Probability and Statistics for Engineering (ESE 326) Final Exam December 15, 2009

This exam contains 20 multiple-choice problems worth two points each and 10 true-false problems worth one point each, for an exam total of 50 points.

Part I. Multiple-Choice (two points each)

Clearly fill in the oval on your answer card which corresponds to the only correct response.

1. Irene is helping her mom decorate for a Christmas party. Her mom gives her a box containing 15 different Christmas decorations and asks her to either place one decoration in the middle of the serving table or choose one decoration for the left side and another for the right. In how many ways can Irene decorate the serving table following her mother's instructions?

(A) 118

(B) 120

(C) 223

(D) 225

(E) 455

(F) 457

(G) 2728

(H) 2730

$15 + 15 \cdot 14$

- or -

${}_{15}P_1 + {}_{15}P_2$

2. If a prospective student first visits Washington University on a day when the weather is good, there is a 60% chance that he or she will apply to WU. If the first visit occurs on a day when the weather is fair, there is a 44% chance that he or she will apply. If the first visit occurs on a day when the weather is bad, there is a 20% chance that he or she will apply. (You may assume that every prospective student visits.) The weather at WU is good 35% of the time, fair 50% of the time, and bad 15% of the time. If a student applies to WU, what is the probability that he or she first visited on a day when the weather was fair?

(A) .0652

(B) .1935

(C) .3548

(D) .4565

(E) .4783

(F) .5217

(G) .5435

(H) .6452

(I) .8065

(J) .9348

G: the weather is good
F: the weather is fair
B: the weather is bad
A: the student applies

Use Bayes' Theorem.

$$P[F/A] = \frac{P[A|F] \cdot P[F]}{P[A/F] \cdot P[F] + P[A|G] \cdot P[G] + P[A|B] \cdot P[B]}$$

$$= \frac{(.44)(.50)}{(.44)(50) + (.60)(.35) + (.20)(.15)} = .4783$$

3. When Snidely answers a question in court, there is a 40% chance that he will answer truthfully. Suppose that Snidely answers 17 questions when he is called to the witness stand in a particular court case. What is the probability that he answers the majority of these questions truthfully?

(A) .0000
(B) .0919
(C) .1070
(D) .1989
(E) .3595
(F) .6405
(G) .8011
(H) .8930
(I) .9081

binomial $n = 17$ $p = .4$

$P[X \geq 9] = 1 - P[X \leq 8]$

$= 1 - .8011$

$= .1989$

4. In Zack's Christmas songsheet, 11 out of 44 songs contain a reference to snow. Zack uses his TI to randomly choose 15 of the 44 songs to sing at a caroling party (at Irene's house). What is the probability that at least 2 of his chosen songs contain a reference to snow?

(A) .7639
(B) .8192
(C) .8629
(D) .9198
(E) .9563
(F) .9608
(G) .9916
(H) .9929
(I) .9955
(J) .9974

hypergeometric $N = 44$ $r = 11$ $n = 15$

$P[X \geq 2] = 1 - P[X = 0] - P[X = 1]$

$$= 1 - \frac{\binom{11}{0}\binom{33}{15}}{\binom{44}{15}} - \frac{\binom{11}{1}\binom{33}{14}}{\binom{44}{15}}$$

$= 1 - .0045 - .0392$

$= .9563$

5. A continuous random variable X has density function $f(x) = \frac{81}{x^4} = 81x^{-4}$, $x \geq 3$. Find Var X.

(A) 3

(B) 4.5

(C) 6.75

(D) 9

(E) 22.5

(F) 27

(G) $\sqrt{4.5}$

(H) $\sqrt{6.75}$

(I) $\sqrt{22.5}$

(J) $\sqrt{27}$

$$E[X] = \int_3^\infty 81x^{-3}dx = \lim_{b\to\infty} -\frac{81}{2}x^{-2}\Big|_3^b$$

$$= \lim_{b\to\infty}\left[\frac{-81}{b^2} + \frac{9}{2}\right] = \frac{9}{2}$$

$$E[X^2] = \int_3^\infty 81x^{-2}dx = \lim_{b\to\infty} -81x^{-1}\Big|_3^b$$

$$= \lim_{b\to\infty} \frac{-81}{b} + 27 = 27$$

$$\text{Var } X = 27 - \left(\frac{9}{2}\right)^2 = 6.75$$

6. The geometric distribution is a special case of what other distribution?

(A) binomial (B) negative binomial (C) hypergeometric

(D) Poisson (discrete) (E) normal (F) gamma

(G) Poisson (continuous) (H) chi-squared (I) Weibull

7. Suppose (X, Y) is a two-dimensional continuous random variable with density function as follows.

$$f_{XY}(x,y) = \frac{1}{48}(x+2y) \qquad 1 \leq x \leq 3 \quad 0 \leq y \leq 4$$

Which of the following integrals would be needed to calculate the marginal density $f_Y(y)$ for Y?

