Introduction to Probability

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 308

1. Limit Theorems
   - Chebyshev's Inequality and Weak Law of Large Numbers
   - The Central Limit Theorem
   - The Strong Law Of Large Numbers
   - Other Inequalities
   - Approximating a Sum of Bernoulli by a Poisson Variable

Subsection 1

Chebyshev's Inequality and Weak Law of Large Numbers

Markov's Inequality

Proposition (Markov's inequality)

If X is a random variable that takes only nonnegative values, then, for any value $a > 0$,

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

- For $a > 0$, let $I = \begin{cases} 1, & \text{if } X \geq a \\ 0, & \text{otherwise} \end{cases}$.

  Note that, since $X \geq 0$, $I \leq \frac{X}{a}$.

  Now we get

$$P\{X \geq a\} = E[I] \leq E\left[\frac{X}{a}\right] \leq \frac{E[X]}{a}.$$

Chebyshev's Inequality

Proposition (Chebyshev's Inequality)

If X is a random variable with finite mean μ and variance σ^2, then, for any value $k > 0$,

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

- Note $(X-\mu)^2$ is a nonnegative random variable.
  So we can apply Markov's inequality (with $a = k^2$) to obtain

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

  But $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$.
  Hence, the last equation is equivalent to

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

Example

- 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?

Let X be the number of items that will be produced in a week.

(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, $P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$.
So the probability that this week's production will be between 40 and 60 is at least 0.75.

Example: Laxity of Bounds

- Suppose X is uniformly distributed over the interval $(0, 10)$.
- We have

$$E[X] = 5 \quad \text{and} \quad \text{Var}(X) = \frac{25}{3}.$$

  Thus, from Chebyshev's inequality,

$$P\{|X - 5| > 4\} \leq \frac{25}{3(16)} \approx 0.52.$$

- The exact result is

$$P\{|X - 5| > 4\} = 0.20.$$

- Although Chebyshev's inequality is correct, the upper bound that it provides is not particularly close to the actual probability.

Example: Laxity of Bounds

- Suppose X is a normal random variable with mean μ and variance σ^2.
- Chebyshev's inequality states that

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

- The actual probability is given by

$$P\{|X-\mu|>2\sigma\} = P\left\{\left|\frac{X-\mu}{\sigma}\right|>2\right\} = 2[1-\Phi(2)] \approx 0.0456.$$

Random Variables With Zero Variance

Proposition

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

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

In other words, the only random variables having variances equal to 0 are those which are constant with probability 1.

- By Chebyshev's Inequality, we have, for any $n \geq 1$,

$$P\left\{|X - \mu| > \frac{1}{n}\right\} = 0.$$

Letting $n \to \infty$ and using the continuity property of probability yields

$$0 = \lim_{n\to\infty} P\left\{|X - \mu| > \frac{1}{n}\right\} = P\left\{\lim_{n\to\infty}\left\{|X - \mu| > \frac{1}{n}\right\}\right\} = P\{X \neq \mu\}.$$

The Weak Law of Large Numbers

Theorem (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 $\varepsilon > 0$,

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

- We shall prove the theorem only under the additional assumption that the random variables have a finite variance σ^2.
  We know that
  $$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \quad \text{and} \quad \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$
  Thus, by Chebyshev's inequality,
  $$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Subsection 2

The Central Limit Theorem

A Technical Lemma

Lemma

Let $Z_1, Z_2, \ldots$ be a sequence of random variables having distribution functions F_{Z_n} and moment generating functions M_{Z_n}, $n \geq 1$. 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.

Example:

Let Z be a standard normal random variable.

Then $M_Z(t) = e^{t^2/2}$.

Suppose that $M_{Z_n}(t) \to e^{t^2/2}$ as $n \to \infty$.

Then, by the lemma, $F_{Z_n}(t) \to \Phi(t)$ as $n \to \infty$.

The Central Limit Theorem

Theorem (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 $-\infty < a < \infty$,

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

- Let us assume at first that $\mu = 0$ and $\sigma^2 = 1$.

  We shall prove the theorem under the assumption that the moment generating function of the X_i, $M(t)$, exists and is finite.

The Central Limit Theorem (Cont'd)

- The moment generating function of $\frac{X_i}{\sqrt{n}}$ is given by

$$E\left[\exp\left\{\frac{tX_i}{\sqrt{n}}\right\}\right] = M\left(\frac{t}{\sqrt{n}}\right).$$

Thus, the moment generating function of $\sum_{i=1}^n \frac{X_i}{\sqrt{n}}$ is $[M(\frac{t}{\sqrt{n}})]^n$.

