第二章 随机变量及其分布

修贤超

https://xianchaoxiu.github.io

目录

随机变量

- 定义: 设随机试验的样本空间为 $S=\{e\}$, $X=X(e)$ 是定义在样本空间 S 上的实值单值函数, 称 $X=X(e)$ 为 随机变量
- 通常用大写字母表示随机变量, 小写字母表示实数
- 如在百分制考试成绩，样本空间为

$$S=\{x \mid 0 \leq x \leq 100\}$$

随机变量为

$$X=x$$

事件考试及格的概率可表示为

$$P\{60 \leq X \leq 100\}$$

例 2

- 在很多情况下，随机试验的结果就是实数, 即样本点 e 本身是实数
- 袋中有编号分别为 1、2、3 的 3 只球. 在袋中任取一球, 放回, 再任取一球. 点数之和的样本空间为

$$S=\{(i,j)\mid i,j=1,2,3\}$$

随机变量为

$$X=X(e)=X((i,j))=i+j$$

概率分布为

X	2	3	4	5	6
P	1/9	2/9	3/9	2/9	1/9

随机变量

- 在很多情况下，随机试验的结果可以表示为实数或映射为一个实数
  - 投掷硬币，正面记为 1，背面记为 0
  - 产品质量检测，合格记为 1，次品记为 0
  - 小球下落遇到钉子，偏向左边为 0，偏向右边为 1

目录

离散型随机变量及其分布律

- 可能取到的值是有限个或可列无限多个的随机变量称为离散型随机变量
  - 掷骰子的数值 $1, 2, \cdots, 6$
  - 某人射击的点数 $0, 1, 2, \cdots, 10$
  - 某城市 120 急救电话收到的呼唤次数
  - 某元件的寿命 t (反例)

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$p_k = P\{X = x_k\}$	p_1	p_2	$\cdots$	p_n	$\cdots$

- 由概率的定义, p_k 满足如下两个条件
  - $p_k \geq 0, \ \ k = 1, 2, \cdots$
  - $\sum_{k=1}^{\infty} p_k = 1$

例 1

- 一汽车在道路上连续经过 4 组红绿灯, 在每组红绿灯前, 遇到的红灯亮概率为 p. 以 X 表示汽车首次停下时, 它已通过红绿灯组数（设各组红绿灯的工作状态相互独立）, 求 X 的分布律

解答 X 的分布律为

X	0	1	2	3	4
p_k	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$

当 $p=1/2$ 时

X	0	1	2	3	4
p_k	0.5	0.25	0.125	0.0625	0.0625

(0-1) 分布

- 定义: 设随机变量 X 只可能取 0 和 1 两个值, 它的分布律是

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

则称 X 服从以 p 为参数的 (0-1) 分布或两点分布

X	0	1
p_k	$1-p$	5

- 如果样本空间中只包含两个元素, 即 $S = \{e_1, e_2\}$, 可以定义一个服从 (0-1) 分布的随机变量

$$X = X(e) = \begin{cases} 0, & \text{当 } e = e_1 \\ 1, & \text{当 } e = e_2 \end{cases}$$

伯努利试验

- 定义: 设试验 E 只有两个可能结果, 则称 E 为 伯努利 (Bernoulli) 试验
- 定义: 将试验 E 独立重复进行 n 次, 则称 n 重伯努利试验
- 设 $P(A)=p,\ P(\bar{A})=1-p$, 则 n 重伯努利试验中事件 A 发生 k 次的概率为

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

于是

$$\begin{aligned}\sum_{k=0}^{n} P\{X=k\} &= \sum_{k=0}^{n} C_n^k p^k(1-p)^{n-k}\\ &= [p+(1-p)]^n = 1\end{aligned}$$

- 称随机变量 X 满足参数为 (n,p) 的 二项分布, 记为 $X\sim b(n,p)$
- 当 $n=1$ 时, 二项分布转化为 (0-1) 分布

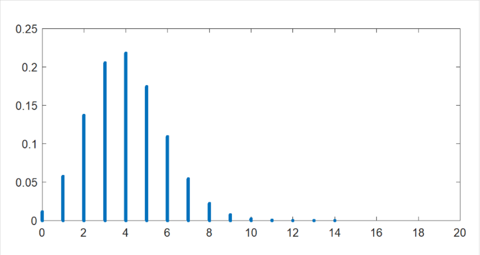
例 2

- 按规定, 某型号电子元件使用寿命超过 1500 小时的为一级品. 已知某批产品一级品率为 0.2, 现在从中随机抽查 20 只, 问恰有 k 只一级品的概率是多少

