第一章：随机向量与随机矩阵的数字特征、多元正态分布及其性质和参数估计

主要介绍了概率统计中关于**随机向量和随机矩阵的数字特征**，特别是**数学期望、协方差矩阵、相关矩阵**，以及**多元正态分布**及其相关性质和参数估计。

第一部分：数字特征 (幻灯片 1-19)

幻灯片 1: 数字特征

这页是标题页，列出了本部分将要介绍的主要内容：

- 数学期望 (均值)
- 协方差矩阵 (协差阵)
- 相关矩阵

幻灯片 2: 数学期望 (均值)

这一页定义了随机向量和随机矩阵的数学期望。

- > **随机向量 $X = (X_1, X_2, \dots, X_p)'$ 的数学期望**
  > 定义为：
  > $E(X) = [E(X_1), E(X_2), \dots, E(X_p)]'$
- 这个期望向量也记为 $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$，其中 $\mu_i = E(X_i)$。
  简单来说，随机向量的期望就是其各个分量期望所组成的向量。
- > **随机矩阵 $X = (X_{ij})$ 的数学期望**
  > 定义为：
  > $E(X) = (E(X_{ij}))$
- 具体地，如果 X 是一个 $n \times p$ 的随机矩阵：
  $$E(X) = E\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} = \begin{pmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1p}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{n1}) & E(X_{n2}) & \cdots & E(X_{np}) \end{pmatrix}$$
  即随机矩阵的期望是其各个元素期望所组成的矩阵。

幻灯片 3: 随机矩阵的数学期望的性质

这里列出了随机矩阵数学期望的一些重要性质：

1. **设 a 为常数，则：**
   $E(aX) = aE(X)$
   常数因子可以提到期望符号外面。
2. > **设 A, B, C 为常数矩阵，则：**
   > $E(AXB + C) = AE(X)B + C$
   > 常数矩阵可以提到期望符号外面，常数矩阵 C 的期望是其自身。

- **特别地，对于随机向量 X，有：**
  $E(AX) = AE(X)$（这里可以看作 $B = I, C = 0$ 的情况，或者 X 是列向量，B 是标量 1）

3. **设 $X_1, X_2, \ldots, X_n$ 为 n 个同阶的随机矩阵，则：**
   $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$
   和的期望等于期望的和。

幻灯片 4: 协方差矩阵 (初步)

这页引入了两个随机变量协方差的概念。

- **两个随机变量的协方差定义为：**
  $\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$
  协方差衡量了两个随机变量之间的线性相关程度。
- **若 $\mathrm{Cov}(X, Y) = 0$，则称 X 和 Y 不相关。**
  不相关意味着两者之间没有线性关系。
- **两个独立的随机变量必然不相关，但两个不相关的随机变量未必独立。**
  这是一个非常重要的概念。独立性是比不相关性更强的条件。不相关只排除了线性依赖关系，但可能存在非线性依赖关系。
- **当 $X = Y$ 时，协方差即为方差，也就是：**
  $\mathrm{Cov}(X, X) = E[(X - E(X))(X - E(X))] = E[(X - E(X))^2] = D(X)$（或 $\mathrm{Var}(X)$）

幻灯片 5: 协方差矩阵 (向量间)

这页定义了两个随机向量之间的协方差矩阵。

> - **设 $X = (X_1, X_2, \ldots, X_p)'$ 和 $Y = (Y_1, Y_2, \ldots, Y_q)'$ 为两个随机向量，它们的协方差矩阵 (简称协差阵) 定义为：**
> $\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))']$
>   这是一个 $p \times q$ 的矩阵，其 (i, j) 元为 $\mathrm{Cov}(X_i, Y_j)$。
> $$\mathrm{Cov}(X, Y) = \begin{pmatrix} \mathrm{Cov}(X_1, Y_1) & \mathrm{Cov}(X_1, Y_2) & \cdots & \mathrm{Cov}(X_1, Y_q) \\ \mathrm{Cov}(X_2, Y_1) & \mathrm{Cov}(X_2, Y_2) & \cdots & \mathrm{Cov}(X_2, Y_q) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, Y_1) & \mathrm{Cov}(X_p, Y_2) & \cdots & \mathrm{Cov}(X_p, Y_q) \end{pmatrix}$$
>   展开形式为：
> $$\mathrm{Cov}(X, Y) = E\left[\begin{pmatrix} X_1 - E(X_1) \\ \vdots \\ X_p - E(X_p) \end{pmatrix} (Y_1 - E(Y_1), \quad \cdots, \quad Y_q - E(Y_q))\right]$$
> $$= \begin{pmatrix} E[(X_1 - E(X_1))(Y_1 - E(Y_1))] & \cdots & E[(X_1 - E(X_1))(Y_q - E(Y_q))] \\ \vdots & \ddots & \vdots \\ E[(X_p - E(X_p))(Y_1 - E(Y_1))] & \cdots & E[(X_p - E(X_p))(Y_q - E(Y_q))] \end{pmatrix}$$

幻灯片 6: 协方差矩阵 (单个向量)

这页讨论了单个随机向量自身的协方差矩阵以及相关概念。

> - **X 和 Y 的协方差矩阵与 Y 和 X 的协差阵互为转置关系，即有：**
> $\mathrm{Cov}(X, Y) = [\mathrm{Cov}(Y, X)]'$

- **若 $\mathrm{Cov}(X, Y) = 0$ (零矩阵)，则称 X 和 Y 不相关。**
- **两个独立的随机向量必然不相关，但两个不相关的随机向量未必独立。** (与标量情况类似)

- **当 $X = Y$ 时的协差阵 $\text{Cov}(X, X)$ 称为 X 的协方差矩阵，记作 $D(X)$ (或 Σ)，即：**
  $D(X) = \text{Cov}(X, X) = E[(X - E(X))(X - E(X))']$
  这是一个 $p \times p$ 的对称矩阵。其对角线元素是 X 各分量的方差，非对角线元素是 X 各分量之间的协方差。
  $$D(X) = \begin{pmatrix} D(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & D(X_2) & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \cdots & D(X_p) \end{pmatrix}$$

- **$D(X)$ 亦记作 $\Sigma = (\sigma_{ij})$，其中 $\sigma_{ij} = \text{Cov}(X_i, X_j)$。因此 $\sigma_{ii} = D(X_i)$。**

幻灯片 7: 协方差矩阵的分块

- **协差阵 Σ 既包含了 X 各分量的方差，也包含了每两个分量之间的协方差。**

- **随机向量一分为二后，其协差阵分为四块：**
  如果随机向量 X 可以分块为 $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$，其中 $X^{(1)}$ 是 q 维的，$X^{(2)}$ 是 $(p-q)$ 维的。
  那么其协方差矩阵 $D(X)$ 可以分块为：
  $$D\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} = \begin{pmatrix} D(X^{(1)}) & \text{Cov}(X^{(1)}, X^{(2)}) \\ \text{Cov}(X^{(2)}, X^{(1)}) & D(X^{(2)}) \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$
  其中，
  - $D(X^{(1)}) = \Sigma_{11}$ 是 $X^{(1)}$ 自身的协方差矩阵。
  - $D(X^{(2)}) = \Sigma_{22}$ 是 $X^{(2)}$ 自身的协方差矩阵。
  - $\text{Cov}(X^{(1)}, X^{(2)}) = \Sigma_{12}$ 是 $X^{(1)}$ 和 $X^{(2)}$ 之间的协方差矩阵。
  - $\text{Cov}(X^{(2)}, X^{(1)}) = \Sigma_{21}$ 是 $X^{(2)}$ 和 $X^{(1)}$ 之间的协方差矩阵，且 $\Sigma_{21} = \Sigma_{12}'$。

- **熟悉这四块子矩阵的含义很有益处。**

幻灯片 8: 协差阵的性质 (1)

- **(1) 协差阵是非负定阵，即 $\Sigma \geq 0$。**

- 这意味着对于任意非零常数向量 $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$，都有 $\alpha'\Sigma\alpha \geq 0$。
  证明：
  $\alpha'\Sigma\alpha = \alpha' E[(X - E(X))(X - E(X))']\alpha$
  $= E[\alpha'(X - E(X))(X - E(X))'\alpha]$
  令 $Y = \alpha'(X - E(X))$，这是一个标量随机变量。则上式变为：
  $= E[Y \cdot Y'] = E[Y^2]$
  因为 Y 是实数，$Y^2 \geq 0$，所以 $E[Y^2] \geq 0$。
  即 $E[(\alpha'(X - E(X)))^2] \geq 0$。

- **注1：若 $|\Sigma| \neq 0$ (即 Σ 可逆或满秩)，则 $\Sigma > 0$ (正定阵)。**
  这意味着 $\alpha'\Sigma\alpha > 0$ 对于所有非零向量 α 成立。

幻灯片 9: 协差阵的性质 (注2)

