distributed random variables.

Theorem 3.2 Central limit theorem for independent random variables: Let $X_1, X_2, \ldots$ be a sequence of independent random variables having respective means and variances $\mu_i = E[X_i], \sigma_i^2 = \mathrm{Var}(X_i)$. If (a) the X_i are uniformly bounded; that is, if for some M, $P\{|X_i| < M\} = 1$ for all i, and (b) $\Sigma_{i=1}^{\infty} \sigma_i^2 = \infty$, then

$$P\left\{\frac{\Sigma_{i=1}^{n}(X_i - \mu_i)}{\sqrt{\Sigma_{i=1}^{n} \sigma_i^2}} \le a\right\} \to \Phi(a) \quad \text{as } n \to \infty$$

8.4 The strong law of large numbers

- The *strong law of large numbers* is probably the best-known result in probability theory.
- It states that the average of a sequence of independent of random variables having a common distribution will, with probability 1, converge to the mean of that distribution.

Theorem 4.1 The strong law of large numbers: Let $X_1, X_2 \dots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as } n \to \infty^{\dagger}$$

Application:

- Suppose that a sequence of independent trials of some experiment is performed.
- Let E be a fixed event of the experiment and denote by $P(E)$ the probability that E

occurs on any particular trial.

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0 & \text{if } E \text{ does not occur on the } i\text{th trial} \end{cases}$$

- $\frac{X_1+X_2+\cdots+X_n}{n} \to E[X] = P(E)$

Proof of the Strong Law of Large Numbers:

- Assume that $\mu = E[X_i] = 0$ and $E[X_i^4] = K < \infty$.
- $S_n = \sum\limits_{i=1}^{n} X_i$
- $E[S_n^4]$: $X_i^4, X_i^3X_j, X_i^2X_j^2, X_i^2X_jX_k, X_iX_jX_kX_l$

$$\begin{aligned} E[X_i^3X_j] &= E[X_i^3]E[X_j] = 0 \\ E[X_i^2X_jX_k] &= E[X_i^2]E[X_j]E[X_k] = 0 \\ E[X_iX_jX_kX_l] &= E[X_i]E[X_j]E[X_k]E[X_l] = 0 \\ E[X_i^2X_j^2] &= E[X_i^2]E[X_j^2] \end{aligned}$$

- $E[S_n^4] = nE[X_i^4] + 6\binom{n}{2}E[X_i^2X_j^2]$
- $\mathrm{Var}(X_i^2) = E[X_i^4] - (E[X_i^2])^2 \geq 0$

- $E[S_n^4] = nK + 3n(n-1)K$
- $E\left[\frac{S_n^4}{n^4}\right] \le \frac{K}{n^3} + \frac{3K}{n^2}$
- $E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] \le \sum_{n=1}^{\infty} \frac{K}{n^3} + \frac{3K}{n^2} < \infty$
- With probability 1, $\lim_{n\to\infty} \frac{S_n^4}{n^4} = 0 \Leftrightarrow \lim_{n\to\infty} \frac{S_n}{n} = 0$.
- If $\mu = E[X_i] \neq 0$, consider $X_i - \mu$.

Illustrations of strong law:

$$p_k = P\{X_i = k\}, \quad E[X_i] = 2.05$$
$$p_0 = .1, p_1 = .2, p_2 = .3, p_3 = .35, p_4 = .05$$

- $n = 100, \bar{X} = 1.89$;
- $n = 1000, \bar{X} = 2.078$;
- $n = 10000, \bar{X} = 2.0416$

Weak law of large numbers v.s. strong law of large numbers:

- WLLN states that for any specified large value n^*, $(X_1 + \cdots + X_{n^*})/n^*$ is likely to be near μ. However, it does not say that $(X_1 + \cdots + X_n)/n$ is bound to stay near μ for all values of n larger than n^*.
- SLLN states that with probability 1, for any positive value ϵ,
$$\left| \sum_1^n \frac{X_i}{n} - \mu \right|$$
will be greater than ϵ only a finite number of times.
- The strong law of large numbers was originally proved, in the special case of Bernoulli random variables, by the French mathematician Borel.
- The general form of the strong law presented in Theorem 4.1 by the Russian mathematician A.N. Kolmogorov.

