Lecture 7: Chebyshev's Inequality, LLN, and the CLT

Sta 111

Colin Rundel

May 22, 2014

Markov's Inequality

For any random variable $X \geq 0$ and constant $a > 0$, then

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

Corollary - Chebyshev's Inequality:

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

"The inequality says that the probability that X is far away from its mean is bounded by a quantity that increases as $Var(X)$ increases."

Derivation of Markov's Inequality

Let X be a random variable such that $X \geq 0$ then

Derivation of Chebyshev's Inequality

Proposition - if $f(x)$ is a non-decreasing function then

$$P(X \geq a) = P\Big(f(X) \geq f(a)\Big).$$

Therefore,

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

If we define the positive valued random variable to be $|X - E(X)|$ and $f(x) = x^2$ then

$$P\Big(|X - E(X)| \geq a\Big) = P\Big((X - E(X))^2 \geq a^2\Big) \leq \frac{E\big((X - E(X))^2\big)}{a^2} = \frac{Var(X)}{a^2}$$

$$P\Big(|X - E(X)| \geq a\Big) \leq \frac{Var(X)}{a^2}$$

Using Markov's and Chebyshev's Inequalities

Suppose that it is known that the number of items produced in a factory during a week is a random variable X with mean 50. (Note that we don't know anything about the distribution of the pmf)

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

(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 units?

Chebyshev's Inequality - Example

Lets use Chebyshev's inequality to make a statement about the bounds for the probability of being with in 1, 2, or 3 standard deviations of the mean for all random variables.

If we define $a = k\sigma$ where $\sigma = \sqrt{Var(X)}$ then

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

Accuracy of Chebyshev's inequality, cont.

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then let $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Empirical Rule:	Chebyshev's inequality:
$1 - P(-1 \leq Z \leq 1) \approx 1 - 0.66 = 0.34$	$1 - P(-1 \leq U \leq 1) \leq \frac{1}{1^1} = 1$
$1 - P(-2 \leq Z \leq 2) \approx 1 - 0.96 = 0.04$	$1 - P(-2 \leq U \leq 2) \leq \frac{1}{2^2} = 0.25$
$1 - P(-3 \leq Z \leq 3) \approx 1 - 0.997 = 0.003$	$1 - P(-3 \leq U \leq 3) \leq \frac{1}{3^2} = 0.11$

Why do we care if it is so inaccurate?

Independent and Identically Distributed (iid)

A collection of random variables that share the same probability distribution and all are mutually independent.

Example

If $X \sim \text{Binom}(n,p)$ then $X = \sum_{i=1}^{n} Y_i$ where $Y_1, \cdots, Y_n \overset{iid}{\sim} \text{Bern}(p)$

Sums of iid Random Variables

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} D$ where D is some probability distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

We defined $S_n = X_1 + X_2 + \cdots + X_n$

Average of iid Random Variables

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} D$ where D is some probability distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

We defined $\overline{X}_n = (X_1 + X_2 + \cdots + X_n)/n = S_n/n$ then

Weak Law of Large Numbers

Based on these results and Markov's Inequality we can show the following:

Therefore, as long as $\sigma^2 < \infty$

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \ge \epsilon) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

Law of Large Numbers

Weak Law of Large Numbers ($\bar{X}_n$ converges in probability to μ):

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

Strong Law of Large Numbers ($\bar{X}_n$ converges almost surely to μ):

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

Strong LLN is a more powerful result (Strong LLN implies Weak LLN), but its proof is more complicated.

LLN - Example

How large a random sample must be taken from a given distribution in order for the probability to be at least 0.99 that the sample mean will be within 2 standard deviations of the mean of the distribution?

What about 0.95 probability to be within 1 standard deviations of the mean?

LLN and CLT

Law of large numbers shows us that

$$\lim_{n\to\infty} \frac{S_n - n\mu}{n} = \lim_{n\to\infty} (\bar{X}_n - \mu) \to 0$$

which shows that for large n, $n \gg \bar{S}_n - n\mu$.

