Lecture 4. Sampling [1]

Econometrics is concerned with inference of parameters based on data. Data typically come as a collection of observations of a vector of random variables for multiple individuals and/or multiple time periods. Such a collection of observations is called a sample. How the sample is collected affects the inference method to be used as well as the properties of the estimators of the parameters.

A basic type of sampling is random sampling:

Definition (Random Sample)**.** *The collection $\{X_1, \ldots, X_n\}$ is a random sample from the population F if they are mutually independent with identitcal marginal distributions F. In this case, we also say that $\{X_1, \ldots, X_n\}$ are i.i.d. (independent and identically distributed).*

Here, n is the sample size, or the number of observations.

The distribution F is called the population distribution, or just population. This is an infinite population and is a mathematical abstraction. We use X without the subscript to denote a generic random variable with distribution F.

X_i may be a random vector.

There are other forms of samplings as well, including

1. Stratified sampling
2. Clustered sampling
3. panel data
4. time series data
5. spatial data

All of these introduce some sort of dependence across observations, which complicates matters.

We could also consider i.n.i.d. (independent, not necessarily identically distributed) to allow different X_i to be drawn from a different F, but that typically does not provide much meaningful insights.

A useful fact that can be derived from the results in the previous lecture is the following: If $\{X_i : i = 1, \ldots, n\}$ are i.i.d. and if $Y_i = g(X_i)$ for a determinstic function g for all i, then $\{Y_i : i = 1, \ldots, n\}$ are i.i.d. as well.

[1] This lecture note is largely adapted from Professor Bruce Hansen's handwritten notes, however, all errors are mine.

1 Using a Random Sample

We use a random sample to build statistics:

Definition (Statistics)**.** *A statistic is a function of the sample* $\{X_i : 1, \ldots, n\}$*.*

For example, the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ is a statistic.

We use statistics to build estimates for parameters:

Definition (Parameter)**.** *A parameter is any function of the population.*

For example, the population mean $\mu = E(X)$.

Definition (Estimator)**.** *An estimator, or a point estimator* $\hat{\theta}$ *for a parameter* θ *is a statistic intended as a guess about* θ*.*

For example, we typically use $\bar{X}_n$ as an estimator of μ. This is the most natural analog estimator since $\bar{X}_n$ is the average of the sample, while $E(X)$ is the average of the population.

More generally, if $\theta = E(g(X))$ for some determinstic function g, we can use the sample analog estimator:

$$\hat{\theta} = n^{-1} \sum_{i=1}^n g(X_i).$$

An even borader class of parameter is of the form:

$$\beta = h(E(g(X)),$$

where h is another determinstic function. For example, the population variance of X is

$$\sigma^2 = E(X - E(X))^2 = E(X^2) - (E(X))^2.$$

Often, we can use the plug-in estimator for β:

$$\hat{\beta} = h(\hat{\theta}) = h\left(n^{-1} \sum_{i=1}^n g(X_i)\right). \tag{1}$$

Since an estimate is a statistics, which is a function of the random variables $X_1, \ldots, X_n$, an estimate is random variables. We want to understand its distribution. There are several approaches

1. Compute exact bias and variaton.
2. Compute exact distribution under normality.
3. Find asymptotic distribution as $n \to \infty$, typically using a law of large numbers and a central limit theorem.

4. Asymptotic expansion, which uses asymptotic tools, but gives a more accurate approximation of the exact distribution.

5. Bootstrap simulation.

2 Exact Bias and Variance Calculation

2.1 Bias

First, let's focus on $\bar{X}_n$, and see if we can find its Exact bias. Let's calculate its expectation first

$$E(\bar{X}_n) = n^{-1}\sum_{i=1}^{n} E(X_i) = n^{-1}\sum_{i=1}^{n} E(X) = E(X).$$

Thus, as an estimator for $E(X)$, the sample average $\bar{X}_n$ has bias $E(\bar{X}_n) - E(X) = 0$, that is, it is unbiased.

