Limits of Sums of Random Variables

Sections 7.2 – 7.3,
(but not Subsections 7.3.2–3)
Limits of Sums of RV’s

- Last time, we defined the sample mean:
  \[ M_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} X_k \]
  where \( X_1, X_2, X_3, \ldots, X_n \) is a sequence of RV’s.

- Today, we show what happens to \( M_n \) as \( n \) gets large.

- First, assuming the \( X_k \)'s have the same mean \( \mu \),
  we see when and how does \( M_n \to \mu \) as \( n \to \infty \).

- This is called a Law of Large Numbers (LLN).

- This is very important for many reasons including estimating the mean of the distribution of \( X \).
Second, we show that under certain conditions, the CDF of $M_n$ (properly normalized) becomes Gaussian: The Central Limit Theorem (CLT).

The CLT is very important for finding the distribution of large samples and is the cornerstone of most basic statistical analysis and of modeling of noise in systems.

We also show how to obtain approximate confidence intervals using these results.

The LLN and CLT are the 2 most fundamental results in Probability Theory.
The Sample Mean and the Laws of Large Numbers

Section 7.2
The Sample Mean

- Last time, we defined the sample mean:
  \[ M_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} X_k \]

  where \( X_1, X_2, X_3, \ldots, X_n \) is a sequence of RV’s.

- Today, we focus on the case where \( X_1, X_2, X_3, \ldots, X_n \) is an iid sequence (independent and identically distributed).

- An iid sequence occurs commonly when we repeat the same experiment independently.

- Note that the sample mean \( M_n \) is a RV.
Expectation of the Sample Mean

- Last time, we showed that expectation of the sample mean is
  \[ E[M_n] = \frac{1}{n} E[S_n] = \frac{1}{n} \sum_{k=1}^{n} E[X_k] \]

- Here we assume that \( X_1, X_2, X_3, \ldots, X_n \) is an iid sequence, so all the \( X_k \)'s have the same mean \( \mu \).

- Thus, \( E[M_n] = \frac{1}{n} \sum_{k=1}^{n} E[X_k] = \frac{1}{n} (n \mu) = \mu \).

- Thinking of \( M_n \) as an estimator of \( \mu \), the property that \( E[M_n] = \mu \) says that \( M_n \) is an “unbiased” estimator, a highly desirable property.
Variance of the Sample Mean

- Since $E[M_n] = \mu$, the variance of $M_n$ is
  \[ \text{VAR}[M_n] = E[(M_n - \mu)^2] \]

- We know that $\text{VAR}[M_n] = \text{VAR}\left[\frac{1}{n} S_n\right] = \frac{1}{n^2} \text{VAR}[S_n]$.

- We showed last time that since the $X_k$’s are iid
  \[ \text{VAR}[S_n] = \sum_{k=1}^{n} \text{VAR}[X_k] = n\sigma^2 \]
  where $\sigma^2$ is the common variance of the $X_k$’s.

- So, $\text{VAR}[M_n] = \frac{1}{n} \sigma^2$
What Happens to $M_n$ for $n$ Large?

- Recall the Chebyshev Inequality: We showed for any RV $X$ and any positive real number $\varepsilon$ that

$$P\left[ \left| X - \mu_X \right| \geq \varepsilon \right] \leq \frac{\sigma_X^2}{\varepsilon^2}$$

- Applying this to $M_n$ we get

$$P\left[ \left| M_n - \mu \right| \geq \varepsilon \right] \leq \frac{\sigma_{M_n}^2}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

- The prob that the RV $M_n$ is more than $\varepsilon$ away from $\mu$ is bounded above by $\frac{\sigma^2}{n\varepsilon^2}$. 

