Advanced Statistics

6. Sample Moments and their Distributions

Christian Aßmann

Chair of Survey Statistics and Data Analysis – Otto-Friedrich-Universität Bamberg

In this chapter we begin to study problems and methods related to **statistical inference**.

In the preceding chapters we discussed fundamental ideas of probability theory and the theory of distributions. There a typical question was:

> *Given the probability space, what can we say about the characteristics and properties of outcomes of an experiment?*

Statistical inference turns this question around:

> *Given the observed characteristics and properties of outcomes of an experiment, what can we say (infer) about the probability space?*

A typical problem of statistical inference is as follows:

- Suppose we seek information about some characteristics of a collection of elements, called **population**.
- For reasons of time or cost we may not wish to study each element of the population. Our object is rather to draw conclusions about the unknown population characteristics on the basis of information on some characteristics of a suitably selected **sample**.
- Formally, let X be a random variable that represents the population under investigation, and let $f(x, \theta)$ denote the parametric family of pdfs of X. The set of possible parameter values is denoted by Ω.
- Then the job of the statistician is to decide on the basis of a sample randomly drawn from the population which member of the family of pdfs $\{f(x, \theta), \theta \in \Omega\}$ can represent the pdf of X.

In the first section we introduce the notions of **random samples** and **sample statistics**.

Often, the collected data of an experiment consist of several observed values of a variable of interest. If the process of data collection is random, it is referred to as random sampling.

Here, we will consider two general random sampling methods of data collection:

1. random sampling from a population distribution (which includes random sampling with replacement);
2. random sampling without replacement.

DEFINITION (RANDOM SAMPLING FROM A POPULATION DISTRIBUTION): Let X be a random variable with pdf $f(x)$. The set of random variables $X_1, ..., X_n$ is called a *random sample of size n from the population distribution with pdf* $f(x)$, if

$$X_1, ..., X_n \text{ are iid random variables with pdf } f(x).$$

The set of observed values $x_1, ..., x_n$ is called *realization of the sample*.

REMARK : According to the definition, we consider a situation where the variable of interest X has a pdf $f(x)$ and where we have repeated observations on this variable.

The first observation x_1 is a realized value of X_1, the second x_2 a realized value of X_2, and so on.

Each x_i represents an observation of the same variable, and each X_i has the same marginal pdf given by $f(x)$. Furthermore, the observations are taken in such a way that the X_is are independent.

(CONTINUES)

REMARK (CONTINUED): Hence, the joint pdf of a random sample from a population distribution $X_1, ..., X_n$ is given by

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i). \quad \diamondsuit$$

REMARK : When a **population with a finite number of elements** is sampled, random sampling from a population distribution is alternatively referred to as **random sampling with replacement**. $\diamondsuit$

EXAMPLE: Consider an urn containing N balls, J red and $N - J$ black balls.

- Let the random variable X represent the color of a ball, with $x = 1$ for a red ball and $x = 0$ for a black ball.

- If we sample n balls with replacement, the **population distribution** is a Bernoulli distribution,

$$f(x; p) = p^x (1 - p)^{1-x}, \qquad \text{with} \quad p = J/N.$$

- The joint pdf of a random sample $X_1, \ldots, X_n$ obtained by drawing n times from the urn with replacement is

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) = p^{\sum_{i=1}^{n} x_i} (1 - p)^{n - \sum_{i=1}^{n} x_i}.$$

Note that if we assume that the population ratio $p = J/N$ is unknown, then estimation of p would be an object of statistical inference.

Note also that the value of the population ratio p has a direct influence on the probability of drawing a red (black) ball. Thus, we have a probabilistic link between the population and the sample characteristics. ||

Random Sampling Without Replacement

- Random sampling without replacement is relevant for populations with finite number of elements.
- It means that once the characteristic of an element has been observed, the element is removed from the population before another element is drawn.
- Removing an element from the population changes the composition of the population and hence the population distribution f for the next draw.

  This implies that the corresponding sample variables $X_1, ..., X_n$ are neither identically distributed nor mutually independent.

  Hence, the joint pdf of the sample variables **can not** be written as $f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$

EXAMPLE: Consider an urn containing N balls, J red and $N - J$ black balls. Let the random variable X represent the color of a ball, with $x = 1$ for a red ball and $x = 0$ for black ball.

- Suppose a random sample $X_1, \ldots, X_n$ without replacement of size $n \leq N$ is drawn from the urn.
- Then, the pdf for the first sample variable X_1 is

$$f(x_1) = \left(\frac{J}{N}\right)^{x_1} \left(\frac{N-J}{N}\right)^{1-x_1}.$$

The pdf of the X_2 given the realization of the first variable x_1

$$f(x_2|x_1) = \underbrace{\left(\frac{J-x_1}{N-1}\right)^{x_2} \left(\frac{N-J-(1-x_1)}{N-1}\right)^{1-x_2}}_{\substack{\text{The 1st draw reduces popalution size by one,} \\ \text{and for } x_1 = 1 \text{ it reduces the number of red balls}}} .$$

Hence, the second sample variable X_2 has a pdf which is different from that of the first one. Furthermore the distribution of X_2 depends on the realization of the first variable.

(CONTINUES)

Example (Continued):

- In general, the pdf of the ℓth variable X_ℓ given the realization of the first $\ell - 1$ variables is

$$f(x_\ell | x_{\ell-1}, \ldots, x_1) = \left(\frac{J - \sum_{i=1}^{\ell-1} x_i}{N - (\ell - 1)} \right)^{x_\ell} \left(\frac{N - J - (\ell - 1 - \sum_{i=1}^{\ell-1} x_i)}{N - (\ell - 1)} \right)^{1 - x_\ell}.$$

Hence, the joint pdf of the random sample obtains as

$$f(x_1, \ldots, x_n) = f(x_1) \cdot f(x_2|x_1) \cdot \ldots \cdot f(x_n|x_{n-1}, \ldots, x_1) \neq \prod_{i=1}^{n} f(x_i). \qquad ||$$

Remark: In the remainder of our course, we will consider primarily random sampling from a population distribution. $\diamondsuit$

Statistics

In statistical inference, we use **functions of the random sample** $X_1, \ldots, X_n$ to map/transform sample information into inferences regarding the population characteristics of interest. The functions used for this mapping are called **statistics**, defined as follows.

DEFINITION (STATISTIC): Let $X_1, \ldots, X_n$ be a random sample from a population and let $T(x_1, \ldots, x_n)$ be a real-valued function which does not depend on unobservable quantities. Then the random variable

$$Y = T(X_1, \ldots, X_n) \quad \text{is called a } \textit{(sample) statistic}.$$

REMARK: The definition requires that the function T does not depend on unobservable quantities, like unknown population parameters. This implies that a statistic is a random variable whose outcomes can be observed.

(CONTINUES)

REMARK (CONTINUED): Two of the most commonly used statistics are the **sample mean** and the **sample variance**, given by

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \text{and} \quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2. \qquad \diamondsuit$$

REMARK: Note that **sample statistics are random variables**, while population characteristics (like the population mean μ or the population variance σ^2) are fixed constants.

Furthermore note that the distribution of sample statistics $T(X_1, ..., X_n)$ depends on the joint distribution of the random sample $X_1, ..., X_n$. $\diamondsuit$

In the following sections we will examine a number of statistics (including the sample mean and variance) that will be useful in statistical inference.

The **empirical distribution function (edf)** provides information about the functional form of the cdf of the underlying population from which a sample is drawn.

DEFINITION (EMPIRICAL DISTRIBUTION FUNCTION): Let $X_1, ..., X_n$ denote a random sample from a population distribution. Then the edf is the following function

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty,t]}(X_i), \qquad t \in (-\infty, \infty).$$

The realization of the random variable $F_n(t)$ is denoted by $\hat{F}_n(t)$.

REMARK: The edf F_n at point t represents the fraction of sample variables that have values $\leq t$.

The edf $F_n(t)$ is the empirical counterpart of the population cdf $F(t)$. Note that the cdf of a population is typically unknown, since it depends on unknown parameters and/or an unknown law of distribution.

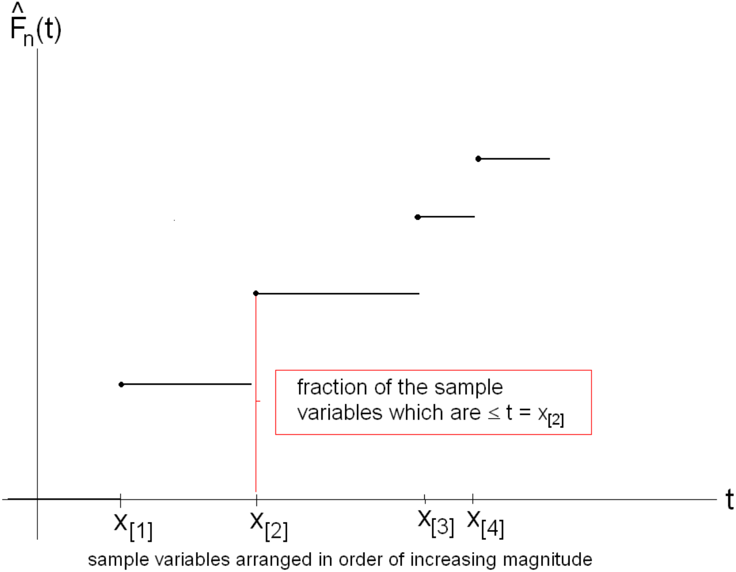

$\hat{F}_n(t)$
fraction of the sample variables which are $\leq t = x_{[2]}$
t
$x_{[1]}$
$x_{[2]}$
$x_{[3]}$
$x_{[4]}$
sample variables arranged in order of increasing magnitude

Fig. 32.