解答　这里是不放回抽样, 但由于总数很大, 因而可以当作放回抽样来处理. 以 X 表示一级品的只数, 则 $X \sim b(20, 0.2)$, 即

$$P\{X = k\} = C_{20}^{k} 0.2^{k} 0.8^{20-k}, \ \ k = 0, 1, \cdots, 20$$

k	P	k	P	k	P
0	0.012	4	0.218	8	0.022
1	0.058	5	0.175	9	0.007
2	0.137	6	0.109	10	0.002
3	0.205	7	0.055		



课堂练习 1

- 已知某工厂生产的螺钉有缺陷的概率是 0.01. 工厂将螺钉 10 个一包卖给客户, 并承诺 10 个中至多出现一个次品, 否则退款. 问退款概率是多少

课堂练习 1

- 已知某工厂生产的螺钉有缺陷的概率是 0.01. 工厂将螺钉 10 个一包卖给客户, 并承诺 10 个中至多出现一个次品, 否则退款. 问退款概率是多少

解答 设次品的数量为 0.004267, 则 $X \sim b(10, 0.01)$, 于是 X 的分布律为

$$P\{X = k\} = C_{10}^{k} 0.01^{k} (1 - 0.01)^{10-k}, \ k = 0, 1, \cdots, 10$$

所求的概率为

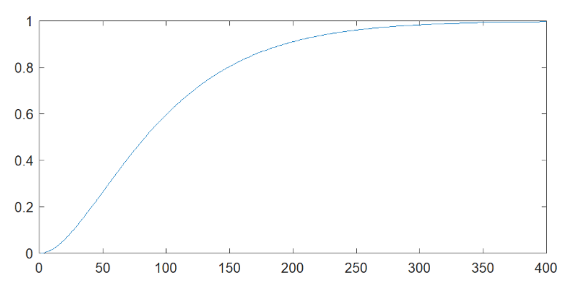
$$\begin{aligned}
&1 - P\{X = 0\} - P\{X = 1\} \\
&= 1 - C_{10}^{0} 0.01^{0} (1 - 0.01)^{10-0} - C_{10}^{1} 0.01^{1} (1 - 0.01)^{10-1} \\
&= 0.004267
\end{aligned}$$

例 3

- 某人单次射击命中率为 0.02，独立射击 400 次求至少击中两次的概率

解答 设击中的次数为 X, 则 $X \sim b(400, 0.02)$, 于是所求的概率为

$$
\begin{aligned}
& 1 - P\{X=0\} - P\{X=1\} \\
& = 1 - C_{400}^{0} 0.02^{0} 0.98^{400} - C_{400}^{1} 0.02^{1} 0.98^{399} \\
& = 0.9972
\end{aligned}
$$



例 4

- 设有 80 台同类型设备，各台工作相互独立，发生故障的概率都是 0.01，且一台设备可由一个人修理．考虑两种配备维修工人的方案：其一是配备 4 人，每人负责 20 台；其二是由 3 人共同维护 80 台．比较两种方法在设备发生故障时不能及时维修的概率

解答　（方案一）设 X 表示同一时刻发生故障的台数，一组（20 台）可以及时维修的概率为

$$
\begin{aligned}
&P\{X=0\}+P\{X=1\}\\
&=C_{20}^{0}0.01^{0}0.99^{20-0}-C_{20}^{1}0.01^{1}0.99^{20-1}\\
&=0.9831
\end{aligned}
$$

则四组设备可以及时维修的概率 $0.9831^4=0.9343$，发生故障时不能及时维修的概率 $10.9831^4=0.0657$

例 4

- 设有 80 台同类型设备，各台工作相互独立，发生故障的概率都是 0.01，且一台设备可由一个人修理. 考虑两种配备维修工人的方案：其一是配备 4 人，每人负责 20 台；其二是由 3 人共同维护 80 台. 比较两种方法在设备发生故障时不能及时维修的概率

解答 （方案二）设 X 表示同一时刻发生故障的台数，由 3 人共同维护 80 台的概率为

$$\begin{aligned}
&1 - P\{X=0\} - P\{X=1\} - P\{X=2\} - P\{X=3\} \\
&= 1 - C_{80}^{0} 0.01^{0} 0.99^{80-0} - C_{80}^{1} 0.01^{1} 0.99^{80-1} \\
&\quad - C_{80}^{2} 0.01^{2} 0.99^{80-2} - C_{80}^{3} 0.01^{3} 0.99^{80-3} \\
&= 0.0087
\end{aligned}$$

