Lecture 3: Limit theorems

Admin: Homework 1 due Thursday

Outline: (each item roughly corresponds to one week’s material)
1. Overview of probability: Probability spaces, random variables, distribution functions, moment generating functions, expectation, conditional probability and expectation, probability inequalities, examples
2. Stochastic processes: Examples, notions of convergence, definition of a stochastic process, independence, zero-one laws, laws of large numbers, central limit theorems, stable laws

**Moment Generating Functions**

Def.: Moment generating function \( \mathbb{E}[e^{\lambda X}] \) as a function of \( \lambda \)

- does not always exist

\[ P(X = i) = \frac{e^{\lambda i}}{i!} \quad \text{Idea: Probabilities don't drop fast enough.} \]

\[ \mathbb{E}[e^{\lambda X}] = \sum_{i=0}^{\infty} \frac{e^{\lambda i} \frac{1}{i!}}{i!} = \infty \text{ for any } \lambda > 0 \]

Characteristic function \( \mathbb{E}[e^{i\lambda X}] \)
- always exists (for real \( \lambda, X \)) since \( |e^{i\lambda X}| = 1 \)

Joint moment generating function \( \mathbb{E}[e^{\lambda X + \mu Y}] = \psi(\lambda, \mu) \)

Fact: The (joint) moment generating function, when it exists, uniquely determines the (joint) distribution.

The (joint) characteristic function uniquely determines the (joint) distribution.

For examples, all the moments \( \mathbb{E}[X^n] \) can be read off \( \mathbb{E}[e^{\lambda X}] \):

\[ \mathbb{E}[X^n] = \left. \frac{d^n}{d\lambda^n} \mathbb{E}[e^{\lambda X}] \right|_{\lambda = 0} \]

\[ = \left. \frac{d^n}{d\lambda^n} \left( \mathbb{E}[e^{\lambda X}] \right) \right|_{\lambda = 0} \text{ provided you can interchange } \frac{d^n}{d\lambda^n} \text{ and } \mathbb{E} \]

Examples:
Examples:

<table>
<thead>
<tr>
<th>Discrete Probability Distribution</th>
<th>Probability Mass Function, ( p(x) )</th>
<th>Moment Generating Function, ( \phi(t) )</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial with parameters ( n, p ), ( 0 \leq p \leq 1 )</td>
<td>( \binom{n}{x} p^x (1-p)^{n-x} )</td>
<td>( (pe^t + (1-p))^n )</td>
<td>( np )</td>
<td>( np(1-p) )</td>
</tr>
<tr>
<td>( x = 0, 1, \ldots, n )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson with parameter ( \lambda &gt; 0 )</td>
<td>( \frac{e^{-\lambda} \lambda^x}{x!} )</td>
<td>( \exp{\lambda(e^t - 1)} )</td>
<td>( \lambda )</td>
<td>( \lambda )</td>
</tr>
<tr>
<td>( x = 0, 1, 2, \ldots )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometric with parameter ( 0 \leq p \leq 1 )</td>
<td>( p(1-p)^{x-1} )</td>
<td>( \frac{pe^t}{1 - (1-p)e^t} )</td>
<td>( \frac{1}{p} )</td>
<td>( \frac{1-p}{p^2} )</td>
</tr>
<tr>
<td>( x = 1, 2, \ldots )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative binomial with parameters ( r, p )</td>
<td>( \binom{x-1}{r-1} p^r (1-p)^{x-r} )</td>
<td>( \frac{pe^t}{1 - (1-p)e^t} )</td>
<td>( \frac{r}{p} )</td>
<td>( \frac{r(1-p)}{p^2} )</td>
</tr>
<tr>
<td>( x = r, r+1, \ldots )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 1.4a:

