目录

第二章 随机变量及其分布 **iii**

2.1 引言 iii

2.2 随机变量的分布函数 iv

2.3 常见的概率分布 vi

2.3.1 常见离散型分布 vi

2.3.2 常见连续型分布 xi

2.4 多维随机变量(随机向量) xiii

2.5 条件分布和独立性 xix

2.5.1 条件分布 xix

2.5.2 随机变量的独立性 xxi

2.6 随机变量的函数的分布 xxii

2.7 总结与讨论 xxx

参考文献 xxxi

第二章　随机变量及其分布

2.1　引言

在很多试验中，处理一个汇总的变量要比处理原来的概率结构方便的多。例如在一次民意调查中，我们调查50个人对某个事情的态度是支持(1)还是反对(0)。那么按照古典概型的处理方式，所有可能的结果有2^{50}个，这是非常大的一个数字。但是如果我们用X={1的个数}，则X的可能取值为$\{0, 1, \cdots, 50\}$，这样处理起来就比原来的概率结构要方便多了.

假设我们随机投两枚骰子，根据两个骰子的点数和决定某个事情。则样本空间为

$$\Omega = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, 2, 3, 4, 5, 6\}\}$$

但是我们其实只对点数和感兴趣，即对函数$S : \Omega \to R$的值感兴趣

$$S(\omega_1, \omega_2) = \omega_1 + \omega_2, \qquad (\omega_1, \omega_2) \in \Omega$$

S的可能值可以通过下表得到:

	ω_2					
ω_1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

我们记事件

$$\{S = k\} = \{(\omega_1, \omega_2) \in \Omega : \quad \omega_1 + \omega_2 = k\}$$

以及其概率为$P(S = k)$.

又若我们是根据两个骰子的最大值决定，则映射

$$M(\omega_1,\omega_2)=\max\{\omega_1,\omega_2\},\quad (\omega_1,\omega_2)\in\Omega$$

类似可以由上表定出由M决定的事件及概率。象S,M这样的函数，我们称其为离散随机变量(discrete random variable).

定义 2.1.1. *一个随机变量(random variable)是从样本空间Ω到实数轴的一个(映射)函数。具体的讲，称样本空间Ω到实数轴$\mathbb{R}$的一个映射$X(\cdot)$为随机变量，如果*

$$\{\omega: X(\omega)<x\}\in\mathcal{F}$$

对任意的实数$x\in\mathbb{R}$。$X(\cdot)$经常简写为X.

例2.1.1. 随机事件的示性函数是随机变量。

2.2 随机变量的分布函数

定义 2.2.1. *称*

$$F(x)=P(X\leq x),\quad for\ all\ x\ \in\mathbb{R}$$

为随机变量X的(累积)分布函数(cdf, cumulative distribution function).

例2.2.1. 投掷三枚硬币, X=正面向上的个数。则X的分布函数为

$$F(x)=\begin{cases}0 & if\ -\infty<x<0\\ \frac{1}{8} & if\ 0\leq x<1\\ \frac{1}{2} & if\ 1\leq x<2\\ \frac{7}{8} & if\ 2\leq x<3\\ 1 & if\ 3\leq x<\infty\end{cases}$$

分布函数$F(x)$满足:

1. $\lim\limits_{x\to-\infty}F(x)=0$和$\lim\limits_{x\to+\infty}F(x)=1$;
2. $F(x)$是一个非降的函数;
3. $F(x)$是一个右连续的函数, 即对任意的x_0有$\lim\limits_{x\downarrow x_0}F(x)=F(x_0)$.

反之，若一个函数满足1-3，则这个函数是一个分布函数。

证明3：对任意的$x_n \downarrow x_0$，事件列$A_n = \{X \le x_n\}$为单调下降的，且有$\bigcap\limits_{x_n \downarrow x_0} A_n = \{X \le x_0\}$。从而有概率的上连续性立得。

根据随机变量取值的性质，我们主要考虑如下两类随机变量。

定义 2.2.2. 若随机变量X的所有可能取值为至多可列个点值，则称X为离散型随机变量。此时，若记X 的所有取值为$\{x_1, x_2, \cdots\}$，则称

$$P(X = x_k) = p_k, \quad k = 1, 2, \cdots$$

或者表示为

$$\begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \qquad \text{(也称此为密度阵)}$$

为随机变量X的分布律(概率函数、密度).

显然，概率函数满足

1. $p_k \ge 0,\ k = 1, 2, \cdots$
2. $\sum\limits_{k=1}^{\infty} p_k = 1$
3. $F(x) = P(X \le x) = \sum\limits_{x_k \le x} P(X = x_k) = \sum\limits_{x_k \le x} p_k$

反之，若有一列实数$\{p_k\}$满足如上1-2两条，则存在某个随机变量使得$\{p_k\}$为其分布律。

按照定义，只有一个可能值的常数c是离散型随机变量，其概率函数为$\binom{c}{1}$。称之为退化分布。

定义 2.2.3. 若存在非负函数$f(x)$使得随机变量X的分布函数

$$F(x) = \int_{-\infty}^{x} f(x)dx, \quad for\ all\ x\ \in \mathbb{R}$$

则称随机变量X为连续型的，并称$f(x)$为X的概率密度函数*(pdf, probability density function)*。

概率密度$f(x)$满足

1. $f(x) \geq 0, forall\ x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. 若$f(x)$在x点连续，则$F'(x) = f(x)$.
4. 对任意的Borel可测集合B, $P(X \in B) = \int_{x \in B} f(x)dx$

反之，若一个函数$f(x)$满足如上1-2两条，则存在某个连续型随机变量使得$f(x)$为其概率密度。

以上连个定义也可以表述为

定义 2.2.4. *称随机变量X为连续型(continuous)的，如果它的分布函数是连续的；称随机变量X为离散型(discrete)的，如果它的分布函数是一个阶梯函数。*

既非连续也非离散的随机变量是存在的，比如

例2.2.2. 设随机变量Y的分布函数为

$$F(y) = \begin{cases} \frac{1-\epsilon}{1+e^{-y}} & if\ y < 0 \\ \epsilon + \frac{1-\epsilon}{1+e^{-y}} & if\ y \geq 0 \end{cases}$$

其中$0 < \epsilon < 1$. 则Y既不是连续的也不是离散的。

2.3 常见的概率分布

2.3.1 常见离散型分布

1. 0-1分布(Bernoulli distribution)

设随机变量X只取两个值，不妨记为0和1，而且$P(X = 1) = p = 1 - P(X = 0)$。其概率函数为

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

或者表示成

$$P(X = i) = p^i(1-p)^{1-i}, \quad i = 0, 1. \tag{2.3.1}$$

则称X的分布为0-1分布并称X服从0-1分布。

在0-1分布中，常把$X = 1$成为一个“成功”。随机变量的示性函数服从的就是0-1分布。

2. 二项分布(Binomial distribution)

定义 2.3.1. *若一个随机试验只有两个可能的结果A和$\bar{A}$，则称这个随机试验为贝努里(Bernoulli)试验。记$p = P(A)(0 < p < 1)$，将这个试验独立的重复做n次，则称这一串独立的试验为n重贝努里试验。在一次试验中，当结果A发生时，称为一次成功。*

在n重贝努里试验中，若以X表示成功的次数(即随机事件A发生的次数)，则X为一离散型随机变量。易知X的概率函数为

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}, \quad k = 0, 1, \cdots, n. \tag{2.3.2}$$

其中$q = 1 - p$。称此概率分布为二项分布，并称X服从二项分布，记为$X \sim B(n, p)$.