尽管方案二每个人的任务重了，但工作效率提高了

泊松分布

- 长期观察, 发现食堂在中午 11 点至 12 点之间平均有 λ 人进入食堂吃饭. 那么某天中午 11 点至 12 点之间有 k 个人进入食堂吃饭的概率是多少

  解答 将 11 点至 12 点这一个小时平均分为 n 个时间段, 每段时间为 $1/n$ 小时且 $n > k$, 且每段时间内最多只有一个人走进食堂, 则在每段时间有人走进食堂的概率是 $p = \lambda/n$. 于是 11 点至 12 点之间有 k 个人走进食堂的概率是

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

  为了避免出现两个人在同一时间段走进食堂的情况, 需要将这一个小时划分成为更小的时间段, 即 $n \to \infty$, 则有

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

泊松分布

- 进一步，可以得到

$$
\begin{aligned}
P\{X=k\} &= \lim_{n\to\infty} C_n^k p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} C_n^k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left[1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k e^{\lambda}}{k!}
\end{aligned}
$$

泊松分布

- 定义: 设随机变量 X 所有可能的取值为 $0, 1, 2, \cdots$, 而取各个值的概率为

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

其中 $\lambda > 0$ 是常数, 则称 X 服从参数为 λ 的泊松分布, 记为 $X \sim \pi(\lambda)$

- 服从泊松分布的例子
  - 一本书中的一页中错误字符的数量
  - 某段时间某路口发生交通事故的数量
  - 一个社区中年龄超过百岁的老人数量
  - 一天中拨错电话号码的数量
  - 一天中进入邮局的顾客数量

泊松分布

- **泊松定理**: 设 $\lambda > 0$ 是一个常数, n 是任意正整数, 设 $np_n = \lambda$, 则对于任一固定的非负整数 k, 有

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

- 根据泊松定理, 当 n 很大, p_n 很小, $\lambda = np_n$ 适中时, 有

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

即用泊松分布概率近似计算二项分布概率

例 5

- 某公司生产芯片，次品率 0.1%，求在 1000 只产品中至少有 2 只次品的概率

解答 设 X 表示次品的数量

- 服从二项分布 $X \sim (1000, 0.001)$，则

$$
\begin{aligned}
P\{X \geq 2\} &= 1 - P\{X = 0\} - P\{X = 1\} \\
&= 1 - C_{1000}^{0} 0.001^{0} 0.999^{1000} - C_{1000}^{1} 0.001^{1} 0.999^{999} \\
&= 0.2642411
\end{aligned}
$$

- 服从泊松分布 $X \sim \pi(1)$，则

$$
\begin{aligned}
P\{X \geq 2\} &= 1 - P\{X = 0\} - P\{X = 1\} \\
&= 1 - \frac{1^{0} e^{-1}}{0!} - \frac{1^{1} e^{-1}}{1!} \\
&= 0.2642411
\end{aligned}
$$

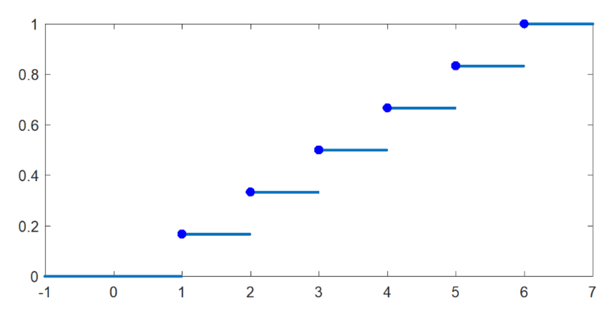
目录

随机变量的分布函数

- 投掷骰子，每个点数的概率分布为

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

$$F(x) = P(X \le x) = \begin{cases} 0, & x < 1 \\ 1/6, & 1 \le x < 2 \\ 1/3, & 2 \le x < 3 \\ 1/2, & 3 \le x < 4 \\ 2/3, & 4 \le x < 5 \\ 1, & 6 \le x \end{cases}$$



随机变量的分布函数

- 定义: 设 X 是一个随机变量, x 是任意实数, 函数

$$F(x) = P(X \leq x), \; -\infty < x < \infty$$

称为 X 的分布函数（Cumulative Distribution Function）

- 性质:
  - F 是非减函数, 即如果 $x_1 \leq x_2$, 则 $F(x_1) \leq F(x_2)$
  - F 是右连续的, 即 $F(x+0) = F(x)$
  - $0 \leq F(x) \leq 1$, 且

$$F(\infty) = \lim_{x\to\infty} F(x) = 1$$
$$F(-\infty) = \lim_{x\to-\infty} F(x) = 0$$

例 2

- 直径 2 米圆靶，随机变量 X 表示中靶时弹着点距离圆心距离, 假设击中靶上同心圆盘的概率与该圆盘面积成正比, 求随机变量 X 的分布函数?

