第二章　随机变量及其分布

上一章研究内容：

　　事件（集合 A）$\longrightarrow$ 概率（数）。

本章将用函数研究概率，函数是数与数的关系，即需要用数反映事件——随机变量。

　　事件（数）$\longrightarrow$ 概率（数）。

§2.1　随机变量及其分布

2.1.1.　随机变量（Random Variable）

随机试验的样本点有些是定量的：如掷骰子掷出的点数，电子元件使用寿命的小时数。有些是定性的：如掷硬币正面或反面，检查产品合格或不合格。对于定性的结果也可以规定其数量性质：如掷硬币，正面记为 1，反面记为 0；检查产品，合格记为 1，不合格记为 0。

随机试验中，可将每一个样本点 ω 都对应于一个实数 $X(\omega)$，在一定条件下该实值函数称为随机变量，常用大写英文字母 X, Y, Z 等表示随机变量，而随机变量的具体取值通常记为小写英文字母 x, y, z。

定义　设 $(\Omega, \mathscr{F}, P)$ 是一个概率空间，样本空间 Ω 到实数集 R 的一个映射 $X:\Omega \to R$，如果对于 R 上的任一 Borel 可测集 B（或者任一开区间），关于映射 X 的原像 $X^{-1}(B) = \{\omega \mid X(\omega) \in B\}$ 都属于 Ω 上的事件域 $\mathscr{F}$，则称之为概率空间 $(\Omega, \mathscr{F}, P)$ 上的一个随机变量。

可见随机变量 X 是样本空间 Ω 到实数集 R 的一个映射 $X:\Omega \to R$，$\omega \to X(\omega)$，且对于 R 上的任一 Borel 可测集 B，根据 X 映射到 B 中的所有样本点的集合都是一个事件（Ω 上事件域 $\mathscr{F}$ 中的元素）。

对于随机变量首先应掌握它的全部可能取值，如

（1）掷硬币，出现正面，记 $X = 1$；出现反面，记 $X = 0$。X 的全部可能取值为 0, 1。

（2）掷两枚骰子，X 表示掷出的点数之和，X 的全部可能取值为 $2, 3, 4, \cdots, 12$。

（3）观察某商店一小时内的进店人数 X，X 的全部可能取值为 $0, 1, 2, 3, \cdots$。

（4）电子元件使用寿命，用 X 表示使用的小时数，X 的全部可能取值为 $[0, +\infty)$。

（5）一场足球比赛（90 分钟），用 X 表示首次进球时间（分钟），若为 $0:0$，记 $X = 100$，X 的全部可能取值为 $(0, 90) \cup \{100\}$。

注： 每个样本点都必须对应于一个实数；不同样本点可以对应于同一个实数；随机变量的每一取值或取值范围都表示一个事件（其原像属于事件域 $\mathscr{F}$）。

应掌握将随机变量的取值或取值范围描述为事件，又能将事件用随机变量表达的方法。

如掷一枚骰子，用 X 表示出现的点数，则 $X = 1$ 表示出现 1 点；$X > 4$ 表示点数大于 4，即出现 5 点或 6 点；$X \le 0$ 为不可能事件；出现奇数点，即 $X = 1, 3, 5$；点数不超过 3，即 $X \le 3$。

X 表示商店一天中某商品的销售件数（顾客的需求件数），则 $X = 0$ 表示没有销售；$X \le 10$ 表示销售不超过 10 件；销售 5 件以上（不含 5 件），即 $X > 5$；若该商店准备了 a 件该商品，事件“能满足顾客需要”，即 $X \le a$。

X 表示一只电子元件的使用寿命（小时），则 $X = 1000$ 表示该元件恰好使用了 1000 小时；$X \ge 800$ 表示该元件使用寿命在 800 小时以上。

90 分钟足球比赛，X 表示首次进球时间（分钟），且 $0:0$ 时，记 $X = 100$，则 $X = 10$ 表示上半场第 10 分钟首次进球；上半场不进球，即 $X > 45$；开场 1 分钟内进球，即 $X \le 1$。

如果随机变量 X 的全部可能取值是有限个或可列无限个，则称为离散随机变量（Discrete Random Variable）。

注： 可列无限个，即可以排成一列，一个一个往下数，如非负整数 $0, 1, 2, 3, \cdots$。

离散随机变量的全部可能取值是实数轴上一些离散的点，而连续随机变量的全部可能取值是实数轴上一个区间或多个区间的并，如电子元件使用寿命 X（小时），全部可能取值是 $[0, +\infty)$。

下面按离散和连续两种主要情况分别进行讨论。

第二章　随机变量及其分布

上一章研究内容：

事件（集合 A ）⟶ 概率（数）。

本章将用函数研究概率，函数是数与数的关系，即需要用数反映事件——随机变量。

事件（数）⟶ 概率（数）。

§2.1　随机变量及其分布

2.1.1.　随机变量（Random Variable）

随机试验的样本点有些是定量的：如掷骰子掷出的点数，电子元件使用寿命的小时数。有些是定性的：如掷硬币正面或反面，检查产品合格或不合格。对于定性的结果也可以规定其数量性质：如掷硬币，正面记为 1，反面记为 0；检查产品，合格记为 1，不合格记为 0。

随机试验中，可将每一个样本点 ω 都对应于一个实数 $X(\omega)$，在一定条件下该实值函数称为随机变量，常用大写英文字母 X, Y, Z 等表示随机变量，而随机变量的具体取值通常记为小写英文字母 x, y, z 。

定义　设 $(\Omega, \mathscr{F}, P)$ 是一个概率空间，样本空间 Ω 到实数集 R 的一个映射 $X: \Omega \to R$，如果对于 R 上的任一 Borel 可测集 B（或者任一开区间），关于映射 X 的原像 $X^{-1}(B) = \{\omega \mid X(\omega) \in B\}$ 都属于 Ω 上的事件域 $\mathscr{F}$，则称之为概率空间 $(\Omega, \mathscr{F}, P)$ 上的一个随机变量。

可见随机变量 X 是样本空间 Ω 到实数集 R 的一个映射 $X: \Omega \to R$，$\omega \to X(\omega)$，且对于 R 上的任一 Borel 可测集 B，根据 X 映射到 B 中的所有样本点的集合都是一个事件（Ω 上事件域 $\mathscr{F}$ 中的元素）。

对于随机变量首先应掌握它的全部可能取值，如

（1）掷硬币，出现正面，记 $X = 1$；出现反面，记 $X = 0$ 。X 的全部可能取值为 0, 1。

（2）掷两枚骰子，X 表示掷出的点数之和，X 的全部可能取值为 $2, 3, 4, \cdots, 12$ 。

（3）观察某商店一小时内的进店人数 X，X 的全部可能取值为 $0, 1, 2, 3, \cdots$ 。

（4）电子元件使用寿命，用 X 表示使用的小时数，X 的全部可能取值为 $[0, +\infty)$ 。

（5）一场足球比赛（90 分钟），用 X 表示首次进球时间（分钟），若为 $0:0$，记 $X = 100$，X 的全部可能取值为 $(0, 90) \cup \{100\}$。

注：每个样本点都必须对应于一个实数；不同样本点可以对应于同一个实数；随机变量的每一取值或取值范围都表示一个事件（其原像属于事件域 $\mathscr{F}$ ）。

应掌握将随机变量的取值或取值范围描述为事件，又能将事件用随机变量表达的方法。

如掷一枚骰子，用 X 表示出现的点数，则 $X = 1$ 表示出现 1 点；$X > 4$ 表示点数大于 4，即出现 5 点或 6 点；$X \le 0$ 为不可能事件；出现奇数点，即 $X = 1, 3, 5$；点数不超过 3，即 $X \le 3$ 。

X 表示商店一天中某商品的销售件数（顾客的需求件数），则 $X = 0$ 表示没有销售；$X \le 10$ 表示销售不超过 10 件；销售 5 件以上（不含 5 件），即 $X > 5$；若该商店准备了 a 件该商品，事件“能满足顾客需要”，即 $X \le a$ 。

X 表示一只电子元件的使用寿命（小时），则 $X = 1000$ 表示该元件恰好使用了 1000 小时；$X \ge 800$ 表示该元件使用寿命在 800 小时以上。

90 分钟足球比赛，X 表示首次进球时间（分钟），且 $0:0$ 时，记 $X = 100$，则 $X = 10$ 表示上半场第 10 分钟首次进球；上半场不进球，即 $X > 45$；开场 1 分钟内进球，即 $X \le 1$。

如果随机变量 X 的全部可能取值是有限个或可列无限个，则称为离散随机变量（Discrete Random Variable）。

注：可列无限个，即可以排成一列，一个一个往下数，如非负整数 $0, 1, 2, 3, \cdots$ 。

离散随机变量的全部可能取值是实数轴上一些离散的点，而连续随机变量的全部可能取值是实数轴上一个区间或多个区间的并，如电子元件使用寿命 X（小时），全部可能取值是 $[0, +\infty)$ 。

下面按离散和连续两种主要情况分别进行讨论。

2.1.2. 离散随机变量的概率分布（Probability Distribution of Discrete Random Variable）

对于随机变量还应该掌握它的每一取值或取值范围表示事件的概率。

定义 如果随机变量 X 的全部可能取值是有限个或可列个，则称为离散随机变量。设离散随机变量 X 的全部可能取值为 $x_1, x_2, \cdots, x_k, \cdots$，则 X 取值 x_k 的概率 $P\{X = x_k\} = p_k$， $k = 1, 2, \cdots$ 称为离散随机变量 X 的概率分布列，简称分布列。直观上又写为

$$\begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_k & \cdots \\ \hline P & p_1 & p_2 & \cdots & p_k & \cdots \end{array} \quad 或 \quad X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix},$$

其函数形式 $P\{X = x_k\} = p_k$， $k = 1, 2, \cdots$ 称为概率质量函数（Probability Mass Function）。

如掷一枚骰子， X 表示出现的点数，其分布列为

$$\begin{array}{c|cccccc} X & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline P & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array}$$

分布列或概率质量函数的基本性质：

（1）非负性， $p(x_k) \ge 0, \quad k = 1, 2, \cdots$；

（2）正则性， $\sum\limits_{k=1}^{\infty} p(x_k) = 1$。

正则性成立是因为事件 $X = x_1, X = x_2, \cdots, X = x_k, \cdots$ 是样本空间 Ω 的一个分割，故

$$p(x_1) + p(x_2) + \cdots + p(x_k) + \cdots = P\{X = x_1\} + P\{X = x_2\} + \cdots + P\{X = x_k\} + \cdots = P(\Omega) = 1 \text{。}$$

例 设盒中有 2 个红球 3 个白球，从中任取 3 球，以 X 表示取得的红球数，求 X 的概率分布列。

解： X 的全部可能取值 $0, 1, 2$，样本点总数为 $n = C_5^3 = 10$。

$X = 0$ 表示“取到 3 个白球”，所含样本点个数为 $k_0 = C_3^3 = 1$，有 $p_0 = \frac{1}{10} = 0.1$； $X = 1$ 表示“取到 1 个红球 2 个白球”，所含样本点个数为 $k_1 = C_3^2 C_2^1 = 6$，有 $p_1 = \frac{6}{10} = 0.6$； $X = 2$ 表示“取到 2 个红球 1 个白球”，所含样本点个数为 $k_2 = C_3^1 C_2^2 = 3$，有 $p_2 = \frac{3}{10} = 0.3$。

故 X 的分布列为

$$\begin{array}{c|ccc} X & 0 & 1 & 2 \\ \hline P & 0.1 & 0.6 & 0.3 \end{array}$$

求离散随机变量 X 的概率分布列步骤：

（1）找出 X 的全部可能取值；

（2）将 X 的每一取值表示为事件；

（3）求出 X 的每一取值的概率。

例 现有 10 件产品，其中有 3 件不合格。若不放回抽取，每次取一件，直到取得合格品为止。用 X 表示抽取次数，求 X 的概率分布列。

解： X 的全部可能取值 $1, 2, 3, 4$。

$X=1$表示“第 1 次就取得合格品”，有 $p_1=\dfrac{7}{10}$；$X=2$表示“第 2 次取得合格品且第 1 次是不合格品”，有 $p_2=\dfrac{3}{10}\cdot\dfrac{7}{9}=\dfrac{7}{30}$；$X=3$表示“第 3 次取得合格品且前两次是不合格品”，有 $p_3=\dfrac{3}{10}\cdot\dfrac{2}{9}\cdot\dfrac{7}{8}=\dfrac{7}{120}$；$X=3$表示“第 4 次取得合格品且前三次是不合格品”，有 $p_4=\dfrac{3}{10}\cdot\dfrac{2}{9}\cdot\dfrac{1}{8}\cdot\dfrac{7}{7}=\dfrac{1}{120}$。

故 X 的分布列为

X	1	2	3	4
P	$\dfrac{7}{10}$	$\dfrac{7}{30}$	$\dfrac{7}{120}$	$\dfrac{1}{120}$

例 上例若改为有放回地抽取，又如何？

解： X 的全部可能取值$1,2,\cdots,n,\cdots$，有

$$p_1=\frac{7}{10}=0.7,\quad p_2=\frac{3}{10}\cdot\frac{7}{10}=0.21,\quad p_3=0.3^2\times0.7,\ \cdots,\quad p_k=0.3^{k-1}\times0.7,\ \cdots。$$

故 X 的概率质量函数为 $p(k)=0.3^{k-1}\times0.7,\quad k=1,2,\cdots$；其分布列为

X	1	2	3	$\cdots$	k	$\cdots$
P	0.7	0.3×0.7	$0.3^2\times0.7$	$\cdots$	$0.3^{k-1}\times0.7$	$\cdots$

例 若离散随机变量 X 的概率质量函数为 $p(k)=\dfrac{C}{k},\quad k=1,2,3,4$，且 C 为常数。求：（1）C 的值；（2）$P\{X=3\}$；（3）$P\{X<3\}$。

解：（1）由正则性知：$p_1+p_2+p_3+p_4=C+\dfrac{C}{2}+\dfrac{C}{3}+\dfrac{C}{4}=1$，即 $\dfrac{25}{12}C=1$，故 $C=\dfrac{12}{25}$；

（2）$P\{X=3\}=p_3=\dfrac{4}{25}$；

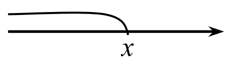
（3）$P\{X<3\}=p_1+p_2=\dfrac{12}{25}+\dfrac{6}{25}=\dfrac{18}{25}$。

2.1.3. 随机变量的分布函数（Cumulative Distribution Function）

连续随机变量在单个点取值概率为零，如电子元件使用寿命恰好为 1000 小时这个事件的概率就等于零，因此连续随机变量不能考虑概率分布。为了用单独一个变量表示一个区间，特别地取区间$(-\infty,x]$。

定义 随机变量 X 与任意实数 x，称

$$F(x)=P\{X\le x\},\quad -\infty<x<+\infty$$



为 X 的累积分布函数，简称分布函数。

随机变量 X 在任一区间上或任一点取值的概率都可用分布函数表示，如

$$P\{a<X\le b\}=P\{X\le b\}-P\{X\le a\}=F(b)-F(a),$$

$$P\{X>a\}=1-P\{X\le a\}=1-F(a),$$

$$P\{X<a\}=\lim_{x\to a^-}P\{X\le x\}=\lim_{x\to a^-}F(x)=F(a-0),$$

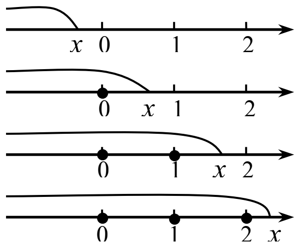
$$P\{X=a\}=P\{X\le a\}-P\{X<a\}=F(a)-F(a-0)。$$

例 已知随机变量 X 的分布列为

X	0	1	2
P	0.2	0.5	0.3

求 X 的分布函数。

解： X 的全部可能取值为$0,1,2$。



当$x<0$时，$F(x)=P\{X\le x\}=P(\varnothing)=0$；

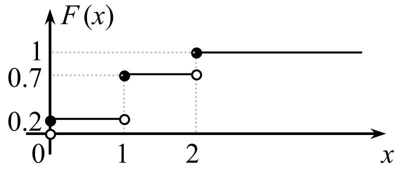
当$0\le x<1$时，$F(x)=P\{X\le x\}=P\{X=0\}=0.2$；

当$1\le x<2$时，$F(x)=P\{X\le x\}=P\{X=0\}+P\{X=1\}=0.7$；

当$x\ge 2$时，$F(x)=P\{X\le x\}=P(\Omega)=1$；

故

$$F(x)=\begin{cases}0, & x<0;\\ 0.2, & 0\le x<1;\\ 0.7, & 1\le x<2;\\ 1, & x\ge 2.\end{cases}$$



若离散随机变量X的全部可能取值为$x_1,x_2,\cdots$，概率分布为$P\{X=x_k\}=p_k$，$k=1,2,\cdots$，则分布函数为$F(x)=P\{X\le x\}=\sum\limits_{x_k\le x}p(x_k)$。且离散随机变量的分布函数$F(x)$是单调不减的阶梯形函数，$X$的每一可能取值$x_k$是$F(x)$的跳跃点，跳跃高度是相应概率$p_k$。

例 已知某离散随机变量X的分布函数为

$$F(x)=\begin{cases}0, & x<-1\\ 0.3, & -1\le x<0,\\ 0.4, & 0\le x<2,\\ 0.6, & 2\le x<5,\\ 1, & x\ge 5,\end{cases}$$

求X的概率分布列。

解： X的全部可能取值是$F(x)$的跳跃点，即$-1,0,2,5$，且跳跃高度依次为：

$P\{X=-1\}=0.3-0=0.3$，$P\{X=0\}=0.4-0.3=0.1$，

$P\{X=2\}=0.6-0.4=0.2$，$P\{X=5\}=1-0.6=0.4$，

故X的分布列为

X	-1	0	2	5
P	0.3	0.1	0.2	0.4

分布函数的基本性质：

（1）单调性（Monotonicity），$F(x)$单调不减，即$x_1<x_2$时，$F(x_1)\le F(x_2)$；

（2）正则性（Regularity），$F(-\infty)=0$，$F(+\infty)=1$；

（3）右连续性（Right-continuity），$F(x)$右连续，即$\lim\limits_{x\to x_0^+}F(x)=F(x_0)$。

证明：（1）当$x_1<x_2$时，$\{X\le x_1\}\subset\{X\le x_2\}$，由概率的单调性可得$F(x_1)\le F(x_2)$；

（2）可以理解为$F(-\infty)=P\{X\le-\infty\}=P(\varnothing)=0$，$F(+\infty)=P\{X<+\infty\}=P(\Omega)=1$；

