Tópicos de matrizes e Distribuição Normal Multivariada

CAPÍTULO 1

Alguns resultados importantes

1.1 definições

I_n : matriz Identidade de dimensão $(n \times n)$
J_n : 11^t

1.1.1 Traço

Sejam $A_{p\times p}, B_{p\times p}, C_{p\times p}, D_{p\times p}, x_{p\times p}$ e α :escalar.

Propriedades

1. $Tr(\alpha) = \alpha$
2. $Tr(A+B) = Tr(A) + Tr(B)$
3. $Tr(\alpha A) = \alpha Tr(A)$
4. $Tr(CD) = Tr(DC) = \sum_{i,j} c_{ij} d_{ji}$
5. $\sum_i x_i^t A x_i = tr(AT)$ onde $T = \sum_i x_i x_i^t$
6. $tr(B^{-1}AB) = tr(A)$

1.1.2 Determinantes

$A_{(p\times p)}$, C (constante)

1. Se A é diagonal ou triangular $|A| = \prod_{i=1}^p a_{ii}$
2. $|CA| = C^p |A|$
3. $|AB| = |A| |B|$
4. $|A| = |A^t|$
5. Se cada elemento de uma linha (coluna) de A é zero, $|A| = 0$
6. Se quaisquer duas linhas (colunas) de A é zero, $|A| = 0$
7. Se quaisquer duas linhas (colunas) de A são id6enticas, $|A| = 0$
8. Se A é não-singular, $|A| = 1/|A^{-1}|$ ou seja, $|A| |A^{-1}| = 1$
9. Se $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ onde A_{11}e A_{22} são matrizes quadradas, $|A| = |A_{11}| \left|A_{22} - A_{21}A_{11}^{-1}A_{12}\right| = |A_{22}| \left|A_{11} - A_{12}A_{22}^{-1}A_{21}\right|$

10. Sejam $B_{(p\times n)}, C_{(n\times p)}$ e $A_{(p\times p)}$ não-singular. Temos $|A+BC| = |A^{-1}|\,|I_p + A^{-1}BC| = |A^{-1}|\,|I_n + CA^{-1}B|$
11. Sejam $b_{(p\times 1)}, A_{(p\times p)}$ não-singular ,$|A+bb^t| = |A|\,|1+b^tA^{-1}b|$
12. Se $B_{(p\times n)}$ e $C_{(n\times p)}$ então $|I_p+BC| = |I_n+CB|$

1.1.3 Inversa

Propriedades

1. $(AB)^{-1} = B^{-1}A^{-1}$
2. A única solução de $Ax = b$ é $x = A^{-1}b$
3. Sejam $A_{(p\times p)}, B_{(p\times n)}, C_{(n\times n)}$ e $D_{(n\times p).}$ Se todas as inversas necessárias existem, então $(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}$
   caso particular $A_{(p\times p)}, b_{(p\times 1)}$ e $c_{(p\times 1)}$, se A^{-1}existe $(A+bc^t)^{-1} = A^{-1} - \frac{A^{-1}bc^tA^{-1}}{1+c^tA^{-1}b}$
4. Se todas as matrizes inversas necessárias existem , então a matriz particionada A^{-1} é dada por :
   $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ $A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}$
   $A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$
   $A^{12} = -A^{11}A_{12}A_{22}^{-1} = -A_{11}^{-1}A_{12}A^{22}$
   $A^{21} = -A^{22}A_{21}A_{11}^{-1} = -A_{22}^{-1}A_{21}A^{11}$
   $A^{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$

1.1.4 Produto de Kronecker

Definição Sejam $A = (a_{ij})$ e $B = (a_{bij})$ matrizes de dimensão $(m\times n)$ e $(p\times q)$, respectivamente. O produto de Kronecker , indicado por $A \bigotimes B = (a_{ij}B) = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m1}B & \dots & a_{mn}B \end{pmatrix}$

Def: Vec(A) Seja A uma matriz de dimensão $(m\times n)$ e $a_{(i)}$a i-ésima coluna de A. $Vec(A)$ é um vetor de dimensão $(mn\times 1)$ definido por :

$$Vec(A) = \begin{pmatrix} a_{(1)} \\ a_{(2)} \\ \vdots \\ a_{(n)} \end{pmatrix}$$

Propriedades A, B, C, D : matrizes, x, y : vetores, α :escalar

1. $\alpha(A \bigotimes B) = (\alpha A) \bigotimes B = A \bigotimes (\alpha B)$
2. $A \bigotimes (B \bigotimes C) = (A \bigotimes B) \bigotimes C = A \bigotimes B \bigotimes C$

3. $(A \bigotimes B)^t = A^t \bigotimes B^t$
4. $(A \bigotimes B)(C \bigotimes D) = (AC) \bigotimes (BD)$
5. $(A \bigotimes B)^{-1} = (A^{-1} \bigotimes B^{-1})$
6. $(A + B) \bigotimes C = (A \bigotimes C) + (B \bigotimes C)$
7. $A \bigotimes (B + C) = (A \bigotimes B) + (A \bigotimes C)$
8. $A_{(p\times p)}, B_{(q\times q)}, |A \bigotimes B| = |A|^q \, |B|^p$
9. $Vec(ABC) = (C^t \bigotimes A)Vec(B)$, se $ABCexiste$
10. $x \bigotimes y = vec(yx^t)$
11. $x \bigotimes y^t = xy^t = y^t \bigotimes x$
12. $(Vec(A))^t vec(B) = tr(A^t B)$
13. $(Vec(A))^t (B \bigotimes C) Vec(D) = tr(A^t C D B^t)$

1.1.5 Matrizes especiais

1. Matrizes ortogonais
   Se A uma matriz quadrada, A é ortogonal se $AA^t = I$
   Propriedades
   (a) $A^{-1} = A^t$
   (b) $A^t A = I$
   (c) $|A| = \pm 1$
   (d) $a^t_{(i)} a_{(j)} = 0$ se $i \neq j$
       $a^t_{(i)} a_{(j)} = 1$ se $i = j$
   (e) Se A e B são ortogonais $C = AB$ é ortogonal.

