	mean	SD.	
Population	μ	σ	unknown fixed
Sample	$\bar{x}$	s	Sampling Variability.

Chapter 7 Normal Probability Approximations

→ **Population:** contains the entire collection of individuals we want to study

→ **Parameter:** characteristic of interest from the population. Value of the parameter is unknown in practice μ, σ

Sample: subset of individuals selected from the population. A sample can provide reliable information about a population if the sample is representative of the population.

Statistic: numerical measure of the sample. We use statistics to estimate the unknown population parameter. Due to sampling variability a statistic takes on different values for different samples.

The **sampling distribution** of a statistic is the distribution of the values of the statistic computed from all possible random samples of a certain size n drawn from the population.

$\bar{x}$
$\vdots$

Ch. 5

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

$n = 100.$

There are many other possible samples of size 5 that the instructor could have drawn. With each random sample he would get

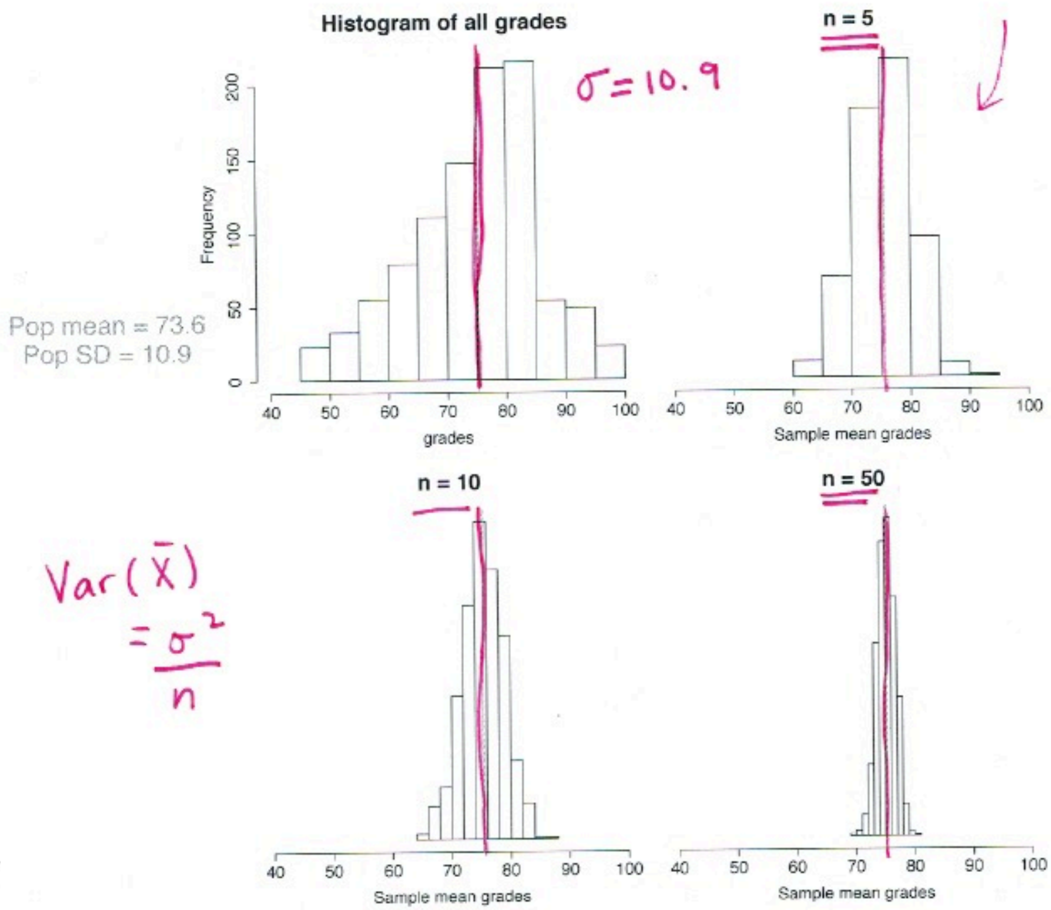
An instructor of large statistics course with 1000 students is interested in finding out the strength of his students' math skills. He thinks their high school calculus grades will be a good indicator of math ability. The instructor drew one random sample of size 5. There are many other possible samples of size 5 that the instructor could have drawn. With each random sample he would get different values of the sample mean. Imagine we repeat the sampling procedure many more times:

Sample	Grades of sampled students	Sample mean
1	74 77 69 69 83	75.4
2	63 83 69 72 80	73.4
3	65 84 69 87 50	71.0
4	72 70 100 76 47	73.0
5	82 65 79 66 84	75.2
6	77 66 78 68 74	72.6
⋮		⋮

Suppose a random sample of n observations is taken from a population with mean μ and standard deviation σ.

- The mean of the sampling distribution of means is equal to μ.
- The standard deviation of the sampling distribution of means is equal to $\frac{\sigma}{\sqrt{n}}$

$E(\bar{X}) = \mu$



$\sigma = 10.9$

$Var(\bar{X}) = \frac{\sigma^2}{n}$

Central Limit Theorem

Let $X_1, X_2, ..., X_n$ be an independent random sample of size n taken from any distribution with mean μ and variance σ^2.
If n is large (book says $n \geq 20$), then

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n} \overset{approx}{\sim} N(\mu, \frac{\sigma^2}{n})$$

Last Class:

Let $X_1, \dots, X_n$ be a sample from an arbitrary population with mean $\underline{\underline{\mu}}$ and Variance $\underline{\underline{\sigma^2}}$.

sequence of independent, identically dist. r.v.

CLT states, when n is large

$$\bar{X} \sim N\left(E(\bar{X}) = \underline{\underline{\mu}},\ Var(\bar{X}) = \frac{\sigma^2}{n} \right) \text{ approx.}$$

$$E\left(\frac{X_1 + \dots + X_n}{n}\right) = \mu \qquad Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

Note that if a question ask about a sum instead of an average, you can still use CLT. Let T be the sum of an independent random sample $X_1, X_2, ..., X_n$.

$$T = X_1 + X_2 + ... + X_n$$

$$T = \sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \implies T = n\bar{X}$$

$$\begin{aligned} E(T) &= E(n(\bar{X})) \\ &= nE((\bar{X})) = n\mu \\ Var(T) &= Var(n\bar{X}) \\ &= n^2 Var(\bar{X}) \\ &= n^2\sigma^2/n \\ &= n\sigma^2 \\ T &\sim N(n\mu, n\sigma^2) \end{aligned}$$

Example 1

Example 7.1 from your text.

A system consists of 25 independent parts connected in such a way that the ith part automatically turnson when the $(i-1)th$ part burns out. The expected lifetime of each part is 10 weeks and the standard deviation is equal to 4 weeks.

(a) Calculate the expected lifetime and standard deviation for the the system.

(b) Calculate the probability that the system will last more than its expected life.

(c) Calculate the probability that the system will last more than 1.1 times its expected life.

(d) What are the (approximate) median life and interquartile range for the system?

X_i denotes the lifetime of the ith part.

a) $E(X) = 10$ weeks $\qquad SD(X) = 4$ weeks.

Let $T = X_1 + X_2 + \ldots + X_{25}$ lifetime of entire system.

$$E(T) = E(X_1 + \ldots + X_{25}) = E(X_1) + E(X_2) + \ldots + E(X_{25})$$
$$= 25\,E(X) = 25 \times 10 = 250$$

$$Var(T) = Var(X_1 + \ldots + X_{25}) = Var(X_1) + \ldots + Var(X_{25})$$
$$= 25 \times 4^2 = 400$$
$$SD(T) = \sqrt{400} = 20$$

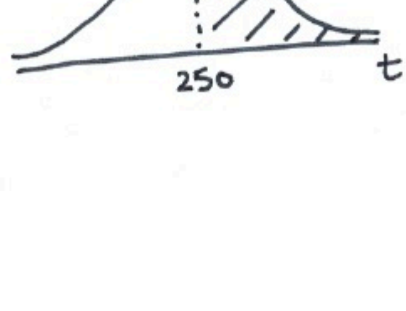
b) $P(T > \underline{E(T)}) = P(T > \underline{\underline{250}})$

$\uparrow$

$\cancel{X} \sim ?$ $\qquad$ CLT says $\bar{X} \sim N\left(\mu = 10, \frac{\sigma^2}{n} = \frac{16}{\underline{\underline{25}}}\right)$

