Elements of Probability Theory

Contents

1 Random Variables and Distributions 2
  1.1 Univariate Random Variables and Distributions . . . . . . 2
  1.2 Bivariate Random Variables and Distributions . . . . . . 3

2 Moments 5
  2.1 Expected Value or Mean . . . . . . . . . . . . . . . . . . . . 5
  2.2 Variance and Standard Deviation . . . . . . . . . . . . . . 6
  2.3 Higher order Moments . . . . . . . . . . . . . . . . . . . . 7
  2.4 Covariance and Correlation . . . . . . . . . . . . . . . . . 7
  2.5 Conditional Expectation and Variance . . . . . . . . . . . 8

3 Random Vectors and Random Matrices 9

4 Important Distributions 10
  4.1 Univariate Normal Distribution . . . . . . . . . . . . . . . 10
  4.2 Bivariate Normal Distribution . . . . . . . . . . . . . . . 10
  4.3 Multivariate Normal Distribution . . . . . . . . . . . . . . 11

1 Random Variables and Distributions

A random variable is a variable whose values are determined by a probability distribution. This is a casual way of defining random variables which is sufficient for our level of analysis. For more advanced probability theory, a random variable will be defined as a real-valued function over some probability space.

In section 1 to 3, a random variable is denoted by capital letters, e.g. \( X \), whereas its realizations are denoted by small letters, e.g. \( x \).

1.1 Univariate Random Variables and Distributions

A univariate discrete random variable is a variable that takes a countable number \( K \) of real numbers with certain probabilities. The probability that the random variable \( X \) takes the value \( x_k \) among the \( K \) possible realizations is given by the probability distribution

\[
P(X = x_k) = P(x_k) = p_k
\]

with \( k = 1, 2, \ldots, K \). \( K \) may be \( \infty \) in some cases. This can also be written as

\[
P(x_k) = \begin{cases} 
p_1 & \text{if } X = x_1 \\
p_2 & \text{if } X = x_2 \\
\vdots \\
p_K & \text{if } X = x_K 
\end{cases}
\]

Note that

\[
\sum_{k=1}^{K} p_k = 1.
\]

A univariate continuous random variable is a variable that takes a continuum of values in the real line. The distribution of a continuous random variable \( X \) can be characterized by a density function or probability
density function (pdf) $f(x)$. The nonnegative function $f(x)$ is such that
\[ P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x) \, dx. \]
defines the probability that $X$ takes a value in the interval $[x_1, x_2]$. Note
that there is no chance that $X$ takes exactly the value $x$, $P(X = x) = 0$.
The probability that $X$ takes any value on the real line is
\[ \int_{-\infty}^{\infty} f(x) \, dx = 1. \]

The distribution of a univariate random variable $X$ is alternatively
described by the cumulative distribution function (cdf)
\[ F(x) = P(X < x). \]
The cdf of a discrete random variable $X$ is
\[ F(x) = \sum_{x_k \leq x} p_k, \]
and of a continuous random variable $X$
\[ F(x) = \int_{-\infty}^{x} f(t) \, dt \]
$F(x)$ has the following properties:
- $F(x)$ is monotonically nondecreasing
- $F(-\infty) = 0$ and $F(\infty) = 1$.
- $F(x)$ is continuous to the left

1.2 Bivariate Random Variables and Distributions

A bivariate continuous random variable is a variable that takes a contin-
umum of values in the plane. The distribution of a bivariate continuous
random variable $(X,Y)$ can be characterized by a joint density function
or joint probability density function, $f(x,y)$. The nonnegative function
$f(x,y)$ is such that
\[ P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x,y) \, dy \, dx \]
defines the probability that $X$ and $Y$ take values in the interval $[x_1, x_2]$ and
$[y_1, y_2]$, respectively. Note that
\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \, dy \, dx = 1. \]

