Complete Data

30M

The annual number of claims is modeled to have a Poisson distribution with mean λ. You observed 3, 1, and 2 claims in the past three years, respectively. Using the data, what is an appropriate estimate of λ?

We are given a sample of three claim counts and are asked to fit a Poisson distribution to the sample. There are many ways to achieve our objective.

Depending on our needs, we could either match the sample moments to the fitted moments or match the sample percentiles to the fitted percentiles in order to estimate λ.

In this section, we will learn about ***maximum likelihood estimation (MLE)***. The MLE finds the set of parameters that maximizes the likelihood of the observations. In our example, we would select the value of λ such that the likelihood of the observations of {3, 1, 2} is at its peak.

Let's start by defining ***likelihood***. "Likelihood" is often used interchangeably with "probability". The likelihood of a data value is the probability function evaluated at the data value. Thus,

The likelihood for year 1 is

$$\Pr(N=3)=\frac{e^{-\lambda}\lambda^3}{3!}$$

The likelihood for year 2 is

$$\Pr(N=1)=e^{-\lambda}\lambda$$

The likelihood for year 3 is

$$\Pr(N=2)=\frac{e^{-\lambda}\lambda^2}{2!}$$

Assuming i.i.d. annual number of claims, the likelihood of the observations is the product of the likelihood of each observation. The ***likelihood function*** is denoted by $L(\lambda)$.

Let n_i be the number of claims observed in the i^{th} year.

$$\begin{aligned}L(\lambda)&=\prod_{i=1}^{3}\Pr(N=n_i)\\&=\frac{e^{-\lambda}\lambda^3}{3!}\left(e^{-\lambda}\lambda\right)\left(\frac{e^{-\lambda}\lambda^2}{2!}\right)\\&=\frac{e^{-3\lambda}\lambda^6}{12}\end{aligned}$$

The idea is to find the value of λ that maximizes the likelihood function.

Coach's Remarks

The likelihood function appears to be equivalent to a joint probability function evaluated at all of the observations. For this exam, you are free to make that association, but the two functions are inherently different. One key distinction is that a likelihood function is a function of the parameter(s), whereas a joint probability function is a function of the random variable outcomes.

Here is a handy trick to maximize a likelihood function more efficiently. If we take the natural logarithm of a function, the result will be positively proportional to the function. That means the logarithm of a function will achieve its maximum value at the same critical point as the function itself. Therefore, maximizing the logarithm of a function is equivalent to maximizing the function itself.

The logarithm of the likelihood function is much easier to take the derivative of because it is a sum, whereas the likelihood function is a product whose derivative will require the product rule.

Thus, take the natural logarithm of the likelihood function and call it the ***log-likelihood function***.

$$\begin{aligned} l(\lambda) &= \ln L(\lambda) \\ &= \ln\left[\frac{e^{-3\lambda}\lambda^6}{12}\right] \\ &= -3\lambda + 6\ln\lambda - \ln 12 \end{aligned}$$

To maximize a function, set its first derivative equal to 0.

$$\begin{aligned} l'(\lambda) &= \frac{\mathrm{d}}{\mathrm{d}\lambda} l(\lambda) \\ &= -3 + \frac{6}{\lambda} = 0 \end{aligned}$$

Solve for λ.

$$\begin{aligned} -3 + \frac{6}{\lambda} &= 0 \\ \hat{\lambda} &= \frac{6}{3} = \mathbf{2} \end{aligned}$$

The "hat" above λ indicates "estimate"; thus $\hat{\lambda}$ is the estimate of λ.

Coach's Remarks

Let's try the same example without the log-likelihood function step. Calculate the first derivative of the likelihood function.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}L(\lambda) &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{e^{-3\lambda}\lambda^6}{12}\right)\\ &= \frac{e^{-3\lambda}\frac{\mathrm{d}}{\mathrm{d}\lambda}(-3\lambda)\lambda^6 + e^{-3\lambda}6\lambda^5}{12}\\ &= \frac{e^{-3\lambda}(-3)\lambda^6 + e^{-3\lambda}6\lambda^5}{12}\\ &= \frac{e^{-3\lambda}\lambda^5}{4}(-\lambda+2)\end{aligned}$$

Set the first derivative equal to zero and solve for λ.

$$\begin{aligned}\frac{e^{-3\lambda}\lambda^5}{4}(-\lambda+2) &= 0\\ -\lambda+2 &= 0\\ \hat{\lambda} &= \mathbf{2}\end{aligned}$$

We still get the same result, but this approach is more complex.

