First Term Exam STA 3000Y1Y
Friday December 6, 2013
10.00 – 13.00

Instructions: Answer all questions in the exam booklets. Be as precise as possible in your answers, stating clearly the results and assumptions you are using. Questions are each worth 25 points.

I gave 6 points for each part of 4-part questions and 8 points for each part of 3-part questions, with a +1 gratis on each.

1. Suppose $Y_0, \ldots, Y_n$ are distributed as a Poisson birth process, i.e. the conditional density of Y_{j+1}, given $Y_j = y_j$, is Poisson with mean θy_j:

$$f(y_{j+1} \mid y_j; \theta) = \frac{(\theta y_j)^{y_{j+1}}}{y_{j+1}!} \exp(-\theta y_j), \quad y_{j+1} = 0, 1, \ldots, \quad \theta > 0.$$

   (a) Assuming Y_0 follows a Poisson(θ) distribution, find the likelihood function for θ based on $(y_0, \ldots, y_n)$.

   (b) Give an expression for the minimal sufficient statistic for θ.

   (c) Give an expression for the maximum likelihood estimate, and describe how you might determine its distribution.

The likelihood function is constructed in the usual way for a Markov process: $L(\theta; y) = f(y_0; \theta) \prod_{j=1}^n f(y_j \mid y_{j-1}; \theta)$, leading to

$$L(\theta; y) \propto \theta^{s_0} \exp(-\theta s_1 + 1), \quad s_0 = \Sigma_0^n y_j, \quad s_1 = \Sigma_0^{n-1} y_j,$$

so the likelihood statistic is $(\Sigma_0^n y_i, y_n)$ or (s_0, s_1), or any one-to-one function of this. This is minimal because it determines the likelihood map; more formally $L(\theta; y)/L(\theta; y')$ is free of θ if and only if $s(y) = s(y')$, by properties of polynomials (θ^{s_0}), and exponentials ($\exp(\theta s_1)$). The maximum likelihood estimator is $\hat\theta = s_0/(s_1 + 1)$, obtained by solving $\ell'(\theta; y) = 0$, where $\ell(\cdot; y) = \log L(\cdot; y)$, and checking that the second derivative is negative. I thought at first that it might be possible to obtain the exact distribution of $\hat\theta$ using properties of the Poisson distribution, but this looks hard. Everyone gave the more obvious answer, that we could use any of our pivotal quantities, for example $(\hat\theta - \theta) j^{1/2}(\hat\theta)$, or $\sqrt{2[\ell(\hat\theta) - \ell(\theta)]}$ for inference, each of these being approximately normally distributed. On reflection, it's not clear that the usual asymptotics applies, because the observations are dependent. We need a central limit theorem for the score function, which can probably be established using the Markov property, and we need $j(\hat{\theta})/i(\theta)$ to converge to 1; this is not guaranteed in dependent data settings, but I haven't tried to work it our for this example. See BNC, Ch 9.3. These considerations did not affect the marking.

2. Suppose $Y_1, \ldots, Y_n$ are independent and identically distributed from a density $f(y;\theta), \theta \in \mathbb{R}$. Denote the log-likelihood function for the sample by $\ell(\theta; y)$, and the total expected Fisher information for θ by $i(\theta)$.

   (a) Using Taylor series expansions show that

   $$2\{\ell(\hat{\theta}) - \ell(\theta)\} = (\hat{\theta} - \theta)^2 i(\theta) + o_p(1),$$

   under the model $f(y;\theta)$, and describe the assumptions on the model needed to establish this. Here $\hat{\theta}$ is the maximum likelihood estimator, assumed to be the unique solution of the score equation $\ell'(\theta) = 0$.

This was pretty straightforward after the agony and ecstasy of HW 3. I was fairly lenient with the treatment of the remainder term, as long as it was sketched plausibly.

   (b) Assume now that $\theta = (\psi, \lambda)$ with $\psi \in \mathbb{R}$, $\lambda \in \mathbb{R}^k$. Define the *profile log-likelihood* function for ψ, and explain how it can be used for approximate inference about ψ.

ditto

   (c) Let $Y_1, \ldots, Y_n$ be independent observations from the gamma distribution with shape parameter α and mean μ_i, where $\mu_i^{-1} = \beta_0 + \beta_1 x_i$. Find an expression for the profile log-likelihood for α. Show that α is orthogonal to (β_0, β_1) with respect to expected Fisher information. What does this tell you about the variance of the limiting distribution of $\hat{\alpha}$?

Sorry I meant to give you the density for a single observation from the gamma: $\Gamma(\alpha)^{-1}(\alpha/\mu_i)^\alpha y_i^{\alpha-1} \exp(-y_i\alpha/\mu_i)$. The constrained maximum likelihood estimators $\hat{\beta}_{0,\alpha}$ and $\hat{\beta}_{1,\alpha}$ cannot be given explicitly, but they solve $\sum y_i = \sum \mu_i(\hat{\beta}_0, \hat{\beta}_1)$ and $\sum y_i x_i = \sum x_i \mu_i(\hat{\beta}_0, \hat{\beta}_1)$, as is usual in generalized linear models, and turn out not to depend on α. I think this implies directly that α and β are orthogonal parameters, but calculating $i(\theta)$ is easy enough as well. The parameters being orthogonal means the limiting variance of $\hat{\alpha}$ is $i_{\alpha\alpha}^{-1}(\theta)$, with no need to worry about adjustments for β. And means the covariance of the limiting joint distribution of $\hat{\alpha}$, $\hat{\beta}$ is 0, i.e. $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically independent. Orthogonality alone doesn't imply that the asymptotic variance of $\hat{\alpha}$ is free of β, although in this example, with the stronger orthogonality $\hat{\beta}_\alpha = \hat{\beta}$, that is indeed the case.