- **注2：$|\Sigma| = 0 \iff X$ 的分量之间存在线性关系 (以概率1)。**

- **含义：** 在实际问题中，有时 $|\Sigma| = 0$ 或 $|\Sigma|$ 近似为0，此时指标之间存在着线性相关关系 (共线性)。
- 通常需要删去"多余"的指标 (变量选择)。
- 不失一般性，总假定 $\Sigma > 0$，保证 Σ^{-1} 存在，从而使得数学问题简化。这在后续的多元正态分布中尤其重要。

幻灯片 10: 协差阵的性质 (2) (3)

- > **(2) 设 A 为常数矩阵，b 为常数向量，则：**
  > $D(AX + b) = AD(X)A'$

- 证明概要：
  $E(AX + b) = AE(X) + b$
  $D(AX + b) = E[(AX + b - E(AX + b))(AX + b - E(AX + b))']$
  $= E[(AX + b - (AE(X) + b))(AX + b - (AE(X) + b))']$
  $= E[(A(X - E(X)))(A(X - E(X)))']$
  $= E[A(X - E(X))(X - E(X))'A']$
  $= AE[(X - E(X))(X - E(X))']A'$
  $= AD(X)A'$

- > **(3) 设 A 和 B 为常数矩阵，则：**
  > $\text{Cov}(AX, BY) = A\text{Cov}(X, Y)B'$

- 证明概要：
  $E(AX) = AE(X), E(BY) = BE(Y)$
  $\text{Cov}(AX, BY) = E[(AX - E(AX))(BY - E(BY))']$
  $= E[(A(X - E(X)))(B(Y - E(Y)))']$
  $= E[A(X - E(X))(Y - E(Y))'B']$
  $= AE[(X - E(X))(Y - E(Y))']B'$
  $= A\text{Cov}(X, Y)B'$

幻灯片 11: 协差阵的性质 (4)

- > **(4) 设 $A_1, A_2, \ldots, A_n$ 和 $B_1, B_2, \ldots, B_m$ 为常数矩阵，则：**
  > $\text{Cov}\left(\sum_{i=1}^n A_i X_i, \sum_{j=1}^m B_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m A_i \text{Cov}(X_i, Y_j) B_j'$

- 这个性质表明协方差算子对于线性组合具有双线性性。

- > **推论：** (当 $A_i = I, B_j = I$ 时，或者 X_i, Y_j 是标量， A_i, B_j 是标量常数)
  > $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$

幻灯片 12: 证明 (性质4的推论及性质4)

这里先证明推论，再利用推论和性质 (3) 证明性质 (4)。

- **先证推论：**
  $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right)$
  $= E\left[\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)\left(\sum_{j=1}^m Y_j - E\left(\sum_{j=1}^m Y_j\right)\right)'\right]$
  $= E\left[\left(\sum_{i=1}^n (X_i - E(X_i))\right)\left(\sum_{j=1}^m (Y_j - E(Y_j))\right)'\right]$
  $= E\left[\sum_{i=1}^n \sum_{j=1}^m (X_i - E(X_i))(Y_j - E(Y_j))'\right]$

$= \sum_{i=1}^{n} \sum_{j=1}^{m} E[(X_i - E(X_i))(Y_j - E(Y_j))']$
$= \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(X_i, Y_j)$

- **再证性质 (4):**
  令 $U_i = A_i X_i$ 和 $V_j = B_j Y_j$。
  $\text{Cov}\left(\sum_{i=1}^{n} A_i X_i, \sum_{j=1}^{m} B_j Y_j\right) = \text{Cov}\left(\sum_{i=1}^{n} U_i, \sum_{j=1}^{m} V_j\right)$
  根据推论:
  $= \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(U_i, V_j)$
  $= \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(A_i X_i, B_j Y_j)$
  根据性质 (3) $\text{Cov}(AX, BY) = A\text{Cov}(X, Y)B'$:
  $= \sum_{i=1}^{n} \sum_{j=1}^{m} A_i \text{Cov}(X_i, Y_j) B_j'$

幻灯片 13: 协差阵的性质 (5) (6)

> - **(5) 设 $k_1, k_2, \ldots, k_n$ 是 n 个常数，$X_1, X_2, \ldots, X_n$ 是 n 个相互独立的 p 维随机向量，则:**
> $D\left(\sum_{i=1}^{n} k_i X_i\right) = \sum_{i=1}^{n} k_i^2 D(X_i)$

- 证明思路：利用 $D(\sum k_i X_i) = \text{Cov}(\sum k_i X_i, \sum k_j X_j) = \sum_i \sum_j k_i k_j \text{Cov}(X_i, X_j)$。由于 X_i, X_j 相互独立（当 $i \neq j$ 时），$\text{Cov}(X_i, X_j) = 0$。当 $i = j$ 时，$\text{Cov}(X_i, X_i) = D(X_i)$。

> - **(6)** $D(X) = E(XX') - (EX)(EX)'$

- 这是计算协方差矩阵的常用公式，类似于一维方差 $D(X) = E(X^2) - (E(X))^2$。
  证明:
  $D(X) = E[(X - EX)(X - EX)']$
  $= E[XX' - X(EX)' - (EX)X' + (EX)(EX)']$
  由于 EX 是常数向量，可以提出期望:
  $= E(XX') - E(X)(EX)' - (EX)E(X)' + (EX)(EX)'$
  $= E(XX') - (EX)(EX)' - (EX)(EX)' + (EX)(EX)'$
  $= E(XX') - (EX)(EX)'$

幻灯片 14: 协差阵的性质 (7)

> - **(7) 设 X 为 p 维随机向量，期望 $E(X) = \mu$ 和协方差矩阵 $D(X) = \Sigma$ 存在，A 为 $p \times p$ 常数阵，则:**
> $E(X'AX) = \text{tr}(A\Sigma) + \mu' A \mu$

- 这里 $X'AX$ 是一个标量 (二次型)。$\text{tr}(\cdot)$ 表示矩阵的迹 (主对角线元素之和)。

  证明方法 (1) (元素法):

  $X'AX = \sum_i \sum_j a_{ij} X_i X_j$
  $E(X'AX) = E\left(\sum_i \sum_j a_{ij} X_i X_j\right) = \sum_i \sum_j a_{ij} E(X_i X_j)$
  我们知道 $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = E(X_i X_j) - \mu_i \mu_j$。
  所以 $E(X_i X_j) = \text{Cov}(X_i, X_j) + \mu_i \mu_j = \sigma_{ij} + \mu_i \mu_j$。
  $E(X'AX) = \sum_i \sum_j a_{ij}(\sigma_{ij} + \mu_i \mu_j) = \sum_i \sum_j a_{ij} \sigma_{ij} + \sum_i \sum_j a_{ij} \mu_i \mu_j$
  注意到 $\sum_i \sum_j a_{ij} \sigma_{ij} = \text{tr}(A\Sigma')$。因为 Σ 是对称的，$\Sigma' = \Sigma$，所以是 $\text{tr}(A\Sigma)$。
  (更准确地说，$\sum_j a_{ij} \sigma_{ji}$ 是 $(A\Sigma)_{ii}$，所以 $\sum_i \sum_j a_{ij} \sigma_{ji} = \text{tr}(A\Sigma)$。而 $\sum_i \sum_j a_{ij} \mu_i \mu_j = \mu' A \mu$ (可以写成 $(\mu' A)\mu$ 或 $\mu'(A\mu)$)。
  所以 $E(X'AX) = \text{tr}(A\Sigma) + \mu' A \mu$。

证明方法 (2) (迹的性质):

$X'AX$ 是一个标量，所以 $X'AX = \mathrm{tr}(X'AX)$。
$E(X'AX) = E(\mathrm{tr}(X'AX))$
利用迹的性质 $\mathrm{tr}(BCD) = \mathrm{tr}(DBC) = \mathrm{tr}(CDB)$，这里 $\mathrm{tr}(X'AX) = \mathrm{tr}(AXX')$。
$= E(\mathrm{tr}(AXX')) = \mathrm{tr}(E(AXX')) = \mathrm{tr}(AE(XX'))$
由性质 (6)，$E(XX') = D(X) + (EX)(EX)' = \Sigma + \mu\mu'$。
$= \mathrm{tr}(A(\Sigma + \mu\mu')) = \mathrm{tr}(A\Sigma + A\mu\mu')$
$= \mathrm{tr}(A\Sigma) + \mathrm{tr}(A\mu\mu')$
再次利用迹的性质 $\mathrm{tr}(A\mu\mu') = \mathrm{tr}(\mu'A\mu)$。因为 $\mu'A\mu$ 是一个标量，其迹就是其自身。
$= \mathrm{tr}(A\Sigma) + \mu'A\mu$

幻灯片 15: 相关矩阵 (定义)

