This module:

1) Discusses several bounds on tail probabilities:
   - Markov inequality
   - Chebyshev inequality
   - Chernoff bound

2) Introduces the central limit theorem and law of large numbers, which describe convergence of empirical averages to their means.

Suggested reading material:

Sections 4.4, 4.6, 4.7 (CLT), and 8.8 in Stark & Woods 4th edition.
3rd edition: 4.4, 4.6, 4.7, 6.8.

Motivation

We only know basic statistics about some random variable — not a precise characterization of the distribution. Using this partial information, we want to bound the probabilities of rare events.

Example application

A router processes 1,000 packets per second. What's the probability that more than 1,500 packets will arrive during one second?
If we design the router's CPU to process, say, 1300 packets per second, then if 1500 arrive the router might be delayed by several seconds. (Or packets might cause overflow of the queue.)

This is why it's useful to understand probabilities of bad events.

We begin with a simple bound.

**Markov Inequality** Let \( X \) be a non-negative RV, then \( \Pr(X \geq n) \leq \frac{E[X]}{n} \)

**Proof**

Define an indicator function,

\[
I\{X \geq n\} = \begin{cases} 1 & X \geq n \\ 0 & \text{else} \end{cases}
\]

over range of valid \( X \), we have \( X \geq n I\{x \geq n\} \).
Therefore, \( E[X] \geq E[n \cdot I_{x=n}] \).

But \( I_{x=n} = 1 \) iff \( X = n \), else it is 0, and therefore

\[
E[I_{x=n}] = 1 \cdot Pr(X = n) + 0 \cdot Pr(X < n) = Pr(X = n).
\]

we can now write

\[
E[X] \geq E[n \cdot I] = n \cdot E[I] = n \cdot Pr(X = n).
\]

In summary,

\[
Pr(X = n) \leq \frac{E[X]}{n}.
\]
Active learning assignment

A router processes $X$ packets per second. Given that $E[X] = 1000$, please compute $\Pr(X > 1500)$.

Note: If the router's CPU can process 1300 per second and 1500 actually arrive, then the queue contains 200 packets.

Note: Modern routers are much faster 😊
The proof of the Markov inequality relied on the observation that \( X \geq n \cdot I\{X=n\} \) for \( X \geq n \).

Let's take this type of idea and extend it:

\[
\frac{1}{n^2}(X-\mu)^2
\]

\[
I(\{X-\mu \geq n\})
\]

\[
\mu-n \quad \mu \quad \mu+n
\]

Define \( f_1(x) = I\{ |x-\mu| \geq n \} \).

This means

\[
f_1(x) = \begin{cases} 
1 & x \geq \mu+n \\
1 & x \leq \mu-n \\
0 & \text{else}
\end{cases}
\]

Also, let \( f_2(x) = \frac{1}{n^2} (X-\mu)^2 \).

Observe that \( f_2(\mu-n) = f_2(\mu+n) = 1 \), just like \( f_1(\mu-n) \) and \( f_1(\mu+n) \).

In general, \( f_2(x) \geq f_1(x) \).
Clearly,
\[ E[f_1(x)] = E[1 \{ |x-\mu| \geq \eta \}] \]
\[ = P_r( |x-\mu| \geq \eta ) + 0 \cdot P_r( |x-\mu| < \eta ) \]
\[ = P_r( |x-\mu| \geq \eta ). \]

We now have
\[ E[f_2(x)] = \frac{1}{\eta^2} E[(X-\mu)^2] \]
\[ \geq E[f_1(x)] \]
\[ = P_r( |x-\mu| \geq \eta ). \]

Therefore,
\[ P_r( |x-\mu| \geq \eta ) \leq \frac{1}{\eta^2} E[(X-\mu)^2]. \]

**How can we utilize this?**

One option is to choose \( \mu = E[X] \) and \( \eta = \alpha \cdot \text{std}(X) \).

\[ P_r( |x-\mu| \geq \eta ) = P_r( |x-E[X]| \geq \alpha \cdot \text{std}(X) ) \]
\[ \leq \frac{1}{\eta^2} E[(X-\mu)^2]. \]

But \( E[(X-\mu)^2] = E[(X-E[X])^2] = \text{Var}(X) \).

Therefore,
\[ \frac{1}{\eta^2} E[(x-\mu)^2] = \frac{\text{Var}(X)}{(\alpha \cdot \text{std}(X))^2} = \frac{1}{\alpha^2}. \]
This result, 
\[ \Pr( |X-\mu| \geq \eta) \leq \frac{1}{\eta^2} \mathbb{E}[ (X-\mu)^2] \]
which can be interpreted:
\[ \Pr( |X-\mathbb{E}[X]| \geq a \cdot \text{std}(X) ) \leq \frac{1}{a^2} \]
is known as the **Chebychev inequality**.

Let's now employ this to solve a midterm question from 2010!
Please write down your first and last name:

(first) 

(last)

Note that there are 4 pages.

1. Consider a random variable $X$ such that $E[X] = 0$ and $E[X^2] = 1$.
   
   (a) Derive an upper bound for $Pr(|X| \geq 2)$. Show your work.