例2.3.1. 按规定, 某种型号的电子元件的使用寿命超过1500小时的为一级品. 已知某一大批产品的一级品率为0.2, 现随机地抽查20件产品, 问这20件产品中恰有$k(k = 0, 1, \cdots, 20)$件一级品的概率是多少。

解: 这是不放回抽样，但是由于产品总数很大，而抽取的20件相对于总数来说很小，故我们可以视为是由放回的抽样。我们检查一个产品相当于作一次试验，抽查20件产品相对于做20次独立的贝努里试验，若以X表示20件中的一级品个数，则X服从二项分布$B(20, 0.2)$，所以

$$P(X = k) = C_{20}^k 0.2^k 0.8^{20-k}, \quad k = 0, 1, \cdots, 20.$$

例2.3.2. 某人进行射击，设每次射击的命中率为0.02。现独立的重复射击400次，求至少击中两次的概率。

解: 设X表示射击400次中的击中的次数，则$X \sim B(400, 0.02)$。所以

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0.98^{400} - 400 * 0.02 * 0.98^{399} = 0.9972.$$

本例说明小概率事件在试验次数足够多时必然发生。

3. 几何分布(Geometric distribution)

定义 2.3.2. *在n重贝努里实验中，当试验次数$n \to \infty$时，称为可列重贝努里试验。*

若以X表示在可列重贝努里试验中结果A出现时的试验次数，即若以“成功”表示结果A发生，则X表示首次成功时的试验次数，所以

$$P(X = k) = q^{k-1} p, \quad k = 1, 2, \cdots. \tag{2.3.3}$$

称此分布为几何分布. 记为$X \sim G(p)$.

例2.3.3. 一个人要开门,他共有n把钥匙。其中仅有一把可以打开门。现随机地有放回的从中选取一把开门，问这人在第S次试开成功的概率。

定理 2.3.1. *以所有正整数为取值集合的随机变量ξ服从几何分布$G(p)$，当且仅当对任何正整数m和n,都有*

$$P(\xi > m+n \mid \xi > m) = P(\xi > n). \tag{2.3.4}$$

这个性质称为几何分布的无记忆性(memoryless property).

证:设随机变量ξ服从几何分布$G(p)$,写$q=1-p$,那么对任何非负整数k,都有

$$P(\xi > k) = \sum_{j=k+1}^{\infty} P(\xi = j) = p \sum_{j=k+1}^{\infty} q^{j-1} = q^k \,.$$

所以对任何正整数m和n,都有

$$\begin{aligned} P(\xi > m+n \mid \xi > m) &= \frac{P(\xi > m+n,\ \xi > m)}{P(\xi > m)} \\ &= \frac{P(\xi > m+n)}{P(\xi > m)} = \frac{q^{m+n}}{q^n} = q^n = P(\xi > n). \end{aligned}$$

故知(2.3.4)式成立.

反之,设对任何正整数m和n,都有(2.3.4)式成立.对非负整数k,我们记$p_k = P(\xi > k)$．于是由(2.3.4)式知,对任何正整数k,都有$p_k > 0$,并且对任何正整数m和n,都有$p_{m+n} = p_m \cdot p_n$．由此等式立知,对任何正整数m,都有$p_m = p_1^m$.由于$p_1 > 0$,而若$p_1 = 1$,则必导致对一切正整数m, 都有$p_m = 1$,此为不可能,所以对某个小于1的正数q,有$p_1 = q$.由此不难得,对任何正整数m,都有

$$P(\xi = m) = P(\xi > m-1) - P(\xi > m) = p_{m-1} - p_m = q^{m-1} - q^m = p\, q^{m-1},$$

其中$p = 1-q$,所以ξ服从几何分布$G(p)$.

我们还可以证明几何分布是唯一的具有无记忆性的取值集合为正整数集的离散型分布.

4. Pascal分布(负二项分布)

在可列重贝努里试验中，若以X_r表示第r次成功发生时的试验次数，则X_r的分布律为

$$\begin{aligned} P(X_r = k) &= P(\{\text{前}k-1\text{次恰有}r\text{次成功且第}k\text{次成功}\}) \\ &= P(\{\text{前}k-1\text{次恰有}r\text{次成功}\})P(\{\text{第}k\text{次成功}\}) \\ &= C_{k-1}^{r-1} p^{r-1} q^{k-r} \cdot p \\ &= C_{k-1}^{r-1} p^r q^{k-r}, \quad k = r, r+1, \cdots . \end{aligned}$$

称此概率分布为Pascal分布。如果记

$$p_k = C_{k-1}^{r-1} p^r q^{k-r}, \quad k = r, r+1, \cdots \tag{2.3.5}$$

那么显然有

$$\sum_{k=r}^{\infty} p_k = \sum_{k=r}^{\infty} C_{k-1}^{r-1} p^{\,r} q^{k-r} = p^{\,r} \sum_{k=0}^{\infty} C_{r+k-1}^{r-1} q^{\,k} = p^{\,r}(1-q)^{-r} = 1\,,$$

所以(2.3.5)式的确是一个离散型随机变量的分布律.我们将其称为参数为p和r的Pascal分布. 又因为上式表明,它可以用负二项展开式中的各项表示,所以又称为负二项分布.

例2.3.4. (Banach火柴问题)某人口袋里放有两盒火柴,每盒装有火柴n根.他每次随机取出一盒,并从中拿出一根火柴使用.试求他取出一盒,发现已空,而此时另一盒中尚余r根火柴的概率.

解:以A表示甲盒已空,而此时乙盒中尚余r根火柴的事件.由对称性知,所求的概率等于$2P(A)$.我们将每取出甲盒一次视为取得一次成功,以ξ表示取得第$n+1$次成功时的取盒次数,则ξ服从参数为0.5和$n+1$ 的Pascal 分布(因为每次取出甲盒的概率是0.5).易知,事件A发生,当且仅当ξ等于$2n-r+1$.所以所求的概率等于

$$2P(A) = 2P(\xi = 2n-r+1) = C_{2n-r}^{n} 2^{\,r-2n}.$$

例2.3.5. 在可列重贝努里试验中，求事件E ={n次成功发生在m次失败之前}的概率。

解: 记F_k={第n次成功发生在第k次试验},则

$$E = \bigcup_{k=n}^{n+m-1} F_k$$

且诸F_k两两互斥，故

$$P(E) = \sum_{k=n}^{n+m-1} P(F_k) = \sum_{k=n}^{n+m-1} C_{k-1}^{n-1} p^n q^{n-k}.$$

5. Possion分布

设随机变量X的概率分布为

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \cdots. \tag{2.3.6}$$

称此分布律为参数为λ的Poisson 分布，并记$X \sim P(\lambda)$。

定理 2.3.2. (Poisson 定理)设有一列二项分布$B(n,\ p_n)$,其中的参数满足条件

$$\lim_{n\to\infty} np_n = \lambda > 0\,, \tag{2.3.7}$$

则对任何非负整数k都有

$$\lim_{n\to\infty} b(k;\ n, p_n) = \lim_{n\to\infty} C_n^k {p_n}^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}\,. \tag{2.3.8}$$

证:易知,对每个固定的非负整数k,有

$$\begin{aligned} C_n^k {p_n}^k (1-p_n)^{n-k} &= \frac{n!}{k!\,(n-k)!}{p_n}^k(1-p_n)^{n-k} \\ &= \frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(1-p_n)^{-k}(np_n)^k(1-p_n)^n\,. \end{aligned}$$

在条件(2.3.7)之下,对固定的非负整数k,显然有

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(1-p_n)^{-k} = 1$$

和$\lim\limits_{n\to\infty}(np_n)^k = \lambda^k$. 因此为证结论, 只需证明

$$\lim_{n\to\infty}(1-p_n)^n = e^{-\lambda}. \tag{2.3.9}$$

众所周知,有$\lim\limits_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$.因此若要证明(2.3.9),就只要证明

$$\lim_{n\to\infty}(1-p_n)^n = \lim_{n\to\infty}(1-\frac{\lambda}{n})^n.$$

显然有$|1-p_n| < 1$,而当n充分大时,也有$|1-\frac{\lambda}{n}| < 1$.我们知道,对$|a| \le 1,\ |b| \le 1$,有初等不等式$|a^n-b^n| \le n|a-b|$成立,结合条件(2.3.7),便知当$n\to\infty$时,有

$$\left|(1-p_n)^n - (1-\frac{\lambda}{n})^n\right| \le n|p_n - \frac{\lambda}{n}| = |np_n-\lambda| \to 0\,.$$

所以(2.3.9)式成立,定理证毕.

