Random Vectors and Matrices

A random matrix is just a matrix of random variables. Their joint probability distribution is the distribution of the random matrix. Random matrices with just one column (say, px1) may be called *random vectors*.

Expected Value

- The expected value of a matrix is defined as the matrix of expected values.

- Denoting the *pxc* random matrix $\boldsymbol{X}$ by $[X_{i,j}]$, $E(\boldsymbol{X}) = [E(X_{i,j})]$

Immediately we have natural properties like

$$
\begin{aligned}
E(\mathbf{X} + \mathbf{Y}) &= E([X_{i,j}] + [Y_{i,j}]) \\
&= [E(X_{i,j} + Y_{i,j})] \\
&= [E(X_{i,j}) + E(Y_{i,j})] \\
&= [E(X_{i,j})] + [E(Y_{i,j})] \\
&= E(\mathbf{X}) + E(\mathbf{Y}).
\end{aligned}
$$

Let $\mathbf{A} = [a_{i,j}]$ be an $r \times p$ matrix of constants, while $\mathbf{X}$ is still a $p \times c$ random matrix. Then

$$
\begin{aligned}
E(\mathbf{AX}) &= E\left(\left[\sum_{k=1}^{p} a_{i,k} X_{k,j}\right]\right) \\
&= \left[E\left(\sum_{k=1}^{p} a_{i,k} X_{k,j}\right)\right] \\
&= \left[\sum_{k=1}^{p} a_{i,k} E(X_{k,j})\right] \\
&= \mathbf{A}E(\mathbf{X}).
\end{aligned}
$$

Similarly, have $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$

Variance-Covariance Matrices

Let X be a px1 random vector with E(X)=mu. The variance-covariance matrix of X (sometimes just called the covariance matrix), denoted by V(X), is defined as

$$V(\mathbf{X}) = E\left\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\right\}$$

$$V(\mathbf{X}) = E\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right\}$$

$$\begin{aligned}
V(\mathbf{X}) &= E\left\{\left[\begin{array}{c} X_1-\mu_1 \\ X_2-\mu_2 \\ X_3-\mu_3 \end{array}\right]\left[\begin{array}{ccc} X_1-\mu_1 & X_2-\mu_2 & X_3-\mu_3 \end{array}\right]\right\} \\
&= E\left\{\left[\begin{array}{lll} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & (X_1-\mu_1)(X_3-\mu_3) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & (X_2-\mu_2)(X_3-\mu_3) \\ (X_3-\mu_3)(X_1-\mu_1) & (X_3-\mu_3)(X_2-\mu_2) & (X_3-\mu_3)^2 \end{array}\right]\right\} \\
&= \left[\begin{array}{lll} E\{(X_1-\mu_1)^2\} & E\{(X_1-\mu_1)(X_2-\mu_2)\} & E\{(X_1-\mu_1)(X_3-\mu_3)\} \\ E\{(X_2-\mu_2)(X_1-\mu_1)\} & E\{(X_2-\mu_2)^2\} & E\{(X_2-\mu_2)(X_3-\mu_3)\} \\ E\{(X_3-\mu_3)(X_1-\mu_1)\} & E\{(X_3-\mu_3)(X_2-\mu_2)\} & E\{(X_3-\mu_3)^2\} \end{array}\right] \\
&= \left[\begin{array}{lll} V(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) \\ Cov(X_1,X_2) & V(X_2) & Cov(X_2,X_3) \\ Cov(X_1,X_3) & Cov(X_2,X_3) & V(X_3) \end{array}\right].
\end{aligned}$$

So, it's a *pxp* symmetric matrix with variances on the main diagonal and covariances on the off-diagonals.

Analogous to Var(aX) = a^2 Var(X)

$$
\begin{aligned}
V(\mathbf{AX}) &= E\{(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})'\} \\
&= E\left\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\left(\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\right)'\right\} \\
&= E\left\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}'\right\} \\
&= \mathbf{A}E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\}\mathbf{A}' \\
&= \mathbf{A}V(\mathbf{X})\mathbf{A}' \\
&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'
\end{aligned}
$$

Multivariate Normal

The $p \times 1$ random vector $\mathbf{X}$ is said to have a *multivariate normal distribution*, and we write $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if $\mathbf{X}$ has (joint) density

$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\mu}$ is $p \times 1$ and $\boldsymbol{\Sigma}$ is $p \times p$ symmetric and positive definite. Positive definite means that for any non-zero $p \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} > 0$.

- Since the one-dimensional random variable $Y = \sum_{i=1}^{p} a_i X_i$ may be written as $Y = \mathbf{a}'\mathbf{X}$ and $Var(Y) = V(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$, it is natural to require that $\boldsymbol{\Sigma}$ be positive definite. All it means is that every non-zero linear combination of $\mathbf{X}$ values has a positive variance.

