协方差与向量内积之间的关系

Shuyong Chen

2023 年 8 月 9 日

1 简介

在下面的提问中
https://www.zhihu.com/question/616068948
提出了一个关于协方差和向量内积之间关系的问题。整理如下：

我觉得协方差和向量内积都是那种两个集合按一定乘法规则遍历乘一遍然后求和的感觉。当然协方差的遍历规则的要求更低，内积需有亚标量的对应。就是这种“相乘”式的规则 (是叫代数结构吗？) 是不是哪个更高级的规则的一些特例呢？有没有专门研究“相乘”的数学领域？

另外它们这种“相乘”之后导出的性质好像也有一些联系的感觉。比如协方差为 0 就是随机变量线性不相关，为 ±1 就是几乎处处满足 $y = kx + b$ (线性关系)；同时向量内积为零就是向量正交 (比向量线性无关还强的性质)；还有就是想问下能从协方差和向量内积这两种计算规则的角度出发解释：向量的线性无关和随机变量的线性不相关之间的联系吗？或者介绍一下两者产生联系的一些场景？

2 问题的回答

你的观察非常有趣，协方差和向量内积确实有一定的相似性。在回答你的问题之前，让我们先回顾一下协方差和向量内积的定义。

协方差 (Covariance) 用于衡量两个随机变量的线性相关性，它的定义为：

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

其中，$E[X]$ 和 $E[Y]$ 分别表示随机变量 X 和 Y 的期望。

向量内积 (Inner product) 是一个二元运算，它将两个向量映射到一个标量。在欧几里得空间中，向量内积的定义为：

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i = |\mathbf{a}||\mathbf{b}| \cos\theta$$

其中，$\mathbf{a}$ 和 $\mathbf{b}$ 是两个 n 维向量，θ 是它们之间的夹角。

从定义上看，协方差和向量内积都可以看作是一种“相乘”操作，所提到的“遍历乘法运算后求和”的计算模式，在数学中被泛化为抽象代数中的双线性形式 (Bilinear Form)。它是线性代数中非常基础和重要的概念。但它们的背景和应用场景有很大不同。协方差主要用于描述随机变量之间的关系，而向量内积主要用于描述向量之间的关系。

接下来，让我们从更高的观点来看待这两个概念。

1. 代数结构：协方差和向量内积都可以看作是某种代数结构的特例。在抽象代数中，有一种叫做双线性型 (Bilinear form) 的代数结构，它接受两个向量作为输入，输出一个标量。向量内积和协方差都可以看作是双线性型的特例。双线性型的定义为：

$$B(\mathbf{a},\mathbf{b})=\sum_{i=1}^{n}\sum_{j=1}^{n}a_iM_{ij}b_j$$

其中，M_{ij} 是一个矩阵，对于向量内积，$M_{ij}=\delta_{ij}$ (克罗内克符号，当 $i=j$ 时为 1，否则为 0)；对于协方差，可以将随机变量 X 和 Y 看作是两个向量，M_{ij} 是协方差矩阵的元素。

2. 典型示例：很多线性代数中的运算，比如向量内积、叉积，都满足双线性形式的条件。而统计学中的协方差，也可以看作是随机变量构成的向量空间上定义的一种双线性形式。从几何的角度看，双线性形式定义了向量空间上的一个二次型，它刻画了空间中不同方向间的关系。当双线性形式等于零，表示对应方向上是正交的，即，零协方差表示随机变量线性无关，零向量内积表示正交。这些都反映了正交方向上两两之间独立的关系。

3. 几何直觉：向量内积可以用来描述向量之间的几何关系，例如夹角、正交性等。而协方差可以用来描述随机变量之间的线性关系，例如相关性、线性回归等。从这个角度看，它们的几何直觉是有联系的。当我们将随机变量看作是高维空间中的点时，协方差和向量内积之间的联系就更加明显了。

4. 线性无关与线性不相关：向量的线性无关是指一组向量不能通过线性组合表示成另一个向量。而随机变量的线性不相关是指两个随机变量之间没有线性关系。从概念上看，这两者有一定的联系。实际上，当我们将随机变量看作是高维空间中的点时，它们之间的线性关系可以通过协方差矩阵来描述。而协方差矩阵又和向量内积有一定的联系，因此可以说，线性无关和线性不相关之间存在一定的联系。

