# MAXIMUM LIKELIHOOD

Mauro Díaz

**15.11**

**Exercise 15.11 (The German tank problem)** During WWII the allied intelligence were faced with an important problem of estimating the total production of certain German tanks, such as the Panther. What turned out to be a successful approach was to estimate the maximum from the serial numbers of the small sample of captured or destroyed tanks (describe the statistical model used).

a. What assumptions were made by using the above model? Do you think they are reasonable assumptions in practice?

b. Show that the plug-in estimate for the maximum (i.e. the maximum of the sample) is a biased estimator.

c. Derive the maximum likelihood estimate of the maximum.

d. Check that the following estimator is not biased: $\hat{n} = \frac{k+1}{k} m - 1$.

First of all, we know that the data will be the serial numbers of the tanks. We are trying to estimate the number $n$ = total production of tanks.

(a) First, we will assume that every serial number is equally likely to appear in the sample, therefore: $\quad Y_i \overset{iid}{\sim} \text{Uniform}(0, \dots, n)$

We are also assuming that the samples will be i.i.d., when in practice might not be true.

Why estimate $n$ with the maximum?

↳ $n$ only affects the upper tail of the ECDF distribution

↳ the maximum carries the information of $n$

(b) If $Y_1, \dots, Y_k \overset{iid}{\sim} \mathcal{U}(0, \dots, n)$ is the sample, the maximum RV will be $\max(Y_1, \dots, Y_k) = M$.

First of all, we need to compute the distribution of $M$, therefore we need to compute $P(M \le x)$ (How prob. is that the value $x$ is maximum of the sample)

But

$$P(M \le x) \iff \{Y_1 \le x, \dots, Y_k \le x\}$$

$$P(Y_1 \le x, \dots, Y_k \le x) = \frac{\overbrace{\binom{x}{k}}^{\text{samples where all numbers are } \le x}}{\underbrace{\binom{n}{k}}_{\text{total possible samples}}}$$

The distribution of $M$ is discrete, so we need to compute the PMF.

$$p_M(x) = P(M \le x) - P(M \le x-1) = \frac{\binom{x}{k} - \binom{x-1}{k}}{\binom{n}{k}} = \frac{\binom{x-1}{k-1}}{\binom{n}{k}}$$

Now, with the PMF, we can compute the expected value of $M$.

$$\mathbb{E}[M] = \sum_{i=k}^{n} i \, \frac{\binom{i-1}{k-1}}{\binom{n}{k}} = \sum_{i=k}^{n} i \, \frac{\frac{(i-1)!}{(k-1)!(i-k)!}}{\binom{n}{k}} \cdot \frac{k}{k} = \sum_{i=k}^{n} \frac{k}{\binom{n}{k}} \, \frac{i(i-1)!}{k(k-1)!(i-k)!} =$$

$$= \frac{k}{\binom{n}{k}} \underbrace{\sum_{i=k}^{n} \binom{i}{k}}_{\text{Hockey-Stick identity}} = \frac{k}{\binom{n}{k}} \binom{n+1}{k+1} \underset{\text{terms cancel}}{=} \frac{k(n+1)}{k+1} \implies \text{Unbiased !}$$

The bias is: $\mathbb{E}[M] - n = \frac{k(n+1)}{k+1} - n = \frac{k-n}{k+1}$

(c)

As we are assuming that $\forall i$: $Y_i$ are i.i.d and $P(Y_i = y_i | n) = \frac{1}{n}$ $\forall y_i \in \{1, \dots, n\}$ therefore:

$$\mathcal{L}(y|n) = P(Y_1 = y_1, \dots, Y_k = y_k | n) = \prod_{i=1}^{n} P(Y_i = y_i | n) = \prod_{i=1}^{n} \frac{1}{n} = n^{-k}$$

This is a decreasing function respect to $n$, and we are assuming that at least $n$ has to be higher than $\max(y_1, \dots, y_k)$ of the sample.
Therefore: $n \geq \max(y_1, \dots, y_k) \implies \text{MLE}(n) = \max(y_1, \dots, y_k)$

(d) $m$ is the maximum estimator and $\mathbb{E}[M=m] = \frac{k(n+1)}{k+1}$

$$\mathbb{E}[\hat{n}] = \frac{k+1}{k}\mathbb{E}[m] - 1 = \frac{\cancel{k+1}}{\cancel{k}} \, \frac{\cancel{k}(n+1)}{\cancel{k+1}} - 1 = (n+1) - 1 = n \to \text{Unbiased}$$