- $\boldsymbol{\Sigma}$ positive definite is equivalent to $\boldsymbol{\Sigma}^{-1}$ positive definite.

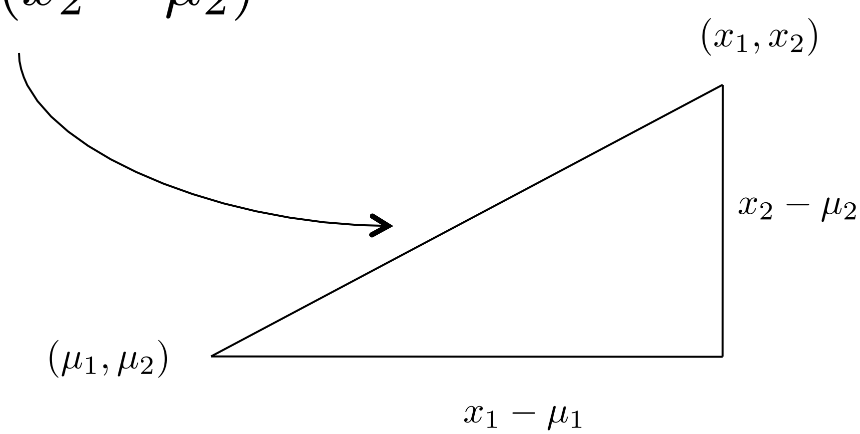
Analogies

- Univariate Normal
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$
  - $\frac{(x-\mu)^2}{\sigma^2}$ is the squared Euclidian distance between x and μ, in a space that is stretched by σ^2.
  - $\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$
- Multivariate Normal
  - $f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{k}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$
  - $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is the squared Euclidian distance between $\mathbf{x}$ and $\boldsymbol{\mu}$, in a space that is warped and stretched by $\boldsymbol{\Sigma}$.
  - $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(k)$

Distance: Suppose $\mathbf{\Sigma} = \mathbf{I}_2$

$$
\begin{aligned}
d^2 &= (\mathbf{X}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\
&= \begin{bmatrix} x_1-\mu_1, & x_2-\mu_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1-\mu_1, & x_2-\mu_2 \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \\
&= (x_1-\mu_1)^2 + (x_2-\mu_2)^2
\end{aligned}
$$

$$
d = \sqrt{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}
$$



The multivariate normal reduces to the univariate normal when $p = 1$. Other properties of the multivariate normal include the following.

1. $E(\mathbf{X}) = \boldsymbol{\mu}$
2. $V(\mathbf{X}) = \boldsymbol{\Sigma}$
3. If $\mathbf{c}$ is a vector of constants, $\mathbf{X} + \mathbf{c} \sim N(\mathbf{c} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$
4. If $\mathbf{A}$ is a matrix of constants, $\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$
5. All the marginals (dimension less than p) of $\mathbf{X}$ are (multivariate) normal, but it is possible in theory to have a collection of univariate normals whose joint distribution is not multivariate normal.
6. For the multivariate normal, zero covariance implies independence. The multivariate normal is the only continuous distribution with this property.
7. The random variable $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has a chi-square distribution with p degrees of freedom.
8. After a bit of work, the multivariate normal likelihood may be written as

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2}(2\pi)^{-np/2} \exp -\frac{n}{2}\left\{tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})\right\}, \quad \text{(A.15)}$$

   where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$ is the sample variance-covariance matrix (it would be unbiased if divided by $n - 1$).

Proof of (7): $(\mathbf{X}-\boldsymbol{\mu})`\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\sim$Chisquare(p)

- Let $\mathbf{Y} = \mathbf{X}-\boldsymbol{\mu} \sim N(\mathbf{0},\boldsymbol{\Sigma})$
- $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}\mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{-1/2}\ \boldsymbol{\Sigma}\ \boldsymbol{\Sigma}^{-1/2})$

  $= N(\mathbf{0}, [\boldsymbol{\Sigma}^{-1/2}\ \boldsymbol{\Sigma}^{1/2}][\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2}])$

  $= N(\mathbf{0},\mathbf{I})$
- $\mathbf{Y}`\boldsymbol{\Sigma}^{-1}\mathbf{Y} = \mathbf{Z}`\mathbf{Z}$ $\sim$Chisquare(p)

Independence of X-bar and S²

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \overline{X} \end{pmatrix} = \mathbf{AX}$$

Show $Cov\left(\overline{X}, (X_i - \overline{X})\right) = 0$ for $i = 1, \dots, n$. (Exercise)

$$\mathbf{Y}_2 = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \end{pmatrix} = \mathbf{BY} \text{ and } \overline{X} = \mathbf{CY} \text{ are independent.}$$

So $S^2 = g(\mathbf{Y}_2)$ and $\overline{X}$ are independent. ■