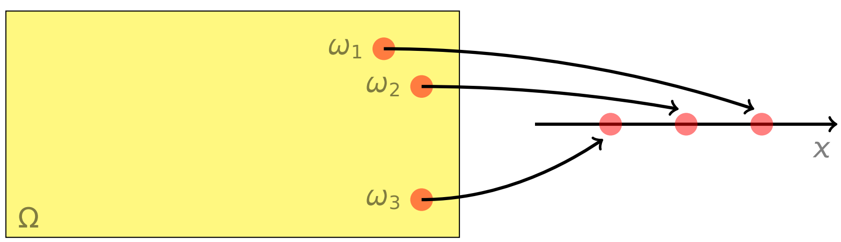
注意：文件只列出了部分要点，没有提到的内容不代表不会考察.

2 一维随机变量及其分布

随机变量

定义 1. 设 $X = X(\omega)$是定义在样本空间Ω上的实值函数，称$X = X(\omega)$为随机变量.

下图给出样本点ω与实数$X = X(\omega)$对应的示意图.



随机变量一般用大写英文字母X、Y、Z或小写希腊字母ξ、η、γ来表示.

一般地，若I是一个实数集，$\{X \in I\}$记为事件B，即

$$B = \{\omega | X(\omega) \in I\},$$

则

$$P\{X \in I\} = P(B) = P\{\omega | X(\omega) \in I\}.$$

按照随机变量可能取值的情况，可以把它们分为两类：离散型随机变量和非离散型随机变量，而非离散型随机变量中最重要的是连续型随机变量.

离散型随机变量

定义 2. 如果随机变量的全部可能取的值只有有限个或可列无限多个，则称这种随机变量为离散型随机变量.

一般地，设离散型随机变量 X 所有可能取的值为x_k $(k = 1, 2, \ldots)$, X 取各个可能值的概率，即事件$\{X = x_k\}$的概率为

$$P\{X = x_k\} = p_k, \; k = 1, 2, \ldots, \tag{1}$$

称(1)式为离散型随机变量X的分布律或概率分布.

分布律也可以用下面的表格来表示：

X	x_1	$x_2,$	$\cdots$	x_n	$\cdots$
p_k	p_1	$p_2,$	$\cdots$	p_n	$\cdots$

由概率的定义，式p_k应满足以下条件：

1. $p_k \geq 0, k = 1, 2 \ldots;$
2. $\sum\limits_{k=1}^{\infty} p_k = 1.$

2.1 常用离散型分布

离散型·两点分布

定义：若随机变量X只能取0或1，其概率分布为：

$$P\{X = 1\} = p, \; P\{X = 0\} = 1 - p \quad (0 < p < 1)$$

则称X服从参数为p的两点分布，记为

$$X \sim b(1, p).$$

离散型·二项分布

定义：如果随机变量X服从以下分布律

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k},$$

其中$0<p<1, 0\leq k\leq n$, 则称X服从参数为n,p的二项分布. 记为

$$X\sim b(n,p).$$

离散型·泊松分布

定义：如果随机变量X服从以下分布律

$$P\{X=m\}=\frac{\lambda^m}{m!}e^{-\lambda},\quad m=0,1,\cdots$$

其中$\lambda>0$, 则称X服从参数为λ的泊松分布, 记为

$$X\sim\pi(\lambda).$$

二项分布的泊松近似

定理 (泊松定理)**.** 在 n 重伯努利试验中, 事件 A 在每次试验中发生概率为 p_n （注意这与实验的次数 n 有关）, 如果$n\to\infty$ 时, $np_n\to\lambda$ (λ为常数), 则对任意给定的非负整数 k, 有

$$\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}.$$

2.2 随机变量的分布函数

随机变量的分布函数

定义：对任何随机变量X, 称函数

$$F(x) := P\{X \le x\}, \quad x \in \mathbb{R}$$

为X的分布函数. 设$F(x)$为某随机变量的分布函数, 则其有以下性质：

1. 广义单增：对任意实数$a < b$, 总有$F(a) \le F(b)$；
2. $0 \le F(x) \le 1$, 且$F(-\infty) = 0, \quad F(+\infty) = 1.$

