Dept Copy

Department of Statistics
University of Wisconsin-Madison
PhD Qualifying Exam Part I
September 2, 2008

12:30–4:30pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do a total of THREE (3) problems.
- Each problem must be done in a separate exam book.
- Please turn in THREE (3) exam books.
- Please write your code name and **NOT** your real name on each exam book.

2. A sample $X_1, X_2, \ldots, X_n$ represents independent and identically distributed random draws from a population with support $\mathcal{A} = \{1, 2, \ldots, A\}$ and probability masses $w = (w_1, w_2, \ldots, w_A)$. A is much larger than n. Define $1[\cdot]$ to be an indicator function. Consider counts $Y_a = \sum_{i=1}^{n} 1[X_i = a]$ for $a \in \mathcal{A}$, and also a Dirichlet prior distribution for w, the density of which is

$$p(w) = \frac{\Gamma(\sum_{a=1}^{A} \theta_a)}{\prod_{a=1}^{A} \Gamma(\theta_a)} \prod_{a=1}^{A} w_a^{\theta_a - 1}$$

for non-negative masses w_a summing to one, and for hyperparameters $\theta_a > 0$. (The density is usually considered with respect to Lebesgue measure on the $A-1$ dimensional space of values for $w_1, w_2, \ldots, w_{A-1}$, rather than the constrained space in which all A masses sum to unity, but the distinction is not critical here.) Recall also that $\Gamma(b) = \int_0^\infty x^{b-1} \exp(-x)\, dx$ for $b > 0$.

(a) For $x_i \in \mathcal{A}$, find $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | w)$.

(b) For possible counts y_a, find $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_A = y_A | w)$.

(c) Suppose $\theta_a = \theta / A$ for all $a \in \mathcal{A}$ and for a fixed $\theta > 0$. Find the unconditional probability $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_A = y_A)$.

(d) i. Write the probability in part (c) in terms of counts $c_j = \sum_{a=1}^{A} 1[y_a = j]$ for $j = 0, 1, \ldots, n$. For instance c_0 would be the number of population elements not observed in the sample.

ii. Show that the associated random vector $(C_1, C_2, \ldots, C_n)$ has the probability mass function

$$P(C_1 = c_1, \ldots, C_n = c_n) = \frac{A! \{\Gamma(\theta/A)\}^{-K}}{(A-K)!} \frac{n!\,\Gamma(\theta)}{\Gamma(\theta + n)} \prod_{j=1}^{n} \frac{1}{c_j!} \left\{ \frac{\Gamma(\theta/A + j)}{j!} \right\}^{c_j}$$

where $K = \sum_{j=1}^{n} c_j$ and we are restricted to counts $c_j \geq 0$ such that $\sum_{j=1}^{n} j c_j = n$.

(e) Derive the limit of the distribution above in ii of part (d) as $A \to \infty$. You may use the fact that $\Gamma(b+1) = b\Gamma(b)$ for $b > 0$.

4. Suppose that $\{X_i\}_{i=1}^n$ are i.i.d. $\sim$ Unif$(-\theta, \theta)$ for some unknown parameter $\theta > 0$. To estimate θ, we consider the estimators $S_n = \{X_{(n)} - X_{(1)}\}/2$ and $T_n = \sqrt{(3/n)\sum_{i=1}^n X_i^2}$, where $X_{(1)} = \min\{X_1, \ldots, X_n\}$ and $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

   (a) Find the limiting distribution of $n(X_{(1)} + \theta, \theta - X_{(n)})$ as $n \to \infty$.

   (b) Determine the limiting distribution of $n(S_n - \theta)$.

   (c) Indicate how an asymptotic confidence interval for θ based on S_n can be constructed.

   (d) Determine the limiting distribution of $\sqrt{n}(T_n - \theta)$.

   (e) Which of S_n or T_n is preferable for estimating θ? Explain.

   (f) Is there another estimate of θ that is asymptotically more accurate than S_n and T_n? If there is, find it and prove its superiority. If not, explain why not.