In order to assess the information content of the edf w.r.t. the underlying population distribution, it is useful to examine the properties of the random variable $F_n(t)$. We begin with the **pdf of the edf**.

THEOREM 6.1 *Let $F_n(t)$ be the edf corresponding to a random sample of size n from a population with cdf $F(t)$. Then the pdf of $F_n(t)$ is*

$$P\left(\hat{F}_n(t) = \frac{j}{n}\right) = \begin{cases} \binom{n}{j}\,[F(t)]^j\,[1-F(t)]^{n-j} & \textit{for} \quad j \in \{0,1,2,\dots,n\}\,, \\ 0 \quad \textit{otherwise.} \end{cases}$$

PROOF: From the definition of $F_n(t)$, it follows that

$$P\left(\hat{F}_n(t) = \tfrac{j}{n}\right) \;=\; P(n\hat{F}_n(t) = j) \;=\; P\left(\textstyle\sum_{i=1}^n \underbrace{\mathbb{I}_{(-\infty,t]}\,(x_i)}_{=\,y_i} = j\right).$$

Note that $\quad Y_i = \mathbb{I}_{(-\infty,t]}\,(X_i) \sim$ Bernoulli, with $P(y_i = 1) = P(x_i \leq t) = F(t)$.
Since the X_i's are iid, it follows that the Y_is are also iid, such that

$$\textstyle\sum_{i=1}^n Y_i \;=\; \sum_{i=1}^n \mathbb{I}_{(-\infty,t]}\,(X_i) \;\sim\; \mathrm{Bin}[n, F(t)]$$

(CONTINUES)

PROOF (CONTINUED): Thus we obtain

$$P\left(\hat{F}_n(t) = \tfrac{j}{n}\right) \;=\; P\left(\textstyle\sum_{i=1}^n \mathbb{I}_{(-\infty,t]}(x_i) = j\right) = \underbrace{\binom{n}{j}[F(t)]^j\,[1-F(t)]^{n-j}}_{\text{binomial pdf evaluated at } j}. \qquad \square$$

REMARK: Based upon the pdf of the edf, it is rather straightforward to derive the mean, the variance and the asymptotic behavior of $F_n(t)$.

- The **mean of the edf** at point t is

$$\mathrm{E}F_n(t) = \mathrm{E}[\tfrac{1}{n}\textstyle\sum_{i=1}^n \mathbb{I}_{(-\infty,t]}(X_i)] = \underbrace{\mathrm{E}(\tfrac{1}{n}\textstyle\sum_{i=1}^n Y_i) = \tfrac{1}{n}[nF(t)]}_{\sum_{i=1}^n Y_i \,\sim\, \text{Binomial with mean } nF(t)} = F(t).$$

  Hence the distribution of $F_n(t)$ is centered on the value of the population cdf $F(t)$. (This means that $F_n(t)$ provides an *unbiased estimate* for the value of $F(t)$.)

(CONTINUES)

Remark (Continued):

- The **variance of the edf** at point t is

$$\mathrm{var}[F_n(t)] = \underbrace{\mathrm{var}(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} Y_i) = \tfrac{1}{n^2}\, nF(t)[1-F(t)]}_{\sum_{i=1}^{n} Y_i \,\sim\, \text{Binomial with variance } nF(t)[1-F(t)]} = \tfrac{1}{n}F(t)[1-F(t)].$$

  Note that the variance of the edf and hence the spread of its distribution decreases as the sample size n increases.

- As to the **probability limit of the sequence of edfs** $\{F_n(t)\}$ at point t: Since $\mathrm{E}F_n(t) = F(t)\ \forall n$, and $\lim_{n\to\infty} \mathrm{var}[F_n(t)] = 0$ it follows that

$$F_n(t) \xrightarrow{m} F(t) \qquad \Rightarrow \qquad \mathrm{plim} F_n(t) = F(t).$$

  This implies that the probability that realizations of $F_n(t)$ differ from $F(t)$ converges to 0 as $n \to \infty$. (This means that $F_n(t)$ provides a *consistent estimate* for the value of $F(t)$.)

(continues)

Remark (Continued):

- Since $F_n(t) = \frac{1}{n}\sum_i Y_i$ is the average of iid Bernoulli variables with mean $\mathrm{E}Y_i = \mu = F(t)$ and variance $\mathrm{var}(Y_i) = \sigma^2 = F(t)[1 - F(t)]$, we can use the CLT of Lindberg-Levy to obtain

$$\frac{\sqrt{n}(\frac{1}{n}\sum_i Y_i - \mu)}{\sigma} = \frac{\sqrt{n}(F_n(t) - F(t))}{\sqrt{F(t)[1 - F(t)]}} \xrightarrow{d} \mathrm{N}(0\ ,\ 1).$$

Hence, the **asymptotic distribution of the edf** at point t is

$$F_n(t) \overset{a}{\sim} \mathrm{N}\Big(F(t)\ ,\ \tfrac{1}{n}F(t)[1 - F(t)]\Big).$$

All in all, these properties possessed by the edf $F_n(t)$ will make it a good statistic to use in providing information about the population cdf $F(t)$. $\diamondsuit$

We have shown that the edf $F_n(t)$ converges in probability to the cdf $F(t)$ for **each value of** t. The **Glivenko-Canetelli Theorem** strengthens this result showing that it is possible to make a probability statement **simultaneously** for all t values.

THEOREM 6.2 (GLIVENKO-CANETELLI THEOREM)

Let[1] $D_n = \sup_{-\infty<t<\infty} \left| F_n(t) - F(t) \right|$. *Then* $P\left(\lim_{n\to\infty} D_n = 0 \right) = 1$.

PROOF: For a proof see Fisz, M. (1976, p. 456-459), *Wahrscheinlichkeitsrechnung und mathematische Statistik,* Berlin, VEB Verlag. □

REMARK: The theorem implies that the *sequence of functions* $\{F_n(t)\}$ converges uniformly as $n \to \infty$ to the *function* $F(t)$ *with probability 1.*

Thus for large enough n, the edf provides a good approximation of the cdf over *its entire domain* (not only for individual points). See Fig. 33. ♢

[1] 'sup S' means the *supremum* or the *least upper bound of the set* S.

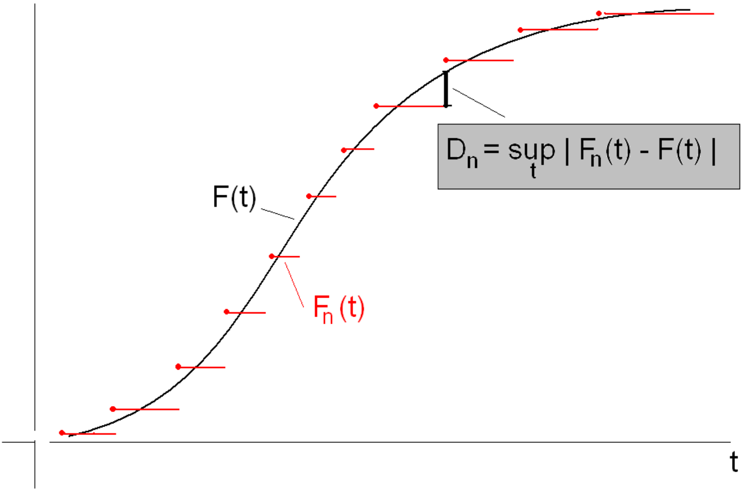


Fig. 33.

Based on random samples, statistics called **sample moments** can be defined that are sample counterparts to the moments of the population distribution (**population moments**) defined in Chapter 3.

The sample moments have properties that make them useful for estimating the values of the corresponding population moments. Therefore, they play a central role for the estimation of parameters.

In the following discussion of sample moments we will assume that the sample variables are from random samples from the population distribution.

Definition (Sample Moments): Let $X_1, ..., X_n$ denote a random sample. Then the **rth order non-central sample moment** (or moment about the origin) is

$$M_r' = \frac{1}{n} \sum_{i=1}^{n} X_i^r.$$

The **rth order central sample moment** (or moment about the mean) is

$$M_r = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^r,$$

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. The realization of the random variables M_r' and M_r are denoted by m_r' and m_r, respectively.

We now discuss some important stochastic properties of the non-central sample moments.

Those properties will suggest that sample moments will be useful for estimating the values of the corresponding population moments.

Let $M'_r = \frac{1}{n}\sum_{i=1}^n X_i^r$ be the rth order non-central sample moment for a random sample $(X_1, ..., X_n)$ from a population distribution. If we assume that the appropriate population moments (denoted by μ'_s) exist, we get the following results:

- For the **mean** of M'_r we obtain

  $$\mathrm{E}M'_r = \frac{1}{n}\sum_{i=1}^n \mathrm{E}X_i^r = \mathrm{E}X_i^r = \mu'_r.$$

  Hence, the mean of the sample moment is equal to the value of the corresponding population moment. (Thus M'_r provides *unbiased estimates* for the value of μ'_r.)

