第 1 章

预 备 知 识

卡尔曼滤波器在工程应用中的重要性得到了广泛的认可，并且建立了严格的数学理论。本书的主要目的是全面讨论卡尔曼滤波的数学理论、计算算法及其在实时跟踪问题中的应用。

为了解释如何得到卡尔曼滤波算法，以及它有哪些优良的性能，必须应用一些矩阵理论的公式和不等式。此外，在实时应用中考虑了系统和量测噪声的统计特性，必须具备一些概率论的基本概念。本章将专门研究这些主题。

1.1 矩阵和行列式初步

用 $\mathbf{R}^n$ 表示所有列向量 $\boldsymbol{x}=[x_1 \quad \cdots \quad x_n]^{\mathrm{T}}$ 的空间，其中 $x_1,\cdots,x_n$ 是实数。对于所有 $\mathbf{R}^n$ 中的非零向量 $\boldsymbol{x}$，如果 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 是一个正数，则称 $n\times n$ 的实矩阵 $\boldsymbol{A}$ 是正定的：如果 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 非负，则称 $\boldsymbol{A}$ 是非负定的。如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是两个任意 $n\times n$ 阶实数矩阵，当 $\boldsymbol{A}-\boldsymbol{B}$ 正定时，可以表示为

$$\boldsymbol{A} > \boldsymbol{B}$$

当 $\boldsymbol{A}-\boldsymbol{B}$ 非负定时，可以表示为

$$\boldsymbol{A} \geqslant \boldsymbol{B}$$

首先来复习施瓦兹不等式(Schwarz inequality)：

$$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leqslant |\boldsymbol{x}|\,|\boldsymbol{y}|, \quad \boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$$

式中，$|\boldsymbol{x}|=(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^{1/2}$。另外，当且仅当 $\boldsymbol{x}$、$\boldsymbol{y}$ 平行时，上面的不等式变成等式，即

$$\boldsymbol{x} = \lambda \boldsymbol{y} \quad 或 \quad \boldsymbol{y} = \lambda \boldsymbol{x}$$

λ 为比例因子。特别地,如果 $\boldsymbol{y} \neq \boldsymbol{0}$,施瓦兹不等式可以写为

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \geqslant (\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y})^{-1}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x})$$

借助这个公式,可以将施瓦兹不等式推广到矩阵形式。

引理 1.1(矩阵施瓦兹不等式(Matrix Schwarz inequality)) 若 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 分别是 $m \times n$ 和 $m \times l$ 矩阵,$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$ 非奇异,则

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} \geqslant (\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q})^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})^{-1}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q}) \tag{1.1}$$

此外,对于某些 $m \times l$ 矩阵 $\boldsymbol{S}$,当且仅当 $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{S}$ 时,式(1.1)的等号成立。 ■

证明如下。

向量形式的施瓦兹不等式的证明比较简单。二次多项式[①]

$$(\boldsymbol{x} - \lambda \boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \lambda \boldsymbol{y}) \geqslant 0, \quad \boldsymbol{y} \neq \boldsymbol{0}$$

关于 λ 的最小值在

$$\lambda = (\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y})^{-1}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x})$$

时得到。将该 λ 值代入上面的不等式,即得施瓦兹不等式。在矩阵形式中,考虑到

$$(\boldsymbol{Q} - \boldsymbol{P}\boldsymbol{S})^{\mathrm{T}}(\boldsymbol{Q} - \boldsymbol{P}\boldsymbol{S}) \geqslant \boldsymbol{0}$$

同时令

$$\boldsymbol{S} = (\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})^{-1}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q})$$

得到

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} \geqslant \boldsymbol{S}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q}) + (\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q})^{\mathrm{T}}\boldsymbol{S} - \boldsymbol{S}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})\boldsymbol{S} = (\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q})^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})^{-1}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q})$$

