541B Midterm 2 Solutions[1]

1. Question 1

Let $\beta_1, \ldots \beta_p$ be real numbers. Show that the following two conditions are equivalent.

- $\beta_1 = \cdots = \beta_p$
- For any $c_1, \ldots, c_p \in \mathbf{R}$ with $\sum_{i=1}^p c_i = 0$, we have
$$\sum_{i=1}^{p} c_i \beta_i = 0.$$

Solution. If the first condition holds, then $\sum_{i=1}^p c_i\beta_i = \beta_1 \sum_{i=1}^p c_i = 0$.

If the second condition holds, then fix any $1 \leq i < j \leq p$, and set $c_i = 1$, $c_j = -1$ and $c_k = 0$ for all other $k \in \{1, \ldots, p\}$. The second condition says $\beta_i - \beta_j = 0$, i.e. $\beta_i = \beta_j$, i.e. the first condition holds.

2. Question 2

Let $\mu \in \mathbf{R}$ and let $0 < \sigma < \infty$. Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables each with mean μ and variance σ^2. Let $h\colon \mathbf{R} \to \mathbf{R}$ be a function such that h' exists and is continuous. Let $\overline{X}_n := (X_1 + \cdots + X_n)/n$. Let $Y_n := h(\overline{X}_n)$.

Show that the jackknife estimator of the variance of Y_n converges almost surely to the same estimate of the variance you get by applying the Delta Method to Y_n.

Solution. Theorem 5.8 in the notes implies that the jackknife variance estimator V_n is consistent, i.e. (when $d = 1$), we have that
$$\frac{V_n}{\frac{1}{n}(h'(\mu))^2 \mathrm{Cov}(X_1, X_1)} = \frac{V_n}{\frac{1}{n}(h'(\mu))^2 \sigma^2}$$
converges almost surely to 1 as $n \to \infty$.

On the other hand, the Delta Method implies that $\sqrt{n}(h(\overline{X}_n) - h(\mu))$ converges in distribution to a mean zero Gaussian with variance $\sigma^2(h'(\mu))^2$ as $n \to \infty$. Consequently, the variance of $h(\overline{X}_n)$ is approximately $\sigma^2(h'(\mu))^2/n$, asymptotically agreeing with the jackknife variance estimate.

3. Question 3

Let $X_1, \ldots, X_n$ be a random sample from a Gaussian random variable with unknown mean $\mu \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$.

Fix $\mu_0 \in \mathbf{R}$. Suppose we want to test that hypothesis H_0 that $\mu = \mu_0$ versus the alternative H_1 that $\mu \neq \mu_0$.

Explicitly describe the rejection region of the generalized likelihood ratio test for this hypothesis.

Solution. For any $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$,
$$f_{\mu,\sigma^2}(x) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

Also, $\Theta = \{(\mu, \sigma)\colon \mu \in \mathbf{R}, \sigma > 0\}$ and $\Theta_0 = \{(\mu_0, \sigma)\colon \sigma > 0\}$.

As we computed in 541A, the MLE is the sample mean and sample variance (dividing by n instead of $n-1$), i.e. for any $x \in \mathbf{R}^n$, if we denote $\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$, $v := \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$

$$\sup_{(\mu,\sigma)\in\Theta} f_{\mu,\sigma^2}(x) = f_{\overline{x},v}(x).$$

Similarly,

$$\sup_{(\mu,\sigma)\in\Theta_0} f_{\mu,\sigma^2}(x) = f_{\mu_0,v'}(x).$$

Here $v' := \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2$. Therefore,

$$\begin{aligned} C &:= \{x \in \mathbf{R}^n \colon \sup_{(\mu,\sigma^2)\in\Theta} f_{\mu,\sigma^2}(x) \geq k \sup_{(\mu,\sigma^2)\in\Theta_0} f_{\mu,\sigma^2}(x)\} \\ &= \Big\{x \in \mathbf{R}^n \colon (v/v')^{n/2} \prod_{i=1}^n e^{-\frac{(x_i - \frac{1}{n}\sum_{j=1}^n x_j)^2}{2v} + \frac{(x_i-\mu_0)^2}{2v'}} \geq k\Big\} \\ &= \Big\{x \in \mathbf{R}^n \colon (v/v')^{n/2} \geq k\Big\} = \Big\{x \in \mathbf{R}^n \colon v/v' \geq k^{2/n}\Big\} \end{aligned}$$

