Math 40 Probability and Statistical Inference

Winter 2021

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 12: Multidimensional Random Vectors (3.10)

Multidimensional Random Vectors (3.10)

- In this lecture, we consider a random vector $\mathbf{X} = (X_1, X_2, ..., X_m)$ of size m, where each X_i is a random variable.
- The bivariate normal distribution (3.5) is an example of a random vector in which each component is a normal distribution.
- Linear Algebra (or Matrix Algebra) is useful to understand random vectors. Please check the appendix (section 10.2) of the textbook.
- As linear algebra is not a prerequisite of this course, I will not ask questions directly related to linear algebra (but still useful to better understand the materials in this course.

Multidimensional Random Vectors (3.10)

- As in the joint cdf case in section 3.1, the cdf of a random vector $\mathbf{X}$ is defined as

$$F(\mathbf{x}) = Pr(X_1 \leq x_1, X_2 \leq x_2, ..., X_m \leq x_m).$$

Note that $\mathbf{x} = (x_1, x_2, ..., x_m)$ is a vector value.

- The pdf is defined as

$$f(\mathbf{x}) = \frac{\partial^m F(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_m}$$

Multidimensional Random Vectors (3.10)

- The mean of **X** is a vector,

$$\boldsymbol{\mu} = E(\mathbf{X}) = (E(X_1), ..., E(X_m)).$$

- How many pairs can you think out of $(X_1, ..., X_m)$?
- The covariance matrix $cov(\mathbf{X})$ is a $m \times m$ matrix

$$cov(\mathbf{X}) = \begin{pmatrix} cov(X_1, X_1) & cov(X_1, X_2) & \cdots & cov(X_1, X_m) \\ cov(X_2, X_1) & cov(X_2, X_2) & \cdots & cov(X_2, X_m) \\ & \vdots & & \vdots \\ cov(X_m, X_1) & cov(X_m, X_2) & \cdots & cov(X_m, X_m) \end{pmatrix}$$

Multidimensional Random Vectors (3.10)

- For two random vectors $\mathbf{X}$ and $\mathbf{Y}$ of the same size m, the covariance matrix $cov(\mathbf{X}, \mathbf{Y})$ is defined as

$$cov(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} cov(X_1, Y_1) & cov(X_1, Y_2) & \cdots & cov(X_1, Y_m) \\ cov(X_2, Y_1) & cov(X_2, Y_2) & \cdots & cov(X_2, Y_m) \\ & \vdots & & \vdots \\ cov(X_m, Y_1) & cov(X_m, Y_2) & \cdots & cov(X_m, Y_m) \end{pmatrix}$$

- $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated if $cov(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$.

Multidimensional Random Vectors (3.10)

Several properties are in order.

- $cov(\mathbf{X}+\mathbf{Y}) = cov(\mathbf{X}) + cov(\mathbf{Y})$ if $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated.

$\mathbf{X}, \mathbf{Y}$, and $\mathbf{Z}$ are random vectors of the same size, and $\mathbf{A}$ and $\mathbf{B}$ are fixed matrices while a and b are scalar values. (theorem 3.79)

- $E(a\mathbf{X}+b\mathbf{Y}) = aE(\mathbf{X}) + bE(\mathbf{Y})$
- $E(A\mathbf{X}+B\mathbf{Y}) = AE(\mathbf{X}) + BE(\mathbf{Y})$
- $cov(\mathbf{X}+\mathbf{Y}, \mathbf{Z}) = cov(\mathbf{X}, \mathbf{Z}) + cov(\mathbf{Y}, \mathbf{Z})$
- $cov(\mathbf{AX}) = \mathbf{A}cov(X)\mathbf{A}'$

Also, the following results hold (theorem 3.83)

- $cov(\mathbf{X}) = E(\mathbf{XX'}) - \mu_{\mathbf{X}}\mu_{\mathbf{X}}'$
- $cov(\mathbf{X}, \mathbf{Y}) = E(\mathbf{XY'}) - \mu_{\mathbf{X}}\mu_{\mathbf{Y}}'$
- $cov(\mathbf{X}) = E((\mathbf{X}-\mu_{\mathbf{X}})\mathbf{X}') = E(\mathbf{X}(\mathbf{X}-\mu_{\mathbf{X}})')$
- $cov(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X}(\mathbf{Y}-\mu_{\mathbf{Y}})') = E((\mathbf{X}-\mu_{\mathbf{X}})\mathbf{Y}')$

Multidimensional Random Vectors (3.10)

Example 3.84 Let X, Y and Z be independent (scalar) random variables. Find the covariance of $\mathbf{X} = (X, Y - X, X + Y + Z)$.

Multidimensional Random Vectors (3.10)

(Related to **Example 3.85**) Let Y is a (scalar) random variable and $\mathbf{1}$ is a vector of size m whose components are all 1. What is the covariance matrix of $Y\mathbf{1} = (Y, Y, ..., Y)$?

