第1章

概率知识

概率论是应用数理统计的重要工具，是研究随机现象统计规律性的数学学科，要用一章的篇幅介绍整个概率论的内容是不现实的，因此本书假定读者已经具备必要的概率论知识，仅仅为了检索或便于复习的考虑，对必备的概率论知识进行简要的归纳和总结，完全没有基础的读者应该参阅相关的概率论教材.

随机现象、随机试验、随机事件是我们进入概率论世界的三把钥匙. 自然界中存在着大量的随机现象，称发生与否具有偶然性的事件为随机事件，通过随机试验可以观察到随机事件. 第一个问题是既然是随机事件，规律性作何解释？这里指的是统计规律，即在大量重复试验时，这些随机事件所显示的必然的本质特征. 第二个问题是为什么说它是一门数学学科？不是研究具体的试验和具体的事件，而是通过数学抽象和概括，用定量化的数学语言去研究一般的规律性，因此是数学的一个分支.

关于概率论的起源简单说明如下：作为机会游戏在纪元前就开始了，毋庸讳言，概率论萌芽于早期的赌博这一今天比较忌讳但又普遍存在的社会现象. 概率论起源于赌博，还可以用 1494 年 Pacciolo 的著名分点问题①(the problem of points)来说明. 围绕着这一问题，Cardano 及随后的很多人进行了更深入的研究，特别是 Pascal 和 Fermat 一生致力于赌博问题的研究并把概率论引入正途. 到 20 世纪初，概率论作为一门数学分支已经牢牢的站住了脚，1933 年 Kolmogorov 提出的概率空间及公理化定义使这一学科进一步完善.

---

① 分点问题：1494 年，在意大利出版的一本计算数学的教科书中，Pacciolo 写道：若一个比赛赢 6 次才算赢，两个赌徒，1 胜 5 次，另一胜 2 次，因故赌博中断，赌金应按 5∶2 分给他们. Cardano 认为第一个赌徒在以后的比赛中只有 5 种可能的结果，胜第 1 次，胜第 2 次，胜第 3 次，胜第 4 次或者全部输掉，因此应按(1+2+3+4)∶1=10∶1 来分. 实际上正确的结果为 15∶1.

1.1 概率的计算

将复杂的事件化为简单事件的关系和运算，有利于对复杂事件统计规律性的描述和研究.

1.1.1 事件的关系与运算

如果用英文字母 A,B,C 等来表示事件，通过集合论知识，可以方便地表示子事件 $A\subset B$，两事件相等 $A=B$，和事件 $A\cup B$，积事件 AB（或 $A\cap B$），差事件 $C=A-B$，互斥事件 $AB=\varnothing$，逆事件 $\overline{A}$，完备事件组 $A_1,A_2,\cdots,A_n$ $\left(\text{两两互斥且} \bigcup\limits_{i=1}^{n} A_i=\Omega\right)$ 等，并可以建立下面的运算规律.

(1) 交换律：$A\cup B=B\cup A,AB=BA$；

(2) 结合律：$(A\cup B)\cup C=A\cup(B\cup C),(AB)C=A(BC)$；

(3) 分配律：$(A\cup B)\cap C=AC\cup BC,(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$；

(4) 德摩根(De Morgan)定理(对偶律)：$\overline{\bigcup\limits_{i=1}^{n} A_i}=\bigcap\limits_{i=1}^{n}\overline{A}_i,\overline{\bigcap\limits_{i=1}^{n} A_i}=\bigcup\limits_{i=1}^{n}\overline{A}_i$；

(5) 差化积：$A-B=A\overline{B}$；

(6) 吸收律：如果 $A\subset B$，则 $A\cup B=B,AB=A$.

注意：当 $AB=\varnothing$ 时，$A\cup B$ 常表示为 $A+B$.

1.1.2 概率的统计定义与公理化定义

在概率论早期的实践中，往往将随机事件的频率 $f_n(A)=\dfrac{r}{n}$ 当作概率. 人们发现，随着试验次数的增加，频率具有一定的稳定性，并且会接近事件的概率. 直到 20 世纪上半叶，一系列极限定理的建立，从理论上支持了这种观点. 建立在这种观点之上的概率定义就称为概率的统计定义.

