Homework 8

36-705

Due: Thursday October 29 by 3pm.

1. Suppose that we have d hypotheses, which are ordered a-priori (i.e. without looking at the data). We use the following procedure:

   (a) If $p_1 \geq \alpha$, we accept all the null hypotheses and stop, else we reject H_{01} and continue to the next step.

   (b) If $p_2 \geq \alpha$, we accept $H_{02}, \ldots, H_{0d}$ and stop, otherwise we reject H_{02} and continue.

   (c) $\vdots$

   Suppose that the p-values are independent: does this procedure control the FWER at α? Would this still be the case if the p-values were dependent? Is this procedure more/less/incomparable in power to the Bonferroni procedure? Explain your answer.

2. Suppose that a scientist plans to do 1000 gene association tests. Before looking at the data she orders the hypotheses according to her belief, placing "promising" genes first and then less promising ones and so on. She then does the following:

   (a) She computes p-values for every hypothesis.

   (b) She rejects the first null hypothesis if the p-value $\leq \alpha/2$, the second null if the p-value is $\leq \alpha/4$ and so on.

   Does her procedure control the FWER at a reasonable level? If not, explain why not and if yes, give a proof.

3. Let $X_1, \ldots, X_n \sim p(x; \theta)$. Let T_n be a test statistic for testing $H_0 : \theta = \theta_0$. Assume that T_n has a continuous distribution and that we reject when T is large. Recall that the p-value is

   $$p = P_{\theta_0}(T_n(X_1^*, \ldots, X_n^*) > T(X_1, \ldots, X_n))$$

   where $X_1^*, \ldots, X_n^* \sim p(x; \theta_0)$. Show that, when H_0 is true, p has a uniform (0,1) distribution.

   Hint: Note that we can write $p = \int_{T(X_1,\ldots,X_n)}^{\infty} f_0(t)dt$ where $f_0(t)$ is the density of $T(X_1, \ldots, X_n)$ under H_0.

4. Suppose that $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$ so the density is

   $$p(x; \lambda) = \frac{1}{\lambda} \exp(-x/\lambda).$$

   (a) Show that

   $$Q(X_1, \ldots, X_n, \lambda) = \frac{2\sum_{i=1}^n X_i}{\lambda}$$

is a valid pivot. You can use the fact that the sum of exponentials has a Gamma distribution. Use this pivot to construct a $(1-\alpha)$-confidence interval for λ.

(b) Construct the Wald confidence interval.

(c) Construct the asymptotic LRT confidence set.

5. A location family is a collection of shifted (mean 0) distributions, i.e. for some fixed distribution F with mean 0, we define:

$$\mathcal{P}_\theta = \{G : G(x) = F(x-\theta)\}\,.$$

Show that $\overline{X} - \theta$ is always a pivot for a location family.