多元统计

陈崇双

西南交通大学数学学院统计系
ccsmars@swjtu.edu.cn

2018-2019学年

1. 多元正态分布
   - 多元分布基本概念
   - 统计距离和马氏距离
   - 多元正态分布
   - 均值向量和协方差矩阵估计
   - 常用抽样分布
   - exercise

本节基本要点：

随机向量

多元分布函数

分布函数和密度函数

独立性

数字特征

随机向量

假定讨论总体具有多个指标，不妨设为p个，为了避免偶然性，进行多次观察，不妨设为n次，即得到如下表形式的数据。

x_{11}	$\cdots$	x_{1j}	$\cdots$	x_{1p}
$\vdots$		$\vdots$		$\vdots$
x_{i1}	$\cdots$	x_{ij}	$\cdots$	x_{ip}
$\vdots$		$\vdots$		$\vdots$
x_{n1}	$\cdots$	x_{nj}	$\cdots$	x_{np}

在该表中，第i行$(x_{i1}, x_{i2}, \cdots, x_{ip})^\mathsf{T} \triangleq \mathbf{x}_{(i)}$表示第$i$个样本的观察值，$i = 1, 2, \cdots, n$。第$j$列$(x_{1j}, x_{2j}, \cdots, x_{nj})^\mathsf{T} \triangleq \mathbf{x}_j$表示第$j$个变量的$n$次观察值，$j = 1, 2, \cdots, p$。

样本观测资料可用矩阵表示

$$\begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p) = \begin{pmatrix} \mathbf{x}_{(1)}^{\mathsf{T}} \\ \mathbf{x}_{(2)}^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_{(n)}^{\mathsf{T}} \end{pmatrix}$$

分布函数与密度函数

回顾：一维随机变量的分布函数、概率分布、密度函数及其性质。

定义1

设$\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$为$p$维随机向量，即每个分量都是随机变量，多元分布函数定义为

$$F(\mathbf{x}) \triangleq F(x_1, x_2, \cdots, x_p) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_p \leqslant x_p)$$

其中$\mathbf{x} \triangleq (x_1, x_2, \cdots, x_p)^T \in \mathbb{R}^p$。

分布函数与密度函数

定义2

设p维随机向量$\mathbf{X}$的分布函数为$F(\mathbf{x})$，若存在非负（p维）函数$f(.)$，使得

$$F(\mathbf{x}) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p} f(t_1, t_2, \ldots, t_p)dt_1dt_2\ldots dt_p$$

则称$\mathbf{X}$为连续型随机向量，且称$\mathbf{X}$有密度函数$f(\mathbf{x})$。

分布函数与密度函数

定义2

设p维随机向量$\mathbf{X}$的分布函数为$F(\mathbf{x})$，若存在非负（p维）函数$f(.)$，使得

$$F(\mathbf{x}) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p} f(t_1, t_2, \ldots, t_p)dt_1dt_2\ldots dt_p$$

则称$\mathbf{X}$为连续型随机向量，且称$\mathbf{X}$有密度函数$f(\mathbf{x})$。

定理1

p维函数$f(.)$是$\mathbb{R}^p$中某个随机向量的密度函数，当且仅当

$$f(\mathbf{x}) \geq 0; \int_{\mathbb{R}^p} f(\mathbf{x})d\mathbf{x} = 1$$

定义3

若两个随机向量$\mathbf{X}$和$\mathbf{Y}$满足

$$P(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y}) = P(\mathbf{X} \leq \mathbf{x})P(\mathbf{Y} \leq \mathbf{y})$$

对于任意的$\mathbf{x}, \mathbf{y}$都成立，则称$\mathbf{X}$和$\mathbf{Y}$相互独立。

独立性

定义3

若两个随机向量$\mathbf{X}$和$\mathbf{Y}$满足

$$P(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y}) = P(\mathbf{X} \leq \mathbf{x})P(\mathbf{Y} \leq \mathbf{y})$$

对于任意的$\mathbf{x}, \mathbf{y}$都成立，则称$\mathbf{X}$和$\mathbf{Y}$相互独立。

命题1

若记$\mathbf{X}$和$\mathbf{Y}$的分布函数分别为$F(\mathbf{x}), G(\mathbf{y})$，密度函数分别为$f(\mathbf{x}), g(\mathbf{y})$，二者的联合分布函数和密度函数分别为$H(\mathbf{x}, \mathbf{y}), h(\mathbf{x}, \mathbf{y})$。则$\mathbf{X}$和$\mathbf{Y}$相互独立，当且仅当$H(\mathbf{x}, \mathbf{y}) = F(\mathbf{x})G(\mathbf{y})$或$h(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})g(\mathbf{y})$ 成立。

独立性

定义3

若两个随机向量$\mathbf{X}$和$\mathbf{Y}$满足

$$P(\mathbf{X} \le \mathbf{x}, \mathbf{Y} \le \mathbf{y}) = P(\mathbf{X} \le \mathbf{x})P(\mathbf{Y} \le \mathbf{y})$$

对于任意的$\mathbf{x}, \mathbf{y}$都成立，则称$\mathbf{X}$和$\mathbf{Y}$相互独立。

命题1

若记$\mathbf{X}$和$\mathbf{Y}$的分布函数分别为$F(\mathbf{x}), G(\mathbf{y})$，密度函数分别为$f(\mathbf{x}), g(\mathbf{y})$，二者的联合分布函数和密度函数分别为$H(\mathbf{x}, \mathbf{y}), h(\mathbf{x}, \mathbf{y})$。则$\mathbf{X}$和$\mathbf{Y}$相互独立，当且仅当$H(\mathbf{x}, \mathbf{y}) = F(\mathbf{x})G(\mathbf{y})$或$h(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})g(\mathbf{y})$ 成立。

注：随机向量的维数不一定相等；定义和性质可扩展到多个情形。

（1）均值向量

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^\mathsf{T}$为$p$维随机向量，若$\mathsf{E}(X_i)=\mu_i$存在，$i=1,2,\cdots,p$，则称$p$维向量$\mu=(\mu_1,\mu_2,\cdots,\mu_p)^\mathsf{T}$为$\mathbf{X}$的均值向量，简记为$\mathsf{E}(\mathbf{X})=\mu$。

（1）均值向量

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^\mathsf{T}$为$p$维随机向量，若$\mathsf{E}(X_i)=\mu_i$存在，$i=1,2,\cdots,p$，则称$p$维向量$\mu=(\mu_1,\mu_2,\cdots,\mu_p)^\mathsf{T}$为$\mathbf{X}$的均值向量，简记为$\mathsf{E}(\mathbf{X})=\mu$。

命题2

对于阶数适合的常矩阵A,B，随机向量均值满足

$$E(A\mathbf{X})=AE(\mathbf{X})$$
$$E(A\mathbf{X}B)=AE(\mathbf{X})B$$
$$E(A\mathbf{X}+B\mathbf{Y})=AE(\mathbf{X})+BE(\mathbf{Y})$$

（2）自协方差阵

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^{\mathsf{T}}$为$p$维随机向量，若协方差$\mathrm{cov}(X_i,X_j)$都存在，$i,j=1,2,\cdots,p$，则称$p$阶方阵

$$\begin{pmatrix} \mathrm{cov}(X_1,X_1) & \cdots & \mathrm{cov}(X_1,X_j) & \cdots & \mathrm{cov}(X_1,X_p) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_i,X_1) & \cdots & \mathrm{cov}(X_i,X_j) & \cdots & \mathrm{cov}(X_i,X_p) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_p,X_1) & \cdots & \mathrm{cov}(X_p,X_j) & \cdots & \mathrm{cov}(X_p,X_p) \end{pmatrix}$$

为$\mathbf{X}$的自协方差阵，记为$D(\mathbf{X})$。$|D(\mathbf{X})|$称之为$\mathbf{X}$的广义方差。

（2）自协方差阵

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^{\mathsf{T}}$为$p$维随机向量，若协方差$\mathrm{cov}(X_i,X_j)$都存在，$i,j=1,2,\cdots,p$，则称$p$阶方阵

$$\begin{pmatrix} \mathrm{cov}(X_1,X_1) & \cdots & \mathrm{cov}(X_1,X_j) & \cdots & \mathrm{cov}(X_1,X_p) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_i,X_1) & \cdots & \mathrm{cov}(X_i,X_j) & \cdots & \mathrm{cov}(X_i,X_p) \\ \vdots & & \vdots & & \vdots \\ \mathrm{cov}(X_p,X_1) & \cdots & \mathrm{cov}(X_p,X_j) & \cdots & \mathrm{cov}(X_p,X_p) \end{pmatrix}$$

为$\mathbf{X}$的自协方差阵，记为$D(\mathbf{X})$。$|D(\mathbf{X})|$称之为$\mathbf{X}$的广义方差。

注：自协方差阵对称、半正定，大多数情形下正定。

回顾：正定矩阵相关

定义4

A $p \times p$ *symmetric real matrix* M *is said to be*

- *positive definite, if* $X^{\mathsf{T}}MX > 0$ *for all* $X \in \mathbb{R}^n - \mathbf{0}$;
- *positive semi-definite, if* $X^{\mathsf{T}}MX \geq 0$ *for all* $X \in \mathbb{R}^n - \mathbf{0}$.

回顾：正定矩阵相关

定义4

A $p \times p$ symmetric real matrix M is said to be

- *positive definite, if $X^TMX > 0$ for all $X \in \mathbb{R}^n - \mathbf{0}$;*
- *positive semi-definite, if $X^TMX \geq 0$ for all $X \in \mathbb{R}^n - \mathbf{0}$.*

命题3

- *Every positive definite matrix is invertible.*
- *All eigenvalues of a positive definite matrix are positive.*
- *All eigenvalues of a positive semi-definite matrix are non-negative.*

https://en.wikipedia.org/wiki/Definiteness_of_a_matrix

（3）协差阵

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^{\mathsf{T}}$和$\mathbf{Y}=(Y_1,Y_2,\cdots,Y_q)^{\mathsf{T}}$分别为$p$维和$q$维的随机向量，若协方差$\mathrm{cov}(X_i,Y_j)$都存在，$i=1,2,\cdots,p$，$j=1,2,\cdots,q$，则称矩阵$(\mathrm{cov}(X_i,Y_j))_{p\times q}$为$\mathbf{X}$和$\mathbf{Y}$的协（方）差阵，记为$\mathrm{cov}(\mathbf{X},\mathbf{Y})$。特别地，若$\mathrm{cov}(\mathbf{X},\mathbf{Y})=\mathbf{0}$，则称$\mathbf{X}$和$\mathbf{Y}$不相关。

（3）协差阵

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^{\mathsf{T}}$和$\mathbf{Y}=(Y_1,Y_2,\cdots,Y_q)^{\mathsf{T}}$分别为$p$维和$q$维的随机向量，若协方差$\mathrm{cov}(X_i,Y_j)$都存在，$i=1,2,\cdots,p$，$j=1,2,\cdots,q$，则称矩阵$(\mathrm{cov}(X_i,Y_j))_{p\times q}$为$\mathbf{X}$和$\mathbf{Y}$的协（方）差阵，记为$\mathrm{cov}(\mathbf{X},\mathbf{Y})$。特别地，若$\mathrm{cov}(\mathbf{X},\mathbf{Y})=\mathbf{0}$，则称$\mathbf{X}$和$\mathbf{Y}$不相关。

命题4

对于阶数适合的常矩阵A,B，协差阵满足

$$
\begin{aligned}
&D(\mathbf{X})=E((\mathbf{X}-E(\mathbf{X}))(\mathbf{X}-E(\mathbf{X}))^T)=E(\mathbf{X}\mathbf{X}^T)-E(\mathbf{X})E(\mathbf{X})^T\\
&D(A\mathbf{X})=AD(\mathbf{X})A^T\\
&cov(A\mathbf{X},B\mathbf{Y})=Acov(\mathbf{X},\mathbf{Y})B^T\\
&E(\mathbf{X}^TA\mathbf{X})=tr(AD(\mathbf{X}))+E(\mathbf{X})^TAE(\mathbf{X})
\end{aligned}
$$

数字特征

（4）相关阵

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^\mathsf{T}$为$p$维随机向量，若自协方差存在且每个分量的方差都大于零，p阶方阵$(r_{i,j})_{p\times p}=(\frac{\mathrm{cov}(X_i,X_j)}{\sqrt{\mathrm{D}(X_i)}\sqrt{\mathrm{D}(X_j)}})_{p\times p}$称之为$\mathbf{X}$的相关阵。$r_{i,j}$称之为分量$X_i,X_j$之间的线性相关系数。

（4）相关阵

设$\mathbf{X}=(X_1,X_2,\cdots,X_p)^{\mathsf{T}}$为$p$维随机向量，若自协方差存在且每个分量的方差都大于零，p阶方阵$(r_{i,j})_{p\times p}=(\frac{\mathrm{cov}(X_i,X_j)}{\sqrt{\mathrm{D}(X_i)}\sqrt{\mathrm{D}(X_j)}})_{p\times p}$称之为$\mathbf{X}$的相关阵。$r_{i,j}$称之为分量$X_i,X_j$之间的线性相关系数。

注：为了克服指标量纲不同对于统计分析的影响，标准化处理: $X_i^*=\frac{X_i-\mathrm{E}(X_i)}{\sqrt{\mathrm{D}(X_i)}}, i=1,2,\ldots,p$，记$\mathbf{X}^*=(X_1^*,X_2^*,\ldots,X_p^*)^{\mathsf{T}}$，则有$\mathrm{E}(\mathbf{X}^*)=\mathbf{0},\mathrm{D}(\mathbf{X}^*)=(r_{i,j})_{p\times p}$。

第二节：统计距离和马氏距离

基本要点:

距离，统计距离，马氏距离

距离

Distance is a numerical measurement of how far apart objects are. A metric is a function that behaves according to a specific set of rules, and is a way of describing what it means for elements of some space to be ”close to” or ”far away from” each other.

https://en.wikipedia.org/wiki/Distance

应用：判别分析，聚类分析

距离

公理化定义:

设E表示点集（如样本集合），从$E \times E$到$[0, +\infty]$的函数$d(.,.)$如果满足以下三条，则称之为距离。

- 非负性：$d(x,y) \geq 0, \forall x, y \in E$，特别地$d(x,y) = 0 \Longleftrightarrow x = y$;
- 对称性：$d(x,y) = d(y,x), \forall x, y \in E$;
- 三角不等式：$d(x,y) \leq d(x,z) + d(z,y), \forall x, y, z \in E$.

