# BIOS 6612 Homework 1: Likelihood theory
## Solutions

1. (**35 points**) Consider the Poisson distribution with mass function

$$f(y) = \frac{\lambda^y \exp(-\lambda)}{y!}, \lambda > 0, y \in \{0, 1, 2, \ldots\}$$

You have a sample of size $n = 50$ from this distribution with the following summary statistics:

- the sample mean of $Y_1, \ldots, Y_n$, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = 1.3$$

- the sample mean of $Y_1^2, \ldots, Y_n^2$, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 = 2.62$$

- the sample mean of $\log(Y_1!), \log(Y_2!), \ldots, \log(Y_n!)$, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} \log(Y_i!) = 0.4148$$

Assume we are interested in making inferences about $\lambda$.

(a) Derive the log-likelihood and score functions for this problem. (**5 points**)
Log-likelihood function is

$$\begin{aligned}
\ell(\lambda) &= \log \prod_{i=1}^{n} f(Y_i) \\
&= \log \prod_{i=1}^{n} \frac{\lambda^{Y_i} \exp(-\lambda)}{Y_i!} \\
&= \sum_{i=1}^{n} \log \left\{ \frac{\lambda_i^{Y} \exp(-\lambda)}{Y_i!} \right\} \\
&= \log \lambda \sum_{i=1}^{n} Y_i - n\lambda - \sum_{i=1}^{n} \log Y_i! \\
&= 50 \left(\log \lambda \cdot 1.3 - \lambda - 0.4148\right)
\end{aligned}$$

Score function is

$$U(\lambda) = \frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left\{ \log \lambda \sum_{i=1}^{n} Y_i - n\lambda - \sum_{i=1}^{n} \log Y_i! \right\}$$
$$= \frac{1}{\lambda} \sum_{i=1}^{n} Y_i - n$$
$$= 50 \left( \frac{1.3}{\lambda} - 1 \right)$$

(b) Compute $\hat{\lambda}$, the maximum likelihood estimator of $\lambda$. (**5 points**)
Set $U(\lambda) = 0$ and solve for $\lambda$:

$$0 = \frac{1}{\lambda} \sum_{i=1}^{n} Y_i - n$$
$$n = \frac{1}{\lambda} \sum_{i=1}^{n} Y_i$$
$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} Y_i = 1.3$$

(c) Derive the expected information about $\lambda$ for this problem. (**5 points**)
The information is the negative expectation of the second derivative of the log-likelihood function:

$$\mathcal{I}(\lambda) = -\mathbb{E} \frac{\partial U(\lambda)}{\partial \lambda}$$
$$= -\mathbb{E} \frac{\partial}{\partial \lambda} \left( \frac{1}{\lambda} \sum_{i=1}^{n} Y_i - n \right)$$
$$= -\mathbb{E} \left( -\frac{1}{\lambda^2} \sum_{i=1}^{n} Y_i \right)$$
$$= \frac{1}{\lambda^2} \sum_{i=1}^{n} \mathbb{E}\, Y_i$$
$$= \frac{n\lambda}{\lambda^2}$$
$$= \frac{n}{\lambda} = \frac{50}{\lambda}$$

(d) Construct a 95% Wald confidence interval for $\lambda$. (**2 points**)
First we need to estimate the standard error of $\hat{\lambda}$, which we do using the information: $\operatorname{Var} \hat{\lambda} \approx 1/\mathcal{I}(\lambda)$. So our 95% confidence interval will be

$$\hat{\lambda} \pm 1.96 \sqrt{\frac{1}{\mathcal{I}(\hat{\lambda})}} = 1.3 \pm 1.96 \sqrt{\frac{1.3}{50}} = (0.98, 1.62)$$

(e) Test $H_0 : \lambda = 1$ versus a two-sided $H_1$ using the score test. Give the test statistic, null distribution, and $p$ value. (**6 points**)

The score test statistic is

$$\frac{U(\lambda_0)^2}{\operatorname{Var} U(\lambda_0)} = \frac{U(\lambda_0)^2}{\mathcal{I}(\lambda_0)} = \frac{\left\{50\left(\frac{1.3}{\lambda_0} - 1\right)\right\}^2}{50/\lambda_0}$$

Plugging in the value $\lambda_0 = 1$ from $H_0$ we find

$$\begin{aligned} \frac{U(\lambda_0)^2}{\operatorname{Var} U(\lambda_0)} &= \frac{\left\{50\left(\frac{1.3}{1} - 1\right)\right\}^2}{50/1} \\ &= 4.5 \end{aligned}$$

Since we are testing the value of 1 parameter, the null distribution is $\chi^2_1$, giving a $p$-value of approximately 0.03.