Note that although $\lambda = 0$ is a valid solution to the equation, since the Poisson mean, λ, must be greater than zero, $\lambda = 0$ is not a valid answer. Additionally, $\lambda = 0$ does not maximize $L(\lambda)$.

In conclusion, the steps to calculate the MLE of the parameter of a distribution (θ in this case) are:

1. Construct the likelihood function.

$$\begin{aligned}L(\theta) &= \prod_{\text{all } i} p(x_i) \qquad \text{(discrete)}\\ &= \prod_{\text{all } i} f(x_i) \qquad \text{(continuous)}\end{aligned}$$

2. Calculate the log-likelihood function.

$$l(\theta) = \ln L(\theta)$$

3. Calculate the first derivative of the log-likelihood function.

$$l'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta)$$

4. Set the first derivative equal to zero. Then, solve for θ.

$$l'(\theta) = 0$$

Coach's Remarks

Recall from Section S2.1.1 that although evaluating a PDF at a single point does not yield a probability, it describes the relative likelihood at that given value. Thus, the product of PDFs is used as a proxy for the probability of a set of observations.

Example S3.1.1.1

You observe 4 claims:

$$20 \quad 50 \quad 150 \quad 380$$

The claim amounts follow an exponential distribution with mean θ.

Calculate the maximum likelihood estimate of θ.

Solution

Step 1: Construct the likelihood function. For continuous random variables, the likelihood is the PDF evaluated at the observation.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{4} f(x_i) \\ &= \left(\frac{1}{\theta}e^{-20/\theta}\right)\left(\frac{1}{\theta}e^{-50/\theta}\right)\left(\frac{1}{\theta}e^{-150/\theta}\right)\left(\frac{1}{\theta}e^{-380/\theta}\right) \\ &= \frac{1}{\theta^4}e^{-600/\theta} \end{aligned}$$

Step 2: Calculate the log-likelihood function.

$$\begin{aligned} l(\theta) &= \ln L(\theta) \\ &= -4\ln\theta - \frac{600}{\theta} \end{aligned}$$

Step 3: Calculate the first derivative of the log-likelihood function.

$$\begin{aligned} l'(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta) \\ &= -\frac{4}{\theta} + \frac{600}{\theta^2} \end{aligned}$$

Step 4: Set the first derivative equal to zero. Then, solve for θ.

$$\begin{aligned} -\frac{4}{\theta} + \frac{600}{\theta^2} &= 0 \\ -4\theta + 600 &= 0 \\ \hat{\theta} &= \frac{600}{4} = \mathbf{150} \end{aligned}$$

Coach's Remarks

As previously mentioned, calculating the log-likelihood function is optional. We can maximize the likelihood function directly.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}L(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1}{\theta^4}e^{-600\,/\,\theta}\right)\\ &= -\frac{4}{\theta^5}e^{-600\,/\,\theta} + \frac{1}{\theta^4}e^{-600\,/\,\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(-\frac{600}{\theta}\right)\\ &= -\frac{4}{\theta^5}e^{-600\,/\,\theta} + \frac{1}{\theta^4}e^{-600\,/\,\theta}\left(\frac{600}{\theta^2}\right)\\ &= \frac{1}{\theta^5}e^{-600\,/\,\theta}\left(-4 + \frac{600}{\theta}\right)\end{aligned}$$

Set the first derivative equal to zero. Then, solve for θ.

$$\begin{aligned}\frac{1}{\theta^5}e^{-600\,/\,\theta}\left(-4 + \frac{600}{\theta}\right) &= 0\\ -4 + \frac{600}{\theta} &= 0\\ \hat{\theta} = \frac{600}{4} &= \mathbf{150}\end{aligned}$$

Example S3.1.1.2

Losses follow a distribution with probability density function:

$$f(x) = \frac{x^3 e^{-x\,/\,\theta}}{6\,\theta^4}, \quad x > 0$$

You observe 5 losses totaling 700.

Calculate the maximum likelihood estimate of θ.

Solution

Step 1: Construct the likelihood function. For continuous random variables, the likelihood is the PDF evaluated at the observations.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{5} \left(\frac{x_i^3 e^{-x_i/\theta}}{6\theta^4} \right) \\
&= \frac{\left(\prod_{i=1}^{5} x_i^3\right) e^{-\sum x_i/\theta}}{6^5 \theta^{20}} \\
&= c \cdot \theta^{-20} e^{-700/\theta}
\end{aligned}
$$

where $c = \dfrac{\prod_{i=1}^{5} x_i^3}{6^5}$.