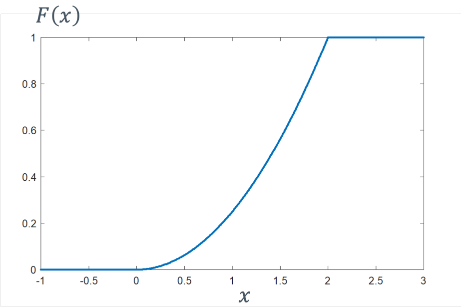
解答　考虑以下三种情况

- 当 $x < 0$ 时, 有 $F(x) = P(X \leq x) = 0$
- 当 $0 \leq x \leq 2$ 时, 有

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(X \leq 0) + P(0 < X \leq x) \\ &= kx^2 \end{aligned}$$

由 $P(X \leq 2) = 1$, 可知 $k = 4$

- 当 $x > 2$ 时, 有 $F(x) = P(X \leq x) = 1$



目录

连续型随机变量及其概率密度

- **定义**: 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

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

则称为 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

- 概率密度的性质
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - 对于任意实数 x_1, x_2 $(x_1 \leq x_2)$, 有

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

  - 若 $f(x)$ 在 x 处连续, 则有 $F'(x) = f(x)$

例 1

- 设随机变量 X 具有概率密度

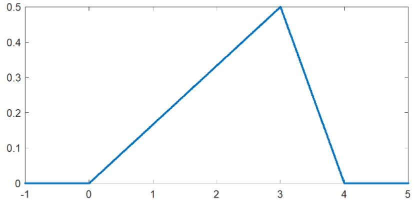
$$f(x)=\begin{cases} kx, & 0\le x<3 \\ 2-\frac{x}{2}, & 3\le x\le 4 \\ 0, & \text{其他} \end{cases}$$

(1) 确定 k; (2) 求 $F(x)$; (3) 求 $P(1<x\le \frac{7}{2})$

解答 (1) 由 $\int_{-\infty}^{\infty} f(x)dx=1$ 得

$$\int_0^3 kxdx+\int_3^4\left(2-\frac{x}{2}\right)dx=\frac{9k}{2}+\frac{1}{4}=1$$

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

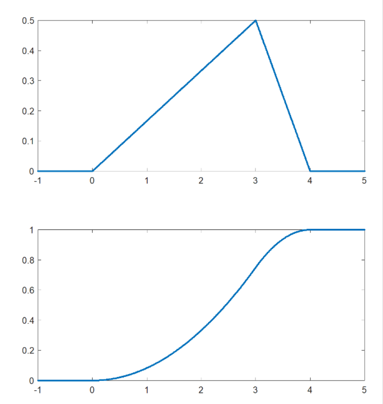


例 1

解答 (2) X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0 \\ \int_0^x \frac{x}{6} dx, & 0 \le x < 3 \\ \int_0^3 \frac{x}{6} dx + \int_3^x \left(2 - \frac{x}{2}\right), & 3 \le x \le 4 \\ 1, & x \ge 4 \end{cases}$$

$$= \begin{cases} 0, & x < 0 \\ \frac{x^2}{12}, & 0 \le x < 3 \\ -3 + 2x - \frac{x^2}{4}, & 3 \le x \le 4 \\ 1, & x \ge 4 \end{cases}$$



(3) $P(1 < x \le \frac{7}{2}) = F(\frac{7}{2}) - F(1) = \frac{41}{48}$

均匀分布

- 定义: 若连续型随机变量 X 具有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

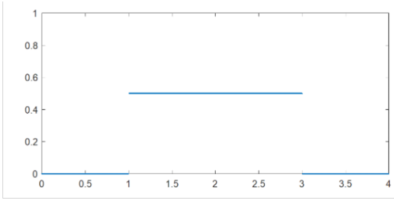
则称 X 在区间 (a, b) 上服从 均匀分布, 记为 $X \sim U(a, b)$

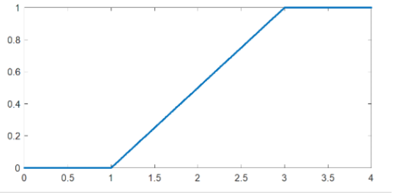
- 易知 $f(x) \geq 0$ 且 $\int_{-\infty}^{\infty} f(x)dx = 1$
- 均匀分布随机变量的分布函数为