例2.3.6. 假设一块放射性物质在单位时间内发射出的α粒子数ξ服从参数为λ的Poisson分布。而每个发射出来的α粒子被记录下来的概率是p，就是说有$q=1-p$的概率被记数器漏记。如果各粒子是否被记数器记录是相互独立的，试求记录下来的α粒子数η的分布。

解: 以事件$\{\xi=n\}, n=0,1,2,\cdots$为划分，则由全概率公式有

$$\begin{aligned} P(\eta=k) &= \sum_{n=0}^{\infty} P(\eta=k|\xi=n)P(\xi=n) \\ &= \sum_{n=k}^{\infty}\binom{n}{k}p^k q^{n-k}\frac{\lambda^n}{n!}e^{-\lambda} \\ &= \sum_{n=k}^{\infty}\frac{(\lambda q)^{n-k}}{k!(n-k)!}e^{-\lambda}(\lambda p)^k = \frac{(\lambda p)^k}{k!}e^{-\lambda p}, \quad k=0,1,2,\cdots.\# \end{aligned}$$

2.3.2 常见连续型分布

1. 均匀分布(Uniform distribution) 设$a<b$,如果分布$F(x)$具有密度函数

$$p(x)=\frac{1}{b-a},\quad a\le x\le b\,, \tag{2.3.10}$$

则称该分布为区间$[a,b]$上的均匀分布,记作$U[a,b]$. 如此定义的$p(x)$显然是一个概率密度函数,容易算出其相应的分布函数为

$$F(x)=\begin{cases} 0, & x\le a,\\ \frac{x-a}{b-a}, & a<x\le b,\\ 1, & x>b.\end{cases}$$

2. 指数分布(Exponential distribution) 如果分布$F(x)$具有密度函数

$$f(x)=\begin{cases} \lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases} \tag{2.3.11}$$

则称该分布为期望为$1/\lambda$的指数分布。记为$exp\{\lambda\}$.

与几何分布类似, 指数分布也是一种"无记忆分布",并且是唯一的无记忆的连续型分布, 对此我们有

定理 2.3.3. *如果ξ为取非负实数值的随机变量, 则ξ服从指数分布, 当且仅当,*

$$P(\xi>s+t\mid\xi>s)=P(\xi>t),\quad \forall\ s>0,\ t>0. \tag{2.3.12}$$

例2.3.7. 设X表示某种电子元件的寿命，$F(x)$为其分布函数。若假设元件无老化，即元件在时刻x正常工作的条件下，其失效率保持为某个常数λ，与x无关。试证明X服从指数分布。

解：失效率即单位时间内失效的概率，因此由题设知

$$P(x\le X\le x+h|X>x)/h=\lambda,\quad h\to 0$$

因为

$$P(x\le X\le x+h|X>x)=\frac{P(\{x\le X\le x+h\}\{X>x\})}{P(X>x)}=\frac{F(x+h)-F(x)}{1-F(x)}$$

所以有

$$\lim_{h\to 0}P(x\le X\le x+h|X>x)/h=\frac{F'(x)}{1-F(x)}=\lambda$$

即得到微分方程$\frac{F'(x)}{1-F(x)}=\lambda$,解此方程得到

$$F(x)=1-e^{-\lambda x}$$

从而结论得证。

3. 正态分布(Normal distribution) 如果连续型分布$F(x)$的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\},\quad x\in\mathbb{R}. \tag{2.3.13}$$

则称该分布为以a和σ^2为参数的正态分布,记作$N(a,\sigma^2)$.正态分布也称为Gauss分布.

在正态分布中,$a=0,\ \sigma=1$的情形具有特别重要的意义,我们将之称为标准正态, 并分别用$\Phi(x)$和$\phi(x)$表示标准正态$N(0,1)$的分布函数和密度函数.

图 2.1: 正态分布密度

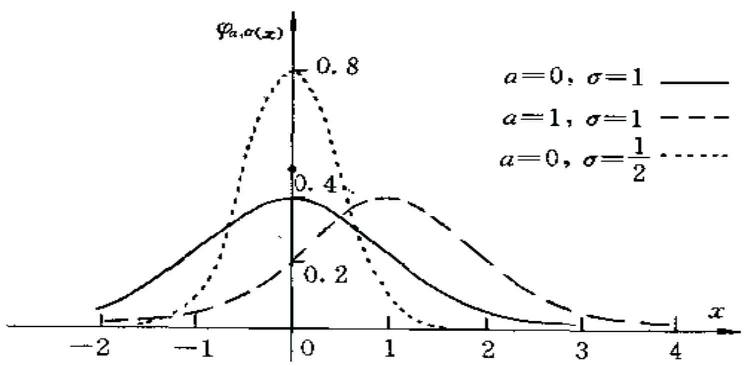

图 4.2 正态分布的密度曲线

通过观察密度函数$f(x)$ 的图形,可以帮助我们初步了解其中两个参数a和σ^2的意义:

1° $f(x)$关于$x=a$对称,即有

$$f(a-x)=f(a+x),\quad x\in\mathbb{R}.$$

特别地,标准正态$N(0,1)$的密度函数

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}},\quad x\in\mathbb{R}$$

关于$x=0$对称,即为偶函数. 我们将a称为位置参数.

2° $f(x)$在$x=a$处达到最大值$\frac{1}{\sqrt{2\pi}\sigma}$. 并且$\sigma$ 的值越小,$f(x)$的峰越陡峭;反之,σ的值越大,$f(x)$的峰越平缓.

我们知道,如果ξ是服从正态分布$N(a,\sigma^2)$的随机变量,那么对任何$x_1<a<x_2$,就有

$$P(x_1<\xi<x_2)=\int_{x_1}^{x_2}\phi_{a,\sigma}(u)du,$$

因此,σ的值越小,那么概率$P(x_1 < \xi < x_2)$的值就越大,反之,σ的值越大,那么概率$P(x_1 < \xi < x_2)$的值就越小. 因此σ的值反映了ξ在$x = a$附近取值的集中程度. 通常将σ 称为正态$N(a, \sigma^2)$的形状参数.

3° $f(x) = \Phi(\frac{x-a}{\sigma})$.

例2.3.8. 如果随机变量ξ服从正态分布$N(a, \sigma^2)$,则有

$$P(|\xi - a| < 3\sigma) = 0.9974\,, \tag{2.3.14}$$

我们把这个性质称为3σ**原则**.

事实上, 不难得知

$$P(|\xi - a| < 3\sigma) = P(-3 < \frac{\xi - a}{\sigma} < 3) = \Phi(3) - \Phi(-3) = 0.9974\,.$$

这一结果反映了正态随机变量的一个十分引人注目的性质.我们已经知道,正态分布$N(a, \sigma^2)$的密度函数$f(x)$处处为正,所以正态随机变量ξ在任何区间中取值的概率都为正数,但是上述结果告诉我们,ξ以大于0.9974的概率取区间$(a - 3\sigma,\ a + 3\sigma)$中的值,从而若以该区间来估计$\xi$ 的取值范围,其误差小于0.003,我们将这种估计称为"3σ 原则". 近来, 3σ 原则被广泛地运用到企业的质量管理上,称为6σ 管理原则(在a的左右各三个σ).

