Stat 513
Emerson, Win 16
Homework #7
March 4, 2016

Written solutions to the homework problems are due on Friday, March 11, 2016 at the beginning of class.

As noted on the syllabus, copying of homework solutions is not allowed and, when detected, will be investigated as an infraction of the academy integrity policy of the University of Washington. While it is permissible to discuss problems with other students, TAs, or the instructor in order to learn how to solve a problem, your written solutions must be prepared without directly referencing any notes or solutions derived from other students or sources found on the internet.

REGULAR PROBLEMS

1. Suppose that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. with density
$$f_Y = e^{-(y-\mu)}\mathbf{1}_{(\mu,\infty)}$$
for $\mu \in (-\infty, \infty)$. Use Basu's theorem to show that $Y_{(1)}$ is independent of the sample variance s^2.

2. Suppose that for data $Y_1, Y_2, \ldots, Y_n$, conditional on the parameter p, have distribution as i.i.d. Bernoulli random variables with $Y_i|p \sim \mathcal{B}(1,p)$, $p \in (0,1)$. Further suppose prior distribution $p \sim \; Beta(\alpha, \beta)$.
   (a) Show that this prior distribution is the conjugate prior distribution.
   (b) Find the posterior distribution of p conditional on $\vec{Y} = (Y_1, \ldots, Y_n)^T$.
   (c) Find the Bayes estimator of p based on squared error loss. Can this estimator be expressed as a weighted average of the data and the prior distribution? If so, do so.
   (d) Can you find a closed form solution for the Bayes estimator of p based on absolute error loss. If so, do so. If not, explain why not.

3. Suppose that for data $Y_1, Y_2, \ldots, Y_n$, conditional on the hazard parameter λ, have distribution as i.i.d. exponential random variables with $Y_i|p \sim \mathcal{E}(\lambda)$, $\lambda \in (0, \infty)$. Further suppose prior distribution $\lambda \sim \; Gamma(\alpha, \beta)$.
   (a) Show that this prior distribution is the conjugate prior distribution.
   (b) Find the posterior distribution of λ conditional on $\vec{Y} = (Y_1, \ldots, Y_n)^T$.
   (c) Find the Bayes estimator of λ based on squared error loss. Can this estimator be expressed as a weighted average of the data and the prior distribution? If so, do so.

(d) Can you find a closed form solution for the Bayes estimator of λ based on absolute error loss. If so, do so. If not, explain why not.

4. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. gamma random variables with $Y_i \sim \Gamma(\gamma, \lambda)$ with $\gamma > 0$ known and $\lambda > 0$ the rate parameter (so $E[Y_i] = \gamma/\lambda$.

   (a) Find the most powerful level α test of the simple hypotheses $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda = \lambda_1 < \lambda_0$. Show that the test is a function of the sufficient statistic, and explicitly provide a definition of the critical value. Is the test unbiased? Is the test consistent?

   (b) Extend the test of part a to the test of the composite hypotheses $H_0 : \lambda \geq \lambda_0$ versus $H_1 : \lambda = \lambda_1 < \lambda_0$. Show that the power function is decreasing in λ. Is the test unbiased? Is the test consistent?

   (c) Define a two-sided $100(1 - \alpha)\%$ confidence interval for λ.

5. Suppose $Y_1, Y_2, \ldots, Y_n$ are independent random variables with $Y_i \sim (\mu_i, \sigma^2)$ and the distribution is otherwise unspecified. Further suppose that for known $x_1, \ldots, x_n$, $\mu_i = \beta_0 + \beta_1 x_i$. We are interested in testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.

   (a) First presuming that σ^2 is known, derive an approximate level α test of H_0 versus H_1, providing an explicit expression for the critical value. Be sure to explain the conditions under which your test will be a good approximation to a level α test.

   (b) Show that your test in part a corresponds to a likelihood ratio test (LRT) when the Y_i's satisfy $Y_i | x_i \sim \mathcal{N}(\mu_i, \sigma^2)$.

   (c) Now modify your test in part a to the case where σ^2 is unknown.

   (d) How would your test in part c differ from a test derived when the data were known to be normally distributed? Are the tests asymptotically equivalent?

   (e) Find an expression for a

6. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. Bernoulli random variables with $Y_i \sim \mathcal{B}(1, p)$, $p \in (0, 1)$.

   (a) Derive an expression for the upper $100(1 - \alpha)\%$ confidence bound for p using the exact distribution of $\vec{Y}$.

   (b) Suppose we observe $Y_i = 0$ for all $i = 1, \ldots, n$. Show that for large n, the bound found in part a is approximately $3/n$ when $\alpha = .05$ and approximately $3.69/n$ when $\alpha = .025$

MORE INVOLVED PROBLEMS

7. Suppose $Y_1, Y_2, \ldots, Y_{n_Y}$ are i.i.d. Bernoulli random variables with $Y_i \sim \mathcal{B}(1, p_Y)$ and $X_1, X_2, \ldots, X_{n_X}$ are i.i.d. Bernoulli random variables with $X_i \sim \mathcal{B}(1, p_X)$, with $p_Y, p_X \in (0, 1)$. Our target of inference is $\theta = p_y - p_x$, interested in approximate two-sided level α tests of $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ and $100(1 - \alpha)\%$ confidence intervals for θ.

