在作业第 23 题中，我们遇到了这样的问题：

23. 设 $(X_1, X_2, \cdots, X_n, X_{n+1})$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$，试求统计量

$$Y = \frac{X_{n+1} - \bar{X}}{S}\sqrt{\frac{n}{n+1}}$$

的抽样分布.

如果你去搜索答案，会得到这样的做法：

> $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$，所以 $X_{n+1} - \bar{X} \sim N\left(0, \frac{n+1}{n}\sigma^2\right)$，所以 $(X_{n+1} - \bar{X})\sqrt{\frac{n}{n+1}} \sim N(0,1)$
> 而 $(n-1)S^2 \sim \chi^2(n-1)$
> 所以原式服从 $t(n-1)$

这个答案给人一些困惑：

- 凭什么 $(n-1)S^2 \sim \chi^2(n-1)$?
- 就算上面是对的，分子里不是涉及到 $\bar{X}$ 吗？怎么保证分子和分母是独立的？（t分布要求分子分母是独立的）

我们来详细讨论这些问题。

正态分布的线性代数视角

如果了解过多元正态分布的一般表达式，会看到这样一个方程：

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|}\exp\left(-\frac{(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)}{2}\right)$$

这是多元正态分布的概率密度函数。其中，$\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ 是正态分布的取值，$\mu = (\mu_1, \mu_2, \ldots, \mu_n)^{\mathrm{T}}$ 是均值向量，而 Σ 是该分布的协方差矩阵。

一个特别的情况是各分量独立。此时，Σ 是一个对角阵。从正态分布中独立同分布地采样也是这种情况。

为了方便分析，从下面开始，我们把所有的随机变量进行中心化：对于随机变量 X，定义 $Z = X - \mu$，则 Z 的期望为 0。则概率密度简化如下：

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}|\Sigma|}\exp\left(-\frac{\mathbf{z}^{\mathrm{T}}\Sigma^{-1}\mathbf{z}}{2}\right)$$

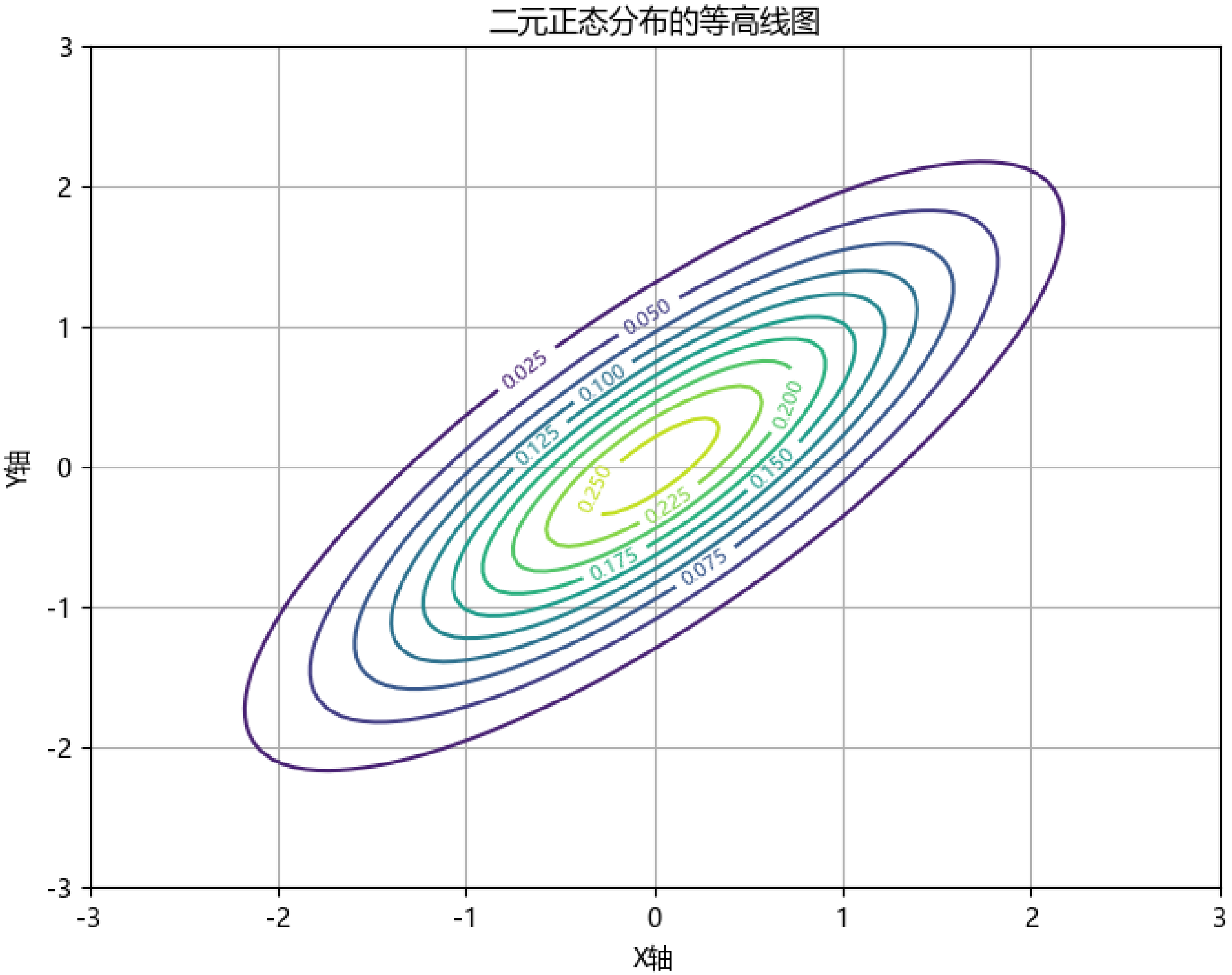
我们不难看出，对于归一化之后的多元正态随机变量，其性质被协方差矩阵 Σ 完全决定。

