Statistical Data Analysis for Physicists
Homework 3

Artur Palha nº46724
Ana Vieira da Silva nº46722

20th October 2002

Exercise 1. Sample statistics for Gaussian variables.

1. **We wish to show that the empirical mean $\bar{X}_n$ follows a normal law $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, given that $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a sample of n i.i.d. drawn from a normal law $\mathcal{N}(\mu, \sigma^2)$.**

   By definition of empirical mean, we have:

$$\bar{X}_n = \frac{1}{n}\sum_i^n X_i$$

Our route will be to find the generating function of $\bar{X}_n$ and, by inspection, conclude it is the generating function of a normal law with the desired parameters. By definition of generating function $\phi(t)$ we mean, as usual:

$$\begin{aligned}\phi_{\bar{X}_n}(t) &= \langle e^{it\bar{X}_n}\rangle \\ &= \langle e^{\frac{it}{n}\sum_i^n X_i}\rangle \\ &= \int_\Omega e^{\frac{it}{n}\sum_i^n X_i}\mathrm{d}x_1 \ldots \mathrm{d}x_n\end{aligned}$$

Noticing that the X_i are i.i.d., we can write:

$$f(x_1, \ldots, x_n) = \prod_i^n f(x_i)$$

Therefore we can rearrange the integrals in this way:

$$\phi_{\bar{X}_n}(t) = \int e^{\frac{it}{n}x_1} f(x_1)\,\mathrm{d}x_1 \ldots \int e^{\frac{it}{n}x_n} f(x_n)\,\mathrm{d}x_n$$

Since all the integrals are identical we can write:

$$\begin{aligned}\phi_{\bar{X}_n}(t) &= \left(\int e^{\frac{it}{n}x} f(x)\,\mathrm{d}x\right)^n \\ &= [\phi_g(t)]^n\end{aligned}$$

Where $\phi_g(t)$ is just the generating function of the normal distribution. Lets compute it.

$$\phi_g(t) = \frac{1}{\sigma\sqrt{2\pi}}\int e^{\frac{it}{n}x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x$$

Making the change of variables $x - \mu = u$, we get:

$$\phi_g(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{it\mu}{n}} \int e^{\frac{itu}{n} - \frac{u^2}{2\sigma^2}} \mathrm{d}u$$

Noticing that:

$$\frac{itu}{n} - \frac{u^2}{2\sigma^2} = -\left(\frac{u}{\sigma\sqrt{2}} - \frac{\sigma}{\sqrt{2}}\frac{it}{n}\right)^2 - \frac{\sigma^2}{2}\frac{t^2}{n^2}$$

We can write:

$$\begin{aligned} \phi_g(t) &= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{it\mu}{n} - \frac{\sigma^2}{2}\frac{t^2}{n^2}} \int e^{-\left(\frac{u}{\sigma\sqrt{2}} - \frac{\sigma}{\sqrt{2}}\frac{it}{n}\right)^2} \mathrm{d}u \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{it\mu}{n} - \frac{\sigma^2}{2}\frac{t^2}{n^2}} \int e^{-\frac{1}{2\sigma^2}\left(u - \sigma\frac{it}{n}\right)^2} \mathrm{d}u \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sqrt{2\pi}\sigma e^{-\left(\frac{\sigma^2}{2}\frac{t^2}{n^2} - \frac{it\mu}{n}\right)} \end{aligned}$$

Hence:

$$\begin{aligned} \phi_{\bar{X}_n}(t) &= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} (2\pi)^{\frac{n}{2}} \sigma^n e^{\frac{it\mu}{n} - \frac{\sigma^2}{2}\frac{t^2}{n^2}} \\ &= e^{\frac{it\mu}{n} - \frac{\sigma^2}{2}\frac{t^2}{n^2}} \end{aligned}$$

Which is just the generating function of a gaussian with mean μ and standard deviation $\frac{\sigma^2}{n}$.

$$\therefore \bar{X}_n \sim \mathcal{N}\left(\mu \frac{\sigma^2}{n}\right)$$

■

2. **An unbiased estimator of the variance is:**

$$S_n^2 = \frac{1}{n-1} \sum_i^n (X_i - \bar{X}_n)^2$$

2.a. **We wish to show that $\frac{(n-1)S_n^2}{\sigma^2}$ follows a χ^2 distribution with $n-1$ degrees of freedom.**

First we will start by stating and demonstrating (exhaustively or just intuitively) some relations that will be needed in solving the exercise.

i. *If $Y = g(X)$, with inverse $X = h(Y)$ single valued, and $f_X(x)$ is the p.d.f. of X then:*

$$f_Y(y) = f_X(x = h(y)) \cdot \left|\frac{\mathrm{d}h(y)}{\mathrm{d}y}\right|$$

Demonstration We know that:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y)$$

It's easy to see that:

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le h(y)) = F_X(h(y)) \tag{1}$$

Having in mind the relation:

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y}$$

and combining it with the last equality of Eq. (1) we get:

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}F_X(h(y))}{\mathrm{d}y} \\ &= f_X(h(y)) \cdot \left|\frac{\mathrm{d}h(y)}{\mathrm{d}y}\right| \end{aligned}$$

where we have used the chain rule and the positiveness of the p.d.f..