（3）任取单调下降且趋于x_0的数列$\{x_n\}$，有$\lim\limits_{n\to+\infty}\{X\le x_n\}=\bigcap\limits_{n=1}^{+\infty}\{X\le x_n\}=\{X\le x_0\}$，根据概率的连续性知$\lim\limits_{n\to+\infty}P\{X\le x_n\}=P\left(\bigcap\limits_{n=1}^{+\infty}\{X\le x_n\}\right)=P\{X\le x_0\}$，即$\lim\limits_{x\to x_0^+}F(x)=F(x_0)$。

注：分布函数 $F(x)$ 不一定左连续。

任取单调增加且趋于 x_0 的数列 $\{x_n\}$，有 $\lim\limits_{n\to+\infty}\{X\le x_n\}=\bigcup\limits_{n=1}^{+\infty}\{X\le x_n\}=\{X<x_0\}$，根据概率的连续性知

$\lim\limits_{n\to+\infty}P\{X\le x_n\}=P\left(\bigcup\limits_{n=1}^{+\infty}\{X\le x_n\}\right)=P\{X<x_0\}$，即 $\lim\limits_{x\to x_0^-}F(x)=P\{X<x_0\}=F(x_0)-P\{X=x_0\}$。

2.1.4.　连续随机变量的概率密度函数（Probability Density Function of Continuous Random Variable）

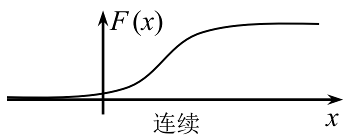
离散随机变量的全部可能取值是有限或可列个点，连续随机变量的全部可能取值是实数区间。但连续随机变量在单独一个点取值的概率为 0，其概率质量函数无实际意义，对于连续随机变量通常只是考虑其在某些区间上取值的概率，这就需要根据分布函数进行讨论。

注：概率为 0 的事件不一定是不可能事件。

定义　随机变量 X 的分布函数 $F(x)$，若存在函数 $p(x)$，使

$$F(x)=\int_{-\infty}^{x}p(u)du,$$

则称 X 为连续随机变量，$p(x)$ 为 X 的概率密度函数，简称密度函数。

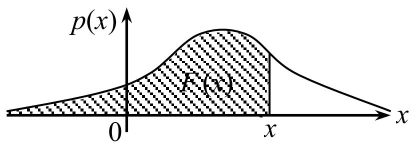


可以理解为 $p(u)$ 为概率密度，$p(u)du$ 为 X 在该小区间内取值的概率，$\int_{-\infty}^{x}$ 为从 $-\infty$ 到 x 无限求和。几何意义为在平面上作出密度函数 $p(x)$ 的图形，则阴影部分的面积即为 $F(x)$ 的值。

概率密度函数基本性质：

（1）非负性，$p(x)\ge 0$；

（2）正则性，$\int_{-\infty}^{+\infty}p(x)dx=1$。

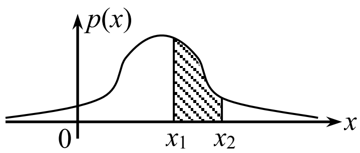


正则性成立是因为 $\int_{-\infty}^{x}p(u)du=F(x)$，有 $\int_{-\infty}^{+\infty}p(x)dx=F(+\infty)=1$。

连续随机变量的性质：设连续随机变量 X 的概率密度函数为 $p(x)$，分布函数为 $F(x)$，则有

（1）$P\{x_1<X\le x_2\}=F(x_2)-F(x_1)=\int_{x_1}^{x_2}p(x)dx$；

（2）当 $p(x)$ 连续时，$p(x)=F'(x)$；



这是因为 $F(x)=\int_{-\infty}^{x}p(u)du$，当 $p(x)$ 连续时，有 $F'(x)=[\int_{-\infty}^{x}p(u)du]'=p(x)$。

（3）X 在单独一个点取值的概率为 0，其分布函数没有跳跃点，为连续函数；

（4）$P\{x_1<X\le x_2\}=P\{x_1\le X\le x_2\}=P\{x_1<X<x_2\}=P\{x_1\le X<x_2\}$，即连续随机变量在某区间内取值的概率与区间开闭无关，而离散随机变量则不成立；

（5）只在有限个点上取值不相同的密度函数对应于同一个分布函数。

例　设 $F(x)=A+B\arctan x$ 为连续随机变量 X 的分布函数。求：（1）参数 A,B；（2）$P\{-1\le X\le\sqrt{3}\}$；（3）密度函数 $p(x)$。

解：（1）由分布函数的正则性 $F(-\infty)=0$，$F(+\infty)=1$，得：

$$\lim_{x\to-\infty}(A+B\arctan x)=A-\frac{\pi}{2}B=0,\quad \lim_{x\to+\infty}(A+B\arctan x)=A+\frac{\pi}{2}B=1,$$

故 $A=\dfrac{1}{2}$，$B=\dfrac{1}{\pi}$；

（2）$F(x)=\frac{1}{2}+\frac{1}{\pi}\arctan x$，得

$$P\{-1\le X\le\sqrt{3}\}=F(\sqrt{3})-F(-1)=\left(\frac{1}{2}+\frac{1}{\pi}\cdot\frac{\pi}{3}\right)-\left[\frac{1}{2}+\frac{1}{\pi}\cdot\left(-\frac{\pi}{4}\right)\right]=\frac{7}{12}。$$

（3）密度函数 $p(x)=F'(x)=\frac{1}{\pi(1+x^2)}$。

例 已知

$$p(x)=\begin{cases}C(x^2-x^3), & 0<x<1;\\ 0, & \text{其它}.\end{cases}$$

是连续随机变量 X 的密度函数，求：（1）C；（2）$P\{-1<X<\frac{1}{2}\}$；（3）分布函数 $F(x)$。

解：（1）由正则性：$\int_{-\infty}^{+\infty}p(x)dx=1$，得

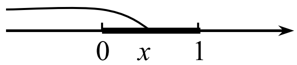
$$\int_0^1 C(x^2-x^3)dx=C\left(\frac{x^3}{3}-\frac{x^4}{4}\right)\Bigg|_0^1=C\left(\frac{1}{3}-\frac{1}{4}\right)-0=\frac{C}{12}=1，$$

故 $C=12$；

（2）$P\{-1<X<\frac{1}{2}\}=\int_{-1}^{\frac{1}{2}}p(x)dx=\int_0^{\frac{1}{2}}12(x^2-x^3)dx=12\left(\frac{x^3}{3}-\frac{x^4}{4}\right)\Bigg|_0^{\frac{1}{2}}=12\left(\frac{1}{24}-\frac{1}{64}\right)=\frac{5}{16}$；

（3）X 的全部可能取值为 $[0,1]$，分段点 0, 1。

当 $x<0$ 时，$F(x)=\int_{-\infty}^{x}p(u)du=0$；



当 $0\le x<1$ 时，$F(x)=\int_{-\infty}^{x}p(u)du=\int_0^x 12(u^2-u^3)du=12\left(\frac{u^3}{3}-\frac{u^4}{4}\right)\Bigg|_0^x=4x^3-3x^4$；

当 $x\ge 1$ 时，$F(x)=\int_{-\infty}^{x}p(u)du=\int_0^1 12(u^2-u^3)du=1$；

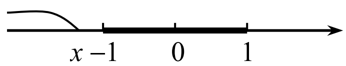
故

$$F(x)=\begin{cases}0, & x<0;\\ 4x^3-3x^4, & 0\le x<1;\\ 1, & x\ge 1.\end{cases}$$

例 已知

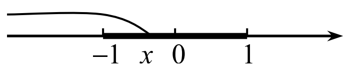
$$p(x)=\begin{cases}|x|, & -1<x<1;\\ 0, & \text{其它}.\end{cases}$$

是连续随机变量 X 的密度函数，求分布函数 $F(x)$。



解：分段点 $-1,0,1$。

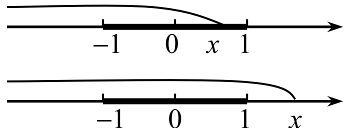
当 $x<-1$ 时，$F(x)=\int_{-\infty}^{x}p(u)du=0$；



当 $-1\le x<0$ 时，$F(x)=\int_{-\infty}^{x}p(u)du=\int_{-1}^{x}(-u)du=-\frac{u^2}{2}\Bigg|_{-1}^{x}=-\frac{x^2}{2}+\frac{1}{2}=\frac{1-x^2}{2}$；

当 $0 \le x < 1$ 时，$F(x)=\int_{-\infty}^{x} p(u)du=\int_{-1}^{0}(-u)du+\int_{0}^{x} udu=-\left.\frac{u^2}{2}\right|_{-1}^{0}+\left.\frac{u^2}{2}\right|_{0}^{x}=0+\frac{1}{2}+\frac{x^2}{2}=\frac{1+x^2}{2}$；

当 $x \ge 1$ 时，$F(x)=\int_{-\infty}^{x} p(u)du=\int_{-1}^{0}(-u)du+\int_{0}^{1} udu=1$；



故

$$F(x)=\begin{cases}0, & x<0;\\ \dfrac{1-x^2}{2}, & -1\le x<0;\\ \dfrac{1+x^2}{2}, & 0\le x<1;\\ 1, & x\ge 1.\end{cases}$$

注：（1）如果连续随机变量 X 的密度函数形式为

$$p(x)=\begin{cases}f(x), & a<x<b;\\ 0, & \text{其它}.\end{cases}$$

则称密度函数 $p(x)$ 取值非零的区间 (a,b) 为随机变量 X 的支撑区间，且密度函数还可简写为

$$p(x)=f(x),\quad a<x<b,$$

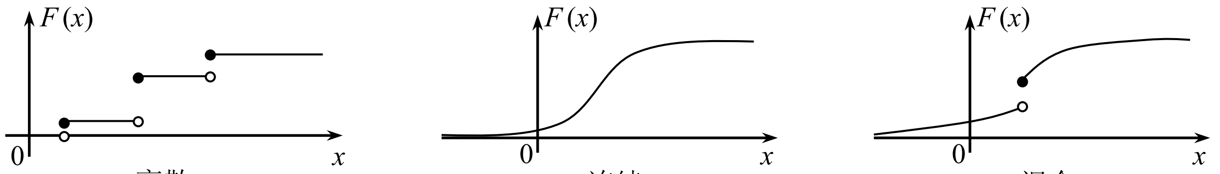
或

$$p(x)=f(x)I_{a<x<b},$$

其中 $I_{a<x<b}$ 为示性函数（Indicator Function），表示在区间 (a,b) 内取值为 1，其它情况取值为 0。一般地，示性函数

$$I_{x\in A}=\begin{cases}1, & x\in A;\\ 0, & x\notin A.\end{cases}$$

（2）对于一个给定的函数，只要满足分布函数的三条基本性质：单调性、正则性、右连续性，都可以成为随机变量的分布函数。离散随机变量的分布函数是单调不减且右连续的阶梯形函数，连续随机变量的分布函数是单调不减的连续函数（不可导的点最多可列个），它们都满足单调性、正则性、右连续性。除了这两种情况外，也还有满足这三条性质的其它情况。如单调不减右连续但不左连续也非阶梯形的函数（不可导的点最多可列个），以这种函数为分布函数的随机变量既有离散成分（间断且跳跃增加）又有连续成分（连续且光滑增加），称为混合随机变量（Mixing Random Variable）。此外还有单调不减且处处连续处处不可导的函数，以这种函数为分布函数的随机变量称为奇异随机变量（Singular Random Variable）。



离散　　　　连续　　　　混合

（3）离散随机变量的概率质量函数、连续随机变量的概率密度函数、所有随机变量共有的累积分布函数，统称为概率分布函数（Probability Distribution Function），简称分布（Distribution）。后面讨论某随机变量的分布，可以是指累积分布函数，也可以是指概率质量函数或概率密度函数。

§2.2 随机变量的数学期望

对于随机变量，还应当掌握反映其平均值、分散程度等的指标，这就需要引入数学期望和方差等概念。

2.2.1. 数学期望（Expectation）

例 甲、乙两个射击选手，在射击训练中甲射了 10 次，其中 3 次 10 环，1 次 9 环，4 次 8 环，2 次 7 环；乙射了 15 次，其中 2 次 10 环，9 次 9 环，2 次 8 环，2 次 7 环。问谁的表现更好？

分析：比较他们射中的平均环数。

甲共射中 $3\times10+1\times9+4\times8+2\times7=85$ 环，平均每次射中 $\dfrac{85}{10}=8.5$ 环；

乙共射中 $2\times10+9\times9+2\times8+2\times7=131$ 环，平均每次射中 $\dfrac{131}{15}\approx8.73$ 环；

故可以认为乙的表现更好。

一般地，若在 n 次试验中，出现了 m_1 次 x_1，m_2 次 x_2，…，m_k 次 x_k，（其中 $m_1+m_2+\cdots+m_k=n$），则平均值为 $\dfrac{m_1x_1+m_2x_2+\cdots+m_kx_k}{n}=\sum\limits_{i=1}^{k}x_i\dfrac{m_i}{n}$，即平均值等于取值与频率乘积之和。

因 n 很大时，频率稳定在概率附近，即平均值将稳定在取值与概率乘积之和附近。

2.2.2. 数学期望的定义

定义 设离散随机变量 X 的分布列是

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

如果级数 $\sum\limits_{k=1}^{\infty}x_kp_k$ 绝对收敛，则称之为 X 的数学期望，记为 $E(X)$。

数学期望的实际意义是反映随机变量的平均取值，是其全部可能取值以相应概率为权数的加权平均。

如 X 的分布列为

X	-2	0	1	4
P	0.3	0.1	0.4	0.2

则 $E(X)=(-2)\times0.3+0\times0.1+1\times0.4+4\times0.2=0.6$。

例 某人有 4 发子弹，现在他向某一目标射击，若命中目标就停止射击，否则直到子弹用完为止。设每发子弹命中率为 0.4，以 X 表示射击次数，求 $E(X)$。

解：先求 X 的分布列，X 全部可能取值为 1, 2, 3, 4。

$X=1$ 表示第一枪就命中，$P\{X=1\}=0.4$；$X=2$ 表示第一枪没有命中，第二枪命中，$P\{X=2\}=0.6\times0.4=0.24$；$X=3$ 表示前两枪没有命中，第三枪命中，$P\{X=3\}=0.6^2\times0.4=0.144$；$X=4$ 表示前三枪没有命中，$P\{X=4\}=0.6^3=0.216$。

X 的分布列为

X	1	2	3	4
P	0.4	0.24	0.144	0.216

故 $E(X)=1\times0.4+2\times0.24+3\times0.144+4\times0.216=2.176$。

例 若 X 的概率质量函数为 $p\left(\frac{(-2)^k}{k}\right)=\frac{1}{2^k}$， $k=1,2,\cdots$，求 $E(X)$。

解： 因 $\sum_{k=1}^{\infty}\frac{(-2)^k}{k}\cdot\frac{1}{2^k}=\sum_{k=1}^{\infty}\frac{(-1)^k}{k}$ 收敛但不是绝对收敛，故 $E(X)$ 不存在。

离散随机变量的数学期望是取值乘概率求和：$\sum_{k=1}^{\infty}x_k p_k$，类似可定义连续随机变量的数学期望是取值乘密度积分：$\int_{-\infty}^{+\infty}xp(x)dx$。

定义 设连续随机变量 X 的密度函数为 $p(x)$。如果反常积分 $\int_{-\infty}^{+\infty}xp(x)dx$ 绝对收敛，则称之为 X 的数学期望，记为 $E(X)$。

例 已知连续随机变量 X 的密度函数为

$$p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & \text{其它}.\end{cases}$$

求 $E(X)$。

解： $E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_0^1 x\cdot 2xdx=2\cdot\left.\frac{x^3}{3}\right|_0^1=\frac{2}{3}$。

例 已知 X 的密度函数为

$$p(x)=\begin{cases}a+bx, & 0<x<2;\\ 0, & \text{其它}.\end{cases}$$

且 $E(X)=\frac{2}{3}$，求 a,b。

解： 由正则性得

$$\int_{-\infty}^{+\infty}p(x)dx=\int_0^2(a+bx)dx=\left.\left(ax+b\cdot\frac{x^2}{2}\right)\right|_0^2=2a+2b=1,$$

且

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_0^2 x(a+bx)dx=\left.\left(a\cdot\frac{x^2}{2}+b\cdot\frac{x^3}{3}\right)\right|_0^2=2a+\frac{8}{3}b=\frac{2}{3},$$

故 $a=1,\ b=-\frac{1}{2}$。

例 已知 X 的密度函数为

$$p(x)=\frac{1}{\pi(1+x^2)},\quad -\infty<x<+\infty,$$

求 $E(X)$。

解： 因

$$\int_{-\infty}^{+\infty}xp(x)dx=\int_{-\infty}^{+\infty}\frac{x}{\pi(1+x^2)}dx=\int_{-\infty}^{+\infty}\frac{1}{\pi(1+x^2)}\cdot\frac{1}{2}d(x^2)=\frac{1}{2\pi}\ln(1+x^2)\Big|_{-\infty}^{+\infty}\text{发散},$$

故 $E(X)$ 不存在。

2.2.3. 数学期望的性质

设 X 为随机变量，$g(x)$ 为函数，则称 $Y=g(X)$ 为随机变量函数，Y 也是一个随机变量。下面不加证明地给出随机变量函数的数学期望计算公式。

定理 设 X 为随机变量，$Y=g(X)$ 为随机变量函数，则

（1）若 X 为离散随机变量，概率质量函数为 $p(x_k)=p_k$，$k=1,2,\cdots$，则 $E(Y)=\sum_{k=1}^{\infty}g(x_k)p(x_k)$；

（2）若 X 为连续随机变量，密度函数为 $p(x)$，则 $E(Y)=\int_{-\infty}^{+\infty}g(x)p(x)dx$。

数学期望具有以下性质：

（1）常数的期望等于其自身，即 $E(c)=c$；

（2）常数因子可移到期望符号外，即 $E(aX)=aE(X)$；

（3）随机变量和的期望等于期望的和，即 $E[g_1(X)+g_2(X)]=E[g_1(X)]+E[g_2(X)]$。

证明：（1）将常数 c 看作是单点分布 $P\{X=c\}=1$，故 $E(c)=c\cdot 1=c$；

（2）以连续为例加以证明，$E(aX)=\int_{-\infty}^{+\infty}axp(x)dx=a\int_{-\infty}^{+\infty}xp(x)dx=aE(X)$；