- For the **variance** of M'_r we obtain

  $$\mathrm{var}(M'_r) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(X_i^r) = \frac{1}{n}\mathrm{var}(X_i^r) = \frac{1}{n}\left[\mu'_{2r} - (\mu'_r)^2\right].$$

  This implies that the variance goes to zero as $n \to \infty$.

- The **probability limit** of M_r' obtains as follows. Since $\mathrm{E}M_r' = \mu_r' \ \forall n$, and $\lim_{n\to\infty} \mathrm{var}(M_r') = 0$, we have

$$M_r' \xrightarrow{m} \mu_r' \qquad \Rightarrow \qquad \mathrm{plim}\, M_r' \ = \ \mu_r'.$$

This implies that the probability that realizations of M_r' differ from μ_r' converges to 0 as $n \to \infty$. (Thus M_r' provides *consistent estimates* for the value of μ_r'.)

- Since $M_r' = \frac{1}{n}\sum_i X_i^r$ is the average of iid variables with mean $\mathrm{E}X_i^r = \mu_r'$ and variance $\mathrm{var}(X_i^r) = \left[\mu_{2r}' - (\mu_r')^2\right]$, we can use the CLT of Lindberg-Levy to obtain

$$\frac{\sqrt{n}\left(\frac{1}{n}\sum_i X_i^r - \mu_r'\right)}{\sqrt{\mu_{2r}' - (\mu_r')^2}} \ \xrightarrow{d} \ \mathrm{N}(0\ ,\ 1).$$

Hence, the **asymptotic distribution** of M_r' is

$$M_r' = \frac{1}{n}\sum_i X_i^r \ \overset{a}{\sim} \ \mathrm{N}\left(\mu_r'\ ,\ \frac{1}{n}[\mu_{2r}' - (\mu_r')^2]\right).$$

(This asymptotic distribution is useful for *testing hypotheses* about the value of μ_r'.)

Two of the most commonly used statistics are the **first-order non-central sample moment** also referred to as the **sample mean** and the **second-order central sample moment** also referred to as the **sample variance** .

DEFINITION (SAMPLE MEAN): Let $X_1, ..., X_n$ denote a random sample. The **sample mean** is

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = M_1'.$$

REMARK: From the discussion of the properties of sample moments, we know that

$$\mathrm{E}\bar{X}_n = \mu, \qquad \mathrm{var}(\bar{X}_n) = \tfrac{1}{n}(\mu_2' - \mu^2) = \tfrac{\sigma^2}{n}, \qquad \mathrm{plim}\bar{X}_n = \mu, \qquad \bar{X}_n \overset{a}{\sim} \mathrm{N}(\mu, \tfrac{1}{n}\sigma^2). \quad \diamondsuit$$

Definition (Sample Variance): Let $X_1, \ldots, X_n$ denote a random sample. The **sample variance** is

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = M_2.$$

Important stochastic properties of the sample variance are summarized in the following theorem.

Theorem 6.3 *Let S_n^2 be the sample variance of a random sample $X_1, \ldots, X_n$ from a population distribution. Then assuming that appropriate population moments exist,*

- **a.** $ES_n^2 = \frac{(n-1)}{n} \sigma^2$,
- **b.** $var(S_n^2) = \frac{1}{n} \left[\left(\frac{n-1}{n} \right)^2 \mu_4 - \frac{(n-1)(n-3)}{n^2} \sigma^4 \right]$,
- **c.** $plim\ S_n^2 = \sigma^2$,
- **d.** $\sqrt{n} \left(S_n^2 - \sigma^2 \right) \xrightarrow{d} N \left(0, \mu_4 - \sigma^4 \right)$,
- **e.** $S_n^2 \overset{a}{\sim} N\left(\sigma^2, \frac{1}{n} (\mu_4 - \sigma^4) \right)$.

PROOF:

a. The mean of the sample variance obtains as follows

$$
\begin{aligned}
\mathrm{E}S_n^2 &= \mathrm{E}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i - \mu + \mu - \bar{X}_n)^2\right] \\
&= \mathrm{E}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i - \mu)^2\right] + \mathrm{E}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^n (\mu - \bar{X}_n)^2\right] + 2\mathrm{E}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i - \mu)(\mu - \bar{X}_n)\right] \\
&= \tfrac{1}{n} n\sigma^2 + \mathrm{E}(\mu - \bar{X}_n)^2 + 2\mathrm{E}\left[(\mu - \bar{X}_n)\tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i - \mu)\right] \\
&= \sigma^2 + \mathrm{E}(\mu - \bar{X}_n)^2 + 2\mathrm{E}\left[(\mu - \bar{X}_n)(\bar{X}_n - \mu)\right] \\
&= \sigma^2 + \mathrm{E}(\mu - \bar{X}_n)^2 - 2\mathrm{E}(\mu - \bar{X}_n)^2 \\
&= \sigma^2 - \mathrm{E}(\mu - \bar{X}_n)^2 \\
&= \sigma^2 - \mathrm{var}(\bar{X}_n) = \sigma^2 - \tfrac{1}{n}\sigma^2 = (1 - \tfrac{1}{n})\sigma^2.
\end{aligned}
$$

Note that in contrast to the expectation of the sample mean, the expectation of the sample variance differs from the corresponding population moment. (Hence S_n^2 provides a *biased estimate* for σ^2.)

(CONTINUES)

PROOF (CONTINUED):

b. The proof for the variance of the sample variance follows from rewriting

$$\text{var}(S_n^2) = \text{E}[S_n^2 - \sigma^2(1 - \tfrac{1}{n})]^2$$

in terms of corresponding sums of the X_is and taking expectation. The corresponding algebra is conceptually straightforward, but tedious (for details, see Rohatgi and Saleh, 2001, p. 315-317).

Note that

$$\lim_{n\to\infty} \text{var}(S_n^2) = \lim_{n\to\infty} \left\{ \frac{1}{n} \left[\underbrace{\left(\frac{n-1}{n}\right)^2}_{\to 1} \mu_4 - \underbrace{\frac{(n-1)(n-3)}{n^2}}_{\to 1} \sigma^4 \right] \right\} = 0.$$

c. Since $\text{E}S_n^2 = \frac{(n-1)}{n}\sigma^2 \to \sigma^2$ and $\text{var}(S_n^2) \to 0$ as $n \to \infty$, we have

$$S_n^2 \overset{m}{\to} \sigma^2 \qquad \Rightarrow \qquad \text{plim} S_n^2 = \sigma^2.$$

(Hence S_n^2 provides consistent estimates for σ^2.)

(CONTINUES)

PROOF (CONTINUED):

d. and e. For the proof for the asymptotic distribution of S_n^2, first note that

$$\begin{aligned} nS_n^2 &= \sum_{i=1}^n (X_i - \mu + \mu - \bar{X}_n)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X}_n) \sum_{i=1}^n (X_i - \mu) + n(\mu - \bar{X}_n)^2. \end{aligned}$$

Subtracting $n\sigma^2$ and dividing by $\sqrt{n}$ yields

$$\sqrt{n}(S_n^2 - \sigma^2) = \underbrace{\frac{1}{\sqrt{n}}[\sum_{i=1}^n (X_i - \mu)^2 - n\sigma^2]}_{Z_n \xrightarrow{d} \mathrm{N}(0,\, \mu_4 - \sigma^4)} + \underbrace{2(\mu - \bar{X}_n)\sqrt{n}\frac{1}{n}\sum_{i=1}^n (X_i - \mu)}_{V_n \xrightarrow{p} 0} + \underbrace{\sqrt{n}(\bar{X}_n - \mu)^2}_{W_n \xrightarrow{p} 0}.$$

- Regarding the limiting behavior of the second term V_n, note that $\sqrt{n}\frac{1}{n}\sum_{i=1}^n (X_i - \mu) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$, by the CLT of Lindberg-Levy, and $\text{plim}(\mu - \bar{X}_n) = \text{plim}\mu - \text{plim}\bar{X}_n = 0$. So by Slutsky's theorem $V_n \xrightarrow{p} 0$.
- The last term W_n can be written as

$$W_n = [n^{1/4}\tfrac{1}{n}(\sum_{i=1}^n X_i - n\mu)]^2 = [\underbrace{\tfrac{1}{n^{3/4}}(\sum_{i=1}^n X_i - n\mu)}_{U_n}]^2$$

where $\mathrm{E}U_n = 0$, $\text{var}(U_n) = \frac{1}{n^{3/2}} n\sigma^2 \to 0$, so that $U_n \xrightarrow{p} 0$ and $\text{plim}\, W_n = [\text{plim}\, U_n]^2 = 0$.

(CONTINUES)

PROOF (CONTINUED):

- The first term Z_n can be rewritten as

$$Z_n = \frac{1}{\sqrt{n}}[\sum_{i=1}^{n}(X_i - \mu)^2 - n\sigma^2] = \sqrt{n}[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2],$$

where $(X_i - \mu)^2$ is an iid variable with $\mathrm{E}(X_i - \mu)^2 = \sigma^2$ and $\mathrm{var}[(X_i - \mu)^2] = \mathrm{E}(X_i - \mu)^4 - [\mathrm{E}(X_i - \mu)^2]^2 = \mu_4 - \sigma^4$. So by Lindberg-Levy's CLT $Z_n \xrightarrow{d} \mathrm{N}(0, \mu_4 - \sigma^4)$.

