概率论知识复习

一、 常见分布及相关知识点

离散型随机变量及分布

离散型随机变量: 如果一个随机变量只可能取有限个值或可列无限个值, 那么称这个随机变量为(一维)离散型随机变量.

离散型随机变量分布的表现形式称为概率函数.

定义　设 $\Omega_X = \{a_1, a_2, \cdots, a_n, \cdots\}$, 且 $P(X = a_i) = p_i$,

其中 p_i 满足: (1) $p_i \geq 0 (i = 1, 2, \cdots)$; (2) $\sum_{i=1}^{\infty} p_i = 1$.

那么称表达式 “ $P(X = a_i) = p_i$, $i = 1, 2, \cdots$ ” 为随机

变量 X 的概率函数或概率分布(律).

随机变量的分布律或概率函数常用表格表示.

$$\begin{array}{c|cccccc} X & a_1 & a_2 & \cdots & \cdots & a_n & \cdots \\ \hline P_i & p_1 & p_2 & \cdots & \cdots & p_n & \cdots \end{array}$$

其中概率为 0 的取值不再罗列.

利用概率函数，可以求出任意随机事件 $\{X \in S\}$ 的概率:

$$P(X \in S) = \sum_{i:\, a_i \in S} P(X = a_i) = \sum_{i:\, a_i \in S} p_i \ .$$

1. $0-1$ 分布

如果随机变量 X 的概率函数为

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

则称 X 服从参数为 p 的 **0-1** 分布，记为 $X \sim B(1,p)$.

或表示为：$P(X=k)=p^k(1-p)^{1-k},\ k=0,1$;

X	1	0
P_r	p	$1-p$

.

凡是样本空间只含有两个样本点的试验或贝努利试验都可以用服从 $0-1$ 分布的随机变量来刻划．如产品的好坏等．

2.二项分布

如果随机变量 X 的概率函数为

$$P(X=k)=C_n^k p^k (1-p)^{n-k},\ k=0,1,2,\cdots,n,$$

则称 X 服从参数为 n,p 的二项分布，记为 $X\sim B(n,p)$.

其中 $0<p<1$.

(1) n 次重复独立试验中，事件 A 发生的次数服从二项分布；

(2)设 $X_1,\cdots,X_n$ 是独立同分布的随机变量，且 $X_i\sim B(1,p)$,

$i=1,2,\cdots,n$. 记 $Y=\sum_{i=1}^{n}X_i$，则 $Y\sim B(n,p)$.

(3)设 $X\sim B(m,p),Y\sim B(n,p)$ 且相互独立，则有

$X+Y\sim B(m+n,p)$. 称为二项分布的可加性.

例　某市的血库急需**AB**型血，需从体检合格的献血者中获得. 已知在体检合格的献血者中，**AB**型血的比例为百分之二，问：至少需要多少位体检合格的献血者才能保证至少获得一份该血型的概率至少为**0.95**？

解　设至少需要n位体检合格的献血者才能保证至少获得一份**AB**型血的概率达到**0.95**，记这n位体检合格的献血者中**AB**型血人数为X，则$X \sim B(n, 0.02)$.

由题意 $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.98^n \geq 0.95$

解得 $n \geq \dfrac{\ln 0.05}{\ln 0.98} \approx 148.3$，故取 $n = 149$.

3.泊松分布

若随机变量 X 的概率函数为

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

则称 X 服从参数为 λ 的泊松分布，记作 $X\sim P(\lambda)$.

(1)泊松分布的概率函数值可以查表得到;

(2)由无穷级数理论可知 $\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\cdot\mathrm{e}^{-\lambda}=1$;

(3)设 $X\sim P(\lambda_1), Y\sim P(\lambda_2)$ 且相互独立，则有

$X+Y\sim P(\lambda_1+\lambda_2)$. 称为泊松分布的可加性.

分布函数定义　给定一个随机变量 X，称定义域为$(-\infty,+\infty)$上的实值函数 $F(x)=P(X\le x)$ 为随机变量 X 的分布函数.