- **随机变量 X 和 Y 的相关系数定义为:**
  $\rho = \rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)D(Y)}}$
  相关系数是标准化的协方差，其值在 $[-1,1]$ 之间。

- **随机向量 $X = (X_1, X_2, \dots, X_p)'$ 和 $Y = (Y_1, Y_2, \dots, Y_q)'$ 的相关阵定义为:**
  (幻灯片这里似乎只写了 X 内部各分量的相关系数，但标题是 X 和 Y。通常相关阵是指一个随机向量内部各分量间的相关系数矩阵。如果指 X 和 Y 之间，则是 $(p \times q)$ 矩阵，元素为 $\rho(X_i, Y_j)$。)
  如果指的是 X 内部的相关阵 R_X，其 (i,j) 元素为:
  $\rho_{ij} = \rho(X_i, X_j) = \frac{\mathrm{Cov}(X_i,X_j)}{\sqrt{D(X_i)D(X_j)}}$

幻灯片 16: 相关矩阵 (性质)

- **若 $\rho(X,Y) = 0$，则表明 X 和 Y 不相关。** (这与 $\mathrm{Cov}(X,Y) = 0$ 等价，只要方差不为0)

- **$X = Y$ 时的相关阵 $\rho(X,X)$ 称为 X 的相关阵，记作 $R = (\rho_{ij})$，这里 $\rho_{ij} = \rho(X_i, X_j)$，$\rho_{ii} = 1$。即:**
  $$R = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}$$
  相关阵是一个对称矩阵，对角线元素全为1。

- **$R = (\rho_{ij})$ 和 $\Sigma = (\sigma_{ij})$ 之间有关系式:**
  $R = D^{-1/2}\Sigma D^{-1/2}$
  其中 D 是一个对角矩阵，其对角线元素是 Σ 的对角线元素 (即各分量的方差):
  $D = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp}) = \mathrm{diag}(D(X_1), D(X_2), \dots, D(X_p))$
  $D^{-1/2}$ 是一个对角矩阵，其对角线元素是 $1/\sqrt{\sigma_{ii}}$:
  $D^{-1/2} = \mathrm{diag}\left(\frac{1}{\sqrt{\sigma_{11}}}, \frac{1}{\sqrt{\sigma_{22}}}, \dots, \frac{1}{\sqrt{\sigma_{pp}}}\right)$

- (幻灯片中写的是 $D = \mathrm{diag}(\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}})$，然后关系式是 $R = D^{-1}\Sigma D^{-1}$。这两种表达是等价的，只是 D 的定义不同。如果按幻灯片 $D = \mathrm{diag}(\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}})$，那么 $D^{-1} = \mathrm{diag}(1/\sqrt{\sigma_{11}}, \dots, 1/\sqrt{\sigma_{pp}})$。则 $(D^{-1}\Sigma D^{-1})_{ij} = (1/\sqrt{\sigma_{ii}})\sigma_{ij}(1/\sqrt{\sigma_{jj}}) = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} = \rho_{ij}$。)

幻灯片 17: R 和 Σ 的相应元素之间的关系式

这一页用矩阵形式明确显示了 $R = D^{-1}\Sigma D^{-1}$ 的关系 (采用幻灯片16中 $D = \text{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{pp}})$ 的定义)。

> - 元素关系：
> $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$
> - 矩阵形式：
> $$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$
>   这与 $R = D^{-1}\Sigma D^{-1}$ 一致。

幻灯片 18: 标准化变换 (定义)

- **在数据处理时，为克服各指标的量纲不同对统计分析结果带来的影响，需要对每个指标作标准化变换，最常用的标准化变换是令：

> - $X_i^* = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad i = 1, 2, \ldots, p$
>   其中 $\mu_i = E(X_i)$，$\sigma_{ii} = D(X_i)$。
>   标准化后的变量 X_i^* 均值为0，方差为1。

- **记 $X^* = (X_1^*, X_2^*, \ldots, X_p^*)'$，则:**
  $E(X^*) = \mathbf{0}$ (零向量)
  $D(X^*) = R$ (原变量 X 的相关矩阵)

幻灯片 19: 标准化变换 (性质)

- 令 $D_\sigma = \text{diag}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \ldots, \sqrt{\sigma_{pp}})$ (注意这里 D_σ 就是幻灯片16中的 D)。
  则标准化变换可以写成向量形式： $X^* = D_\sigma^{-1}(X - E(X))$。
  于是有：
  $E(X^*) = E[D_\sigma^{-1}(X - E(X))] = D_\sigma^{-1}E(X - E(X)) = D_\sigma^{-1}(\mathbf{0}) = \mathbf{0}$
  $D(X^*) = D[D_\sigma^{-1}(X - E(X))]$
  利用 $D(AY + b) = AD(Y)A'$ 的性质 (这里 $A = D_\sigma^{-1}$，$Y = X$，$b = -D_\sigma^{-1}E(X)$，但更简单的是 $D(A(X-\mu)) = AD(X-\mu)A' = AD(X)A'$)：
  $D(X^*) = D_\sigma^{-1}D(X)(D_\sigma^{-1})'$
  因为 D_σ^{-1} 是对角阵，所以 $(D_\sigma^{-1})' = D_\sigma^{-1}$。
  $D(X^*) = D_\sigma^{-1}\Sigma D_\sigma^{-1} = R$

> - **标准化数据的协差阵正好是原指标的相关阵。**

> - **相关阵 R 也是一个非负定阵。** (因为它是标准化向量的协方差矩阵，协方差矩阵是非负定矩阵)

第二部分：多元正态分布 (幻灯片 20-38)

幻灯片 20: 多元正态分布的定义

- **一元正态分布 $N(\mu, \sigma^2)$ 的概率密度函数为：**
  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  也可以写成：
  $f(x) = (2\pi)^{-1/2}(\sigma^2)^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right], \quad -\infty < x < \infty$

- > **若随机向量 $X = (X_1, X_2, \ldots, X_p)'$ 的概率密度函数为：**
  > $f(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right]$
  > 其中 $\mathbf{x} = (x_1, \ldots, x_p)'$, μ 是 p 维均值向量，Σ 是 $p \times p$ 正定协方差矩阵。
  > 则称 X 服从 p 元正态分布，记作 $X \sim N_p(\mu, \Sigma)$。

- **可以证明：参数 μ 和 Σ 分别为 X 的均值和协差阵。**

幻灯片 21: 多元正态分布的等价定义 (线性变换)

- **随机向量 $Y = (Y_1, Y_2, \ldots, Y_q)'$, $Y_i \sim N(0, 1)$ i.i.d. (独立同分布)。**
  这意味着 $E(Y) = \mathbf{0}$ 且 $D(Y) = I_q$ (q 阶单位矩阵)。

- > **令 $X = \mu_{p\times 1} + A_{p\times q}Y$，则称 X 服从多元正态分布。**
  > 这里 μ 是一个 p 维常数向量，A 是一个 $p \times q$ 常数矩阵。
  > 此时，$E(X) = \mu + AE(Y) = \mu$。
  > $D(X) = AD(Y)A' = AI_qA' = AA'$。
  > 记为：$X \sim N_p(\mu, AA')$ 或 $X \sim N_p(\mu, \Sigma)$，其中 $\Sigma = AA'$。
  > (注意：要使得 Σ 是 $p \times p$ 正定矩阵从而有密度函数，通常要求 A 是 $p \times p$ 的满秩矩阵，即 $q = p$ 且 A 可逆。如果 $q < p$ 或者 A 不是满秩，$\Sigma = AA'$ 将是奇异的，此时 X 服从退化的多元正态分布，没有通常意义下的密度函数。)

- **定理2.1：$X \sim N_p(\mu, \Sigma) \implies E(X) = \mu, D(X) = \Sigma$。**
  这个定义方法直接给出了均值和协方差矩阵。

- **证明 (基于此定义)：**
  $E(Y) = \mathbf{0}, D(Y) = I$
  $E(X) = E(\mu + AY) = \mu + AE(Y) = \mu + A\mathbf{0} = \mu$
  $D(X) = D(\mu + AY) = D(AY) = AD(Y)A' = AIA' = AA' = \Sigma$

幻灯片 22: 多元正态分布的等价定义 (从PDF推导线性变换)