\( X \sim \text{Normal}(\mu, \sigma^2) \)

\[
\mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \cdot e^{tx}
\]

\[
= \frac{1}{\sqrt{2\pi} \sigma^2} \int_{-\infty}^{\infty} dx \cdot e^{-\frac{1}{2\sigma^2} (x-\mu + \frac{1}{\sigma^2} t x)^2 - \frac{1}{2} \sigma^2 t x}
\]

\[
= e^{\mu t + \frac{1}{2} \sigma^2 t^2}
\]
<table>
<thead>
<tr>
<th>Continuous Probability Distribution</th>
<th>Probability Density Function, ( f(x) )</th>
<th>Moment Generating Function, ( \phi(t) )</th>
<th>Mean ( \mu )</th>
<th>Variance ( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform over ((a, b)) ( a &lt; x &lt; b )</td>
<td>( \frac{1}{b - a} )</td>
<td>( e^b - e^a )</td>
<td>( a + b )</td>
<td>( (b - a)^2 )</td>
</tr>
<tr>
<td>Exponential with parameter ( \lambda &gt; 0 ) ( \lambda e^{-\lambda x}, x \geq 0 )</td>
<td>( \frac{\lambda}{\lambda - t} )</td>
<td>( \frac{1}{\lambda} )</td>
<td>( 1 )</td>
<td>( \lambda^2 )</td>
</tr>
<tr>
<td>Gamma with parameters ((n, \lambda), \lambda &gt; 0 ) ( \lambda e^{-\lambda x}(\lambda x)^{n-1} \frac{1}{(n-1)!}, x \geq 0 )</td>
<td>( \frac{\lambda}{\lambda - t} )</td>
<td>( n )</td>
<td>( \frac{n}{\lambda} )</td>
<td>( \frac{n}{\lambda^2} )</td>
</tr>
<tr>
<td>Normal with parameters ((\mu, \sigma^2)) ( \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty &lt; x &lt; \infty )</td>
<td>( \exp\left(\frac{\mu t}{\lambda} + \frac{\sigma^2 t^2}{2}\right) )</td>
<td>( \mu )</td>
<td>( \sigma^2 )</td>
<td></td>
</tr>
<tr>
<td>Beta with parameters (a, b, a &gt; 0, b &gt; 0) ( cx^{a-1}(1-x)^{b-1}, 0 &lt; x &lt; 1 )</td>
<td>( \frac{\Gamma(a+b)}{a \Gamma(a) \Gamma(b)} )</td>
<td>( \frac{a}{a + b} )</td>
<td>( \frac{ab}{(a + b)^2(a + b + 1)} )</td>
<td></td>
</tr>
</tbody>
</table>

\( X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2) \) independent

\[
\begin{align*}
\mathbb{E}[e^{t(X+Y)}] &= \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] \\
&= \exp((\mu_X+\mu_Y)t + \frac{1}{2}t^2(\sigma_X^2+\sigma_Y^2))
\end{align*}
\]

\( \Rightarrow \) by uniqueness,

\( X + Y \sim N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2) \).

**Example 1.5b: Trapped miner's random walk**

\[\begin{array}{c}
\text{safety} \\
\text{3} \\
\text{2} \\
\text{1} \\
\text{0} \\
\end{array}\]

\( X = \text{time he reaches safety} \)

**Q:** What is \( \mathbb{E}[X] \)? \( \text{Var}(X) \)?

**Answer:**

\( Y = 1^{st} \text{ door chosen} \)

\[
\begin{align*}
\mathbb{E}[e^{tX}] &= \frac{1}{3} \mathbb{E}[e^{tX} | Y = 1] + \frac{1}{3} \mathbb{E}[e^{tX} | Y = 2] + \frac{1}{3} \mathbb{E}[e^{tX} | Y = 3] \\
&= \frac{1}{3} \left( e^{2t} + (e^{3t} + e^{5t}) \mathbb{E}[e^{tX}] \right) - \frac{1}{3} e^{2t}
\end{align*}
\]
\[ E[X] = \frac{d}{dt} E[1] \bigg|_{t=0} = 0 \]

\[ E[X^2] = \frac{d^2}{dt^2} E[1] \bigg|_{t=0} = 198 \quad \Rightarrow \text{Var}(X) = 98 \]

\[
D\left[ \frac{1}{3} e^{2t} \right] \bigg|_{t=0} \quad \text{and} \quad D\left[ \frac{1}{3} e^{2t} \right] \bigg|_{t=0} \]