由定义可知，对任意实数 a,b，$-\infty<a<b<+\infty$，总有

$$P(a<X\le b)=F(b)-F(a);$$

分布函数的性质： 设 $F(x)$是随机变量 X 的分布函数，则

(1) $0\le F(x)\le 1$；

(2)分布函数单调不减;

(3)对任意的 $x\in(-\infty,+\infty)$，分布函数右连续;

(4) $\lim\limits_{x\to-\infty}F(x)=0,\ \lim\limits_{x\to+\infty}F(x)=1.$

连续型随机变量及分布

若随机变量的取值范围是一个区间或者若干个区间的并集,那么就称这类随机变量为连续型随机变量.

描述连续型随机变量分布特点的量就是概率密度函数.

定义　给定一个连续型随机变量 X，如果存在一个定义域为$(-\infty,+\infty)$的非负实值函数$f(x)$，使得X的分布函数$F(x)$可以表示为$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t,\ -\infty<x<+\infty$，那么称$f(x)$为连续型随机变量 X 的概率密度函数.

概率密度函数需满足下面两个条件:

(1) $f(x)\geq 0, -\infty<x<+\infty$；(2) $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$.

且有下列结论:

连续型随机变量的性质　设 X 是任意一个连续型随机变量, $F(x)$ 与 $f(x)$ 分别是它的分布函数与概率密度函数, 则有

(1) $F(x)$是连续函数且在 $f(x)$ 的连续点处有 $F'(x)=f(x)$；

(2)对任意常数 $c\in(-\infty,+\infty)$，有 $P(X=c)=0$；

(3)对任意的两个常数 $a,b,\ -\infty<a<b<+\infty$，有

$$P(a<X\le b)=P(a\le X\le b)=P(a<X<b)$$

$$=P(a\le X<b)=F(b)-F(a)=\int_a^b f(x)\mathrm{d}x.$$

特别地，有 $P(X>a)=\int_a^{+\infty} f(x)\mathrm{d}x=1-F(a)$，

$$P(X\le b)=\int_{-\infty}^{b} f(x)\mathrm{d}x=F(b)\quad.$$

更一般地，对实数轴上任意集合S，$P(X\in S)=\int_S f(x)\mathrm{d}x$．

1.均匀分布

若随机变量 X 的密度函数为 $f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其余}.\end{cases}$

则称 X 服从区间 $[a,b]$ 上的**均匀分布**，记作 $X\sim R(a,b)$.

均匀分布的分布函数为 $F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x<b,\\ 1, & x\ge b.\end{cases}$

2.指数分布

如果随机变量 X 的密度函数为 $f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x}, & x>0, \\ 0, & 其余.\end{cases}$

则称 X 服从参数为 λ 的指数分布, 记为 $X \sim E(\lambda), \lambda>0$.

指数分布的分布函数为 $F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x \geq 0, \\ 0, & x<0.\end{cases}$

若 $X \sim E(\lambda), 0 \leq a<b$, 则 $P(a<X \leq b)=\mathrm{e}^{-\lambda a}-\mathrm{e}^{-\lambda b}$.

例　设顾客在某银行的窗口等待服务的时间服从参数为 **0.2** 的指数分布. 某顾客在窗口等待服务，若超过**10**分钟就离开.

⑴设某顾客某天去银行，求他没等到服务就离开的概率.

⑵设某顾客一个月要去银行五次，求五次中最多有一次没有等到服务而离开的概率.

解　⑴设 X 表示等候服务的时间，则 $X \sim E(0.2)$.

