Stat 111 Week 3 Presentation 3
P. Everson

3. **Vector Random Variables** (Rice 14.4.1, Blitzstein 7.5)

   a) Introduce the concept of an n-dimensional random vector $\mathbf{X}$, and the generalizations of the mean and variance to the $n \times 1$ mean vector $\boldsymbol{\mu}$ and $n \times n$ covariance matrix $\mathbf{V}$ (Rice, 14.4.1). Show how the usual definition of $\text{Cov}(X_i, X_j)$ arises as the expectation of the i, jth element (and the j, ith element) of the random $n \times n$ matrix $(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}$. Give $\boldsymbol{\mu}$ and $\mathbf{V}$ for an $n \times 1$ vector of independent random variables with means μ_i and variances σ_i^2, for $i = 1, \ldots, n$.

   b) Suppose $\mathbf{X}$ is $n \times 1$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$. Explain the formulas for the mean vector and covariance matrix for linear transformations:

   $$E(\mathbf{AX} + \mathbf{C}) \; = \; \mathbf{A}E(\mathbf{X}) + \mathbf{C}, \qquad \text{Var}(\mathbf{AX} + \mathbf{C}) \; = \; \mathbf{AVA}^{\mathrm{T}}$$

   where $\mathbf{Y} = \mathbf{AX} + \mathbf{C}$, for $\mathbf{A}$ an $m \times n$ and $\mathbf{C}$ an $m \times 1$ matrix of constants.

   c) Show that

   $$\text{Var}\left(\sum_{i=1}^{n} X_i\right) \; = \; \left(\sum_{i=1}^{n} \text{Var}(X_i)\right) + 2\left(\sum_{i<j} \text{Cov}(X_i, X_j)\right).$$

   which is the sum of all of the elements in the covariance matrix $\mathbf{V}$.

   d) Show that $\text{Cov}(aX + b, c + dY) = ad\text{Cov}(X, Y)$ and $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

   e) For $X_1, \ldots, X_n$ iid random variables, show that $\bar{X}$ is uncorrelated with $X_i - \bar{X}$, for any $i = 1, \ldots, n$. Use the vector matrix formulation, and also show a scalar argument.

The mean and variance summarize the center and spread of the distribution of a random variable, and covariance/correlation summarizes the strength and direction of the (linear) association between two random variables. For describing n random variables $X_1, \ldots, X_n$ ($n \geq 2$) it is convenient to organize the variables as an $n \times 1$ vector and treat this as a vector random variable. The $n \times 1$ vector of mean values $\boldsymbol{\mu}$ is then the mean vector, and the symmetric $n \times n$ matrix of covariances $\mathbf{V}$, with $\mathbf{V}_{[ij]} = \text{Cov}(X_i, X_j)$ is the variance/covariance matrix (diagonal entries of $\mathbf{V}$ are the variances).

$$\mathbf{V} \; = \; \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \ldots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & \ldots & \text{Cov}(X_2, X_n) \\ \vdots & & \ddots & \vdots \\ \text{Cov}(X_1, X_n) & \ldots & & \text{Var}(X_n) \end{pmatrix}$$

The mean vector and covariance matrix are natural extensions of the usual scalar mean and variance. The $n \times 1$ mean vector $\boldsymbol{\mu}$ is the expectation of the $n \times 1$ vector random variable $\mathbf{X}$, with expectation defined element-wise. Similarly, the variance matrix $\mathbf{V}$ is the expectation of the $n \times n$ matrix $(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}$, generalizing the definition of $\text{Var}(X) = E((X - \mu)^2)$:

$$(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}} \; = \; \begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) & \ldots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_1 - \mu_1)(X_2 - \mu_2) & (X_2 - \mu_2)(X_2 - \mu_2) & \ldots & (X_2 - \mu_2)(X_n - \mu_n) \\ \vdots & & \ddots & \vdots \\ (X_1 - \mu_1)(X_n - \mu_n) & \ldots & & (X_n - \mu_n)(X_n - \mu_n) \end{pmatrix}$$

Often we assume iid random variables, all with mean μ and variance σ^2, in which case these two scalar parameters completely specify the mean vector and covariance matrix: $\boldsymbol{\mu} = \mu\mathbf{1}$, for $\mathbf{1}$ an $n \times 1$ vector of 1's, and $\mathbf{V} = \sigma^2\mathbf{I}$, for $\mathbf{I}$ the $n \times n$ identity matrix. For general independent random variables (not iid), the covariance matrix is the diagonal matrix of variances.