和式(1.1)相同。对于 $n \times l$ 矩阵 $\boldsymbol{S}$,当且仅当

$$(\boldsymbol{Q} - \boldsymbol{P}\boldsymbol{S})^{\mathrm{T}}(\boldsymbol{Q} - \boldsymbol{P}\boldsymbol{S}) = \boldsymbol{0}$$

即 $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{S}$ 时,该不等式的等号成立。这就完成了该引理的证明。

下面来看矩阵求逆引理:

引理 1.2(矩阵求逆引理(Matrix inversion lemma)) 令

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$$

式中,$\boldsymbol{A}_{11}$ 和 $\boldsymbol{A}_{22}$ 分别是 $n \times n$ 和 $m \times m$ 的非奇异子矩阵。这样,

$$(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}) \text{ 和 } (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})$$

都是非奇异的,所以 $\boldsymbol{A}$ 非奇异,

$$\begin{aligned}
\boldsymbol{A}^{-1} &= \begin{bmatrix} \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \\ -(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} & -(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} & \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \end{bmatrix}
\end{aligned} \tag{1.2}$$

---

① 下式译者作了微小修订,补充了"$\geqslant 0$"。

特别地，

$$(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}=\boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} \tag{1.3}$$

及

$$\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}=(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \tag{1.4}$$

更进一步有

$$\begin{aligned}\det\boldsymbol{A}&=(\det\boldsymbol{A}_{11})\det(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})\\&=(\det\boldsymbol{A}_{22})\det(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})\end{aligned} \tag{1.5}$$ ■

证明如下。

可以将 $\boldsymbol{A}$ 表示为

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_m\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{bmatrix}$$

或

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{I}_n & \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\\ \boldsymbol{0} & \boldsymbol{I}_m\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{0}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}$$

取行列式，得到(1.5)式。特别地，从这里我们获得

$$\det\boldsymbol{A}\neq 0$$

即 $\boldsymbol{A}$ 是非奇异的。接下来，注意到

$$\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\\ \boldsymbol{0} & (\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}\end{bmatrix}$$

及

$$\begin{bmatrix}\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_m\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{I}_n & \boldsymbol{0}\\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_m\end{bmatrix}$$

可得

$$\boldsymbol{A}^{-1}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_m\end{bmatrix}^{-1}$$

这样就可以得到式(1.2)的第一部分。式(1.2)的第二部分可以用同样的方法证明。式(1.3)和式(1.4)可以根据式(1.2)的相应矩阵块相等得到。

直接应用引理 1.2 可以得到下面的结果。

引理 1.3　如果 $\boldsymbol{P}\geqslant\boldsymbol{Q}>\boldsymbol{0}$，则 $\boldsymbol{Q}^{-1}\geqslant\boldsymbol{P}^{-1}>\boldsymbol{0}$。 ■

证明如下。

令 $\boldsymbol{P}(\varepsilon)=\boldsymbol{P}+\varepsilon\boldsymbol{I}$，其中 $\varepsilon>0$，则 $\boldsymbol{P}(\varepsilon)-\boldsymbol{Q}>\boldsymbol{0}$。根据引理 1.2，有

$$\begin{aligned}\boldsymbol{P}^{-1}(\varepsilon) &= [\boldsymbol{Q}+(\boldsymbol{P}(\varepsilon)-\boldsymbol{Q})]^{-1} \\ &= \boldsymbol{Q}^{-1}-\boldsymbol{Q}^{-1}[(\boldsymbol{P}(\varepsilon)-\boldsymbol{Q})^{-1}+\boldsymbol{Q}^{-1}]^{-1}\boldsymbol{Q}^{-1}\end{aligned}$$

可得

$$\boldsymbol{Q}^{-1}-\boldsymbol{P}^{-1}(\varepsilon)=\boldsymbol{Q}^{-1}[(\boldsymbol{P}(\varepsilon)-\boldsymbol{Q})^{-1}+\boldsymbol{Q}^{-1}]^{-1}\boldsymbol{Q}^{-1}>\mathbf{0}$$

