2. Basic Properties of Random Vectors

2.1 CDF's and PDF's

- Let $\underset{\sim}{x} = (x_1, \dots, x_p)^T$ be a random vector
- The CDF associated with $\underset{\sim}{x}$ is given by:

$$F(\underset{\sim}{x}) = \mathbb{P}(\underset{\sim}{x} \leq \underset{\sim}{x}^\circ) = \mathbb{P}(x_1 \leq x_1^\circ, \dots, x_p \leq x_p^\circ)$$

- Two important cases of CDF's are:

1. **Absolute continuous:** We say that $\underset{\sim}{x}$ is absolute continuous if there exists a PDF associated with $\underset{\sim}{x}$ which satisfies

$$F(\underset{\sim}{x}) = \int_{-\infty}^{\underset{\sim}{x}} f(\underset{\sim}{u}) \, d\underset{\sim}{u}$$

where $d\underset{\sim}{u} = du_1 \dots du_p$ is the product of p differential elements.

Note that, for any measurable set $D \subseteq \mathbb{R}^p$

$$\mathbb{P}(\underset{\sim}{x} \in D) = \int_D f(\underset{\sim}{u}) \, d\underset{\sim}{u} \qquad \text{and}$$

$$\int_{-\infty}^{\infty} f(\underset{\sim}{u}) \, d\underset{\sim}{u} = 1$$

2. Discrete R.V. : For a discrete random vector $\underset{\sim}{x}$ the probability density is concentrated on a countable (or finite) set of points $\{x_j : j = 1, 2, \dots\}$. It's pdf is given by

$$\mathbb{P}(\underset{\sim}{x} = \underset{\sim}{x}_j) = \begin{cases} f(\underset{\sim}{x}_j), & j = 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

and the integrals in the absolute continuous case can be replaced by

$$\mathbb{P}(\underset{\sim}{x} \in D) = \sum_{j : x_j \in D} f(\underset{\sim}{x}_j)$$

- The support S of $\underset{\sim}{x}$ is defined as $S = \{\underset{\sim}{x} \in \mathbb{R}^p : f(\underset{\sim}{x}) > 0\}$

Marginal and conditional distributions

- Consider the following vector $\underset{\sim}{x}^T = (\underset{\sim}{x}_1^T, \underset{\sim}{x}_2^T)$, where $\underset{\sim}{x}_1$ has k elements and $\underset{\sim}{x}_2$ $(p-k)$ elements. $(k < p)$

$$\mathbb{P}(\underset{\sim}{x}_1 \le \underset{\sim}{x}_1^0) = F(x_1^0, x_2^0, \dots) \quad \text{marginal cdf of } \underset{\sim}{x}_1$$

$$f_1(\underset{\sim}{x}_1) = \int_{-\infty}^{\infty} f(\underset{\sim}{x}_1, \underset{\sim}{x}_2)\, d\underset{\sim}{x}_2 \quad \text{marginal pdf of } \underset{\sim}{x}_1$$

- For a given value $\underset{\sim}{x}_1 = \underset{\sim}{x}_1^\circ$, the conditional pdf of $\underset{\sim}{x}_2$ is given by

$$f(\underset{\sim}{x}_2 \mid \underset{\sim}{x}_1 = \underset{\sim}{x}_1^\circ) = \frac{f(\underset{\sim}{x}_1^\circ, \underset{\sim}{x}_2)}{f_1(\underset{\sim}{x}_1^\circ)}$$

conditional pdf

Independence

- When the conditional $f(\underset{\sim}{x}_2 \mid \underset{\sim}{x}_1 = \underset{\sim}{x}_1^\circ)$ is the same for every value of $\underset{\sim}{x}_1^\circ$, then we say that $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are statistically independent of each other

Theorem 2.1.1 If $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are statistically indep. then $f(\underset{\sim}{x}) = f_1(\underset{\sim}{x}_1)\, f_2(\underset{\sim}{x}_2) \Rightarrow f(\underset{\sim}{x}_2 \mid \underset{\sim}{x}_1 = \underset{\sim}{x}_1^\circ) = f_2(\underset{\sim}{x}_2)$.