这一页和下一页展示了如何从 $Y_i \sim N(0, 1)$ i.i.d. 通过变换得到 $X \sim N_p(\mu, \Sigma)$ 的密度函数。

- **$X \sim N_p(\mu, \Sigma), \Sigma > 0, X$ 的概率密度函数为：**
  $f(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right]$
- **因为 $\Sigma > 0$ (正定)，可以进行 Cholesky 分解或特征值分解，使得 $\Sigma = TT'$，其中 T 是一个 $p \times p$ 的可逆矩阵。**
  (例如，如果 $\Sigma = PDP'$ 是谱分解，P 是正交阵，D 是对角阵，则 $T = PD^{1/2}$。)
- **令 $X = TY + \mu$，其中 $Y_i \sim N(0, 1)$ i.i.d., $i = 1, 2, \ldots, p$。**
  则 $Y = T^{-1}(X - \mu)$。
  Y 的联合密度函数为 $f_Y(\mathbf{y}) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2} = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right)$。
- 利用多维随机变量的密度函数变换公式：$f_X(\mathbf{x}) = f_Y(\mathbf{y}(\mathbf{x}))|J(\mathbf{y} \to \mathbf{x})|$ 或者 $f_X(\mathbf{x}) = f_Y(\mathbf{y}(\mathbf{x}))/|J(\mathbf{x} \to \mathbf{y})|$。
  其中 $J(\mathbf{x} \to \mathbf{y})$ 是从 $\mathbf{x}$ 到 $\mathbf{y}$ 变换的雅可比行列式，即 $J(\mathbf{x} \to \mathbf{y}) = |\det(\frac{\partial y_i}{\partial x_j})|$。
  $f_X(\mathbf{x}) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right) \frac{1}{|J(\mathbf{x}\to\mathbf{y})|}$

幻灯片 23: 多元正态分布的等价定义 (雅可比行列式)

- **雅可比行列式 $J(\mathbf{x} \to \mathbf{y})$ 的计算:**
  $\mathbf{y} = T^{-1}(\mathbf{x} - \mu)$。
  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = T^{-1}$ (这是一个矩阵，其 (i, j) 元为 $\partial y_i / \partial x_j$)。
  所以 $|J(\mathbf{x} \to \mathbf{y})| = |\det(T^{-1})| = |\det(T)|^{-1}$。
  因为 $\Sigma = TT'$, 则 $|\Sigma| = |\det(T)\det(T')| = (\det(T))^2$。
  所以 $|\det(T)| = |\Sigma|^{1/2}$。
  因此 $|J(\mathbf{x} \to \mathbf{y})| = |\Sigma|^{-1/2}$。
  幻灯片上的 $|J(\mathbf{x} \to \mathbf{y})| = \left|\frac{\partial \mathbf{x}'}{\partial \mathbf{y}}\right|_+ = |T'|_+ = |TT'|^{1/2} = |\Sigma|^{1/2}$ 似乎是 $J(\mathbf{y} \to \mathbf{x})$ 的行列式，即 $\mathbf{x} = T\mathbf{y} + \mu \implies \frac{\partial \mathbf{x}}{\partial \mathbf{y}'} = T$。所以 $|J(\mathbf{y} \to \mathbf{x})| = |\det(T)| = |\Sigma|^{1/2}$。
  那么 $f_X(\mathbf{x}) = f_Y(\mathbf{y}(\mathbf{x}))|J(\mathbf{y} \to \mathbf{x})|^{-1} = f_Y(\mathbf{y}(\mathbf{x}))|J(\mathbf{x} \to \mathbf{y})|$.
- **代入 $f_X(\mathbf{x})$ 的表达式:**
  $\mathbf{y}'\mathbf{y} = (T^{-1}(\mathbf{x} - \mu))'(T^{-1}(\mathbf{x} - \mu)) = (\mathbf{x} - \mu)'(T^{-1})'T^{-1}(\mathbf{x} - \mu)$
  由于 $\Sigma = TT'$, $\Sigma^{-1} = (TT')^{-1} = (T')^{-1}T^{-1} = (T^{-1})'T^{-1}$。
  所以 $\mathbf{y}'\mathbf{y} = (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$。
  $f_X(\mathbf{x}) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right)|\Sigma|^{-1/2}$
  这与幻灯片20的定义一致。

幻灯片 24: 多元正态分布 (补充说明)

- **需要明确的是，更广泛的多元正态分布的定义可以借助特征函数来定义。**
  一个 p 维随机向量 X 服从多元正态分布，如果对任意 p 维常数向量 t， $t'X$ 服从一元正态分布 (或为常数)。其特征函数为 $\phi_X(t) = E(e^{it'X}) = \exp(it'\mu - \frac{1}{2}t'\Sigma t)$。
- **可以证明，这三类定义是等价的** (PDF 定义, 线性变换定义, 特征函数定义)。

> - **另外，在密度函数定义中，如果 $|\Sigma| = 0$，则 Σ^{-1} 不存在，X 不存在通常意义下的密度函数。**
>   这种情况称为退化 (singular) 正态分布，随机向量 X 的分量间存在线性关系，其分布集中在一个低维的超平面上。

幻灯片 25: 二元正态分布

这里详细给出了 $p = 2$ 的情况，即二元正态分布。

- **设 $X \sim N_2(\mu, \Sigma)$，这里:**
  $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$
  其中 $\sigma_1^2 = D(X_1), \sigma_2^2 = D(X_2), \rho = \rho(X_1, X_2)$ 是 X_1 和 X_2 的相关系数。
- **易见，ρ 是 X_1 和 X_2 的相关系数。**

> - **当 $|\rho| < 1$ 时 (即 Σ 正定，行列式 $|\Sigma| = \sigma_1^2\sigma_2^2(1 - \rho^2) > 0$)，可得 X 的概率密度函数为:**
> $f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$

幻灯片 26: 二元正态分布的密度曲面图

- **下图是当 $\sigma_1^2 = \sigma_2^2 = 1, \rho = 0.75$ 时二元正态分布的钟形密度曲面图。**
  图中展示了典型的钟形曲面，峰值在 (μ_1, μ_2) 处。$\rho = 0.75$ 表明 X_1, X_2 正相关，密度函数的等高线（投影到 $x_1 - x_2$ 平面）会是沿 $y = x$ 方向拉长的椭圆。

幻灯片 27: 二元正态分布等高线

- **等高 (椭圆) 线:**
  密度函数 $f(x_1, x_2)$ 为常数的点构成的曲线。即指数部分为常数:
  $\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 = c^2$
  这是一个二次型，表示中心在 (μ_1, μ_2) 的椭圆。椭圆的形状和方向由 ρ, σ_1, σ_2 决定。
- **上述等高线上的密度值:**
  $f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{c^2}{2(1-\rho^2)}\right]$
  对于给定的 c^2，密度值是恒定的。

幻灯片 28: 二元正态分布的密度等高线族

- 图中展示了两种情况下的等高线族:
  - $\rho > 0$: 椭圆的长轴大致沿 $y = x$ 方向 (或 $x_2 = kx_1$，$k > 0$)，表示 X_1, X_2 正相关。
  - $\rho < 0$: 椭圆的长轴大致沿 $y = -x$ 方向 (或 $x_2 = kx_1$，$k < 0$)，表示 X_1, X_2 负相关。
    等高线越往外，对应的 c^2 越大，密度值越小。

幻灯片 29: 多元正态分布的性质 (1) 线性变换

- **(1) 设 $X \sim N_p(\mu, \Sigma)$，$Y = CX + b$，其中 C 为 $r \times p$ 常数矩阵，b 为 r 维常数向量，则:**
  $Y \sim N_r(C\mu + b, C\Sigma C')$
  这是多元正态分布一个非常重要的性质：**正态变量的线性变换仍为正态变量。**
  如果 $C\Sigma C'$ 是奇异的，则 Y 服从退化的 r 元正态分布。
- **特别地：设 $X \sim N_p(\mu, \Sigma)$，a 为 p 维常数向量，则 $a'X$ (一个标量) 服从一元正态分布:**
  $a'X \sim N(a'\mu, a'\Sigma a)$
  (这里 $C = a', b = 0$)。如果 $a'\Sigma a = 0$，则 $a'X$ 是一个常数 $a'\mu$ (以概率1)。

幻灯片 30: 多元正态分布的性质 (2) 边缘分布

- **(2) 设 $X \sim N_p(\mu, \Sigma)$，则 X 的任何子向量也服从 (多元) 正态分布，其均值为 μ 的相应子向量，协方差矩阵为 Σ 的相应子矩阵。**
  对 $X, \mu, \Sigma(> 0)$ 作如下的剖分:
  $$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}$$
  $q \quad p-q$
  则:
  $X^{(1)} \sim N_q(\mu^{(1)}, \Sigma_{11})$
  $X^{(2)} \sim N_{p-q}(\mu^{(2)}, \Sigma_{22})$
  这个性质说明：**多元正态分布的任何边缘分布仍为 (多元) 正态分布。**
  例如，$X^{(1)}$ 可以通过令 $C = [I_q \vdots \mathbf{0}_{q\times(p-q)}]$ 得到，$Y = X^{(1)} = CX$。

