Systems and Signals 344

2025

Lecture 24

Central Limit Theorem & The Sample Mean & Deviation of RV from the Expected Value

Section 9.4, 10.1, 10.2

Recap: Expected Values of Sums

For any set of random variables $X_1, \ldots, X_n$, the sum $W_n = X_1 + \cdots + X_n$ has expected value and variance:

$$\mathrm{E}[W_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]$$

$$\mathrm{Var}[W_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{Cov}[X_i, X_j]$$

When $X_1, \ldots, X_n$ are uncorrelated

$$\mathrm{Var}[W_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$$

Recap: MGF of the Sum of Independent RVs

For a set of independent random variables $X_1, \ldots, X_n$, the moment generating function of $W = X_1 + \cdots + X_n$ is:

$$\phi_W(s) = \phi_{X_1}(s)\phi_{X_2}(s)\ldots\phi_{X_n}(s)$$

When $X_1, \ldots, X_n$ are *iid*, each with MGF $\phi_{X_i}(s) = \phi_X(s)$,

$$\phi_W(s) = [\phi_X(s)]^n$$

Central Limit Theorem

The ***central limit theorem*** *states that the CDF of the the sum of n independent random variables converges to a Gaussian CDF as n grows without bound.*

For values of n encountered in many applications, the approximate Gaussian model provides a very close approximation to the actual model.

Using the Gaussian approximation is far more efficient computationally than working with the exact probability model of a sum of random variables.

Central Limit Theorem

- Very many practical situations produces PMFs and PDFs which approximately follow a **bell-shaped curve**.
- The **central limit theorem** explains why so many practical phenomena produce data that can be modeled as a Gaussian distribution.
- We can use the central limit theorem to *estimate* probabilities associated with the *iid* sum $W_n = X_1 + \cdots + X_n$.

