L E C T U R E 5

§6.2. RANDOM SAMPLES AND THE EXPECTED VALUES OF THE SAMPLE MEAN AND COVARIANCE MATRIX

In order to study the sampling variability of statistics like $\overline{x}$ and S_n with the ultimate aim of making inferences, we need to make assumptions about the variables whose observed values constitute the data matrix $\mathbf{X}$.

Suppose, then, that the data have not yet been observed, but we intend to collect n sets of measurements on m variables. Before the measurements are made, their values cannot, in general, be predicted exactly. Consequently, we treat them as random variables. In this context, let the (j,k)-th entry in the data matrix be the random variable X_{jk}. Each set of measurements $\mathbf{X_j}$ on m variables is a random vector, and we have the random matrix

$$\mathbf{X} = \begin{bmatrix} X_{11}\ X_{12}\ \ldots\ X_{1m} \\ X_{21}\ X_{22}\ \ldots\ X_{2m} \\ \ldots\ldots\ \ldots\ldots\ \ldots\ \ldots\ldots \\ X_{n1}\ X_{n2}\ \ldots\ X_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \\ \vdots \\ \mathbf{X}_n' \end{bmatrix}. \tag{6.1}$$

A random sample can now be defined.

If the row vectors $\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_n'$ in (6.1) represent independent observations from a common joint distribution with density function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_m)$, then $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are said to form a random sample from $f(\mathbf{x})$. Mathematically, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ form a random sample if their joint density function is given by the product $f(\mathbf{x}_1)f(\mathbf{x}_2)\cdots f(\mathbf{x}_n)$, where $f(\mathbf{x}_j) = f(x_{j1}, x_{j2}, \ldots, x_{jm})$ is the density function for the jth row vector.

Two points connected with the definition of random sample merit special attention:

1. The measurements of the m variables in a single trial, such as $\mathbf{X}_j' = [X_{j1}, X_{j2}, \ldots, X_{jm}]$, will usually be correlated. Indeed, we expect this to be the case. The measurements from different trials must, however, be independent.

2. The independence of measurements from trial to trial may not hold when the variables are likely to drift over time, as with sets of m stock prices or m economic indicators. Violations of the tentative assumption of independence can have a serious impact on the

quality of statistical inferences.

As we have argued heuristically, the notion of statistical independence has important implications for measuring distance. Euclidean distance appears appropriate if the components of a vector are independent and have the same variances. Suppose we consider the location of the kth column $\mathbf{Y}_k' = [X_{1k}, X_{2k}, \ldots, X_{nk}]$ of $\mathbf{X}$, regarded as a point in n dimensions. The location of this point is determined by the joint probability distribution $f(\mathbf{y}_k) = f(x_{1k}, x_{2k}, \ldots, x_{nk})$. When the measurements $X_{1k}, X_{2k}, \ldots, X_{nk}$ are a random sample, $f(\mathbf{y}_k) = f(x_{1k}, x_{2k}, \ldots, x_{nk}) = f_k(x_{1k})\, f_k(x_{2k}) \cdots f_k(x_{nk})$ and, consequently, each coordinate x_{jk} contributes equally to the location through the identical marginal distributions $f_k(x_{jk})$.

If the n components are not independent or the marginal distributions are not identical, the influence of individual measurements (coordinates) on location is asymmetrical. We would then be led to consider a distance function in which the coordinates were weighted unequally, as in the "statistical" distances or quadratic forms introduced in Section 4.

Certain conclusions can be reached concerning the sampling distributions of $\overline{\mathbf{X}}$ and $\mathbf{S}_n$ without making further assumptions regarding the form of the underlying joint distribution of the variables. In particular, we can see how $\overline{\mathbf{X}}$ and $\mathbf{S}_n$ are as point estimators of the corresponding population mean vector μ and covariance matrix Σ.