众所周知，数学在 20 世纪初开始了一场公理化革命，作为数学分支的概率论也不例外，在 1933 年苏联著名数学家 Kolmogorov 提出了概率的公理化定义.

定义 1.1.1 设 Ω 是试验 E 的样本空间，对于试验 E 的每个随机事件 $A(A\subset\Omega)$，有一个实数 $P(A)$ 与之对应，且 $P(A)$ 满足

公理 1(非负性)：$P(A)\geqslant 0$；

公理 2(规范性)：$P(\Omega)=1$；

公理 3(可列可加性)：$P\left(\sum\limits_{i=1}^{\infty} A_i\right)=\sum\limits_{i=1}^{\infty} P(A_i)$；

则称 $P(A)$为事件 A 的概率.

利用这三条公理,可以推导出概率的许多性质,下面是其中的一些,证明请读者自行补上或参阅相关教材.

(1) 不可能事件的概率为零 $P(\varnothing)=0$;

(2) (有限可加性) $P\left(\sum_{i=1}^{m} A_i\right)=\sum_{i=1}^{m} P(A_i)$;

(3) 设 A 为任一随机事件,则 $P(A)=1-P(\overline{A})$;

(4) 设 A,B 为任两随机事件,则 $P(B-A)=P(B)-P(AB)$;

(5) (单调性)若 $A\subset B$,则 $P(A)\leqslant P(B)$;

(6) 设 A,B 为任两事件,则 $P(A\cup B)=P(A)+P(B)-P(AB)$.

1.1.3 古典概型与古典概率计算

古典概型是这样的随机试验:①基本事件总数是有限的;②基本事件发生的可能性相等.在古典概型下,概率的计算很简单,为

$$P(A)=\frac{r}{n}=\frac{A\text{ 包含的基本事件数}}{\text{基本事件总数}} \tag{1.1.1}$$

某些时候或者对于某些特定的问题,n 和 r 的求取会很困难,需要用到排列组合的知识,对这方面不熟悉的读者请参考相关教材.对于古典概型中等可能的判断可以通过对称性、均衡性、平凡性等作出.古典概率具有如下性质:

(1) 非负有界性 $0\leqslant P(A)\leqslant 1$;

(2) 规范性 $P(\Omega)=1$;

(3) 有限可加性 $P\left(\sum_{i=1}^{m} A_i\right)=\sum_{i=1}^{m} P(A_i)$.

例 1.1.1 设有 40 件产品,其中有 10 件为次品其余为正品,现从中任取 5 件,求取出的 5 件产品中至少有 4 件次品的概率.

解 "至少有 4 件次品"这一事件可表示成"恰有 4 件次品"与"5 件全是次品"这两个互斥事件的和事件.

设 A="恰有 4 件次品",B="5 件全是次品",C="至少有 4 件次品",则

$$C=A\cup B,\quad 且\quad AB=\varnothing$$

利用概率的有限可加性,有

$$P(C)=P(A\cup B)=P(A)+P(B)$$

而基本事件总数 $n=\mathrm{C}_{40}^{5}$(这里 C_n^m 表示从 n 个元素中抽取 m 个元素的全部组合数),A 所含基本事件数为 $\mathrm{C}_{10}^{4}\mathrm{C}_{30}^{1}$,$B$ 所含基本事件数为 C_{10}^{5},故

$$P(A)=\frac{\mathrm{C}_{10}^{4}\mathrm{C}_{30}^{1}}{\mathrm{C}_{40}^{5}}=\frac{175}{18\,278}\approx 0.0096$$

$$P(B)=\frac{C_{10}^{5}}{C_{40}^{5}}=\frac{7}{18\,278}\approx 0.0004$$

因此

$$P(C)=P(A)+P(B)\approx 0.01$$

1.1.4 几何概型与几何概率计算

与古典概型相比，几何概型的基本试验结果可以是无穷多个，但它要求试验结果在空间区域中是均匀的，同时，该空间区域是可测的，整个概率空间具有有限测度. 在几何概型下，几何概率的计算公式如下：

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{A\text{ 的几何测度}}{\text{概率空间的几何测度}} \tag{1.1.2}$$

几何概率也有如下性质：

(1) 非负有界性 $0\leqslant P(A)\leqslant 1$；

(2) 规范性 $P(\Omega)=1$；

(3) 可列可加性 $P\left(\sum\limits_{i=1}^{\infty}A_i\right)=\sum\limits_{i=1}^{\infty}P(A_i)$.