(a) Provide an expression for the Wald test (including critical value).
Ans: For this regular probability model, we find the likelihood, log likelihood, score, and information in the full (unrestricted) model. For notational convenience, define $Y = \sum_{i=1}^{n} Y_i$ and $X = \sum_{i=1}^{n} X_i$.

$$L(\vec{p}|\vec{Y}, \vec{X}) = p_Y^Y (1 - p_Y)^{n_Y - Y} p_X^X (1 - p_X)^{n_X - X}$$
$$\mathcal{L}(\vec{p}|\vec{Y}, \vec{X}) = Y \log(p_Y) + (n_Y - Y) \log(1 - p_Y) + X \log(p_X) + (n_X - X) \log(1 - p_X)$$
$$\mathcal{U}_1(\vec{p}|\vec{Y}, \vec{X}) = \frac{\partial}{\partial p_Y} \mathcal{L}(\vec{p}) = \frac{Y}{p_Y} - \frac{n_Y - Y}{1 - p_Y} = \frac{Y - n_Y p_Y}{p_Y(1 - p_Y)}$$
$$\mathcal{U}_2(\vec{p}|\vec{Y}, \vec{X}) = \frac{\partial}{\partial p_X} \mathcal{L}(\vec{p}) = \frac{X - n_X p_X}{p_X(1 - p_X)} \qquad \text{(by symmetry)}$$
$$\vec{\mathcal{U}}(\hat{\vec{p}}) = \vec{0} \qquad \Rightarrow \qquad \hat{\vec{p}} = \begin{pmatrix} \overline{Y} \\ \overline{X} \end{pmatrix}$$
$$\mathbf{J}(\hat{\vec{p}}) = \left[Var(\hat{\vec{p}}) \right]^{-1} = \begin{pmatrix} \frac{p_Y(1-p_Y)}{n_Y} & 0 \\ 0 & \frac{p_X(1-p_X)}{n_X} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{n_Y}{p_Y(1-p_Y)} & 0 \\ 0 & \frac{n_X}{p_X(1-p_X)} \end{pmatrix}$$

where I used the known relationship between the information matrix and the variance of the MLE to more easily find the information matrix.
Using the results for the asymptotic normality of MLEs in regular problems (or alternatively, the CLT)

$$\begin{pmatrix} \overline{Y} \\ \overline{X} \end{pmatrix} \dot{\sim} \mathcal{N} \left(\begin{pmatrix} p_Y \\ p_X \end{pmatrix}, \begin{pmatrix} \frac{p_Y(1-p_Y)}{n_Y} & 0 \\ 0 & \frac{p_X(1-p_X)}{n_X} \end{pmatrix} \right).$$

By the invariance of MLEs, we know $\hat{\theta} = \hat{p}_Y - \hat{p}_X$, and thus using the linearity of the normal distribution

$$\hat{\theta} = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \overline{Y} \\ \overline{X} \end{pmatrix} \dot{\sim} \mathcal{N} \left(\theta = p_Y - p_X, \ \frac{p_Y(1 - p_Y)}{n_Y} + \frac{p_X(1 - p_X)}{n_X} \right).$$

So a Wald test would use $\hat{\vec{p}}$ in the expression for the variance to obtain

$$Z = \frac{\overline{Y} - \overline{X}}{\sqrt{\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y} + \frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \dot{\sim}_{H_0} \mathcal{N}(0, 1),$$

and we reject when

$$Z \leq z_{\alpha/2} \qquad \text{or} \qquad Z \geq z_{1-\alpha/2}$$

(b) Provide an expression for the score test (including critical value).
Ans: We find the MLE for $\vec{p}$ under the restriction $H_0 : p_Y = p_X$. We choose to use p_Y as the common parameter, and using the log likelihood derived above

$$\mathcal{L}^{(0)}(\vec{p}|\vec{Y},\vec{X}) = Y\log(p_Y) + (n_Y - Y)\log(1-p_Y) + X\log(p_Y) + (n_X - X)\log(1-p_Y)$$

$$\mathcal{U}_1^{(0)}(\vec{p}|\vec{Y},\vec{X}) = \frac{\partial}{\partial p_Y}\mathcal{L}(\vec{p}) = \frac{Y+X}{p_Y} - \frac{n_Y+n_X-Y-X}{1-p_Y} = \frac{Y+X-(n_Y+n_X)p_Y}{p_Y(1-p_Y)}$$

$$\vec{\mathcal{U}}^{(0)}(\hat{p}_Y^{(0)}) = \vec{0} \qquad \Rightarrow \qquad \hat{p}_Y^{(0)} = \frac{Y+X}{n_Y+n_X}$$

