概率论与信息论

概率论为不确定性建模及处理提供了强有力的工具。概率论广泛用于机器学习中，尤其是分类模型的构建及分析。

1.1 概率论基础

1.1.1 随机试验、随机事件

自然界与社会生活中存在两类现象，其中：结果确定的现象为确定性现象；而结果不确定的现象为不确定性现象。而概率统计的研究对象就是随机现象的数量规律。

对随机现象的观察、记录、试验统称为**随机试验**。随机试验的特性表现为：可以在相同条件下重复进行；事先知道可能出现的所有结果；试验前并不知道哪种结果会发生。

随机试验E的所有可能结果构成一个集合，记为E的**样本空间**$\Omega = \{e\}$,其中e为样本空间的**基本事件**或样本点。

对于样本空间Ω的一个子集A，记为随机试验E的**随机事件**A；当且仅当A所包含的一个基本事件发生，称事件A发生。对于事件A，若每次试验时，事件A均发生，则称A为**必然事件**；为便于描述，记∅为**不可能事件**，不可能事件中不含任何基本事件。

对于随机试验E下的两个随机事件A、B，若事件A发生一定导致事件B发生，则记为A⊂B；事件A=B，当且仅当A⊂B并且B⊂A。

随机事件A、B的**和事件**，记为$A \cup B = \{x|x \in A或x \in B\}$，它表示A、B至少有一个事件发生。

随机事件A、B的**积事件**，记为$A \cap B = \{x|x \in A并且x \in B\}$，它表示A、B同时发生，也可表示为：AB，或者$A \cdot B$。

对于随机试验E下的随机事件$A_1, A_2, \dots, A_n$：至少有一个事件发生，记为$\cup_{i=1}^n A_i$；随机事件$A_1, A_2, \dots, A_n$同时发生，记为$\cap_{i=1}^n A_i$。

对于随机事件A、B，若$AB = \emptyset$，则称A、B**互不相容**，或**互斥**。

随机事件A的逆事件记为$\overline{A}$。显然：$A\overline{A} = \emptyset$，并且$A \cup \overline{A} = \Omega$。

1.1.2 概率空间与概率

我们所提及的概率通常是指随机事件发生的概率。例如：明天下午下雨的概率。为此，首先，以三元组$(\Omega, \mathcal{F}, P)$的形式给出**概率空间(probability space)**的定义，其中：

- Ω为样本空间(或结果空间)，该空间包含随机试验 E 所有可能的结果；
- $\mathcal{F}$为事件空间，某随机事件A所有输出结果$\subseteq \Omega$；
- P为概率测度，使得$\mathcal{F} \rightarrow \Re$，并满足：(I)非负性 $\mathrm{P(A)} \geq 0$；(II)$\mathrm{P}(\Omega) = 1$；(III)若随机事件 A、B 互斥，则 $\mathrm{P(A \cup B) = P(A) + P(B)}$。

随机事件发生的概率具有如下性质：

- 对于随机事件 A、B，如果$\mathrm{A} \subseteq \mathrm{B}$(即：事件 A 蕴含事件 B)，则有 $\mathrm{P(A)} \leqslant \mathrm{P(B)}$
- $\mathrm{P(A \cap B) \leq \min(P(A), P(B))}$
- $\mathrm{P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)}$
- $\mathrm{P(\overline{A}) = P(\Omega / A) = 1 - P(A)}$
- 如果事件$\mathrm{A_1, A_2, \ldots, A_n}$彼此互斥，并且$\bigcup_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{A_i} = \Omega$，则有$\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{P(A_i)} = 1$

1.1.3 条件概率

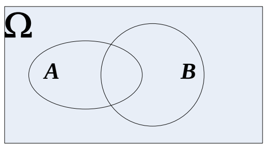

图1. 样本空间Ω、随机事件A及B

对于事件A和B，在观察到事件B后，A发生的概率记为**条件概率$\mathbf{P(A|B)}$**。

由图1可以设定$\mathrm{P(A|B)} = x$. 由于$\mathrm{P(B):P(AB)} = 1{:}\,x$, 所以

$$x = \mathrm{P(A|B)} = \frac{\mathrm{P(AB)}}{\mathrm{P(B)}} \tag{1.1}$$

即：$\mathrm{P(A|B)} = \frac{\mathrm{P(AB)}}{\mathrm{P(B)}}$。

显然，条件概率具有概率的所有性质，例如：

- $\mathrm{P(A|B) = 1 - P(\overline{A}|B)}$
- $\mathrm{P(A \cup C|B) = P(A|B) + P(C|B) - P(AC|B)}$

