概率统计第四讲：
随机变量

史灵生　清华数学系

1、本讲提要

1. 随机变量
   - 离散型
   - 连续型
   - 分布函数
2. 数学期望
   - 定义
   - 函数的期望
   - 性质

2、二点（Bernoulli）分布

考虑一个只有两种可能结果的随机试验，则$\Omega = \{\omega_1, \omega_2\}$。

$$P(\{\omega_1\}) = p, \quad P(\{\omega_2\}) = 1 - p,$$

$$X(\omega) = \begin{cases} 1 & \omega = \omega_1, \\ 0 & \omega = \omega_2. \end{cases}$$

它的概率分布是：

X	0	1
P	$1-p$	p

此分布称为二点（*Bernoulli*）分布，记为$X \sim b(1, p)$。

3、二项分布

- Bernoulli试验是指相继独立地重复一个试验，在每一次试验中，只有两个不同结果，它们出现的概率分别是p与q，$p+q=1$。这一概率模型称为Bernoulli概型。
- 连续独立地抛掷一个硬币n次。记N是正面出现的次数。假设每次抛掷正面出现的概率是p，则
$$P(\{\omega\in\Omega : N(\omega)=k\}) = \binom{n}{k}p^k(1-p)^{n-k} \quad (k=0,1,\ldots,n).$$

N的概率分布是：

N	0	1	$\ldots$	k	$\ldots$	n
P	$(1-p)^n$	$np(1-p)^{n-1}$	$\ldots$	$\binom{n}{k}p^k(1-p)^{n-k}$	$\ldots$	p^n

此分布称为二项分布，记为$N\sim b(n,p)$。

注：$P\left(\bigcup_{k=0}^{n}\{\omega : N(\omega)=k\}\right) = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = 1$。

4、几何分布

- 连续独立地抛掷一个硬币，直到正面出现为止。假设每次抛正面出现的概率是p。记M是在正面首次出现时所抛的次数.
- 令A_i：第i次抛掷为正面，则

$$
\begin{aligned}
P(\{\omega \in \Omega : M(\omega) = 1\}) &= P(A_1) = p, \\
P(\{\omega \in \Omega : M(\omega) = k\}) &= P\left(\bigcap_{i=1}^{k-1} \bar{A}_i A_k\right) \\
&= (1-p)^{k-1} p, \quad (k \in \mathbb{N}).
\end{aligned}
$$

- M的概率分布是：

M	1	2	$\dots$	k	$\dots$
P	p	$(1-p)p$	$\dots$	$(1-p)^{k-1}p$	$\dots$

此分布称为几何分布，记为$M \sim G(p)$。

注: $P\left(\bigcup_{k=1}^{\infty} \{\omega : M(\omega) = k\}\right) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = \frac{p}{1-(1-p)} = 1$。

5、离散随机变量

- 随机变量仅可能取有限或可列个值[illegible]为离散型随机变量.
- 一个离散型随机变量X的分布列可以表示成：

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

- 或记成

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}.$$

分布列的基本性质：

1. 正性：$P(\{\omega : X(\omega) = x_n\}) = p_n > 0$;
2. 正则性：$\sum_n p_n = P\left(\bigcup_n \{\omega : X(\omega) = x_n\}\right) = P(\Omega) = 1$。

注：$\{x_1, x_2, \dots, x_n, \dots\}$为某个随机变量的分布$\Leftrightarrow$它满足(1)与(2)。

6、离散随机变量

一个离散型随机变量X的分布列可以表示成：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

注：

- 对于离散型随机变量，如果知道了它的概率分布，也就知道了该随机变量取值的概率规律。
- 在这个意义上，我们说：**离散型随机变量由它的概率分布唯一确定**。这和我们将要学习的连续型随机变量有本质区别！（连续型由它的概率密度唯一确定）
- 注意二点分布、二项分布和几何分布的关系。

7、系统的寿命

例

在可靠性理论中，若设备的失效率为常数$\lambda > 0$且将设备的寿命记为X，则它的统计规律如何？

解：

已知$P(X > t) = e^{-\lambda t}$，所以，对$0 \leq s < t$，有

$$\begin{aligned} P(s < X \leq t) &= P(X > s) - P(X > t) \\ &= e^{-\lambda s} - e^{-\lambda t} \\ &= \int_s^t \lambda e^{-\lambda x} \mathrm{d}x. \end{aligned}$$

8、连续型随机变量

定义

称随机变量X是连续型的，指存在非负可积函数$p : \mathbb{R} \to \mathbb{R}$（即$\int_{-\infty}^{\infty} p(x)\mathrm{d}x$收敛）使得

$$P(a < X \leq b) = \int_a^b p(x)\mathrm{d}x, \qquad \forall a < b.$$

这时称p为X的一个概率密度函数（简称密度）。

注：失效率为常数λ的系统的寿命X密度为$\lambda e^{-\lambda x} I_{[0,\infty)}(x)$。

密度的基本性质：

1. 非负性：$p(x) \geq 0$；
2. 正则性：$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$。