令 $\varepsilon\to 0$,得到 $\boldsymbol{Q}^{-1}-\boldsymbol{P}^{-1}\geqslant\mathbf{0}$,所以

$$\boldsymbol{Q}^{-1}\geqslant\boldsymbol{P}^{-1}>\mathbf{0}$$

下面讨论 $n\times n$ 矩阵 $\boldsymbol{A}$ 的迹。矩阵 $\boldsymbol{A}$ 的迹可以表示为 $\mathrm{tr}\boldsymbol{A}$,定义为矩阵 $\boldsymbol{A}$ 对角线元素的和,即

$$\mathrm{tr}\boldsymbol{A}=\sum_{i=1}^{n}a_{ii}$$

式中,$\boldsymbol{A}=[a_{ij}]$。首先来介绍一些基本性质。

引理 1.4 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是 $n\times n$ 阶矩阵,则

$$\mathrm{tr}\boldsymbol{A}^{\mathrm{T}}=\mathrm{tr}\boldsymbol{A} \tag{1.6}$$

$$\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})=\mathrm{tr}\boldsymbol{A}+\mathrm{tr}\boldsymbol{B} \tag{1.7}$$

$$\mathrm{tr}(\lambda\boldsymbol{A})=\lambda\mathrm{tr}\boldsymbol{A} \tag{1.8}$$

如果 $\boldsymbol{A}$ 是 $n\times m$ 阶矩阵,$\boldsymbol{B}$ 是 $m\times n$ 阶矩阵,则

$$\mathrm{tr}\boldsymbol{AB}=\mathrm{tr}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}=\mathrm{tr}\boldsymbol{BA}=\mathrm{tr}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}} \tag{1.9}$$

$$\mathrm{tr}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\sum_{i=1}^{n}\sum_{j=1}^{m}a_{ij}^{2} \tag{1.10}$$ ■

上面恒等式的证明可以从定义直接得到,作为练习(见练习 1.1)留给读者自己完成。下面的结论很重要。

引理 1.5 令 $n\times n$ 阶矩阵 $\boldsymbol{A}$ 的特征值是 $\lambda_1,\cdots,\lambda_n$,重根也列入其中,则

$$\mathrm{tr}\boldsymbol{A}=\sum_{i=1}^{n}\lambda_i \tag{1.11}$$ ■

证明如下。

将 $\boldsymbol{A}$ 写为 $\boldsymbol{A}=\boldsymbol{UJU}^{-1}$,其中 $\boldsymbol{J}$ 是矩阵 $\boldsymbol{A}$ 的若尔当标准型(Jordan canonical form),$\boldsymbol{U}$ 为非奇异矩阵。应用式(1.9)可以得到

$$\mathrm{tr}\boldsymbol{A}=\mathrm{tr}(\boldsymbol{AU})\boldsymbol{U}^{-1}=\mathrm{tr}\boldsymbol{U}^{-1}(\boldsymbol{AU})=\mathrm{tr}\boldsymbol{J}=\sum_{i=1}^{n}\lambda_i$$

根据这个引理,如果 $\boldsymbol{A}>\mathbf{0}$,则 $\mathrm{tr}\boldsymbol{A}>0$;如果 $\boldsymbol{A}\geqslant\mathbf{0}$,则 $\mathrm{tr}\boldsymbol{A}\geqslant 0$。

下面介绍一些关于迹的有用不等式。

引理 1.6 令 $\boldsymbol{A}$ 为 $n\times n$ 阶矩阵,则

$$\mathrm{tr}\boldsymbol{A}\leqslant(n\mathrm{tr}\boldsymbol{AA}^{\mathrm{T}})^{1/2} \tag{1.12}$$ ■