Let $L(t) = \log M(t)$.

Note that

$$\begin{aligned} L(0) &= 0; \\ L'(0) &= \frac{M'(0)}{M(0)} = \mu = 0; \\ L''(0) &= \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = E[X^2] = 1. \end{aligned}$$

The Central Limit Theorem (Cont'd)

- We must now show that $[M(\frac{t}{\sqrt{n}})]^n \to e^{t^2/2}$ as $n \to \infty$.

  Equivalently, that $nL(\frac{t}{\sqrt{n}}) \to \frac{t^2}{2}$ as $n \to \infty$.

  We have

$$\begin{array}{rcll}
\lim_{n\to\infty} \frac{L(\frac{t}{\sqrt{n}})}{\frac{1}{n}} & = & \lim_{n\to\infty} \frac{-L'(\frac{t}{\sqrt{n}})\frac{t}{\sqrt{n^3}}}{-\frac{2}{n^2}} & \text{(by L'Hôpital's rule)} \\
& = & \lim_{n\to\infty} \left[\frac{L'(\frac{t}{\sqrt{n}})t}{\frac{2}{\sqrt{n}}}\right] & \\
& = & \lim_{n\to\infty} \left[\frac{-L''(\frac{t}{\sqrt{n}})\frac{t^2}{\sqrt{n^3}}}{-\frac{2}{\sqrt{n^3}}}\right] & \text{(by L'Hôpital's rule)} \\
& = & \lim_{n\to\infty} \left[L''(\frac{t}{\sqrt{n}})\frac{t^2}{2}\right] = \frac{t^2}{2}. &
\end{array}$$

  Thus, the central limit theorem is proven when $\mu = 0$ and $\sigma^2 = 1$.

  The general case follows by considering $X_i^* = \frac{X_i - \mu}{\sigma}$ and applying the preceding result, since $E[X_i^*] = 0,\ \mathrm{Var}(X_i^*) = 1$.

Example

- An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star.

  Because of changing atmospheric conditions and normal error, a measurement does not yield the exact distance, but only an estimate.

  The astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance.

  Suppose the astronomer believes that the values of the measurements are independent and identically distributed random variables having:
  - a common mean d (the actual distance);
  - a common variance of 4 (light-years).

  How many measurements need he make to be reasonably sure that his estimated distance is accurate to within ± 0.5 light-year?

Example (Cont'd)

- Suppose that the astronomer decides to make n observations.
  Let $X_1, X_2, \ldots, X_n$ be the n measurements.
  From the central limit theorem, $Z_n = \frac{\sum_{i=1}^n X_i - nd}{2\sqrt{n}}$ has approximately a standard normal distribution:

$$\begin{array}{rcl} P\left\{-0.5 \leq \frac{\sum_{i=1}^n X_i}{n} - d \leq 0.5\right\} & = & P\left\{-0.5\frac{\sqrt{n}}{2} \leq Z_n \leq 0.5\frac{\sqrt{n}}{2}\right\} \\ & \approx & \Phi(\frac{\sqrt{n}}{4}) - \Phi(-\frac{\sqrt{n}}{4}) \\ & = & 2\Phi(\frac{\sqrt{n}}{4}) - 1. \end{array}$$

  Suppose this probability is to be 95 percent.
  Then the number n^* of measurements must be such that $2\Phi(\frac{\sqrt{n^*}}{4}) - 1 = 0.95$ or $\Phi(\frac{\sqrt{n^*}}{4}) = 0.975$.
  Consulting a table, we get $\frac{\sqrt{n^*}}{4} = 1.96$ or $n^* = (7.84)^2 \approx 61.47$.
  As n^* is not integral valued, he should make 62 observations.

Example (Cont'd)

- The preceding analysis has been done under the assumption that the normal approximation will be a good approximation when $n = 62$.

  In general the question of how large n need be before the approximation is "good" depends on the distribution of the X_i.

  If this is a point of concern, to avoid taking any chances, apply Chebyshev's inequality.

  We have $E[\sum_{i=1}^n \frac{X_i}{n}] = d$ and $\mathrm{Var}(\sum_{i=1}^n \frac{X_i}{n}) = \frac{4}{n}$.

