概率论与数理统计

第十讲

连续型随机变量及其分布

1. 连续型随机变量

- **回顾**

随机变量的统计规律可用分布函数刻画. 离散型随机变量有更方便的刻画——分布律.

问题：对于取值不可列的随机变量如何？

例（污染问题）得到对应 PM2.5 值 X，其分布函数可表为

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

经空气质量指数标准测试：

一般a=0.0, b=120.4

如果令 $f(x)=\begin{cases}\dfrac{1}{b-a} & a<x\le b\\ 0 & \text{其他}\end{cases}$

则有 $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$

问：是否有普遍性.

例（电压问题）设某个系统的输出电压 X 为随机变量，电压超过 $x(x>0)$ 的概率与 $(1+x)$ 反比，即有

$$P\{X>x\}=1-F(x)=\frac{k}{1+x} \quad \text{由} \quad F(-\infty)=F(0)=0 \Rightarrow k=1.$$

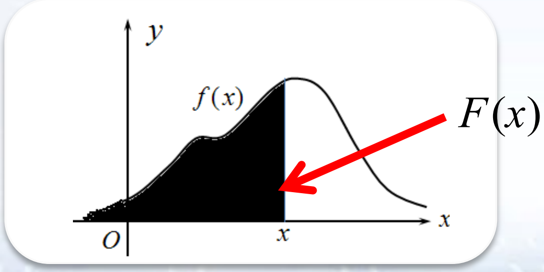
令 $f(x)=\begin{cases}\dfrac{1}{(1+x)^2} & x>0\\ 0 & x\le 0\end{cases}$ 则 $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t.$

定义 设随机变量 X 的分布函数 $F(x)$ 可表成

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$$

其中 $f(x)\ge 0$, 则称X是连续型随机变量，$f(x)$ 称为是 X 的（概率）密度函数 (density function).

几何意义



2. 密度函数性质

密度函数满足：

① $f(x) \geq 0$

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

③ $\forall\ x_1 < x_2$ 有

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)\mathrm{d}t$$

④ 在 $f(x)$ 的连续点处，有 $f(x) = F'(x)$

因 $F'(x) = \lim\limits_{\Delta x \to 0} \dfrac{F(x+\Delta x) - F(x)}{\Delta x} = \lim\limits_{\Delta x \to 0} \dfrac{1}{\Delta x} \int_{x}^{x+\Delta x} f(t)\mathrm{d}t$

$$= \lim_{\Delta x \to 0} \frac{1}{\Delta x} f(\xi)\Delta x = f(x)$$

注

1. 连续型随机变量的分布函数总是连续的；

2. 对于连续型随机变量 X, $P\{X=x\}=0, \forall x \in R$

因为 $P\{X=x\}=F(x)-F(x-0)=0$;

3. 密度函数不唯一, 事实上改变 $f(x)$ “个别值”（零测度集上的值）不影响积分值.

例 设 $f(x)=\begin{cases} kx^{-\frac{1}{3}} & 0<x<1 \\ 0 & \text{其他} \end{cases}$

为一密度函数，求

a) 常数k; b) 相应的分布函数; c) $P\{\frac{1}{4}<X\le 2\}$.

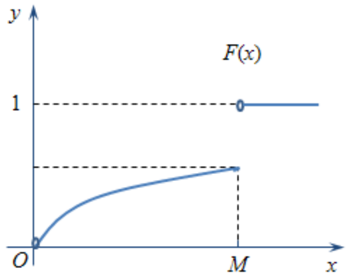
答：a) $k=\frac{2}{3}$; b) $F(x)=\begin{cases} 0 & x\le 0 \\ x^{2/3} & 0<x<1 \\ 1 & x\ge 1 \end{cases}$ c) $1-\frac{1}{2\sqrt[3]{2}}$.

反之，对分布函数逐段微分，也可得到原密度函数。

注 并非所有随机变量非离散型即连续型，也有两者皆不是的.