(f) Repeat question **??** using the likelihood ratio test. (**6 points**)

Test statistic is

$$\begin{aligned} 2\left(\ell(\hat\lambda) - \ell(\lambda_0)\right) &= 2\left\{\left(\log\hat\lambda\sum_{i=1}^n Y_i - n\hat\lambda - \sum_{i=1}^n \log Y_i!\right) - \left(\log\lambda_0\sum_{i=1}^n Y_i - n\lambda_0 - \sum_{i=1}^n \log Y_i!\right)\right\} \\ &= 2\left\{\left(\log\hat\lambda\sum_{i=1}^n Y_i - n\hat\lambda\right) - \left(\log\lambda_0\sum_{i=1}^n Y_i - n\lambda_0\right)\right\} \\ &= 2\left\{\log\frac{\hat\lambda}{\lambda_0}\sum_{i=1}^n Y_i + n\left(\lambda_0 - \hat\lambda\right)\right\} \\ &= 2\left\{\log\frac{1.3}{1}\cdot 50\cdot 1.3 + 50\,(1 - 1.3)\right\} \\ &\approx 4.11 \end{aligned}$$

Null distribution is still $\chi^2_1$ so $p$-value is approximately 0.04.

(g) Repeat question **??** using the Wald test. (**6 points**)

The Wald test statistic is

$$\frac{\left(\hat\lambda - \lambda_0\right)^2}{\operatorname{Var}\hat\lambda} = \frac{\left(\hat\lambda - \lambda_0\right)^2}{1/\mathcal{I}(\hat\lambda)} = \frac{\left(\hat\lambda - \lambda_0\right)^2}{\hat\lambda/n}$$

Plugging in the value $\lambda_0 = 1$ from $H_0$ and the estimated $\hat\lambda = 1.3$ we find

$$\frac{\left(\hat\lambda - \lambda_0\right)^2}{\operatorname{Var}\hat\lambda} = \frac{(1.3-1)^2}{1.3/50} \approx 3.46$$

Null distribution is still $\chi^2_1$ so $p$-value is approximately 0.06.

2. (**20 points**) Suppose now we only are able to observe $Y_i^* = \mathbb{1}(Y_i > 0)$, i.e., a value of 1 if $Y_i > 0$ and 0 if $Y_i = 0$. As before, assume that $Y_i$ has the Poisson distribution with mean $\lambda$.

   (a) Give the likelihood function required to estimate $\lambda$ based on a sample of $Y_1^*, Y_2^*, \ldots, Y_n^*$. (**4 points**)
   Note that $Y_i^*$ is just a Bernoulli random variable, taking the value 1 with probability $p$ and 0 with probability $1-p$. We can find the value $p$ using the Poisson mass function:

$$P(Y_i^* = 1) = P(Y_i > 0) = 1 - P(Y_i = 0) = 1 - \exp(-\lambda).$$

   Recall from class that the binomial likelihood is

$$\prod_{i=1}^{n} p^{Y_i^*}(1-p)^{1-Y_i^*}.$$

   Substituting in the value of $p$ we found above, the likelihood function for $\lambda$ is

$$L(\lambda) = \prod_{i=1}^{n} \left(1 - \exp(-\lambda)\right)^{Y_i^*} \left(\exp(-\lambda)\right)^{1-Y_i^*}.$$

   (b) Find the MLE of $\lambda$ based on this likelihood function. *Hint: Your answer should be a function of the sample mean of the $Y_i^*$'s.* (**6 points**)
   The score function for this problem is the derivative of the log-likelihood above with respect to $\lambda$:

$$\begin{aligned}
U(\lambda) &= \frac{\partial \log L(\lambda)}{\partial \lambda} \\
&= \frac{\partial}{\partial \lambda}\left\{\log\left(1 - e^{-\lambda}\right)\sum_{i=1}^{n} Y_i^* - \lambda \sum_{i=1}^{n}(1 - Y_i^*)\right\} \\
&= \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{i=1}^{n} Y_i^* - \sum_{i=1}^{n}(1 - Y_i^*).
\end{aligned}$$

   To get the MLE of $\lambda$ we set this equal to 0 and solve for $\lambda$:

$$\begin{aligned}
0 &= \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{i=1}^{n} Y_i^* - \sum_{i=1}^{n}(1 - Y_i^*) \\
\sum_{i=1}^{n}(1 - Y_i^*) &= \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{i=1}^{n} Y_i^* \\
\frac{\sum_{i=1}^{n}(1 - Y_i^*)}{\sum_{i=1}^{n} Y_i^*} &= \frac{e^{-\lambda}}{1 - e^{-\lambda}}.
\end{aligned}$$