上面不等式的证明作为练习(见练习 1.2)留给读者自己完成。

引理 1.7　如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别是 $n\times m$ 和 $m\times l$ 阶矩阵，则

$$\operatorname{tr}(\boldsymbol{AB})(\boldsymbol{AB})^{\mathrm{T}} \leqslant (\operatorname{tr}\boldsymbol{AA}^{\mathrm{T}})(\operatorname{tr}\boldsymbol{BB}^{\mathrm{T}})$$

相应的对于任意适当维数的矩阵 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_p$，

$$\operatorname{tr}(\boldsymbol{A}_1\cdots\boldsymbol{A}_p)(\boldsymbol{A}_1\cdots\boldsymbol{A}_p)^{\mathrm{T}} \leqslant (\operatorname{tr}\boldsymbol{A}_1\boldsymbol{A}_1^{\mathrm{T}})\cdots(\operatorname{tr}\boldsymbol{A}_p\boldsymbol{A}_p^{\mathrm{T}}) \tag{1.13}$$ ■

证明如下。

如果 $\boldsymbol{A}=[a_{ij}]$、$\boldsymbol{B}=[b_{ij}]$，根据施瓦兹不等式，则

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{AB})(\boldsymbol{AB})^{\mathrm{T}} &= \operatorname{tr}\left[\sum_{k=1}^{m} a_{ik}b_{kj}\right]\left[\sum_{k=1}^{m} a_{jk}b_{ki}\right] \\
&= \operatorname{tr}\begin{bmatrix} \sum\limits_{p=1}^{l}\left(\sum\limits_{k=1}^{m} a_{1k}b_{kp}\right)^2 & & * \\ & \ddots & \\ * & & \sum\limits_{p=1}^{l}\left(\sum\limits_{k=1}^{m} a_{nk}b_{kp}\right)^2 \end{bmatrix} \\
&= \sum_{i=1}^{n}\sum_{p=1}^{l}\left(\sum_{k=1}^{m} a_{ik}b_{kp}\right)^2 \leqslant \sum_{i=1}^{n}\sum_{p=1}^{l}\sum_{k=1}^{m} a_{ik}^2 \sum_{k=1}^{m} b_{kp}^2 \\
&= \left(\sum_{i=1}^{n}\sum_{k=1}^{m} a_{ik}^2\right)\left(\sum_{p=1}^{l}\sum_{k=1}^{m} b_{kp}^2\right) \\
&= (\operatorname{tr}\boldsymbol{AA}^{\mathrm{T}})(\operatorname{tr}\boldsymbol{BB}^{\mathrm{T}})
\end{aligned}$$

这就完成了该引理的证明。

必须注意，对于 $\boldsymbol{A}\geqslant\boldsymbol{B}>\boldsymbol{0}$，不一定就有 $\operatorname{tr}\boldsymbol{AA}^{\mathrm{T}}\geqslant\operatorname{tr}\boldsymbol{BB}^{\mathrm{T}}$，一个例子为

$$\boldsymbol{A}=\begin{bmatrix}\frac{12}{5} & 0 \\ 0 & 1\end{bmatrix},\quad \boldsymbol{B}=\begin{bmatrix}2 & -1 \\ 1 & 1\end{bmatrix}$$

很显然，$\boldsymbol{A}-\boldsymbol{B}\geqslant\boldsymbol{0}$ 和 $\boldsymbol{B}>\boldsymbol{0}$，但是

$$\operatorname{tr}\boldsymbol{AA}^{\mathrm{T}}=\frac{169}{25}<7=\operatorname{tr}\boldsymbol{BB}^{\mathrm{T}}$$

（见练习 1.3）。

对于对称矩阵，可以用下面的方法得到需要的结论。

引理 1.8　令 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是非负定对称矩阵，且 $\boldsymbol{A}\geqslant\boldsymbol{B}$，则 $\operatorname{tr}\boldsymbol{AA}^{\mathrm{T}}\geqslant\operatorname{tr}\boldsymbol{BB}^{\mathrm{T}}$，或 $\operatorname{tr}\boldsymbol{A}^2\geqslant\operatorname{tr}\boldsymbol{B}^2$。 ■