Collecting all terms we have by Slutsky's theorem

$$\sqrt{n}(S_n^2 - \sigma^2) = (Z_n + V_n + W_n) \xrightarrow{d} \mathrm{N}(0, \mu_4 - \sigma^4),$$

$$\text{so that} \qquad S_n^2 \overset{a}{\sim} \mathrm{N}(\sigma^2, \frac{1}{n}[\mu_4 - \sigma^4]).$$

(This asymptotic distribution is useful for *testing hypotheses* about the value of σ^2.) □

Sample Covariance

So far we considered sample moments for random samples with scalar random variables. For random samples with multivariate variables, the **joint sample moments** between pairs of variables become relevant.

One of the most commonly used joint sample moment is the **sample covariance**. It is the sample counterpart of the population covariance.

> DEFINITION (SAMPLE COVARIANCE): Let $(X_1, Y_1), ..., (X_n, Y_n)$ denote a random sample. Then the **sample covariance** is
>
> $$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}_n \bar{Y}_n.$$

We now examine important properties of the sample covariance as the sample counterpart of the population covariance σ_{XY}.

Let S_{XY} be the sample covariance for a random sample $(X_1, Y_1), ..., (X_n, Y_n)$ from a joint population distribution. If we assume that the appropriate population moments exist, we get the following results:

- For the **mean** of S_{XY} we obtain

$$\mathrm{E}S_{XY} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \mathrm{E}[(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)].$$

where

$$\mathrm{E}[(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)]$$
$$= \mathrm{E}\Big[X_i Y_i - \underbrace{\tfrac{1}{n} X_i \textstyle\sum_{i=1}^{n} Y_i}_{\frac{1}{n} X_i (Y_1 + \ldots + Y_n)} - \underbrace{\tfrac{1}{n} Y_i \textstyle\sum_{i=1}^{n} X_i}_{\frac{1}{n} Y_i (X_1 + \ldots + X_n)} + \underbrace{\tfrac{1}{n^2}(\textstyle\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}_{\frac{1}{n^2}(X_1 + \ldots + X_n)(Y_1 + \ldots + Y_n)}\Big].$$

Since $(X_1, Y_1), ..., (X_n, Y_n)$ are iid X_i is independent of X_j and Y_j for $i \neq j$, and vice versa, we obtain

$$\begin{aligned}\mathrm{E}[(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)]& \\ &= \mu'_{1,1} - 2\tfrac{1}{n}\Big[\mu'_{1,1} + (n-1)\mu_X\mu_Y\Big] + \tfrac{1}{n^2}\Big[n\mu'_{1,1} + (n^2 - n)\mu_X\mu_Y\Big]. \\ &= \mu'_{1,1}(1 - \tfrac{1}{n}) - (1 - \tfrac{1}{n})\mu_X\mu_Y \quad = \quad (1 - \tfrac{1}{n})\sigma_{XY} \qquad .\end{aligned}$$

Using this result, we find that

$$\mathrm{E}S_{XY} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)] = (\frac{n-1}{n})\sigma_{XY}.$$

Hence, as it is the case for the sample variance, the expectation of the sample covariance differs from the corresponding population moment. However, the difference goes to 0 as $n \to \infty$, so that the distribution of S_{XY} becomes centered on σ_{XY}.

- The **variance** of S_{XY} has the form

$$\mathrm{var}(S_{XY}) = \frac{1}{n}[\mu_{2,2} - (\mu_{1,1})^2] + o(\frac{1}{n}).$$

This result obtains from a Taylor series expansion (see Kendall M., Stuart, A., (1994, Chap. 13) *The Advanced Theory of Statistics, Vol. 1*, New York, Wiley).

Note, that the variance of S_{XY} disappears as $n \to \infty$, so that its distribution increasingly concentrates within a small neighborhood of σ_{XY}.

- Since $\mathrm{E}S_{XY} \to \sigma_{XY}$ and $\lim_{n\to\infty}\mathrm{var}(S_{XY}) = 0$, we have

$$S_{XY} \overset{m}{\to} \sigma_{XY} \qquad \Rightarrow \qquad \mathrm{plim} S_{XY} = \sigma_{XY}.$$

- In order to obtain the **asymptotic distribution** of S_{XY}, represent S_{XY} as

$$S_{XY} \;=\; \underbrace{\tfrac{1}{n}\sum_{i=1}^{n} X_i Y_i}_{=M'_{1,1}} - \underbrace{\left(\tfrac{1}{n}\sum_{i=1}^{n} X_i\right)}_{=\bar{X}_n}\underbrace{\left(\tfrac{1}{n}\sum_{i=1}^{n} Y_i\right)}_{=\bar{X}_n} \;=\; g(M'_{1,1}, \bar{X}_n, \bar{Y}_n).$$

Note that $M'_{1,1}$, $\bar{X}_n$, and $\bar{Y}_n$ are averages of the iid variables $(X_i Y_i)$, X_i, and Y_i respectively, with mean and covariance matrix

$$\mathrm{E}\begin{bmatrix} X_i Y_i \\ X_i \\ Y_i \end{bmatrix} \;=\; \begin{bmatrix} \mu'_{1,1} \\ \mu_X \\ \mu_Y \end{bmatrix} = \boldsymbol{\mu}.$$

$$\mathbf{Cov}\begin{bmatrix} X_i Y_i \\ X_i \\ Y_i \end{bmatrix} \;=\; \begin{bmatrix} \mu'_{2,2} - (\mu'_{1,1})^2 & \mu'_{2,1} - \mu_{1,1}\mu_X & \mu'_{1,2} - \mu_{1,1}\mu_Y \\ \cdot & \sigma_x^2 & \sigma_{XY} \\ \cdot & \cdot & \sigma_Y^2 \end{bmatrix} = \boldsymbol{\Sigma}.$$

Then by the multivariate CLT of Lindberg-Levy

$$\sqrt{n}\begin{bmatrix} M'_{1,1} - \mu'_{1,1} \\ \bar{X}_n - \mu_X \\ \bar{Y}_n - \mu_Y \end{bmatrix} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Since $S_{XY} = g(M'_{1,1}, \bar{X}_n, \bar{Y}_n) = M'_{1,1} - \bar{X}_n \cdot \bar{Y}_n$ is a differentiable function of asymptotically normally distributed variables, S_{XY} itself is by Theorem 5.18

asymptotically normally distributed

$$S_{XY} = g(M'_{1,1}, \bar{X}_n, \bar{Y}_n) \overset{a}{\sim} \mathsf{N}\Big[\underbrace{(\mu'_{1,1} - \mu_X\mu_Y)}_{g(\boldsymbol{\mu}) = \sigma_{XY}},\ \tfrac{1}{n}\boldsymbol{G\Sigma G}'\Big]$$

where

$$\begin{aligned} \boldsymbol{G} &= (1, -\mu_Y, -\mu_X) \qquad\qquad \text{(gradient vector of } g \text{ evaluated at } \boldsymbol{\mu}) \\ \tfrac{1}{n}\boldsymbol{G\Sigma G}' &= \frac{1}{n}\Big[\mu_{2,2} - (\mu'_{1,1})^2\Big]. \end{aligned}$$

Thus the asymptotic distribution of the sample covariance is

$$S_{XY} \overset{a}{\sim} \mathsf{N}\Big(\sigma_{XY}, \frac{1}{n}[\mu_{2,2} - (\mu'_{1,1})^2]\Big).$$

As we have similarly argued in the case of the sample mean and variance, the properties of the sample covariance S_n^2 derived above are useful for estimating its population counterpart, the covariance σ_{XY}.

In the previous section we investigated the properties of the sample mean $\bar{X}_n$ and sample variance S_n^2 for a random sample, without assuming a specific population distribution for the random sample.

This section deals with **additional properties** of $\bar{X}_n$ and S_n^2 that arise when random sampling is from a **Normal distribution** - still one of the most widely used statistical models.

In particular, we will show that $\bar{X}_n$ and S_n^2 are then *independent* random variables, $\bar{X}_n$ is then *normally distributed*, and S_n^2 is then *Gamma distributed*.

In order to show the independence of $\bar{X}_n$ and S_n^2 when the random sample is from a Normal distribution, the following result is useful.

THEOREM 6.4 *Let*

$\boldsymbol{B}$: *real* $(q \times n)$ *matrix*, $\boldsymbol{A}$: *real symmetric* $(n \times n)$ *matrix with rank* p,

$\boldsymbol{X}$: $(n \times 1)$ *random vector with a* $N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$*-distribution.*

Then the linear form $\boldsymbol{BX}$ *and the quadratic form* $\boldsymbol{X}'\boldsymbol{AX}$ *are independent, if* $\boldsymbol{BA} = \boldsymbol{0}$.

PROOF: Let the $(n \times n)$ diagonal matrix of the **eigenvalues** of $\boldsymbol{A}$ be denoted by[2]

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \qquad \text{where } \lambda_i : i\text{th eigenvalue of } \boldsymbol{A},$$

and the $(n \times n)$ matrix of the corresponding **eigenvectors** by

$$\boldsymbol{P} = (\boldsymbol{P}_1, \ldots, \boldsymbol{P}_n), \qquad \text{where } P_i : i\text{th eigenvector of } \boldsymbol{A},$$

with $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}$, so that $\boldsymbol{P}' = \boldsymbol{P}^{-1}$ and $\boldsymbol{PP}' = \boldsymbol{PP}^{-1} = \boldsymbol{I}$.

