Answers of Master's Exam, Fall, 2007

Probability

1) a) There are 1 C, 5 A's, 2 R's, 1 D, 2 B's
$\binom{1}{1}\binom{5}{1}\binom{2}{1}\binom{1}{1}/\binom{11}{4}$ = (1)(5)(2)(1) /330
b) $(9/11)^5$ (2/11)
c) Let X = (# cards with R), X ~ Binomial(605, 2/11), E(X) = 110, Var(X) = 90, z = (131.5 – 110)/$\sqrt{90}$ = 2.2663, P(X > 131) = 1 - Φ(z) = 0.0117.
d) 0.1103
e) 5! 2! 2! 1! !! /11! = (5)(2)(2)(4)(1)(3)(1)(2)(2)(1)(1)/ (11!)

2) a) $X_1 = U^{1/2}$
b) For $-1 \le s < 0$, let z = 1 + s. . $F_S(s) = z^2 - z^4 / 2 + (4/3)z^3\, y - z^2 y^2$.=
$z^3(1 - s) - z^4 / 2 + (4/3)z^3\, y$
For $0 \le s \le 1$, $F_S(s) = 1 - F_S(-s)$ by symmetry around zero.
Comment: This problem can take much time to solve accurately. I advise students to indicate the double integrals involved (over triangles) without simplifying them.

c) $F_Y(y) = F_X((1 + y^{1/2})/2) - F_X(((1 - y^{1/2})/2) = (1/4)[(1 + y^{1/2})^2 - (1 - y^{1/2})^2 = y^{1/2}$ for $0 \le y \le 1$, 0 for y < 0, 1 for y > 1

d) Let $\sigma^{2-} = Var(X_1) = Var(X_2) = Var(X_3)$, $Z_1 = X_2 + 2 X_3$, $Z_2 = 5 X_1 + 4 X_2 - 3 X_3$. Then $Cov(Z_1, Z_2) = -2 \sigma^2$, $Var(Z_1) = 5 \sigma^2$, $Var(Z_2) = 50 \sigma^2$, $\rho(Z_1, Z_2) = -2 \sigma^2 /\sqrt{(5)(50) \sigma^4} = -2/\sqrt{250}$. Notice that the correlation coefficient does not depend on σ^2 . I saved some time in writing by defining the symbols Z_1, Z_2..

e) By independence $E(W) = E(X_1^2) E(1/X_2) E(1/X_3) = (2/3)2^2 = 8/3$**. Don't make the mistake** of thinking that $E(1/X_2)$ is equal to $1/E(X_2)$.

3) a) $\lim_{n\to\infty} F_n(x)$ = F(x) for every x on the real line at which F is continuous.

b) $\lim_{n\to\infty} F_n(x)$ = 0 for x < 0, 1 for x ≥ 1. The distribution with cdf F(x) = 0 for x < 1, 1 for x ≥ 1 (mass one at 1) is therefore the limiting distribution of the sequence $\{F_n\}$. Though F is discontinuous at x = 1, we need not have $\lim_{n \to \infty} F_n(1)$, though we do for this example. If, for example, $F_n(x) = 0$ for $x \le 1$, $n(x - 1)$ for $1 < x \le 1 + 1/n$, 1 for $x \ge 1 + 1/n$, then $\lim_{n \to \infty} F_n(1) = 0$, but the limiting cdf is F.

c) Let $X_1, \ldots, X_n$ be iid with mean μ, variance σ^2 . Let $S_n = X_1 + \ldots + X_n$.
Let $Z_n = (S_n - n\mu)/\sqrt{n \sigma^2}$. Then for any real number z $\lim_{n \to \infty} P(Z_n \le z) = \Phi(z)$, the cdf of the N(0, 1) distribution.

4) a) $f_X(x) = (2/\pi) \sqrt{1 - x^2}$ for $0 \le x \le 1$

b) $f_{Y|X}(y \mid x) = (1/\pi)/f_X(x)$ for $0 \le y \le \sqrt{1 - x^2}$, $- 1 < x < 1$.

Don't forget to give the domain of functions!

c) Given X = x, Y is uniformly distributed on the interval $[- \theta, \theta]$, where $\theta = \sqrt{1 - x^2}$ . .The expected square of such the $U(- \theta, + \theta)$ distribution is $\theta^2/3$.

Thus $E(Y^2 \mid X = x) = (1 - x^2)/3$, so $E(Y^2 \mid X) = (1 - X^2)/3$.

$V = E(X + Y^2|X) = X + (1 - X^2)/3$.

5) a) Y takes the values -2 and +2. $P(Y = +2) = \int_0^1 3 x^2 E(Y|X = x) dx$

$= \int_0^1 3 x^2 x \, dx = ¾$. $P(Y = -2) = ¼$.

b) $f_{X|Y}(x \mid y) = 3 x^2 (1 - x)/ (1/4) = 12 x^2 (1 - x)$ for $0 \le x < 1$, o otherwise.