Note that

$$\begin{aligned} v' &= \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x} + \overline{x} - \mu_0)^2 \\ &= \frac{1}{n}\sum_{i=1}^n [(x_i - \overline{x})^2 + (\overline{x} - \mu_0)^2] = v + \frac{1}{n}\sum_{i=1}^n (\overline{x} - \mu_0)^2 = v + (\overline{x} - \mu_0)^2 \end{aligned}$$

Using this fact, and that the inverse of $x \mapsto 1/(1+x)$ is $x \mapsto -1 + 1/x$,

$$\begin{aligned} C &= \Big\{x \in \mathbf{R}^n \colon \frac{v}{v + (\overline{x} - \mu_0)^2} \geq k^{2/n}\Big\} = \Big\{x \in \mathbf{R}^n \colon \frac{1}{1 + \frac{(\overline{x}-\mu_0)^2}{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2}} \geq k^{2/n}\Big\} \\ &= \Big\{x \in \mathbf{R}^n \colon \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2}{(\overline{x} - \mu_0)^2} - 1 \leq k^{-2/n} - 1\Big\} = \Big\{x \in \mathbf{R}^n \colon \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2}{(\overline{x} - \mu_0)^2} \leq k^{-2/n}\Big\}. \end{aligned}$$

4. Question 4

Let X_1, X_2, X_3 be i.i.d. continuous random variables such that X_1 has PDF $\{f_\theta \colon \theta \in \Theta\}$. Let W_1, W_2, W_3 be a bootstrap sample from X_1, X_2, X_3. Let Y denote the sample median of X_1, X_2, X_3. (That is, Y is the middle value among X_1, X_2, X_3, which is unique with probability one since the random variables are continuous.)

- Describe the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$.
- Describe the bootstrap estimator of Y.

Solution. Since X_1, X_2, X_3 are all distinct with probability one, we have

$$\mathbf{P}(W_1 = X_i, W_2 = X_j, W_3 = X_k \,|\, X_1, X_2, X_3) = (1/3)^3, \qquad \forall\, 1 \leq i, j, k \leq 3.$$

That is, in describing the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$, we may as well assume that $X_{(1)} = 1, X_{(2)} = 2, X_{(3)} = 3$, and W_1, W_2, W_3 are i.i.d. uniform in $\{1, 2, 3\}$. (We are satisfied with this description of the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$.)

Now, as covered e.g. in Exercise 2.19 in the notes, by considering Y which is the number of indices $1 \leq j \leq 3$ such that $W_j \leq X_{(i)}$, we have

$$\mathbf{P}(W_{(2)} \leq X_{(i)} \,|\, X_1, X_2, X_3) = \sum_{k=2}^{3} \binom{3}{k} p_i^k (1-p_i)^{n-k},$$

where $p_i = i/3$ for all $1 \leq i \leq 3$. (This follows since Y is a binomial random variable with parameters 3 and p_i.) That is,

$$\mathbf{P}(W_{(2)} \leq X_{(i)} \,|\, X_1, X_2, X_3) = \sum_{k=2}^{3} \binom{3}{k} (i/3)^k (1-i/3)^{n-k},$$

Therefore, for all $1 \leq i \leq 3$, we have

$$\begin{aligned}\mathbf{P}(W_{(2)} = X_{(i)} \,|\, X_1, X_2, X_3) &= \mathbf{P}(W_{(2)} \leq X_{(i)} \,|\, X_1, X_2, X_3) - \mathbf{P}(W_{(2)} \leq X_{(i-1)} \,|\, X_1, X_2, X_3) \\ &= \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1-i/3)^{n-k} - ((i-1)/3)^k (1-(i-1)/3)^{n-k} \Big).\end{aligned}$$