(CONTINUES)

[2]The eigenvalues and - vectors of $\boldsymbol{A}$ obtain from the n nontrivial solutions of $(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{P}_i = 0$ under the normalizing restrictions $\boldsymbol{P}_i'\boldsymbol{P}_i = 1$ and $\boldsymbol{P}_i'\boldsymbol{P}_j = 0\ \forall i \neq j$.

PROOF (CONTINUED): Then, the spectral decomposition of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{P\Lambda P}', \quad \text{so that} \quad \boldsymbol{P}'\boldsymbol{AP} = \underbrace{\boldsymbol{\Lambda} = \left[\begin{array}{ccc|c} \lambda_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \lambda_p & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]}_{rk(A) = p \Rightarrow p = (\#\text{ of eigenvalues} \neq 0)} \stackrel{\text{(say)}}{=} \left[\begin{array}{c|c} \underset{(p \times p)}{\boldsymbol{D}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right].$$

Now let $\boldsymbol{BA} = \mathbf{0}$, so that

$$\boldsymbol{BAP} = \mathbf{0} \quad \text{and} \quad \boldsymbol{B}\underbrace{\boldsymbol{PP}'}_{\boldsymbol{I}}\boldsymbol{AP} = \mathbf{0},$$

and let

$$\underset{(q \times n)}{\boldsymbol{C}} = \underset{(q \times n)}{\boldsymbol{B}}\underset{(n \times n)}{\boldsymbol{P}}, \quad \text{so that} \quad \underbrace{\boldsymbol{BP}}_{\boldsymbol{C}}\underbrace{\boldsymbol{P}'\boldsymbol{AP}}_{\boldsymbol{\Lambda}} = \boldsymbol{C\Lambda} = \mathbf{0}.$$

Partitioning $\boldsymbol{C}$ conformably with $\boldsymbol{\Lambda}$, we get

$$\boldsymbol{C\Lambda} = \left[\begin{array}{c|c} \underset{(q \times p)}{\boldsymbol{C}_1} & \underset{(q \times n-p)}{\boldsymbol{C}_2} \end{array}\right] \cdot \left[\begin{array}{c|c} \boldsymbol{D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{C}_1\boldsymbol{D} + \boldsymbol{C}_2\mathbf{0} & \boldsymbol{C}_1\mathbf{0} + \boldsymbol{C}_2\mathbf{0} \end{array}\right] = \mathbf{0}.$$

This implies that $\boldsymbol{C}_1\boldsymbol{D} = \mathbf{0}$ and, since $\boldsymbol{D} \neq \mathbf{0}$, that $\boldsymbol{C}_1 = \mathbf{0}$. Thus the matrix $\boldsymbol{C} = \boldsymbol{BP}$ must have the form

$$\boldsymbol{C} = \left[\begin{array}{c|c} \mathbf{0} & \boldsymbol{C}_2 \end{array}\right].$$

(CONTINUES)

PROOF (CONTINUED): Now use the eigenvector matrix $\boldsymbol{P}$ and $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ to define

$$\underset{(n \times 1)}{\boldsymbol{Z}} = \boldsymbol{P}'\boldsymbol{X}, \qquad \text{where} \qquad \boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{P}'\boldsymbol{\mu}, \sigma^2 \boldsymbol{P}'\boldsymbol{P}) = \mathrm{N}(\boldsymbol{P}'\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}).$$

Note that the elements in $\boldsymbol{Z}$, say $(Z_1, \ldots, Z_n)$, are independent variables since they are uncorrelated and normally distributed.

Collecting all terms, we obtain for the quadratic form of $\boldsymbol{X}$

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \underbrace{\boldsymbol{X}\boldsymbol{P}}_{\boldsymbol{Z}'} \underbrace{\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}}_{\boldsymbol{\Lambda}} \underbrace{\boldsymbol{P}'\boldsymbol{X}}_{\boldsymbol{Z}} = \boldsymbol{Z}'\boldsymbol{\Lambda}\boldsymbol{Z} = \sum_{i=1}^{p} \lambda_i Z_i^2.$$

For the linear form of $\boldsymbol{X}$ we obtain

$$\boldsymbol{B}\boldsymbol{X} = \underbrace{\boldsymbol{B}\boldsymbol{P}}_{\boldsymbol{C}} \underbrace{\boldsymbol{P}'\boldsymbol{X}}_{\boldsymbol{Z}} = \boldsymbol{C}\boldsymbol{Z} = \left[\begin{array}{c|c} \underset{(q \times p)}{\boldsymbol{0}} & \underset{(q \times n-p)}{\boldsymbol{C}_2} \end{array} \right] \cdot \left[\begin{array}{c} Z_1 \\ \vdots \\ Z_p \\ \hline Z_{p+1} \\ \vdots \\ Z_n \end{array} \right] = \boldsymbol{C}_2 \cdot \left[\begin{array}{c} Z_{p+1} \\ \vdots \\ Z_n \end{array} \right].$$

Thus

$$\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = g_1(Z_1, \ldots, Z_p) \qquad \text{and} \qquad \boldsymbol{B}\boldsymbol{X} = g_2(Z_{p+1}, \ldots, Z_n),$$

and because $(Z_1, \ldots, Z_p)$ and $(Z_{p+1}, \ldots, Z_n)$ are independent, $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{B}\boldsymbol{X}$ are independent. □

We now use the preceding theorem to establish the independence of $\bar{X}_n$ and S_n^2 and the distribution of nS_n^2/σ^2 when the random sample is from a Normal distribution.

Theorem 6.5 *Let $\bar{X}_n$ and S_n^2 are the sample mean and sample variance of a size-n random sample from a $N(\mu, \sigma^2)$-distribution. Then*

a. $\bar{X}_n \sim N(\mu, \frac{1}{n}\sigma^2)$,

b. $\bar{X}_n$ *and* S_n^2 *are independent,*

c. $nS_n^2/\sigma^2 \sim \chi^2_{n-1}$.

Proof: **a.** The normality of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ follows immediately from the fact that $\bar{X}_n$ is a linear combination of iid Gaussian random variables.

(Hence $\bar{X}_n$ is not only asymptotically normally distributed but also exactly normally distributed when we sample from a Normal distribution.)

(continues)

PROOF (CONTINUED): **b.** Write the sample mean as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \underbrace{[\frac{1}{n}, \ldots, \frac{1}{n}]}_{B_{(1\times n)}} \underbrace{\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}}_{X\ \sim \mathrm{N}(\mathbf{0},\, \sigma^2 I)} = BX.$$

Also write the vector of differences between the sample variables and the sample mean as

$$\begin{bmatrix} X_1 - \bar{X}_n \\ \vdots \\ X_n - \bar{X}_n \end{bmatrix} = IX - \underbrace{\begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}}_{H_{(n\times n)}} X = (I - H)X,$$

where $(I - H)$ is symmetric and idempotent, i.e. $(I - H)(I - H) = (I - H)$. Using the matrix $(I - H)$, we can write the sample variance as

$$\begin{aligned} S_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}[(X_1 - \bar{X}_n), \ldots, (X_n - \bar{X}_n)] \cdot \begin{bmatrix} X_1 - \bar{X}_n \\ \vdots \\ X_n - \bar{X}_n \end{bmatrix} \\ &= \frac{1}{n}X'(I - H)'(I - H)X = X' \underbrace{\frac{1}{n}(I - H)}_{A} X = X'AX. \end{aligned}$$

(CONTINUES)

PROOF (CONTINUED): It follows from Theorem 6.4 that the sample mean $\bar{X}_n = \boldsymbol{BX}$ and the sample variance $S_n^2 = \boldsymbol{X}'\boldsymbol{AX}$ are independent since

$$\begin{aligned} \boldsymbol{BA} &= \boldsymbol{B}\frac{1}{n}(\boldsymbol{I} - \boldsymbol{H}) = \frac{1}{n}(\boldsymbol{B} - \boldsymbol{BH}) \\ &= \frac{1}{n}(\boldsymbol{B} - \boldsymbol{B}) = \boldsymbol{0}. \qquad \left(\boldsymbol{BH} = [\tfrac{1}{n}, \ldots, \tfrac{1}{n}] \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} = [n\frac{1}{n^2}, \ldots, n\frac{1}{n^2}] = \boldsymbol{B}\right) \end{aligned}$$

(CONTINUES)

PROOF (CONTINUED): **c.** In order to obtain the distribution of nS_n^2/σ^2, we represent this variable as

$$\frac{nS_n^2}{\sigma^2} = \frac{1}{\sigma^2}\boldsymbol{X}'(\boldsymbol{I}-\boldsymbol{H})'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X}.$$

Furthermore, let

$$\boldsymbol{\mu_X} = (\mu, \ldots, \mu)' \qquad \text{and note that} \qquad (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\mu_X} = \boldsymbol{\mu_X} - \boldsymbol{\mu_X} = \boldsymbol{0}.$$

This allows us to write

$$\begin{aligned}
\frac{nS_n^2}{\sigma^2} &= \frac{1}{\sigma^2}\Big[\boldsymbol{X}'(\boldsymbol{I}-\boldsymbol{H})' - \underbrace{\boldsymbol{\mu_X}'(\boldsymbol{I}-\boldsymbol{H})'}_{\boldsymbol{0}}\Big]\Big[(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X} - \underbrace{(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\mu_X}}_{\boldsymbol{0}}\Big] \\
&= \frac{1}{\sigma^2}\Big[(\boldsymbol{X}-\boldsymbol{\mu_X})'(\boldsymbol{I}-\boldsymbol{H})'(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{X}-\boldsymbol{\mu_X})\Big] \\
&= \frac{1}{\sigma^2}\Big[(\boldsymbol{X}-\boldsymbol{\mu_X})'(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{X}-\boldsymbol{\mu_X})\Big].
\end{aligned}$$

Since the $(n \times n)$ matrix $(\boldsymbol{I}-\boldsymbol{H})$ is **symmetric** and **idempotent**, it implies that[3]

$$\text{rk}(\boldsymbol{I}-\boldsymbol{H}) = \text{trace}(\boldsymbol{I}-\boldsymbol{H}) = \text{trace}(\boldsymbol{I}) - \text{trace}(\boldsymbol{H}) = n-1,$$

and that

$$\text{the eigenvalues of } (\boldsymbol{I}-\boldsymbol{H}) \text{ are either 0 or 1, with}$$
$$(\#\text{ eigenvalues} = 1) = \text{rk}(\boldsymbol{I}-\boldsymbol{H}) = n-1.$$

(CONTINUES)

[3] See Lütkepohl (1996, Chap. 5.2), *Handbook of Matrices*, Chichester, Wiley.