The marginal density function or marginal probability density function
is given by
\[ f(x) = \int_{-\infty}^{\infty} f(x,y) \, dy \]
such that
\[ P(x_1 \leq X \leq x_2) = P(x_1 \leq X \leq x_2, -\infty \leq Y \leq \infty) = \int_{x_1}^{x_2} f(x) \, dx. \]

The conditional density function or conditional probability density func-
tion with respect to the event \{Y = y\} is given by
\[ f(y|x) = \frac{f(x,y)}{f(x)} \]
provided that $f(x) > 0$. Note that
\[ \int_{-\infty}^{\infty} f(y|x) \, dy = 1. \]

Two random variables $X$ and $Y$ are called independent, if and only if
\[ f(x,y) = f(x) \cdot f(y) \]
If $X$ and $Y$ are independent, then:
- $f(y|x) = f(y)$
- $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(x_1 \leq X \leq x_2) \cdot P(y_1 \leq Y \leq y_2)$
More generally, if a finite set of $n$ continuous random variables $X_1, X_2, X_3, ..., X_n$ are mutually independent, then

$$f(x_1, x_2, x_3, ..., x_n) = f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot ... \cdot f(x_n).$$

### 2 Moments

#### 2.1 Expected Value or Mean

The expected value or mean of a discrete random variable with probability distribution $P(x_k)$ and $k = 1, 2, ..., K$ is defined as

$$E[X] = \sum_{k=1}^{K} x_k P(x_k)$$

if the series converges absolutely.

The expected value or mean of a continuous univariate random variable with density function $f(x)$ is defined as

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

if the integral exists.

For a random variable $Z$ which is a continuous function $\phi$ of a discrete random variable $X$, we have:

$$E[Z] = E[\phi(X)] = \sum_{k=1}^{K} \phi(x_k) P(x_k)$$

For a random variable $Z$ which is a continuous function $\phi$ of the continuous random variables $X$ and $Y$, we have:

$$E[Z] = E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) f(x) dx$$

$$E[Z] = E[\phi(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x,y) f(x,y) dx dy$$

#### 2.2 Variance and Standard Deviation

The variance of a univariate random variable $X$ is defined as

$$V[X] = E [(X - E[X])^2] = E[X^2] - (E[X])^2$$

The variance has the following properties:

- $V[X] \geq 0$
- $V[X] = 0$ if and only if $X = E[X]$

The following rules hold in general, i.e. for discrete, continuous and mixed types of random variables:

- $V[\alpha X + \beta Y] = \alpha^2 V[X]$
- $V[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} V[X_i]$ if $X_i$ and $X_j$ independent for all $i \neq j$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are constants.

Instead of the variance, one often considers the standard deviation

$$\sigma_X = \sqrt{V[X]}.$$
2.3 Higher order Moments

The $j$-th moment around zero is defined as

$$E[(X - E[X])^j].$$

2.4 Covariance and Correlation

The Covariance between two random variables $X$ and $Y$ is defined as:

$$\text{Cov}[X,Y] = E[(X - E[X])(Y - E[Y])].$$

$$= E[XY] - E[X]E[Y]$$

$$= E[(X - E[X])Y] = E[X(Y - E[Y])].$$

The following rules hold in general, i.e. for discrete, continuous and mixed types of random variables:

- $\text{Cov}[\alpha X + \gamma, \beta Y + \mu] = \alpha \beta \text{Cov}[X,Y]$ 
- $\text{Cov}[X_1 + X_2, Y_1 + Y_2]$
  $$= \text{Cov}[X_1, Y_1] + \text{Cov}[X_1, Y_2] + \text{Cov}[X_2, Y_1] + \text{Cov}[X_2, Y_2]$$
- $\text{Cov}[X,Y] = 0$ if $X$ and $Y$ are independent

where $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}$ and $\mu \in \mathbb{R}$ are constants.