例 1.1.2(蒲丰投针问题) 在平面上画一些距离都等于 a 的平行线，向此平面上任投一长为 $l(l<a)$的针，试求此针与任一平行线相交的概率.

解 以 x 表示针的中点 M 到最近一条平行线的距离，φ 表示针与平行线的交角. 据针与平行线的位置关系(如图 1.1.1)，显然有

$$0\leqslant x\leqslant \frac{a}{2},\quad 0\leqslant \varphi\leqslant \pi$$

满足上面两个不等式的点(φ,x)在 φ-x 平面上构成一个矩形 G，如图 1.1.2.

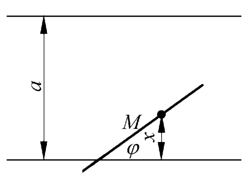

图 1.1.1

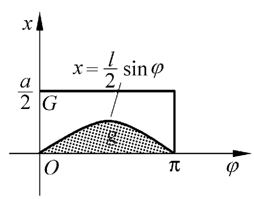

图 1.1.2

为使针与平行线相交，充要条件是

$$x\leqslant \frac{l}{2}\sin\varphi$$

此不等式构成图 1.1.2 中 G 的一个子区域 g. 于是，投针问题就等价于向 G 内投掷一点，求点落入 g 内的概率问题. 这是一个几何概率问题，于是得

$$P(\text{针与任一平行线相交}) = \frac{g\text{ 的面积}}{G\text{ 的面积}} = \frac{\frac{1}{2}\int_0^{\pi} l\sin\varphi \mathrm{d}\varphi}{\frac{1}{2}\pi a} = \frac{2l}{\pi a}$$

通过这个公式，Fox 在 1884 年取 $a=1,l=0.75$，投针 1030 次试验，实际相交次数 489 次，先计算出“针与任一平行线的相交”频率作为其概率的近似值，代入上面的公式，求出 π 的近似值为 3.1595.

1.2 一维随机变量

描述随机试验结果的变量 X 取什么值，在每次试验之前是不能确定的(试验前只知道它所有可能的取值或取值范围)，它们的取值依赖于试验的结果，由于试验的结果是随机的，所以它们的取值具有随机性. 像这种取值由随机试验的结果所确定的变量称为随机变量. 随机变量主要有离散型随机变量与连续型随机变量两大类.

1.2.1 一维离散型随机变量及其分布列

设 X 是一个随机变量，如果 X 可能取的值只有有限个或可数个，则称 X 是一维离散型随机变量，简称离散型随机变量.

设离散型随机变量 X 可能取值为 $x_1,x_2,\cdots$(其中任何两个都不相同)，X 取各个值的概率分别是 $p_1,p_2,\cdots$，记成

$$P(X=x_k)=p_k,\quad k=1,2,\cdots \tag{1.2.1}$$

称 p_k 为离散型随机变量 X 的分布列，也称为概率分布. 显然分布列满足如下条件：

$$p_k \geqslant 0(k=1,2,\cdots) \quad 且 \quad \sum_{k=1}^{\infty} p_k = 1$$

1.2.2 一维连续型随机变量及其密度函数

对于随机变量 X，如果存在一个非负可积函数 $f(x)$，对任意实数 x，有

$$P(X \leqslant x) = \int_{-\infty}^{x} f(u)\mathrm{d}u \tag{1.2.2}$$

则称 X 为连续型随机变量，称 $f(x)$为 X 的密度函数或分布密度. 密度函数具有如下两个性质：

(1) $f(x)\geqslant 0, x\in(-\infty,+\infty)$；

(2) $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$.

