Ma 3b Practical – Recitation 6

February 20, 2025

Recall definitions of likelihood function and maximal likelihood estimation. Also introduce the log-likelihood function and show why it's important (when doing derivative it's easy for calculation).

Exercise 1. (Discrete MLE)

Suppose one wishes to determine just how biased an unfair coin is. Call the probability of tossing a 'head' p.

Suppose the outcome is 49 heads and 31 tails, and suppose the coin was taken from a box containing three coins: one which gives heads with probability $p = 1/3$, one which gives heads with probability $p = 1/2$ and another which gives heads with probability $p = 2/3$. The coins have lost their labels, so which one it was is unknown. What is the maximum likelihood estimator for p?

Exercise 2. (Continuous MLE)

If $X_1, X_2, \ldots, X_n$ are i.i.d. $N\left(\mu, \sigma^2\right)$, find the maximum likelihood estimator for μ and σ.

Remark: Bernoulli, Normal, Poisson samples all satisfy that the MLE is the sample mean, but it's not true for all distributions. We will see a counterexample.

Exercise 3. (Discret MLE)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with the probability density function $f(x \mid \theta)$, where if $\theta = 0$, then

$$f(x \mid \theta) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases},$$

while if $\theta = 1$, then

$$f(x \mid \theta) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the MLE of θ.

Exercise 4. (An Example of Non-Expotential Family)

The Cauchy distribution is the probability distribution with the following probability density function (PDF)

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi}\left[\frac{\gamma}{(x - x_0)^2 + \gamma^2}\right]$$

Equivalently, we could use cumulative distribution function to describe it by

$$F(x; x_0, \gamma) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$

Still we have a size n samples $X_1, \cdots, X_n$ i.i.d. with PDF $f(x; x_0, \gamma)$. Use log-maximal likelihood function estimation to give a system of parameter x_0, γ. Prove that Cauchy distribution doesn't lie in the exponential family, i.e., $\widehat{x_0} \neq \frac{x_1 + \cdots + x_n}{n}$ in general.

Exercise 5. (Mean square error and Bias)

Still consider $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$.

1. To estimate $\theta = \mu$, we consider the estimator $\hat{\theta} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i)$. Is it biased? What is the MSE of this estimator?
2. To estimate σ^2, we consider consider $S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$. Is it biased?

Remark on MSE: If the true value of the quantity being measured is denoted by x_0, *the measurement,* X, *is modeled as*

$$X = x_0 + \beta + \varepsilon$$

where β *is the constant, or systematic, error and* ε *is the random component of the error;* ε *is a random variable with* $E(\varepsilon) = 0$ *and* $\mathrm{Var}(\varepsilon) = \sigma^2$. *We then have* $E(X) = x_0 + \beta$ *and* $\mathrm{Var}(X) = \sigma^2$. *Here* β *is often called the bias of the measurement procedure.*

Remark: one can show that $MSE = \sigma^2 + \beta^2$ *by considering variance of the variable* $X - x_0$.

Solution. Exercise 1

$$\mathbb{P}\left[H = 49 \,\middle|\, p = \frac{1}{3}\right] = \binom{80}{49}\left(\frac{1}{3}\right)^{49}\left(1 - \frac{1}{3}\right)^{31} \approx 0.000$$
$$\mathbb{P}\left[H = 49 \,\middle|\, p = \frac{1}{2}\right] = \binom{80}{49}\left(\frac{1}{2}\right)^{49}\left(1 - \frac{1}{2}\right)^{31} \approx 0.012$$
$$\mathbb{P}\left[H = 49 \,\middle|\, p = \frac{2}{3}\right] = \binom{80}{49}\left(\frac{2}{3}\right)^{49}\left(1 - \frac{2}{3}\right)^{31} \approx 0.054$$

The likelihood is maximized when $p = \frac{2}{3}$, and so this is the maximum likelihood estimate for p.

Solution. Exercise 2

Their joint density is the product of their marginal densities:

$$f(x_1, x_2, \ldots, x_n \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i - \mu}{\sigma}\right]^2\right)$$

Regarded as a function of μ and σ, this is the likelihood function. The log likelihood is thus

$$l(\mu, \sigma) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

The partials with respect to μ and σ are

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

Setting the first partial equal to zero and solving for the mle, we obtain

$$\hat{\mu} = \bar{X}$$

Setting the second partial equal to zero and substituting the mle for μ, we find that the MLE for σ is