将该引理的证明留作练习（见练习 1.4）。

引理 1.9　令 $\boldsymbol{B}$ 是 $n\times n$ 阶非负定对称矩阵，则

$$\operatorname{tr}\boldsymbol{B}^2\leqslant(\operatorname{tr}\boldsymbol{B})^2 \tag{1.14}$$

若 $\boldsymbol{A}$ 是另一个 $n\times n$ 阶非负定对称矩阵，且 $\boldsymbol{B}\leqslant\boldsymbol{A}$，则

$$\mathrm{tr}\boldsymbol{B}^2 \leqslant (\mathrm{tr}\boldsymbol{A})^2 \tag{1.15}$$ ■

为了证明式(1.14),令 $\boldsymbol{B}$ 的特征值为 $\lambda_1,\cdots,\lambda_n$,则 $\boldsymbol{B}^2$ 的特征值为 $\lambda_1^2,\cdots,\lambda_n^2$。因为 $\lambda_1,\cdots,\lambda_n$ 是非负的,根据引理1.5有

$$\mathrm{tr}\boldsymbol{B}^2 = \sum_{i=1}^{n}\lambda_i^2 \leqslant \left(\sum_{i=1}^{n}\lambda_i\right)^2 = (\mathrm{tr}\boldsymbol{B})^2$$

又 $\boldsymbol{B}\leqslant\boldsymbol{A}$,则 $\mathrm{tr}\boldsymbol{B}\leqslant\mathrm{tr}\boldsymbol{A}$,可得式(1.15)。

还有下面在以后章节中非常有用的结论。

引理1.10 设 $\boldsymbol{F}$ 是一个特征值为 $\lambda_1,\cdots,\lambda_n$ 的 $n\times n$ 阶矩阵。若

$$\lambda := \max(|\lambda_1|,\cdots,|\lambda_n|) < 1$$

则存在满足 $0<r<1$ 的实数 r 和常数 C,使得

$$|\mathrm{tr}\boldsymbol{F}^k(\boldsymbol{F}^k)^{\mathrm{T}}| \leqslant Cr^k,\quad k=1,2,\cdots$$ ■

证明如下。

令 $\boldsymbol{J}$ 是矩阵 $\boldsymbol{F}$ 的若尔当标准型,则存在非奇异矩阵 $\boldsymbol{U}$,使 $\boldsymbol{F}=\boldsymbol{U}\boldsymbol{J}\boldsymbol{U}^{-1}$。根据式(1.13),得

$$\begin{aligned}|\mathrm{tr}\boldsymbol{F}^k(\boldsymbol{F}^k)^{\mathrm{T}}| &= |\mathrm{tr}\boldsymbol{U}\boldsymbol{J}^k\boldsymbol{U}^{-1}(\boldsymbol{U}^{-1})^{\mathrm{T}}(\boldsymbol{J}^k)^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}| \\ &\leqslant |\mathrm{tr}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\|\mathrm{tr}\boldsymbol{J}^k(\boldsymbol{J}^k)^{\mathrm{T}}\|\mathrm{tr}\boldsymbol{U}^{-1}(\boldsymbol{U}^{-1})^{\mathrm{T}}| \\ &\leqslant p(k)\lambda^{2k}\end{aligned}$$

式中,$p(k)$是 k 的多项式。

取任意的 r 满足 $\lambda^2<r<1$,并对所有的 k,选择正常数 C 满足

$$p(k)\left(\frac{\lambda^2}{r}\right)^k \leqslant C$$

就可以得到所需结果。

1.2 概率论初步

投掷均匀硬币的试验,在每次投掷时,结果为正面朝上(H),或反面朝上(T)。将试验实际的结果称为试验的输出,所有可能的输出集合称为试验的样本空间(S)。例如,如果投掷一枚均匀的硬币两次,则两次投掷的结果可能为 HH、TT、HT 或 TH,集合$\{HH,TT,HT,TH\}$是一个样本空间 S。此外,该样本的任意子集称为一个事件;如果某个事件只有一个可能的结果,则称为基本事件。