- 若随机事件$A \subseteq C$，则有$P(A|B) \leq P(C|B)$

若$P(B|A)P(A)=P(A)P(B)$，则称随机事件A和B**相互独立**。

若如下条件概率均有意义，则有：

- $P(AB) = P(B|A)P(A) = P(A|B)P(B)$
- $P(ABC) = P(AB)P(C|AB) = P(A)P(B|A)P(C|AB)$
- $P(A_1A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1})$

1.1.4 全概率公式与贝叶斯公式

设$\boldsymbol{\Omega}$为随机试验E的样本空间，$B_1, B_2, \ldots, B_n$为试验E的一组事件。若：

(I) $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$;

(II) $\forall i \neq j$, 总有$B_i \cap B_j = \emptyset$

则称$B_1, B_2, \ldots, B_n$为Ω的一个划分，或称之为一个完备事件组。

显然，"事件$B_1, B_2, \ldots, B_n$两两同时发生"为不可能事件，"$B_1, B_2, \ldots, B_n$至少有一个事件发生"为必然事件。

对于随机试验 E 的一个随机事件 A, 及其样本空间Ω的一个划分$B_1, B_2, \ldots, B_n$，其中$P(B_i) > 0$, $i = 1,2, \ldots, n$, 有**全概率公式**：

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \qquad (1.2)$$

由于$A = A\Omega = \cup_{i=1}^{n} AB_i$，并且对于$\forall i \neq j$，$AB_i$与 AB_j互斥，所以：

$$P(A) = P(\cup_{i=1}^{n} AB_i) = \sum_{i=1}^{n} P(AB_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

进一步，得条件概率：

$$P(B_j|A) = \frac{P(AB_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \qquad (1.3)$$

记 $P(B_j|A) = \dfrac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$为**贝叶斯公式**。

例：贝叶斯公式的理解。已知框中含有 N 个苹果和 M 个梨子，苹果为黄色的概率为 20%，梨子为黄色的概率为 80%，问：黄色水果是梨子的概率为多少。

1.1.5 随机事件的独立性

对于随机事件 A、B，并且$P(A) > 0$，$P(B) > 0$，若$P(A|B) = P(A)$或$P(B|A) = P(B)$，即：$P(AB) = P(A)P(B)$, 则称**事件 A、B 相互独立**。

显然，A、B 相互独立$\Leftrightarrow \overline{A}$、B 相互独立$\Leftrightarrow$ A、$\overline{B}$相互独立$\Leftrightarrow \overline{A}$、$\overline{B}$相互独立。

给定 n 个随机事件$A_1, A_2, \ldots, A_n$，若对于$2 \le k \le n$, 总有$P(A_{i_1}A_{i_{12}} \cdots A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$, 则称 n 个随机事件$A_1, A_2, \ldots, A_n$**相互独立**。

1.2 随机变量及分布

1.2.1 随机变量

随机变量在概率论中扮演重要角色。**随机变量**不是变量，而是一种将样本空间的结果映射至实值的**函数**。由于该函数可以随机取不同值，因此称其为随机变量，常以大写字母表示，如随机变量X。一个随机变量只是对可能的状态进行描述，需要借助概率分布来指定每个状态的可能性。

以掷骰子为例。若以X表达不同投掷结果的随机变量，其中随机变量$X = 1$，表示将事件“投掷 1 点”映射到实值 1。当然，也可以指定其它映射，比如$X = 1$表示将“投掷点数为奇数”映射到 1；而$X = 2$表示将“投掷点数为偶数”映射到 2。

若随机变量取值为有限个可能状态、或者可数的无限多状态，则称其为**离散随机变量**；若随机变量取值是连续的，称其为**连续随机变量**。

例：离散随机变量X取某个值a的概率记为$P(X = a)$或$P_X(a)$；连续随机变量X取值不超出实数 b 的概率记为$P(X \le b)$。

1.2.2 随机变量的累积分布函数

随机变量X的**累积分布函数(Cumulative Distribution Function ,CDF)** 为$F_X: \mathbb{R} \to [0,1]$，并且$F_X(x) = P(X \le x)$. 利用累积分布函数，可以计算任何事件的概率。$F_X(x)$具有如下性质：

