2.2 连续型随机变量及其分布

北京化工大学数学系

苏贵福

在实际问题中, 存在着与离散型随机变量取值形式不同的 另外一类随机变量. 它们可以在整个实数轴, 或实数轴的某个区间 上取值. 因此, 这类随机变量的概率分布规律, 就不可能用离散型随机变量 的概率分布律来描述.

一. 连续型随机变量

定义1 如果对于随机变量X的分布函数$F(x)$, 存在非负可积函数$f(x)$, 使对于任意实数x都有

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

则称X为<u>连续型随机变量</u>, $f(x)$称为X的<u>概率密度</u>.

♣ 据数学分析的知识知连续型随机变量的 分布函数是连续函数.

♣ 据定义知改变概率密度$f(x)$在个别点 的函数值不改变分布函数的取值. 因此, 在实际讨论中并不在乎概率密度在个别点处的值.

概率密度函数$f(x)$具有如下性质:

① $f(x) \geq 0$, 其中$-\infty < x < +\infty$.

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

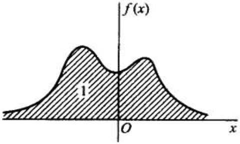
③ 对于任意实数$x_1, x_2 (x_1 \leq x_2)$, 有

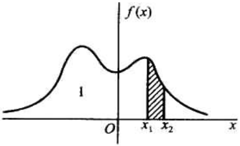
$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx.$$

④ 若$f(x)$在点x处连续, 则有$F'(x) = f(x)$.

♣ 反之, 若$f(x)$具备性质①和②, 则函数$H(x) = \int_{-\infty}^{x} f(t)dt$一定是某一随机变量$X$的分布函数, 而$f(x)$是相应的概率密度.

- 由②可知介于曲线$f(x)$和x轴之间的面积等于1, 如左图所示.





- 由③可知X落在区间$(x_1, x_2]$的概率$P\{x_1 < X \le x_2\}$ 等于区间$(x_1, x_2]$上曲线$f(x)$之下的曲边梯形面积, 如右图所示.

- 由④可知在$f(x)$的连续点x处有

$$f(x) = \lim_{\Delta x \to 0^+} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{P\{x < X \le x + \Delta x\}}{\Delta x}$$

因此X落在$(x, x+\Delta x]$上的概率$P\{x < X \le x + \Delta x\} \approx f(x)\Delta x$.

例1 设随机变量X具有概率密度

$$f(x)=\begin{cases} kx, & 0\le x<3 \\ 2-\dfrac{x}{2}, & 3\le x\le 4 \\ 0, & \text{其他} \end{cases}$$

(1) 确定常数k; (2) 求X的分布函数$F(x)$; (3) 求$P\{1<X\le \frac{7}{2}\}$.

解 (1) 由$\int_{-\infty}^{+\infty} f(x)dx=1$得

$$\int_0^3 kxdx+\int_3^4 (2-\frac{x}{2})dx=1$$

解得$k=\frac{1}{6}$. 于是X的概率密度为:

$$f(x)=\begin{cases}\dfrac{x}{6}, & 0\le x<3\\ 2-\dfrac{x}{2}, & 3\le x\le 4\\ 0, & \text{其他}\end{cases}$$

(2) X的分布函数为

$$F(x)=\begin{cases}0, & x<0\\ \displaystyle\int_0^x \frac{x}{6}dx, & 0\le x<3\\ \displaystyle\int_0^3 \frac{x}{6}dx+\int_3^x (2-\frac{x}{2})dx, & 3\le x\le 4\\ 1, & x\ge 4\end{cases}$$

也就是

$$F(x)=\begin{cases} 0, & x<0 \\ \dfrac{x^2}{12}, & 0\le x<3 \\ -3+2x-\dfrac{x^2}{4}, & 3\le x\le 4 \\ 1, & x\ge 4 \end{cases}$$

由此可以求得

$$\begin{aligned} P\{1<X\le \frac{7}{2}\} &= F(\frac{7}{2})-F(1) \\ &= (-3+2\cdot\frac{7}{2}-\frac{1}{4}\cdot(\frac{7}{2})^2)-\frac{1}{12}\cdot 1 \\ &= \frac{41}{48}. \end{aligned}$$

♠ 注释

- 对于连续性随机变量X来说，它取任一指定实数值a的概率均为0，即$P\{X=a\}=0$.