只要得到了连续型随机变量 X 的密度函数 $f(x)$，就可以解决任何事件的概率计算问题，比如

$$P(a < X < b) = \int_a^b f(x)\,\mathrm{d}x, \quad \text{对 } a, b \text{ 取任意实数} \tag{1.2.3}$$

1.2.3 一维随机变量的分布函数

设 X 为随机变量，x 为任意实数，称 $F(x)=P(X\leqslant x)$ 为随机变量 X 的分布函数，简称分布. 分布函数 $F(x)$ 与离散型随机变量 X 的分布列以及连续型随机变量 X 的密度函数 $f(x)$ 有如下关系：

$$F(x) = P(X \leqslant x) = \begin{cases} \sum\limits_{x_k \leqslant x} P(X = x_k), & X \text{ 是离散随机变量} \\ \int_{-\infty}^{x} f(x)\,\mathrm{d}x, & X \text{ 是连续随机变量} \end{cases} \tag{1.2.4}$$

分布函数 $F(x)$ 有如下性质：

(1) $0\leqslant F(x)\leqslant 1$；

(2) $F(-\infty)=0, F(+\infty)=1$；

(3) $F(x)$ 是单调递增函数；

(4) $F(x)$ 是右连续函数.

特别地，连续型随机变量 X 的分布有如下性质：

(1) 若 $f(x)$ 在 x 点连续，则 $F(x)$ 在 x 点可导，且 $F'(x)=f(x)$；

(2) $F(x)$ 是连续函数.

例 1.2.1 现有 7 件产品，其中一等品 4 件，二等品 3 件，从中任取 3 件(不放回抽取). 求：

(1) 抽取的 3 件产品中含一等品件数 X 的分布列；

(2) X 的分布函数；

(3) 抽取的 3 件产品中至少含一件一等品的概率.

解 (1) X 的可能取值为：0,1,2,3，利用古典概率及性质计算：

$$P(X=0) = \frac{\mathrm{C}_4^0\mathrm{C}_3^3}{\mathrm{C}_7^3} = \frac{1}{35}, \quad P(X=1) = \frac{\mathrm{C}_4^1\mathrm{C}_3^2}{\mathrm{C}_7^3} = \frac{12}{35}$$

$$P(X=2) = \frac{\mathrm{C}_4^2\mathrm{C}_3^1}{\mathrm{C}_7^3} = \frac{18}{35}, \quad P(X=3) = \frac{\mathrm{C}_4^3\mathrm{C}_3^0}{\mathrm{C}_7^3} = \frac{4}{35}$$

X 的分布列为

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{35} & \frac{12}{35} & \frac{18}{35} & \frac{4}{35} \end{pmatrix}$$

显然，该分布列满足性质：$p_k \geqslant 0(k=1,2,\cdots)$ 和 $\sum\limits_{k=1}^{\infty} p_k = 1$.

(2) 根据分布函数定义,即参照公式(1.2.4),计算 X 的分布函数如下:

$$F(x)=P(X\leqslant x)=\begin{cases}0, & x<0\\ \dfrac{1}{35}, & 0\leqslant x<1\\ \dfrac{13}{35}, & 1\leqslant x<2\\ \dfrac{31}{35}, & 2\leqslant x<3\\ 1, & x\geqslant 3\end{cases}$$

(3) 所求概率可以表示为$\{X\geqslant 1\}$,即

$$P(X\geqslant 1)=1-P(X<1)=1-P(X=0)=\frac{34}{35}$$

例 1.2.2 设连续型随机变量 X 的分布函数为

$$F(x)=\begin{cases}Ae^{x}, & x<0\\ B, & 0\leqslant x<1\\ 1-Ae^{-(x-1)}, & x\geqslant 1\end{cases}$$

求:(1)常数 A,B;(2)X 的密度函数;(3)概率 $P\left(X>\dfrac{1}{3}\right)$.

解 (1) 由于 X 是连续型随机变量,故 $F(x)$是连续函数,所以 $F(x)$在 $x=0$ 和 $x=1$ 两点连续,由

$$\lim_{x\to 0^-}F(x)=\lim_{x\to 0^-}Ae^{x}=A,\quad \lim_{x\to 0^+}F(x)=\lim_{x\to 0^+}B=B$$

可知 $A=B$. 又由于

$$\lim_{x\to 1^-}F(x)=\lim_{x\to 1^-}B=B$$

$$\lim_{x\to 1^+}F(x)=\lim_{x\to 1^+}(1-Ae^{-(x-1)})=1-A$$

可知 $B=1-A$,故有 $A=B=\dfrac{1}{2}$,于是

$$F(x)=\begin{cases}\dfrac{1}{2}e^{x}, & x<0\\ \dfrac{1}{2}, & 0\leqslant x<1\\ 1-\dfrac{1}{2}e^{-(x-1)}, & x\geqslant 1\end{cases}$$