距离

常用的距离：

不妨设$\mathbf{x} = (x_1, x_2, \cdots, x_p)^\mathsf{T}, \mathbf{y} = (y_1, y_2, \cdots, y_p)^\mathsf{T}$。

- Euclidean distance：$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{p} |x_i - y_i|^2}$;
- Manhattan distance：$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{p} |x_i - y_i|$;
- Chebyshev distance：$d(\mathbf{x}, \mathbf{y}) = \max_{1 \le i \le p} |x_i - y_i|$;
- q-norm：$d(\mathbf{x}, \mathbf{y}) = \sqrt[1/q]{\sum_{i=1}^{p} |x_i - y_i|^q}, q > 0$。

距离

The 2-norm distance: a generalization of the Pythagorean theorem to more than two coordinates. It is what would be obtained if the distance between two points were measured with a ruler: the ”intuitive” idea of distance.

The 1-norm distance is more colourfully called the taxicab norm or Manhattan distance, because it is the distance a car would drive in a city laid out in square blocks (if there are no one-way streets).

The infinity norm distance: In 2D, it is the minimum number of moves kings require to travel between two squares on a chessboard.

The q-norm is rarely used for values of p other than 1, 2, and infinity, but see super ellipse.

`https://en.wikipedia.org/wiki/Distance`

距离

上述几种距离的不足：

（1）每个坐标对距离的贡献同等。

如当坐标轴表示测量值时，它们往往带有大小不等的随机波动。波动程度较大的分量，对于最终的距离贡献很大，以至于其他分量的影响微乎其微。

上述几种距离的不足：

（1）每个坐标对距离的贡献同等。

如当坐标轴表示测量值时，它们往往带有大小不等的随机波动。波动程度较大的分量，对于最终的距离贡献很大，以至于其他分量的影响微乎其微。

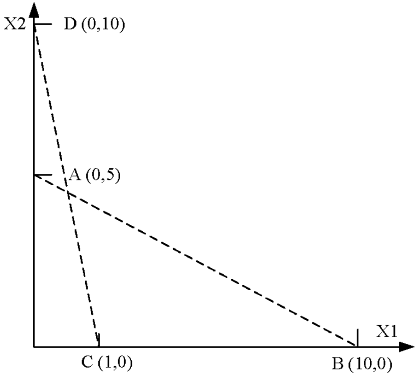
因而，合理的办法是对坐标加权，使得变化较大的坐标比变化小的坐标有较小的权系数。

上述几种距离的不足：

例1

例1：若x_1 取单位kg, x_2 取单位cm, A, B, C, D 是四个二维样品点分布如下图所示，根据欧式距离定义有：

$$d(A,B) = \sqrt{10^2 + 5^2} = \sqrt{125}$$
$$d(C,D) = \sqrt{1^2 + 10^2} = \sqrt{101}$$
$$d(A,B) > d(C,D)$$

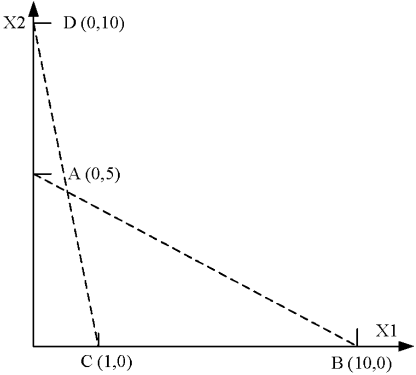


距离

上述几种距离的不足：

如果x_2取单位mm。四个样品点分布仍如图所示，根据欧式距离定义有：

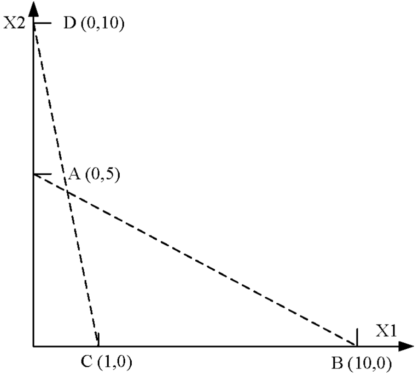
$$d(A,B)=\sqrt{10^2+50^2}=\sqrt{2600}$$
$$d(C,D)=\sqrt{1^2+100^2}=\sqrt{10001}$$
$$d(A,B)<d(C,D)$$



距离

上述几种距离的不足：

如果x_2取单位mm。四个样品点分布仍如图所示，根据欧式距离定义有：



$$d(A,B) = \sqrt{10^2 + 50^2} = \sqrt{2600}$$

$$d(C,D) = \sqrt{1^2 + 100^2} = \sqrt{10001}$$

$$d(A,B) < d(C,D)$$

（2）“距离”的大小竟然与指标的单位有关。

距离

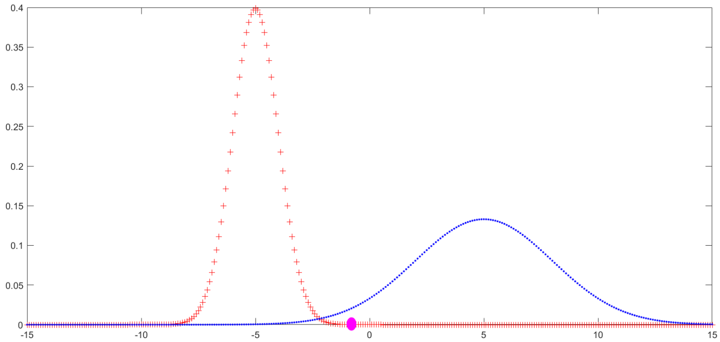
前例说明：

1. 单位选取不同（影响方差），相同样品的距离大小关系竟然得出相反结论；
2. 各指标的量纲可能不同，造成距离的量纲比较复杂。

针对上述问题提出“统计距离”，以区别传统的欧式距离。

例2

设有两个一维正态总体$G_1 : N(-5, 1)$和$G_2 : N(5, 3^2)$，样本点$A(-1, 0)$。A距离哪个总体更近一些呢？



统计距离

在绝对距离上，$A(-1,0)$与$G_1: N(-5,1)$的均值距离为4，小于与$G_2: N(5,3^2)$的均值距离为6。同时，A位于G_1右侧的4倍标准差处，又位于G_2左侧的2倍标准差处，即来自后者的可能性更大，与G_2的“距离”更近。

统计距离

在绝对距离上，$A(-1,0)$与$G_1 : N(-5,1)$的均值距离为4，小于与$G_2 : N(5,3^2)$的均值距离为6。同时，A位于G_1右侧的4倍标准差处，又位于G_2左侧的2倍标准差处，即来自后者的可能性更大，与G_2的“距离”更近。

概率上的“距离”应该比欧式距离更合理。究其原因在于，考虑了体现波动的方差项。即坐标差绝对值除以均方差（或乘均方差的倒数），从而化为无量纲数。如果推广到多维的话，应该需要引进协方差阵。

统计距离

多维情形：

设$\mathbf{x} = (x_1, x_2, \cdots, x_p)^\mathsf{T}, \mathbf{y} = (y_1, y_2, \cdots, y_p)^\mathsf{T}$。假定$\mathbf{x}$固定，$\mathbf{y}$且各分量可以独立变动，记样本方差分别为$S_1^2, S_2^2, \cdots, S_p^2$。$\mathbf{x}$与$\mathbf{y}$的统计距离为

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{(x_1 - y_1)^2}{S_1^2} + \frac{(x_2 - y_2)^2}{S_2^2} + \cdots + \frac{(x_p - y_p)^2}{S_p^2}}$$

统计距离

多维情形：

设$\mathbf{x} = (x_1, x_2, \cdots, x_p)^\mathsf{T}, \mathbf{y} = (y_1, y_2, \cdots, y_p)^\mathsf{T}$。假定$\mathbf{x}$固定，$\mathbf{y}$且各分量可以独立变动，记样本方差分别为$S_1^2, S_2^2, \cdots, S_p^2$。$\mathbf{x}$与$\mathbf{y}$的统计距离为

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{(x_1 - y_1)^2}{S_1^2} + \frac{(x_2 - y_2)^2}{S_2^2} + \cdots + \frac{(x_p - y_p)^2}{S_p^2}}$$

注1：上述统计距离无量纲。

统计距离

多维情形：

设$\mathbf{x}=(x_1,x_2,\cdots,x_p)^\mathsf{T}$, $\mathbf{y}=(y_1,y_2,\cdots,y_p)^\mathsf{T}$。假定$\mathbf{x}$固定，$\mathbf{y}$且各分量可以独立变动，记样本方差分别为$S_1^2,S_2^2,\cdots,S_p^2$。$\mathbf{x}$与$\mathbf{y}$的统计距离为

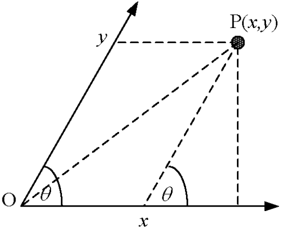
$$d(\mathbf{x},\mathbf{y})=\sqrt{\frac{(x_1-y_1)^2}{S_1^2}+\frac{(x_2-y_2)^2}{S_2^2}+\cdots+\frac{(x_p-y_p)^2}{S_p^2}}$$

注1：上述统计距离无量纲。

注2：特别地，当$S_1^2=S_2^2=\cdots=S_p^2$时，即各分量的权重都相等，用欧氏距离是方便可行的。

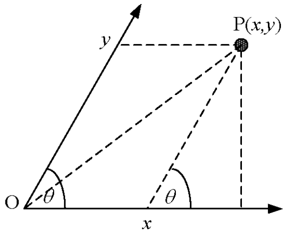
统计距离

一般地，多维向量的各分量不能独立变动，即存在相关关系。以二维为例进行说明。点P距离原点O的欧式距离。



统计距离

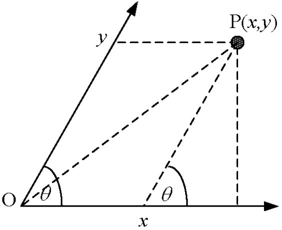
一般地，多维向量的各分量不能独立变动，即存在相关关系。以二维为例进行说明。点P距离原点O的欧式距离。



$$
\begin{aligned}
d(O,P) &= \sqrt{(x+y\cos(\theta))^2+(y\sin(\theta))^2} \\
&= \sqrt{x^2+2\cos(\theta)xy+y^2}
\end{aligned}
$$

统计距离

一般地，多维向量的各分量不能独立变动，即存在相关关系。以二维为例进行说明。点P距离原点O的欧式距离。



$$d(O,P) = \sqrt{(x + y\cos(\theta))^2 + (y\sin(\theta))^2}$$
$$= \sqrt{x^2 + 2\cos(\theta)xy + y^2}$$

一般地，二者的统计距离定义为$d(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^{\mathsf{T}}\mathbf{A}(\mathbf{x}-\mathbf{y})}$，其中$\mathbf{A}$为$p$阶对称、正定矩阵。

马氏距离

最常见的统计距离，由印度统计学家P. C. Mahalanobis于1930年提出。

定义5

设$\mathbf{x},\mathbf{y}$是来自均值向量为μ协方差阵为Σ的总体G的两个样本。

$\mathbf{x}$与$\mathbf{y}$的马氏距离为$d(\mathbf{x},\mathbf{y})=\sqrt{(\mathbf{x}-\mathbf{y})^T\Sigma^{-1}(\mathbf{x}-\mathbf{y})}$，

$\mathbf{x}$与总体G的马氏距离为$d(\mathbf{x},G)=\sqrt{(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}$。

马氏距离

Prasanta Chandra Mahalanobis (1893–1972) was an Indian scientist and applied statistician. He is best remembered for the Mahalanobis distance, a statistical measure, and for being one of the members of the first Planning Commission of free India. He made pioneering studies in anthropometry in India. He founded the Indian Statistical Institute, and contributed to the design of large-scale sample surveys. For his contributions, Mahalanobis has been considered the father of modern statistics in India.