Definition 1 (Unbiasedness)**.** *An estimator $\hat{\theta}_n$ of a parameter θ is unbiased if $E(\hat{\theta}_n) = \theta$ regardless of the population F.*

Example. *Suppose that we use $\tilde{\mu}_n = \bar{X}_n/2$ as an estimator of $\mu = E(X)$. Then this is not an unbiased estimator because it is only when $\mu = 0$, do we have $E(\tilde{\theta}_n) = 0/2 = 0 = \mu$. With any other value μ, $E(\tilde{\theta}_n) \neq \mu$.*

Example. *Suppose we use $\bar{X}_n$ as an estimator of $m = median(X)$. This is also not an unbiased estimator because it is only when F is a symmetric distribution, do we have $E[\bar{X}_n] = \mu = m$. When F is not a symmetric distribution, $E[\bar{X}_n] = \mu \neq m$.*

Affine transformations preserves unbiasedness: If $\hat{\theta}$ is an unbiased estimator of θ, then $\hat{\beta} = a\hat{\theta} + b$ also is an unbiased estimator of $\beta = a\theta + b$.

Nonlinear transformations generally do not preserve unbiasedness: If $\hat{\theta}$ is an unbiased estimator of θ, then in general $\hat{\beta} = h(\hat{\theta})$ is NOT an unbiased estimator of $\beta = h(\theta)$. In fact, if h is concave (e.g. $h(\theta) = \log(\theta)$), we know that $\hat{\beta}$ is biased downward as an estimator of β:

$$E(\hat{\beta}) = E[h(\hat{\theta})] < h(\theta) = \beta.$$

If h is convex ($h(\theta) = \theta^2$), we know that $\hat{\beta}$ is biased upward as an estimator of β:

$$E(\hat{\beta}) = E[h(\hat{\theta})] > h(\theta) = \beta.$$

Example (Truncated Estimator)**.** *Suppose that prior knowledge of F tells us that $\mu = E(X) \geq 0$. A reasonable estimator for μ would be*

$$\hat{\mu} = \begin{cases} \bar{X}_n & \text{if } \bar{X}_n \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Though this estimator makes sense and has some other good properties, it is biased upward:

$$E(\hat{\mu}) > \mu,$$

because $h(\theta) = \theta \times 1(\theta \geq 0)$ is convex.

2.2 Variance

For the estimator $\bar{X}_n$ of μ, we can calculate its exact variance as well:

$$\begin{aligned} Var(\bar{X}_n) &= E(\bar{X}_n - E(\bar{X}_n))^2 \\ &= E\left(n^{-1}\sum_{i=1}^{n}(X_i - \mu)\right)^2 \\ &= n^{-2}E\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2 \\ &= n^{-2}\sum_{i=1}^{n}E(X_i - \mu)^2 && \text{by def of random sample} \\ &= n^{-2}\sum \sigma_X^2 \\ &= \sigma_X^2/n. \end{aligned}$$

Note that the variance drops with n. Thus, $\bar{X}_n$ is a better estimator that $(2/n)\sum_{i=1}^{n/2} X_i$ even though the latter is also an unbiased estimator of μ.

For affine transformations $\hat{\beta} = a\bar{X}_n + b$ as an estimator of $\beta = a\mu + b$, we have

$$Var(\hat{\beta}) = a^2 Var(\bar{X}_n).$$

For nonlinear transformations $h(\bar{X}_n)$ we typically cannot exactly calculate the variance.

2.3 Standard Error

Finding the formula of the exact variance is only half way through understanding the accuracy of an estimator. This is because the exact variance depends on the unknown population F, and thus is unknown. Based on a random sample, we can estimate the variance, and compute the standard error. We can then assess the estimation accuracy of an estimator based on the standard error.

