Answers of Master's Exam, Spring, 2008

To: Master's degree students.
From: Jim Stapleton, Professor Emeritus

I discovered, as I did the problems below, that some of these problems are quite difficult. I suggest that you not try to complete all of them. For example, 4d) and 5b) took me a long time to complete.

1) a) $\dfrac{\binom{32}{16}\binom{32}{16}}{\binom{64}{32}}$ b) $\dfrac{\binom{32}{3}\binom{32}{7}}{\binom{64}{10}}$

2) a) 0.04187 b) 0.47412 c) P(Exactly One) = 0.25803
d) Using ½ correction and normal approximation: 0.0870

3) a) $\lim_{n\to\infty} F_n(x)$ = F(x) for every x on the real line at which F is continuous. F is the cdf For X.
b) $F_n(x) = 0$ for $x < 0$, $1 - (1 - x/n)^n$ for $0 \le x \le n$, 1 for $x \ge n$.
$\lim_{n\to\infty} F_n(x)$ = 0 for $x < 0$, $1 - e^{-x}$ for $x \ge 0$.. The limit function is
F(x) = 0 for $x < 0$, $1 - e^{-x}$ for $x \ge 0$. Thus $\lim_{n\to\infty} F_n(x)$ = F(x) for every x.

c) For every $\varepsilon > 0$ $\lim_{n\to\infty} P(\,|X_n - 1| < \varepsilon\,) = 1$.

4) a) $f(x, y) = 2(1 - y)^{1/2}$ for $0 \le x \le (1 - y)^{1/2}$, $0 < y \le 1$

b) $(4/3)(1 - x^3)$, $0 < x \le 1$

c) $f_{Y|X}(\,y \mid x\,) = 2(\,1 - y)^{1/2}/((4/3)(1 - x^3))$ for $0 < x \le (1 - y)^{1/2}$, $0 < x \le 1$.

d) $(3/(2(1 - x^3))(4/15 - (2/3)x^3 + (2/5)x^5)$ for $0 < x \le 1$. E(Y | X) is obtained by replacing x by X.
e) $V(x) = E(\,X^3 + X Y \mid X = x) = x^3 + x\,E(\,Y \mid X = x) = x^3 + x(4/3)(1 - x^3)$. Replace x by X to obtain V = V(X).

5) a) $(2\,U - 1)^{1/5}$.
b) $F_M(m) = 2(m^5/4 - 5\,m^5 \ln(m)/4) + ½$ for $0 \le m \le 1$. By symmetry,
$F_M(m) = 1 - F_M(-m)$ for $-1 \le m < 0$. Of course, $F_M(m) = 0$ for $m < -1$, 1 for $m > 1$. This is a difficult problem, far too long for this exam. Don't spend much time on it.

The derivative with respect to m is the density. It's messy, too messy to write here.

c) $F_Y(y) = (\ln(y))^5$ for $1 < y \le e$, 0 for $y \le 1$, 1 for $y > e$.
$f_Y(y) = 5 \ln(y)/y$ for $1 \le y \le e$, 0 otherwise
This is a difficult problem, far too long for this exam. Don't spend much time on it.

d) Since the rv's have the same variance, σ^2 say, the correlation coefficient does not depend on it. The covariance of the two linear combinations is 57. The variances are 130 and 99. The correlation coefficient is $57 / [130(99)]^{1/2}$.

e) Since $X_2^{1/3}$ has a symmetric distribution about zero, $E(X_2^{1/3}) = 0$, so that by independence, $E(W) = 0$. We should also note that $E(X_1^2)$ and $E(1/X_3^4)$ exist so that $E(W)$ exists.

6) a) MLE is $\hat{\alpha} = \max(X_1, \ldots, X_n)$

b) $\mu = E(X_1) = (2/3)\alpha$, so $\alpha = (3/2)\mu$, $\hat{\alpha}_{mom} = (3/2)\bar{X}$

c) $E(\hat{\alpha}) = \alpha\, 2n/(2n+1)$, so bias $= -\alpha/(2n+1)$

7) a) $\lambda(x) = (x^3/4)/(x/2) = x^2/2$ for $0 < x \le 2$. Reject for large $\lambda(x)$, therefore for $x \ge k$, for some k. We want $P(X \ge k \mid H_0) = 1 - F_0(k) = 1 - (k/4)^2 = \alpha$, $k = 2(1 - \alpha)^{1/2}$.

b) Neyman-Pearson

c) Power $= P(X \ge k \mid H_1) = 1 - k^4 = 1 - (1 - \alpha)^2$.

8) a) Let D_i = Expenditure – Intake for Player i
Suppose the D_i 's are a random sample from some the $N(\mu_D, \sigma_D^2)$ distribution.. Let

$H_0: \mu_D = 0$, $H_1: \mu_D \ne 0$. We observe $\bar{D} = 1.5857$, $S_D^2 = 1.9414$, $t = 3.011$. Since the 0.975 quantile of the t-distribution with 6 df is 2.447, we reject H_0 at the 0.05 level.

b) We could perform a sign test. The number of observations less than zero is 2. Therefore the observed p-value is 2(29/128) = 58/128.
We could also use Wilcoxon's signed rank test. The sum of the ranks of the negative ranks is $W_- = 1 + 3 = 4$. (Remember that we are ranking the 7 absolute values.) . Since $P(W_- \le 4 \mid H_0) = 7/128$, the p-value for a two-sided test is 14/128. We do not reject H_0.

9) Let $X_1, \ldots, X_5$ be the control observations. Let $Y_1, \ldots, Y_5$ be the steroid observations. Suppose that the X_i's constitute a random sample from a cdf F, and the Y_i's constitute a random sample from a cdf G. We wish to test $H_0: F(x) = G(x)$ for all x vs H_1: H_0 not true. Suppose all 10 rv's are independent.

The Wilcoxon rank sum statistic, the sum of the ranks of the steroid observations is $W = 1 + 10 + 9 + 8 + 7 = 35$. $P(W \ge 35 \mid H_0) = 17 / \binom{10}{5}$, so the p-value is $34/\binom{10}{5} = 34/252 > 0.05$. Do not reject H_0.

10) a) Let $Q(\beta) = \Sigma (Y_i - \beta x_i^2)$. Taking the partial derivative with respect to β and setting the result equal to zero, we get $\hat{\beta}$.

b) $\hat{\beta} = \Sigma x_i^2 (\beta x_i^2 + \varepsilon_i)/ \Sigma x_i^4 = \beta + \Sigma x_i^2 \varepsilon_i / \Sigma x_i^4$. The second term has expectation zero because $E(\varepsilon_i) = 0$ for each i.

c) $Var(\hat{\beta}) = Var(\Sigma x_i^2 \varepsilon_i / \Sigma x_i^4) = \sigma^2 \Sigma x_i^4 / (\Sigma x_i^4)^2 = \sigma^2 / \Sigma x_i^4$..

d) $\hat{\beta} = 3$, $\hat{Y}_1 = 3$, $\hat{Y}_2 = 12$, $\hat{Y}_3 = 3$, $\hat{Y}_4 = 12$, so the residuals e_i are 1, -2, -1, 2. The estimate of σ^2 is therefore $s^2 = 10/3$, and the 95% confidence interval is 3 ± 1.8289.

11) The matrix of estimates of the cell expected frequencies under the null hypothesis of independence of Size and Distance is

	[,1]	[,2]	[,3]
[1,]	10	25.40	14.60
[2,]	12	30.48	17.52
[3,]	20	50.80	29.20
[4,]	8	20.32	11.68

Pearson's chi-square statistic is 13.823. Since df = (4-1)(3-1) = 6, the observed p-value is 0.0317. Reject at the 0.1 level (or any level ≥ 0.0313). The 0.90 quantile of the chi-square distribution with 6 df is 10.645

12) a) Let X_1 = (# who want decrease at time t_1),
X_2 = (# who want decrease at time t_2),
$\hat{p}_1 = X_1/n_1$, $\hat{p}_2 = X_2/n_2$,
$\hat{\Delta} = \hat{p}_1 - 2\hat{p}_2$. Then $E(\hat{\Delta}) = \Delta$, and $Var(\hat{\Delta}) = p_1 q_1/n_1 + 4 p_2 q_2/n_2$.
We can replace p_1, p_2, q_1, q_2 by their estimates to obtain an estimate of $Var(\hat{\Delta})$. Call this $\hat{\sigma}^2$. Then a 95% confidence interval on Δ is given by

$$[\hat{\Delta} \pm 1.96\,\hat{\sigma}].$$

We have assumed that sampling is with replacement. For samples of 500 and 1000 it makes little difference whether sampling is with or without replacement.

b) [-0.9 ± 0.0016]

c) For independent repetitions of this experiment, 95% all repetitions will produce random intervals given by the formula in a) which will contain Δ. Thus, in 10,000 repetitions we would expect 9500 to contain Δ. The particular interval in b) was one such interval.

d) The sample size needed if sampling were with replacement is 48020. Sampling with replacement would be silly, of course. From the sampling course, you may have learned that the sample size needed for without replacement sampling (“simple random sampling”) would therefore be $\frac{n}{1 + n/N} = \frac{48020}{1+48020/45000} = 23230$.