**LIMIT THEOREMS**

Recall:

- **Chebyshev's inequality:**
  \[
P\left[ |X - \mu_X| \geq k \sigma_X \right] \leq \frac{1}{k^2}
  \]
  where \( \mu_X = E[X] \), \( \sigma_X^2 = E[(X - \mu_X)^2] \)

- **Chernoff/Hoeffding Lemma:**
  For \( X_1, X_2, \ldots, X_n \) independent, \( a_i \leq X_i \leq b_i \),
  let \( X = \sum X_i \) and \( \mu = E[X] \). Then,
  \[
P\left[ |X - \mu| \geq \delta \right] \leq 2 \exp \left( -\frac{2\delta^2}{\sum (b_i - a_i) \cdot \sigma_X^2} \right)
  \]

**Example:** \( n \) independent, fair coin flips

\[
X_i = \begin{cases} 0 & \text{w/ prob. } \frac{1}{2} \\ 1 & \text{w/ prob. } \frac{1}{2} \end{cases}
\]

\[
X = \sum X_i
\]

\[
\Rightarrow E[X] = \frac{1}{2} = \mu \quad \text{and} \quad \sigma = \sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{\sqrt{n}}
\]

- \( P[X > 3n/4] \) ?
  \[
  \frac{4}{n} \leq \frac{3n}{4} \Rightarrow \frac{2}{n} = \frac{2}{n} \quad \text{(Markov)}
  \]
\[
\begin{align*}
\frac{n}{2} & \leq \frac{n/2}{3n/4} = \frac{2}{3} \quad \text{(Markov)} \\
\frac{4}{n} & \leq \frac{4}{n} \quad \text{(Chebyshev, } k = \frac{1}{2} \sqrt{n}) \\
2e^{-2n/8} & \leq 2e^{-2n/8} = 2e^{-2} \quad \text{(Chernoff, } f = \frac{n}{4})
\end{align*}
\]

\[\Pr[X > n/2 + \sqrt{n}] \leq \frac{n/2}{n/2 + \sqrt{n}} \approx 1 - \frac{2}{\sqrt{n}} \quad \text{(Markov)} \]

\[\leq \frac{1}{4} \quad \text{(Chebyshev, } k = 2) \]

\[\leq 2e^{-2n/8} = 2e^{-2} \approx 0.27 \quad \text{(Chernoff)}\]

The tails of the CDF have exponentially small probability outside \((1 \pm \frac{n}{2})\).

But what is the correct asymptotic behavior near \(n/2\), e.g., at \(n/2 \pm k\sqrt{n}\)?

**Theorem:** Central limit theorem

\[X_1, X_2, \ldots \text{ i.i.d. w/ mean } \mu, \text{ variance } \sigma^2,\]

\[\lim_{n \to \infty} \Pr\left[ \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n} \cdot \sigma} \leq a \right] = \int_{-\infty}^{\infty} dx \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \]

\(\text{CDF for normal distribution}\)

\[\Rightarrow \text{In the previous example,} \]

\[\Pr[X > n/2 + \sqrt{n}] = 1 - \Pr[\frac{X - n/2}{\sqrt{n}} \leq 2] \]

\[\approx 1 - \int_{-\infty}^{2} dx \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \]

\[\approx 0.023\]
\[ 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx \approx N \]

**Matlab:**

```matlab
>> help normcdf

normcdf Normal cumulative distribution function (cdf).
    P = normcdf(X,MU,SIGMA) returns the cdf of the normal distribution with
    mean MU and standard deviation SIGMA, evaluated at the values in X.
    The size of P is the common size of X, MU and SIGMA. A scalar input
    functions as a constant matrix of the same size as the other inputs.

Default values for MU and SIGMA are 0 and 1, respectively.

[P,PLO,PUP] = normcdf(X,MU,SIGMA,PCOV,ALPHA) produces confidence bounds
for P when the input parameters MU and SIGMA are estimates. PCOV is a
2-by-2 matrix containing the covariance matrix of the estimated parameters.
ALPHA has a default value of 0.05, and specifies 100*(1-ALPHA)% confidence
bounds. PLO and PUP are arrays of the same size as P containing the lower
and upper confidence bounds.

See also erf, erfc, normfit, norminv, normlike, normpdf, normrnd, normstat.

Reference page in Help browser
    help normcdf
```

```matlab
>> [normcdf(0) normcdf(1) normcdf(2) normcdf(3)]

ans =

    0.5000    0.8413    0.9772    0.9987
```