2.4 多维随机变量(随机向量)

在实际应用中，对所考虑的问题需要多个变量来描述。如研究身高体重关系等。简单的讲，我们把多个随机变量放在一起组成向量，称为随机向量或者多维随机变量。

定义 2.4.1. 称$X = (X_1, \ldots, X_n)$为n维随机变量或者随机向量，如果X的每一维$X_i (i = 1, \cdots, n)$都是一个随机变量。

对随机向量，我们也可以仿照随机变量分布函数的定义来定义随机向量的分布函数。

定义 2.4.2. 设$X = (X_1, \ldots, X_n)$为n维随机变量，称

$$F(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n) = P(\prod_{i=1}^{n} \{\omega : X_i(\omega) \le x_i\}),\quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n \tag{2.4.1}$$

为随机向量X的分布函数或者随机变量$X_1,\dots,X_n$的联合分布函数。而称$F_{X_i}(x_i)=P(X_i\le x_i)$为随机变量$X_i(i=1,\dots,n)$的边际(边缘)分布函数。

分布函数的性质 n维分布$F(x_1,\cdots,x_n)$具有下述性质:
1°. $F(x_1,\cdots,x_n)$对每个变元非降;
2°. $F(x_1,\cdots,x_n)$对每个变元右连续;
3°. $\lim\limits_{x_j\to-\infty}F(x_1,\cdots,x_n)=0\,,\quad\forall\,1\le j\le n,$
$\lim\limits_{x_1\to\infty,\cdots,x_n\to\infty}F(x_1,\cdots,x_n)=1\,;$

根据多维随机变量取值的特点，我们考虑离散型随机向量和连续型随机向量。

定义 2.4.3. 称$X=(X_1,\dots,X_n)$为n维离散随机变量或者随机向量，如果X的每一维$X_i(i=1,\cdots,n)$ 都是一个离散型随机变量。设X_i的所有可能取值为$\{a_{i1},a_{i2},\cdots\}(i=1,\dots,n)$,则称

$$P(X_1=a_{1j_1},\dots,X_n=a_{nj_n})=p(j_1,\dots,j_n),\quad j_i=1,2,\cdots(i=1,2,\dots,n).\tag{2.4.2}$$

为离散型随机向量X的概率函数或者n维离散型随机变量$X_1,\cdots,X_n$的联合概率函数(分布律)。而称$P(X_i=a_{ij_i})=p_{j_i}(j_i=1,2,\cdots.)$为离散型随机变量$X_i(i=1,\dots,n)$的边际概率函数(分布律).

分布律的性质

1. $p(j_1,\dots,j_n)\ge 0,\quad j_i=1,2,\cdots(i=1,2,\dots,n).$
2. $\sum\limits_{j_1,\cdots,j_n}p(j_1,\dots,j_n)$
3. $p_{j_i}=\sum\limits_{j_1,\cdots,j_{i-1},j_{i+1},\cdots,j_n}p(j_1,\dots,j_n)$

对连续型随机向量，类似于连续型随机变量的定义有

定义 2.4.4. 称$X=(X_1,\dots,X_n)$为n维连续型随机变量或者随机向量，如果存在$\mathbb{R}^n$上的非负函数$f(x_1,\dots,x_n)$，使得对任意的$Borel$可测集$A\subset\mathbb{R}^n$，有

$$P(X\in A)=\int\cdots\int_A f(x_1,\dots,x_n)dx_1\cdots dx_n\tag{2.4.3}$$

并称f为随机向量X的概率密度或者随机变量$X_1,\dots,X_n$的联合概率密度。而称$f_i(x_i)$ 为连续型随机变量$X_i(i=1,\dots,n)$的边际概率密度。

概率密度的性质

1. $f(x_1,\ldots,x_n)\geq 0,\quad \forall(x_1,\cdots,x_n)\in\mathbb{R}^n$
2. $\underset{\mathbb{R}^n}{\int\cdots\int} f(x_1,\ldots,x_n)dx_1\cdots dx_n=1$
3. $f_i(x_i)=\underset{\mathbb{R}^{n-1}}{\int\cdots\int} f(x_1,\ldots,x_n)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_n$

例2.4.1. 设$A_1,\cdots,A_n$为某一实验下的完备事件群，即$A_1,\cdots,A_n$两两互斥且和为Ω。记$p_k=P(A_k)(k=1,\ldots,n)$，则$p_k\geq 0,p_1+\cdots+p_n=1$。现将实验独立的重复作$N$次，分别用$X_i$表示事件$A_i$出现的次数$(i=1,\cdots,n)$。则$X=(X_1,\ldots,X_n)$ 为一离散型随机向量，试求X的概率函数。此分布律称为多项分布, 记为$M(N;p_1,\ldots,p_n)$.

解: 由于试验独立进行, 总的结果数为N，记结果A_i出现的次数为k_i，则$k_1+\cdots+k_n=N$。因此相当于多组组合，所以

$$\begin{aligned}
P(X_1=k_1,\cdots,X_n=k_n) &= \frac{N!}{k_1!\cdots k_n!}P(\underline{A_1\cdots A_1}\ldots\underline{A_n\cdots A_n})\\
&= \frac{N!}{k_1!\cdots k_n!}p_1^{k_1}\cdots p_n^{k_n},\\
&\quad \text{其中 } k_1,\ldots,k_n \text{ 为非负整数且 } k_1+\cdots+k_n=N.
\end{aligned}$$

我们来看一下X_i的分布：此时我们把试验结果分为两类, A_i和$\bar{A}_i$，则显然就是一个N重贝努里试验，因此

$$P(X_i=k_i)=\binom{N}{k_i}p_i^{k_i}(1-p_i)^{N-k_i},\quad k_i=1,\cdots,N.$$

类似我们也可以找出$(X_i,X_j)(i\neq j)$的联合分布律，即为$M(N,p_i,p_j,1-p_i-p_j)$.

例2.4.2. 设$X=(X_1,\cdots,X_n)$有概率密度

$$f(x_1,\cdots,x_n)=\begin{cases}\frac{1}{m(\Omega)}, & (x_1,\cdots,x_n)\in\Omega\\ 0, & otherwise\end{cases}$$

其中$\Omega\subset\mathbb{R}^n$且$0<m(\Omega)<\infty$。称此概率密度为$\Omega$上的均匀分布。

二维随机变量

1. 二维离散型随机变量

设二维离散型随机变量(X,Y)的所有可能取值为$\{(x_i,y_j):i,j=1,2,\cdots\}$，则$(X,Y)$的联合分布律为

$$P(X=x_i,Y=y_j)=p_{ij}\quad i,j=1,2,\cdots$$

或者表示成

Y\ X	x_1	x_2	$\cdots$	x_n	$\cdots$	$P(Y=y_j)$
y_1	p_{11}	p_{21}	$\cdots$	p_{n1}	$\cdots$	$p_{\cdot 1}$
y_2	p_{12}	p_{22}	$\cdots$	p_{n2}	$\cdots$	$p_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_m	p_{1m}	p_{2m}	$\vdots$	p_{nm}	$\vdots$	$p_{\cdot m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$P(X=x_i)$	$p_{1\cdot}$	$p_{2\cdot}$	$\cdots$	$p_{n\cdot}$	$\cdots$	1

且X的边际分布律为

$$\begin{aligned}
P(X=x_i) &= P(\{\omega: X(\omega)=x_i\}\sum_j\{\omega: Y(\omega)=y_j\}) \\
&= P(\sum_j\{\omega: X(\omega)=x_i, Y(\omega)=y_j\}) \\
&= \sum_j P(X=x_i, Y=y_j) = \sum_j p_{ij} = p_{i\cdot} \quad i=1,2,\cdots
\end{aligned}$$

类似可以得到Y的边际分布律

$$P(Y=y_j) = \sum_i p_{ij} = p_{\cdot j} \quad j=1,2,\cdots$$

同时(X,Y)的联合分布函数为

$$F(x,y) = P(X\le x, Y\le y) = \sum_{x_i\le x, y_j\le y} P(X=x_i, Y=y_j) = \sum_{i,j: x_i\le x, y_j\le y} p_{ij}$$

例2.4.3. *袋中有5张外形相同的卡片，其中3张写上数字"0"，另2张写上"1"。现从袋中任取两张卡片，分别以ξ,η表示第一张和第二张卡片上的数字，试求分别在有放回和不放回两种情形下(ξ,η)的联合分布律及边际分布律.*

解：简单计算得到

$\eta\backslash\xi$	0	1	$p_{\cdot j}$
0	9/25	6/25	3/5
1	6/25	4/25	2/5
$p_{i\cdot}$	3/5	2/5	1

$\eta\backslash\xi$	0	1	$p_{\cdot j}$
0	6/20	6/20	3/5
1	6/20	2/20	2/5
$p_{i\cdot}$	3/5	2/5	1

这个例子说明**边际分布律不能决定联合分布律**。

2. 二维连续型随机变量

定义 2.4.5. 设(X,Y)为二维随机变量，其分布函数为$F(x,y)$。如果存在$\mathbb{R}^2$上的非负函数$f(x,y)$，使得

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dvdu,\qquad \forall(x,y)\in\mathbb{R}^2$$

则称(X,Y)为二维连续型随机变量，并称$f(x,y)$为X和Y的联合概率密度函数。而称$F_X(x),f_X(x),F_Y(y),f_Y(y)$分别为$X$和$Y$的边际分布函数和边际概率密度函数。

设(X,Y)为二维连续型随机变量，则在上述定义中的记号下

$$\begin{aligned}F_X(t)=P(X\le t) &= P(X\le t,-\infty<Y<\infty)=\int_{-\infty}^{t}\int_{-\infty}^{\infty}f(x,y)dydx\\ &= \int_{-\infty}^{t}f_X(x)du\end{aligned}$$

这里

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy$$

按照连续型随机变量的定义知X的边际概率密度函数就是$f_X(x)$. 类似可以到Y的边际概率密度为

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx$$

例2.4.4. 考虑两个概率密度函数

$$\begin{aligned}p(x,y) &= x+y,\qquad 0<x,y<1\\ q(x,y) &= (x+\frac{1}{2})(y+\frac{1}{2}),\qquad 0<x,y<1\end{aligned}$$

试求边际概率密度。

解：易得所求边际概率密度都是如下形式

$$f(t)=t+\frac{1}{2},\quad 0<t<1.$$

说明**边际概率密度不能决定联合概率密度**。

例2.4.5. 设(X,Y)的联合概率密度有形式$(\forall(x,y)\in\mathbb{R}^2)$

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

其中$-\infty<a,b<\infty;0<\sigma_1,\sigma_2<\infty;-1\le\rho\le1$. 则称$(X,Y)$服从参数为$a,b,\sigma_1,\sigma_2,\rho$的二元正态分布，记为$N(a,b,\sigma_1^2,\sigma_2^2,\rho)$. 试计算$X$和$Y$的边际概率密度。

解:

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a)^2}{\sigma_1^2} - 2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2}\right]\right\} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} exp\left\{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right\} dv \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1} exp\{-\frac{1}{2}[(\frac{v-\rho u}{\sqrt{1-\rho^2}})^2 + u^2]\} dv \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} exp\{-\frac{(x-a)^2}{2\sigma_1^2}\}
\end{aligned}
$$

即$X \sim N(a, \sigma_1^2)$. 类似可得$Y \sim N(b, \sigma_2^2)$, 其边际概率密度为$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} exp\{-\frac{(y-b)^2}{2\sigma_2^2}\}$.

2.5 条件分布和独立性

2.5.1 条件分布

一个随机变量或向量的条件概率分布，就是在给定某种条件下的概率分布。

1. 离散型随机变量的条件分布

设(X,Y)为二维离散型随机变量，其全部的可能取值为$\{(x_i,y_j):i,j=1,2,\cdots\}$。记其联合分布律为

$$p_{ij}=P(X=x_i,Y=y_j),\qquad i,j=1,2,\cdots$$

若对给定的事件$\{Y=y_j\}$，其概率$P(Y=y_j)>0$，则称

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots$$

为X在给定$Y=y_j$的条件下的条件分布律(概率函数)。类似的，若$P(X=x_i)>0$，则称

$$P(Y=y_j|X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots$$

为Y在给定条件$X=x_i$下的条件分布律。

例2.5.1. 设$X=(X_1,X_2,\cdots,X_n)\sim M(N;p_1,p_2,\ldots,p_n)$，试求$X_1$在给定$X_2=k$的条件下的条件分布律。

解：由于易知$(X_1,X_2)\sim M(N;p_1,p_2,1-p_1-p_2)$，即其联合分布律为

$$P(X_1=i,X_2=j)=\frac{N!}{i!j!(N-i-j)!}p_1^ip_2^j(1-p_1-p_2)^{N-i-j},\quad 0\le i,j\le N\ \&\ 0\le i+j\le N.$$

并且$X_2\sim B(N,p_2)$.

因此

$$\begin{aligned}P(X_1=i|X_2=k) &= \frac{P(X_1=i,X_2=k)}{P(X_2=k)}\\ &= \frac{N!}{i!k!(N-i-k)!}p_1^ip_2^k(1-p_1-p_2)^{N-i-k}\Big/C_N^kp_2^k(1-p_2)^{N-k}\\ &= \frac{(N-k)!}{i!(N-k-i)!}\left(\frac{p_1}{1-p_2}\right)^i\left(1-\frac{p_1}{1-p_2}\right)^{N-k-i},\ i=0,1,\cdots,N-k.\end{aligned}$$

即X_1在给定$X_2=k$的条件下服从二项分布$B(N-k,p_1/(1-p_2))$.

2. 连续型随机变量的条件分布

设(X,Y)有概率密度$f(x,y)$，我们考虑在给定$a\le Y\le b$的条件下X的条件分布函数

$$\begin{aligned}P(X\le x|a\le Y\le b) &= \frac{P(X\le x,a\le Y\le b)}{P(a\le Y\le b)}\\ &= \int_{-\infty}^{x}\int_a^b f(u,y)dydu\Big/\int_a^b f_Y(y)dy\\ &= \int_{-\infty}^{x}\frac{\int_a^b f(u,y)dy}{\int_a^b f_Y(y)dy}du\end{aligned}$$

因此X在给定条件$a\le Y\le b$下的条件概率密度为

$$f_1(x|a\le Y\le b)=\frac{\int_a^b f(x,y)dy}{\int_a^b f_Y(y)dy}$$

我们更感兴趣的是当$a=b$的情形。由极限性质(若极限存在)的到

$$\begin{aligned}f_1(x|Y=a) &= \lim_{h\to 0} f_1(x|a\le Y\le a+h)\\ &= \lim_{h\to 0}\frac{f(x,\xi_1)}{f_Y(\xi_2)}=\frac{f(x,a)}{f_Y(a)},\quad if\ f_Y(a)>0\end{aligned}$$

我们把上式记为

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)},\quad if\ f_Y(y)>0$$

成为X在给定$Y=y$的条件下的条件概率密度。类似地有Y在给定$X=x$的条件下的条件概率密度

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)},\quad if\ f_X(x)>0$$

例2.5.2. 设(X,Y)服从二元正态分布$N(a,b,\sigma_1^2,\sigma_2^2,\rho)$，试求$X|Y=y$的条件概率密度。

解:

$$\begin{aligned}f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)}\\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\{-\frac{(x-(a+\rho\sigma_1\sigma_2^{-1}(y-b)))^2}{2\sigma_1^2(1-\rho^2)}\}\end{aligned}$$

即$X|Y=y\sim N(a+\rho\sigma_1\sigma_2^{-1}(y-b),\sigma_1^2(1-\rho^2))$。

例2.5.3. 设X,Y服从单位圆上的均匀分布，试求$f_{X|Y}(x|y)$和$f_{Y|X}(y|x)$。

解: 由题设知(X,Y)的联合概率密度为

$$f(x,y)=\begin{cases}\frac{1}{\pi} & if\ x^2+y^2\le 1\\ 0 & otherwise\end{cases}$$

易知
$$f_X(x)=\begin{cases}\frac{2}{\pi}\sqrt{1-x^2} & if -1\le x\le 1\\ 0 & otherwise\end{cases}$$
所以
$$f_{X|Y}(x|y)=\begin{cases}\frac{1}{2\sqrt{1-y^2}} & if -\sqrt{1-y^2}\le x\le \sqrt{1-y^2}\\ 0 & otherwise\end{cases}$$
只需要把x,y互换，就可以得到$f_{Y|X}(y|x)$。

