By signing below I pledge that I have not communicated with anyone about this quiz other than the instructor.

Signature: ______________________________

1(a). Let $Y_1, \dots, Y_n \sim$ i.i.d. Poisson(θ), with density $p(y|\theta) = \theta^y e^{-\theta}/y!$. Write out the log likelihood, find the likelihood equation and the MLE. Show all of your steps.

1(b). Identify the asymptotic distribution of the MLE. Specifically, identify the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{MLE} - \theta)$ and justify your result.

1(c). Suppose we are interested in estimating the probability that an observation from this population will be zero. In other words, we are interested in $g(\theta) = \Pr(Y = 0|\theta)$, where $Y \sim$ Poisson(θ). Find the MLE of $g(\theta)$ based on a sample $Y_1, \dots, Y_n$, and also specify its asymptotic distribution.

1(d). Let X_i be the indicator that $Y_i = 0$, so that $X_i = 1$ if $Y_i = 0$ and $X_i = 0$ otherwise. Find a CAN estimator of $g(\theta)$ based on $X_1, \dots, X_n$ and derive its asymptotic distribution. You may use results from class but explain what results you are using.

1(e). Compare the asymptotic variance of the estimator in part (c) to that of the estimator in part (d).

Solution

1(a). The log-likelihood is given by

$$\ell_n(\theta) = \sum_{i=1}^n [y_i \log\theta - \theta - \log(y_i!)] = \sum_{i=1}^n y_i \log\theta - n\theta - \sum_{i=1}^n \log(y_i!)$$

Taking the derivative of log-likelihood and setting it equal to 0, we have

$$\frac{\partial \ell_n}{\partial \theta} = \frac{\sum_{i=1}^n y_i}{\theta} - n = 0$$

Therefore, the MLE is $\hat{\theta}_{\text{MLE}} = \sum_{i=1}^n Y_i/n = \bar{Y}$.

1(b). Since $Y_i \overset{\text{iid}}{\sim} \mathsf{Po}(\theta)$, we know that $\mathsf{E}[Y_i] = \mathsf{Var}(Y_i) = \theta$. Therefore, using CLT, we have

$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \xrightarrow{\text{d}} \mathsf{N}(0, \theta)$$

1(c). We have $g(\theta) = \mathsf{Pr}(Y = 0 \mid \theta) = \theta^0 e^{-\theta}/0! = e^{-\theta}$. Since the MLE for θ is $\bar{Y}$, the MLE of $g(\theta)$ is $g(\bar{Y}) = \exp(-\bar{Y})$. Since

$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \xrightarrow{\text{d}} \mathsf{N}(0, \theta)$$

Using Delta's method, we have

$$\sqrt{n}\left(g(\hat{\theta}_{\text{MLE}}) - g(\theta)\right) \xrightarrow{\text{d}} \mathsf{N}\left(0, |g'(\theta)|^2\theta\right) \stackrel{\text{d}}{=} \mathsf{N}\left(0, e^{-2\theta}\theta\right)$$

1(d). Here $X \sim \mathsf{Ber}(g(\theta))$. A natural choice of estimator of $g(\theta)$ is the sample mean $\widehat{g(\theta)} = \sum_{i=1}^n X_i = \bar{X}$. We know that $\mathsf{E}[X] = g(\theta), \mathsf{Var}(X) = g(\theta)\left(1 - g(\theta)\right)$. Due to CLT, we have

$$\sqrt{n}\left(\widehat{g(\theta)} - g(\theta)\right) \xrightarrow{\text{d}} \mathsf{N}\left(0, e^{-\theta}(1 - e^{-\theta})\right)$$

Therefore, $\widehat{g(\theta)}$ is a CAN estimator.

1(e). It suffices to compare θ and $e^\theta - 1$. Clearly, using Taylor's expansion, we have $\theta < e^\theta - 1$ for any $\theta > 0$. Therefore, the asymptotic variance of $g(\hat{\theta}_{\text{MLE}})$ is smaller than that of $\widehat{g(\theta)}$.

For each $j = 1, \ldots, m$, let p_j be a valid p-value for testing a null hypothesis H_j.

2(a). Suppose we reject hypothesis j when $p_j < \alpha_E$. In the case that the p-values are mutually independent, derive the value of α_E that makes the global type I error rate (the probability that one or more H_j's are rejected assuming they are all true) exactly equal to a specified value α_G.

2(b). Find a value of α_F such that if we reject hypothesis j when $p_j < \alpha_F$, then the global error rate is less than or equal to α_G, even if the p-values are dependent, and prove your result.

2(c). Which threshold, α_E or α_F, is preferable in terms of power if the p-values are truly independent, and why?