  Thus, Chebyshev's inequality yields

$$P\left\{\left|\sum_{i=1}^n \frac{X_i}{n} - d\right| > 0.5\right\} \leq \frac{4}{n(0.5)^2} = \frac{16}{n}.$$

  Hence, if $n = \frac{16}{0.05} = 320$ observations are made, there is 95 percent assurance that the estimate will be accurate to within 0.5 light-year.

Example

- The number of students who enroll in a psychology course is a Poisson random variable with mean 100.

  The Dean has decided that:
  - If the number enrolling is 120 or more, two separate sections will run;
  - If fewer than 120 students enroll, a single section will run.

  What is the probability that two sections are created?

  The exact solution is
  $$e^{-100}\sum_{i=120}^{\infty}\frac{(100)^i}{i!}.$$

  However, it does not readily yield a numerical answer.

Example (Cont'd)

- Recall that a Poisson random variable with mean 100 is the sum of 100 independent Poisson random variables, each with mean 1.

  So we can make use of the central limit theorem to obtain an approximate solution.

  Let X denote the number of students that enroll in the course:

$$\begin{aligned} P\{X \geq 120\} &= P\{X \geq 119.5\} \text{ (the continuity correction)} \\ &= P\left\{\frac{X-100}{\sqrt{100}} \geq \frac{119.5-100}{\sqrt{100}}\right\} \\ &\approx 1-\Phi(1.95) \approx 0.0256. \end{aligned}$$

  Here we have used the fact that the variance of a Poisson random variable is equal to its mean.

Example

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

  Let X_i denote the value of the ith die, $i = 1, 2, \ldots, 10$.

  We have

$$E(X_i) = \frac{7}{2}, \quad \text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{35}{12}.$$

  Thus, by the Central Limit Theorem,

$$\begin{aligned} P\{29.5 \leq X \leq 40.5\} &= P\left\{\frac{29.5-35}{\sqrt{\frac{350}{12}}} \leq \frac{X-35}{\sqrt{\frac{350}{12}}} \leq \frac{40.5-35}{\sqrt{\frac{350}{12}}}\right\} \\ &\approx 2\Phi(1.0184) - 1 \\ &\approx 0.692. \end{aligned}$$

Example

- Let X_i, $i = 1, \ldots, 10$, be independent random variables, each uniformly distributed over $(0,1)$.

  Calculate an approximation to $P\{\sum_{i=1}^{10} X_i > 6\}$.

  We have

$$E[X_i] = \frac{1}{2} \quad \text{and} \quad \mathrm{Var}(X_i) = \frac{1}{12}.$$

  Thus, by the Central Limit Theorem,

$$\begin{aligned} P\{\textstyle\sum_1^{10} X_i > 6\} &= P\left\{\frac{\sum_1^{10} X_i - 5}{\sqrt{10(\frac{1}{12})}} > \frac{6-5}{\sqrt{10(\frac{1}{12})}}\right\} \\ &\approx 1 - \Phi(\sqrt{1.2}) \\ &\approx 0.1367. \end{aligned}$$

  Hence, $\sum_{i=1}^{10} X_i$ will be greater than 6 only 14 percent of the time.

Example

- An instructor has 50 exams that will be graded in sequence.

  The times required to grade the 50 exams are independent, with a common distribution that has mean 20 minutes and standard deviation 4 minutes.

  Approximate the probability that the instructor will grade at least 25 of the exams in the first 450 minutes of work.

  Let X_i be the time that it takes to grade exam i.

  Then the time it takes to grade the first 25 exams is $X = \sum_{i=1}^{25} X_i$.

  The instructor will grade at least 25 exams in the first 450 minutes if the time it takes to grade the first 25 exams is ≤ 450.

  Thus, the desired probability is $P\{X \leq 450\}$.

Example (Cont'd)

- To approximate $P\{X \leq 450\}$, we apply the Central Limit Theorem. We have

$$\begin{aligned} E[X] &= \textstyle\sum_{i=1}^{25} E[X_i] = 25 \cdot 20 = 500; \\ \mathrm{Var}(X) &= \textstyle\sum_{i=1}^{25} \mathrm{Var}(X_i) = 25 \cdot 16 = 400. \end{aligned}$$

  Consequently, with Z being a standard normal random variable, we have

$$\begin{aligned} P\{X \leq 450\} &= P\left\{\frac{X-500}{\sqrt{400}} \leq \frac{450-500}{\sqrt{400}}\right\} \\ &\approx P\{Z \leq -2.5\} \\ &= P\{Z \geq 2.5\} \\ &= 1 - \Phi(2.5) = 0.006. \end{aligned}$$

Another Version of the Central Limit Theorem

- Central limit theorems also exist when the X_i are independent, but not necessarily identically distributed random variables.
- One version, by no means the most general, is as follows.

