Example 2 (cont'd)

Indeed, let $Y_i \sim N(\mu, \sigma^2)$. Then $\overline{Y}$ and S^2 are independent and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

It follows that $\text{var}(S^2) = 2(\sigma^2)^2/(n-1)$. Furthermore, combining this with the results of Example 1, we have that

$$\frac{\frac{\sqrt{n}(\overline{Y}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\sqrt{n}(\overline{Y}-\mu)}{S} \sim t_{n-1}.$$

Quantities s.t. $(n-1)S^2/\sigma^2$ or $\sqrt{n}(\overline{Y}-\mu)/S^2$, which are not statistics but have a known distribution, are called pivotal quantities.

They play a key role in the development of confidence intervals and hypothesis testing as we will see in L5 and L6.

Example 3

Let $Y_1, \ldots, Y_n$ be an iid random sample with $Y_i \sim \text{Unif}(0, \theta)$, with $\theta > 0$. For the statistics $\overline{Y}$ and $Y_{(n)}$, let's compute some of their features.

For the sample average we have

$$E(\overline{Y}) = E(Y_1) = \theta/2, \quad \text{var}(\overline{Y}) = E((\overline{Y})^2) - E(\overline{Y})^2 = \theta^2/(12n).$$

For the maximum we have

$$E(X_{(n)}) = \int_{-\infty}^{+\infty} t f_{X_{(n)}}(t)dt = \int_0^{\theta} t \cdot nF(t)^{n-1} f(t)dt$$
$$= \int_0^{\theta} tn(t/\theta)^{n-1}(1/\theta)dt = \theta n/(n+1),$$

and

$$\text{var}(Y_{(n)}) = \theta^2 n/(n+1)^2.$$

Note that in the case of $Y_{(n)}$ we had to use it's pdf in order to compute the two moments.

Some statistics target θ the parameter of a distribution (or a component of it).

In that case, we call them estimators and denote them by a Greek letter with a hat, e.g. $\widehat{\theta}$.

An estimator thus is a function of the random sample, and can tell us something useful about the parameter θ.

For example, the sample average is useful when we want to learn about the population average μ, the sample median Q_2 is useful for learning about the population median and so on.

→ hist is an estimate of the pdf

We now look at methods for computing estimators and then we will see methods for comparing estimators.

Method of Moments

Useful when we want to estimate θ that can be expressed as a function of moments of Y.

Is one of the oldest statistical estimation methods, dating back to 1936.

The method consists in equating sample moments, e.g. $\overline{Y}$, $\overline{Y^2}$, $\overline{Y^3}, \ldots$ with the corresponding population moments, e.g. $E(Y)$, $E(Y^2)$, $E(Y^3)$ and solving these equations in terms of the parameter θ.

Example 4

Suppose $Y_1, \ldots, Y_n$ is an iid sample from some distribution F_θ and let $E(Y_1) = \mu$ be the unknown parameter.

Equating the sample moment with the corresponding population moment leads to

$$\overline{Y} = E(Y) = \mu.$$

So $\overline{Y}$ is the method of moment estimator for μ, i.e. $\widehat{\mu}_{MM} = \overline{Y}$ (reads: the method of moments estimator for μ is $\overline{Y}$).

Example 5

Let $Y_1, \ldots, Y_n$ be an iid random sample with $E(Y_1) = \mu$ and $\text{var}(Y_1) = \sigma^2$, with μ, σ^2 unknown.

First note that $\sigma^2 = E(Y^2) - E(Y)^2 = E(Y^2) - \mu^2$. Equating the first two moments leads to

estimator for μ

$$\overline{Y} = E(Y) = \mu,$$

$$E(Y^2) = \sigma^2 + \mu^2 = \overline{Y^2}.$$

"method of moments"

estimator for σ^2

We conclude that

$$\widehat{\sigma}^2_{MM} = \overline{Y^2} - \overline{Y}^2 = \widehat{\sigma}^2,$$

(we have encountered this estimator previously.) and $\widehat{\mu}_{MM} = \overline{Y}$.

So $(\overline{Y}, \widehat{\sigma}^2)$ is the method of moments estimator for (μ, σ^2).

A signal+noise problem

M.M. not useful for not i.i.d. samples ⇒ used as a starting point.

In many situations, we have measurements Y_i, which can be thought of as the sum of a physical signal $g_i(\beta)$ and a noise ϵ_i, i.e.

$$Y_i = g_i(\beta) + \epsilon_i,$$

where ϵ_i is some unknown component whereas $g_i(\beta)$ is the signal, which may depend on some unknown parameter β.

For instance, when measuring the resistance of an electronic circuit with a multimeter, we observe a realisation of Y_i, say $y_i = 12\,\Omega$.

This value depends, in part, on the real resistance of the equipment $g_i(\beta)$ (e.g. the sum of the resistance of all its component) and for the other part, on the accuracy of the instrument; see also L2 for another example.