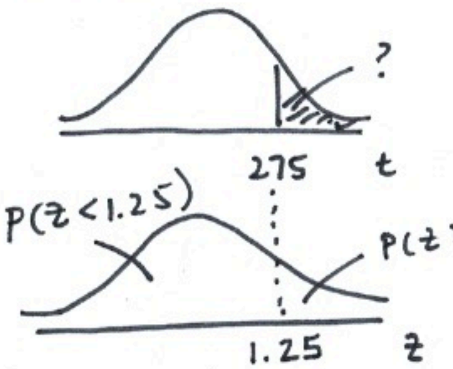
Alternatively $\underline{\underline{T}} \overset{\text{approx.}}{\sim} N(n\mu = 250, n\sigma^2 = 400)$

$P(T > 250) = \frac{1}{2}$



c) $P(T > 1.1 \times 250) = ?$

$T \sim N(250, 400)$



$$P(T > 275) = P\left(z > \frac{275-250}{\sqrt{400}}\right)$$

$$= P(z > 1.25)$$

$$= 1 - P(z < 1.25)$$

$$= 1 - 0.8944$$

$$= 0.1056$$

d) $T \sim N(250, 400)$

median $= 250$

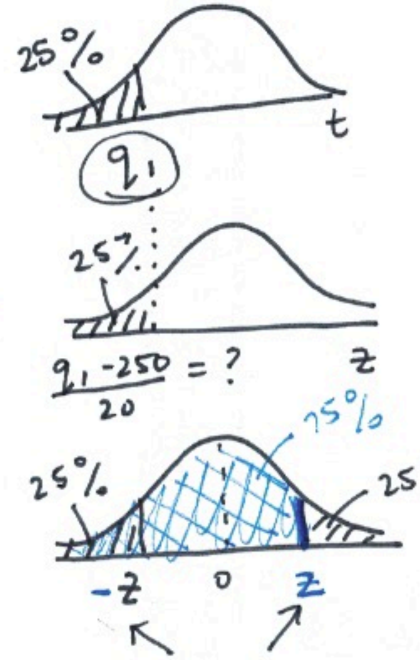
$IQR = q_3 - q_1$

Let q_1 lower quartile

Let q_3 upper " "

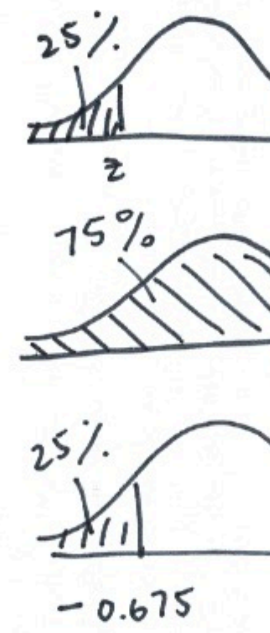
$P(T < q_1) = 0.25$

$$P\left(Z < \frac{q_1 - 250}{20}\right) = 0.25$$



$$\frac{q_1 - 250}{20} = -0.675$$

$q_1 = 236.5$



$P(T < q_3) = 0.75$

$$P\left(Z < \frac{q_3 - 250}{20}\right) = P(Z < 0.675)$$

$$\frac{q_3 - 250}{20} = 0.675$$

$q_3 = 263.50$

$IQR = q_3 - q_1 = 263.50 -$

$= 27$

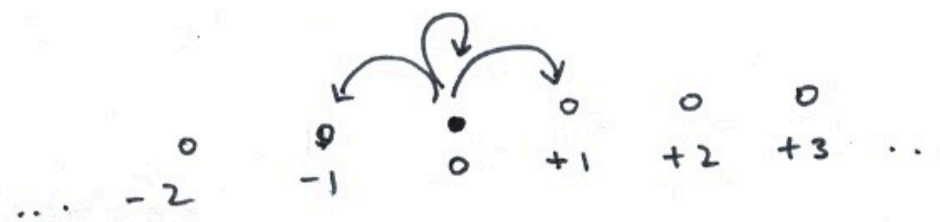
Example 2

In physics, random walks are used to model the process of dif or random motion, such as the random movement of molecul liquids and gases. The position, S_n of a particle at time n ca thought of as a sum of displacements, $X_1, X_2, ..., X_n$. Assum the displacements are independent and identically distributed, can use the CLT to solve questions on random walks. Suppos

each step, a particle moving on sites labelled by integers is eq likely to move one step to the right, or one step to the left, o where it is. Find approximately the probability that after 10,0 steps, the particle ends up:

(a) more than 50 sites to the right of starting point?

(b) more than 50 sites to the left of starting point?