总结一下，协方差和向量内积确实有一定的相似性，它们都可以看作是双线性型的特例，刻画了向量 (或随机变量) 空间中的正交关系，反映了线性相关性。从几何直觉的角度看，它们都可以用来描述空间中的关系，例如向量之间的夹角、正交性，以及随机变量之间的线性关系。从这个角度看，向量的线性无关和随机变量的线性不相关之间确实存在一定的联系。

另外，判断向量线性相关的一个重要工具就是组成矩阵后计算并判断这个矩阵的秩。这也适用于判断随机变量的线性相关性。

下面，我们将从二次型 (Quadratic Forms) 与主轴方面的特性进行扩展理解。

3 二次型矩阵优化定理

定理 1 (在单位向量上优化二次型矩阵). *设 $A\in\mathbb{R}^{n\times n}$ 与相应的二次型矩阵 $\mathcal{Q}(\vec{x}):=\vec{x}^\top A\vec{x}$ 对称。设置 $M=\max\left\{\vec{x}^\top A\vec{x}:\|\vec{x}\|=1\right\}$ 和 $m=\min\left\{\vec{x}^\top A\vec{x}:\|\vec{x}\|=1\right\}$。则 M 是 A 的最大特征值 λ_1，m 是 A 的最小特征值 λ_n，这些值是通过将向量 $\vec{x}$ 设为相应的单位特征向量得到的。*

证明. 将 $A=PDP^{-1}$ 正交对角化，并用 P 作为变量矩阵的变换，消除了在 $\mathcal{Q}$ 中的交叉项，$\vec{x}=P\vec{y}\Longrightarrow\vec{x}^\top A\vec{x}=\vec{y}^\top D\vec{y}$。注意：$\|\vec{x}\|=\vec{x}^\top\vec{x}=(P\vec{y})^\top(P\vec{y})\cdots$ □

我们可以将二次型矩阵 $\mathcal{Q}$ 的其它特征值解释为关于附加约束的最大化 $\mathcal{Q}$。

定理 2 (附加约束条件). 设 $A \in \mathbb{R}^{n\times n}$ 与相应的二次型矩阵 $\mathcal{Q}(\vec{x}) := \vec{x}^\top A\vec{x}$ 对称，并可正交对角化 $A = PDP^{-1}$。不失一般性，D 的对角项按降序排列 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$，所以 $P = [\vec{u}_1 \cdots \vec{u}_n]$，其中 $\vec{u}_i$ 是对应于 λ_i 的特征向量。则对于 $k = 2, 3, \ldots n$，$\mathcal{Q}(\vec{x}) := \vec{x}^\top A\vec{x}$ 的最大值服从于

$$\vec{x}^\top\vec{x} = 1, \quad \vec{x}^\top\vec{u}_1 = 0, \quad \ldots, \quad \vec{x}^\top\vec{u}_{k-1} = 0$$

条件的约束，即对于特征值 λ_k，当 $\vec{x} = \vec{u}_k$ 时获得。

上述定理 1，描述的就是协方差矩阵，由 P，也就是特征向量决定扩散方向，由 λ_i，也就是特征值决定各个方向的扩散大小。定理 2，描述的就是随机变量的线性不相关，但是以与特征向量的线性无关方式描述。

4 直观理解向量内积

我们可以从柯西-施瓦兹不等式的向量形式直观理解向量内积。

4.1 定理证明

柯西-施瓦兹不等式的向量形式： 若有向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，则

$$\left|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right| \leq \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|$$

$$-1 \leq \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \leq 1$$

当且仅当其中一个向量为 $\mathbf{0}$，或一个向量为另外一个向量的倍数时，等号成立。

证明： 有向量 $\boldsymbol{\epsilon} = \boldsymbol{x} - \lambda\boldsymbol{y}$，我们计算它的范数为：

$$\begin{aligned}
\|\boldsymbol{\epsilon}\|^2 &\geq 0 \\
(\boldsymbol{x} - \lambda\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \lambda\boldsymbol{y}) &\geq 0 \\
\left(\boldsymbol{x}^{\mathrm{T}} - \lambda\boldsymbol{y}^{\mathrm{T}}\right)(\boldsymbol{x} - \lambda\boldsymbol{y}) &\geq 0 \\
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\lambda\boldsymbol{y} - \lambda\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} + \lambda^2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} &\geq 0 \\
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - \lambda\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right) + \lambda^2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} &\geq 0 \\
\text{Let } a = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y},\ b = \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right),\ c = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}& \\
a\lambda^2 - b\lambda + c &\geq 0
\end{aligned}$$

