14.381 STATISTICS 11/12/2004

HANDOUT # 7

TA: JOSÉ TESSADA
(tessada@mit.edu)

Suppose you observe the following: you collect the number of goals scored by the Chilean national team against Argentina in soccer matches (unfortunately I do not expect to observe a very high number of goals). You are asked to give an estimation of the number of goals the Chileans will score in the next game, what would you do?

The previous situation basically describes a problem where you want to use data to gain some knowledge about the true distribution (you want to make some inference).
For now, we will deal with situation where you assume the data comes from a population described by a pdf or pmf $f(x|\theta)$, so by knowing θ we can describe the whole population.

Definition 7.1.1 A point estimator is any function $W(X_1, \ldots, X_n)$ of a sample; that is, any statistic is a point estimator.

This is not particularly useful, because it doesn't give us any guidance about what functions are more "relevant".
$\Rightarrow$ we will also care about ways to obtain estimators that are "relatively good" in a certain sense.

① <u>Method of Moments</u>: has the virtue of simplicity, but the estimators you obtain generally require some additional "refinement".

Def.: Let $X_1, \ldots, X_n$ be a sample from a population with pdf or pmf $f(x \mid \theta_1, \ldots, \theta_k)$. Method of moments estimators are found by equating the first k sample moments to the corresponding k population moments, and solving the resulting system of simultaneous equations.

Note: you need to be able to compute the moments (you need the moments to exist) in order to estimate the parameters — ex. question 3 in problem set 7.

② <u>Maximum Likelihood Estimators</u>: for each sample point x, let $\theta(x)$ be a parameter value at which $L(\theta \mid x)$ attains its maximum as a function of θ, with x held fixed. A maximum likelihood estimator (MLE) of the parameter θ based on a sample X is $\hat{\theta}(X)$ [Definition 7.2.4]

Invariance property of maximum likelihood estimators: suppose we are interested in estimating $\tau(\theta)$ instead of the parameter θ itself; the invariance property tells us that if $\hat{\theta}$ is the MLE of θ, then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$. [Theorem 7.2.10]

Note: for those of you who are really interested in the topic, C&B cover some technical issues about ML estimation in ch. 7.

Methods of Evaluating Estimators

The purpose here is to find certain criteria to compare different estimators. As you will see it is not true that we can find "the" estimator, but we can define certain elements to guide us in the selection or evaluation process.

Definition 7.3.1: The mean squared error (MSE) of an estimator W of a parameter θ is the function of θ defined by $E_\theta(W-\theta)^2$.

We can rewrite the MSE as:

$$E_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + (\mathrm{Bias}_\theta W)^2$$

Definition 7.3.2: the bias of a point estimator W of a parameter θ is the difference between the expected value of W and θ, that is, $\mathrm{Bias}_\theta = E_\theta W - \theta$. An estimator whose bias is identically (in θ) equal to 0 is called unbiased.

if $E_\theta W = \theta \quad \forall\, \theta \in \Theta \Rightarrow W$ is unbiased

EXERCISES

① Let $X_1, \dots, X_n$ be a random sample drawn from a Poisson(λ) population.

(a) find a method of moments estimator;

(b) find the MLE of λ.

(a) we know that $E(X_i) = \lambda$

$$\Rightarrow \hat{\lambda}_{MM} = \sum_i \frac{X_i}{n}$$

(b) Write the likelihood as follows:

$$L(\theta|x) = \frac{e^{-\lambda n} \lambda^{\sum_i X_i}}{\prod_i X_i!}$$

$$\Rightarrow \log L = -\lambda n + \sum_i X_i \log(\lambda) - \sum_i \log(X_i!)$$

$$\frac{\partial \log L}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_i X_i$$

set $\frac{\partial \log L}{\partial \lambda} = 0$ and obtain

$$\hat{\lambda}_{ML} = \frac{\sum_i X_i}{n} = \bar{X}$$

② A random sample $X_1, \dots, X_n$ is drawn from a population that is $N(\theta, \theta)$, $\theta > 0$.

Show that the MLE of θ, $\hat{\theta}$, is a root of the quadratic equation $\theta^2 + \theta - W = 0$, where $W = \frac{1}{n} \sum_{i=1}^{n} X_i^2$, and determine which root equals the MLE.