幻灯片 31: 多元正态分布的性质 (3) 独立性

- **(3) 子向量 $X^{(1)}$ 和 $X^{(2)}$ 相互独立，当且仅当 $\Sigma_{12} = 0$。**
  回顾幻灯片7，$\Sigma_{12} = \text{Cov}(X^{(1)}, X^{(2)})$。
  所以，对于多元正态分布，**不相关 ($\Sigma_{12} = 0$) 等价于相互独立。** 这是正态分布特有的重要性质。
  一般情况下，不相关不意味着独立，但对于正态分布，它们是等价的。
  $\leftarrow$ 箭头表示 $\Sigma_{12} = 0 \implies$ 独立。
  如果 $\Sigma_{12} = 0$ (从而 $\Sigma_{21} = 0'$), 则协方差矩阵 $\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$。
  $|\Sigma| = |\Sigma_{11}||\Sigma_{22}|$。
  $\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$。
  指数项 $(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$ 可以分解为两部分之和:

$(\mathbf{x}^{(1)} - \mu^{(1)})'\Sigma_{11}^{-1}(\mathbf{x}^{(1)} - \mu^{(1)}) + (\mathbf{x}^{(2)} - \mu^{(2)})'\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \mu^{(2)})$。
因此，联合密度函数 $f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ 可以分解为 $f_1(\mathbf{x}^{(1)})f_2(\mathbf{x}^{(2)})$ 的形式，表明 $X^{(1)}$ 和 $X^{(2)}$ 独立。
$f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = (2\pi)^{-p/2}|\Sigma_{11}|^{-1/2}|\Sigma_{22}|^{-1/2} \exp\left[-\frac{1}{2}(\ldots)\right]$
$= \left\{(2\pi)^{-q/2}|\Sigma_{11}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}^{(1)} - \mu^{(1)})'\Sigma_{11}^{-1}(\mathbf{x}^{(1)} - \mu^{(1)})\right]\right\} \times \left\{(2\pi)^{-(p-q)/2}|\Sigma_{22}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}^{(2)} - \mu^{(2)})'\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \mu^{(2)})\right]\right\}$

幻灯片 32: 多元正态分布的性质 (推广与注意)

- 若 $X = (X_1, X_2, \ldots, X_p)' \sim N_p(\mu, \Sigma)$，则 $X_1, X_2, \ldots, X_p$ 相互独立当且仅当 $\Sigma_{ij} = 0$ 对一切 $i \neq j$ 成立 (即 Σ 为对角矩阵)。
- **即多元正态变量，其子向量之间互不相关和相互独立是等价的。**
  例： 设 $X \sim N_3(\mu, \Sigma)$，其中
  $\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 5 & -1 \\ 0 & -1 & 1 \end{pmatrix}$
  - $\text{Cov}(X_2, X_3) = \Sigma_{23} = -1 \neq 0$，所以 X_2 和 X_3 不独立 (也不相关)。
  - X_1 和 $(X_2, X_3)'$ 对应的协方差子矩阵 $\Sigma_{1,(23)} = \begin{pmatrix} \Sigma_{12} \\ \Sigma_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$。所以 X_1 和 $(X_2, X_3)'$ 相互独立。
- **需注意：随机向量的任何边缘分布皆为 (多元) 正态分布未必表明该随机向量就服从多元正态分布。**
  即，所有分量 X_i 都是正态的，甚至所有子向量 $X^{(k)}$ 都是多元正态的，也不能保证 X 本身是多元正态的。PPT中提到 "[补充例P15] 就是这样的一个反例"。

幻灯片 33: 例子 (边缘分布)

设 $X \sim N_4(\mu, \Sigma)$，这里
$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}$
则:

- 一维边缘分布：$X_i \sim N(\mu_i, \sigma_{ii}), \quad i = 1, 2, 3, 4$
- 二维边缘分布：例如，向量 $\begin{pmatrix} X_1 \\ X_4 \end{pmatrix}$ 服从
  $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_4 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{14} \\ \sigma_{41} & \sigma_{44} \end{pmatrix}\right)$
- 三维边缘分布：例如，向量 $\begin{pmatrix} X_4 \\ X_1 \\ X_3 \end{pmatrix}$ (注意顺序) 服从
  $N_3\left(\begin{pmatrix} \mu_4 \\ \mu_1 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{44} & \sigma_{41} & \sigma_{43} \\ \sigma_{14} & \sigma_{11} & \sigma_{13} \\ \sigma_{34} & \sigma_{31} & \sigma_{33} \end{pmatrix}\right)$

幻灯片 34-37: 例子 (线性变换和分块)

【例 2.4】若 $X = (X_1, X_2, X_3)' \sim N_3(\mu, \Sigma)$
其中，
$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$
设 $a = (0, 1, 0)', A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$。

(1) (幻灯片 35) $a'X$ 的分布
$a'X = (0, 1, 0) \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = X_2$
根据性质 (1) 特例：$a'X \sim N(a'\mu, a'\Sigma a)$。
$a'\mu = (0, 1, 0) \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \mu_2$

$$a'\Sigma a = (0,1,0)\begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix} = (0,1,0)\begin{pmatrix}\sigma_{12}\\ \sigma_{22}\\ \sigma_{32}\end{pmatrix} = \sigma_{22}$$

所以 $X_2 \sim N(\mu_2, \sigma_{22})$，这与边缘分布的结论一致。

(2) (幻灯片 36) AX 的分布

$$AX = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & -1\end{pmatrix}\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix} = \begin{pmatrix}X_1\\ -X_3\end{pmatrix}$$

根据性质 (1)： $AX \sim N_2(A\mu, A\Sigma A')$。

$$A\mu = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & -1\end{pmatrix}\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix} = \begin{pmatrix}\mu_1\\ -\mu_3\end{pmatrix}$$

$$A\Sigma A' = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & -1\end{pmatrix}\begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\\ 0 & -1\end{pmatrix}$$

$$= \begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ -\sigma_{31} & -\sigma_{32} & -\sigma_{33}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\\ 0 & -1\end{pmatrix} = \begin{pmatrix}\sigma_{11} & -\sigma_{13}\\ -\sigma_{31} & \sigma_{33}\end{pmatrix}$$

所以 $\begin{pmatrix}X_1\\ -X_3\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu_1\\ -\mu_3\end{pmatrix}, \begin{pmatrix}\sigma_{11} & -\sigma_{13}\\ -\sigma_{31} & \sigma_{33}\end{pmatrix}\right)$。

(3) (幻灯片 37) 子向量的分布

记 $X = \begin{pmatrix}X^{(1)}\\ X^{(2)}\end{pmatrix}$，其中 $X^{(1)} = \begin{pmatrix}X_1\\ X_2\end{pmatrix}$ (2维)，$X^{(2)} = (X_3)$ (1维)。

$\mu = \begin{pmatrix}\mu^{(1)}\\ \mu^{(2)}\end{pmatrix}$，其中 $\mu^{(1)} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \mu^{(2)} = (\mu_3)$。

$\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$，其中

$\Sigma_{11} = \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}, \Sigma_{12} = \begin{pmatrix}\sigma_{13}\\ \sigma_{23}\end{pmatrix}, \Sigma_{21} = (\sigma_{31}, \sigma_{32}), \Sigma_{22} = (\sigma_{33})$。

则 $X^{(1)} = \begin{pmatrix}X_1\\ X_2\end{pmatrix} \sim N_2(\mu^{(1)}, \Sigma_{11}) = N_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}\right)$。

幻灯片 38: 边缘分布的应用与提醒

- **注意到：如果随机向量服从多元正态分布，则它的每个分量必服从一元正态分布。**
- **把某个分量的 n 个样本值作频率直方图，如果断定不呈正态分布，则就可以断定随机向量也不可能服从多元正态分布。**
  这是检验多元正态性的一个必要条件，但不是充分条件。即所有分量都服从一元正态分布，并不能保证整体服从多元正态分布。
- **后续的探索性数据分析中再详加讨论!**

第三部分：多元正态分布的参数估计 (幻灯片 39-59)

幻灯片 39: 多元正态分布的参数估计 (概述)

本部分将讨论以下内容：

- 多元样本 $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 的联合概率密度
- μ 和 Σ 的极大似然估计
- 相关系数的极大似然估计
- 估计量的性质
- Wishart 分布

幻灯片 40: 简单随机样本和数据矩阵

- 设 $X \sim N_p(\mu, \Sigma), \Sigma > 0$。 $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 是从总体 X 中抽取的一个简单随机样本，即满足：
  1. $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ 相互独立。
  2. 每个 $X_{(a)}$ $(a = 1, \ldots, n)$ 都与总体 X 同分布，即 $X_{(a)} \sim N_p(\mu, \Sigma)$。

- **令数据矩阵 (样本资料阵) 为:**

$$\mathbf{X}_{n\times p} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} = \begin{pmatrix} X'_{(1)} \\ X'_{(2)} \\ \vdots \\ X'_{(n)} \end{pmatrix}$$