2. Matriz de equicorrelação
   $E = (1-\rho)I_\rho + \rho J_p,$
   $$E = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}, \rho : \text{número real}$$
   $E^{-1} = (1-\rho)^{-1}[(I_\rho - \rho\{1 + (p-1)\rho\}^{-1} J_p$
   $|E| = (1-\rho)^{p-1}\{1 + \rho(p-1)\}$
   De uma forma mais geral,
   $$A = \begin{pmatrix} c+b & c & \dots & c \\ c & c+b & \dots & c \\ \vdots & \vdots & \vdots & \vdots \\ c & c & \dots & c+b \end{pmatrix}, A = cJ_n + bI_n$$

$A^{-1} = \frac{1}{b} I_n - \frac{c}{(nc+b)b} J_n$

3. Matriz Idempotente

   A é idempotente de $A^2 = A$

4. Matriz Positiva definida e positiva semi-definida

   A é positiva definida se $x^t A x > 0, \forall x \neq 0$

   A é positiva semi-definida se $x^t A x \geq 0, \forall x \neq 0$

1.1.6 Posto de uma matriz

O posto de uma matriz $A_{(n \times p)}$ é definida como o número máximo linhas (colunas) linearmente independentes de A; ou é a ordem da maior submatriz quadrada de A com determinante não-nulo. $Posto(A) : r(A)$

Propriedades Seja uma matriz $A_{(n \times p)}$

1. $0 \leq r(A) \leq \min(n, p)$
2. $r(A) = r(A^t)$
3. $r(A + B) \leq r(A) + r(B)$
4. $r(AB) \leq \min\{r(A), r(B)\}$
5. $r(A^t A) = r(AA^t) = r(A)$
6. Se $B_{(n \times n)}$ e $C_{(p \times p)}$ são não-singular, então $r(BAC) = r(A)$
7. Se $n = p$ então $r(A) = p$ se, e somente se, A é não singular.

   Posto de alguns matrizes

1. $A = diag(a_i), r(A) =$números de $a_i' s \neq 0$
2. $r(H) = n - 1$
3. A idempotente, $r(A) = tr(A)$

1.1.7 Autovalores e autovetores

Definição

Autovalores. Seja A uma matriz de dimensão $(p \times p)$.$\lambda_1, \ldots \lambda_p$ que satisfazem a equação $|A - \lambda I_p| = 0$ são denominados autovalores da matriz A.Os autovalores podem ser complexos ou múltiplos.

Autovetores. Para todo autovalor λ_i existe um vetor $\gamma \neq 0$ tal que A$\gamma = \lambda_i \gamma$ onde γ é denominado autovetor de A associado ao autovalor $\lambda_{i.}$

Em geral vamos usar os autovetores normalizados ou seja $\gamma^t \gamma = 1$

$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \lambda_1 = 0; \lambda_2 = 5$

$\gamma_1 = \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}$ e $\gamma_2 = \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}$

Propriedades

1. Seja $C_{(p\times p)}$ uma matriz não-singular A e CAC^{-1} tem os mesmos autovalores. Se γ é um autovetor de A para λ_i então $\nu = C\gamma$ é um autovetor de CAC^{-1} para λ_i.

Prova. $CAC^{-1} - \lambda I = CAC^{-1} - \lambda CC^{-1}CC^{-1} = C(A - \lambda I)C^{-1}$
$|CAC^{-1} - \lambda I| = |C|\,|A - \lambda I|\,|C^{-1}| = |A - \lambda I|$
$A\gamma = \lambda_i \gamma$
$CA\gamma = \lambda_i C\gamma$
$CAC^{-1}C\gamma = \lambda_i C\gamma$
$CAC^{-1}\nu = \lambda_i \nu \quad \square$

1. Seja α escalar. Então $A + \alpha I$ tem autovalores $\lambda_i + \alpha$. Além disso, A e $A + \alpha I$ tem os mesmos autovetores.
2. Se $A_{(p\times p)}$ é simétrica então todos os autovalores são reais.

1.2 Decomposição Espectral

Qualquer matriz simétrica $A_{(p\times p})$ pode ser escrita como $A = \Gamma\Lambda\Gamma^t = \sum_{i=1}^{p} \lambda_i \gamma_{(i)}\gamma_{(i)}^t$ onde Λ é a matriz diagonal dos autovalores de A e Γ é uma matriz ortogonal cujas colunas são os autovetores normalizados de A.

1.2.1 Propriedade

1. Se $A_{(p\times p)}$ é uma matriz simétrica não-singular então para qualquer inteiro n, $\Lambda^n = diag(\lambda_i^n)$ e $A^n = \Gamma\Lambda^n\Gamma^t$.
2. Se todos os autovalores de A são positivos, $A^{r/s} = \Gamma\Lambda^{r/s}\Gamma^t$ onde $\Lambda^{r/s} = diag(\lambda_i^{r/s})$, para inteiros $s > 0$ e r.

   obs: Se alguns dos autovalores de A são iguais a zero, entõ os resultados anteriores são válidos se os expoentes forem não-negativos.

Prova. por indução
$\square$

Casos Especiais $\quad A^2 = \Gamma\Lambda^2\Gamma^t$; $A^{-1} = \Gamma\Lambda^{-1}\Gamma^t$; $A^{-1/2} = \Gamma\Lambda^{-1/2}\Gamma^t$.

Propriedades de $A^{-1/2}$

1. $(A^{-1/2})^t = A^{-1/2}$
2. $A^{1/2}A^{1/2} = A$
3. $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = I$
4. $A^{-1/2}A^{-1/2} = A^{-1}$
5. Seja A simétrica então o posto de A é igual ao número de autovalores não nulo de A

Prova. $A = \Gamma\Lambda\Gamma^t$

$r(A) = r(\Gamma\Lambda\Gamma^t) = r(\Lambda)$

□

1. Se $A_{(p\times p)}$ é simétrica , então :
2. $tr(A) = \sum_{i=1}^{p} \lambda_i$
3. $\det(A) = \Pi_{i=1}^{p}\lambda_i$
4. Uma matriz simétrica A tem posto 1 se, e somente se, $A = xx^t$ para algum x. Então, o único autovalor de A não-nulo é dado por $tr(A) = tr(xx^t) = xx^t$.
5. Seja $J = 11^t$. Temos que $r(J) = 1$ e que o único autovalor não-nulo de J é $1^t1 = p$ e o autovetor correspondente é 1_p.

