Review of Random Variables, Vectors, and Matrices

Subrata Paul

6/4/2020

Probability Basics

The **sample space** Ω is the set of possible outcomes of an experiment.

Points ω in Ω are called **sample outcomes**, **realizations**, or **elements**.

Subsets of Ω are **events**.

A function P that assigns a real number $P(A)$ to every event A is a probability distribution if it satisfies three properties:

- $P(A) \geq 0$ for all $A \in \Omega$
- $P(\Omega) = 1$
- If $A_1, A_2, … $ are disjoint, then $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Independence

A set of events $\{A_i : i \in I\}$ is independent if

$$P\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

for every finite subset J of I.

Random Variables

A **random variable** Y is a mapping

$$Y : \Omega \to \mathbb{R}$$

that assigns a real number $Y(\omega)$ to each outcome ω.

The **cumulative distribution function (CDF)** of Y is a function $F_Y : \mathbb{R} \to [0, 1]$ defined by $F_Y(y) = P(Y \le y)$.

Y is a **discrete** random variable if it takes countably many values $\{x_1, x_2, \dots\}$.

The **probability mass function (pmf)** for Y is $f_Y(y) = P(Y = y)$.

Random Variables

Y is a continuous random variable if there exists a function $f_Y(y)$ such that: - $f_Y(y) \geq 0$ for all y, - $\int_{-\infty}^{\infty} f_Y(y)dy = 1$, - and for $a \leq b$, $P(a < Y < b) = \int_a^b f_Y(y)dy$.

The function f_Y is called the **probability density function (pdf)**.

Additionally, $F_Y(y) = \int_{-\infty}^{y} f_Y(y)dy$ and $f_Y(y) = F'_Y(y)$ at any point y at which F_Y is differentiable.

Expected value and Variance

The expected value, mean, or first moment of Y is defined as

$$E(Y) = \begin{cases} \sum_y yf(y) & \text{if } Y \text{ is discrete} \\ \int yf(y)dy & \text{if } Y \text{ is continuous} \end{cases}$$

assuming the sum and integral are well-defined.

The **variance** of Y is defined by

$$var(Y) = E(Y - E(Y))^2$$

and the **standard deviation** of Y is

$$SD(Y) = \sqrt{var(Y)}$$

.

Bivariate distributions

Consider two random variables X and Y. Let S be the joint **support** of X and Y (all the possible combinations of X and Y).

The joint CDF of the random variables is

$$F(x, y) = P(X \leq x, Y \leq y)$$

.

If X and Y are jointly discrete, the joint pmf $f(x, y)$ specifies $P(X = x, Y = y)$ and satisfies the following properties:

- $0 \leq f(x, y) \leq 1$
- $\sum \sum_{(x,y) \in S} f(x, y) = 1$
- $P((X, Y) \in A) = \sum \sum_{(x,y) \in A} f(x, y)$.

Marginal Density of Discrete RV

If X and Y are jointly discrete with joint pmf $f(x, y)$, then the marginal pmf of X, $f_X(x)$ is obtained via the formula

$$f_X(x) = \sum_{\text{all } y} f(x, y)$$

and

$$E(XY) = \sum_{\text{all } x} \sum_{\text{all } y} xyf(x, y).$$

Joint of continuous RV

If X and Y are jointly continuous, $f(x, y)$ is the joint pdf if:

- f$(x, y) \geq 0$ for all $(x, y) \in S$
- $\int_{-\infty}^{\infty} \int_{-infty}^{\infty} f(x, y) dxdy = 1$
- $P((X, Y) \in A) = \int \int_{x,y \in A} f(x, y) dxdy.$

If X and Y are jointly continuous with joint pdf $f(x, y)$, then the marginal density of X, $f_X(x)$ is obtained via the formula

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dydx.$$

Covariance

The covariance between random variables Y and Z is

$$cov(Y, Z) = E[(Y - E(Y))(Z - E(Z))] = E(YZ) - E(Y)E(Z).$$

Properties of Variance and Covariance

Let a and b be scalar constants. Then:

- $E(aY) = aE(Y)$
- $E(a + Y) = a + E(Y)$
- $E(aY + bZ) = aE(Y) + bE(Z)$
- $var(aY) = a^2 var(Y)$
- $var(a + Y) = var(Y)$
- $cov(aY, bZ) = abcov(Y, Z)$.
- $var(Y + Z) = var(Y) + var(Z) + 2cov(Y, Z)$.