离散型随机变量X的概率分布 $p_k = P\{X = x_k\}$ 满足

1. $p_k \ge 0, \quad k = 1, 2, \cdots$
2. $\sum_k p_k = 1$
3. $F(x) = \sum_{x_k \le x} p_k$

2.3 连续型随机变量及其概率密度函数

连续型随机变量

设X为连续性随机变量, 则对每个实数a, 总有

$$P\{X = a\} = 0.$$

任意区间上概率的计算：由概率密度函数的定义可知,

$$P\{X \in (a, b]\} = F(b) - F(a) = \int_a^b f(x)dx.$$

上式中的区间$(a, b]$改为(a, b), $[a, b)$或$[a, b]$后等式仍成立.

连续型随机变量X的概率密度 $f(x)$ 满足

1. $f(x) \geq 0, \quad x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$
3. $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$

连续型·均匀分布

定义：若随机变量X有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}, \qquad (a < b)$$

则称X服从区间$[a, b]$上的均匀分布, 记为

$$X \sim U[a, b].$$

连续型·指数分布

定义：如果随机变量X有以下概率密度

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

其中$\lambda > 0$, 则称X服从参数为λ的指数分布. 其分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

连续型·正态分布

定义：如果随机变量X有以下概率密度

$$\varphi_{\mu,\sigma^2}(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

其中μ,σ为常数且$\sigma>0$, 则称X服从正态分布. 简记为

$$X\sim N(\mu,\sigma^2).$$

称$N(0,1)$为标准正态分布, 并简写$\varphi_{0,1}(x)$为$\varphi(x)$.

正态分布的分布函数为

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

该函数不是初等函数. 标准正态分布的分布函数简记为$\Phi(x)$.

标准正态分布的分布函数的性质

$$\Phi(-x)=1-\Phi(x).$$

服从正态分布随机变量的标准化：若$X\sim N(\mu,\sigma^2)$, 则

$$\frac{X-\mu}{\sigma}\sim N(0,1).$$

2.4 随机变量函数的分布

离散型随机变量函数的分布

设离散型随机变量X的概率分布为

$$P\{X=x_k\}=p_k,\quad k=1,2,\cdots$$

令$Y=g(X)$, 则Y也是一个离散型随机变量, 其分布可按如下步骤求得

1. 根据函数关系列出Y的所有可能值；
2. 对Y的每个可能值y, $P\{Y=y\}$等于所有满足$g(x_k)=y$的p_k之和.

连续型随机变量函数的分布

对连续型随机变量X, 求$Y=g(X)$的密度函数的基本方法是

1. 根据函数关系先求Y的分布函数
$$F_Y(y)=P\{Y\le y\}=P\{g(X)\le y\}$$
2. 然后对$F_Y(y)$求导可得Y的概率密度.

定理 1. 设随机变量X具有概率密度$f_X(x), -\infty<x<+\infty$; 函数$g(x)$处处可导且恒有$g'(x)>0$(或恒有$g'(x)<0$) 则$Y=g(X)$是连续型随机变量, 其概率密度为
$$f_Y(y)=\begin{cases} f_X[h(y)]\cdot|h'(y)|, & y\in(\alpha,\beta) \\ 0, & otherwise. \end{cases}$$
其中 $h(y)$是$g(x)$的反函数,
$$\alpha=\min(g(-\infty),g(+\infty)),\ \beta=\max(g(-\infty),g(+\infty)).$$

设随机变量 $X\sim N(\mu,\sigma^2)$, $Y=aX+b$, 则 $Y\sim N\left(a\mu+b,(a\sigma)^2\right)$.