Chapter 8.3. Maximum Likelihood Estimation

Prof. Tesler

Math 283
Fall 2019

Estimating parameters

- Let Y be a random variable with a distribution of known type but unknown parameter value θ.
  - **Bernoulli or geometric** with unknown p.
  - **Poisson** with unknown mean μ.
- Denote the pdf of Y by

$$P_Y(y;\theta)$$

  to emphasize that there is a parameter θ.
- Do n independent trials to get data $y_1, y_2, y_3, \ldots, y_n$.
  The joint pdf is

$$P_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n;\theta) = P_Y(y_1;\theta)\cdots P_Y(y_n;\theta)$$

- **Goal:** Use the data to estimate θ.

Likelihood function

- Previously, we knew the parameter θ and regarded the y's as unknowns (occurring with certain probabilities).
- Define the *likelihood* of θ given data $y_1, \ldots, y_n$ to be

$$L(\theta; y_1, \ldots, y_n) = P_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n; \theta) = P_Y(y_1; \theta) \cdots P_Y(y_n; \theta)$$

- It's the exact same formula as the joint pdf; the difference is the interpretation. Now the data $y_1, \ldots, y_n$ is given while θ is unknown.

Definition (Maximum Likelihood Estimate, or MLE)

The value $\theta = \widehat{\theta}$ that maximizes L is the *Maximum Likelihood Estimate*. Often, it is found using Calculus by locating a critical point:

$$\frac{dL}{d\theta} = 0 \qquad \frac{d^2L}{d\theta^2} < 0$$

However, be sure to check for complications such as discontinuities and boundary values of θ.

MLE for the Poisson distribution

- Y has a Poisson distribution with unknown parameter $\mu \geqslant 0$.
- Collect data from independent trials:

$$Y_1 = y_1, \; Y_2 = y_2, \; \cdots, \; Y_n = y_n$$

- **Likelihood:**

$$L(\mu; y_1, \ldots, y_n) = \prod_{i=1}^{n} e^{-\mu} \frac{\mu^{y_i}}{y_i!} = \frac{e^{-n\mu} \mu^{y_1 + \cdots + y_n}}{y_1! \cdots y_n!}$$

- **Log likelihood** is maximized at the same μ and is easier to use:

$$\ln L(\mu; y_1, \ldots, y_n) = -n\mu + (y_1 + \cdots + y_n) \ln \mu - \ln(y_1! \cdots y_n!)$$

- **Critical point:** Solve $d(\ln L)/d\mu = 0$:

$$\frac{d(\ln L)}{d\mu} = -n + \frac{y_1 + \cdots + y_n}{\mu} = 0 \qquad \text{so} \qquad \mu = \frac{y_1 + \cdots + y_n}{n}$$

MLE for the Poisson distribution

- **Log likelihood** is maximized at the same μ and is easier to use:

$$\ln L(\mu; y_1, \ldots, y_n) = -n\mu + (y_1 + \cdots + y_n) \ln \mu - \ln(y_1! \cdots y_n!)$$

- **Critical point:** Solve $d(\ln L)/d\mu = 0$:

$$\frac{d(\ln L)}{d\mu} = -n + \frac{y_1 + \cdots + y_n}{\mu} = 0 \qquad \text{so} \qquad \mu = \frac{y_1 + \cdots + y_n}{n}$$

- **Check second derivative is negative:**

$$\frac{d^2(\ln L)}{d\mu^2} = -\frac{y_1 + \cdots + y_n}{\mu^2} = -\frac{n^2}{y_1 + \cdots + y_n} < 0$$

provided $y_1 + \cdots + y_n > 0$. So it's a max unless $y_1 + \cdots + y_n = 0$.

- **Boundaries for range $\mu \geq 0$:** Must check $\mu \to 0^+$ and $\mu \to \infty$. Both send $\ln L \to -\infty$, so the μ identified above gives the max.

The Maximum Likelihood Estimate for the Poisson distribution

$$\hat{\mu} = \frac{y_1 + \cdots + y_n}{n} = \frac{0(\text{\# of 0's}) + 1(\text{\# of 1's}) + 2(\text{\# of 2's}) + \cdots}{n}$$

MLE for the Poisson distribution

- The exceptional case on the previous slide was $y_1 + \cdots + y_n = 0$, giving $y_1 = \cdots = y_n = 0$ (since all $y_i \geqslant 0$).