Theorem 6.1. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a joint distribution that has mean vector μ and covariance matrix Σ. Then $\overline{\mathbf{X}}$ is an unbiased estimator of μ and its covariance matrix is

$$\frac{1}{n}\Sigma$$

That is,

$$E(\overline{\mathbf{X}}) = \mu \qquad \text{(population mean vector)}$$

$$Cov(\overline{\mathbf{X}}) = \frac{1}{n}\Sigma \quad \text{(population variance-covariance matrix divided by sample size)} \tag{6.2}$$

For the covariance matrix $\mathbf{S}_n$,

$$E(\mathbf{S_n}) = \frac{n-1}{n}\Sigma = \Sigma - \frac{1}{n}\Sigma$$

Thus,

$$E(\frac{n}{n-1}\mathbf{S_n}) = \Sigma \tag{6.3}$$

so $[n/(n-1)]\mathbf{S_n}$ is an unbiased estimator of Σ, while $\mathbf{S_n}$ is a biased estimator with $(bias) = E(\mathbf{S_n}) - \Sigma = -(1/n)\Sigma$.

Proof. Now, $\overline{\mathbf{X}} = (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X_n})/n$. Using additivity property of expectation for two vectors gives

$$E(\overline{\mathbf{X}}) = E\left(\frac{1}{n}\mathbf{X}_1 + \frac{1}{n}\mathbf{X}_2 + \ldots + \frac{1}{n}\mathbf{X}_n\right)$$

$$= E\left(\frac{1}{n}\mathbf{X}_1\right) + E\left(\frac{1}{n}\mathbf{X}_2\right) + \ldots + E\left(\frac{1}{n}\mathbf{X}_n\right)$$

$$= \frac{1}{n}E(\mathbf{X}_1) + \frac{1}{n}E(\mathbf{X}_2) + \ldots + \frac{1}{n}E(\mathbf{X}_n) = \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu = \mu.$$

Next,

$$(\overline{\mathbf{X}} - \mu)(\overline{\mathbf{X}} - \mu)' = \left(\frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \mu)\right)\left(\frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_l - \mu)\right)'$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(\mathbf{X}_j - \mu)(\mathbf{X}_l - \mu)'$$

so

$$Cov(\overline{\mathbf{X}}) = E(\overline{\mathbf{X}} - \mu)(\overline{\mathbf{X}} - \mu)' = \frac{1}{n^2}\left(\sum_{j=1}^{n}\sum_{j=1}^{n}E(\mathbf{X}_j - \mu)(\mathbf{X}_l - \mu)'\right)$$

For $j \neq l$, each entry in $E(\mathbf{X}_j - \mu)(\mathbf{X}_l - \mu)'$ is zero because the entry is the covariance between a component of $\mathbf{X_j}$ and a component of $\mathbf{X_l}$, and these are independent. Therefore,

$$Cov(\overline{\mathbf{X}}) = \frac{1}{n^2}\left(\sum_{j=1}^{n}E(\mathbf{X}_j - \mu)(\mathbf{X}_j - \mu)'\right)$$

Since $\Sigma = E(\mathbf{X}_j - \mu)(\mathbf{X}_j - \mu)'$ is the common population covariance matrix for each $\mathbf{X}_j$, we have

$$Cov(\overline{\mathbf{X}}) = \frac{1}{n^2}\left(\sum_{j=1}^{n}E(\mathbf{X}_j - \mu)(\mathbf{X}_j - \mu)'\right) = \frac{1}{n^2}\sum_{j=1}^{n}\Sigma$$

$$= \frac{1}{n^2}(n\Sigma) = \frac{1}{n}\Sigma$$

To obtain the expected value of $\mathbf{S}_n$, we first note that $(X_{ji} - \overline{X}_i)(X_{jk} - \overline{X}_k)$ is the (i,k)th element of $(\mathbf{X_j} - \overline{\mathbf{X}})$ $(\mathbf{X_j} - \overline{\mathbf{X}})$. The matrix representing sums of squares and cross products can then be written as

$$\sum (X_j - \overline{X})(X_j - \overline{X})' = \sum (X_j - \overline{X})X_j' + (\sum (X_j - \overline{X}))(-\overline{X})'$$

$$= \sum X_j X_j' - n\overline{X}\,\overline{X}'$$

since $\sum (X_j - \overline{X}) = 0$ and $n\overline{X}' = \sum X_j'$. Therefore, its expected value is

$$E(\sum X_j X_j' - n\overline{X}\overline{X}') = \sum E(\overline{X}\overline{X}') - nE(\overline{X}\,\overline{X}')$$