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 $\mu_i = E[X_i]$, $\sigma_i^2 = \mathrm{Var}(X_i)$, and such that:

(a) The X_i are uniformly bounded - that is, if for some M, $P\{|X_i| < M\} = 1$ for all i;

(b) $\sum_{i=1}^{\infty} \sigma_i^2 = \infty$.

Then

$$P\left\{\frac{\sum_{i=1}^{n}(X_i - \mu_i)}{\sqrt{\sum_{i=1}^{n}\sigma_i^2}} \leq a\right\} \xrightarrow{n\to\infty} \Phi(a).$$

Subsection 3

The Strong Law Of Large Numbers

The Strong Law of Large Numbers

Theorem (The Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{n\to\infty} \mu.$$

- That is, the strong law of large numbers states that

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

Example

- 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.
- Let
$$X_i = \begin{cases} 1, & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0, & \text{if } E \text{ does not occur on the } i\text{th trial} \end{cases}$$
- By the strong law of large numbers, with probability 1,
$$\frac{X_1 + \cdots + X_n}{n} \to E[X] = P(E).$$
- $X_1 + \cdots + X_n$ is the number of times that the event E occurs in the first n trials.

  Thus, we may interpret this equation as stating that, with probability 1, the limiting proportion of time that the event E occurs is $P(E)$.

Proof of the Strong Law Of Large Numbers

- We assume that the random variables X_i have a finite fourth moment. However, the theorem can be proven without this assumption.
- Suppose that $E[X_i^4] = K < \infty$.

  Assume that μ, the mean of the X_i, is equal to 0.

  Let $S_n = \sum_{i=1}^n X_i$ and consider

$$\begin{aligned} E[S_n^4] &= E[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n) \\ &\quad \times (X_1 + \cdots + X_n)(X_1 + \cdots + X_n)]. \end{aligned}$$

  Expanding the right side of the preceding equation results in terms of the form

$$X_i^4, \quad X_i^3 X_j, \quad X_i^2 X_j^2, \quad X_i^2 X_j X_k, \quad X_i X_j X_k X_\ell,$$

  where i, j, k and ℓ are all different.

Proof of the Strong Law Of Large Numbers (Cont'd)

- All the X_i have mean 0.

  Thus, by independence,

$$\begin{aligned} E[X_i^3 X_j] &= E[X_i^3]E[X_j] = 0; \\ E[X_i^2 X_j X_k] &= E[X_i^2]E[X_j]E[X_k] = 0; \\ E[X_i X_j X_k X_\ell] &= 0. \end{aligned}$$

  Now, for a given pair i and j, there will be $\binom{4}{2} = 6$ terms in the expansion that will equal $X_i^2 X_j^2$.

  We expanding the preceding product and take expectations (using, once more, independence):

$$\begin{aligned} E[S_n^4] &= nE[X_i^4] + 6\binom{n}{2}E[X_i^2 X_j^2] \\ &= nK + 3n(n-1)E[X_i^2]E[X_j^2]. \end{aligned}$$

Proof of the Strong Law Of Large Numbers (Cont'd)

- But $0 \leq \mathrm{Var}(X_i^2) = E[X_i^4] - (E[X_i^2])^2$.
  Thus,
  $$(E[X_i^2])^2 \leq E[X_i^4] = K.$$
  Therefore, from $E[S_n^4] = nK + 3n(n-1)E[X_i^2]E[X_j^2]$, we obtain
  $$E[S_n^4] \leq nK + 3n(n-1)K;$$
  $$E\left[\frac{S_n^4}{n^4}\right] \leq \frac{K}{n^3} + \frac{3K}{n^2};$$
  $$E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] = \sum_{n=1}^{\infty} E\left[\frac{S_n^4}{n^4}\right] < \infty.$$
  Hence, with probability 1, $\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty$.
  By the convergence criterion, the nth term goes to 0.
  So, with probability 1, $\lim_{n\to\infty} \frac{S_n^4}{n^4} = 0$.
  But if $\frac{S_n^4}{n^4} = (\frac{S_n}{n})^4$ goes to 0, then so must $\frac{S_n}{n}$.
  Hence, with probability 1, $\frac{S_n}{n} \to 0$ as $n \to \infty$.

