Announcements

Exam 1 on Friday 2/28 during lecture (50 min)

- Format: mostly short answer w/ calculations and a few multiple choice and/or fill-in-the blank questions
- Covers up to and including section 4.8
- Review class on Wednesday
- Practice exam available on Learn@UW
- Bring formula sheet – double-sided 8.5”x11” paper; **hand-written** notes of definitions and formulas (no photocopies)
- Standard normal table (or portion thereof) will be provided
- Bring a (scientific or graphing) calculator to the exam
- No homework due next Friday 2/28 (exam day)

Central Limit Theorem

Keegan Korthauer
Department of Statistics
UW Madison

CENTRAL LIMIT THEOREM

Central limit theorem

Normal approximation to binomial

Normal approximation to Poisson

Distribution of the Mean of a Normal RV

Recall that for $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

for any sample size n

Normal Distribution and the CLT

- Why is the normal distribution so important?
  - The **Central Limit Theorem (CLT)** allows us to apply the normal distribution to the sample mean in certain situations where we do not know the population distribution

- Simply stated, the CLT says that the **mean** of a large simple random sample is approximately normally distributed - ***even if the population distribution is not normal!***

- This lets us compute probabilities with the normal table when we have no idea about the underlying distribution – as long as our sample size is big

Central Limit Theorem

- Let $X_1,\dots,X_n$ be a simple random sample from a population with mean μ and variance σ^2

- Let $\overline{X} = (X_1+\dots+X_n)/n$ be the sample mean

- Let $S_n = X_1+\dots+X_n$ be the sum of the sample observations

- Then if n is sufficiently large,

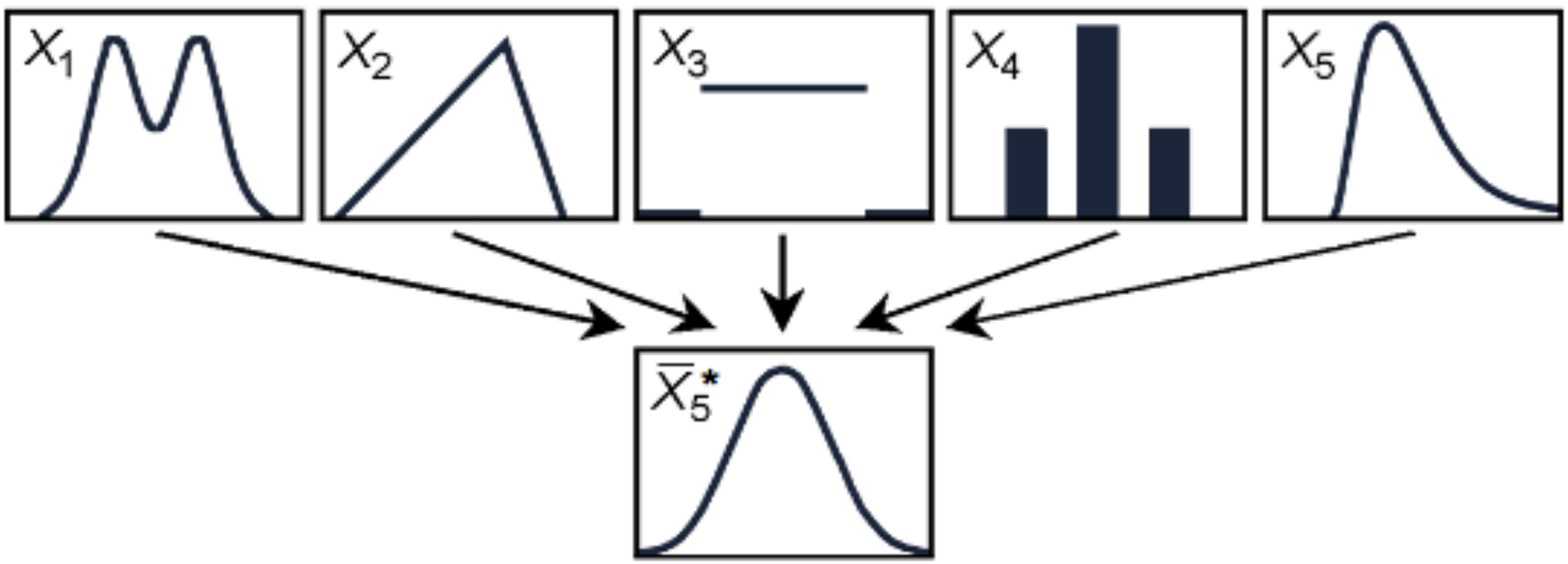
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{approximately}$$