(1) $0 \le F_X(x) \le 1$

(2) $\lim_{x \to -\infty} F_X(x) = 0$

(3) $\lim_{x \to \infty} F_X(x) = 1$

(4) 若$x_1 \le x_2$，则$F_X(x_1) \le F_X(x_2)$

1.2.3 离散随机变量及其分布

离散随机变量的概率分布以**概率质量函数(Probability Mass Function, PMF)**描述，每个随机变量都有不同的概率质量函数，该函数将随机变量每个可能的状态映射至随机变量取得该状态的概率，以P 表示概率质量函数。

设离散随机变量X的各种可能状态取值构成集合$\mathrm{Val}(X)$，则X为某个取值$x \in \mathrm{Val}(X)$的概率记为：$\mathrm{P}(X=x)=\mathrm{P}_X(x)$。概率$\mathrm{P}_X(x)$必须满足如下条件：(1) $0 \le \mathrm{P}_X(x) \le 1$，以及(2) $\sum_{x\in \mathrm{Val}(X)} \mathrm{P}_X(x)=1$。

概率质量函数也可同时作用于多个随机变量，此时多个随机变量的概率分布为联合概率分布。例如，随机变量$X=x$和随机变量$Y=y$同时发生的概率为$\mathrm{P}(X=x,Y=y)=\mathrm{P}_{X,\mathrm{Y}}(x,y)$。

例：数字图像的归一化灰度直方图。

1.2.4 连续随机变量及其分布

以**概率密度函数**(Probability Density Function, PDF)描述连续随机变量的概率分布。若连续随机变量X的概率密度函数为f_X，则其必满足如下条件：

(I)函数f_X的定义域需包含随机变量X所有可能的取值，记为$\mathrm{Val}(X)$；

(II)对于$\forall x \in \mathrm{Val}(X)$，总有$f_X(x) \ge 0$;

(III)$\int_{x\in \mathrm{Val}(X),} f_X(x)dx=1.$

相应地，连续随机变量X的累积分布函数$\mathrm{F_X}(x)$为：

$$\mathrm{F_X}(x) = \mathrm{P}(X \le x)=\int_{-\infty}^{x} f_X(t)dt \tag{1.4}$$

显然，分布函数具有如下特性：

(I)对于任意$\forall x_1 \le x_2$，总有$\mathrm{F_X}(x_1) \le \mathrm{F_X}(x_2)$

(II)$\mathrm{P}(x_1 \le X \le x_2)=\mathrm{F_X}(x_2)-\mathrm{F_X}(x_1)=\int_{x_1}^{x_2} f_X(t)dt$

(III)在$f_X(x)$的连续点x处，$\mathrm{F_X}'(x) = f_X(x)$

1.2.5 数学期望

离散随机变量X的数学期望：

$$\mu=\mathrm{E}[X]=\sum_{x\in \mathrm{Val}(X)} x\mathrm{P}(X=x)=\sum_{x\in \mathrm{Val}(X)} x\mathrm{P}_X(x) \tag{1.5}$$

离散随机变量X的函数$f(X)$的数学期望：

$$\mathrm{E}[f(X)]=\sum_{x\in \mathrm{Val}(X)} f(x)\mathrm{P}_X(x) \tag{1.6}$$

连续随机变量X的数学期望：

$$\mu=\mathrm{E}[X]=\int_{\mathrm{Val}(X)} x f_X(x)dx \tag{1.7}$$

连续随机变量X的函数$g(X)$的数学期望：

$$\mathrm{E}[g(X)]=\int_{\mathrm{Val}(X)} g(x) f_X(x)dx \tag{1.8}$$

期望具有如下性质：

(1) 对于任意常数a, $\mathrm{E}[a]=a$；

(2) 对于两个任意常数a_1, a_2以及 $g_1(X), g_2(X)$,总有

$$\mathrm{E}[a_1 g_1(X)+a_2 g_2(X)]=a_1\mathrm{E}[g_1(X)]+a_2\mathrm{E}[g_2(X)] \tag{1.9}$$

1.2.6 方差

离散随机变量X的方差：

$$\sigma^2=\mathrm{Var}[X]=\mathrm{E}[(X-\mu)^2]=\sum_{x\in\mathrm{Val}(X)}(x-\mu)^2\mathrm{P}_X(x) \tag{1.10}$$

