连续型随机变量及其概率密度

◆ 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有

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

则称X为**连续型随机变量**，$f(x)$称为X的**概率密度函数**，简称**概率密度。**

连续型随机变量及其概率密度

◆ 概率密度$f(x)$具有以下**性质**：

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- 对于任意实数$x_1, x_2(x_1 \leq x_2)$,

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

- 若$f(x)$在点x处连续，则有$F'(x) = f(x)$

连续型随机变量及其概率密度

例：随机变量X具有概率密度

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

(1)确定参数k.(2)求X的分布函数F(x).(3)求$P\left\{1<X\le\frac{7}{2}\right\}$

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$k=\frac{1}{6}$

连续型随机变量及其概率密度

(2) X概率密度为

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

则X得分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ \int_0^x \frac{x}{6}\mathrm{d}x, & 0\leqslant x<3,\\ \int_0^3 \frac{x}{6}\mathrm{d}x+\int_3^x\left(2-\frac{x}{2}\right)\mathrm{d}x, & 3\leqslant x<4,\\ 1, & x\geqslant 4.\end{cases} \Rightarrow F(x)=\begin{cases}0, & x<0,\\ \frac{x^2}{12}, & 0\leqslant x<3,\\ -3+2x-\frac{x^2}{4}, & 3\leqslant x<4,\\ 1, & x\geqslant 4.\end{cases}$$

(3) $P\left\{1<X\leq\frac{7}{2}\right\}=F\left(\frac{7}{2}\right)-F(1)=\frac{41}{48}$

连续型随机变量及其概率密度

- 对于连续型随机变量X，取任一指定实数值a的概率均为0，

$$即P\{X=\mathrm{a}\}=0.$$

- 在计算连续型随机变量落在某一区间的概率时，可以不必区分是开区间或闭区间或半闭区间，即

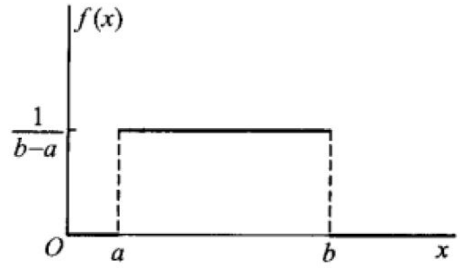
$$P\{\mathrm{a}<X\le b\}=P\{\mathrm{a}\le X\le b\}=P\{\mathrm{a}<X<b\}$$

- 当提到随机变量X的“概率分布”时，指的是分布函数；或当X是连续型随机变量时，指的是概率密度，当X是离散型随机变量时，指的是分布律

◆ 均匀分布

若连续型随机变量X具有概率密度

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

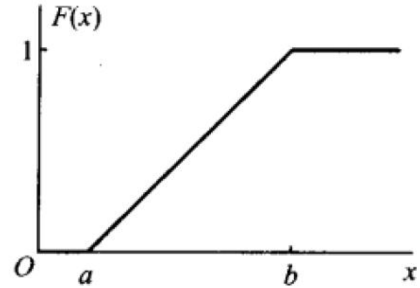


则称X在区间(a, b)上服从**均匀分布**，记为$X\sim U(a,b)$.

注：落在(a, b) 中任意等长度子区间内的可能性是一样的。

X的分布函数为：

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



连续型随机变量及其概率密度

例：设电阻值R是一个随机变量，均匀分布在900~1100.求R的概率密度及落在950~1050的概率。

解：R的概率密度为

$$f(r)=\begin{cases}\dfrac{1}{1100-900}, 900<x<1100\\ 0, others\end{cases}$$

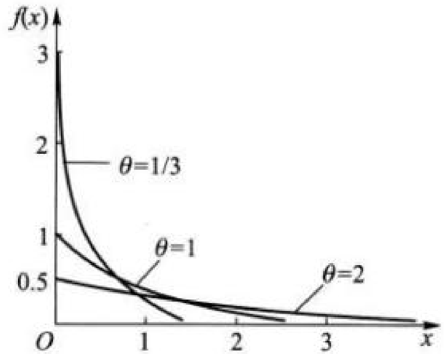
故有

$$P\{950<\mathrm{R}\leq 1050\}=\int_{950}^{1050}\frac{1}{200}dr=0.5$$

连续型随机变量及其概率密度

◆ 指数分布

若连续型随机变量X具有概率密度为



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

其中$\theta > 0$为常数，则称X服从参数为θ的**指数分布**。

X的分布函数为：

$$F(x)=\begin{cases}1-e^{-x/\theta}, x>0\\0, others\end{cases}$$

连续型随机变量及其概率密度

服从指数分布的随机变量X具有**无记忆性**：

对于任意$s, t > 0$有

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

$$P\{X > s + t \mid X > s\} = \frac{P\{(X > s + t) \cap (X > t)\}}{P\{X > s\}}$$

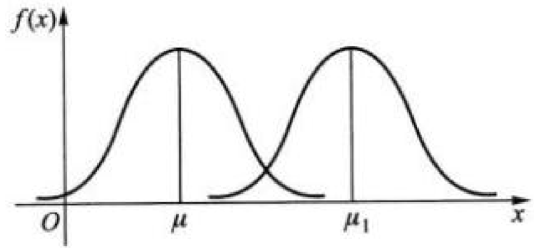
$$= \frac{P\{X > s + t\}}{P\{X > s\}} = \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-(s+t)/\theta}}{e^{-s/\theta}} = e^{-t/\theta} = P\{X > t\}$$