Definition 2 (Standard Error)**.** *A standard error $s(\hat{\theta}_n)$ for an estimator $\hat{\theta}$ of a parameter θ, is the square root of an estimator $\hat{V}$ of $V = Var(\hat{\theta})$.*

For example, for the estimator $\bar{X}_n$ of the population expectation μ, the exact variance is σ_X^2/n. One possible estimator of $\sigma_X^2 = E(X-\mu)^2$ is the direct sample analogue of σ_X^2:

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

If we use this as the estimator of σ_X^2, then the implied standard error of the estimator $\bar{X}_n$ of μ is

$$s(\bar{X}_n) = \hat{\sigma}_n / n^{1/2}$$

We can evaluate whether $\hat{\sigma}_n^2$ is a good estimator of σ_X^2 by calculating its exact bias:

$$\begin{aligned} E(\hat{\sigma}_n^2) - \sigma_X^2 &= E(X_i-\mu)^2 - E(\bar{X}_n - \mu)^2 - \sigma_X^2 \\ &= \sigma_X^2 - n^{-1}\sigma_X^2 - \sigma_X^2 \\ &= \frac{n-1}{n}\sigma_X^2 - \sigma_X^2 \\ &= -n^{-1}\sigma_X^2. \end{aligned} \tag{3}$$

The direct sample analogue estimator is biased downward (and toward zero).

The bias calculation motivates the unbiased estimator for σ_X^2:

$$s^2 = \frac{n}{n-1}\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2. \tag{4}$$

The above calculation indicates that $E(s^2) = \sigma_X^2$. If we use this as the estimator of σ_X^2, then the implied standard error of the estimator $\bar{X}_n$ of μ is

$$s(\bar{X}_n) = s/\sqrt{n}. \tag{5}$$

This is the typical choice in statistics and econometrics due to the unbiasedness of s^2 as an estimator of σ_X^2. On the other hand, the difference between $s/\sqrt{n}$ and $\hat{\sigma}_n/\sqrt{n}$ is small when n is large, in which case both are reasonable choices.

3 Assuming Normality of F

Without much assumption on the population F, we cannot say much about the distribution of $\bar{X}_n$ (or estimators of other parameters) beyond its bias and variance. Now, we assume that F is a normal distribution and see how much more we can get.

3.1 Review of Normal Distribution

Before we go on, let's first review the normal distributions. A normal distribution is characterized by its mean μ and variance σ^2, and often denoted $N(\mu, \sigma^2)$. A normal distribution with zero mean and unit variance is called the standard normal distribution. The PDF of a standard normal distribution is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \qquad x \in R.$$

It's usual notation is $\phi(x)$. Section 3.3 of the textbook has a very nice proof for this to integrate up to one: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx = 1$. Now we demonstrate that it has zero mean and unit variance:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2)dx \\
&= \int_{-\infty}^{0} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2)dx + \int_{0}^{\infty} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2)dx \\
&= \int_{0}^{\infty} \frac{-y}{\sqrt{2\pi}} \exp(-y^2/2)dy + \int_{0}^{\infty} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2)dx \qquad \text{(use the transformation:} y = -x) \\
&= 0.
\end{aligned}$$

Note that this is the trick you could use for integrating any odd function. Similarly, you can find that, for any odd number m, we have

$$E(X^m) = 0.$$

Now, the variance:

$$\begin{aligned}
Var(X) = E(X^2) &= \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp(-x^2/2)dx \\
&= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2)d\frac{x^2}{2} \\
&= -\int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} d\exp(-x^2/2)
\end{aligned}$$

$$= 0 + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx \qquad \text{(integration by part)}$$

$$= 1. \qquad \left(\text{since } \phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \text{ is a PDF}\right)$$

All moments of the standard normal distribution exist. Moreoever, the moment generating function exists. These are because the tail of the PDF decreases very very fast.

The CDF is written as

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt.$$

There is no closed form expression for this CDF.

A normal distribution $N(\mu, \sigma^2)$ is obtained when a linear transformation is applied on a standard normal random variable: If $Z \sim N(0, 1)$, then $X = \sigma Z + \mu \sim N(\mu, \sigma^2)$. The PDF of X is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x-\mu)^2/(2\sigma^2)),$$

which one can obtain using the transformation formula. Using the relationship $X = \sigma Z + \mu$, we see that

$$E(X) = \sigma E(Z) + \mu = \mu$$
$$Var(X) = \sigma^2 Var(Z) = \sigma^2.$$

Conversely, if $X \sim N(\mu, \sigma^2)$, then we can define the standardized version of it: $Z = (X - \mu)/\sigma$, and we should have $Z \sim N(0, 1)$.