   Seja $E = (1-p)I + \rho J$, os autovalores de E são $\lambda_1 = 1 + (p-1)\rho, \lambda_2 = \ldots = \lambda_p = 1 - \rho$ e seu autovetores de E são os mesmos de J
6. Se A, é simétrica e idempotente então $\lambda_i = 0$ ou $1, \forall i$.

1.3 Formas Quadráticas

Definição

Uma forma quadrática no vetor x é uma função da forma :

$Q(x) = x^tAx = \sum_i \sum_j x_i a_{ij} x_i$ onde A é um matriz simétrica.

Propriedades

1. $Q(0) = 0$
2. $Q(x)$ é positiva definida se $Q(x) > 0$, $\forall x \neq 0$
3. A simétrica é p.d. (p.s.d) se $Q(x)$ é p.d. (p.s.d.)

   Para qualquer matriz simétrica A, existe um transformação ortogonal $y = \Gamma^t x$ tal que $x^tAx = \sum_i \lambda_i y_i^2$.

   Prova. Sabemos que $A = \Gamma\Lambda\Gamma^t$, seja $y = \Gamma^t x$. Logo

   $\Gamma\ y = \Gamma\Gamma^t x \Rightarrow x^t = y^t\Gamma^t$.

   $x^tAx = y^t\Gamma^tA\Gamma\ y = y^t\Gamma^t\Gamma\Lambda\Gamma^t\Gamma y = y^t\Lambda y = \sum_i \lambda_i y_i^2$.

   □
4. Se $A > 0$ então $\lambda_i > 0, \forall i$
5. Se $A \geq 0$ então $\lambda_i \geq 0, \forall i$
6. Se $A > 0$ então A é não-singular e $|A| > 0$
7. Se $A > 0$ então $A^{-1} > 0$

   Prova. $A^{-1} = \Gamma\Lambda^{-1}\Gamma^t$

   $y = \Gamma^t x$ tal que $x^tA^{-1}x = \sum_i \frac{1}{\lambda_i} y_i^2$.

$x^tA^{-1}x = y^t\Gamma^tA^{-1}\Gamma\ y = y^t\Gamma^t\Gamma\Lambda^{-1}\Gamma^t\Gamma y = y^t\Lambda^{-1}y = \sum_i \frac{1}{\lambda_i}y_i^2. > 0$, pois $\lambda_i > 0, y \neq 0$.

□

8. Qualquer matriz $A \geq 0$ pode ser escrita como $A = B^2$ onde B é uma matriz simétrica.

   Prova. Sabemos que $A = \Gamma\Lambda\Gamma^t$, seja $B = \Gamma\Lambda^{1/2}\Gamma^t$ então $B^2 = \Gamma\Lambda^{1/2}\Gamma^t\Gamma\Lambda^{1/2}\Gamma^t = \Gamma\Lambda\Gamma^t = A$

   □

9. Se $A \geq 0, A_{(p\times p)}$ então para qualquer matriz C de ordem $(p\times n)$ temos $C^tAC \geq 0$
10. Se $A > 0$ e C não-singular $(p = n)$ então $C^tAC > 0$
11. Se $A \geq 0$ e $B > 0$ matrizes de ordem $(p \times p)$ então todas as raízes características não-nulas de $B^{-1}A$ são positivas.

Interpretação Geométrica Seja A uma matriz positiva definida. Então $(x-\alpha)^tA^{-1}(x-\alpha) = C^2$ representa um elipsóide em dimensão p. O centro do elipsoíde é $x = \alpha$.

1.4 Inversa Generalizada

Definição : Seja a matriz $A_{(n\times p)}$. A^- é a g-inversa ou inversa generalizada de A se $AA^-A = A$. A g-inversa sempre existe, embora possa não ser única

1. Se $r(A) = r$ e $A_{(n\times p)}$ então as linhas e colunas podem ser rearranjadas de modo que $A_{11}(r \times r)$ seja não-singular . Logo uma g-inversa é dada por

   $A^- = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$
2. Se $A_{(p\times p)}$ é não-singular então $A^- = A^{-1}$ e é única

Teorema Seja G uma g-inversa de X^tX

1. G^t é uma g-inversa de X^tX
2. GX^t é uma g-inversa de X
3. XGX^t não varia com G
4. XGX^t é simétrica mesmo que G não o seja.

1.5 Diferenciação de Vetores e Matrizes

1. Seja a um vetor de constantes

   $a^tx = x^ta = \lambda; \lambda = a_1x_1 + a_2x_2 + \ldots + a_px_p$

   $\frac{\partial\lambda}{\partial x} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}$

2. $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ e $A = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \vdots & \vdots \\ a_{p1} & \dots & a_{pp} \end{pmatrix}$

   $x^t A = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_p \end{pmatrix}$ onde $\lambda_i = \sum_{j=1}^{p} x_j a_{ji}$

   $\frac{\partial x^t A}{\partial x} = \begin{pmatrix} \frac{\partial \lambda_1}{\partial x_1} & \dots & \frac{\partial \lambda_p}{\partial x_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial \lambda_1}{\partial x_p} & \dots & \frac{\partial \lambda_p}{\partial x_p} \end{pmatrix} = A$

3. Formas Quadráticas

   $Q(x) = x^t A x$, A simétrica

   $x^t A x = a_{11} x_1^2 + \ldots + a_{pp} x_p^2 + 2a_{12} x_1 x_2 + \ldots + 2a_{(p-1)p} x_{p-1} x_p$