Step 2: Calculate the log-likelihood function.

$$
l(\theta) = \ln c - 20 \ln \theta - \frac{700}{\theta}
$$

Step 3: Calculate the first derivative of the log-likelihood function.

$$
l'(\theta) = 0 - \frac{20}{\theta} + \frac{700}{\theta^2}
$$

Step 4: Set the first derivative equal to zero. Then, solve for θ.

$$-\frac{20}{\theta}+\frac{700}{\theta^2}=0$$
$$-20\theta+700=0$$
$$\hat{\theta}=\frac{700}{20}=\mathbf{35}$$

Coach's Remarks

In the context of a likelihood function, a "constant" is something that does not depend on the parameter and can be factored out of the likelihood, i.e. it is multiplicative and not additive. The likelihood function often contains complex constant terms like what was denoted as c in Step 1. With the knowledge that these terms will eventually be cancelled out, we can drop them from the beginning.

A word of caution: this is allowed **only** because dropping multiplicative constants does not affect the MLE of the parameter(s). Dropping constants simplifies the computation. However, if the actual log-likelihood function is needed (we will see this in the Likelihood Ratio Hypothesis

What happens if we want to estimate two parameters? Take partial derivatives with respect to each parameter, set them both equal to zero, then solve for the two parameters.

Coach's Remarks

It is unlikely for an exam question to require candidates to perform the calculations for estimating two parameters, as the process can be tedious. However, Example S3.1.1.3 illustrates how it would be done.

Example S3.1.1.3

The random variable X follows a distribution with parameters θ_1 and θ_2 that has the following PDF:

$$f(x) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x-\theta_1)^2}{2\theta_2}}, \qquad -\infty < x < \infty$$

n values of X are observed: $x_1, x_2, \ldots, x_n$.

Calculate the maximum likelihood estimators of θ_1 and θ_2.

Solution

The likelihood function is the product of the PDFs.

$$\begin{aligned} L(\theta_1, \theta_2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}} \\ &= c \cdot (\theta_2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\theta_1)^2}{2\theta_2}} \end{aligned}$$

where $c = \left(\sqrt{2\pi}\right)^{-n}$.

Calculate the log-likelihood function.

$$l(\theta_1, \theta_2) = \ln c - \frac{n}{2} \ln\theta_2 - \frac{\sum_{i=1}^{n}(x_i-\theta_1)^2}{2\theta_2}$$

To find the values of the two parameters that maximize the function, take partial derivatives with respect to each parameter. Start with θ_1.

$$\begin{aligned}\frac{\partial}{\partial \theta_1} l(\theta_1,\, \theta_2) &= 0 - \frac{\sum_{i=1}^{n} \left[2(x_i-\theta_1)(-1)\right]}{2\theta_2} \\ &= \frac{\sum_{i=1}^{n} (x_i-\theta_1)}{\theta_2} \\ &= \frac{\left(\sum_{i=1}^{n} x_i\right) - n\theta_1}{\theta_2}\end{aligned}$$

Now, take the partial derivative with respect to θ_2.

$$\frac{\partial}{\partial \theta_2} l(\theta_1,\, \theta_2) = -\frac{n}{2\theta_2} + \frac{\sum_{i=1}^{n} (x_i-\theta_1)^2}{2{\theta_2}^2}$$

Set both partial derivatives equal to zero. Now, we have two equations to solve for two parameters.

$$\begin{aligned}\frac{\sum_{i=1}^{n} x_i - n\theta_1}{\theta_2} &= 0 \qquad\qquad -(1) \\ -\frac{n}{2\theta_2} + \frac{\sum_{i=1}^{n} (x_i-\theta_1)^2}{2{\theta_2}^2} &= 0 \qquad\qquad -(2)\end{aligned}$$

Multiply (1) by θ_2.

$$\begin{aligned}\sum_{i=1}^{n} x_i - n\theta_1 = 0 \\ \hat{\theta}_1 = \boldsymbol{\frac{\sum_{i=1}^{n} x_i}{n}}\end{aligned}$$

Multiply (2) by $2{\theta_2}^2$.

$$-n\theta_2 + \sum_{i=1}^{n} (x_i - \theta_1)^2 = 0$$

$$\hat{\theta}_2 = \frac{\sum_{i=1}^{n} \left(x_i - \hat{\theta}_1\right)^2}{n}$$