第二章 统计决策方法

苏智勇
可视计算研究组
南京理工大学
suzhiyong@njust.edu.cn
https://zhiyongsu.github.io

概率统计复习

一、概率

- **频率**：$f_n(A)=\dfrac{n_A}{n}=\dfrac{\text{A发生次数}}{\text{总次数}}$，表示事件$A$发生的频数/频率
- **概率**：表示事件A出现的可能性大小，$P(\cdot)$满足如下条件
  - **非负性**：$P(A)\geq 0$
  - **规范性/正则性**：$P(\Omega)=1$
  - **可列可加性**：若$A_1,A_2,\ldots\ldots,A_n$互不相容，则$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P\left(A_i\right)$

概率统计复习

一、概率

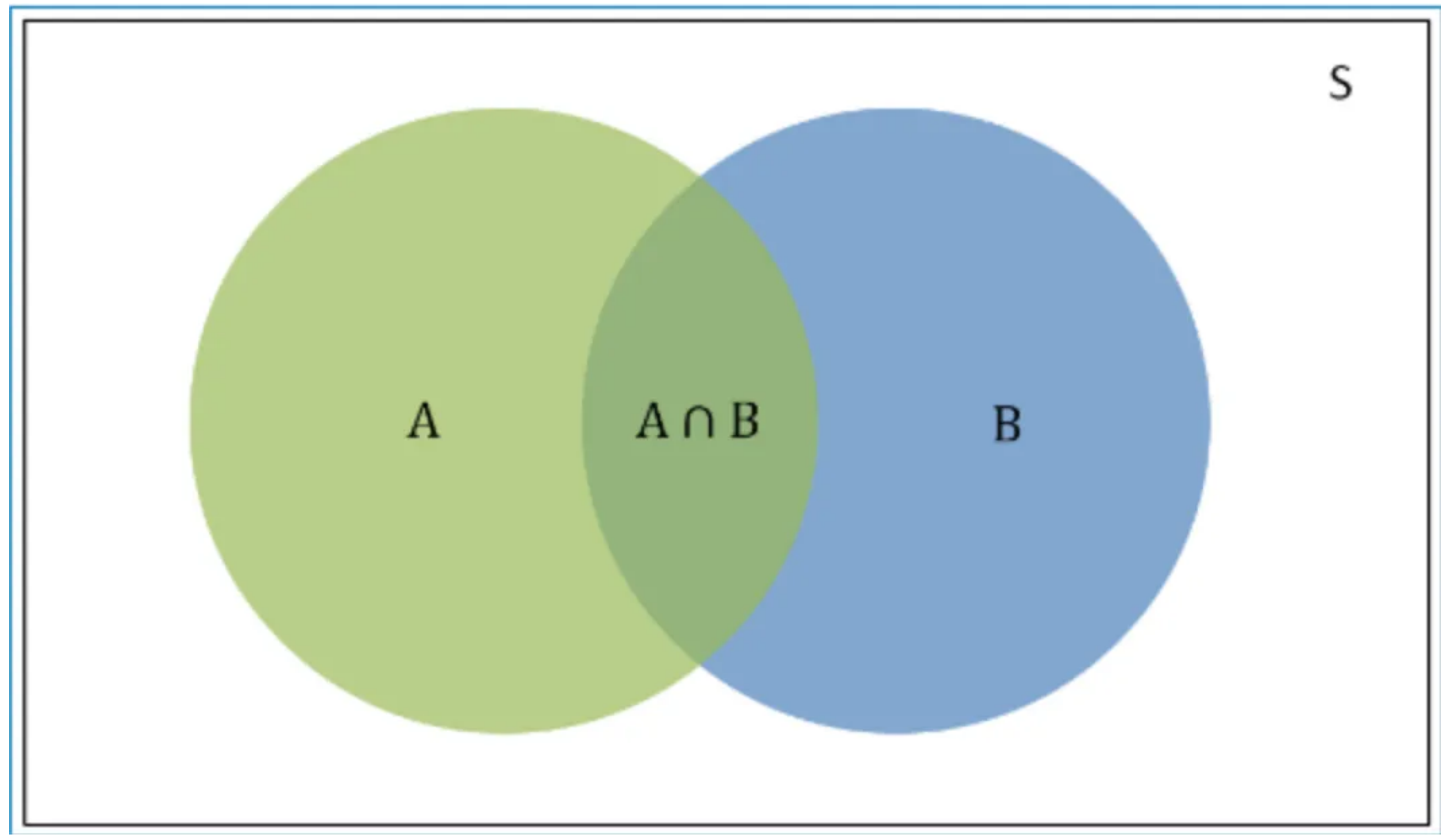
– 概率的性质

- $P(\overline{A})=1-P(A)$
- $P(A\bigcup B)=P(A)+P(B)-P(AB)$
- $P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_iA_j)+\sum_{1\le i<j<k\le n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_iA_2\cdots A_n)$

概率统计复习

二、条件概率

对于事件A和B，若$P(A)>0$时，则称$P(B|A)=\dfrac{P(AB)}{P(A)}$为在$A$出现的条件下，$B$出现的条件概率。



- **乘法公式**

✓$P(AB)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)$

✓$P(ABC)=P(A)P(B|A)P(C|AB)$

✓$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1})$

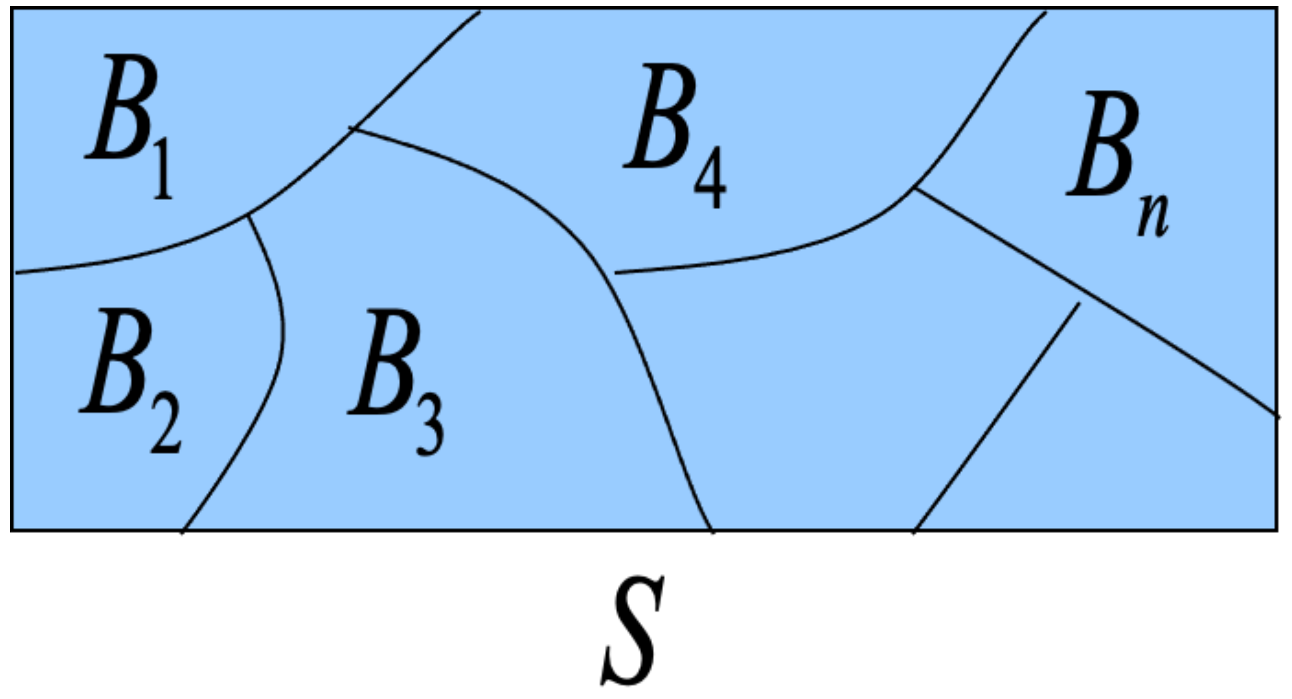
概率统计复习

二、条件概率

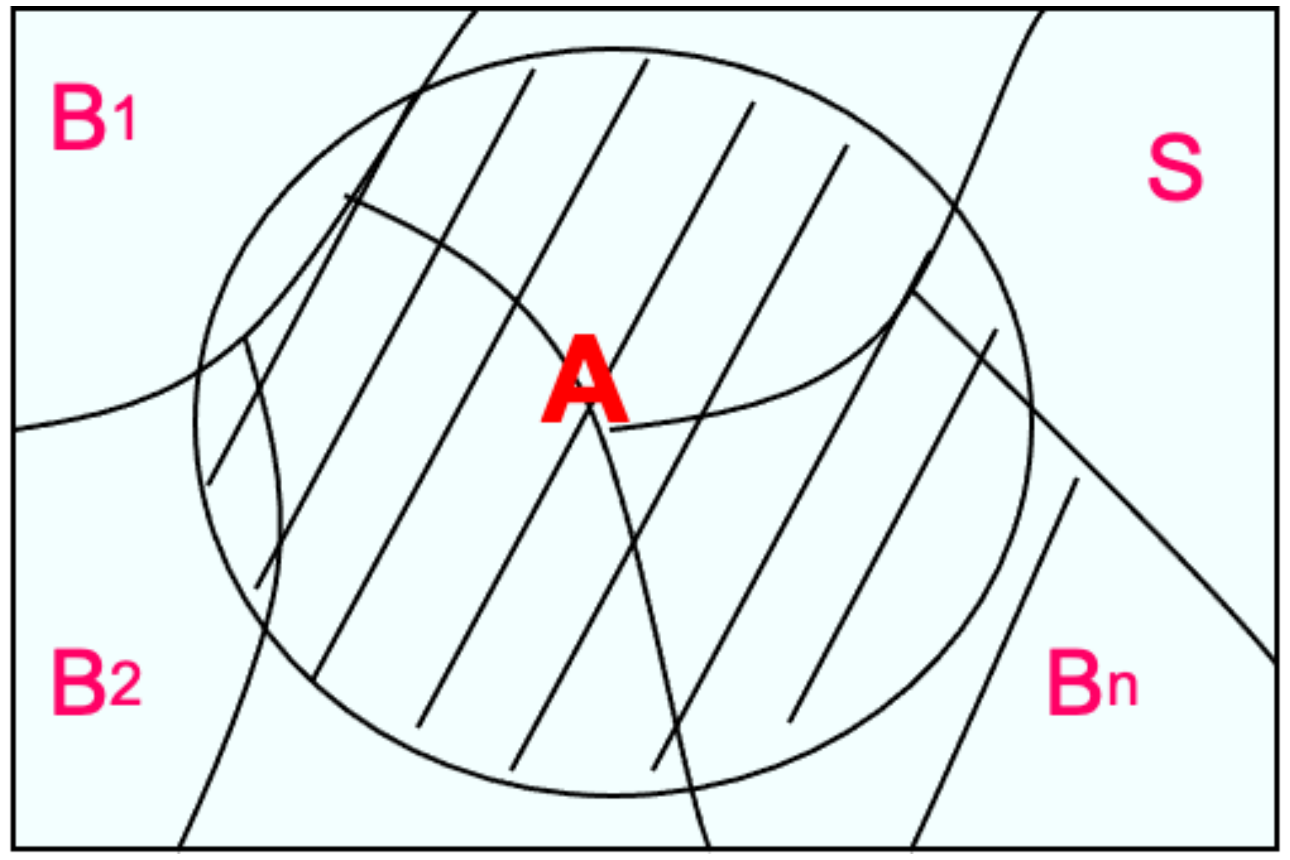
– 全概率公式

设事件$B_1, B_2, \cdots, B_n$为样本空间S的一个划分，且$P(B_i) > 0$，则：

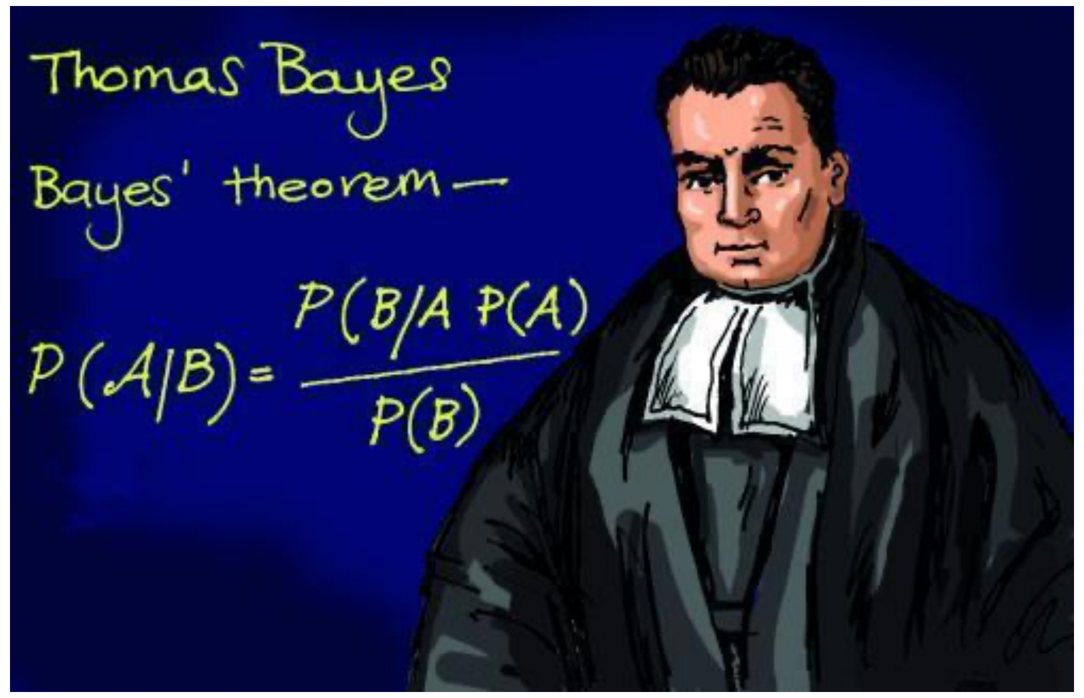
$$P(A) = \sum_{j=1}^{n} P(AB_j) = \sum_{j=1}^{n} P(B_j) \cdot P(A \mid B_j)$$



- 属于**先验概率**：A为“结果”，B_j为导致A发生的诸多“原因”之一，全概率公式可视为“**由因求果**”问题中的”因”出现的概率。
- **直观理解**：这件事还没有发生，根据以往的经验和数据推断出这件事会发生的概率（例如抛硬币）。



概率统计复习



二、条件概率

– 贝叶斯公式

设事件$B_1, B_2, \cdots, B_n$为样本空间S的一个划分，且$P(A) > 0$，$P(B_i) > 0$，则：

$$P(B_i \mid A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}$$

- 属于**后验概率**：在得到“结果”A的信息后重新修正的概率，是“**执果寻因**”问题中的”果”。
- **类概率密度**$P(A \mid B_i)$：在类B_i条件下的概率密度，即类B_i模式A的概率密度分布。
- **直观理解**：事件A已经发生，但是导致A发生的原因可能有多种（B_j），计算A发生的原因是由某个因素引起的概率。

概率统计复习

二、条件概率

- **先验概率** $P(B_i)$
  - 以全事件（一般是统计获得）为背景下，事件B_i发生的概率，$P(B_i) = P(B_i \mid \Omega)$
  - 根据以往经验和分析得到的概率，称为先验概率（没有实验前的概率），如**全概率公式**
- **后验概率** $P(B_i \mid A)$
  - 以新事件A为背景下，事件B_i发生的概率， $P(B_i \mid A)$
  - 新事件一般是实验，如实验A，此时的事件背景从全事件变成了A，该事件A可能对B_i的概率有影响，从$P(B_i \mid \Omega)$变为$P(B_i \mid A)$，称为后验概率，即试验(事件A发生)后的概率