PROOF (CONTINUED): The spectral decomposition of $(\boldsymbol{I}-\boldsymbol{H})$ into the matrices of eigenvalues $(\boldsymbol{\Lambda})$ and eigenvectors $(\boldsymbol{P})$, i.e.

$$\boldsymbol{I}-\boldsymbol{H}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}',$$

then yields

$$\boldsymbol{P}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{P}=\boldsymbol{\Lambda}=\left[\begin{array}{c|c}\boldsymbol{I}_{(n-1)} & \mathbf{0} \\ \hline \mathbf{0} & 0\end{array}\right].$$

If this is accounted for in the last equation for nS_n^2/σ^2 we get

$$\begin{aligned}\frac{nS_n^2}{\sigma^2} &= \frac{1}{\sigma^2}\left[(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})'(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})\right]=\frac{1}{\sigma^2}\left[(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})\right] \\ &= \boldsymbol{Z}'\boldsymbol{\Lambda}\boldsymbol{Z},\end{aligned}$$

$$\text{where} \quad \boldsymbol{Z}=\boldsymbol{P}'\underbrace{\left(\frac{\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}}{\sigma}\right)}_{\substack{\text{vector of iidN(0,1)} \\ \text{variables}}} \quad \sim \quad \mathrm{N}(\mathbf{0},\boldsymbol{P}'\boldsymbol{P})=\mathrm{N}(\mathbf{0},\boldsymbol{I}).$$

Thus

$$\frac{nS_n^2}{\sigma^2}=\boldsymbol{Z}'\boldsymbol{\Lambda}\boldsymbol{Z}=\sum_{i=1}^{n-1}Z_i^2 \quad \sim \quad \chi^2_{(n-1)}. \qquad \square$$

From the fact that $nS_n^2/\sigma^2 \sim \chi^2_{(n-1)}$ it follows that the sample variance S_n^2 is Gamma-distributed as stated in the following theorem.

Theorem 6.6 *Under the assumptions of Theorem 6.5,*

$$S_n^2 \sim \textit{Gamma with} \qquad \alpha = \frac{n-1}{2},\ \beta = \frac{2\sigma^2}{n}.$$

Proof: Let

$$Y = \frac{nS_n^2}{\sigma^2} \sim \chi^2_{(n-1)}, \qquad \text{so that} \qquad S_n^2 = \frac{Y\sigma^2}{n}.$$

Then the MGF for S_n^2 is obtained as

$$\begin{aligned} M_{S_n^2}(t) &\overset{\text{(def)}}{=} \mathrm{E}\exp\{S_n^2 t\} = \mathrm{E}\exp\{Y\underbrace{\left(\frac{\sigma^2}{n}t\right)}_{t^*}\} = \underbrace{(1-2t^*)^{-\frac{(n-1)}{2}}}_{\text{MGF of } Y \sim \chi^2_{(n-1)}} \\ &= (1-\underbrace{2\frac{\sigma^2}{n}}_{\beta}t)^{-\frac{(n-1)}{2}}, \end{aligned}$$

which is the MGF associated with the Gamma distribution having $\alpha = \frac{n-1}{2}$ and $\beta = \frac{2\sigma^2}{n}$. $\square$

REMARK: The theorem implies that the sample variance S_n^2 is exactly Gamma distributed when we sample from a Normal distribution.

Hence if the normal model is supposed to be the correct model, we do not need to rely on the normal approximation for the S_n^2 distribution which is given in Theorem 6.3.
$\diamondsuit$

In the preceding section, we examined a number of statistics (sample mean, sample variance, etc.), which are useful in a number of statistical inference problems.

However, one needs to be concerned with a much larger variety of functions of random samples to deal with the variety of inference problems that arise in practice.

Furthermore, in order to assess the properties of statistical procedures, it will be necessary to identify the distribution for functions of random samples that are used as estimators or as hypothesis-testing statistics.

In this section, we will discuss three approaches, which can be used to derive the **pdf for functions of a random sample**:

1.) MGF Approach,
2.) Equivalent Events Approach,
3.) Change of Variable Approach.

MGF Approach

- Let $X_1, ..., X_n$ be a random sample from a known population distribution, and let $Y = g(X_1, ..., X_n)$ denote the function of interest.
- Then one can attempt to derive the MGF of Y, i.e.

$$M_Y(t) = \mathrm{E}e^{Yt} = \mathrm{E}e^{g(X_1,...,X_n)t},$$

and identify the distribution associated with that MGF.

EXAMPLE: Consider the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, when $X_i \sim$ iid Exponential(θ).

The MGF of $\bar{X}_n$ is

$$M_{\bar{X}_n}(t) = \mathrm{E}\exp\left\{\left(\tfrac{1}{n}\sum_{i=1}^n X_i\right)t\right\} = \mathrm{E}\prod_{i=1}^n \exp\left\{X_i \tfrac{t}{n}\right\} \overset{\text{(indep.)}}{=} \prod_{i=1}^n \underbrace{\mathrm{E}\exp\left\{X_i \tfrac{t}{n}\right\}}_{\substack{\text{MGF of } X_i \sim \\ \text{Exponential}(\theta) \\ \text{evaluated at } \frac{t}{n}}}$$

$$= \prod_{i=1}^n \left(1-\theta\tfrac{t}{n}\right)^{-1} = \left(1-\theta\tfrac{t}{n}\right)^{-n}, \quad \text{for} \quad t < \tfrac{n}{\theta},$$

which is the MGF of a Gamma distribution with $\alpha = n$ and $\beta = \frac{\theta}{n}$. Thus the sampling distribution of $\bar{X}_n$ is gamma. $||$

REMARK : Note that the MGF approach is applicable only if the MGF does exist and if we know a distribution which corresponds to the MGF when the MGF does exist. $\diamondsuit$

Equivalent Events Approach (Discrete Case)

In the case of **discrete random variables**, we can use the **equivalent-events approach** for deriving the distribution of functions of random variables.

- Let $Y = g(X)$ be the function of interest, where X represents a discrete variable with pdf f_x;
- Consider the set of elementary events x generating a particular elementary event y, i.e,

$$A_y = \{x : y = g(x)\ ,\ x \in R(X)\};$$

- Then the probability for the elementary event y can be written as

$$P_Y(y) = P_X(x \in A_Y) = \sum_{\{x \in A_Y\}} f_X(x) = f_Y(y),$$

  which defines the pdf of Y.
- The extension to the case of multivariate variables is straightforward.

EXAMPLE: Let $\boldsymbol{X} = (X_1, X_2, X_3)'$ have a joint discrete pdf given by

$\boldsymbol{x}$	(0,0,0)	(0,0,1)	(0,1,1)	(1,0,1)	(1,1,0)	(1,1,1)
$f_{\boldsymbol{X}}(\boldsymbol{x})$	1/8	3/8	1/8	1/8	1/8	1/8

What is the joint pdf of $\boldsymbol{Y} = (Y_1, Y_2)$ with $Y_1 = X_1 + X_2 + X_3$ and $Y_2 = |X_3 - X_2|$?

The mapping between the outcomes in the range of $\boldsymbol{X}$ and that of $\boldsymbol{Y}$ is

$\boldsymbol{x}$	(0,0,0)	(0,0,1)	(0,1,1)	(1,0,1)	(1,1,0)	(1,1,1)
$\boldsymbol{y}$	(0,0)	(1,1)	(2,0)	(2,1)	(2,1)	(3,0)

Hence the joint pdf of $\boldsymbol{Y} = (Y_1, Y_2)$ obtains as

$\boldsymbol{y}$	(0,0)	(1,1)	(2,0)	(2,1)	(3,0)
$f_{\boldsymbol{Y}}(\boldsymbol{y})$	1/8	3/8	1/8	2/8	1/8

||

Change of Variables Approach (Continuous case)

A useful technique for deriving the pdf of functions of **continuous random variables** is the **change-of-variables technique**.

If the function of interest $y = g(x)$ is monotone and continuously differentiable, the expression for the pdf of Y is given in the following theorem.