What happens if we divide by something that grows slower than n like $\sqrt{n}$?

$$\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n}} = \lim_{n\to\infty} \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

This is the Central Limit Theorem, of which the DeMoivre-Laplace theorem for the normal approximation to the binomial is a special case. Hopefully by the end of this class we will have the tools to prove this.

Equivalence of de Moivre-Laplace with CLT

We have already seen that a Binomial random variable is equivalent to the sum of n iid Bernoulli random variables.

Let $X \sim$ Binom(n,p) where $X = \sum_{i=1}^n Y_i$ with $Y_1, \cdots, Y_n \overset{iid}{\sim}$ Bern(p) and $E(Y_i) = p$, $Var(Y_i) = p(1-p)$.

de Moivre-Laplace tells us:

$$P(a \le X \le b) = P\left(\frac{a-np}{\sqrt{np(1-p)}} \le \frac{X-np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}}\right)$$
$$\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

Equivalence of de Moivre-Laplace with CLT, cont.

The Central Limit theorem gives us,

$$P\left(c \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le d\right) \approx \Phi(d) - \Phi(c)$$

then for $X = S_n = \sum Y_i$

$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-n\mu}{\sqrt{n}\sigma} \le \frac{X-n\mu}{\sqrt{n}\sigma} \le \frac{b-n\mu}{\sqrt{n}\sigma}\right) \\ &= P\left(\frac{a-np}{\sqrt{np(1-p)}} \le \frac{S_n-np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Why do we care? (Statistics)

In general we are interested in making statements about how the world works, we usually do this based on empirical evidence. This tends to involve observing or measuring something a bunch of times.

- LLN tells us that if we average our results we'll get closer and closer to the true value as we take more measurements
- CLT tells us the distribution of our average is normal if we take enough measurements, which tells us our uncertainty about the 'true' value

Example - Polling: I can't interview everyone about their opinions but if I interview enough my estimate should be close and I can use the CLT to approximate my margin of error.

Astronomy Example

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 measurement 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 value of the actual distance.

If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light years^2), how many measurements should he make to be reasonably sure that his estimated distance is accurate to within ± 0.5 light years?

Astronomy Example, cont.

We need to decide on a level to account for being "reasonably sure", usually taken to be 95% but the choice somewhat arbitrary.

Astronomy Example, cont.

Our previous estimate depends on how long it takes for the distribution of the average of the measurements to converge to the normal distribution. CLT only guarantees convergence as $n \to \infty$, but most distributions converge much more quickly (think about the $np \geq 10, np \geq 10$ requirement for Normal approximation to Binomial).

We can also solve this problem using Chebyshev's inequality

Moment Generating Function - Properties

If X and Y are independent random variables then the moment generating function for the distribution of $X+Y$ is

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$

Similarly, the moment generating function for S_n, the sum of iid random variables $X_1, X_2, \ldots, X_n$ is

$$M_{S_n}(t) = [M_{X_i}(t)]^n$$

Moment Generating Function - Unit Normal

Let $Z \sim \mathcal{N}(0,1)$ then

$$\begin{aligned}
M_Z(t) = E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-tx}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}+\frac{t^2}{2}} dx \\
&= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx \\
&= e^{t^2/2}
\end{aligned}$$

Central Limit Theorem

Let $X_1, \ldots, X_n$ 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 unit normal as $n \to \infty$.

That is, for $-\infty < a < \infty$,

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

Sketch of Proof

Proposition

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables and $S_n = X_1 + \cdots + X_n$. The distribution of S_n is given by the distribution function f_{S_n} which has a moment generating function M_{S_n} with $n \ge 1$.

Let Z being a random variable with distribution function f_Z and moment generating function M_Z.

If $M_{S_n}(t) \to M_Z(t)$ for all t, then $f_{S_n}(t) \to f_Z(t)$ for all t at which $f_Z(t)$ is continuous.