这个**完全决定**贡献了这么一条性质：

- 如果两个服从正态分布的随机变量**不相关**，那么他们独立。

这并不难理解：如果这两个变量不相关，则协方差矩阵是对角阵。这和独立的情况是一样的。既然不相关和独立有

而协方差矩阵是一个**非负实对称矩阵**，和我们比较熟悉的**二次型**比较相似。准确来说，对应的二次型**总是圆（超球）或椭圆（超椭球）**。事实上，圆和椭圆正是正态分布密度函数的等值线图：



上图是 $\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$ 的等值线图。

这个椭圆具体是怎么确定的？方向和伸缩比如何决定？我们来做一些推导。

既然已经知道 Σ 是一个非负实对称矩阵，一个自然的想法是对它做特征值分解：

$$
\begin{aligned}
\Sigma &= P\Lambda P^{\mathrm{T}} \\
\Sigma^{-1} &= P^{\mathrm{T}}\Lambda^{-1}P
\end{aligned}
$$

从线性代数我们知道，P 必然可以取为一个规范正交阵（$PP^{\mathrm{T}} = I$）。

带入表达式：

$$
f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}|\Lambda|}\exp\left(-\frac{(P\mathbf{z})^{\mathrm{T}}\Lambda^{-1}P\mathbf{z}}{2}\right)
$$

我特地写成了 $(P\mathbf{z})^{\mathrm{T}}\Lambda^{-1}(P\mathbf{z})$，是因为这样的事实：

- 如果 Z 遵循协方差为 $\Sigma = P\Lambda P^{\mathrm{T}}$ 的多元正态分布，则 PZ 遵循协方差为 Λ 的多元正态分布。考虑到 Λ 是一个对角阵，也就是说，PZ 的各分量独立。

这个证明很简单。因为 PZ 关于 Z 的雅可比矩阵就是 P，而 P 作为规范正交阵，行列式绝对值为 1，所以：

$$
f(P\mathbf{z}) = |P|f(\mathbf{z}) = f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}|\Lambda|}\exp\left(-\frac{(P\mathbf{z})^{\mathrm{T}}\Lambda^{-1}P\mathbf{z}}{2}\right)
$$

考虑这个结论的几何意义：规范正交阵 P 的作用效果其实是对向量进行旋转。

以上面的图为例：原来的 X 与 Y 并不独立，是因为椭圆的短轴和长轴并不在坐标轴上，而是在他们之间。对协方差矩阵做特征值分解：

$$\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1.8 & 0 \\ 0 & 0.2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

按照特征值分解结果，如果定义新的随机变量 $Z_1 = \frac{1}{\sqrt{2}}(X+Y), Z_2 = \frac{1}{\sqrt{2}}(X-Y)$，则 Z_1 与 Z_2 相互独立，方差分别为 1.8 和 0.2。

从图像上也能看出这个结果：椭圆的长轴方向向量是 $(1,1)^{\mathrm{T}}$，短轴则是 $(1,-1)^{\mathrm{T}}$，所以如果把坐标轴旋转到这两个方向，新的椭圆就是正的，对应独立的情况。

如果我们沿着这条路继续走下去，就可以得到主成分分析在正态分布的特殊情况。（老实说，我也不知道这个内容更接近概率还是数基，也许这个情况会对大家理解主成分分析有帮助）

不过我们的目标不在这里。我上面写那么多，其实是希望大家对于正态分布和二次型矩阵的关系有一定的认识。总的来说，要理解这些点：

- 中心化之后的正态分布被一个二次型矩阵（协方差矩阵）决定
- 可以对正态分布进行线性变换（可逆矩阵乘），得到的结果仍然是正态分布
- 有相关性的正态分布可以通过旋转变换变成不相关的正态分布（主成分分析）
- 因为正态分布被协方差矩阵决定，所以不相关就是独立
- 从上一点可以得出，如果协方差矩阵是一个分块对角阵：
  - $\Sigma = \begin{bmatrix} \Sigma_k & O \\ O & \Sigma_{n-k} \end{bmatrix}$
  - 那么前 k 个分量构成的随机向量 Z_k 和后 $n-k$ 个分量构成的随机向量 Z_{n-k} 是独立的，进一步，由前 k 个分量的函数结果与后 $n-k$ 个分量的函数结果也是独立的。