概率统计复习

三、随机变量

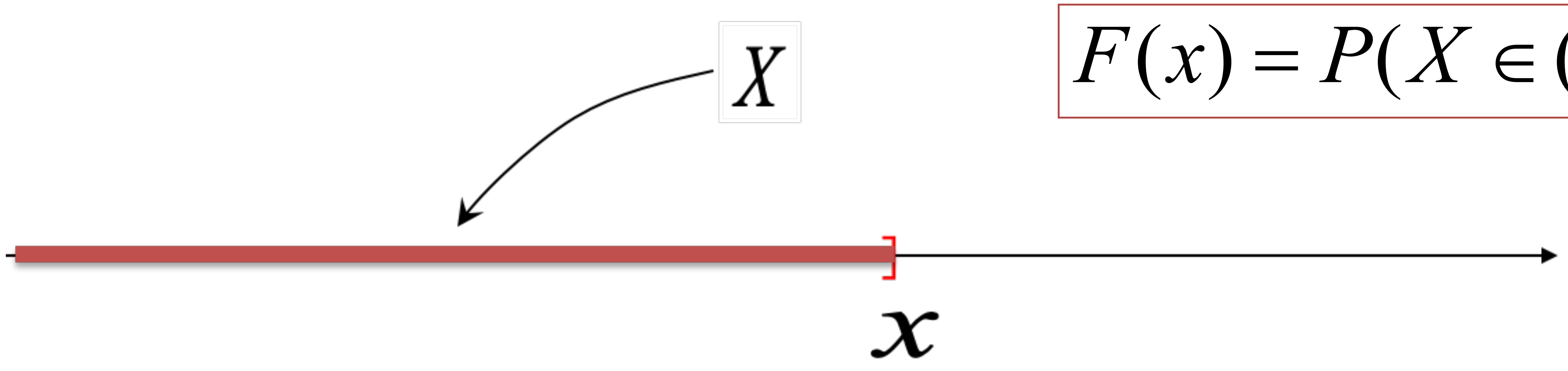
- **定义**：若随机变量X的取值为有限个或可数个，则称X为**离散型随机变量**
- **性质**：
  - $P(X=x_k)=p_k,\ \ k=1,2,\cdots$
  - $p_k \geq 0,\ \ \sum_{k=1}^{+\infty} p_k = 1$

X	x_1	x_2	…	x_k	…
P	p_1	p_2	…	p_k	…

概率统计复习

四、概率分布函数

- **定义**：随机变随机变量X, 对任意实数x, 称函数$F(x) = P(X \le x)$为X的**概率分布函数**，简称**分布函数**。
- **几何意义**：



$$\boxed{F(x) = P(X \in (-\infty, x])}$$

概率统计复习

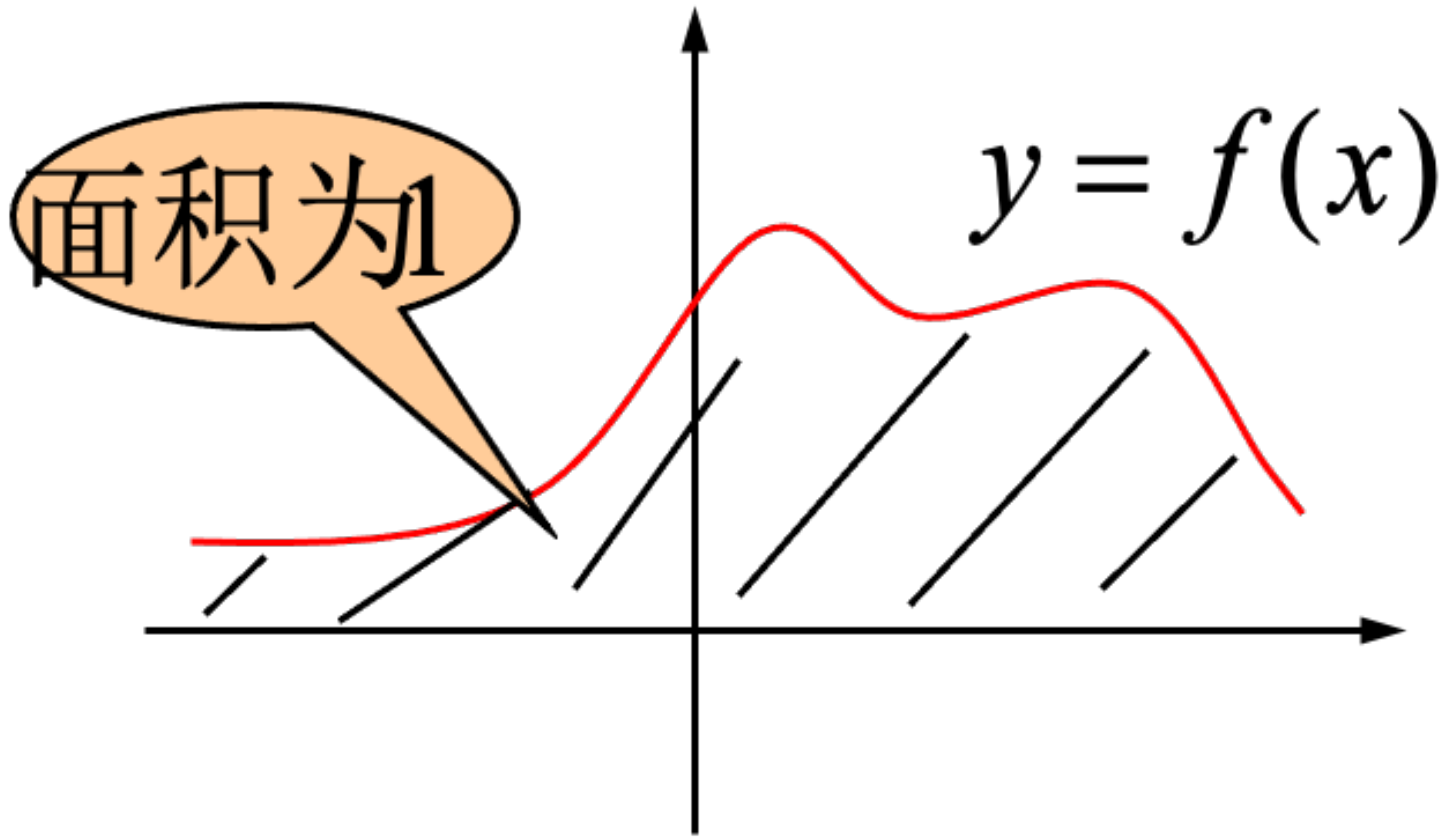
五、概率密度函数

- **定义**：对于随机变量X的分布函数$F(x)$，若存在非负的函数$f(x)$，使对于任意实数x有：$F(x)=\int_{-\infty}^{x} f(t)dt$，则称$X$为连续型随机变量，其中$f(x)$称为$X$的**概率密度函数**，简称**概率密度**。

概率统计复习

五、概率密度函数

- 性质
  - $f(x) \geq 0$
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$
  - $P(X \in D)=\int_D f(x)dx$，任意$D \subset R$
  - 在$f(x)$连续点x，$F'(x) = f(x)$