We can prove the CLT by letting $Z \sim \mathcal{N}(0,1)$, $M_Z(t) = e^{t^2/2}$ and then showing for any S_n that $M_{S_n/\sqrt{n}} \to e^{t^2/2}$ as $n \to \infty$.

Proof of the CLT

Some simplifying assumptions and notation:

- $E(X_i) = 0$
- $Var(X_i) = 1$
- $M_{X_i}(t)$ exists and is finite
- $L_X(t) = \log M_X(t)$

Also, remember L'Hospital's Rule:

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)}$$

Proof of the CLT, cont.

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

$$M_{X_i/\sqrt{n}}(t) = E\left[\exp\left(\frac{tX_i}{\sqrt{n}}\right)\right] = M_{X_i}\left(\frac{t}{\sqrt{n}}\right)$$

and this the moment generating function of $S_n/\sqrt{n} = \sum_{i=1}^n X_i/\sqrt{n}$ is given by

$$M_{S_n/\sqrt{n}}(t) = \left[M_{X_i}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Therefore in order to show $M_{S_n/\sqrt{n}} \to M_Z(t)$ we need to show

$$\left[M_{X_i}\left(\frac{t}{\sqrt{n}}\right)\right]^n \to e^{t^2/2}$$

Proof of the CLT, cont.

$$L_{X_i}(t) = \log M_{X_i}(t)$$
$$L_{X_i}(0) = \log M_{X_i}(0) = \log 1 = 0$$

$$L'_{X_i}(t) = \frac{d}{dt}\log M_{X_i}(t) = \frac{M'_{X_i}(t)}{M_{X_i}(t)}$$
$$L'_{X_i}(0) = \frac{M'_{X_i}(0)}{M_{X_i}(0)} = \frac{\mu}{1} = 0$$

$$L''_{X_i}(t) = \frac{d}{dt}\frac{M'_{X_i}(t)}{M_{X_i}(t)} = \frac{M_{X_i}(t)M''_{X_i}(t) - [M'_{X_i}(t)]^2}{[M_{X_i}(t)]^2}$$
$$L''_{X_i}(0) = \frac{M_{X_i}(0)M''_{X_i}(0) - [M'_{X_i}(0)]^2}{[M_{X_i}(0)]^2}$$
$$= \frac{E(X_i^0)E(X_i^2) - E(X_i)^2}{E(X_i^0)^2} = E(X_i^2) - E(X_i)^2 = \sigma^2 = 1$$

Proof of the CLT, cont.

$$[M_{X_i}(t/\sqrt{n})]^n \to e^{t^2/2}$$
$$nL_{X_i}(t/\sqrt{n}) \to t^2/2$$

$$\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{L'(t/\sqrt{n})(-\frac{1}{2}tn^{-3/2})}{-n^{-2}} && \text{by L'Hospital's rule} \\
&= \lim_{n\to\infty} \frac{L'(t/\sqrt{n})t}{2n^{-1/2}} \\
&= \lim_{n\to\infty} \frac{L''(t/\sqrt{n})t(-\frac{1}{2}tn^{-3/2})}{-n^{-3/2}} && \text{by L'Hospital's rule} \\
&= \lim_{n\to\infty} L''(t/\sqrt{n})\frac{t^2}{2} \\
&= \frac{t^2}{2}
\end{aligned}$$

Proof of the CLT, Final Comments

The preceding proof assumes that $E(X_i) = 0$ and $Var(X_i) = 1$.

We can generalize this result to any collection of random variables Y_i by considering the standardized form $Y_i^* = (Y_i - \mu)/\sigma$.

$$\begin{aligned} \frac{Y_1 + \cdots + Y_n - n\mu}{\sigma\sqrt{n}} &= \left(\frac{Y_1 - \mu}{\sigma} + \cdots + \frac{Y_n - \mu}{\sigma}\right) \Big/ \sqrt{n} \\ &= \left(Y_1^* + \cdots + Y_n^*\right) / \sqrt{n} \end{aligned}$$

$$\begin{aligned} E(Y_i^*) &= 0 \\ Var(Y_i^*) &= 1 \end{aligned}$$