2.5.2 随机变量的独立性

条件分布等于无条件分布的情况，即称是独立的。定义如下

定义 2.5.1. 设$X_1,\cdots,X_n$为概率空间$(\Omega,\mathcal{F},P)$上的n个随机变量，如果它们的联合分布函数等于各自边际分布函数的乘积，即

$$F(x_1,\cdots,x_n)=F_1(x_1)\cdots F_n(x_n)$$

则称随机变量$X_1,\cdots,X_n$相互独立。一族无限多的随机变量称为是独立的，如果其中任意有限个相互独立。

例2.5.4. 如果随机变量$X_1,\cdots,X_n$相互独立，则其中任何一部分随机变量也相互独立。

证明显然，此性质的逆不真。反例如下

例2.5.5. 若ξ,η相互独立,都服从-1和1这两点上的等可能分布，而$\zeta=\xi\eta$。求证ζ,ξ,η两两独立而不相互独立。

对离散型随机变量和连续型随机变量，我们有

定理 2.5.4. 设X,Y为离散型随机变量，其所有可能取值为$\{(x_i,y_j):i,j=1,2,\cdots\}$。则$X,Y$相互独立当且仅当

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)\quad\forall\, i,j=1,2,\cdots$$

一般地，

离散型随机变量$X_1,\cdots,X_n$相互独立，当且仅当它们的联合分布律等于各自的边际分布律的乘积，即

$$P(X_1=x_1,\ldots,X_n=x_n)=P(X_1=x_1)\cdots P(X_n=x_n)\quad forall\ \ (x_1,\ldots,x_n)$$

证明: 充分性显然，下证必要性。由于X,Y相互独立，所以有

$$F(x,y) = F_1(x)F_2(y)$$

因此对任意的取值对(x_i,y_j)(不妨假定我们已把X,Y的可能取值从小到大排列)，有

$$\begin{aligned} P(X=x_i,Y=y_j) &= F(x_i,y_j) - F(x_{i-1},y_i) - F(x_i,y_{j-1}) + F(x_{i-1},y_{j-1}) \\ &= [F_1(x_i) - F_1(x_{i-1})]F_2(y_j) - [F_1(x_i) - F_1(x_{i-1})]F_2(y_{j-1}) \\ &= P(X=x_i)P(Y=y_j) \end{aligned}$$

从而定理得证。对n的随机变量的独立性等价性条件用归纳法可证。

定理 2.5.5. 设X,Y为连续型随机变量，则X,Y相互独立当且仅当它们的联合概率密度等于各自的边际概率密度的乘积，即

$$f(x,y) = f_1(x)f_2(y), \quad \forall (x,y) \in \mathbb{R}^2$$

一般地，

> 连续型随机变量$X_1,\cdots,X_n$相互独立，当且仅当它们的联合密度等于各自的边际密度的乘积，即
>
> $$f(x_1,\ldots,x_n) = f_1(x_1)\cdots f_n(x_n), \quad \forall\ (x_1,\ldots,x_n) \in \mathbb{R}^n$$

证明显然。

例2.5.6. 证明二元正态随机变量相互独立的充要条件是$\rho = 0$。

2.6 随机变量的函数的分布

1. 离散型随机变量的情形

设X的分布律为

$$P(X=x_i) = p_i, \quad i = 1,2,\cdots$$

$g: R \to R$，令$Y = g(X)$，则Y的分布律为

$$P(Y=y_j) = P(g(X)=y_j) = \sum_{x_i: g(x_i)=y_j} P(X=x_i) = \sum_{i: g(x_i)=y_j} p_i$$

例2.6.1. 设X的概率函数为

X	-1	0	1	2
P	1/4	1/2	1/8	1/8

试求$Y=X^2,\quad Z=X^3+1$的分布律。

上述结论可以推广到多维随机变量的情形。

设随机向量X的分布律为$P(X=x)$，则X的函数$Y=g(X)$的分布律为

$$P(Y=y)=P(g(X)=y)=\sum_{x:g(x)=y}P(X=x)$$

特别当ξ,η是相互独立的非负整值随机变量，各有分布$\{a_k\}$与$\{b_k\}$. 那么$\xi+\eta$有分布

$$P(\xi+\eta=n)=\sum_{k=0}^{n}a_kb_{n-k}$$

称此公式为**离散卷积公式**

例2.6.2. 设$(X_1,\cdots,X_n)$服从多项分布$M(N;p_1,\ldots,p_n)$，求$Y=X_1+X_2$的分布律。

解：可以由多项分布的定义直接得到结果。下面我们直接计算。由于(X_1,X_2)服从三项分布$M(N;p_1,p_2,1-p_1-p_2)$，所以

$$\begin{aligned}P(Y=k) &= P(X_1+X_2=k)=\sum_{i=0}^{k}P(X_1=i,X_2=k-i)\\ &= \sum_{i=0}^{k}\frac{N!}{i!(k-i)!(N-k)!}p_1^ip_2^{k-i}(1-p_1-p_2)^{N-k}\\ &= \binom{N}{k}(p_1+p_2)^k(1-p_1-p_2)^{N-k},\quad k=0,1,\cdots,N.\end{aligned}$$

即$Y\sim B(N,p_1+p_2)$.

例2.6.3. 设$X\sim B(n,p)$，$Y\sim B(m,p)$且X和Y相互独立，则$X+Y\sim B(n+m,p)$。这种性质称为再生性。

2. 连续型随机变量的情形

定理 2.6.6. [**密度变换公式**] *设$r.v.X$有pdf $f(x)$ $x\in(a,b)$(a,b可以∞), 而$y=g(x)$在$x\in(a,b)$上是严格单调的连续函数，存在唯一的反函数$x=h(y),y\in(\alpha,\beta)$并且$h'(y)$存在且连续，那么$Y=g(X)$也是连续型$r.v.$且有$pdf$*

$$p(y)=f(h(y))|h'(y)|,\quad y\in(\alpha,\beta).$$

例2.6.4. 设$r.v.\ X \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$，求$Y = tgX$的$pdf$。

由密度变换公式知Y的pdf为

$$f(y) = \frac{1}{\pi} arctg'(y) = \frac{1}{\pi(1+y^2)}, \quad -\infty < y < \infty$$

此分布称为Cauchy分布。Cauchy分布的概率密度一般形式为

$$p(x) = \frac{\lambda^2}{\lambda^2 + (x-\mu)^2}, \quad \lambda \neq 0, -\infty < \mu < \infty.$$

本题我们也可以用一般的方法求解，即先求出分布函数，然后对分布函数求导数得到。

$$\begin{aligned} F(y) &= P(Y \leq y) = P(tg(X) \leq y) \\ &= P(X \leq arctg(y)) = \int_{-\frac{\pi}{2}}^{arctg(y)} \frac{1}{\pi} dy = \frac{1}{\pi} arctg(y) + \frac{1}{2}. \end{aligned}$$

所以Y的概率密度为

$$f(y) = F'(y) = \frac{1}{\pi(1+y^2)}.$$

这种方法更具有一般性。

当g不是在全区间上单调而是逐段单调时，密度变换公式为

定理 2.6.7. *设随机变量ξ的密度函数为$p_\xi(x)$，$a < x < b$. 如果可以把(a,b)分割为一些(有限个或可列个)子区间$(a,b) = \bigcup_j I_j$，使得函数$u = g(t)$，$t \in (a,b)$在每个子区间上有唯一的反函数$h_j(u)$，并且$h_j^{'}(u)$存在连续，则$\eta = g(\xi)$是连续型随机变量，其密度函数为：*

$$p_\eta(x) = \sum_j p_\xi(h_j(x))|h_j^{'}(x)| \ . \tag{2.6.1}$$

例2.6.5. 设$X \sim N(0,1)$，求$Y = X^2$的概率密度。

解：由于函数$y = x^2$在$(-\infty, 0)$和$[0, \infty)$上严格单调，因此由上述定理知Y的概率密度为

$$\begin{aligned} f(y) &= \phi(-\sqrt{y})|-\sqrt{y}'|I_{\{y>0\}} + \phi(\sqrt{y})|\sqrt{y}'|I_{\{y>0\}} \\ &= \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} I_{\{y>0\}} \end{aligned}$$

例2.6.6. 设$r.v.\ X$有概率密度

$$f(x) = \begin{cases} \frac{2x}{\pi^2} & 0 < x < \pi \\ 0 & otherwise \end{cases}$$