The correlation coefficient between two random variables $X$ and $Y$ is defined as:

$$\rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sigma_X \sigma_Y}$$

where $\sigma_X$ and $\sigma_Y$ denote the corresponding standard deviations. The correlation coefficient has the following property:

- $-1 \leq \rho_{X,Y} \leq 1$

The following rule holds:

- $\rho_{\alpha X + \gamma, \beta Y + \mu} = \rho_{X,Y}$
- $\rho_{X,Y} = 0$ if $X$ and $Y$ are independent

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are constants.

2.5 Conditional Expectation and Variance

Let $(X,Y)$ be a bivariate discrete random variable and $P(y_k|X)$ the conditional probability of $Y = y_k$ given $X$. Then the conditional expected value or conditional mean of $Y$ given $X$ is

$$E[Y|X] = E_{Y|X}[Y] = \sum_{k=1}^{K} y_k P(y_k|X).$$

Let $(X,Y)$ be a bivariate continuous random variable and $f(y|x)$ the conditional density of $Y$ given $X$. Then the conditional expected value or conditional mean of $Y$ given $X$ is

$$E[Y|X] = E_{Y|X}[Y] = \int_{-\infty}^{\infty} y f(y|X)dy.$$
3 Random Vectors and Random Matrices

In this section we denote matrices (random or non-random) by bold capital letters, e.g. \( \mathbf{X} \) and vectors by small letters, e.g. \( \mathbf{x} \).

Let \( \mathbf{x} = (x_1, \ldots, x_n)' \) be a \((n \times 1)\)-dimensional vector such that each element \( x_i \) is a random variable. Let \( \mathbf{X} \) be a \((n \times k)\)-dimensional matrix such that each element \( x_{ij} \) is a random variable. Let \( \mathbf{a} = (a_1, \ldots, a_n)' \) be a \(n \times 1\)-dimensional vector of constants and \( \mathbf{A} \) a \((m \times n)\) matrix of constants.

The expectation of a random vector, \( \mathbb{E}[\mathbf{x}] \), and of a random matrix, \( \mathbb{E}[\mathbf{X}] \), summarizes the expected values of its elements, respectively:

\[
\mathbb{E}[\mathbf{x}] = \left( \begin{array}{c}
\mathbb{E}[x_1] \\
\mathbb{E}[x_2] \\
\vdots \\
\mathbb{E}[x_n]
\end{array} \right) \quad \text{and} \quad \mathbb{E}[\mathbf{X}] = \left( \begin{array}{ccc}
\mathbb{E}[x_{11}] & \mathbb{E}[x_{12}] & \cdots & \mathbb{E}[x_{1k}] \\
\mathbb{E}[x_{21}] & \mathbb{E}[x_{22}] & \cdots & \mathbb{E}[x_{2k}] \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}[x_{n1}] & \mathbb{E}[x_{n2}] & \cdots & \mathbb{E}[x_{nk}]
\end{array} \right).
\]

The following rules hold:

- \( \mathbb{E}[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\mathbb{E}[\mathbf{x}] \)
- \( \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] \)
- \( \mathbb{E}[\mathbf{A}\mathbf{X}] = \mathbf{A}\mathbb{E}[\mathbf{X}] \)
- \( \mathbb{E}[\mathbf{tr}(\mathbf{X})] = \mathbf{tr}(\mathbb{E}[\mathbf{X}]) \) for \( \mathbf{X} \) a quadratic matrix

The variance-Covariance matrix of a random vector, \( \mathbb{V}[\mathbf{x}] \), summarizes all variances and Covariances of its elements:

\[
\mathbb{V}[\mathbf{x}] = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])'] = \mathbb{E}[\mathbf{xx}'] - (\mathbb{E}[\mathbf{x}])(\mathbb{E}[\mathbf{x}])'
= \left( \begin{array}{cccc}
\mathbb{V}[x_1] & \text{Cov}[x_1, x_2] & \cdots & \text{Cov}[x_1, x_n] \\
\text{Cov}[x_2, x_1] & \mathbb{V}[x_2] & \cdots & \text{Cov}[x_2, x_n] \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}[x_n, x_1] & \text{Cov}[x_n, x_2] & \cdots & \mathbb{V}[x_n]
\end{array} \right).
\]

The following rules hold:

- \( \mathbb{V}[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\mathbb{V}[\mathbf{x}]\mathbf{a} \)
- \( \mathbb{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}[\mathbf{x}]\mathbf{A}' \)

where the \((m \times n)\) dimensional matrix \( \mathbf{A} \) with \( m \leq n \) has full row rank.