Solution

2(a). Denote H_0 : All null hypotheses are true. Then the type I error rate is

$$\begin{aligned}\mathsf{Pr}\left(\text{Reject } H_0 \mid H_0\right) &= 1 - \mathsf{Pr}\left(\text{Accept all } H_j,\ j = 1, \ldots, m \mid H_0\right) \\ &= 1 - (1 - \alpha_E)^m\end{aligned}$$

The second step is due to the facts that (1) all p-values are mutually independent and (2) each p-value follows a uniform distribution under the null hypothesis.

Let $1 - (1 - \alpha_E)^m = \alpha_G$, we have $\alpha_E = 1 - (1 - \alpha_G)^{1/m}$.

2(b). Let $\alpha_F = \alpha_G/m$ and we claim that this choice of α_F will lead to a global type I error rate which is less than or equal to α_G. A proof is as follows.

$$\begin{aligned}\mathsf{Pr}(\text{Reject } H_0 \mid H_0) &= \mathsf{Pr}(\text{Reject at least one } H_j,\ j = 1, \ldots, m \mid H_0) \\ &= \mathsf{Pr}\left(\bigcup_{j=1}^{m} \{\text{Reject } H_j\} \mid H_0\right) \\ &\leqslant \sum_{j=1}^{m} \mathsf{Pr}\left(\text{Reject } H_j \mid H_0\right) \\ &= \sum_{j=1}^{m} \alpha_F = m\alpha_F = \alpha_G\end{aligned}$$

The final step is again due to the fact that each p-value follows a uniform distribution under the null hypothesis.

2(c). Conceptually, when p-values are truly independent, the threshold α_E is preferable as it is derived under the exactly same assumption and the global type I error rate is exactly α_G. Compared to α_E, α_F is derived under more relaxed assumption and hence will not be as sharp as α_E. Also, the global type I error rate is not exactly equal to α_G. It is more conservative.

Mathematically, let $f(x) = x^d - dx + d - 1$ where $x \in (0, 1)$. It is clear that $f(1) = 0$ for any $d \in (0, 1)$. Since $f'(x) = d(x^{d-1} - 1) < 0$ for any $x \in (0, 1)$ and any $d \in (0, 1)$, $f(x)$ is a monotone decreasing function on $(0, 1)$. Therefore,

$$
\begin{aligned}
& (1-\alpha_G)^{1/m} - \frac{1-\alpha_G}{m} + \frac{1}{m} - 1 < 0 \quad \text{for all } \alpha_G \in (0,1) \text{ and } m > 1 \\
\Longrightarrow\ & \frac{\alpha_G}{m} < 1 - (1-\alpha_G)^{1/m} \quad \text{for all } \alpha_G \in (0,1) \text{ and } m > 1 \\
\Longrightarrow\ & \alpha_F < \alpha_E \quad \text{for all } \alpha_G \in (0,1) \text{ and } m > 1
\end{aligned}
$$

When $m = 1$, two thresholds are exactly same. When $m > 1$, α_F is smaller than α_E. Therefore, using α_E results in more rejections than using α_F. From the perspective of power comparison, using α_E leads to a more powerful test.

3(a). Let $Y_1, \ldots, Y_n \sim$ i.i.d. P_θ for some unknown $\theta \in \mathbb{R}^+$, where the density for P_θ is $p(y|\theta) = e^{-y/\theta}/\theta$. Write out the log likelihood $l(\theta)$ for θ based on observed values $y_1, \ldots, y_n$, find the MLE $\hat{\theta}$, and write out the formula for $l(\hat{\theta})$, the maximized value of the log likelihood.

3(b). Suppose we have two independent observations Y_A and Y_B, one from each of two populations, so that $Y_A \sim P_{\theta_A}$ and $Y_B \sim P_{\theta_B}$. Write out the formula for $l(\theta_A, \theta_B)$, the log likelihood of (θ_A, θ_B) based on observed values y_A and y_B.

3(c). We wish to test the hypothesis $H_0 : \theta_A = \theta_B$ versus $H_1 : \theta_A \neq \theta_B$. Compute the minus two log likelihood ratio statistic for this test, which is $t(Y_A, Y_B) = -2 \times (l_0 - l_1)$, where l_1 is the maximum value of $l(\theta_A, \theta_B)$ and l_0 is the maximum subject to the restriction that $\theta_A = \theta_B$. (Hint: Use the result from (a) for some cases where $n = 1$ and $n = 2$.)

3(d). Write $t(Y_A, Y_B)$ in terms of the geometric mean $(Y_A Y_B)^{1/2}$ and the arithmetic mean $(Y_A + Y_B)/2$. Provide some intuition as to why this statistic makes sense for evaluating evidence against H_0.