8.5 Other inequalities

Only μ and σ^2 are known:

Chebyshev inequality:

$$P\{X-\mu \geq a\} \leq P\{|X-\mu| \geq a\} \leq \frac{\sigma^2}{a^2}$$

Proposition 5.1 One-sided Chebyshev inequality: If X is a random variable with mean 0 and finite variance σ^2, then for any $a > 0$,

$$P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

- If $b > 0$, then $X \geq a \iff X + b \geq a + b$.
- $P\{X \geq a\} \leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{\sigma^2+b^2}{(a+b)^2}$
- $\min_b \frac{\sigma^2+b^2}{(a+b)^2}$ occurs at $b = \sigma^2/a$.
- Let $b = \sigma^2/a$, then $P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2+a^2}$.

Example 8.5a. If the number of items produced in a factory during a week is a random variable with mean 100 and variance 400, compute an upper bound on the probability that this week's production will be at least 120.

- One-sided Chebyshev inequality:
  $P\{X \geq 120\} = P\{X - 100 \geq 20\} \leq \frac{400}{400+(20)^2} = \frac{1}{2}$
- Markov's inequality:
  $P\{X \geq 120\} \leq \frac{E(X)}{120} = \frac{5}{6}$

Corollary 5.1: If $E[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, then for $a > 0$,

$$P\{X \geq \mu + a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P\{X \leq \mu - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Example 8.5b. A set of 200 people, consisting of 100 men and 100 women, is randomly divided into 100 pairs of 2 each. Give an upper bound to the probability that at most 30 of these pairs will consist of a man and a woman.

- $$X_i = \begin{cases} 1 & \text{if man } i \text{ is paired with a woman} \\ 0 & \text{otherwise} \end{cases}$$
- $X = \sum_{i=1}^{100} X_i$
- $E[X_i] = P\{X_i = 1\} = \frac{100}{199}$
- Similarly, for $i \neq j$,
$$\begin{aligned} E[X_iX_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\}P\{X_j = 1 | X_i = 1\} = \frac{100}{199}\frac{99}{197} \end{aligned}$$
- $$\begin{aligned} E[X] &= \sum_{i=1}^{100} E[X_i] \\ &= (100)\frac{100}{199} \\ &\approx 50.25 \end{aligned}$$
- $$\begin{aligned} \mathrm{Var}(X) &= \sum_{i=1}^{100} \mathrm{Var}(X_i) + 2\sum\sum_{i<j} \mathrm{Cov}(X_i, X_j) \\ &= 100\frac{100}{199}\frac{99}{199} + 2\binom{100}{2}\left[\frac{100}{199}\frac{99}{197} - \left(\frac{100}{199}\right)^2\right] \\ &\approx 25.126 \end{aligned}$$

- The Chebyshev inequality yields that

  $P\{X \le 30\} \le P\{|X - 50.25| \ge 20.25\} \le \dfrac{25.126}{(20.25)^2} \approx .061$

- $$\begin{aligned} P\{X \le 30\} &= P\{X \le 50.25 - 20.25\} \\ &\le \frac{25.126}{25.126 + (20.25)^2} \\ &\approx .058 \end{aligned}$$

- $M(t) = E[e^{tX}]$

- For $t > 0$

  $$\begin{aligned} P\{X \ge a\} &= P\{e^{tX} \ge e^{ta}\} \\ &\le E[e^{tX}]e^{-ta} \quad \text{by Markov's inequality} \end{aligned}$$

- Similarly, for $t < 0$,

  $$\begin{aligned} P\{X \le a\} &= P\{e^{tX} \ge e^{ta}\} \\ &\le E[e^{tX}]e^{-ta} \end{aligned}$$

Proposition 5.2 Chernoff bounds:

$$\begin{aligned} P\{X \ge a\} &\le e^{-ta}M(t) \quad \text{for all } t > 0 \\ P\{X \le a\} &\le e^{-ta}M(t) \quad \text{for all } t < 0 \end{aligned}$$

The best bound on $P\{X \geq a\}$ uses the t that minimizes $e^{-ta}M(t)$.

Example 8.5c. *Chernoff bounds for the standard normal random variable.* If Z is a standard normal random variable, then its moment generating function is $M(t) = e^{t^2/2}$.