Statistics

6) a) $\hat{a} = \min(X_1, \dots , X_n)$ $P(\hat{a} \le x) = 1 - P(\hat{a} > x) = 1 - [e^{-2(x - a)}]^n = 1 - e^{-2n(x - a)}$ for $x \ge a$

b) $\mu = a + ½$, so the MOM Est. of a is $\bar{X} - ½$.

c) $E(\hat{a}) = a + (1/2)/n$, so the bias is $1/(2n)$.

7) Let D_i = (Before Value – After Value) for patient i, i = 1, … , 10.

Suppose the D_i are a random sample from the $N(\mu_D, \sigma_D^2)$ distribution.

Test $H_0: \mu_D \le 0$ vs $H_a: \mu_D > 0$.

The Di are : 2.7, 2.1, 2.2, 2.6, 2.4, -0.5, -1.0, 1.3, -1.8, -2.0.

We observe $\bar{D} = 1.2$, $S^2 = 2.76$, t = 2.28. The 0.99 quantile of the t-distribution with 9 df is 2.82, so we should not reject H_0 at the $\alpha = 0.01$ level.

b) Either the sign test or the Wilcoxon signed rank test could be used. The p-value for the sign test is $P(X \le 4) = 386/2^{10} = 0.37695$. Do not reject.

The Wilcoxon signed rank statistic is $W_+ = 55$ – (sum of ranks of negative values) = 55 – (1+2+4) = 48. $P(W_+ \geq 48) = P(W_- \leq 7) = 1 + 7 + 5 + 3 + 1 + 2)/2^{10} = 20/1024 > 0.01$, so again we do not reject at the $\alpha = 0.01$ level.

8) a) $X_1 \sim \text{Binomial}(n_1 = 1000, p_1)$, $X_2 \sim \text{Binomial}(n_2 = 800, p_2)$, independent.

Let $\hat{p}_1 = X_1 / n_1$, $\hat{p}_2 = X_2 / n_2$. $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$. 99% confidence interval on Δ is

$[\hat{\Delta} \pm 2.58 \sqrt{\hat{p}_1 (1 - \hat{p}_1)/n_1 + \hat{p}_2 (1 - \hat{p}_2)/n_2}\ ]$.

b) [-0.07 – 0.0456, - 0.07 + .0456]

c) These samples were taken and the corresponding interval determined in such a way that repeated sampling would determine intervals which contain the true population difference in population proportions (first proportion minus the second proportion) for 99% of all repetitions.

d) 8295, using the "worst case", for which $p_1 = p_2 = 0.5$.

9) a) Reject H_0 for Level of Significance = $P(T > 25.5 \mid \mu = 2)$

$= 1 - \Phi((25.5 - 18)/\sqrt{9} = 1 - \Phi(2.5) = 0.0062$

b) Use the Neyman-Pearson Lemma, with the likelihood for $\mu = \mu_0 > 2$ in the numerator, and the likelihood for $\mu = 0$ in the denominator. The test reduces to critical region $T \geq k$ for some k, equivalently $\overline{X} > k^*$ for some constant k^*.

c) Power = $1 - \Phi((25.5 - 27)/. \sqrt{9}\ ) = 1 - \Phi(-0.5) = \Phi(0.5) = 0.691$

10) a) Let $Q(\beta) = \Sigma (Y_i - \beta/x_i)^2$, so $\frac{\delta}{\delta\beta} Q(\beta) = \Sigma (- 1/x_i) (Y_i - \beta/x_i)$.. Setting this equal to zero and solving for β, we get $\hat{\beta}$ as given.

b) $\hat{\beta} = (\Sigma Y_i / x_i) /(\Sigma 1/x_i^2) = \beta + \Sigma (\varepsilon_i/x_i) / (\Sigma 1/x_i^2)$. Since $E(\varepsilon_i) = 0$ for each i, $E(\hat{\beta}) = \beta$.

c) $\text{Var}(\hat{\beta}) = \text{Var}(\Sigma (\varepsilon_i/x_i) / (\Sigma 1/x_i^2).) = (1 /(\Sigma 1/x_i^2))^2 \Sigma (1/x_i^2) \sigma^2 = \sigma^2 / (\Sigma 1/x_i^2)$.

d) $\hat{\beta} = 0.8$, $S^2 = 48.6/3 = 16.2$ 90% CI on β: 0.8 ± 4.89.

11) Let W = (# women on the jury) Under random sampling W has a hypergeometric distribution.

$$P(W \le 1) = P(W = 0) + P(W = 1) = \frac{\binom{12}{7}\binom{6}{0}}{\binom{18}{7}} + \frac{\binom{12}{6}\binom{6}{1}}{\binom{18}{7}}.$$

12) W = 11, $P(W \le 11 \mid H_0) = 2/\binom{9}{4} = 1/42$, so the p-value for a 2 sided test is 1/21.