9、连续型随机变量

注：

1. 称$B\subset\mathbb{R}$是一个*Borel*集，如果它可以由可数多个区间经交、并、补运算得到。
2. 对连续型随机变量X和任意Borel集$B\subset\mathbb{R}$，
$$P(X\in B)=\int_B p(x)\mathrm{d}x.$$
3. 特别地，如果p在a处连续，则
$$P(a<X\le a+\Delta x)=\int_a^{a+\Delta x}p(u)\mathrm{d}u\approx p(a)\Delta x$$
从而$P(a<X\le a+\Delta x)/\Delta x\to p(a)$。
4. 一个连续型随机变量的概率密度函数不唯一，它们可以在个别处取不同的值。
5. 对连续型随机变量X，$P(X=a)=0$（$\forall a\in\mathbb{R}$）。

10、概率分布函数

定义

- 设$(\Omega, \mathcal{F}, P)$是一个概率空间。称函数$X : \Omega \to \mathbb{R}$是一个随机变量（*random variable*），如果对任何$x \in \mathbb{R}$，$\{\omega \in \Omega : X(\omega) \leq x\}$（简记为$\{X \leq x\}$或$\{X \in (-\infty, x]\}$）是$\mathcal{F}$中的事件。
- 这时称函数$F_X : \mathbb{R} \to \mathbb{R}$，

$$F_X(x) = P(X \leq x)$$

为随机变量X的概率分布函数（*probability distribution function*，简称分布函数）。

11、概率分布函数

注：

1. 随机变量X为离散型当且仅当它的分布函数是阶梯函数。且所有可能跳跃处就是它所有可能取值处，对应的跃度就是取该值的概率大小。
2. 当随机变量X为连续型时，有
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(u)\mathrm{d}u.$$
故$F(x)$不仅是连续的而且分段可导；反之，若随机变量的分布函数分段可导则它是连续型的。此时导函数就是密度。
3. 连续型随机变量的分布函数必是连续的，但只具有连续分布函数的随机变量未必是连续型的。(参考Brown运动轨道[1])

12、分布函数的性质

定理

1. F_X是单调不减、右连续函数，满足 $F_X(-\infty) = 0 \leq F_X(x) \leq 1 = F_X(+\infty)$；
2. 对任何区间$B \subset \mathbb{R}$，$P(X \in B)$可用F_X表达。特别是，$P(X = x) = F_X(x) - \lim_{z \to x-} F_X(z)$，如果$F_X$在$x$连续，则$P(X = x) = 0$。

- 设$-\infty < a < b < +\infty$。则

$$F_X(b) - F_X(a) = P(a < X \leq b) \geq 0.$$

- 因此F_X单调不减、有界，所以F_X在任何$x \in \mathbb{R}$处有左、右极限，在$-\infty$和$+\infty$有极限。

13、分布函数的性质

- 对$-\infty \le c < a \le b < d \le +\infty$，

$$F_X(b) - \lim_{a\to c+} F_X(a) = \lim_{a\to c+} P(a < X \le b) = P\left(\bigcup_{a\to c+} \{a < X \le b\}\right) = P(c < X \le b),$$

- $$\lim_{b\to d-} F_X(b) - F_X(a) = \lim_{b\to d-} P(a < X \le b) = P\left(\bigcup_{b\to d-} \{a < X \le b\}\right) = P(a < X < d)。$$

- 如果c, d是有限数，则

$$\lim_{a\to c+} F_X(a) = F_X(c), \qquad \lim_{b\to d-} F_X(b) = P(X < d),$$

- 如果$c = -\infty, d = +\infty$，则

$$\lim_{a\to -\infty} F_X(a) = 0, \qquad \lim_{b\to +\infty} F_X(b) = 1。$$

14、分布函数的性质

如果c,d有限，则$\lim_{a\to c+} F_X(a)=F_X(c)$, $\lim_{b\to d-} F_X(b)=P(X<d)$.

- 故$P(X=b)=F_X(b)-P(X<b)=F_X(b)-\lim_{a\to b-} F_X(a)$;
- 对任何区间$B\subset\mathbb{R}$，$P(X\in B)$可用F_X表达，比如，$P(a\le X\le b)=F_X(b)-P(X<a)=F_X(b)-\lim_{x\to a-} F_X(x)$.