(2) 因为 X 的密度函数在连续点处与分布函数有关系 $f(x)=F'(x)$,所以 X 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2}e^{x}, & x<0\\ 0, & 0\leqslant x<1\\ \dfrac{1}{2}e^{-(x-1)}, & x\geqslant 1\end{cases}$$

(3) $P\left(X>\dfrac{1}{3}\right)=1-P\left(X\leqslant\dfrac{1}{3}\right)=1-F\left(\dfrac{1}{3}\right)=1-\dfrac{1}{2}=\dfrac{1}{2}$

或

$$P\left(X>\frac{1}{3}\right)=\int_{\frac{1}{3}}^{+\infty}f(x)\mathrm{d}x=\int_{\frac{1}{3}}^{1}0\mathrm{d}x+\int_{1}^{+\infty}\frac{1}{2}e^{-(x-1)}\mathrm{d}x=\frac{1}{2}$$

1.2.4 常见的分布列或密度函数

1. 二项分布

若随机变量 X 满足 $P(X=k)=C_n^k p^k q^{n-k}$, $k=0,1,\cdots,n$,其中 $0<p<1$, $q=1-p$,则称 X 服从参数为 (n,p) 的二项分布,记作 $X\sim B(n,p)$. 当 $n=1$ 时,称 $B(1,p)$ 为两点分布(或 0-1 分布). 图 1.2.1 给出了 $B(10,0.2)$ 的概率分布图形.

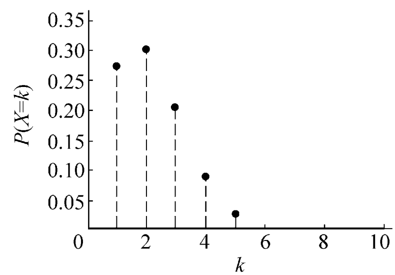

图 1.2.1 $B(10,0.2)$概率分布

二项分布的实际背景:在 n 重 Bernoulli 试验(n 次重复独立试验,每次试验只有两种可能结果)中,若每次试验事件 A 出现的概率为 p,则事件 A 出现的次数 X 就服从参数为 (n,p) 的二项分布. 例如在产品检验中,观察次品数 X 的分布状况;又如观察某幢大楼电梯的运行状况,需要了解每年电梯出故障次数 X 的分布情况等.

二项分布具有线性可加性:若 $X_1\sim B(n_1,p)$, $X_2\sim B(n_2,p)$,并且 X_1 与 X_2 独立,则 $X_1+X_2\sim B(n_1+n_2,p)$;特别地,若 $X_1,X_2,\cdots,X_n$ 独立同分布,均服从 0-1 分布,则 $\sum\limits_{i=1}^{n}X_i\sim B(n,p)$.

2. Poisson 分布

若随机变量 X 满足 $P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\cdots$, $\lambda>0$,则称 X 服从参数为 λ 的 Poisson 分布,记作 $X\sim P(\lambda)$. 图 1.2.2 给出了 $\lambda=1$ 的 Poisson 分布图形.

常见的例子有:一块钢板上出现的气泡数 X,一本书上的印刷错误数 X,某超市排队等候的人数 X 等都服从 Poisson 分布.

3. 几何分布

若随机变量 X 满足 $P(X=k)=pq^{k-1}$，$k=1,2,\cdots$，其中 $0<p<1$，$q=1-p$，称 X 服从参数为 p 的几何分布，记作 $X\sim G(p)$.

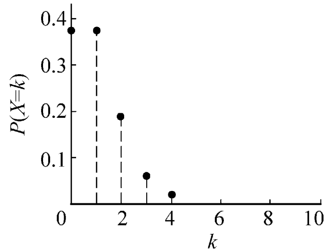

图 1.2.2　**P(1)概率分布**

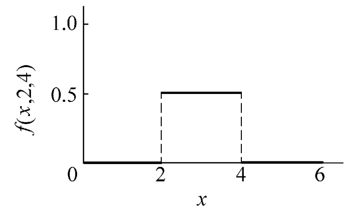

图 1.2.3　**U[2,4]分布密度函数曲线**

4. 均匀分布

若随机变量 X 的分布密度为

$$f(x,a,b)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b\\ 0, & 其他\end{cases}$$

则称 X 服从区间 $[a,b]$ 上的均匀分布，简记为 $X\sim U[a,b]$. 图 1.2.3 给出了区间 $[2,4]$ 上的均匀分布密度函数的图形.