（3）以连续为例加以证明，

$$E[g_1(X)+g_2(X)]=\int_{-\infty}^{+\infty}[g_1(x)+g_2(x)]p(x)dx=\int_{-\infty}^{+\infty}g_1(x)p(x)dx+\int_{-\infty}^{+\infty}g_2(x)p(x)dx$$

$$=E[g_1(X)+g_2(X)]。$$

由性质（2）、（3）知随机变量线性组合的期望等于期望的线性组合，可见数学期望具有线性性质。

例 设 X 的分布列为

X	-1	0	1	2
P	0.2	0.1	0.4	0.3

求 $E(2X+1)$，$E(X^2)$。

解：所求期望为

$$E(2X+1)=(-1)\times 0.2+1\times 0.1+3\times 0.4+5\times 0.3=2.6；$$

$$E(X^2)=1\times 0.2+0\times 0.1+1\times 0.4+4\times 0.3=1.8。$$

例 已知圆的半径 X 是一个随机变量，密度函数为

$$p(x)=\begin{cases}\frac{1}{2}, & 1<x<3;\\ 0, & 其他.\end{cases}$$

求圆面积 Y 的数学期望。

解：圆面积 $Y=\pi X^2$，故

$$E(Y)=\int_{-\infty}^{+\infty}\pi x^2p(x)dx=\int_1^3\pi x^2\cdot\frac{1}{2}dx=\frac{\pi}{2}\cdot\frac{x^3}{3}\bigg|_1^3=\frac{13\pi}{3}。$$

例 设国际市场对我国某种出口商品的需求量 X（吨）的密度函数为

$$p(x)=\begin{cases}\frac{1}{2000}, & 2000<x<4000;\\ 0, & 其他.\end{cases}$$

设每售出一吨，可获利 3 万美元，但若销售不出，每积压一吨将亏损 1 万美元，问如何计划年出口量，能使国家获利的期望最大。

解：设计划年出口量为 a 吨，每年获利 Y 万美元。当 $X \ge a$ 时，销售 a 吨，获利 $3a$ 万美元；当 $X < a$ 时，销售 X 吨，积压 $a-X$ 吨，获利 $3X-(a-X)=4X-a$ 万美元；即

$$Y = g(X) = \begin{cases} 3a, & a \le X \le 4000; \\ 4X-a, & 2000 \le X < a. \end{cases}$$

则

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} g(x)p(x)dx = \int_{2000}^{a} (4x-a)\cdot\frac{1}{2000}dx + \int_{a}^{4000} 3a\cdot\frac{1}{2000}dx \\ &= \frac{1}{2000}(2x^2-ax)\Big|_{2000}^{a} + \frac{3a}{2000}x\Big|_{a}^{4000} \\ &= -\frac{1}{1000}a^2 + 7a - 4000 = -\frac{1}{1000}(a-3500)^2 + 8250, \end{aligned}$$

故计划年出口量为 3500 吨时，使国家获利的期望最大。

§2.3 随机变量的方差与标准差

数学期望反映平均值，方差反映波动程度。

如甲、乙两台包装机，要求包装重量为每袋 500 克，现各取 5 袋，重量分布为

甲：498，499，500，501，502；

乙：490，495，500，505，510。

二者平均值相同都是 500 克，但显然甲比乙好。此时比较的是它们的偏差（即取值与平均值之差）。偏差分布为

甲：–2，–1，0，1，2；

乙：–10，–5，0，5，10。

2.3.1. 方差与标准差（Variance & Standard Deviation）

随机变量X与其数学期望$E(X)$之差$X-E(X)$称为偏差。偏差有大有小，可正可负，比较时需要去掉符号，但绝对值函数进行微积分处理不方便，因此考虑偏差平方的数学期望。

定义 随机变量X，若$E[X-E(X)]^2$存在，则称之为X的方差（Variance），记为$\mathrm{Var}(X)$或$D(X)$。即

$$\mathrm{Var}(X)=E[X-E(X)]^2。$$

显然方差$\mathrm{Var}(X)\ge 0$，称$\sqrt{\mathrm{Var}(X)}$为X的标准差（Standard Deviation）。在实际问题中，标准差与随机变量有相同的量纲。

方差与标准差反映波动程度。方差越大，取值越分散；方差越小，取值越集中。

例 设X的分布列为

X	1	2	3
P	0.2	0.4	0.4

求期望$E(X)$，方差$\mathrm{Var}(X)$。

解：期望与方差分别为

$$E(X)=1\times 0.2+2\times 0.4+3\times 0.4=2.2；$$

$$\mathrm{Var}(X)=(-1.2)^2\times 0.2+(-0.2)^2\times 0.4+0.8^2\times 0.4=0.56。$$

例 已知X的密度函数为

$$p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$$

求期望$E(X)$，方差$\mathrm{Var}(X)$。

解：期望与方差分别为

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_0^1 x\cdot 2xdx=2\cdot\frac{x^3}{3}\bigg|_0^1=\frac{2}{3}；$$

$$\mathrm{Var}(X)=\int_{-\infty}^{+\infty}(x-\frac{2}{3})^2p(x)dx=\int_0^1(2x^3-\frac{8}{3}x^2+\frac{8}{9}x)dx=\left(\frac{2}{4}x^4-\frac{8}{9}x^3+\frac{4}{9}x^2\right)\bigg|_0^1=\frac{1}{2}-\frac{8}{9}+\frac{4}{9}=\frac{1}{18}。$$

例 已知X的全部可能取值为 0, 1, 2，且$E(X)=1.3$，$\mathrm{Var}(X)=0.81$。求X的分布。

解：设X的分布列为

X	0	1	2
P	a	b	c

由正则性可得$a+b+c=1$，且

$$E(X)=0\times a+1\times b+2\times c=b+2c=1.3,$$

$$\mathrm{Var}(X)=(-1.3)^2\times a+(-0.3)^2\times b+0.7^2\times c=1.69a+0.09b+0.49c=0.81,$$

解得$a=0.3, b=0.1, c=0.6$，故X的分布列为

X	0	1	2
P	0.3	0.1	0.6

2.3.2. 方差的性质

方差具有以下性质：

（1）方差计算公式，$\mathrm{Var}(X)=E(X^2)-[E(X)]^2$；

（2）常数的方差等于零，即$\mathrm{Var}(c)=0$；

（3）设a,b为常数，则$\mathrm{Var}(aX+b)=a^2\,\mathrm{Var}(X)$。

证明：（1）由方差的定义得

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

$$=E(X^2)-2E(X)\cdot E(X)+[E(X)]^2=E(X^2)-[E(X)]^2;$$

（2）$\mathrm{Var}(c)=E[c-E(c)]^2=E(c-c)^2=E(0)=0$；

（3）$\mathrm{Var}(aX+b)=E[aX+b-E(aX+b)]^2=E[aX+b-aE(X)-b]^2=a^2E[X-E(X)]^2=a^2\,\mathrm{Var}(X)$。

由性质（1），显然有以下推论：

推论 对于随机变量X，如果$E(X^2)$存在，则$E(X^2)\geq[E(X)]^2$。

以后常利用方差计算公式$\mathrm{Var}(X)=E(X^2)-[E(X)]^2$计算随机变量的方差。通常用公式计算比直接用定义计算方差要方便。

由性质（3）可得$\mathrm{Var}(aX)=a^2\,\mathrm{Var}(X)$且$\mathrm{Var}(X+b)=\mathrm{Var}(X)$，可知方差具有平方性质以及平移不变性。

例 设X的分布列为

X	1	2	3
P	0.2	0.4	0.4

求方差$\mathrm{Var}(X)$。

解：前面已求得$E(X)=2.2$，又因$E(X^2)=0^2\times 0.2+1^2\times 0.4+2^2\times 0.4=5.4$，故

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=5.4-2.2^2=0.56。$$

例　已知 X 的密度函数为

$$p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$$

求方差 $\mathrm{Var}(X)$。

解：前面已求得 $E(X)=\dfrac{2}{3}$，又因 $E(X^2)=\int_0^1 x^2\cdot 2xdx=2\cdot\dfrac{x^4}{4}\Big|_0^1=\dfrac{1}{2}$，故

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}\text{。}$$

对于随机变量 X，若方差 $\mathrm{Var}(X)$ 存在，且 $\mathrm{Var}(X)>0$。令

$$X^*=\frac{X-E(X)}{\sqrt{\mathrm{Var}(X)}}\text{，}$$

有

$$E(X^*)=E\left(\frac{X-E(X)}{\sqrt{\mathrm{Var}(X)}}\right)=\frac{1}{\sqrt{\mathrm{Var}(X)}}E[X-E(X)]=\frac{1}{\sqrt{\mathrm{Var}(X)}}[E(X)-E(X)]=0\text{；}$$

$$\mathrm{Var}(X^*)=\mathrm{Var}\left(\frac{X-E(X)}{\sqrt{\mathrm{Var}(X)}}\right)=\frac{1}{\mathrm{Var}(X)}\mathrm{Var}[X-E(X)]=\frac{1}{\mathrm{Var}(X)}\mathrm{Var}(X)=1\text{。}$$

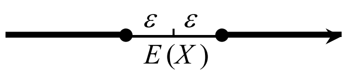
称 X^* 为 X 的标准化随机变量。

2.3.3.　切比雪夫不等式（Chebyshev Inequality）

方差反映随机变量的分散程度，切比雪夫不等式给出其定量标准。切比雪夫不等式表明大偏差概率的上限与方差成正比。

定理　设 X 为随机变量，且方差 $\mathrm{Var}(X)$ 存在，则对于任何正数 ε，都有

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{\mathrm{Var}(X)}{\varepsilon^2}\text{。}$$



证明：以连续随机变量为例证明，设 X 的密度函数为 $p(x)$，有

$$P\{|X-E(X)|\geq\varepsilon\}=\int_{|x-E(X)|\geq\varepsilon}p(x)dx \text{ 且 } \frac{\mathrm{Var}(X)}{\varepsilon^2}=\frac{1}{\varepsilon^2}E[X-E(X)]^2=\int_{-\infty}^{+\infty}\frac{[x-E(X)]^2}{\varepsilon^2}p(x)dx\text{，}$$

故

$$P\{|X-E(X)|\geq\varepsilon\}\leq\int_{|x-E(X)|\geq\varepsilon}\frac{[x-E(X)]^2}{\varepsilon^2}p(x)dx\leq\int_{-\infty}^{+\infty}\frac{[x-E(X)]^2}{\varepsilon^2}p(x)dx=\frac{\mathrm{Var}(X)}{\varepsilon^2}\text{。}$$

注：切比雪夫不等式的等价形式

$$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{\mathrm{Var}(X)}{\varepsilon^2}\text{。}$$

如随机变量 X 的数学期望为 $E(X)=10$，$\mathrm{Var}(X)=1$，则由切比雪夫不等式可得

$$P\{8<X<12\}=P\{|X-10|<2\}\geq 1-\frac{1}{2^2}=\frac{3}{4}\text{。}$$

例　设随机变量 X 的全部可能取值为 $[0,+\infty)$，且数学期望 $E(X)$ 存在，试证：对任何正数 a，都有

$$P\{X \ge a\} \le \frac{1}{a} E(X) 。$$

证明：以连续随机变量为例证明，设 X 的密度函数为 $p(x)$，有

$$P\{X \ge a\} = \int_a^{+\infty} p(x)dx，\quad \frac{1}{a}E(X) = \frac{1}{a}\int_{-\infty}^{+\infty} xp(x)dx = \int_0^{+\infty} \frac{x}{a} p(x)dx，$$

故

$$P\{X \ge a\} \le \int_a^{+\infty} \frac{x}{a} p(x)dx \le \int_0^{+\infty} \frac{x}{a} p(x)dx = \frac{1}{a}E(X) 。$$

定理　设随机变量 X 的方差存在，则 $\mathrm{Var}(X) = 0$ 的充分必要条件是存在常数 b，使得 X 几乎处处收敛于 b，即 $P\{X = b\} = 1$。

证明：充分性，设存在常数 b，使得 $P\{X = b\} = 1$，有 $P\{X \ne b\} = 0$，即 $E(X) = bP\{X = b\} = b$，故

$$\mathrm{Var}(X) = E[X - E(X)]^2 = E(X - b)^2 = 0 \times P\{X = b\} = 0；$$

必要性，设 X 的方差 $\mathrm{Var}(X) = 0$。因事件

$$\{|X - E(X)| > 0\} = \bigcup_{n=1}^{+\infty}\left\{|X - E(X)| \ge \frac{1}{n}\right\} = \lim_{n\to+\infty}\left\{|X - E(X)| \ge \frac{1}{n}\right\}，$$

则

$$P\{|X - E(X)| > 0\} = P\left(\bigcup_{n=1}^{+\infty}\left\{|X - E(X)| \ge \frac{1}{n}\right\}\right) = \lim_{n\to+\infty} P\left\{|X - E(X)| \ge \frac{1}{n}\right\} \le \lim_{n\to+\infty}\frac{\mathrm{Var}(X)}{\left(\frac{1}{n}\right)^2} = 0，$$

可得 $P\{|X - E(X)| > 0\} = 0$，即 $P\{|X - E(X)| = 0\} = 1$，取 $b = E(X)$，有 b 为常数，故 $P\{X = b\} = 1$。

注：如果 $P\{X = b\} = 1$，记为 $X = b$，a.e.（或 a.s.），称为 $X = b$ 几乎处处成立（或几乎必然成立）。这里，a.e.就是 almost everywhere 的缩写，a.s.就是 almost surely 的缩写。意味着不成立的情况是一个测度（或概率）等于零的集合（或事件）。

§2.4　常用离散分布

对于一个给定的函数，只要满足概率质量函数的两条基本性质：非负性、正则性，都可以成为某个离散随机变量的概率质量函数。但绝大多数在实际工作中并不常见，下面是几种常用的离散分布。

2.4.1.　两点分布与二项分布

一．两点分布（Two-point Distribution）

两点分布只可能在两个点取值，通常就是 0 或 1。

定义　随机变量 X 的可能取值只有两个：0 或 1，且分布列为 $P\{X=1\}=p$，$P\{X=0\}=1-p$，其中 $0<p<1$，则称 X 服从两点分布或 0-1 分布，记为 $X\sim(0\text{-}1)$。

两点分布实际背景是一次伯努利试验中事件 A 的发生次数。

基本性质：非负性，$p>0$，$1-p>0$；正则性，$p+(1-p)=1$。

两点分布的数学期望为

$$E(X)=1\times p+0\times(1-p)=p。$$

又因

$$E(X^2)=1^2\times p+0^2\times(1-p)=p，$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=p-p^2=p(1-p)。$$

二．二项分布（Binomial Distribution）

在 n 重伯努利试验中，以 X 表示事件 A 的发生次数，则 A 的全部可能取值为 $0,1,2,\cdots,n$，且概率分布为

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k}。$$

定义　若离散随机变量 X 的概率质量函数为

$$p(k)=C_n^k p^k(1-p)^{n-k},\quad k=0,1,2,\cdots,n；\ 0<p<1，$$

则称 X 服从二项分布，记为 $X\sim b(n,p)$。

二项分布的实际背景是 n 重伯努利试验中事件 A 的发生次数。当 $n=1$ 时，二项分布就是两点分布。

基本性质：非负性，$C_n^k p^k(1-p)^{n-k}>0$；正则性，

$$\sum_{k=1}^{n}C_n^k p^k(1-p)^{n-k}=[p+(1-p)]^n=1。$$

例　掷三枚硬币，X 表示正面朝上的次数，求 X 的分布。

解：X 的全部可能取值为 $0,1,2,3$，将掷每一枚硬币看作一次试验。每次试验两种结果：正面 A，反面 $\overline{A}$；相互独立；概率 $P(A)=0.5$ 不变。n 重伯努利试验，$n=3$，$p=0.5$，有 $X\sim b(3,0.5)$，可得

$$P\{X=0\}=0.5^3=0.125，$$

$$P\{X=1\}=C_3^1\times0.5^1\times0.5^2=0.375，$$

$$P\{X=2\}=C_3^2\times0.5^2\times0.5^1=0.375，$$

$$P\{X=3\}=0.5^3=0.125。$$

例 现有 5 台机床，每台机床一小时内平均开动 18 分钟，且是否开动相互独立，以 X 表示同一时刻开动的机床数，求 X 的分布。

解： X 的全部可能取值为 0, 1, 2, 3, 4, 5，将每台机床是否开动看作一次试验。每次试验两种结果：开动 A，不开动 $\overline{A}$；相互独立；概率 $P(A)=0.3$ 不变。n 重伯努利试验，$n=5$，$p=0.3$，有 $X\sim b(5,0.3)$，可得

$$P\{X=0\}=0.7^5=0.16807,$$

$$P\{X=1\}=C_5^1\times 0.3^1\times 0.7^4=0.36015,$$

$$P\{X=2\}=C_5^2\times 0.3^2\times 0.7^3=0.3087,$$

$$P\{X=3\}=C_5^3\times 0.3^3\times 0.7^2=0.1323,$$

$$P\{X=4\}=C_5^4\times 0.3^4\times 0.7^1=0.02835,$$

$$P\{X=5\}=0.3^5=0.00243。$$

一般地，如果 X 服从二项分布，概率 $P\{X=k\}$ 将随着 k 的增加，先逐渐增加，达到最大值后，又逐渐减少。

三．二项分布的数学期望和方差

组合数公式

$$C_n^k=\frac{n!}{k!\cdot(n-k)!}=\frac{n}{k}\cdot\frac{(n-1)!}{(k-1)!\cdot(n-k)!}=\frac{n}{k}\cdot C_{n-1}^{k-1}，（n\ge k>0）。$$