所求概率为 $P(X > 10) = 1 - F(10) = \mathrm{e}^{-2} \approx 0.135$;

⑵记 Y 为该顾客在这五次当中因未等到服务而离开的次数，则 $Y \sim B(5, 0.135)$. 因此所求概率即为

$$P(Y \le 1) = C_5^0(1-0.135)^5 + C_5^1 0.135 \times (1-0.135)^4 \approx 0.86.$$

3.正态分布

设随机变量 X 的概率密度函数为 $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，

$-\infty<x<+\infty$．其中参数 $\mu\in R$，而 $\sigma>0$．则称 X 服从

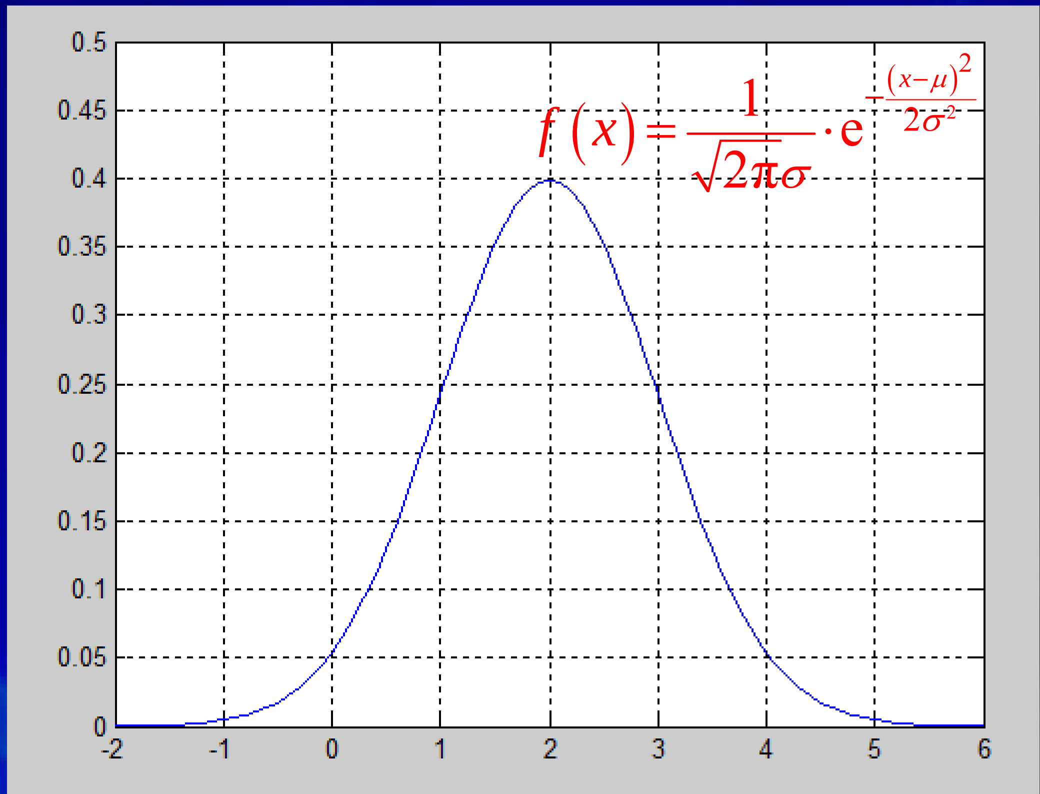
服从参数为 μ 和 σ^2 的正态分布, 记为 $X\sim N(\mu,\sigma^2)$.

服从正态分布的随机变量统称为正态随机变量.