对下式求极值

$$a\lambda^2 - b\lambda + c = 0$$

这是开口向上的抛物线函数，有极值

$$\begin{aligned}
2a\lambda - b &= 0 \\
\lambda &= \frac{b}{2a}
\end{aligned}$$

将 λ 代入上式得

$$
\begin{aligned}
a\left(\frac{b}{2a}\right)^2 - b\left(\frac{b}{2a}\right) + c &\geq 0 \\
\frac{b^2}{4a} - \frac{b^2}{2a} + c &\geq 0 \\
c &\geq \frac{b^2}{4a}
\end{aligned}
$$

根据向量的内积公式，我们有 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}$，并且向量 $\boldsymbol{x}$ 的欧几里德范数为：$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$。将上述数值代回公式得

$$
\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} &\geq \frac{\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right)^2}{4\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}} \\
\because\ & \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} \\
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} &\geq \frac{4\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right)^2}{4\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}} \\
\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\right) &\geq \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right)^2 \\
\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 &\geq \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right)^2 \\
\|\boldsymbol{x}\|\|\boldsymbol{y}\| &\geq \left\|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right\|
\end{aligned}
$$

我们还有

$$
\begin{aligned}
\lambda &= \frac{b}{2a} \\
&= \left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right)\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\right)^{-1} \\
&= \left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\right)^{-1}
\end{aligned}
$$

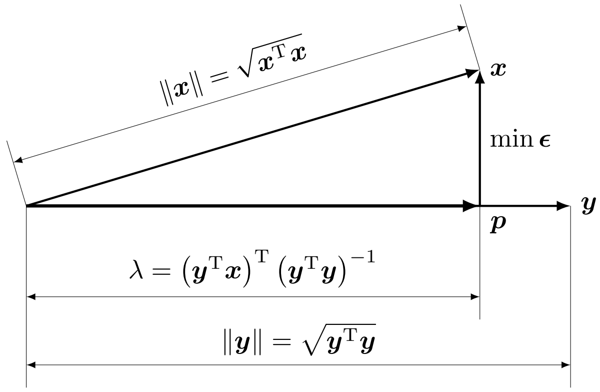
同时向量 $\boldsymbol{x}$ 到 $\boldsymbol{y}$ 的向量投影 (vector projection) $\boldsymbol{p}$ 为

$$
\boldsymbol{p} = \frac{\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}\boldsymbol{y} = \lambda\boldsymbol{y}
$$

可以看到，λ 是向量的投影系数。

4.2 几何解释

从几何上，有向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，λ 代表比例因子，$\boldsymbol{\epsilon} = \boldsymbol{x} - \lambda\boldsymbol{y}$ 则代表从 $\lambda\boldsymbol{y}$ 指向 $\boldsymbol{x}$ 的向量。柯西-施瓦兹不等式的含义是调整 λ 的大小，找到向量 $\boldsymbol{x} - \lambda\boldsymbol{y}$ 的最小范数，就是向量 $\lambda\boldsymbol{y}$ 到向量 $\boldsymbol{x}$ 的最短距离。当向量 $\boldsymbol{x}$ 和向量 $\boldsymbol{y}$ 共线时，该距离为 0。而 $\lambda = \left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\right)^{\mathrm{T}}\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\right)^{-1}$，则意味着当 $\lambda\boldsymbol{y}$ 为向量 $\boldsymbol{x}$ 到 $\boldsymbol{y}$ 的向量投影时 ($\boldsymbol{\epsilon}\perp\boldsymbol{y}$)，$\boldsymbol{\epsilon}$ 有最小范数。