12/1/2014
ECSE-2500 Lecture 24: Limits
Weak Law of Large Numbers

- Chebyshev says $P\left[\left|M_n - \mu\right| \geq \varepsilon \right] \leq \frac{\sigma^2}{n\varepsilon^2}$

- Thus we have, for any iid sequence of RV’s, each with mean $\mu$ and variance $\sigma^2$, for a fixed $\varepsilon > 0$,

$$P\left[\left|M_n - \mu\right| \geq \varepsilon \right] \leq \frac{1}{n} \left(\frac{\sigma^2}{\varepsilon^2}\right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$ 

- Another way of looking at this is that

$$P\left[\left|M_n - \mu\right| < \varepsilon \right] = 1 - P\left[\left|M_n - \mu\right| \geq \varepsilon \right] \geq \left[1 - \frac{1}{n} \left(\frac{\sigma^2}{\varepsilon^2}\right)\right] \rightarrow 1 \text{ as } n \rightarrow \infty.$$ 

- So $\lim_{n \rightarrow \infty} P\left[\left|M_n - \mu\right| < \varepsilon \right] = 1$. 

12/1/2014

ECSE-2500 Lecture 24: Limits
Weak Law of Large Numbers continued

- While our proof required a finite variance for $X_k$, a more general proof yields:

- **WLLN**: For any iid sequence of RV’s with finite mean $\mu$ and for any fixed $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left[ \left| M_n - \mu \right| < \varepsilon \right] = 1.$$ 

- Thus, we know that if we take a large enough sample $n$, then the sample mean is as close as we want to the true mean $\mu$ with probability that goes to one.
WLLN Example 1: Bernoulli Trials

- Let $X_1, X_2, X_3, \ldots, X_n$ be independent Bernoulli RV’s with prob of success $= p$.

- $\mu = p$ here, so we have

$$\lim_{n \to \infty} P \left( \left| \frac{1}{n} \sum_{k=1}^{n} X_k - p \right| < \varepsilon \right) = 1.$$ 

- The average number of successes in $n$ trials is very likely to be very close to $p$ as $n$ gets large.
WLLN Example 2: One Die

- Let $X_1, X_2, X_3, \ldots, X_n$ be independent discrete uniform RV's with $S_X = \{1, 2, 3, 4, 5, 6\}$, each $X_k$ representing one independent roll of a fair die.

- We know $\mu = \sum_{x=1}^{6} x p_X(x) = \sum_{x=1}^{6} x \frac{1}{6} = 3.5$

- So, WLLN says $\lim_{{n \to \infty}} P\left[\left|\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) - 3.5\right| < \varepsilon\right] = 1$.

- The average value on the face of the die in $n$ tosses is very likely to be very close to 3.5 as $n$ gets large.
WLLN Example 2: One Die

Sample Mean $M_n$ vs. $n$ Tosses of a Fair Die
How close are $M_n$ and $\mu$?

- If $M_n$ is an estimator of $\mu$, Chebyshev can also be used for a level of confidence in our estimator.

- For example, if $\varepsilon = 0.1$ and $\sigma = 0.1$, then $\frac{\sigma^2}{n\varepsilon^2} = \frac{1}{n}$.

- So $P[|M_n - \mu| \geq 0.1] \leq \frac{1}{n}$.

- In other words, if we collect $n = 100$ samples, we expect that 99 out of 100 times, we will know that $M_n$ is within 0.1 of the true value of $\mu$.

- If we collect $n = 1000$ samples, we expect that 999 out of 1000 times, we will know that $M_n$ is within 0.1 of the true value of $\mu$. 
How close are $M_n$ and $\mu$? continued

- If instead $\varepsilon = 0.01$, $\sigma = 0.1$, then $\frac{\sigma^2}{n\varepsilon^2} = \frac{100}{n}$

- Then $P[|M_n - \mu| \geq 0.01] \leq \frac{100}{n}$

- We now need to collect $n = 10,000$ samples, if we want to expect guaranteed closeness in 99 out of 100 times, but in return the guaranteed closeness is that $M_n$ is within 0.01 of the true value of $\mu$.

- If we collect $n = 10^6$ samples, then 999 out of 1000 times, we expect that $M_n$ is within 0.01 of the true value of $\mu$. 

Ex: How close are $M_n$ and $\mu$?

- Let $X_1, X_2, X_3, \ldots, X_n$ be independent Bernoulli RV's with prob of success $= p = 0.5$.