二项分布 $b(n,p)$ 的数学期望为

$$E(X)=\sum_{k=0}^{n}k\cdot C_n^k p^k(1-p)^{n-k}=\sum_{k=1}^{n}k\cdot\frac{n}{k}\cdot C_{n-1}^{k-1}p^k(1-p)^{n-k}=np\sum_{k=1}^{n}C_{n-1}^{k-1}p^{k-1}(1-p)^{n-k}$$

$$=np[p+(1-p)]^{n-1}=np。$$

又因

$$E(X^2)=\sum_{k=0}^{n}k^2\cdot C_n^k p^k(1-p)^{n-k}=\sum_{k=0}^{n}(k^2-k)\cdot C_n^k p^k(1-p)^{n-k}+\sum_{k=0}^{n}k\cdot C_n^k p^k(1-p)^{n-k}$$

$$=\sum_{k=2}^{n}(k^2-k)\cdot\frac{n(n-1)}{k(k-1)}C_{n-2}^{k-2}p^k(1-p)^{n-k}+E(X)$$

$$=n(n-1)p^2\sum_{k=2}^{n}C_{n-2}^{k-2}p^{k-2}(1-p)^{n-k}+np$$

$$=n(n-1)p^2[p+(1-p)]^{n-2}+np=(n^2-n)p^2+np,$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=(n^2-n)p^2+np-(np)^2=-np^2+np=np(1-p)。$$

2.4.2. 泊松分布（Poisson Distribution）

一．泊松分布

泊松分布是一种理论推导的极限分布。

定义 若随机变量 X 的概率质量函数为

$$p(k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots;\ \ \lambda>0,$$

则称 X 服从参数为 λ 的泊松分布，记为 $X\sim P(\lambda)$。

泊松分布的实际背景是在一段连续的范围内，某种事件随时可能发生，考虑该事件的实际发生次数。如保险事故的发生次数、足球比赛进球数、商店进店人数、电话接听次数等。

基本性质：非负性，$\lambda>0$ 时，$\frac{\lambda^k}{k!}e^{-\lambda}>0$；正则性，

$$\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda}\cdot e^{-\lambda}=1。$$

泊松分布求概率时，可以查附表 1。

例 已知一场足球比赛的进球数 X 服从参数 $\lambda=2.5$ 的泊松分布，求比分为 $0:0$，$1:0$ 以及总进球数超过 5 个的概率。

解：因 $X\sim P(2.5)$，有

$$P\{X=k\}=\frac{2.5^k}{k!}e^{-2.5},\quad k=0,1,2,\cdots。$$

比分 $0:0$，$P\{X=0\}=\frac{2.5^0}{0!}e^{-2.5}=e^{-2.5}=0.082$；

比分 $1:0$，$P\{X=1\}=\frac{2.5^1}{1!}e^{-2.5}=2.5e^{-2.5}=0.287-0.082=0.205$；

总进球数超过 5 个，$P\{X>5\}=\sum_{k=6}^{+\infty}\frac{2.5^k}{k!}e^{-2.5}=1-\sum_{k=0}^{5}\frac{2.5^k}{k!}e^{-2.5}=1-0.958=0.042$。

例 已知某公用电话每小时内打电话的人数 X 服从参数为 $\lambda=8$ 的泊松分布。求某一小时内无人打电话的概率，恰有 10 人打电话的概率，至少有 10 人打电话的概率。

解：因 $X\sim P(8)$，有

$$P\{X=k\}=\frac{8^k}{k!}e^{-8},\quad k=0,1,2,\cdots。$$

无人打电话，$P\{X=0\}=\frac{8^0}{0!}e^{-8}=e^{-8}=0.0003$，

恰有 10 人打电话，$P\{X=10\}=\frac{8^{10}}{10!}e^{-8}=0.816-0.717=0.099$，

至少有 10 人打电话，$P\{X\ge 10\}=\sum_{k=10}^{+\infty}\frac{8^k}{k!}e^{-8}=1-\sum_{k=0}^{9}\frac{8^k}{k!}e^{-8}=1-0.717=0.283$。

例 已知某商店一天中某种贵重商品的销售件数 X 服从泊松分布 $P(7)$，问该商店每天应该准备多少件该商品才能以 99.9%以上的概率满足顾客需要？

解：设准备了 a 件该商品，$X\sim P(7)$，有

$$P\{X=k\}=\frac{7^k}{k!}e^{-7},\quad k=0,1,2,\cdots。$$

事件“满足顾客需要”，即 $X\le a$，有 $P\{X\le a\}\ge 0.999$，故查表可得 $a=16$。

二．泊松分布的数学期望和方差

泊松分布 $P(\lambda)$ 的数学期望为

$$E(X)=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\mathrm{e}^{-\lambda}=\lambda\,\mathrm{e}^{-\lambda}\cdot\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda\,\mathrm{e}^{-\lambda}\cdot\mathrm{e}^{\lambda}=\lambda\text{，}$$

即泊松分布的参数 λ 反映平均发生次数。又因

$$E(X^2)=\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\sum_{k=0}^{\infty}(k^2-k)\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}+\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\lambda^2\,\mathrm{e}^{-\lambda}\cdot\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+E(X)$$

$$=\lambda^2\,\mathrm{e}^{-\lambda}\cdot\mathrm{e}^{\lambda}+\lambda=\lambda^2+\lambda\text{，}$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\lambda^2+\lambda-(\lambda)^2=\lambda\text{。}$$

三．二项分布的泊松近似

二项分布与泊松分布的实际背景都是反映发生次数问题。下面的定理说明了二者之间的联系，泊松分布是二项分布的一种极限分布。

定理 （泊松定理）在 n 重伯努利试验中，记事件 A 在每次试验中发生的概率为与试验次数 n 有关的数 p_n，如果当 $n\to+\infty$ 时，有 $np_n\to\lambda$，则

$$\lim_{n\to+\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\text{，}$$

证明：记 $\lambda_n=np_n$，有 $\lim\limits_{n\to+\infty}\lambda_n=\lambda$，因

$$(1-p_n)^{n-k}=\left(1-\frac{\lambda_n}{n}\right)^{n-k}=\left(1+\frac{-\lambda_n}{n}\right)^{\frac{n}{-\lambda_n}\cdot\frac{-\lambda_n(n-k)}{n}}\text{，}$$

且

$$\lim_{n\to+\infty}\left(1+\frac{-\lambda_n}{n}\right)^{\frac{n}{-\lambda_n}}=\mathrm{e}\text{，}\quad\lim_{n\to+\infty}\frac{-\lambda_n(n-k)}{n}=-\lambda\text{，}$$

则

$$\lim_{n\to+\infty}(1-p_n)^{n-k}=\lim_{n\to+\infty}\left(1+\frac{-\lambda_n}{n}\right)^{\frac{n}{-\lambda_n}\cdot\frac{-\lambda_n(n-k)}{n}}=\mathrm{e}^{-\lambda}\text{，}$$

又因

$$C_n^k=\frac{n(n-1)\cdots(n-k+1)}{k!}=\frac{n^k}{k!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\text{，且}\lim_{n\to+\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1\text{，}$$

故

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

$$=\lim_{n\to+\infty}\frac{(np_n)^k}{k!}\cdot\lim_{n\to+\infty}(1-p_n)^{n-k}\cdot\lim_{n\to+\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=\frac{\lambda^k}{k!}e^{-\lambda}\text{。}$$

此定理表明对于二项分布 $b(n,p)$，当 n 很大，p 很小时，可用泊松分布 $P(\lambda)$ 近似，其中 $\lambda=np$。

例 某地区每年人口意外死亡率为 0.0001，现有 60000 人投保人身意外保险，求一年内因投保人意外死亡恰好赔付 8 人的概率以及赔付不超过 5 人的概率。

解：设 X 表示“一年内因投保人意外死亡而赔付的人数”，$X\sim b(60000,0.0001)$。则

$$P\{X=8\}=C_{60000}^{8}\times 0.0001^{8}\times 0.9999^{59992}，$$

$$P\{X\le 5\}=\sum_{k=0}^{5}C_{60000}^{k}\times 0.0001^{k}\times 0.9999^{60000-k}。$$

但显然计算很繁琐，为便于计算，用泊松分布近似。因 $n=60000$ 很大，$p=0.0001$ 很小，$\lambda=np=6$，有 $X\dot{\sim}P(6)$，故

$$P\{X=8\}\approx\frac{6^{8}}{8!}e^{-6}=0.847-0.744=0.103，$$

$$P\{X\le 5\}\approx\sum_{k=0}^{5}\frac{6^{k}}{k!}e^{-6}=0.446。$$

2.4.3. 超几何分布（Hypergeometric Distribution）

一．超几何分布

N 个元素分成两类，第一类有 M 个，第二类有 $N-M$ 个。从中不放回地任取 n 个，以 X 表示取得第一类的个数。设 X 取值为 k，一方面，显然有 $k\le n$ 且 $k\le M$，即 $k\le\min\{n,M\}$，另一方面，有 $k\ge 0$ 且 $n-k\le N-M$，可得 $k\ge M+n-N$，即 $k\ge\max\{0,M+n-N\}$。

这样 X 的全部可能取值为 $l,l+1,\cdots,r$，其中 $l=\max\{0,M+n-N\}$，$r=\min\{n,M\}$，且

$$P\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}。$$

定义 若随机变量 X 的概率质量函数为

$$p(k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},\quad k=l,l+1,\cdots,r；\quad M<N，\quad n<N，$$

其中 $l=\max\{0,M+n-N\}$，$r=\min\{n,M\}$，则称 X 服从超几何分布，记为 $X\sim h(n,N,M)$。

超几何分布的实际背景是古典概型中的不放回抽样检验问题。

注：有放回检验抽样问题对应的是二项分布。

基本性质：非负性，$\dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}>0$；正则性，

$$\sum_{k=l}^{r}\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}=\frac{1}{C_N^n}\sum_{k=l}^{r}C_M^k C_{N-M}^{n-k}=\frac{C_N^n}{C_N^n}=1。$$

注：比较 $(1+x)^M(1+x)^{N-M}$ 与 $(1+x)^N$ 中 x^n 的系数可以证明 $\sum\limits_{k=l}^{r}C_M^k C_{N-M}^{n-k}=C_N^n$。

例 一袋中有 3 个红球，2 个白球，不放回地取出 3 个球，X 表示取得的红球数。求 X 的分布。

解：不放回抽样，$N=3$，$M=2$，$n=3$，有 $X\sim h(3,5,3)$，则 $\max\{0,M+n-N\}=1$，$\min\{n,M\}=3$，可知 X 的全部可能取值为 1, 2, 3，故

$$P\{X=1\}=\frac{C_3^1C_2^2}{C_5^3}=0.3，\quad P\{X=2\}=\frac{C_3^2C_2^1}{C_5^3}=0.6，\quad P\{X=3\}=\frac{C_3^3C_2^0}{C_5^3}=0.1。$$

二．超几何分布的数学期望和方差

超几何分布 $h(n,N,M)$ 的数学期望为

$$E(X)=\sum_{k=l}^{r}k\cdot\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}=\sum_{k=l}^{r}k\cdot\frac{\frac{M}{k}C_{M-1}^{k-1}C_{N-M}^{n-k}}{\frac{N}{n}C_{N-1}^{n-1}}=\frac{nM}{N}\cdot\sum_{k=l}^{r}\frac{C_{M-1}^{k-1}C_{N-M}^{n-k}}{C_{N-1}^{n-1}}=\frac{nM}{N},$$

又因

$$E(X^2)=\sum_{k=l}^{r}k^2\cdot\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}=\sum_{k=l}^{r}(k^2-k)\cdot\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}+\sum_{k=l}^{r}k\cdot\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$$

$$=\sum_{k=l}^{r}(k^2-k)\cdot\frac{\frac{M(M-1)}{k(k-1)}C_{M-2}^{k-2}C_{N-M}^{n-k}}{\frac{N(N-1)}{n(n-1)}C_{N-2}^{n-2}}+E(X)$$

$$=\frac{n(n-1)M(M-1)}{N(N-1)}\cdot\sum_{k=l}^{r}\frac{C_{M-2}^{k-2}C_{N-M}^{n-k}}{C_{N-2}^{n-2}}+\frac{nM}{N}=\frac{n(n-1)M(M-1)}{N(N-1)}+\frac{nM}{N},$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{nM(n-1)(M-1)}{N(N-1)}+\frac{nM}{N}-\frac{n^2M^2}{N^2}=\frac{nM(N-M)(N-n)}{N^2(N-1)}\text{。}$$

为了便于记忆，可将超几何分布与二项分布的数学期望和方差进行比较。

$b(n,p)$： $E(X)=np$， $\mathrm{Var}(X)=np(1-p)$；

$h(n,N,M)$： $E(X)=n\dfrac{M}{N}$， $\mathrm{Var}(X)=n\dfrac{M}{N}\left(1-\dfrac{M}{N}\right)\dfrac{N-n}{N-1}$；

可见分布 $h(n,N,M)$ 中的 $\dfrac{M}{N}$ 相当于二项分布 $b(n,p)$ 中的 p，方差修正因子为 $\dfrac{N-n}{N-1}$。

三．超几何分布的二项近似

直观上，当抽样个数 n 远小于 M 及 $N-M$ 时，不放回抽样问题可近似看作有放回抽样问题，也就是此时超几何分布可用二项分布近似。

定理 如果当 $N\to+\infty$ 时，$\dfrac{M}{N}\to p,\ \ (0<p<1)$，则

$$\lim_{N\to+\infty}\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}=C_n^kp^k(1-p)^{n-k}\text{。}$$

证明：因

$$C_N^n=\frac{N(N-1)\cdots(N-n+1)}{n!}=\frac{N^n}{n!}\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right),$$

$$C_M^k=\frac{M^k}{k!}\left(1-\frac{1}{M}\right)\cdots\left(1-\frac{k-1}{M}\right),\quad C_{N-M}^{n-k}=\frac{(N-M)^{n-k}}{(n-k)!}\left(1-\frac{1}{N-M}\right)\cdots\left(1-\frac{n-k-1}{N-M}\right),$$

故

$$\lim_{N\to+\infty}\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}=\lim_{N\to+\infty}\frac{\frac{M^k}{k!}\left(1-\frac{1}{M}\right)\cdots\left(1-\frac{k-1}{M}\right)\cdot\frac{(N-M)^{n-k}}{(n-k)!}\left(1-\frac{1}{N-M}\right)\cdots\left(1-\frac{n-k-1}{N-M}\right)}{\frac{N^n}{n!}\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}$$

$$=\frac{n!}{k!(n-k)!}\cdot\lim_{N\to+\infty}\frac{M^k(N-M)^{n-k}}{N^n}\cdot\frac{\left(1-\frac{1}{M}\right)\cdots\left(1-\frac{k-1}{M}\right)\cdot\left(1-\frac{1}{N-M}\right)\cdots\left(1-\frac{n-k-1}{N-M}\right)}{\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}$$

$$=C_n^k\cdot\lim_{N\to+\infty}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}\cdot\lim_{N\to+\infty}\frac{\left(1-\frac{1}{M}\right)\cdots\left(1-\frac{k-1}{M}\right)\cdot\left(1-\frac{1}{N-M}\right)\cdots\left(1-\frac{n-k-1}{N-M}\right)}{\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}$$

$$=C_n^k p^k(1-p)^{n-k}\text{。}$$

此定理表明对于超几何分布 $h(n,N,M)$，当抽样个数 n 远小于 M 及 $N-M$ 时，可用二项分布 $b(n,p)$ 近似，其中 $p=\frac{M}{N}$。

例 某校有 20000 名学生，其中男生 8000 人，女生 12000 人，从中任选 6 人。求选到 2 个男生与 4 个女生的概率。

解： 设 X 表示“选到的男生数”，有 $X\sim h(6,20000,8000)$，可得

$$P\{X=2\}=\frac{C_{8000}^2C_{12000}^4}{C_{20000}^6}\text{。}$$

但显然计算很繁琐，为便于计算，用二项分布近似。因 $n=6$ 较小，远小于 $M=8000$ 与 $N-M=12000$，且 $p=\frac{M}{N}=0.4$，有 $X\dot\sim b(6,0.4)$，故

$$P\{X=2\}\approx C_6^2\times0.4^2\times0.6^4=0.31104\text{。}$$

2.4.4. 几何分布与负二项分布

一．几何分布（Geometric Distribution）

在伯努利试验中，以 X 表示事件 A 首次发生时的试验次数，则 X 的全部可能取值为 $1,2,\cdots$，且

$$P\{X=k\}=(1-p)^{k-1}p\text{。}$$

定义 若随机变量 X 的概率质量函数为

$$p(k)=(1-p)^{k-1}p,\quad k=1,2,\cdots;\quad 0<p<1,$$

则称 X 服从几何分布，记为 $X\sim Ge(p)$。

几何分布的实际背景是事件首次发生时的试验次数。

基本性质：非负性，$(1-p)^{k-1}p>0$；正则性，

$$\sum_{k=1}^{+\infty}(1-p)^{k-1}p=\frac{p}{1-(1-p)}=1\text{。}$$

二．几何分布的数学期望和方差

令 $q=1-p$，有 $P\{X=k\}=q^{k-1}p,\quad k=1,2,\cdots$。

几何分布 $Ge(p)$ 的数学期望为

$$E(X)=\sum_{k=1}^{+\infty}kq^{k-1}p=p\cdot\sum_{k=0}^{+\infty}\frac{d(q^k)}{dq}=p\cdot\frac{d}{dq}\left(\sum_{k=0}^{+\infty}q^k\right)=p\cdot\frac{d}{dq}\left(\frac{1}{1-q}\right)=p\cdot\frac{1}{(1-q)^2}=\frac{1}{p},$$

又因

$$E(X^2)=\sum_{k=1}^{+\infty}k^2q^{k-1}p=\sum_{k=1}^{+\infty}(k^2+k)q^{k-1}p-\sum_{k=1}^{+\infty}kq^{k-1}p=p\cdot\sum_{k=0}^{+\infty}\frac{d^2(q^{k+1})}{dq^2}-E(X)$$

$$=p\cdot\frac{d^2}{dq^2}\left(\sum_{k=0}^{+\infty}q^{k+1}\right)-\frac{1}{p}=p\cdot\frac{d^2}{dq^2}\left(\frac{q}{1-q}\right)-\frac{1}{p}=p\cdot\frac{2}{(1-q)^3}-\frac{1}{p}=\frac{2-p}{p^2},$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{2-p}{p^2}-\left(\frac{1}{p}\right)^2=\frac{1-p}{p^2}。$$