Linear transformations are easily represented using matrix algebra. Expectation is a linear operator, so the mean vector for a linear transform is simply the linear transform applied to the mean:

$$E(\mathbf{AX} + \mathbf{C}) \;=\; \mathbf{A}E(\mathbf{X}) + \mathbf{C}$$

Adding a constant vector $\mathbf{C}$ does not affect variance (or covariance), and multiplying a random variable by a constant produces a variance multiplied by the constant squared. The analog to squaring a matrix is to pre-multiply $\mathbf{V}$ by the matrix, and then post-multiply by the matrix transpose. So if $\mathbf{X}$ has covariance matrix $\mathbf{V}$, then

$$\mathrm{Var}(\mathbf{AX} + \mathbf{C}) \;=\; \mathbf{AVA}^{\mathrm{T}}$$

A simple example is to set $\mathbf{A}$ be a $1 \times n$ vector of 1's, so that $\mathbf{AX} = \sum X_i$. The covariance matrix becomes $\mathbf{AVA}^{\mathrm{T}}$ which is simply the sum of every element in the matrix $\mathbf{V}$. We could represent this as

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \;=\; \left(\sum_{i=1}^{n} \mathrm{Var}(X_i)\right) + 2\left(\sum_{i<j} \mathrm{Cov}(X_i, X_j)\right)$$

We can also see nice distributive properties for covariance. For a vector $\mathbf{T} = \begin{pmatrix} X \\ Y \end{pmatrix}$ with covariance matrix $\mathbf{V}$, consider $\mathbf{C} + \mathbf{AT}$ for $\mathbf{C} = \begin{pmatrix} b \\ c \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$, so that $\mathbf{T} = \begin{pmatrix} aX + b \\ dY + c \end{pmatrix}$. The new covariance matrix is

$$\mathbf{AVA}^{\mathrm{T}} \;=\; \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \;=\; \begin{pmatrix} a^2\sigma_x^2 & ad\sigma_{xy} \\ ad\sigma_{xy} & d^2\sigma^2 y \end{pmatrix}$$

showing that $\mathrm{Cov}(aX + b, c + dY) = ad\mathrm{Cov}(X, Y)$. Similarly, if $\mathbf{T} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ has covariance matrix $\mathbf{V}$ and $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ so that $\mathbf{AT} = \begin{pmatrix} X + Y \\ Z \end{pmatrix}$, then the covariance matrix for $\mathbf{AT}$ is

$$\begin{aligned}
\mathbf{AVA}^{\mathrm{T}} \;&=\; \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \;=\; \begin{pmatrix} \sigma_x^2 + \sigma_{xy} & \sigma_{xy} + \sigma_y^2 & \sigma_{xz} + \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&=\; \begin{pmatrix} \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy} & \sigma_{xz} + \sigma_{yz} \\ \sigma_{xz} + \sigma_{yz} & \sigma_z^2 \end{pmatrix}
\end{aligned}$$

showing that $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$. Finally, we show that for $X_1, \ldots X_n$ iid random variables, the average $\bar{X} = \frac{1}{n}\sum X_i$ is uncorrelated with any deviation from the average $X_i - \bar{X}$. Without loss of generality, consider $i = 1$. Consider $\mathbf{AX}$ for $\mathbf{A} = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 1 - \frac{1}{n} & -\frac{1}{n} & \ldots, & -\frac{1}{n} \end{pmatrix}$ so that $\mathbf{AX} = \begin{pmatrix} \bar{X} \\ X_1 - \bar{X} \end{pmatrix}$. We have $\mathbf{V} = \sigma^2\mathbf{I}$, so $\mathrm{Cov}(\mathbf{AX})$ is $\sigma^2\mathbf{AA}^{\mathrm{T}} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{n-1}{n} \end{pmatrix}$, showing $\bar{X}$ and $X_1 - \bar{X}$ are uncorrelated. To see this in scalar form, write $\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}\sum_{j=2}^{n} X_j$ and use the distributive property:

$$\begin{aligned}
\mathrm{Cov}(\bar{X}, X_1 - \bar{X}) &= \mathrm{Cov}(\bar{X}, X_1) - \mathrm{Cov}(\bar{X}, \bar{X}) = \mathrm{Cov}(\frac{1}{n}X_1, X_1) + \mathrm{Cov}(\frac{1}{n}\sum_{j=2}^{n} X_j, X_1) - \mathrm{Var}(\bar{X}) \\
&= \frac{1}{n}\mathrm{Var}(X_1) + 0 - \frac{\sigma^2}{n} = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0.
\end{aligned}$$

More generally, we can say that $\bar{X}$ is uncorrelated with the vector of deviations $X_i - \bar{X}$, for $i = 1, \ldots, n$. As we shall see, for multivariate Normal variables, uncorrelated implies independent. So for an iid sample of Normal variables, $\bar{X}$ is independent of the set of deviations, and so is independent of the sample variance $s^2 = \sum \frac{(X_i - \bar{X})^2}{n-1}$, which is a function of the deviations only. This is important for deriving the sampling distribution of the t statistic.