Mahalanobis Distance is one of the most widely used metric to find how much a point diverges from a distribution, based on measurements in multiple dimensions. It is widely used in the field of cluster analysis and classification.

https://en.wikipedia.org/wiki/Prasanta_Chandra_Mahalanobis

第三节：多元正态分布

资源：统计之都《正态分布的前世今生》–靳志辉

http://cos.name/2013/01/story-of-normal-distribution-1/

http://cos.name/2013/01/story-of-normal-distribution-2/

第三节：多元正态分布

资源：统计之都《正态分布的前世今生》—靳志辉

http://cos.name/2013/01/story-of-normal-distribution-1/

http://cos.name/2013/01/story-of-normal-distribution-2/

重要性：（1）许多实际问题服从或近似服从多元正态分布；（2）多元分析的主要理论分析，都建立在多元正态分布基础上。

第三节：多元正态分布

资源：统计之都《正态分布的前世今生》–靳志辉

http://cos.name/2013/01/story-of-normal-distribution-1/

http://cos.name/2013/01/story-of-normal-distribution-2/

重要性：（1）许多实际问题服从或近似服从多元正态分布；（2）多元分析的主要理论分析，都建立在多元正态分布基础上。

多元正态分布是一元正态分布的推广。其他常见的多元分布包括：多项分布，多元对数正态分布，多元χ^2分布，多元t分布，多元指数分布，等。

本节主要内容：多元正态分布的定义、性质、条件分布和独立性。

多元正态分布定义

回顾一元正态分布的密度函数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \forall x \in \mathbb{R}$$

上式可以改写成

$$f(x) = \frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}} \mathrm{exp}\{-\frac{1}{2}(x-\mu)^{\mathsf{T}}(\sigma^2)^{-1}(x-\mu)\}, \forall x \in \mathbb{R}$$

将其推广，给出多元正态分布的定义。

多元正态分布定义

定义6

若p维随机向量$\mathbf{X}=(X_1,X_2,\cdots,X_p)^T$的概率密度函数为

$$f(\mathbf{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\},\forall\mathbf{x}\in\mathbb{R}^p$$

其中协方差阵$\Sigma>0$，则称$\mathbf{X}$服从p维正态分布，记为$\mathbf{X}\sim N_p(\mu,\Sigma)$。

多元正态分布定义

定义6

若p维随机向量$\mathbf{X}=(X_1,X_2,\cdots,X_p)^T$的概率密度函数为

$$f(\mathbf{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\},\forall\mathbf{x}\in\mathbb{R}^p$$

其中协方差阵$\Sigma>0$，则称$\mathbf{X}$服从p维正态分布，记为$\mathbf{X}\sim N_p(\mu,\Sigma)$。

注1：上述定义中$|\Sigma|\neq 0$，否则不存在通常意义下的密度函数。

多元正态分布定义

定义6

若p维随机向量$\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$的概率密度函数为

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\}, \forall \mathbf{x} \in \mathbb{R}^p$$

其中协方差阵$\Sigma > 0$，则称$\mathbf{X}$服从p维正态分布，记为$\mathbf{X} \sim N_p(\mu, \Sigma)$。

注1：上述定义中$|\Sigma| \neq 0$，否则不存在通常意义下的密度函数。

注2：多元正态分布定义不止前述，其他如下述，相互等价。

多元正态分布定义

定义7

若q维随机向量$\mathbf{U}=(U_1,U_2,\cdots,U_q)^T$的每个分量$U_i, i=1,2,\cdots,q$独立同$N(0,1)$。设$A$为$p\times q$阶常矩阵，$\mu$为$p$维常向量，则称$\mathbf{X}=A\mathbf{U}+\mu$服从$p$维正态分布，记为$\mathbf{X}\sim N_p(\mu,AA^T)$。

多元正态分布定义

定义7

若q维随机向量$\mathbf{U} = (U_1, U_2, \cdots, U_q)^T$的每个分量$U_i, i = 1, 2, \cdots, q$独立同$N(0, 1)$。设$A$为$p \times q$阶常矩阵，$\mu$为$p$维常向量，则称$\mathbf{X} = A\mathbf{U} + \mu$服从$p$维正态分布，记为$\mathbf{X} \sim N_p(\mu, AA^T)$。

注：即由标准正态分布的线性组合构成。

多元正态分布定义

定义7

若q维随机向量$\mathbf{U}=(U_1,U_2,\cdots,U_q)^T$的每个分量$U_i, i=1,2,\cdots,q$独立同$N(0,1)$。设$A$为$p\times q$阶常矩阵，$\mu$为$p$维常向量，则称$\mathbf{X}=A\mathbf{U}+\mu$服从$p$维正态分布，记为$\mathbf{X}\sim N_p(\mu, AA^T)$。

注：即由标准正态分布的线性组合构成。

定义8

若p维随机向量$\mathbf{X}=(X_1,X_2,\cdots,X_p)^T$的特征函数为

$$\Phi_X(\mathbf{t})=E[exp\{i\mathbf{t}^T\mathbf{X}\}]=exp\{i\mathbf{t}^T\mu-\frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\}$$

其中$\mathbf{t}\in\mathbb{R}^p$，则称$\mathbf{X}$服从$p$维正态分布，记为$\mathbf{X}\sim N_p(\mu,\Sigma)$。

多元正态分布定义

注3：特别当$p=2$时，也记为$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,r)$

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2 r\\ \sigma_1\sigma_2 r & \sigma_2^2\end{pmatrix}$$

其中σ_1^2,σ_2^2分别为X_1,X_2的方差，$r\neq\pm1$为二者的相关系数。

多元正态分布定义

注3：特别当$p=2$时，也记为$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,r)$

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2 r\\ \sigma_1\sigma_2 r & \sigma_2^2\end{pmatrix}$$

其中σ_1^2,σ_2^2分别为X_1,X_2的方差，$r\neq\pm1$为二者的相关系数。从而有

$$|\Sigma|=\sigma_1^2\sigma_2^2(1-r^2)$$

$$\Sigma^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-r^2)}\begin{pmatrix}\sigma_2^2 & -\sigma_1\sigma_2 r\\ -\sigma_1\sigma_2 r & \sigma_1^2\end{pmatrix}$$

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\{-\frac{1}{2(1-r^2)}[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2r\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}]\}$$

上述密度函数与《概率论》中的结果一致。

多元正态分布定义

注3：特别当$p=2$时，也记为$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,r)$

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2 r\\ \sigma_1\sigma_2 r & \sigma_2^2\end{pmatrix}$$

其中σ_1^2,σ_2^2分别为X_1,X_2的方差，$r\neq\pm1$为二者的相关系数。从而有

$$|\Sigma|=\sigma_1^2\sigma_2^2(1-r^2)$$

$$\Sigma^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-r^2)}\begin{pmatrix}\sigma_2^2 & -\sigma_1\sigma_2 r\\ -\sigma_1\sigma_2 r & \sigma_1^2\end{pmatrix}$$

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\{-\frac{1}{2(1-r^2)}[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2r\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}]\}$$

上述密度函数与《概率论》中的结果一致。若$r=0$，则X_1,X_2相互独立；若$r>0$，则X_1,X_2正相关；若$r<0$，则X_1,X_2负相关。

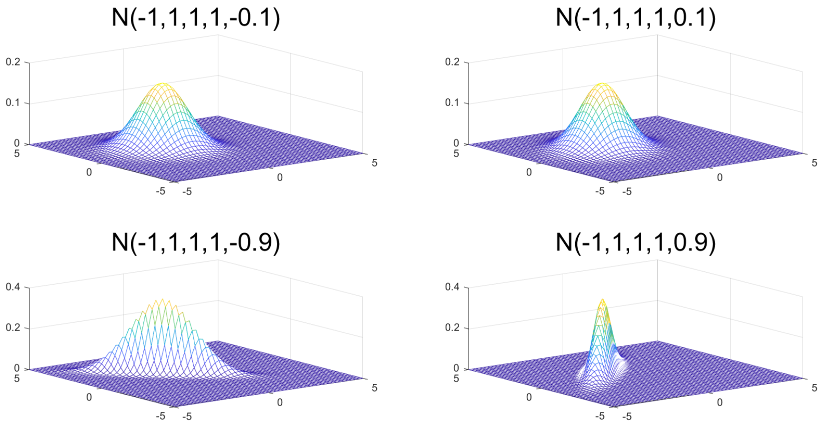


图: 二维正态分布密度函数

多元正态分布性质

定理2

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$E(\mathbf{X}) = \mu, D(\mathbf{X}) = \Sigma$。

多元正态分布性质

定理2

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$E(\mathbf{X}) = \mu, D(\mathbf{X}) = \Sigma$。

命题5

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，且$\Sigma$是对角阵，则$\mathbf{X}$的各分量相互独立。

多元正态分布性质

定理2

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$E(\mathbf{X}) = \mu, D(\mathbf{X}) = \Sigma$。

命题5

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，且$\Sigma$是对角阵，则$\mathbf{X}$的各分量相互独立。

命题6

对于服从正态分布的随机向量$\mathbf{X}$，其分量的任何一个子集构成的向量服从的分布（边缘分布），仍然为正态分布。

定理2

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$E(\mathbf{X}) = \mu, D(\mathbf{X}) = \Sigma$。

命题5

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，且$\Sigma$是对角阵，则$\mathbf{X}$的各分量相互独立。

命题6

对于服从正态分布的随机向量$\mathbf{X}$，其分量的任何一个子集构成的向量服从的分布（边缘分布），仍然为正态分布。

注：逆命题不成立。如$f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)}[1 + x_1 x_2 e^{-\frac{1}{2}(x_1^2+x_2^2)}]$。
容易验证：$X_1 \sim N(0, 1), X_2 \sim N(0, 1)$

多元正态分布性质

命题7

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，$A$为$q \times p$阶常矩阵，$b$为$q$维常向量，则其线性变换$\mathbf{Z} = A\mathbf{X} + \mathbf{b} \sim N_q(A\mu + b, A\Sigma A^T)$。

命题7

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，$A$为$q \times p$阶常矩阵，$b$为$q$维常向量，则其线性变换$\mathbf{Z} = A\mathbf{X} + \mathbf{b} \sim N_q(A\mu + b, A\Sigma A^T)$。

命题8

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$d^2 = (\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi^2(p)$。

多元正态分布性质

命题7

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，$A$为$q \times p$阶常矩阵，$b$为$q$维常向量，则其线性变换$\mathbf{Z} = A\mathbf{X} + \mathbf{b} \sim N_q(A\mu + b, A\Sigma A^T)$。

命题8

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$d^2 = (\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi^2(p)$。

注1：关于证明。记$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \mu)$。根据正态分布的线性性质有$\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$。从而$d^2 = \mathbf{Z}^{\mathsf{T}}\mathbf{Z} \sim \chi^2(p)$。

多元正态分布性质

命题7

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，$A$为$q \times p$阶常矩阵，$b$为$q$维常向量，则其线性变换$\mathbf{Z} = A\mathbf{X} + \mathbf{b} \sim N_q(A\mu + b, A\Sigma A^T)$。

命题8

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$d^2 = (\mathbf{X} - \mu)^T\Sigma^{-1}(\mathbf{X} - \mu) \sim \chi^2(p)$。

注1：关于证明。记$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \mu)$。根据正态分布的线性性质有$\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$。从而$d^2 = \mathbf{Z}^{\mathsf{T}}\mathbf{Z} \sim \chi^2(p)$。

注2：对于给定的d^2，d^2是$\mathbf{X}$的密度函数的等值面。

多元正态分布性质

命题7

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，$A$为$q \times p$阶常矩阵，$b$为$q$维常向量，则其线性变换$\mathbf{Z} = A\mathbf{X} + \mathbf{b} \sim N_q(A\mu + b, A\Sigma A^T)$。

命题8

若随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$d^2 = (\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi^2(p)$。

注1：关于证明。记$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \mu)$。根据正态分布的线性性质有$\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$。从而$d^2 = \mathbf{Z}^\mathsf{T}\mathbf{Z} \sim \chi^2(p)$。

注2：对于给定的d^2，d^2是$\mathbf{X}$的密度函数的等值面。

注3：对于给定的$\mathbf{X}$，d^2是$\mathbf{X}$到总体的马氏距离的平方。

条件分布和独立性

回顾《概率论》中的条件密度：

条件分布和独立性

回顾《概率论》中的条件密度：

X_1, X_2的联合密度函数为$f(x_1, x_2)$，X_2的边缘密度函数为$f_2(x_2)$，则给定$X_2 = x_2$时X_1的条件密度函数为$f(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)}$。

