# Chapter 4: Sample Theory and Sample Distributions

*By* Joan Llull*

Probability and Statistics.
QEM Erasmus Mundus Master. Fall 2016

**Main references:**

— Mood: IV:1, IV:2.1, IV:2.2-2, IV:3.1, IV:4

— Lindgren: 7.1, 7.3, 7.4, 7.5

## I. Random samples

The objective of this chapter is to make inference about some characteristics of a population from a set of observations in the data. The population is described by a probabilistic model like those seen in previous chapters. The observations in the data are considered as realizations from the probabilistic model. Recall that one of the main features of a random experiment is that it can be replicated under the same conditions.

The process through which we obtain our data is called ***sampling***. There are several ways of sampling. In Chapter 2 we introduced sampling in finite sets as an example to illustrate the use of combinatorial analysis to compute probabilities.

Simple random sampling is the easiest way of selecting a sample. It is not always the best way we can do it in Economics, but its simplicity puts it as the starting point for all others. A collection of random variables (or random vectors) $(X_1, ..., X_N)$ is a (simple) ***random sample*** from $F_X$ if $(X_1, ..., X_N)$ are independent and identically distributed (i.i.d) with cdf $F_X$. We can use the word sample to refer both to this random vector $(X_1, ..., X_N)$, and to the realization of it $(x_1, ..., x_N)$. Each of the elements of this vector is known as an ***observation***.

Given that the observations are i.i.d., the cdf of the sample is:

$$F_{X_1 \dots X_N}(x_1, ..., x_N) = \prod_{i=1}^{N} F_X(x_i), \tag{1}$$

and, thus:

$$f_{X_1 \dots X_N}(x_1, ..., x_N) = \prod_{i=1}^{N} f_X(x_i), \tag{2}$$

* Departament d'Economia i Història Econòmica. Universitat Autònoma de Barcelona. Facultat d'Economia, Edifici B, Campus de Bellaterra, 08193, Cerdanyola del Vallès, Barcelona (Spain). E-mail: joan.llull[at]movebarcelona[dot]eu. URL: http://pareto.uab.cat/jllull.

where $f_X$ is the pmf of the sample if $X$ is discrete, and the corresponding pdf if $X$ is continuous.

For example, consider a Bernoulli random variable with pmf equal to:

$$f_X(x) = \begin{cases} \frac{2}{3} & \text{if } x = 0 \\ \frac{1}{3} & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Now consider a random sample of three observations obtained from this population. As we discussed in Chapter 2, there are $2^3 = 8$ possible permutations, and the pmf is given by:

| Sample: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $x_1$: | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| $x_2$: | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| $x_3$: | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| $f_{X1 \ldots X_3}(x_1, x_2, x_3)$: | $\frac{8}{27}$ | $\frac{4}{27}$ | $\frac{4}{27}$ | $\frac{2}{27}$ | $\frac{4}{27}$ | $\frac{2}{27}$ | $\frac{2}{27}$ | $\frac{1}{27}$ |

## II. Sample mean and variance

A ***statistic*** is a single measure of some attribute of a sample. It is calculated by applying a function to the values of the items of the sample. Any of the synthetic measures that we computed in Chapter 1 were statistics. In that chapter we were using them to summarize the data. Now, we are going to use them to infer some properties of the probability model that generated the data.

As a transformation of random variables, a statistic is a random variable. As such, it has a probability distribution. This probability distribution is called ***sample distribution***.

The first of the statistics that we introduced in Chapter 1 is the ***sample mean***. In a simple random sample, the weights used to compute the sample mean are all equal, and thus equal to $\frac{1}{N}$. Therefore, here we define the sample mean as:

$$\bar{X}_N \equiv \frac{1}{N} \sum_{\text{i=1}}^{\text{N}} X_i. \tag{4}$$

In the example before, for each of the possible samples we would obtain a different sample mean:

| Sample: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\bar{x}$: | 0 | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ | $\frac{2}{3}$ | 1 |

Note that all combinations of the same inputs give the same sample mean, so we could alternatively count the number of combinations instead of permutations.