   $\mu = 0$
   $\eta = 2$

   \[
   Pr(|X - \mu| \geq \eta) = Pr(|X - 0| \geq 2) = Pr(|X| \geq 2) \\
   \leq \frac{1}{\eta^2} E[(X - \mu)^2] \\
   = \frac{1}{4^2} E[X^2] = \frac{1}{4} \cdot 1 = \frac{1}{4}
   \]

   (b) Find a distribution for $X$ such that the bound in Part 1a holds with equality.

   Intuition - put $\frac{1}{4}$ of probability mass at $\pm 2$ such that expected value is 0:
   \[
   Pr(X = -2) = Pr(X = +2) = \frac{1}{8}, \quad Pr(X = 0) = \frac{3}{4}
   \]

   Check:
   $E[X] = -2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 0 \cdot \frac{3}{4} = 0$, $E[X^2] = \frac{1}{8} (-2)^2 + \frac{1}{8} (2)^2 = 1$

   (c) Derive an upper bound for $Pr(|X - \mu| \geq 2)$. Show your work.

   $\mu = 1$
   $\eta = 2$

   \[
   Pr(|X - \mu| \geq \eta) \leq \frac{1}{\eta^2} E[(X - \mu)^2] \\
   = \frac{1}{4} (E[X^2] - 2E[X] + E[\mu^2]) \\
   = \frac{1}{4} (1 - 2 \cdot 0 + 1) = \frac{1}{2}
   \]

   (d) Find a distribution for $X$ such that the bound in Part 1c holds with equality.

   Same intuition: put probability $\frac{1}{2}$ at $X = -1$ and $X = +3$. Putting $\frac{1}{2}$ at $X = -1$ requires the other $\frac{1}{2}$ of probability mass to be at $X = +1$.

   $\Rightarrow Pr(X = \pm 1) = \frac{1}{2}$ each.

   \[
   E[X] = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot (+1) = 0 \\
   E[X^2] = \frac{1}{2} \cdot (1^2) + \frac{1}{2} \cdot (-1)^2 = 1 \\
   Pr(|X - \mu| \geq \eta) = \frac{1}{2}
   \]
Now take $f_1(x) = 1_{\{X = n\}}$ and
$f_2(x) = e^{a(x-n)}$ for $a > 0$.
We can see that $f_2(x) \geq f_1(x)$, and so

$$E[f_2(x)] = E[e^{a(x-n)}]$$
$$\geq E[f_1(x)]$$
$$= P(X \geq n).$$

But

$$E[e^{a(x-n)}] = E[e^{-an \cdot e^{ax}}]$$
$$= e^{-an} \cdot E[e^{ax}]$$
$$= e^{-an} \cdot M_X(a),$$

where $M_X(a)$ is moment generating function of
$X$ evaluated at $a$. 

The diagram illustrates the function $e^{a(x-n)}$ for $x = n$, showing the growth of the function as $x$ increases.
Now that we know \( \Pr(X \geq \eta) \leq e^{-an \cdot M_X(a)} \), we optimize over \( a > 0 \) to get the tightest bound. That is, the smallest upper bound.

**Example (4.6-1)**

\( X \sim N(\mu, \sigma^2) \)

We will compute \( \Pr(X \geq \eta) \), where \( \eta \geq \mu \).

It's given that \( M_X(a) = \exp\{\mu a + \frac{\sigma^2 a^2}{2}\} \)

We know \( \Pr(X \geq \eta) \leq e^{-an \cdot M_X(a)} \)

Define \( f(a) = e^{-an \cdot M_X(a)} \)

\[
= \exp\{-an + \mu a + \frac{\sigma^2 a^2}{2}\}
\]

Now let \( g(a) = \ln(f(a)) \)

\[
= a(\mu - \eta) + \frac{\sigma^2 a^2}{2}
\]

Taking the derivative

\[
g'(a) = \mu - \eta + \sigma^2 a
\]
The optimal $a^*$ satisfies $g'(a^*) = 0,$

$$0 = \mu - \eta + \sigma^2 a^*$$

$$\Rightarrow a^* = \frac{\eta - \mu}{\sigma^2}$$

Now compute the bound,

$$f(a^*) = \exp \left\{ -a^* \eta + \mu a^* + \sigma^2 \left( \frac{a^*}{2} \right)^2 \right\}$$

$$= \exp \left\{ \frac{\eta - \mu}{\sigma^2} (\mu - \eta) + \frac{\sigma^2}{2} \left( \frac{\eta - \mu}{\sigma^2} \right)^2 \right\}$$

$$= \exp \left\{ - \left( \frac{\eta - \mu}{\sigma^2} \right)^2 + \frac{1}{2} \left( \frac{\eta - \mu}{\sigma^2} \right)^2 \right\}$$

$$= \exp \left\{ - \frac{1}{2} \left( \frac{\eta - \mu}{\sigma^2} \right)^2 \right\}$$

Therefore, $\Pr(X \geq \eta) \leq e^{-\frac{1}{2} \left( \frac{\eta - \mu}{\sigma^2} \right)^2}.$