因为没有预测输出结果的方法,需要针对每一个事件定义一个0和1之间的实数 P,来描述某个确定事件发生的概率。这可以由一个实值函数来描述,称

为定义在样本空间的随机变量。在上面的例子中，如果随机变量 $X=X(s)$，$s\in S$，表示 s 次试验中正面朝上的次数（H 发生），则 $P=P(X(s))$ 给出了 s 次试验中正面朝上的概率。令 S 为样本空间，X：$S\to\mathbf{R}^1$ 为随机变量。对于任何一个可测集合 $A\subset\mathbf{R}^1$（在本例中，$A=\{0\}$、$\{1\}$、$\{2\}$，分别表示没有出现正面、出现一次正面、出现两次正面），定义一个 P：$\{$事件$\}\to[0,1]$，其中每个事件是一个集合 $\{s\in S\colon X(s)\in A\subset\mathbf{R}^1\}:=\{X\in A\}$，满足以下条件：

(1) 对于任意可测集 $A\subset\mathbf{R}^1$，$P(X\in A)\geqslant 0$；

(2) $P(X\in\mathbf{R}^1)=1$；

(3) 对 $\mathbf{R}^1$ 中的任意两两互不相容的可测集 A_i

$$P\left(X\in\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(X\in A_i)$$

则称 P 为随机变量 X 的概率分布（或者概率分布函数）。

如果存在一个非负可积函数[①] f，使得

$$P(X\in A)=\int_A f(x)\mathrm{d}x \tag{1.16}$$

对所有的可测集 A 成立，则称 P 为连续概率分布，f 为随机变量 X 的概率密度函数。注意到我们已经定义 $f(x)\mathrm{d}x=\mathrm{d}\lambda$，其中 λ 是一个测度（比如阶梯函数），那么投掷硬币的离散型情形也包含在其中了。

如果相应的概率密度函数 f 定义为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2},\quad \sigma>0,\mu\in\mathbf{R} \tag{1.17}$$

则称为高斯（或者正态）概率密度函数，P 为随机变量 X 的正态分布，记为 $X\sim N(\mu,\sigma^2)$，可以简单地证明正态分布 P 为一个概率分布。实际上，(1) 因为 $f(x)>0$，对任意可测集 $A\subset\mathbf{R}$，有 $P(X\in A)=\int_A f(x)\mathrm{d}x\geqslant 0$；(2) 通过变换 $y=(x-\mu)/(\sqrt{2}\sigma)$，得

$$P(X\in\mathbf{R}^1)=\int_{-\infty}^{\infty}f(x)\mathrm{d}x=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2}\mathrm{d}y=1$$

（见练习题 1.5）；(3) 因为

$$\int_{\bigcup_i A_i}f(x)\mathrm{d}x=\sum_i\int_{A_i}f(x)\mathrm{d}x$$

对任意两两互不相交的可测集 $A_i\subset\mathbf{R}^1$ 有

---

① 译者注：原文中为“可积函数”。

$$P\left(X \in \bigcup_i A_i\right) = \sum_i P(X \in A_i)$$

令 X 为随机变量,则 X 的数学期望,即 X 值的均值,可以定义为

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \tag{1.18}$$

对任意以 f 为概率密度函数的随机变量 X,$E(X)$为实数。对于正态分布,再次做变换 $y=(x-\mu)/(\sqrt{2}\sigma)$,可得

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2} \mathrm{d}x \\ &= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (\sqrt{2}\sigma y + \mu)\mathrm{e}^{-y^2} \mathrm{d}y \\ &= \mu \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-y^2} \mathrm{d}y \\ &= \mu \end{aligned} \tag{1.19}$$