Proof of the Strong Law Of Large Numbers (Cont'd)

- In general, suppose μ, the mean of the X_i, is not equal to 0.

  We can apply the preceding argument to the random variables $X_i - \mu$.

  We obtain that with probability 1,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{(X_i - \mu)}{n} = 0.$$

  Equivalently, with probability 1,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i}{n} = \mu.$$

Weak versus Strong Law of Large Numbers

- The Weak Law of Large Numbers states that, for any specified large value n^*, $\frac{X_1+\cdots+X_{n^*}}{n^*}$ is likely to be near μ.

  However, it does not say that $\frac{X_1+\cdots+X_n}{n}$ is bound to stay near μ for all values of n larger than n^*.

  Thus, it leaves open the possibility that large values of $\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|$ can occur infinitely often (though at infrequent intervals).
- The Strong Law shows that this cannot occur.

  It implies that, with probability 1, for any positive value ε,
  $$\left|\sum_{i=1}^{n}\frac{X_i}{n}-\mu\right|$$
  will be greater than ε only a finite number of times.

Subsection 4

Other Inequalities

One-Sided Chebyshev Inequality

Proposition (One-Sided Chebyshev Inequality)

If X is a random variable with mean 0 and finite variance σ^2, then, for any $a > 0$,

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

- Let $b > 0$. Note that $X \geq a$ is equivalent to $X + b \geq a + b$.

  Hence,

  $$\begin{array}{rcl} P\{X \geq a\} & = & P\{X + b \geq a + b\} \\ & \leq & P\{(X+b)^2 \geq (a+b)^2\}. \end{array}$$

  Here the inequality is obtained by noting that since $a + b > 0$, $X + b \geq a + b$ implies that $(X+b)^2 \geq (a+b)^2$.

One-Sided Chebyshev Inequality (Cont'd)

- We obtained $P\{X \geq a\} \leq P\{(X+b)^2 \geq (a+b)^2\}$.

  Upon applying Markov's inequality, the preceding yields that

  $$P\{X \geq a\} \leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{\sigma^2+b^2}{(a+b)^2}.$$

  The function $f(b) = \frac{\sigma^2+b^2}{(a+b)^2}$ has derivative

  $$f'(b) = \frac{2b(a+b)^2 - 2(a+b)(\sigma^2+b^2)}{(a+b)^4} = \frac{2ab-2\sigma^2}{(a+b)^3}.$$

  So it is minimized by $b = \frac{\sigma^2}{a}$.

  Thus,

  $$P\{X \geq a\} \leq \frac{\sigma^2 + (\frac{\sigma^2}{a})^2}{(a+\frac{\sigma^2}{a})^2} = \frac{\frac{a^2\sigma^2+\sigma^4}{a^2}}{\frac{(a^2+\sigma^2)^2}{a^2}} = \frac{\sigma^2}{\sigma^2+a^2}.$$

Example

- Suppose 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.

  The one-sided Chebyshev inequality yields

  $$P\{X \geq 120\} = P\{X - 100 \geq 20\} \leq \frac{400}{400 + (20)^2} = \frac{1}{2}.$$

  Hence, the probability that this week's production will be 120 or more is at most $\frac{1}{2}$.
- The bound obtained by applying Markov's inequality is

  $$P\{X \geq 120\} \leq \frac{E(X)}{120} = \frac{5}{6}.$$

  Note that this is a far weaker bound than the preceding one.

A Corollary

- Suppose that X has mean μ and variance σ^2.
- Then both $X-\mu$ and $\mu-X$ have mean 0 and variance σ^2.
- Thus, by the one-sided Chebyshev inequality, for $a>0$,

$$\begin{aligned} P\{X-\mu \geq a\} &\leq \frac{\sigma^2}{\sigma^2+a^2}; \\ P\{\mu-X \geq a\} &\leq \frac{\sigma^2}{\sigma^2+a^2}. \end{aligned}$$

Corollary

If $E[X]=\mu$ and $\text{Var}(X)=\sigma^2$, then, for $a>0$,

$$\begin{aligned} P\{X \geq \mu+a\} &\leq \frac{\sigma^2}{\sigma^2+a^2}, \\ P\{X \leq \mu-a\} &\leq \frac{\sigma^2}{\sigma^2+a^2}. \end{aligned}$$

Example

- 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.