$\underline{X_1, X_2, \ldots, X_n}$ $\qquad$ $\underline{\underline{S_n}} = \underline{X_1} + \ldots + \underline{X_{10000}}$

Let X_i single step takes values:

-1 moves left

0 stays put

$+1$ moves right.

	-1	0	$+1$
$P(X=x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$E(X) = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$$

$$Var(x) = E(X^2) - E(X)^2$$

$$= \left[(-1)^2 \frac{1}{3} + 0^2 \left(\frac{1}{3}\right) + 1^2 \frac{1}{3} \right] - 0^2 = \frac{2}{3}.$$

a) $P(S > 50)$ $\qquad S \overset{approx}{\sim} N\left(0, \frac{20000}{3}\right)$

method ①

$$S = X_1 + \dots + X_{10000}$$

$$E(S) = E(X_1 + \dots + X_{10000}) = 10000\, E(X) = 0$$

$$Var(S) = Var(X_1 + \dots + X_{10000}) = 10000\, Var(X) \quad \text{indep.}$$

$$= 10000 \times \frac{2}{3} = \frac{20000}{3}$$

method ②

$S = X_1 + \dots + X_{10000}$ $\qquad \bar{X} = \frac{X_1 + \dots + X_{10000}}{n}$

Alternate:

$$P(S > 50) = P\left(\frac{S}{n} > \frac{50}{10000}\right) = P\left(\bar{X} > \frac{50}{10000}\right)$$

① $P(S>50) = P\left(Z > \dfrac{50-0}{\sqrt{\frac{20000}{3}}}\right) = P(Z > 0.61)$

$= 1 - P(Z < 0.61)$

$= 1 - 0.7291$ ← standard norm table

$= 0.2709$

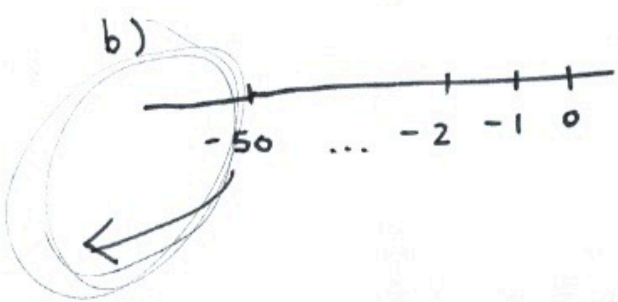
② $P\left(\bar{X} > \dfrac{50}{10000}\right)$

by CLT, $n > 20$ large

$\bar{X} \sim N\left(0, Var(\bar{x}) = \dfrac{\sigma^2}{n} = \dfrac{2/3}{10000}\right.$

$P(\bar{X} > 0.005) = P\left(Z > \dfrac{0.005-0}{\sqrt{2/30000}}\right) = P(Z > 0.61)$

b)

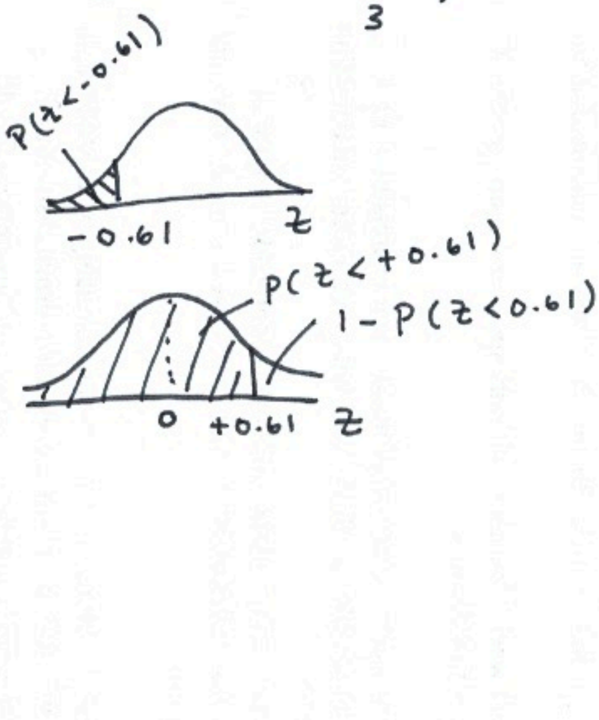


$$P(S<-50) = P\left(Z < \frac{-50-0}{\sqrt{\frac{20000}{3}}}\right) = P($$

$$S \sim N\left(0, \frac{20000}{3}\right)$$

$$= 1 - P($$

$$= 0.27$$



Example 3

Suppose that $\frac{1}{3}$ of computer chips manufactured by a certain company are defective. Suppose we randomly inspect $n = 36$ chips. What is the probability that in such a sample more tha chips will be defective?

$$X \sim Bin(n = 36, p = \frac{1}{3})$$

Find $P(X > 13)$.