概率统计复习

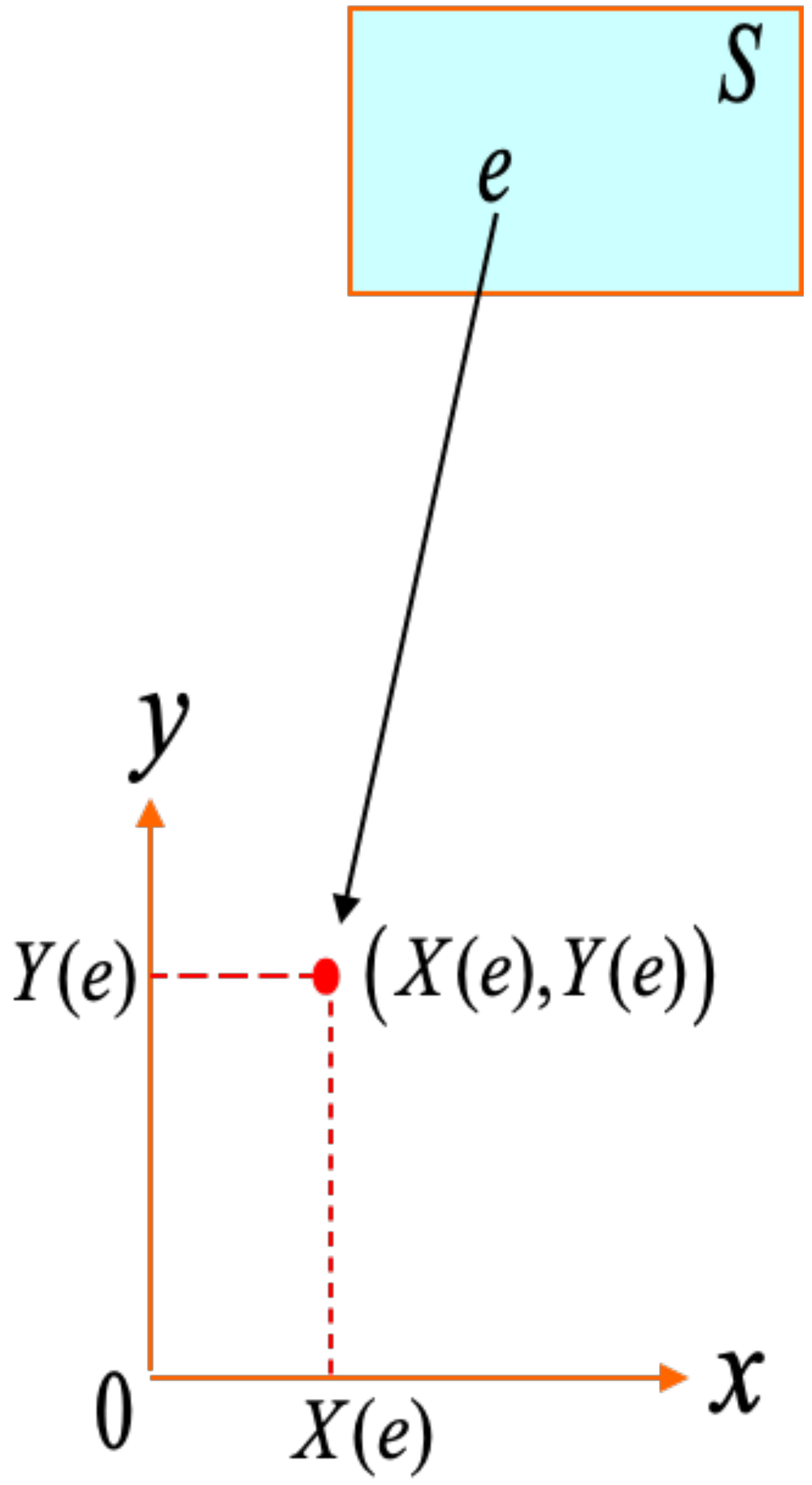
六、二元随机变量

- **定义**：设E是一个随机试验，样本空间

  $S=\{e\}$；设$X=X(e)$和$Y=Y(e)$是定义

  在S上的随机变量，由它们构成的向量

  (X,Y)称为**二维随机向量或二元随机变量**。



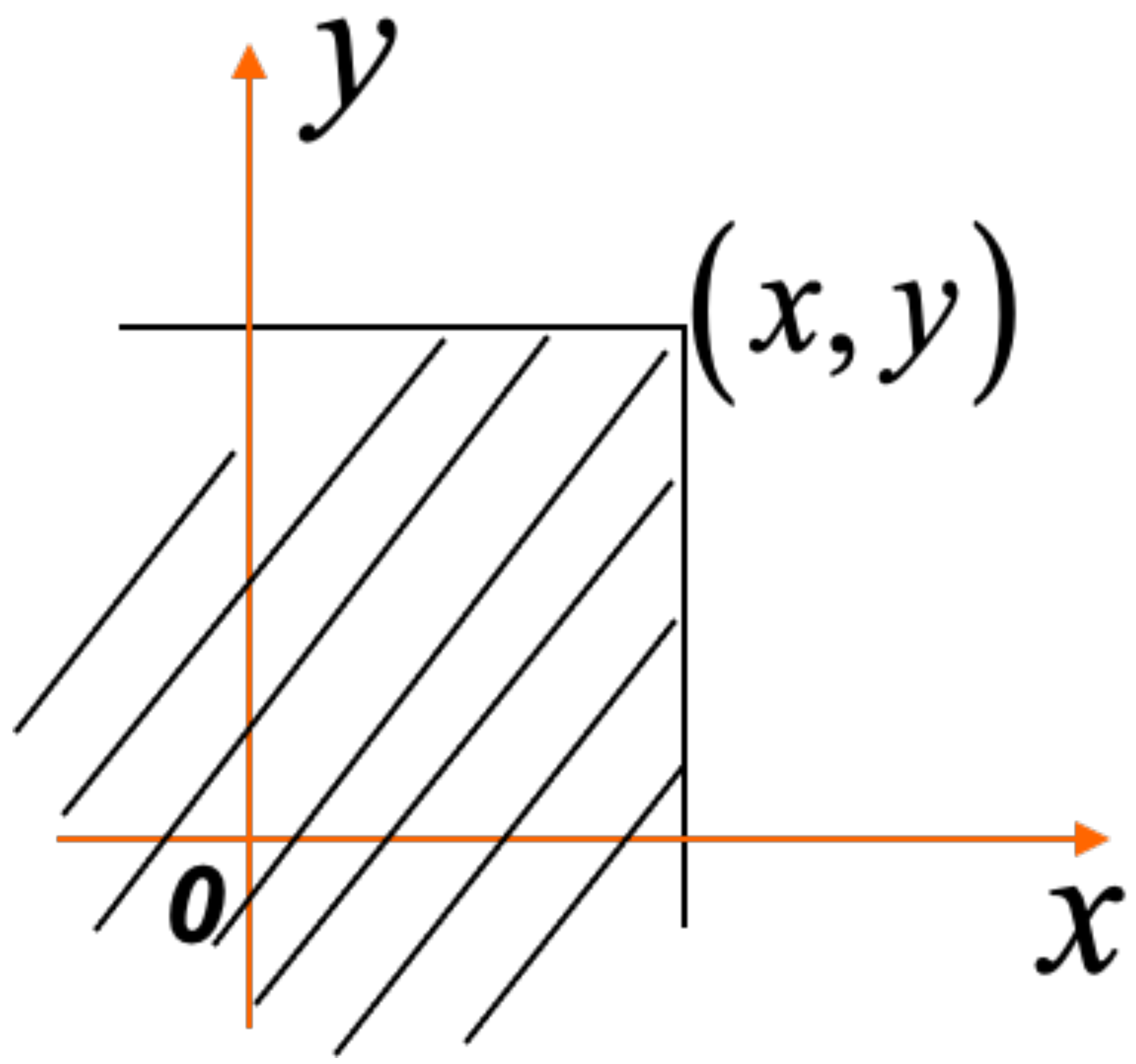
概率统计复习

七、联合分布函数

– **定义**：设(X, Y)是二元随机变量，对于任意实数 x, y，二元函数

$$F(x, y) = P(X \leq x, Y \leq y)$$

称为二元随机变量(X, Y)的**联合分布函数**



概率统计复习

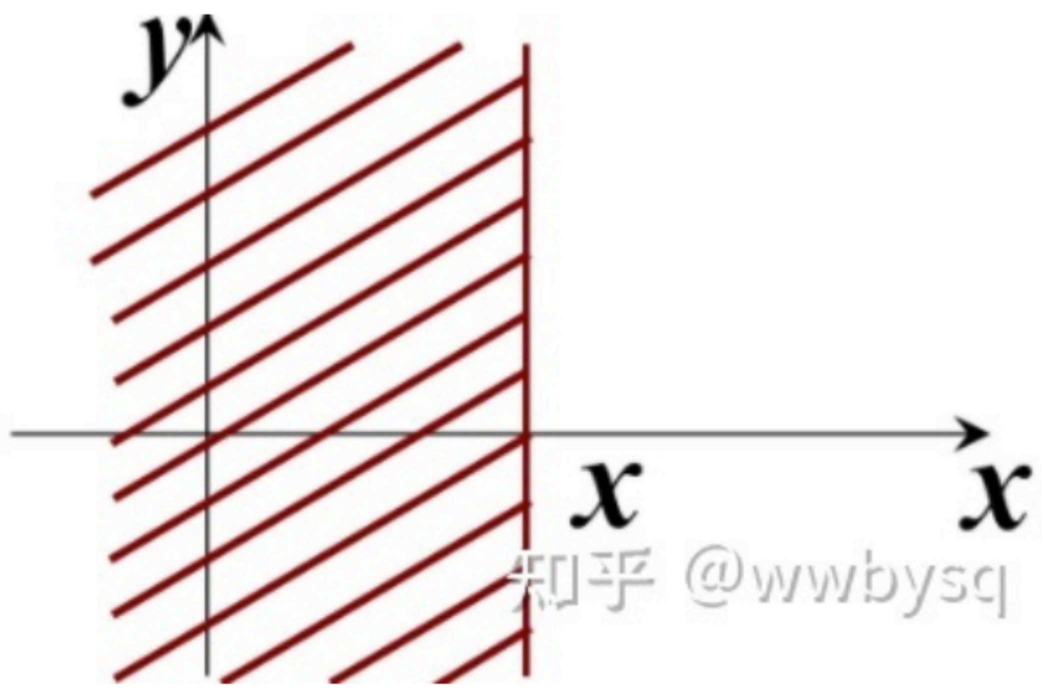
八、边际/边缘分布函数

– **定义**：X和Y也有它们自己的分布函数，分别记为 $F_X(x)$，$F_Y(y)$并称他们为边际/边缘分布函数

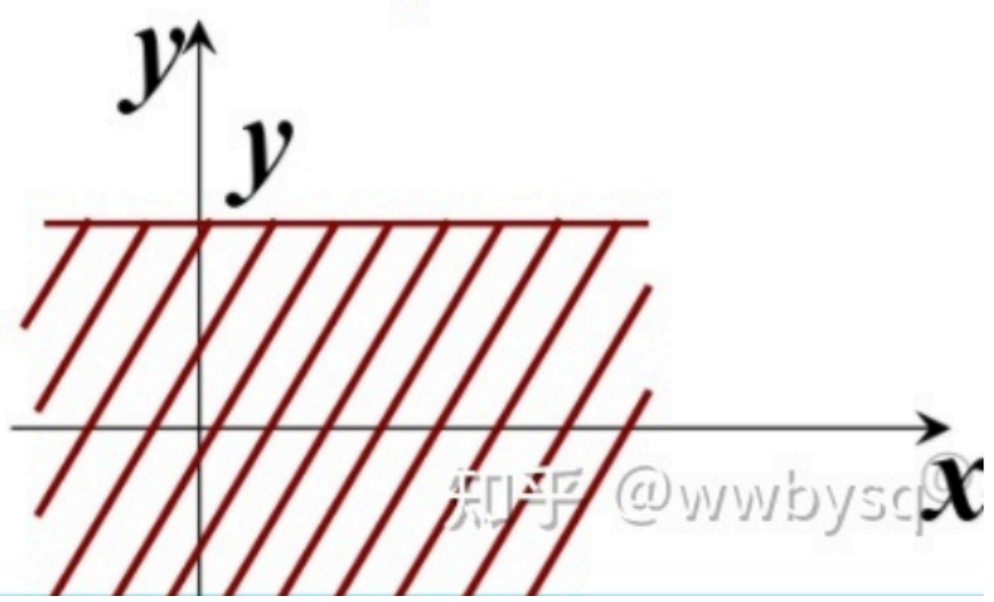
$$F_X(x) = P(X \leq x) = F(x, +\infty)$$

$$F_Y(y) = P(Y \leq y) = F(+\infty, y)$$

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(X \leq x, Y < +\infty) \\ &= F(x, +\infty) \end{aligned}$$



$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X < +\infty, Y \leq y) \\ &= F(+\infty, y) \end{aligned}$$



概率统计复习

九、条件分布函数

– **定义：** 若$P(Y=y)>0$，则在$Y=y$条件下，X的条件分布函数为

$$F_{X|Y}(x\mid y)=P(X\le x\mid Y=y)=\frac{P(X\le x, Y=y)}{P(Y=y)}$$

概率统计复习

十、联合概率密度函数

- **定义**：对于二元随机变量(X, Y)的分布函数$F(x, y)$，如果存在非负函数$f(x, y)$，使对于任意x, y，有

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)dudv$$

则称(X, Y)为二元连续型随机变量，并称$f(x, y)$为二元随机变量(X, Y)的**联合概率密度函数**。

概率统计复习

十一、数字特征

期望（数学期望，均值）

$$E(X)=\begin{cases}\sum_i x_i p_i, & 若X是离散型\\ \int_{-\infty}^{\infty} xf(x)dx, & 若X是连续型\end{cases}$$

$$E(g(X))=\begin{cases}\sum_i g(x_i)p_i, & 若X是离散型\\ \int_{-\infty}^{\infty} g(x)f(x)dx, & 若X是连续型\end{cases}$$

$$E(g(X,Y))=\begin{cases}\sum_i g(x_i,y_j)p_{ij}, & 若(X,Y)是离散型\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy, & 若(X,Y)是连续型\end{cases}$$

特别地，若(X,Y)是连续型，则

$$E(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy$$

$$E(Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$$

概率统计复习

十一、数字特征

方差：$D(X)=Var(X)=E\left\{[X-E(X)]^2\right\}.$ $=E(X^2)-[E(X)]^2$

$$\Rightarrow E(X^2)=Var(X)+[E(X)]^2$$

标准差：$\sigma(X)=\sqrt{Var(X)}$

协方差：$Cov(X,Y)=E\left\{[X-E(X)][Y-E(Y)]\right\}$ $=E(XY)-E(X)E(Y)$

$$\Rightarrow E(XY)=Cov(X,Y)+E(X)E(Y)$$

主要内容

2.1 引言

2.2 最小错误率贝叶斯决策

2.3 最小风险贝叶斯决策

2.4 Neyman-Pearson决策

2.5 正态分布时的统计决策

2.6 错误率

2.1 引言

- **贝叶斯定理有什么用**

根据**有限**的**历史**数据预测未来事情发生的概率，帮助我们更好地决策。

正向概率

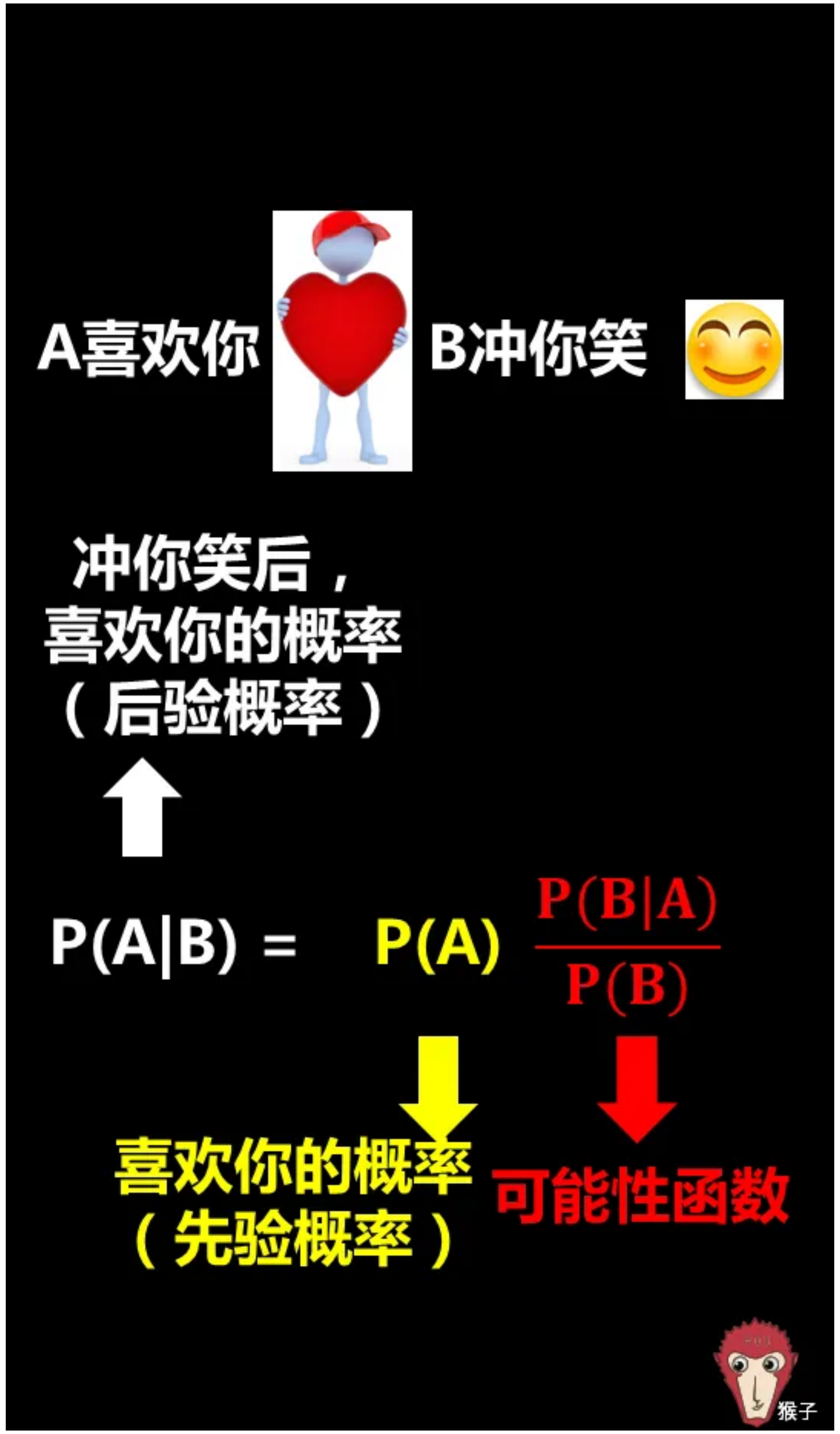
逆概率

2.1 引言

- **贝叶斯定理怎么用**
  - **思路**：在主观判断的基础上，先估计一个值（**先验概率**），然后根据观察的新信息不断修正(**可能性函数**)。

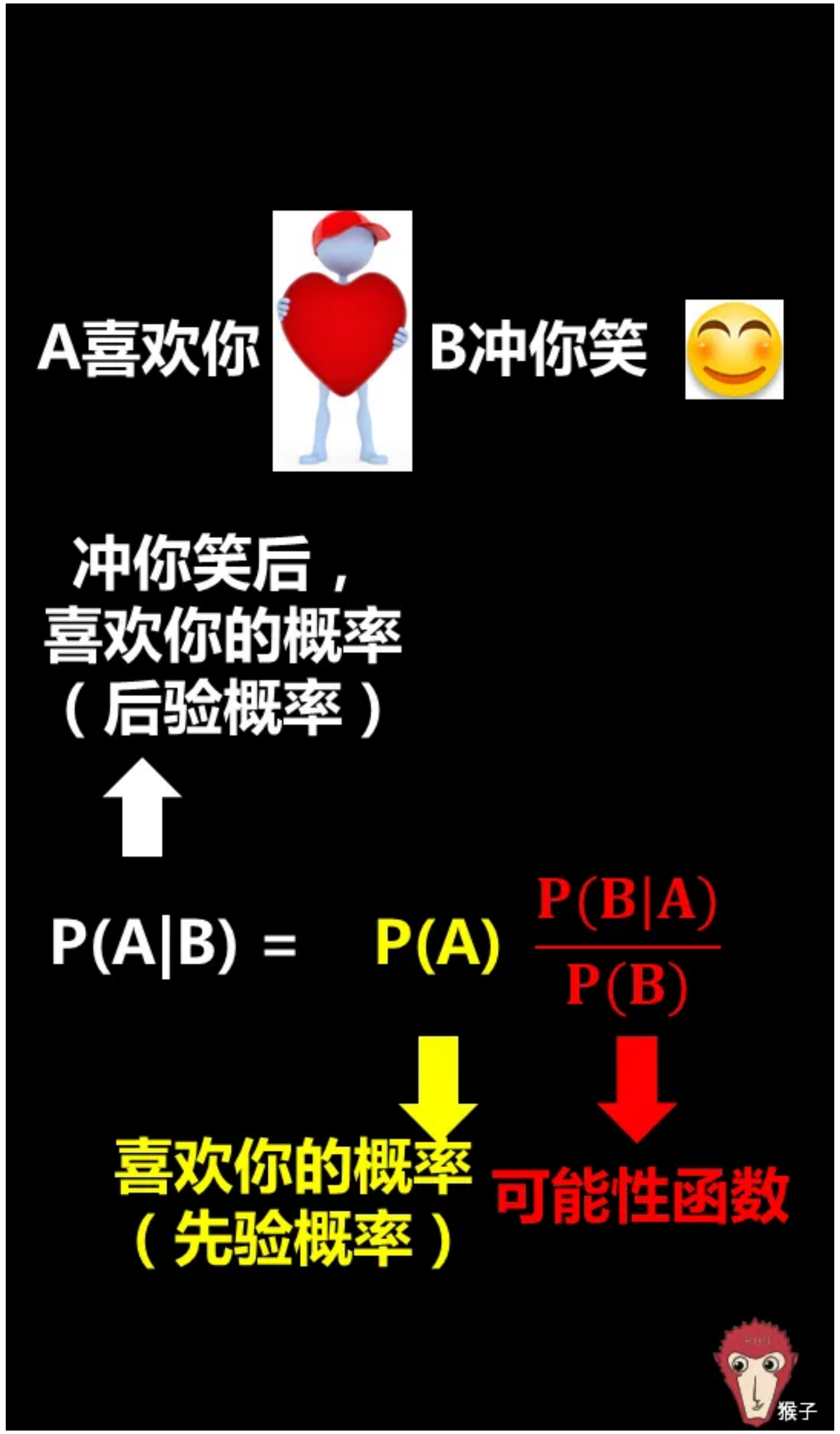
例：我的朋友小鹿说，他的女神每次看到他的时候都冲他笑，他现在想知道女神是不是喜欢他呢?