条件分布和独立性

回顾《概率论》中的条件密度：

X_1, X_2的联合密度函数为$f(x_1, x_2)$，X_2的边缘密度函数为$f_2(x_2)$，则给定$X_2 = x_2$时X_1的条件密度函数为$f(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)}$。

特别地对于二维正态分布$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r)$，即有联合密度函数

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\text{exp}\{-\frac{1}{2(1-r^2)}[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}]\}$$

则给定$X_2 = x_2$时X_1的条件密度函数为

$$f(x_1|x_2) \sim N(\mu_1 + \frac{r\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - r^2)\sigma_1^2)$$

条件分布和独立性

回顾《概率论》中的条件密度：

X_1, X_2的联合密度函数为$f(x_1, x_2)$，X_2的边缘密度函数为$f_2(x_2)$，则给定$X_2 = x_2$时X_1的条件密度函数为$f(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)}$。

特别地对于二维正态分布$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r)$，即有联合密度函数

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\{-\frac{1}{2(1-r^2)}[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}]\}$$

则给定$X_2 = x_2$时X_1的条件密度函数为

$$f(x_1|x_2) \sim N(\mu_1 + \frac{r\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - r^2)\sigma_1^2)$$

注：条件密度函数公式验证。

定理3

设随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，其中$p \geq 2$，若$\Sigma > 0$，将$\mathbf{X}, \mu, \Sigma$作如下剖分

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中$X^{(1)}, \mu^{(1)}, \Sigma_{11}$的阶数分别为$q \times 1, q \times 1, q \times q, q < p$。则给定$\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$时$\mathbf{X}^{(1)}$的条件分布：$(\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}) \sim N_q(\mu_{1.2}, \Sigma_{11.2})$，其中$\mu_{1.2} = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。

定理3

设随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，其中$p \geq 2$，若$\Sigma > 0$，将$\mathbf{X}, \mu, \Sigma$作如下剖分

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中$X^{(1)}, \mu^{(1)}, \Sigma_{11}$的阶数分别为$q \times 1, q \times 1, q \times q, q < p$。则给定$\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$时$\mathbf{X}^{(1)}$的条件分布：$(\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}) \sim N_q(\mu_{1.2}, \Sigma_{11.2})$，其中$\mu_{1.2} = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。

注1：由于$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \geqslant 0$，故$\Sigma_{11.2} \leqslant \Sigma_{11}$。即条件分布的离散程度相比无条件分布降低。=成立$\Longleftrightarrow \Sigma_{12} = 0$。

条件分布和独立性

定理3

设随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，其中$p \geq 2$，若$\Sigma > 0$，将$\mathbf{X}, \mu, \Sigma$作如下剖分

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中$X^{(1)}, \mu^{(1)}, \Sigma_{11}$的阶数分别为$q \times 1, q \times 1, q \times q, q < p$。则给定$\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$时$\mathbf{X}^{(1)}$的条件分布：$(\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}) \sim N_q(\mu_{1.2}, \Sigma_{11.2})$，其中$\mu_{1.2} = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。

注1：由于$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \geqslant 0$，故$\Sigma_{11.2} \leqslant \Sigma_{11}$。即条件分布的离散程度相比无条件分布降低。=成立$\Longleftrightarrow \Sigma_{12} = 0$。

注2：$\mathbf{X}^{(2)}$与$\mathbf{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)}$不相关。（只需求二者的协方差阵。）

定理3

设随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，其中$p \geq 2$，若$\Sigma > 0$，将$\mathbf{X}, \mu, \Sigma$作如下剖分

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中$X^{(1)}, \mu^{(1)}, \Sigma_{11}$的阶数分别为$q \times 1, q \times 1, q \times q, q < p$。则给定$\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$时$\mathbf{X}^{(1)}$的条件分布：$(\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}) \sim N_q(\mu_{1.2}, \Sigma_{11.2})$，其中$\mu_{1.2} = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。

注1：由于$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \geqslant 0$，故$\Sigma_{11.2} \leqslant \Sigma_{11}$。即条件分布的离散程度相比无条件分布降低。=成立$\Longleftrightarrow \Sigma_{12} = 0$。

注2：$\mathbf{X}^{(2)}$与$\mathbf{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)}$不相关。（只需求二者的协方差阵。）

注3：$\mathbf{X}^{(2)}$与$\mathbf{X}^{(1)}$对称，二者交换也有类似结论。

注4：证明。记$\mathbf{Z}=\begin{pmatrix}\mathbf{Z}^{(1)}\\ \mathbf{Z}^{(2)}\end{pmatrix}=\begin{pmatrix}\mathbf{I}_q & -\Sigma_{12}\Sigma_{22}^{-1}\\ \mathbf{0} & \mathbf{I}_{(p-q)}\end{pmatrix}\begin{pmatrix}\mathbf{X}^{(1)}\\ \mathbf{X}^{(2)}\end{pmatrix}\triangleq B\mathbf{X}$，则

注4：证明。记$\mathbf{Z}=\begin{pmatrix}\mathbf{Z}^{(1)}\\ \mathbf{Z}^{(2)}\end{pmatrix}=\begin{pmatrix}\mathbf{I}_q & -\Sigma_{12}\Sigma_{22}^{-1}\\ \mathbf{0} & \mathbf{I}_{(p-q)}\end{pmatrix}\begin{pmatrix}\mathbf{X}^{(1)}\\ \mathbf{X}^{(2)}\end{pmatrix}\triangleq B\mathbf{X}$，则

$$\mathbf{Z}=\begin{pmatrix}\mathbf{X}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)}\\ \mathbf{X}^{(2)}\end{pmatrix}\sim N_p\left(\begin{bmatrix}\mu^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\\ \mu^{(2)}\end{bmatrix},\begin{bmatrix}\Sigma_{11.2} & \mathbf{0}\\ \mathbf{0} & \Sigma_{22}\end{bmatrix}\right)$$

条件分布和独立性

注4：证明。记$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}^{(1)} \\ \mathbf{Z}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_{(p-q)} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \triangleq B\mathbf{X}$，则

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)} \\ \mathbf{X}^{(2)} \end{pmatrix} \sim N_p\left(\begin{bmatrix} \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)} \\ \mu^{(2)} \end{bmatrix}, \begin{bmatrix} \Sigma_{11.2} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix} \right)$$

故$\mathbf{Z}^{(1)}$与$\mathbf{Z}^{(2)}$相互独立，则二者的联合密度函数$g(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = g_1(\mathbf{z}^{(1)})g_2(\mathbf{z}^{(2)})$。

注4：证明。记$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}^{(1)} \\ \mathbf{Z}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_{(p-q)} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \triangleq B\mathbf{X}$，则

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)} \\ \mathbf{X}^{(2)} \end{pmatrix} \sim N_p\left(\begin{bmatrix} \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)} \\ \mu^{(2)} \end{bmatrix}, \begin{bmatrix} \Sigma_{11.2} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix} \right)$$

故$\mathbf{Z}^{(1)}$与$\mathbf{Z}^{(2)}$相互独立，则二者的联合密度函数$g(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = g_1(\mathbf{z}^{(1)})g_2(\mathbf{z}^{(2)})$。

而$\mathbf{X}^{(1)}$与$\mathbf{X}^{(2)}$的联合密度函数

注4：证明。记$\mathbf{Z}=\begin{pmatrix}\mathbf{Z}^{(1)}\\ \mathbf{Z}^{(2)}\end{pmatrix}=\begin{pmatrix}\mathbf{I}_q & -\Sigma_{12}\Sigma_{22}^{-1}\\ \mathbf{0} & \mathbf{I}_{(p-q)}\end{pmatrix}\begin{pmatrix}\mathbf{X}^{(1)}\\ \mathbf{X}^{(2)}\end{pmatrix}\triangleq B\mathbf{X}$，则

$$\mathbf{Z}=\begin{pmatrix}\mathbf{X}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)}\\ \mathbf{X}^{(2)}\end{pmatrix}\sim N_p\left(\begin{bmatrix}\mu^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}\\ \mu^{(2)}\end{bmatrix},\begin{bmatrix}\Sigma_{11.2} & \mathbf{0}\\ \mathbf{0} & \Sigma_{22}\end{bmatrix}\right)$$

故$\mathbf{Z}^{(1)}$与$\mathbf{Z}^{(2)}$相互独立，则二者的联合密度函数$g(\mathbf{z}^{(1)},\mathbf{z}^{(2)})=g_1(\mathbf{z}^{(1)})g_2(\mathbf{z}^{(2)})$。

而$\mathbf{X}^{(1)}$与$\mathbf{X}^{(2)}$的联合密度函数

$$\begin{aligned}f(\mathbf{x}^{(1)},\mathbf{x}^{(2)}) &= g(B\mathbf{X})|\frac{\partial\mathbf{z}}{\partial\mathbf{x}}|_+\\ &= g_1(\mathbf{x}^{(1)}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}^{(2)})g_2(\mathbf{x}^{(2)})|det(B)|\end{aligned}$$

联系：(一维)随机变量函数分布，$f_Y(y)=f_X(h(x))|h'(x)|$；向量导数知识。

定理4

设随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，若$\Sigma > 0$，将$\mathbf{X}, \mu, \Sigma$作如下剖分

$$\mathbf{X} = \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(k)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \vdots \\ \mu^{(k)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} \end{pmatrix}$$

其中$X^{(j)}, \mu^{(j)}, \Sigma_{jj}$的阶数分别为$s_j \times 1, s_j \times 1, s_j \times s_j$，$s_j \geq 0, j = 1, 2, \cdots, k$，$\sum_{j=1}^{k} s_j = p$。则$X^{(1)}, X^{(2)}, \cdots, X^{(k)}$相互独立，当且仅当$\Sigma_{ij} = 0, \forall i \neq j$。

定理4

设随机向量$\mathbf{X} \sim N_p(\mu, \Sigma)$，若$\Sigma > 0$，将$\mathbf{X}, \mu, \Sigma$作如下剖分

$$\mathbf{X} = \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(k)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \vdots \\ \mu^{(k)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} \end{pmatrix}$$

其中$X^{(j)}, \mu^{(j)}, \Sigma_{jj}$的阶数分别为$s_j \times 1, s_j \times 1, s_j \times s_j$，$s_j \geq 0, j = 1, 2, \cdots, k$, $\sum_{j=1}^{k} s_j = p$。则$X^{(1)}, X^{(2)}, \cdots, X^{(k)}$相互独立，当且仅当$\Sigma_{ij} = 0, \forall i \neq j$。

注：该定理也说明，$X^{(i)}, X^{(j)}$相互独立$\Longleftrightarrow X^{(i)}, X^{(j)}$不相关。

第四节:均值向量和协方差矩阵估计

重要性：实际中的很多问题，都可以假定研究对象服从或近似服从多元正态分布，但分布所涉及两个参数（μ, Σ）是未知的。同一元情形，一般做法是通过样本去估计。

主要内容：均值向量估计；协方差矩阵估计。

均值向量和协方差矩阵估计

设多元正态分布$\mathbf{X} \sim N_p(\mu, \Sigma)$，简单随机抽样$\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \cdots, \mathbf{X}_{(n)}$，即满足独立同分布（iid.），观测值为$\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \cdots, \mathbf{x}_{(n)}$。而且$n > p, \Sigma > 0$。构造似然函数

$$
\begin{aligned}
L(\mu, \Sigma) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}_{(i)} - \mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_{(i)} - \mu)\} \\
&= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_{(i)} - \mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_{(i)} - \mu)\}
\end{aligned}
$$

均值向量和协方差矩阵估计

设多元正态分布$\mathbf{X} \sim N_p(\mu, \Sigma)$，简单随机抽样$\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \cdots, \mathbf{X}_{(n)}$，即满足独立同分布（iid.），观测值为$\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \cdots, \mathbf{x}_{(n)}$。而且$n > p, \Sigma > 0$。构造似然函数

$$
\begin{aligned}
L(\mu, \Sigma) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}_{(i)} - \mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_{(i)} - \mu)\} \\
&= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_{(i)} - \mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_{(i)} - \mu)\}
\end{aligned}
$$

对数似然函数

$$
lnL(\mu, \Sigma) = -\frac{np}{2} ln(2\pi) - \frac{n}{2} ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_{(i)} - \mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_{(i)} - \mu)
$$

复习：矩阵导数

定义9

设$f(\mathbf{x})$为实向量$\mathbf{x}=(x_1,\cdots,x_p)^T$的实函数，则$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq (\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_p})^T$。若

$$\mathbf{x}=\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_{1p}} \\ \vdots & & \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_{n1}} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_{np}} \end{pmatrix}$$

