5.1 - Distribution of Sample Mean Vector

5.1 - Distribution of Sample Mean Vector

As noted previously $\bar{\mathbf{x}}$ is a function of random data, and hence $\bar{\mathbf{x}}$ is also a random vector with a mean, a variance-covariance matrix, and a distribution. We have already seen that the mean of the sample mean vector is equal to the population mean vector $\boldsymbol{\mu}$.

Variance

Before considering the sample variance-covariance matrix for the mean vector $\bar{\mathbf{x}}$, let us revisit the univariate setting.

Univariate Setting

You should recall from introductory statistics that the population variance of the sample mean, generated from independent samples of size *n,* is equal to the population variance, σ_2 divided by *n*.

$$\mathrm{var}(\bar{x}) = \frac{\sigma^2}{n}$$

This, of course, is a function of the unknown population variance σ_2. We can estimate this by simply substituting s^2 in the sample variance σ_2 yielding our estimate for the variance of the population mean as shown below:

$$\widehat{\mathrm{var}}(\bar{x}) = \frac{s^2}{n}$$

If we were to take the square root of this quantity we would obtain the standard error of the mean. The standard error of the mean is a measure of the uncertainty of our estimate of the population mean. If the standard error is large, then we are *less confident* of our estimate of the mean. Conversely, if the standard error is small, then we are *more confident* in our estimate. What is meant by large or small depends on the application at hand. But in any case, because the standard error is a decreasing function of sample size, the larger our sample the more confident we can be of our estimate of the population mean.

Multivariate Setting

The population variance-covariance matrix replaces the variance of the $\bar{x}$'s generated from independent samples of size *n*, taking a form similar to the univariate setting as shown below:

$$\mathrm{var}(\bar{\mathbf{x}}) = \frac{1}{n}\Sigma$$

Again, this is a function of the unknown population variance-covariance matrix Σ. An estimate of the variance-covariance matrix of $\bar{\mathbf{x}}$ can be obtained by substituting the sample variance-covariance matrix **S** for the population variance-covariance matrix Σ, yielding the estimate as shown below:

$$\widehat{\mathrm{var}}(\bar{\mathbf{x}}) = \frac{1}{n}\mathbf{S}$$

Distribution

Let's consider the distribution of the sample mean vector, first looking at the univariate setting and comparing this to the multivariate setting.

Univariate Setting

Here we are going to make the additional assumption that $X_1, X_{2,}\ldots X_n$ are independently sampled from a normal distribution with mean μ and variance σ_2. In this case, $\bar{x}$ is normally distributed as

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Multivariate Setting

Similarly, for the multivariate setting, we are going to assume that the data vectors $X_1, X_{2,}\ldots X_n$ are independently sampled from a multivariate normal distribution with mean vector μ and variance-covariance matrix Σ. Then, in this case, the sample mean vector, $\bar{\mathbf{x}}$, is distributed as multivariate normal with mean vector μ and variance-covariance matrix $\frac{1}{n}\Sigma$, the variance-covariance matrix for $\bar{\mathbf{x}}$. In statistical notation we write:

$$\bar{\mathbf{x}} \sim N\left(\mu, \frac{1}{n}\Sigma\right)$$

Law of Large Numbers

At this point, we will drop the assumption that the individual observations are sampled from a normal distribution and look at the laws of large numbers. These will hold regardless of the distribution of the individual observations.

Univariate Setting

In the univariate setting, we see that if the data are independently sampled, then the sample mean, $\bar{x}$, is going to converge (in probability) to the population mean μ. What does this mean exactly? It means that as the sample size gets larger and larger the sample mean will tend to approach the true value for a population μ.

Multivariate Setting

A similar result is involved in the multivariate setting, the sample mean vector, $\bar{\mathbf{x}}$, will also converge (in probability) to the mean vector μ As our sample size gets larger and larger, each of the individual components of that vector, $\bar{x}_j$, will converge to the corresponding mean, μ_j.

$$\bar{x}_j \xrightarrow{p} \mu_j$$

Central Limit Theorem

Just as in the univariate setting we also have a multivariate Central Limit Theorem. But first, let's review the univariate Central Limit Theorem.

Univariate Setting

If all of our individual observations, $X_1, X_2, \ldots X_n$, are independently sampled from a population with mean μ and variance σ_2, then, the sample mean, $\bar{x}$, is *approximately* normally distributed with mean μ and variance σ_2.

Note! In the distribution property described above, normality was a requirement. Under normality, even for small samples, the data are normally distributed. The Central Limit Theorem is a more general result that holds regardless of the distribution of the original data. The significance of CLT lies in the fact that the sample mean is approximately normally distributed for large samples whatever the distribution of the individual observations.

Multivariate Setting

A similar result is available in the multivariate setting. If our data vectors $X_1, X_2, \ldots X_n$, are independently sampled from a population with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, then the sample mean vector, $\bar{\mathbf{x}}$, is going to be *approximately* normally distributed with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\frac{1}{n}\boldsymbol{\Sigma}$.

This Central Limit Theorem is a key result that we will take advantage of later on in this course when we talk about hypothesis tests for individual mean vectors or collections of mean vectors under different treatment regimens.

Legend	
[1]	Link
↕	Has Tooltip/Popover
[ ]	Toggleable Visibility

Source: https://www.google.com/

Links: