Lecture 8: Using the LLN and CLT, Moments of Distributions

Sta230/Mth230

Colin Rundel

February 12, 2014

Markov's and Chebyshev's Inequalities

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

Markov's Inequality:

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

Chebyshev's Inequality:

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

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?

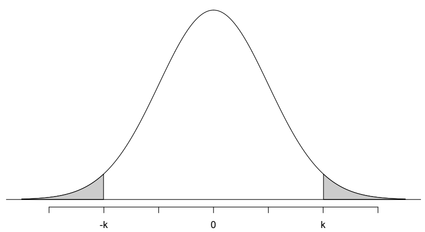
Accuracy of Chebyshev's inequality

How tight the inequality is depends on the distribution, but we can look at the Normal distribution since it is easy to evaluate. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ then

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

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

$$P\left(\frac{X - \mu}{\sigma} \leq -k \text{ or } \frac{X - \mu}{\sigma} \geq k\right) \leq \frac{1}{k^2}$$



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:
$1 - P(-1 \le Z \le 1) \approx 1 - 0.66 = 0.34$	$1 - P(-1 \le U \le 1) \le \frac{1}{1^1} = 1$
$1 - P(-2 \le Z \le 2) \approx 1 - 0.96 = 0.04$	$1 - P(-2 \le U \le 2) \le \frac{1}{2^2} = 0.25$
$1 - P(-3 \le Z \le 3) \approx 1 - 0.997 = 0.003$	$1 - P(-3 \le U \le 3) \le \frac{1}{3^2} = 0.11$

Why do we care if it is so inaccurate?

Independent and Identically Distributed (iid)

A collection of independent random variables that share the same probability distribution.

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 the sum of the random variables to be $S_n = X_1 + X_2 + \cdots + X_n$

$$\begin{aligned} E(S_n) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= \mu + \mu + \cdots + \mu = n\mu \end{aligned}$$

$$\begin{aligned} Var(S_n) &= \sum_{i=1}^n Var(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \ne j}}^n Cov(X_i, X_j) \\ &= \sum_{i=1}^n Var(X_i) = \sum_{i=1}^n \sigma^2 \\ &= n\sigma^2 \end{aligned}$$

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 define the average of the random variables to be

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

then,

$$E(\overline{X}_n) = E(S_n/n) = E(S_n)/n = \mu$$

$$\begin{aligned} Var(\overline{X}_n) &= Var(S_n/n) = \frac{1}{n^2} Var(S_n) \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

Weak Law of Large Numbers

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

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

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

Law of Large Numbers

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

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

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

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

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

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

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 $\sqrt{n}$?

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

This is the Central Limit Theorem, of which the DeMoivre-Laplace theorem for the normal approximation to the binomial is a special case. For large n,

$$\overline{X}_n \sim N\left(\mu,\ \sigma^2/n\right)$$

LLN and CLT

Law of large numbers:

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

Central Limit Theorem:

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

$$P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = P\left(a \le \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le b\right) \approx \Phi(b) - \Phi(a)$$

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:

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

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

Moments

Some definitions,

Raw moment:

$$\mu'_n = E(X^n)$$

Central moment:

$$\mu_n = E[(X-\mu)^n]$$

Normalized / Standardized moment:

$$\frac{\mu_n}{\sigma^n}$$

Properties of Moments

Zeroth Moment:

$$\mu_0' = \mu_0 = 1$$

First Moment:

$$\mu_1' = E(X) = \mu$$
$$\mu_1 = E(X-\mu) = 0$$

Second Moment:

$$\mu_2 = E[(X-\mu)^2] = Var(X)$$
$$\mu_2' - (\mu_1')^2 = Var(X)$$

Third Moment:

$$\text{Skewness}(X) = \frac{\mu_3}{\sigma^3}$$

Fourth Moment:

$$\text{Kurtosis}(X) = \frac{\mu_4}{\sigma^4}$$
$$\text{Ex. Kurtosis}(X) = \frac{\mu_4}{\sigma^4} - 3$$

Note that some moments do not exist, which is the case when $E(X^n)$ does not converge.

Third and Forth Moments

$$\begin{aligned}\mu_3 = E[(X-\mu)^3] &= E(X^3 - 3X^2\mu + 3X\mu^2 - \mu^3)\\ &= E(X^3) - 3\mu E(X^2) + 3\mu^3 - \mu^3\\ &= E(X^3) - 3\mu\sigma^2 - \mu^3\\ &= \mu_3' - 3\mu\sigma^2 - \mu^3\end{aligned}$$

$$\begin{aligned}\mu_4 = E[(X-\mu)^4] &= E(X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4)\\ &= E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 4\mu^4 + \mu^4\\ &= E(X^4) - 4\mu E(X^3) + 2\mu^2(\sigma^2+\mu^2) + 4\mu^2\sigma^2 + \mu^4\\ &= \mu_4' - 4\mu\mu_3' + 6\mu^2\sigma^2 + 3\mu^4\end{aligned}$$

Moment Generating Function

The moment generating function of a discrete random variable X is defined for all real values of t by

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} P(X=x)$$

This is called the moment generating function because we can obtain the moments of X by successively differentiating $M_X(t)$ and evaluating at $t=0$.

$$\begin{aligned}&M_X(0) = E[e^0] = 1 = \mu_0'\\ &M_X'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}e^{tX}\right] = E[Xe^{tX}]\\ &M_X'(0) = E[Xe^0] = E[X] = \mu_1'\\ &M_X''(t) = \frac{d}{dt}M_X'(t) = \frac{d}{dt}E[Xe^{tX}] = E\left[\frac{d}{dt}(Xe^{tX})\right] = E[X^2e^{tX}]\\ &M_X''(0) = E[X^2e^0] = E[X^2] = \mu_2'\end{aligned}$$

Moment Generating Function - Poisson

Let $X \sim \text{Pois}(\lambda)$ then

Moment Generating Function - Poisson Skewness

Moment Generating Function - Poisson Kurtosis

Moment Generating Function - Normal

Moment Generating Function - Normal, cont.

$$M'_X(t) = \exp(\mu t + \frac{t^2\sigma^2}{2})(\mu + t\sigma^2)$$
$$M'_X(0) = \mu'_1 = \mu$$

$$M''_X(t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\sigma^2 + \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)(\mu + t\sigma^2)^2$$
$$M''_X(0) = \mu'_2 = \mu^2 + \sigma^2$$

Moment Generating Function - Normal, cont.

$$\mu_3' = M_X'''(0) = \mu^3 + 3\mu\sigma^2$$
$$\mu_4' = M_X''''(0) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

$$\begin{aligned}\text{Skewness}(X) &= \frac{\mu_3}{\sigma^3} = \frac{\mu_3' - 3\mu\sigma^2 - \mu^3}{\sigma^3} \\ &= \frac{\mu^3 + 3\mu\sigma^2 - 3\mu\sigma^2 - \mu^3}{\sigma^3} = 0\end{aligned}$$

$$\begin{aligned}\text{Kurtosis}(X) &= \frac{\mu_4}{\sigma^4} = \frac{\mu_4' - 4\mu\mu_3' + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} \\ &= \frac{\mu_4' - 4\mu(\mu^3 + 3\mu\sigma^2) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} \\ &= \frac{(\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4) - 6\mu^2\sigma^2 - 1\mu^4}{\sigma^4} = \frac{3\sigma^4}{\sigma^4} = 3 \\ \text{Ex. Kurtosis}(X) &= 0\end{aligned}$$