3. Let $Y_1, \ldots Y_n$ be i.i.d. from the density

$$f(y;\theta) = \exp\{-(y-\theta)\}, \quad y > \theta.$$

To test the hypothesis $H_0 : \theta = 1$ against $H_1 : \theta > 1$ a critical region of the form $w_\alpha = \{y_{(1)} > c\}$ is proposed, where $y_{(1)} = \min(y_1, \ldots y_n)$.

(a) Determine c so this critical region has size α.

(b) Sketch the power function of this test.

(c) Justify the choice of the critical region above, or suggest a better one.

The joint density is

$$f(\underline{y};\theta) = \exp(\Sigma y_i - n\theta) 1\{y_{(1)} > \theta\},$$

where $1\{\cdot\}$ is the indicator function. From this we see that $y_{(1)}$ is minimal sufficient, which essentially answers (c). The value of c is determined from the distribution of $Y_{(1)}$, which has $1 - F_{Y_{(1)}}(y_{(1)}) = \exp\{-n(y_{(1)} - \theta)\} 1\{y_{(1)} > \theta\}$, since $\Pr(Y_{(1)} > c) = \prod \Pr(Y_i > c)$. Substituting $\theta_0 = 1$ gives the critical region: $c_\alpha = 1 - n^{-1}\log\alpha$. Substituing θ gives the power function, which is exponentially increasing, until it hits 1 and stays there, at $\theta =?$ (your task).

4. Suppose that $(Y_{1i}, Y_{2i}), i = 1, \ldots, n$ follow the bivariate normal distribution with mean (θ_1, θ_2) and identity covariance matrix: the joint distribution of the sample $(y_1, y_2) = (y_{11}, \ldots, y_{1n}, y_{21}, \ldots, y_{2n})$ is then

$$f(y_1, y_2; \theta) = \left(\frac{1}{2\pi}\right)^n \exp -\frac{1}{2}\{\sum (y_{1i} - \theta_1)^2 + \sum (y_{2i} - \theta_2)^2\}.$$

The parameter space is restricted to $\theta_1 \geq 0, \theta_2 \geq 0$.

(a) Show that the maximum likelihood estimators of θ_1 and θ_2 are given by

$$\hat{\theta}_1 = \max(\bar{y}_1, 0), \quad \hat{\theta}_2 = \max(\bar{y}_2, 0).$$

(b) Derive the form of the log-likelihood ratio statistic $w(\theta_0) = 2\{\ell(\hat{\theta}) - \ell(\theta_0)\}$ for testing $H_0 : \theta = \theta_0 = (0, 0)$. (It has a different expression for $(\bar{y}_1, \bar{y}_2)$ in each of the 4 quadrants of the plane.)

(c) By considering these four quadrants, argue that the distribution of $w(\theta_0)$, under H_0, is

$$\frac{1}{4}\delta_{\{0\}} + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2,$$

where $\delta_{\{0\}}$ is a point mass at 0.

This question was on an old comprehensive exam. It's the only example we did of a non-regular problem, where the maximum is on the boundary of the parameter space if the sample mean(s) are negative. Thus the exact distribution of the log-likelihood ratio statistic $w(\theta_0)$ is a mixture of χ^2's, with differing degrees of freedom (sometimes a point mass at 0 is identified with a χ_0^2). A more realistic example is the random effects model, the simplest version of which is

$$y_{ij} = \mu + b_i + \epsilon_{ij}, \quad j = 1, \ldots, m; i = 1, \ldots, k,$$

with $b_i \sim N(0, \sigma_b^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$, all independent of each other. The solutions of the score equations for σ_b^2 and σ^2 can lead, in some samples, to negative estimates of σ_b^2, so the maximum likelihood estimate is on the boundary of the parameter space, and the log-likelihood ratio statistic for σ_b^2 has a distribution with a point mass at 0, and follows a χ^2 distribution otherwise. The classic reference for this is Self and Liang (1987, *JASA*).

5. **Bonus Question (5 bonus points):** Assume that Y follows a distribution $f(y; \theta_1, \theta_2)$, for $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$. Prove that if $T_1 = t_1(Y)$ is sufficient for θ_1 when θ_2 is known, and $T_2 = t_2(Y)$ is sufficient for θ_2 when θ_1 is known, then (T_1, T_2) is sufficient for θ.

This is also from an old comprehensive exam. This solution thanks to David Soave uses the factorization theorem. First, fix $\theta_2 = \theta_2^0$. (This notation isn't strictly necessary, but makes things clearer.) Then by the first condition,

$$f(y; \theta_1, \theta_2^0) \propto h_1(t_1(y); \theta_1, \theta_2^0) g_1(y; \theta_2^0).$$

By the second condition,

$$g_1(y; \theta_2^0) = h_2(t_2(y); \theta_2^0) g_2(y).$$

Putting these together and allowing θ_2^0 to be arbitrary gives the result. It's also possible to get the result from the conditional distribution of y, given t_1, t_2, but I think this version is easier.