Let $\hat{\theta} = \frac{\sum_{i=1}^n (1-Y_i^*)}{\sum_{i=1}^n Y_i^*}$. Now,

$$\begin{aligned}
\hat{\theta} &= \frac{e^{-\lambda}}{1-e^{-\lambda}} \\
\hat{\theta} - \hat{\theta}e^{-\lambda} &= e^{-\lambda} \\
\hat{\theta} &= e^{-\lambda}\left(1+\hat{\theta}\right) \\
e^{-\lambda} &= \frac{\hat{\theta}}{1+\hat{\theta}} \\
\hat{\lambda} &= -\log\frac{\hat{\theta}}{1+\hat{\theta}} \\
&= \log\left(1+\hat{\theta}\right) - \log\left(\hat{\theta}\right) \\
&= \log\left(1+\frac{\sum_{i=1}^n (1-Y_i^*)}{\sum_{i=1}^n Y_i^*}\right) - \log\left(\frac{\sum_{i=1}^n (1-Y_i^*)}{\sum_{i=1}^n Y_i^*}\right) \\
&= \log\left(\frac{n}{\sum_{i=1}^n Y_i^*}\right) - \log\left(\frac{n-\sum_{i=1}^n Y_i^*}{\sum_{i=1}^n Y_i^*}\right).
\end{aligned}$$

Let $\overline{Y^*} = \frac{1}{n}\sum_{i=1}^n Y_i^*$. Then,

$$\begin{aligned}
\hat{\lambda} &= \log\left(\frac{1}{\overline{Y^*}}\right) - \log\left(\frac{n/n-\sum_{i=1}^n Y_i^*/n}{\sum_{i=1}^n Y_i^*/n}\right) \\
&= -\log\overline{Y^*} - \log\left(1-\overline{Y^*}\right) + \log\overline{Y^*} \\
&= -\log\left(1-\overline{Y^*}\right).
\end{aligned}$$

(c) Derive the information about $\lambda$ based on this sample and hence find the variance of the MLE. (**6 points**)

The information is the negative expectation of the derivative of the score function:

$$\begin{aligned}
\mathcal{I}(\lambda) &= -\,\mathbb{E}\,\frac{\partial U(\lambda)}{\partial \lambda} \\
&= -\,\mathbb{E}\,\frac{\partial}{\partial \lambda}\left\{\frac{e^{-\lambda}}{1-e^{-\lambda}}\sum_{i=1}^{n} Y_i^* - \sum_{i=1}^{n}(1-Y_i^*)\right\} \\
&= -\frac{\partial}{\partial \lambda}\left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right)\mathbb{E}\sum_{i=1}^{n} Y_i^* \\
&= -n\left(1-e^{-\lambda}\right)\frac{\partial}{\partial \lambda}\left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right) \\
&= -n\left(1-e^{-\lambda}\right)\left(-\frac{e^{-\lambda}}{1-e^{-\lambda}} + \frac{e^{-2\lambda}}{(1-e^{-\lambda})^2}\right) \\
&= -n\left(1-e^{-\lambda}\right)\left(-\frac{e^{-\lambda}}{(1-e^{-\lambda})^2}\right) \\
&= n \cdot \frac{e^{-\lambda}}{1-e^{-\lambda}}
\end{aligned}$$

(d) Using the invariance property of MLEs, explain an easier method for finding the MLE of $\lambda$ in this problem. Recall that this property states that the MLE of a function $g(\theta)$ is the function applied to the MLE of $\theta$, i.e., $g(\hat{\theta})$. (**4 points**)
Let $\pi = P(Y_i > 0) = P(Y_i^* = 1) = 1 - \exp(-\lambda)$. Recall that the MLE of $\pi$ based on $Y_i^*$'s is the sample mean $\frac{1}{n}\sum_{i=1}^{n} Y_i^*$. Let $g(\pi) = -\log(1-\pi)$ By the invariance property of MLEs, $-\log(1-\hat{\pi})$ is the MLE of $\lambda$.

3. (**20 points**) Now consider estimation of $\zeta = P(Y_1 = 0) = \exp(-\lambda)$ based on $Y_1, \ldots, Y_n$.

(a) One possible estimator is the proportion of 0's in the sample, i.e.,

$$\tilde{\zeta} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i = 0).$$

Find the variance of this estimator. (**10 points**)
We know that an individual observation of $\mathbb{1}(Y_i = 0)$ is distributed as Bernoulli with mean $\exp(-\lambda)$, so we have

$$\begin{aligned}
\mathrm{Var}(\tilde{\zeta}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i = 0)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left\{\mathbb{1}(Y_i = 0)\right\} \\
&= \frac{1}{n^2} n \cdot \exp(-\lambda)\left\{1 - \exp(-\lambda)\right\} \\
&= \frac{\exp(-\lambda)\left\{1 - \exp(-\lambda)\right\}}{n}.
\end{aligned}$$

(b) Another possible estimator is a transformation of the MLE, i.e.,

$$\hat{\zeta} = \exp(-\hat{\lambda}).$$

Find the approximate variance of this estimator. *Hint: Use the delta method.* (**10 points**)

Our transformation of $\hat{\lambda}$ is $g(\lambda) = \exp(-\lambda)$, with derivative $g'(\lambda) = -\exp(-\lambda)$. By the delta method, $\operatorname{Var} g(\hat{\lambda}) \approx \{g'(\lambda)\}^2 \operatorname{Var} \hat{\lambda} = \exp(-2\lambda)/\mathcal{I}(\lambda) = \exp(-2\lambda)\lambda/n$