Multivariate Conditional Distribution (3.10.1)

Let $\mathbf{Y}$ is a random vector of size p while $\mathbf{X}$ is a random vector of size q.

As a prediction of $\mathbf{Y}$ for a given $\mathbf{X}$, we choose the regression, i.e. the conditional expectation of $\mathbf{Y}$ given $\mathbf{X}$. Note that this minimizes the expected square of the error $E((\mathbf{Y} - r(\mathbf{X})^2)$.

Note that $r(\mathbf{X})$ is $\mu(\mathbf{X})$ in Example 3.88.

Multivariate Conditional Distribution (3.10.1)

To calculate the conditional expectation, we need to know the conditional density,

$$f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{f(\mathbf{x},\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}.$$

Using the conditional density, the conditional expectation is given by

$$E(\mathbf{Y}|\mathbf{X} = \mathbf{X} = \mathbf{x}) = \int_{\mathbb{R}^p} \mathbf{y} f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) d\mathbf{y} = \int_{\mathbb{R}^p} \mathbf{y} \frac{f(\mathbf{x},\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})} d\mathbf{y}$$

Multivariate MGF (3.10.2)

For a random vector $\mathbf{X}$, the MGF of $\mathbf{X}$ is defined by

$$M(\mathbf{t}) = E(e^{\mathbf{t}'\mathbf{X}})$$

Here $\mathbf{t}$ is a vector of the same length as $\mathbf{X}$.
Note that $\mathbf{t}'\mathbf{X}$ is a scalar. That is,

$$\mathbf{t}'\mathbf{X} = t_1X_1 + t_2X_2 + \cdots + t_mX_m.$$

where $\mathbf{t} = (t_1, t_2, ..., t_m)$ and $\mathbf{X} = (X_1, X_2, ..., X_m)$.

Multinomial Distribution (3.10.4)

- n-tosses of a coin can be modeled as a Binomial distribution (section 1.6).
- What about n-tosses of a dice? A toss of a dice has six possible outcomes. This can be modeled as a multinomial distribution (section 3.10.4).
- As a general definition, let's assume that we do n experiments of a special dice with m outcomes.
- Let p_i be the corresponding probability of the i-th outcome.
- Also let X_i be the number of the i-th outcome, which is a random variable.

Multinomial Distribution (3.10.4)

- The probability mass function

$$Pr(X_1 = x_1, X_2 = x_2, ..., X_m = x_m) = \frac{n!}{x_1!x_2!\cdots x_m!} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m},$$

  where $x_1 + x_2 + \cdots x_m = 1$.
- $E(X_j) = np_j$
- $Var(X_j) = np_i(1 - p_j)$
- $E(X_j X_k) = n(n-1)p_j p_k$
- $cov(X_j, X_k) = -np_j p_k$

Multinomial Distribution (3.10.4)

- The probability mass function

$$Pr(X_1 = x_1, X_2 = x_2, ..., X_m = x_m) = \frac{n!}{x_1!x_2!\cdots x_m!} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m},$$

where $x_1 + x_2 + \cdots x_m = 1$.

- $E(X_j) = np_j$
- $Var(X_j) = np_i(1 - p_j)$
- $E(X_j X_k) = n(n-1)p_j p_k$
- $cov(X_j, X_k) = -np_j p_k$

All these properties can be derived from the MGF function

$$M(\mathbf{t}) = \left(\sum_{i=1}^{k} p_i e^{t_i} \right)^n$$

Multinomial Distribution (3.10.4)

- $E(X_j) = \frac{M(\mathbf{t})}{\partial t_j}|_{\mathbf{t}=\mathbf{0}} = n\left(\sum_{i=1}^{k} p_i e^{t_i}\right)^{n-1} p_j e^{t_j}|_{\mathbf{t}=\mathbf{0}} = np_j$
- $E(X_j X_k) = n(n-1)\left(\sum_{i=1}^{k} p_i e^{t_i}\right)^{n-2} p_j e^{t_j} p_k e^{t_k}|_{\mathbf{t}=\mathbf{0}} = n(n-1)p_j p_k$
- etc.

Multinomial Distribution (3.10.4)

- $E(X_j) = \frac{M(\mathbf{t})}{\partial t_j}|_{\mathbf{t}=\mathbf{0}} = n\left(\sum_{i=1}^{k} p_i e^{t_i}\right)^{n-1} p_j e^{t_j}|_{\mathbf{t}=\mathbf{0}} = np_j$
- $E(X_j X_k) = n(n-1)\left(\sum_{i=1}^{k} p_i e^{t_i}\right)^{n-2} p_j e^{t_j} p_k e^{t_k}|_{\mathbf{t}=\mathbf{0}} = n(n-1)p_j p_k$
- etc.

So, the MGF is important. How do you calculate the MGF?