三．几何分布的无记忆性（Memoryless）

定理 设 X 服从几何分布 $Ge(p)$，则对任意正整数 m 与 n 有

$$P\{X>m+n\,|\,X>m\}=P\{X>n\}。$$

证明：因对于正整数 k，有 $P\{X>k\}=\sum_{i=k+1}^{+\infty}(1-p)^{i-1}p=\frac{(1-p)^k p}{1-(1-p)}=(1-p)^k$，故

$$P\{X>m+n\,|\,X>m\}=\frac{P\{X>m+n\}}{P\{X>m\}}=\frac{(1-p)^{m+n}}{(1-p)^m}=(1-p)^n=P\{X>n\}。$$

此定理在已经试验 m 次事件 A 没有发生的条件下，继续试验 n 次仍没有发生的条件概率，等于试验 n 次 A 没有发生的概率。这称之为几何分布的无记忆性。

四．负二项分布（Negative Binomial Distribution）

在伯努利试验中，以 X 表示事件 A 第 r 次发生时的试验次数，则 X 的全部可能取值为 $r, r+1,\cdots$，且

$$P\{X=k\}=C_{k-1}^{r-1}(1-p)^{k-r}p^r。$$

定义 若随机变量 X 的概率质量函数为

$$p(k)=C_{k-1}^{r-1}(1-p)^{k-r}p^r,\quad k=r,r+1,\cdots;\ \ 0<p<1,$$

则称 X 服从负二项分布，又称帕斯卡分布（Pascal Distribution），记为 $X\sim Nb(r,p)$。

实际背景是事件第 r 次发生时的试验次数。当 $r=1$ 时，负二项分布 $Nb(1,p)$ 就是几何分布 $Ge(p)$。

注：二项分布是已知实验次数时，发生次数的分布；负二项分布是已知发生次数时，试验次数的分布。

基本性质：非负性，$C_{k-1}^{r-1}(1-p)^{k-r}p^r>0$；正则性，

$$\sum_{k=r}^{+\infty}C_{k-1}^{r-1}(1-p)^{k-r}p^r=\frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty}(k-1)\cdots(k-r+1)q^{k-r}=\frac{p^r}{(r-1)!}\sum_{k=1}^{+\infty}\frac{d^{r-1}(q^{k-1})}{dq^{r-1}}$$

$$=\frac{p^r}{(r-1)!}\cdot\frac{d^{r-1}}{dq^{r-1}}\sum_{k=1}^{+\infty}q^{k-1}=\frac{p^r}{(r-1)!}\cdot\frac{d^{r-1}}{dq^{r-1}}\left(\frac{1}{1-q}\right)=\frac{p^r}{(r-1)!}\cdot\frac{(r-1)!}{(1-q)^r}=1。$$

负二项分布 $Nb(r,p)$ 的数学期望为

$$E(X)=\sum_{k=r}^{+\infty}k\cdot C_{k-1}^{r-1}(1-p)^{k-r}p^r=\frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty}k(k-1)\cdots(k-r+1)q^{k-r}=\frac{p^r}{(r-1)!}\sum_{k=0}^{+\infty}\frac{d^r(q^k)}{dq^r}$$

$$=\frac{p^r}{(r-1)!}\cdot\frac{d^r}{dq^r}\sum_{k=0}^{+\infty}q^k=\frac{p^r}{(r-1)!}\cdot\frac{d^r}{dq^r}\left(\frac{1}{1-q}\right)=\frac{p^r}{(r-1)!}\cdot\frac{r!}{(1-q)^{r+1}}=\frac{r}{p}。$$

又因

$$E(X^2)=\sum_{k=r}^{+\infty}k^2\cdot C_{k-1}^{r-1}(1-p)^{k-r}p^r=\sum_{k=r}^{+\infty}(k^2+k)\cdot C_{k-1}^{r-1}(1-p)^{k-r}p^r-\sum_{k=r}^{+\infty}k\cdot C_{k-1}^{r-1}(1-p)^{k-r}p^r$$

$$=\frac{p^r}{(r-1)!}\sum_{k=r}^{+\infty}(k+1)k(k-1)\cdots(k-r+1)q^{k-r}-E(X)=\frac{p^r}{(r-1)!}\sum_{k=0}^{+\infty}\frac{d^{r+1}(q^{k+1})}{dq^{r+1}}-\frac{r}{p}$$

$$=\frac{p^r}{(r-1)!}\cdot\frac{d^{r+1}}{dq^{r+1}}\sum_{k=0}^{+\infty}q^{k+1}-\frac{r}{p}=\frac{p^r}{(r-1)!}\cdot\frac{d^{r+1}}{dq^{r+1}}\left(\frac{q}{1-q}\right)-\frac{r}{p}=\frac{p^r}{(r-1)!}\cdot\frac{(r+1)!}{(1-q)^{r+2}}-\frac{r}{p}$$

$$=\frac{r(r+1)-rp}{p^2},$$

故方差为

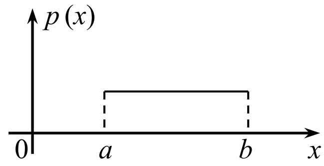
$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{r^2+r-rp}{p^2}-\left(\frac{r}{p}\right)^2=\frac{r(1-p)}{p^2}。$$

§2.5 常用连续分布

2.5.1. 均匀分布（Uniform Distribution）

一．均匀分布的密度函数和分布函数

某些随机变量分布在一个区间内，且其中处处都是等可能的。



定义 若连续随机变量 X 的密度函数为

$$p(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b;\\ 0, & 其它.\end{cases}\quad (a<b),$$

则称 X 服从区间 (a,b) 上的均匀分布，记为 $X\sim U(a,b)$。其分布函数为

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

实际背景是一维几何概型。

基本性质：非负性，$a<b$ 时，$\dfrac{1}{b-a}>0$；正则性，

$$\int_{-\infty}^{+\infty}p(x)dx=\int_a^b\frac{1}{b-a}dx=\frac{1}{b-a}x\Big|_a^b=\frac{1}{b-a}(b-a)=1。$$

如果 $X\sim U(a,b)$，且 $(c,d)\subset(a,b)$，则

$$P\{c<X<d\}=\int_c^d p(x)dx=\int_c^d\frac{1}{b-a}dx=\frac{1}{b-a}x\Big|_c^d=\frac{d-c}{b-a},$$

即均匀分布求概率公式是区间长度之比，这就是几何概型的求概率公式是 $P(A)=\dfrac{m(A)}{m(\Omega)}$。

例 若随机变量 $X\sim U(0,5)$，求方程 $4t^2+4Xt+(X+2)=0$ 有实根的概率。

解：因方程有实根，则判别式

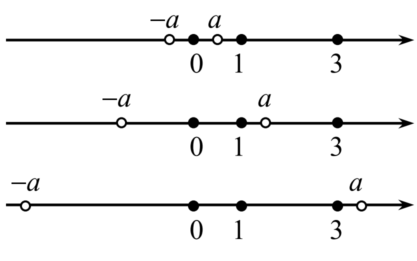
$$\Delta=(4X)^2-4\times4\times(X+2)=16X^2-16X-32=16(X+1)(X-2)\ge0,$$

即 $X\le-1$ 或 $X\ge2$。因 $X\sim U(0,5)$，即 X 的全部可能取值为 $(0,5)$，故所求概率为

$$P\{2\le X<5\}=\frac{5-2}{5-0}=0.6。$$

例 已知随机变量 $X\sim U(-a,a)$，且 $P\{1<X<3\}=\dfrac{1}{4}$，求常数 a。

解：显然 $a>0$。



当 $0<a<1$ 时，$P\{1<X<3\}=0$，矛盾；

当 $1\le a<3$ 时，$P\{1<X<3\}=\dfrac{a-1}{2a}=\dfrac{1}{4}$，得 $a=2$；

当 $a\ge3$ 时，$P\{1<X<3\}=\dfrac{2}{2a}=\dfrac{1}{4}$，得 $a=4$；

故 $a=2$ 或 4。

二．均匀分布的数学期望和方差

均匀分布 $U(a,b)$ 的数学期望为

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\frac{1}{2}x^2\Big|_a^b=\frac{1}{b-a}\cdot\frac{b^2-a^2}{2}=\frac{a+b}{2},$$

方差为

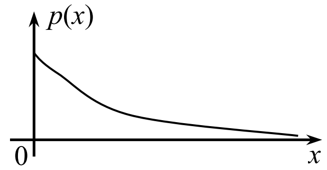
$$\mathrm{Var}(X)=E[X-E(X)]^2=\int_{-\infty}^{+\infty}\left(x-\frac{a+b}{2}\right)^2 p(x)dx=\int_a^b\left(x-\frac{a+b}{2}\right)^2\cdot\frac{1}{b-a}dx$$

$$=\frac{1}{b-a}\cdot\frac{1}{3}\left(x-\frac{a+b}{2}\right)^3\Bigg|_a^b=\frac{1}{3(b-a)}\cdot 2\left(\frac{b-a}{2}\right)^3=\frac{(b-a)^2}{12}\text{。}$$

2.5.2. 指数分布（Exponential Distribution）

一．指数分布的密度函数和分布函数

现实生活中，有些现象随着时间长度的增加发生的可能性逐渐减少。



定义 若连续随机变量 X 的密度函数为

$$p(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}\quad(\lambda>0)\text{，}$$

则称 X 服从参数为 λ 的指数分布，记为 $X\sim Exp(\lambda)$ 或 $e(\lambda)$ 。其分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x\ge 0;\\ 0, & x<0.\end{cases}$$

实际背景是发生时间问题，如首次发生时间、间隔时间、某些元件的寿命等。

基本性质：非负性， $\lambda>0$ 时， $\lambda\mathrm{e}^{-\lambda x}>0$ ；正则性，

$$\int_{-\infty}^{+\infty}p(x)dx=\int_0^{+\infty}\lambda\mathrm{e}^{-\lambda x}dx=(-\mathrm{e}^{-\lambda x})\Big|_0^{+\infty}=0-(-1)=1\text{。}$$

例 设一场足球比赛首次进球时间 X （分钟）服从参数 $\lambda=\dfrac{1}{36}$ 的指数分布，密度函数为

$$p(x)=\begin{cases}\dfrac{1}{36}\mathrm{e}^{-\frac{x}{36}}, & x>0;\\ 0, & x\le 0.\end{cases}$$

求：（1）上半场不进球的概率；（2）90 分钟内打成 $0:0$ 的概率。

解：（1）上半场不进球，即 $X>45$ ，

$$P\{X>45\}=\int_{45}^{+\infty}p(x)dx=\int_{45}^{+\infty}\frac{1}{36}\mathrm{e}^{-\frac{x}{36}}dx=-\mathrm{e}^{-\frac{x}{36}}\Big|_{45}^{+\infty}=0+\mathrm{e}^{-\frac{45}{36}}=\mathrm{e}^{-1.25}=0.2865\text{；}$$

（2）90 分钟内打成 $0:0$ ，即 $X>90$ ，

$$P\{X>90\}=\int_{90}^{+\infty}p(x)dx=\int_{90}^{+\infty}\frac{1}{36}\mathrm{e}^{-\frac{x}{36}}dx=-\mathrm{e}^{-\frac{x}{36}}\Big|_{90}^{+\infty}=0+\mathrm{e}^{-\frac{90}{36}}=\mathrm{e}^{-2.5}=0.0821\text{。}$$

例 设某元件的寿命 X （小时）的密度函数为

$$p(x)=\begin{cases}\dfrac{1}{600}\mathrm{e}^{-\frac{x}{600}}, & x\ge 0;\\ 0, & x<0.\end{cases}$$

求：（1）该元件寿命超过 300 小时的概率，以及不到 600 小时的概率；（2）常数 a 使得概率 $P\{X<a\}=0.5$ 。

解：（1）所求概率为

$$P\{X>300\}=\int_{300}^{+\infty}p(x)dx=\int_{300}^{+\infty}\frac{1}{600}\mathrm{e}^{-\frac{x}{600}}dx=-\mathrm{e}^{-\frac{x}{600}}\Big|_{300}^{+\infty}=0+\mathrm{e}^{-0.5}=0.6065\text{；}$$

$$P\{X<600\}=\int_{-\infty}^{600}p(x)dx=\int_0^{600}\frac{1}{600}\mathrm{e}^{-\frac{x}{600}}dx=-\mathrm{e}^{-\frac{x}{600}}\bigg|_0^{600}=-\mathrm{e}^{-1}+1=0.6321；$$

（2）因

$$P\{X<a\}=\int_{-\infty}^{a}p(x)dx=\int_0^{a}\frac{1}{600}\mathrm{e}^{-\frac{x}{600}}dx=-\mathrm{e}^{-\frac{x}{600}}\bigg|_0^{a}=-\mathrm{e}^{-\frac{a}{600}}+1=0.5，$$

故 $a=-600\ln 0.5=600\ln 2$ 。

二．指数分布的数学期望和方差

指数分布 $Exp(\lambda)$ 的数学期望为

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_0^{+\infty}x\cdot\lambda\mathrm{e}^{-\lambda x}dx=\int_0^{+\infty}x\cdot d(-\mathrm{e}^{-\lambda x})=-x\mathrm{e}^{-\lambda x}\Big|_0^{+\infty}+\int_0^{+\infty}\mathrm{e}^{-\lambda x}dx$$

$$=0+\left(-\frac{1}{\lambda}\mathrm{e}^{-\lambda x}\right)\bigg|_0^{+\infty}=\frac{1}{\lambda}，$$

又因

$$E(X^2)=\int_{-\infty}^{+\infty}x^2p(x)dx=\int_0^{+\infty}x^2\cdot\lambda\mathrm{e}^{-\lambda x}dx=\int_0^{+\infty}x^2\cdot d(-\mathrm{e}^{-\lambda x})=-x^2\mathrm{e}^{-\lambda x}\Big|_0^{+\infty}+\int_0^{+\infty}\mathrm{e}^{-\lambda x}\cdot 2xdx$$

$$=0+\frac{2}{\lambda}\int_0^{+\infty}x\cdot\lambda\mathrm{e}^{-\lambda x}dx=\frac{2}{\lambda}E(X)=\frac{2}{\lambda^2}，$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}。$$

三．指数分布的无记忆性

几何分布是首次发生的试验次数的离散分布，而指数分布是首次发生的时间的连续分布，指数分布与几何分布类似具有无记忆性。

定理　设 X 服从指数分布 $Exp(\lambda)$，则对任意 $s>0,t>0$，有

$P\{X>s+t\,|\,X>s\}=P\{X>t\}$ 。

证明：因对于 $\tau>0$，有

$$P\{X>\tau\}=\int_\tau^{+\infty}\lambda\mathrm{e}^{-\lambda x}dx=(-\mathrm{e}^{-\lambda x})\Big|_\tau^{+\infty}=\mathrm{e}^{-\lambda\tau}，$$

故

$$P\{X>s+t\,|\,X>s\}=\frac{P\{X>s+t\}}{P\{X>s\}}=\frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda s}}=\mathrm{e}^{-\lambda t}=P\{X>t\}。$$

此定理表明在已经经过 s 单位时间事件 A 没有发生的条件下，继续经过 t 单位时间仍没有发生的条件概率，等于直接经过 t 单位时间 A 没有发生的概率。这称之为指数分布的无记忆性。

例　设有 A,B,C 三个新的机器，它们的使用寿命相互独立且服从相同的指数分布，先同时独立使用 A,B，当其中一个损坏时换上新的 C，则 C 是最后一个损坏的概率是多少？

解：当 A,B 其中一个损坏时，根据指数分布的无记忆性可知所剩下的一个其使用寿命与前面的试验无关，则其分布与新换上的 C 是一样的，即 C 后损坏与 C 先损坏的概率是相等的，都是 0.5，故 C 是最后一个损坏的概率是 0.5。

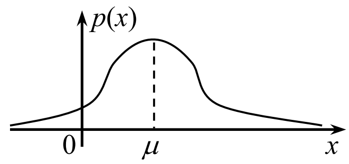
注：此题使用寿命的分布若改为几何分布，虽然仍可根据几何分布的无记忆性可得 C 后损坏与 C 先损坏的概率相等，但由于几何分布是离散分布，C 与 A,B 中后坏的一个同时损坏的概率不等于零，因此 C 后损坏与 C 先损坏的概率相等，但都小于 0.5。

2.5.3.　正态分布（Normal Distribution）

一．正态分布的密度函数和分布函数

现实生活中只要考虑的样本数量较大，通常情况下，就会体现出“两头小，中间大”的特点，如全国性考试成绩等。

定义　若连续随机变量 X 的密度函数为



$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty;\quad (\sigma>0),$$

则称 X 服从参数为 μ,σ^2 的正态分布，又称高斯分布（Gaussian Distribution），记为 $X\sim N(\mu,\sigma^2)$。其分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(u-\mu)^2}{2\sigma^2}}du\ 。$$

实际背景是现实生活中最常见的“两头小，中间大”的随机现象。

基本性质：非负性，$\sigma>0$ 时，$\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}>0$；正则性，令 $y=\frac{x-\mu}{\sqrt{2}\sigma}$，有 $x=\sqrt{2}\sigma y+\mu$，$dx=\sqrt{2}\sigma dy$，且 $x\to-\infty$ 时，$y\to-\infty$；$x\to+\infty$ 时，$y\to+\infty$，故

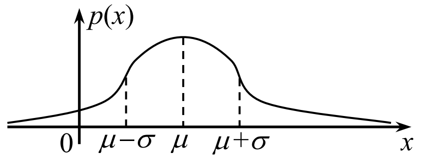
$$\int_{-\infty}^{+\infty}p(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-y^2}\cdot\sqrt{2}\sigma dy=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-y^2}dy=\frac{1}{\sqrt{\pi}}\cdot\sqrt{\pi}=1\ 。$$

注： 令 $I=\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}dx$，有 $I^2=\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}dx\int_{-\infty}^{+\infty}\mathrm{e}^{-y^2}dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2-y^2}dxdy=\int_0^{2\pi}d\theta\int_0^{+\infty}\mathrm{e}^{-r^2}rdr=2\pi\cdot\frac{1}{2}=\pi$，故 $I=\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}dx=\sqrt{\pi}$。

单调性与极值：因 $p'(x)=-\frac{x-\mu}{\sqrt{2\pi}\,\sigma^3}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$，当 $x<\mu$ 时，有 $p'(x)>0$，$p(x)$ 单调增加；当 $x>\mu$ 时，有 $p'(x)<0$，$p(x)$ 单调减少；可得 $x=\mu$ 为 $p(x)$ 的极大值点，$p(\mu)=\frac{1}{\sqrt{2\pi}\,\sigma}$ 为极大值。

