Probability and Stochastic Processes I - Lecture 16

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

III.4 Expectations for Random Vectors

Definition III.4.1 For random vector $\mathbf{X} \in R^k$, the *mean vector* of $\mathbf{X}$ is

$$\begin{aligned} \mu_{\mathbf{X}} &= E(\mathbf{X}) = \left(\; E(X_1), \quad E(X_2), \quad \ldots, \quad E(X_k) \; \right)' \\ &= \left(\; \mu_1, \quad \mu_2, \quad \ldots, \quad \mu_k \; \right)' \end{aligned}$$

provided each $E(X_i) = \mu_i$ exists. If each $E(X_i)$ is finite (so $E(\mathbf{X}) \in R^k$) then the *variance matrix* of $\mathbf{X}$ is given by

$$\begin{aligned} \Sigma_{\mathbf{X}} &= Var(\mathbf{X}) \\ &= \begin{pmatrix} E\left((X_1 - \mu_1)^2\right) & \cdots & E\left((X_1 - \mu_1)(X_k - \mu_k)\right) \\ E\left((X_2 - \mu_2)(X_1 - \mu_1)\right) & \cdots & E\left((X_2 - \mu_2)(X_k - \mu_k)\right) \\ \vdots & \vdots & \vdots \\ E\left((X_k - \mu_k)(X_1 - \mu_1)\right) & \cdots & E\left((X_k - \mu_k)^2\right) \end{pmatrix} \end{aligned}$$

provided each $E\left((X_i - \mu_i)(X_j - \mu_j)\right)$ for $i \neq j$ exists. ■

notes

1. If $E|X_i| < \infty$ for $i = 1, , k$ then $\boldsymbol{\mu}_{\mathbf{X}} \in R^k$.

2. The *covariance* between r.v.'s X_i and X_j is defined by

$$Cov(X_i, X_j) = E\left((X_i - \mu_i)(X_j - \mu_j)\right)$$

and so $Cov(X_i, X_j) = Cov(X_j, X_i)$ (Cov is symmetric) and $Cov(X_i, X_i) = Var(X_i)$ and so

$$\Sigma_{\mathbf{X}} = (Cov(X_i, X_j))$$

where we have written the matrix in terms of its (i, j)-th element.

3. If $Cov(X_i, X_j)$ is finite for every i and j, then $\Sigma_{\mathbf{X}} \in R^{k\times k}$ and it is symmetric.

4. If $X = (X_{ij}) \in R^{k\times l}$ is a matrix of r.v.'s, then the expected value of this random matrix is defined to be $E(X) = (E(X_{ij}))$ when each $E(X_{ij})$ exists and $E(X) \in R^{k\times l}$ when each $E(X_{ij})$ is finite.

5.

$$
\begin{aligned}
\Sigma_{\mathbf{X}} &= Var(\mathbf{X}) \\
&= E\begin{pmatrix} (X_1-\mu_1)^2 & \cdots & (X_1-\mu_2)(X_k-\mu_k) \\ (X_2-\mu_2)(X_1-\mu_1) & \cdots & (X_2-\mu_2)(X_k-\mu_k) \\ \vdots & \vdots & \vdots \\ (X_k-\mu_k)(X_1-\mu_1) & \cdots & (X_k-\mu_k)^2 \end{pmatrix} \\
&= E\left((\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})'\right)
\end{aligned}
$$

6. **Exercise III.4.1** When X is a r.v. prove that $E(X^2) < \infty$ implies that $E(X)$ is finite. When X and Y are r.v.'s and $E(X^2) < \infty, E(Y^2) < \infty$ prove that $E(XY)$ is finite. Use these results to prove that if $E(X_i^2) < \infty$ for all $i = 1, \ldots, k$, then $\Sigma_{\mathbf{X}} \in R^{k \times k}$.