连续随机变量X的方差：

$$\sigma^2=\mathrm{Var}[X]=\mathrm{E}[(X-\mu)^2]=\int_{\mathrm{Val}(X)}(x-\mu)^2 f_X(x)dx \tag{1.11}$$

随机变量X的方差满足如下关系：

$$\mathrm{Var}[X]=\mathrm{E}[X^2]-(\mathrm{E}[X])^2 \tag{1.12}$$

随机变量的方差为概率质量函数(或概率密度函数)的转动惯量，具有非负性，只有当全部概率质量(或概率密度值)集中于一点时，方差取值才为 0。其中，σ为随机变量X的标准差，标准差是随机变量关于均值偏离程度的一种度量。

方差具有如下性质：

(1) 对于任意常数a, $\mathrm{Var}[a]=0$；

(2) 对于两个任意常数a_1, a_2以及 $f_1(X), f_2(X)$,总有

$$\mathrm{Var}[a_1 f_1(X)+a_2 f_2(X)]=a_1{}^2\mathrm{Var}[f_1(X)]+a_2{}^2\mathrm{Var}[f_2(X)] \tag{1.13}$$

显然，对于任意常数a,b，总有 $\mathrm{Var}[aX+\mathrm{b}]=a^2\mathrm{Var}[X]$。

1.3 随机变量的典型分布

伯努利分布

二项分布

泊松分布

均匀分布

正态分布

分布名称	参数	分布律或概率密度	期望	方差
伯努利分布(两点分布, 0-1分布)	$0<p<1$	$P\{X=k\}=p^k(1-p)^{1-k},\quad k=0,1$	p	p(1-p)
二项分布 $\sim B(n,p)$	$n\geq 1, 0<p<1$	$P\{X=k\}=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\cdots,n$	np	$np(1-p)$
负二项分布 $\sim NB(n,p)$	$r\geq 1, 0<p<1$	$P\{X=k\}=\binom{k-1}{r-1}p^r(1-p)^{k-r},\quad k=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
泊松分布 $\sim \pi(\lambda)$或$\sim P(\lambda)$	$\lambda>0$	$P\{X=k\}=e^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,\cdots$	λ	λ
均匀分布 $\sim U(a,b)$	$a<b$	$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b \\ 0, & 其他\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
正态分布(高斯分布) $\sim N(\mu,\sigma^2)$	μ $\sigma>0$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
标准正态分布 $\sim N(0,1)$	$\mu=0$ $\sigma=1$	$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	0	1
对数正态分布 $\sim LN(\mu,\sigma^2)$	μ $\sigma>0$	$f(x)=\begin{cases}\frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, & x>0 \\ 0, & 其他\end{cases}$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
指数分布 $\sim e(\mu,\lambda)$	$\lambda>0$	$f(x)=\begin{cases}\lambda e^{-\lambda(x-\mu)}, & x\geq\mu \\ 0, & 其他\end{cases}$	$\mu+\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
几何分布 $\sim G(p)$	$0<p<1$	$P\{X=k\}=p(1-p)^{k-1},\quad k=1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
超几何分布 $\sim H(n,M,N)$	N,M,n $(n\leq M)$	$P\{X=k\}=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,\cdots,n$	$\frac{nM}{N}$	$\frac{nM}{N}(1-\frac{M}{N})(\frac{N-n}{N-1})$
χ^2分布 $\sim\chi^2(n)$或$\sim\chi^2_n$	$n\geq 1$	$f(x)=\begin{cases}\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}, & x>0 \\ 0, & 其他\end{cases}$	n	$2n$
Γ分布 $\sim\Gamma(\alpha,\beta)$	$\alpha>0$ $\beta>0$	$f(x)=\begin{cases}\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}, & x>0 \\ 0, & 其他\end{cases}$	$\alpha\beta$	$\alpha\beta^2$
瑞利分布 $\sim R(\sigma)$	$\sigma>0$	$f(x)=\begin{cases}\frac{x}{\sigma^2}e^{-x^2/(2\sigma^2)}, & x>0 \\ 0, & 其他\end{cases}$	$\sqrt{\frac{\pi}{2}}\sigma$	$\frac{4-\pi}{2}\sigma^2$
β分布 $\sim\beta(\alpha,\beta)$	$\alpha>0$ $\beta>0$	$f(x)=\begin{cases}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & 0<x<1 \\ 0, & 其他\end{cases}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
柯西分布 $\sim C(\lambda,\alpha)$	α $\lambda>0$	$f(x)=\frac{1}{\pi}\frac{1}{\lambda^2+(x-\alpha)^2}$	不存在	不存在
t分布 $\sim t(n)$	$n\geq 1$	$f(x)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$	0	$\frac{n}{n-2}(n>2)$
F分布 $\sim F(n_1,n_2)$	n_1,n_2	$f(x)=\begin{cases}\frac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)}\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}x\right)^{(n_1+n_2)/2}\cdot\left(1+\frac{n_1}{n_2}x\right)^{-(n_1+n_2)/2}, & x>0 \\ 0, & 其他\end{cases}$	$\frac{n_2}{n_2-2}$ $(n_2>2)$	$\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ $(n_2>4)$