□

ii. *In certain cases the inverse might not be single valued. Suppose that for a given y there exist n values of x, let's say: $x_1, \ldots, x_n$. Then, the p.d.f. of Y will be given by:*

$$f_Y(y) = \sum_i^n f_X(x_i = h(y)) \cdot \left|\frac{\mathrm{d}h(y)}{\mathrm{d}y}\right| \tag{2}$$

This result is intuitive since if we can get Y throughout all the X_i then, the "probability" of obtaining Y will be the sum of the "probabilities" of obtaining the X_i. Hence, the result is, more or less, obvious.

□

Now we will begin solving the exercise. First we will establish the defining equation of a χ_1^2 (one degree of freedom). Second, it's generating function will be computed, followed by the generating function of a χ_n^2 (n degrees of freedom). Third and finally we will show that S_n^2 follows a χ_{n-1}^2, using for that a change a variables and a small drive-through quadratic forms.

I. Suppose, for simplicity, that $X \sim \mathcal{N}(0,1)$. Therefore:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{3}$$

Suppose that $Y = g(X) = X^2$, and the inverse function $X = h(Y) = \pm\sqrt{Y}$. We have also:

$$\left|\frac{\mathrm{d}h(y)}{\mathrm{d}y}\right| = \frac{1}{2\sqrt{y}}$$

Combining Eq. (2) with Eq. (3) gives:

$$\begin{aligned} f_Y(y) &= \frac{f_X\left(x = \sqrt{y}\right)}{2\sqrt{y}} + \frac{f_X\left(x = -\sqrt{y}\right)}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \end{aligned}$$

By definition this p.d.f. is a χ_1^2, therefore Y follows a χ_1^2.

II. It's generating function is, by definition of generating function:

$$\phi_Y(t) = \int_0^{+\infty} f_Y(y)\, e^{ity} \mathrm{d}y$$

for simplicity of the notation we will use $\phi_Y \equiv \phi_Y(t)$ and we will drop the limits of integration, focusing that, implicitly, their are still there. They will be reinserted whenever there is an ambiguity. Then:

$$\phi_Y = \frac{1}{\sqrt{2\pi}} \int e^{y(it-\frac{1}{2})} y^{-\frac{1}{2}} \mathrm{d}y$$

Consider the integral:

$$I(a,\gamma) = \int e^{y(it-a)} y^{\gamma-1} \mathrm{d}y, \quad (a>0)$$

Let $x = y(a-it)$, therefore:

$$\begin{aligned} I(a,\gamma) &= \int \frac{e^{-x} x^{\gamma-1}}{(a-it)^{\gamma}} \mathrm{d}x \\ &= \frac{\Gamma\gamma}{(a-it)^{\gamma}} \end{aligned}$$

Hence:

$$\begin{aligned} \phi_Y &= \frac{1}{\sqrt{2\pi}} I\left(\frac{1}{2},\frac{1}{2}\right) \\ &= (1-2it)^{\frac{1}{2}} \end{aligned}$$

Since the X_i are independent, we have:

$$\begin{aligned} \phi_{Y_n} &= (\phi_Y)^n \\ &= (1-2it)^{\frac{n}{2}} \end{aligned}$$

The corresponding p.d.f. is:

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi} \int_0^{+\infty} e^{-ity} \phi_{Y_n} \mathrm{d}t \\ &= \frac{1}{2\pi} \int_0^{+\infty} e^{-ity} (1-2it)^{\frac{n}{2}} \mathrm{d}t \\ &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}, \qquad (y \geq 0) \end{aligned}$$

And 0 if $y \leq 0$. Where we have used the standard result:

$$\begin{aligned} \textstyle\int_0^{+\infty} e^{-ity}(1-2it)^{-\delta} \mathrm{d}t \quad &= 2\pi y^{\delta-1} e^{-ay}, \quad (y \geq 0,\ \mathrm{Re}(a) > 0) \\ &= 0, \qquad\qquad\qquad (y < 0,\ \mathrm{Re}(\delta) > 0) \end{aligned}$$

By definition, this is a χ^2 p.d.f. with n degrees of freedom.

III. To show that:

$$\frac{1}{\sigma^2} \sum_i^n \left(X_i - \bar{X}_n\right)^2$$

follows a χ^2 with $n-1$ we have to enter a little in the field of quadratic forms. Here we will not demonstrate an important result that we will use, we just state it (you can see the proof in any algebra book).