复习：矩阵导数

定义9

设$f(\mathbf{x})$为实向量$\mathbf{x}=(x_1,\cdots,x_p)^T$的实函数，则$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq (\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_p})^T$。若

$$\mathbf{x}=\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_{1p}} \\ \vdots & & \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_{n1}} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_{np}} \end{pmatrix}$$

命题9

- $\frac{\partial(\mathbf{a}^T\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}, \forall \mathbf{a} \in \mathbb{R}^P$
- $\frac{\partial|\mathbf{x}|}{\partial \mathbf{x}} = |\mathbf{x}|\mathbf{x}^{-T}$
- $\frac{\partial \mathbf{x}^T A\mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$, 其中$A$为$p$阶实对称矩阵
- $\frac{\partial \mathbf{x}^T A\mathbf{x}}{\partial A} = \mathbf{x}^T\mathbf{x}$, $\frac{\partial \mathbf{x}^T A^{-1}\mathbf{x}}{\partial A} = -\mathbf{x}^T\mathbf{x}A^{-1}A^{-1}$, 其中$A$为$p$阶实对称矩阵

均值向量和协方差矩阵估计

对数似然函数分别关于μ, Σ求偏导数有

$$\frac{\partial lnL(\mu,\Sigma)}{\partial \mu} = \sum_{i=1}^{n} \Sigma^{-1}(\mathbf{x}_{(i)} - \mu) = 0$$

$$\frac{\partial lnL(\mu,\Sigma)}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_{(i)} - \mu)^{\mathsf{T}}(\mathbf{x}_{(i)} - \mu)\Sigma^{-1}\Sigma^{-1} = 0$$

均值向量和协方差矩阵估计

均值向量μ的极大似然估计为样本均值向量，即

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ip} \end{pmatrix} \triangleq \begin{pmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_p \end{pmatrix} \triangleq \overline{\mathbf{X}}$$

均值向量和协方差矩阵估计

均值向量μ的极大似然估计为样本均值向量，即

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ip} \end{pmatrix} \triangleq \begin{pmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_p \end{pmatrix} \triangleq \overline{\mathbf{X}}$$

注1：$\overline{\mathbf{X}}$也是μ的无偏估计，充分统计量，强相合估计。

均值向量μ的极大似然估计为样本均值向量，即

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ip} \end{pmatrix} \triangleq \begin{pmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_p \end{pmatrix} \triangleq \overline{\mathbf{X}}$$

注1：$\overline{\mathbf{X}}$也是μ的无偏估计，充分统计量，强相合估计。

注2：若$\mathbf{X} \sim N_p(\mu, \Sigma)$，则$\overline{\mathbf{X}} \sim N_p(\mu, \frac{1}{n}\Sigma)$。提示：根据正态分布的性质以及关系$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)}$即可完成证明。

补充：充分统计量

In statistics, a statistic is sufficient with respect to a statistical model and its associated unknown parameter, if no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter. In particular, a statistic is sufficient for a family of probability distributions, if the sample from which it is calculated gives no additional information than does the statistic, as to which of those probability distributions is that of the population from which the sample was taken.

https://en.wikipedia.org/wiki/Sufficient_statistic

补充：充分统计量

直观来说，不损失信息的统计量，就是充分统计量。

直观来说，不损失信息的统计量，就是充分统计量。

为了推断包含未知参数θ的总体，需要从中抽取样本$X_1, X_2, \ldots, X_n$，并将其加工即统计量$T = T(X_1, X_2, \ldots, X_n)$。既然样本包含了参数$\theta$的信息，则统计量$T$也包含了参数$\theta$的信息，但二者的信息是否一样多呢？统计量$T$肯定不会增加信息，不减少就不错了。

补充：充分统计量

直观来说，不损失信息的统计量，就是充分统计量。

为了推断包含未知参数θ的总体，需要从中抽取样本$X_1, X_2, \ldots, X_n$，并将其加工即统计量$T = T(X_1, X_2, \ldots, X_n)$。既然样本包含了参数$\theta$的信息，则统计量$T$也包含了参数$\theta$的信息，但二者的信息是否一样多呢？统计量$T$肯定不会增加信息，不减少就不错了。假如能由统计量的值恢复样本，则就不会损失参数的信息。即验证条件分布

$$P_\theta(X_1 = x_1, \ldots, X_n = x_n | T = t)$$

不依赖于参数θ。

例3

设样本$X_1, X_2, \ldots, X_n$来自$(0-1)$分布，验证$T=\sum_{i=1}^{n} X_i$的充分性。

例3

设样本$X_1, X_2, \ldots, X_n$来自$(0-1)$分布，验证$T=\sum_{i=1}^n X_i$的充分性。

由于$T \sim B(n,\theta)$，则

$$
\begin{aligned}
& P_\theta(X_1 = x_1, \ldots, X_n = x_n | T = t) \\
=& \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P_\theta(T = t)} \\
=& \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}}{C_n^t \theta^t (1-\theta)^{n-t}} \\
=& \frac{1}{C_n^t}
\end{aligned}
$$

补充：相合估计

In statistics, a consistent estimator or asymptotically consistent estimator is an estimator—a rule for computing estimates of a parameter θ having the property that as the number of data points used increases indefinitely, the resulting sequence of estimates converges in probability to θ. This means that the distributions of the estimates become more and more concentrated near the true value of the parameter being estimated, so that the probability of the estimator being arbitrarily close to θ converges to one.

https://en.wikipedia.org/wiki/Consistent_estimator

定义10

设$g_n = g(X_1, X_2, \ldots, X_n)$是$g(\theta)$的基于样本$X_1, X_2, \ldots, X_n$估计量，若$\theta \in \theta$对都满足

$$\lim_{n\to\infty} P_\theta\{|g_n - g(\theta)| \le \varepsilon\} = 1, \forall \varepsilon > 0$$

则称g_n为$g(\theta)$的弱相合估计量，简记为$g_n \xrightarrow{P} g(\theta)$。

$$P_\theta\{\lim_{n\to\infty} |g_n - g(\theta)| \le \varepsilon\} = 1, \forall \varepsilon > 0$$

则称g_n为$g(\theta)$的强相合估计量，简记为$g_n \to g(\theta), a.s.$。

补充：相合估计

关于随机变量的收敛性：

严继高,吴建楠,张耀.随机变量序列两种收敛性的一个注记.大学数学,2017,33(4):66-68.

李上桐.随机变量的四种收敛性.鄂西大学学报(自然科学报),1987(3):70-74.

均值向量和协方差矩阵估计

协方差阵Σ的极大似然估计为

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{(i)} - \overline{\mathbf{X}})(\mathbf{X}_{(i)} - \overline{\mathbf{X}})^{\mathsf{T}} \triangleq \frac{1}{n}\mathbf{L}$$

均值向量和协方差矩阵估计

协方差阵Σ的极大似然估计为

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{(i)} - \overline{\mathbf{X}})(\mathbf{X}_{(i)} - \overline{\mathbf{X}})^{\mathsf{T}} \triangleq \frac{1}{n}\mathbf{L}$$

注1：$\mathbf{L}$为$p \times p$的对称矩阵，称之为样本离差阵，其中第i行第j列元素$l_{ij} = \sum_{k=1}^{n}(X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)$。

均值向量和协方差矩阵估计

协方差阵Σ的极大似然估计为

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{(i)} - \overline{\mathbf{X}})(\mathbf{X}_{(i)} - \overline{\mathbf{X}})^{\mathsf{T}} \triangleq \frac{1}{n}\mathbf{L}$$

注1：$\mathbf{L}$为$p \times p$的对称矩阵，称之为样本离差阵，其中第i行第j列元素$l_{ij} = \sum_{k=1}^{n}(X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)$。

注2：称$\mathbf{S} = \frac{1}{n-1}\mathbf{L}$为样本协方差阵。特别地当$i = j$时，$\mathbf{S}_{ii} = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)^2$正好为$X_i$的样本方差。

协方差阵Σ的极大似然估计为

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{(i)} - \overline{\mathbf{X}})(\mathbf{X}_{(i)} - \overline{\mathbf{X}})^{\mathsf{T}} \triangleq \frac{1}{n}\mathbf{L}$$

注1：$\mathbf{L}$为$p \times p$的对称矩阵，称之为样本离差阵，其中第i行第j列元素$l_{ij} = \sum_{k=1}^{n}(X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)$。

注2：称$\mathbf{S} = \frac{1}{n-1}\mathbf{L}$为样本协方差阵。特别地当$i = j$时，$\mathbf{S}_{ii} = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)^2$正好为$X_i$的样本方差。

注3：$\frac{1}{n}\mathbf{L}$是Σ的强相合估计，$\frac{1}{n-1}\mathbf{L}$才是Σ的无偏估计。

均值向量和协方差矩阵估计

回忆：数理统计中关于矩的介绍：

总体的k阶中心矩/原点矩；

均值向量和协方差矩阵估计

回忆：数理统计中关于矩的介绍：

总体的k阶中心矩/原点矩；

样本的k阶中心矩/原点矩$\Rightarrow$ 统计量；

均值向量和协方差矩阵估计

回忆：数理统计中关于矩的介绍：

总体的k阶中心矩/原点矩；

样本的k阶中心矩/原点矩$\Rightarrow$ 统计量；

样本的2阶混合中心矩$\Rightarrow$ 总体的协方差

必要性：（1）当总体特征未知时，一般通过抽样进行了解。（2）尽管每个样本在一定程度上体现了总体的特征，但是分散不集中，需要对其加工浓缩，即统计量来加以刻画。统计量的分布称为抽样分布。

必要性：（1）当总体特征未知时，一般通过抽样进行了解。（2）尽管每个样本在一定程度上体现了总体的特征，但是分散不集中，需要对其加工浓缩，即统计量来加以刻画。统计量的分布称为抽样分布。

在《数理统计》中已经讲授了χ^2分布，t分布，F分布。本节将介绍：Wishart分布，T^2分布，Wilks分布。

Wishart分布是关于随机矩阵的分布，是一元情形χ^2分布的推广，目的在于研究样本离差阵$\mathbf{L}=\sum_{i=1}^{n}(\mathbf{X}_{(i)}-\overline{\mathbf{X}})(\mathbf{X}_{(i)}-\overline{\mathbf{X}})^{\mathsf{T}}$，为$p\times p$阶随机矩阵，现考察其分布。

回顾χ^2分布

In probability theory and statistics, the chi-squared distribution (also chi-square) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. The chi-squared distribution is a special case of the gamma distribution and is one of the most widely used probability distributions in inferential statistics, notably in hypothesis testing or in construction of confidence intervals. When it is being distinguished from the more general noncentral chi-squared distribution, this distribution is sometimes called the central chi-squared distribution.

The chi-squared distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one, the independence of two criteria of classification of qualitative data, and in confidence interval estimation for a population standard deviation of a normal distribution from a sample standard deviation.

https://en.wikipedia.org/wiki/Chi-squared_distribution

回顾χ^2分布

定义11

若一维随机变量$X_i \sim N(0,1), i = 1,2,\cdots,n$，且相互独立，则称$\sum_{i=1}^n X_i^2$服从自由度为$n$的$\chi^2$分布，记为$\sum_{i=1}^n X_i^2 \sim \chi^2(n)$。

回顾χ^2分布

定义11

若一维随机变量$X_i \sim N(0,1), i=1,2,\cdots,n$，且相互独立，则称$\sum_{i=1}^n X_i^2$服从自由度为$n$的$\chi^2$分布，记为$\sum_{i=1}^n X_i^2 \sim \chi^2(n)$。

命题10

若$\chi_i^2 \sim \chi^2(n_i), i=1,2,\cdots,k$，且相互独立，则$\sum_{i=1}^k \chi_i^2 \sim \chi^2(\sum_{i=1}^k n_i)$。

回顾χ^2分布

定义11

若一维随机变量$X_i \sim N(0,1), i=1,2,\cdots,n$，且相互独立，则称$\sum_{i=1}^n X_i^2$服从自由度为$n$的$\chi^2$分布，记为$\sum_{i=1}^n X_i^2 \sim \chi^2(n)$。

命题10

若$\chi_i^2 \sim \chi^2(n_i), i=1,2,\cdots,k$，且相互独立，则$\sum_{i=1}^k \chi_i^2 \sim \chi^2(\sum_{i=1}^k n_i)$。

命题11

从一元正态总体$N(\mu,\sigma^2)$中抽样$X_1,\cdots,X_n$，则$\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$，样本均值$\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$与样本方差$S^2=\frac{1}{n-1}\sum_{i=1}^2 (X_i-\overline{X})^2$相互独立，且$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$。

Wishart分布

定义12

设随机向量$\mathbf{X}_{(k)} \sim N_p(\mu_k, \Sigma), k = 1, 2, \cdots, n$，且相互独立，则$p \times p$阶随机矩阵$\mathbf{W} = \sum_{k=1}^{n} \mathbf{X}_{(k)} \mathbf{X}_{(k)}^T$服从自由度为$n$的$p$维非中心*Wishart*分布，记为$\mathbf{W} \sim W_p(n, \Sigma, \mathbf{Z})$，其中$\mathbf{Z} = \sum_{k=1}^{n} \mu_k \mu_k^T$为非中心参数。