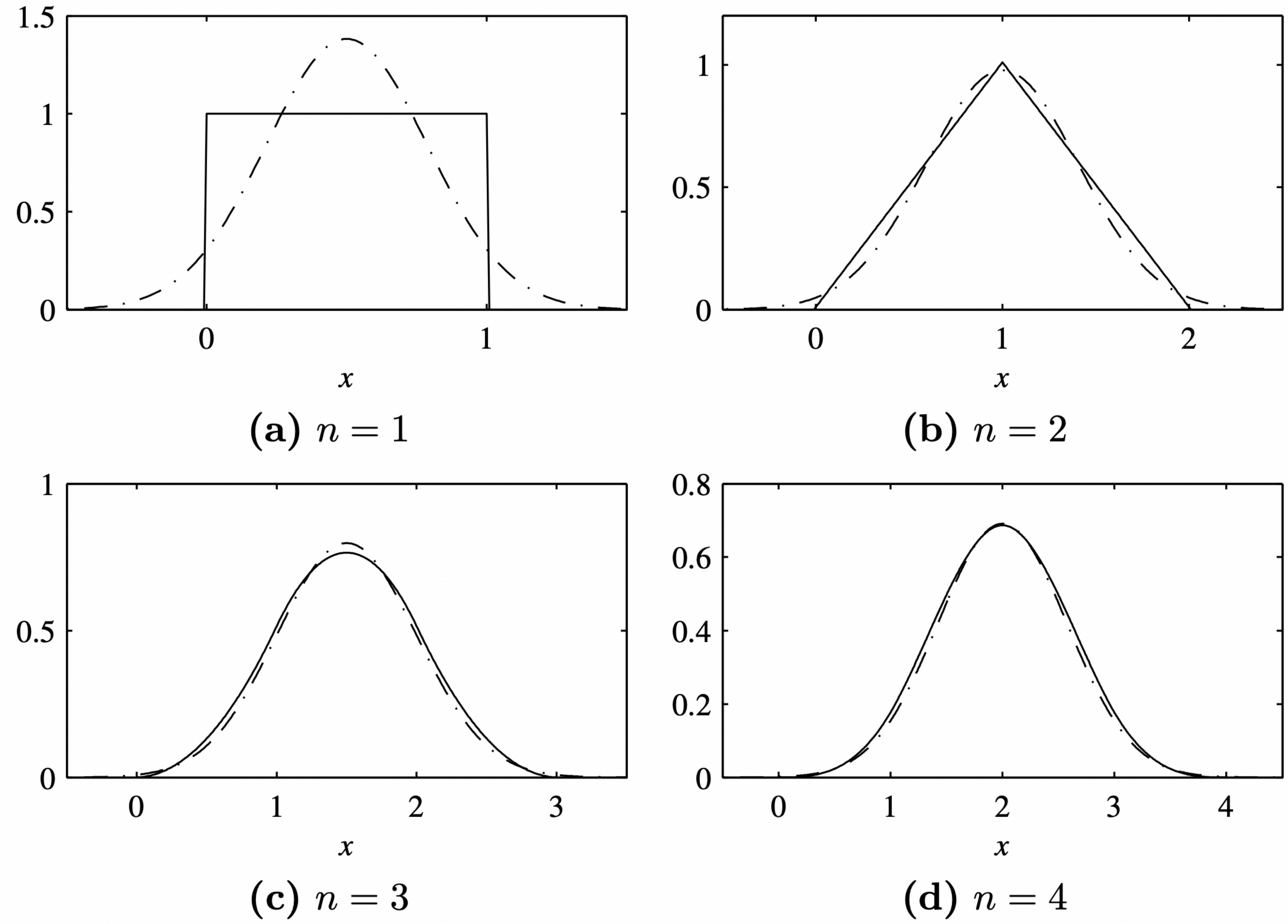
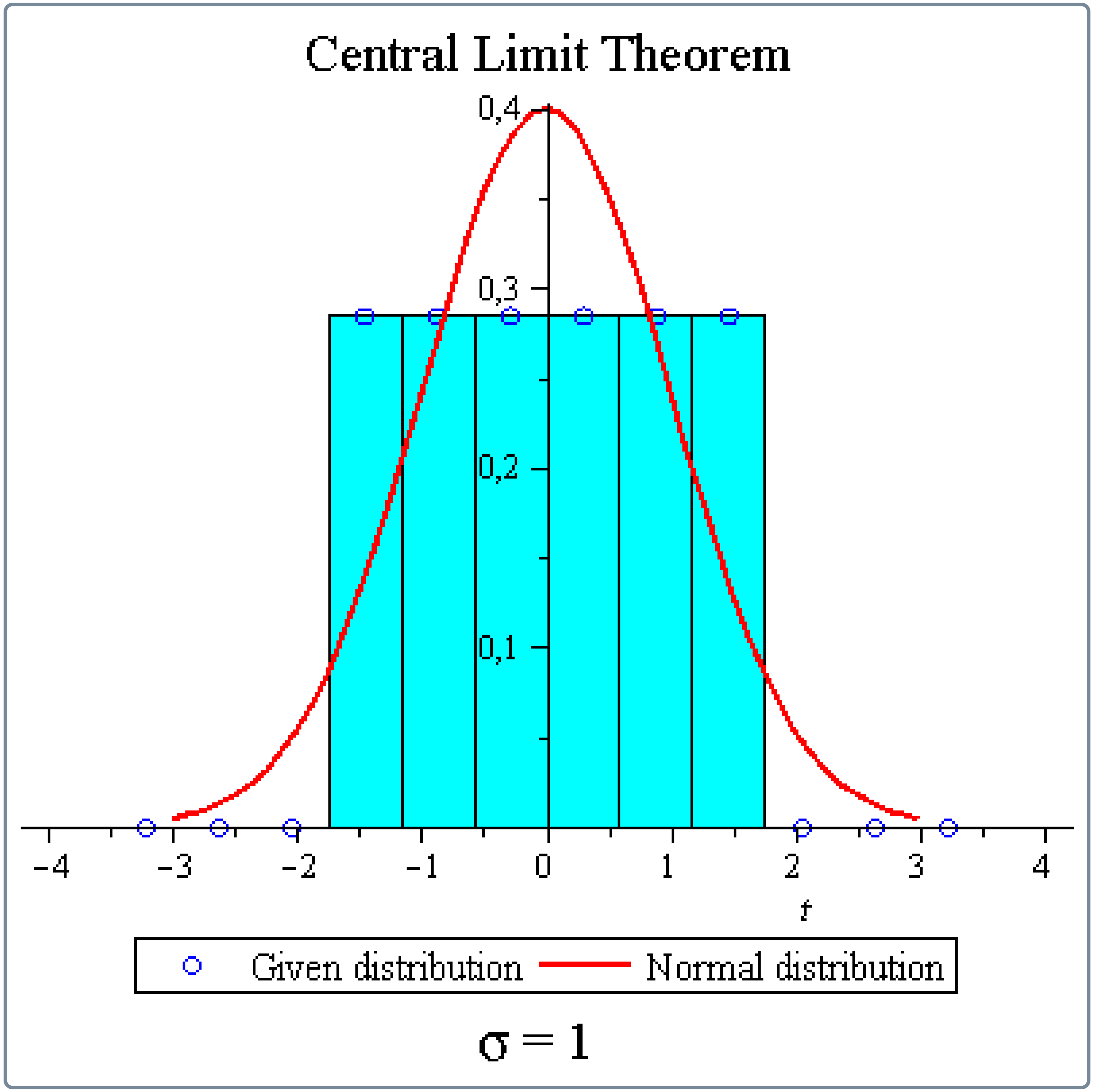