5 直观理解协方差矩阵

协方差矩阵为实对称半正定矩阵，有很多相关特性。最重要的特性就是协方差矩阵是可正交对角化的。

协方差矩阵正交对角化形式为

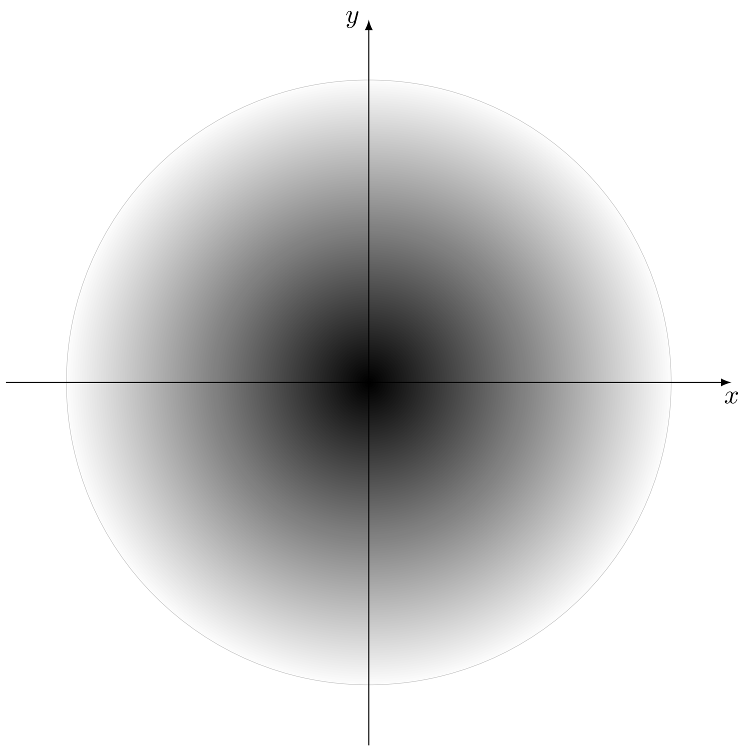
$$\mathrm{K}_{\mathbf{XX}} = \boldsymbol{R}\boldsymbol{\Sigma}^2\boldsymbol{R}^{\mathrm{T}}$$

其中，$\boldsymbol{R}$ 是以特征向量为列向量构成的正交矩阵，并且有特性 $\boldsymbol{R}^{-1} = \boldsymbol{R}^{\mathrm{T}}$，$\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{I}$，$\det(\boldsymbol{R}) = 1$，因此 $\boldsymbol{R}$ 实际上是 $\mathrm{SO}(n)$ 的旋转矩阵；$\boldsymbol{\Sigma}^2$ 是由特征值，在这里是方差 $\sigma_1^2, \ldots, \sigma_n^2$ 构成的对角矩阵，一般从大到小排列。

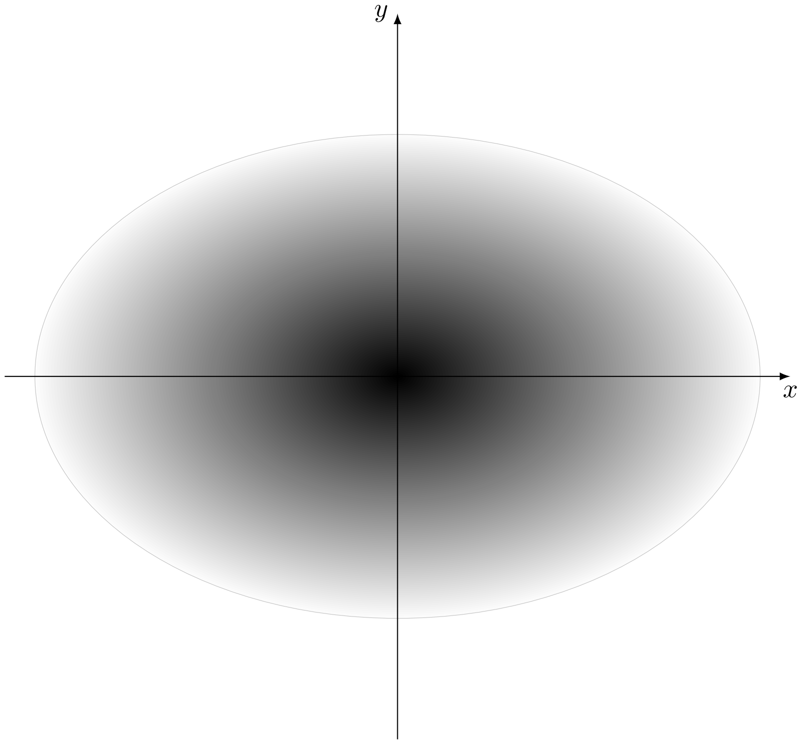
当协方差矩阵只有变量自身的方差而无变量之间的协方差时，矩阵可表示为

$$\mathrm{K}_{\mathbf{XX}} = \boldsymbol{I}\boldsymbol{\Sigma}^2\boldsymbol{I}^{\mathrm{T}}$$