Wishart分布

定义12

设随机向量$\mathbf{X}_{(k)} \sim N_p(\mu_k, \Sigma), k = 1, 2, \cdots, n$，且相互独立，则$p \times p$阶随机矩阵$\mathbf{W} = \sum_{k=1}^{n} \mathbf{X}_{(k)} \mathbf{X}_{(k)}^T$服从自由度为$n$的$p$维非中心*Wishart*分布，记为$\mathbf{W} \sim W_p(n, \Sigma, \mathbf{Z})$，其中$\mathbf{Z} = \sum_{k=1}^{n} \mu_k \mu_k^T$为非中心参数。特别地若$\mu_k = 0$时，$k = 1, 2, \cdots, n$，则$\mathbf{Z} = \mathbf{0}$，称服从中心*Wishart*分布，简记为$\mathbf{W} \sim W_p(n, \Sigma)$。

Wishart分布

定义12

设随机向量$\mathbf{X}_{(k)} \sim N_p(\mu_k, \Sigma), k = 1, 2, \cdots, n$，且相互独立，则$p \times p$阶随机矩阵$\mathbf{W} = \sum_{k=1}^{n} \mathbf{X}_{(k)} \mathbf{X}_{(k)}^T$服从自由度为$n$的$p$维非中心*Wishart*分布，记为$\mathbf{W} \sim W_p(n, \Sigma, \mathbf{Z})$，其中$\mathbf{Z} = \sum_{k=1}^{n} \mu_k \mu_k^T$为非中心参数。特别地若$\mu_k = 0$时，$k = 1, 2, \cdots, n$，则$\mathbf{Z} = \mathbf{0}$，称服从中心*Wishart*分布，简记为$\mathbf{W} \sim W_p(n, \Sigma)$。

注1：随机矩阵的代表元素$W_{ij} = \sum_{k=1}^{n} X_{ki} X_{kj}$，非中心参数Z也为$p \times p$阶矩阵，代表元素$Z_{ij} = \sum_{k=1}^{n} \mu_{ki} \mu_{kj}$，$i, j = 1, 2, \cdots, p$。

注2：当$n > p$，$\Sigma > 0$时，中心Wishart分布的密度函数（$p * p$元）

$$f(\mathbf{w}) = \begin{cases} \frac{|\mathbf{w}|^{(n-p-1)/2} e^{-\mathrm{tr}(\Sigma^{-1}\mathbf{w})/2}}{2^{np/2}\pi^{p(p-1)/4}|\Sigma|^{n/2}\prod_{i=1}^{p}\Gamma(\frac{n-i+1}{2})}, & \mathbf{w}\text{对称正定;} \\ 0, & \text{otherwise.} \end{cases}$$

其中$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}e^{-x}dx$。特别地当$p = 1, \Sigma = I_1$时，

$$f(w) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} w^{n/2-1} e^{-w/2}, & w > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Wishart分布

该分布由苏格兰数学家和农业统计学家John Wishart推导出。他于1928年在Biometrika上发表《The generalised product moment distribution in samples from a normal multivariate population》。一般也就将这个时间视为多元统计分支的诞生。

THE GENERALISED PRODUCT MOMENT DISTRIBUTION IN SAMPLES FROM A NORMAL MULTIVARIATE POPULATION.

BY JOHN WISHART, M.A., B.Sc. Statistical Department, Rothamsted Experimental Station.

1. *Introduction.*

For some years prior to 1915, various writers struggled with the problems that arise when samples are taken from uni-variate and bi-variate populations, assumed in most cases for simplicity to be normal. Thus "Student," in 1908*, by considering the first four moments, was led by K. Pearson's methods to infer the distribution of standard deviations, in samples from a normal population. His results, for comparison with others to be deduced later, will be stated in the form

$$dp = \frac{1}{\Gamma\left(\frac{N-1}{2}\right)} A^{\frac{N-1}{2}} \,.\, e^{-Aa} \,.\, a^{\frac{N-3}{2}} \, da \quad \ldots\ldots\ldots\ldots\ldots\ldots (1),$$

where N is the size of the sample, and

$$A = \frac{N}{2\sigma^2}, \quad a = s^2,$$

σ being the standard deviation of the sampled population, and s that estimated from the sample. Thus, if $x_1, x_2, \ldots x_N$ are the sample values, John Wishart(1898 –1956) worked successively at University College London with Karl Pearson, at Rothamsted Experimental Station with Ronald Fisher, and then as a leader in statistics in the University of Cambridge where he became the first Director of the Statistical Laboratory in 1953. He was elected a Fellow of the Royal Society of Edinburgh in 1931, and edited Biometrika from 1937. In 1950 he was elected as a Fellow of the American Statistical Association. He first formulated a generalised product-moment distribution named the Wishart distribution in his honour, in 1928.

https://en.wikipedia.org/wiki/John_Wishart_(statistician)

Discipline	Statistics
Language	English
Publication details	
Publication history	1838–present
Publisher	Wiley-Blackwell (United Kingdom)
Impact factor (2013)	1.573 (Series A) 5.721 (Series B) 1.418 (Series C)
Standard abbreviations	
ISO 4	*J. Royal Stat. Soc.*
Indexing	
ISSN	0964-1998
LCCN	sn99023416
OCLC no.	18305542

图: Journal of the Royal Statistical Society

Discipline	Statistics
Language	English
Publication details	
Publication history	1888–present
Publisher	American Statistical Association (USA)
Frequency	quarterly
Impact factor (2017)	2.297
Standard abbreviations	
ISO 4	*J. Am. Stat. Assoc.*
MathSciNet	*J. Amer. Statist. Assoc.*
Indexing	
ISSN	0162-1459 (print) 1537-274X (web)
LCCN	sn99-23377
JSTOR	01621459
OCLC no.	1480864

图: Journal of the American Statistical Association

Discipline	Statistics
Language	English
Edited by	Paul Fearnhead
Publication details	
Publication history	1901–present
Publisher	Oxford University Press on behalf of the Biometrika Trust
Frequency	Quarterly
Impact factor (2017)	1.669
Standard abbreviations	
ISO 4	*Biometrika*
Indexing	
CODEN	BIOKAX
ISSN	0006-3444 (print) 1464-3510 (web)
LCCN	25005151
JSTOR	00063444
OCLC no.	934140435

图: Biometrika

Discipline	Statistics
Language	English
Publication details	
Publication history	1973–present
Publisher	Institute of Mathematical Statistics (United States)
Standard abbreviations	
ISO 4	*Ann. Stat.*
MathSciNet	*Ann. Statist.*
Indexing	
ISSN	0090-5364
JSTOR	00905364

图: Annals of Statistics

Wishart分布

注3：考察随机矩阵$\mathbf{W} = \sum_{k=1}^{n} \mathbf{X}_{(k)}\mathbf{X}_{(k)}^{\mathsf{T}}$在对角线上的第$i$个元素，$W_{ii} = \sum_{k=1}^{n} X_{ki}^2$，$i = 1, 2, \cdots, p$。记$\Sigma = (\sigma_{ii})_{p\times p}$，若$\mu = \mathbf{0}$，由于$X_{1i}, X_{2i}, \cdots, X_{ni}$独立同分布，故$\frac{W_{ii}}{\sigma_{ii}} \sim \chi^2(n)$。

Wishart分布

注3：考察随机矩阵$\mathbf{W}=\sum_{k=1}^{n}\mathbf{X}_{(k)}\mathbf{X}_{(k)}^{\mathsf{T}}$在对角线上的第$i$个元素，$W_{ii}=\sum_{k=1}^{n}X_{ki}^2$，$i=1,2,\cdots,p$。记$\Sigma=(\sigma_{ii})_{p\times p}$，若$\mu=\mathbf{0}$，由于$X_{1i},X_{2i},\cdots,X_{ni}$独立同分布，故$\frac{W_{ii}}{\sigma_{ii}}\sim\chi^2(n)$。

注4：特别地当$p=1,\mu=0,\Sigma=\sigma^2$时，即$X_k\sim N(0,\sigma^2)$相互独立，$k=1,2,\cdots,n$，则$\mathbf{W}=\frac{\sum_{k=1}^{n}X_k^2}{\sigma^2}\sim\chi^2(n)$。同时$\sum_{k=1}^{n}X_k^2\sim W_1(n,\sigma^2)$。从而$\frac{W_1(n,\sigma^2)}{\sigma^2}\sim\chi^2(n)$。即Wishart分布是$\chi^2$分布在多元情形的推广。

Wishart分布

命题12

若$\mathbf{X}_{(k)} \sim N_p(\mu, \Sigma), k = 1, 2, \cdots, n$，且相互独立，
则$\sum_{k=1}^{n}(\mathbf{X}_{(k)} - \mu)(\mathbf{X}_{(k)} - \mu)^T \sim W_p(n, \Sigma)$，而样本离差阵$\mathbf{L} = \sum_{k=1}^{n}(\mathbf{X}_{(k)} - \overline{\mathbf{X}})(\mathbf{X}_{(k)} - \overline{\mathbf{X}})^T \sim W_p(n-1, \Sigma)$，且与$\overline{\mathbf{X}}$相互独立。

该性质对应于一维情形中样本方差S^2的抽样分布。

Wishart分布

命题12

若$\mathbf{X}_{(k)} \sim N_p(\mu, \Sigma), k = 1, 2, \cdots, n$，且相互独立，
则$\sum_{k=1}^{n}(\mathbf{X}_{(k)} - \mu)(\mathbf{X}_{(k)} - \mu)^T \sim W_p(n, \Sigma)$，而样本离差阵$\mathbf{L} = \sum_{k=1}^{n}(\mathbf{X}_{(k)} - \overline{\mathbf{X}})(\mathbf{X}_{(k)} - \overline{\mathbf{X}})^T \sim W_p(n-1, \Sigma)$，且与$\overline{\mathbf{X}}$相互独立。

该性质对应于一维情形中样本方差S^2的抽样分布。

命题13

若$\mathbf{W}_k \sim W_p(n_k, \Sigma), k = 1, 2, \cdots, q$，且相互独立，
则$\sum_{k=1}^{q} \mathbf{W}_k \sim W_p(\sum_{k=1}^{q} n_k, \Sigma)$。

该性质对应于χ^2分布的可加性。

Wishart分布

命题14

若$\mathbf{W} \sim W_p(n, \Sigma)$，$C$为非奇异的$p \times p$阶常矩阵，

则$C\mathbf{W}C^T \sim W_p(n, C\Sigma C^T)$。

命题14

若$\mathbf{W} \sim W_p(n, \Sigma)$，$C$为非奇异的$p \times p$阶常矩阵，
则$C\mathbf{W}C^T \sim W_p(n, C\Sigma C^T)$。

证明：根据Wishart分布定义，令$\mathbf{W} = \sum_{i=1}^n \mathbf{X}_{(i)}\mathbf{X}_{(i)}^\mathsf{T}$，其中$\mathbf{X}_{(i)} \sim N_p(\mathbf{0}, \Sigma)$，$i = 1, 2, \cdots, n$，且相互独立。
则$C\mathbf{W}C^\mathsf{T} = \sum_{i=1}^n (C\mathbf{X}_{(i)})(C\mathbf{X}_{(i)})^\mathsf{T}$，其中$C\mathbf{X}_{(i)} \sim N_p(\mathbf{0}, C\Sigma C^\mathsf{T})$。得证。

Wishart分布

命题15

若$\mathbf{W} \sim W_p(n,\Sigma)$，$p$维常向量$A$满足$A^T\Sigma A \neq 0$，则$\frac{A^T\mathbf{W}A}{A^T\Sigma A} \sim \chi^2(n)$。

Wishart分布

命题15

若$\mathbf{W} \sim W_p(n, \Sigma)$，$p$维常向量$A$满足$A^T\Sigma A \neq 0$，则$\frac{A^T\mathbf{W}A}{A^T\Sigma A} \sim \chi^2(n)$。

证明：根据Wishart分布定义，令$\mathbf{W} = \sum_{i=1}^n \mathbf{X}_{(i)}\mathbf{X}_{(i)}^{\mathsf{T}}$，其中$\mathbf{X}_{(i)} \sim N_p(\mathbf{0}, \Sigma)$，$i = 1, 2, \cdots, n$，且相互独立。

Wishart分布

命题15

若$\mathbf{W} \sim W_p(n, \Sigma)$，$p$维常向量$A$满足$A^T\Sigma A \neq 0$，则$\frac{A^T\mathbf{W}A}{A^T\Sigma A} \sim \chi^2(n)$。