For any random vector $\mathbf{V}$ with $E(\mathbf{V}) = \mu_v$ and $Cov(\mathbf{V}) = \Sigma_v$, we have $E(\mathbf{V}\,\mathbf{V}') = \Sigma_v + \mu_v\,\mu_v'$. Consequently,

$$E(\mathbf{X_j X_j'}) = \Sigma + \mu\,\mu' \quad \text{and} \quad E(\overline{\mathbf{X}}\,\overline{\mathbf{X}}') = \frac{1}{n}\Sigma + \mu\,\mu'$$

Using these results, we obtain

$$\sum E(X_j X_j') - nE(\overline{X}\overline{X}') = n\Sigma - n\mu\mu' - n(\frac{1}{n}\Sigma + \mu\mu') = (n-1)\Sigma$$

and thus, since $\mathbf{S}_n = (1/n)(\sum X_j X_j' - n\overline{X}\overline{X}')$, it follows immediately that

$$E(\mathbf{S}_n) = \frac{(n-1)}{n}\Sigma$$

Theorem 6.1 shows that the (i,k)th entry, $(n-1)^{-1}\sum (X_{ji} - \overline{X}_i)(X_{jk} - \overline{X}_k)$, of $[n/(n-1)]\mathbf{S}_n$ is an unbiased estimator of σ_{ik}. However, the individual sample standard deviations $\sqrt{s_{ii}}$, calculated with either n or $n-1$ as a divisor, are not unbiased estimators of the corresponding population quantities $\sqrt{\sigma_{ii}}$. Moreover, the correlation coefficients r_{ik} are not unbiased estimators of the population quantities P_{ik}. However, the bias $E(\sqrt{s_{ii}}) - \sqrt{\sigma_{ii}}$ or $E(r_{ik}) - \rho_{ik}$, can usually be ignored if the sample size n is moderately large. Consideration of bias motivates a slightly modified definition of the sample variance-covariance matrix. Theorem 6.1 provides us with an unbiased estimator $\mathbf{S}$ of Σ: (Unbiased) Sample Variance-Covariance Matrix

$$\mathbf{S} = (\frac{n}{n-1})\mathbf{S_n} = \frac{1}{n-1}\sum (X_j - \overline{X})(X_j - \overline{X}') \tag{6.4}$$

Here $\mathbf{S}$, without a subscript, has (i,k)th entry $(n-1)^{-1}\sum_{j=1}^{n}(X_{ji}-\overline{X}_i)(X_{jk}-\overline{X}_k)$.

This definition of sample covariance is commonly used in many multivariate test statistics.

§6.3. GENERALIZED VARIANCE.

With a single variable, the sample variance is often used to describe the amount of variation in the measurements on that variable. When m variables are observed on each unit, the variation is described by the sample variance-covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1m} \\ s_{12} & s_{22} & \dots & s_{2m} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ s_{1m} & s_{2m} & \dots & s_{mm} \end{bmatrix} = \left\{ s_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji}-\overline{x_i})(x_{jk}-\overline{x_k}) \right\}.$$

The sample covariance matrix contains m variances and $\frac{1}{2}m(m-1)$ potentially different covariances. Sometimes it is desirable to assign a single numerical value for the variation expressed by S. One choice for a value is the determinant of S, which reduces to the usual sample variance of a single characteristic when $m = 1$. This determinant is called the generalized sample variance:

$$\text{Generalized sample variance} = |S|.$$

The generalized sample variance provides one way of writing the information on all variances and covariances as a single number. Of course, when $m > 1$, some information about the sample is lost in the process. A geometrical interpretation of $|S|$ will help us appreciate its strengths and weaknesses as a descriptive summary.

§6.4 The Gamma Distributions.

The *gamma distribution* is important because it includes a wide class of specific distributions, some of which underlie fundamental statistical procedures. In addition to serving as a utility distribution, the gamma provides probabilities for yet another random variable associated with Poisson processes (the exponential distribution itself is a member of the gamma distributions).

The Gamma Function.

In order to describe the gamma distribution in detail, we must first consider a useful function, Gamma Function:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}\, e^{-x}\, dx, \qquad \alpha > 0.$$

The symbol Γ (Greek uppercase *gamma*) is reserved for this function. The integration by parts of $\Gamma(\alpha)$ yields that

$$\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha).$$

Note, that for any nonnegative integer k we have

$$\Gamma(k+1) = k!,$$

In particular, $\Gamma(1) = 1$.

