概率论与数理统计

第七讲

随机变量与分布函数

1. 为什么要引入随机变量

- **随机变量**：对于随机试验的每一结果（样本）对应一个实数.

随机现象（所关心的） ⟷ **实值函数（所熟悉的）**

随机变量

纽 带

✓ 有些问题我们关心的不是随机试验的结果,而是联系该结果的数量,引入随机变量有助于研究所关心的问题.

例 (射击评估) 记录某人进行射击训练的情况，以对其射击能力进行评估，假设给其5发子弹，命中记为H，否则记为N，考虑

a) 样本空间如何表示；

b) 此问题中最感兴趣的是什么？

答：a) $S=\{\underbrace{NNNNN, NNNNH, \cdots, HHHHH}_{32个}\}$

b) 更感兴趣的是每个样本S中“H”的个数.

可定义：$X(s)=s$ 中“H”的个数.

例　(网店问题) 某网店出售一种商品. 若商品合格则每件赚n元，若不合格则每件赔m元，设合格与否是随机的. 问能否赚钱取决于什么因素.

答： 合格品的概率与m，n 的大小.

此时　$S=\{合格,\ 不合格\}$

可定义　$X(合格)=n,\ X(不合格)=-m$

注： 今后可在此基础上引入期望的概念解决盈亏计算问题.

✓ 有些问题样本空间本身就是实数集，引入随机变量，对关心的随机事件描述更简洁准确.

例 (血糖值问题) 设空腹血糖高6.10mol/L为高血糖，随机抽一人检测血糖，结果事件A“高血糖”发生.

可定义，$X(s)=s$

于是，$A=\{s|X(s)\geq 6.10\}=\underline{\underline{\{X\geq 6.10\}}}$

今后常用！

✓ 有些问题的背景不同但数学本质完全一样。引入随机变量，可以在更抽象和一般的层次上研究.

如前面的射击评估问题，可用于“类似的”随机变量的研究.

包括

- 试验中成功的次数
- 产品中合格品的件数
- 治病中治愈的人数
- ……

2. 随机变量定义

- **定义** 称一个（可测）映射 $X: S \longrightarrow R^1$ 为一个**随机变量**（random variable).

- **注**："可测"的数学含义及相关理论不作介绍，它为了保证由 X 刻画的子集 $\{X \le x\}$, $\{X > x\}$, $\{X = x\}$, $\{x_1 < X \le x_2\}$, …… 都是事件.

3. 随机变量的分布函数

问题　对于随机变量，我们感兴趣什么？

回忆：对实数区间上定义的函数 $f(x)$，我们关心连续型、可微性……

对随机变量，我们关心X的取值所体现的统计规律，即事件 $\{X \le x\}, \{X > x\}, \{x_1 < X \le x_2\}, \ldots\ldots$ 发生的概率.

注意到：$\{X > x\} = \overline{\{X \le x\}}$，$\{x_1 < X \le x_2\} = \{X \le x_2\} - \{X \le x_1\}$

只须研究形如$\{X \le x\}$的事件.

定义

对于随机变量X，定义函数 $F(x)=P\{X\leq x\},\forall x\in R$，称为$X$的分布函数 (distribution function).

注1 与随机变量不同，分布函数是定义在R上的普通实函数.

2 $X(s)$可以看成是 s 在实数轴上的“坐标”，$F(x)$就可以看成是坐标$X(s)$落在$(-\infty,x]$中的概率.

利用 X 的分布函数, 可以表示许多由 X 刻画的随机事件的概率.

例 用 X 的分布函数 $F(x)$ 表示下列事件的概率.

a) $\{X > a\}$;　　b) $\{a < X \le b\}$.

