CS156 Session 6.2 PCW

1. Maximum Likelihood Estimation of a Bernoulli Distribution

Suppose we have a random sample $X_1, X_2, ..., X_n$ where:

- $X_i \ = 0$ if a randomly selected student does not eat meat, and
- $X_i \ = 1$ if a randomly selected student does eat meat.

Assuming that the X_i are independent Bernoulli random variables with unknown parameter p, derive the maximum likelihood estimator of p, the proportion of students who eat meat.

Answer:

Using the PMF for a bernoulli random variable, the likelihood can be defined thus:

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$

To simplify things we can take the natural log of the likelihood as was done in the readings. Again, as per the readings, this is fine because the natural log is a monotonically increasing function:

$$LL(p) = \sum_{i=1}^{n} log\ p^{X_i}(1-p)^{1-X_i} = \sum_{i=1}^{n} X_i log\ p + (1-X_i) log(1-p)$$

Note: Multiplication turns to addition in logarithm for

$$= \sum_{i=1}^{n} X_i log\, p + \sum_{i=1}^{n} (1 - X_i) log(1 - p)$$

To make things more concise, let $\sum_{i=1}^{n} X_i = K$. Thus:

$$LL(p) = K log\, p + (n - K) log(1 - p)$$

Now we differentiate with respect to p and equate to expression to 0

$$\frac{\partial LL(p)}{\partial p} = \frac{K}{p} - \frac{n-K}{1-p} = 0$$

$$\frac{K}{p} = \frac{n-K}{1-p}$$

$$K - Kp = pn - Kp$$

$$p = \frac{K}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Hence, the MLE is the sample mean

---

2. Maximum Likelihood Estimation for a Normal Distribution

Derive the maximum likelihood estimate for a group of observations that is normally distributed, but with unknown mean and variance. How does the maximum likelihood estimate compare to the standard estimates of an unknown mean and variance?

Answer:

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

To simplify things as we saw in the readings, we can take the log likelihood and are guaranteed to find the maximum value for the parameters

$$LL(\mu, \sigma) = \sum_{i=1}^{n} \left[ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2 \right]$$

We can simplify again with the laws of logarithm

$$LL(\mu, \sigma) = \sum_{i=1}^{n} \left[ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}(X_i - \mu)^2 \right]$$

Next we find the MLE of μ. To do so, we differentiate with respect to μ and equate the expression to 0

MLE of μ

$$\frac{\partial LL(\mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \left[-\frac{1}{2\sigma^2}(X_i - \mu)^2 \right] = 0$$

Using chain rule, we get:

$$\frac{\partial LL(\mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \left[-\frac{1}{\sigma^2}(X_i - \mu) \right] = 0$$

$$\frac{\partial LL(\mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} -\frac{1}{\sigma^2} X_i \sum_{i=1}^{n} \frac{1}{\sigma^2} \mu = 0$$

Move the left term to the other-side and divide b/s by $-1/\sigma^2$

$$\sum_{i=1}^{n} \mu = \sum_{i=1}^{n} X_i$$

Dividing both sides by n, we get:

$$\mu = \frac{\sum_{i=1}^{n} X_i}{n}$$

Not so sure what the standard estimate of an unknown mean is but this is my interpretation:

Assuming we are interested in the population mean, this is often difficult to determine (i.e. it's true value is unkown). Hence to estimate it, we get a sample and compute the mean of the sample. This is exactly what the MLE estimate of the mean does. Hence, in my interpretation the MLE estimate is equal to the standard estimate of an unknown mean.

MLE of σ

$$LL(\mu, \sigma) = \sum_{i=1}^{n} \left[ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}(X_i - \mu)^2 \right]$$

Applying the sum/difference rule, we get

$$\frac{\partial LL(\mu, \sigma)}{\partial \sigma} = \sum_{i=1}^{n} \left[\frac{1}{\sigma} - \frac{\partial}{\partial \sigma} \left(\frac{1}{2\sigma^2}(X_i - \mu)^2 \right) \right] = 0$$

Treat X_i and μ as constants:

$$\frac{\partial LL(\mu,\sigma)}{\partial\sigma} = \sum_{i=1}^{n}\left[\frac{1}{\sigma} - \frac{(X_i-\mu)^2}{2}\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma^2}\right)\right] = 0$$

Using the power rule, we get:

$$\frac{\partial LL(\mu,\sigma)}{\partial\sigma} = \sum_{i=1}^{n}\left[\frac{1}{\sigma} - \frac{(X_i-\mu)^2}{\sigma^3}\right] = 0$$

$$\frac{\partial LL(\mu,\sigma)}{\partial\sigma} = \sum_{i=1}^{n}\frac{\sigma^2-(X_i-\mu)^2}{\sigma^3} = 0$$

Multiply b/s by σ^3

$$= n\sigma^2 = \sum_{i=1}^{n}(X_i-\mu)^2$$

$$\therefore \sigma^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$$

The MLE estimate variance is almost close to the standard estimate for the variance. However, the denominator is different: n as against n-1. It is well known that the denominator n underestimates the variance thus the MLE estimate also underestimates this parameter.