Let XAX^t be the quadratic form associated with a real symmetric matrix A, and let C be an orthogonal matrix that converts A to a diagonal matrix $\Lambda = C^t AC$. Then we have:

$$XAX^t = Y\Lambda Y^t = \sum_i^n \lambda_i y_i^2$$

where $Y = [y_1, \ldots, y_n]$ is the row matrix $Y = XC$, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

Checking our expression, we see (we will drop the limits of the sums having in mind that they are all from 1 to n):

$$\begin{aligned}
\sum_i \left(X_i - \frac{1}{n} \sum_j X_j \right)^2 &= \sum_i \left(X_i^2 - \frac{2}{n} X_i \sum_j X_j + \frac{1}{n^2} \sum_{lm} X_l X_m \right) \\
&= \sum_i X_i^2 - \frac{2}{n} \sum_{ij} X_i X_j + \frac{1}{n} \sum_{lm} X_l X_m \\
&= \sum_{ij} \left(\delta_{ij} - \frac{2}{n} + \frac{1}{n} \right) X_i X_j \\
&= \sum_{ij} \left(\delta_{ij} - \frac{1}{n} \right) X_i X_j
\end{aligned}$$

Therefore, the matrix associated to this quadratic form is:

$$\begin{bmatrix} 1+a & a & a & \cdots \\ a & 1+a & a & \cdots \\ a & a & \ddots & \\ \vdots & \vdots & & \ddots \end{bmatrix}$$

where $a = -\frac{1}{n}$. It's easy to see that a matrix of this type has the following eigenvalues: $\lambda_1 = \ldots = \lambda_{n-1} = 1$ and $\lambda_n = na + 1$. Since in our case $a = -\frac{1}{n}$, we have that $\lambda_n = 0$. Therefore there exist only $n-1$ eigenvalues different from zero. Hence this quadratic form will be equivalent to another quadratic form with one less dimension (degree of freedom). In other words, we can make a transformation of coordinates such that we can reduce the number of degrees of freedom. Such a transformation is called Helmert's transformation, which consists in:

$$\begin{aligned}
u_1 &= \frac{X_1 - X_2}{\sqrt{2}} \\
u_2 &= \frac{X_1 + X_2 - 2X_3}{\sqrt{6}} \\
u_3 &= \frac{X_1 + X_2 + X_3 - 3X_4}{\sqrt{12}} \\
\cdots \\
u_{n-1} &= \frac{X_1 + X_2 + \ldots + X_{n-1} - (n-1)X_n}{\sqrt{n(n-1)}} \\
u_n &= \bar{X}_n \sqrt{n}
\end{aligned}$$

This $\{u_i\}$ have the same distribution as the $\{X_i\}$, meaning they are i.i.d. also.

We easily see that:

$$\begin{aligned}
\sum_i (X_i - \bar{X}_n)^2 &= \sum_i X_i^2 - n\bar{X}_n \\
&= \sum_i u_i^2 - n\bar{X}_n \\
&= \sum_{i=1}^{n-1} u_i^2
\end{aligned}$$

Therefore $\frac{(n-1)S_n^2}{\sigma^2}$ follows a χ^2 with $n-1$ degrees of freedom.

■

2.b. **Now we wish to find** $\text{Var}\left(S_n^2\right)$ **and** $\text{Var}\left(S_n'^2 = \frac{n-1}{n} S_n^2\right)$**.**

i. **Computation of** $\text{Var}\left(S_n^2\right)$**.**
We already showed that $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$. It's a standard result that the expected value of a χ^2_{n-1} random variable is just $n-1$ and its variance is $2(n-1)$. Therefore:

$$\text{Var}\left(\frac{n-1}{\sigma^2} S_n^2\right) = 2(n-1)$$

Using the properties of the variance we can write:

$$\frac{(n-1)^2}{(\sigma^2)^2} \text{Var}\left(S_n^2\right) = 2(n-1)$$

Which yields:

$$\text{Var}\left(S_n^2\right) = 2\frac{(\sigma^2)^2}{n-1}$$

□

ii. **Computation of** $\text{Var}\left(S_n'^2 = \frac{n-1}{n} S_n^2\right)$**.**
We just have to notice that:

$$\begin{aligned} \text{Var}\left(S_n'^2\right) &= \text{Var}\left(\frac{n-1}{n} S_n^2\right) \\ &= \frac{(n-1)^2}{n^2} \text{Var}\left(S_n^2\right) \\ &= 2\frac{n-1}{n^2}(\sigma^2)^2 \end{aligned}$$

■

3. **We wish to show that** $\bar{X}_n$ **e** S_n^2 **are independent random variables.**

We verified that $(n-1)S_n^2$ follows a χ^2 with $n-1$ degrees off freedom. $(n-1)S_n^2$ (therefore S_n^2) is independent of $\bar{X}_n$ because the distribution of the u_i is $ke^{-\frac{1}{2}\left((n-1)S_n^2 - u_n^2\right)}$, where, as we have seen, $u_n = \sqrt{n}\bar{X}_n$. Therefore $\bar{X}_n$ and S_n^2 are independent random variables.