  - **问题**：女神是否喜欢你，记为A事件
  - **已知**：女神经常冲你笑，记为B事件
  - **求解**：P(A|B)，表示女神经常冲你笑这个事件(B)发生后，女神喜欢你（A）的概率



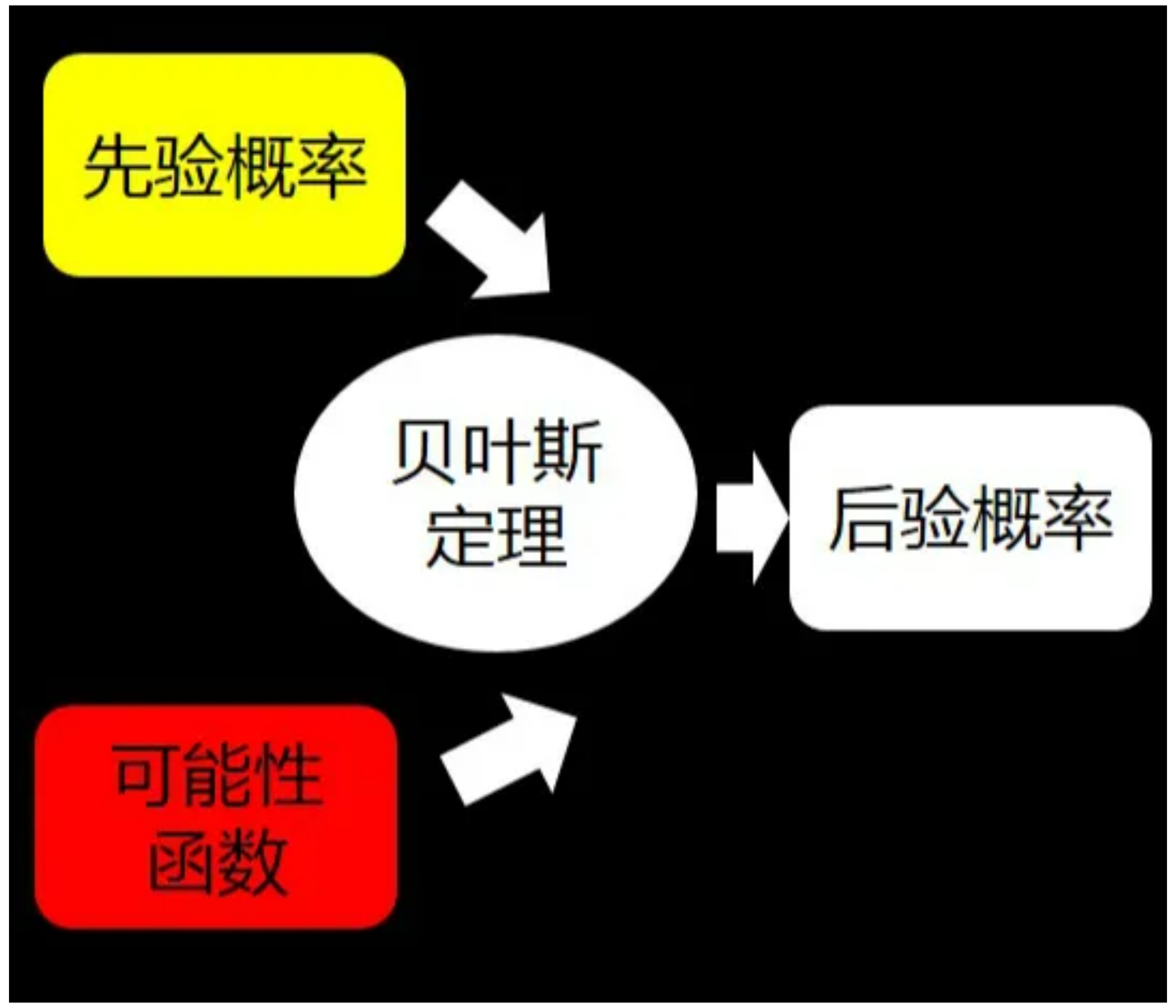
2.1 引言

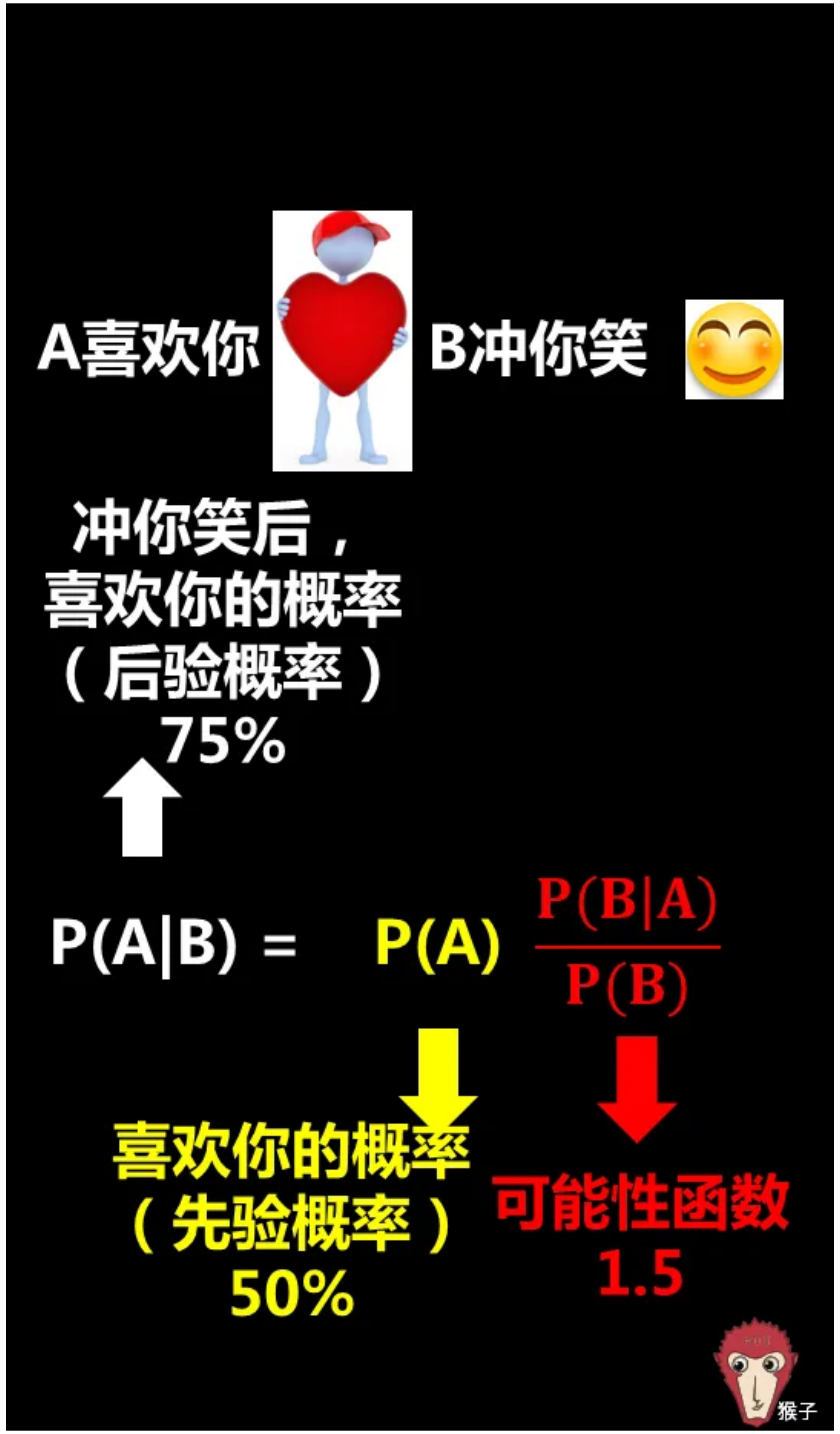
- **先验概率 P(A)**
  - 在不知道B事件的前提下，我们对A事件概率的一个主观判断。
  - 假设P(A)=50%，也就是不喜欢你，可能不喜欢你的概率都是一半
- **可能性函数 P(B|A)/P(B)**
  - 调整因子，新信息B带来的调整，作用是将先验概率（之前的主观判断）调整到更接近真实概率
  - P(B|A)/P(B)=1.5：走访调研女神的闺蜜，发现女神平日比较高冷，很少对人笑，也就是对你有好感的可能性比较大（可能性函数>1）



2.1 引言

- **后验概率 P(A|B)**
  - 在B事件发生之后，我们对A事件概率的重新评估。这个例子里就是在女神冲你笑后，对女神喜欢你的概率重新预测
  - $P(A|B) = P(A)* P(B|A)/P(B)=50\% *1.5=75\%$





2.1 引言

给一个硬币 x，没有任何信息，猜是5角还是1角。

- 设：$P(\omega_1)$为1角的概率，$P(\omega_2)$为5角的概率；$P(\omega_i)$为先验概率 (priori probability)，没有对样本进行任何观测情况下的概率。

- **错误率**：在所有可能出现的样本上类别决策错误的概率

$$P(\text{error}) = 1 - P(\omega_1) = P(\omega_2) \quad (\text{如果决策 } x \in \omega_1 \text{ })$$

- **决策规则**（最小错误率准则）：

$$x \begin{cases} \in \omega_1 & P(\omega_1) \geq P(\omega_2) \\ \in \omega_2 & P(\omega_1) < P(\omega_2) \end{cases}$$

2.1 引言

给一个硬币，已知重量 x，猜是5角还是1角。

- 设：$P(\omega_i|x)$为已知硬币重量为x的情况下硬币属于各类的概率，即后验概率 (posterior probability)。
- 根据贝叶斯公式求解$P(\omega_i|x)$：

$$P(\omega_i|x) = \frac{P(\omega_i, x)}{P(x)} = \frac{P(x|\omega_i)P(\omega_i)}{P(x)}$$

- **决策规则**：

$$x \begin{cases} \in \omega_1 & P\left(\omega_1 \mid x\right) \geq P\left(\omega_2 \mid x\right) \\ \in \omega_2 & P\left(\omega_1 \mid x\right) < P\left(\omega_2 \mid x\right) \end{cases}$$

- **错误率**：

$$P(\text{error} \mid x) = 1 - P(\omega_1 \mid x) = P(\omega_2 \mid x) \quad (\text{如果决策}x \in \omega_1)$$

2.2 最小错误率贝叶斯决策

- **定义**

最小错误率贝叶斯决策从最小错误率出发，利用概率论中的贝叶斯公式，得出使错误率最小的分类决策。

$$\min P(\text{error}) = \min \int P(\text{error} \mid x) P(x) dx \Rightarrow \min P(\text{error} \mid x) \quad \forall x$$

2.2 最小错误率贝叶斯决策

- **二分类最小错误率贝叶斯决策**

$$\boxed{P(\omega_i|x) = \frac{P(\omega_i, x)}{P(x)} = \frac{P(x|\omega_i)P(\omega_i)}{P(x)}}$$

以二分类为例，对应的**最小错误率贝叶斯决策**如下：

$$\text{如果} P(\omega_1 \mid x) > P(\omega_2 \mid x)\text{，则} x \in \omega_1\text{；反之，} x \in \omega_2$$

简记作：

$$\text{如果} P(\omega_1 \mid x) \lessgtr P(\omega_2 \mid x)\text{，则 } x \begin{cases} \in \omega_1 \\ \in \omega_2 \end{cases}$$

等价为：

$$\text{若\textbf{似然比} } l(x) = \frac{p\left(x \mid \omega_1\right)}{p\left(x \mid \omega_2\right)} \lessgtr \lambda = \frac{p(\omega_2)}{p(\omega_1)}, \quad \text{则 } x \begin{cases} \in \omega_1 \\ \in \omega_2 \end{cases}$$

2.2 最小错误率贝叶斯决策

癌细胞识别问题：ω_1正常细胞，ω_2癌细胞。某地区，经大量统计获先验概率$P(\omega_1)$，$P(\omega_2)$。若给定该地区某人细胞 x ，判断该细胞种类。假定特征 x 表示细胞核总的光密度，维数为1。

- **利用先验概率判断**

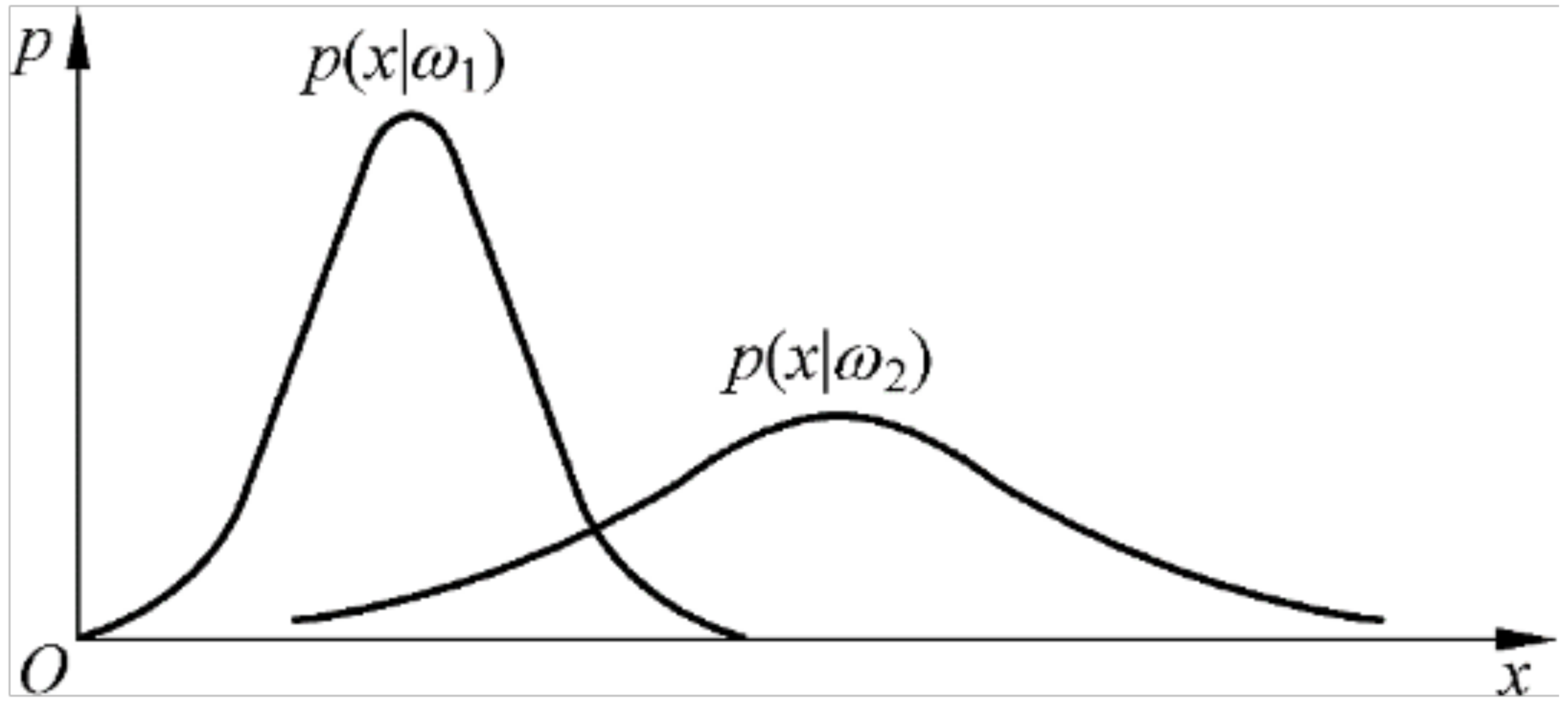
  如果$P(\omega_1) > P(\omega_2)$，$x \in \omega_1$；反之，$x \in \omega_2$。

  - 无意义分类器决策：先验概率$P(\omega_i)$只能提供整体上两类细胞出现比例的估计，不能用于对个体的判断。

2.2 最小错误率贝叶斯决策

- **先计算后验概率$P(\omega_1 \mid x)$，$P(\omega_2 \mid x)$，然后再判断类别**
    - 已知：先验概率$P(\omega_1)$，$P(\omega_2)$
    - 计算类条件概率密度$P(x \mid \omega_i)$：细胞ω_i光密度值的概率密度，可以根据之前收集的细胞图像数据计算获得

$$\boxed{P(\omega_i | x) = \frac{P(\omega_i, x)}{P(x)} = \frac{P(x | \omega_i)P(\omega_i)}{P(x)}}$$