(A) $\int_1^3 \frac{1}{48}(x+2y)\,dx$

(B) $\int_1^3 \frac{1}{48}x(x+2y)\,dx$

(C) $\int_1^3 \frac{1}{48}y(x+2y)\,dx$

(D) $\int_0^4 \frac{1}{48}(x+2y)\,dy$

(E) $\int_0^4 \frac{1}{48}x(x+2y)\,dy$

(F) $\int_0^4 \frac{1}{48}y(x+2y)\,dy$

(G) $\int_0^4\int_1^3 \frac{1}{48}(x+2y)\,dx\,dy$

(H) $\int_0^4\int_1^3 \frac{1}{48}x(x+2y)\,dx\,dy$

(I) $\int_0^4\int_1^3 \frac{1}{48}y(x+2y)\,dx\,dy$

8. Suppose (X, Y) is a discrete random variable with the following density table. Find $\mu_{Y|x=2}$.

x/y	4	6	10	$f_X(x)$
0	.1	.1	0	.2
2	.1	.3	.1	.5
3	.2	0	.1	.3

$$f_{Y|x=2} = \frac{f_{XY}(2,y)}{f_X(2)}$$

(A) 2.8

(B) 3.2

(C) 3.6

(D) 4.0

(E) 4.4

(F) 4.8

(G) 5.2

(H) 5.6

(I) 6.0

(J) 6.4

y	4	6	10
$f_{Y\|x=2}$	.2	.6	.2

$$\mu_{Y|x=2} = (4)(.2) + (6)(.6) + (10)(.2) = 6.4$$

9. Nine normal curves are shown below. Suppose that curve (e) pictures the distribution of a certain random variable X. Then which of the nine curves pictures the distribution of the sample means $\overline{X}$? (For example, if this distribution is the same, select curve (e). If it differs in the size of its mean and/or standard deviation, select accordingly.) (You may assume $n > 1$.)

(A) curve (a)

(B) curve (b)

(C) curve (c)

(D) curve (d)

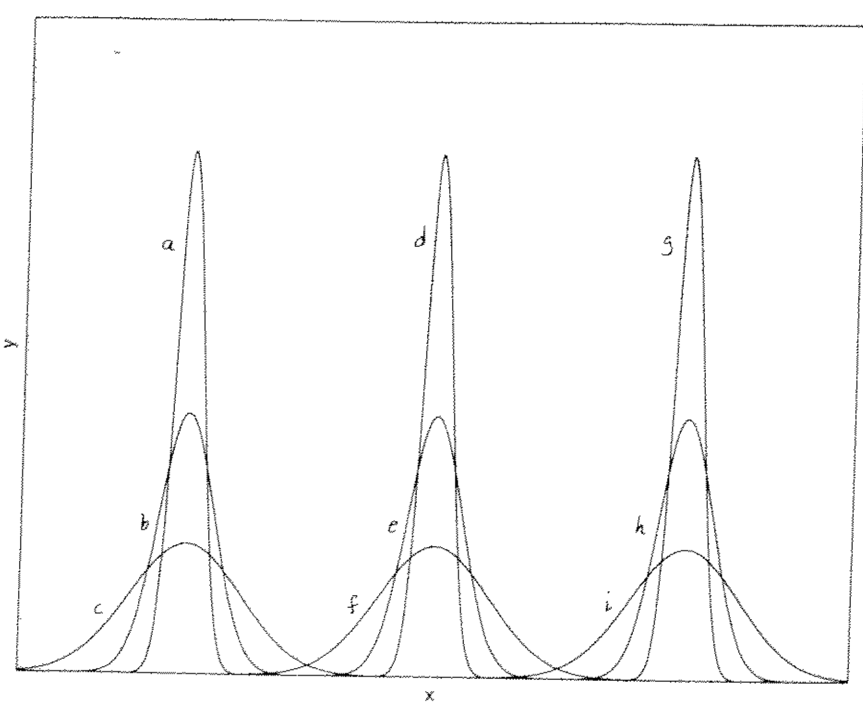
(E) curve (e)

(F) curve (f)

(G) curve (g)

(H) curve (h)

(I) curve (i)



same mean; smaller standard deviation

10. Suppose X and Y are independent normal random variables with parameters as follows.

$\mu_X = 15 \qquad \sigma_X = 4 \qquad \mu_Y = 25 \qquad \sigma_Y = 3$

Find $P[X + Y \geq 48]$.