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}$$

Solution. Exercise 3

We started with the calculation of likelihood function

$$l(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \begin{cases} 1 & \text{if } 0 < x_i < 1, \forall i \quad \& \quad \theta = 0 \\ \frac{1}{2^n \prod_{i=1}^{n} \sqrt{x_i}} & \text{if } 0 < x_i < 1, \forall i \quad \& \quad \theta = 1 \\ 0 & \text{otherwise} \end{cases},$$

Then we get the MLE of θ to be 0 if $0 < x_i < 1$ for all i and $1 \geqslant \frac{1}{2^n \prod_{i=1}^{n} \sqrt{x_i}}$, equivalently, $\prod_{i=1}^{n} x_i \geqslant 4^{-n}$. Conversely, the MLF of θ would be 1 if $0 < x_i < 1$ for all i and $\prod_{i=1}^{n} x_i \leqslant 4^{-n}$. For other parts, the likelihood function would always be zero and MLF would be meaningless (when you choose θ to be 0 or 1 you get the same likelihood function).

Solution. Exericse 4

Now we consider the likelihood function to be

$$l(x_1, \cdots, x_n|x_0, \gamma) = \frac{\gamma^n}{\pi^n} \prod_{i=1}^{n} \frac{1}{(x_i - x_0)^2 + \gamma^2}.$$

Then it's easy to get the log-likelihood function $\hat{l} = n\log(\gamma) - n\log(\pi) - \sum_{i=1}^{n} \log(\gamma^2 + (x_i - x_0)^2) = -n\log(\gamma\pi) - \sum_{i=1}^{n} \log\left(1 + \left(\frac{x_i - x_0}{\gamma}\right)^2\right)$. By taking the derivative of γ and x_0, we get the estimation system

$$\frac{d\hat{l}}{dx_0} = \sum_{i=1}^{n} \frac{2\,(x_i - x_0)}{\gamma^2 + (x_i - x_0)^2} = 0$$

$$\frac{d\hat{l}}{d\gamma} = \sum_{i=1}^{n} \frac{2\,(x_i - x_0)^2}{\gamma\left(\gamma^2 + (x_i - x_0)^2\right)} - \frac{n}{\gamma} = 0$$

Generally, we choose $x_1 = \cdots = x_{n-1} = 0$ and $x_n = 1$, then if the MLF of $\widehat{x_0}$ is $1/n$, we have from the first equation

$$2 * (-\frac{n-1}{n})\frac{1}{\gamma^2 + \frac{1}{n^2}} + 2 * (\frac{n-1}{n}) + \frac{1}{\gamma^2 + \frac{(n-1)^2}{n^2}} = 0,$$

which leads to contradiction if $n \geqslant 3$. Thus, Cauchy distribution doesn't lie in the expotential family.

Solution. Exercise 5

For iid normal distribution

- To estimate μ, the estimator has $E = \mu$ so it's not biased.

$$\mathrm{MSE}(\bar{X}) = E\left((\bar{X} - \mu)^2\right) = \left(\frac{\sigma}{\sqrt{n}}\right)^2$$

- To estimate σ^2,

  1. First we see an alternative formula

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right)$$

Proof.

$$\begin{aligned} S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} X_i^2 - 2\bar{X}_n \sum_{i=1}^{n} X_i + n\bar{X}_n^2\right\} \\ &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} X_i^2 - 2\bar{X}_n \cdot n\bar{X}_n + n\bar{X}_n^2\right\} \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right). \end{aligned}$$

Then we can compute the expectation of this estimator and find that it's not biased.

$$\begin{aligned} E\left[S_n^2\right] &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} E\left[X_i^2\right] - nE\left[\bar{X}_n^2\right]\right\} \\ &= \frac{1}{n-1}\left\{\sum_{i=1}^{n}\left(\mathrm{Var}\left(X_i\right) + \left(E\left[X_i\right]\right)^2\right) - n\left(\mathrm{Var}\left(\bar{X}_n\right) + \left(E\left[\bar{X}_n\right]\right)^2\right)\right\} \\ &= \frac{1}{n-1}\left\{\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\} \\ &= \frac{1}{n-1}\left\{(n-1)\sigma^2\right\} \\ &= \sigma^2 \end{aligned}$$

One can also compute MSE : if $X_1, \cdots, X_n$ come from a normal distribution with variance σ^2, then the sample variance S^2 is defined as

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}$$

It can be shown that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. From the properties of χ^2 distribution, we have

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1 \Rightarrow E\left(S^2\right) = \sigma^2$$

and

$$\mathrm{Var}\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1) \Rightarrow \mathrm{Var}\left(S^2\right) = \frac{2\sigma^4}{n-1}$$

remark: we only used here these variables are iid, with same expectation and variance, then this estimator is always unbiased.