7. **Exercise III.4.2** When r.v.'s X and Y satisfy $E(X^2) < \infty, E(Y^2) < \infty$ prove that $Cov(X, Y) = E(XY) - E(X)E(Y)$. Extend this result to random vectors $\mathbf{X}$ to show that $\Sigma_{\mathbf{X}} = Var(\mathbf{X}) = E(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{X}}'$.

Proposition III.4.1 Suppose $\mathbf{X} \in R^k$ is a random vector and $\mathbf{Y} = \mathbf{a} + C\mathbf{X}$ where $\mathbf{a} \in R^l$, $C \in R^{l \times k}$ are constant.

(i) If $\mu_{\mathbf{X}} \in R^k$, then $\mu_{\mathbf{Y}} = \mathbf{a} + C\mu_{\mathbf{X}} \in R^l$.

(ii) If $\Sigma_{\mathbf{X}} \in R^{k \times k}$, then $\Sigma_{\mathbf{Y}} = C\Sigma_{\mathbf{X}}C' \in R^{l \times l}$.

Proof: (i) $\mu_{\mathbf{Y}} = E(\mathbf{Y}) = E(\mathbf{a} + C\mathbf{X}) = \mathbf{a} + \mathbf{C}E(\mathbf{X})$ since $E(a_i + \sum_{j=1}^k c_{ij}X_j) = a_i + \sum_{j=1}^k c_{ij}E(X_j)$ by the linearity of E (using the fact here that $E(X_j) \in R^1$) which establishes the result.

(ii)

$$\begin{aligned}
\Sigma_{\mathbf{Y}} &= Var(\mathbf{Y}) = E((\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y} - \mu_{\mathbf{Y}})') \\
&= E((\mathbf{a} + C\mathbf{X} - (\mathbf{a} + \mathbf{C}\mu_{\mathbf{X}}))(\mathbf{a} + C\mathbf{X} - (\mathbf{a} + \mathbf{C}\mu_{\mathbf{X}}))') \\
&= E(C(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})'C') \\
&= CE((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})'C') \text{ using linearity of } E \\
&= CE((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})')C' \text{ using linearity of } E \\
&= C\Sigma_{\mathbf{X}}C'. \blacksquare
\end{aligned}$$

Proposition III.4.2 (i) If X is a r.v. and $Var(X) = 0$, then

$$P(X = \mu_X) = 1$$

so X has a probability distribution degenerate at a constant.

(ii) If $\mathbf{X} \in R^k$ is a random vector $\Sigma_{\mathbf{X}} \in R^{k \times k}$ and $\mathbf{c} \in R^k$ is constant then $\mathbf{c}'\Sigma_{\mathbf{X}}\mathbf{c} \geq 0$. So any variance matrix is *positive semidefinite* (p.s.d.).

(iii) If $\mathbf{c}'\Sigma_{\mathbf{X}}\mathbf{c} = 0$ for some $\mathbf{c} \neq \mathbf{0}$, then the probability distribution of $\mathbf{X}$ is concentrated on the affine plane $\boldsymbol{\mu}_{\mathbf{X}} + L^{\perp}\{\mathbf{c}\}$.

Proof: (i) $Var(X) = E((X - \mu_X)^2) = 0$ iff
$1 = P((X - \mu_X)^2 = 0) = P(X - \mu_X = 0) = P(X = \mu_X)$.

(ii) Consider r.v. $Y = \mathbf{c}'\mathbf{X}$. Then, by Prop. III.4.1(ii), $Var(Y) = \mathbf{c}'\Sigma_{\mathbf{X}}\mathbf{c} \geq 0$ since a variance is always nonnegative.