Clearly this would take a long time...

$$\begin{aligned} P(X > 13) &= 1 - P(X \leq 13) \\ &= 1 - [P(X = 0) + P(X = 1) + ... + P(X = 13 \\ &= 1 - [\binom{36}{0}(1/3)^0(2/3)^{36} + ... + \binom{36}{13}(1/3)^13(\\ &= 0.2933 \end{aligned}$$

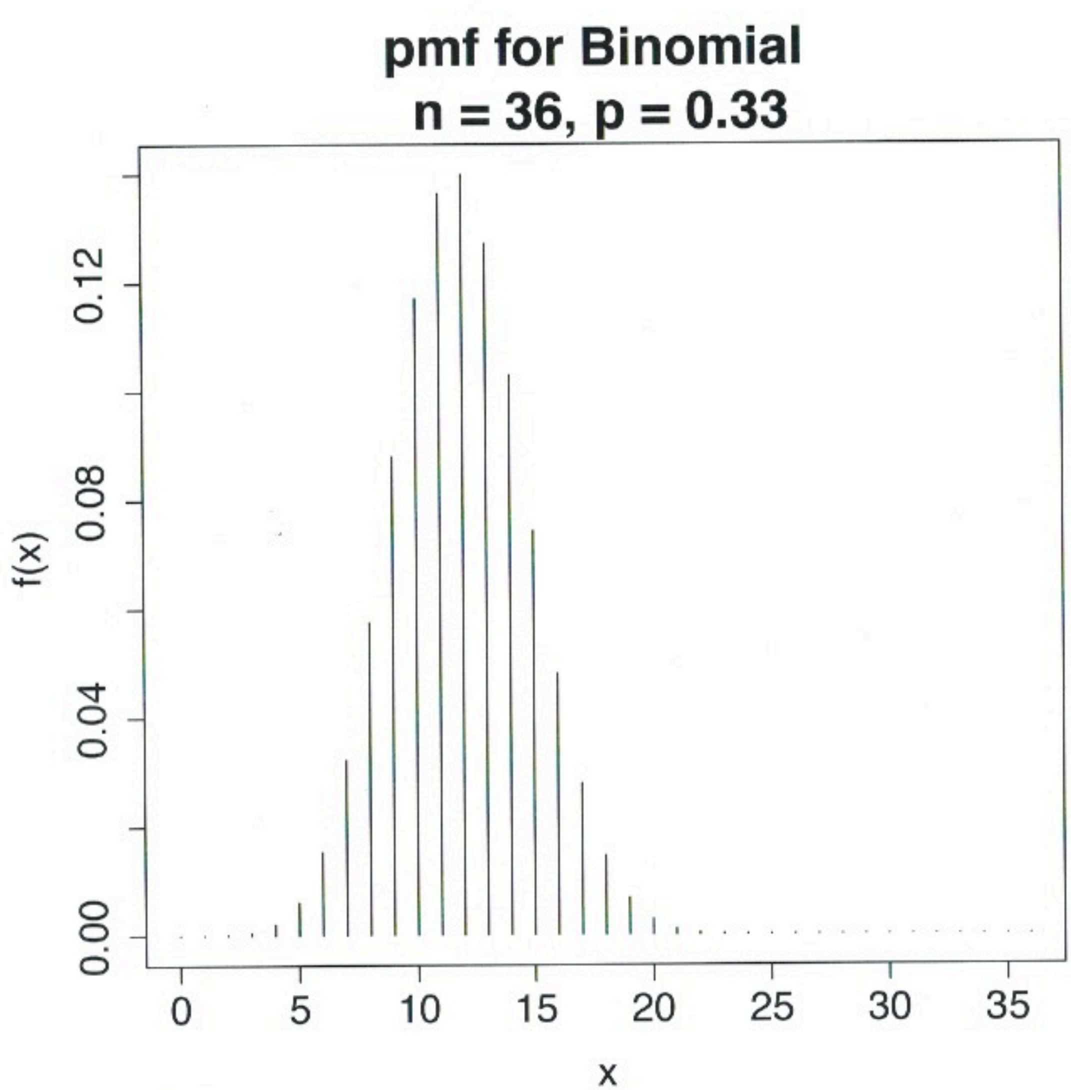
Normal Approximation to the Binomial

$q = 1 - p$

If $X \sim Bin(n, p)$ and if n is large so that $\underline{np \geq 5}$, $\underline{nq \geq 5}$, we use the normal distribution to get an approximate answer