求$Y = sinX$的概率密度。

在多元随机变量场合，我们有

定理 2.6.8. 如果$(\xi_1,\cdots,\xi_n)$是n维连续型随机向量, 具有联合密度函数$p(x_1,\cdots,x_n)$. 假设存在n个$Borel$可测的函数

$$y_j = f_j(x_1,\cdots,x_n), \quad j=1,\cdots,n,$$

使得

$$\zeta_j = f_j(\xi_1,\cdots,\xi_n), \quad j=1,\cdots,n,$$

并且对于$(\zeta_1,\cdots,\zeta_n)$的每一组可能值$(y_1,\cdots,y_n)$, 方程组

$$\left\{\begin{array}{c} y_1 = f_1(x_1,\cdots,x_n) \\ \cdots\cdots\cdots\cdots \\ y_n = f_n(x_1,\cdots,x_n) \end{array}\right. \tag{2.6.2}$$

都有唯一解

$$\left\{\begin{array}{c} x_1 = h_1(y_1,\cdots,y_n) \\ \cdots\cdots\cdots\cdots \\ x_n = h_n(y_1,\cdots,y_n) \end{array}\right. , \tag{2.6.3}$$

其中每个$h_j(y_1,\cdots,y_n)$都有一阶连续偏导数, 那么随机向量$(\zeta_1,\cdots,\zeta_n)$是连续型的, 具有联合密度函数

$$q(y_1,\cdots,y_n) = \left\{\begin{array}{cc} p\left(h_1(y_1,\cdots,y_n),\cdots,h_n(y_1,\cdots,y_n)\right)|J|, & (y_1,\cdots,y_n)\in\mathcal{D}, \\ 0, & (y_1,\cdots,y_n)\notin\mathcal{D}, \end{array}\right. \tag{2.6.4}$$

其中$\mathcal{D}$是随机向量$(\zeta_1,\cdots,\zeta_n)$的所有可能值的集合, $|J|$是变换的$Jaccobi$行列式，即

$$J = \left|\begin{array}{ccc} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{array}\right|$$

证: 由于随机向量$(\zeta_1,\cdots,\zeta_n)$的联合分布为

$$\begin{aligned} &F(z_1,\cdots,z_n) = P\left(\zeta_1\le z_1,\cdots,\zeta_n\le z_n\right) \\ &= \int_{f_1(x_1,\cdots,x_n)\le z_1}\cdots\int_{f_n(x_1,\cdots,x_n)\le z_n} p(x_1,\cdots,x_n)dx_1\cdots dx_n \\ &= \int_{-\infty}^{z_1}\cdots\int_{-\infty}^{z_n} p\left(h_1(y_1,\cdots,y_n),\cdots,h_n(y_1,\cdots,y_n)\right)|J|dy_1\cdots dy_n, \end{aligned}$$

其中最后一步得自变量替换(2.6.3), 所以随机向量$(\zeta_1,\cdots,\zeta_n)$的联合密度函数恰如(2.6.4)式所示. #

例2.6.7. 在直角坐标平面上随机选取一点, 分别以随机变量ξ与η表示其横坐标和纵坐标, 可以认为ξ与η相互独立. 如果ξ与η都服从正态分布$\mathbb{N}(0,1)$, 试求其极坐标(ρ,θ)的分布.

解: 易知

$$\begin{cases} x = r\cos t \\ y = r\sin t \end{cases}$$

是$(0,\infty)\times[0,2\pi)$与$\mathcal{R}^2$(原点除外)之间的一一变换, 变换的Jaccobi行列式

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial t} \end{vmatrix} = \begin{vmatrix} \cos t & -r\sin t \\ \sin t & r\cos t \end{vmatrix} = r\ .$$

由于(ξ,η)的联合密度为

$$p(x,y) = \frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\},$$

所以由(2.6.4)式得知, (ρ,θ)的联合密度为

$$q(r,t) = \frac{1}{2\pi} r\exp\left\{-\frac{r^2}{2}\right\}, \quad r>0,\ \ t\in[0,2\pi). \tag{2.6.5}$$

这一结果表明: θ与ρ相互独立, 其中θ服从$[0,2\pi)$ 上的均匀分布; 而ρ的边缘密度函数为

$$q_1(r) = r\exp\left\{-\frac{r^2}{2}\right\}, \quad r>0. \quad \#$$

在计算两个随机变量之和时，我们还经常用到如下定理

定理 2.6.9. 设X,Y的联合概率密度为$f(x,y)$，则$X+Y$的概率密度$p(z)$为

$$p(z) = \int_{-\infty}^{\infty} f(x,z-x)dx = \int_{-\infty}^{\infty} f(z-y,y)dy$$

证: 先求$X+Y$的分布函数$F(z)$. 我们有

$$\begin{aligned} F(z) = P(X+Y\le z) = \int\int_{x+y<z} f(x,y)dxdy = \int_{-\infty}^{\infty} dx\int_{-\infty}^{z-x} f(x,y)dy \\ = \int_{-\infty}^{\infty} du\int_{-\infty}^{z} f(x,t-x)dt = \int_{-\infty}^{z}\left\{\int_{-\infty}^{\infty} f(x,t-x)dx\right\}dt. \end{aligned}$$

这就说明,$X+Y$的分布函数$F(z)$是其中的花括弧中的函数在区间$(-\infty,z)$上的积分,所以$X+Y$是连续型随机变量,其密度函数如定理所述。#

特别，

当X与Y独立时，分别记X和Y的概率密度为$f_1(x)$和$f_2(y)$，则$X+Y$的概率密度为

$$p(z) = \int_{-\infty}^{\infty} f_1(x)f_2(z-x)dx = \int_{-\infty}^{\infty} f_1(z-y)f_2(y)dy$$

称此公式为**卷积公式**。

例2.6.8. 设X服从期望为2的指数分布，$Y \sim U(0,1)$，且X和Y相互独立。求$X-Y$的概率密度和$P(X \le Y)$。

解：解法1:

由题设知$-Y \sim U(-1,0)$，并记X和$-Y$的密度分别为f_1和f_2，从而由卷积公式有

$$
\begin{aligned}
f_{X-Y}(z) &= \int_{-\infty}^{\infty} f_1(x) f_2(z-x) dy \\
&= \begin{cases} \int_z^{z+1} f_1(x)dx & z \ge 0 \\ \int_0^{z+1} f_1(x)dx & -1 < z < 0 \\ 0 & otherwise \end{cases} \\
&= \begin{cases} e^{-\frac{z}{2}}(1-e^{-\frac{1}{2}}) & z \ge 0 \\ 1-e^{-\frac{z+1}{2}} & -1 < z < 0 \\ 0 & otherwise \end{cases}
\end{aligned}
$$

所以$P(X \le Y) = P(X-Y \le 0) = 1-e^{-\frac{1}{2}}$。

解法2: 由于

$$
\begin{aligned}
P(X-Y \le z) &= \int P(X \le z+y | Y=y) f(y) dy \\
&= \begin{cases} \int_0^1 P(X \le z+y)dy & z \ge 0 \\ \int_{-z}^1 P(X \le z+y)dy & -1 < z < 0 \\ 0 & otherwise \end{cases} \\
&= \begin{cases} 1-2e^{-z/2}(1-e^{-1/2}) & z \ge 0 \\ z+2e^{-(z+1)/2}-1 & -1 < z < 0 \\ 0 & otherwise \end{cases}
\end{aligned}
$$

再对分布函数求导数即得所求.

对一些连续型随机变量，也有再生性性质。

例2.6.9. 设$X \sim N(a,\sigma_1^2)$, $Y \sim N(b,\sigma_2^2)$且X与Y相互独立，则$X+Y \sim N(a+b, \sigma_1^2+\sigma_2^2)$.