$E(X)$也称为概率密度函数 f 的一阶矩。二阶矩给出了 X 的方差,定义为

$$\operatorname{Var}(X) = E(X - E(X))^2 = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) \mathrm{d}x \tag{1.20}$$

这个值表示随机变量 X 偏离其数学期望 $E(X)$ 的离散程度。对于正态分布,再作变换 $y = (x-\mu)/(\sqrt{2}\sigma)$,代入等式$\int_{-\infty}^{\infty} y^2 \mathrm{e}^{-y^2} \mathrm{d}y = \sqrt{\pi}/2$(见练习 1.6),可得

$$\begin{aligned} \operatorname{Var}(X) &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2} \mathrm{d}x \\ &= \frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} y^2 \mathrm{e}^{-y^2} \mathrm{d}y \\ &= \sigma^2 \end{aligned} \tag{1.21}$$

下面来看由随机变量组成的随机向量。记 n 维随机向量 $\boldsymbol{X}=[X_1 \quad \cdots \quad X_n]^{\mathrm{T}}$,其中 $X_i(s) \in \mathbf{R}^1$,$s \in S$。

令 P 表示 $\boldsymbol{X}$ 的连续概率分布函数,则

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = \int_{A_1} \cdots \int_{A_n} f(x_1, \cdots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{1.22}$$

其中 $A_1,\cdots,A_n$ 为 $\mathbf{R}^1$ 上的可测集，f 为可积函数，则 f 称为随机向量 $\boldsymbol{X}$ 的联合概率密度函数，P 为随机向量 $\boldsymbol{X}$ 的联合概率分布(函数)。对于每个 $i(i=1,\cdots,n)$，定义

$$f_i(x)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_{i-1},x,x_{i+1},\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_{i-1}\mathrm{d}x_{i+1}\cdots\mathrm{d}x_n \tag{1.23}$$

显然$\int_{-\infty}^{\infty}f_i(x)\mathrm{d}x=1$，$f_i$ 为随机向量 $\boldsymbol{X}$ 的联合概率密度函数 $f(x_1,\cdots,x_n)$ 的第 i 个边缘概率密度函数。同理，按照定义 f_i 的方法，定义 f_{ij}、f_{ijk} 为分别去掉 x_i、x_j 与 x_i、x_j、x_k 的概率密度函数。如果

$$f(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}(\det\boldsymbol{R})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\underline{\mu})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{x}-\underline{\mu})\right\} \tag{1.24}$$

式中$\underline{\mu}$ 是一个 n 维常值向量，$\boldsymbol{R}$ 为对称正定矩阵，则 $f(\boldsymbol{x})$是随机向量 $\boldsymbol{X}$ 的高斯(或正态)概率密度函数。可以证明：

$$\int_{-\infty}^{+\infty}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}:=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(\boldsymbol{x})\mathrm{d}x_1\cdots\mathrm{d}x_n=1 \tag{1.25}$$

$$\begin{aligned}E(\boldsymbol{X})&=\int_{-\infty}^{+\infty}\boldsymbol{x}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\\&:=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}f(\boldsymbol{x})\mathrm{d}x_1\cdots\mathrm{d}x_n\\&=\underline{\mu}\end{aligned} \tag{1.26}$$

及

$$\mathrm{Var}(\boldsymbol{X})=E(\boldsymbol{X}-\underline{\mu})(\boldsymbol{X}-\underline{\mu})^{\mathrm{T}}=\boldsymbol{R} \tag{1.27}$$