We thus find score statistic using

$$\hat{\vec{p}}^{(0)} = \begin{pmatrix} \frac{Y+X}{n_Y+n_X} \\ \\ \frac{Y+X}{n_Y+n_X} \end{pmatrix}$$

$$\vec{\mathcal{U}}(\hat{\vec{p}}^{(0)}) = \begin{pmatrix} \frac{(n_Y+n_X)(n_XY-n_YX)}{(Y+X)(n_Y+n_X-Y-X)} \\ \\ \frac{(n_Y+n_X)(n_YX-n_XY)}{(Y+X)(n_Y+n_X-Y-X)} \end{pmatrix}$$

$$\mathbf{J}(\hat{\vec{p}}^{(0)}) = \begin{pmatrix} \frac{n_Y(n_Y+n_X)^2}{(Y+X)(n_Y+n_X-Y-X)} & 0 \\ \\ 0 & \frac{n_X(n_Y+n_X)^2}{(Y+X)(n_Y+n_X-Y-X)} \end{pmatrix}$$

$$\mathbf{J}^{-1}(\hat{\vec{p}}^{(0)}) = \begin{pmatrix} \frac{(Y+X)(n_Y+n_X-Y-X)}{n_Y(n_Y+n_X)^2} & 0 \\ \\ 0 & \frac{(Y+X)(n_Y+n_X-Y-X)}{n_X(n_Y+n_X)^2} \end{pmatrix}$$

So

$$T = \vec{\mathcal{U}}^T(\hat{\vec{p}}^{(0)})\mathbf{J}^{-1}(\hat{\vec{p}}^{(0)})\vec{\mathcal{U}}(\hat{\vec{p}}^{(0)}) = \frac{(\overline{Y}-\overline{X})^2}{\overline{p}(1-\overline{p})\left(\frac{1}{n_Y}+\frac{1}{n_X}\right)},$$

where for notational convenience we define $\overline{p} = (Y+X)/(n_Y+n_X)$. Under the null hypothesis T has a χ_1^2 distribution. So we reject when

$$T > \chi^2_{1,\alpha}$$

(c) Provide an expression for the likelihood ratio test (including critical value).
Ans: We compute

$$\begin{aligned}\Lambda &= 2(\mathcal{L}(\hat{\vec{p}}) - \mathcal{L}(\hat{\vec{p}}^{(0)})) \\ &= 2(\ Y\log(\overline{Y}) + (n_Y - Y)\log(1-\overline{Y}) + X\log(\overline{X}) + (n_X - X)\log(1-\overline{X}) \\ &- (X+Y)\log(\overline{p}) - (n_Y + n_x - X - Y)\log(\overline{p})\ )\end{aligned}$$

Under the null hypothesis Λ has a χ_1^2 distribution. So we reject when

$$\Lambda > \chi^2_{1,\alpha}$$

(d) Provide an expression for a CI based on the Wald test.
Ans: An approximate $100(1-\alpha)\%$ CI would be

$$\hat{p}_Y - \hat{p}_X \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y} + \frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}$$

(e) Describe the steps involved in finding CI based on score or LRT.
Ans: We invert the score or LRT, by considering testing null hypotheses $H_0 : \theta = \theta_1$ (so $p_X = p_Y - \theta_1$). For each such θ_1, we have to maximize the restricted likelihood

$$\mathcal{L}^{(0)}(\vec{p}|\vec{Y},\vec{X}) = Y\log(p_Y) + (n_Y - Y)\log(1-p_Y) + X\log(p_Y - \theta_1) + (n_X - X)\log(1 - p_Y + \theta_1)$$

Our CI is thus the set of all θ_1 that would not be rejected in a level α test. This is most easily done in a computerized searching which for a given observation of $\vec{Y}$ and $\vec{X}$, we search for the value of θ_L and θ_U such that the test statistic is exactly the critical value of the χ^2_1 distribution.

8. (optional) Suppose we observe data $Y_1, Y_2, \ldots, Y_n$ that are i.i.d. nonnegative random variables with $Y_i \sim (\theta, \gamma\theta^2)$. Our target of interest is to decide whether Y_i is distributed according to a gamma distribution or a lognormal distribution. That is we want a most powerful level α test of H_0 vs H_1 with

$$H_0 : Y \sim f_Y(y \,|\, \eta, \gamma) = \frac{1}{\Gamma(\eta)\gamma^\eta} y^{\eta-1} e^{-y/\gamma} \qquad \text{for } \eta = \theta/\gamma$$

$$H_1 : Y \sim f_Y(y \,|\, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{(\log(y)-\mu)^2}{2\sigma^2}\right\} \qquad \text{for } \begin{cases} \sigma^2 = \log(\gamma+1) \\ \eta = \log(\theta) - \frac{\sigma^2}{2} \end{cases}$$

(a) Suppose it is known $\mu = 0$ and $\eta = 1$ (so $\gamma = (1+\sqrt{(5)})/2$ and $\sigma^2 = 2\log(\gamma)$). Derive the MP-α test, describing how the critical value could be found.

(b) Provide an expression for the MP-α test when all distributional parameters are unknown.