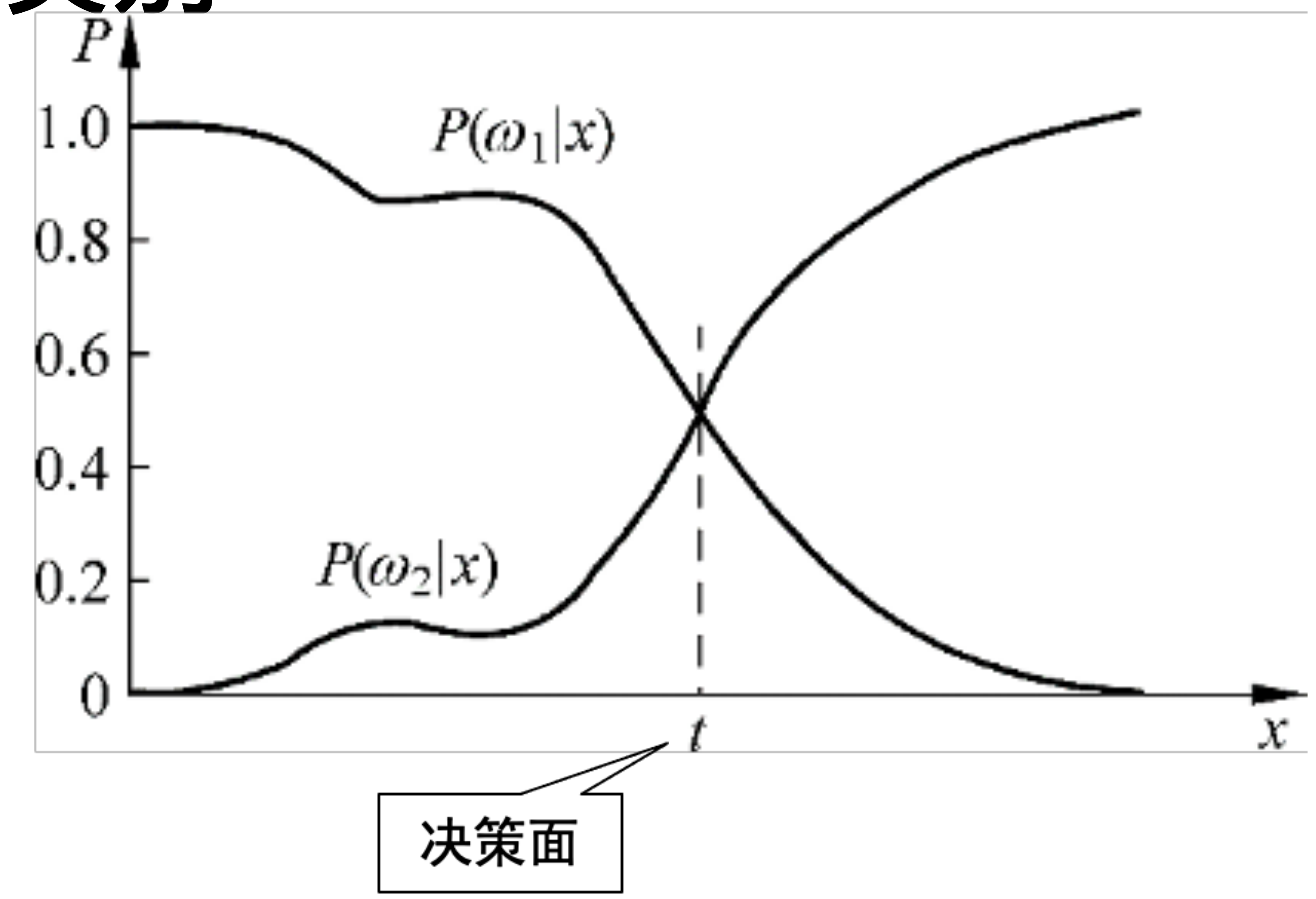
2.2 最小错误率贝叶斯决策

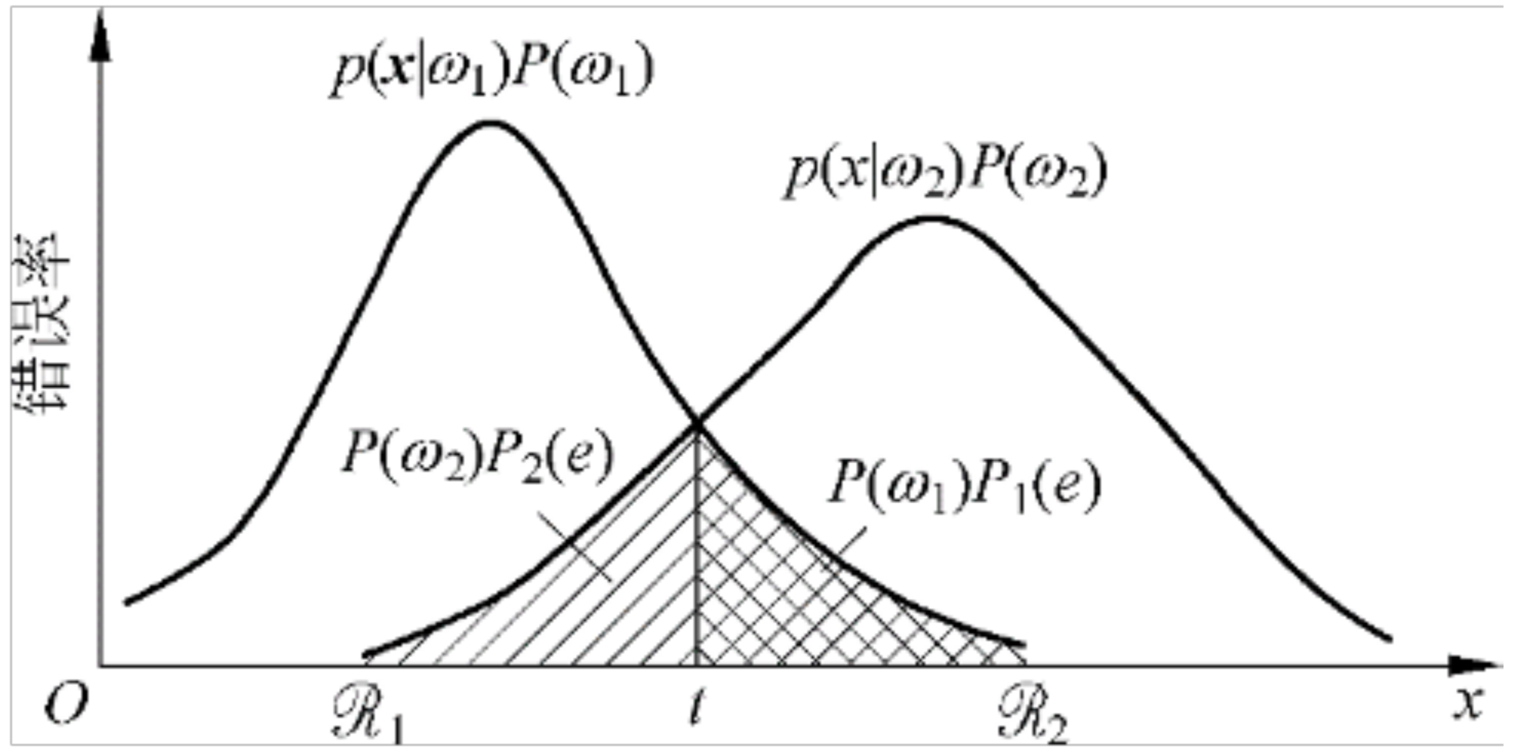
- 先计算后验概率$P(\omega_1 \mid x)$，$P(\omega_2 \mid x)$，然后再判断类别
  - 计算后验概率$P(\omega_1 \mid x)$，$P(\omega_2 \mid x)$

$$P(\omega_i \mid x) = \frac{P(x\omega_i)}{P(x)} = \frac{P(x \mid \omega_i)P\left(\omega_i\right)}{\sum_{j=1}^{2} P(x \mid \omega_j)P(\omega_j)}$$

  - 决策规则：选择后验概率大的一类

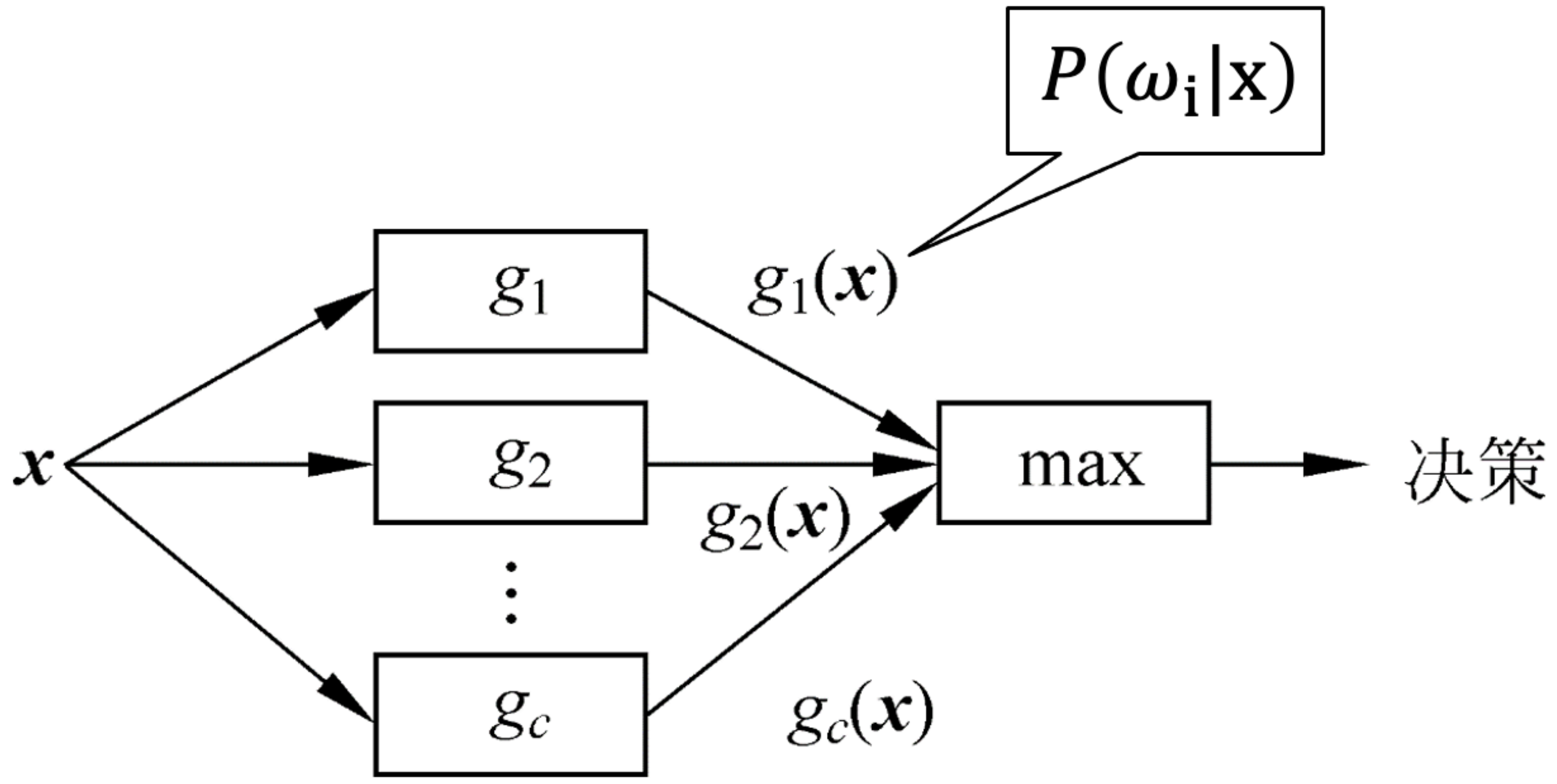
$$\begin{cases} P(\omega_1 \mid x) > P(\omega_2 \mid x), & x \in \omega_1 \\ P(\omega_1 \mid x) < P(\omega_2 \mid x), & x \in \omega_2 \end{cases}$$





2.2 最小错误率贝叶斯决策

- **多分类最小错误率贝叶斯决策**



设类别数为 c，则对应的**最小错误率贝叶斯决策**如下：

$$P(\omega_i|x)=\max_{j=1,\cdots,c} P(\omega_j|x) \Rightarrow x \in \omega_i$$

由于其贝叶斯公式分母相同，只需要比较分子，因此，可以等价于：

$$P(x|\omega_i)P(\omega_i)=\max_{j=1,\cdots,c} P\left(x|\omega_j\right)P(\omega_j) \Rightarrow x \in \omega_i$$

2.3 最小风险贝叶斯决策

- **决策原则**

  考虑各种错误决策造成损失不同时的一种最优决策。

- **风险 / 期望损失**

$$R(\alpha_i \mid x) = \sum_{j=1}^{c} \lambda(\alpha_i, \omega_j) P(\omega_j \mid x), \quad i = 1, \cdots, k$$

- $\lambda(\alpha_i, \omega_j)$：损失函数，对状态为$\omega_j$的向量 x 采取策略 α_i 所带来的损失；由具体应用确定
- x：样本，d 维特征向量
- ω_j：状态，共有 c 个可能的状态，构成状态空间
- α_i：决策，共有 k 个决策，组成决策空间

2.3 最小风险贝叶斯决策

- **最小风险**

$$R(\alpha_i \mid x) = \begin{cases} \lambda(\alpha_1, \omega_2) \cdot P(\omega_2 \mid x) & (\text{如果决策 } x \in \omega_1) \\ \lambda(\alpha_2, \omega_1) \cdot P(\omega_1 \mid x) & (\text{如果决策 } x \in \omega_2) \end{cases} \qquad \min R(\alpha) = \min \int R(\alpha(x) \mid x) P(x) dx$$

- **最小风险贝叶斯决策**
  - 考虑各种错误造成损失不同时的一种最优决策：

$$\alpha = \underset{i=1,\cdots,k}{\operatorname{argmin}} \ R(\alpha_i \mid x)$$

  - 决策表不同会导致不同的决策结果，因此，需要人为合理地确定决策表

2.4 Neyman-Pearson决策

状态 \ 决策	阳性	阴性
阳性	真阳性 (TP)	假阳性 (FP)
阴性	假阴性 (FN)	真阴性 (TN)

- **两类错误率及评价指标**
  - 两类：阳性 (positive)、阴性 (negative)
  - **第一类错误率** (Type-I error rate)：假阳性/误报/虚警率 $\alpha = FP/(FP+TN)$
    - 假阳性样本占总阴性样本的比例
  - **第二类错误率** (Type-II error rate)：假阴性/漏报率 $\beta = FN/(FN+TP)$
    - 假阴性样本占总阳性样本的比例

2.4 Neyman-Pearson决策

状态 / 决策	阳性	阴性
阳性	真阳性 (TP)	假阳性 (FP)
阴性	假阴性 (FN)	真阴性 (TN)

- **两类错误率及评价指标**
  - **灵敏度** (sensitivity)：$S_n = TP/(TP+FN) = 1-\beta$
    - 在真正的阳性样本中有多少比例能被正确检测出来，或没有被误判
  - **特异度** (specificity)：$S_p = TN/(TN+FP) = 1-\alpha$
    - 在真正的阴性样本中有多少比例能被正确检测出来，或没有被误判

2.4 Neyman-Pearson决策

- **其他常用评价指标**
  - **正确率** (accuracy)：$\text{Acc} = \dfrac{TP + TN}{TP + TN + FP + FN}$
  - **召回率** (recall) ：$\text{Rec} = \dfrac{TP}{TP + FN}$
  - **正确率** (precision)：$\text{Pre} = \dfrac{TP}{TP + FP}$
  - **F度量** (F-measure) ：$\text{F} = \dfrac{2Rec \cdot Pre}{Rec + Pre}$

状态 / 决策	阳性	阴性
阳性	真阳性 (TP)	假阳性 (FP)
阴性	假阴性 (FN)	真阴性 (TN)