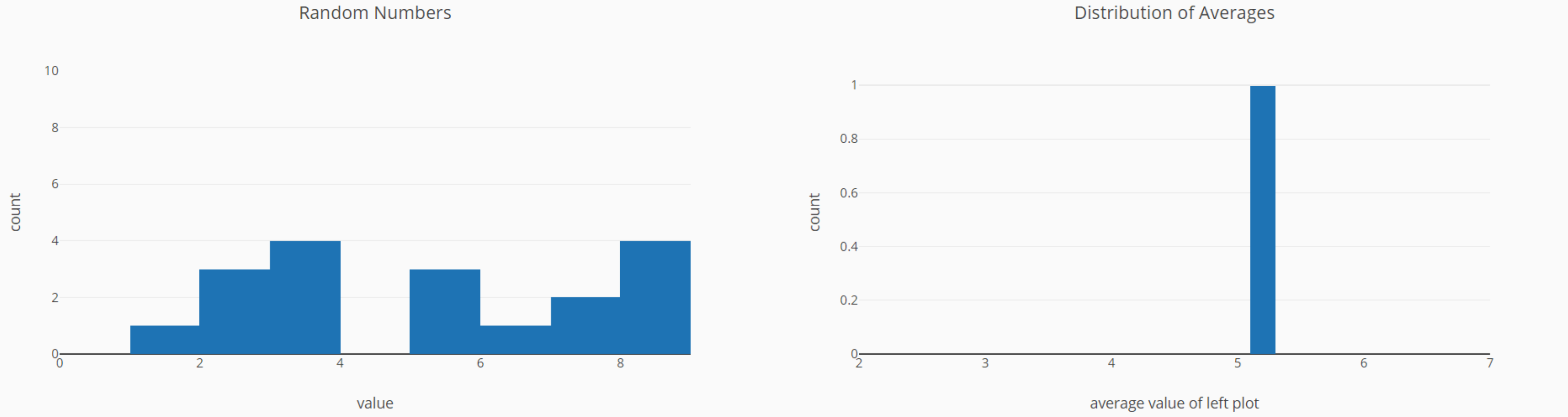