  Number the men arbitrarily from 1 to 100.

  For $i = 1, 2, \ldots, 100$, let

  $$X_i = \begin{cases} 1, & \text{if man } i \text{ is paired with a woman} \\ 0, & \text{otherwise} \end{cases}$$

  Then the number X of man-woman pairs is $X = \sum_{i=1}^{100} X_i$.

  Because man i is equally likely to be paired with any of the other 199 people, of which 100 are women, we have

  $$E[X_i] = P\{X_i = 1\} = \frac{100}{199}.$$

Example (Cont'd)

- Similarly, for $i \neq j$,

$$\begin{aligned} E[X_i X_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\} P\{X_j = 1 | X_i = 1\} = \tfrac{100}{199} \tfrac{99}{197}. \end{aligned}$$

Here $P\{X_j = 1 | X_i = 1\} = \frac{99}{197}$, since, given that man i is paired with a woman, man j is equally likely to be paired with any of the remaining 197 people, of which 99 are women.

Hence, we obtain

$$\begin{aligned} E[X] &= \textstyle\sum_{i=1}^{100} E[X_i] = 100 \cdot \frac{100}{199} \approx 50.25; \\ \mathrm{Var}(X) &= \textstyle\sum_{i=1}^{100} \mathrm{Var}(X_i) + 2 \sum\limits_{i<j}\sum \mathrm{Cov}(X_i, X_j) \\ &= 100 \tfrac{100}{199} \tfrac{99}{199} + 2 \binom{100}{2} [\tfrac{100}{199} \tfrac{99}{197} - (\tfrac{100}{199})^2] \\ &\approx 25.126. \end{aligned}$$

Example (Cont'd)

- The Chebyshev inequality then yields

$$P\{X \leq 30\} = P\{|X - 50.25| \geq 20.25\} \leq \frac{25.126}{(20.25)^2} \approx 0.061.$$

  Thus, there are fewer than 6 chances in a hundred that fewer than 30 men will be paired with women.

- We can improve on this bound by using the one-sided Chebyshev inequality:

$$\begin{aligned} P\{X \leq 30\} &= P\{X \leq 50.25 - 20.25\} \\ &\leq \frac{25.126}{25.126 + (20.25)^2} \\ &\approx 0.058. \end{aligned}$$

Chernoff Bounds

- When the moment generating function of the random variable X is known, we can obtain even more effective bounds on $P\{X \geq a\}$.
- Let $M(t) = E[e^{tX}]$ be the moment generating function of X.
  - For $t > 0$, $P\{X \geq a\} = P\{e^{tX} \geq e^{ta}\} \overset{\text{Markov}}{\leq} E[e^{tX}]e^{-ta}$;
  - For $t < 0$, $P\{X \leq a\} = P\{e^{tX} \geq e^{ta}\} \leq E[e^{tX}]e^{-ta}$.

Proposition (Chernoff Bounds)

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

- The Chernoff bounds hold for all t, either positive or negative.

  So the best bound on $P\{X \geq a\}$ is obtained by using the t that minimizes $e^{-ta}M(t)$.

Chernoff Bounds for Standard Normal Random Variable

- If Z is a standard normal random variable, then its moment generating function is $M(t) = e^{t^2/2}$.
- So the Chernoff bound on $P\{Z \geq a\}$ is given by

$$P\{Z \geq a\} \leq e^{-ta} e^{t^2/2}, \text{ for all } t > 0.$$

- The value of t, $t > 0$, that minimizes $e^{t^2/2 - ta}$ is the value that minimizes $\frac{t^2}{2} - ta$, which is $t = a$.
- Thus, for $a > 0$, we have

$$P\{Z \geq a\} \leq e^{-a^2/2}.$$

- Similarly, we can show that, for $a < 0$,

$$P\{Z \leq a\} \leq e^{-a^2/2}.$$

Chernoff Bounds for the Poisson Random Variable

- If X is a Poisson random variable with parameter λ, then its moment generating function is $M(t) = e^{\lambda(e^t - 1)}$.
- Hence, the 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 preceding inequality is equivalent to minimizing $\lambda(e^t - 1) - it$.