Importantly, note that our statistic (the sample mean) has a sample distribution:

$$f_{\bar{X}_N}(\bar{x}) = \begin{cases} \frac{8}{27} & \text{if } \bar{x} = 0 \\ \frac{12}{27} & \text{if } \bar{x} = \frac{1}{3} \\ \frac{6}{27} & \text{if } \bar{x} = \frac{2}{3} \\ \frac{1}{27} & \text{if } \bar{x} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Given Equation (5), we could compute $\mathbb{E}[\bar{X}_N]$ and $\text{Var}(\bar{X}_N)$:

$$\mathbb{E}[\bar{X}_N] = \frac{8}{27} \cdot 0 + \frac{12}{27} \cdot \frac{1}{3} + \frac{6}{27} \cdot \frac{2}{3} + \frac{1}{27} \cdot 1 = \frac{1}{3}, \tag{6}$$

and:

$$\begin{aligned} \text{Var}(\bar{X}_N) &= \mathbb{E}[\bar{X}_N^2] - \mathbb{E}[\bar{X}_N]^2 \\ &= \left[\frac{8}{27} \cdot 0^2 + \frac{12}{27} \cdot \left(\frac{1}{3}\right)^2 + \frac{6}{27} \cdot \left(\frac{2}{3}\right)^2 + \frac{1}{27} \cdot 1^2\right] - \left(\frac{1}{3}\right)^2 \\ &= \frac{2}{27} = \frac{2/9}{3}. \end{aligned} \tag{7}$$

Note that, for this variable, $\mathbb{E}[X] = p = 1/3$, and $\text{Var}(X) = p(1-p) = 2/9$. Therefore, at least in this example, $\mathbb{E}[\bar{X}_N] = \mathbb{E}[X]$ and $\text{Var}(\bar{X}_N) = \text{Var}(X)/N$. This result is general, as discussed in the following paragraph.

Let $(X_1, ..., X_N)$ be a random sample from a population described by the cdf $F_X$ which has mean $\mathbb{E}[X] = \mu$ and variance $\text{Var}(X) = \sigma^2$. Let $\bar{X}_N$ denote the sample mean of this sample. Then, $\mathbb{E}[\bar{X}_N] = \mu$, and $\text{Var}(\bar{X}_N) = \sigma^2/N$. Let us check that:

$$\mathbb{E}[\bar{X}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{\text{i=1}}^{\text{N}} X_i\right] = \frac{1}{N}\sum_{\text{i=1}}^{\text{N}} \mathbb{E}[X_i] = \frac{1}{N}\sum_{\text{i=1}}^{\text{N}} \mu = \frac{1}{N} N\mu = \mu. \tag{8}$$

And for the variance:

$$\begin{aligned} \text{Var}(\bar{X}_N) &= \mathbb{E}[(\bar{X}_N - \mu)^2] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{\text{i=1}}^{\text{N}}(X_i - \mu)\right)^2\right] \\ &= \mathbb{E}\left[\frac{1}{N^2}\left(\sum_{\text{i=1}}^{\text{N}}(X_i - \mu)^2 + \sum_{\text{i=1}}^{\text{N}}\sum_{\text{j=1}}^{\text{N}}\underset{j\neq i}{}(X_i - \mu)(X_j - \mu)\right)\right] \\ &= \frac{1}{N^2}\sum_{\text{i=1}}^{\text{N}}\mathbb{E}\left[(X_i - \mu)^2\right] + \frac{1}{N^2}\sum_{\text{i=1}}^{\text{N}}\sum_{\text{j=1}}^{\text{N}}\underset{j\neq i}{}\mathbb{E}\left[(X_i - \mu)(X_j - \mu)\right] \\ &= \left[\frac{1}{N^2}\sum_{\text{i=1}}^{\text{N}}\mathbb{E}\left((X_i - \mu)^2\right)\right] = \frac{1}{N}N\sigma^2 = \frac{\sigma^2}{N}, \end{aligned} \tag{9}$$

where, from the third to the fourth line, we used the fact that, given that the observations are i.i.d., the covariance between $X_i$ and $X_j$ is equal to zero.