Linar regression

The signal $g_i(\beta)$ may depend on some known features $x_{i1}, \ldots, x_{ip}$, through the linear function → not suitable for all applications

$$g_i(\beta) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \quad i = 1, \ldots, n,$$

with unknown parameters $\beta = (\beta_0, \ldots, \beta_p)$.

The aim is to understand the impact of x_{ij}'s on Y_i, i.e. β_i.

We have the pair $y_i, g_i(\beta)$, where y_i is what we measure and $g_i(\beta)$ is how the system should behave according to modeller's view.

This is commonly known as a linear regression problem, and one of the points is how to estimate β using $y_i, g_i(\beta)$ for all i.

Method of Least Squares

For any fixed β, the deviances $y_i - g_i(\beta)$, tell's us by how much our model $g_i(\beta)$ misses the observed value. It seems intuitive then to look for a β that leads to smallest deviances.

The method of least squares consists in estimating β by the value that leads to smallest sum of squared deviances.

That is, the least squares estimator is defined by

$$\widehat{\beta}_{LS} = \arg \inf_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - g_i(\beta))^2.$$

Example 6

Let $y_1, \ldots, Y_n$ be counts of bacteria in a culture of cells, measured at time points $t_1, \ldots, t_n$. Aim: study bacteria growth rate.

A possible model for this problem is

$$\begin{aligned} Y_i &= g(t_i; \beta) + \epsilon_i, \\ g(t_i; \theta) &= \beta_0 + \theta_1 t_i, \quad i = 1, \ldots, n, \end{aligned}$$

where $\beta = (\beta_0, \beta_1) \in \mathbb{R}^2$ is unknown.

In words: (we assume) the counts follow a linear equation with time, and we wish to learn about the parameters of this line.

To find the LS estimator we solve in β_0, β_1 the system

$$\frac{d}{d\beta} \sum_{i=1}^{n} (y_i - g_i(\theta))^2 = 0,$$

→ to find the minimum

Example 6 (cont'd)

This system is

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1}) = 0$$

$$\sum_{i=1}^{n} x_{i1}(y_i - \beta_0 - \beta_1 x_{i1}) = 0$$

and the solution is

$$\widehat{\beta}_0 = \overline{y} - \frac{s_{y,x}}{s_x^2}\overline{x},$$

and

$$\widehat{\beta}_1 = \frac{s_{y,x}}{s_x^2}.$$

covariance / variance

The vector $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)$ as above is the least squares estimator for β.

Method of Maximum Likelihood

Let $Y_1, \ldots, Y_n$ be an iid random sample from with $Y_i \sim F_\theta$ and pdf f and unknown parameter θ.

The Maximum Likelihood Estimator (MLE) is defined by

$$\widehat{\theta} = \arg \inf_{\theta \in \Theta} L(\theta).$$

Under standard regularity conditions, the MLE is also defined as the solution to the likelihood equation

$$\frac{d\ell(\theta)}{d\theta} = 0.$$

Note that θ may be d-dimensional vector, in which case the likelihood equation consists in d simultaneous equations.

Example 7

Let $Y_1, \ldots, Y_n$ be an iid random sample with $Y_i \sim \text{Ber}(\theta)$, with θ unknown. Let's compute the maximum likelihood estimator for θ.

$$L(\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}$$
$$= \theta^{\sum_i y_i}(1-\theta)^{n-\sum_i y_i},$$

Since $L(\theta) > 0$, for all θ, we can apply the log to get the log-likelihood function. So we solve the likelihood equation $d\ell(\theta)/d\theta = 0$, i.e.

$$\frac{\sum_i y_i}{\theta} - \frac{n-\sum_i y_i}{1-\theta} = 0,$$

to get the solution $\widehat{\theta} = \overline{y}$

Furthermore, $d^2\ell(\theta)/d\theta$ at $\theta = \widehat{\theta}$ is negative, so θ is a local maximum, thus it's the MLE of θ. In this case, the MLE coincides with the MME.

Example 8

Let $Y_1, \ldots, Y_m$ be a random vector with distribution $\text{Mn}(n, \theta_1, \ldots, \theta_m)$ with $0 < \theta_i < 1$ for all i, $\sum_i \theta_i = 1$ and $n = \sum_i Y_i$.

For example, in a sample of Chinese population of Hong Kong in 1937, blood types occur with the following frequencies, where M and N are red cell antigens

	Blood Type			
	M	MN	N	Total
Frequency	342	500	187	1029

Intuitively, the estimated probability of each blood type is the ratio of the observed frequency divided by n, i.e. $\widehat{\theta_i} = y_i/n$ for all i. This is the MLE of θ.

Indeed, the log-likelihood is

$$\ell(\theta) = \log n! - \sum_{i=1}^{3} \log y_i! + \sum_{i=1}^{3} y_i \log \theta_i.$$