Γ函数：$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$　　β函数：$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$

2.1 二维随机变量

2.1.1 联合累积分布和边缘累积分布

累积分布函数同样适合两个随机变量X和Y.

定义随机变量X和Y的**联合累积分布函数**：$F_{XY}(x,y)=P(X\leq x,Y\leq y)$.

利用联合累积分布函数，可以计算涉及X和Y的任何概率。

进一步，利用联合累积分布函数，可以定义关于随机变量X或Y的**边缘累积分布函数**，分别定义如下：

$$F_X(x)=P(X\leq x)=P(X\leq x,Y<+\infty)=F_{XY}(x,+\infty)=\lim_{y\to+\infty}F_{XY}(x,y)\qquad(1.14)$$

$$F_Y(y)=P(Y\leq y)=P(X<+\infty,Y\leq y)=F_{XY}(+\infty,y)=\lim_{x\to+\infty}F_{XY}(x,y)\qquad(1.15)$$

$F_{XY}(x,y)$具有如下性质：

(1) $0\leq F_{XY}(x,y)\leq 1$

(2) $\lim\limits_{x,y\to-\infty} \mathrm{F_{XY}}(\mathrm{x},\mathrm{y}) = 0$

(3) $\lim\limits_{x,y\to\infty} \mathrm{F_{XY}}(\mathrm{x},\mathrm{y}) = 1$

(4) $\mathrm{F_X}(\mathrm{x}) = \lim\limits_{y\to+\infty} \mathrm{F_{XY}}(\mathrm{x},\mathrm{y})$, $\mathrm{F_Y}(\mathrm{y}) = \lim\limits_{x\to+\infty} \mathrm{F_{XY}}(\mathrm{x},\mathrm{y})$

(5) 对 于 任 意 $x_1 \le x_2, y_1 \le y_2$，总有$\mathrm{F_{XY}}(x_1, y) \le \mathrm{F_{XY}}(x_2, y)$ ，$\mathrm{F_{XY}}(x, y_1) \le \mathrm{F_{XY}}(x, y_1)$

2.1.2 联合概率质量函数和边缘概率质量函数

离散随机变量X和Y的联合概率分布以**联合概率质量函数**表示，定义为$\mathrm{P_{XY}}$：$\mathbb{R} \times \mathbb{R} \to [0,1]$，并且$\mathrm{P_{XY}}(x, y) = \mathrm{P}(X = x, Y = y)$。$\mathrm{P_{XY}}(x, y)$表示 $X = x$和$Y = y$同时取值的概率。

联合概率$\mathrm{P_{XY}}(x, y)$必须满足如下条件：(1) $0 \le \mathrm{P_{XY}}(x, y) \le 1$，以及(2) $\sum_{(\mathrm{x},y)\in\mathrm{Val}(X,Y)} \mathrm{P_{XY}}(x, y) = 1$。

进一步，利用X和Y联合概率质量函数可定义随机变量X或Y的**边缘概率质量函数**$\mathrm{P_X}(x)$或$\mathrm{P_Y}(y)$：

$$\mathrm{P_X}(x) = \sum_{y\in\mathrm{Val}(Y)} \mathrm{P_{XY}}(x, y) \tag{1.16}$$

$$\mathrm{P_Y}(y) = \sum_{x\in\mathrm{Val}(X)} \mathrm{P_{XY}}(x, y) \tag{1.17}$$

2.1.3 联合密度函数和边缘概率密度函数

以**联合概率密度函数**描述连续随机变量X和Y的联合概率分布。若连续随机变量X和Y的联合概率密度函数为f_{XY}，则其必满足如下条件：

(I)函数f_{XY}的定义域需包含随机变量X和Y所有可能的取值，记为$\mathrm{Val}(X, \mathrm{Y})$；

(II)对于$\forall(x, y) \in \mathrm{Val}(X, Y)$，总有$f_{\mathrm{XY}}(x, y) \ge 0$;