(A) .0111
(B) .0266
(C) .0548
(D) .1265
(E) .3745
(F) .6255
(G) .8735
(H) .9452
(I) .9734
(J) .9889

$\mu_{X+Y} = \mu_X + \mu_Y = 40$

$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y = 25$ (independence)

$\sigma_{X+Y} = 5$

$X+Y$ is normal (independence)

$P[X+Y \geq 48] = P\left[Z \geq \frac{48-40}{5}\right]$

$= P[Z \geq 1.6] = .0548$

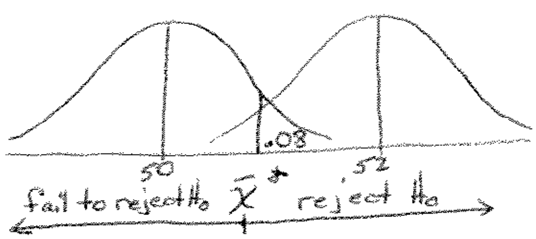
11. Let X be a normal random variable, and consider a hypothesis test on the mean μ of X having the following characteristics.

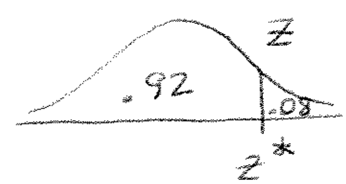
$n = 16 \qquad \alpha = .08 \qquad \sigma = 3$

$H_0\colon\ \mu = 50 \qquad H_1\colon\ \mu > 50$

Find the probability of committing a Type II error given that $\mu = 52$.

(A) .0011
(B) .0472
(C) .0800
(D) .1035
(E) .1333
(F) .1543
(G) .1798
(H) .2151
(I) .2880
(J) .3379





$z^* = 1.405072$

$\frac{\bar{x}^* - 50}{3/\sqrt{16}} = 1.405072 \qquad \bar{x}^* = 51.053804$

$\beta = P[\bar{X} < 51.053804 \mid \mu = 52]$

$= P\left[Z < \frac{51.053804 - 52}{3/\sqrt{16}}\right] = P[Z < -1.261595]$

$= .1035$

12. Let X be a random variable, and suppose $[L_1, L_2]$ is a confidence interval for the mean μ of X, where $\alpha = .05$. Which of the following is an appropriate statement to make about this interval?

(A) We are 90% confident that μ lies in this interval.

(B) We are 90% confident that μ does not lie in this interval.

(C) We are 90% confident that $\overline{x}$ lies in this interval.

(D) We are 90% confident that $\overline{x}$ does not lie in this interval.

(E) We are 95% confident that μ lies in this interval.

(F) We are 95% confident that μ does not lie in this interval.

(G) We are 95% confident that $\overline{x}$ lies in this interval.

(H) We are 95% confident that $\overline{x}$ does not lie in this interval.

13. A nutritionist wants to determine the proportion of people in her community who understand the difference between good cholesterol and bad cholesterol. What sample size should she use in order to estimate this proportion to within .05 with 88% confidence? (Choose the best answer — neither too low to accomplish the goal nor higher than necessary.)

(A) 34

(B) 35

(C) 60

(D) 61

(E) 138

(F) 139

(G) 155

(H) 156

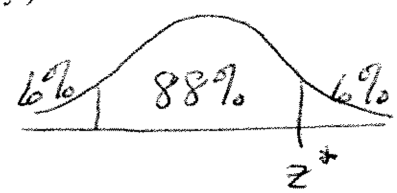
(I) 241

(J) 242

$$n \geq \frac{(z^*)^2}{4d^2}$$

$$n \geq \frac{(1.554774)^2}{4(.05)^2}$$

$$= 241.7$$



$$z^* = 1.554774$$

14. Which of the following statements about the slope of a regression line is correct?

(A) $E[\beta_1] = B_1$

(B) $E[\beta_1] = b_1$

(C) $E[B_1] = b_1$

(D) $E[B_1] = \beta_1$

(E) $E[b_1] = \beta_1$

(F) $E[b_1] = B_1$

The random variable B_1 is unbiased for the parameter β_1.