and

$$S_n \sim N\left(n\mu, n\sigma^2\right) \quad \text{approximately}$$

Starting Distribution Doesn't Matter

Even if we start with a discrete or skewed or bimodal population distribution, the Central Limit Theorem still applies



http://value-at-risk.net/central-limit-theorem/

Rule of Thumb

- What does “sufficiently large” mean?
- This can depend on the shape of the underlying population distribution
- Approximation gets better as we increase the sample size
- Generally a sample size of **at least 30** works well enough

Example – 4.70

Let X be the number of flaws in a 1 inch length of copper wire. The PMF of X is:

x	$P(X = x)$
0	0.48
1	0.39
2	0.12
3	0.01

We sample 100 wires from this population. What is the probability that the average number of flaws per wire in this sample is less than 0.5?

Combining CLT and Linear Combinations

- Recall that in section 4.5 we learned that linear combinations of independent normal RVs are normal

- Combine that result with the CLT and we can find probabilities of linear combinations of sample means and sample sums

Example – Commute Times

Recall our commute time example:

- Let X_1 represent the time it takes (in minutes) to walk from my house to the bus stop. Assume $E(X_1)=3$, $Var(X_1)=1$.
- Let X_2 represent the time it takes the bus to travel between the bus stop and campus. Assume $E(X_2)=8$, $Var(X_2)=4$.
- X_1 and X_2 are independent

Say I take a random sample of 50 days and measure the commute times. What is the probability that the average total commute time will be greater than 11.5 minutes?

$P(\overline{Y} > 11.5) = 0.0571$

Normal Approximation to Binomial

- Recall that if X ~ Bin(n, p), then we can write X as a sum of independent and identically distributed RVs from a Bernoulli(p) population:

$$X = Y_1 + \ldots + Y_n$$

where $Y_1, \ldots, Y_n$ ~ Bern(p) (with mean p and variance p(1-p))

- Also note that $\hat{p} = \frac{X}{n} = \frac{Y_1 + \ldots + Y_n}{n} = \bar{Y}$

- Then by the CLT if n is large enough,

$$\hat{p} \sim N(p,\ p(1-p)/n) \text{ and } X \sim N(np,\ np(1-p))$$

(approximately)

Normal Approximation to Binomial

- In the case of the binomial, the accuracy of the CLT approximation depends on p and n

- Need large enough number of successes **and** failures (large enough np **and** $n(1-p)$)

- Rules of thumb:

$$np > 10 \textbf{ and } n(1-p) > 10$$

Normal Approximation to Binomial

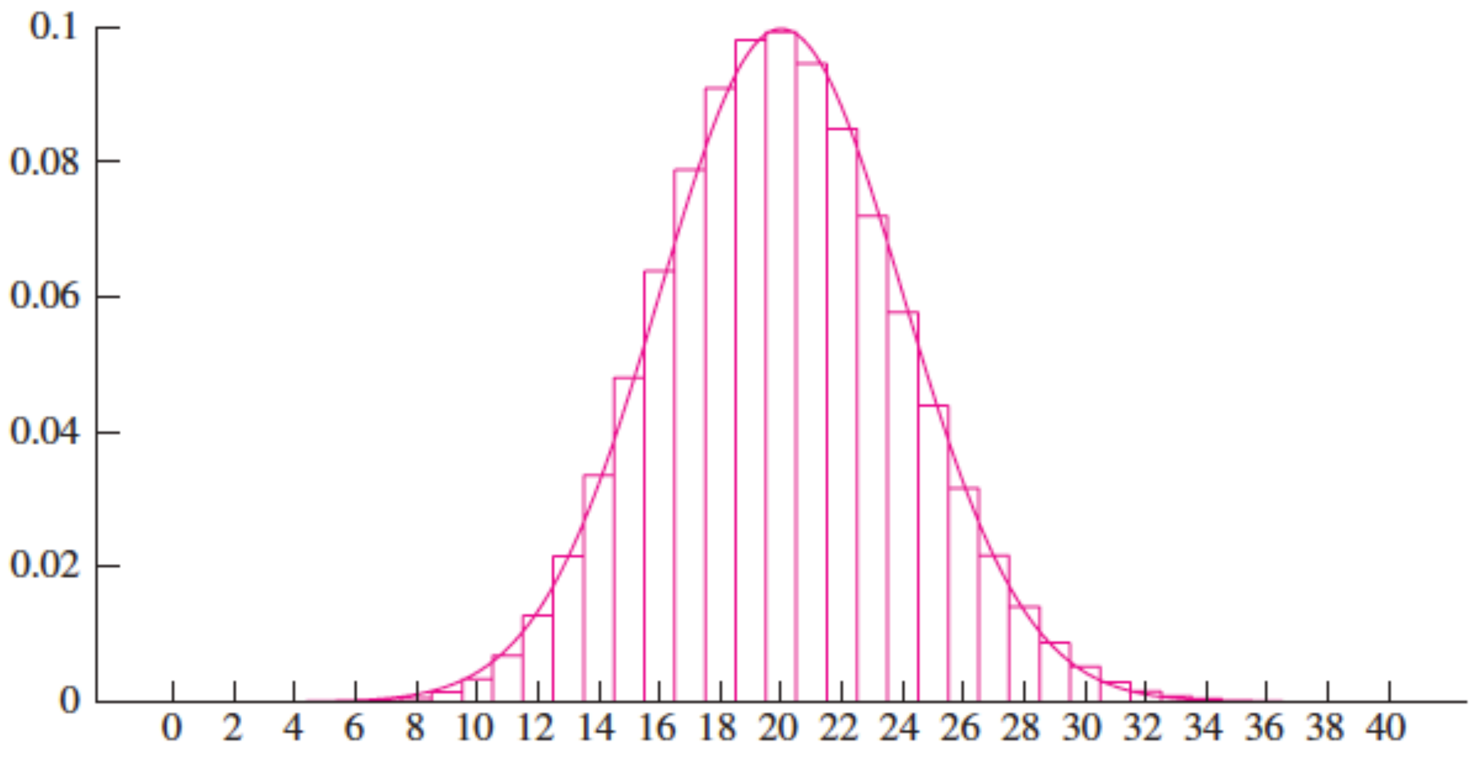


FIGURE 4.27 The Bin(100, 0.2) probability histogram, with the $N(20, 16)$ probability density function superimposed.