(iii) Suppose $\mathbf{c}'\Sigma_{\mathbf{X}}\mathbf{c} = 0$ and consider $Y = \mathbf{c}'\mathbf{X}$. Then by (i) and (ii)

$$\begin{aligned} 1 &= P(Y = \mu_Y) = P(\mathbf{c}'\mathbf{X} = \mathbf{c}'\boldsymbol{\mu}_{\mathbf{X}}) = P(\mathbf{c}'(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) = 0) \\ &= P(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}} \in L^{\perp}\{\mathbf{c}\}) = P_{\mathbf{X}}(\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{L}^{\perp}\{\mathbf{c}\}). \blacksquare \end{aligned}$$

note

- since $\Sigma_{\mathbf{X}} \in R^{k\times k}$ is p.s.d. then the spectral decomposition gives $\Sigma_{\mathbf{X}} = Q\Lambda Q'$ where $Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_k \end{pmatrix} \in R^{k\times k}$ is orthogonal and $\Lambda = diag(\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$

- if $\mathbf{c} \in R^k$ then $\mathbf{c} = Q\mathbf{a} = \sum_{i=1}^k a_i \mathbf{q}_i$ and so $0 \leq \mathbf{c}'\Sigma\mathbf{c} = \sum_{i=1}^k \lambda_i a_i^2$ and $\mathbf{c}'\Sigma\mathbf{c} = 0$ iff $a_i = 0$ whenever $\lambda_i > 0$

- therefore, if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l > 0$ and $\lambda_{l+1} = \cdots = \lambda_k = 0$, then $\mathbf{c}'\Sigma\mathbf{c} = 0$ iff $\mathbf{c} \in L\{\mathbf{q}_{l+1}, \ldots, \mathbf{q}_k\}$

- this implies $P_{\mathbf{X}}(\mu_{\mathbf{X}} + L\{\mathbf{q}_1, \ldots, \mathbf{q}_l\}) = 1$

- so $\Sigma_{\mathbf{X}}$ is p.d. iff $\lambda_k > 0$ which holds iff $\Sigma_{\mathbf{X}}$ is invertible

Exercise III.4.3 Prove that, if $X \in R^{k\times l}$ is a random matrix such that each $E(X_{ij})$ is finite and $A \in R^{p\times q}, B \in R^{p\times k}, C \in R^{l\times q}$ are fixed, then $E(A + BXC) = A + BE(X)C$.

Example III.4.1 $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$

- consider first $\mathbf{Z} \sim N_k(\mathbf{0}, I)$ which has density

$$(2\pi)^{-k/2} \exp(-\mathbf{z}'\mathbf{z}/2) = \prod_{i=1}^{k} (2\pi)^{-1/2} \exp(-z_i^2/2)$$

and so $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} N(0,1)$ which implies $E(Z_i) = 0, Var(Z_i) = 1$

- also, when $i \neq j$

$$Cov(Z_i, Z_j) = E(Z_i Z_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{z_i z_j}{2\pi} \exp(-(z_i^2 + z_j^2)/2)\, dz_i\, dz_j$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} z_i (2\pi)^{-1/2} \exp(-z_i^2/2)\, dz_i \int_{-\infty}^{\infty} z_j (2\pi)^{-1/2} \exp(-z_j^2/2)\, dz_j \\ &= E(Z_i)E(Z_j) = 0 \end{aligned}$$

- so $E(\mathbf{Z}) = \mathbf{0}, Var(\mathbf{Z}) = I$

- if $\Sigma = \Sigma^{1/2}\Sigma^{1/2}, \mathbf{Z} \sim N_k(\mathbf{0}, I)$ then $\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Z} \sim N_k(\boldsymbol{\mu}, \Sigma)$ and

$$E(\mathbf{X}) = \boldsymbol{\mu} + \Sigma^{1/2} E(\mathbf{Z}) = \boldsymbol{\mu}, \quad Var(\mathbf{X}) = \Sigma^{1/2} Var(\mathbf{Z}) \Sigma^{1/2} = \Sigma^{1/2}\Sigma^{1/2} = \Sigma$$

Exercise III.4.4 Suppose $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$. Determine $E(\mathbf{X}'\mathbf{X})$.