2.4 Neyman-Pearson决策

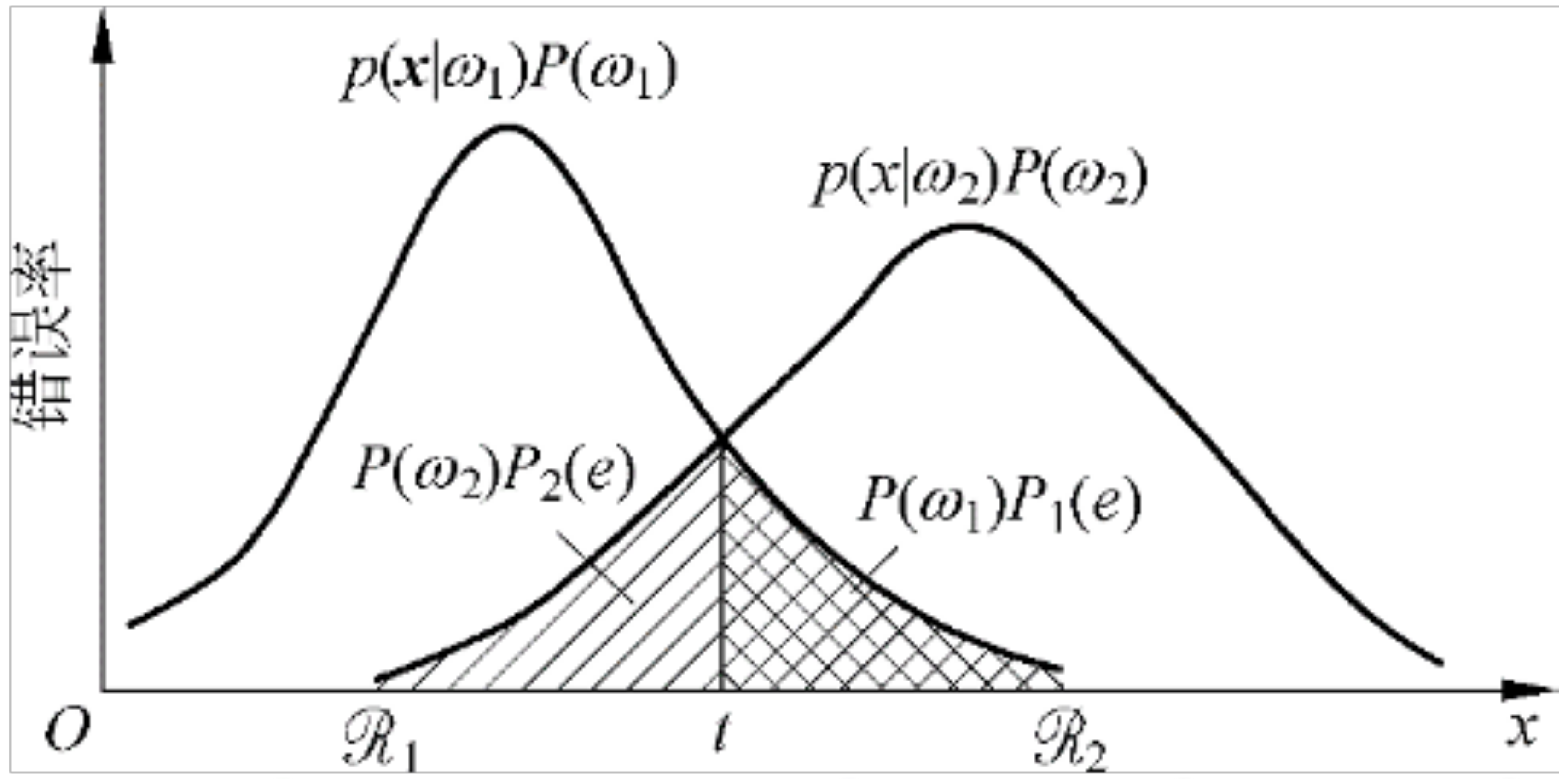
- **Neyman-Pearson决策**
  - **决策的引入**：保证某一类错误率不变的前提下，最小化另一类错误率。假设R_1，R_2分别为第一、二两类的决策域，ω_1类为阴性，ω_2类为阳性，t 为分界面，第一类错误率和第二类错误率分别表示为：

$$P_1(e) = \int_{R_2} p\left(x \mid \omega_1\right) \mathrm{d}x, \qquad P_2(e) = \int_{R_1} p\left(x \mid \omega_2\right) \mathrm{d}x$$

那么决策可以转化为如下约束最优化问题：

$$\min\ P_1(e)$$

$$\mathrm{s.t.}\ \mathrm{P}_2(e) - \varepsilon_0 = 0$$



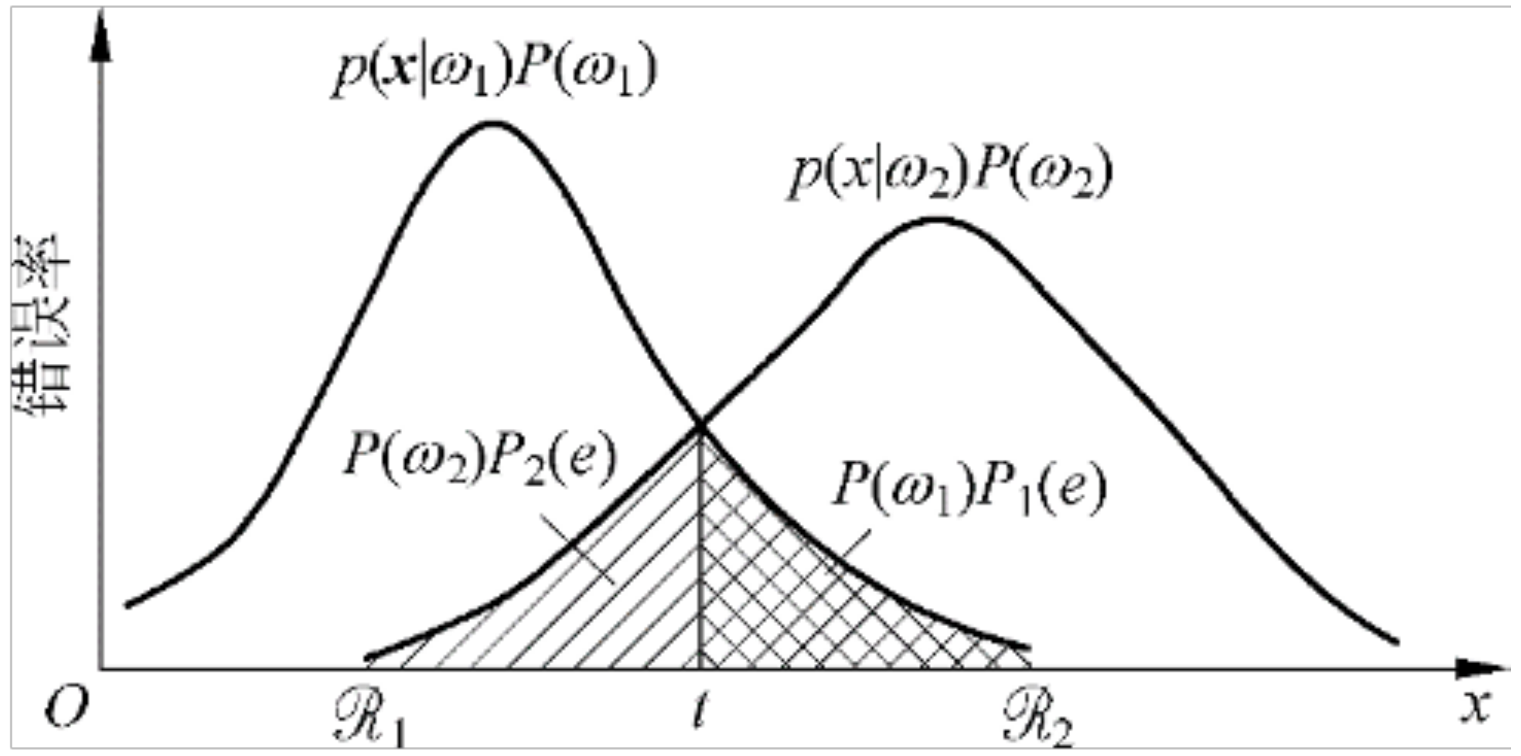
2.4 Neyman-Pearson决策

- **Neyman-Pearson决策**
  - **转化**：利用拉格朗日乘子法，将上述约束优化问题转化为无约束优化问题（拉格朗日函数），其中 λ 是拉格朗日乘子

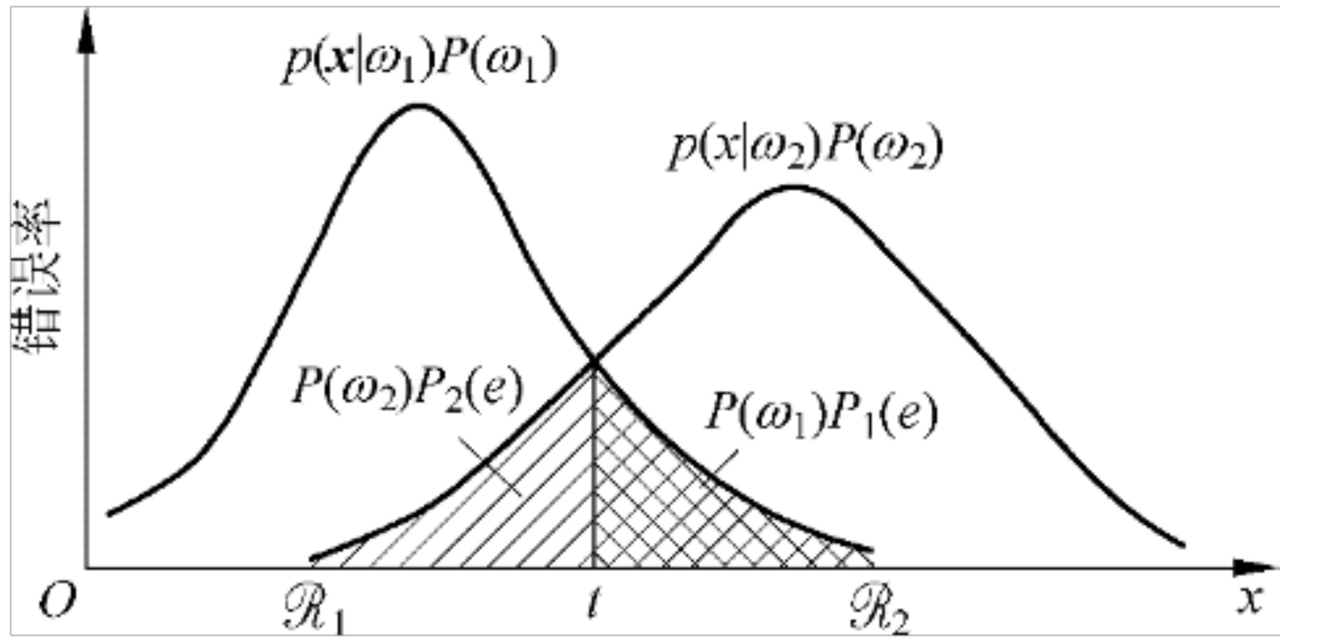
$$\min \quad \gamma = P_1(e) + \lambda\left(P_2(e) - \varepsilon_0\right)$$

$$\gamma(x,\lambda) = \int_{\mathrm{R}_2} p\left(x \mid \omega_1\right) \mathrm{d}x + \lambda\left[\int_{\mathrm{R}_1} p\left(x \mid \omega_2\right) \mathrm{d}x - \varepsilon_0\right]$$

$$= \left(1 - \lambda\varepsilon_0\right) + \int_{\mathrm{R}_1}\left[\lambda p\left(x \mid \omega_2\right) - p\left(x \mid \omega_1\right)\right] \mathrm{d}x$$



2.4 Neyman-Pearson决策



- **Neyman-Pearson决策**

  优化目标：求解使$\gamma(x,\lambda)$最小的决策边界 x = t，分别对 λ 和 x 分别求导

$$\frac{\partial \gamma}{\partial x} = \lambda p\left(x \mid \omega_2\right) - \rho\left(x \mid \omega_1\right), \qquad \frac{\partial \gamma}{\partial \lambda} = \int_{\mathrm{R}_1} p\left(x \mid \omega_2\right) \mathrm{d}x - \varepsilon_0$$

  在γ的极值处这两个导数都应该为0，因此，在决策边界上应满足：

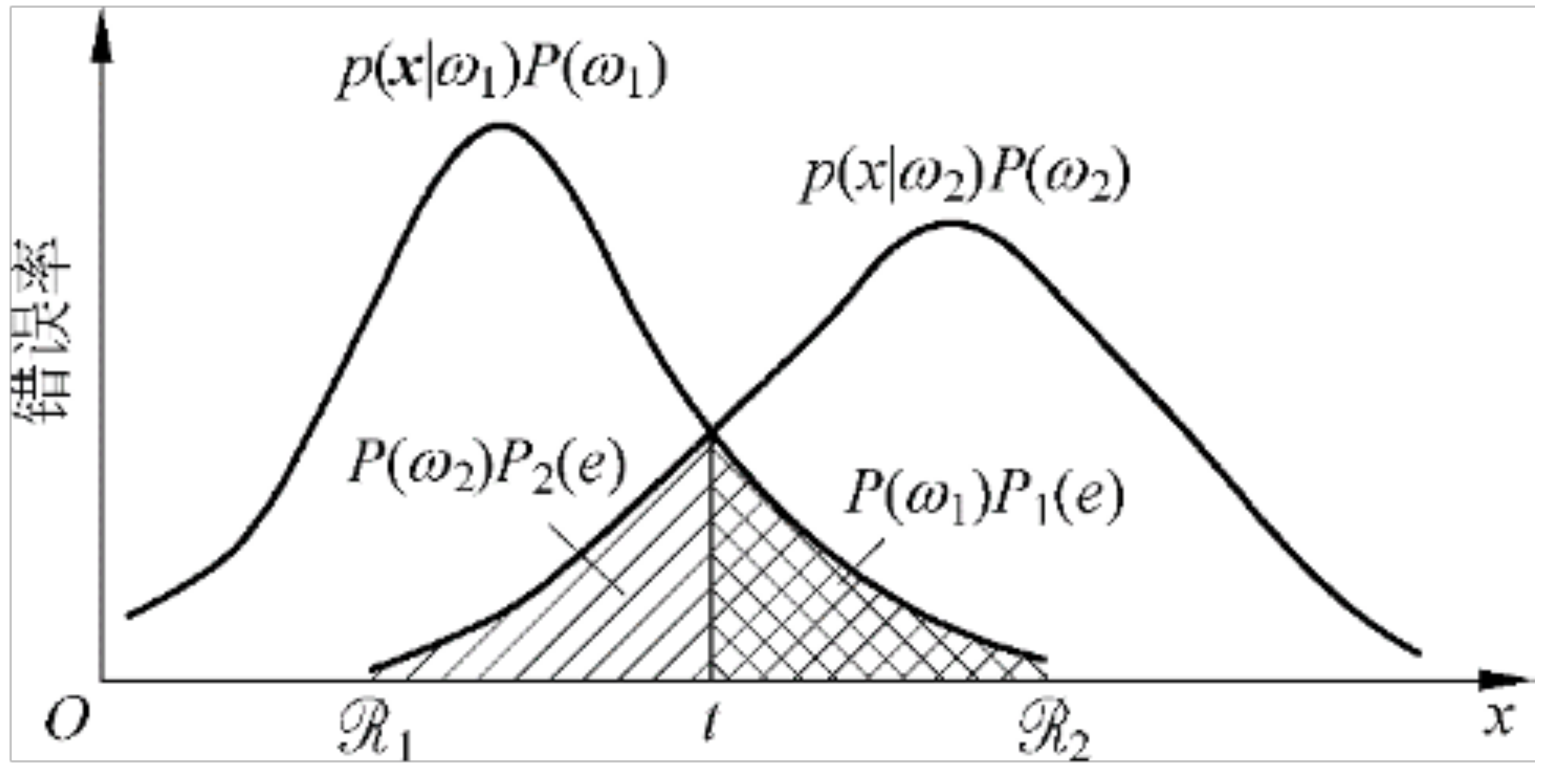
$$\lambda = \frac{p\left(x \mid \omega_1\right)}{p\left(x \mid \omega_2\right)}, \qquad \int_{\mathrm{R}_1} p\left(x \mid \omega_2\right) \mathrm{d}x = \varepsilon_0$$

  要使$\gamma(x,\lambda)$最小，应选择R_1使积分项内全为负值，因此，R_1应使得下式成立的 x 组成的区域：

$$\lambda p\left(x \mid \omega_2\right) - p\left(x \mid \omega_1\right) < 0$$

2.4 Neyman-Pearson决策

- **Neyman-Pearson决策**
  - **决策规则：**



$$l(x)=\frac{p\left(x \mid \omega_1\right)}{p\left(x \mid \omega_2\right)} \lessgtr \lambda,\text{ 则}x \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$$

利用数值方法，基于似然比$l(x)$和似然比密度函数$P\left(l \mid \omega_2\right)$求解 λ：

$$P_2(e)=1-\int_0^\lambda p\left(l \mid \omega_2\right) \mathrm{d} l=\varepsilon_0$$

- $p\left(l \mid \omega_2\right) \geq 0$，$P_2(e)$是 λ 的单调函数
- 采用试探法寻找一个合适的λ值，满足$P_2(e)=\varepsilon_0$且$\mathrm{P}_1(e)$尽可能小