THEOREM 6.7 *Let X be a continuous random variable with a pdf $f(x)$ with support $\Xi = \{x : f(x) > 0\}$. Suppose that $y = g(x)$ is a continuously differentiable function with*

$\frac{dg(x)}{dx} \neq 0 \;\; \forall \;\; x$ *in some open interval* Δ *containing* Ξ,

and an inverse $x = g^{-1}(y)$ *defined* $\forall \;\; y \in \Psi = \{y : y = g(x), x \in \Xi\}$.

Then the pdf of $Y = g(X)$ is given by

$$h(y) = f\left(g^{-1}(y)\right) \left| \frac{dg^{-1}(y)}{dy} \right| \quad \text{for} \quad y \in \Psi.$$

PROOF: The fact that $g(x)$ is continuously differentiable with $\frac{dg(x)}{dx} \neq 0$ implies that g is either monotonically increasing (case a) or monotonically decreasing (case b) - see Fig. 34.

Case a., where $\frac{dg(x)}{dx} > 0$: In this case we have (see Fig. 34)

$$P(y \leq b) = P\left(x \leq g^{-1}(b)\right).$$

Thus the cdf for Y obtains as

$$H(b) \stackrel{\text{(def.)}}{=} P(y \leq b) = P\left(x \leq g^{-1}(b)\right) = \int_{-\infty}^{g^{-1}(b)} f(x)dx.$$

Then the pdf of Y obtains by differentiation of H as

$$h(b) \stackrel{\text{(def.)}}{=} \frac{dH(b)}{db} = \frac{d\int_{-\infty}^{g^{-1}(b)} f(x)dx}{db} \stackrel{\text{(chain rule)}}{=} f\left(g^{-1}(b)\right) \underbrace{\frac{dg^{-1}(b)}{db}}_{\substack{\text{note that} \\ \text{this term} \\ \text{is } > 0}} .$$

(CONTINUES)

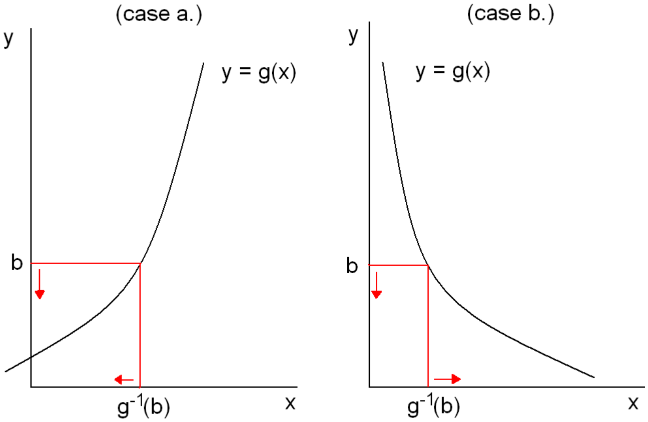


Fig. 34.

Proof (Continued): **Case b.**, where $\frac{dg(x)}{dx} < 0$: In this case the cdf of Y is given by (see Fig. 34)

$$H(b) = P(y \leq b) = P\left(x \geq g^{-1}(b)\right) = \int_{g^{-1}(b)}^{\infty} f(x)dx.$$

Thus the pdf of Y obtains as

$$h(b) = -\, f\left(g^{-1}(b)\right) \underbrace{\frac{dg^{-1}(b)}{db}}_{\substack{\text{note that} \\ \text{this term} \\ \text{is } < 0}} .$$

Combining both cases (increasing and decreasing g), the pdf $h(b)$ can be expressed concisely as

$$h(b) \;=\; f\left(g^{-1}(b)\right)\left|\frac{dg^{-1}(b)}{db}\right|. \qquad \square$$

EXAMPLE: Consider the Cobb-Douglas production function defined as the following product of weighted input factors:

$$Q = \beta_0 \prod_{i=1}^{k} x_i^{\beta_1} \; e^W,$$

where Q : output, $x_i > 0$: deterministic quantities of input factors, β_i : corresponding partial production elasticities, $\beta_0 > 0$: efficiency parameter, and $W \sim \mathrm{N}(0, \sigma^2)$: stochastic error term.

What is the pdf of Q? In order to answer this question rewrite Q as

$$Q = \exp\{\underbrace{\ln \beta_0 + \textstyle\sum_{i=1}^{n} \beta_i \ln x_i + W}_{Z}\} = \exp Z,$$

$$\text{where} \qquad Z \sim \mathrm{N}(\underbrace{\ln \beta_0 + \textstyle\sum_{i=1}^{n} \beta_i \ln x_i}_{\mu_Z}\,,\ \sigma^2) = \mathrm{N}(\mu_Z\,,\ \sigma^2).$$

The function $q = \exp z$ is monotonically increasing with $\frac{dq}{dz} = \exp z > 0 \;\forall\; z$. The inverse is $z = \ln q$ with $\frac{dz}{dq} = \frac{1}{q} > 0$. Thus Theorem 6.7 applies, and the pdf for Q is

$$h(q) = \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma} \exp\{-\frac{(\ln q - \mu_Z)^2}{2\sigma^2}\}}_{f_Z(g^{-1}(q))} \cdot \underbrace{\left(\frac{1}{q}\right)}_{\frac{dg^{-1}(q)}{dq}}, \quad \text{for} \quad q > 0.$$

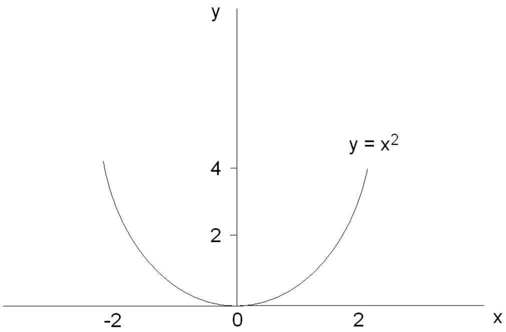
This is the density of a **lognormal distribution**. $\diamondsuit$

Remark : It is important to note that Theorem 6.7 **does not** apply to cases where the function g does not have an inverse such as, for example, the function $y = x^2$ (see Fig. 35).

However, it can be generalized to cases with *piecewise invertible* functions (see Mittelhammer, 1996, p. 338).

Furthermore, the results of Theorem 6.7 can can be extended to the **multivariate case**, as stated in the following theorem. $\diamondsuit$

Fig. 35.



THEOREM 6.8 *Let* $\boldsymbol{X}$ *be a continuous* $(n \times 1)$ *random vector with joint pdf* $f(\boldsymbol{x})$ *with support* Ξ. *Furthermore, let* $\mathbf{g}(\boldsymbol{x})$ *be a* $(n \times 1)$ *vector function which is*

continuously differentiable $\forall\ \boldsymbol{x}$ *in some open rectangle,* Δ, *containing* Ξ,

and with an inverse $\boldsymbol{x} = \boldsymbol{g}^{-1}(\boldsymbol{y})$, *which exists* $\forall\ \boldsymbol{y} \in \boldsymbol{\Psi} = \{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}), \boldsymbol{x} \in \Xi\}$.

Assume that the Jacobian matrix

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial g_1^{-1}(\boldsymbol{y})}{\partial y_1} & \cdots & \frac{\partial g_1^{-1}(\boldsymbol{y})}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n^{-1}(\boldsymbol{y})}{\partial y_1} & \cdots & \frac{\partial g_n^{-1}(\boldsymbol{y})}{\partial y_n} \end{bmatrix} \quad \textit{satisfies} \quad det(\boldsymbol{J}) \neq 0,$$

and assume that all partial derivatives in $\boldsymbol{J}$ *are continuous* $\forall\ \boldsymbol{y} \in \boldsymbol{\Psi}$. *Then the joint pdf of* $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$ *is given by*

$$h(\boldsymbol{y}) = f\left(g_1^{-1}(\boldsymbol{y}), \ldots, g_n^{-1}(\boldsymbol{y})\right) |det\,(\boldsymbol{J})| \quad \textit{for} \quad \boldsymbol{y} \in \boldsymbol{\Psi}.$$

PROOF: See Rohatgi and Saleh (2001), p. 133-134. $\square$

REMARK : In the multivariate change-of-variable Theorem 6.8, there are as many coordinates in $\boldsymbol{y}$ as there are elements in the argument $\boldsymbol{x}$, i.e., $\dim(\boldsymbol{y}) = \dim(\boldsymbol{x}) = n$.

In cases where $\dim(\boldsymbol{y}) < \dim(\boldsymbol{x}) = n$, we need to introduce *auxiliary variables* to obtain an invertible function having n coordinates and to apply Theorem 6.8.

Then, in a second step, the auxiliary variables are integrated out from the derived joint pdf. $\diamondsuit$

We now illustrate the multivariate change-of-variable approach in order to derive the Student-t density and the F-density.

Both densities play an important role for hypothesis-testing procedures when random sampling is from a normal population.

t-Density

THEOREM 6.9 *Let Let $Z \sim N(0,1)$, let $Y \sim \chi^2_v$, and let Z and Y be independent. Then*

$$T = \frac{Z}{\sqrt{Y/v}} \quad \textit{has the t-density with } v \textit{ degrees of freedom,}$$

defined as

$$f(t;v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}.$$

PROOF: The proof is based on the multivariate change-of-variable Theorem 6.8. Define the (2×1) vector function $\boldsymbol{g}$ as

$$\begin{bmatrix} t \\ w \end{bmatrix} = \begin{bmatrix} g_1(z,y) \\ g_2(z,y) \end{bmatrix} = \begin{bmatrix} \frac{z}{\sqrt{y/v}} \\ y \end{bmatrix},$$

where g_2 is an *auxiliary function* introduced to allow the use of Theorem 6.8.