   $\frac{\partial x^t A x}{\partial x} = \begin{pmatrix} \frac{\partial x^t A x}{\partial x_1} \\ \vdots \\ \frac{\partial x^t A x}{\partial x_p} \end{pmatrix} = \begin{pmatrix} 2a_{11} x_1 + 2a_{12} x_2 + \ldots + 2a_{1p} x_p \\ \vdots \\ 2a_{p1} x_1 + 2a_{p2} x_2 + \ldots + 2a_{pp} x_p \end{pmatrix} = 2AX$

1.5.1 Resultados

1. $\partial A = 0$
2. $\partial(\alpha U) = \alpha U$
3. $\partial(U \pm V) = \partial U \pm \partial V$
4. $\partial(UV) = (\partial U)V + U(\partial V)$
5. $\partial U^t = (\partial U)^t$
6. $\partial vec(U) = vec(\partial U)$
7. $\partial tr(U) = tr(\partial U)$
8. $\partial A^{-1} = -A^{-1} \partial A A^{-1}$
9. $\partial(B^t X) = B$
10. $\partial(X^t A y) = Ay$
11. $\partial(X^t X) = 2X$
12. $\partial(X^t A X) = 2AX$ se X é simétrica
13. $\partial(Y^t A X) = Y B^t$
14. $\partial(Y^t X Y) = Y Y^t$

    A é simétrica e a um vetor

1. $\partial(a^t X A X^t a) = 2aa^t X A$
2. $\partial(a^t X A Y^t a) = aa^t Y A$
3. $\partial(tr(YX)) = Y^t$
4. $\partial(tr(YXB)) = Y^t B^t$

CAPÍTULO 2

Vetores Aleatórios

Um vetor aleatório é um vetor cujos elementos são variáveis aleatórias. Similarmente, uma matriz aleatória é uma matriz cujos elementos são variáveis aleatórias. Os vetores aleatórios são também chamados de variáveis aleatórias multidimensionais. O Valor esperado de uma matriz aleatória é uma matriz consistindo dos valores esperados de cada um de seus elementos. Seja X uma matriz aleatória $p \times n$, $X = (X_{ij})$, se existem os valores esperados $E(X_{ij})$,

Se (a matriz de valores esperados)X e Y têm mesma dimensão $p \times n$ e são matrizes aleatórias e A e B são adequadas matrizes constantes,

$E(X + Y) = E(X) + E(Y)$

$E(AXB) = AE(X)B$

2.1 Vetor de médias e Matriz de covariâncias

Seja X, $p \times 1$ e $E(X_i) = \mu_i$, $i = 1, \ldots, p, Cov(X_i, X_j) = \sigma_{ij}, i, j = 1, 2, \ldots, p$, então denotamos $E(X)$ por $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$ e $Cov(X)$ por $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{21} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{pmatrix}$.

Se X_i e X_j são independentes então $Cov(X_{i,}X_j) = 0$. Há situações em que $Cov(X_i, X_j) = 0$ mas X_i e X_j não são independentes.

Por definição $E(X) = E \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \mu$ e $Cov(X) =$

$$E(X - \mu)(X - \mu)^t = E \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{pmatrix} \begin{pmatrix} (X_1 - \mu_1) & (X_2 - \mu_2) & \ldots & (X_p - \mu_p) \end{pmatrix} =$$

$$= E \begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) & \ldots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)(X_2 - \mu_2) & \ldots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \vdots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \ldots & (X_p - \mu_p)(X_p - \mu_p) \end{pmatrix} =$$

$$=E\begin{pmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \dots & (X_1-\mu_1)(X_p-\mu_p) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \dots & (X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \vdots & \vdots \\ (X_p-\mu_p)(X_1-\mu_1) & (X_p-\mu_p)(X_2-\mu_2) & \dots & (X_p-\mu_p)^2 \end{pmatrix} =$$

$$\begin{pmatrix} \sigma_{11} & \sigma_{11} & \dots & \sigma_{11} \\ \sigma_{11} & & \dots & \sigma_{11} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{11} & \sigma_{11} & \dots & \sigma_{11} \end{pmatrix}$$

onde $\sigma_{ii} = \sigma_i^2, \sigma_{ij} = \sigma_{ji}$.

2.1.1 Matriz de Correlação

Uma medida de correlação linear entre X_i e X_j é dada pelo coeficiente de correlação linear simples $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$. O coeficiente de correlação é obtido da matriz de covariância-variância Σ. A Matriz de correlação $\rho = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \vdots & 1 & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}$ pode ser obtida por $\rho = \left[(V^{1/2})^{-1}\Sigma(V^{1/2})^{-1}\right]$ onde

$$V^{1/2} = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_{pp}} \end{pmatrix} e$$

$$V^{-1/2} = \begin{pmatrix} \sigma_{11}^{-1/2} & 0 & \dots & 0 \\ 0 & \sigma_{22}^{-1/2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp}^{-1/2} \end{pmatrix}.$$

Outra relação importante é $\Sigma = V^{1/2}\rho V^{1/2}$.Assim Σ pode ser obtida de Σ pode ser obtida de ρ e $V^{1/2}$ enquanto ρ pode ser obtida de Σ.