连续型随机变量及其概率密度

◆ 正态分布

若连续型随机变量X具有概率密度为

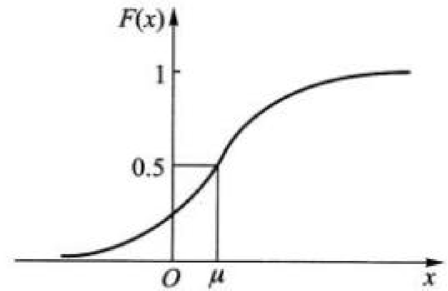
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$



其中μ, $\sigma(\sigma>0)$为常数，则称X服从参数为μ, σ的**正态分布**或**高斯分布**，记为$X\sim N(\mu,\sigma^2)$。

X的分布函数为：

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



连续型随机变量及其概率密度

显然$f(x)\ge 0$, 下面证明$\int_{-\infty}^{+\infty} f(x)dx = 1$。

令$\frac{x-\mu}{\sigma}=t$，则有$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-t^2/2}dt$

记$I=\int_{-\infty}^{+\infty}e^{-t^2/2}dt$，则有$I^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(t^2+u^2)/2}dtdu$

利用极坐标化成累次积分得到$I^2=\int_0^{2\pi}\int_0^{\infty}\mathrm{r}e^{-r^2/2}drd\theta=2\pi$

而$I>0$，故有$I=\sqrt{2\pi}$，即$\int_{-\infty}^{+\infty}e^{-t^2/2}dt=\sqrt{2\pi}$

于是

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

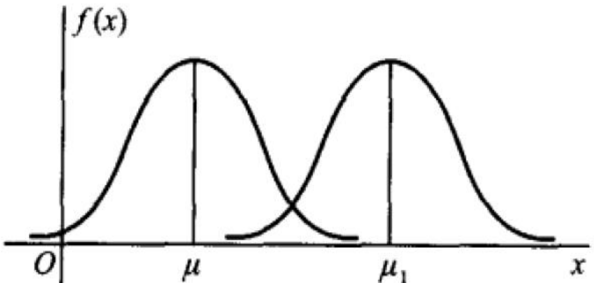
连续型随机变量及其概率密度

◆ 服从正态分布的随机变量X的概率密度具有以下**性质**：

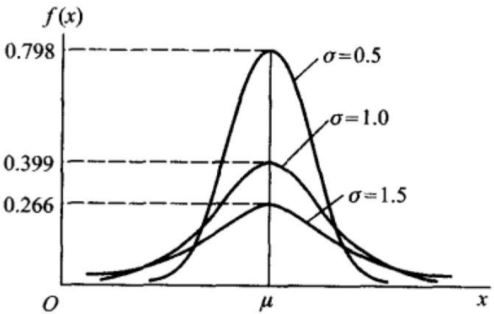
➢ 曲线关于$x=\mu$对称，对于任意$h>0$有

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

➢ 当$x=\mu$时取到最大值，$f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$



μ控制了函数位置
函数整体关于$x=\mu$对称， $x=\mu$时取最大值，离μ越远，其函数值越小。



σ控制了函数的形状
σ越小越往中心位置μ集中，图形变得越尖，x落在μ附近的概率越大。

连续型随机变量及其概率密度

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$\varphi(x), \Phi(x)$表示，有

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

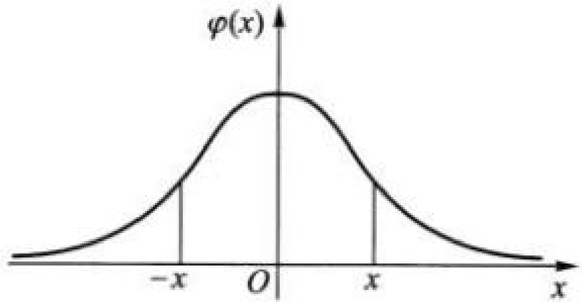
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

且$\Phi(-x) = 1 - \Phi(x)$

$\Phi(0) = 0.5$

$P\{|X| \le x\} = 2\Phi(x) - 1$

$P\{|X| \ge x\} = 2[1 - \Phi(x)]$



连续型随机变量及其概率密度

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

证： $Z=\frac{X-\mu}{\sigma}$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$\frac{t-\mu}{\sigma}=u$，得$P\{Z \le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du=\Phi(x)$

由此知$Z=\frac{X-\mu}{\sigma} \sim N(0,1)$。

若随机变量$X \sim N(\mu,\sigma^2)$, 则分布函数为

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

对任意区间$(x_1,\ x_2]$, 有

$$P\{x_1 < X \le x_2\}=\left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right\}=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right)$$