**Mathematica:**
Remark: How fast is the convergence?

Berry-Esseen theorem:

\[
\left| P \left[ \frac{X_1 + \cdots + X_n - n \mu}{\sigma \sqrt{n}} < a \right] - \Phi(a) \right| \leq \frac{E |X - \mu|^3}{2 \sigma^3 \sqrt{n}}
\]

if this exists

Remark: Central limit theorem: sum of independent increments \( \Rightarrow \) Gaussian

\( \Rightarrow \) Gaussians are everywhere!

More limit theorems:

Theorem: Weak law of large numbers
For $X_1, X_2, \ldots$ i.i.d. with mean $\mu$, $\forall \varepsilon > 0$,

$$\mathbb{P}\left[ \lim_{n \to \infty} \frac{1}{n} (X_1 + \cdots + X_n) - \mu \right] = 0$$

**Theorem** \textbf{Strong law of large numbers:}

For $X_1, X_2, \ldots$ i.i.d. w/ mean $\mu$,

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

(“almost surely,” $\lim_{n \to \infty} \frac{1}{n} (X_1 + \cdots + X_n) = \mu$)

**Remark:** These theorems hold under more general conditions than the Chernoff bounds or CLT, e.g., for the strong law $\operatorname{Var}(X)$ could be $\infty$, but they do not guarantee the speed of convergence.

**Easy exercise:**

- Give an example of a random variable $X$ with $\mathbb{E}[X] = \infty$.
- Give an example with $\mathbb{E}[X] = 0$, $\operatorname{Var}(X) = \mathbb{E}[X^2] = \infty$.

**Answer:** For example,

$$\mathbb{P}[X = \pm j] = \frac{1/2^j}{\sum_{j=1}^{\infty} 1/2^j} \quad \text{for } j = 1, 2, \ldots$$

$$\mathbb{E}[X] = \frac{1}{2} \sum_{j=1}^{\infty} \frac{1}{j} = \infty.$$  

Or, $\mathbb{P}[X = \pm j] = \frac{1/2^j}{2 \sum_{j=1}^{\infty} 1/2^j} \quad \text{for } j = \pm 1, \pm 2, \pm 3, \ldots$

$$\mathbb{E}[X] = 0 \quad \text{and} \quad \mathbb{E}[X^2] = \frac{1}{2} \sum_{j=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{j} = \infty.$$  

**Example:**
Figure 1.5: The CDF $F_{S_n}(s)$ of $S_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are typical IID rv's and $n$ takes the values 4, 20, and 50. The particular rv $X$ in the figure is binary with $p_X(1) = 1/4$, $p_X(0) = 3/4$. Note that the mean of $S_n$ is proportional to $n$ and the standard deviation to $\sqrt{n}$. [Gallager, Ch. 1]

Figure 1.9: The same CDF as Figure 1.5, scaled differently to give the CDF of the sample average $Z_n$. It can be visualized that as $n$ increases, the CDF of $Z_n$ becomes increasingly close to a unit step at the mean, 0.25, of the variables $X$ being summed.
Figure 1.11: The same CDF’s as Figure 1.5 normalized to 0 mean and unit standard deviation, i.e., the CDF’s of $Z_n = \frac{S_n - n\bar{X}}{\sigma\sqrt{n}}$ for $n = 4, 20, 50$. Note that as $n$ increases, the CDF of $Z_n$ slowly starts to resemble the normal CDF.

