随机变量

> 参考：林家浩, 张亚辉. 随机振动的虚拟激励法. 北京: 科学出版社, 2004.

随机变量基本概念

随机变量(random variable)

离散型随机变量(discrete random variable)

连续型随机变量(continuous random variable)

概率(probability)

$$p_i(x_i) = P(X = x_i)(i = 1, 2, \cdots)$$

概率分布(probability distribution)

平均概率密度(average probability density)

$$\frac{P(x \le X < x + \Delta x)}{\Delta x}$$

概率密度函数(probability density function)

$$\lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x)}{\Delta x} = p(x)$$

概率分布函数(probability distribution function)

$$F(x) = P(X \le x) \quad (-\infty < x < +\infty)$$

$$F(x) = \int_{-\infty}^{x} p(\xi) d\xi$$

概率分布函数的性质：

- $0 \le F(x) \le 1$；
- 概率分布函数是单调上升的；
- 左极限($x \to -\infty$)为0，右极限($x \to +\infty$)为1；
- 对于连续型随机变量，有

$$p(x)=\frac{d}{dx}F(x)$$

联合概率分布函数(joint probability distribution function)

$$F(x,y)=P(X\leq x;Y\leq y)$$

联合概率密度函数(joint probability density function)

$$p(x,y)=\frac{\partial^2}{\partial x\partial y}F(x,y)$$

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(\xi,\eta)d\xi d\eta$$

若X、Y是独立的，则有

$$p(x,y)=p(x)p(y)$$

n个随机变量的联合概率密度函数

$$p(x_1,x_2,\cdots,x_n)=\frac{\partial^n}{\partial x_1\partial x_2\cdots\partial x_n}F(x_1,x_2,\cdots,x_n)$$

随机变量的数字特征

数学期望(expected value)、**均值(mean value)** 描述了随机变量取值的平均值。

- 离散型随机变量的数学期望

$$E[X]=\mu=\sum_i x_ip_i$$

  其中$x_i(i=1,2,\cdots)$为随机变量X可能取的数值，其分布列为$p_i=P(X=x_i)(i=1,2,\cdots)$。

- 连续型随机变量的数学期望

$$E[x]=\mu=\int_{-\infty}^{+\infty}xp(x)dx$$

- 两个连续型随机变量乘积的数学期望

$$E[XY]=\int_{-\infty}^{+\infty}xydF(x,y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyp(x,y)dxdy$$

其中$F(x, y)$、$p(x, y)$分别为X和Y的联合概率分布函数和联合概率密度函数。

方差(variance) 描述了随机变量取值与其均值的偏离程度。

- 离散型随机变量的方差

$$D[X] = \sigma^2 = E[(X - \mu)^2] = \sum_i (x_i - \mu)^2 p_i$$

- 连续型随机变量的方差

$$D[X] = \sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx$$

σ称为**标准差(standard deviation)**、**标准离差**或**均方差(mean square deviation)**。

变异系数(coefficient of variation)

$$\xi = \sigma / \mu$$

变异系数是一个无量纲量，在工程中常用以表示随机变量偏离平均值的程度。一般要求$\sigma \ll \mu$。

n阶原点矩(n-th moment)

- 离散型随机变量的n阶原点矩

$$m_n = E[X^n] = \sum_i x_i^n p_i$$

- 连续型随机变量的n阶原点矩

$$m_n = E[X^n] = \int_{-\infty}^{+\infty} x^n p(x) dx$$

当$n = 2$时，$m_2 = E[X^2]$称为**均方值**(mean square value)或**二阶原点矩**。其平方根称为**均方根值**(mean square root)。

n阶中心矩(n-th central moment)

- 离散型随机变量的n阶中心矩

$$K_n = E[(X-\mu)^n] = \sum_i (x_i - \mu)^n p_i$$

- 连续型随机变量的n阶中心矩

$$K_n = E[(X-\mu)^n] = \int_{-\infty}^{+\infty} (x-\mu)^n p(x) dx$$

当$n=2$时，$K_2 = E[(X-\mu)^2] = \sigma^2$，因此方差又称为**二阶中心矩**。

协方差(covariance)与**相关系数**(correlation coefficient)

- 协方差

$$cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

- 相关系数(规格化协方差)

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

相关系数为协方差的无量纲表达，协方差与相关系数是X与Y之间关系“密切程度”的表征。

平均值与方差的性质:

1. $D[X] = E[X^2] - (E[X])^2$
2. $E[a] = a$, $D[a] = 0$
3. $E[cX] = cE[X]$
4. $E[XY] = E[X]E[Y] + E[(X-\mu_X)(Y-\mu_Y)]$
5. 若$X_1, X_2, \cdots, X_n$为n个随机变量，则

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

$$D[X_1+X_2+\cdots+X_n]=\sum_{i,j=1}^{n}E[(X_i-\mu_{X_i})(X_j-\mu_{X_j})]$$

6. 设$Y=g(X)$是随机变量X的连续函数，则:
   如X是离散型随机变量，其分布列是$p_k=P(X=x_k),k=1,2,\cdots$，且$\sum_{k=1}^{\infty}|g(x_k)|p_k$收敛，则$Y$的数学期望为

$$E[Y]=E[g(X)]=\sum_{k=1}^{\infty}g(x_k)p_k$$

   如X是连续型随机变量，其概率密度函数为$p(x)$，且$\int_{-\infty}^{+\infty}|g(x)|p(x)dx$收敛，则$Y$的数学期望为

$$E[Y]=E[g(X)]=\int_{-\infty}^{+\infty}g(x)p(x)dx$$

几种重要分布

正态分布(nomal distribution)、**高斯分布**(Gaussian distribution)

- 概率密度函数

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

- 概率分布函数

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

- n阶中心矩

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

  - $K_0=1$
  - n为奇数时，该积分值为0
  - n为其他偶数时

$$K_n=1\times 3\times 5\times\cdots\times(n-1)\sigma^n$$

- 正太分布随机变量区间概率

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

变量替换$t = \frac{x-\mu}{\sigma\sqrt{2}}$，于是

$$P(\alpha < X < \beta) = \frac{1}{\sqrt{\pi}} \int_{\frac{\alpha-\mu}{\sigma\sqrt{2}}}^{\frac{\beta-\mu}{\sigma\sqrt{2}}} e^{-t^2} dx$$

引入拉普拉斯函数

$$\mathit{\Phi}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

于是

$$P(\alpha < X < \beta) = \frac{1}{2} \left[\mathit{\Phi}\left[\frac{\beta-\mu}{\sigma\sqrt{2}}\right] - \mathit{\Phi}\left[\frac{\alpha-\mu}{\sigma\sqrt{2}}\right] \right]$$

拉普拉斯函数$\mathit{\Phi}$是x的奇函数，且$\mathit{\Phi}(0) = 0, \mathit{\Phi}(\infty) = 1$，拉普拉斯函数一般预先制成表用于查询。

> Chrome插件MathJax Plugin for github只认识 `\Phi` 不认识 `\varPhi`

- **3σ法则**：$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$
- 联合正态分布
  两个随机变量X和Y若服从**二元正态分布**或**二元联合正态分布**(two dimensional joint normal distribution)，则其分布密度为

$$p(X,Y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}$$

  其中ρ为X与Y的相关系数。

  > 射击命中点的位置一般服从二元正态分布。

  Z为若干互相独立正态随机变量的线性组合$Z = \sum_{i=1}^n a_i X_i$，则Z也服从正态分布，且

$$\mu_Z = \sum_{i=1}^n a_i \mu_{X_i}, \quad \sigma_Z^2 = \sum_{i=1}^n a_i^2 \sigma_{X_i}^2$$

  其中μ_{X_i}和$\sigma_{X_i}^2$分别是X_i的均值和方差。

瑞利分布(Rayleigh distribution)

- 概率密度函数

$$p(x)=\begin{cases}\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}} & x>0\\ 0 & x\le 0\end{cases}$$

> 如果射击命中点的位置服从二元正态分布，则命中点离靶心的距离服从瑞利分布。

泊松分布(Poisson distribution)

- 概率分布

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\quad k=0,1,2,\cdots$$

其中$\lambda>0$。
泊松分布的均值和方差均为λ。

韦布尔分布(Weibull distribution)

- 概率密度函数

$$p(x)=\begin{cases}\frac{\beta}{\eta}(\frac{x}{\eta})^{\beta-1}e^{(\frac{x}{\eta})^{\beta}} & x>0\\ 0 & x\le 0\end{cases}$$

β、η均为正常数。

> 大量的试验表明，许多产品的寿命服从韦尔分布。

平均分布(uniform distribution)

- 概率密度函数

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

其期望为$(a+b)/2$，方差为$(b-a)^3/12$。

> 平均分布的典型例子是随机初相位角，它一般被假定为在$[0,2\pi)$区间内是均匀分布的。