Exercise III.4.5 Suppose $\mathbf{X} \sim \text{multinomial}(n, p_1, \ldots, p_k)$. Determine $\boldsymbol{\mu}_{\mathbf{X}}$ and $\Sigma_{\mathbf{X}}$.

Exercise III.4.6 The correlation between r.v.'s X and Y is defined by

$$\rho_{XY} = Corr(X, Y) = \frac{Cov(X, Y)}{Sd(X)Sd(Y)}$$

where $Sd(X) = \sqrt{Var(X)}$ is the *standard deviation* of X.
(i) What has to hold for ρ_{XY} to exist and provide sufficient conditions.
(ii) Prove that for constants a, b, c, d then

$$Corr(a + bX, c + dY) = Corr(X, Y)$$

provided $b > 0, d > 0$. What happens when $b = 0$? What happens when $b < 0, d > 0$ and when $b < 0, d < 0$?
(iii) Suppose $Y \overset{wp1}{=} a + bX$. What is $Corr(X, Y)$?
(iv) Suppose $X \sim U(0, 1)$ and $Y = X^2$. Determine $Corr(X, Y)$.
(v) Suppose $X \sim U(-1, 1)$ and $Y = X^2$. Determine $Corr(X, Y)$. Are X and Y independent?

- if we have two collections of r.v.'s $\{X_s : s \in S\}, \{Y_t : t \in T\}$ then these collections are statistically independent if for any finite subsets $\{s_1, \ldots, s_m\} \subset S, \{t_1, \ldots, t_n\} \subset T$, the joint cdf satisfies

$$\begin{aligned} & F_{(X_{s_1},\ldots,X_{s_m},Y_{t_1},\ldots,Y_{t_n})}(x_1, \ldots, x_m, y_1, \ldots, y_n) \\ = \; & F_{(X_{s_1},\ldots,X_{s_m})}(x_1, \ldots, x_m) F_{(Y_{t_1},\ldots,Y_{t_n})}(y_1, \ldots, y_n) \end{aligned}$$

for all $x_1, \ldots, x_m, y_1, \ldots, y_n \in R^1$

- recall that the Extension Thm then implies

$$P_{(X_{s_1},\ldots,X_{s_m},Y_{t_1},\ldots,Y_{t_n})}(B_1 \times B_2) = P_{(X_{s_1},\ldots,X_{s_m})}(B_1) P_{(Y_{t_1},\ldots,Y_{t_n})}(B_2)$$

for any $B_1 \in \mathcal{B}^m, B_2 \in \mathcal{B}^n$

Proposition III.5.1 If $\mathbf{X} \in R^k$ and $\mathbf{Y} \in R^l$ are statistically independent independent random vectors and

$$h_1 : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1), h_2 : (R^l, \mathcal{B}^l) \to (R^1, \mathcal{B}^1)$$

then $h_1(\mathbf{X})$ and $h_2(\mathbf{Y})$ are statistically independent and when $E(h_1^2(\mathbf{X})) < \infty$, $E(h_2^2(\mathbf{Y})) < \infty$ then

$$E(h_1(\mathbf{X})h_2(\mathbf{Y})) = E(h_1(\mathbf{X}))E(h_2(\mathbf{Y})).$$

Proof: We have

$$\begin{aligned} & F_{(h_1(\mathbf{X}),h_2(\mathbf{Y}))}(x, y) = P(h_1(\mathbf{X}) \le x, h_2(\mathbf{Y}) \le y) \\ = \; & P(\mathbf{X} \in h_1^{-1}(-\infty, x], \mathbf{Y} \in h_2^{-1}(-\infty, y]) \\ = \; & P_{(\mathbf{X},\mathbf{Y})}(h_1^{-1}(-\infty, x] \times h_2^{-1}(-\infty, y]) \\ = \; & P_{\mathbf{X}}(h_1^{-1}(-\infty, x])P_{\mathbf{Y}}(h_2^{-1}(-\infty, y]) = F_{h_1(\mathbf{X})}(x)F_{h_2(\mathbf{Y})}(y) \end{aligned}$$

for every x and y so $h_1(\mathbf{X})$ and $h_2(\mathbf{Y})$ are statistically independent.