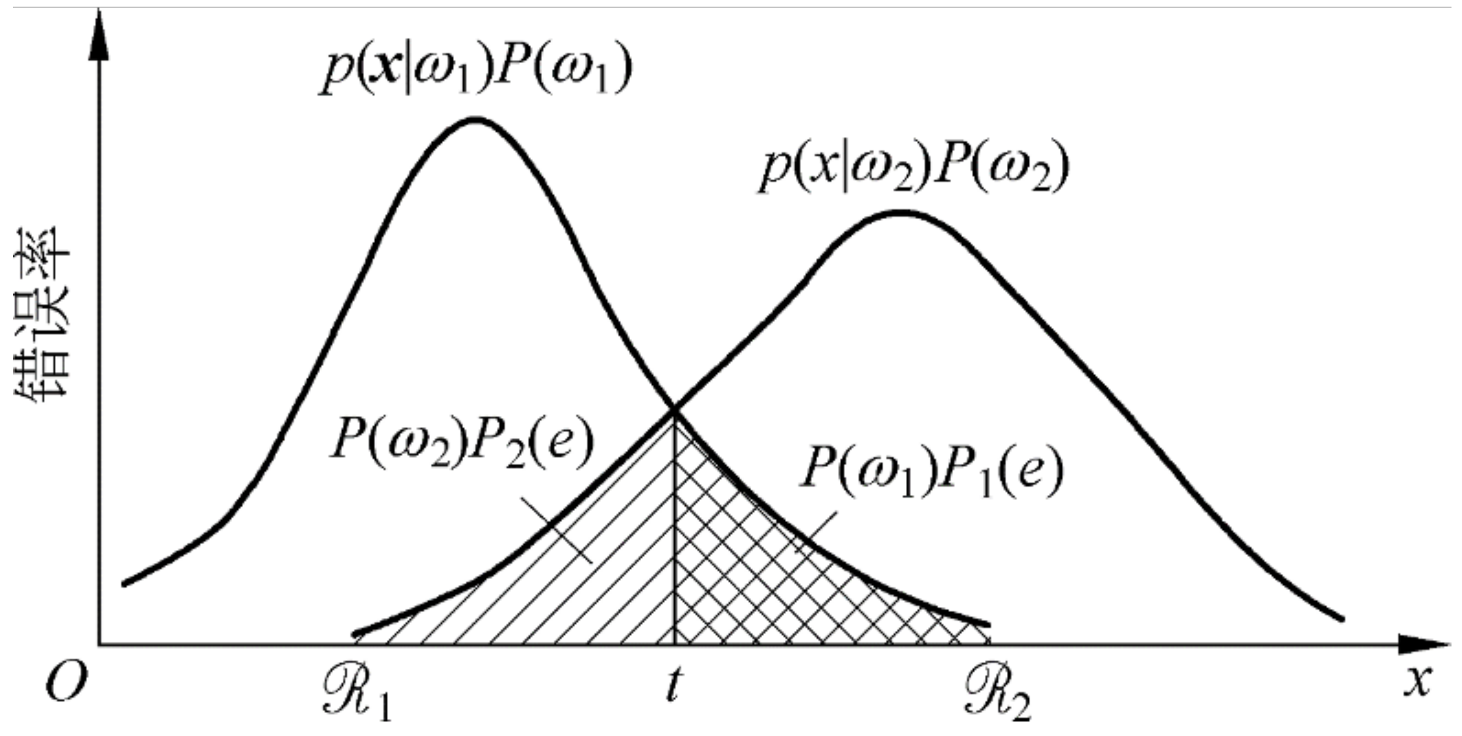
2.4 Neyman-Pearson决策

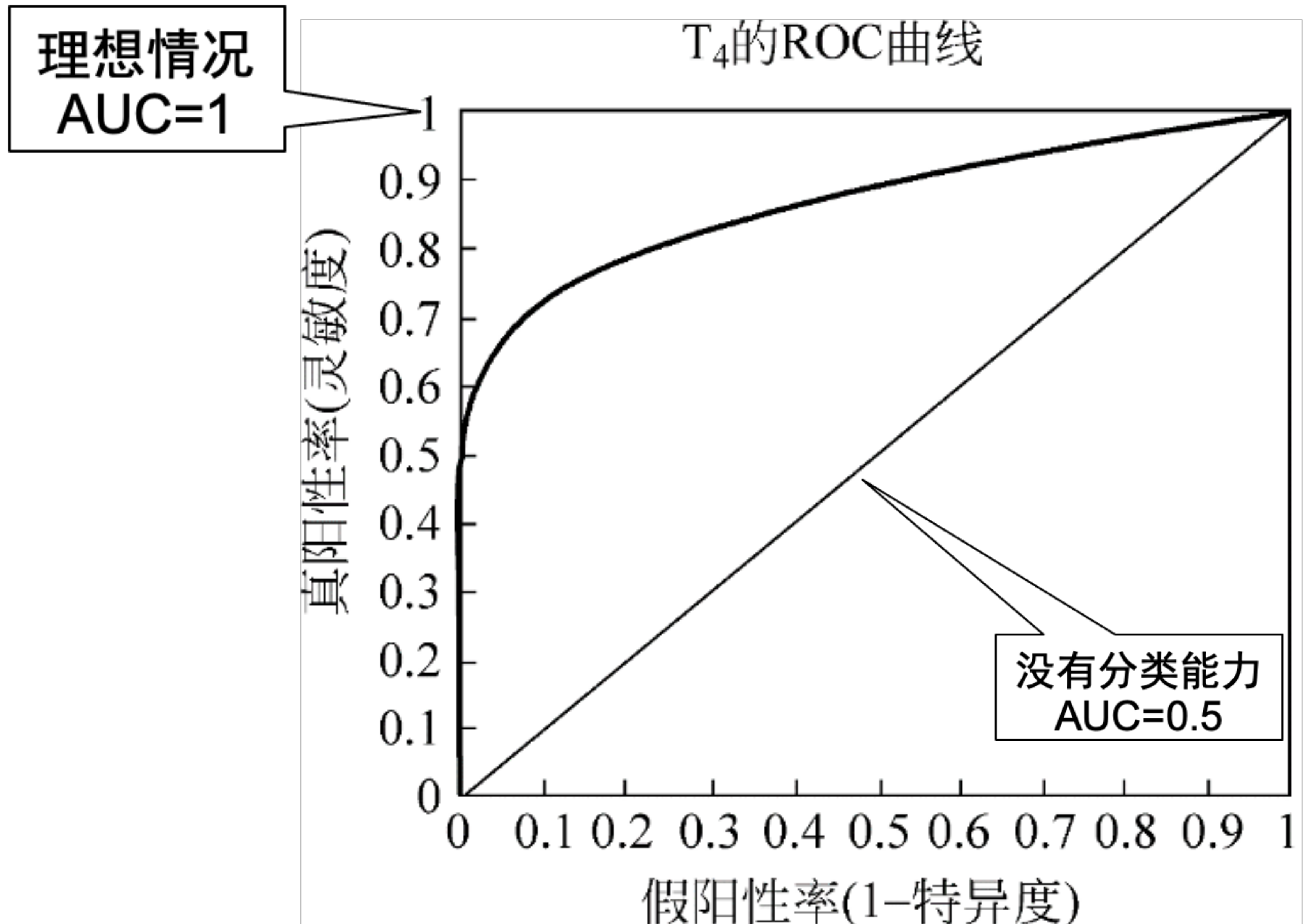
- 三种决策规则的比较
  - 本质：采用不同的决策阈值，可以达到不同的错误率。

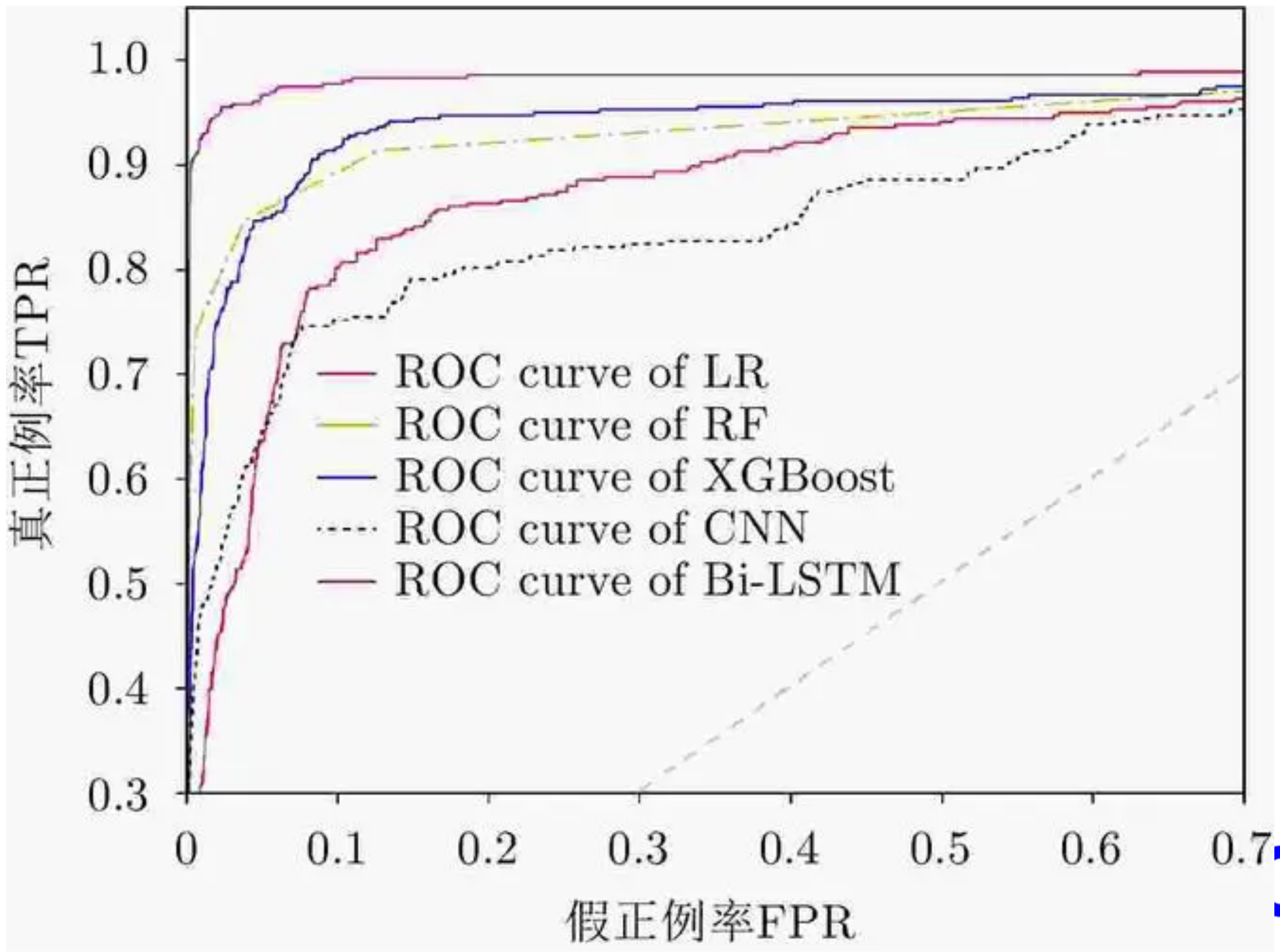
决策名称	含义
最小错误率决策	先验概率比$P(w_2)/P(w_1)$作阈值，达到总的错误率最小，即两类错误率加权之和最小
最小风险决策	阈值中考虑了对两类错误率不同的惩罚，实现风险最小
Neyman-Pearson决策	通过调整阈值，使一类的错误率为指定数值，而另一类的错误率求最小

2.4 Neyman-Pearson决策

- **ROC曲线**
  - ROC (Receiver Operating Characteristic) 曲线
    - 用于阈值选择和模型比较
  - AUC (Area Under ROC curves) 曲线下面积







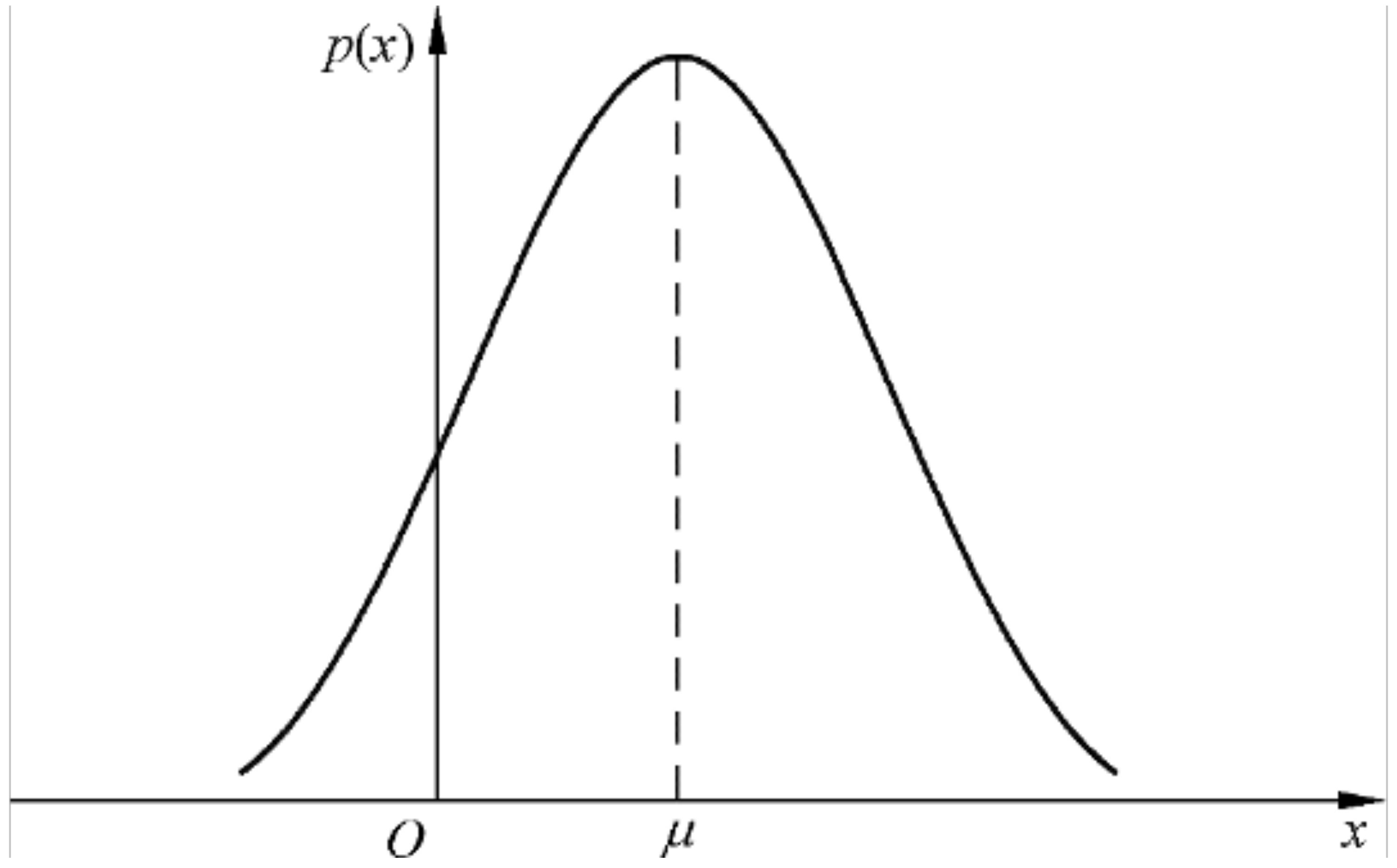
2.5 正态分布时的贝叶斯决策

2.5.1 正态分布及其性质回顾

- **单变量正态分布** $N\left(\mu, \sigma^2\right)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

