24) Suppose that $X_1, \dots, X_n$ are iid binomial$(1, p)$.

Then $Y = \sum_{i=1}^{n} X_i$ is binomial(n, p)

$E(X_i) = 1(p) = p$, $V(X_i) = (1)p(1-p) = p(1-p)$.

By the central limit theorem (with $\bar{X} = \frac{1}{n}Y$, $E\bar{X} = p$, $V(\bar{X}) = \frac{\sigma^2}{n} = \frac{p(1-p)}{n}$)

$$\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\frac{1}{n}Y - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1) \text{ for large } n.$$

25) p379 If Y is binomial(n, p),

then $Y \approx$ normal with mean $\mu_Y = np$ and variance $\sigma_Y^2 = np(1-p)$.

26) p380 Use this approximation if

$$n > 9 \frac{(\text{larger of } p \text{ and } 1-p)}{(\text{smaller of } p \text{ and } 1-p)} = \frac{9 \max(p, 1-p)}{\min(p, 1-p)}$$

Know equivalently $n > \frac{9p}{1-p}$ and $n > \frac{9(1-p)}{p}$.

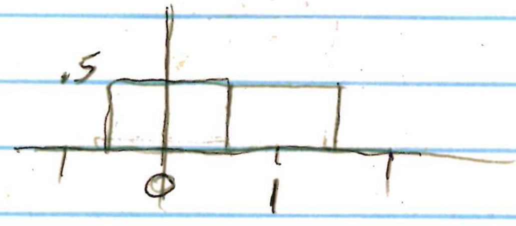
Otherwise, do not use the normal approximation to the binomial.

ex) If $p = .5$, need $n > 9(1) = 9$

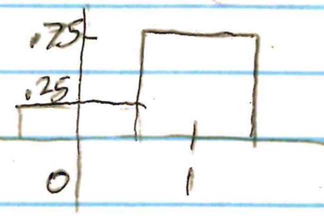
If $p = .25$, want $n > 9\left(\frac{.25}{.75}\right)$ and $n > 9\left(\frac{.75}{.25}\right) = 27$

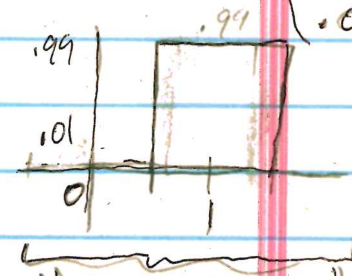
48.5

If $p = .01$, want $n > 9\left(\frac{.01}{.99}\right)$ and $n > 9\left(\frac{.99}{.01}\right) = 891$



symmetric

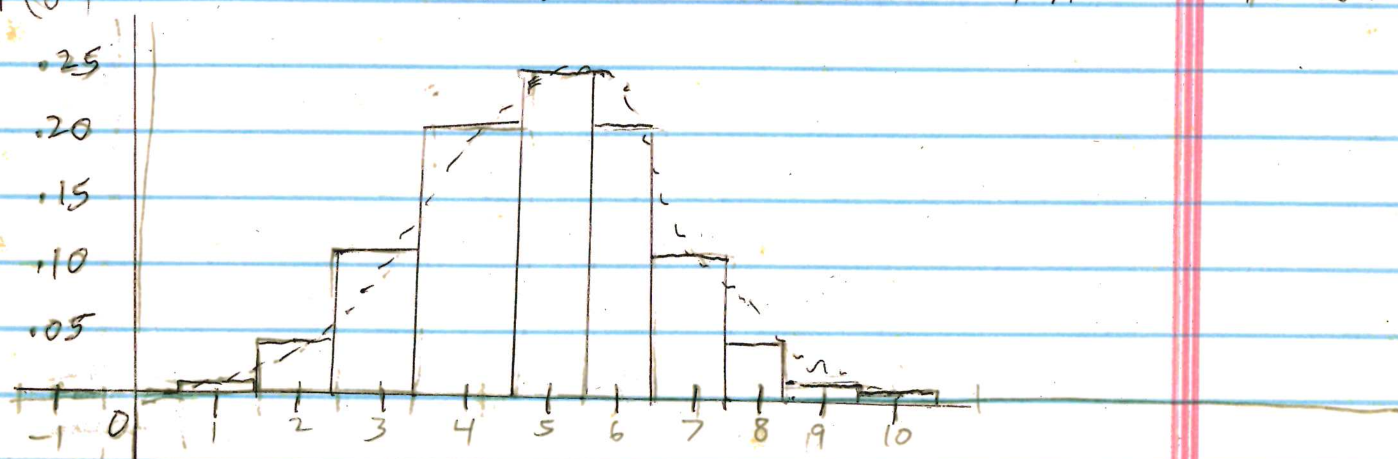




"left skewed"

ex] $Y \sim bin(10, \frac{1}{2})$

y	0	1	2	3	4	5	6	7	8	9	10
$p(y)$	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010