- $\sigma^2_x = p(1-p) = 0.25$ here, so we have for $\varepsilon = 0.1$

  $$P\left[|M_n - 0.5| \geq 0.1\right] \leq \frac{25}{n}$$

- If we collect $n = 10,000$ samples, then we have

  $$P\left[|M_n - 0.5| \geq 0.1\right] \leq 0.0025.$$  

- For $\varepsilon = 0.01$ and $n = 10^6$ samples, then

  $$P\left[|M_n - 0.5| \geq 0.01\right] \leq 0.0025.$$
The Strong Law of Large Numbers

- The Strong Law of Large Number (SLLN) looks a lot like the WLLN, but the convergence is stronger:

- **SLLN**: For any iid sequence of RV’s with finite mean $\mu$ and finite variance, we have
  \[ P\left[ \lim_{n \to \infty} M_n = \mu \right] = 1. \]

- This says that for essentially any outcome in the sample space, $M_n$ is guaranteed to converge to $\mu$.

- This is a stronger form of convergence than the WLLN which gave us: 
  \[ \lim_{n \to \infty} P\left[ |M_n - \mu| < \varepsilon \right] = 1. \]

but the SLLN does not give us concrete bounds.
The Infinite Monkey Theorem

- “A monkey hitting keys at random on a typewriter for an infinite amount of time will eventually type the complete works of William Shakespeare.”

- This is a “Folk Theorem” (something people generally believe, but cannot be proven) which is really based on the Law of Large Numbers.

- “We've heard that a million monkeys at a million keyboards could produce the complete works of Shakespeare. Now, thanks to the Internet, we know that is not true.” –Robert Wilensky (Prof. of EECS, UC Berkeley)
The “Law” of Averages

- A folk theorem loosely based on the Law of Large Numbers, but usually misapplied to imply that things should “even out” in the short term, not just the long term.

- For example, if we have gotten 10 heads in a row on a fair coin, people will apply the “law” of averages to say that a tail is more likely on the next toss.

- Sometimes referred to as “we are due for a tails, after so many heads in a row”.

- If the tosses are independent, then this is false.

- We only know that eventually things will even out.
The Central Limit Theorem

Section 7.3

(but not Subsections 7.3.2–3)
The Central Limit Theorem

- The Weak and Strong Laws of Large Numbers tell us what happens to the value of the sample mean of iid RV’s as $n$ gets large.

- The Central Limit Theorem (CLT) tells us what happens to the Cumulative Distribution Function (CDF) of the sample mean of iid RV’s (when appropriately normalized):

It converges to a Gaussian CDF!
Central Limit Theorem Set Up

- For the LLNs, we normalized $S_n = \sum_{k=1}^{n} X_k$ by dividing by $n$ to get $M_n$.

- This gave us $E[M_n] = E[\frac{1}{n} S_n] = \mu$.

- For the CLT, we need to normalize the mean to zero and the variance to one.

- We have shown that for iid $X_k$’s, $E[S_n] = n \mu$ and $\text{VAR}[S_n] = n \sigma^2$.

- So the normalized $S_n$ is $Z_n = \frac{S_n - n \mu}{\sigma \sqrt{n}}$

- So, $E[Z_n] = 0$, $\text{VAR}[Z_n] = 1$, for every $n = 1, 2, 3, \ldots$

- This normalization is very important!
The Central Limit Theorem

- Let $X_1, X_2, X_3, \ldots, X_n, \ldots$ be independent and identically distributed random variables with finite mean $\mu$ and finite variance $\sigma^2$.

- Let $Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}}$, the normalized sum of the $X_k$'s.

- Then $\lim_{n \to \infty} P[Z_n \leq z] = \lim_{n \to \infty} F_{Z_n}(z) = \Phi(z)$, where

  $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$ is the standard Gaussian CDF.

- So as $n$ gets large, the CDF of the normalized sum of the $X_k$'s converges to a standard Gaussian CDF.
The Central Limit Theorem continued

- As \( n \) gets large, the CDF of the normalized sum of the \( X_k \)'s converges to a Gaussian CDF.
- This is a remarkable result in part because it asks nothing of the CDF of the \( X_k \)'s.
  - The \( X_k \)'s and hence \( S_n \) can even be discrete RVs!
The CLT continued

- As \( n \) gets large, the CDF of the normalized sum of the \( X_k \)'s converges to a Gaussian CDF.
- We CANNOT say that the PDF of \( S_n \) converges to a Gaussian PDF for discrete RVs.
- However, the PMF can “look” a lot like a Gaussian PDF.
- Consider \( M_n \) for rolling a fair die \( n \) times:

This “looks Gaussian” but it is discrete
The Central Limit Theorem continued

- As $n$ gets large, the CDF of the normalized sum of the $X_k$’s converges to a Gaussian CDF.