3.2 Random Samples from $N(\mu, \sigma^2)$

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from $N(\mu, \sigma^2)$. Then they have the joint PDF:

$$\begin{aligned} f(x_1, \ldots, x_n) &= f(x_1)f(x_2)\cdots f(x_n) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \Pi_{i=1}^n \exp(-(x_i-\mu)^2/(2\sigma^2)) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right) \end{aligned}$$

Like in the previous lecture, we use the notation $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$, and $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$, and we can write

$$f(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu})'(\mathbf{x}-\boldsymbol{\mu})\right)$$

This is a special case of a $n-$variate normal distribution.

Definition 3 (Multi-variate Normal Distribution)**.** *The $n-$vector $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ with a $n \times 1$ mean vector $\boldsymbol{\mu}$ and a $n \times n$ variance-covariance matrix Σ if it has PDF:*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(det(\Sigma))^{1/2}} \exp(-(1/2)(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})).$$

We emphasized in the previous lecture that zero correlation between two random variables does not imply independence between them. But if we have the additional knowledge that the two random variables (say, X, Y) are jointly normal, that is, $(X,Y)'$ is a bivariate normal random vector, then zero correlation between X and Y implies independence. This actually applies a bit more generally to two normal random vectors $\mathbf{X}$ and $\mathbf{Y}$ are themselves normal random vectors, in which case, we have the following theorem:

Theorem 4. *If $(\mathbf{X}', \mathbf{Y}')' \sim N(\mu, \Sigma)$, and $Cov(\mathbf{X}, \mathbf{Y}) := E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)'] = \mathbf{0}$, then $\mathbf{X}$ and $\mathbf{Y}$ are independent.*

Proof. The fact that $Cov(\mathbf{X}, \mathbf{Y}) = 0$ implies that Σ is of the following block diagonal structre:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22}, \end{pmatrix}$$

where $\Sigma_{11} = Var(\mathbf{X})$ and $\Sigma_{22} = Var(\mathbf{Y})$. Thus, the joint PDF of $(\mathbf{X}', \mathbf{Y}')'$ is

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= \frac{1}{(2\pi)^{n/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\begin{pmatrix} \mathbf{x} - \mu_{\mathbf{X}} \\ \mathbf{y} - \mu_{\mathbf{Y}} \end{pmatrix}' \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22}, \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x} - \mu_{\mathbf{X}} \\ \mathbf{y} - \mu_{\mathbf{Y}} \end{pmatrix}\right) \\ &= \frac{1}{(2\pi)^{n_X/2}\det(\Sigma_{11})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_{\mathbf{X}})'\Sigma_{11}^{-1}(\mathbf{x} - \mu_{\mathbf{X}})\right) \times \\ &\quad \frac{1}{(2\pi)^{n_Y/2}\det(\Sigma_{22})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu_{\mathbf{Y}})'\Sigma_{22}^{-1}(\mathbf{y} - \mu_{\mathbf{Y}})\right). \end{aligned}$$

That is, $f(\mathbf{x}, \mathbf{y})$ factors into the products of a function of only $\mathbf{x}$ and a function of only $\mathbf{y}$, which implies that $\mathbf{X}$ and $\mathbf{Y}$ are independent. $\square$

Moment Generating Function. For a multivariate normal distribution, we can find its momeng generating function:

$$\begin{aligned} &M_{N(\boldsymbol{\mu},\Sigma)}(t) \\ &= E(\exp(t'\mathbf{X}) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}det(\Sigma)^{1/2}} \exp\left(t'\mathbf{x} - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) dx_1 \ldots dx_n \end{aligned}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2} det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \Sigma t)'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu} - \Sigma t) + \frac{1}{2}t'\Sigma t + \mu' t\right) dx_1 \ldots dx_n$$
$$= \exp(\frac{1}{2}t'\Sigma t + \mu' t) \times$$
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2} det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \Sigma t)'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu} - \Sigma t)\right) dx_1 \ldots dx_n$$
$$= \exp(\frac{1}{2}t'\Sigma t + \mu' t).$$