There are three main conclusions to extract from this general result. The first one is that $\mathbb{E}[\bar{X}_N]$, and $\mathrm{Var}(\bar{X}_N)$ do not depend of the form of $F_X$, they only depend on its first two moments. The second one is that $\bar{X}_N$ is "centered" around the population mean $\mu$. And the third one is that the dispersion of $\bar{X}_N$ is reduced when we increase $N$, tending to zero when $N \to \infty$. We care about the variance of $\bar{X}_N$ as an indicator of the (inverse of the) ***precision*** of $\bar{X}_N$ as a proxy for $\mu$: the smaller $\mathrm{Var}(\bar{X}_N)$, the more likely is that $\bar{X}_N$ is "close" to $\mu$. Thus, the larger the sample, the more "accurate" is $\bar{X}_N$ as an approximation to $\mu$. We will discuss extensively all this in the following chapters.

A similar analysis can be performed with respect to another of the statistics that we introduced in Chapter 1: the ***sample variance***. Again, given the observations are obtained from a random sample, the weight we give to each observation is equal for all of them, and equal to $\frac{1}{N}$. Thus, the sample variance, which we denote as $\hat{\sigma}_N^2$, is defined as:

$$\hat{\sigma}_N^2 \equiv \frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}(X_i - \bar{X}_N)^2. \tag{10}$$

Let us first compute the expectation:

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}_N^2] &= \mathbb{E}\left[\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}(X_i - \bar{X}_N)^2\right] \\
&= \mathbb{E}\left[\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}(X_i - \mu - (\bar{X}_N - \mu))^2\right] \\
&= \frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}\mathbb{E}\left[(X_i - \mu)^2\right] - \mathbb{E}\left[(\bar{X}_N - \mu))^2\right] \\
&= \mathrm{Var}(X) - \mathrm{Var}(\bar{X}_N) = \sigma^2 - \frac{\sigma^2}{N} = \frac{(N-1)}{N}\,\sigma^2.
\end{aligned} \tag{11}$$

Thus, with the sample variance we expect to obtain less dispersion that the dispersion in the population, except when $N \to \infty$.

We often propose an alternative statistic to measure dispersion in the sample, the ***corrected sample variance***, which is defined as:

$$s_N^2 \equiv \frac{N}{N-1}\,\hat{\sigma}_N^2 = \frac{1}{N-1}\sum_{\mathrm{i=1}}^{\mathrm{N}}(X_i - \bar{X}_N)^2. \tag{12}$$

Easily we can check that $\mathbb{E}[s_N^2] = \sigma^2$. Therefore, unlike the sample variance, the corrected sample variance is centered around the population dispersion of the data. This is a desirable property when we want to make inference about the population, and, thus, $s_N^2$ is commonly used instead of $\hat{\sigma}_N^2$.

Even though we are not going to prove it (it is recommended as an exercise), it is messy but easy to show that:

$$\mathrm{Var}(s_N^2) = \frac{2\sigma^4}{N-1} + \frac{\mu_4 - 3\sigma^4}{N}. \tag{13}$$

There exists an alternative measure of dispersion that has lower variance (i.e. that is more precise) than $s_N^2$:

$$\tilde{\sigma}_N^2 \equiv \frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}(X_i - \mu)^2. \tag{14}$$

Trivially, we can check that $\mathbb{E}[\tilde{\sigma}_N^2] = \sigma^2$. To compute the variance of $\tilde{\sigma}_N^2$, we only need to compute $\mathbb{E}[(\tilde{\sigma}_N^2)^2]$. To do so, define $Z_i \equiv X_i - mu$, so that notation is less messy:

$$\begin{aligned}\mathbb{E}[(\tilde{\sigma}_N^2)^2] &= \mathbb{E}\left[\left(\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}} Z_i^2\right)^2\right] = \frac{1}{N^2}\mathbb{E}\left[\sum_{\mathrm{i=1}}^{\mathrm{N}} Z_i^4 + \sum_{\mathrm{i=1}}^{\mathrm{N}}\sum_{\mathrm{j=1}\,i\neq j}^{\mathrm{N}} Z_i^2 Z_j^2\right] \\ &= \frac{1}{N^2}[N\mu_4 + (N^2-N)\sigma^4] = \frac{1}{N}[\mu_4 - (N-1)\sigma^4]. \end{aligned} \tag{15}$$

And, hence:

$$\mathrm{Var}(\tilde{\sigma}_N^2) = \frac{1}{N}[\mu_4 - (N-1)\sigma^4] - \sigma^4 = \frac{1}{N}[\mu_4 - \sigma^4] < \mathrm{Var}(s_N^2). \tag{16}$$

Therefore, this statistic would be preferred to the previous two to make inference about the variance of the distribution of $X$ because it is centered at $\sigma^2$, like $s_N^2$, but it is more precise. However, this is an ***unfeasible estimator***, which means that, in general, we cannot compute it, because we do not know $\mu$.