(III)$\iint_{(x,y)\in\mathrm{Val}(X,Y)} f_{\mathrm{XY}}(x, y)dxdy = 1.$

类似于一维连续随机变量，有：$\mathrm{P}\{(x, y) \in \mathrm{A}\} = \iint_{(x,y)\in\mathrm{A}} f_{\mathrm{XY}}(x, y)dxdy$。

并且，连续随机变量X和Y的联合累积分布函数$\mathrm{F_{XY}}(x, y)$满足：

$$\mathrm{F_{XY}}(x, \mathrm{y}) = \mathrm{P}(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{\mathrm{XY}}(x, y)dxdy \tag{1.18}$$

密度函数为：$f_{\mathrm{XY}}(x,y)=\frac{\partial^2 \mathrm{F}_{\mathrm{XY}}(x,y)}{\partial \mathrm{x}\,\partial \mathrm{y}}$。

进一步，利用二维随机变量X和Y联合概率密度函数可定义随机变量X或Y的**边缘密度函数**$\mathrm{f}_{\mathrm{X}}(x)$或$\mathrm{f}_{\mathrm{Y}}(y)$：

$$f_X(x)=\int_{-\infty}^{\infty} f_{\mathrm{XY}}(x,y)dy \tag{1.19}$$

$$f_Y(y)=\int_{-\infty}^{\infty} f_{\mathrm{XY}}(x,y)dx \tag{1.20}$$

2.1.4 条件分布

(1)离散随机变量$(\boldsymbol{X},\boldsymbol{Y})$条件分布

以**条件概率质量函数**描述离散随机变量(X,Y)的条件分布。

若离散随机变量(X,Y)在(x,y)处的联合概率为$\mathrm{P}_{\mathrm{XY}}(x,y)$，并且$\mathrm{P}_{\mathrm{Y}}(y)>0$，那么已知$Y=y$的条件下,$X=x$的**条件概率**为：

$$\mathrm{P}_{\mathrm{X|Y}}(x|y)=\frac{\mathrm{P}_{\mathrm{XY}}(x,y)}{\mathrm{P}_{\mathrm{Y}}(y)} \tag{1.21}$$

条件分布$\mathrm{P}_{\mathrm{X|Y}}$反映了在已知Y条件下，随机变量X的概率分布情况。

离散随机变量X的条件分布$\mathrm{P}_{\mathrm{X|Y}}$具有如下性质：

- $\mathrm{P}_{\mathrm{X|Y}}(x|y)\geq 0$
- $\sum_{x\in \mathrm{Val}(X)} \mathrm{P}_{\mathrm{X|Y}}(x|y)=1$

(2)连续随机变量$(\boldsymbol{X},\boldsymbol{Y})$条件分布

以**条件概率密度函数**描述连续随机变量(X,Y)的条件分布。

若连续随机变量(X,Y)在(x,y)处的联合概率密度函数为f_{XY}，并且边缘密度$f_Y(y)>0$，那么已知$Y=y$的条件下,$X=x$的**条件密度**为：

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)} \tag{1.22}$$

条件密度$f_{X|Y}$反映了在已知Y条件下，随机变量X的概率分布情况。

连续随机变量X的条件密度$f_{X|Y}$具有如下性质：

- $f_{X|Y}(x|y)\geq 0$
- $\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx=1$

2.1.5 贝叶斯公式

在上述条件分布定义的前提下，这里分别给出两种情况下的贝叶斯公式。

(1)以**条件概率质量函数**描述离散随机变量(X,Y)的条件分布

由于 $\mathrm{P_{XY}}(x,y)=\mathrm{P_{Y|X}}(y|x)\mathrm{P_X}(x)$，并且 $\mathrm{P_Y}(y)=\sum_{\mathrm{x}'\in\mathrm{Val}(X)}\mathrm{P_{XY}}(x',y)=\sum_{\mathrm{x}'\in\mathrm{Val}(X)}\mathrm{P_{Y|X}}(y|x')\mathrm{P_X}(x')$，因此，以贝叶斯公式表示条件概率$\mathrm{P_{X|Y}}(x|y)$：

$$\mathrm{P_{X|Y}}(x|y)=\frac{\mathrm{P_{XY}}(x,y)}{\mathrm{P_Y}(y)}=\frac{\mathrm{P_{Y|X}}(y|x)\mathrm{P_X}(x)}{\sum_{\mathrm{x}'\in\mathrm{Val}(X)}\mathrm{P_{Y|X}}(y|x')\mathrm{P_X}(x')} \tag{1.23}$$