When MGF of a random vector exists, it uniquely determines the distribution of the random vector. Thus, the MGF function for the normal distribution gives us another way to show that <u>linear transformations of normal random vectors are also random vectors</u>. For example, let $\mathbf{Y} = B\mathbf{X} + \mathbf{a}$ for an $m \times n$ matrix B and an $m \times 1$ vector $\mathbf{a}$. Then the MGF of $\mathbf{Y}$ is

$$\begin{aligned}
M_{\mathbf{Y}}(z) &= E(\exp(z'\mathbf{Y})) \\
&= E(\exp(z'B\mathbf{X} + z'\mathbf{a}) \\
&= \exp(z'\mathbf{a})E(\exp((B'z)'\mathbf{X})) \\
&= \exp(z'\mathbf{a})M_{\mathbf{X}}(B'z) \\
&= \exp(z'\mathbf{a}) \exp\left(\frac{1}{2}z'B\Sigma B'z + \mu' B'z\right) \\
&= \exp\left(\frac{1}{2}z'B\Sigma B'z + (\mathbf{a} + B\mu)'z\right).
\end{aligned}$$

That is the MGF of $N(B\mu + \mathbf{a}, B\Sigma B')$. Therefore, $\mathbf{Y} \sim N(B\mu + \mathbf{a}, B\Sigma B')$.

Example. *If* $\{X_i\}_{i=1}^n$ *are i.i.d.* $N(\mu, \sigma^2)$. *Then* $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n)$.

Distribution of Variance Estimator. When the random sample $\{X_i\}_{i=1}^n$ is drawn from the normal distribution $N(\mu, \sigma^2)$, we can figure out the distribution of the variance estimator: $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

In fact, $(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$, where $\chi^2(k)$ denote the distribution of the sum of k mutually independent $\chi^2(1)$ random variables, or equivalently, the distribution of the sum of the squares of k mutually independent standard normal random variables. And it is called the chi-squared distribution with k-degrees of freedom.

Proof. We can write s^2 in matrix notation:

$$\begin{aligned}
(n-1)s^2 &= \left(\left(\sum_{i=1}(X_i - \mu)^2\right) - n(\bar{X}_n - \mu)^2\right) \\
&= \left((\mathbf{X} - \boldsymbol{\mu})'(\mathbf{X} - \boldsymbol{\mu}) - n^{-1}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{1}_n\mathbf{1}_n'(\mathbf{X} - \boldsymbol{\mu})\right)
\end{aligned}$$

$$= (\mathbf{X} - \boldsymbol{\mu})' \left(I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'\right)(\mathbf{X} - \boldsymbol{\mu})$$
$$= (\mathbf{X} - \boldsymbol{\mu})' \left(I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'\right)(\mathbf{X} - \boldsymbol{\mu}),$$

where $\mathbf{1}_n$ is the $n \times 1$ vectors with all elements being 1, and $\boldsymbol{\mu} = \mu\mathbf{1}n$.

The matrix $I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'$ is idempotent, that is it multiplied itself equals itself. Idemponent matrices have a nice property: its eigenvalues are either zero or one. That means, it admits an eigenvalue-decomposition of the form:

$$I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n' = \begin{pmatrix} Q_1 & Q_0 \end{pmatrix} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1 & Q_0 \end{pmatrix}', \tag{6}$$

where r is the rank of the matrix $I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'$, Q_1 is a $n \times r$ matrix the columns of which are the orthonormalized eigenvalues of $I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'$ corresponding to the eigenvalue 1, and Q_0 is the $n \times (n-r)$ matrix the columns of which are the orthonormalized eigenvectors corresponding to the eigenvalue 0. By definition, $Q_1'Q_1 = I_r$.