注[2]：

- 如果函数$F:\mathbb{R}\to\mathbb{R}$满足上述命题中的性质(1)，则一定存在随机变量$X$以$F$为分布函数。
- 对一个随机变量$X:\Omega\to\mathbb{R}$和任何一个Borel集$B\subset\mathbb{R}$，$\{X\in B\}$是事件，并且$P(X\in B)$的值可由F_X计算得到。
- 给定$X:\Omega\to\mathbb{R}$，那么可以证明$\{X^{-1}(B): B\subset\mathbb{R}$是Borel集$\}$是$\Omega$上的一个事件域，在此之上构造概率空间，就可以使$X$成为一个随机变量。这是一个常用手法。

15、数学期望

定义

设X是一个离散型随机变量，它的分布列为

X	x_1	$\cdots$	x_n	$\cdots$
P	p_1	$\cdots$	p_n	$\cdots$

如果级数

$$\sum_n x_n p_n$$

绝对收敛，即

$$\sum_n |x_n| p_n < +\infty,$$

则称X有数学期望，并且记$EX = \sum_n x_n p_n$，称它是X的数学期望（*expectation*）。

16、数学期望的注记

注：

1. X的数学期望就是按X的不同取值的概率计算X所有可能取值的加权平均。
2. 定义中要求$\sum_n x_n p_n$绝对收敛，不仅保证了数学期望作为级数的和是存在的，而且保证了这个和与级数的求和顺序无关。（这是因为指定x_n为第n个取值的作法是人为的，而绝对收敛就保证了级数的和不依赖于不同的求和方式，从而EX反映了X的本质属性。）
3. 从上述定义我们不难发现，数学期望是分布的性质，与随机变量的具体形式无关。
4. X的数学期望是k阶（原点）矩$E(X^k)$的特例（$k=1$）。

17、数学期望

定义

设X是一个连续型随机变量，它的概率密度函数为p。如果积分$\int_{-\infty}^{\infty} xp(x)\mathrm{d}x$ 绝对收敛，即$\int_{-\infty}^{\infty} |x|p(x)\mathrm{d}x < +\infty$，则称$X$有数学期望，并且记$EX = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x$，称它为$X$的数学期望（*expectation*）。

- 连续型随机变量的数学期望的积分表达是基于概率微元$P(x \le X < x + \Delta x) \approx p(x)\Delta x$以及如下极限思想

$$\lim_{\max \triangle x_i \to 0} \sum_i x_i P(x_i \le X < x_i + \triangle x_i).$$

- 对一般的（不必是离散型或连续型）随机变量X，我们也可以利用分布函数来定义数学期望：$EX = \int_{-\infty}^{\infty} x\mathrm{d}F(x)$，但这里的积分是Lebesgue-Stieltjes积分[2]。

18、例（快速验血）

为了从人群中筛查某种疾病的患病者，需要对人群做血样检测。如果按照老办法，需要对每个人进行单独验血。为提高筛查的效率，有人提出一个新方法：将待检人群分组，每组k个人；对每组人的混合血样进行一次检查；对检查呈阳性的组，再对该组中的每个人分别进行单独的血液检验。新方法是否真的提高了验血的效率呢？(疾病的患病率为$0 < p < 1$)

解：记M为按新办法一组人的验血次数，则

M	1	k	$k+1$
P	$(1-p)^k$	$p(1-p)^{k-1}$	$1-(1-p)^{k-1}$

$$\begin{aligned} EM &= (1-p)^k + kp(1-p)^{k-1} + (k+1)[1-(1-p)^{k-1}] \\ &= 1 - k(1-p)^k + k - p(1-p)^{k-1}, \end{aligned}$$

故 每个人的平均验血次数为

$$EM/k = 1/k - (1-p)^k + 1 - p(1-p)^{k-1}/k。$$

19、数学期望公式

性质

离散型：$EX = \sum_{n=0}^{\infty} P(X > n) - \sum_{n=0}^{\infty} P(X < -n)$；（$X(\Omega) \subset \mathbb{Z}$）

连续型：$EX = \int_0^{+\infty} P(X > x)\mathrm{d}x - \int_{-\infty}^{0} P(X < x)\mathrm{d}x$。

设$X(\Omega) \subset \mathbb{N}$，则$EX = \sum_{n=0}^{\infty} np_n = p_1 + 2p_2 + \cdots + np_n + \cdots$。

$$\begin{array}{llllll}
P(X > 0) & p_1 & p_2 & \cdots & p_{n+1} & \cdots \\
P(X > 1) & & p_2 & \cdots & p_{n+1} & \cdots \\
\vdots & & & \ddots & \vdots & \vdots \\
P(X > n) & & & & p_{n+1} & \cdots \\
\cdots & & & & & \cdots
\end{array}$$