这里 $X_{(a)} = (X_{a1}, X_{a2}, \dots, X_{ap})'$ 是第 a 个观测向量。
(幻灯片中 $X_{(1)}, \dots, X_p$ 作为列向量，这似乎是把数据矩阵的列向量表示成 $X_j = (X_{1j}, \dots, X_{nj})'$。但通常 $(X_1, \dots, X_p)$ 表示一个 p 维随机向量的 p 个分量，而 $X_{(a)}$ 表示第 a 个样本观测。从 $X'_{(a)}$ 的表示来看，行是样本，列是变量/维度。)
幻灯片右侧的 $(X_1, X_2, \dots, X_p)$ 应该是指数据矩阵的 p 个列向量，每个列向量 X_j 包含该变量的 n 次观测。
而等号后的 $\begin{pmatrix} X'_{(1)} \\ \vdots \\ X'_{(n)} \end{pmatrix}$ 是标准的样本数据矩阵表示，每一行是一个 p 维观测。

幻灯片 41: 样本数据数字特征 (样本均值向量)

- 设 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 为来自 p 元总体的样本，其中 $X_{(a)} = (X_{a1}, X_{a2}, \dots, X_{ap})'$，$a = 1, 2, \dots, n$。
- **样本均值向量定义为:**
  $\bar{X} = \frac{1}{n} \sum_{a=1}^{n} X_{(a)} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)'$
  其中 $\bar{X}_j = \frac{1}{n} \sum_{a=1}^{n} X_{aj}$ 是第 j 个分量的样本均值。
  矩阵表示:
  $$\bar{X} = \frac{1}{n} \begin{pmatrix} \sum X_{a1} \\ \sum X_{a2} \\ \vdots \\ \sum X_{ap} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} X_{11} + X_{21} + \dots + X_{n1} \\ X_{12} + X_{22} + \dots + X_{n2} \\ \vdots \\ X_{1p} + X_{2p} + \dots + X_{np} \end{pmatrix}$$
  如果用数据矩阵 $\mathbf{X} = \begin{pmatrix} X'_{(1)} \\ \vdots \\ X'_{(n)} \end{pmatrix}$，则 $\bar{X} = \frac{1}{n} \mathbf{X}' \mathbf{1}_n$，其中 $\mathbf{1}_n$ 是 n 维全1列向量。

幻灯片 42: 样本数据数字特征 (样本离差阵)

- **样本离差阵 (Sum of Squares and Cross-products matrix, SSCP) 定义为:**
  $S_{p\times p}(\text{或 } A_{p\times p}) = \sum_{a=1}^{n} (X_{(a)} - \bar{X})(X_{(a)} - \bar{X})' = (s_{ij})_{p\times p}$
  其中 $s_{ij} = \sum_{a=1}^{n} (X_{ai} - \bar{X}_i)(X_{aj} - \bar{X}_j)$。
  这是一个 $p \times p$ 的对称矩阵。
  对角线元素 $s_{ii} = \sum_{a=1}^{n} (X_{ai} - \bar{X}_i)^2$ 是第 i 个分量的离差平方和。
  非对角线元素 $s_{ij} = \sum_{a=1}^{n} (X_{ai} - \bar{X}_i)(X_{aj} - \bar{X}_j)$ 是第 i 和第 j 分量的离差乘积之和。

幻灯片 43: 样本数据数字特征 (样本协差阵)

- **样本协差阵定义为:**
  $V_{p\times p} = \frac{1}{n-1} S = \frac{1}{n-1} \sum_{a=1}^{n} (X_{(a)} - \bar{X})(X_{(a)} - \bar{X})' = (v_{ij})_{p\times p}$
  其中 $v_{ij} = \frac{1}{n-1} \sum_{a=1}^{n} (X_{ai} - \bar{X}_i)(X_{aj} - \bar{X}_j)$。
  V 是总体协方差矩阵 Σ 的无偏估计 (如果样本来自同一总体)。
  (有时也会用 $1/n$ 作为分母，称为 Σ 的极大似然估计，记为 $\hat{\Sigma}_{MLE}$ 或 $\tilde{S}$。)

幻灯片 44: 样本数据数字特征 (样本相关阵)

- **样本相关阵定义为:**
  $R_{p\times p} = (r_{ij})_{p\times p}$
  其中 r_{ij} 是第 i 个分量和第 j 个分量的样本相关系数:
  $r_{ij} = \frac{v_{ij}}{\sqrt{v_{ii}}\sqrt{v_{jj}}} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$
  R 是总体相关矩阵的估计。对角线元素 $r_{ii} = 1$。

幻灯片 45: 样本均值向量用样本资料阵表示

- 令 $\mathbf{1}_n = (1, 1, \ldots, 1)'$ 为 n 维全1列向量。
  数据矩阵 $\mathbf{X}_{n\times p} = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{pmatrix} = \begin{pmatrix} X'_{(1)} \\ \vdots \\ X'_{(n)} \end{pmatrix}$。
  则样本均值向量 $\bar{X}_{p\times 1}$:
  $\bar{X} = \dfrac{1}{n} \begin{pmatrix} \sum X_{a1} \\ \vdots \\ \sum X_{ap} \end{pmatrix} = \dfrac{1}{n}\mathbf{X}'\mathbf{1}_n$
  幻灯片上的展开形式:
  $\bar{X}_{p\times 1} = \dfrac{1}{n} \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1p} & X_{2p} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \dfrac{1}{n}\mathbf{X}'_{\text{cols as vars}}\mathbf{1}_n$ (这里$\mathbf{X}_{\text{cols as vars}}$是 $n\times p$ 数据矩阵，列是变量，行是样本)
  或者 $\bar{X}_{p\times 1} = \frac{1}{n}[X_{(1)}, X_{(2)}, \ldots, X_{(n)}]\mathbf{1}_n$ (这里 $X_{(a)}$ 是列向量)
  $\bar{X} = \frac{1}{n}\sum X_{(a)}$
  幻灯片最后 $\frac{1}{n}\mathbf{X}'\mathbf{1}_n$ 是正确的，如果 $\mathbf{X}$ 定义为行是样本，列是变量的 $n\times p$ 矩阵。

幻灯片 46: 样本离差阵用样本资料阵表示

- 样本离差阵 $S = \sum_{a=1}^{n}(X_{(a)} - \bar{X})(X_{(a)} - \bar{X})'$
  可以展开为:
  $S = \sum_{a=1}^{n} X_{(a)}X'_{(a)} - \sum_{a=1}^{n} X_{(a)}\bar{X}' - \sum_{a=1}^{n} \bar{X}X'_{(a)} + \sum_{a=1}^{n} \bar{X}\bar{X}'$
  $S = \sum_{a=1}^{n} X_{(a)}X'_{(a)} - (\sum_{a=1}^{n} X_{(a)})\bar{X}' - \bar{X}(\sum_{a=1}^{n} X_{(a)})' + n\bar{X}\bar{X}'$
  由于 $\sum X_{(a)} = n\bar{X}$:
  $S = \sum_{a=1}^{n} X_{(a)}X'_{(a)} - n\bar{X}\bar{X}' - n\bar{X}\bar{X}' + n\bar{X}\bar{X}'$
  $S = \sum_{a=1}^{n} X_{(a)}X'_{(a)} - n\bar{X}\bar{X}'$
  如果 $\mathbf{X}$ 是 $n\times p$ 数据矩阵 (行是样本 $X'_{(a)}$，列是变量)，则 $\sum X_{(a)}X'_{(a)} = \mathbf{X}'\mathbf{X}$ (这里需要 $X_{(a)}$ 是列向量，$\mathbf{X}$ 的列是 $X_{(a)}$，即 $p\times n$ 矩阵)。
  如果 $\mathbf{X}_{n\times p} = \begin{pmatrix} X'_{(1)} \\ \vdots \\ X'_{(n)} \end{pmatrix}$, 那么 $\mathbf{X}'\mathbf{X} = \sum_{a=1}^{n} X_{(a)}X'_{(a)}$。
  并且 $\bar{X} = \frac{1}{n}\mathbf{X}'\mathbf{1}_n$。
  $n\bar{X}\bar{X}' = n(\frac{1}{n}\mathbf{X}'\mathbf{1}_n)(\frac{1}{n}\mathbf{1}'_n\mathbf{X}) = \frac{1}{n}\mathbf{X}'\mathbf{1}_n\mathbf{1}'_n\mathbf{X}$
  所以 $S = \mathbf{X}'\mathbf{X} - \frac{1}{n}\mathbf{X}'\mathbf{1}_n\mathbf{1}'_n\mathbf{X} = \mathbf{X}'(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n)\mathbf{X}$
  $I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$ 是一个中心化矩阵。
  (幻灯片中的 $X'X - n\bar{X}\bar{X}'$ 是正确的，如果 X 是 $p\times n$ 矩阵，列为样本 $X_{(a)}$。如果 X 是 $n\times p$ 矩阵，行为样本 $X'_{(a)}$，则 $S = X'X - n\bar{X}\bar{X}'$ 也是对的，但 $\bar{X}$ 的定义要匹配。)
  更清晰的写法是 $S = \sum X_{(a)}X'_{(a)} - n\bar{X}\bar{X}'$。