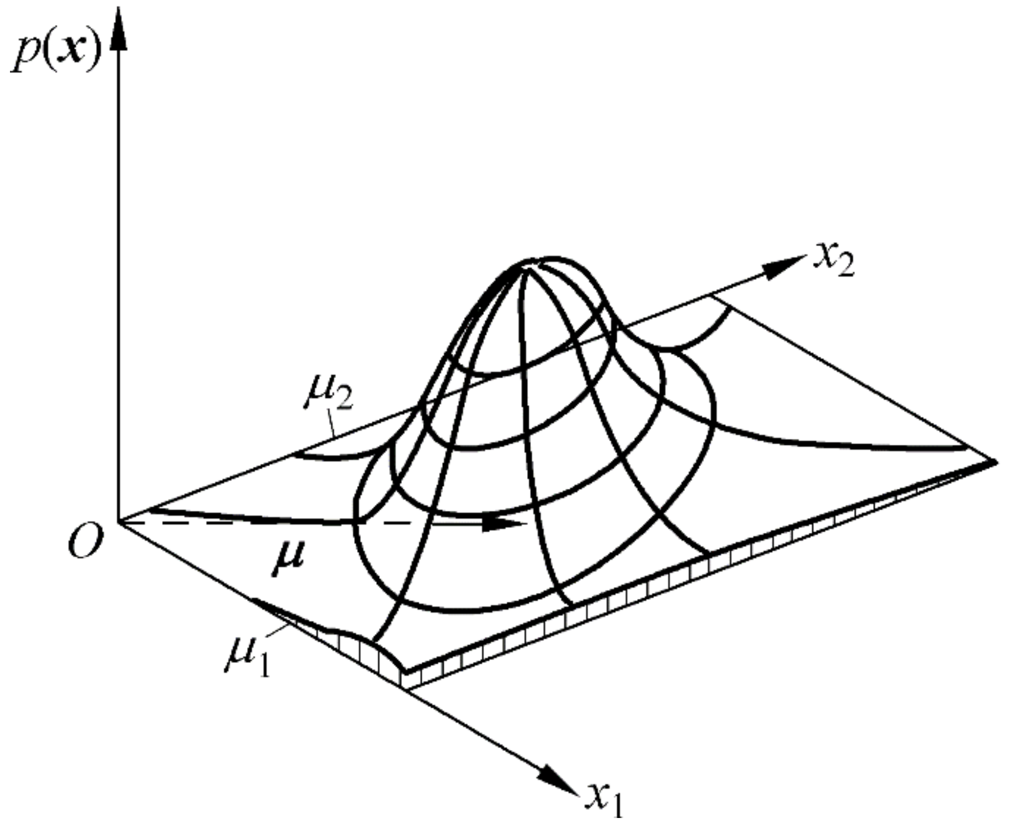
$$\mu = E\{x\} = \int_{-\infty}^{\infty} xp(x)dx,\ \ \sigma^2 = \int_{-\infty}^{\infty}(x-\mu)^2 p(x)dx$$

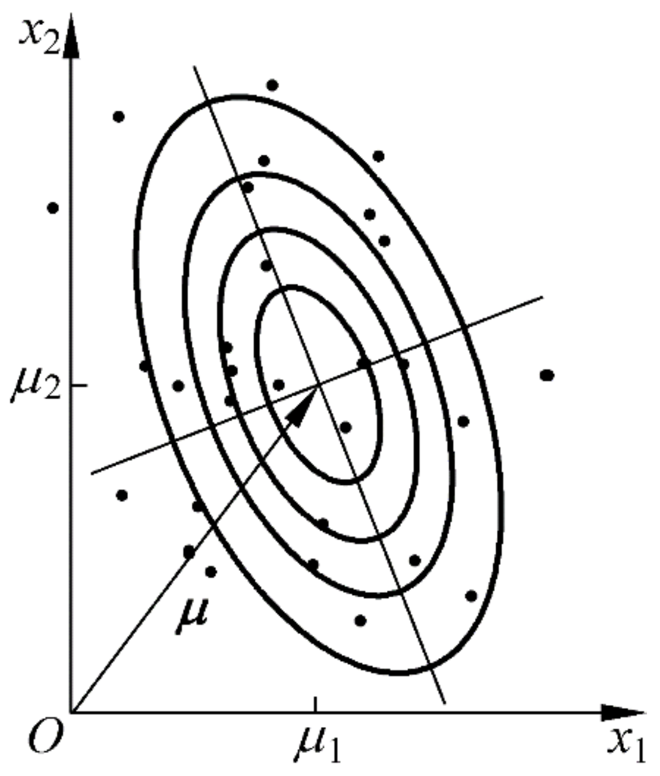


- **多元正态分布** $N\left(\mu, \Sigma\right)$

$$p(x) = \frac{1}{(2\pi)^{d/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(x-\mu\right)^T\Sigma^{-1}\left(x-\mu\right)\right\}$$

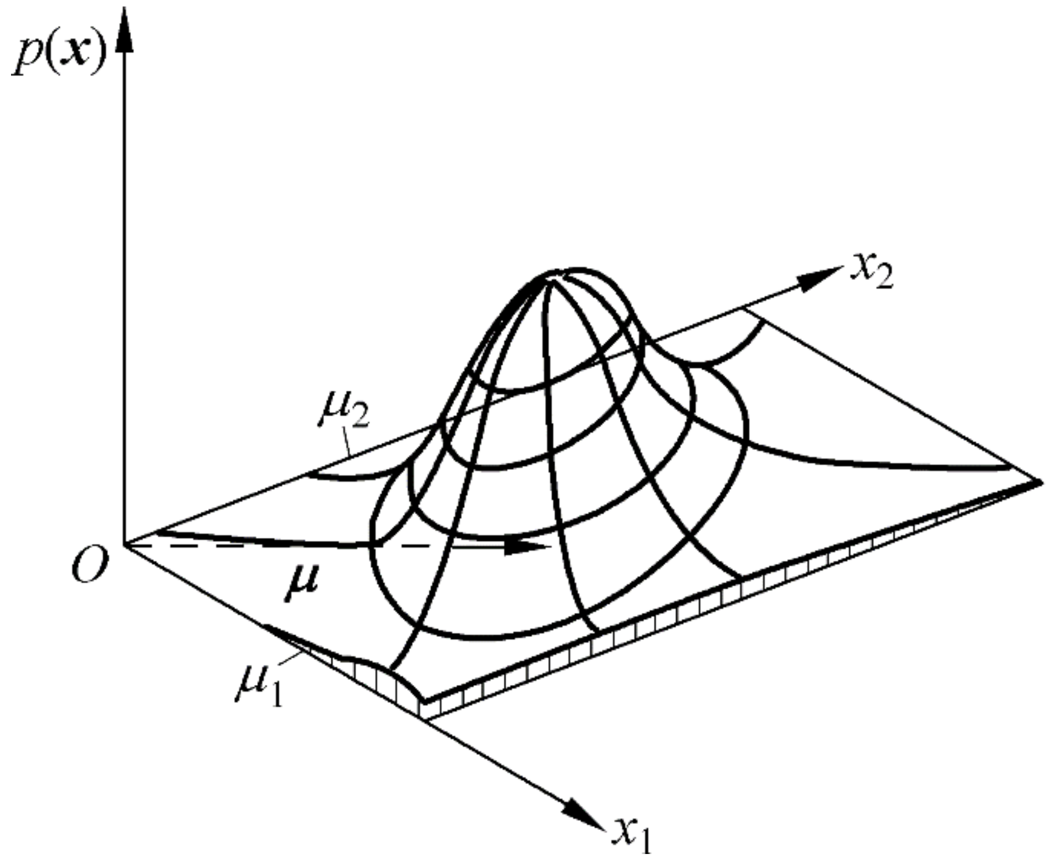
$$\mu = E\{x\},\ \ \Sigma = E\left\{\left(x-\mu\right)\left(x-\mu\right)^T\right\}$$

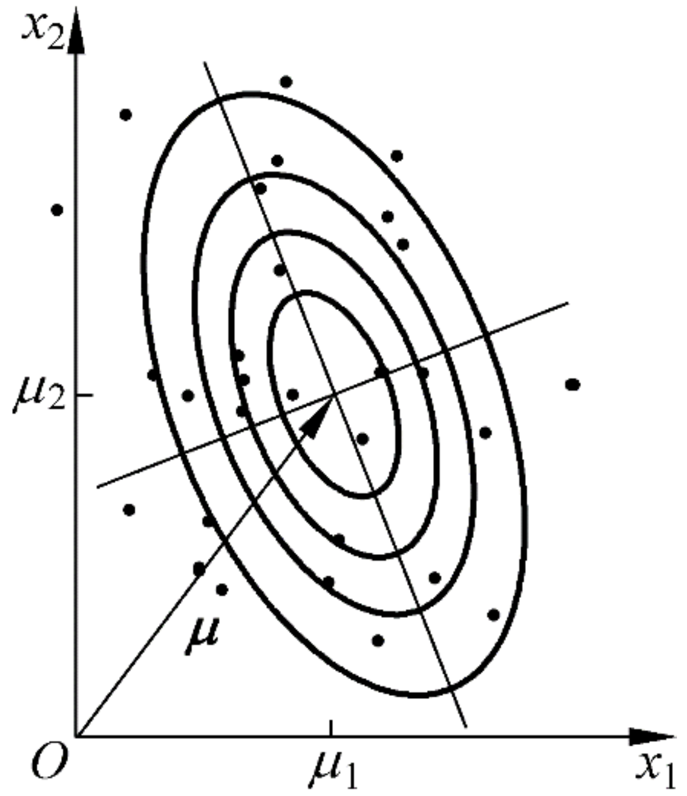




2.5.1 正态分布及其性质回顾

- 多元正态分布由均值μ和协方差Σ完全确定
- 等密度点形成超椭球面
- 不相关性等价于独立性
  - 若分量x_i和分量x_j不相关 → x_i和x_j相互独立
  - 若协方差矩阵Σ为对角阵 → 各分量相互独立





2.5.1 正态分布及其性质回顾

- 多元正态分布的边缘分布是正态分布 $p(x_i) \sim N(\mu_i, \sigma_{ii}^2)$
- 多元正态分布的条件分布是正态分布
- 多元正态随机向量的线性变换仍为多元正态分布

$$\begin{cases} p(x) \sim N(\mu, \Sigma) \\ y = Ax \end{cases} \Rightarrow \quad p(y) \sim N\left(A\mu, A\Sigma A^T\right)$$

- 多元正态随机向量的线性组合为一维正态随机变量

2.5.2 正态分布下的贝叶斯决策

$$P(\omega_i|x) = \frac{P(\omega_i, x)}{P(x)} = \frac{P(x|\omega_i)P(\omega_i)}{P(x)}$$

- **判别函数**
  - **概率密度函数** $p\left(x \mid \omega_i\right) \sim N\left(\mu_i, \Sigma_i\right)$

$$p\left(x \mid \omega_i\right) = \frac{1}{(2\pi)^{d/2} \mid \Sigma_i \mid^{1/2}} \exp\left\{-\frac{1}{2}\left(x-\mu_i\right)^T \Sigma_i^{-1}\left(x-\mu_i\right)\right\}$$

  - **判别函数**：只关注分子，取对数形式

$$g_i(x) = \ln\left(p\left(x|\omega_i\right)p\left(\omega_i\right)\right) = -\frac{1}{2}\left(x-\mu_i\right)^T \Sigma_i^{-1}\left(x-\mu_i\right) + \ln\frac{1}{(2\pi)^{\frac{d}{2}}\left|\Sigma_i\right|^{\frac{1}{2}}} + \ln p\left(\omega_i\right)$$

- **决策面方程**

$$g_i(x) = g_j(x) \Longrightarrow -\frac{1}{2}\left[\left(x-\mu_i\right)^T \Sigma_i^{-1}(x-\mu_i) - \left(x-\mu_j\right)^T \Sigma_j^{-1}\left(x-\mu_j\right)\right] - \frac{1}{2}\ln\frac{\left|\Sigma_i\right|}{\left|\Sigma_j\right|} + \ln\frac{p(\omega_i)}{p(\omega_j)} = 0$$

特殊情况下的正态分布的贝叶斯决策

- **各类协方差矩阵相等、且各特征独立、方差σ^2相等：** $\Sigma_i = \sigma^2\mathbf{I}$
  - 各类先验概率不相等 $p(\omega_i) \neq p(\omega_j)$，则有

$$g_i(x) \triangleq -\frac{1}{2\sigma^2}\left(x-\mu_i\right)^T(x-\mu_i) + \ln p\left(\omega_i\right) = -\frac{1}{2\sigma^2}\parallel x-\mu_i \parallel^2 + \ln p\left(\omega_i\right)$$

$$g_i(x) = \frac{\mu_i^T}{\sigma^2}x - \frac{1}{2\sigma^2}\mu_\text{i}^T\mu_i + \ln p\left(\omega_i\right)$$

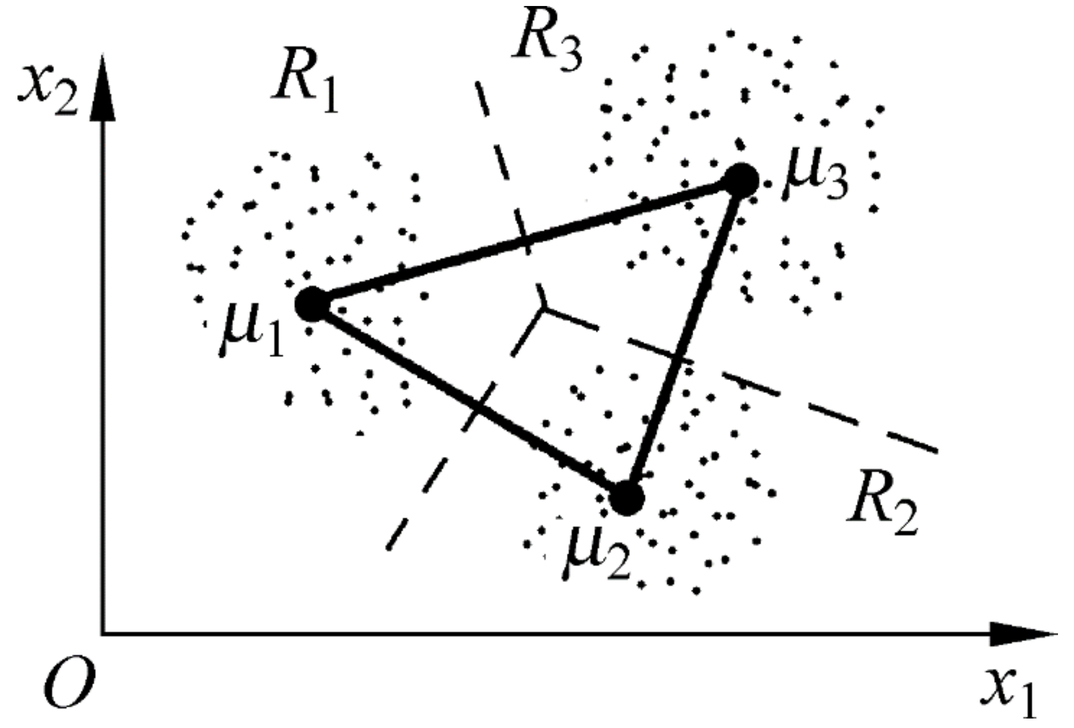
  - 球状分布（各类样本落入以μ_i为中心的同样大小的超球体内），分类只取决于样本到各类中心的距离

特殊情况下的正态分布的贝叶斯决策

- **各类协方差矩阵相等、且各特征独立、方差σ^2相等：** $\Sigma_i = \sigma^2 \mathbf{I}$
  - 各类先验概率相等 $p(\omega_i) = p(\omega_j),\ i, j = 1,2,\cdots c$，则有

$$g_i(x) \triangleq -\frac{1}{2\sigma^2}\left(x-\mu_i\right)^T(x-\mu_i) = -\frac{1}{2\sigma^2}\ \| x-\mu_i \|^2$$

  - 决策面与先验概率相等时的决策面平行，只是向先验概率小的方向偏移
    - 先验概率大的一类要占据更大的决策空间



特殊情况下的正态分布的贝叶斯决策

- **各类协方差矩阵都相等** $\Sigma_i = \Sigma$
  - 各类样本集中于以该类均值 μ_i 为中心的同样大小和形状的超椭球内

$$g_i(x) \triangleq \ln\left(p\left(x\,|\,\omega_i\right)p\left(\omega_i\right)\right) = -\frac{1}{2}\underbrace{\left(x-\mu_i\right)^T\Sigma^{-1}\left(x-\mu_i\right)}_{\text{Mahalanobis 距离（马氏距离的平方）,记为}\gamma^2} + \ln p\left(\omega_i\right)$$

  - 若各类的先验概率相等，决策面通过μ_i和μ_j连线中点
  - 若各类的先验概率不相等，决策面向先验概率小的均值点偏移

2.6 错误率

- 错误率反映了分类问题固有复杂性的程度
- 在分类器设计出来后，通常是以错误率大小来衡量其性能优劣
- 通常是以错误率大小作为比较方案的标准

$$P(e)=P(\omega_1)\int P(x|\omega_1)dx+P(\omega_2)\int P(x|\omega_2)dx=P(\omega_1)P_1(e)+P(\omega_2)P_2(e)$$

- 实际中，按理论公式计算错误率很困难