$$X \overset{approx}{\sim} N(np, np(1-p))$$

Idea: When n is large and $np \geq 5$ and $n(1-p) \geq 5$, the shap
the binomial distribution is approximately symmetrical.



$X \sim Bin(n = 36, p = \frac{1}{3})$

Check:

$np = 36 \times \frac{1}{3} = 12$ ≥ 5 use the normal approx to binom.

$n(1-p) = 24$

$E(X) = np = 12$

$Var(X) = np(1-p) = 36 \times \frac{1}{3} \times \frac{2}{3} = 8$

$X \sim N(12, 8)$

$P(X > 13) = 1 - P(X \leq 13)$

$= 1 - P\left(Z \leq \frac{13 - 12}{\sqrt{8}}\right) = 1 - P(Z \leq 0.35$

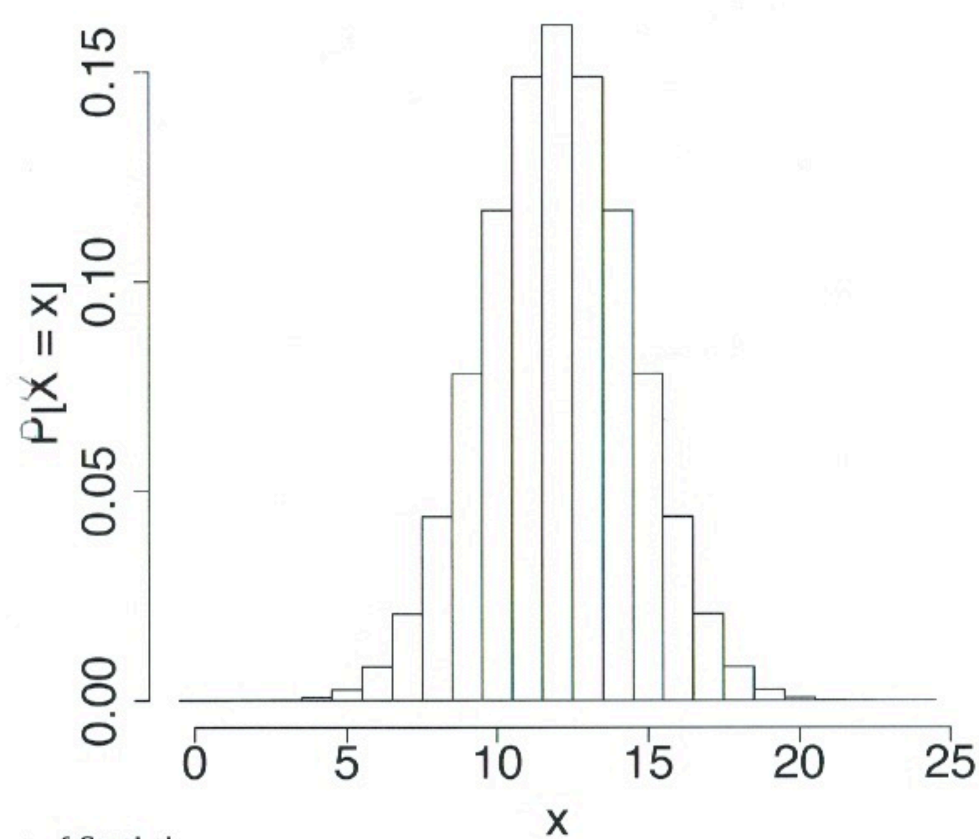
$= 0.362 \leftarrow$ close but

Continuity Correction

discrete → $\underline{\text{continuous}}$.

$P(X = a) =$

When using the normal model to approximate Binomial or Po
distributions, we can make more accurate approximations if w
a continuity correction.