The bootstrap estimator of Y is then

$$\begin{aligned}\mathbf{E}[W_{(2)} \,|\, X_1, X_2, X_3] &= \sum_{i=1}^{3} X_{(i)} \mathbf{P}(W_{(2)} = X_{(i)} \,|\, X_1, X_2, X_3) \\ &= \sum_{i=1}^{3} X_{(i)} \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1-i/3)^{n-k} - ((i-1)/3)^k (1-(i-1)/3)^{n-k} \Big).\end{aligned}$$

5. Question 5

Let $X = (X_1, \ldots, X_n)$ be a random sample of size n from a family of distributions $\{f_\theta \colon \theta \in \Theta\}$. Fix $\theta_0 \in \Theta \subseteq \mathbf{R}$. Suppose we test the hypothesis H_0 that $\{\theta = \theta_0\}$ versus the alternative $\{\theta \neq \theta_0\}$. Suppose the Fisher information of X_1 exists, is finite and nonzero, and the MLE exists, is unique, and is consitent.

Let $\lambda(X) := \frac{\sup_{\theta \in \Theta} f_\theta(X)}{\sup_{\theta \in \Theta_0} f_\theta(X)}$ denote the generalized likelihood ratio statistic.

Give a sketch of the proof of the following: If H_0 is true, then $2 \log \lambda(X)$ converges in distribution as $n \to \infty$ to a chi-squared random variable with one degree of freedom.

(This Theorem, which we sketched in the notes, is known as Wilks' Theorem.)

(Hint: Perform a second order Taylor expansion of the log-likelihood $\ell(\theta)$ at the point Y where $Y = Y_n$ is the MLE of θ, and recall that $\mathbf{E}_{\theta_0} \ell''(\theta_0) = -n I_{X_1}(\theta_0)$.)

(You are allowed in your proof sketch to ignore technicalities, e.g. you can ignore error terms in the Taylor expansion, and you can freely assume that the MLE is consistent. You can also freely use that $\sqrt{n}(Y_n - \theta_0)$ converges in distribution to a mean zero Gaussian as $n \to \infty$, with variance $1/I_{X_1}(\theta_0)$.)

Solution. Recall that $\ell(\theta) := \log f_\theta(x)$. Suppose we expand $\ell(\theta)$ in a Taylor series around the random point Y, i.e. assume there exists $h \colon \mathbf{R} \to \mathbf{R}$ such that $\lim_{z \to 0} \frac{h(z)}{z^2} = 0$ and, for all $\theta_0 \in \mathbf{R}$,

$$\ell(\theta_0) = \ell(Y) + \ell'(Y)(\theta_0 - Y) + (1/2)\ell''(Y)(\theta_0 - Y)^2 + h(Y - \theta_0).$$

By definition of Y, $\ell'(Y) = 0$. Since $2\log\lambda(X) = -2\ell(\theta_0) + 2\ell(Y)$, we rearrange the equality to get

$$2\log\lambda(X) \approx -\ell''(Y)(\theta_0 - Y)^2.$$

As mentioned in the hint, $\mathbf{E}_{\theta_0}\ell''(\theta_0) = -I_X(\theta_0) = -nI_{X_1}(\theta_0)$. From the hint, we may assume that, $Y = Y_n$ converges in probability to the constant θ_0 with respect to $\mathbf{P}_{\theta_0}$ as $n \to \infty$. So, we can approximate $\ell''(Y)$ by $\ell''(\theta_0) \approx -nI_{X_1}(\theta_0)$. That is,

$$2\log\lambda(X) \approx nI_{X_1}(\theta_0)(\theta_0 - Y)^2.$$

From the hint (the Theorem in the notes about the MLE converging in distribution), $\sqrt{n}(Y - \theta_0)$ converges in distribution to a mean zero Gaussian with variance $1/I_{X_1}(\theta_0)$ as $n \to \infty$. Therefore, $2\log\lambda(X)$ converges in distribution to a chi-squared random variable with one degree of freedom as $n \to \infty$.