The function $\boldsymbol{g}$ is continuously differentiable with an inverse function $\boldsymbol{g}^{-1}$ which obtains by solving for z and y as

(CONTINUES)

Proof (Continued):

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} g_1^{-1}(t,w) \\ g_2^{-1}(t,w) \end{bmatrix} = \begin{bmatrix} t\sqrt{\frac{w}{v}} \\ w \end{bmatrix},$$

The Jacobian of the inverse $\boldsymbol{g}^{-1}$ is thus

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial g_1^{-1}(\cdot)}{\partial t} & \frac{\partial g_1^{-1}(\cdot)}{\partial w} \\ \frac{\partial g_2^{-1}(\cdot)}{\partial t} & \frac{\partial g_2^{-1}(\cdot)}{\partial w} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{w}{v}} & \frac{t}{2}\sqrt{\frac{v}{w}}\frac{1}{v} \\ 0 & 1 \end{bmatrix}, \qquad \text{with} \qquad |\det(J)| = \sqrt{\frac{w}{v}}.$$

The assumed joint pdf of (Z, Y) is

$$f(z,y) \overset{\text{indep.}}{=} f(z)f(y) = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}}_{Z \sim \mathrm{N}(0,1)} \cdot \underbrace{\frac{1}{2^{v/2}\Gamma(v/2)} y^{(v/2)-1} e^{-y/2}}_{Y \sim \chi^2_v}.$$

Then by Theorem 6.8 the **joint pdf of (T, W)** is given by

$$h(t,w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2\frac{w}{v}} \cdot \frac{1}{2^{v/2}\Gamma(v/2)} w^{(v/2)-1} e^{-w/2} \cdot \sqrt{\frac{w}{v}} = \frac{w^{(v-1)/2} e^{-w[1+(t^2/v)]/2}}{\Gamma(v/2)\sqrt{\pi v}\,2^{(v+1)/2}}$$

The **marginal pdf of T** obtains by integrating w out from the joint pdf $h(t,w)$, i.e.,

(continues)

PROOF (CONTINUED):

$$f(t) = \int_0^\infty h(t,w)dw = \int_0^\infty \frac{w^{(v-1)/2}e^{-w[1+(t^2/v)]/2}}{\Gamma(v/2)\sqrt{\pi v}2^{(v+1)/2}}dw.$$

Making an appropriate variable substitution in this integral yields[4]

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v/2)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}. \qquad \square$$

REMARK : The t-distribution with a pdf as given in Theorem 6.9 has the following moments

$$\mu = 0 \text{ for } v > 1, \qquad \sigma^2 = \frac{v}{v-2} \text{ for } v > 2, \qquad \mu_3 = 0 \text{ for } v > 3,$$

where the parameter $v > 0$ is referred to as *the degrees of freedom*. The MGF does not exist.

As Fig. 36 shows, the t-density is symmetric about 0 and has fatter tails than a standard normal density. $\diamondsuit$

[4] For details, see Mittelhammer (1996, p. 342).

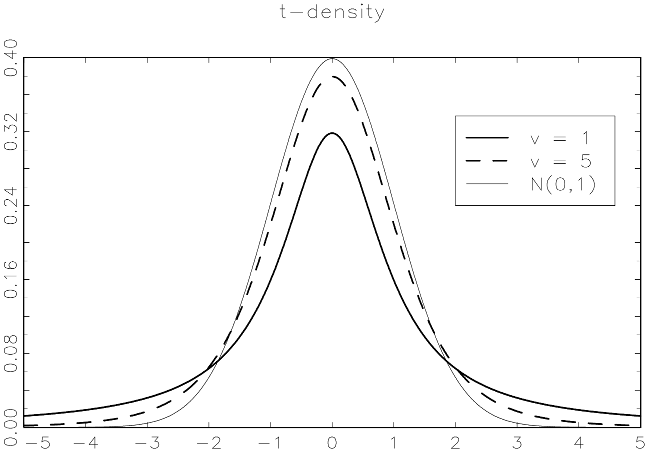


Fig. 36.

Theorem 6.9 defining the t-distribution facilitates the derivation of the pdf of the so called t-statistic when the random sample is from a normal distribution, as stated in the following theorem.

THEOREM 6.10 *Let $(X_1, ..., X_n)$ be a random sample from a $N(\mu, \sigma^2)$ - distribution. Then the t-statistic given by*

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

$$\text{where} \qquad \bar{X}_n = \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i, \qquad \hat{\sigma}_n^2 = \tfrac{n}{n-1} S_n^2, \qquad S_n^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

follows a t-distribution with $v = n - 1$ degrees of freedom.

PROOF: Rewrite the t-statistic as

$$T = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}\sqrt{\frac{\hat{\sigma}_n^2}{\sigma^2}}} = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}\sqrt{\frac{nS_n^2}{\sigma^2}/(n-1)}},$$

where $\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0,1)$, and $\frac{nS_n^2}{\sigma^2} \sim \chi^2_{n-1}$, and where $\bar{X}_n$ and S_n^2 are independent (see Theorem 6.5). Thus Theorem 6.9 applies, so that $T \sim t_{n-1}$. $\square$

An important property of the t_v-distribution is that it converges to a Normal distribution as $v \to \infty$, as stated in the following theorem (see also Fig. 36).

THEOREM 6.11 *Let $Z \sim N(0,1)$, let $Y \sim \chi_v^2$, and let Z and Y be independent so that*

$$T_v = \frac{Z}{\sqrt{Y/v}} \sim t_v. \qquad \text{Then} \qquad T_v \xrightarrow{d} N(0,1), \quad \text{for} \quad v \to \infty.$$

PROOF: Since $Y \sim \chi_v^2$, we have $\mathrm{E}Y = v$ and $\mathrm{var}(Y) = 2v$, so that

$$\mathrm{E}(\tfrac{Y}{v}) = 1, \qquad \mathrm{var}(\tfrac{Y}{v}) = \tfrac{2}{v} \to 0, \quad \text{for} \quad v \to \infty.$$

Hence

$$\tfrac{Y}{v} \xrightarrow{m} 1 \qquad \Rightarrow \qquad \mathrm{plim}(\tfrac{Y}{v}) = 1.$$

Also note that since $Z \sim \mathrm{N}(0,1)\ \forall v$, it follows that $Z \xrightarrow{d} \mathrm{N}(0,1)$. Then by Slutsky's theorem,

$$T_v = \frac{Z}{\sqrt{Y/v}} \xrightarrow{d} \frac{Z}{1} \sim \mathrm{N}(0,1). \qquad \square$$

F-Density

The F-density arises as the density of a ratio of to independent χ^2 random variables.

THEOREM 6.12 *Let $Y_1 \sim \chi^2_{v_1}$, let $Y_2 \sim \chi^2_{v_2}$, and let Y_1 and Y_2 be independent. Then*

$$F = \frac{Y_1/v_1}{Y_2/v_2}$$

has the F-density with v_1 numerator and v_2 denominator degrees of freedom, defined as

$$m(f; v_1, v_2) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{v_1/2} f^{(v_1/2)-1} \left(1 + \frac{v_1}{v_2} f\right)^{-(1/2)(v_1+v_2)} \mathbb{I}_{(0,\infty)}(f).$$

PROOF: The proof based upon the multivariate change-of-variable technique (Theorem 6.8) is straightforward (see Mittelhammer, 1996, p. 345). □

REMARK : The F-distribution with a pdf as given in Theorem 6.12 has the following moments

$$\begin{aligned}
\mu &= \frac{v_2}{v_2 - 2} \quad \text{for} \quad v_2 > 2 \\
\sigma^2 &= \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)} \quad \text{for} \quad v_2 > 4 \\
\mu_3 &= \frac{v_2^3 8v_1(v_1 + v_2 - 2)(2v_1 + v_2 - 2)}{v_1^3(v_2 - 2)^3(v_2 - 4)(v_2 - 6)} > 0 \quad \text{for} \quad v > 6,
\end{aligned}$$

where the degrees-of-freedom parameters are $v_1 > 0$ and $v_2 > 0$. The MGF does not exist.

As Fig. 37 shows, the F-density is skewed to the right. $\diamondsuit$

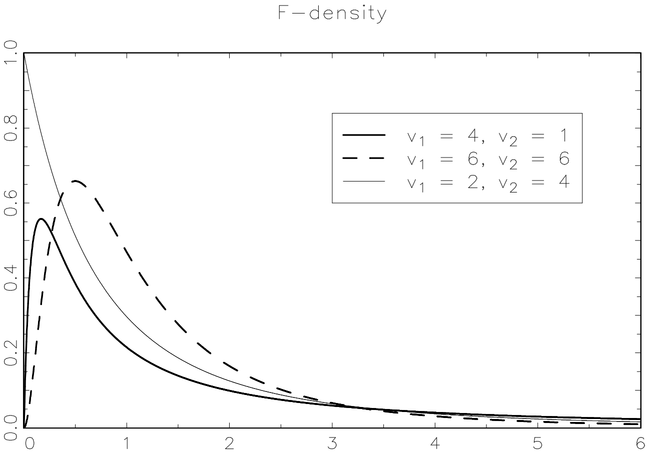


Fig. 37.