■

4. **We wish to show that** $T_n = \frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu)$ **follows a Student's law.**

By definition, if $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2_s$, independent, then the random variable:

$$T_s = \frac{X}{\frac{Y}{s}}$$

is said to have a Student's t-distribution with r degrees of freedom.

We can show that, in fact, the pdf of T_s is, in fact, the expression that exist on tables of pdf's. We think that we are just demonstrating a definition. We will just obtain the pdf of T_s.

The joint probability of X and Y is, since they are independent:

$$f_{XY}(x,y) = C_s e^{-\frac{x^2}{2}} y^{\frac{s}{2}-1} e^{-\frac{y}{2}}, \qquad \text{for } x \in \mathbb{R}, y > 0$$

Where:

$$C_s = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{s}{2}\right) 2^{\frac{s}{2}}}$$

With the transformation $t = x\left(\frac{y}{s}\right)^{-\frac{1}{2}}$, the joint pdf of T and Y is:

$$f_{TY}(t,y) = f_{XY}\left(t\left(\frac{y}{s}\right)^{\frac{1}{2}}, y\right)\left(\frac{y}{s}\right)^{\frac{1}{2}}, \qquad \text{for } x \in \mathbb{R}, y > 0$$

since the Jacobian of the transformation is $\left(\frac{y}{s}\right)^{\frac{1}{2}}$. The marginal of T is then obtained by integrating out y, for $0 < y < +\infty$.

The calculation becomes:

$$f_T(t) = s^{-\frac{1}{2}} C_s \int_0^{+\infty} e^{-\frac{y}{2}\ 1+\frac{t^2}{s}}\, y^{\left(\frac{s+1}{2}-1\right)} \mathrm{d}y$$

After making the substitution $u = \frac{y}{2}\left(1+\frac{t^2}{s}\right)$, this reduces to:

$$\begin{aligned} f_T(t) &= s^{-\frac{1}{2}} C_s \left(\frac{2}{1+\frac{t^2}{s}}\right)^{\frac{s+1}{2}} \int_0^{+\infty} e^{-u} u^{\frac{s+1}{2}-1} \mathrm{d}u \\ &= s^{-\frac{1}{2}} C_s \left(\frac{2}{1+\frac{t^2}{s}}\right)^{\frac{s+1}{2}} \Gamma\left(\frac{s+1}{2}\right) \end{aligned}$$

Which simplifies to:

$$f_T(t) = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{s}{2}\right)} \frac{1}{(s\pi)^{\frac{1}{2}}} \frac{1}{\left(1+\frac{t^2}{s}\right)^{\frac{s+1}{2}}}, \qquad \text{for } t \in \mathbb{R}$$

□

We will now prove that:

$$T = \frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}}$$

has a Student's t-distribution with $n-1$ degrees of freedom.

Since, as we know $\frac{n-1}{\sigma 2} S_n^2 \sim \chi_{n-1}$ and $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ and these random variables are independent, we can write:

$$T = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \frac{\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{S_n^2}{n}}}$$

Which is just:

$$T = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \frac{1}{\frac{S_n^2}{\sigma^2}}$$

Is trivial the following expression:

$$T = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \frac{1}{\frac{S_n^2}{\sigma^2}\frac{n-1}{n-1}}$$

We just make the change of variables $Y = \frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$ and $V = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)$, which is, by definition, a Student's t-distribution with $n-1$ degrees of freedom.

■

Exercise 2. Fisher Information.

1. 1.a. **We wish to show that the Fisher Information of a sample of n observations is n times the information of the first observation, in the case of the empirical mean $\bar{X}_n$ of a sample of normally distributed data.**

By definition the *Fisher Information* of one parameter of a distribution is:

$$I(\theta) = -\mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right]$$

that is to say,

$$I(\theta) = \int_\Omega \left(-\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2 \mathcal{L}(\boldsymbol{x}, \theta)\,\mathrm{d}\boldsymbol{x}$$

Where $\mathcal{L}(\boldsymbol{X}, \theta) = \prod_i^n f_{Xi}(x_i|\theta)$ is called the *Likelihood Function.* As $\mathcal{L}$ is a joint function,

$$\int_\Omega \mathcal{L}\,\mathrm{d}x_1 \ldots \mathrm{d}x_n = 1$$

Differentiating both sides with respect to θ:

$$\begin{aligned}\frac{\partial}{\partial \theta}\int_\Omega \mathcal{L}\,\mathrm{d}x_1 \ldots \mathrm{d}x_n &= 0\\ \int_\Omega \frac{\partial \mathcal{L}}{\partial \theta}\mathrm{d}x_1 \ldots \mathrm{d}x_n &= \int_\Omega \frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial \theta}\mathcal{L}\,\mathrm{d}x_1 \ldots \mathrm{d}x_n = \int_\Omega \frac{\partial \ln \mathcal{L}}{\partial \theta}\mathcal{L}\mathrm{d}x_1 \ldots \mathrm{d}x_n = 0 \qquad (4)\end{aligned}$$