5. 指数分布

若随机变量 X 的分布密度为

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

其中 $\lambda>0$ 为常数，则称 X 服从参数为 λ 的指数分布，记为 $X\sim\Gamma(1,\lambda)$. $\lambda=1$ 或 2 时密度函数图形如图 1.2.4 所示.

应用范围广泛的一类寿命问题如电子元件的寿命、设备的寿命、人的寿命等，其分布特征都与指数分布相近.

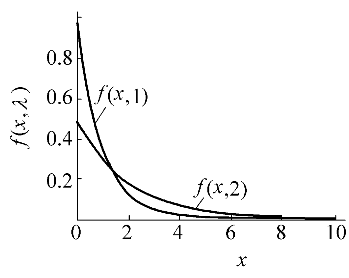

图 1.2.4　**指数分布密度函数曲线**

6. 正态分布

若随机变量 X 的分布密度为

$$f(x,\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

其中 $\sigma>0$,μ 为常数,则称 X 服从参数为(μ,σ^2)的正态分布,记作 $X\sim N(\mu,\sigma^2)$. 特别地,当 $\mu=0$,$\sigma=1$ 时,则 $X\sim N(0,1)$是正态分布的特殊情形,称为标准正态分布,其密度函数为

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

它的分布函数记为 $\Phi(x)$. 图 1.2.5 给出了参数为(2,1)和(2,3)时正态分布密度函数的图形.

正态分布在数理统计中起着特别重要的作用,其应用例子举不胜举,如人的身高、动物的睡眠时间、产品的某规格尺寸上的测量误差、某地区粮食产量等都是随机变量,其分布特征服从或近似服从正态分布.

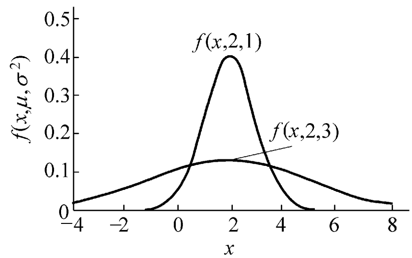

图 1.2.5　正态分布密度函数曲线

正态分布具有如下性质.

① 分布密度 $f(x)$ 关于轴 $x=\mu$ 对称,如图 1.2.5 所示.

② 密度曲线呈单峰状,最大值为$(\sqrt{2\pi}\sigma)^{-1}$,随 σ 的值增加而减少.

③ 一般正态分布函数 $F(x)$与标准正态分布函数 $\Phi(x)$具有如下关系:

$$F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

④ 正态分布具有线性性质,即若 $X\sim N(\mu_1,\sigma_1^2)$,$Y\sim N(\mu_2,\sigma_2^2)$,并且 X 与 Y 独立,则

$$cX\pm dY\sim N(c\mu_1\pm d\mu_2,c^2\sigma_1^2+d^2\sigma_2^2)$$

例 1.2.3　设顾客到某银行窗口等待服务的时间 X(单位:min)服从指数分布,其密度函数为

$$f(x)=\begin{cases}\frac{1}{5}e^{-\frac{x}{5}}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

某顾客在窗口等待服务,如超过 10min 他就离开. 他一个月要到银行 5 次,以 Y 表示一个月内未等到服务而离开窗口的次数,试写出 Y 的分布列,并求 $P(Y\geqslant 1)$.

解　Y 的取值为 0,1,2,3,4,5,且 $Y\sim B(5,p)$(二项分布),而

$$p=P(X>10)=\int_{10}^{+\infty}f(x)\mathrm{d}x=\int_{10}^{+\infty}\frac{1}{5}e^{-\frac{x}{5}}\mathrm{d}x=e^{-2}$$

故 Y 的分布列为

$$P(Y=k)=C_5^k e^{-2k}(1-e^{-2})^{5-k},\quad k=0,1,2,3,4,5$$