事实上，因为 $\boldsymbol{R}$ 是一个对称正定矩阵，存在单位矩阵 $\boldsymbol{U}$ 满足 $\boldsymbol{R}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{U}$，其中 $\boldsymbol{J}=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$，且 $\lambda_1,\cdots,\lambda_n>0$。令 $y=\frac{1}{\sqrt{2}}\mathrm{diag}[\sqrt{\lambda_1},\cdots,\sqrt{\lambda_n}]\boldsymbol{U}(\boldsymbol{X}-\underline{\mu})$，则

$$\int_{-\infty}^{+\infty}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=\frac{2^{n/2}\sqrt{\lambda_1}\cdots\sqrt{\lambda_n}}{(2\pi)^{n/2}(\lambda_1\cdots\lambda_n)^{1/2}}\int_{-\infty}^{+\infty}\mathrm{e}^{-y_1^2}\mathrm{d}y_1\cdots\int_{-\infty}^{+\infty}\mathrm{e}^{-y_n^2}\mathrm{d}y_n=1$$

方程(1.26)和方程(1.27)的证明可以应用前面标量情况下采用的代换来完成(见式(1.21)和练习 1.7)。

下面介绍条件概率的概念。考虑这样一个试验，一个罐子中装有 M_1 个白球和 M_2 个黑球。当采用不放回抽取时，即第一个球取出后不再放入原来的罐子，求在第一次取到黑球(事件 A_1)的条件下，第二次又取到黑球(事件 A_2)的概率。

对于这个简单问题,我们这样解释:因为从罐中取到的第一个球是黑色的,则第二次取球前,罐中剩余 M_1 个白球和 M_2-1 个黑球,此时取得黑球的概率为

$$\frac{M_2-1}{M_1+M_2-1}$$

又

$$\frac{M_2-1}{M_1+M_2-1}=\frac{M_2}{M_1+M_2}\cdot\frac{M_2-1}{M_1+M_2-1}\Big/\frac{M_2}{M_1+M_2}$$

式中 $\frac{M_2}{M_1+M_2}$ 为第一次取球时取到黑球的概率,$\frac{M_2}{M_1+M_2}\cdot\frac{M_2-1}{M_1+M_2-1}$ 为第一次和第二次都取到黑球的概率。

根据上述例子,引出条件概率的定义:在 $X_2\in A_2$ 的条件下,$X_1\in A_1$ 的条件概率为

$$P(X_1\in A_1\mid X_2\in A_2)=\frac{P(X_1\in A_1,X_2\in A_2)}{P(X_2\in A_2)}\tag{1.28}$$

假定 P 为联合概率密度函数 f 所对应的连续概率分布函数,则式(1.28)变为

$$P(X_1\in A_1\mid X_2\in A_2)=\frac{\int_{A_1}\int_{A_2}f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2}{\int_{A_2}f_2(x_2)\mathrm{d}x_2}$$

式中 f_2 定义为

$$f_2(x_2)=\int_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_1$$

为 f 的第二个边缘概率密度函数。

令 $f(x_1|x_2)$ 表示 $P(X_1\in A_1|X_2\in A_2)$ 所对应的概率密度函数。$f(x_1|x_2)$ 被称为条件概率分布函数 $P(X_1\in A_1|X_2\in A_2)$ 所对应的条件概率密度函数。显然,

$$f(x_1\mid x_2)=\frac{f(x_1,x_2)}{f_2(x_2)}\tag{1.29}$$

这称作贝叶斯公式(参见《概率论》A. N. Shiryayev(1984))。同理,贝叶斯公式也可以写为

$$f(x_1,x_2)=f(x_1\mid x_2)f_2(x_2)=f(x_2\mid x_1)f_1(x_1)\tag{1.30}$$

上述贝叶斯公式对于随机向量 $\boldsymbol{X}_1,\boldsymbol{X}_2$ 也成立。

如果 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 分别是 n 维和 m 维随机向量,则 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的协方差是一个 $n\times m$ 矩阵

$$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})=E[(\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{Y}-E(\boldsymbol{Y}))^{\mathrm{T}}]\tag{1.31}$$