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

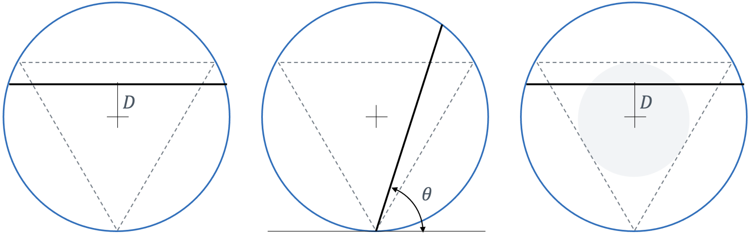
均匀分布

- 例如$X \sim U(1,3)$





- Bertrand 悖论



指数分布

- 在某段时间，平均每隔 θ 时间，有一人进入食堂吃饭. 连续两个人进入食堂间隔时间不超过 t 的概率是多少

  解答 时间 t 内有 k 人进入食堂的概率是

  $$P\{X=k\}=\frac{(\frac{t}{\theta})^k e^{-\frac{t}{\theta}}}{k!}$$

  时间 t 内有 0 人进入食堂的概率是

  $$P\{X=0\}=\frac{(\frac{t}{\theta})^0 e^{-\frac{t}{\theta}}}{0!}=e^{-\frac{t}{\theta}}$$

  连续两个人进入食堂间隔时间为 t 的概率是

  $$F(t)=P\{T\leq t\}=1-P(T>t)=1-P(X=0)=1-e^{-\frac{t}{\theta}}$$

  $$\Longrightarrow f(t)=F'(t)=\frac{1}{\theta}e^{-\frac{t}{\theta}}$$

指数分布

- 定义: 若连续型随机变量 X 具有概率密度

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

  其中 $\theta>0$ 是常数, 则称 X 服从参数为 θ 的 指数分布

- 易知 $f(x)\geq 0$ 且 $\int_{-\infty}^{\infty}f(x)dx=1$
- 指数分布随机变量的分布函数为

$$F(x)=\begin{cases}1-e^{-\frac{x}{\theta}}, & x>0\\ 0, & \text{其他}\end{cases}$$

- 试画出指数分布的概率密度和分布函数

指数分布

- 指数分布随机变量具有无记忆性: 对于任意 $s, t > 0$ 有

$$\begin{aligned}
P\{X > s + t \mid X > s\} &= \frac{P\{(X > s + t) \cap (X > s)\}}{P\{X > s\}} \\
&= \frac{P\{(X > s + t)\}}{P\{X > s\}} = \frac{1 - F(s + t)}{1 - F(s)} \\
&= \frac{e^{-\frac{s+t}{\theta}}}{e^{-\frac{s}{\theta}}} = e^{-\frac{t}{\theta}} \\
&= P\{X > t\}
\end{aligned}$$

- 指数分布在可靠性理论与排队论中有广泛的应用
  - 电子元器件的失效时间分布
  - 大型复杂系统的平均故障间隔时间的失效分布
  - 从现在起到战争、地震等不平常事件发生的时间

课堂练习 1

- 假设电话通话时间是一个指数分布随机变量, $\theta = 10$ 分钟. 你和某人前后脚来到电话亭, 你排在某人后边.

  (1) 求你的等待时间超过 10 分钟的概率
  (2) 假如你已等待 5 分钟, 求还需等待超过 10 分钟的概率

课堂练习 1

- 假设电话通话时间是一个指数分布随机变量, $\theta = 10$ 分钟. 你和某人前后脚来到电话亭, 你排在某人后边.
  (1) 求你的等待时间超过 10 分钟的概率
  (2) 假如你已等待 5 分钟, 求还需等待超过 10 分钟的概率

解答 设 X 表示等待时间, 则

$$\begin{aligned} P\{X > 10\} &= 1 - F(10) \\ &= e^{-\frac{10}{10}} = e^{-1} = 0.368 \end{aligned}$$

$$\begin{aligned} P\{X > 15 | X > 5\} &= \frac{1 - F(15)}{1 - F(5)} \\ &= \frac{e^{-\frac{15}{\theta}}}{e^{-\frac{5}{\theta}}} = e^{-1} = 0.368 \end{aligned}$$