## III. Sampling form a normal population: $\chi^2$, $t$, and $F$ distributions

Let $(X_1, ..., X_N)$ be a random sample from the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. From previous chapters we know that, as a linear combination of normal random variables, the sample mean is also normally distributed. And, from previous section, we know the parameters of this normal distribution:

$$\bar{X}_N \sim \mathcal{N}(\mu, \sigma^2/N). \tag{17}$$

Also using the materials from previous chapters, we also know the distribution of the following transformation:

$$Z \equiv \frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1). \tag{18}$$

This result will allow us to make inference about $\mu$ based on $\bar{X}_N$ in future chapters, provided that $\sigma^2$ is known, because, given $\sigma$, the distribution of this statistic is known (standard normal).

One alternative is to replace $\sigma^2$ by $s_N^2$, but $s_N^2$ is itself a random variable, and, hence, the distribution above is altered. To see how, we first need to derive the distribution of $s_N^2$, and, to do that, we have to introduce some intermediate results:

1) Let $\tilde{Z} \equiv (\tilde{Z}_1, ..., \tilde{Z}_K)'$ be a vector of $K$ i.i.d. random variables, with $\tilde{Z}_i \sim \mathcal{N}(0,1)$. Then, we say that $\tilde{W} = \tilde{Z}_1^2 + ... + \tilde{Z}_K^2 = \tilde{Z}'\tilde{Z}$ is distributed as a ***chi-squared*** with $K$ ***degrees of freedom***: $\tilde{W} \sim \chi_K^2$. The degrees of freedom are the number of independent squared standard normal distributions that are adding. The support of this distribution is $\mathbb{R}^+$. Interesting results for this distribution are that $\mathbb{E}[\tilde{W}] = K$ and $\text{Var}(\tilde{W}) = 2K$ (you are strongly encouraged to prove them).

2) Let $\tilde{X} \sim \mathcal{N}_N(0, \Sigma)$. Then, $\tilde{X}'\Sigma^{-1}\tilde{X} \sim \chi_N^2$. To see it, decompose $\Sigma = \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}$ as in previous chapter. Thus, $\tilde{X}'\Sigma^{-1}\tilde{X} = (\tilde{X}'\Sigma^{-\frac{1}{2}})(\Sigma^{-\frac{1}{2}}\tilde{X}) = \tilde{Z}'\tilde{Z}$, where $\tilde{Z} \sim \mathcal{N}_N(0, I)$, which is equivalent to say that all its elements are independently distributed as a standard normal. Given the definition of the chi-squared distribution in the previous bullet, we therefore know that $\tilde{Z}'\tilde{Z} \sim \chi_N^2$, completing the proof.

3) Let $M$ be a size $K \times K$ idempotent (satisfies $MM = M$) and symmetric (satisfies $M' = M$) matrix, with $\text{rank}(M) = R \leq K$. Because it is idempotent, $M$ is singular (with the only exception of $M = I$), it is also diagonalizable, and its eigenvalues are either 0 or 1. In particular, it can always be diagonalized as $M = C'\Lambda C$ such that $C'C = I$, and $\Lambda$ is a matrix that include ones in the first $R$ elements of the diagonal and zeros elsewhere. As a result, the trace of $M$ (the sum of its diagonal elements) is equal to its rank (and thus always a natural number).

4) Let $\tilde{Z} \sim \mathcal{N}_K(0, I)$, and $M$ be a size $K \times K$ idempotent and symmetric matrix with $\text{rank}(M) = R \leq K$. Then $\tilde{Z}'M\tilde{Z} \sim \chi_R^2$. To prove it, consider the diagonalization above: $\tilde{Z}'C'\Lambda C\tilde{Z}$. If we let $C$ be the equivalent to $\Sigma^{\frac{1}{2}}$ above, $\tilde{Z}C \sim \mathcal{N}_K(0, C'C) = \mathcal{N}_K(0, I)$. Therefore, $\tilde{Z}'M\tilde{Z}$ is a sum of

$R$ independent squared standard normals(given that $\Lambda$ has $R$ elements in the diagonal that are equal to one, and the rest are equal to zero), and thus $\tilde{Z}'M\tilde{Z} \sim \chi^2_R$.