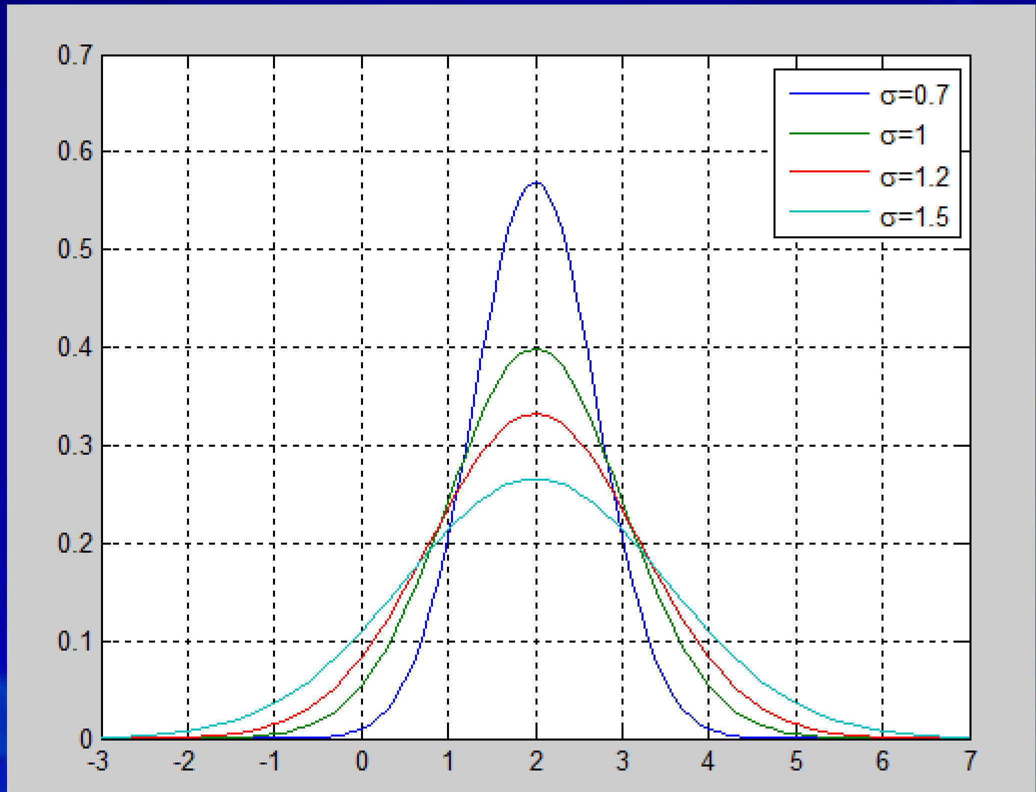
$\mu=0,\sigma^2=1$时的正态分布称为标准正态分布. 其概率密度

函数和分布函数分别为 $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}},-\infty<x<+\infty$；