Continuity Correction

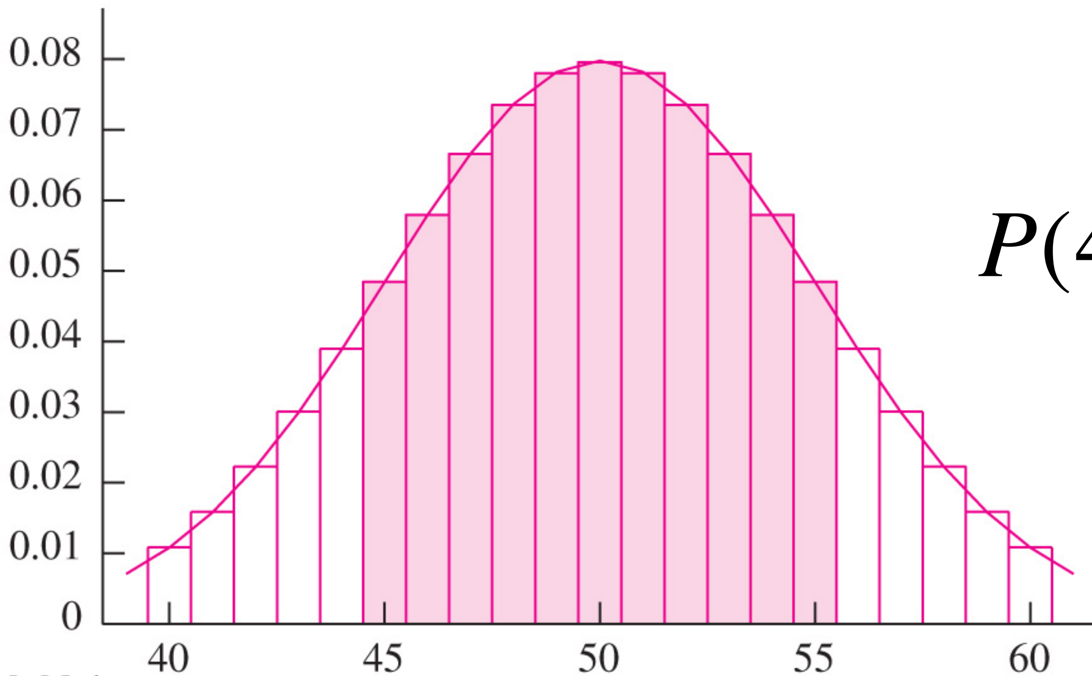
- Recall that for continuous random variables

$$P(a \le X \le b) = P(a < X < b)$$

- But this is **not** true for discrete random variables
- When approximating a discrete RV with the continuous normal distribution we have to worry about what to do with the endpoints
- We apply a **continuity correction** to improve the accuracy*

