Math 362, Problem set 5

Due 3/16/10

1. (3.7.6) Another estimating chi-square: Let the result of a random experiment be classifies as one of the mutually exclusive and exhaustive ways A_1, A_2, A_3 and also as one of the mutually exclusive and exhaustive ways B_1, B_2, B_3, B_4. Two hundred independent trials of the experiment result in the following data

	B_1	B_2	B_3	B_4
A_1	10	21	15	6
A_2	11	27	21	13
A_3	6	19	27	24

Test, at the 0.05 significance level the hypothesis of independence of the A and B attribute, namely $H_0 : \mathbb{P}(A_i \cap B_i) = \mathbb{P}(A_i)\mathbb{P}(B_i)$, $i = 1, 2, 3$ and $j = 1, 2, 3, 4$ against the alternative of dependence.

Answer:

We estimate $\mathbb{P}(A_1) = \frac{10+21+15+6}{200} = \frac{13}{50}$. Likewise we estimate:

$$\mathbb{P}(A_2) = \frac{18}{50} \qquad \mathbb{P}(A_3) = \frac{19}{50} \qquad \mathbb{P}(B_1) = \frac{27}{200}$$
$$\mathbb{P}(B_2) = \frac{67}{200} \qquad \mathbb{P}(B_3) = \frac{63}{200} \qquad \mathbb{P}(B_4) = \frac{43}{200}$$

Note that, really, we've only estimated 5 probabilities, because $\mathbb{P}(A_3)$ is determined once $\mathbb{P}(A_2)$ and $\mathbb{P}(A_1)$ is estimated, and likewise for $\mathbb{P}(B_4)$.

Let n_{ij} denote the number of events in $A_i \cap B_j$ (so $n_{23} = 21$). We compute

$$\sum \frac{(n_{ij} - 200\mathbb{P}(A_i)\mathbb{P}(B_j))^2}{200\mathbb{P}(A_i)\mathbb{P}(B_j)} \approx 12.941.$$

Since we estimated 5 parameters, we compare to a $\chi^2(12 - 1 - 5) = \chi^2(6)$ random variable. Since $12.941 > 12.592$, we reject H_0.

2. (5.8.5) Determine a method to generate random observations for the following pdf: $f(x) = 4x^3$ for $0 < x < 1$, zero elsewhere.

3. (5.8.18) For $\alpha > 0$ and $\beta > 0$, consider the following accept/reject algorithm:

   (1) Generate U_1 and U_2 iid uniform(0,1) random variables. Set $V_1 = U_1^{1/\alpha}$ and $V_2 = U_2^{1/\beta}$.

   (2) Set $W = V_1 + V_2$. If $W \le 1$, set $X = V_1/W$, else goto Step(1).

   (3) Deliver X.

   Show that X has a beta distribution with parameters α and β.

   Note/Hint: The analysis is quite similar to the analysis of the algorithm we did in class. That is, look for the cdf $\mathbb{P}(X \le x)$ and note that it is some conditional probability.

   Answer: Our goal is to show that $f_X(x) = Cx^{\alpha-1}(1-x)^{\beta-1}$ for some constant C. We note that

$$\begin{aligned} F_X(x) = \mathbb{P}(X \le x) &= \mathbb{P}(V_1/W \le x | W \le 1) \\ &= \frac{\mathbb{P}(V_1 \le xW, W \le 1)}{\mathbb{P}(W \le 1)} \\ &= c\mathbb{P}(V_1 \le xW, W \le 1), \end{aligned}$$

   for some constant C. We find the joint distribution of V_1 and W as follows. First we find that $f_{V_1}(v_1) = \alpha v_1^{\alpha-1}$ for $0 < v_1 < 1$, and likewise $f_{V_2}(v_2) = \alpha v_1^{\alpha-1}$. Noting that $V_2 = W - V_1$, we have that the Jacobian of the transformation $V = V_1$ and $W = V_1 + V_2$ is 1. Thus the joint distribution of $f_{V_1,W}(v,w) = \alpha\beta v^{\alpha-1}(w-v)^{\beta-1}$ for $0 < v < 1$, and $v < w < v+1$. We have that

$$F_X(x) = c\int_{w=0}^{1}\int_{v=0}^{wx} \alpha\beta v^{\alpha-1}(w-v)^{\beta-1} dvdw$$

   Taking the derivative and applying the fundamental theorem of calculus, we have

$$f_X(x) = c\int_{w=0}^{1} \alpha\beta(wx)^{\alpha-1}(w-wx)^{\beta-1}dw = x^{\alpha-1}(1-x)^{\beta-1}c\int_{w=0}^{1} \alpha\beta w^{\alpha+\beta-2}dw$$

   Since this is of the proper form, the normalizing constant must be correct and this is a beta distribution as desired.