- We can get PDF convergence for continuous RVs:

- Consider $X_k$ uniform on $[0,1]$ for $n = 1, 2, 3, 4, 8, 16, 32$.

- $\bar{X}$ (aka “Xbar”) is another common name for $M_n$. 

![Graph showing the distribution of $X$]
The Central Limit Theorem continued

- Here is $X_k$ uniform on $[0,1]$ for $n = 1, 2, \ldots, 6$.

- The $x$-axis in each fig is $[0,n]$.
Example: CLT Where $X_k$ is Bimodal
Ex: CLT Where \( f(x) \) is Asymmetric

\[
\begin{align*}
n &= 1 & n &= 2 \\
n &= 3 & n &= 4
\end{align*}
\]
The Central Limit Theorem and Confidence Intervals
The CLT and Confidence Intervals

- Chebyshev gave us this bound: \( P\left[ |M_n - \mu| \geq \varepsilon \right] \leq \frac{\sigma^2}{n\varepsilon^2} \)

- CLT says that for large \( n \), the CDF of \( Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \) is approximately Gaussian(0,1): \( \Phi(z) \).

- Thus, \( P[Z_n \leq z] \approx \Phi(z) \), so \( P\left[ \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right] \approx \Phi(z) \),

- We can rewrite \( Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{M_n - \mu}{\sigma/\sqrt{n}} \)

- So \( P\left[ (M_n - \mu) \leq \frac{z\sigma}{\sqrt{n}} \right] \approx \Phi(z) \).
The CLT and Confidence Intervals

- CLT says that for large $n$, \( P \left[ \left( M_n - \mu \right) \leq \frac{z\sigma}{\sqrt{n}} \right] \approx \Phi(z). \)

- This also says \( P \left[ \left| M_n - \mu \right| \leq \frac{z\sigma}{\sqrt{n}} \right] \approx \Phi(z) - \Phi(-z). \)

- Recall the $Q$-function defined in Chapter 4:
  \[ Q(z) = 1 - \Phi(z) \]

- By the symmetry of the Gaussian, we have
  \[ P \left[ \left| M_n - \mu \right| \leq \frac{z\sigma}{\sqrt{n}} \right] \approx 1 - 2Q(z). \]

- This can be viewed as an approximate confidence interval for $M_n$ as an estimator of $\mu$. 
Example: Confidence Intervals

- Let $X_1, X_2, X_3, \ldots, X_n$ be independent Bernoulli RV's with prob of success $= p = 0.5$. $\sigma = \sqrt{p(1-p)} = 0.5$

- How close is $M_n$ to $\mu (= p = 0.5)$?

- For $z = 1$ and $n = 10,000$,

$$P\left[|M_n - \mu| \leq \frac{z\sigma}{\sqrt{n}} = \frac{0.5}{100} = 0.005\right] \approx 1 - 2Q(1) = 0.682.$$  

- Suppose we want this prob to be 0.9 (a 90\% confidence interval), then (from Table 4.2):

$$1 - 2Q(z) = 0.9 \quad \text{if} \quad Q(z) = 0.05, \text{i.e. if} \quad z = 1.65.$$  

- So $P\left[|M_n - \mu| \leq \frac{z\sigma}{\sqrt{n}} = 0.00825\right] \approx 0.9.$
Example: Confidence Intervals continued

- How close is $M_n$ to $\mu (= p = 0.5)$?
- For $z = 1.65$ and $n = 10,000$,
- We just showed that a 90% confidence interval here is

$$P\left[\left|M_n - \mu\right| \leq \frac{z\sigma}{\sqrt{n}} = 0.00825\right] \approx 0.9.$$ 

- In other words, we are 90% confident that the true value of the mean is in the interval

$$\left[M_n - 0.00825, M_n + 0.00825\right].$$