Stat 571
10-10-13

The sample

(1)

n observations of a p-dimensional random vector

Data matrix:

$$X_{n \times p} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

Note: each row is an observation
each column consists of all obs on a single random variable.

(2)

Example: $X = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}_{3 \times 2}$

This represents 3 observations of a 2-dim r.v.

Write $X_{n \times p} = \begin{bmatrix} \vec{x}_1' \\ \hline \vec{x}_2' \\ \hline \vdots \\ \hline \vec{x}_n' \end{bmatrix}$ and $[\vec{y}_1 | \vec{y}_2 | \cdots | \vec{y}_p]$

③

Defn: The sample mean vector is

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{p\times 1}, \text{ where } \bar{x}_j = \sum_{i=1}^{n} X_{ij} / n$$

Defn: $$s_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (X_{ki} - \bar{x}_i)(X_{kj} - \bar{x}_j)$$

Note: s_{ii} is the sample variance of the i^{th} column of X.
s_{ij} is the sample covar. between i^{th} & j^{th} cols.

④

Defn. $S = \{s_{ij}\}_{p\times p}$ is the sample covariance matrix.

Defn: $$S_n = \frac{n-1}{n} S$$

Theorem: Suppose that $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ are independent random vectors, identically distributed (i.i.d), each with mean $\vec{\mu}$ and covariance Σ.

Then $E[\bar{X}] = \vec{\mu}$, $Cov[\bar{X}] = \frac{1}{n}\Sigma$, (5)

and $E[S] = \Sigma$.

Proof: $\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{p\times 1}$, $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$

So $\bar{X} = \left(\frac{1}{n}\sum_{i=1}^{n} \vec{x}_i'\right)'$

$= \frac{1}{n}\sum_{i=1}^{n}\vec{x}_i$

(6)

Then $E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n}\vec{x}_i\right]$

$= \frac{1}{n}\sum_{i=1}^{n} E[\vec{x}_i]$

$= \frac{1}{n}\sum_{i=1}^{n}\vec{\mu} = \vec{\mu}$

$Cov[\bar{X}] = E[(\bar{X}-\vec{\mu})(\bar{X}-\vec{\mu})']$

$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i - \vec{\mu})\right)\left(\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i-\vec{\mu})\right)'\right]$

(7)

$$= E\left[\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\vec{x}_i - \vec{\mu})(\vec{x}_j - \vec{\mu})' \right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} E\left[(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})' \right]$$

($i \neq j$ terms vanish by independence)

$$= \frac{1}{n^2} \sum_{i=1}^{n} \Sigma$$

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

(8)

Consider S_n.

Its $(i,j)^{th}$ entry is

$$\frac{1}{n} \sum_{k=1}^{n} \underbrace{(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}_{t_{ijk}}$$

Consider $\underset{p\times 1}{(\vec{x}_k - \bar{x})}\ \underset{1\times p}{(\vec{x}_k - \bar{x})'}$

$$= \begin{bmatrix} x_{k1} - \bar{x}_1 \\ x_{k2} - \bar{x}_2 \\ \vdots \\ x_{kp} - \bar{x}_p \end{bmatrix} \left[x_{k1} - \bar{x}_1, x_{k2} - \bar{x}_2, \ldots, x_{kp} - \bar{x}_p \right]$$

(9)

Its $(i,j)^{\text{th}}$ entry is

$$(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j) = t_{ijk}$$

Notice that

$$S_n = \frac{1}{n}\sum_{k=1}^{n}(\vec{X}_k-\bar{X})(\vec{X}_k-\bar{X})'$$

$$= \frac{1}{n}\left[\sum_{k=1}^{n}\vec{X}_k\vec{X}_k' - \overbrace{\left(\sum_{k=1}^{n}\vec{X}_k\right)}^{n\bar{X}}\bar{X}' - \bar{X}\underbrace{\sum_{k=1}^{n}\vec{X}_k'}_{n\bar{X}'} + n\bar{X}\bar{X}'\right]$$

(10)

$$S_n = \frac{1}{n}\sum_{k=1}^{n}\vec{X}_k\vec{X}_k' - \bar{X}\bar{X}'$$

$$E[S_n] = \frac{1}{n}\sum_{k=1}^{n}E[\vec{X}_k\vec{X}_k'] - E[\bar{X}\bar{X}']$$

Let $\vec{V}$ be a random vector

$$\mathrm{Cov}(\vec{V}) = E[(\vec{V}-\vec{\mu})(\vec{V}-\vec{\mu})']$$

$$= E[\vec{V}\vec{V}' - \vec{V}\vec{\mu}' - \vec{\mu}\vec{V}' + \vec{\mu}\vec{\mu}']$$

$$= E[\vec{V}\vec{V}'] - E[\vec{V}]\vec{\mu}' - \cancel{\vec{\mu}E[\vec{V}']} + \cancel{\vec{\mu}\vec{\mu}'}$$

(11)

$$\mathrm{Cov}(\vec{V}) = E[\vec{V}\vec{V}'] - \vec{\mu}\vec{\mu}'$$

$$\underline{\text{So } E[\vec{V}\vec{V}'] = \mathrm{Cov}(\vec{V}) + \vec{\mu}\vec{\mu}'}$$

Thus,

$$E[S_n] = \frac{1}{n}\sum_{k=1}^{n}\left[\mathrm{Cov}(\vec{X}_k) + \vec{\mu}\vec{\mu}'\right] - \left[\mathrm{Cov}(\bar{X}) + \vec{\mu}_{\bar{X}}\vec{\mu}_{\bar{X}}'\right]$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left[\Sigma + \vec{\mu}\vec{\mu}'\right] - \left[\frac{1}{n}\Sigma + \vec{\mu}\vec{\mu}'\right]$$

(12)

$$= \left(1 - \frac{1}{n}\right)\Sigma = \frac{n-1}{n}\Sigma$$

$$\therefore E\left(\frac{n}{n-1} S_n\right) = \Sigma$$

$$\shortparallel$$

$$E[S]$$

//

HW #2 due 10/17

p.144 #6, 7, 8