BC203 Final Exam Crib Sheet

Probability

For any events A and B:

1. $P(A^c) = 1 - P(A)$
2. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
3. $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$
4. $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$
5. For disjoint $A_1, A_2, ..., A_n$ then $P(B) = \sum_i P(B|A_i)P(A_i)$
6. $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$
7. For disjoint $A_1, A_2, ..., A_n$ then $P(A_1|B) = P(B|A_1)P(A_1)/\sum_i P(B|A_i)P(A_i)$
8. A and B are independent if $P(A|B) = P(A)$ or $P(A \text{ and } B) = P(A)P(B)$
9. A and B are mutually exculsive if $P(A \text{ and } B) = 0$

Models

1. $X \sim \text{Bin}(n, \theta)$ where $P(X = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$
   $E[X] = n\theta$ and $Var[X] = n\theta(1-\theta)$
2. $X \sim \text{Poiss}(\lambda)$ where $P(X = k) = e^{-\lambda}\lambda^k/k!$
   $E[X] = \lambda$ and $Var[X] = \lambda$
3. $X \sim \text{Hyper}(N, n, C)$ where $P(X = k) = \binom{C}{k}\binom{N-C}{n-k} / \binom{N}{n}$
   $E[X] = Cn/N$ and $Var[X] = Cn(N-n)(N-c)/(N^2(N-1))$
4. $X \sim \text{Normal}(\mu, \sigma^2)$ where $P(Z > k)$ is given by the table
   $E[X] = \mu$ and $Var[X] = \sigma^2$

Standardize

$$Z = \frac{X - E[X]}{\sqrt{Var[X]}} \text{ or } Z = \frac{\overline{X}_n - E[\overline{X}_n]}{\sqrt{Var[\overline{X}_n]}}$$

Expected Values and Variances

For random variables X, Y and constants a, c:

1. $E[aX + c] = aE[X] + c$
2. $Var[aX + c] = a^2 Var[X]$
3. $Corr[X, Y] = Cov[X, Y]/\sqrt{Var[X]Var[Y]}$
4. $E[X + Y] = E[X] + E[Y]$
5. $Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$
6. $Var[X - Y] = Var[X] + Var[Y] - 2Cov[X, Y]$

Sample means

For independent random variables $X_1, X_2, \ldots, X_n$:

1. $E[\overline{X}_n] = E[X]$
2. $Var[\overline{X}_n] = Var[X]/n$

Boxplots

1. median = 50^{th} percentile – middle observation or average of two middle obsn's
2. upper hinge = 75^{th} percentile – $(3(n + 1)/4)^{th}$ observation (round up)
3. lower hinge = 25^{th} percentile – $((n + 1)/4)^{th}$ observation (round down)
4. IQR = 75^{th}% observation - 25^{th}% observation
5. upper fence = upper hinge + 1.5*IQR
6. lower fence = lower hinge - 1.5*IQR

Rates

Crude rate = number of events/total population
If the statum specific rate is r_i and the relative frequency of the population is w_i, then for k strata the Crude rate is a weighted average,

$$cr = \sum_{i=1}^{k} r_i w_i$$

1. Direct adjustment uses the standard population relative frequencies (w_i).
2. Indirect adjustment uses the standard population rates (r_i).

Chi-Square test statistic for 2x2 tables

$$T^* = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2(df=1)$$

where O_i's are the observed cell counts (a,b,c, or d in our notation) and the E_i are the expected cell counts obtained by multipling the row and colum totals and then dividing by the the grand total N. here the 'a' count is assumed to be the top left cell count, 'b' the top right cell count, 'c' the bottom left cell count, and 'd' the bottom right cell count.

McNemar's test statistic for 2x2 tables

$$T^* = \frac{(b-c)^2}{(b+c)} \sim \chi^2(df=1)$$

One Way ANOVA

k groups each with a mean $\overline{x_i}$ and sample variance s_i^2 and n_i observations in each group where $i = 1, ..., k$; The total number of observations is $n = \sum_{i=1}^{k} n_i$ with a grand mean of $\overline{x}$.

$$T^* = \frac{s_b^2}{s_w^2} = \frac{\sum_{i=1}^{k} n_i(\overline{x_i} - \overline{x})^2/(k-1)}{\sum_{i=1}^{k}(n_i - 1)s_i^2/(n-k)} \sim F(df_1 = k-1, df_2 = n-k)$$

Additions

1. CI for log odds ratio in a 2x2 table (be sure this is the 'correct' odds ratio)

$$log\left(\frac{ad}{bc}\right) \pm Z_{\alpha/2}\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

2. Sample variance

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

3. Binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Power and Sample Size

1. Normal Mean – One Sample

$$N = \frac{\sigma^2 (Z_{\alpha/2} + Z_\beta)^2}{(\mu_0 - \mu_A)^2}$$

2. Proportion – One Sample

$$N = \frac{\left(Z_{\alpha/2}\sqrt{\theta_0(1-\theta_0)} + Z_\beta\sqrt{\theta_A(1-\theta_A)}\right)^2}{(\theta_0 - \theta_A)^2}$$

3. Normal Mean – Two Sample ($r = n_1/n_2$ is the allocation ratio of sample sizes)

$$n_1 = \frac{(Z_{\alpha/2} + Z_\beta)^2}{(\mu_{d0} - \mu_{dA})^2}(\sigma_1^2 + r\sigma_2^2)$$

CI Length and Sample Size

1. Normal Mean – One Sample

$$N = \frac{4\sigma^2 \left(Z_{\alpha/2}\right)^2}{L^2}$$

2. Proportion – One Sample

$$N = \frac{4\theta(1-\theta)\left(Z_{\alpha/2}\right)^2}{L^2}$$

3. Normal Mean – Two Sample ($r = n_1/n_2$ is the allocation ratio of sample sizes)

$$n_1 = \frac{4(\sigma_1^2 + r\sigma_2^2)\left(Z_{\alpha/2}\right)^2}{L^2}$$