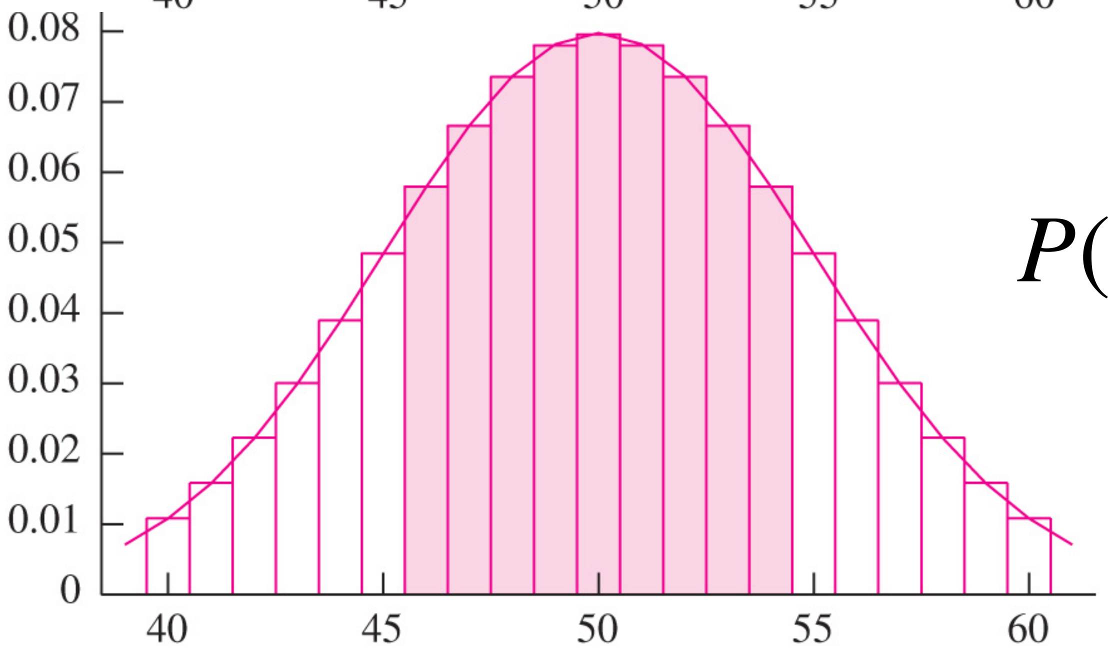
Example - Continuity Correction

$X \sim Bin(100, 0.5)$



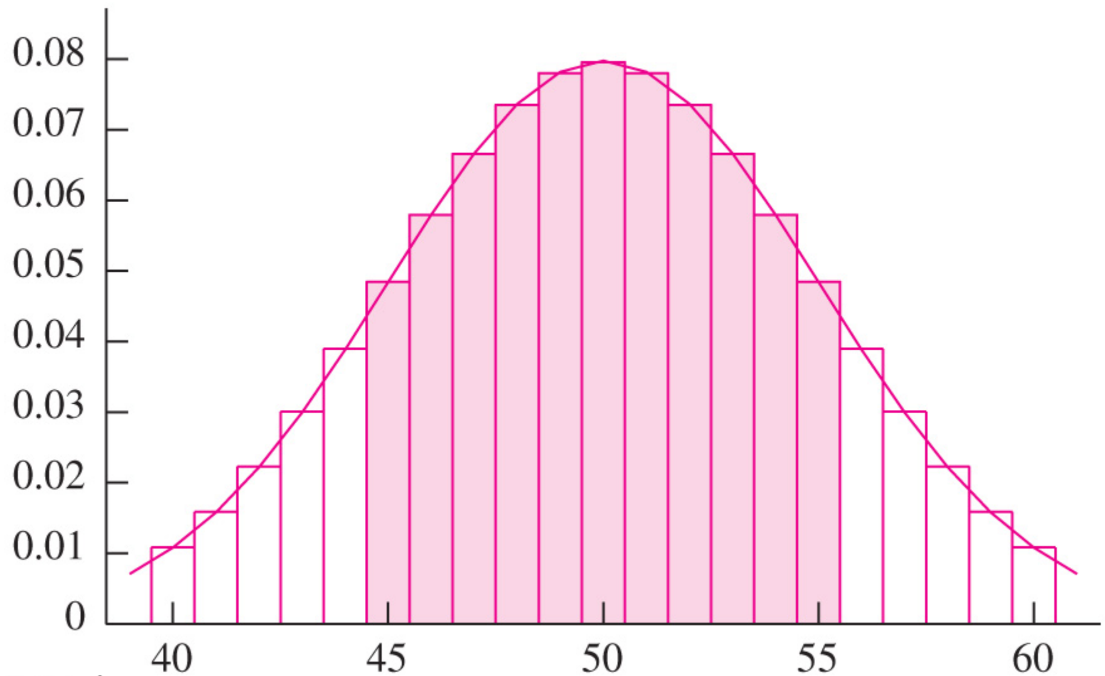
$P(45 \le X \le 55)$

$X \sim N(50, 25)$



$P(45 < X < 55)$

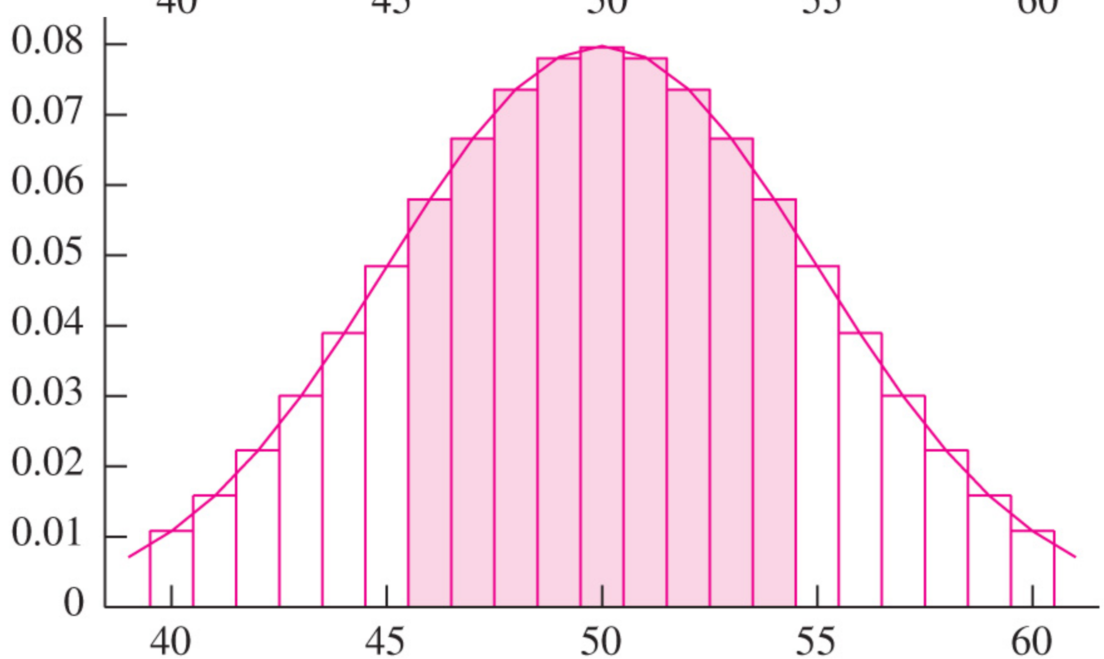
Solution



If we want to approximate

$$P(45 \leq X \leq 55)$$

we should integrate the approximated normal curve from 44.5 to 55.5



If we want to approximate

$$P(45 < X < 55)$$

we should integrate the approximated normal curve from 45.5 to 54.5

Example – Binomial Approximation

- A manufactured component meets its specifications 78% of the time.
- In a random sample of 500 components, what is the probability that at least 400 meet the specifications?

Let X be the number of components meeting specifications. Then $X \sim Bin(500, 0.78)$. Since np and n(1-p) > 10, we can use the normal approximation: $X \sim N(np, np(1-p)) = N(390, 85.8)$.

We want $P(X \geq 400)$ which **includes** the endpoint 400, so we want to calculate

$$P(X \geq 399.5) = 1 - P(X < 399.5) = 1 - P(Z < (399.5-390)/\text{sqrt}(85.8))$$
$$= 1 - P(Z < 1.026) = 1 - 0.847 = 0.153$$

Normal Approximation to Poisson

- Recall the connection between Poisson and Binomial
  - we can approximate Poisson with Binomial when n is large and p is small where $\lambda = np$
- Also recall that the mean and variance of a Poisson RV are both λ