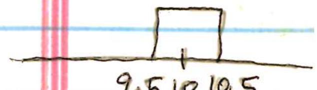
p380

27] ✱ Correction for continuity

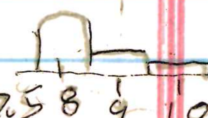
When you use the normal approx for the binomial, you are approximating the histogram.

Let Y be $bin(n, p)$, let y be an integer $\in [0, n]$, and let X be normal $\mu = np$ $\sigma = \sqrt{np(1-p)}$

i) $P(Y = y) \approx P\left[y - \frac{1}{2} \le X \le y + \frac{1}{2}\right]$,



ii) $P(Y \ge y) \approx P(X \ge y - .5)$,



iii) $P(Y \le y) \approx P(X \le y + .5)$,

28] Common E4 problem Use the normal approx for the binomial. 483 49

Step *] check that $n > 9\frac{p}{1-p}$ and $n > 9\frac{1-p}{p}$.

Step 0] Find $\mu_X = np$, $\sigma_X = \sqrt{np(1-p)}$

Step i] Make the continuity correction and draw X picture

Step ii] z scores

Step iii] z picture

Step iv] table

ex] Y is binomial $n = 15$, $p = 0.4$

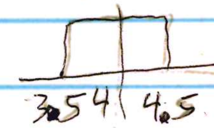
a) Find $P(Y=4)$ with the normal approx

Step *) $9\frac{.4}{.6} < \frac{9(.6)}{.4} = 13.5 < n = 15$

Step 0) $\mu_X = np = 15(.4) = 6$

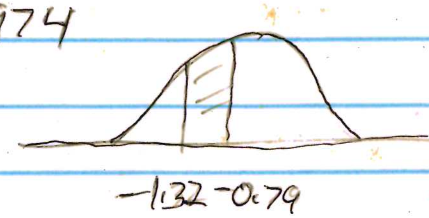
$\sigma_X = \sqrt{np(1-p)} = \sqrt{15(.4)\,.6} = \sqrt{3.6} = 1.8974$

i) $P(Y=4) \simeq P(3.5 < X < 4.5)$



ii) $\frac{3.5-6}{1.8974} = -1.32$, $\frac{4.5-6}{1.8974} = -0.79$

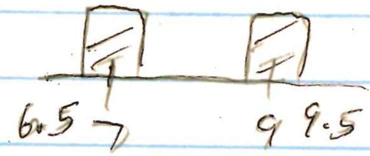
iii)



iv) $= P(-1.32 < z < -0.79) = P(z > 0.79) - P(z > 1.32)$

$= .2148 - .0934 = .1214$ (exact value is .1268)

b) $P(7 \leq Y \leq 9)$

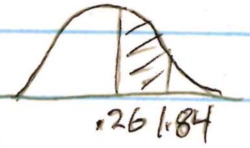


i)

×

ii) $\frac{6.5-6}{1.8974} = .26$, $\frac{9.5-6}{1.8974} = 1.84$

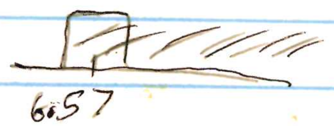
iii)



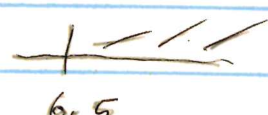
note
$P(7 < Y < 9) = P(Y = 8)$

iv) $= P(Z > .26) - P(Z > 1.84) = .3974 - .0329$
$= .3645$

c) $P(7 \leq Y)$



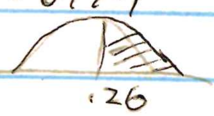
i)



note
$P(7 < Y)$
$= P(8 \leq Y)$
$\approx P(7.5 < \cancel{X})$

ii) $\frac{6.5-6}{1.8974} = .26$

iii)



iv) $P(Z > .26) = \boxed{.3974}$

This is much easier than finding $\sum_{i=7}^{15} b(15, i, 0.4)$

or $1 - \sum_{i=0}^{6} b(15, i, 0.4)$.