例（电压问题）前述电压问题，若电压表的里程到某M为止, 设X为表测量值, 则其分布函数应为

$$F(x)=\begin{cases} 0 & x \le 0 \\ \dfrac{x}{1+x} & 0<x<M \\ 1 & x \ge M \end{cases}$$



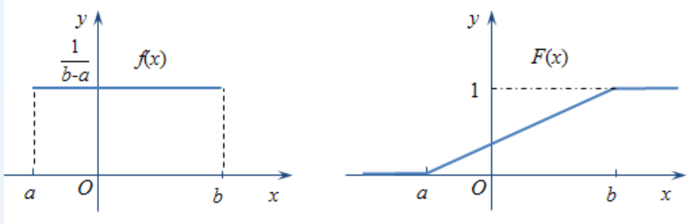
X既不能用密度函数，也不能用分布律刻画，但能用分布函数刻画.

3. 几种常用的连续型分布

① 均匀分布

定义 设随机变量X具有如下形式的密度函数

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



则称 X 服从区间$[a, b]$上均匀(uniformly)分布，记为$X\sim U(a, b)$.

✓ 前面所举的PM2.5值, 就服从 [0.0, 120.4]上的均匀分布.

✓ 均匀分布通常用来刻画区间上的等可能问题：舍入误差、命中率……；$[a, b]$上的均匀分布在$[a, b]$内任意小区间取值的概率只与小区间的长度有关，而与其位置无关。

② 指数分布

定义 设随机变量X具有如下形式的密度函数

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, x>0\\ 0\ ,\ x\le 0\end{cases}\qquad (\theta>0)$$

则称X服从参数为θ的指数分布, 记为$X\sim EXP(\theta)$.

问：这是一个密度函数吗？

第10讲 连续型随机变量及其分布

- 其分布函数：

$$F(x)=\begin{cases}1-e^{-\frac{x}{\theta}}, & x\geq 0\\ 0 \quad , & x<0\end{cases}$$

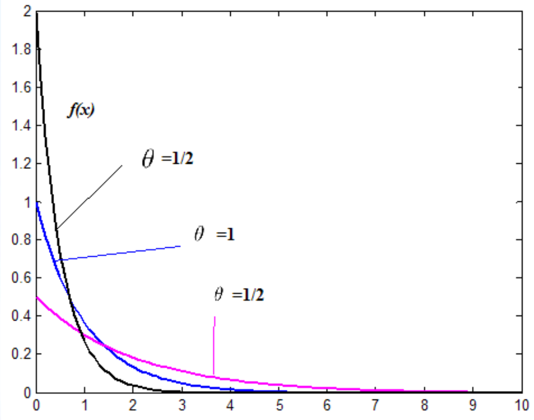
指数分布通常用来描述生命周期（生物、产品……）.

问：θ的含义是什么？

性质：无记忆性：设$X\sim EXP(\theta)$，

则对于$t, s>0$,

$P\{X>t+s \mid X>s\}=P\{X>t\}$.



指数分布与泊松流

泊松流：对于源源不断依次随机而来的质点（粒子、顾客...），考虑其数量，如果满足

1. 在任意时间t内质点数$X_t \sim P(\lambda t)$;
2. 在不同时段中的质点数相互独立. 就称$\{X_t: t>0\}$是个泊松流.

问题：泊松流中质点出来的间隔服从什么分布？

设第一个质点出来的时刻为0，下一质点出来时刻为Y，对于任意的$t>0$，考虑

$$P\{Y>t\}=P\{[0,t]\text{时间内有0个质点出来}\}$$
$$=e^{-\lambda t}\qquad (t>0)$$

故 Y 的分布函数为

$$F_Y(t)=P\{Y\le t\}=1-e^{-\lambda t}\qquad (t>0)$$

即 $Y\sim EXP(\lambda^{-1})$.

- 为何用它来解释电子元件的寿命？