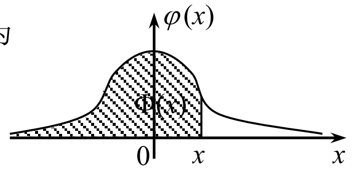
凹凸性与拐点：因 $p''(x)=-\frac{1}{\sqrt{2\pi}\,\sigma^3}\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$，当 $x<\mu-\sigma$ 或 $x>\mu+\sigma$ 时，有 $p''(x)>0$，$p(x)$ 为凸；当 $\mu-\sigma<x<\mu+\sigma$ 时，有 $p''(x)<0$，$p(x)$ 为凹；可得 $x=\mu\pm\sigma$ 为 $p(x)$ 的拐点横坐标。

密度函数 $p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 的图形特点是以 $x=\mu$ 为对称轴，且 $p(\mu)=\frac{1}{\sqrt{2\pi}\,\sigma}$ 为 $p(x)$ 的最大值，且 σ 越大，最大值 $p(\mu)$ 越小，图形越平坦；σ 越小，最大值 $p(\mu)$ 越大，图形越陡峭。



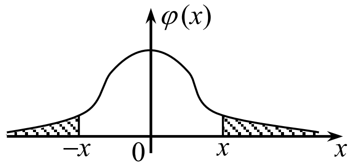
二．标准正态分布（Standard Normal Distribution）

当参数 $\mu=0$，$\sigma^2=1$时，$N(0,1)$ 称为标准正态分布。其密度函数记为



$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},\quad -\infty<x<+\infty\text{；}$$

分布函数记为



$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}du\text{。}$$

从图形直观上，$\varphi(x)$ 是偶函数，关于 y 轴对称。$\varphi(0)=\frac{1}{\sqrt{2\pi}}$ 为$\varphi(x)$ 的最大值，且

$$\varphi(-x)=\varphi(x)\text{，}\Phi(0)=0.5\text{，}\Phi(-x)=1-\Phi(x)\text{。}$$

标准正态分布求概率时，可以查附表 2。

例 已知 $X\sim N(0,1)$，求 $P\{X<1.28\}$，$P\{1.25<X<2.12\}$，$P\{|X|<1.96\}$，$P\{-\frac{1}{3}<X<6\}$。

解： 因 $X\sim N(0,1)$，有分布函数 $\Phi(x)=P\{X\le x\}$，故

$P\{X<1.28\}=\Phi(1.28)=0.8997$；

$P\{1.25<X<2.12\}=\Phi(2.12)-\Phi(1.25)=0.9830-0.8944=0.0886$；

$P\{|X|<1.96\}=\Phi(1.96)-\Phi(-1.96)=2\Phi(1.96)-1=2\times0.9750-1=0.95$；

$$P\{-\frac{1}{3}<X<6\}=\Phi(6)-\Phi\left(-\frac{1}{3}\right)=1-\left[1-\Phi\left(\frac{1}{3}\right)\right]=\Phi\left(\frac{1}{3}\right)\approx\Phi(0.33)+\frac{1}{3}[\Phi(0.34)-\Phi(0.33)]$$

$$=0.6293+\frac{1}{3}\times0.0038\approx0.6306\text{。}$$

注： 最后一问是查附表 2 并利用线性插值（Linear Interpolation）进行近似计算的。设自变量 x 介于 x_1 与 x_2 之间，且函数值 $f(x_1)$ 与 $f(x_2)$ 都已知，线性插值就是将 x_1 与 x_2 两点之间用线性函数近似，即当 $x=x_1+\alpha(x_2-x_1)$ 时，取 $f(x)=f(x_1)+\alpha[f(x_2)-f(x_1)]$。此外，也可用 MATLAB 中的命令 normcdf 进行计算。

三．一般正态分布的标准化

对于一般正态分布 $N(\mu,\sigma^2)$，可通过标准化成为标准正态分布 $N(0,1)$。若 $X\sim N(\mu,\sigma^2)$，其密度函数为 $p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$，分布函数为 $F(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(u-\mu)^2}{2\sigma^2}}du$，令 $t=\frac{u-\mu}{\sigma}$，有 $dt=\frac{1}{\sigma}du$；且 $u\to-\infty$ 时，$t\to-\infty$；$u=x$ 时，$t=\frac{x-\mu}{\sigma}$，则

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma}}\mathrm{e}^{-\frac{t^2}{2}}dt=\Phi\left(\frac{x-\mu}{\sigma}\right)\text{。}$$

例 若 $X\sim N(3,16)$，求 $P\{X>0\}$，$P\{1<X<10\}$，$P\{|X|<9\}$。

解： $\mu=3$，$\sigma^2=16$，即 $\sigma=4$，故

$$P\{X>0\}=1-F(0)=1-\Phi\left(\frac{0-3}{4}\right)=1-\Phi(-0.75)=\Phi(0.75)=0.7734\text{；}$$

$$P\{1<X<10\}=F(10)-F(1)=\Phi\left(\frac{10-3}{4}\right)-\Phi\left(\frac{1-3}{4}\right)=\Phi(1.75)-\Phi(-0.5)$$

$$=\Phi(1.75)-1+\Phi(0.5)=0.9599-1+0.6915=0.6514\text{；}$$

$$P\{|X|<9\}=F(9)-F(-9)=\Phi\left(\frac{9-3}{4}\right)-\Phi\left(\frac{-9-3}{4}\right)=\Phi(1.5)-\Phi(-3)$$

$$=0.9332-1+0.9986=0.9318\text{。}$$

例 已知随机变量X的密度函数$p(x)=C\,\mathrm{e}^{-x^2-x}$，求：（1）常数$C$；（2）$P\{X<0\}$；（3）常数$a$，使得$\int_{-\infty}^{a}p(x)dx=2\int_{a}^{+\infty}p(x)dx$。

解：（1）X密度函数

$$p(x)=C\,\mathrm{e}^{-x^2-x}=C\,\mathrm{e}^{-(x^2+x+\frac{1}{4}-\frac{1}{4})}=C\,\mathrm{e}^{-(x+\frac{1}{2})^2+\frac{1}{4}}=C\,\mathrm{e}^{\frac{1}{4}}\cdot\mathrm{e}^{-(x+\frac{1}{2})^2}\text{，}$$

与正态分布$N(\mu,\sigma^2)$的密度函数$p(x)=\dfrac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$比较，可得

$$\mu=-\frac{1}{2}\text{，}\quad\sigma^2=\frac{1}{2}\text{，且}\,C\,\mathrm{e}^{\frac{1}{4}}=\frac{1}{\sqrt{2\pi}\,\sigma}=\frac{1}{\sqrt{\pi}}\text{，}$$

故$C=\dfrac{1}{\sqrt{\pi}}\mathrm{e}^{-\frac{1}{4}}$；

（2）所求概率为

$$P\{X<0\}=F(0)=\Phi\left(\frac{0+1/2}{1/\sqrt{2}}\right)=\Phi(0.7071)\approx\Phi(0.70)+0.71\times[\Phi(0.71)-\Phi(0.70)]$$

$$=0.7580+0.71\times0.0031=0.7602\text{；}$$

（3）因$\int_{-\infty}^{a}p(x)dx=2\int_{a}^{+\infty}p(x)dx$，有$F(a)=2[1-F(a)]$，得$F(a)=\dfrac{2}{3}\approx0.6667$，则

$$F(a)=\Phi\left(\frac{a+0.5}{1/\sqrt{2}}\right)=0.6667\text{，}\quad\frac{a+0.5}{1/\sqrt{2}}=0.43+\frac{0.0003}{0.0036}\times(0.44-0.43)=0.4308\text{，}$$

故$a=-0.1954$。

四．正态分布的数学期望和方差

正态分布$N(\mu,\sigma^2)$的数学期望为

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_{-\infty}^{+\infty}(x-\mu+\mu)p(x)dx=\int_{-\infty}^{+\infty}(x-\mu)\cdot\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}dx+\int_{-\infty}^{+\infty}\mu\,p(x)dx$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}y\,\mathrm{e}^{-\frac{y^2}{2\sigma^2}}dy+\mu\int_{-\infty}^{+\infty}p(x)dx=0+\mu=\mu\text{，}$$

方差为

$$\mathrm{Var}(X)=E[X-E(X)]^2=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)(-\sigma^2)d\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$=\frac{-\sigma^2}{\sqrt{2\pi}\sigma}(x-\mu)\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\Bigg|_{-\infty}^{+\infty}+\frac{\sigma^2}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=0+\sigma^2\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\sigma^2\text{。}$$

五．正态分布的 3σ 原则

设 $X\sim N(\mu,\sigma^2)$，且 $k=1,2,3$，则

$$P\{|X-\mu|<k\sigma\}=\Phi(\frac{\mu+k\sigma-\mu}{\sigma})-\Phi(\frac{\mu-k\sigma-\mu}{\sigma})=\Phi(k)-\Phi(-k)=2\Phi(k)-1\text{，}$$

当 $k=1$ 时，$P\{|X-\mu|<\sigma\}=2\Phi(1)-1=2\times0.8413-1=0.6826$；

当 $k=2$ 时，$P\{|X-\mu|<2\sigma\}=2\Phi(2)-1=2\times0.9772-1=0.9544$；

当 $k=3$ 时，$P\{|X-\mu|<3\sigma\}=2\Phi(3)-1=2\times0.9987-1=0.9974$。

可见，正态随机变量 X 的取值几乎总是落在 $(\mu-3\sigma,\mu+3\sigma)$ 之间，称之为正态分布的 3σ 原则。

例 某企业有 400 名工人，每人都有 20%的可能性使用某种设备，且是否使用互不影响，试问准备多少台该种设备比较合适？

分析：由第四章的中心极限定理知随机因素较多时，可用正态分布近似反映随机现象，这里先用正态分布的 3σ 原则估计使用该种设备的工人数，由此决定需要准备的设备数量。

解：设 X 表示“使用该种设备的工人数”，有 X 服从二项分布 $b(400,0.2)$，则

$$\mu=E(X)=np=400\times0.2=80\text{，}\quad\sigma^2=\mathrm{Var}(X)=np(1-p)=400\times0.2\times0.8=64\text{，}$$

由 3σ 原则知，使用该种设备的工人数 X 几乎总是在 $(\mu\pm3\sigma)=(80\pm3\times8)=(56,104)$ 之间，故准备 104 台该种设备比较合适。

2.5.4. 伽玛分布（Gamma Distribution）

一．伽玛函数

含参变量积分

$$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}\mathrm{e}^{-x}dx,\quad\alpha>0\text{，}$$

称为伽玛函数（Γ 函数）。伽玛函数具有以下性质：

（1）$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$，特别是 $\alpha=n$ 为正整数时，$\Gamma(n+1)=n\Gamma(n)=n!$；

（2）$\Gamma\left(\frac{1}{2}\right)=\int_0^{+\infty}x^{-\frac{1}{2}}\mathrm{e}^{-x}dx=\sqrt{\pi}$。

证明：（1）$\Gamma(\alpha+1)=\int_0^{+\infty}x^{\alpha}\mathrm{e}^{-x}dx=\int_0^{+\infty}x^{\alpha}d(-\mathrm{e}^{-x})=-x^{\alpha}\mathrm{e}^{-x}\Big|_0^{+\infty}+\int_0^{+\infty}\mathrm{e}^{-x}\cdot\alpha x^{\alpha-1}dx=0+\alpha\Gamma(\alpha)=\alpha\Gamma(\alpha)$。

当 $\alpha=n$ 为正整数时，

$$\Gamma(n+1)=n\Gamma(n)=\Lambda=n(n-1)\Lambda\ 2\cdot1\cdot\Gamma(1)=n!\cdot\int_0^{+\infty}\mathrm{e}^{-x}dx=n!\cdot(-\mathrm{e}^{-x})\Big|_0^{+\infty}=n!\text{；}$$

（2）对于 $\int_0^{+\infty}x^{-\frac{1}{2}}\mathrm{e}^{-x}dx$，令 $y=x^{\frac{1}{2}}$，有 $x=y^2$，$dx=2ydy$，且 $x=0$ 时，$y=0$；$x\to+\infty$ 时，$y\to+\infty$，

$$\Gamma\left(\frac{1}{2}\right)=\int_0^{+\infty}x^{-\frac{1}{2}}\mathrm{e}^{-x}dx=\int_0^{+\infty}y^{-1}\mathrm{e}^{-y^2}\cdot2ydy=2\int_0^{+\infty}\mathrm{e}^{-y^2}dy=\int_{-\infty}^{+\infty}\mathrm{e}^{-y^2}dy=\sqrt{\pi}\text{。}$$

二．伽玛分布

定义 若连续随机变量 X 的密度函数为

$$p(x)=\begin{cases}\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\,\mathrm{e}^{-\lambda x}, & x\geq 0;\\ 0, & x<0.\end{cases}\quad (\alpha>0,\lambda>0)\,,$$

则称 X 服从伽玛分布（Γ 分布），记为 $X\sim Ga(\alpha,\lambda)$。

基本性质：非负性，$\alpha>0,\lambda>0$ 时，$\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\,\mathrm{e}^{-\lambda x}>0$；正则性，当 $r>0$ 时，对于 $\int_0^{+\infty}x^{r-1}\,\mathrm{e}^{-\lambda x}\,dx$，

令 $y=\lambda x$，有 $x=\dfrac{y}{\lambda}$，$dx=\dfrac{1}{\lambda}dy$，且 $x=0$ 时，$y=0$；$x\to+\infty$ 时，$y\to+\infty$，则

$$\int_0^{+\infty}x^{r-1}\,\mathrm{e}^{-\lambda x}\,dx=\int_0^{+\infty}\frac{y^{r-1}}{\lambda^{r-1}}\mathrm{e}^{-y}\cdot\frac{1}{\lambda}dy=\frac{1}{\lambda^{r}}\int_0^{+\infty}y^{r-1}\,\mathrm{e}^{-y}\,dy=\frac{\Gamma(r)}{\lambda^{r}},$$

故

$$\int_{-\infty}^{+\infty}p(x)dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{+\infty}x^{\alpha-1}\,\mathrm{e}^{-\lambda x}\,dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha)}{\lambda^{\alpha}}=1\text{。}$$

三．伽玛分布的数学期望和方差

当 $r>0$ 时，有 $\int_0^{+\infty}x^{r-1}\,\mathrm{e}^{-\lambda x}\,dx=\dfrac{\Gamma(r)}{\lambda^{r}}$，可得伽玛分布 $Ga(\alpha,\lambda)$ 的数学期望为

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_0^{+\infty}x\cdot\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\,\mathrm{e}^{-\lambda x}\,dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{+\infty}x^{\alpha}\,\mathrm{e}^{-\lambda x}\,dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}=\frac{\alpha}{\lambda},$$

又因

$$E(X^2)=\int_0^{+\infty}x^2\cdot\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\,\mathrm{e}^{-\lambda x}\,dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{+\infty}x^{\alpha+1}\,\mathrm{e}^{-\lambda x}\,dx=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}}=\frac{\alpha(\alpha+1)}{\lambda^2},$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{\alpha(\alpha+1)}{\lambda^2}-\left(\frac{\alpha}{\lambda}\right)^2=\frac{\alpha}{\lambda^2}\text{。}$$

四．伽玛分布的特例

（1）当 $\alpha=1$ 时，伽玛分布 $Ga(1,\lambda)$ 的密度函数为

$$p(x)=\begin{cases}\lambda\,\mathrm{e}^{-\lambda x}, & x\geq 0;\\ 0, & x<0.\end{cases}\quad(\lambda>0)\,,$$

可见伽玛分布 $Ga(1,\lambda)$ 就是指数分布 $Exp(\lambda)$。

（2）当 $\alpha=\dfrac{n}{2},\lambda=\dfrac{1}{2}$ 时，伽玛分布 $Ga\left(\dfrac{n}{2},\dfrac{1}{2}\right)$ 是自由度为 n 的 χ^2 分布 $\chi^2(n)$，密度函数为

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}\,\mathrm{e}^{-\frac{x}{2}}, & x\geq 0,\\ 0, & x<0.\end{cases}\quad n\text{ 为正整数，}$$

可以证明n个独立标准正态变量的平方和服从$\chi^2(n)$。χ^2分布是统计学中三种重要分布之一。

（3）当$\alpha = n$为正整数时，伽玛分布$Ga(n, \lambda)$表示泊松过程的质点流中第n个质点到达时间的分布。

例　一个质点流在$(0, t]$时间内出现的质点个数为$N(t)\quad (t > 0)$，设$N(t)$服从参数为λt的泊松分布$P(\lambda t)$，求第n个质点到达时间S_n的分布。

解： 因$N(t)$服从参数为λt的泊松分布$P(\lambda t)$，有$N(t)$的概率分布为

$$P\{N(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \cdots 。$$

当$t < 0$时，S_n的分布函数$F_S(t) = P\{S_n \le t\} = 0$，密度函数$p_S(t) = F_S'(t) = 0$；当$t \ge 0$时，$S_n$的分布函数

$$F_S(t) = P\{S_n \le t\} = P\{N(t) \ge n\} = \sum_{k=n}^{+\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = e^{-\lambda t}\left(\frac{\lambda^n t^n}{n!} + \frac{\lambda^{n+1} t^{n+1}}{(n+1)!} + \frac{\lambda^{n+2} t^{n+2}}{(n+2)!} + \cdots\right),$$

密度函数

$$p_S(t) = F_S'(t) = -\lambda e^{-\lambda t}\left(\frac{\lambda^n t^n}{n!} + \frac{\lambda^{n+1} t^{n+1}}{(n+1)!} + \frac{\lambda^{n+2} t^{n+2}}{(n+2)!} + \cdots\right) + e^{-\lambda t}\left(\frac{\lambda^n t^{n-1}}{(n-1)!} + \frac{\lambda^{n+1} t^n}{n!} + \frac{\lambda^{n+2} t^{n+1}}{(n+1)!} + \cdots\right)$$

$$= e^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!} = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t},$$

故

$$p_S(t) = \begin{cases} \dfrac{\lambda^n}{\Gamma(n)} t^{n-1} \mathrm{e}^{-\lambda t}, & t \ge 0; \\ 0, & t < 0. \end{cases} \quad (\lambda > 0),$$

即S_n服从伽玛分布$Ga(n, \lambda)$。

2.5.5.　贝塔分布（Beta Distribution）

一．贝塔函数

含参变量积分

$$\mathrm{B}(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx, \quad a > 0, b > 0$$

