Math 281C Homework 8 (Last)

Due: 5:00pm, June 3rd

1. Let $X_1, \ldots, X_n$ be i.i.d. random variables from the beta distribution with pdf $\theta x^{\theta-1}\mathbb{1}(0 < x < 1)$, and independent of X_i's, let $Y_1, \ldots, Y_m$ be i.i.d. from the beta distribution with pdf $\mu x^{\mu-1}\mathbb{1}(0 < x < 1)$. For testing $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$, find the forms of the LR test, Wald's test, and Rao's score test.

2. Suppose that $X = (X_1, \ldots, X_k)^T$ has the multinomial distribution with a known size n and an unknown probability vector $\mathbf{p} = (p_1, \ldots, p_k)^T$. Consider the problem of testing

$$H_0 : \mathbf{p} = \mathbf{p}_0 \quad \text{versus} \quad H_1 : \mathbf{p} \neq \mathbf{p}_0,$$

where $\mathbf{p}_0 = (p_{01}, \ldots, p_{0k})^T$ is a known probability vector, that is, $\sum_{j=1}^k p_{0j} = 1$ and $p_{0j} \in (0,1)$. Find the forms of Wald's test and Rao's score test.

3. Suppose that $X_1, \ldots, X_n$ are iid from the Beta$(\mu, 1)$ distribution, and $Y_1, \ldots, Y_m$ are iid from the Beta$(\theta, 1)$ distribution. Assume that X_i's and Y_j's are independent.
   (i) Find the LRT for testing $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$.
   (ii) Show that the test in part (i) can be based on the statistic

$$T = \frac{\sum_{i=1}^n \log X_i}{\sum_{i=1}^n \log X_i + \sum_{j=1}^m \log Y_j}.$$

   (iii) Find the distribution of T when H_0 is true, and show how to get a test of size $\alpha = 0.1$.

4. The following example comes from genetics. There is a particular characteristic of human blood (the so-called MN blood group) that has three types: M, MN, and N. Under idealized circumstances known as Hardy-Weinberg equilibrium, these three types occur in the population with probabilities $p_1 = \pi_M^2$, $p_2 = 2\pi_M\pi_N$ and $p_3 = \pi_N^2$, respectively, where π_M is the frequency of the M allele in the population and $\pi_N = 1 - \pi_M$ is the frequency of the N allele.
We observe data $X_1, \ldots, X_n$, where X_i has one the three possible values: $(1,0,0)^T$, $(0,1,0)^T$, or $(0,0,1)^T$, depending on whether the ith individual has the M, MN, or N blood type. Denote the total number of individuals of each of the three types by n_1, n_2, and n_3; that is, $n_j = n\bar{X}_j$ for each j.
If the value of π_M were known, then we already know that the Pearson χ^2 statistic converges in distribution to a chi-square distribution with 2 degrees of freedom. However, in practice we usually don't know π_M. Instead, we estimate it using the maximum likelihood estimator $\widehat{\pi}_M = (2n_1 + n_2)/(2n)$. By the invariance principle of maximum likelihood estimation, this gives $\widehat{p} = (\widehat{\pi}_M^2, 2\widehat{\pi}_M\widehat{\pi}_N, \widehat{\pi}_N^2)^T$ as the maximum likelihood estimator of $p = (p_1, p_2, p_3)^T$.
   (a) Define $Z_n = \sqrt{n}(\bar{X} - \widehat{p})$. Use the delta method to derive the asymptotic distribution of $D^{-1/2}Z_n$, where $D = \mathrm{diag}(p_1, p_2, p_3)$.
   (b) Define $\widehat{D}$ to be the diagonal matrix with entries $\widehat{p}_1, \widehat{p}_2, \widehat{p}_3$ along its diagonal. Derive the asymptotic distribution of $\widehat{D}^{-1/2}Z_n$.
   (c) Derive the asymptotic distribution of the Pearson chi-square statistic

$$\chi^2 = \sum_{j=1}^n \frac{(n_j - n\widehat{p}_j)^2}{n\widehat{p}_j}.$$