(2)以**条件概率密度函数**描述连续随机变量(X,Y)的条件分布

由于 $f_{XY}(x,y)=f_{Y|X}(y|x)f_X(x)$，并且 $f_Y(y)=\int_{x'\in\mathrm{Val}(X)}f_{XY}(x',y)dx'=\int_{x'\in\mathrm{Val}(X)}f_{Y|X}(y|x')f_X(x')dx'$，因此，以贝叶斯公式表示条件密度$f_{X|Y}(x|y)$：

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{f_{Y|X}(y|x)f_X(x)}{\int_{x'\in\mathrm{Val}(X)}f_{Y|X}(y|x')f_X(x')dx'} \tag{1.24}$$

2.1.6 条件累积分布

对于两个随机变量X和Y，给定$Y=y$的条件下，定义随机变量X的**条件累积分布函数**：$\mathrm{F_{X|Y}(x|y)=P(X\le x|Y=y)}$.

对于离散随机变量X和Y，有：

$$\mathrm{F_{X|Y}(x|y)=P(X\le x|Y=y)}=\sum_{x'\le \mathrm{x}}\mathrm{P_{X|Y}}(x'|y)=\sum_{x'\le\mathrm{x}}\frac{\mathrm{P_{XY}}(x',y)}{\mathrm{P_Y}(y)}=\frac{\sum_{x'\le\mathrm{x}}\mathrm{P_{XY}}(x',y)}{\mathrm{P_Y}(y)}$$

对于连续随机变量X和Y，有：

$$\mathrm{F_{X|Y}(x|y)=P(X\le x|Y=y)}=\int_{-\infty}^{x}f_{X|Y}(t|y)dt=\int_{-\infty}^{x}\frac{f_{XY}(t,y)}{f_Y(y)}dt=\frac{\int_{-\infty}^{x}f_{XY}(t,y)dt}{f_Y(y)}$$

2.1.7 随机变量的独立性及条件独立

对于随机变量X,Y，若对于任意实数x,y，总有$\mathrm{F_{XY}}(x,y)=\mathrm{F_X}(x)\mathrm{F_Y}(y)$，即：$\mathrm{P(X}\le x,Y\le y)=\mathrm{P(X}\le x)\mathrm{P}(Y\le y)$,则称$\boldsymbol{X},\boldsymbol{Y}$**相互独立**。

若X,Y为离散随机变量，并且(X,Y)所有可能取值构成$\mathrm{Val}(X,Y)$，则X,Y相互独立的充要条件是：对于任意$(x,y)\in\mathrm{Val}(X,Y)$，总有$\mathrm{P_{XY}}(x,y)=\mathrm{P_X}(x)\mathrm{P_Y}(y)$,或者$\mathrm{P_X}(x)=\mathrm{P_{X|Y}}(x|y)$, $\mathrm{P_Y}(y)=\mathrm{P_{Y|X}}(y|x)$

若X,Y为连续随机变量，并且$f_{XY}(x,y)$处处连续，则X,Y相互独立的充要条件

是：对于任意(x,y)，总有$f_{XY}(x,y)=f_X(x)f_Y(y)$，或者$f_X(x)=f_{X|Y}(x|y), f_Y(y)=f_{Y|X}(y|x)$.

同理，我们可以定义条件独立。对于随机变量X,Y,Z，在已知$Z=z$的条件下若对于任意实数x,y，总有$\mathrm{F}_{\mathrm{XY|Z}}(x,y|z)=\mathrm{F}_{\mathrm{X|Z}}(x|z)\mathrm{F}_{\mathrm{Y|Z}}(y|z)$，即：$\mathrm{P}(\mathrm{X}\le x, Y\le y)=\mathrm{P}(\mathrm{X}\le x|z)\mathrm{P}(Y\le y|z)$,则称$\boldsymbol{X},\boldsymbol{Y}$在$Z=z$条件下的**相互独立**。