证明：根据Wishart分布定义，令$\mathbf{W} = \sum_{i=1}^n \mathbf{X}_{(i)}\mathbf{X}_{(i)}^\mathsf{T}$，其中$\mathbf{X}_{(i)} \sim N_p(\mathbf{0}, \Sigma)$，$i = 1, 2, \cdots, n$，且相互独立。由于$A^\mathsf{T}\mathbf{X}_{(i)} \sim N(0, A^\mathsf{T}\Sigma A)$，则$Y_i = (A^\mathsf{T}\Sigma A)^{-1/2}A^\mathsf{T}\mathbf{X}_{(i)} \sim N(0, 1)$。从而有

命题15

若$\mathbf{W} \sim W_p(n, \Sigma)$，$p$维常向量$A$满足$A^T \Sigma A \neq 0$，则$\frac{A^T \mathbf{W} A}{A^T \Sigma A} \sim \chi^2(n)$。

证明：根据Wishart分布定义，令$\mathbf{W} = \sum_{i=1}^n \mathbf{X}_{(i)} \mathbf{X}_{(i)}^\mathsf{T}$，其中$\mathbf{X}_{(i)} \sim N_p(\mathbf{0}, \Sigma)$，$i = 1, 2, \cdots, n$，且相互独立。由于$A^\mathsf{T} \mathbf{X}_{(i)} \sim N(0, A^\mathsf{T} \Sigma A)$，则$Y_i = (A^\mathsf{T} \Sigma A)^{-1/2} A^\mathsf{T} \mathbf{X}_{(i)} \sim N(0, 1)$。从而有

$$
\begin{aligned}
\frac{A^\mathsf{T} \mathbf{W} A}{A^\mathsf{T} \Sigma A} &= \frac{A^\mathsf{T} \sum_{i=1}^n \mathbf{X}_{(i)} \mathbf{X}_{(i)}^\mathsf{T} A}{A^\mathsf{T} \Sigma A} \\
&= \frac{\sum_{i=1}^n A^\mathsf{T} \mathbf{X}_{(i)} (A^\mathsf{T} \mathbf{X}_{(i)})^\mathsf{T}}{(A^\mathsf{T} \Sigma A)^{1/2} (A^\mathsf{T} \Sigma A)^{1/2}} \\
&= \sum_{i=1}^n Y_i^2
\end{aligned}
$$

Wishart分布

定义13

若随机矩阵$\mathbf{W}$的逆矩阵$\mathbf{W}^{-1}$满足$\mathbf{W}^{-1} \sim W_p(n, \Sigma^{-1})$，则称$\mathbf{W}$服从自由度为$n$的$p$维逆*Wishart*分布，记为$\mathbf{W} \sim W_p^{-1}(n, \Sigma)$。

定义13

若随机矩阵$\mathbf{W}$的逆矩阵$\mathbf{W}^{-1}$满足$\mathbf{W}^{-1} \sim W_p(n, \Sigma^{-1})$，则称$\mathbf{W}$服从自由度为$n$的$p$维逆*Wishart*分布，记为$\mathbf{W} \sim W_p^{-1}(n, \Sigma)$。

命题16

若$\mathbf{W} \sim W_p(n, \Sigma)$，$A$为$p$维非零常向量，则$\frac{A^T\Sigma^{-1}A}{A^T\mathbf{W}^{-1}A} \sim \chi^2(n-p+1)$。

Wishart分布

定义13

若随机矩阵$\mathbf{W}$的逆矩阵$\mathbf{W}^{-1}$满足$\mathbf{W}^{-1} \sim W_p(n, \Sigma^{-1})$，则称$\mathbf{W}$服从自由度为$n$的$p$维逆*Wishart*分布，记为$\mathbf{W} \sim W_p^{-1}(n, \Sigma)$。

命题16

若$\mathbf{W} \sim W_p(n, \Sigma)$，$A$为$p$维非零常向量，则$\frac{A^T\Sigma^{-1}A}{A^T\mathbf{W}^{-1}A} \sim \chi^2(n-p+1)$。

证明较复杂，略。可参照教材《Multivariate Statistics: A Vector Space Approach》P.312.

回顾t分布

定义14

若$X \sim N(0,1), Y \sim \chi^2(n)$，且$X, Y$相互独立，则称$T = \frac{X}{\sqrt{Y/n}}$服从自由度$n$为的$t$分布，记为$T \sim t(n)$。

回顾t分布

定义14

若$X \sim N(0,1), Y \sim \chi^2(n)$，且$X, Y$相互独立，则称$T = \frac{X}{\sqrt{Y/n}}$服从自由度$n$为的$t$分布，记为$T \sim t(n)$。

命题17

若$X_1, X_2, \cdots, X_n$是来自一元正态总体$N(\mu, \sigma^2)$的样本，
则$\frac{\overline{X}-\mu}{\sqrt{S^2/n}} = \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1)$。

回顾t分布

定义14

若$X \sim N(0,1), Y \sim \chi^2(n)$，且$X, Y$相互独立，则称$T = \frac{X}{\sqrt{Y/n}}$服从自由度$n$为的$t$分布，记为$T \sim t(n)$。

命题17

若$X_1, X_2, \cdots, X_n$是来自一元正态总体$N(\mu, \sigma^2)$的样本，则$\frac{\overline{X}-\mu}{\sqrt{S^2/n}} = \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1)$。

注：$T^2 = \frac{X^2/1}{Y/n} \sim F(1,n)$。若将$X$从一维随机变量拓展到多维随机向量，改写$T^2 = nX^{\mathsf{T}}Y^{-1}X$，服从何种分布呢？

T^2分布

定义15

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mathbf{0}, c\Sigma)$，且二者相互独立，其中$c > 0$，$n \geq p$，$\Sigma > 0$，则称随机变量$T^2 = \frac{n}{c}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}$服从第一自由度为$p$第二自由度为$n$的$T^2$分布或者Hotelling分布，记为$T^2(p, n)$。

T^2分布

定义15

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mathbf{0}, c\Sigma)$，且二者相互独立，其中$c > 0$，$n \geq p$，$\Sigma > 0$，则称随机变量$T^2 = \frac{n}{c}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}$服从第一自由度为$p$第二自由度为$n$的$T^2$分布或者*Hotelling*分布，记为$T^2(p, n)$。

注：T^2分布由Harold Hotelling推导出，他于1931年在Annals of Mathematical Statistics上发表《The generalization of student's ratio》。

T^2分布

THE GENERALIZATION OF STUDENT'S RATIO*

By

HAROLD HOTELLING

The accuracy of an estimate of a normally distributed quantity is judged by reference to its variance, or rather, to an estimate of the variance based on the available sample. In 1908 "Student" examined the ratio of the mean to the standard deviation of a sample.[1] The distribution at which he arrived was obtained in a more rigorous manner in 1925 by R. A. Fisher,[2] who at the same time showed how to extend the application of the distribution beyond the problem of the significance of means, which had been its original object, and applied it to examine regression coefficients and other quantities obtained by least squares, testing not only the deviation of a statistic from a hypothetical value but also the difference between two statistics.

T^2分布

Harold Hotelling (1895-1973) was an American mathematical statistician and an influential economic theorist, known for Hotelling's law, Hotelling's lemma, and Hotelling's rule in economics, as well as Hotelling's T-squared distribution in statistics. He also developed and named the principal component analysis method widely used in statistics and computer science.

https://en.wikipedia.org/wiki/Harold_Hotelling

命题18

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mathbf{0}, \Sigma)$，且二者相互独立，其中$n \geq p$，$\Sigma > 0$。令$T^2 = n\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}$则，则$\frac{n-p+1}{np}T^2(p, n) \sim F(p, n-p+1)$。

T^2分布

命题18

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mathbf{0}, \Sigma)$，且二者相互独立，其中$n \geq p$，$\Sigma > 0$。令$T^2 = n\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}$则，则$\frac{n-p+1}{np}T^2(p, n) \sim F(p, n-p+1)$。

注1：$\frac{n-p+1}{np}T^2(p, n) = \dfrac{\mathbf{X}^\mathsf{T}\Sigma^{-1}\mathbf{X}/p}{\frac{\mathbf{X}^\mathsf{T}\Sigma^{-1}\mathbf{X}}{\mathbf{X}^\mathsf{T}\mathbf{W}^{-1}\mathbf{X}}/(n-p+1)} \sim F(p, n-p+1)$

命题18

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mathbf{0}, \Sigma)$，且二者相互独立，其中$n \geq p$，$\Sigma > 0$。令$T^2 = n\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}$则，则$\frac{n-p+1}{np}T^2(p, n) \sim F(p, n-p+1)$。

注1：$\frac{n-p+1}{np}T^2(p, n) = \dfrac{\mathbf{X}^\mathsf{T}\Sigma^{-1}\mathbf{X}/p}{\frac{\mathbf{X}^\mathsf{T}\Sigma^{-1}\mathbf{X}}{\mathbf{X}^\mathsf{T}\mathbf{W}^{-1}\mathbf{X}}/(n-p+1)} \sim F(p, n-p+1)$

注2：特别地当$p = 1$时，即有$T^2(1, n) \sim F(1, n)$。

命题18

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mathbf{0}, \Sigma)$，且二者相互独立，其中$n \geq p$，$\Sigma > 0$。令$T^2 = n\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}$则，则$\frac{n-p+1}{np}T^2(p, n) \sim F(p, n-p+1)$。

注1：$\frac{n-p+1}{np}T^2(p,n) = \dfrac{\mathbf{X}^\mathsf{T}\Sigma^{-1}\mathbf{X}/p}{\frac{\mathbf{X}^\mathsf{T}\Sigma^{-1}\mathbf{X}}{\mathbf{X}^\mathsf{T}\mathbf{W}^{-1}\mathbf{X}}/(n-p+1)} \sim F(p, n-p+1)$

注2：特别地当$p = 1$时，即有$T^2(1, n) \sim F(1, n)$。

注3：在涉及到Hotelling分布的概率计算时，可借助于容易计算的F分布实现。

T^2分布

命题19

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mu, \Sigma)$，且二者相互独立，其中$n \geq p$，$\Sigma > 0$，则$n(\mathbf{X} - \mu)^T \mathbf{W}^{-1} (\mathbf{X} - \mu) \sim T^2(p, n)$。

命题19

设$\mathbf{W} \sim W_p(n, \Sigma)$，$\mathbf{X} \sim N_p(\mu, \Sigma)$，且二者相互独立，其中$n \geq p$，$\Sigma > 0$，则$n(\mathbf{X} - \mu)^T \mathbf{W}^{-1} (\mathbf{X} - \mu) \sim T^2(p, n)$。

证明提示：$\mathbf{X} - \mu \sim N_p(\mathbf{0}, \Sigma)$，根据Hotelling分布定义即可。

命题20

设从p维正态总体$N_p(\mu,\Sigma)$中抽取n个样本$\mathbf{X}_{(i)} = (X_{i1}, X_{i2}, \cdots, X_{ip})^T$, $i = 1, 2, \cdots, n$，样本均值向量为$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)}$，样本离差阵为$\mathbf{L} = \sum_{i=1}^{n}(\mathbf{X}_{(i)} - \overline{\mathbf{X}})(\mathbf{X}_{(i)} - \overline{\mathbf{X}})^T$,
则$n(n-1)(\overline{\mathbf{X}} - \mu)^T\mathbf{L}^{-1}(\overline{\mathbf{X}} - \mu) \sim T^2(p, n-1)$。

命题20

设从p维正态总体$N_p(\mu,\Sigma)$中抽取n个样本$\mathbf{X}_{(i)}=(X_{i1},X_{i2},\cdots,X_{ip})^T$，$i=1,2,\cdots,n$，样本均值向量为$\overline{\mathbf{X}}=\frac{1}{n}\sum_{i=1}^n\mathbf{X}_{(i)}$，样本离差阵为$\mathbf{L}=\sum_{i=1}^n(\mathbf{X}_{(i)}-\overline{\mathbf{X}})(\mathbf{X}_{(i)}-\overline{\mathbf{X}})^T$，
则$n(n-1)(\overline{\mathbf{X}}-\mu)^T\mathbf{L}^{-1}(\overline{\mathbf{X}}-\mu)\sim T^2(p,n-1)$。

注1：关于证明。因为$\overline{\mathbf{X}}\sim N_p(\mu,\frac{1}{n}\Sigma)$，则$\sqrt{n}(\overline{\mathbf{X}}-\mu)\sim N_p(\mathbf{0},\Sigma)$。同时$\mathbf{L}\sim W_p(n-1,\Sigma)$，且与$\overline{\mathbf{X}}$相互独立。

T^2分布

命题20

设从p维正态总体$N_p(\mu,\Sigma)$中抽取n个样本$\mathbf{X}_{(i)}=(X_{i1},X_{i2},\cdots,X_{ip})^T$，
$i=1,2,\cdots,n$，样本均值向量为$\overline{\mathbf{X}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)}$，样本离差阵
为$\mathbf{L}=\sum_{i=1}^{n}(\mathbf{X}_{(i)}-\overline{\mathbf{X}})(\mathbf{X}_{(i)}-\overline{\mathbf{X}})^T$，
则$n(n-1)(\overline{\mathbf{X}}-\mu)^T\mathbf{L}^{-1}(\overline{\mathbf{X}}-\mu)\sim T^2(p,n-1)$。