- In this case,

$$\begin{aligned} \ln L(\mu; y_1, \ldots, y_n) &= -n\mu + (y_1 + \cdots + y_n) \ln \mu - \ln(y_1! \cdots y_n!) \\ &= -n\mu + 0 \ln \mu - \ln(0! \cdots 0!) \\ &= -n\mu \end{aligned}$$

- On the range $\mu \geqslant 0$, this is maximized at $\hat{\mu} = 0$, which agrees with the main formula:

$$\hat{\mu} = \frac{y_1 + \cdots + y_n}{n} = \frac{0 + \cdots + 0}{n} = 0$$

Repeating the estimation gives different results

- **Scenario:** In a lab class, each student does 10 trials of an experiment and averages them. How do their results compare?
- A does n trials $y_{A1}, y_{A2}, \ldots, y_{An}$, leading to MLE $\widehat{\theta}_A$,
  B does n trials $y_{B1}, y_{B2}, \ldots, y_{Bn}$, leading to MLE $\widehat{\theta}_B$, etc.
  How do $\widehat{\theta}_A, \widehat{\theta}_B, \ldots$ compare?
- Treat the n trials in each experiment as random variables $Y_1, \ldots, Y_n$ and the MLE as a random variable $\widehat{\Theta}$.

Estimate Poisson parameter with $n = 10$ trials (secret: $\mu = 1.23$)

Experiment	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	$\widehat{\Theta}$
A	1	0	0	0	3	0	2	2	0	2	1.0
B	1	2	0	1	1	3	0	0	0	1	0.9
C	3	2	2	1	1	1	1	2	1	1	1.5
D	1	2	1	2	1	4	2	3	2	1	1.9
E	0	3	0	1	1	0	0	1	2	2	1.0
Mean	1.2	1.8	0.6	1	1.4	1.6	1	1.6	1	1.4	1.26

Desireable properties of an estimator $\widehat{\Theta}$

- $\widehat{\Theta}$ should be narrowly distributed around the correct value of θ.
- Increasing n should improve the estimate.
- The distribution of $\widehat{\Theta}$ should be known.

The MLE often does this (though not always!).

Bias

- Suppose Y is Poisson with secret parameter μ.
- Poisson MLE from data is

$$\hat{\mu} = \frac{Y_1 + \cdots + Y_n}{n}$$

- If many MLEs are computed from independent data sets, the average tends to

$$E(\hat{\mu}) = E\left(\frac{Y_1 + \cdots + Y_n}{n}\right) = \frac{E(Y_1) + \cdots + E(Y_n)}{n}$$
$$= \frac{\mu + \cdots + \mu}{n} = \frac{n\mu}{n} = \mu$$

- Since $E(\hat{\mu}) = \mu$, we say $\hat{\mu}$ is an *unbiased estimator* of μ.

Bias

- If $E(\hat{\mu}) = \mu$, then $\hat{\mu}$ is an *unbiased estimator* of μ.
  But if $E(\hat{\mu}) \neq \mu$, then $\hat{\mu}$ is a *biased estimator* of μ.
- **Contrived example:** Estimator $\hat{\mu}' = 2Y_1$ has $E(\hat{\mu}') = 2\mu$, so it's biased (unless $\mu = 0$).
- We will soon see an example (normal distribution) where the MLE gives a biased estimator.

Efficiency (want estimates to have small spread)

Increasing n

- Continue with Poisson MLE $\hat{\mu} = \frac{Y_1 + \cdots + Y_n}{n}$ and secret mean μ.
- The variance is

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\frac{Y_1 + \cdots + Y_n}{n}\right) = \frac{\mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_n)}{n^2}$$

$$= \frac{n\,\mathrm{Var}(Y_1)}{n^2} = \frac{\mathrm{Var}(Y_1)}{n} = \frac{\mu}{n}$$

Increasing n makes the variance smaller ($\hat{\mu}$ is more *efficient*).

Another estimator

- Set $\hat{\mu}' = \frac{Y_1 + 2Y_2}{3}$ (and ignore $Y_3, \ldots, Y_n$).

$$E(\hat{\mu}') = \frac{\mu + 2\mu}{3} = \mu \qquad \text{so unbiased}$$

$$\mathrm{Var}(\hat{\mu}') = \frac{\mathrm{Var}(Y_1) + 4\,\mathrm{Var}(Y_2)}{9} = \frac{\mu + 4\mu}{9} = \frac{5\mu}{9}$$

so it has higher variance (less efficient) than the MLE.