Matriz de covariância Particionada

Frequentemente as características observadas num experimento podem ser classificadas em dois grupos. Por exemplo, em observando-se estudantes as variáveis sócio-econômicas podem formar um grupo, enquanto o desempenho acadêmico é composto por outro grupo de variáveis. Em geral, particionando o vetor X em dois grupos de variáveis, digamos, $X^{(1)}, (q \times 1)$ e $X^{(2)}, (p-q) \times 1$, obtém-se

$$E(X) = E\begin{pmatrix} X^{(1)} \\ \dots \\ X^{(2)} \end{pmatrix} = \begin{pmatrix} E(X^{(1)}) \\ \dots \\ E(X^{(2)}) \end{pmatrix} = \begin{pmatrix} \mu^{(1)} \\ \dots \\ \mu^{(2)} \end{pmatrix}$$

$$Cov(X) = Cov \begin{pmatrix} X^{(1)} \\ \dots \\ X^{(2)} \end{pmatrix} = E(X-\mu)(X-\mu)^t$$

$$= E \begin{pmatrix} X^{(1)} - \mu^{(1)} \\ \dots \\ X^{(2)} - \mu^{(2)} \end{pmatrix} \Big((X^{(1)} - \mu^{(1)})^t \vdots (X^{(2)} - \mu^{(2)})^t \Big)$$

$$=E \begin{pmatrix} (X^{(1)} - \mu^{(1)})(X^{(1)} - \mu^{(1)})^t & \vdots & (X^{(1)} - \mu^{(1)})(X^{(2)} - \mu^{(2)})^t \\ \dots & \dots & \dots \\ (X^{(2)} - \mu^{(2)})(X^{(1)} - \mu^{(1)})^t & \vdots & (X^{(2)} - \mu^{(2)})(X^{(2)} - \mu^{(2)})^t \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \dots & \dots & \dots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{pmatrix}$$

$$\Sigma_{12} = \Sigma_{12}^t$$

2.2 Lista de exercícios

1. Seja a variável aleatória bidimensional X, $p \times 1, p = 2$. X_1 e X_2 são v.a. discretas independentes com as seguintes funções de probabilidade,

(a)

x_1	-1	0	1
$p(x_i)$	$0,3$	$0,3$	$0,4$

x_2	0	1
$p(x_i)$	$0,8$	$0,2$

Calcule :

(b) $E(X), Cov(X)$

(c) $E(AX), Cov(AX)$ para $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

(d) ρ_x

Comente

1. Verifique que para o vetor aleatório $X = (X_1, \ldots, X_p)^t$

(a) $Cov(X_i + a, X_j + b) = Cov(X_i, X_j)$, a e b constantes

(b) $Cov(aX_i, bX_j) = abCov(X_i, X_j)$, a e b constantes

(c) Para combinações lineares das variáveis componentes de X, a^tX e $b^tX, Cov(a^tX, b^tX) = a^t \sum b$, forma bilinear.

2. Se A e B são matrizes constantes $(r \times p)$ e $(s \times p)$, respectivamente e $Y = AX$, $Z = BX$ são duas transformações da variável aleatória X então :
$Cov(Y, Y) = A\Sigma A^t$, $Cov(Z, Z) = B\Sigma B^t$, $Cov(Y, Z) = A\Sigma B^t$

3. Dado $E(X) = \mu$, $Var(X) = \Sigma$, $Cov(X_i, X_j) = 0, \forall i \neq j$
Calcule $|\Sigma|$, $(X - \mu)^t\Sigma^{-1}(X - \mu)$
Verifique que $\rho = (V^{(1/2)})^{-1}\Sigma(V^{(1/2)})^{-1}$ com $V^{(1/2)} = diag(\sqrt{\sigma_{ii}})$, $i = 1, \ldots, p$

4. Seja X tal que $\Sigma = \begin{pmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{pmatrix}$

(a) Calcule ρ , $V^{(1/2)}$ e Σ^{-1}.

(b) Encontre os valores e vetores próprios de Σ

(c) Verifique que $\Sigma = V^{(1/2)}\rho V^{(1/2)}$

(d) Encontre a correlação entre X_1 e $\frac{X_2+X_3}{2}$.

CAPÍTULO 3

Distribuição Normal Multivariada

A generalização da familiar densidade normal para várias dimensões tem um fundamental papel na análise multivariada. Enquanto dados reais nunca são exatamente normal multivariados, a densidade normal é frequentemente uma útil aproximação para a veradadeira distribuição da população.

Uma vantagem da distribuição normal multivariada é que ela é matamaticamente atrativa, dela obtendo-se excelentes resultados. Mas estatísticamente, duas outras razões são as que indicam o uso da distribuição normal.

Primeira, distribuições amostrais de muitos estatísticos multivariados são aproximadamente normais, devido ao efeito do teorema do limite central. Em segundo lugar, a distribuição normal serve como modelo aproximado para certos fenômenos naturais.

3.1 Densidade e propriedades da distribuição normal multivariada

3.1.1 Definição 1

Sabemos que a distribuição normal univariada, com média μ e variância σ^2, tem função de densidade de probabilidade $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$, $-\infty < x < \infty$.

$X \sim N(\mu, \sigma^2)$ implica que $P(\mu - \sigma \leq X \leq \mu + \sigma) \cong 0,68$ e $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \cong 0,95$.

A densidade normal multivariada é a generalização da densidade normal univariada para dimensões $p \geq 2$. O termo $\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$ é generalizado para $(x-\mu)^t\Sigma^{-1}(x-\mu)$ que é a distância quadrada generalizada (distância de Mahalanobis), quando Σ admite inversa. Em outro caso a densidade não é bem definida. Também o termo $\frac{1}{\sqrt{2\pi\sigma^2}} = (2\pi)^{-1/2}(\sigma^2)^{-1/2}$ deve ser modificado para uma constante mais geral para tornar o ' volume$'$ (no caso multivariado as probabilidades são representadas por volumes sob a superfícies na região definida) sob a superfície da função de densidade multivariada unitária para qualquer p. Essa constante será $(2\pi)^{-p/2}\left|\Sigma\right|^{-1/2}$.

Consequentemente, para Σ definida positiva, a função de densidade de uma variável $X \sim N_p(\mu, \Sigma)$ será

$$
\begin{aligned}
f(x) &= (2\pi)^{-p/2} \left|\Sigma\right|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}, \\
-\infty &< x_i < \infty,\ i = 1, 2, \ldots, p.
\end{aligned}
$$

3.1.2 Definição 2

Dizemos que X tem uma distribuição normal multivariada $p-$variada se e somente se $a^t x$ tem distribuição normal univariada para todo a fixado.

Se X tem distribuição normal multivariada $p-$variada então cada um dos elementos de X, ou seja, X_i , $i = 1, \ldots, p$ tem distribuição normal univariada.