5) Let $\tilde{Z} \sim \mathcal{N}_K(0, I)$, and $M$ be a size $K \times K$ idempotent and symmetric matrix with $\text{rank}(M) = R \leq K$. Also let $P$ be a $Q \times N$ matrix such that $PM = 0$. Then $\tilde{Z}'M\tilde{Z}$ and $P\tilde{Z}$ are independent. To prove it, note that, as linear combinations of a standard normal vector, both $M\tilde{Z}$ and $P\tilde{Z}$ are normal (thus, independence and absence of correlation are equivalent, as we saw in Chapter 3). Additionally:

$$\text{Cov}(P\tilde{Z}, M\tilde{Z}) = P\,\text{Cov}(\tilde{Z}, \tilde{Z})M = P\,\text{Var}(\tilde{Z})M = PIM = PM = 0. \quad (19)$$

(last step by assumption). Because $M$ is idempotent and symmetric, $\tilde{Z}'M\tilde{Z} = \tilde{Z}'M'M\tilde{Z} = (M\tilde{Z})'M\tilde{Z}$. Thus, $\tilde{Z}'M\tilde{Z}$ is a function of $M\tilde{Z}$ so, since $M\tilde{Z}$ and $P\tilde{Z}$ are independent, $\tilde{Z}'M\tilde{Z}$ and $P\tilde{Z}$ are independent.

We now can use these intermediate results to derive the distribution of $s^2_N$. Define $\iota \equiv (1, ..., 1)'$ a size $N$ vector of ones. Clearly, $\iota'\iota = N$, and, thus:

$$\bar{X} = \frac{1}{N}\sum_{\text{i=1}}^{\text{N}} X_i = (\iota'\iota)^{-1}\iota'X \equiv PX. \quad (20)$$

Similarly:

$$X - \iota\bar{X} = \begin{pmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_N - \bar{X} \end{pmatrix} = (I - \iota(\iota'\iota)^{-1}\iota')X \equiv MX. \quad (21)$$

Trivially, the matrix $M$ is symmetric and idempotent. Thus, we can write:

$$\sum_{\text{i=1}}^{\text{N}}(X_i - \bar{X})^2 = X'M'MX = X'MX. \quad (22)$$

This result implies that:

$$\frac{(N-1)s^2_N}{\sigma^2} = \frac{\sum_{\text{i=1}}^{\text{N}}(X_i - \bar{X})^2}{\sigma^2} = \frac{1}{\sigma^2}X'MX = \tilde{Z}'M\tilde{Z}. \quad (23)$$

where $\tilde{Z}_i \equiv \frac{X_i - \mu}{\sigma}$. The last equality is obtained by noting that $\bar{\tilde{Z}} \equiv N^{-1}\sum_{\text{i=1}}^{\text{N}} \tilde{Z}_i = \frac{1}{\sigma}\bar{X} - \frac{\mu}{\sigma}$, and thus:

$$M\tilde{Z} = \tilde{Z}_i - \bar{\tilde{Z}} = \tilde{Z}_i + \frac{\mu}{\sigma} - \left(\bar{\tilde{Z}} + \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}MX. \quad (24)$$

Therefore, since $\tilde{Z} \sim \mathcal{N}_N(0, I)$ and $\text{rank}(M) = \text{tr}(M) = N-1$, we conclude that:[1]

$$W \equiv \frac{(N-1)s_N^2}{\sigma^2} \sim \chi^2_{N-1}. \tag{25}$$

Finally, we introduce a new distribution: the ***Student-t***. Let $Z \sim \mathcal{N}(0,1)$ and $W \sim \chi^2_K$, with $Z$ and $W$ being independent. Then:

$$t \equiv \frac{Z}{\sqrt{\frac{W}{K}}} \sim t_K, \tag{26}$$

which we read $t$ follows a Student-$t$ distribution with $K$ degrees of freedom. The pdf of this distribution is symmetric with respect to zero, and its support is the real line. Also, $\mathbb{E}[t] = 0$ and $\text{Var}(t) = \frac{K}{K-2}$ for $K > 2$ (with $K \leq 2$ then the variance does not converge). When $K \to \infty$, the distribution is very similar to a normal distribution.