S^2 和 $\bar{X}$

假设我们有 n 个独立同分布的正态样本：

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$$

其中：

- μ 是总体均值
- σ^2 是总体方差
- 样本均值为：

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- 样本方差为（无偏）：

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

我们把 X 标准化（期望为 0，方差为 1）：

$$Z_i = \frac{X_i - \mu}{\sigma}$$

因为 $X_i \sim N(\mu, \sigma^2)$，所以：

- $\bar{X} = \mu + \sigma \cdot \bar{Z}$
- $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$

把 S^2 的表达式全部换成 Z，得到:

$$\begin{aligned}\frac{(n-1)S^2}{\sigma^2} &= \sum_{i=1}^{n}(Z_i - \bar{Z})^2 \\ &= \left(\sum_{i=1}^{n} Z_i^2\right) - n\bar{Z}^2\end{aligned}$$

我们发现，右边的式子实际是 Z_i 的一个二次型。记 $\mathbf{z} = (Z_1, Z_2, \dots, Z_n)$，则:

$$\frac{(n-1)S^2}{\sigma^2} = \mathbf{z}^{\mathrm{T}}\left(I - \frac{1}{n}A\right)\mathbf{z}$$

其中 A 是全 1 矩阵。对 $I - \frac{A}{n}$ 特征值分解，得到:

- 一个特征值为0，特征向量为 $\mathbf{1} = (1, 1, \dots, 1)^{\mathrm{T}}$
- 其他特征值为1，特征向量可以是任意满足 $\sum x_i = 0$ 的向量。

因为 $I - \frac{A}{n}$ 是实对称矩阵，正交的特征向量必然存在。那么，0 特征值的特征向量取为 $u = \frac{1}{\sqrt{n}}(1, 1, \dots, 1)^{\mathrm{T}}$，而 1 的正交特征向量设为 $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$。$I - \frac{1}{n}A$ 特征值分解如下:

$$\begin{aligned}I - \frac{1}{n}A &= P\Lambda P^{\mathrm{T}} \\ &= \begin{bmatrix}\mathbf{u} & \mathbf{v}_1 & \dots & \mathbf{v_{n-1}}\end{bmatrix}\operatorname{diag}([0, 1, 1, \dots, 1])\begin{bmatrix}\mathbf{u}^{\mathrm{T}} \\ \mathbf{v_1}^{\mathrm{T}} \\ \dots \\ \mathbf{v_{n-1}}^{\mathrm{T}}\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v_{n-1}}\end{bmatrix}\begin{bmatrix}\mathbf{v_1}^{\mathrm{T}} \\ \mathbf{v_2}^{\mathrm{T}} \\ \dots \\ \mathbf{v_{n-1}}^{\mathrm{T}}\end{bmatrix} \\ &= VV^{\mathrm{T}}\end{aligned}$$

所以:

$$\frac{(n-1)S^2}{\sigma^2} = \mathbf{z}^{\mathrm{T}}VV^{\mathrm{T}}\mathbf{z} = \|V^{\mathrm{T}}\mathbf{z}\|^2$$

而又有:

$$\frac{\bar{X} - \mu}{\sigma} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \sqrt{n}\mathbf{u}^{\mathrm{T}}\mathbf{z}$$

所以 S^2 是 $V^{\mathrm{T}}\mathbf{z}$ 的函数，而 $\bar{X}$ 是 $\mathbf{u}^{\mathrm{T}}\mathbf{z}$ 的函数。只要 $V^{\mathrm{T}}\mathbf{z}$ 和 $\mathbf{u}^{\mathrm{T}}\mathbf{z}$ 是独立的，S^2 和 $\bar{X}$ 也就是独立的。

结合我们关于多元正态分布的分析:

$\mathbf{z}$ 遵循协方差矩阵为 I 的多元正态分布，所以 $P^{\mathrm{T}}z$ 遵循协方差矩阵为 PP^{T} 的多元正态分布，但是 $PP^{\mathrm{T}} = I$。也就是说，$P^{\mathrm{T}}\mathbf{z}$ 的各个分量仍然是独立的。特别地，第一个分量 $\mathbf{u}^{\mathrm{T}}\mathbf{z}$ 和后 $n-1$ 个分量 $V^{\mathrm{T}}\mathbf{z}$ 是独立的。

证明完毕，S^2 和 $\bar{X}$ 确实是独立的。

而且更进一步地，$V^{\mathrm{T}}\mathbf{z}$ 既然是 $P^{\mathrm{T}}\mathbf{z}$ 的后 $n-1$ 个分量，而 $P^{\mathrm{T}}\mathbf{z}$ 遵循互相独立的 n 维标准正态分布，则 $V^{\mathrm{T}}\mathbf{z}$ 遵循互相独立的 $n-1$ 维标准正态分布。

所以，$\|V^{\mathrm{T}}\mathbf{z}\|^2$ 是 $n-1$ 个分量的平方和，则 $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$。

两个疑问解答完毕，最开始的问题也得到了解决。