Simplifying (6), we have

$$I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n' = Q_1Q_1'.$$

Thus, $(n-1)s^2/\sigma^2 = (\mathbf{X} - \boldsymbol{\mu})'Q_1Q_1'(\mathbf{X} - \boldsymbol{\mu})/\sigma^2 = \mathbf{Z}'\mathbf{Z}$ where $\mathbf{Z} = Q_1'(\mathbf{X} - \boldsymbol{\mu})/\sigma$. Now, observe that $\mathbf{Z}$ is a linear transformation of $(\mathbf{X} - \boldsymbol{\mu})/\sigma$, and $(\mathbf{X} - \boldsymbol{\mu})/\sigma \sim N(\mathbf{0}, I_n)$. Thus

$$\mathbf{Z} \sim N(0, Q_1'I_nQ_1) = N(0, Q_1'Q_1) = N(0, I_r).$$

Thus

$$(n-1)s^2/\sigma^2 = \sum_{j=1}^{r} Z_j^2,$$

where $\{Z_j\}_{j=1}^r$ are the elements of $\mathbf{Z}$. That is, $(n-1)s^2$ is the sum of the squares of r independent standard normal random variables. Thus,

$$(n-1)s^2/\sigma^2 \sim \chi^2(r).$$

It is still left to determine r. For idempotent matrices, the rank equals the trace. Thus

$$r = trace(I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n') = n - 1.$$

Therefore,

$$(n-1)s^2/\sigma^2 \sim \chi^2(n-1).$$

□

Distribution of t-ratio. An important statistic is the t ratio

$$t = \frac{\sqrt{n}(\bar{X}_n - \mu)}{s}.$$

Can we figure out the distribution of t? In fact, we already know the distribution of its two components:

$$\begin{aligned} \sqrt{n}(\bar{X}_n - \mu)/\sigma &\sim N(0,1). \\ (n-1)s^2/\sigma^2 &\sim \chi^2(n-1). \end{aligned}$$

We do not yet know the relationship between the two components. Next we show that the two components are independent. Once that is shown, we will have a full characterization of the distribution of t:

$$t \sim \frac{N(0,1)}{\sqrt{\chi^2(n-1)/n-1}},$$

where the $N(0,1)$ random variable and the $\chi^2(n-1)$ random variable are independent. Such a distribution is called a student-t distribution with $n-1$ degrees of freedom. This distribution is discovered by Gossett (1908).

Now we show that $\bar{X}_n - \mu$ and s^2 are independent. Observe that s^2 is a function of the random vector $(\mathbf{X} - \bar{X}_n \mathbf{1}_n)$. Thus, it suffices to show that $\bar{X}_n - \mu$ is independent of $(\mathbf{X} - \bar{X}_n \mathbf{1}_n)$. Moreover, $\bar{X}_n - \mu$ and $(\mathbf{X} - \bar{X}_n \mathbf{1}_n)$ are jointly normal because together they form a random vector that is a linear function of the normal random vector $\mathbf{X}$. Therefore, by the theorem above, it suffices to show that the two vectors are uncorrelated.

It is sufficient to look at the covariance between $\bar{X}_n - \mu$ and $X_i - \bar{X}_n$ for each i. Observe that

$$\begin{aligned} Cov(\bar{X}_n - \mu, X_i - \bar{X}_n) &= E[(\bar{X}_n - \mu)(X_i - \bar{X}_n)] \\ &= E[(X_i - \mu + \mu - \bar{X}_n)(\bar{X}_n - \mu)] \\ &= E[(X_i - \mu)(\bar{X}_n - \mu)] - Var(\bar{X}_n) \\ &= n^{-1} \sum_{j=1}^{n} E(X_i - \mu)(X_j - \mu) - \sigma^2/n \\ &= n^{-1} E((X_i - \mu)^2) - \sigma^2/n \\ &= \sigma^2/n - \sigma^2/n = 0. \end{aligned}$$

Therefore, $Cov(\bar{X}_n - \mu, (\mathbf{X} - \bar{X}_n \mathbf{1}_n)) = \mathbf{0}$, proving that $\bar{X}_n - \mu$ is independent of $(\mathbf{X} - \bar{X}_n \mathbf{1}_n)$, which in turn proves that $\bar{X}_n - \mu$ and s^2 are independent.