我们把具有再生性性质的分布总结一下为

- 二项分布(关于试验次数具有再生性)
- $Poisson$分布(关于参数λ具有再生性)
- $Pascal$分布(关于成功次数r具有再生性)
- 正态分布(关于两个参数都具有再生性)
- 具有再生性的连续型分布还有χ^2分布和Γ分布

有时我们还会碰到计算随机变量之商的pdf. 我们有

定理 2.6.10. 如果(ξ,η)是二维连续型随机向量, 它们的联合密度为$p(x,y)$, 则它们的商$\frac{\xi}{\eta}$是连续型随机变量, 具有密度函数

$$p_{\frac{\xi}{\eta}}(x)=\int_{-\infty}^{\infty}|t|p(xt,t)dt=\int_{-\infty}^{\infty}|u|p(u,xu)du,\quad \forall\, x\in\mathcal{R}. \tag{2.6.6}$$

证: 我们来求$\frac{\xi}{\eta}$的分布函数$F(x)$. 我们有

$$\begin{aligned}
F(x)=P\left(\frac{\xi}{\eta}\le x\right)&=\int\int_{\frac{u}{v}\le x}p(u,v)dudv\\
&=\int\int_{u\le xv,\ v>0}p(u,v)dudv+\int\int_{u\ge xv,\ v<0}p(u,v)dudv\\
&=\int_0^{\infty}dv\int_{-\infty}^{xv}p(u,v)du+\int_{-\infty}^{0}dv\int_{xv}^{\infty}p(u,v)du\\
&=\int_0^{\infty}dv\int_{-\infty}^{x}v\,p(tv,v)dt+\int_{-\infty}^{0}dv\int_{x}^{\infty}v\,p(tv,v)dt\\
&=\int_{-\infty}^{x}\left\{\int_{-\infty}^{\infty}|v|p(tv,v)dv\right\}dt.
\end{aligned}$$

这就说明,$\frac{\xi}{\eta}$的分布函数$F(x)$是其中的花括弧中的函数在区间$(-\infty,x)$上的积分,所以$\frac{\xi}{\eta}$是连续型随机变量,其密度函数如(2.6.6)中的第一式所示; 改变积分顺序, 可得(2.6.6)中的第二式. #

例2.6.10. 设随机变量ξ与η相互独立,同服从参数$\lambda=1$的指数分布, 试求$\frac{\xi}{\eta}$的密度函数.

解: 我们利用(2.6.6)式求$p_{\frac{\xi}{\eta}}(x)$.由于(ξ,η) 的联合密度为

$$p(u,v)=e^{-u-v},\quad u>0,\ v>0,$$

所以欲(2.6.6)式中的被积函数$|t|p(xt,t)\neq 0$, 当且仅当, $t>0$和$xt>0$, 从而知有

$$p_{\frac{\xi}{\eta}}(x)=\begin{cases}\int_0^{\infty}t\,e^{-xt-t}dt=\frac{1}{(1+x)^2} & x>0;\\ 0 & x\le 0.\end{cases}\qquad \#$$

3. 极小值和极大值的分布

对于定义在同一个概率空间$(\Omega, \mathcal{F}, P)$上的n个随机变量$\xi_1, ..., \xi_n$,我们可以考察它们的最大值和最小值:

$$\eta_1 = \max\{\xi_1, ..., \xi_n\},$$

$$\eta_2 = \min\{\xi_1, ..., \xi_n\}.$$

其含义是:

$$\eta_1(\omega) = \max\{\xi_1(\omega), ..., \xi_n(\omega)\}, \ \ \omega \in \Omega,$$

$$\eta_2(\omega) = \min\{\xi_1(\omega), ..., \xi_n(\omega)\}, \ \ \omega \in \Omega.$$

如此定义的η_1与η_2当然也是随机变量.事实上, 我们有

$$\begin{aligned}(\eta_1 \le x) &= (\max\{\xi_1, \ \dots, \xi_n\} \le x) \\ &= (\xi_1 \le x, \ \dots, \xi_n \le x) = \bigcap_{k=1}^{n} (\xi_k \le x) \in \mathcal{F},\end{aligned} \tag{2.6.7}$$

$$(\eta_2 \le x) = (\min\{\xi_1, \ \dots, \xi_n\} \le x) = \bigcup_{k=1}^{n} (\xi_k \le x) \in \mathcal{F}. \tag{2.6.8}$$

当$\xi_1, ..., \xi_n$相互独立时, 我们不难利用它们的分布函数$F_1(x), ..., F_n(x)$求出η_1与η_2 的分布函数$F_{\eta_1}(x)$和$F_{\eta_2}(x)$.

事实上, 此时由(2.6.7), 我们立即得到

$$\begin{aligned}F_{\eta_1}(x) &= P(\eta_1 \le x) = P\left(\bigcap_{k=1}^{n} (\xi_k \le x)\right) \\ &= \prod_{k=1}^{n} P(\xi_k \le x) = \prod_{k=1}^{n} F_k(x);\end{aligned} \tag{2.6.9}$$

而利用关系式

$$(\eta_2 > x) = (\xi_1 > x, \ \dots, \xi_n > x) = \bigcap_{k=1}^{n} (\xi_k > x)$$

可得

$$\begin{aligned}F_{\eta_2}(x) &= P(\eta_2 \le x) = 1 - P(\eta_2 > x) = 1 - P\left(\bigcap_{k=1}^{n} (\xi_k > x)\right) \\ &= 1 - \prod_{k=1}^{n} P(\xi_k > x) = 1 - \prod_{k=1}^{n} (1 - F_k(x)).\end{aligned} \tag{2.6.10}$$

比较有趣的是考察(η_1, η_2)的联合分布. 对此我们仅给出独立同分布场合下的结果.

例2.6.11. 如果$\xi_1, \cdots, \xi_n$是独立同分布的随机变量, 它们的共同分布为$F(x)$. 试求它们的最大值η_1与最小值η_2的联合分布. 如果进一步, 假设$F(x)$为连续型分布,具有密度函数$p(x)$, 试求(η_1, η_2)的联合密度.

解: 我们以$G(x,y)$表示(η_1, η_2)的联合分布. 由(2.6.7)和(2.6.8)式, 容易看出: 当$x \le y$时, 有

$$G(x,y) = P(\eta_1 \le x,\ \eta_2 \le y) = P(\eta_1 \le x) = P\left(\bigcap_{k=1}^{n}(\xi_k \le x)\right) = F^n(x);$$

当$x > y$时, 有

$$\begin{aligned} G(x,y) &= P(\eta_1 \le x,\ \eta_2 \le y) = P(\eta_1 \le x) - P(\eta_1 \le x,\ \eta_2 > y) \\ &= P\left(\bigcap_{k=1}^{n}(\xi_k \le x)\right) - P\left(\bigcap_{k=1}^{n}(y < \xi_k \le x)\right) \\ &= F^n(x) - (F(x) - F(y))^n. \end{aligned}$$

所以(η_1, η_2)的联合分布为

$$G(x,y) = \begin{cases} F^n(x) - \{F(x) - F(y)\}^n, & x > y, \\ F^n(x), & x \le y. \end{cases} \tag{2.6.11}$$

当$F(x)$为连续型分布,具有密度函数$p(x)$时, 我们以$q(x,y)$表示(η_1, η_2) 的联合密度, 由(2.6.11)式, 知有

$$q(x,y) = \frac{\partial^2}{\partial x \partial y} G(x,y) = n(n-1)\{F(x) - F(y)\}^{n-2} p(x)p(y), \quad x > y. \tag{2.6.12}$$

2.7 总结与讨论

目前我们接触到的分布的关系为

- n个独立同分布$B(1,p)$的0-1分布随机变量之和为二项分布$B(n,p)$;
- 有限个独立二项随机变量(成功的概率相同)之和仍为二项分布;
- 有限个独立的$Poisson$分布随机变量之和服从$Poisson$分布，参数相加；
- r个独立同分布几何分布$G(p)$的随机变量之和服从参数为r和p的$Pascal$分布；
- 任意有限个独立的正态分布随机变量的线性组合仍然服从正态分布;

更详细的常见分布之间的关系，可以参考网页上的文件。

参考文献

[1] 杨振明., 概率论, 南开大学数学教学丛书. 北京: 科学出版社, 2001.

[2] 苏淳., 概率论, 北京: 科学出版社, 2004.

[3] 陈希孺., 概率论与数理统计., 合肥: 中国科学技术大学出版社, 1995.