The choice of $Z$ and $W$ as a notation for this definition are not coincidental. The $Z$ and $W$ respectively defined in Equations (18) and (25) satisfy $Z \sim \mathcal{N}(0,1)$ and $W \sim \chi^2_K$, as we proved above. Thus, we only need to prove that they are independent to be able to use the ***t-statistic*** from Equation (26) for our $Z$ and $W$. To do so, we start by checking that $PM = 0$:

$$PM = \iota(\iota'\iota)^{-1}\iota'(I - \iota(\iota'\iota)^{-1}\iota') = \iota(\iota'\iota)^{-1}\iota' - \iota(\iota'\iota)^{-1}\iota' = 0. \tag{27}$$

Also, we note that $Z = \sqrt{N}P\tilde{Z}$, with $\tilde{Z} \sim \mathcal{N}_N(0, I)$. Thus, given that $W = \tilde{Z}'M\tilde{Z}$, and using the intermediate result number 5 above, we conclude that $P\tilde{Z}$ and $W$ are independent, as so are $Z$ and $W$. Therefore:

$$\frac{Z}{\sqrt{\frac{W}{N-1}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{N}}}{\sqrt{\frac{(N-1)s_N^2/\sigma^2}{N-1}}} = \frac{(\bar{X}-\mu)}{s/\sqrt{N}} \sim t_{N-1}. \tag{28}$$

Hence, with this statistic, we can make inference about $\mu$ without knowing $\sigma^2$.

There is another distribution that is useful to make inference about the variance. Even though we will not enter into the details of it, let us define it. Let $W_1$ and $W_2$ be two independent random variables such that $W_1 \sim \chi^2_K$ and $W_2 \sim \chi^2_Q$. Then:

$$F \equiv \frac{W_1/K}{W_2/Q} \sim F_{K,Q}, \tag{29}$$

or, in words, the statistic $F$ follows a ***F-distribution*** with $K$ and $Q$ degrees of freedom. This distribution satisfies that $\mathbb{E}[F] = \frac{Q}{Q-2}$ (for $Q > 2$, otherwise the

[1] To prove that $\text{tr}(M) = N-1$, note that $\text{tr}(I_N) = N$, and $\text{tr}(P) = \text{tr}(\iota(\iota'\iota)^{-1}\iota') = \text{tr}(\iota'\iota(\iota'\iota)^{-1}) = \text{tr}(1) = 1$ (since $\text{tr}(AB) = \text{tr}(BA)$), and, thus, $\text{tr}(M) = \text{tr}(I_N) - \text{tr}(P) = N-1$.

integral does not converge). Also, $(t_K)^2 \sim F_{1,K}$, since the numerator is one squared normal (i.e. a chi-squared with one degree of freedom), and the denominator is a chi-squared with $K$ degrees of freedom divided by $K$.

## IV. Bivariate and Multivariate Sampling

So far we have analyzed the case in which we sample from a univariate distribution. However, we can also sample from a multivariate distribution. Let $X$ be a size $K$ random variable with joint pdf equal to $f_X(x)$. Now, we extract a random sample $(X_1, ..., X_N)$ where $X_i$ for $i = 1..., N$ are random vectors. Given the random sampling, the joint pdf is given by:

$$f_{X_1 \ldots X_N}(x_1, ..., x_k) = \prod_{i=1}^{N} f_{X_i}(x_i). \tag{30}$$

Thus, we can define the following "joint" statistics:

- Sample mean: $\mathbb{E}[X]$.
- Sample variance-covariance matrix: $\frac{1}{N} \sum_{\mathrm{i=1}}^{\mathrm{N}} (X_i - \bar{X})(X_i - \bar{X})'$.

The sample variance-covariance matrix includes variances and covariances. We showed above that the expectation of the sample variance was not equal to the population variance, and thus we created a corrected variance. Should we do the same thing with the covariance? The answer is yes. The proof is analogous to the univariate case discussed above (one only needs to know that the expectation of the matrix is the matrix of the expectations, as we also discussed, and then operate individually).