正态分布

- 定义: 若连续型随机变量 X 的概率密度为

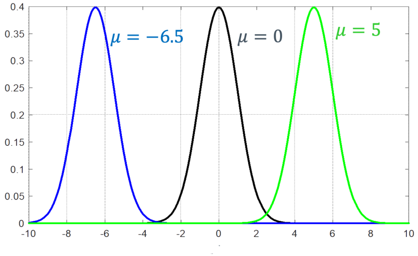
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \infty < x < \infty$$

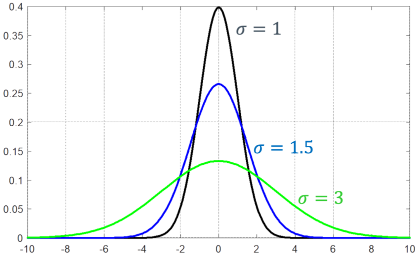
  其中 μ, σ $(\sigma > 0)$ 为常数, 则称 X 服从参数为 μ, σ 的 正态分布或高斯分布, 记为 $X \sim N(\mu, \sigma^2)$

- 中心极限定理说明: 许多实际中的随机变量服从, 或近似服从正态分布, 例如
  - 成年男子身高
  - 许多物理测量误差
  - 分子沿某一方向的速度

正态分布

- 正态分布概率密度函数为钟形（Bell-shaped）曲线，关于 $X = \mu$ 对称
- 参数 μ 决定曲线的峰值位置，峰值大小为 $\frac{1}{\sqrt{s\pi\sigma}}$
- 参数 σ 决定峰的宽度，$x = \mu \pm \sigma$ 为曲线拐点位置
- 在两侧，曲线尾部趋近于 x 轴





正态分布

- **正态分布的概率密度满足** $f(x) \geq 0$ **且** $\int_{-\infty}^{\infty} f(x)dx = 1$

证明 **显然** $f(x) \geq 0$, **下面证明** $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$. **令** $y = \frac{x=\mu}{\sigma}$, **有**

$$I = \frac{1}{\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy$$

令 $x = rcos\theta, y = rsin\theta, dydx = rd\theta dr$, **则**

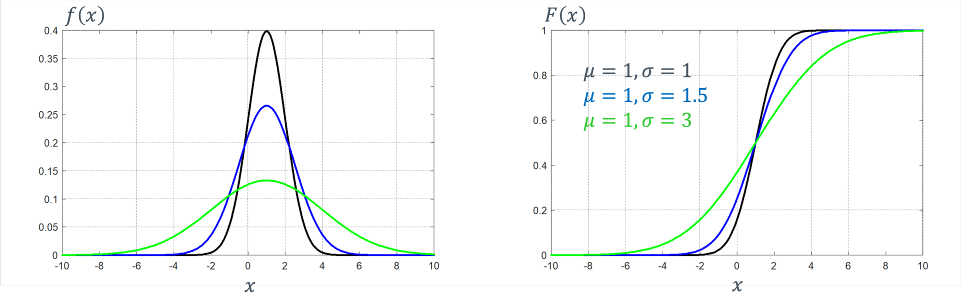
$$I^2 = \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta dr = 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = -2\pi e^{-\frac{r^2}{2}} \Big|_0^{\infty} = 2\pi$$

于是 $I = \sqrt{2\pi}$

正态分布

- 正态分布随机变量的分布函数

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



正态分布

- 如果 $X \sim N(\mu, \sigma^2)$, $Z = aX + b$, 则有 $Z \sim N(a\mu + b, (a\sigma)^2)$

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

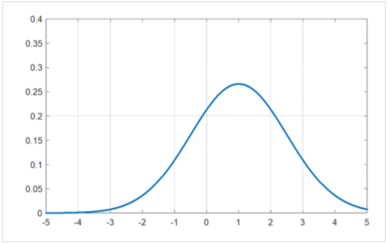
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

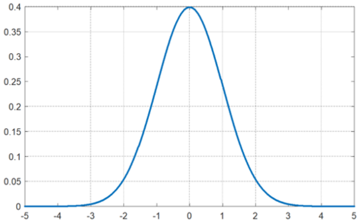
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$Z \sim N(0, 1)$$

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

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



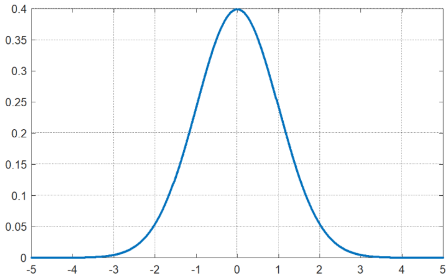


标准正态分布

- 标准正态分布, 又称单位正态分布

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

- $\Phi(-z) = 1 - \Phi(z)$
- $P\{\mu - \sigma < X < \mu + \sigma\}$
  $= \Phi(1) - \Phi(-1) = 0.6826$
- $P\{\mu - 2\sigma < X < \mu + 2\sigma\}$
  $= \Phi(2) - \Phi(-2) = 0.9544$
- $P\{\mu - 3\sigma < X < \mu + 3\sigma\}$
  $= \Phi(3) - \Phi(-3) = 0.9974$