- Then if λ is sufficiently large ($\lambda > 10$) we can approximate $X \sim \text{Poisson}(\lambda)$ with a binomial (and $np > 10$)
- Under these conditions, Poisson is approximately binomial and binomial is approximately normal, so **Poisson is approximately normal** as well!

Normal Approximation to Poisson

Formally, if $X \sim \text{Poisson}(\lambda)$ where $\lambda > 10$, then

$$X \sim N(\lambda, \lambda) \quad \text{approximately}$$

The same continuity issue applies, but a standard correction can make tail areas less accurate, so we will not worry about a continuity correction with the Poisson

Example – 4.76

- The number of hits on a website follows a Poisson distribution with mean 27 hits per hour.

- Find the probability that there will be 90 or more hits in three hours.

$$P(X \geq 90) = 0.1587$$

Next

- Intro to R
- Exam 1 on Friday 2/28 during lecture (50 min)
  - Format: mostly short answer w/ calculations and a few multiple choice and/or fill-in-the blank questions
  - Review class on Wednesday
  - Practice exam available on Learn@UW
  - Formula sheet – double-sided 8.5”x11” paper; hand-written notes of definitions and formulas (no photocopies)
  - Standard normal table (or portion thereof) will be provided
  - Bring a (scientific or graphing) calculator to the exam
  - No homework due next Friday 2/28 (exam day)