Independence

Y and Z are independent if $F(y, z) = F_Y(y)F_Z(z)$.

If Y and Z are independent, then:

- E(YZ)=E(Y)E(Z)
- cov(Y,Z)=0.

Properties of random vectors

Let $\mathbf{y} = (Y_1, Y_2, \ldots, Y_n)^T$ be an $n \times 1$ vector of random variables. $\mathbf{y}$ is a random vector.

- A vector is always defined to be a column vector, even if the notation is ambiguous.

$$E(\mathbf{y}) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix}$$

Properties of random vectors

Let $\mathbf{y} = (Y_1, Y_2, \ldots, Y_n)^T$ be an $n \times 1$ vector of random variables. $\mathbf{y}$ is a random vector.

$$
\begin{aligned}
var(\mathbf{y}) =& E(\mathbf{y}\mathbf{y}^T) - E(\mathbf{y})E(\mathbf{y})^T \\
=& \begin{pmatrix} var(Y_1) & cov(Y_1, Y_2) & \ldots & cov(Y_1, Y_n) \\ cov(Y_2, Y_1) & var(Y_2) & \ldots & cov(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ cov(Y_n, Y_1) & cov(Y_n, Y_2) & \ldots & var(Y_n) \end{pmatrix}.
\end{aligned}
$$

Important Properties

Define:

- A to be an m×n matrix of constants
- $\mathbf{x} = (X_1, X_2, \ldots, X_n)^T$ and $\mathbf{z} = (Z_1, Z_2, \ldots, Z_n)^T$ to be $n \times 1$ random vectors.

Formally,

$$cov(\mathbf{x}, \mathbf{y}) = E(\mathbf{x}\mathbf{y}^T) - E(\mathbf{x})E(\mathbf{y})^T.$$

Important Properties

Additionally:

- $E(A\mathbf{y}) = AE(\mathbf{y}), E(\mathbf{y}A^T) = E(\mathbf{y})A^T.$
- $E(\mathbf{x} + \mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y})$
- $var(A\mathbf{y}) = A\ var(\mathbf{y})A^T$
- $cov(\mathbf{x} + \mathbf{y}, \mathbf{z}) = cov(\mathbf{x}, \mathbf{z}) + cov(\mathbf{y}, \mathbf{z})$
- $cov(\mathbf{x}, \mathbf{y} + \mathbf{z}) = cov(\mathbf{x}, \mathbf{y}) + cov(\mathbf{x}, \mathbf{z})$
- $cov(A\mathbf{x}, \mathbf{y}) = A\ cov(\mathbf{x}, \mathbf{y})$ and $cov(\mathbf{x}, A\mathbf{y}) = cov(\mathbf{x}, \mathbf{y})A^T.$

Important Properties

Let $\mathbf{a}$ is an $n \times 1$ vector of constants and $\mathbf{0}_{n\times n}$ be an $n \times n$ matrix of zeros, then

$$var(a) = 0_{n\times n},$$

$$cov(\mathbf{a}, \mathbf{y}) = 0_{n\times n},$$

and

$$var(\mathbf{a} + \mathbf{y}) = var(\mathbf{y}).$$

Multivariate normal (Gaussian) distribution

$\mathbf{y} = (Y_1, \ldots, Y_n)^T$ has a multivariate normal distribution with mean μ (an $n \times 1$ vector) and covariance Σ (an $n \times n$ matrix) if the joint pdf is

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu) \right).$$

Note that Σ must be symmetric and positive definite.

We would denote this as $\mathbf{y} \sim N(\mu, \Sigma)$.