4. (5.9.1) Let $x_1, \ldots, x_n$ be the values of a random sample. A bootstrap sample $\mathbf{x}^{*'} = (x_1^*, \ldots, x_n^*)$ is a random sample of $x_1, \ldots, x_n$ drawn with replacement.

   (a) Show that $\mathbb{E}[x_i^*] = \bar{x}$

   (b) If n is odd, show that median $\{x_i^*\} = x_{((n+1)/2)}$

   (c) Show that $\mathrm{Var}(x_i^*) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Notes: There was a part (a) on the book problem, which I cut as it's obvious but hard to write up nicely. It asks you to show that the x_i^* are independent (which they are because I chose with replacement) and have the CDF $\hat{F}_n$ which is $\hat{F}_n(x) = \frac{1}{n} \times (\#\text{ of } x_i < x)$, which is obvious because I select the x_i^* uniformly from $x_1, \ldots, x_n$. Note that in (c), you have $\frac{1}{n}$ and not $\frac{1}{n-1}$ because $\bar{x}$ is the *real* mean of the x_i^* as opposed to the sample mean of the x_i^*.

Answer:

$$\mathbb{E}[x_i^*] = \sum x_i \mathbb{P}(x_i^* = x_i) = \frac{1}{n} \sum x_i = \bar{x}.$$

For (b), since half of the values are above, and half are below $x_{(n+1)/2}$, this is the median as desired.

For (c), we have

$$\mathrm{Var}(x_i^*) = \mathbb{E}[(x_i^* - \mathbb{E}[x_i^*])^2] = \frac{1}{n} \sum (x_i^* - \bar{x})^2.$$

5. (5.9.2) Let $X_1, \ldots, X_n$ be a random sample from a $\Gamma(1, \beta)$ distribution.

   (a.) Show that the confidence interval $(2n\bar{X}/(\chi^2_{2n})^{(1-(\alpha/2))}, 2n\bar{X}/(\chi^2_{2n})^{\alpha/2})$ is an exact $(1-\alpha)100\%$ confidence interval for β.

   Notation note: $(\chi^2_{2n})^{(1-(\alpha/2))}$ is the books (rather confusing) way of saying 'the number such that $\mathbb{P}(\chi^2(2n) < (\chi^2_{2n})^{(1-(\alpha/2))}) = 1 - \alpha/2$; for instance for $\alpha = .5$ and $n = 4$, so that $2n = 8$, we have (by our chart) $(\chi^2_8)^{(1-(\alpha/2))} = 17.535$. Unfortunately the notation makes the problem look harder than it is.

   (b.) Using part (a), show that the 90% confidence interval discussed in Example 5.9.1 is $(64.99, 136.69)$.

   Answer: For (a) note that $2n\bar{X}$ has a $\Gamma(n, 2\beta)$ distribution and $\frac{2n\bar{X}}{\beta}$ has a $\Gamma(n, 2) = \chi^2(2n)$ distribution. Thus:

$$\mathbb{P}((\chi^2_{2n})^{\alpha/2} \leq \frac{2n\bar{X}}{\beta} \leq (\chi^2_2 n)^{1-\alpha/2}) = 1 - \alpha.$$

   Solving the interval for β, we get the desired answer.

   For (b), we get that $(\chi^2_{40})^{.05} = 26.5$ and $(\chi^2_{40})^{.95} = 55.8$ (sadly from the internet). Then we compute to see the hypothesized interval.

6. (5.9.10) Let z^* be drawn at random from the discrete distribution which has mass n^{-1} at each point $z_i = x_i - \bar{x} + \mu_0$, where $(x_1, \ldots, x_n)$ is the realization of a random sample. Determine $\mathbb{E}[z^*]$ and $\mathrm{Var}(z^*)$.

   Answer:

$$\mathbb{E}[z^*] = \frac{1}{n}\sum(x_i - \bar{x} + \mu_0) = \mu_0.$$

Likewise

$$\mathrm{Var}(z^*) = \frac{1}{n}\sum(x_i - \bar{x} + \mu_0 - \mu_0)^2 = \frac{1}{n}\sum(x_i - \bar{x})^* = \mathrm{Var}(x^*).$$

where x^* is as in problem 3.