Figure 9.2 The PDF of W_n, the sum of n uniform $(0, 1)$ random variables, and the corresponding central limit theorem approximation for $n = 1, 2, 3, 4$. The solid — line denotes the PDF $f_{W_n}(w)$, and the broken $- \cdot -$ line denotes the Gaussian approximation.

Remember convolutions create these shapes.

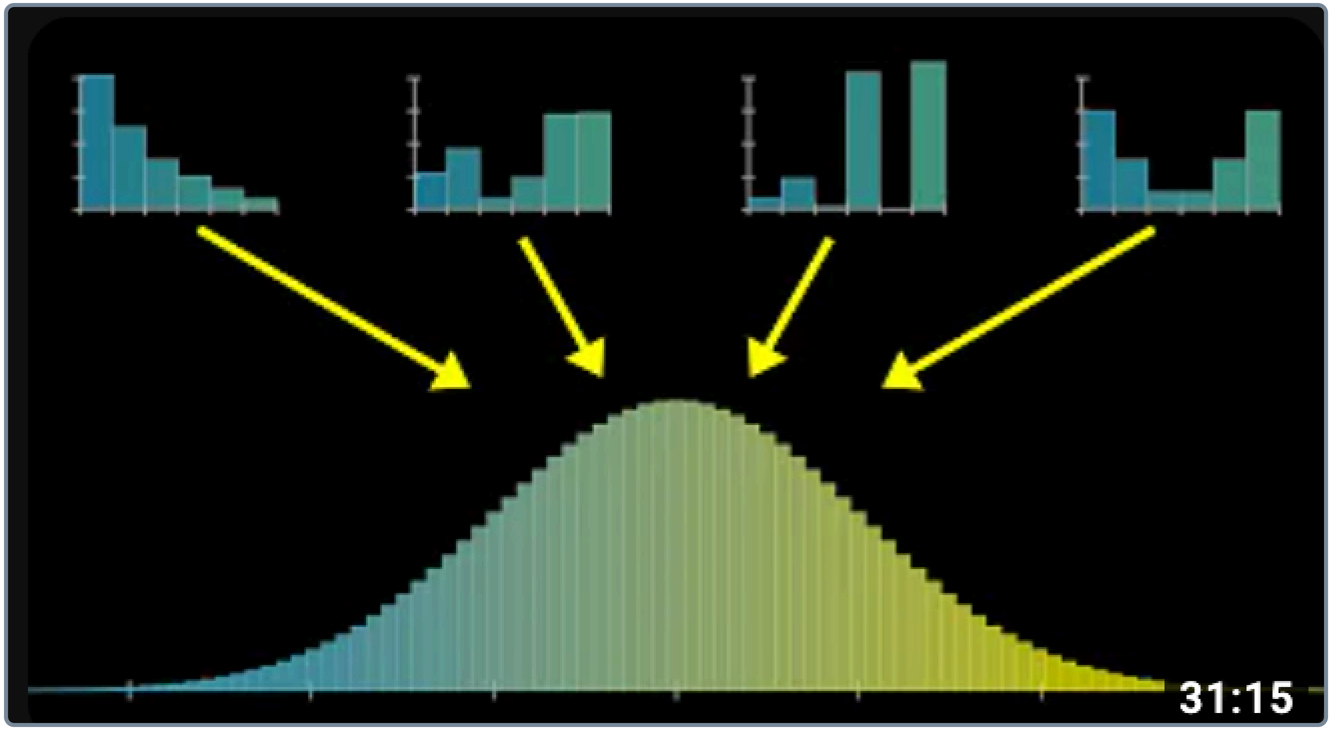
Even for *discrete* random variables





Generate 20 random numbers from 0 and 9. Find their average. Repeat 1000 times. The averages will approximate a normal distribution (bell curve) centered at 4.5.

Great video by 3Blue1Brown



https://www.youtube.com/watch?v=zeJD6dqJ5lo

Central Limit Theorem

- We can use the central limit theorem to *estimate* probabilities associated with the *iid* sum $W_n = X_1 + \cdots + X_n$.
  - Note there are no restrictions on the nature of the random variables X_i in the sum (can be continuous or discrete).
  - However, as n approaches infinity, $\mathbf{E}[W_n] = n\mu_X$ and $\mathbf{Var}[W_n] = n\mathbf{Var}[X]$ approach infinity, which makes it difficult to make a mathematical statement about the convergence of the CDF $F_{W_n}(w)$...

Central Limit Theorem

Therefore, our formal statement of the **central limit theorem** will be in terms of the standardized random variable instead of W_n. Thus:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu_X}{\sqrt{n\sigma_X^2}}$$

This results in:

$$\mathbf{E}[Z_n] = 0 \quad \text{and} \quad \mathbf{Var}[Z_n] = 1$$

Which has the property:

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z)$$

$$Z_n = \frac{\sum_{i=0}^{n} X_i - n\mu_X}{\sqrt{n\sigma_X^2}} = \frac{W_n - n\mu_X}{\sqrt{n\sigma_X^2}}$$

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z)$$

Therefore:

$$F_{W_n}(w) = P(\sqrt{n\sigma_X^2} Z_n + n\mu_X \leq w) = F_{Z_n}\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right)$$

$$\approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right)$$

Note that it will asymptotically approach $\Phi(z)$ for large n

Central Limit Theorem

Let $W_n = X_1 + \cdots + X_n$ be the sum of n iid random variables, each with $\mathbf{E}[X] = \mu_X$ and $\mathbf{Var}[X] = \sigma_X^2$. The central limit theorem approximation to the CDF of W_n is

$$F_{W_n}(w) \approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right)$$

We call this the Gaussian approximation of $F_{W_n}(w)$

Although the central limit theorem approximation provides a useful means of calculating events related to complicated probability models, it has to be used with caution. When the events of interest are confined to outcomes *at the edge of the range* of a random variable, the central limit theorem approximation can be quite inaccurate.