Differentiating again:

$$\int_\Omega \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} + \frac{\partial \ln \mathcal{L}}{\partial \theta}\frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial \theta}\right)\mathcal{L}\mathrm{d}x_1 \ldots \mathrm{d}x_n = 0$$

which is just the definition of the expectation:

$$\mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} + \left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right] = 0$$

This way we can redefine the *Fisher Information* as:

$$I(\theta) = \mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right]$$

In normally distributed data,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

the *Likelihood function* of the parameter $\theta = \mu$ of a sample of n observations is:

$$\mathcal{L}_n(\boldsymbol{X}|\mu) = \prod_i^n f_{Xi}(x_i|\mu) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Considering it´s logarithm we obtain:

$$\begin{aligned}\ln \mathcal{L}_n(\boldsymbol{X}|\mu) &= -n\ln\sigma\sqrt{2\pi} - \sum_i^n \frac{(x_i-\mu)^2}{2\sigma^2}\\ &= -n\ln\sigma\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_i^n x_i^2 - \frac{1}{2\sigma^2}n\mu^2 + \frac{1}{\sigma^2}n\bar{x}\mu\end{aligned}$$

Calculating the first two derivatives in order of μ:

$$\frac{\partial \ln \mathcal{L}_n}{\partial \mu} = -\frac{n}{\sigma^2}\mu + \frac{n\bar{x}}{\sigma^2}$$
$$\frac{\partial^2 \ln \mathcal{L}_n}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

Taking it´s expectation:

$$I(\mu) = -\mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}_n}{\partial \mu^2}\right] = \frac{n}{\sigma^2}$$

we obtain the *Fisher Information* of the sample of n observations.

Considering the first element of the sample as a unit sample, its *Likelihood function* is just its density function and so:

$$\ln \mathcal{L}_1(X_1, \mu) = -\ln \sqrt{2\pi}\sigma - \frac{(x_1 - \mu)^2}{2\sigma^2}$$

Calculating again its derivatives and its expectation we obtain:

$$\frac{\partial^2 \ln \mathcal{L}_1}{\partial \mu^2} = -\frac{1}{\sigma^2}$$
$$\mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}_1}{\partial \mu^2}\right] = -\frac{1}{\sigma^2}$$
$$I_1(\mu) = \frac{1}{\sigma^2}$$

And so we have proven that $I_n(\theta) = nI_1(\theta)$

■

1.b. **We wish to show how does the minimum variance bound (MVB) behaves with the number of observations n.**

Consider the expectation of the estimator t of a function of a distribution's parameter:

$$\mathrm{E}\left[t - \tau(\theta)\right] = b$$

Using the linearity of the expected value and it's definition:

$$\int_\Omega t\mathcal{L}\mathrm{d}x_1 \ldots \mathrm{d}x_n = b + \tau(\theta)$$

Differentiating both sides in respect to θ

$$\frac{\partial}{\partial \theta}\int_\Omega t\mathcal{L}\mathrm{d}x_1 \ldots \mathrm{d}x_n = \int_\Omega t\frac{\partial \mathcal{L}}{\partial \theta}\mathrm{d}x_1 \ldots \mathrm{d}x_n = \tau'(\theta)$$

Rearranging the integral we obtain:

$$\int_\Omega t\frac{\partial \ln(\mathcal{L})\,\mathcal{L}}{\partial \theta}\mathrm{d}x_1 \ldots \mathrm{d}x_n = \tau'(\theta)$$

Equation (4) states that:

$$\mathrm{E}\left[\frac{\partial \ln \mathcal{L}}{\partial \theta}\right] = 0$$

This means we can rewrite the equation and obtain:

$$\int_\Omega \left[t - \tau(\theta)\right]\frac{\partial \ln \mathcal{L}}{\partial \theta}\mathcal{L}\mathrm{d}x_1 \ldots \mathrm{d}x_n = \tau'(\theta)$$

Using the Cauchy-Schwarz inequality:

$$\mathrm{E}\left[X_1X_2\right]^2 \leq \mathrm{E}\left[X_1^2\right]\mathrm{E}\left[X_2^2\right]$$

Therefore:

$$\mathrm{E}\left[\tau'(\theta)\right]^2 = \mathrm{E}\left[\left[t-\tau(\theta)\right]\frac{\partial \ln \mathcal{L}}{\partial \theta}\mathcal{L}\right]^2$$

combined with the inequality stated, gives:

$$\mathrm{E}\left[\tau'(\theta)\right]^2 \leq \mathrm{E}\left[\left[t-\tau(\theta)\right]^2\right]\mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\mathcal{L}\right)^2\right]$$

Hence, using the definition of $\mathrm{Var}(t)$ and of *Fisher Information*:

$$\mathrm{Var}(t) \geq \frac{\tau'(\theta)^2}{I(\theta)}$$

This means that the variance has a lower limit – MVB.
In our case $\tau(\theta)$ is just $\theta = \mu$ and $t = \bar{X}_n$. So the MVB is:

$$\mathrm{MVB} = I(\theta)^{-1} = \frac{\sigma^2}{n}$$

So, as n increases the MVB decreases.