这是因为：设X的分布函数为$F(x)$，$\Delta x>0$，则由$\{X=a\}\subset\{a-\Delta x<X\le a\}$得

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

在上述不等式中令$\Delta x\to 0$，则有$P\{X=a\}=0$.

- 计算连续性随机变量X落在某一区间的概率时，可以不比区分区间是开区间或闭区间或半闭区间．即

$$P\{a<X\le b\}=P\{a\le X\le b\}=P\{a<X<b\}.$$

二. 几种重要分布

定义2 若连续型随机变量X具有概率密度(如下图左所示)

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

则称X在(a,b)上服从均匀分布, 记作$X\sim U(a,b)$.

在区间(a,b)上服从均匀分布的随机变量X, 具有 下述性质:

- 它落在区间(a,b)中任意等长度的子区间内的可能性是相同的.

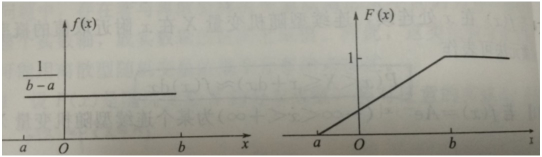
或者说它落在(a,b)的子区间内的概率只依赖于子区间的长度, 而与子区间的位置无关.

这是由于: 对于任一长度为l的子区间$(c, c+l) \subseteq (a, b)$, 有

$$P\{c < X \leq c+l\} = \int_c^{c+l} f(x)dx = \int_c^{c+l} \frac{1}{b-a}dx = \frac{l}{b-a}.$$

根据连续型随机变量的定义, X的分布函数为(如下图右所示)

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



例2 设X服从$[1,6]$上的均匀分布，求方程$t^2+Xt+1=0$ 有实根的概率.

解 由已知条件可知X的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{5}, & x\in[1,6]\\ 0, & x\text{其他}\end{cases}$$

易知$t^2+Xt+1=0$有实根$\Leftrightarrow\Delta=X^2-4\geq 0\Leftrightarrow|X|\geq 2\Leftrightarrow X\geq 2$
或$X\leq -2$. 于是所求的概率为

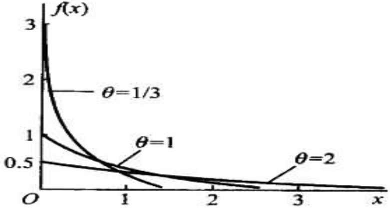
$$P\{X\geq 2\}+P\{X\leq -2\}=\int_2^6\frac{1}{5}dx+0=\frac{4}{5}.$$

即方程$t^2+Xt+1=0$有实根的概率为$\frac{4}{5}$.

定义3 若连续型随机变量X具有概率密度

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

其中$\theta>0$为常数, 则称X服从参数为θ的指数分布.



上图画出了$\theta=\frac{1}{3}$, $\theta=1$和$\theta=2$时指数分布密度函数$f(x)$的图像.

♣ 服从指数分布的随机变量X的分布函数为

$$F(x)=\begin{cases}1-e^{-\frac{x}{\theta}}, & x>0\\ 0, & \text{其他}\end{cases}$$

♣ 若X服从指数分布, 那么对任意的$s,t>0$, 有

$$P\{X>s+t|X>s\}=P\{X>t\}. \quad (\underline{\underline{\text{无记忆性}}})$$

这是因为

$$P\{X>s+t|X>s\}=\frac{P\{(X>s+t)\cap(X>s)\}}{P\{X>s\}}=\frac{P\{X>s+t\}}{P\{X>s\}}$$
$$=\frac{1-F(s+t)}{1-F(s)}=\frac{e^{-(s+t)/\theta}}{e^{-s/\theta}}=e^{-t/\theta}=P\{X>t\}.$$

① "无记忆性"说明: 元件对它已使用过s小时没有记忆, 这一性质是指数分布被广泛应用的重要原因. 指数分布在可靠性理论与排队论中有着广发的应用.

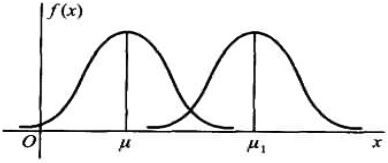
② 指数分布常用来描述设备或元件的寿命, 而实际生活中元件的寿命不会无记忆的, 因此指数分布只能粗略近似地刻画寿命问题.

定义4 若连续型随机变量X具有概率密度

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

其中$\mu,\sigma(\sigma>0)$为常数, 则称X服从参数为μ,σ的正态分布, 记作$X\sim N(\mu,\sigma^2)$.