解: a) $P\{X > a\} = P\{-\infty < X < +\infty\} - P\{X \le a\}$

$$= 1 - F(a)$$

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

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

例 (收入分布)据有关研究资料, 我国2012年家庭人均收入如下表:

x (千元)	1	2	4.5	9	15.9	25.8	34.3
收入低于x的家庭比例	0.05	0.1	0.25	0.5	0.75	0.9	0.95

设 X 是随机抽取的一个家庭的人均收入, $F(x)$为其分布函数, 试用分布函数表示下列事件的概率.

a) $\{X \le 1\}$;　　b) $\{X > 34.3\}$;　　c) $\{2 < X \le 25.8\}$.

解： a) $F(1) = 0.05$;

b) $1 - F(34.3) = 1 - 0.95 = 0.05$;

c) $F(25.8) - F(2) = 0.9 - 0.1 = 0.8$.

分布函数的基本性质：

① $F(x)$是单调不减函数;

提示：注意到对于$x_1<x_2$，$\{X \le x_1\} \subset \{X \le x_2\}$.

② $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$

$F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$

③ $F(x)$ 为右连续函数，即

$F(x)=F(x+0) = \lim\limits_{y \to x^+} F(y)$

利用概率的定义, 我们可以证明②, ③.

证： ② 由可列可加性

$$1 = P\{-\infty < X < +\infty\} = P\{\bigcup_{n=-\infty}^{+\infty}(n < X \le n+1)\}$$

$$= \sum_{n=-\infty}^{+\infty}(F(n+1) - F(n))$$

$$= \lim_{n\to+\infty} F(n) - \lim_{m\to-\infty} F(m).$$

由于$F(x)$单调, 有

$$F(+\infty)=\lim_{x\to+\infty}F(x)=\lim_{n\to+\infty}F(n)$$

$$F(-\infty)=\lim_{x\to-\infty}F(x)=\lim_{m\to-\infty}F(m)$$

再由概率的非负与规范性, 知

$$F(+\infty)=1,\quad F(-\infty)=0.$$

对于③, 可取一列 $x_n\downarrow x$, 同样利用可列可加性, 得

$$F(x+0)=\lim_{y\to x^+}F(y)=F(x).$$

思考 若分布函数定义为

$$F(x) = P\{X < x\}, \quad \forall x \in R$$

则形式③有何变化？

问题 如何用分布函数表示下列事件的概率？

a) $P\{X = a\}$; b) $P\{a \le X \le b\}$.

答： a) $P\{X = a\} = F(a) - F(a-0)$;

b) $P\{a \le X \le b\} = F(b) - F(a-0)$.

例 (污染问题, 续)已知空气中PM2.5一般在0.0-120.4 $(\mu g/m^3)$之间, 根据有关指数标准, PM2.5含量在100.5以上为对人体有害, 设PM2.5的值在任一小区间$[a,b]\subset[0.0,120.4]$中的概率与区间长b-a成正比, 随机抽检空气质量, 求

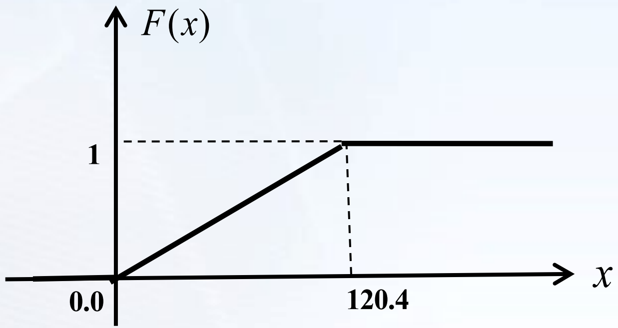
a) PM2.5值 X 的分布函数并作图；

b) 求空气质量正常的概率.

解：

a) 由题意知: 当$0.0<x<120.4$时，$P\{X \le x\} = \dfrac{x-0.0}{120.4-0.0}$,

当$x<0.0$时, $F(x)=0$, 当$x>120.4$时, $F(x)=1$.



b) 空气质量正常的概率：$P\{X \le 100.5\} = \dfrac{100.5-0.0}{120.4-0.0} \approx 0.8347.$