4 Problems

Most of the problems assume a random sample $\{X_1, X_2, \ldots, X_n\}$ from a common distribution F with density f such that $E(X) = \mu$ and $Var(X) = \sigma^2$ for a generic random varaible $X \sim F$. The sample mean and variances are denoted $\bar{X}_n$ and $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, with the bias-corrected variance $s_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$.

1. Suppose that another observation X_{n+1} becomes available. Show that

   (a) $\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1)$

   (b) $s_{n+1}^2 = \left((n-1)s_n^2 + (n/(n+1))(X_{n+1} - \bar{X}_n)^2\right)/n$

2. For some integer k, set $\mu_k = E(X^k)$. Construct an unbiased estimator $\hat{\mu}_k$ for μ_k, and show its unbiasedness.

3. Consider the central moment $m_k = E((X-\mu)^k)$. Construct an estimator $\hat{m}_k$ for m_k without assuming a known μ. In general, do you expect $\hat{m}_k$ to be biased or unbiased?

4. Calculate the variance of $\hat{\mu}_k$ that you proposed abve, and call it $Var(\hat{\mu}_k)$.

5. Propose an estimator of $Var(\hat{\mu}_k)$. Does an unbiased version exist?

6. Show that $E(s_n) \leq \sigma$. (Hint: Use Jensen's inequality, CB Theorem 4.7.7).

7. Calculate $E((\bar{X}_n - \mu)^3)$, the skewness of $\bar{X}_n$. Under what condition is it zero?

8. Show algebraically that $\hat{\sigma}^2 = n^{-1} \sum_{i-1}^{n} (X_i - \mu)^2 - (\bar{X}_n - \mu)^2$.

9. Find the covariance of $\hat{\sigma}^2$ and $\bar{X}_n$. Under what condition is this zero. (Hint: Use the form obtained in #9. This exercise shows that the zero correlation between the numerator and the denominator of the $t-$ ratio does not always hold when the random sample is not from a normal distribution.)

10. Suppose that X_i are i.n.i.d. (Independent but not necessarily identically distributed) with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$.

    (a) Find $E(\bar{X}_n)$.

    (b) Find $Var(\bar{X}_n)$.

11. Assume $X \sim N(0,1)$ which has density $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

    (a) Use integration by parts to show that $E(X^4) = 3$.

    (b) Use the moment generating function of X to verify that $E(X^2) = 1$ and $E(X^4) = 3$.

(c) Compare the well-known $P(|X| > 1.96) = 0.05$ with the bound for $P(|X| > 1.96)$ from Chebychev's Inequality.

12. Show that if $Q \sim \chi^2_r$, then $E(Q) = r$ and $Var(Q) = 2r$. (Hint: use the representation: $Q = \sum_{i=1}^n X_i^2$ with X_i being i.i.d. $N(0, 1)$.

13. Suppose that $X_i : i = 1, \ldots, n$ are independent $N(\mu_i, \sigma_i^2)$. Find the distribution of the weighted sum $\sum_{i=1}^n w_i X_i$.

14. Suppose that $X_i \sim N(\mu_X, \sigma_X^2) : i = 1, \ldots, n_1$ and $Y_i \sim N(\mu_Y, \sigma_Y^2)$, $i = 1, \ldots, n_2$ are mutually independent. Set $\bar{X}_n = n_1^{-1} \sum_{i=1}^{n_1} X_i$ and $\bar{Y}_n = n_2^{-1} \sum_{i=1}^{n_2} Y_i$.

    (a) Find $E(\bar{X}_n - \bar{Y}_n)$.

    (b) Find $Var(\bar{X}_n - \bar{Y}_n)$.

    (c) Find the distribution of $\bar{X}_n - \bar{Y}_n$.