CSE 312: Foundations of Computing II

Section 9: Concentration Inequalities and Maximum Likelihood Solutions

0. Get an Inch, Take a MLE

Suppose $x_1, \ldots, x_n$ are iid realizations from density

$$f_X(x \mid \theta) = \begin{cases} \frac{\theta x^{\theta-1}}{3^\theta}, & 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

Find the MLE for θ.

Solution:

$$\begin{aligned}
L(x_1, \ldots, x_n \mid \theta) &= \prod_{i=1}^{n} \frac{\theta x_i^{\theta-1}}{3^\theta} \\
\ln L(x_1, \ldots, x_n \mid \theta) &= \sum_{i=1}^{n} (\ln\theta + (\theta-1)\ln x_i - \theta \ln 3) \\
\frac{\partial}{\partial\theta} \ln L(x_1, \ldots, x_n \mid \theta) &= \sum_{i=1}^{n} \left(\frac{1}{\theta} + \ln x_i - \ln 3\right) = 0 \\
\frac{n}{\hat{\theta}} + \sum_{i=1}^{n} \ln x_i - n\ln 3 &= 0 \\
\frac{n}{\hat{\theta}} &= n\ln 3 - \sum_{i=1}^{n} \ln x_i \\
\hat{\theta}_{\text{MLE}} &= \frac{n}{n\ln 3 - \sum_{i=1}^{n} \ln x_i}
\end{aligned}$$

Check that it is a maximum by showing the second derivative is negative for all values of θ.

$$\frac{\partial^2}{\partial\theta^2} \ln L(x_1, \ldots, x_n \mid \theta) = \sum_{i=1}^{n} \left(-\frac{1}{\theta^2}\right) = -\frac{n}{\theta^2} < 0$$

so $\ln L(x_1, \ldots, x_n \mid \theta)$ is concave downward everywhere.
Since it's concave downward everywhere, the only criticcal point is a maximum.

1. Independent Shreds, You Say?

You are given 100 independent samples $x_1, x_2, \ldots, x_{100}$ from Bernoulli(p), where p is unknown. These 100 samples sum to 30. You would like to estimate the distribution's parameter p. Give all answers to 3 significant digits.

(a) What is the maximum likelihood estimator $\hat{p}$ of p?

Solution:

Note that $\Sigma_{i\in[n]} x_i = 30$, as given in the problem spec. Therefore, there are 30 heads and 70 tails. Therefore, we can setup L as follows,

$$\begin{aligned}
L(x_1, \ldots, x_n \mid p) &= (1-p)^{70} p^{30} \\
\ln L(x_1, \ldots, x_n \mid p) &= 70 \ln(1-p) + 30 \ln p \\
\frac{\partial}{\partial p} \ln L(x_1, \ldots, x_n \mid p) &= -\frac{70}{1-p} + \frac{30}{p} = 0 \\
\frac{30}{\hat{p}} &= \frac{70}{1-\hat{p}} \\
30 - 30\hat{p} &= 70\hat{p} \\
\hat{p} &= \frac{30}{100}
\end{aligned}$$

(b) Is $\hat{p}$ an unbiased estimator of p?

Solution:

$$\begin{aligned}
\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{1}{100}\sum_{i=1}^{100} x_i\right] & \\
= \frac{1}{100}\sum_{i=1}^{100} \mathbb{E}[x_i] & \\
= \frac{1}{100} \cdot 100p & = p.
\end{aligned}$$

so it is unbiased.

2. Y Me?

Let $Y_1, Y_2, ... Y_n$ be i.i.d. random variables with density function

$$f_Y(y|\sigma) = \frac{1}{2\sigma}\exp(-\frac{|y|}{\sigma})$$

.

Find the the MLE for σ in terms of $|y_i|$.

Solution:

$$\begin{aligned}
L(y_1, \ldots, y_n \mid \sigma) &= \prod_{i=1}^{n} \frac{1}{2\sigma}\exp(-\frac{y_i}{\sigma}) \\
\ln L(y_1, \ldots, y_n \mid \sigma) &= \sum_{i=1}^{n}\left[-\ln 2 - \ln \sigma - \frac{|y_i|}{\sigma}\right] \\
\frac{\partial}{\partial \sigma} \ln L(y_1, \ldots, y_n \mid \sigma) &= \sum_{i=1}^{n}\left[-\frac{1}{\sigma} + \frac{|y_i|}{\sigma^2}\right] = 0 \\
-\frac{n}{\hat{\sigma}} + \sum_{i=1}^{n} \frac{\Sigma_{i=1}^{n}|y_i|}{\hat{\sigma}^2} &= 0 \\
\hat{\sigma} = \frac{\Sigma_{i=1}^{n}|y_i|}{n} &
\end{aligned}$$

3. It Means Nothing

(a) Suppose $x_1, x_2, \ldots, x_n$ are samples from a normal distribution whose mean is known to be zero, but whose variance is unknown. What is the maximum likelihood estimator for its variance?

Solution:

Before we begin, we should note that this derivation will have to be with respect to σ^2, not σ. Therefore, we want to analyze the function $L(x_1, \ldots, x_n \mid \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x-\mu)^2}{2\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x-\mu)^2}{2\sigma^2}$.

$$\begin{aligned}
L(x_1, \ldots, x_n \mid \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x-\mu)^2}{2\sigma^2} \\
\ln L(x_1, \ldots, x_n \mid \sigma^2) &= \sum_{i=1}^{n} -\ln\sqrt{2\pi\sigma^2} - \frac{x_i^2}{2\sigma^2} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\ln 2\pi\sigma^2 - \frac{x_i^2}{2\sigma^2} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{x_i^2}{2\sigma^2} \\
&= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{\Sigma_{i=1}^{n} x_i^2}{2\sigma^2} \\
\frac{\partial}{\partial\sigma^2}\ln L(x_1, \ldots, x_n \mid \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{\Sigma_{i=1}^{n} x_i^2}{2\sigma^4} = 0 \\
\frac{\Sigma_{i=1}^{n} x_i^2}{2\sigma^4} &= \frac{n}{2\sigma^2} \\
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2
\end{aligned}$$

(b) Suppose the mean is known to be μ but the variance is unknown. How does the maximum likelihood estimator for the variance differ from the maximum likelihood estimator when both mean and variance are unknown?

Solution:

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

vs.

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

(The former turns out to be unbiased, the latter biased.)