■

2. **We wish to study the middle point estimator for the mean of the uniform distribution.**

   i. Using the definition of the *likelihood function*:

$$\begin{aligned}\mathcal{L}_n(\boldsymbol{x},\theta) &= \begin{cases}\prod_i^n \frac{1}{\theta} & 0 \leq x_i \leq \theta \\ 0 & if not.\end{cases} \\ &= \begin{cases}\frac{1}{\theta^n} & 0 \leq x_i \leq \theta \\ 0 & if not.\end{cases}\end{aligned}$$

   In Figure (1) we plot the *likelihood function* for two values of n.

□

   ii. The log-likelihood of this function is:

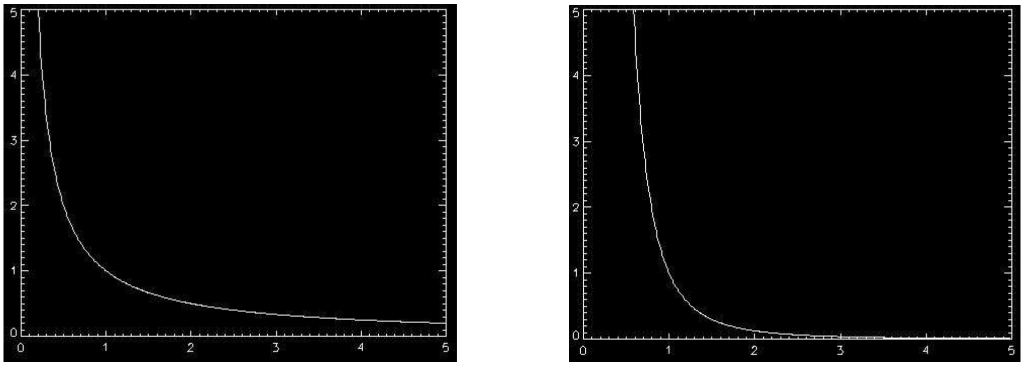
$$\ln \mathcal{L}_n = -n \ln \theta, \qquad 0 \leq x_i \leq \theta$$

   only defined for this values of x_i.
   Computing the derivative:

$$\frac{\partial \ln \mathcal{L}_n}{\partial\left(\frac{\theta}{2}\right)} = -\frac{n}{\frac{\theta}{2}}$$

   The *Fisher Information* of this estimator is:

$$\begin{aligned}I_n\left(\frac{\theta}{2}\right) &= \mathrm{E}\left[\left(\frac{\partial \ln \mathcal{L}_n}{\partial\left(\frac{\theta}{2}\right)}\right)^2\right] \\ &= \mathrm{E}\left[\left(-\frac{n}{\frac{\theta}{2}}\right)^2\right] \\ &= \left(\frac{n}{\frac{\theta}{2}}\right)^2\end{aligned}$$



(a) n=1 (b) n=3

Figura 1: Likelihood of theta

□

iii. In the case of the unit sample:

$$
\begin{aligned}
\mathcal{L}_1(\boldsymbol{x},\theta) &= \begin{cases} \frac{1}{\theta} \; 0 \leq x_i \leq \theta \\ 0 \;\; ifnot. \end{cases} \\
&= \begin{cases} \frac{1}{\theta^n} \; 0 \leq x_i \leq \theta \\ 0 \quad ifnot. \end{cases}
\end{aligned}
$$

And so:

$$\frac{\partial \ln \mathcal{L}_1}{\partial \left(\frac{\theta}{2}\right)} = -\frac{1}{\frac{\theta}{2}}$$

Which means that:

$$I_1\left(\frac{\theta}{2}\right) = \frac{1}{\left(\frac{\theta}{2}\right)^2}$$

This means that the statement:

$$I_2\left(\frac{\theta}{2}\right) = nI_1\left(\frac{\theta}{2}\right)$$

does not hold in this case.

□

iv. Using the definition of the Variance:

$$
\begin{aligned}
\mathrm{Var}\left(\frac{\theta}{2}\right) &= \mathrm{E}\left[\left(\frac{\theta}{2} - \theta\right)^2\right] \\
&= \left(\frac{\theta}{2}\right)^2
\end{aligned}
$$

The variance of this estimator is independent of the size of the sample, which means that this is not a consistent estimator, that is to say that its accuracy doesn't improve when we make more observations. We can also realize that this is not the MVB estimator of this distribution, noting that:

$$\mathrm{MVB} = I_2\left(\frac{\theta}{2}\right)^{-1} = \frac{\left(\frac{\theta}{2}\right)^2}{n}$$

■

Exercise 3. Sufficient statistics.

Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a sample of n i.i.d. drawn from a normal law $\mathcal{N}(\mu, \mu^2)$.

1. **We wish to show that the empirical mean $\bar{X}_n$ is not a sufficient statistics for μ.**

   The necessary and sufficient condition for $\boldsymbol{t}$ to be a sufficient estimator of the parameter $\boldsymbol{\theta}$:

$$\mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) = g(\boldsymbol{t}|\boldsymbol{\theta})\, h(\boldsymbol{x})$$

   We obtained in Exercise 2 the *likelihood function* of the mean of a normal distribution:

$$\begin{aligned}
\mathcal{L}_n(\boldsymbol{X}|\mu) &= \prod_i^n \frac{1}{\mu\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\mu^2}} \\
&= \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\sum_i^n \frac{(x_i-\mu)^2}{2\mu^2}} \\
&= \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\mu^2}\left(\sum_i^n x_i^2 + n\mu^2 - 2\mu\sum_i^n x_i\right)} \qquad (5) \\
&= \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\frac{n}{2}} e^{\frac{2n\bar{x}}{\mu}} e^{-\frac{\sum_i^n x_i^2}{2\mu^2}} \\
&= g(t|\theta)\, h(x, \theta)
\end{aligned}$$

   So we are not able to factorize the *likelihood function* into two factors, one depending explicitly only on the observations and another on the mean μ and the estimator we are considering for it - $\bar{X}_n$.

   ■

2. **We will find a two-dimensional sufficient statistics for μ.**

   If we also estimate the population variance μ^2 with the statistics empirical variance:

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

   Using Equation (5)

$$\begin{aligned}
\mathcal{L}\big(\boldsymbol{x}|(\mu, \mu^2)\big) &= \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\mu^2}\left(\sum_i^n x_i^2 + n\mu^2 - 2\mu\sum_i^n x_i\right)} \\
&= \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\mu^2}\left(ns^2 + n\bar{x}^2 + n\mu^2 - 2n\mu\bar{x}\right)} \\
&= \left(\frac{1}{\mu\sqrt{2\pi}}\right)^n e^{-\frac{n}{2\mu^2}\left(s^2 + (\bar{x}-\mu)^2\right)} \\
&= g\big((\mu, \mu^2)|(\bar{x}, s^2)\big)
\end{aligned}$$

   And so $\mathcal{L}\big(\boldsymbol{x}|(\mu, mu^2)\big) = g\big((\mu, \mu^2)|(\bar{x}, s^2)\big) \cdot 1$.

   This way a jointly sufficient statistics for μ is:

$$\boldsymbol{t} = (\bar{X}_n, S_n^2)$$

   ■

Exercise 4. Cosmological parameters.

There exist measures d_i of the distance d and z_i of the redshift z. Related by the expression:

$$d = \frac{c}{H_0}\left[z + \frac{1}{2}\left(1 - q_0\right)z^2 + \mathcal{O}\left(z^3\right)\right]$$

1. **We wish to derive the Fisher information matrix for the parameters (H_0, q_0) given we have performed n measurements of d and z**

   Assuming that the differences between the distances measured directly and the distances obtained with the experimental values of the red shift given the model (the errors):

   $$E_i = d_i - d(z_i)$$

   follow a $\mathcal{N}^2(0, \sigma^2)$, we have by definition of the *(*likelihood function):

   $$\mathcal{L}(\boldsymbol{e}|\boldsymbol{\theta}) = \prod_i^n f_{Ei}(e_i)$$

   where $\boldsymbol{\theta} = (H_0, q_0)$.Taking the logarithm, we get:

   $$\begin{aligned}\ln \mathcal{L}(\boldsymbol{e}|\boldsymbol{\theta}) &= \sum_i^n \ln f_{Ei}(e_i) \\ &= -n \ln \sigma\sqrt{2\pi} - \sum_i^n \frac{(e_i)^2}{2\sigma^2}\end{aligned}$$

   Using the fact that:

   $$E_i = d_i - \frac{c}{H_0}\left(z_i - \frac{1}{2}(1 - q_0)z_i^2\right)$$

   We obtain the following expression:

   $$\ln \mathcal{L}(\boldsymbol{e}|\boldsymbol{\theta}) = -n \ln \sigma\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_i^n \left(d_i - \frac{c}{H_0}[z_i + \frac{z_i^2}{2}(1 - q_0)]\right)^2$$