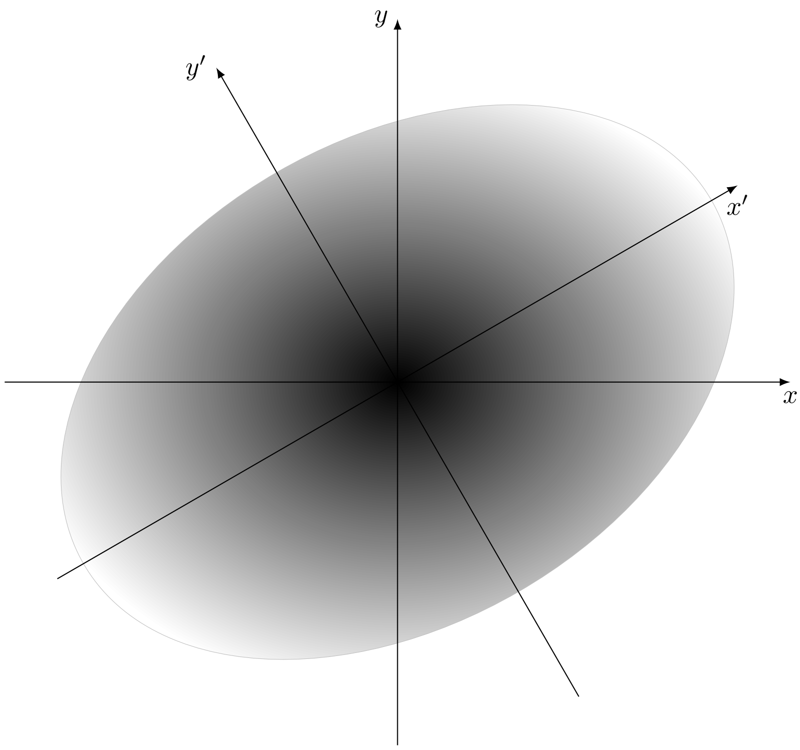
如果我们有高斯白色数据集合，表示为一个圆或一个球，



经过上式的线性变换，圆或球被拉伸为椭圆或椭球，数据在各轴的缩放程度由各轴的方差 σ_i^2 决定。



如果协方差矩阵变量之间的协方差不为零，则椭圆或椭球发生了旋转，因为 $\det(\boldsymbol{R}) = 1$，所以各个数据点到椭圆或椭球的中心的距离不会发生改变，也就是椭圆或椭球的形状不会发生改变。



对于高斯白色数据，协方差矩阵中的变量方差将圆拉伸为椭圆，变量协方差将椭圆旋转到新的

方向，$\boldsymbol{R}$ 中的列向量为新坐标系的正交基，$\boldsymbol{R}$ 就是新坐标系相对于旧坐标系的姿态，类似于向量夹角关系。最终的椭圆的形状，由 $\boldsymbol{R}$，也就是特征向量决定扩散方向，由 σ_i^2，也就是特征值决定扩散大小。

6 总结

首先，你的观察非常独到。协方差和向量内积确实有一些相似之处，它们都可以看作是某种形式的"相乘"并求和。这种操作在数学中被称为"内积"，并且是线性代数中的重要概念。

内积可以定义在任何向量空间上，其中向量可以是实数、复数、函数、随机变量等等。内积给出了向量空间中向量的几何结构，它可以用来定义长度 (范数)、角度 (夹角) 和正交性。

协方差和向量内积都可以看作是特殊的内积。协方差是随机变量向量空间中的内积，它衡量的是两个随机变量之间的线性关系。向量内积则是欧几里得空间中的内积，它衡量的是两个向量之间的几何关系。

协方差为 0 意味着随机变量线性无关，这与向量内积为 0 (向量正交) 的含义相似。但是，这两种无关性是不同的。向量的线性无关是指没有线性关系，而随机变量的线性无关是指没有线性相关性。这是因为随机变量的线性关系是指存在一种确定的关系，而向量的线性关系是指存在一种可能的关系。

你提到的"相乘"的规则 (代数结构) 在数学中被称为"环"或"域"。环是一个包含加法和乘法两种运算的代数结构，域则进一步要求乘法还满足除法。实数、复数、矩阵、函数等都可以构成环或域。

关于你的最后一个问题，向量的线性无关和随机变量的线性无关之间的联系，我认为一个可能的联系是：它们都可以通过某种内积 (协方差或向量内积) 来判定。如果一组向量的任意两个向量都正交 (内积为 0)，那么这组向量就是线性无关的；如果一组随机变量的任意两个变量都不相关 (协方差为 0)，那么这组随机变量就是线性无关的。这是因为在这两种情况下，没有一个向量 (或随机变量) 可以表示为其他向量 (或随机变量) 的线性组合。