MODULE 1.5
Second-order properties of random variables and vectors

by
S.Lakshmivarahan
School of Computer Science
University of Oklahoma
Norman, OK-73019, USA
varahan@ou.edu

L_2 - space of square integrable random variables

- Let (Ω, Γ, P) be a probability space
- $L_2 = L_2(\Omega, \Gamma, P)$ denote the family of square integrable random variables
- Say that $x(\omega) \in L_2$ if

$$\int_\Omega |x(\omega)|^2 dP(\omega) < \infty \tag{1}$$

- L_2 is a real vector space-closed under addition and multiplication by a real constant

- Let $x \in L_2$, with mean $\mu_x = E[x] < \infty$
- Variance of x: $var(x) = \sigma_x^2 = E[(x - \mu)^2] < \infty$
- Covariance between $x, y \in L_2$:

$$cov(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

- x, y are uncorrelated if cov(x, y)$= 0$
- Correlation between $x, y \in L_2$:

$$corr(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

- $|corr(x, y)| \leq 1$

A Geometric view of random variables in L_2

- Let $x, y \in L_2$
- Inner product of x and y: $< x, y >= E(xy)$
- Norm of x: $||x|| =< x, x >^{\frac{1}{2}} = [E(x^2)]^{\frac{1}{2}}$
- Distance between x and y:

$$dist(x, y) = ||x - y|| = [E(x - y)^2]^{\frac{1}{2}}$$

- x and y are orthogonal if $< x, y >= E(x, y) = 0$
- For mean zero random variables: orthogonality implies uncorrelated

Orthogonal projection in L_2

- Let x be a random variable defined on (Ω, Γ, P)
- Let y be a random variable on a subspace (Ω, Y, P) where Y is a sub σ-field of Γ
- Orthogonal projection theorem: For $x \in L_2(\Omega, \Gamma, P)$ there exists an unique $\hat{x} \in L_2(\Omega, Y, P)$ such that
  $(a) ||x - \hat{x}|| = min\{||x - y|| : y \in L_2(\Omega, Y, P)\}$
  $(b) <x - \hat{x}, y> = 0$ for all $y \in L_2(\Omega, Y, P)$

- Let $x \in R^m$ be a random vector with $x = (x_1, x_2, x_3, \ldots, x_m)$
- Say $x \in L_2$ if each component $x_i \in L_2$
- Mean of $x = \mu_x = E(x) = (\mu_1, \mu_2, \mu_3, \ldots, \mu_m)^T$ where $\mu_i = E(x_i)$
- $cov(x_i, x_j) = \sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$
- $var(x_i) = \sigma_i^2 = E[(x_i - \mu_i)^2]$
- $cov(x) = E[(x - \mu)(x - \mu)^T] = [\sigma_{ij}] = \Sigma \in R^{m \times m}$
- $var(x) = \sum_{i=1}^{n} var(x_i)$
  $= E[(x - \mu)^T (x - \mu)] = \sum_{i=1}^{n} \sigma_i^2 = tr(\Sigma)$
  where $tr(A)$ is called the trace of A.

A geometric view of random vectors in L_2

- Let $x, y \in R^m$ be two random vectors in L_2
- Inner product: $< x, y >= E[x^T y] = \sum_{i=1}^{m} E(x_i y_i)$
- Norm: $||x||^2 =< x, x >= E[x^T x] = \sum_{i=1}^{m} E(x_i)^2$
- Distance: $||x - y||^2 =< x - y, x - y >$
  $= E[(x - y)^T (x - y)] = \sum_{i=1}^{m} E(x_i - y_i)^2$
- Orthogonal: x and y are orthogonal if $< x, y >= 0$
- For mean zero random vectors orthogonality implies uncorrelated

Orthogonal projection

- The statement of orthogonal projection theorem carries over verbatim if we replace random variables by random vectors
- This projection theorem is the basis for generating optimal prediction, optimal estimation in Time Series Analysis, Spatial and Spatio-temporal statistics.
- It also plays a key role in the Principal Component Analysis (PCA) and in the development of Empirical orthogonal functions (EOF)