例 3

- 将一温度调节器放置在储存着某种液体的容器内, 调节器整定在 $d\ ^\circ C$, 液体的温度 X 是一个随机变量, 且 $X \sim N(d, 0.5^2)$

  (1) 若 $d = 90\ ^\circ C$, 求 X 小于 $89\ ^\circ C$ 的概率
  (2) 若要求液体的温度至少为 $80\ ^\circ C$ 的概率不低于 0.99, 问 d 至少为多少

解答 (1) 所求概率为

$$\begin{aligned}
P\{X < 89\} &= P\left\{\frac{X-90}{0.5} < \frac{89-90}{0.5}\right\} \\
&= \Phi\left(\frac{89-90}{0.5}\right) = \Phi(-2) \\
&= 1 - \Phi(2) \\
&= 1 - 0.9772 \\
&= 0.0228
\end{aligned}$$

例 3

- 将一温度调节器放置在储存着某种液体的容器内，调节器整定在 $d\ ^\circ C$，液体的温度 X 是一个随机变量，且 $X \sim N(d, 0.5^2)$
  (1) 若 $d = 90\ ^\circ C$，求 X 小于 $89\ ^\circ C$ 的概率
  (2) 若要求液体的温度至少为 $80\ ^\circ C$ 的概率不低于 0.99，问 d 至少为多少

解答 (2) 求 d 应满足

$$
\begin{aligned}
0.99 \le P\{X \ge 80\} &= P\left\{\frac{X-d}{0.5} \ge \frac{80-90}{0.5}\right\} \\
&= 1 - P\left\{\frac{X-d}{0.5} < \frac{80-90}{0.5}\right\} \\
&= 1 - \Phi\left(\frac{80-d}{0.5}\right)
\end{aligned}
$$

即 $\Phi\left(\frac{d-80}{0.5}\right) \ge 0.99 = \Phi(2.327) \implies d \ge 81.1635$

课堂练习 2

- 一场“好”的考试被认为其分数应该服从正态分布，依据分数可求出其均值 μ 与方差 σ^2. 那么，应该如何将分数转化为五级分数制 (A, B, C, D, F)

课堂练习 2

- 一场“好”的考试被认为其分数应该服从正态分布，依据分数可求出其均值 μ 与方差 σ^2. 那么，应该如何将分数转化为五级分数制 (A, B, C, D, F)

  解答　设 X 服从正态分布，则

  - A: $P\{X > \mu + \sigma\} = 1 - \Phi(1) = 0.1587$
  - B: $P\{\mu < X < \mu + \sigma\} = \Phi(1) - \Phi(0) = 0.3413$
  - C: $P\{\mu - \sigma < X < \mu\} = \Phi(0) - \Phi(-1) = 0.3413$
  - D: $P\{\mu - 2\sigma < X < \mu - \sigma\} = \Phi(-1) - \Phi(-2) = 0.1359$
  - E: $P\{X < \mu - 2\sigma\} = \Phi(-2) = 0.0228$

目录

例 1

- 设离散型随机变量 X 具有如下分布律，求 $Y=(X-1)^2$ 的分布律

X	-1	0	1	2
p_k	0.2	0.3	0.1	0.4

解答 Y 的可能取值 $0,1,4$，有

$$P(Y=4)=P(X=-1)=0.2$$
$$P(Y=1)=P(X=0)+P(X=2)=0.7$$
$$P(Y=0)=P(X=1)=0.1$$

于是 $Y=(X-1)^2$ 的分布律为

Y	0	1	4
p_k	0.1	0.7	0.2

例 2

- 设连续型随机变量 X 具有概率密度, 求 $Y=2X+8$ 的概率密度

$$f_X(x)=\begin{cases}\frac{x}{8}, & 0<x<4\\ 0, & 其他\end{cases}$$

解答 分别记 X,Y 的分布函数为 $F_X(x),F_Y(y)$, 有

$$F_Y(y)=P\{Y\le y\}=P\{2X+8\le y\}=P\left\{X\le\frac{y-8}{2}\right\}=F_X\left(\frac{y-8}{2}\right)$$