An important class involves values with halves. We have

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

and for any positive integer k

$$\Gamma\left(k+\frac{1}{2}\right) = \frac{(2k-1)!!}{2^k}\,\sqrt{\pi},$$

where $(2k-1)!! = 1\cdot 3\cdot 5\cdot \ldots \cdot (2k-1)$.

The Density Function of Gamma Random Variable

The following expression gives the density function for a gamma distribution.

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda\, x} & \text{if} \quad x > 0. \end{cases}$$

The two parameters λ and α may be any positive values ($\lambda > 0$ and $\alpha > 0$).

A special case of this function occurs when $\alpha = 1$. We have

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \lambda\, e^{-\lambda\, x} & \text{if} \quad x > 0 \end{cases}$$

which is the density function for the exponential distribution.

The expectation and Variance of gamma distribution have the forms:

$$E\eta = \frac{\alpha}{\lambda} \qquad \text{and} \qquad \text{Var}\,(\eta) = \frac{\alpha}{\lambda^2}.$$

When α is natural, say $\alpha = n$, the gamma distribution with parameters (n, λ) often arises in practice:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{\lambda^n}{(n-1)!}\, x^{n-1}\, e^{-\lambda\, x} & \text{if} \quad x > 0. \end{cases}$$

This distribution is often referred to in the literature as the n-Erlang distribution. Note that when $n = 1$, this distribution reduces to the exponential.

The gamma distribution with $\lambda = 1/2$ and $\alpha = n/2$ (n is natural) is called χ^2_n (read "chi-squared") distribution with n degrees of freedom:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{1}{2^{n/2}\,\Gamma(n/2)}\, x^{n/2-1}\, e^{-x/2} & \text{if} \quad x > 0. \end{cases}$$

We have

$$E\chi^2_n = n \qquad \text{and} \qquad \text{Var}\chi^2_n = 2n.$$

If η_1, η_2, ..., η_n are independent standard normal random variables, then

$$\xi = \sum_{i=1}^{n} \eta_i^2$$

is said to have the *chi-squared* (sometimes seen as χ^2) distribution with n degrees of freedom. Let us compute its density function. When $n = 1$, $\xi = \eta_1^2$, and its density function is given by

$$f_{\eta_1^2}(x) = \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{1}{2\sqrt{x}}\,[f_{\eta_1}(\sqrt{x}) + f_{\eta_1}(-\sqrt{x})] & \text{if} \quad x > 0, \end{cases} =$$

$$= \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{1}{2\sqrt{x}}\,\dfrac{2}{\sqrt{2\,\pi}}\, exp(-x/2) & \text{if} \quad x > 0, \end{cases} =$$

$$= \begin{cases} 0 & \text{if} \quad x \le 0, \\ \dfrac{1}{2\sqrt{\pi}}\,\left(\dfrac{x}{2}\right)^{-1/2}\, exp(-x/2) & \text{if} \quad x > 0. \end{cases}$$

But we recognize the above as the gamma distribution with parameters $(1/2, 1/2)$ (A by-product of this analysis is that $\Gamma(1/2) = \sqrt{\pi}$). But as each η_i^2 is gamma distribution with parameters $(1/2, 1/2)$, we obtain that the χ^2 distribution with n degrees of freedom is just the gamma distribution with parameters $(n/2, 1/2)$ and has a probability density function given by

$$f_{\chi^2}(x) = \begin{cases} 0 & \text{if} \quad x \leq 0, \\ \dfrac{x^{n/2-1}\, exp\left(-\dfrac{x}{2}\right)}{2^{n/2}\,\Gamma\left(\dfrac{n}{2}\right)} & \text{if} \quad x > 0. \end{cases}$$

When n is even, $\Gamma(n/2) = (n/2 - 1)!$, whereas when n is odd, $\Gamma(n/2)$ can be obtained from iterating the relationship $\Gamma(\lambda) = (\lambda - 1)\,\Gamma(\lambda - 1)$ and then using the previously obtained result that $\Gamma(1/2) = \sqrt{\pi}$. For example,

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\,\Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\,\sqrt{\pi}.$$

The chi-squared distribution often arises in practice as being the distribution of the square of the error involved when one attempts to hit a target in n-dimensional space when the coordinate errors are taken to be independent until normal random variables. It is also important in statistical analysis.