  Calculus shows that the minimal value occurs when $e^t = \frac{i}{\lambda}$.

  Provided that $\frac{i}{\lambda} > 1$, this minimizing value of t will be positive.
- Therefore, assuming that $i > \lambda$ and letting $e^t = \frac{i}{\lambda}$ in the Chernoff bound yields

$$P\{X \geq i\} \leq e^{\lambda(i/\lambda - 1)} \left(\frac{\lambda}{i}\right)^i = \frac{e^{-\lambda}(e\lambda)^i}{i^i}.$$

Example

- Consider a gambler who is equally likely to either win or lose 1 unit on every play, independently of his past results.

  Let X_i be the gambler's winnings on the ith play.

  The X_i are independent and

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

  Let $S_n = \sum_{i=1}^{n} X_i$ denote the gambler's winnings after n plays.

  We will use the Chernoff bound on $P\{S_n \geq a\}$.

  Note, first, that the moment generating function of X_i is

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

Example (Cont'd)

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

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

Therefore, $E[e^{tX}] \leq e^{t^2/2}$.

Example (Cont'd)

- But the moment generating function of the sum of independent random variables is the product of their moment generating functions. Hence, we have
$$E[e^{tS_n}] = (E[e^{tX}])^n \leq e^{nt^2/2}.$$
- Using the preceding result along with the Chernoff bound gives
$$P\{S_n \geq a\} \leq M(t)e^{-ta} \leq e^{-ta}e^{nt^2/2}, \quad t > 0.$$
- The value of t that minimizes the right side of the preceding is the value that minimizes $\frac{nt^2}{2} - ta$, which is $t = \frac{a}{n}$.
- Supposing that $a > 0$ (so that the minimizing t is positive) and letting $t = \frac{a}{n}$ in the preceding inequality yields
$$P\{S_n \geq a\} \leq e^{-a^2/2n}, \quad a > 0.$$

Convex and Concave Functions

Definition

A twice-differentiable real-valued function $f(x)$ is said to be:

- **convex** if $f''(x) \geq 0$, for all x;
- **concave** if $f''(x) \leq 0$, for all x.

- Some examples of convex functions:
  - $f(x) = x^2$;
  - $f(x) = e^{ax}$;
  - $f(x) = -x^{1/n}$, for $x \geq 0$.
- A useful observation is that

$$f(x) \text{ is convex if and only if } g(x) = -f(x) \text{ is concave.}$$

Jensen's Inequality

Proposition (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.

- Expand $f(x)$ in a Taylor series expansion about $\mu = E[X]$:

  $$f(x) = f(\mu) + f'(\mu)(x-\mu) + \frac{f''(\xi)(x-\mu)^2}{2},$$

  where ξ is some value between x and μ.
  Since $f''(\xi) \geq 0$, we obtain $f(x) \geq f(\mu) + f'(\mu)(x-\mu)$.
  Hence, $f(X) \geq f(\mu) + f'(\mu)(X-\mu)$.
  Taking expectations yields

  $$E[f(X)] \geq f(\mu) + f'(\mu)E[X-\mu] = f(\mu).$$

Example

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

  Denote by:
  - R her return;
  - u her utility function.

  Suppose that her decision will be made on the basis of maximizing the expected value of $u(R)$.

  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.
  - If u is convex, then $E[u(X)] \geq u(m)$.
    So the risky investment alternative would be preferred.

Subsection 5

Approximating a Sum of Bernoulli by a Poisson Variable

Strategy

- Suppose that we want to approximate the sum of independent Bernoulli random variables with respective means $p_1, p_2, \ldots, p_n$.
- Start with a sequence $Y_1, \ldots, Y_n$ of independent Poisson random variables, with Y_i having mean p_i.
- We construct a sequence of independent Bernoulli random variables $X_1, \ldots, X_n$ with parameters $p_1, \ldots, p_n$, such that

$$P\{X_i \neq Y_i\} \leq p_i^2, \text{ for each } i.$$

- Letting $X = \sum_{i=1}^n X_i$ and $Y = \sum_{i=1}^n Y_i$, we use the preceding inequality to conclude that

$$P\{X \neq Y\} \leq \sum_{i=1}^n p_i^2.$$

Strategy (Cont'd)

- Finally, we show that the preceding inequality implies that, for any set of real numbers A,

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

- But, by hypothesis,
  - X is the sum of independent Bernoulli random variables;
  - Y is a Poisson random variable.

  So, the latter inequality yields the desired bound.

Random Variables Y_i and U_i

- Let Y_i, $i = 1, \ldots, n$, be independent Poisson random variables with respective means p_i.
- Recall the inequality
$$e^{-p} \geq 1 - p.$$
It implies that $(1 - p_i)e^{p_i} \leq 1$.
- Let $U_1, \ldots, U_n$ be random variables, that are:
  - independent;
  - independent of the Y_i's;
  - such that
$$U_i = \begin{cases} 0, & \text{with probability } (1 - p_i)e^{p_i} \\ 1, & \text{with probability } 1 - (1 - p_i)e^{p_i} \end{cases}$$

Random Variables X_i

- Define the random variables X_i, $i = 1, \ldots, n$, by

$$X_i = \begin{cases} 0, & \text{if } Y_i = U_i = 0 \\ 1, & \text{otherwise} \end{cases}$$

- Note that:

$$\begin{aligned} P\{X_i = 0\} &= P\{Y_i = 0\}P\{U_i = 0\} \\ &= e^{-p_i}(1 - p_i)e^{p_i} \\ &= 1 - p_i; \\ P\{X_i = 1\} &= 1 - P\{X_i = 0\} \\ &= p_i. \end{aligned}$$

Probability of $X_i \neq Y_i$

- If X_i is equal to 0, then so must Y_i equal 0 (by the definition of X_i).
- Therefore,

$$\begin{aligned}
P\{X_i \neq Y_i\} &= P\{X_i = 1, Y_i \neq 1\} \\
&= P\{Y_i = 0, X_i = 1\} + P\{Y_i > 1\} \\
&= P\{Y_i = 0, U_i = 1\} + P\{Y_i > 1\} \\
&= e^{-p_i}[1 - (1 - p_i)e^{p_i}] + 1 - e^{-p_i} - p_i e^{-p_i} \\
&= p_i - p_i e^{-p_i} \\
&\leq p_i^2 \quad (\text{since } 1 - e^{-p} \leq p)
\end{aligned}$$

Probability of $X \neq Y$

- Let $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{n} Y_i$.
- Note that:
  - X is the sum of independent Bernoulli random variables;
  - Y is Poisson with the expected value $E[Y] = E[X] = \sum_{i=1}^{n} p_i$.
- Note also that the inequality $X \neq Y$ implies that $X_i \neq Y_i$ for some i.
- So

$$\begin{array}{rcll} P\{X \neq Y\} & \leq & P\{X_i \neq Y_i \text{ for some } i\} & \\ & \leq & \sum_{i=1}^{n} P\{X_i \neq Y_i\} & \text{(Boole's Inequality)} \\ & \leq & \sum_{i=1}^{n} p_i^2. & \end{array}$$

Final Step

- For any event B, let I_B, the indicator variable for the event B, be defined by

$$I_B = \begin{cases} 1, & \text{if } B \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

- Let A be any set of real numbers.
  - $I_{\{X \in A\}} - I_{\{Y \in A\}} = 1$ implies $I_{\{X \in A\}} = 1$ and $I_{\{Y \in A\}} = 0$;
  - This implies that $X \in A$ and $Y \notin A$;
  - This implies that $X \neq Y$, i.e., $I_{\{X \neq Y\}} = 1$.

  Thus, we get

$$I_{\{X \in A\}} - I_{\{Y \in A\}} \leq I_{\{X \neq Y\}}.$$

- Upon taking expectations of the preceding inequality, we obtain

$$P\{X \in A\} - P\{Y \in A\} \leq P\{X \neq Y\}.$$

Final Step (Cont'd)

- We showed that

$$P\{X \in A\} - P\{Y \in A\} \leq P\{X \neq Y\}.$$

- By reversing X and Y, we obtain, in the same manner,

$$P\{Y \in A\} - P\{X \in A\} \leq P\{X \neq Y\}.$$

- Thus, we can conclude that

$$|P\{X \in A\} - P\{Y \in A\}| \leq P\{X \neq Y\}.$$

- Therefore, with $\lambda = \sum_{i=1}^{n} p_i$,

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

Case of Binomial Random Variables

- When all the p_i are equal to p, X is a binomial random variable.
- Hence, the preceding inequality shows that, for any set of nonnegative integers A,

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