Lecture 06: Continuous Random Variables

$\boxed{\text{Definition 6.1}}$ 设存在一个非负的可积函数 $p(x)$，若 X 的分布函数 $F(x)=\int_{-\infty}^{x}p(u)du$，则称 X 为连续型随机变量，$p(x)$ 为密度函数。

根据定义，我们可以得到

$$P(X\le a)=F(a)=\int_{-\infty}^{a}p(x)dx$$
$$P(a<X\le b)=F(b)-F(a)=\int_{a}^{b}p(x)dx$$

一些注意点：

- $F(x)$ 连续。

> 显然 $p(x)$ 有界，设 $p(x)\le M<+\infty$，对于固定的 x_0，
>
> $$|F(x)-F(x_0)|=\left|\int_{x_0}^{x}p(u)d(u)\right|\le\int_{x_0}^{x}|p(u)|du\le|x-x_0|M$$
>
> $\forall\varepsilon>0$，当 $|x-x_0|\le\frac{\varepsilon}{M}$ 时，$|F(x)-F(x_0)\le\varepsilon$。

- $p(x)$ 并不能表示 $P(X=x)$。事实上对于任意 x_0，$P(X=x_0)=0$。$p(x)$ 只能理解为 X 在 x 的一个小邻域内的概率与该邻域的长度的近似比，即

  $$P(x<X<x\le\Delta x)=\int_{x}^{x+\Delta x}p(u)du\approx p(x)\Delta x$$

  关于 $P(X=x_0)=0$ 的说明：

  $$0\le P(X=x_0)\le P(x_0-\Delta x<X\le x_0)=F(x_0)-F(x_0-\Delta x)\xrightarrow{\Delta x\to 0}0$$

- 若 $p(x)$ 在 x_0 连续，则 $F(x)$ 在 x_0 处可导，且 $F'(x_0)=p(x_0)$。

> $$\begin{aligned}F(x)-F(x_0)&=\int_{x_0}^{x}p(u)du\\&=\int_{x_0}^{x}(p(u)-p(x_0))du+\int_{x_0}^{x}p(x_0)du\\&=\int_{x_0}^{x}(p(u)-p(x_0))du+p(x_0)(x-x_0)\end{aligned}$$
>
> 于是
>
> $$\left|\frac{F(x)-F(x_0)}{x-x_0}-p(x_0)\right|\le\frac{1}{|x-x_0|}\int_{x_0}^{x}|p(u)-p(x_0)|du\qquad(*)$$
>
> $\forall\varepsilon>0,\exists\delta,|x-x_0|<\delta\Rightarrow|p(u)-p(x_0)|<\varepsilon\Rightarrow(*)\le\frac{1}{|x-x_0|}\cdot\varepsilon|x-x_0|=\varepsilon$。

【例】(均匀分布) $[a,b]$ 的均匀分布 $U[a,b]$ 中，X 的密度函数和分布函数为

$$p(x)=\begin{cases}\frac{1}{b-a} & ,a\le x\le b\\ 0 & ,otherwise\end{cases}$$

$$F(x)=\begin{cases}0 & ,x<a\\ \int_a^x \frac{du}{b-a}=\frac{x-a}{b-a} & ,a\le x\le b\\ 1 & ,x>b\end{cases}$$

【例】(正态分布) X 服从正态分布，记为 $X\sim N(\mu,\sigma^2)$，若

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

> $$\begin{aligned}\left(\int_{-\infty}^{+\infty}p(x)dx\right)^2&=\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\right)\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}dy\right)\\ &\overset{u=\frac{x-\mu}{\sigma},v=\frac{y-\mu}{\sigma}}{=}\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{u^2+v^2}{2}}dudv\\ &\overset{u=r\cos\theta,v=r\sin\theta}{=}\frac{1}{2\pi}\int_0^{2\pi}\int_0^{+\infty}e^{-\frac{r^2}{2}}rdrd\theta\\ &=1\end{aligned}$$

正态分布 $p(x)$ 的图像有以下性质：

- $p(x)$ 关于 $x=\mu$ 轴对称。$F(\mu-a)+F(\mu+a)=1$。
- σ^2 越大，$p(x)$ 的图像越平。(最大值变小，两头变高，方差 $\uparrow$)

当 $\mu=0,\sigma=1$ 时，X 称为标准正态分布：

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

$$\Phi(x)=\int_{-\infty}^{x}\varphi(u)du$$

$\boxed{\text{Theorem 6.2}}$ 设 $X\sim N(\mu,\sigma^2)$，则 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$。

> $$\begin{aligned}F_Z(x)&=P(\frac{X-\mu}{\sigma}\le x)=P(X\le\sigma x+\mu)\\ &=\int_{-\infty}^{\sigma x+\mu}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(v-\mu)^2}{2\sigma^2}}dv\\ &\overset{\frac{v-\mu}{\sigma}=z}{=}\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{z^2}{2}}d(\sigma z+\mu)\\ &=\int_{-\infty}^{x}\varphi(u)du\end{aligned}$$

利用该定理，我们可以将所有的正态分布转换为标准正态分布的计算：

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

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

"3-σ" 准则：

$$P(|X-\mu|\le\sigma)=\Phi(1)-\Phi(-1)\approx 0.6826$$

$$P(|X-\mu|\le 2\sigma)\approx 0.9544$$

$$P(|X-\mu|\le 3\sigma)\approx 0.9974$$

当某个样本超过均值三个标准差以上时，可以怀疑它存在一些问题。

$\boxed{\text{Theorem 6.3}}$

$$(\frac{1}{x}-\frac{1}{x^3})\varphi(x)\le 1-\Phi(x)\le \frac{1}{x}\varphi(x)$$

且当 x 充分大时，$1-\Phi(x)\approx \frac{1}{x}\varphi(x)$。

【例】(指数分布) 称 X 服从参数为 $\lambda(\lambda>0)$ 的指数分布，记为 $X\sim E(\lambda)$，若

$$p(x)=\begin{cases}\lambda e^{-\lambda x} & ,x\ge 0\\ 0 & ,x<0\end{cases}$$

$$F(x)=\begin{cases}\int_0^x \lambda e^{-\lambda u}du=1-e^{-\lambda x} & ,x>0\\ 0 & ,x<0\end{cases}$$

注：指数分布具有无记忆性，i.e. $P(X>s+t|x>t)=P(X>s)$。指数分布和几何分布是唯二具有无记忆性的分布。