2.2. Population moments

Expectation and correlation

- If $\underset{\sim}{x}$ is a random vector with pdf $f(\underset{\sim}{x})$, then

$$\mathbb{E}[g(\underset{\sim}{x})] = \int_{-\infty}^{\infty} g(\underset{\sim}{x})\, f(\underset{\sim}{x})\, d\underset{\sim}{x}$$

expected value of a scalar-valued function

- We assume that all necessary integrals converge, so expectations are finite
- We have the following properties of the $\mathbb{E}[\cdot]$:

1. Linearity

$$\mathbb{E}[a_1 g_1(\underset{\sim}{x}) + a_2 g_2(\underset{\sim}{x})] = a_1 \mathbb{E}[g_1(\underset{\sim}{x})] + a_2 \mathbb{E}[g_2(\underset{\sim}{x})]$$

2. Partition $\underset{\sim}{x}^T = (\underset{\sim}{x}_1^T, \underset{\sim}{x}_2^T)$. The expectation of a function of $\underset{\sim}{x}_1$ may be written in terms of the marginal distribution

$$\mathbb{E}[g(\underset{\sim}{x}_1)] = \int_{-\infty}^{\infty} g(\underset{\sim}{x}_1) f(\underset{\sim}{x}) d\underset{\sim}{x} = \int_{-\infty}^{\infty} g(\underset{\sim}{x}_1) f_1(\underset{\sim}{x}_1) d\underset{\sim}{x}_1$$

when f_1 is known.

3. If $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are independent and $g_i(\underset{\sim}{x}_i)$ is a function of $\underset{\sim}{x}_i$ only $(i=1,2)$, then

$$\mathbb{E}[g_1(\underset{\sim}{x}_1) g_2(\underset{\sim}{x}_2)] = \mathbb{E}[g_1(\underset{\sim}{x}_1)] \mathbb{E}[g_2(\underset{\sim}{x}_2)]$$

- More generally, the expectation of a matrix-valued (or vector-valued) function of $\underset{\sim}{x}$, $G(\underset{\sim}{x}) = (g_{ij}(\underset{\sim}{x}))$ is defined as the following matrix

$$\mathbb{E}[G(\underset{\sim}{x})] = (\mathbb{E}[g_{ij}(\underset{\sim}{x})])$$

Population mean vector and covariance matrix

- The vector $\mathbb{E}[\underset{\sim}{x}] = \underset{\sim}{\mu}$ is called the population mean vector of $\underset{\sim}{x}$

$$\mu_i = \int_{-\infty}^{\infty} x_i f(\underset{\sim}{x})\, d\underset{\sim}{x} \quad , \; i = 1, \ldots, p$$

which possesses the linearity property

$$\mathbb{E}[A\underset{\sim}{x} + \underset{\sim}{b}] = \underset{q \times p}{A}\, \mathbb{E}[\underset{\sim}{x}] + \underset{q \times 1}{\underset{\sim}{b}}$$

- The matrix

$$\mathbb{E}[(\underset{\sim}{x} - \underset{\sim}{\mu})(\underset{\sim}{x} - \underset{\sim}{\mu})^T] = \Sigma = V(\underset{\sim}{x})$$

is called the covariance matrix of $\underset{\sim}{x}$.

- For conciseness, we write

$$\underset{\sim}{x} \sim (\underset{\sim}{\mu}, \Sigma)$$

to describe a random vector.