Centering

- Let $x \in R^m$ be a random vector with mean $\mu \in R^m$ and $cov(x) = \Sigma \in R^{m \times m}$
- Then, $y = x - \mu$ is called the centered version of x
- Clearly: E(y) $= 0$ and $cov(y) = \Sigma$

Normalization of $x \in R^n$

- Let $x \in R^m$ be a random vector with mean $\mu \in R^m$ and $cov(x) = \Sigma$ with $y = x - \mu$
- $z_i = \frac{x_i - \mu_i}{\sigma_i} = \frac{y_i}{\sigma_i}$ is the normalized version of y_i
- Mean$(z_i) = 0$, Var$(z_i) = 1$
- $z = (z_1, z_2, \ldots, z_m)^T$ is the centered and normalized version of x

- Let $D = Diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ - the diagonal matrix with the diagonal of Σ
- Define square root, $D^{1/2} : D = D^{1/2} D^{1/2}$ where $D^{1/2} = Diag(\sigma_1, \sigma_2, \ldots, \sigma_m)$
- Define $z = D^{-1/2} Y = D^{-1/2}(x - \mu)$
- $cov(z) = E(zz^T) = D^{-1/2} E[(x-\mu)(x-\mu)^T] D^{-1/2}$ $= D^{-1/2} \Sigma D^{-1/2} = R = corr(z)$
- Correlation matrix: $R = [R_{ij}]$ and $R_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$
- $|R_{ij}| \leq 1$

- Let $A \in R^{m \times m}$ and $b \in R^m$
- Define $\xi = Ax + b$
- Mean: $E(\xi) = A\mu + b$ where $\mu = E(x)$
- $$\begin{aligned} cov(\xi) &= E[(\xi - E(\xi))(\xi - E(\xi))^T] \\ &= E[(A(x-\mu))(A(x-\mu))^T] \\ &= AE[(x-\mu)(x-\mu)^T]A^T = A\Sigma A^T \end{aligned}$$
- Thus, if $x \sim N(m, \Sigma)$, $\xi \sim N(Am + b, A\Sigma A^T)$

A special linear transformation

- Let $x \in R^m$ with mean μ and $cov(x) = \Sigma$, SPD
- Define square root of $\Sigma : \Sigma = \Sigma^{1/2}\Sigma^{1/2}$
- let $\xi = \Sigma^{-1/2}(x - \mu)$
- Then $E(\xi) = 0$
- $$\begin{aligned} cov(\xi) &= E[(\Sigma^{-1/2}(x-\mu))(\Sigma^{-1/2}(x-\mu))^T] \\ &= \Sigma^{-1/2}E[(x-\mu)(x-\mu)^T]\Sigma^{-1/2} \\ &= \Sigma^{-1/2}\Sigma\Sigma^{-1/2} = I \end{aligned}$$
- That is: $var(\xi_i) = 1$ and $cov(\xi_i, \xi_j) = 0$ for i $\neq$ j
- This is known as Whitening transformation

Linear functional of x

- Let $a \in R^m$ and $x \in R^m$ with mean μ and $cov(x)$
- Define $\eta = a^T x$, a real random variable
- $E(\eta) = a^T \mu$
- $$\begin{aligned} var(\eta) &= E[(a^T(x-\mu))^2] = E[(a^T(x-\mu))(a^T(x-\mu))] \\ &= E[a^T(x-\mu)(x-\mu)^T a^T] \\ &= a^T \Sigma a \end{aligned}$$
- Clearly, η is a non-degenerate random variable (that is, var(η) $>$0) for all $a \in R^n$, if and only if Σ is SPD

References

- M.Grigoriu (2002) Stochastic Calculus, Brikhauser, Basel contains a good introduction to basic Probability theory and L_2 spaces

Exercise

1. Prove that $|R_{ij}| = |\frac{\sigma_{ij}}{\sigma_i \sigma_j}|$