Se todas as $p(p-1)/2$ covariâncias são nulas, as p componentes de x são independentemente distribuídas e

$$f(x) = f_1(x_1)f_2(x_2)\ldots f_p(x_p), \text{ consequentemente,}$$

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(x)dx_1 \ldots dx_p =$$

$$= \int_{-\infty}^{x_1} f(x_1)dx_1 \int_{-\infty}^{x_2} f(x_2)dx_2 \cdots \int_{-\infty}^{x_p} f(x_p)dx_p =$$

$$= F_1(x_1)F_1(x_1)\ldots F_p(x_p).$$

A densidade normal multivariada é constante nas superfícies onde a distância $(x-\mu)^t\Sigma^{-1}(x-\mu)$ é constante. Esse corte é chamado de *contorno*.

O contorno de uma densidade de probabilidade constante é a superfície de um elipsóide centrado em μ e é igual ao conjunto de pontos $\{x : (x-\mu)^t\Sigma^{-1}(x-\mu) = c^2\}$. Esses elipsóides têm eixos $\pm c\sqrt{\lambda_i}e_i$, onde $(\lambda_i e_i)$ é um par de autovalor-autovetor da matriz Σ.

As figuras a_1 e a_2 mostram as distribuições de duas binormais, na primeira X_1 e X_2 são independentes, na segunda X_1 e X_2 tem correlação de 0,75. As figuras b_1 e b_2 são contornos de 50% e 90% para duas $N_2(\mu,\Sigma)$, X_1 e X_2 independentes ou correlacionadas. A figura b_3 mostra contorno de densidade constante para uma normal bivariada com $\sigma_{11} = \sigma_{22}$ e $\rho_{12} > 0$.

O elipsóide sólido dos x tais que $(x-\mu)^t\Sigma^{-1}(x-\mu) \leq \chi^2_p(\alpha)$ tem probabilidade $(1-\alpha)$, para $X \sim N_p(\mu,\Sigma)$ e χ^2_p sendo o $\alpha-$quantil superior da qui-quadrado com p graus de liberdade.

3.1.3 Propriedades da Distribuição normal multivariada

Seja $X \sim N_p(\mu,\Sigma)$, então são verdades

1. Combinações lineares de componentes de X são distribuídas normalmente :

$$X \sim N_p(\mu,\Sigma) \Rightarrow a^tX \sim N(a^t\mu, a^t\Sigma a)$$

2. Todo subconjunto de componentes de X tem distribuição normal multivariada.

$$X \sim N_p(\mu,\Sigma), A_{(q\times p)} \Rightarrow AX \sim N_q(A\mu, A\Sigma A^t)$$

3. Covariância zero implica que as correspondentes componentes são independentemente distribuídas.

$$X \sim N_p(\mu,\Sigma) \Rightarrow [cov(X_i,X_j) \Longleftrightarrow X_i \text{ independente } X_j$$

   (a) Seja $X \sim N_p(\mu,\Sigma)$ e $Y = \Sigma^{-1/2}(X-\mu)$ onde $\Sigma^{-1/2}$ é a raiz quadrada simétrica positiva definida de Σ^{-1}. Então $Y_1, Y_2, \ldots, Y_p$ são independentes e $Y_i \sim N(0,1)$ para todo i.
   (b) Se $X \sim N_p(\mu,\Sigma)$ então $E(X) = \mu$, $Var(X) = \Sigma$
   (c) Se $X \sim N_p(\mu,\Sigma)$ então $U = (X-\mu)^t\Sigma^{-1}(X-\mu) \sim \chi^2_p$
   (d) Se $X \sim N_p(\mu,\Sigma)$, $A_{(q\times p)}$, $C_{(q\times 1)}$ e $Y = AX + C \Rightarrow Y \sim N_q(A\mu + C, A\Sigma A^t)$

   Função característica

   Seja X um vetor aleatório $(p\times 1)$. A função característica de X é definida como :

$$\phi_x(s) = E(e^{is^tx})$$

4. Seja $X^t = (X_1^t, X_2^t)$. Os vetores aleatórios X_1 e X_2 são independentes se, e somente se,

$$\phi_x(s) = \phi_{X_1}(s_1)\phi_{X_2}(s_2) \text{ onde } s^t = (s_1^t, s_2^t).$$

5. Se X e Y são vetores aleatórios $(p \times 1)$ independentes então $\phi_{X+Y}(s) = \phi_X(s) + \phi_Y(s)$
6. Se $X \sim N_p(\mu, \Sigma) \iff \phi_X(s) = \exp\left\{is^t\mu - \frac{1}{2}s^t\Sigma s\right\}$
7. Dois vetores conjuntamente multinormais são independentes se, e somente se são não correlacionados.
8. Se $X \sim N_p(\mu, \Sigma)$ então AX e BX são independentes se, e somente se $A\Sigma B^t = 0$.
9. Distribuições condicionais das componentes são multinormais :

$$\begin{aligned} X \sim N_p(\mu, \Sigma) \Rightarrow X^{(1)} \mid X^{(2)} \sim N_q(\mu_{1|2}, \Sigma_{1|2}), \\ \text{onde } \mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \\ \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