- More generally, we can define the covariance between two random vectors as

$$C(\underset{\sim}{x}, \underset{\sim}{y}) = \mathbb{E}[(\underset{\sim}{x} - \mu_{\underset{\sim}{x}})(\underset{\sim}{y} - \mu_{\underset{\sim}{y}})^T]$$

- Note the following properties of the covariance matrix

1. $\sigma_{ij} = C(x_i, x_j)$, $i \neq j$

   $\sigma_{ii} = V(x_i) = \sigma^2_i$

2. $\Sigma = \mathbb{E}[\underset{\sim}{x}\underset{\sim}{x}^T] - \underset{\sim}{\mu}\,\underset{\sim}{\mu}^T$

3. $V(\underset{\sim}{a}^T\underset{\sim}{x}) = \underset{\sim}{a}^T V(\underset{\sim}{x})\,\underset{\sim}{a}$

4. $\Sigma \geqslant 0$

5. $V(A\underset{\sim}{x} + \underset{\sim}{b}) = A\,V(\underset{\sim}{x})\,A^T$

6. $C(\underset{\sim}{x}, \underset{\sim}{x}) = V(\underset{\sim}{x})$

7. $C(\underset{\sim}{x}, \underset{\sim}{y}) = C(\underset{\sim}{y}, \underset{\sim}{x})$

8. $C(\underset{\sim}{x}_1 + \underset{\sim}{x}_2, \underset{\sim}{y}) = C(\underset{\sim}{x}_1, \underset{\sim}{y}) + C(\underset{\sim}{x}_2, \underset{\sim}{y})$

9. If $p = q$, $V(\underset{\sim}{x} + \underset{\sim}{y}) = V(\underset{\sim}{x}) + C(\underset{\sim}{x}, \underset{\sim}{y}) + C(\underset{\sim}{y}, \underset{\sim}{x}) + V(\underset{\sim}{y})$

10. $C(A\underset{\sim}{x}, B\underset{\sim}{y}) = A\,C(\underset{\sim}{x}, \underset{\sim}{y})\,B^T$

11. If $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are independent, then $C(\underset{\sim}{x}, \underset{\sim}{y}) = 0$, however the converse is not true!

Correlation matrix

- Denote the correlation coefficient as follows

$$\rho_{ij} = \sigma_{ij} / \sigma_i \sigma_j \quad , \quad i \neq j$$

and the correlation matrix

$$P = (\rho_{ij}) \quad , \quad \text{where} \quad \rho_{ii} = 1$$ pop. correl. matrix

- Let $\Delta = \text{diag}(\sigma_i)$, then

$$P = \Delta^{-1} \Sigma \Delta^{-1}$$

- Since $\Sigma \geq 0$ and Δ is symmetric $\Rightarrow P \geq 0$
- Furthermore,

$|\Sigma|$ generalized variance

$\text{trace}(\Sigma)$ Total variance

Mahalanobis space

- Recall that if x and y are two points in space, then

$$\Delta^2_{\Sigma}(x, y) = (x - y)^T \Sigma^{-1} (x - y)$$ Mahalanobis distance pop. ?

Conditional Moments

- The **regression curve** of $\underset{\sim}{x}_1$ on $\underset{\sim}{x}_2$ is defined by

  $E[\underset{\sim}{x}_1 | \underset{\sim}{x}_2]$ conditional mean (expected) value

  defined on the support of $\underset{\sim}{x}_2$.

- If $E[\underset{\sim}{x}_1 | \underset{\sim}{x}_2]$ is linear on $\underset{\sim}{x}_2$, then the regression is called **linear**

- The **conditional variance** is given by

  $V(\underset{\sim}{x}_1 | \underset{\sim}{x}_2)$

  which defines the **scedastic curve** of $\underset{\sim}{x}_1$ on $\underset{\sim}{x}_2$.