15. Suppose we are testing the hypothesis that the slope of a regression line is different from zero (given the four standard assumptions on the "error" random variables E_i). Given the following information, find the P value for this hypothesis test.

$H_0\colon\ \beta_1 = 0 \qquad H_1\colon\ \beta_1 \neq 0$

$n = 20 \qquad \alpha = .05 \qquad b_1 = -2.1 \qquad s = 2.357 \qquad S_{xx} = 40$

(A) 2.7×10^{-6}
(B) 4.1×10^{-6}
(C) 5.4×10^{-6}
(D) 8.3×10^{-6}
(E) 9.8×10^{-6}
(F) 1.2×10^{-5}
(G) 2.0×10^{-5}
(H) 2.4×10^{-5}
(I) 3.8×10^{-5}
(J) 7.6×10^{-5}

$$2P[B_1 \leq -2.1 \mid \beta_1 = 0]$$
$$= 2P\left[T_{18} \leq \frac{-2.1 - 0}{\frac{2.357}{\sqrt{40}}}\right]$$
$$= 2P[T_{18} \leq -5.634945]$$
$$= 2(1.2 \times 10^{-5}) = 2.4 \times 10^{-5}$$

16. Consider the following data.

x	1	2	3	4	5	6
y	47	42	41	37	31	28

Find the right endpoint L_2 for a 95% confidence interval for $\mu_{Y|x=3.5}$ (given the four standard assumptions on the "error" random variables E_i).

(A) 39.10
(B) 39.43
(C) 39.47
(D) 39.88
(E) 40.44
(F) 40.76
(G) 40.81
(H) 41.21
(I) 41.47
(J) 41.60

regression line: $y = 50.866667 - 3.771429x$

point estimate:

$$\hat{\mu}_{Y|x=3.5} = 50.866667 - 3.771429(3.5)$$
$$= 37.666666$$

$t^* = 2.776445$

$SSE = 6.419048$

$s^2 = \frac{SSE}{4} = 1.604762 \qquad s = 1.266792$

$$L_2 = 37.6667 + (2.7764)(1.2668)\sqrt{\tfrac{1}{6}}$$
$$= 39.10$$

17. A 3-sigma $\overline{X}$ control chart is to be set up to monitor the sample means of a normal random variable X, using samples of size $n = 4$. When the process is in control, the mean and standard deviation of X are $\mu_0 = 17.0$ and $\sigma = 3.0$. The control limits can then be computed to be as follows. (You do not need to verify these.)

$$\text{LCL} = 12.5 \qquad \text{UCL} = 21.5$$

If the mean shifts to $\mu = 19$, what is the average run length, to the nearest integer?

(A) 6
(B) 8
(C) 10
(D) 12
(E) 15
(F) 21
(G) 59
(H) 211
(I) 720
(J) 2330

First find the probability of a signal:

$P[\bar{X} < 12.5 \text{ or } \bar{X} > 21.5 \mid \mu = 19]$

$= P\left[Z < \frac{12.5 - 19}{3/\sqrt{4}} \text{ or } Z > \frac{21.5 - 19}{3/\sqrt{4}}\right]$

$= P[Z < -4.3333 \text{ or } Z > 1.6667] = .0478$

average run length: $\frac{1}{.0478} \approx 21$

18. A fast-food chain beginning a quality control program decides to monitor the length of its french fries. In order to set up a 3-sigma $\overline{X}$ control chart, the quality control engineer takes six random samples, each of size four. (He is blissfully unaware that at least 20 such samples should be taken.) The following data are obtained. Find the value of $\bar{r}$.

sample number	length of french fry (inches)				r
1	4.2	5.3	4.0	2.8	2.5
2	3.3	4.2	5.1	4.6	1.8
3	4.4	3.7	5.0	4.2	1.3
4	2.9	3.7	3.3	4.3	1.4
5	6.1	2.6	4.1	4.2	3.5
6	5.6	3.2	5.3	5.9	2.7

(A) 1.6
(B) 1.75
(C) 1.85
(D) 2.2
(E) 2.25
(F) 2.775
(G) 3.175
(H) 3.375
(I) 3.6

$\bar{r} = 2.2$

19. The value of $\overline{x}$ in the above problem is $\overline{x} = 4.25$. (You do not need to verify this.) The value of $\overline{r}$ above is not $\overline{r} = 2.6$, but let's pretend that it is. (In other words, do not use the value you got for $\overline{r}$ in problem 18 in your calculations for problem 19, but instead use $\overline{r} = 2.6$. This way, your success on this problem will not depend on your success on the previous problem.) Under this assumption, together with the assumption that french fry length is normally-distributed, what is the upper control limit for the 3-sigma $\overline{X}$ chart?

