Key to Sample Stat 224 Final

Prof. Callan

1. a) Because of independence, $P(A \cap B) = P(A)P(B) = (.6)(.4) = .24$. Next, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .6 + .4 - .24 = .76$. For (iii), $P(A|B) = P(A \cap B)/P(B) = .6$, but the quick way is to use independence which says $P(A|B) = P(A) = .6$. For part (iv), "exactly one'' means A occurs and B, C do not, OR B occurs and A, C do not, OR C occurs and A, B do not. So the required probability is $P(A \cap B' \cap C') + P(B \cap A' \cap C') + P(C \cap A \cap B') = (.6)(.6)(.7) + (.4)(.4)(.7) + (.4)(.6)(.3) = .436$ (again using independence). Alternatively, you could construct a Venn diagram and read off all the answers except part (iii).

b) (i) True

(ii) False (t_α *decreases* to z_α as the number of degrees of freedom increases.)

(iii) True

(iv) True (one reason $n - 1$ rather than n appears in the denominator of *sample* variance is precisely to make this statement true.)

(v) False (we want a very low probability of convicting an innocent defendant—even at the risk of acquitting a guilty defendant. So Type I error, convicting an innocent defendant, is considered more serious.)

2. a) In 36 plays, on average you win all but 6 times (when the dice are equal), in other words you win on average 30 times out of 36, or 5 times out of 6, so you win $5 \times 12 = \$60$ per 6 plays or \$10 per play. The fair price is thus \$10.

b) ${}_{11}P_3 = 11 \times 10 \times 9 = 990$

c) $\binom{20}{4}\binom{20}{2} = 4845 \times 190 = 920,550$ which is a little under a million.

3. Say B_1 is the event "(random) person has disease" and B_2 is the event "does not have disease". Also, A = "test comes up positive". So $P(B_1) = .0001$, $P(B_2 = .9999$, $P(A|B_1) = .98$, and $P(A|B_2) = .01$.

The Bayes formula then says

$$P(B_1|A) = \frac{.98 \times .0001}{.98 \times .0001 + .01 \times .9999} = .0097,$$

the answer to part a) (good news for a worried patient)

Now take as A the event "test comes up negative". So now $P(A|B_1) = .02$, and $P(A|B_2) = .99$.

The Bayes formula now says

$$P(B_1|A) = \frac{.02 \times .0001}{.02 \times .0001 + .99 \times .9999} = .00000202,$$

only about 1 in 500,000.

4. a) Working in units of a thousand dollars, the confidence interval for the mean is

$$18.5 \pm 1.96 \times 1.7/\sqrt{45} = 18.5 \pm .50 = (18.0, 19.0).$$

b) The formula gives

$$\frac{145}{200} - \frac{200}{360} \pm \sqrt{\frac{\frac{145}{200}\frac{55}{200}}{200} + \frac{\frac{200}{360}\frac{160}{360}}{360}} = .17 \pm .08 = (.09, .25)$$

5. a) Since we have no prior information, use $p = 1/2$ in the formula $n = (z_{\alpha/2}/E)^2 p(1-p)$ to get $(1.96/.04)^2/4 = 600.25$, so the answer is 601.

b) Calculate $d = |\mu - \mu_0|/\sigma = |28 - 30|/2.7 = .74$. Then Table 8a) gives $\beta \approx .05$.

6. The frequency for $X = 0$ should read 23 (not 29—the given frequencies don't add up to 100). First compute the sample mean $\overline{X} = (23\times0+35\times1+26\times2+11\times3+5\times4)/100 = 1.4$. From "$\mu = np$" for a binomial distribution, estimate p as $1.4/4 = .35$. Using the cumulative $P(X \leq k)$ tables in the text, construct the following table of expected frequencies:

k	$P(X \leq k)$	$P(X = k)$	expected freq.
0	.1785	.1785	17.85
1	.5630	.3845	38.45
2	.8735	.3105	31.05
3	.9850	.1115	11.15
4	1.0000	.0150	1.50

The last two rows (cells) must be combined to ensure each expected frequency is at least 5, and the χ^2 test statistic is then $\sum_i (e_i - o_i)^2/e_i = 3.504$. The number of degrees of freedom is $\nu = 4($ # cells$) - 2$ (two parameters, sample size 100 and p, gotten from the data were used to get the expected frequencies)$= 2$ and $\chi^2_{.05}$ at $\nu = 2$ is 5.991. Since 3.504 is not > 5.991, we conclude the data are consistent with a binomial distribution.

7. Calculate $b = S_{xy}/S_{xx} = .837$, SSE $= S_{yy} - S^2_{xy}/S_{xx} = 2.000$, $s_e = \sqrt{\text{SSE}/(n-2)} = .160$. For 98% confidence, $z_{\alpha/2} = z_{.01} = 2.33$. So the 98% confidence interval for the slope β is

$$.837 \pm 2.33(.160)/\sqrt{26.4} = .837 \pm .073 = (.764, .910)$$

8. a) We need $\chi^2_{.025} = 106.629$ and $\chi^2_{.975} = 57.153$ (both at $\nu = 80$). Also $(n-1)s^2 = 8.090$. So the required confidence interval is

$$(\sqrt{8.090/106.629}, \sqrt{8.090/57.153}) = (.275, .376)$$

b) SSE = SST-SS(Tr) = 66.7, MSE = 66.7/60 = 1.112, MS(Tr) = 13.2/4 = 3.3. So the test statistic is $F = 3.3/1.112 = 2.96$. And $F_{.05}$ with 4 d.f. in the numerator, and 60 d.f. in the denominator is 2.53. Since $2.96 > 2.53$, we reject the null hypothesis and conclude there is a significant difference between the treatments.