10. $X \sim N_p(\mu, \Sigma), d$ um vetor de constantes $\Rightarrow (X + d) \sim N_p(\mu + d, \Sigma)$.
11. Todos os subconjuntos de componentes de X são normalmente distribuídas. Se particionamos X, seu vetor de médias μ e matriz de covariância Σ. Seja X_1 e X_2 com dimensão q e p-q respectivamente, isto é
$X = \begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \cdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{pmatrix}$ então $X_1 \sim N_q(\mu_1, \Sigma_{11})$ e $X_2 \sim N_q(\mu_2, \Sigma_{22})$
12. X_1 e X_2 são independentes se e somente se $Cov(X_1\ , X_2) = \Sigma_{12} = 0$.
13. Se X_1 e X_2 são independentes e $X_1 \sim N_{q_1}(\mu_1, \Sigma_{11})$ e $X_2 \sim N_{q_2}(\mu_2, \Sigma_{22})$ respectivamente, então $\begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \mu_1 \\ \cdots \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \Sigma_{22} \end{pmatrix}\right)$.
14. Se $X \sim N_p(\mu, \Sigma)$ e $|\Sigma| > 0$, então $(x - \mu)^t\Sigma^{-1}(x - \mu) \sim \chi^2_p$
15. Se $X \sim N_p(\mu, \Sigma)$ e $|\Sigma| > 0$, então o elipsóide sólido $\{\ (x-\mu)^t\Sigma^{-1}(x-\mu) \leq \chi^2_p(\alpha)\}$ tem probabilidade $(1-\alpha)$, com $\chi^2_p(\alpha)$ sendo o $\alpha-$quantil superior da distribuição qui-quadrado com p graus de liberdade.
16. Se $X_1, X_2, \ldots, X_n$ são mutuamente independentes com $X_j \sim N_p(\mu_j, \Sigma)$, com mesma matriz de covariância Σ então $V_1 = c_1X_1+c_2X_2+\ldots+c_nX_n \sim N_p(\mu_{V_1}, \Sigma_{V_1})$ com $\mu_{V_1} = \sum_{j=1}^n c_j\mu_j$ e $\sum_{V_1} = (\sum_{j=1}^n c_j^2)\Sigma$. Além do mais V_1 e $V_2 = \sum_{j=1}^n b_jX_j$ são conjuntamente normais multivariadas com matriz de covariâncias

$$\begin{pmatrix} (\sum_{j=1}^n c_j^2)\Sigma & b^tc\Sigma \\ b^tc\Sigma & (\sum_{j=1}^n b_j^2)\Sigma \end{pmatrix}$$

Consequentemente V_1 e V_2 são independentes se $b^t c = \sum_{j=1}^{n} b_j c_j = 0$,isto é , os vetores b e c são perpendiculares.

Considerando todos os possíveis valores de x_2, podemos escrever a variável $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ como predição da distribuição condicional de X_1. A diferença entre X_1 e a predição da média da distribuição condicional de X_1 é o vetor $X_{1.2}$ é chamado de conjunto de variáveis residuais.

$$X_{1.2} = X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

Em populações multinormais as variáveis residuais e as fixadas são distribuidas independentemente.

3.2 Lista 2 de exercícios de Análise Multivariada

1. Considere uma população normal bivariada com $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ e $\rho_{12} = 0,5$.

   (a) Escreva a densidade desta normal

   (b) Apresente a expressão da distância quadrada generalizada $(x-\mu)^t\Sigma^{-1}(x-\mu)$ como uma função de x_1 e x_2.

   (c) Determine o contorno de densidade constante que contém 50% de probabilidade. Esboce o gráfico do contorno.

   (d) Especifique a distribuição condicional de X_1, dado $X_2 = x_2$ para a distribuição

2. Sejam $X_1 \sim N(0,1)$ e $X_2 = \begin{cases} -X_1 & , \ \ se\ -1 \leq X_1 \leq 1 \\ X_1 & , \ \ em\ outro\ caso. \end{cases}$

   (a) Mostre que X_2 tem também distribuição normal

   (b) Mostre que X_1 e X_2 não tem distribuição normal bivariada

3. Seja $X \sim N_3(\mu, \Sigma)$ com $\mu = (2, -3, 1)^t$ e $\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}$

   (a) Encontre a distribuição de $3X_1 - 2X_2 + X_3$

   (b) Determine um vetor $a_{(2\times1)}$, tal que X_2 e $\left[X_2 - a^t \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}\right]$ são independentes.

   (c) Determine a distribuição de X_3 dado $X_1 = x_1$ e $X_2 = x_2$

   (d) Verifique que na questão acima (X_1, X_2) é independente da variável residual.

CAPÍTULO 4

Amostras Aleatórias

4.1 Introdução

Uma observação multivariada é o conjunto de medidas de p diferentes variáveis na mesma unidade de análise. Tomando-se n observações, a massa de dados pode ser arranjada em uma matriz de dados X como

$$X_{(p\times n)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{11} & x_{11} & \cdots & x_{11} \\ \cdots & \cdots & \vdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} = \begin{pmatrix} x_1, & x_{2,} & \ldots & x_n \end{pmatrix}$$

Cada coluna de X representa uma observação multivariada e a matriz X é uma amostra de tamanho n de uma população n de uma população $p-$variada. Cada coluna reoresenta um ponto num espaço $p-$dimensional, fornecendo informação sobre sua locação e variabilidade além de associação linear.

O vetor média amostral $\overline{x}$ é obtido como combinação linear das colunas de X, ou seja,

$$\overline{x}_{(p\times 1)} = \frac{1}{n}\sum_{i=1}^{n} x_i = X \begin{pmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{pmatrix} = \frac{1}{n} X\ 1$$

Se os pontos são considerados esferoídes o vetor de médias, $\overline{x}$, é o centro de gravidade. A matriz S de variância e covariâncias amostral indica a variação nas várias direções do sistema. O determinante da matriz de variância e covariâncias amostral é uma medida númerica da variabilidade total.

$$S = \left[\frac{1}{n-1}\sum_{i=1}^{n} x_i x_i^t - n\overline{x}\overline{x}^t\right] = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(x_i - \overline{x}\right)^t = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{pmatrix}$$

A matriz de covariância amostral contém p variâncias e $\frac{1}{2}p(p-1)$ covariâncias. A variância amostral generalizada é o determinande de S e representa a variação expressa em S. A fragilidade da variância generalizada pode ser mostrada nas seguintes três matrizes de covariâncias as quais tem mesma variância generalizada e diferente estrutura de correlação, não detectada por $\det(S)$,

$$\begin{aligned} S_1 &= \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, S_2 = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}, S_3 = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \\ \rho_1 &> 0,\ \rho_2 < 0 \text{ e } \rho_3 = 0 \end{aligned}$$

A matriz de dados X pode ser considerada como uma observação da matriz aleatória $\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{pmatrix}$ composta dos vetores colunas $\begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}$.