Example 9.8

A compact disc (CD) contains digitized samples of an acoustic waveform. In a CD player with a "one bit digital to analog converter," each digital sample is represented to an accuracy of ± 0.5 mV and is uniformly distributed in this range.

The CD player oversamples the waveform by making eight independent measurements corresponding to each sample. The CD player obtains a waveform sample by calculating the average (sample mean) of the eight measurements.

What is the probability that the error in the waveform sample is greater than 0.1 mV?

Chapter 10

The Sample Mean

The Sample Mean

- Up until now we have considered the properties of probability models.
- In referring to applications of probability theory, we have assumed prior knowledge of the *probability model* that governs the outcomes of an experiment.
- In practice, however, we encounter many situations in which the probability model is not known in advance and experimenters *collect data* in order to learn about the model.
- In doing so, they apply principles of **statistical inference**, a body of knowledge that governs the use of measurements to discover the properties of a probability model.

Sampling

- When we take a measurement of something, we **sample** it.
- For example measuring the voltage.
  - This sample differs from the *true value* due to effects such as voltage drift and measurement noise.
  - Thus the sample is an *estimate* of the true value.
- We can improve the estimate by taking more than one sample.
  - If we had a large number n identical circuits, we could take *simultaneous* measurements from each and estimate the true voltage by calculating the mean.
  - In practice we usually have just one circuit, and therefore assume that the circuit does not change with time, and then take n *sequential* samples and determine the mean.

Sampling

- The aforementioned setup implicitly assumes:
    1. That successive samples are *statistically independent*
    2. That each sample is *drawn from the same distribution* that we are trying to estimate.

The Sample Mean

We can consider sampling as n trials, producing n iid random variables $X_1, \ldots, X_n$ with PDF $f_X(x)$, the sample mean of X is the RV:

$$M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Since $M_n(X)$ is a function of random variables, it is a random variable itself.
- This stands separate from $\mathbf{E}[X]$ which is just a number and is sometimes called the *mean value* of random variable X, but due to possible confusion it is better to call it the *expected value*.
- However, as $n \to \infty$ then $M_n(X)$ will predictably approach $\mathbf{E}[X]$.

Note that in previous years $\hat{\bar{X}}_N$ was used instead of $M_n(X)$.

The Sample Mean

The sample mean $M_n(X)$ has expected value and variance

$$\mathbf{E}[M_n(X)] = \mathbf{E}[X]$$

$$\mathbf{Var}[M_n(X)] = \frac{\mathbf{Var}[X]}{n}$$

Therefore, the typical value of the sample mean (regardless of n) will be the expected value.

As $n \to \infty$ then the variance goes to 0, in other words as $n \to \infty$ then $M_n(X)$ converges to the expected value $\mathbf{E}[X]$.

The Sample Mean

In statistics, the **bias** of an estimator is the difference between this estimator's *expected value* and the *true value* of the parameter being estimated.

$$\mathbf{E}[M_n(X)] = \mathbf{E}[X]$$

Due to this definition we call the sample mean an **unbiased estimator**, since the bias is 0.

Deviation of a RV from the Expected Value

*The **Chebyshev inequality** is an upper bound on the probability $P(|X - \mu_X| > c)$.*

*We use the Chebyshev inequality to derive the **Laws of Large Number**s and the parameter-estimation techniques.*

*The **Chebyshev inequality** is derived from the **Markov inequality**, a looser upper bound.*

*The **Chernoff bound** is a more accurate inequality using the complete probability model of X.*

- The analysis of the convergence of $M_n(X)$ to $\mathbf{E}[X]$ begins with a study of the random variable $|X - \mu_X|$ (the absolute difference between a random variable X and its expected value).
- This study leads to the **Chebyshev inequality**, which states that the probability of a large deviation from the expected value is inversely proportional to the square of the deviation.
- The derivation of the Chebyshev inequality begins with the **Markov inequality**, an upper bound on the probability that a sample value of a nonnegative random variable exceeds the expected value by any arbitrary factor.
- The **Chernoff bound** is a third inequality used to estimate the probability that a random sample differs substantially from its expected value.
    - The Chernoff bound is more accurate than the Chebyshev and Markov inequalities because it takes into account more information about the probability model of X.