Continuity correction:

$$P(X = x) = P(x - 0.5 \leq X \leq x + 0.5)$$

$$P(X > x) = P(X \geq x + 0.5)$$

$$P(X \geq x) = P(X \geq x - 0.5)$$

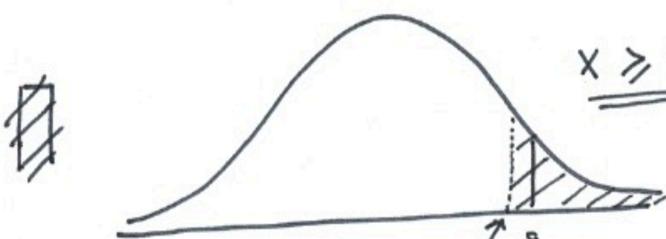
$$P(X < x) = P(X \leq x - 0.5)$$

$$P(X \leq x) = P(X \leq x + 0.5)$$

$P(X=10) = P(9.5 \le X \le 10.5)$

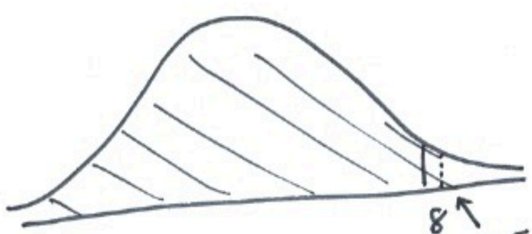
continuous.

$P(X \geq 8) \quad P(X > 8)$



$X \geq 8$

$P(X \geq 7.5) = P(X \geq 8)$



$P(X \le 8) = P(X \le 8.5)$

Let's continue example 3, using the Normal approximation.
Suppose that $\frac{1}{3}$ of computer chips manufactured by a certain
company are 3 defective. Suppose we randomly inspect $n = 3$
chips. What is the probability that in such a sample more tha
chips will be defective?

use Normal approx to binomial

$X \sim Bin(36, p = \frac{1}{3})$

$E(X) = 12$

$Var(X) = 8$

$P(X > 13) = P(X \geq 13.5)$ ap co

$= P\left(Z \geq \frac{13.5 - 12}{\sqrt{8}}\right)$

$= 1 - P(Z \leq 0.53)$

$= 1 - 0.7019$ ← table

$= 0.298$ ← closer.

Normal Approximation to the Poisson

If $X \sim pois(\lambda t)$ and λ is large $(\lambda t \geq 20)$, we can use the nor
distribution to approximate the Poisson distribution.

$$X \sim N(\lambda t, \lambda t)$$

Example 4

A radioactive element disintegrates such that it follows a Pois
distribution. If the mean number of particles emitted is recor
a 1 second interval as 25, what is the probability that

(a) Exactly 27 particles are emitted in 1 second?

(b) Between 24 and 27 particles inclusive are emitted in 1 se

$P(X=27)$ X = # particles emitted in 1 sec.

$X \sim Pois(25)$

Actual calculation:

$$P(X=27) = \frac{e^{-25}\, 25^{27}}{27!} = 0.0708$$

Use Normal approx:

$\lambda t = 25 \geq 20$ ✓ $\quad X \overset{\text{approx}}{\sim} N(25, 25)$

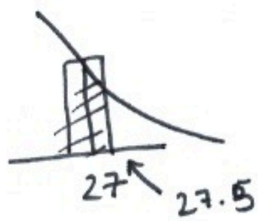
$$P(X = 27) = P(26.5 \leq X \leq 27.5) \quad \text{continuity correction}$$

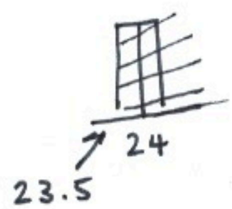
$$= P\left(\frac{26.5 - 25}{5} \leq Z \leq \frac{27.5 - 25}{5}\right)$$

$$= P(Z \leq 0.5) - P(Z \leq 0.3)$$

$$= 0.6915 - 0.6179 = 0.0736$$

b) $P(24 \le X \le 27) = P(23.5 \le X \le 27.5)$ continuity





$= P\left(\frac{23.5-25}{5} \le z \le \frac{27.5-2}{5}\right.$

$= P(-0.3 \le z \le 0.5)$

$= P(z \le 0.5) - P(z \le -0.3)$

$= P(z \le 0.5) - (1 - P(z \le 0.3$

$= 0.6915 - (1 - 0.6179)$

$= 0.3094$