Se os vetores colunas $X_1, \ldots, X_n$ representam independentes observações de uma distribuição comum, com função de densidade $f(x) = f(x_1, \ldots, x_p)$, então $X_1, X_2, \ldots, X_n$ formam uma amostra aleatória de $f(x)$. Então $f(x_1, \ldots, x_n) = f(x_1).f(x_2). \ldots .f(x_n)$ onde $f(x_j) = f(x_{1j}, x_{2j}, \ldots, x_{pj})$

As medidas das p variáveis em uma uínica observação $X_j^t = (X_{1j}, \ldots, X_{pj})$, serão em geral correlacionadas.

As medidas de diferentes observações devem ser no entanto independentes.

A violação da hipótese de independência entre cada observação pode causar sérios impactos na qualidade da inferência estatística. Observações através do tempo são um exemplo desta situação.

4.1.1 Resultados sobre a variância generalizada , $|S|$:

1. Em qualquer análise estatística, $|S| = 0$ significa que as medidas de algumas variáveis devem ser removidas do estudo.
2. Se $n \leq p$ (isto é, o número de observações é menor ou igual ao número de variáveis observadas), então $|S| = 0$ para todas as amostras.
3. Se a combinação linear $a^t X_j$ tem variância positiva para cada vetor constante $a \neq 0$ e se $p < n$, então S tem posto completo com probabilidade 1 e $|S| > 0$.

4.1.2 Variância Total Amostral

Outra generalização da variância é definida como a soma dos elementos sa diagonal principal e é chamada de variância total amostral, $\sum_{i=1}^{p} s_{ii} = s_{11} + s_{12} + \ldots + s_{pp}$.

4.2 Amostras Aleatórias de uma Distribuição Multinormal

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população $p-$variada com o vetor de médias μ e matriz de covariância Σ. Desde que $X_1, X_2, \ldots, X_n$ são mutuamente independentes e com uma distribuição comum $N_p(\mu, \Sigma)$, a função de densidade conjunta de todas as observações é o produto das densidades normais marginais,

$$\begin{aligned} f_{X_1,\ldots X_p}(x_1,\ldots,x_n) &= f_{X_1}(x_1).f_{X_2}(x_2).\ldots.f_{X_n}(x_n) = \prod_{j=1}^{n} f_{X_j}(x_j) = \\ &= \prod_{j=1}^{n}\left\{\frac{1}{(2\pi)^{p/2}\left|\Sigma\right|^{n/2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu)^t\Sigma^{-1}(x_j-\mu)\right\}\right\} \end{aligned}$$

Quando considerada como uma função de μ e Σ esta função de densidade conjunta é a função de verossimilhança.

4.3 Estimação de Máxima verossimilhaça de μ e Σ para $N_p(\mu, \Sigma)$.

Consideremos uma amostra aleatória de uma $N_p(\mu, \Sigma)$. A função de verossimilhança dada acima será denotada por $L(\mu, \Sigma)$ para ressaltar que é uma função de μ e Σ. Após algumas manipulações algébricas, podemos reescrever esta função como

$$L(\mu,\Sigma) = (2\pi)^{-np/2}\left|\Sigma\right|^{-n/2}\exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\overline{x})(x_j-\overline{x})^t + n(\overline{x}-\mu)(\overline{x}-\mu)^t\right)\right]\right\}$$

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população normal com média μ e covariância Σ. Então ,

$$\widehat{\mu} = \overline{X} \text{ e } \widehat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(X_j-\overline{X})(X_j-\overline{X})^t = \frac{n-1}{n}S,$$

são os <u>estimadores de máxima verossimilhança</u> de μ e Σ, respectivamente. Seus valores observados são

$$\overline{x} \text{ e } \frac{1}{n}\sum_{j=1}^{n}(x_j-\overline{x})(x_j-\overline{x})^t$$

Pela propriedade da invariância, se $\widehat{\theta}$ é um estimador máxima verossimilhança de θ, então o estimador máxima verossimilhança de uma função de θ, seja $h(\theta)$, é dado por $h(\widehat{\theta})$. Assim sendo o estimador máxima verossimilhança de ρ, mattriz de correlação de X é $\widehat{\rho}$,ou seja,

$$\begin{aligned} \rho &= diag(\sigma_{ii}^{-1/2})\Sigma diag(\sigma_{ii}^{-1/2}) = f(\Sigma) \\ \widehat{\rho} &= f(\widehat{\Sigma}) = diag(\widehat{\sigma}_{ii}^{-1/2})\Sigma diag(\widehat{\sigma}_{ii}^{-1/2}), \text{ onde} \\ \widehat{\rho}_{ij} &= \frac{\widehat{\sigma}_{ij}}{(\widehat{\sigma}_{ii}\widehat{\sigma}_{ii})^{1/2}} \end{aligned}$$

Propriedades

1. Se $X_{(p\times n)}$ é uma matriz de dados da $N_p(\mu, \Sigma)$ e $\overline{X} = n^{-1}X1$ então $\overline{X} = N_p(\mu, n^{-1}\Sigma)$
2. $E(\overline{X}) = \mu$ e $Var(\overline{X}) = \frac{1}{n}\Sigma$
3. $E(\widehat{\Sigma}) = \frac{n}{n-1}\Sigma$, $E(S) = \Sigma$
4. Se X é uma matriz de dados da $N_p(\mu, \Sigma)$ e se $Y = AXB$ e $Z = CXD$, então os elementos de Y são independentes dos de Z se, e somente se,

$$B\Sigma D^t = 0 \text{ ou } A^t C = 0$$

Teorema do Limite Central

Sejam $X_1, X_2, \ldots,$ uma sequência infinita de vetores aleatórios indenticamente independentemente distribuidas de uma distribuição com média μ e Σ. Então

$$n^{-1/2}\sum_{r=1}^{n}(X_r - \mu) = n^{-1/2}(\overline{x} - \mu) \xrightarrow{D} N_p(0, \Sigma)$$