- Chernoff bound on $P\{Z \geq a\}$ is given by
$$P\{Z \geq a\} \leq e^{-ta}e^{t^2/2} \quad \text{for all } t > 0$$
- Now the value of t, $t > 0$, that minimizes $e^{t^2/2-ta}$ is the value that minimizes $t^2/2 - ta$, which is $t = a$.
- Thus for $a > 0$ we see that
$$P\{Z \geq a\} \leq e^{-a^2/2}$$
- Similarly, we can show that for $a < 0$,
$$P\{Z \leq a\} \leq e^{-a^2/2}$$

Example 8.5d. *Chernoff bounds for the Poisson random variable.* If X is a Poisson random variable with parameter λ, then its moment generating function is $M(t) = e^{\lambda(e^t - 1)}$.

- Chernoff bound on $P\{X \geq i\}$ is
$$P\{X \geq i\} \leq e^{\lambda(e^t-1)} e^{-it} \quad t > 0$$
- Minimizing the right side of the above is equivalent to minimizing $\lambda(e^t - 1) - it$, and calculus shows that the minimal value occurs when $e^t = i/\lambda$.
- Provided that $i/\lambda > 1$, this minimizing values of i will be positive.
- Therefore, assuming that $i > \lambda$ and letting $e^t = i/\lambda$ in the Chernoff bound yields that
$$P\{X \geq i\} \leq e^{\lambda(i/\lambda - 1)} \left(\frac{\lambda}{i}\right)^i$$
or, equivalently,
$$P\{X \geq i\} \leq \frac{e^{-\lambda} (e\lambda)^i}{i^i}$$

Example 8.5e. Consider a gambler who on every play is equally likely, independent of the past, to either win or lose 1 unit. That is, if X_i is the gambler's

$$P\{X_i = 1\} = P\{X_i = -1\} = \frac{1}{2}$$

- $S_n = \Sigma_{i=1}^n X_i$ denote the gambler's winnings after n plays.
- Use the Chernoff bound on $P\{S_n \geq a\}$.
- The moment generating function of X_i is

$$E[e^{tX}] = \frac{e^t + e^{-t}}{2}$$

- Now, using the McLaurin expansions of e^t and e^{-t}

$$\begin{aligned}
e^t + e^{-t} &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots + \left(1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \cdots\right) \\
&= 2\left\{1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots\right\} \\
&= 2 \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} \\
&\leq 2 \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} \quad \text{since } (2n)! \geq n!2^n \\
&= 2e^{t^2/2}
\end{aligned}$$

- Therefore, $E[e^{tX}] \leq e^{t^2/2}$.
- Since the moment generating function of the sum of independent random variables is the product of their moment generating functions, we have that

$$\begin{aligned} E[e^{tS_n}] &= (E[e^{tX}])^n \\ &\leq e^{nt^2/2} \end{aligned}$$

- Using the result above along with the Chernoff bound given that

$$P\{S_n \geq a\} \leq e^{-ta} e^{nt^2/2} \quad t > 0$$

- The value of t that minimizes the right side of the above is the value that minimizes $nt^2/2 - ta$, and this values is $t = a/n$.
- Supposing that $a > 0$ (so that this minimizing t is positive) and letting $t = a/n$ in the preceding inequality yields that

$$P\{S_n \geq n\} \leq e^{-a^2/2n} \quad a > 0$$

- For instance, this inequality yields that

$$P\{S_{10} \geq 6\} \leq e^{-36/20} \approx .1653$$

whereas the exact probability is

$$\begin{aligned} P\{S_{10} \geq 6\} &= P\{\text{gambler wins at least 8 of the first 10 games}\} \\ &= \frac{\binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} = \frac{56}{1024} \approx .0547 \end{aligned}$$

Definition: A twice-differentiable real-valued function $f(x)$ is said to be *convex* if $f''(x) \geq 0$ for all x; similarly, it is said to be *concave* if $f''(x) \leq 0$.

- Convex functions: $f(x) = x^2, e^{ax}, -x^{1/n}$ for $x \geq 0$.
- If $f(x)$ is convex, then $g(x) = -f(x)$ is concave.