- The regression of $\underset{\sim}{x}_1$ on $\underset{\sim}{x}_2$ is called **homoscedastic** if $V(\underset{\sim}{x}_1 | \underset{\sim}{x}_2)$ is a constant matrix

Example

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & , \ 0 \le x_1, x_2 \le 1 \\ 0 & , \ \text{otherwise} \end{cases}$$

$\underset{\sim}{\mu}$

Σ

$E[x_1 | x_2]$

$V(x_1 | x_2)$

Q

$$E[x_1] = \int_0^1 x_1 f_1(x_1) \, dx_1$$

$$f_1(x_1) = \int_0^1 f(x_1, x_2) \, dx_2 = \int_0^1 x_1 + x_2 \, dx_2$$

$=$

2.5 The Multinomial Distributions

- The most important multivariate distribution is the multivariate normal.

$$X \sim N(\mu, \sigma^2) \Rightarrow f(x) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right\}$$

$$\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma) \Rightarrow f(\underset{\sim}{x}) = |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right\} \quad *$$

where $\Sigma > 0$

Def (p-variate normal dist.) The random vector $\underset{\sim}{x}$ is said to have a p-variate normal dist. with mean vector $\underset{\sim}{\mu}$ and covariance matrix Σ if its p.d.f. is given by $*$.

- The p.d.f. of the p-variate normal dist may also be written in terms of correlations

Theorem 2.5.1 Let $\underset{\sim}{x}$ have the p.d.f. given by $*$, and let

$$\underset{\sim}{y} = \Sigma^{-1/2}(\underset{\sim}{x}-\underset{\sim}{\mu}) \sim N(\underset{\sim}{0}, I)$$

Then, $y_1, \dots, y_p$ are i.i.d. $N(0,1)$ variables.

Corollary 2.5.1.1 If $\underset{\sim}{x}$ has p.d.f. given by $*$, then

$$E[\underset{\sim}{x}] = \underset{\sim}{\mu} \qquad V(\underset{\sim}{x}) = \Sigma$$

Geometry

- The multivariate normal distribution in p dimensions has constant density on ellipses or ellipsoids of the form

$$(\underset{\sim}{x}-\underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu}) = c^2$$

where c is a constant.

- For $\underset{\sim}{\mu} = \underset{\sim}{0}$ and $\Sigma = I$, the ellipsoids are hyperspheres

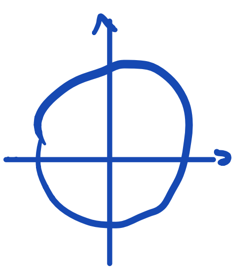
- The principal component transformation facilitates the interpretation of the ellipsoids of equal concentrations

- Using the spectral decomposition theorem write

$$\Sigma = \Gamma \Lambda \Gamma^T$$

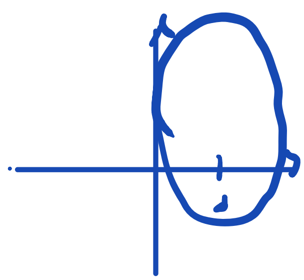
Define the principal component transformation by

$$\underset{\sim}{y} = \Gamma^T(\underset{\sim}{x}-\underset{\sim}{\mu}) \Rightarrow \sum_{i=1}^{p} y_i^2/\lambda_i = c^2$$



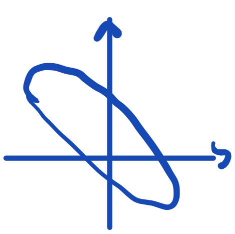
$\underset{\sim}{\mu} = (0,0)^T$

$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$



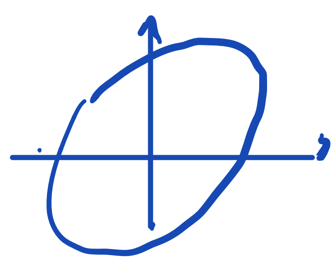
$\underset{\sim}{\mu} = (1,1)$

$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$



$\underset{\sim}{\mu} = (0,0)^T$

$\Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$



$\underset{\sim}{\mu} = (0,0)^T$

$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Properties

- If $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then we have the following results

Theorem 2.5.2 $U = (\underset{\sim}{x} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}) \sim \chi^2_p$

Theorem 2.5.3 All non-trivial linear combinations of the elements of $\underset{\sim}{x}$ are univariate normal