Suppose $h_1 = \sum_i a_i I_{A_i}, h_2 = \sum_j b_j I_{b_j}$ are simple functions. Then

$$h_1(\mathbf{x})h_2(\mathbf{y}) = \sum_{i,j} a_i b_j I_{A_i}(\mathbf{x}) I_{B_j}(\mathbf{y}) = \sum_{i,j} a_i b_j I_{A_i \times B_j}(\mathbf{x}, \mathbf{y})$$

is also simple and

$$\begin{aligned} E(h_1(\mathbf{X})h_2(\mathbf{Y})) &= \sum_{i,j} a_i b_j P_{(\mathbf{X},\mathbf{Y})}(A_i \times B_j) \\ &= \sum_{i,j} a_i b_j P_{\mathbf{X}}(A_i) P_{\mathbf{Y}}(B_j) \\ &= \sum_{i} a_i P_{\mathbf{X}}(A_i) \sum_{j} b_j P_{\mathbf{Y}}(B_j) \\ &= E(h_1(\mathbf{X}))E(h_2(\mathbf{Y})) \end{aligned}$$

as required. The result then follows by proceeding to nonnegative h_1, h_2 by limits and then to general $h_1 = h_{1+} - h_{1-}, h_2 = h_{2+} - h_{2-}$. ■

Corollary III.5.2 $Cov(h_1(\mathbf{X}), h_2(\mathbf{Y})) = 0$.

Proof: **Exercise III.5.1**

Exercise III.5.2 For random vectors $\mathbf{X} \in R^k$ and $\mathbf{Y} \in R^l$ define $Cov(\mathbf{X}, \mathbf{Y}) = E((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})')$ provided all the relevant expectations exist.
(i) Give conditions under which $Cov(\mathbf{X}, \mathbf{Y}) \in R^{k \times l}$.
(ii) Assuming $Cov(\mathbf{X}, \mathbf{Y}) \in R^{k \times l}$ and $\mathbf{a} \in R^p, \mathbf{b} \in R^q, A \in R^{p \times k}, B \in R^{q \times l}$ are constant then determine $Cov(\mathbf{a} + A\mathbf{X}, \mathbf{b} + B\mathbf{Y})$.
(iii) Assuming $Cov(\mathbf{X}, \mathbf{Y}) \in R^{k \times l}$ and $\mathbf{X}$ and $\mathbf{Y}$ are statistically independent, then determine $Cov(\mathbf{X}, \mathbf{Y})$.

Exercise III.5.3 For random vector $\mathbf{X} \in R^k$ with $\Sigma_{\mathbf{X}} \in R^{k \times k}$ the *correlation matrix* is defined by $Corr(\mathbf{X}) = R_{\mathbf{X}} = D_{\mathbf{X}}^{-1} \Sigma_{\mathbf{X}} D_{\mathbf{X}}^{-1}$ where

$$D_{\mathbf{X}} = diag(Sd(X_1), \ldots, Sd(X_1)) = diag(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{kk}}).$$

(i) Show that the (i, j)-th element of $R_{\mathbf{X}}$ is $Corr(X_i, X_j)$.
(ii) Suppose $\mathbf{Y} = D\mathbf{X}$ where $D = diag(d_1, \ldots, d_k)$ with $d_i > 0$ for $i = 1, \ldots, k$. Show $Corr(\mathbf{Y}) = Corr(\mathbf{X})$.
(iii) Suppose in (ii) that D is not diagonal with positive diagonal, is it true that $Corr(\mathbf{Y}) = Corr(\mathbf{X})$?