若X,Y,Z为离散随机变量，则X,Y在$Z=z$条件下**相互独立**的充要条件是：总有$\mathrm{P}_{\mathrm{XY|Z}}(x,y|\mathrm{z})=\mathrm{P}_{\mathrm{X|Z}}(x|z)\mathrm{P}_{\mathrm{Y|Z}}(y|z)$。

若X,Y,Z为连续随机变量，并且$f_{XY|\mathrm{Z}}(x,y|z)$处处连续，则X,Y在$Z=z$条件下**相互独立**的充要条件是：对于任意(x,y)，总有$f_{XY|Z}(x,y|\mathrm{z})=f_{X|Z}(x|z)f_{Y|Z}(y|z)$。

2.1.8 期望、方差、协方差

(1)期望

给定二维离散随机变量X和 Y 的联合概率$\mathrm{P}_{XY}(x,y)$或边缘概率$\mathrm{P}_X(x)$，可定义随机变量X的数学期望：

$$\mu_{\mathrm{X}}=\mathrm{E}[X]=\sum_{(x,y)\in\mathrm{Val}(X,Y)}x\mathrm{P}_{XY}(x,y)=\sum_{x\in\mathrm{Val}(X)}x\mathrm{P}_X(x) \tag{1.25}$$

离散随机变量(X,Y)的实值函数$\mathrm{g}(X,Y)$的数学期望：

$$\mathrm{E}[\mathrm{g}(X,Y)]=\sum_{(x,y)\in\mathrm{Val}(X,Y)}\mathrm{g}(x,y)\mathrm{P}_{XY}(x,y) \tag{1.26}$$

给定二维连续随机变量X和 Y 的联合密度$f_{XY}(x,y)$或边缘密度$f_X(x)$，可定义随机变量X的数学期望：

$$\mu_{\mathrm{X}}=\mathrm{E}[X]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{\mathrm{XY}}(x,y)dxdy=\int_{-\infty}^{\infty}xf_{\mathrm{X}}(x)dx \tag{1.27}$$

连续随机变量(X,Y)的实值函数$\mathrm{g}(X,Y)$的数学期望：

$$\mathrm{E}[\mathrm{g}(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{g}(x,y)f_{\mathrm{XY}}(x,y)dxdy \tag{1.28}$$

(2)方差

利用随机变量的函数的数学期望定义随机变量的方差。

给定二维离散随机变量X和 Y 的联合概率$\mathrm{P}_{XY}(x,y)$或边缘概率$\mathrm{P}_X(x)$，可定义随机变量X的方差：

$$\sigma_X^2 = E[(X-\mu_X)^2] = \sum_{x,y}(x-\mu_X)^2 P_{XY}(x,y) = \sum_x (x-\mu_X)^2 P_X(x)$$

给定二维连续随机变量X和 Y 的联合密度$f_{XY}(x,y)$或边缘密度$f_X(x)$，可定义随机变量X的数学期望：

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)^2 f_{XY}(x,y)dxdy = \int_{-\infty}^{\infty}(x-\mu_X)^2 f_X(x)dx$$

(3)协方差

利用协方差可以直观描述随机变量之间的相关性。二维离散随机变量X和 Y 的协方差$\mathrm{Cov}(X,Y) = \sigma_{XY}$：

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_{x,y}(x-\mu_X)(y-\mu_Y)P_{XY}(x,y)$$

二维连续随机变量X和 Y的协方差：

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{XY}(x,y)dxdy$$

协方差满足：$E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - E[X]E[Y]$。若$\sigma_{XY} = 0$，即$E[XY] = E[X]E[Y]$，称$X$和 Y 互不相关。

若X和 Y相互独立，则有$\sigma_{XY} = 0$，从而$E[XY] = E[X]E[Y]$。因此，两随机变量相互独立，二者必互不相关。

进一步，利用两随机变量X和 Y之间的协方差、以及各自方差，定义X和 Y的**互相关系数**：

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

显然，$\rho_{XY} \in [-1,1]$。

有关性质

(I) 对于两个任意常数a_1, a_2以及实值函数 $g_1(X,Y), g_2(X,Y)$,总有

$$E[a_1 g_1(X,Y) + a_2 g_2(X,Y)] = a_1 E[g_1(X,Y)] + a_2 E[g_2(X,Y)]$$

(II) $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}(X,Y)$

(III) X 与 Y 相互独立，则$\mathrm{Cov}(X,Y) = \sigma_{XY} = 0$

(IV) X 与 Y 相互独立，则$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$