注1：关于证明。因为$\overline{\mathbf{X}}\sim N_p(\mu,\frac{1}{n}\Sigma)$，则$\sqrt{n}(\overline{\mathbf{X}}-\mu)\sim N_p(\mathbf{0},\Sigma)$。同时$\mathbf{L}\sim W_p(n-1,\Sigma)$，且与$\overline{\mathbf{X}}$相互独立。
则$(n-1)[\sqrt{n}(\overline{\mathbf{X}}-\mu)^T]\mathbf{L}^{-1}[\sqrt{n}(\overline{\mathbf{X}}-\mu)]=n(n-1)(\overline{\mathbf{X}}-\mu)^T\mathbf{L}^{-1}(\overline{\mathbf{X}}-\mu)$。
根据Hotelling分布定义即可。

T^2分布

命题20

设从p维正态总体$N_p(\mu,\Sigma)$中抽取n个样本$\mathbf{X}_{(i)}=(X_{i1},X_{i2},\cdots,X_{ip})^T$，$i=1,2,\cdots,n$，样本均值向量为$\overline{\mathbf{X}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{(i)}$，样本离差阵为$\mathbf{L}=\sum_{i=1}^{n}(\mathbf{X}_{(i)}-\overline{\mathbf{X}})(\mathbf{X}_{(i)}-\overline{\mathbf{X}})^T$，
则$n(n-1)(\overline{\mathbf{X}}-\mu)^T\mathbf{L}^{-1}(\overline{\mathbf{X}}-\mu)\sim T^2(p,n-1)$。

注1：关于证明。因为$\overline{\mathbf{X}}\sim N_p(\mu,\frac{1}{n}\Sigma)$，则$\sqrt{n}(\overline{\mathbf{X}}-\mu)\sim N_p(\mathbf{0},\Sigma)$。同时$\mathbf{L}\sim W_p(n-1,\Sigma)$，且与$\overline{\mathbf{X}}$相互独立。
则$(n-1)[\sqrt{n}(\overline{\mathbf{X}}-\mu)^T]\mathbf{L}^{-1}[\sqrt{n}(\overline{\mathbf{X}}-\mu)]=n(n-1)(\overline{\mathbf{X}}-\mu)^T\mathbf{L}^{-1}(\overline{\mathbf{X}}-\mu)$。
根据Hotelling分布定义即可。

注2：若$p=1$时，$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\sim N(0,1)$，$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，$L=(n-1)S^2$，则$L^{-1}=\frac{1}{(n-1)S^2}$。

命题21

从两个正态总体$N_p(\mu_i, \Sigma)$，分别抽样$\mathbf{X}_{(1)}, \cdots, \mathbf{X}_{(n_1)}$；$\mathbf{Y}_{(1)}, \cdots, \mathbf{Y}_{(n_2)}$，$i = 1, 2$。样本均值向量分别为$\overline{\mathbf{X}}, \overline{\mathbf{Y}}$，样本离差阵分别为$\mathbf{L}_1, \mathbf{L}_2$。若$\mu_1 = \mu_2$，则$\frac{n_1 n_2}{n_1+n_2}(n_1 + n_2 - 2)(\overline{\mathbf{X}} - \overline{\mathbf{Y}})^T(\mathbf{L}_1 + \mathbf{L}_2)^{-1}(\overline{\mathbf{X}} - \overline{\mathbf{Y}}) \sim T^2(p, n_1 + n_2 - 2)$。

命题21

从两个正态总体$N_p(\mu_i,\Sigma)$，分别抽样$\mathbf{X}_{(1)},\cdots,\mathbf{X}_{(n_1)}$；$\mathbf{Y}_{(1)},\cdots,\mathbf{Y}_{(n_2)}$，$i=1,2$。样本均值向量分别为$\overline{\mathbf{X}},\overline{\mathbf{Y}}$，样本离差阵分别为$\mathbf{L}_1,\mathbf{L}_2$。若$\mu_1=\mu_2$，则$\frac{n_1n_2}{n_1+n_2}(n_1+n_2-2)(\overline{\mathbf{X}}-\overline{\mathbf{Y}})^T(\mathbf{L}_1+\mathbf{L}_2)^{-1}(\overline{\mathbf{X}}-\overline{\mathbf{Y}})\sim T^2(p,n_1+n_2-2)$。

注1：$\overline{\mathbf{X}}-\overline{\mathbf{Y}}\sim N_p(\mathbf{0},(\frac{1}{n_1}+\frac{1}{n_2})\Sigma)\Longrightarrow\sqrt{\frac{n_1n_2}{n_1+n_2}}(\overline{\mathbf{X}}-\overline{\mathbf{Y}})\sim N_p(\mathbf{0},\Sigma)$。同时，$\mathbf{L}_i\sim W_p(n_i-1,\Sigma)$，$i=1,2$，从而$\mathbf{L}_1+\mathbf{L}_2\sim W_p(n_1+n_2-2,\Sigma)$，则由Hotelling分布的定义得证。

命题21

从两个正态总体$N_p(\mu_i,\Sigma)$，分别抽样$\mathbf{X}_{(1)},\cdots,\mathbf{X}_{(n_1)}$；$\mathbf{Y}_{(1)},\cdots,\mathbf{Y}_{(n_2)}$，$i=1,2$。样本均值向量分别为$\overline{\mathbf{X}},\overline{\mathbf{Y}}$，样本离差阵分别为$\mathbf{L}_1,\mathbf{L}_2$。若$\mu_1=\mu_2$，则$\frac{n_1n_2}{n_1+n_2}(n_1+n_2-2)(\overline{\mathbf{X}}-\overline{\mathbf{Y}})^T(\mathbf{L}_1+\mathbf{L}_2)^{-1}(\overline{\mathbf{X}}-\overline{\mathbf{Y}})\sim T^2(p,n_1+n_2-2)$。

注1：$\overline{\mathbf{X}}-\overline{\mathbf{Y}}\sim N_p(\mathbf{0},(\frac{1}{n_1}+\frac{1}{n_2})\Sigma)\Longrightarrow\sqrt{\frac{n_1n_2}{n_1+n_2}}(\overline{\mathbf{X}}-\overline{\mathbf{Y}})\sim N_p(\mathbf{0},\Sigma)$。同时，$\mathbf{L}_i\sim W_p(n_i-1,\Sigma)$，$i=1,2$，从而$\mathbf{L}_1+\mathbf{L}_2\sim W_p(n_1+n_2-2,\Sigma)$，则由Hotelling分布的定义得证。

注2：若$p=1$时，$\Sigma=\sigma^2$，$\mathbf{L}_1=(n_1-1)S_1^2$，$\mathbf{L}_2=(n_2-1)S_2^2$。

$$\sqrt{\frac{n_1n_2}{n_1+n_2}}\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\mathbf{L}_1+\mathbf{L}_2}{n_1+n_2-2}}}=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}}\sim t(n_1+n_2-2)。$$

回顾F分布

若$X\sim\chi^2(m),Y\sim\chi^2(n)$且相互独立，则称$F=\frac{X/m}{Y/n}$服从第一自由度为$m$第二自由度为$n$的$F$分布，记为$F\sim F(m,n)$。

回顾F分布

若$X\sim\chi^2(m), Y\sim\chi^2(n)$且相互独立，则称$F=\frac{X/m}{Y/n}$服从第一自由度为$m$第二自由度为$n$的$F$分布，记为$F\sim F(m,n)$。

设样本$X_1,\cdots,X_{n_1}$来自$N(\mu_1,\sigma_1^2)$，$Y_1,\cdots,Y_{n_2}$来自$N(\mu_2,\sigma_2^2)$。样本均值分别为$\overline{X},\overline{Y}$，样本方差阵分别为$S_1^2,S_2^2$。则$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$。

Wilks分布

F分布能够刻画两个正态总体的样本方差比的分布。如果研究对象由一维扩展到多维，而刻画多元总体取值变异程度的是样本离差阵（或样本协差阵），是一个矩阵而非一个数字。若要将其表达成数字之比的形式，就必须将矩阵加工成一个有代表性的量。

Wilks分布

F分布能够刻画两个正态总体的样本方差比的分布。如果研究对象由一维扩展到多维，而刻画多元总体取值变异程度的是样本离差阵（或样本协差阵），是一个矩阵而非一个数字。若要将其表达成数字之比的形式，就必须将矩阵加工成一个有代表性的量。

常见的加工方式有：矩阵的行列式、迹、最大或最小特征根等，统称为广义方差。Wilks分布就是基于矩阵的行列式给出。

Wilks分布

定义16

设$\mathbf{W}_1 \sim W_p(n_1,\Sigma)$，$\mathbf{W}_2 \sim W_p(n_2,\Sigma)$，且二者相互独立，称$\Lambda = \frac{|\mathbf{W}_1|}{|\mathbf{W}_1+\mathbf{W}_2|}$服从$p$维第一自由度为$n_1$第二自由度为$n_2$的Wilks分布，记为$\Lambda \sim \Lambda(p,n_1,n_2)$。

Wilks分布

定义16

设$\mathbf{W}_1 \sim W_p(n_1, \Sigma)$，$\mathbf{W}_2 \sim W_p(n_2, \Sigma)$，且二者相互独立，称$\Lambda = \frac{|\mathbf{W}_1|}{|\mathbf{W}_1+\mathbf{W}_2|}$服从$p$维第一自由度为$n_1$第二自由度为$n_2$的Wilks分布，记为$\Lambda \sim \Lambda(p, n_1, n_2)$。

Wilks分布比较复杂，涉及到概率计算时，通常用F分布或者χ^2分布近似。详见教材§1.5.3。

Wilks分布

Samuel Stanley Wilks(1906-1964) was an American mathematician and academic who played an important role in the development of mathematical statistics, especially in regard to practical applications.

https://en.wikipedia.org/wiki/Samuel_S._Wilks

他于1932年在Biometrika上发表《Certain Generalizations in the Analysis of Variance》。

https://www.jstor.org/stable/2331979

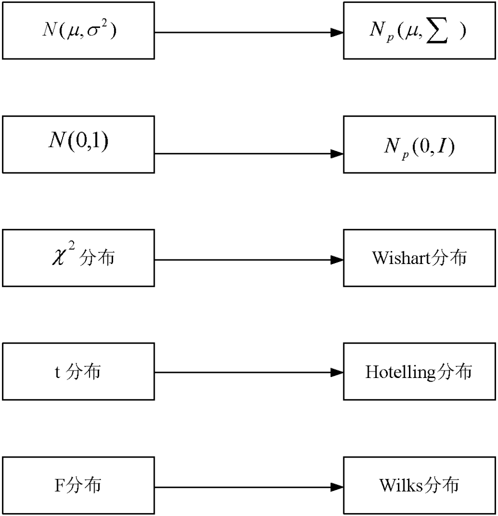


图: 分布之间的关系1：一元vs多元

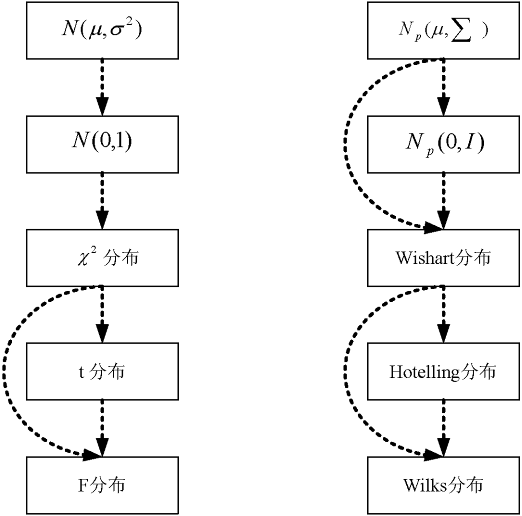


图: 分布之间的关系2：关联

第一章习题

（1）若$\mathbf{X} \sim N_p(\mu, \Sigma)$，且$\Sigma$是对角阵，则$\mathbf{X}$的各分量相互独立。

（2）若$(X_1, X_2)^\mathsf{T} \sim N_2(\mu, \Sigma)$，其中$\mu = (\mu_1, \mu_2)^\mathsf{T}$，
$\Sigma = \sigma^2 \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$，$\sigma^2 > 0, 0 < r < 1$。证明$X_1 + X_2$和$X_1 - X_2$相互独立；并求$X_1 + X_2$和$X_1 - X_2$的联合分布。

（3）若$\mathbf{W}_i \sim W_p(n_i, \Sigma), i = 1, 2, \cdots, k$，且相互独立。证明：
$\sum_{i=1}^k \mathbf{W}_i \sim W_p(\sum_{i=1}^k n_i, \Sigma)$。

（4）教材第一章习题第5题。