29] $\boxed{\text{Common}^{E4} \text{ problem}}$ Y bin $n = 20$ $p = .2$
Use normal approx to find $P(Y > 15)$
if possible.
$9\frac{p}{1-p} = 9\frac{.2}{.8} = 2.25 < 20$, but, $9\frac{.8}{.2} = 36 > 20$, not possible

END Actuarial Exam P material 483 50

Ch 8] 1] p391 An estimator of a parameter is a statistic based on the sample $Y_1, \dots, Y_n$.

2] In probability, the parameters θ are known and the focus is on how the sample will behave
eg $P(|X-\mu| < 3\sigma)$.

In statistics, a sample is observed and θ is unknown. want to estimate θ to gain knowledge about the population.

3] p391 A point estimator gives a numerical estimate of a population parameter

population	sample
mean μ	mean $\bar{Y}$
variance σ^2	variance S^2
SD σ	SD S
proportion p	proportion $\hat{p} = \frac{Y}{n} = \bar{W}$ where W_i is a 0 or 1.

$\bar{Y}, S^2, S,$ and $\hat{p}$ are all point estimators.
(while) $\mu, \sigma^2, \sigma,$ and p are all population parameters.

§8.2 4] p392 Suppose that $\hat{\theta}_n$ is a point estimator for θ based on a sample $Y_1, \dots, Y_n$.
Want $E\hat{\theta}_n \to \theta$ and want $V(\hat{\theta}_n) \to 0$ as $n \to \infty$.

5] If $E(\hat{\theta}) = \theta$ (for any n) then $\hat{\theta}$ is an unbiased estimator of θ. Otherwise, $\underbrace{\hat{\theta}}_{\text{"theta hat"}}$ is biased.

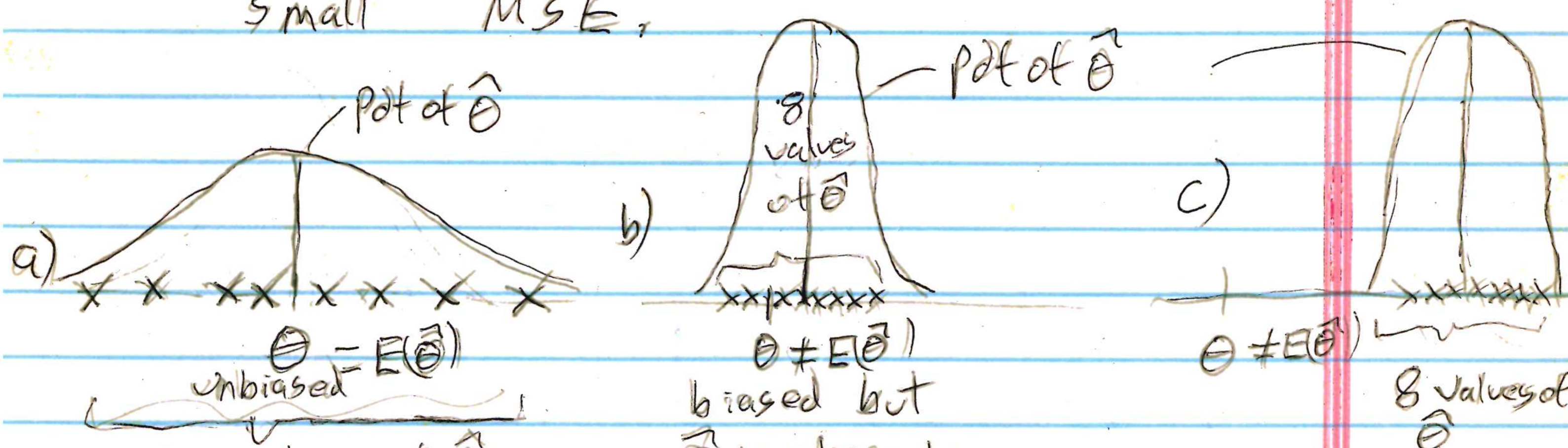
6] The bias of a point estimator is $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