(A) 5.51

(B) 5.62

(C) 5.76

(D) 5.80

(E) 5.93

(F) 6.14

(G) 6.71

(H) 7.43

(I) 7.87

(J) 8.68

$d_2 = 2.059$

$UCL = 4.25 + 3\frac{2.6}{(2.059)\sqrt{4}}$

$= 6.144$

TABLE XII
Control chart constants

Number of observations in sample, n	d_2	d_3
2	1.128	0.853
3	1.693	0.888
4	2.059	0.880
5	2.326	0.864
6	2.534	0.848
7	2.704	0.833
8	2.847	0.820
9	2.970	0.808
10	3.078	0.797
11	3.173	0.787
12	3.258	0.778
13	3.336	0.770
14	3.407	0.762
15	3.472	0.755
16	3.532	0.749
17	3.588	0.743
18	3.640	0.738

20. In the movie "Significance Tests," a hypothesis test was performed to test the hypothesis that a recently-discovered poem attributed to William Shakespeare was in fact not written by Shakespeare. What aspect of the poem was compared to known Shakespearean poems in this hypothesis test?

(A) the alliteration

(B) the average number of letters in the words

(C) the handwriting

(D) the imagery

(E) the metric style

(F) the number of lines

(G) the number of new words

(H) the number of verses

(I) the percentage of rhyming lines

(J) the spelling

Part II. True-False (one point each)

Mark "A" on your answer card if the statement is true; mark "B" if it is false.

21. Suppose A_1 and A_2 are events such that $P[A_1] = .7$, $P[A_2] = .4$, and $P[A_1 \cup A_2] = .82$. Then A_1 and A_2 are independent.

$$P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$$

true

$$.82 = .7 + .4 - P[A_1 \cap A_2]$$

$$P[A_1 \cap A_2] = .28 = P[A_1] \cdot P[A_2]$$

22. The following function qualifies as a density for a discrete random variable.

$f(x) = (3.75)(.4)^x \qquad x = 2, 3, 4, \ldots$

true

x	2	3	4	$\cdots$
$f(x)$	$.6$	$(.6)(.4)$	$(.6)(.4)^2$	

geometric

$a = .6 \quad r = .4 \qquad \frac{a}{1-r} = \frac{.6}{1-.4} = 1$

23. For any random variable X, $E[X^2 + X] = \left(E[X]\right)^2 + E[X]$.

false

$$E[X^2 + X] = E[X^2] + E[X] \neq \left(E[X]\right)^2 + E[X]$$

24. Suppose X is a normal random variable with $\sigma^2 = 10$. Then $2X$ is a normal random variable with $\sigma^2 = 20$.

false

$$\sigma^2 = 40$$

25. Suppose that in a class of 100 students, the mean for an exam is exactly 80%. Then, (assuming no individual student scored exactly 80%), 50 students scored above 80% and 50 students scored below 80%.

false

If the median were 80%, then we could conclude that 50 students scored above 80% and 50 scored below.

26. If the hazard rate function for a certain system is $\rho(t) = \frac{1}{2}t^{-1/2}$, then the system is more and more likely to fail as time goes by.

false $\beta < 1$, so the system is less likely to fail as time passes.

27. Suppose (X, Y) is a two-dimensional random variable. If the correlation ρ_{XY} equals zero, then X and Y are independent. (Of course, the "ρ" in problem 26 and the "ρ" in problem 27 have nothing to do with one another.)

false However, if X and Y are independent, then $\rho_{XY} = 0$.

28. In order to construct a confidence interval for the unknown variance σ^2 of a normal random variable X, one should use T_{n-1} for the prototype distribution.

false Use χ^2.

29. Suppose you are performing a hypothesis test. All other things being equal, if the value of α increases, then the P value increases.

false A change in α does not affect the P value. In fact, α does not even need to be known for the P value to be computed.

30. The sum of the residuals for a regression line always equals zero.

true