MATH 281C – Qualifying Exam 2007

Problem 1. Let $X_1, \ldots, X_n$ be i.i.d. with the exponential distribution with rate $\lambda > 0$, with cumulative distribution function

$$P(X_i \le x) = 1 - e^{-\lambda\, x}, \qquad x > 0$$

1. What is the distribution of $n \min(X_1, \ldots, X_n)$?

2. Show that
$$\frac{\max(X_1, \ldots, X_n)}{\log(n)} \to^P \frac{1}{\lambda}, \qquad n \to \infty$$
where the convergence is in probability.

3. Show that
$$\max(X_1, \ldots, X_n) - \frac{\log(n)}{\lambda}$$
converges in distribution as $n \to \infty$ and compute the cumulative distribution function of the limiting distribution.

Problem 2. Let $X_1, \ldots, X_m$ be i.i.d. with the exponential distribution with rate $\lambda > 0$ and $Y_1, \ldots, Y_n$ be i.i.d. with the exponential distribution with rate $\mu > 0$. Consider the Mann-Whitney U statistic:

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$$

where $D_{ij} = 1$ if $X_i > Y_j$ and $D_{ij} = 0$ otherwise. Compute $E(U)$ and $\mathrm{var}(U)$ as functions of λ and μ.

Problem 3. Suppose we observe the following two samples:

```
(X): 1.27  -0.03   0.86   4.20  -0.92   0.71  -0.13
(Y): 0.38   0.80  -0.08   0.18  -2.03
```

1. Perform a two-sided median test at the 5% level (use the normal approximation).

2. Perform a two-sided Wilcoxon rank sum test at the 5% level (use the normal approximation).

Note: $P(\mathcal{N}(0,1) > 1.96) \approx 2.5\%$

Question 4. (a) Define completeness in the context of a statistical estimation problem.

(b) A one-parameter subproblem of the usual two-parameter normal problem is gotten by fixing the coefficient of variation $\left(\frac{\sigma}{\mu} = c_0 \text{ (known)}\right)$. Show that this problem is an exponential family, and that i.i.d. data $X_1, \ldots, X_n$ $(n \geq 2)$ can be reduced to a two-dimensional sufficient statistic T. Show how the usual criterion for proving completeness in exponential families fails for T in this case. Prove or disprove that T is in fact complete.

(c) State the Lehmann-Scheffé theorem.

(d) $X_1, X_2 \sim$ iid geometric(p) $\left(P[X_i = x] = (1-p)^{x-1} p \, ; \right.$

$\left. x = 1, 2, \ldots \, ; \; p \in (0,1) \right)$.

Find (with justification) UMVUEs for $\frac{1}{p}$ and p.

Question 5. $X_1, \ldots, X_n \sim$ iid $N(\theta, 1)$. Find an unbiased estimator of θ (randomized if you wish) that is consistent but not L_2-consistent. Is it $\sqrt{n}$-consistent?

Question 6. (a) Outline the Bayesian estimation framework. You will have to define symbols for parameter/s, data, and prior and posterior distributions.

(b) What is meant by the "Bayes procedure" for a given loss function? Show that if the loss is squared error the Bayes procedure for an estimation problem is to simply take the posterior expectation.

(c) X and Y have some joint distribution. Prove the 2nd-order iterated expectation formula:

$\operatorname{var} X = E \operatorname{var}[X|Y] + \operatorname{var} E[X|Y]$, using the 1st-order one: $EX = E[E[X|Y]]$. (It is notationally less messy to write a symbol "$E_Y X$" in place of the usual $E[X|Y]$.)

(d) What is the intuitive idea behind the rough statement that a posterior variance "should be" less than a prior variance?

(e) Find a Bayesian estimation problem in which the posterior variance is less than the prior variance with probability 1.

Find another one in which this happens with probability less than 1. Show that there is no problem in which it can happen with probability 0.

Question 7. (a) Define the generalized likelihood ratio test (GLRT) in a "regular" parametric setting, and state the big theorem on the asymptotic distribution of the generalized likelihood ratio statistic. (Skip the details on the actual regularity conditions.)

(b) $X_i \sim_{\text{indep}} \text{binomial}(n_i, p_i)$; $i = 1, 2$.

Show that as $n_1, n_2 \to \infty$ the GLRT for $H_0: p_1 = p_2$ is asymptotically the same as the chi-square test with a statistic of the form $\sum_{i=1}^{2} (\text{observed}_i - \text{expected}_i)^2 / \text{expected}_i$ (written in the style of a non-rigorous undergraduate statistics text!). How many degrees of freedom does the chi-square null distribution have?

(c) Give a good asymptotic confidence interval for $p_1 - p_2$. What would be the test procedure for $H_0: p_1 = p_2$ based on that interval through the "duality principle"? Is the basic GLRT one-sided or two-sided? How can it be adapted to the other form?