所以，$EX = \sum_{n=0}^{\infty} P(X > n)$。

20、数学期望公式

性质

离散型：$EX = \sum\limits_{n=0}^{\infty} P(X > n) - \sum\limits_{n=0}^{\infty} P(X < -n)$； （$X(\Omega) \subset \mathbb{Z}$）

连续型：$EX = \int_0^{+\infty} P(X > x)\mathrm{d}x - \int_{-\infty}^{0} P(X < x)\mathrm{d}x$。

$$\int_0^{+\infty} P(X > x)\mathrm{d}x - \int_{-\infty}^{0} P(X < x)\mathrm{d}x$$

$$\begin{aligned}
&= \int_0^{+\infty}\int_{\mathbb{R}} p(y) I_{y>x>0}\mathrm{d}y\mathrm{d}x - \int_{-\infty}^{0}\int_{\mathbb{R}} p(y) I_{y<x<0}\mathrm{d}y\mathrm{d}x \\
&= \int_{\mathbb{R}}\int_0^{+\infty} p(y) I_{y>x>0}\mathrm{d}x\mathrm{d}y - \int_{\mathbb{R}}\int_{-\infty}^{0} p(y) I_{y<x<0}\mathrm{d}x\mathrm{d}y \\
&= \int_{\mathbb{R}} y p(y) I_{y>0}\mathrm{d}y - \int_{\mathbb{R}} (-y) p(y) I_{y<0}\mathrm{d}y = EX.
\end{aligned}$$

21、随机变量的函数

- 设X是一个离散型随机变量，它的分布为：

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

- 若$g:\mathbb{R}\to\mathbb{R}$，则$g(X)$也是一个离散型随机变量，且

$$P(\{\omega: g(X(\omega))=y\})=\sum_{i:g(x_i)=y} p_i,\ \ y=g(x_1),g(x_2),\ldots.$$

定理

$$Eg(X)=\sum_i g(x_i)p_i.$$

证明：

$$Eg(X)=\sum_j y_j P[g(X)=y_j]=\sum_j y_j \sum_{i:g(x_i)=y_j} p_i = \sum_i g(x_i)p_i。$$

22、函数的数学期望

定理

设X是连续型随机变量，概率密度为$p(x)$，$g:\mathbb{R}\to\mathbb{R}$是Borel函数，$Y=g(X)$存在数学期望$EY$，则$Eg(X)=\int_{-\infty}^{+\infty}g(x)p(x)\mathrm{d}x$.

$$\begin{aligned}
Eg(X) &= \int_0^{+\infty} P(g(X)>y)\mathrm{d}y - \int_{-\infty}^{0} P(g(X)<y)\mathrm{d}y \\
&= \int_0^{+\infty}\int_{\mathbb{R}} p(x)I_{g(x)>y>0}\mathrm{d}x\mathrm{d}y - \int_{-\infty}^{0}\int_{\mathbb{R}} p(x)I_{g(x)<y<0}\mathrm{d}x\mathrm{d}y \\
&= \int_{\mathbb{R}}\int_0^{+\infty} p(x)I_{g(x)>y>0}\mathrm{d}y\mathrm{d}x - \int_{\mathbb{R}}\int_{-\infty}^{0} p(x)I_{g(x)<y<0}\mathrm{d}y\mathrm{d}x \\
&= \int_{\mathbb{R}} g(x)p(x)I_{g(x)>0}\mathrm{d}x + \int_{\mathbb{R}} g(x)p(x)I_{g(x)<0}\mathrm{d}x \\
&= \int_{\mathbb{R}} g(x)p(x)\mathrm{d}x.
\end{aligned}$$

23、数学期望的性质

性质

1. $\forall a, b \in \mathbb{R},\ E(aX+b) = aEX + b$；
2. $|EX| \le E|X|$；
3. 若$a \le X \le b$，则EX存在，且$a \le EX \le b$；
4. 若$X \le Y$，EX, EY存在，则$EX \le EY$。

证明：

1. $E(aX+b) = \sum_i (ax_i + b)p_i = a\sum_i x_i p_i + b = aEX + b$；
2. $|EX| = |\sum_i x_i p_i| \le \sum_i |x_i| p_i = E|X|$
3. $\qquad \le \sum_i p_i \max\{|a|, |b|\} = \max\{|a|, |b|\} < \infty$；

   故EX存在且 $a = \sum_i ap_i \le \sum_i x_i p_i \le \sum_i bp_i = b$。
4. 提示：由（3）和期望的线性性质可得。

24、参考文献

- 1、Kai Lai Chung and Farid AitSahlia，Elementary probability theory：with stochastic processes and an introduction to mathematical finance，4th ed. New York：Springer，2003.
- 2、Erhan Çinlar, Probability and Stochastics, New York, NY : Springer Science+Business Media, LLC, 2011.