幻灯片 47: 均值向量与协差阵的极大似然估计 (概述)

- 多元正态分布有两组参数：均值 μ 和协差阵 Σ。
- 在许多问题中它们是未知的，需要通过样本数据来估计。
- 参数估计方法：极大似然估计法 (MLE)。

幻灯片 48: 多元样本的似然函数

- **极大似然估计** 是通过似然函数来求得的。
- **似然函数** 可以是样本联合概率密度 $f(\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(n)})$ 的任意正常数倍，记为 $L(\mu, \Sigma)$。

- **多元正态假定下:**
由于样本 $X_{(1)},\dots,X_{(n)}$ 独立同分布于 $N_p(\mu,\Sigma)$，联合密度函数 (即似然函数) 为:
$L(\mu,\Sigma)=f(\mathbf{x}_{(1)},\dots,\mathbf{x}_{(n)})=\prod_{a=1}^n f(\mathbf{x}_{(a)})$
$=\prod_{a=1}^n(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{x}_{(a)}-\mu)'\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)\right]$
$=(2\pi)^{-np/2}|\Sigma|^{-n/2}\exp\left[-\frac{1}{2}\sum_{a=1}^n(\mathbf{x}_{(a)}-\mu)'\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)\right]$

幻灯片 49: 极大似然估计 (对数似然和矩阵导数)

- **对数似然函数:**
$\ln L(\mu,\Sigma)=-\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\Sigma|-\frac{1}{2}\sum_{a=1}^n(\mathbf{x}_{(a)}-\mu)'\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)$
- **根据矩阵代数有关运算 (矩阵求导法则):**
  - $\frac{\partial(x'Ax)}{\partial x}=2Ax$ (如果 A 对称) 或 $(A+A')x$ (一般情况)。对于 $(\mathbf{x}_{(a)}-\mu)'\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)$ 对 μ 求导，这里 $A=\Sigma^{-1}$ (对称)，$x=(\mathbf{x}_{(a)}-\mu)$，对 μ 求导会多一个负号。
  - $\frac{\partial\mathrm{tr}(AB)}{\partial A}=B'$。
  - $\frac{\partial\ln|A|}{\partial A}=(A^{-1})'$ (如果 A 对称，则是 A^{-1})。
  - 幻灯片上的 $\frac{\partial(X'AX)}{\partial X}=2AX$ (假设 A 对称)。
  - $\frac{\partial\mathrm{tr}(X'AX)}{\partial A}=XX'$ (这里 X 被看作常数)。实际上是 $\frac{\partial\mathbf{x}'A\mathbf{x}}{\partial A}=\mathbf{x}\mathbf{x}'$。
  - $\frac{\partial\ln|A|}{\partial A}=A^{-1}$ (如果 A 对称且可逆)。

  指数项可以写成迹的形式: $(\mathbf{x}_{(a)}-\mu)'\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)=\mathrm{tr}((\mathbf{x}_{(a)}-\mu)'\Sigma^{-1}(\mathbf{x}_{(a)}-\mu))=\mathrm{tr}(\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)(\mathbf{x}_{(a)}-\mu)')$.

幻灯片 50: 极大似然估计 (结果)

对 $\ln L(\mu,\Sigma)$ 分别对 μ 和 Σ^{-1} (或 Σ) 求导，并令导数为0。

- **对 μ 求导并令其为0:**
$\frac{\partial\ln L(\mu,\Sigma)}{\partial\mu}=-\frac{1}{2}\sum_{a=1}^n[-2\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)]=\sum_{a=1}^n\Sigma^{-1}(\mathbf{x}_{(a)}-\mu)=0$
由于 Σ^{-1} 可逆，所以 $\sum_{a=1}^n(\mathbf{x}_{(a)}-\mu)=0 \implies n\bar{X}-n\mu=0 \implies \hat{\mu}=\bar{X}$。
$\hat{\mu}_{MLE}=\bar{X}=\frac{1}{n}\sum_{a=1}^n X_{(a)}$
- **对 Σ^{-1} 求导并令其为0 (或用更复杂的对 Σ 求导):**
$\frac{\partial\ln L(\mu,\Sigma)}{\partial\Sigma}=-\frac{n}{2}\Sigma^{-1}+\frac{1}{2}\Sigma^{-1}\left(\sum_{a=1}^n(\mathbf{x}_{(a)}-\mu)(\mathbf{x}_{(a)}-\mu)'\right)\Sigma^{-1}=0$
(这里使用了 $\frac{\partial\ln|\Sigma|}{\partial\Sigma}=\Sigma^{-1}$ 和 $\frac{\partial\mathrm{tr}(A\Sigma^{-1}B)}{\partial\Sigma}=-(\Sigma^{-1}BA\Sigma^{-1})'$。)
这会导致 $\hat{\Sigma}_{MLE}=\frac{1}{n}\sum_{a=1}^n(X_{(a)}-\hat{\mu})(X_{(a)}-\hat{\mu})'=\frac{1}{n}S$ (其中 S 是样本离差阵)。
$\hat{\Sigma}_{MLE}=\frac{1}{n}S=\frac{1}{n}\sum_{a=1}^n(X_{(a)}-\bar{X})(X_{(a)}-\bar{X})'$
- **多元正态总体的均值向量的极大似然估计量就是样本均值向量。**
- **协差阵的极大似然估计就是样本离差阵 S/n。** (这是一个有偏估计，但具有一致性)

幻灯片 51: 估计量的性质

μ 和 Σ 的估计量有如下基本性质:

1. $E(\bar{X})=\mu$，即 $\bar{X}$ 是 μ 的无偏估计。
$E(\frac{1}{n}S)=\frac{n-1}{n}\Sigma$，即 $\frac{1}{n}S$ 不是 Σ 的无偏估计 (是有偏的)。
而 $E(\frac{1}{n-1}S)=\Sigma$，即样本协方差矩阵 $V=\frac{1}{n-1}S$ 是 Σ 的无偏估计。
2. $\bar{X},\frac{1}{n-1}S$ 分别是 μ,Σ 的有效估计 (最小方差无偏估计，MVUE，在某些条件下)。
3. $\bar{X},\frac{1}{n}S$ (或 $\frac{1}{n-1}S$) 分别是 μ,Σ 的一致估计 (强相合估计)。
一致性意味着当样本量 $n\to\infty$ 时，估计量收敛到真值。

幻灯片 52: 估计量性质的证明 (部分)

- $E(\bar{X})=E\left[\frac{1}{n}\sum_a X_{(a)}\right]=\frac{1}{n}\sum_a E[X_{(a)}]=\frac{1}{n}\sum_a\mu=\frac{1}{n}n\mu=\mu$

- $E(\frac{1}{n}S) = \frac{1}{n}E\left[\sum_a (X_{(a)} - \bar{X})(X_{(a)} - \bar{X})'\right]$
  这里用了一个技巧:
  $S = \sum_a (X_{(a)} - \mu - (\bar{X} - \mu))(X_{(a)} - \mu - (\bar{X} - \mu))'$
  $= \sum_a (X_{(a)} - \mu)(X_{(a)} - \mu)' - n(\bar{X} - \mu)(\bar{X} - \mu)'$
  $E(S) = \sum_a E[(X_{(a)} - \mu)(X_{(a)} - \mu)'] - nE[(\bar{X} - \mu)(\bar{X} - \mu)']$
  $E[(X_{(a)} - \mu)(X_{(a)} - \mu)'] = D(X_{(a)}) = \Sigma$
  $E[(\bar{X} - \mu)(\bar{X} - \mu)'] = D(\bar{X}) = D(\frac{1}{n}\sum X_{(a)}) = \frac{1}{n^2}\sum D(X_{(a)})$ (因为独立) $= \frac{1}{n^2}n\Sigma = \frac{1}{n}\Sigma$
  所以 $E(S) = n\Sigma - n(\frac{1}{n}\Sigma) = n\Sigma - \Sigma = (n-1)\Sigma$。
  因此 $E(\frac{1}{n}S) = \frac{n-1}{n}\Sigma$。
  幻灯片中的推导步骤:
  $E(\frac{1}{n}S) = \frac{1}{n}E[\sum_a (X_{(a)} - \mu)(X_{(a)} - \mu)' - n(\bar{X} - \mu)(\bar{X} - \mu)']$
  $= \frac{1}{n}[\sum_a D(X_{(a)}) - nD(\bar{X})]$
  $= \frac{1}{n}[n\Sigma - n(\frac{1}{n}\Sigma)] = \frac{1}{n}[n\Sigma - \Sigma] = \frac{n-1}{n}\Sigma$。
  这个推导是正确的。