$\boxed{15.2}$

**Exercise 15.2 (Multivariate normal distribution)**

a. Derive the maximum likelihood estimate for the mean and covariance matrix of the multivariate normal.

b. Simulate $n = 40$ samples from a bivariate normal distribution (choose non-trivial parameters, that is, mean $\neq 0$ and covariance $\neq 0$). Compute the MLE for the sample. Overlay the data with an ellipse that is determined by the MLE and an ellipse that is determined by the chosen true parameters.

c. Repeat b. several times and observe how the estimates (ellipses) vary around the true value.

Hint: For the derivation of MLE, these identities will be helpful: $\frac{\partial b^T a}{\partial a} = \frac{\partial a^T b}{\partial a} = b$, $\frac{\partial a^T A a}{\partial a} = (A + A^T)a$, $\frac{\partial \operatorname{tr}(BA)}{\partial A} = B^T$, $\frac{\partial \ln|A|}{\partial A} = (A^{-1})^T$, $a^T A a = \operatorname{tr}(a^T A a) = \operatorname{tr}(a a^T A) = \operatorname{tr}(A a a^T)$.

(a) We have $Y_1, \dots, Y_n \overset{iid}{\sim} N(\mu, \Sigma)$ $k$-random vectors and for each of them the density function is:

$$f(x_1, \dots, x_k) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

Therefore,

$$\mathcal{L}(y;\mu,\Sigma^{-1}) = \prod_{i=1}^{n} f(y_i \mid \mu,\Sigma^{-1}) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)} =$$

$$= \frac{1}{(2\pi)^{nk/2}|\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)}$$

$$\ell(y;\mu,\Sigma^{-1}) = \log\mathcal{L}(y;\mu,\Sigma^{-1}) =$$

$$= -\frac{nk}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)$$

Now, taking derivatives, we can find the maximum for each parameter

$\boxed{\text{Mean } \mu}$

$$\frac{\partial}{\partial\mu}\ell(y;\mu,\Sigma^{-1}) = \frac{\partial}{\partial\mu}\left(-\frac{nk}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial\mu}\left((y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right) = -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial\mu}\left(y_i^T\Sigma^{-1}y_i - \mu^T\Sigma^{-1}y_i - y_i^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right)$$

applying properties of the hints

$\Sigma$ symmetric

$$= -\frac{1}{2}\sum_{i=1}^{n}\left(-\Sigma^{-1}y_i - y_i\Sigma^{-1} + \Sigma^{-1}\mu + (\Sigma^{-1})^T\mu\right) = -\Sigma^{-1}\sum_{i=1}^{n}\mu - y_i = -\Sigma^{-1}\left(n\mu - \sum_{i=1}^{n}y_i\right)$$

and $\frac{\partial}{\partial\mu}\ell(y;\mu,\Sigma^{-1}) = 0 \iff \boxed{\mu = \frac{1}{n}\sum_{i=1}^{n}y_i}$

$\boxed{\text{Covariance } \Sigma}$

Due by the linearity with $\Sigma^{-1}$, we will derive respect the inverse:

$$\frac{\partial}{\partial\Sigma^{-1}}\ell(y;\mu,\Sigma^{-1}) = \frac{\partial}{\partial\Sigma^{-1}}\left(-\frac{n}{2}\overbrace{\log(|\Sigma|)}^{-\log(|\Sigma^{-1}|)} - \frac{1}{2}\sum_{i=1}^{n}\underbrace{(y_i-\mu)^T}_{a}\Sigma^{-1}\underbrace{(y_i-\mu)}_{a}\right) =$$

$$= \frac{\partial}{\partial\Sigma^{-1}}\left(+\frac{n}{2}\log(|\Sigma^{-1}|) - \frac{1}{2}\sum_{i=1}^{n}\operatorname{tr}\left((y_i-\mu)(y_i-\mu)^T\Sigma^{-1}\right)\right) =$$

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

and $\frac{\partial}{\partial\Sigma^{-1}}\ell(y;\mu,\Sigma^{-1}) = 0 \iff \boxed{\Sigma = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)(y_i-\mu)^T}$