Proofs of the limit theorems:

Markov ✓ Chebyshev ✓ Chernoff ✓

Weak law of large numbers

For $X_1, X_2, \ldots$ i.i.d. with mean $\mu$ and $\mathbb{E}[X_i^2] < \infty$,

$$\mathbb{P}\left[ \left| \frac{1}{n}(X_1 + \cdots + X_n) - \mu \right| > \varepsilon \right] \xrightarrow{n \to \infty} 0$$

Proof: $\text{Var}\left( \frac{1}{n}(X_1 + \cdots + X_n) \right) = \frac{\sigma^2}{n}$
Chebyshev \[ \Rightarrow \Pr \left[ \frac{1}{n}(X_1 + \cdots + X_n) - \mu \geq \frac{\sigma^2}{\varepsilon} \right] \leq \frac{\varepsilon^2}{\sigma^2/n} \rightarrow 0 \]

Remark: To prove this theorem without assuming \( \mathbb{E}[X_i^2] < \infty \), use a truncation argument: Apply the above theorem to the variables

\[ Y_i = \begin{cases} X_i & \text{if } \mu - b < X_i < \mu + b \\ \mu - b & \text{if } X_i \leq \mu - b \\ \mu + b & \text{if } X_i \geq \mu + b \end{cases} \]

Then let \( b \to \infty \) carefully with \( n \), using a Markov inequality (and union bound) to bound \( \Pr[\text{any } Y_i \neq X_i] \).

Strong law of large numbers

For \( X_1, X_2, \ldots \) i.i.d. with mean \( \mu \) and \( \mathbb{E}[X_i^2] < \infty \)

\[ \Pr \left[ \lim_{n \to \infty} \frac{1}{n}(X_1 + \cdots + X_n) = \mu \right] = 1. \]

Proof:

Assume that \( \mu = 0 \)!

(If \( \mu \neq 0 \), replace each \( X_i \) with \( X_i - \mu \).)

Let \( S_n = X_1 + \cdots + X_n \).

\[ \mathbb{E}[S_n^4] = \mathbb{E}[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)] \]

\[ = \sum_{i,j,k,l} \mathbb{E}[X_i X_j X_k X_l] \]

This sum has terms

\[ \mathbb{E}[X_i^4], \mathbb{E}[X_i^3 X_j], \mathbb{E}[X_i^2 X_j^2], \mathbb{E}[X_i X_j X_k X_l] \]

with \( i, j, k, l \) all different.
with \( i, j, k, l \) all different

\[
= n \cdot \mathbb{E}[X_i^4] + \binom{n}{2} \cdot \binom{4}{2} \cdot \mathbb{E}[X_i^2 X_i^2] \\
= n \cdot K + 3n(n-1) \cdot \sigma^4
\]

\[
\Rightarrow E \left[ \sum_{n=1}^{\infty} \frac{S_n^4}{n^4} \right] < \infty \quad \checkmark
\]

The proof now follows by

**Lemma**: \( Z_1, Z_2, Z_3, \ldots \)

If \( \sum_{n=1}^{\infty} \mathbb{E}[|Z_n|] < \infty \), then \( \mathbb{P} [ \lim_{n \to \infty} Z_n = 0 ] = 1 \).

Because \( \lim_{n \to \infty} (\frac{S_n}{n})^4 = 0 \Rightarrow \lim_{n \to \infty} \frac{S_n}{n} = 0 \), as desired.

**Proof of the lemma**:

Consider

\[
\sum_{n} \mathbb{P}(|Z_n| > \varepsilon) \leq \sum_{n} \frac{\mathbb{E}[|Z_n|]}{\varepsilon} < \infty
\]

\[
\Rightarrow \text{ for only finitely many } n, \ |Z_n| > \varepsilon
\]

**Borel-Cantelli Lemma** (w/ prob. 1)

ie., \( \lim_{n \to \infty} Z_n = 0 \quad \checkmark \)

**Remark**: Why did we need the 4th power?

- \( \mathbb{E}[S_n^4] \) doesn't work. We can't apply Markov's inequality in the lemma since \( S_n^4 \) can be negative.
- \( \mathbb{E}[S_n^2] = n \cdot \mathbb{E}[X_i^2] = n \sigma^2 \),
  so \( \sum_{n=1}^{\infty} (\frac{S_n}{n})^2 = \infty \); Borel-Cantelli does not apply.