$\Phi(x)=\int_{-\infty}^{x}\varphi(t)\mathrm{d}t=P(X\le x),-\infty<x<+\infty.$



正态分布的密度函数曲线图形



不同σ的正态分布密度函数图形

(1)当$x>0$时，$\Phi(x)$ 的值可以查附表1得到 **(P295-296)**，且

$$P(a<X\le b)=\Phi(b)-\Phi(a)$$

(2)当$x<0$时，有 $\Phi(x)=1-\Phi(-x)$．可推出$\Phi(0)=0.5$；

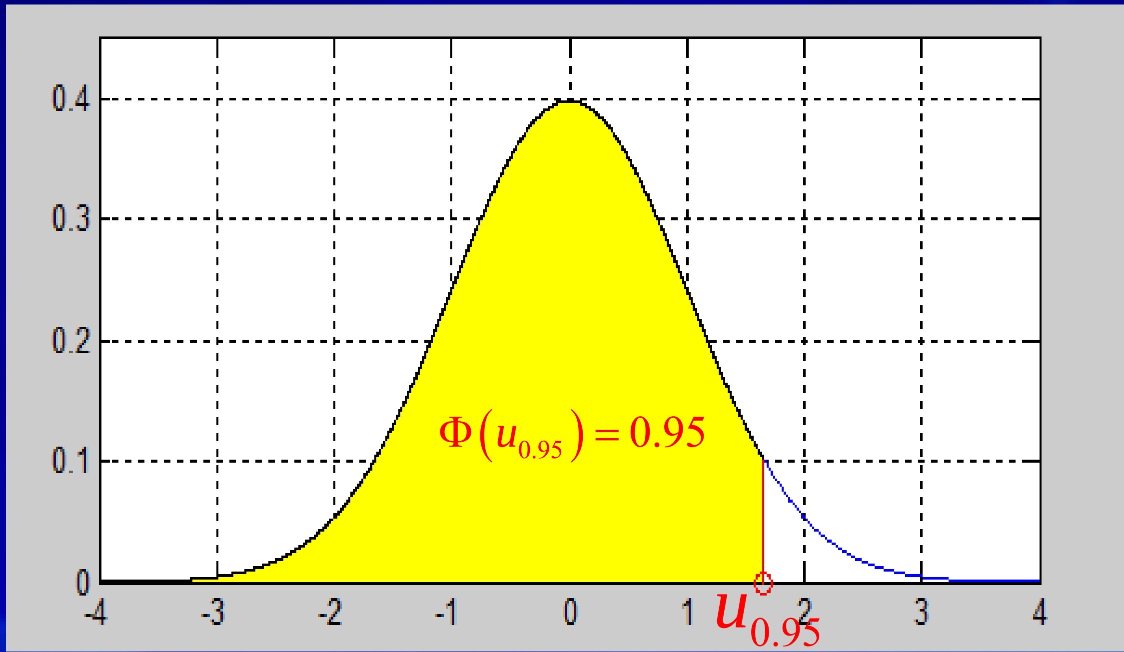
(3)若$X\sim N(\mu,\sigma^2)$，则$P(a<X\le b)=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$；

特别地 $P(X>a)=1-\Phi(\frac{a-\mu}{\sigma})$, $P(X\le b)=\Phi(\frac{b-\mu}{\sigma})$；

(4)满足$\Phi(u_p)=\int_{-\infty}^{u_p}\varphi(x)\mathrm{d}x=P(X\le u_p)=p, 0<p<1$的

数值 u_p称为是标准正态分布X 的$p-$分位数.

且当$p\ge\frac{1}{2}$时，$u_p\ge 0$；当 $p<\frac{1}{2}$ 时，$u_p=-u_{1-p}<0$．



分位数的几何意义

例　设正态随机变量 $X \sim N(\mu, 4^2)$，$Y \sim N(\mu, 5^2)$，记 $p_1 = P(X \le \mu - 4), p_2 = P(Y \ge \mu + 5)$，则对任意实数 μ，试比较概率值 p_1 和 p_2 .

解 $p_1 = \Phi(\dfrac{\mu - 4 - \mu}{4}) = \Phi(-1), p_2 = 1 - \Phi(\dfrac{\mu + 5 - \mu}{5}) = 1 - \Phi(1)$

可知对任意实数 μ，都有 $p_1 = p_2$.

(5)设 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 且相互独立，则有 $kX + lY + b \sim N(k\mu_1 + l\mu_2 + b, k^2\sigma_1^2 + l^2\sigma_2^2)$.

随机变量的相互独立性

若 $r.v.(X,Y)$ 的联合密度函数 $f(x,y)=f_X(x)\cdot f_Y(y)$ 对所有的连续点 x,y 成立, 则称连续型随机变量 X,Y 相互独立.

在离散型随机变量的情形, 独立性定义为下列形式:

$P(X=x,Y=y)=P(X=x)\cdot P(Y=y)$ 对任意 x,y 成立.

更一般地, 若 X,Y 的联合分布函数 $F(x,y)=F_X(x)\cdot F_Y(y)$ 对所有的 $x,y\in R$ 成立, 则称随机变量 X,Y 相互独立.

设 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量，对任意正整数 $m(1 \le m \le n-1)$，随机变量 $g(X_1, X_2, \cdots, X_m)$ 与随机变量 $h(X_{m+1}, \cdots, X_n)$ 相互独立，其中 g 与 h 都是单值函数.

由此可得，若 X, Y 相互独立，则 X^2 与 Y^2, $X+a$ 与 $Y+b$ 等都相互独立.

更一般地，若 X, Y 相互独立，则 $g(X)$ 与 $h(Y)$ 也相互独立.