称为贝塔函数（β函数）。贝塔函数具有以下性质：

（1）$\mathrm{B}(a, b) = \mathrm{B}(b, a)$；

（2）贝塔函数与伽玛函数的关系为

$$\mathrm{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} 。$$

证明：（1）对于$\mathrm{B}(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$，令$y = 1 - x$，有$x = 1 - y$，$dx = -dy$，且$x = 0$时，$y = 1$；$x = 1$时，$y = 0$；故

$$\mathrm{B}(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = \int_1^0 (1-y)^{a-1} y^{b-1}(-dy) = \int_0^1 y^{b-1}(1-y)^{a-1} dy = \mathrm{B}(b, a);$$

（2）因

$$\Gamma(a)\Gamma(b)=\int_0^{+\infty} x^{a-1}\,\mathrm{e}^{-x}\,dx\cdot\int_0^{+\infty} y^{b-1}\,\mathrm{e}^{-y}\,dy=\int_0^{+\infty}\int_0^{+\infty} x^{a-1}y^{b-1}\,\mathrm{e}^{-(x+y)}\,dxdy\,,$$

令 $x=uv$， $y=u(1-v)$，有

$$J=\begin{vmatrix}\dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v}\\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix} v & u\\ 1-v & -u\end{vmatrix}=-u\,,$$

即 $dxdy=|J|\,dudv=|u|\,dudv$，且 $x\ge 0, y\ge 0$ 时，有 $uv\ge 0, u(1-v)\ge 0$，可得 $u\ge 0, 0\le v\le 1$，故

$$\Gamma(a)\Gamma(b)=\int_0^{+\infty}\int_0^1 (uv)^{a-1}[u(1-v)]^{b-1}\,\mathrm{e}^{-u}\cdot u\,dudv$$

$$=\int_0^{+\infty} u^{a+b-1}\,\mathrm{e}^{-u}\,du\cdot\int_0^1 v^{a-1}(1-v)^{b-1}dv=\Gamma(a+b)\cdot\mathrm{B}(a,b)\,,$$

得证。

二．贝塔分布

定义 若连续随机变量 X 的密度函数为

$$p(x)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, & 0<x<1;\\ 0, & \text{其他.}\end{cases}\quad (a>0, b>0)\,,$$

则称 X 服从贝塔分布（ β 分布），记为 $X\sim Be(a,b)$。

由于贝塔分布的支撑区间为 $(0,1)$，故常用贝塔分布反映比率的分布。

基本性质：非负性， $a>0, b>0$ 时， $\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}>0$；正则性，

$$\int_{-\infty}^{+\infty} p(x)dx=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\mathrm{B}(a,b)=1\text{。}$$

三．贝塔分布的数学期望和方差

贝塔分布 $Be(a,b)$ 的数学期望为

$$E(X)=\int_{-\infty}^{+\infty} xp(x)dx=\int_0^{+\infty} x\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^{+\infty} x^a(1-x)^{b-1}dx$$

$$=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\mathrm{B}(a+1,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}=\frac{a}{a+b}\,,$$

又因

$$E(X^2)=\int_{-\infty}^{+\infty} x^2p(x)dx=\int_0^{+\infty} x^2\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^{+\infty} x^{a+1}(1-x)^{b-1}dx$$

$$=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\mathrm{B}(a+2,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)}=\frac{a(a+1)}{(a+b)(a+b+1)}\,,$$

故方差为

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\frac{a(a+1)}{(a+b)(a+b+1)}-\left(\frac{a}{a+b}\right)^2=\frac{ab}{(a+b)^2(a+b+1)}\text{。}$$

常用分布

分布		记号	概率分布或密度函数	实际背景	期望	方差
离散	两点	$(0\text{-}1)$	$P\{X=1\}=p$，$P\{X=0\}=1-p$	一次伯努利试验	p	$p(1-p)$
	二项	$b(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$，$k=0,1,2,\cdots,n$	n 重伯努利试验	np	$np(1-p)$
	泊松	$P(\lambda)$	$P\{X=k\}=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$，$k=0,1,2,\cdots$	连续范围内的发生次数问题	λ	λ
	超几何	$h(n,N,M)$	$P\{X=k\}=\dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$，$k=l,l+1,\cdots,r$	不放回抽样问题	$\dfrac{nM}{N}$	$n\dfrac{M}{N}(1-\dfrac{M}{N})\dfrac{N-n}{N-1}$
	几何	$Ge(p)$	$P\{X=k\}=(1-p)^{k-1}p$，$k=1,2,\cdots$	首次发生时的试验次数	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
	负二项	$Nb(r,p)$	$P\{X=k\}=C_{k-1}^{r-1}(1-p)^{k-r}p^r$，$k=r,r+1,\cdots$	第 r 次发生时的试验次数	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$
连续	均匀	$U(a,b)$	$p(x)=\dfrac{1}{b-a}$，$a<x<b$	一维几何概型	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
	指数	$Exp(\lambda)$	$p(x)=\lambda\mathrm{e}^{-\lambda x}$，$x>0$	发生时间问题	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
	正态	$N(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	大数量	μ	σ^2
	伽玛	$Ga(\alpha,\lambda)$	$p(x)=\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}$，$x>0$	伽玛函数	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$
	贝塔	$Be(a,b)$	$p(x)=\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$，$0<x<1$	贝塔函数 比率问题	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$

§2.6 随机变量函数的分布

设 X 为一个随机变量，$y=g(x)$ 为一个函数，则称 $Y=g(X)$ 为随机变量函数。根据随机变量 X 的分布推导出其函数 Y 的分布。

2.6.1. 离散随机变量函数

设 X 为离散随机变量，全部可能取值为 $x_1, x_2, \cdots, x_k, \cdots$，则其函数 $Y=g(X)$ 也为离散随机变量，全部可能取值为 $g(x_1), g(x_2), \cdots, g(x_k), \cdots$，直接求出 $Y=g(x_k),\quad k=1,2,\cdots$ 的概率，可得 Y 的概率质量函数，此方法称为质量函数法。

若 X 的概率质量函数为 $p(x_k)=p_k,\quad k=1,2,\cdots$，即分布列为

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

当 $y_k=g(x_k),\quad k=1,2,\cdots$ 为一一对应时，Y 的概率质量函数为 $p[g(x_k)]=p_k,\quad k=1,2,\cdots$，即分布列为

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_k)$	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

当 $y_k=g(x_k),\quad k=1,2,\cdots$ 有重复时，应将重复取值合并，对应概率相加。

例 已知 X 的分布列为

X	-2	-1	0	1	2
P	0.1	0.3	0.3	0.2	0.1

求 $2X+1$，X^2 的分布。

解： 因 X 的全部可能取值为 $-2,-1,0,1,2$，可得 $2X+1$ 的全部可能取值为 $-3,-1,1,3,5$，X^2 的全部可能取值为 $4,1,0,1,4$，故 $2X+1$ 的分布列为

$2X+1$	-3	-1	1	3	5
P	0.1	0.3	0.3	0.2	0.1

且 X^2 的分布列为

X^2	0	1	4
P	0.3	0.5	0.2

2.6.2. 连续随机变量函数

一．当函数 $g(x)$ 的取值为有限个或可列个时，采用质量函数法

当函数 $y=g(x)$ 的全部可能取值为 $y_1, y_2, \cdots, y_k, \cdots$ 时，虽然此时 X 为连续随机变量，但 $Y=g(X)$ 却为离散随机变量，直接求出由 $Y=y_k,\quad k=1,2,\cdots$ 的概率，可得 Y 的概率质量函数。

例 设随机变量 X 服从均匀分布 $U(-1,2)$，求 $Y=\mathrm{sgn}(X)$ 的分布。

解： 因符号函数

$$y=\operatorname{sgn}(x)=\begin{cases}-1, & x<0;\\ 0, & x=0;\\ 1, & x>0.\end{cases}$$

则 $P\{Y=-1\}=P\{X<0\}=\dfrac{1}{3}$，$P\{Y=0\}=P\{X=0\}=0$，$P\{Y=1\}=P\{X>0\}=\dfrac{2}{3}$，故 $Y=\operatorname{sgn}(X)$ 的分布列为

$\operatorname{sgn}(X)$	-1	1
P	$\dfrac{1}{3}$	$\dfrac{2}{3}$

二．当函数 $g(x)$ 为一般形式时，采用分布函数法

设 X 为连续随机变量，对于一般函数 $Y=g(X)$，由 Y 分布函数中的不等式 $Y=g(X)\le y$，解得 X 的取值范围 $X\in A$，从而求得概率 $P\{Y\le y\}=P\{X\in A\}$，得到 Y 的分布函数 $F_Y(y)$，此方法称为分布函数法。

若分布函数 $F_Y(y)$ 连续（不可导的点最多可列个），再求导得 Y 的密度函数 $p_Y(y)$。

分布函数法的关键是应首先根据函数 $y=g(x)$ 的值域以及 X 的分布确定 y 的分段点，再在 y 的各个分段区间内求解不等式 $g(X)\le y$。

例 设随机变量 X 服从标准正态分布 $N(0,1)$，求 $Y=X^2$ 的分布。

解：由函数 $y=x^2$ 的值域得分段点 $y=0$。

当 $y<0$ 时，$F_Y(y)=P\{X^2\le y\}=P(\varnothing)=0$；

当 $y\ge 0$ 时，$F_Y(y)=P\{X^2\le y\}=P\{-\sqrt{y}\le X\le\sqrt{y}\}=\Phi(\sqrt{y})-\Phi(-\sqrt{y})=2\Phi(\sqrt{y})-1$；

可知 $F_Y(y)$ 连续，Y 是连续随机变量，因当 $y>0$ 时，密度函数

$$p_Y(y)=F_Y'(y)=2\varphi(\sqrt{y})\cdot\frac{1}{2\sqrt{y}}=\frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}\mathrm{e}^{-\frac{y}{2}},$$

故 $Y=X^2$ 的密度函数为

$$p_Y(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}\mathrm{e}^{-\frac{y}{2}}, & y>0;\\ 0, & y\le 0.\end{cases}$$

即 Y 服从伽马分布 $\Gamma\left(\dfrac{1}{2},\dfrac{1}{2}\right)$。

例 设随机变量 X 的密度函数为

$$p_X(x)=\begin{cases}\dfrac{2+2x}{9}, & -1<x<2;\\ 0, & \text{其它}.\end{cases}$$

求 $Y=|X|$ 的分布。

解：由函数 $y=|x|$ 的值域得分段点 $y=0$，由 X 的分布得分段点 $y=1,2$。

当 $y<0$ 时，$F_Y(y)=P\{|X|\le y\}=P(\varnothing)=0$；

当$0\le y<1$时，$F_Y(y)=P\{|X|\le y\}=P\{-y\le X\le y\}=\int_{-y}^{y}\frac{2+2x}{9}dx=\left.\frac{2x+x^2}{9}\right|_{-y}^{y}=\frac{4y}{9}$；

当$1\le y<2$时，$F_Y(y)=P\{|X|\le y\}=P\{-1\le X\le y\}=\int_{-1}^{y}\frac{2+2x}{9}dx=\left.\frac{2x+x^2}{9}\right|_{-1}^{y}=\frac{2y+y^2+1}{9}$；

当$y\ge 2$时，$F_Y(y)=P\{|X|\le y\}=P(\Omega)=1$；

可知$F_Y(y)$连续，Y是连续随机变量，故$Y=|X|$的密度函数为

$$p_Y(y)=F_Y'(y)=\begin{cases}\frac{4}{9}, & 0<y<1;\\ \frac{2+2y}{9}, & 1\le y<2;\\ 0, & \text{其他.}\end{cases}$$

例　设某元件的使用寿命X（年）服从指数分布$Exp(\lambda)$，现在开始使用一只该元件，当其损坏后或用满两年后将其更换，求更换时间Y的分布。

解：X的密度函数与分布函数分别为

$$p_X(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}\qquad F_X(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}$$

更换时间

$$Y=\begin{cases}X, & X\le 2;\\ 2, & X>2.\end{cases}=\min\{X,2\},$$

由函数$y=\min\{x,2\}$的值域得分段点$y=2$，由X的分布得分段点$y=0$，

当$y<0$时，$F_Y(y)=P\{\min\{X,2\}\le y\}=P(\varnothing)=0$；

当$0\le y<2$时，$F_Y(y)=P\{\min\{X,2\}\le y\}=P\{X\le y\}=F_X(y)=1-\mathrm{e}^{-\lambda y}$；

当$y\ge 2$时，$F_Y(y)=P\{\min\{X,2\}\le y\}=P(\Omega)=1$；

故$Y=\min\{X,2\}$的分布函数为

$$F_Y(y)=\begin{cases}0, & y<0;\\ 1-\mathrm{e}^{-\lambda y}, & 0\le y<2;\\ 1, & y\ge 2.\end{cases}$$

可知$F_Y(y)$不是连续函数，即Y不是连续随机变量，而是混合型随机变量。

注：此题Y的分布既不是离散，也不是连续，而是混合型，不存在质量函数或密度函数。

定理　设X的分布函数$F_X(x)$为严格单调增加的连续函数，则$Y=F_X(X)$服从均匀分布$U(0,1)$。

证明：由函数$y=F_X(x)$的值域得分段点$y=0,1$，

当$y<0$时，$F_Y(y)=P\{F_X(X)\le y\}=P(\varnothing)=0$；

当$0 \le y < 1$时，$F_Y(y) = P\{F_X(X) \le y\} = P\{X \le F_X^{-1}(y)\} = F_X[F_X^{-1}(y)] = y$；

当$y \ge 1$时，$F_Y(y) = P\{F_X(X) \le y\} = P(\Omega) = 1$；

可知$F_Y(y)$连续，Y是连续随机变量，故

$$p_Y(y) = F_Y'(y) = \begin{cases} 1, & 0 < y < 1; \\ 0, & 其他. \end{cases}$$

即$Y = F_X(X)$服从均匀分布$U(0,1)$。

三．当函数$g(x)$严格单调且可导时，采用密度函数法

设X为连续随机变量，当$g(x)$严格单调且可导时，可根据以下定理直接求得$Y = g(X)$的密度函数。此方法称为密度函数法。

定理　设连续随机变量X的密度函数为$p_X(x)$，若$y = g(x)$为严格单调函数，其反函数$x = h(y)$有连续导数，则随机变量函数$Y = g(X)$的密度函数为

$$p_Y(y) = p_X[h(y)] \cdot |h'(y)|。$$

证明：若$y = g(x)$严格单调增加，则

$$F_Y(y) = P\{g(X) \le y\} = P\{X \le h(y)\} = F_X[h(y)]，$$

可知$F_Y(y)$连续，Y是连续随机变量，可得$p_Y(y) = F_Y'(y) = F_X'[h(y)] \cdot h'(y) = p_X[h(y)] \cdot h'(y)$；

若$y = g(x)$严格单调减少，则

$$F_Y(y) = P\{g(X) \le y\} = P\{X \ge h(y)\} = 1 - F_X[h(y)]，$$

可知$F_Y(y)$连续，Y是连续随机变量，可得

$$p_Y(y) = F_Y'(y) = -F_X'[h(y)] \cdot h'(y) = -p_X[h(y)] \cdot h'(y)。$$

由于$y = g(x)$严格单调增加时，$h'(y) \ge 0$；严格单调减少时，$h'(y) \le 0$；故

$$p_Y(y) = p_X[h(y)] \cdot |h'(y)|。$$

注：只要$y = g(x)$在随机变量X的支撑区间内为严格单调函数，此定理仍然适用。

例　设X的密度函数为

$$p_X(x) = \begin{cases} 2x, & 0 < x < 1; \\ 0, & 其他. \end{cases}$$

求$Y = 3 - 2X$与$Z = -\ln X$的分布。

解：对于$Y = 3 - 2X$，有$y = 3 - 2x$严格单调减少，反函数$x = h(y) = \dfrac{3-y}{2}$，导数$h'(y) = -\dfrac{1}{2}$，且当$0 < x < 1$时，有$1 < y < 3$，故

$$p_Y(y) = 2 \cdot \frac{3-y}{2} \cdot \left| -\frac{1}{2} \right| = \frac{3-y}{2},\quad 1 < y < 3；$$

对于 $Z=-\ln X$，有 $z=-\ln x$ 严格单减，反函数 $x=h(z)=\mathrm{e}^{-z}$，导数 $h'(z)=-\mathrm{e}^{-z}$，且当 $0<x<1$ 时，有 $z>0$，故

$$p_Z(z)=2\mathrm{e}^{-z}\cdot|-\mathrm{e}^{-z}|=2\mathrm{e}^{-2z},\quad z>0\text{。}$$

例 设 $X\sim U\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$，求 $Y=\sin X$ 与 $Z=\tan X$ 的密度函数。

解：因 $X\sim U\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$，有 X 的密度函数为 $p_X(x)=\dfrac{1}{\pi},\quad -\dfrac{\pi}{2}<x<\dfrac{\pi}{2}$。

对于 $Y=\sin X$，有 $y=\sin x,\quad -\dfrac{\pi}{2}<x<\dfrac{\pi}{2}$ 严格单增，反函数 $x=h(y)=\arcsin y$，导数 $h'(y)=\dfrac{1}{\sqrt{1-y^2}}$，且当 $-\dfrac{\pi}{2}<x<\dfrac{\pi}{2}$，有 $-1<y<1$，故

$$p_Y(y)=\frac{1}{\pi}\cdot\frac{1}{\sqrt{1-y^2}}=\frac{1}{\pi\sqrt{1-y^2}},\quad -1<y<1\text{；}$$

对于 $Z=\tan X$，有 $z=\tan x,\quad -\dfrac{\pi}{2}<x<\dfrac{\pi}{2}$ 严格单增，反函数 $x=h(z)=\arctan z$，导数 $h'(z)=\dfrac{1}{1+z^2}$，且当 $-\dfrac{\pi}{2}<x<\dfrac{\pi}{2}$，有 $-\infty<z<+\infty$，故

$$p_Z(z)=\frac{1}{\pi}\cdot\frac{1}{1+z^2}=\frac{1}{\pi(1+z^2)},\quad -\infty<z<+\infty\text{。}$$

以下是一些重要结论：

（1）设随机变量 X 服从正态分布 $N(\mu,\sigma^2)$，则 $Y=aX+b,\quad(a\neq 0)$ 服从正态分布 $N(a\mu+b,a^2\sigma^2)$；