3(e). To actually implement a level-α test of H_0 with this test statistic, you would have to know, approximate, or simulate its null distribution, which is the distribution of $t(Y_A, Y_B)$ when when θ_A and θ_B are equal to some common value, say θ. But how could you do this, if you don't know θ? To solve this problem, show that the distribution of $t(Y_A, Y_b)$ under the null hypothesis doesn't depend on the particular value of θ. (Hint: You may use the fact that if $Y \sim P_\theta$ then $Y \stackrel{d}{=} \theta Z$, where $Z \sim P_1$.)

Solution

3(a). The log-likelihood is

$$\ell_n(\theta) = \sum_{i=1}^{n} (-y_i/\theta - \log\theta) = -\frac{\sum_{i=1}^{n} y_i}{\theta} - n\log\theta$$

Taking derivative of log-likelihood and setting it equal to 0, we have

$$\frac{\partial \ell_n}{\partial \theta} = \frac{\sum_{i=1}^{n} y_i}{\theta^2} - \frac{n}{\theta} = 0$$

Therefore, $\hat{\theta}_{\text{MLE}} = \sum_{i=1}^{n} Y_i/n = \bar{Y}$. The maximized value of the log-likelihood is

$$\ell_n(\hat{\theta}_{\text{MLE}}) = -\frac{n\bar{y}}{\bar{y}} - n\log\bar{y} = -n - n\log\bar{y}.$$

3(b). Since two observations are independent, the joint log-likelihood of θ_A and θ_B is

$$\ell(\theta_A, \theta_B) = -y_A/\theta_A - \log\theta_A - y_B/\theta_B - \log\theta_B$$

3(c). The test statistic is

$$\begin{aligned}
t(Y_A, Y_B) &= -2\log\frac{\max_\theta L_0(\theta)}{\max_{\theta_A,\theta_B} L_1(\theta_A,\theta_B)} \\
&= -2\left(\max_\theta \ell_0(\theta) - \max_{\theta_A,\theta_B} \ell_1(\theta_A,\theta_B)\right) \\
&= -2\left(-2 - 2\log\frac{y_A+y_B}{2} - (-1 - \log y_A - 1 - \log y_B)\right) \qquad (3) \\
&= 4\left(\log\frac{y_A+y_B}{2} - \log\left(y_A y_B\right)^{1/2}\right)
\end{aligned}$$

Equation (3) is derived using the result in part (a) with $n = 1$ or 2.

3(d). Let $M_A = (Y_A + Y_B)/2$ and $M_G = (Y_A Y_B)^{1/2}$. Then

$$t(Y_A, Y_B) = 4\log\frac{M_A}{M_G}$$

We know that $M_A \geqslant M_G$ and the equality holds if and only if $Y_A = Y_B$. Let $Y_A/Y_B = m$, then

$$t(Y_A, Y_B) = 4\log\frac{(m+1)/2}{\sqrt{m}}$$

When the null hypothesis is true, $\theta_A = \theta_B$, we expect to observe that Y_A and Y_B are similar and hence $m \approx 1$. In this case $(m+1)/2 \approx \sqrt{m}$, the value of test statistic is small. However, when the null hypothesis is wrong, m could be very small or large. In either case, the ratio between $(m+1)/2$ and $\sqrt{m}$ is large and hence the value of test statistic is large.

Therefore, the behavior of the test statistic makes intuitive sense. It evaluates the evidence against the null hypothesis.

3(e). By change of variable, it is easy to show that if $Y \sim P_\theta$, then $Y \overset{\mathrm{d}}{=} \theta Z$ where $Z \sim P_1$. Under the null hypothesis, we have $\theta_A = \theta_B = \theta$ and hence $Y_A \overset{\mathrm{d}}{=} \theta Z_A$ and $Y_B \overset{\mathrm{d}}{=} \theta Z_B$ where $Z_A, Z_B \overset{\mathrm{iid}}{\sim} P_1$.

Therefore, the test statistic can be rewritten as

$$\begin{aligned} t(Y_A, Y_B) &= 4\left(\log \frac{y_A + y_B}{2} - \log\left(y_A y_B\right)^{1/2}\right) \\ &= 4\left(\log \frac{\theta Z_A + \theta Z_B}{2} - \log\left(\theta Z_A \theta Z_B\right)^{1/2}\right) \\ &= 4\left(\log \frac{Z_A + Z_B}{2} - \log\left(Z_A Z_B\right)^{1/2}\right) \end{aligned}$$

Since $Z_A, Z_B \overset{\mathrm{iid}}{\sim} P_1$, the null distribution does not depend on actual value of θ.