参数μ,σ的意义将在第三章说明. 密度函数$f(x)$的图像如下所示.



并且具有下列性质:

♣ 曲线关于$x=\mu$对称. 这表明对任意的$t>0$有

$$P\{\mu-t<X\leq\mu\}=P\{\mu<X\leq\mu+t\}.$$

♣ 当$x=\mu$时$f(x)$取到最大值$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$. 而且 x离μ越远, $f(x)$的值越小. 这表明对于同样长度的区间, 当区间离μ越远, X落在该区间的概率就越小.

♣ 如果固定σ改变μ, 则图形沿着x轴平移, 而不改变其形状. 如果固定μ改变σ, 那么X落在μ附近的概率大小与σ的取值成反比.

服从正态分布的随机变量X的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

特别地，当$\mu=0$，$\sigma=1$时称随机变量X服从标准正态分布. 其概率密度和分布函数分别用 $\varphi(x)$和$\Phi(x)$表示，即有

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

不难验证$\Phi(-x)=1-\Phi(x)$. 而标准正态分布的值可以通过表予以查询. 因此将一般正态分布化为标准正态分布是一个关键问题.

定理1 若随机变量$X \sim N(\mu, \sigma^2)$, 则$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

证明 $Z = \frac{X-\mu}{\sigma}$的分布函数为

$$P\{Z \le x\} = P\left\{\frac{X-\mu}{\sigma} \le x\right\} = P\{X \le \mu + \sigma x\}$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt.$$

令$\frac{t-\mu}{\sigma} = u$, 则有

$$P\{Z \le x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du = \Phi(x).$$

由此证得$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. ■

♣ 由定理1知, 若$X \sim N(\mu, \sigma^2)$, 则其分布函数$F(x)$可以改写为

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

♣ 对任意区间$(x_1, x_2]$, 有

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{x_2-\mu}{\sigma}\right\} \\ &= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right). \end{aligned}$$

♣ 设$X \sim N(1, 4)$, 则查表得

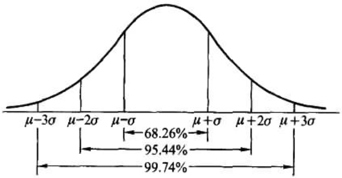
$$\begin{aligned} P\{0 < X \leq 1.6\} &= \Phi\left(\frac{1.6-1}{2}\right) - \Phi\left(\frac{0-1}{2}\right) = \Phi(0.3) - \Phi(-0.5) \\ &= 0.6179 - [1 - \Phi(0.5)] = 0.6179 - 1 + 0.6915 = 0.3094. \end{aligned}$$

♣ 若$X \sim N(\mu,\sigma^2)$, 由$\Phi(x)$的函数表可得

$$P\{\mu-\sigma<X<\mu+\sigma\}=\Phi(1)-\Phi(-1)=\frac{68.26}{100}$$

$$P\{\mu-2\sigma<X<\mu+2\sigma\}=\Phi(2)-\Phi(-2)=\frac{95.44}{100}$$

$$P\{\mu-3\sigma<X<\mu+3\sigma\}=\Phi(3)-\Phi(-3)=\frac{99.74}{100}$$



尽管正态变量的取值范围是$(-\infty,+\infty)$, 但它的值落在$(\mu-3\sigma,\mu+3\sigma)$内几乎是肯定的. 这就是"$3\sigma$法则".

例3 轴的长度$X \sim N(10, 0.01)$, 如果轴的长度 在$(10-0.2, 10+0.2)$范围内算合格. 现有四根轴, 试求(1)恰有三根轴长度合格的概率; (2) 至少有三根轴长度合格的概率.

解 轴的长度X合格, 即X应满足$10-0.2 < X < 10+0.2$. 于是

$$P\{10-0.2 < X < 10+0.2\} = P\left\{\left|\frac{X-10}{0.1}\right| < 2\right\} = 2\Phi(2)-1$$

查表得$P\{10-0.2 < X < 10+0.2\} = 0.9544$. 故

(1) 恰有三根轴长度合格的概率为$\binom{4}{3} \cdot 0.9544^3 \cdot 0.0456 \approx 0.1586$.

(2) 至少有三根轴长度合格的概率为

$$\binom{4}{3} \cdot 0.9544^3 \cdot 0.0456 + 0.9544^4 \approx 0.9883.$$