If the variance-Covariance matrix \( \mathbb{V}[\mathbf{x}] \) is positive definite (p.d.) then all random elements and all linear combinations of its random elements have strictly positive variance:

\[ \mathbb{V}[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\mathbb{V}[\mathbf{x}]\mathbf{a} > 0 \text{ for all } \mathbf{a} \neq 0. \]

4 Important Distributions

4.1 Univariate Normal Distribution

The density of the univariate normal distribution is given by:

\[
f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.
\]

The normal distribution is characterized by the two parameters \( \mu \) and \( \sigma \). The mean of the normal distribution is \( \mathbb{E}[\mathbf{X}] = \mu \) and the variance \( \mathbb{V}[\mathbf{X}] = \sigma^2 \). We write \( \mathbf{X} \sim N(\mu,\sigma^2) \).

The univariate normal distribution with mean \( \mu = 0 \) and variance \( \sigma^2 = 1 \) is called the standard normal distribution \( N(0,1) \).

4.2 Bivariate Normal Distribution

The density of the bivariate normal distribution is

\[
f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) \right] \right\}.
\]
If \((X, Y)\) follows a bivariate normal distribution, then:

- The marginal densities \(f(x)\) and \(f(y)\) are univariate normal.
- The conditional densities \(f(x|y)\) and \(f(y|x)\) are univariate normal.
- \(E[X] = \mu_X, V[X] = \sigma_X^2, E[Y] = \mu_Y, V[Y] = \sigma_Y^2\).
- The correlation coefficient between \(X\) and \(Y\) is \(\rho_{X,Y} = \rho\).
- \(E[Y|X] = \mu_Y + \rho_{X,Y} \sigma_Y \sigma_X^{-1} (X - \mu_X)\) and \(V[Y|X] = \sigma_Y^2 (1 - \rho^2)\).

The above properties characterize the normal distribution. It is the only distribution with all these properties.

Further important properties:

- If \((X, Y)\) follows a bivariate normal distribution, then \(aX + bY\) is also normally distributed:
  \[aX + bY \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho \sigma_X \sigma_Y).\]

  The reverse implication is not true.

- If \(X\) and \(Y\) are bivariate normally distributed with \(\text{Cov}[X, Y] = 0\), then \(X\) and \(Y\) are independent.

### 4.3 Multivariate Normal Distribution

Let \(\mathbf{x} = (x_1, \ldots, x_n)'\) be a \(n\)-dimensional vector such that each element \(x_i\) is a random variable. In addition let \(E[\mathbf{x}] = \mu = (\mu_1, \ldots, \mu_n)\) and \(V[\mathbf{x}] = \Sigma\) with

\[
\Sigma = \begin{pmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}
\end{pmatrix}
\]

where \(\sigma_{ij} = \text{Cov}[x_i, x_j]\).

A \(n\)-dimensional random variable \(\mathbf{x}\) is multivariate normally distributed with mean \(\mu\) and variance-Covariance matrix \(\Sigma\), \(\mathbf{x} \sim N(\mu, \Sigma)\) if its density is:

\[
f(\mathbf{x}) = (2\pi)^{-n/2} (\text{det} \, \Sigma)^{-1/2} \exp \left( -\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right).\]

Let \(\mathbf{x} \sim N(\mu, \Sigma)\) and \(\mathbf{A}\) a \((m \times n)\) matrix with \(m \leq n\) and \(m\) linearly independent rows then we have

\[\mathbf{Ax} \sim N(\mathbf{A} \mu, \mathbf{A} \Sigma \mathbf{A}').\]

### References