Properties

Important fact: A linear function of a multivariate normal random vector (i.e., $a + A\mathbf{y}$) is also multivariate normal (though it could collapse to a single random variable).

Application: Suppose that $\mathbf{y} \sim N(\mu, \Sigma)$. For an $m \times n$ matrix of constants A, $A\mathbf{y} \sim N(A\mu, A\Sigma A^T)$.

Example

Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let Y_1 denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, Y_1 varies from week to week. Let Y_2 denote the proportion of the capacity of the bulk tank that is sold during the week. Because Y_1 and Y_2 are both proportions, both variables are between 0 and 1. Further, the amount sold, y_2, cannot exceed the amount available, y_1. Suppose the joint density function for Y_1 and Y_2 is given by

$$f(y_1, y_2) = 3y_1;\ 0 \le y_2 \le y_1 \le 1.$$

Problem 1

Determine $P(0 \leq Y_1 \leq 0.5;\ 0.25 \leq Y_2)$

Problem 2

Determine f_{Y_1} and f_{Y_2}

Problem 3

Determine $E(Y_1)$ and $E(Y_2)$

Problem 4

Determine $var(Y_1)$ and $var(Y_2)$

Problem 5

Determine $E(Y_1 Y_2)$

Problem 6

Determine $cov(Y_1, Y_2)$

Problem 7

Determine the mean and variance of $a^T y$, where $a = (1, -1)^T$ and $y = (Y_1, Y_2)^T$. This is the expectation and variance of the different between the amount of gas available and the amount of gas sold:

Matrix Differentiation

Matrix Differentiation 1

Let

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = A$$

Matrix Differentiation 1 (Proof)

Since ith element of $\mathbf{y}$ is given by

$$y_i = \sum_{k=1}^{n} a_{ik} x_k,$$

it follows that

$$\frac{\partial y_i}{\partial x_j} = a_{ij}$$

for all $i = 1, \ldots, m, \quad j = 1, \ldots, n$. Hence

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = A$$

Matrix Differentiation 2

Let the scalar α be defined by

$$\alpha = \mathbf{y}^T \mathbf{A} \mathbf{x},$$

where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$ and $\mathbf{y}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^T \mathbf{A}$$

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^T \mathbf{A}^T$$

Matrix Differentiation 2 (Proof)

Define $\mathbf{w}^T = \mathbf{y}^T \mathbf{A}$ and note that $\alpha = \mathbf{w}^T \mathbf{x}$

Hence,

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{w}^T = \mathbf{y}^T \mathbf{A}.$$

Since α is a scalar we can write

$$\alpha = \alpha^T = \mathbf{x}^T \mathbf{A}^T \mathbf{y}$$

hence,

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^T \mathbf{A}^T$$

Matrix Differentiation 3

For the special case in which the scalar α is given by the quadratic form

$$\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T (\mathbf{A} + \mathbf{A}^T)$$

Matrix Differentiation 3 (Proof)

By definition,

$$\alpha = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

Differentiating with respect to the kth element of x we have

$$\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$$

for all $k = 1, \ldots, n$, and consequently,

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A} = \mathbf{x}^T (A^T + A)$$

Matrix Differentiation 3.5

For the special case where $\mathbb{A}$ is a symmetric matrix and

$$\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$, and $\mathbf{A}$ does not depend on $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}^T \mathbb{A}.$$

Matrix Differentiation 4

Let the scalar α be defined by

$$\alpha = \mathbf{y}^T \mathbf{x}$$

where $\mathbf{y}$ is $n \times 1$, $\mathbf{x}$ is $n \times 1$, and both $\mathbf{y}$ and $\mathbf{x}$ are functions of the vector $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^T \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^T \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Matrix Differentiation 4.5

Let the scalar α be defined by

$$\alpha = \mathbf{x}^T \mathbf{x}$$

where $\mathbf{x}$ is $n \times 1$, and $\mathbf{x}$ is a function of the vector $\mathbf{z}$. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = 2\mathbf{x}^T \frac{\partial \mathbf{x}}{\partial \mathbf{z}}.$$