Markov inequality

The **Markov inequality** states that for a random variable X, such that $P(X < 0) = 0$ (taking on nonnegative values only), and a constant $c > 0$:

$$P(X \geq c) \leq \frac{\mathrm{E}[X]}{c}$$

This gives the *upper bound* of the probability of seeing a sample value greater than some value c.

And note that the distribution of X is not important.

Markov inequality

Example 10.1

Let X represent the height (in feet) of a storm surge following a hurricane. If the expected height is $\mathrm{E}[X] = 5.5$, then the Markov inequality states that an upper bound on the probability of a storm surge at least 11 feet high is

$$\mathrm{P}[X \geq 11] \leq 5.5/11 = 1/2. \tag{10.5}$$

This is clearly a very loose bound...

Chebyshev Inequality

The **Chebyshev Inequality** states that for an arbitrary random variable X and constant $c > 0$

$$P(|X - \mu_X| \geq c) \leq \frac{\mathrm{Var}[X]}{c^2}$$

The Chebyshev inequality states that the difference between X and $\mathbf{E}[X]$ is somehow limited by $\mathbf{Var}[X]$.

It provides the upper bound of the *probability* that the value of X lies outside some margin c about the expected value.

The Chebyshev inequality generally provides a tighter bound than the Markov inequality. In particular, when the variance of X is very small.

Chebyshev Inequality

Example 10.3

If the height X of a storm surge following a hurricane has expected value $\mathrm{E}[X] = 5.5$ feet and standard deviation $\sigma_X = 1$ foot, use the Chebyshev inequality to to find an upper bound on $\mathrm{P}[X \geq 11]$.

Since a height X is nonnegative, the probability that $X \geq 11$ can be written as

$$\mathrm{P}\left[X \geq 11\right] = \mathrm{P}\left[X - \mu_X \geq 11 - \mu_X\right] = \mathrm{P}\left[|X - \mu_X| \geq 5.5\right]. \qquad (10.8)$$

Now we use the Chebyshev inequality to obtain

$$\mathrm{P}\left[X \geq 11\right] = \mathrm{P}\left[|X - \mu_X| \geq 5.5\right] \leq \mathrm{Var}[X]/(5.5)^2 = 0.033 \approx 1/30. \qquad (10.9)$$

Although this bound is better than the Markov bound, it is also loose. $\mathrm{P}[X \geq 11]$ is seven orders of magnitude lower than $1/30$.

Chernoff Bound

The **Chernoff Bound** states that for an arbitrary random variable X and constant c

$$P(X \geq c) \leq \min_{s\geq 0} e^{-sc}\phi_X(s)$$

With $\phi_X(s)$ the moment generating function.

The Chernoff bound can be applied to any random variable. However, for small values of c, $e^{-sc}\phi_X(s)$ will be minimized by a negative value of s. In this case, the minimizing nonnegative s is $s = 0$, and the Chernoff bound gives the trivial answer $P(X \geq c) \leq 1$.

Chernoff Bound

Example 10.4

If the probability model of the height X, measured in feet, of a storm surge following a hurricane at a certain location is the Gaussian (5.5, 1) random variable, use the Chernoff bound to find an upper bound on $\mathrm{P}[X \geq 11]$.

In Table 9.1 the MGF of X is

$$\phi_X(s) = e^{(11s+s^2)/2}. \tag{10.12}$$

Thus the Chernoff bound is

$$\mathrm{P}[X \geq 11] \leq \min_{s \geq 0} e^{-11s} e^{(11s+s^2)/2} = \min_{s \geq 0} e^{(s^2-11s)/2}. \tag{10.13}$$

To find the minimizing s, it is sufficient to choose s to minimize $h(s) = s^2 - 11s$. Setting the derivative $dh(s)/ds = 2s - 11 = 0$ yields $s = 5.5$. Applying $s = 5.5$ to the bound yields

$$\mathrm{P}[X \geq 11] \leq e^{(s^2-11s)/2}\Big|_{s=5.5} = e^{-(5.5)^2/2} = 2.7 \times 10^{-7}. \tag{10.14}$$

The true probability is $1 - \Phi(5.5) = 1.9 \times 10^{-8}$

Comparison

- Markov Inequality only uses the expected value.
- The Chebyshev inequality uses the expected value and the variance.
- But the much more accurate Chernoff bound is based on knowledge of the complete probability model expressed by the MGF.

We will apply this in the next lecture to discuss the convergence of $M_n(X)$ to $\mathbf{E}[X]$