（2）设随机变量 X 服从正态分布 $N(\mu,\sigma^2)$，则 $Y=\mathrm{e}^X$ 服从对数正态分布 $LN(\mu,\sigma^2)$，密度函数为

$$p_Y(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma y}\mathrm{e}^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

（3）设随机变量 X 服从伽玛分布 $Ga(\alpha,\lambda)$，则 $Y=kX,\quad(k>0)$ 服从伽玛分布 $Ga(\alpha,\dfrac{\lambda}{k})$。

证明：（1）因 X 服从正态分布 $N(\mu,\sigma^2)$，有 X 的密度函数为 $p_X(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$。

对于 $Y=aX+b$，有 $y=ax+b,\quad(a\neq 0)$ 严格单调，反函数 $x=h(y)=\dfrac{y-b}{a}$，导数 $h'(y)=\dfrac{1}{a}$，且当 $-\infty<x<+\infty$ 时，有 $-\infty<y<+\infty$，则

$$p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2\sigma^2}\left(\frac{y-b}{a}-\mu\right)^2}\cdot\left|\frac{1}{a}\right|=\frac{1}{\sqrt{2\pi}\sigma|a|}\mathrm{e}^{-\frac{(y-a\mu-b)^2}{2\sigma^2a^2}},\quad -\infty<y<+\infty\text{，}$$

故 $Y=aX+b$ 服从正态分布 $N(a\mu+b,a^2\sigma^2)$；

注：由此结论可知，若 X 服从正态分布 $N(\mu,\sigma^2)$，则 $Y=\dfrac{X-\mu}{\sigma}$ 服从标准正态分布 $N(0,1)$，这称为正态变量的标准化。

（2）因 X 服从正态分布 $N(\mu,\sigma^2)$，有 X 的密度函数为 $p_X(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$， $-\infty<x<+\infty$ 。

对于 $Y=\mathrm{e}^X$，有 $y=\mathrm{e}^x$ 严格单增，反函数 $x=h(y)=\ln y$，导数 $h'(y)=\dfrac{1}{y}$，且当 $-\infty<x<+\infty$ 时，有 $0<y<+\infty$，故

$$p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}\cdot\left|\frac{1}{y}\right|=\frac{1}{\sqrt{2\pi}\sigma y}\mathrm{e}^{-\frac{(\ln y-\mu)^2}{2\sigma^2}},\quad 0<y<+\infty\text{；}$$

（3）因 X 服从伽玛分布 $Ga(\alpha,\lambda)$，有 X 的密度函数为 $p_X(x)=\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}$， $0<x<+\infty$ 。

对于 $Y=kX$，有 $y=kx,\ (k>0)$ 严格单增，反函数 $x=h(y)=\dfrac{y}{k}$，导数 $h'(y)=\dfrac{1}{k}$，且当 $0<x<+\infty$ 时，有 $0<y<+\infty$，则

$$p_Y(y)=\frac{\lambda^\alpha}{\Gamma(\alpha)}\left(\frac{y}{k}\right)^{\alpha-1}\mathrm{e}^{-\lambda\frac{y}{k}}\cdot\left|\frac{1}{k}\right|=\frac{\left(\frac{\lambda}{k}\right)^\alpha}{\Gamma(\alpha)}y^{\alpha-1}\mathrm{e}^{-\frac{\lambda}{k}y},\quad 0<y<+\infty\text{，}$$

故 $Y=kX$ 服从伽玛分布 $Ga(\alpha,\dfrac{\lambda}{k})$。

注：若 $g(x)$ 在 X 的支撑区间内不是严格单调函数，但可分解成若干个严格单调区间并且可导，仍然可以在每个严格单调区间上分别运用上面的定理，再将所得的各个结果重复取值合并，对应密度相加。

例 设随机变量 X 的密度函数为

$$p_X(x)=\begin{cases}\dfrac{2+2x}{9}, & -1<x<2;\\ 0, & \text{其它.}\end{cases}$$

求 $Y=|X|$ 的分布。

解：函数 $y=|x|$ 将 X 的支撑区间分成两个严格单调区间 $(-1,0)$、$(0,2)$。因 $y=|x|,\ -1<x<0$ 严格单减，反函数 $x=h(y)=-y$，导数 $h'(y)=-1$，且当 $-1<x<0$ 时，有 $0<y<1$，故

$$p_Y^{(1)}(y)=\frac{2-2y}{9}\cdot|-1|=\frac{2-2y}{9},\quad 0<y<1\text{；}$$

因 $y=|x|,\ 0<x<2$ 严格单增，反函数 $x=h(y)=y$，导数 $h'(y)=1$，且当 $0<x<2$ 时，有 $0<y<2$，故

$$p_Y^{(2)}(y)=\frac{2+2y}{9}\cdot 1=\frac{2+2y}{9},\quad 0<y<2\text{；}$$

在 $0<y<1$ 时，两部分重复，对应密度相加，故 $Y=|X|$ 的密度函数为

$$p_Y(y)=F_Y'(y)=\begin{cases}\dfrac{4}{9}, & 0<y<1;\\ \dfrac{2+2y}{9}, & 1\le y<2;\\ 0, & \text{其他.}\end{cases}$$

§2.7 分布的其他特征数

数学期望与方差是随机变量最重要的两个特征数（Characteristic Number），此外还有一些其他的特征数。

2.7.1. 矩（Moment）

定义 设 X 为随机变量，k 为正整数。

若 $E(X^k)$ 存在，则称之为 X 的 k 阶原点矩（Origin Moment），记为 μ_k。

若 $E[X-E(X)]^k$ 存在，则称之为 X 的 k 阶中心矩（Central Moment），记为 ν_k。

显然一阶原点矩 μ_1 就是数学期望 $E(X)$，一阶中心矩 ν_1 恒等于零，二阶中心矩 ν_2 就是方差 $\mathrm{Var}(X)$。

由于 $|Y|^{k-1}\le \max\{|Y|^k,1\}\le |Y|^k+1$，可见当 X 的 k 阶矩存在时，其 $k-1$ 阶矩也存在，从而低于 k 的各阶矩都存在。

中心矩 ν_k 与原点矩 μ_k 的关系为

$$\nu_k = E(X-\mu_1)^k = \sum_{i=0}^{k} C_k^i(-\mu_1)^i E(X^{k-i}) = \sum_{i=0}^{k} C_k^i(-\mu_1)^i \cdot \mu_{k-i} \quad （这里取 \mu_0 = 1）$$

$$= \mu_k - k\mu_1\mu_{k-1} + C_k^2\mu_1^2\mu_{k-2} + \cdots + (-1)^{k-2}C_k^{k-2}\mu_1^{k-2}\mu_2 + (-1)^{k-1}k\mu_1^{k-1}\mu_1 + (-1)^k\mu_1^k$$

$$= \mu_k - k\mu_1\mu_{k-1} + C_k^2\mu_1^2\mu_{k-2} + \cdots + (-1)^{k-2}C_k^{k-2}\mu_1^{k-2}\mu_2 + (-1)^{k-1}(k-1)\mu_1^k,$$

特别是前四阶中心矩分别为

$$\nu_1 = 0；\ \nu_2 = \mu_2 - \mu_1^2；\ \nu_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3；\ \nu_4 = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4。$$

例 设随机变量 X 服从正态分布 $N(\mu,\sigma^2)$，求 X 的 k 阶中心矩 ν_k。

解： 由定义

$$\nu_k = E[X-E(X)]^k = E(X-\mu)^k = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)^k \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}dx,$$

令 $y=\dfrac{x-\mu}{\sqrt{2}\sigma}$，有 $x-\mu=\sqrt{2}\sigma y$，$dx=\sqrt{2}\sigma dy$，且当 $x\to-\infty$ 时，$y\to-\infty$，当 $x\to+\infty$ 时，$y\to+\infty$，则

$$\nu_k = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(\sqrt{2}\sigma y)^k \mathrm{e}^{-y^2}\sqrt{2}\sigma dy = \frac{(\sqrt{2}\sigma)^k}{\sqrt{\pi}}\int_{-\infty}^{+\infty}y^k \mathrm{e}^{-y^2}dy,$$

当 k 为奇数时，被积函数 $y^k \mathrm{e}^{-y^2}$ 为奇函数，有 $\nu_k=0,\quad k=1,3,5,\Lambda$；当 k 为偶数时，被积函数 $y^k \mathrm{e}^{-y^2}$ 为偶函数，有

$$\nu_k = \frac{(\sqrt{2}\sigma)^k}{\sqrt{\pi}}\cdot 2\int_0^{+\infty}y^k \mathrm{e}^{-y^2}dy,$$

令 $z=y^2$，有 $y=\sqrt{z}$，$dy=\dfrac{1}{2\sqrt{z}}dz$，且当 $y=0$ 时，$z=0$，当 $y\to+\infty$ 时，$z\to+\infty$，故

$$\nu_k=\frac{(\sqrt{2}\sigma)^k}{\sqrt{\pi}}\cdot 2\int_0^{+\infty} z^{\frac{k}{2}}\,\mathrm{e}^{-z}\cdot\frac{1}{2\sqrt{z}}dz=\frac{(\sqrt{2}\sigma)^k}{\sqrt{\pi}}\int_0^{+\infty} z^{\frac{k-1}{2}}\,\mathrm{e}^{-z}dz=\frac{(\sqrt{2}\sigma)^k}{\sqrt{\pi}}\Gamma\left(\frac{k+1}{2}\right)$$

$$=\frac{(\sqrt{2}\sigma)^k}{\sqrt{\pi}}\cdot\frac{k-1}{2}\cdot\frac{k-3}{2}\cdots\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=(k-1)(k-3)\cdots 1\cdot\sigma^k,\quad k=2,4,6,\cdots。$$

特别是

$$\nu_1=0；\ \nu_2=\sigma^2；\ \nu_3=0；\ \nu_4=3\sigma^4；\ \nu_5=0；\ \nu_6=15\sigma^6。$$

2.7.2. 变异系数（Coefficient of Variation）

方差和标准差用于反映随机变量取值的波动程度，但不同的随机变量可能有不同的量纲，两个随机变量的波动程度不便于直接比较，通常使用无量纲的变异系数进行比较。

定义 设随机变量 X 的二阶矩存在，且数学期望 $E(X)\neq 0$，则称标准差与数学期望绝对值的比值

$$\frac{\sqrt{\mathrm{Var}(X)}}{|E(X)|}$$

为 X 的变异系数，记为 $\mathrm{CV}(X)$。

例 设随机变量 X 服从指数分布 $Exp(\lambda)$，求 X 的变异系数。

解：因 $E(X)=\dfrac{1}{\lambda}\neq 0$， $\mathrm{Var}(X)=\dfrac{1}{\lambda^2}$，故变异系数 $\mathrm{CV}(X)=\dfrac{\sqrt{\mathrm{Var}(X)}}{|E(X)|}=1$。

例 设随机变量 X 服从均匀分布 $U(0,\theta),\quad \theta>0$，求 X 的变异系数。

解：因 $E(X)=\dfrac{\theta}{2}\neq 0$， $\mathrm{Var}(X)=\dfrac{\theta^2}{12}$，故变异系数 $\mathrm{CV}(X)=\dfrac{\sqrt{\mathrm{Var}(X)}}{|E(X)|}=\dfrac{1}{\sqrt{3}}$。

2.7.3. 分位数（Quantile Fractile）

随机变量的分布函数是根据取值求得不超过该取值的概率，而分位数是反过来根据指定的概率求得该取值。

定义 设 X 为随机变量，$0<p<1$，若 x_p 满足

$$P\{X<x_p\}\leq p\leq P\{X\leq x_p\},$$

则称 x_p 为 X 的（下侧） p 分位数。

注：若 X 为连续随机变量， x_p 为 X 的 p 分位数，则 $F(x_p)=P\{X\leq x_p\}=p$。

标准正态分布 $N(0,1)$ 的 p 分位数记为 u_p，有 $\Phi(u_p)=p$，常见的标准正态分布 p 分位数有

$$u_{0.9}=1.28，\ u_{0.95}=1.645，\ u_{0.975}=1.96，\ u_{0.99}=2.33，$$

可通过附表 2 查到。

定义 设 X 为随机变量，$0<p<1$，若 x_p 满足

$$P\{X>x_p^*\}\leq p\leq P\{X\geq x_p^*\},$$

则称 x_p^* 为 X 的（上侧） p 分位数。

注： 若 X 为连续随机变量， x_p^* 为 X 的上侧 p 分位数，则 $1-F(x_p^*)=P\{X>x_p^*\}=p$ 。

显然上侧 p 分位数与下侧 p 分位数的关系为 $x_p^*=x_{1-p}$ 。

2.7.4. 中位数（Median）

随机变量 X 的 0.5 分位数 $x_{0.5}$ 称为中位数。

例 设随机变量 X 服从指数分布 $Exp(\lambda)$ ，求 X 的中位数 $x_{0.5}$ 。

解： 因 $F(x_{0.5})=1-\mathrm{e}^{-\lambda x_{0.5}}=0.5$ ，有 $\mathrm{e}^{-\lambda x_{0.5}}=0.5$ ，故中位数 $x_{0.5}=\dfrac{\ln 2}{\lambda}$ 。

2.7.5. 偏度系数（Skewness）

偏度系数用于反映随机变量分布曲线的偏斜程度。对于连续随机变量，若密度函数为单峰函数（即只有唯一的极值），则根据平均值与峰值的位置确定偏斜程度，平均值与峰值相距越远表明偏斜程度越大。如果平均值大于峰值，称为正偏或右偏，此时密度函数图形的“尾部”右端比左端长；如果平均值小于峰值，称为负偏或左偏，此时密度函数图形的“尾部”左端比右端长。偏度系数用三阶中心矩刻画。

定义 设随机变量 X 的三阶矩存在，且非单点分布，则称三阶中心矩与标准差三次方的比值

$$\beta_S=\frac{\nu_3}{\nu_2^{3/2}}$$

为 X 的偏度系数。

当偏度系数 $\beta_S>0$ 时，则分布为正偏；当偏度系数 $\beta_S<0$ 时，则分布为负偏。

例 设随机变量 X 服从指数分布 $Exp(\lambda)$ ，求 X 的偏度系数 β_S 。

解： 因 X 的密度函数为 $p(x)=\lambda\mathrm{e}^{-\lambda x},\quad x>0$ ，则

$$\mu_k=E(X^k)=\int_0^{+\infty}x^k\cdot\lambda\,\mathrm{e}^{-\lambda x}\,dx=\int_0^{+\infty}\left(\frac{y}{\lambda}\right)^k\cdot\lambda\,\mathrm{e}^{-y}\cdot\frac{1}{\lambda}\,dy=\frac{1}{\lambda^k}\int_0^{+\infty}y^k\,\mathrm{e}^{-y}\,dy=\frac{1}{\lambda^k}\Gamma(k+1)=\frac{k!}{\lambda^k},$$

可得

$$\nu_2=\mu_2-\mu_1^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2},\quad \nu_3=\mu_3-3\mu_1\mu_2+2\mu_1^3=\frac{6}{\lambda^3}-3\cdot\frac{1}{\lambda}\cdot\frac{2}{\lambda^2}+2\cdot\left(\frac{1}{\lambda}\right)^3=\frac{2}{\lambda^3},$$

故偏度系数

$$\beta_S=\frac{\nu_3}{\nu_2^{3/2}}=2。$$

2.7.6.　峰度系数（Kurtosis）

峰度系数用于反映随机变量分布曲线的峰峭程度，即“峰顶”的尖峭或扁平程度，这不能用二阶矩刻画。如正态分布 $N(0,1)$、均匀分布 $U(-\sqrt{3},\sqrt{3})$、指数分布 $Exp(\lambda)$ 的方差都是 1，但它们的峰峭程度显然不同。峰度系数用四阶中心矩刻画。

定义　设随机变量 X 的四阶矩存在，且非单点分布，则称四阶中心矩与标准差四次方的比值减去 3 之差

$$\beta_K=\frac{\nu_4}{\nu_2^2}-3$$

为 X 的峰度系数。

因正态分布的二阶中心矩 $\nu_2=\sigma^2$，四阶中心矩 $\nu_4=3\sigma^4$，则正态分布的峰度系数 $\beta_K=0$。事实上，在峰度系数的定义中减去 3 就是以正态分布的“峰顶”为标准反映峰峭程度。当峰度系数 $\beta_K>0$ 时，则分布曲线的“峰顶”比正态分布更尖峭；当峰度系数 $\beta_K<0$ 时，则分布曲线的“峰顶”比正态分布更扁平。

例　设随机变量 X 服从指数分布 $Exp(\lambda)$，求 X 的峰度系数 β_K。

解： 因 $\mu_k=\dfrac{k!}{\lambda^k}$，有

$$\nu_4=\mu_4-4\mu_1\mu_3+6\mu_1^2\mu_2-3\mu_1^4=\frac{24}{\lambda^4}-4\cdot\frac{1}{\lambda}\cdot\frac{6}{\lambda^3}+6\cdot\left(\frac{1}{\lambda}\right)^2\cdot\frac{2}{\lambda^2}-3\cdot\left(\frac{1}{\lambda}\right)^4=\frac{9}{\lambda^4},$$

故峰度系数

$$\beta_K=\frac{\nu_4}{\nu_2^2}-3=6。$$

例　设随机变量 X 服从均匀分布 $U(a,b)$，求 X 的峰度系数 β_K。

解： 因 $E(X)=\dfrac{a+b}{2}$，则

$$\nu_k=E\left(X-\frac{a+b}{2}\right)^k=\int_a^b\left(x-\frac{a+b}{2}\right)^k\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\frac{1}{k+1}\left(x-\frac{a+b}{2}\right)^{k+1}\Bigg|_a^b$$

$$=\frac{1}{b-a}\cdot\frac{1}{k+1}\left[\left(\frac{b-a}{2}\right)^{k+1}-\left(-\frac{b-a}{2}\right)^{k+1}\right]=\begin{cases}\dfrac{(b-a)^k}{2^k(k+1)}, & k\text{ 为偶数；}\\ 0, & k\text{ 为奇数.}\end{cases}$$

即 $\nu_2=\dfrac{(b-a)^2}{12}$，$\nu_4=\dfrac{(b-a)^4}{80}$，故峰度系数

$$\beta_K=\frac{\nu_4}{\nu_2^2}-3=\frac{144}{80}-3=-\frac{6}{5}。$$