Proposition 5.3 Jensen's inequality: If $f(x)$ is a convex function, then

$$E[f(X)] \geq f(E[X])$$

provided that the expectations exist and are finite.

- $f(x) = f(\mu) + f'(\mu)(x-\mu) + f''(\xi)(x-\mu)^2/2$
- $f(x) \geq f(\mu) + f'(\mu)(x-\mu)$ since $f''(\xi) \geq 0$.
- $E[f(X)] \geq f(\mu) + f'(\mu)E[X-\mu] = f(\mu)$

Example 8.5f. An investor is faced with the following choices: She can either invest all of her money in a risky proposition that would lead to a random return X that has mean m; or she can put the money into a risk-free venture that will lead to a return of m with probability 1.

- Suppose that her decision will be made on the basis of maximizing the expected value of $u(R)$, where R is her return and u is her utility function.
- By Jensen's inequality it follows that if u is a concave function, then $E[u(X)] \leq u(m)$, so the risk-free alternative is preferable; whereas if u is convex, then $E[u(X)] \geq u(m)$, so the

risk investment alternative would be preferred.

8.6 Bounding the error probability when approximating a sum of independent Bernoulli random variables by a Poisson random variable

- $X_i \sim \text{Bernoulli}(p_i)$
- $Y_i \sim \text{Poisson}(p_i)$
- Construct a sequence of independent Bernoulli random variables $X_1, \ldots, X_n$ with parameters $p_1, \ldots, p_n$ such that $P\{X_i \neq Y_i\} \leq p_i^2$ for each i.
- $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{n} Y_i$.
- $P\{X \neq Y\} \leq \sum_{i=1}^{n} p_i^2$
- Next we will show that

$$|P\{X \in A\} - P\{Y \in A\}| \leq \sum_{i=1}^{n} p_i^2$$

  – $U_i \sim \text{Bernoulli}(1 - (1 - p_i)e^{p_i})$

– $(1-p_i)e^{p_i} \le 1$ since $e^{-p} \ge 1-p$.

– $X_i = 0$ if $Y_i = U_i = 0$ and 1 otherwise.

– $P\{X_i = 0\} = P\{Y_i = 0\}P\{U_i = 0\} = 1 - p_i$

– $P\{X_i = 1\} = p_i$

$$\begin{aligned} P\{X_i \neq Y_i\} &= P\{X_i = 1, Y_i \neq 1\} \\ &= P\{Y_i = 0, X_i = 1\} + P\{Y_i > 1\} \\ &= P\{Y_i = 0, U_i = 1\} + P\{Y_i > 1\} \\ &= p_i - p_i e^{-p_i} \\ &= p_i^2 \end{aligned}$$

– $X \neq Y$ implies that $X_i \neq Y_i$ for some i.

$$\begin{aligned} P\{X \neq Y\} &\le P\{X_i \neq Y_i \text{ for some } i\} \\ &\le \sum_{i=1}^{n} P\{X_i \neq Y_i\} \\ &\le \sum_{i=1}^{n} p_i^2 \end{aligned}$$

– For any event B, $I_B = 1$ if B occurs and 0 otherwise.

– $I_{\{X \in A\}} - I_{\{Y \in A\}} \le I_{\{X \neq Y\}}$

– $P\{X \in A\} - P\{Y \in A\} \le P\{X \neq Y\}$

- $|P\{X \in A\} - P\{Y \in A\}| \leq P\{X \neq Y\}$
- If $\lambda = \sum_{i=1}^{n} p_i$,

$$\left| P\left\{ \sum_{i=1}^{n} X_i \in A \right\} - \sum_{i \in A} \frac{e^{-\lambda}\lambda^i}{i!} \right| \leq \sum_{i=1}^{n} p_i^2$$

- If $p_i = p$ and $X \sim \text{Binomial}(p)$, then

$$\left| \sum_{i \in A} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i \in A} \frac{e^{-np}(np)^i}{i!} \right| \leq np^2$$

Summary

- **Markov inequality**:

$$P\{X \geq a\} \leq \frac{E[X]}{a} \qquad a > 0$$

- **Chebyshev inequality**:

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2} \qquad k > 0$$

- **Strong law of large numbers**

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \qquad \text{as } n \to \infty$$

• Central limit theorem

$$\lim_{n\to\infty} P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}\,dx$$