7] The mean square error MSE of an estimator $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + [B(\hat{\theta})]^2.$$

← absolute

8] p393 A good estimator has small |bias| and small variance. So a good estimator has a small MSE.



unbiased, large variance

biased but $\hat{\theta}$ is closer to θ than in a)
small bias
small variance

large bias, small variance

See Fig 8.2 & 8.3

9] $\bar{Y}$ is an unbiased estimator of μ

S^2 σ^2

$\hat{p} = \frac{Y}{n} = \bar{W}$ where W_i is a 0 or a 1 p

S is not an unbiased estimator of σ.

§8.5 10] A confidence interval (CI) for θ is an interval estimator of θ $[\hat{\theta}_L, \hat{\theta}_U]$ where $\hat{\theta}_L \le \hat{\theta}_U$.

11] Idea: Suppose $Y_1, \ldots, Y_n$ are iid from a population where σ is known, but μ is unknown. Suppose that n is large enough so that the CLT holds.

i) $\overline{Y}$ is approx normal with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Let $A = P\left(\mu - 2\frac{\sigma}{\sqrt{n}} \le \overline{Y} \le \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx .95$

since $P(\overline{Y}$ is within $2\,\sigma_{\overline{Y}}$ of $\mu_{\overline{Y}}) \approx 0.95$ by the 68-95-99.7 empirical rule.

So $A = P\left(\frac{-2\sigma}{\sqrt{n}} \le \overline{Y} - \mu \le 2\frac{\sigma}{\sqrt{n}}\right)$

$= P\left(\overline{Y} - \frac{2\sigma}{\sqrt{n}} \le \mu \le \overline{Y} + \frac{2\sigma}{\sqrt{n}}\right) \approx .95$

and $\left[\overline{y} - \frac{2\sigma}{\sqrt{n}}, \overline{y} + \frac{2\sigma}{\sqrt{n}}\right]$ is an approximate 95% CI for μ.

based on observed values $y_1, \ldots, y_n$

ii) Suppose $\sigma = 10$, $n = 100$ and the CLT holds. Then $\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} = 1$ and $2\,\sigma_{\overline{Y}} = 2$.

want to know for which values of μ is it "reasonable" that $\bar{Y}$ came from a distribution with mean μ and SD $\sigma = 10$? Fix μ and take "reasonable" to mean that the probability that $\bar{Y}$ is within $2\sigma_{\bar{Y}}$ of μ is at least 95%:

$$P(|\bar{Y}-\mu| \le 2\sigma_{\bar{Y}}) = P(|\bar{Y}-\mu| \le 2) \ge 0.95$$

eg suppose $\bar{Y} = 90$

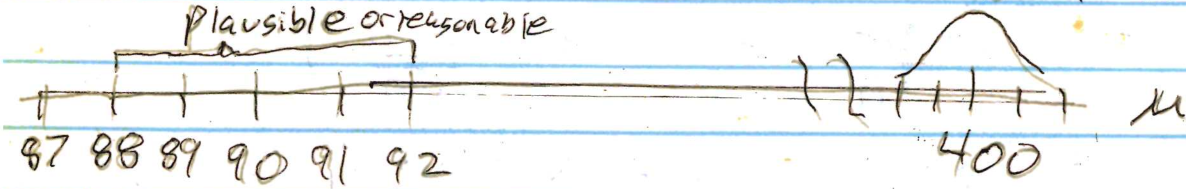
$\mu_1 = 90$, $\mu_2 = 400$, $\mu_3 = 88$, $\mu_4 = 87$ and $\mu_5 = 92$.

Then $\mu_3 = 88$ is the smallest "reasonable" value (by the 68-95-99.7 empirical rule) and $\mu_5 = 92$ is the largest reasonable value. So the 95% CI for μ is $\approx$

$$\bar{Y} \pm 2 = \underbrace{[88, 92]}_{\text{reasonable values for } \mu \text{ if } \bar{Y}=90}$$

1.96 is better

$\sigma = 10$ and $n = 100$.



if $\bar{Y} \in [398, 402]$

then $\mu = 400$ would be reasonable

12] Why 95% <u>Confidence</u> instead of 95% probability?

A CI is computed after the experiment. After the