$$\log L(\theta|X) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \theta - \frac{1}{2\theta} \sum (X_i - \theta)^2$$

$$\frac{\partial \log L}{\partial \theta} = -\frac{n}{2}\frac{1}{\theta} + \frac{1}{2}\frac{1}{\theta^2}\sum(x_i-\theta)^2 + \frac{1}{\theta}\sum(x_i-\theta)$$

$$-\frac{n}{2}\frac{1}{\hat\theta} + \frac{1}{2}\frac{1}{\hat\theta^2}\sum(x_i-\hat\theta)^2 + \frac{1}{\hat\theta}\sum(x_i-\hat\theta) = 0$$

$$-\frac{n}{2}\hat\theta + \frac{1}{2}\sum(x_i-\hat\theta)^2 + \hat\theta\sum(x_i-\hat\theta) = 0$$

$$-\frac{n}{2}\hat\theta + \frac{1}{2}\sum x_i^2 - \cancel{\hat\theta\sum x_i} + \frac{1}{2}n\hat\theta^2 + \cancel{\hat\theta\sum x_i} - n\hat\theta^2 = 0$$

$$-\frac{n}{2}\hat\theta + \frac{1}{2}\sum x_i^2 - \frac{1}{2}n\hat\theta^2 = 0$$

$$-\hat\theta + \frac{\sum x_i^2}{n} - \hat\theta^2 = 0$$

$$\hat\theta^2 + \hat\theta - \frac{\sum x_i^2}{n} = 0$$

roots are $\dfrac{-1 \pm \sqrt{1+4W}}{2}$ $\qquad W = \dfrac{\sum x_i^2}{n}$ as defined before

$\hat\theta$ is the positive root of that equation

③ C&B 7.13

Let $X_1, \ldots, X_n$ be a sample from a population with double exponential pdf.

$$f(x|\theta) = \frac{1}{2}\exp\{-|x-\theta|\}, \quad \begin{matrix} -\infty < x < \infty \\ -\infty < \theta < \infty \end{matrix}$$

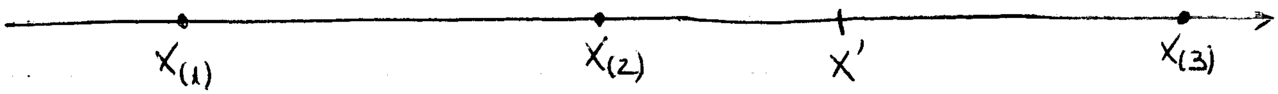
Find the MLE of θ

$$L(\theta|x) = \prod_i \frac{1}{2}\exp\{-|x_i-\theta|\} = \frac{1}{2^n}\exp\{-\sum_i |x_i-\theta|\}$$

so the MLE minimizes $\sum_i |x_i-\theta| = \sum_i |x_{(i)}-\theta|$,

where $X_{(1)}, \ldots, X_{(n)}$ are the order statistics.

Now comes the somewhat complicated part:
suppose $n=3$



Choose x', there $\sum |X_i - x'| = x' - X_{(1)} + x' - X_{(2)} + X_{(3)} - x'$
Now choose $x'' = x' - \varepsilon$, $\varepsilon > 0$, $\varepsilon < x' - x_{(2)}$

Now
$$|x'' - X_{(1)}| = (x' - X_{(1)}) - \varepsilon$$
$$|x'' - X_{(2)}| = (x' - X_{(2)}) - \varepsilon$$
$$|x'' - X_{(3)}| = (X_{(3)} - x') + \varepsilon$$

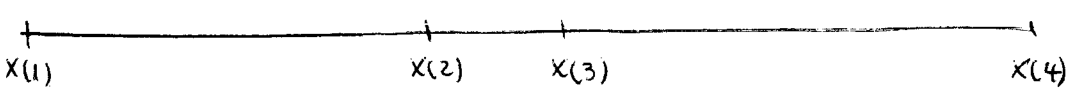
net change : $-\varepsilon$

Now, this is true for any $x \in (X_{(2)}, x']$, where x' is an arbitrary point in $(X_{(2)}, X_{(3)}]$

The reverse is true in the case of $x' \in [X_{(1)}, X_{(2)})$, in that case you want to move to the right until you reach $X_{(2)}$.

Then $X_{(2)}$ is the only point such that the change in the obj. function is >0 as you move away

Think now about $n=4$



In this case the previous reasoning applies to any point in the intervals $[X_{(1)}, X_{(2)})$ and $(X_{(3)}, X_{(4)}]$. However for any $x \in [X_{(2)}, X_{(3)}]$, the obj. function is completely flat $\Rightarrow$ the whole set minimizes the sum of the absolute deviations.
$\Rightarrow$ if n is even, MLE is any point between $X_{(n/2)}$ and $X_{(n/2)+1}$; if n is odd, $\hat{\theta} = X_{((n/2 + 1/2))}$.

④ C&B 7.20

$Y_1, \dots, Y_n$ random variables

$$Y_i = \beta X_i + \varepsilon_i, \qquad i = 1, \dots, n$$

where $x_1, \dots, x_n$ are fixed constants, and $\varepsilon_1, \dots, \varepsilon_n$ are iid $N(0, \sigma^2)$, σ^2 unknown.

(a) Show that $\sum Y_i / \sum X_i$ is an unbiased estimator of β

(b) Calculate the exact variance of $\dfrac{\sum Y_i}{\sum X_i}$