幻灯片 53: 估计量的性质 (定理 2.2)

定理 2.2 设 $\bar{X}$ 和 S 分别是正态总体 $N_p(\mu, \Sigma)$ 的样本均值向量和离差阵，则:

1. $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$
   (因为 $\bar{X}$ 是独立同正态分布随机向量的线性组合，所以仍服从正态分布。其均值和协方差矩阵如上页计算)。
2. 离差阵 S 可以写为 $S = \sum_{a=1}^{n-1} Z_a Z_a'$
   其中，$Z_1, \dots, Z_{n-1}$ 独立同分布于 $N_p(\mathbf{0}, \Sigma)$。
   这表明 S 服从 Wishart 分布 $W_p(n-1, \Sigma)$ (见后续幻灯片)。
3. $\bar{X}$ 和 S 相互独立。
   这是正态样本的一个非常重要的性质，类似于一元情况下样本均值和样本方差的独立性。
4. S 为正定阵的充要条件是 $n > p$ (或 $n-1 \geq p$ 使得Wishart分布非退化)。
   如果 $n \leq p$，则 S 几乎必然是奇异的。

幻灯片 54: 相关系数的极大似然估计

- 相关系数 $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$ 的极大似然估计为:
  利用极大似然估计的不变性，将 Σ 的MLE $\hat{\Sigma}_{MLE} = \frac{1}{n}S = (\frac{s_{ij}}{n})$ 代入。
  $\hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} = \frac{s_{ij}/n}{\sqrt{s_{ii}/n}\sqrt{s_{jj}/n}} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$
  这与样本相关系数 r_{ij} 的定义一致。
  $\hat{r}_{ij} = \frac{\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)}{\sqrt{\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)^2}\sqrt{\sum_{k=1}^{n}(X_{kj} - \bar{X}_j)^2}}$
- 其中 $\hat{\Sigma} = (\hat{\sigma}_{ij})$ 是 Σ 的MLE，$\bar{X} = (\bar{X}_1, \dots, \bar{X}_p)'$, $S = (s_{ij})$。
- S 为样本离差矩阵，r_{ij} 为样本相关系数，$R = (r_{ij})$ 为样本相关矩阵。

幻灯片 55: $\bar{X}$ 的抽样分布

- **正态总体:**
  设 $X \sim N_p(\mu, \Sigma), \Sigma > 0$。$X_{(1)}, \dots, X_{(n)}$ 是来自该总体的样本，则
  $\bar{X} \sim N_p\left(\mu, \frac{1}{n}\Sigma\right)$
  (已在定理2.2中给出)
- **非正态总体 (中心极限定理):**
  设 $X_{(1)}, \dots, X_{(n)}$ 是来自某总体 X (不一定是正态的，但有均值 μ 和协方差矩阵 Σ) 的一个样本。
  则当 n 很大且相对于 p 也很大时，上述 $\bar{X} \sim N_p\left(\mu, \frac{1}{n}\Sigma\right)$ 近似地成立。
  这是多元中心极限定理。

幻灯片 56: S 的分布——Wishart分布 (定义)

- 维希特 (Wishart) 分布是 Wishart 在1928年推导出来的。
- 是一元 χ^2 (卡方) 分布的推广，也是构成其它重要分布 (如Hotelling's T^2) 的基础。
- **定义2.10：** 设 $X_{(a)} = (X_{a1}, \ldots, X_{ap})' \sim N_p(\mu_a, \Sigma), a = 1, 2, \ldots, n$，且相互独立。
  则由 $X_{(a)}$ 组成的随机矩阵：
  $W_{p\times p} = \sum_{a=1}^n X_{(a)} X_{(a)}'$
  的分布称为 **非中心Wishart分布**，记为 $W_p(n, \Sigma, Z)$ (这里的 Z 是非中心参数矩阵，与 μ_a 有关)。
  非中心参数矩阵 $Z = (\sum_{a=1}^n \mu_{a,i}\mu_{a,j})$ 或更通常地通过 $\mathbf{M}'\mathbf{M}$ 定义，其中 $\mathbf{M}_{n\times p} = (\mu_1, \ldots, \mu_n)'$。
  幻灯片中 $Z = (\mu_{a1}, \ldots, \mu_{ap})(\mu_{a1}, \ldots, \mu_{ap})'$ 应该是 $Z = \sum_{a=1}^n \mu_a \mu_a'$。

幻灯片 57: Wishart分布 (中心Wishart)

- 当 $\mu_a = \mathbf{0}$ 对所有 a 成立时 (即 $X_{(a)} \sim N_p(\mathbf{0}, \Sigma)$)，称为 **中心Wishart分布**，记为 $W_p(n, \Sigma)$。
  此时 $W = \sum_{a=1}^n X_{(a)} X_{(a)}'$。
  这里的 n 称为自由度。
- 当 $n \geq p, \Sigma > 0$，$W_p(n, \Sigma)$ 有密度存在，其表达式为：
  $f(w) = \frac{|w|^{(n-p-1)/2} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}w))}{2^{np/2}\pi^{p(p-1)/4}|\Sigma|^{n/2}\prod_{i=1}^p \Gamma(\frac{n-i+1}{2})}$
  当 w 为正定阵时。其它情况 $f(w) = 0$。
  $\Gamma(\cdot)$ 是伽马函数。
- **显然，当 $p = 1, \Sigma = \sigma^2$ 时，$X_{(a)}$ 是一维正态 $N(0, \sigma^2)$。**
  $W = \sum_{a=1}^n X_{(a)}^2$。
  $\frac{W}{\sigma^2} = \sum_{a=1}^n (X_{(a)}/\sigma)^2$。由于 $X_{(a)}/\sigma \sim N(0,1)$，所以 $\frac{W}{\sigma^2} \sim \chi^2(n)$。
  $f(w)$ 就是 $\sigma^2\chi^2(n)$ 的分布密度 (经过变量变换 $w \to w/\sigma^2$)。
- **因此，Wishart分布是 χ^2 分布在 p 维正态情况下的推广。**
  样本离差阵 $S = \sum_{a=1}^{n-1} Z_a Z_a'$ (定理2.2)，其中 $Z_a \sim N_p(\mathbf{0}, \Sigma)$ i.i.d.，所以 $S \sim W_p(n-1, \Sigma)$。

幻灯片 58: Wishart分布的基本性质

1. 若 $X_{(a)} \sim N_p(\mu, \Sigma), a = 1, \ldots, n$ 且相互独立，则样本离差阵
   $S = \sum_{a=1}^n (X_{(a)} - \bar{X})(X_{(a)} - \bar{X})' \sim W_p(n-1, \Sigma)$。
   其中 $\bar{X} = \frac{1}{n}\sum X_{(a)}$。
   自由度为 $n-1$ 而不是 n 是因为减去了样本均值 $\bar{X}$，损失了一个自由度。
2. 若 $S_i \sim W_p(n_i, \Sigma), i = 1, \ldots, k$，且相互独立，则
   $\sum_{i=1}^k S_i \sim W_p\left(\sum_{i=1}^k n_i, \Sigma\right)$
   Wishart 分布具有可加性 (对于相同的 Σ)。
3. 若 $X_{p\times p} \sim W_p(n, \Sigma), C_{p\times p}$ 为非奇异阵 (可逆矩阵)，则
   $CXC' \sim W_p(n, C\Sigma C')$
   这是 Wishart 分布的线性变换性质。

幻灯片 59: 随机向量 $\to$ 随机矩阵的分布

- 可以利用已知向量分布的定义给出矩阵分布的定义。
- 设随机矩阵：
  $$X_{n\times p} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}$$
- 将该矩阵的列向量 (或行向量) 一个接一个地连接起来，组成一个长的向量，即拉直向量 (vectorization, vec operator)：
  $\mathrm{vec}(X) = (X_{11}, X_{21}, \ldots, X_{n1}, X_{12}, X_{22}, \ldots, X_{n2}, \ldots, X_{1p}, X_{2p}, \ldots, X_{np})'$ (按列拉直)
  这个 $np \times 1$ 维的拉直向量的分布，定义为该阵的分布。
  例如，如果 $\mathrm{vec}(X)$ 服从 np 元正态分布，则称矩阵 X 服从矩阵正态分布。

幻灯片 60: 多元正态分布参数估计的计算机实现

这页只是一个标题页，表明后续可能涉及：

- (一) 均值向量的估计
- (二) 协差阵的估计
  在实际软件 (如 R, Python, SAS, SPSS) 中如何计算这些估计量。

学习这些内容时，关键在于理解每个概念的定义、几何意义 (如协方差和相关性如何影响数据点的散布形状) 以及它们之间的联系 (如协方差阵与相关阵的关系，正态分布中独立与不相关的等价性等)。祝您学习顺利！