   The *Fisher information* of a multidimensional parameter $\boldsymbol{\theta}$ is a matrix with the following elements:

   $$I_{ij} = \mathrm{E}\left[\frac{\partial \ln \mathcal{L}}{\partial \theta_i} \cdot \frac{\partial \ln \mathcal{L}}{\partial \theta_j}\right]$$

   In the present case it becomes:

   $$I = \begin{bmatrix} \mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial H_0}\right)^2\right] & \mathrm{E}\left[\frac{\partial \mathcal{L}}{\partial H_0} \cdot \frac{\partial \mathcal{L}}{\partial q_0}\right] \\ \mathrm{E}\left[\frac{\partial \mathcal{L}}{\partial q_0} \cdot \frac{\partial \mathcal{L}}{\partial H_0}\right] & \mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial q_0}\right)^2\right] \end{bmatrix}$$

   Computing the two derivatives:

   $$\begin{aligned}\frac{\partial \mathcal{L}}{\partial H_0} &= \frac{1}{H_0\sigma^2}\sum_i^n \left(\frac{c}{H_0}[z_i - \frac{z_i^2}{2}(1 - q_0)] - d_i\right) \cdot \left(z_i - \frac{z_i^2}{2}(1 - q_0)\right) \\ \frac{\partial \mathcal{L}}{\partial q_0} &= \frac{c}{H_0\sigma^2}\sum_i^n \left(\frac{c}{H_0}[z_i - \frac{z_i^2}{2}(1 - q_0)] - d_i\right) \cdot \frac{z_i^2}{2}\end{aligned}$$

Calculating the elements of the *Fisher Information Matrix*:

$$
\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial H_0}\right)^2\right] &= \mathrm{E}\left[\frac{1}{\sigma^4 H_0^2}\left(\sum_{i,j}^{n}\left(\frac{c}{H_0}[z_i+\frac{z_i^2}{2}(1-q_0)]-d_i\right)\cdot\left(\frac{c}{H_0}[z_i+\frac{z_i^2}{2}(1-q_0)]\right)\times\right.\right.\\
&\left.\left.\times\left(\frac{c}{H_0}[z_j+\frac{z_j^2}{2}(1-q_0)]-d_j\right)\cdot\left(\frac{c}{H_0}[z_j+\frac{z_j^2}{2}(1-q_0)]\right)\right)\right]\\
\mathrm{E}\left[\frac{\partial \mathcal{L}}{\partial q_0}\cdot\frac{\partial \mathcal{L}}{\partial H_0}\right] &= \mathrm{E}\left[\frac{c}{2\sigma^4 H_0^2}\left(\sum_{i,j}^{n}\left(\frac{c}{H_0}[z_i+\frac{z_i^2}{2}(1-q_0)]-d_i\right)\cdot\frac{c}{H_0}[z_i+\frac{z_i^2}{2}(1-q_0)]\times\right.\right.\\
&\left.\left.\times\left(\frac{c}{H_0}[z_j+\frac{z_j^2}{2}(1-q_0)]-d_j\right)\cdot z_j^2\right)\right]\\
\mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial q_0}\right)^2\right] &= \mathrm{E}\left[\frac{c^2}{4\sigma^4 H_0^2}\left(\sum_{i,j}^{n}\left(\frac{c}{H_0}[z_i+\frac{z_i^2}{2}(1-q_0)]-d_i\right)\cdot z_i^2\cdot\left(\frac{c}{H_0}[z_j+\frac{z_j^2}{2}(1-q_0)]-d_j\right)\cdot z_j^2\right)\right]
\end{aligned}
$$

And identifying the various terms:

$$
\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial H_0}\right)^2\right] &= \frac{1}{\sigma^4 H_0^2}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot(e_i+d_i)\cdot e_j\cdot(e_j+d_j)\right]\\
\mathrm{E}\left[\frac{\partial \mathcal{L}}{\partial q_0}\cdot\frac{\partial \mathcal{L}}{\partial H_0}\right] &= \frac{c}{2\sigma^4 H_0^2}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot(e_i+d_i)\cdot e_j\cdot z_j^2\right]\\
\mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial q_0}\right)^2\right] &= \frac{c^2}{4\sigma^4 H_0^2}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot z_i^2\cdot e_j\cdot z_j^2\right]
\end{aligned}
$$

In the matrix form it becomes:

$$
I(H_0,q_0)=\frac{1}{\sigma^4 H_0^2}\begin{bmatrix}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot(e_i+d_i)\cdot e_j\cdot(e_j+d_j)\right] & \frac{c}{2}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot(e_i+d_i)\cdot e_j\cdot z_j^2\right]\\ \frac{c}{2}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot(e_i+d_i)\cdot e_j\cdot z_j^2\right] & \frac{c^2}{4}\sum_{i,j}^{n}\mathrm{E}\left[e_i\cdot z_i^2\cdot e_j\cdot z_j^2\right]\end{bmatrix}
$$

■