将 $F_Y(y)$ 关于 y 求导, 得到

$$f_Y(y)=f_X\left(\frac{y-8}{2}\right)\left(\frac{y-8}{2}\right)'=\begin{cases}\frac{1}{8}\times\frac{y-8}{2}\times\frac{1}{2}, & 0<\frac{y-8}{2}<4\\ 0, & 其他\end{cases}$$

$$=\begin{cases}\frac{y-8}{32}, & 8<y<16\\ 0, & 其他\end{cases}$$

例 3

- 设随机变量 X 具有概率密度 $f_X(x)$, $-\infty < x < \infty$, 求 $Y = X^2$ 的概率密度

  解答　分别记 X, Y 的分布函数为 $F_X(x), F_Y(y)$. 由于 $Y = X^2 \geq 0$, 当 $y \leq 0$ 时 $F_Y(y) = 0$. 当 $y > 0$ 时有

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{X^2 \leq 2\} \\ &= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

  将 $F_Y(y)$ 关于 y 求导, 得到

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], & y > 0 \\ 0, & y \leq 0 \end{cases}$$

课堂练习 1

- 设 $X \sim (0,1)$, 其概率密度为

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

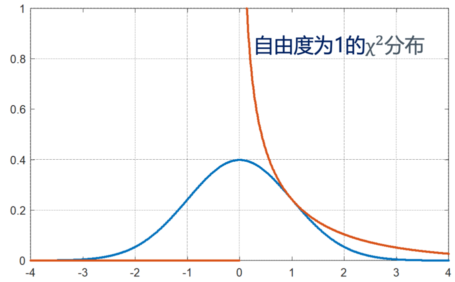
求 $Y = X^2$ 的概率密度

课堂练习 1

■ **解答** 根据例 3 知

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[\varphi(\sqrt{y}) + \varphi(-\sqrt{y})], & y > 0 \\ 0, & y \leq 0 \end{cases}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$



随机变量的函数的分布

- **定理**: 设随机变量 X 具有概率密度 $f_X(x)$, $-\infty < x < \infty$, 又设函数 $g(x)$ 处处可导数且恒有 $g(x)' > 0$（或恒有 $g(x)' < 0$），则随机变量 $Y = g(x)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)]|h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中

- $\alpha = \min\{g(-\infty), g(\infty)\}$
- $\beta = \max\{g(-\infty), g(\infty)\}$
- $h(y)$ 是 $g(x)$ 的反函数

随机变量的函数的分布

- **证明**: 考 虑 $y=g(x)$ 严格单调增的情况, 有 $Y=g(X)$ 且

$$\begin{aligned} F_Y(y) &= P\{g(X)\leq y\} \\ &= P\{X\leq g^{-1}(y)\} = F_X[g^{-1}(y)] \end{aligned}$$

将 $F_Y(y)$ 关于 y 求导, 得到

$$\begin{aligned} f_Y(y) &= f_X[g^{-1}(y)]\frac{d}{dy}g^{-1}(y) \\ &= f_X[g^{-1}(y)]\left|\frac{d}{dy}g^{-1}(y)\right| \end{aligned}$$

当 $y\neq g(x)$ 时, $F_Y(y)$ 为 0 或 1 且 $f_Y(y)=0$

例 5

- **设电压 $V = Asin\Theta$，其中 A 是一个已知正常数，相角 Θ 是一个随机变量，且有 $\Theta \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$，试求电压 V 的概率密度**

  解答 **Θ 的概率密度为**

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi}, & -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0, & \text{其他} \end{cases}$$

因 $V = Asin\Theta$ 在区间 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 严格单调，有

$$\theta = g^{-1}(v) = arcsin\frac{v}{A}, \ h'(v) = \frac{1}{\sqrt{A^2 - v^2}}$$

于是 V 的概率密度为

$$f_V(v) = \begin{cases} \frac{1}{\pi}\frac{1}{\sqrt{A^2-v^2}}, & -A < v < A \\ 0, & \text{其他} \end{cases}$$

小结

- **考试内容**
  - 随机变量、随机变量的分布函数、离散型随机变量的概率分布、连续型随机变量的概率密度、常见随机变量的分布、随机变量函数的分布
- **考试要求**
  - 理解随机变量的概念，理解分布函数的概念及性质，会计算与随机变量相联系的事件的概率
  - 理解离散型随机变量及其概率分布的概念，掌握 0-1 分布、二项分布、几何分布、超几何分布、泊松（Poisson）分布及其应用
  - 了解泊松定理的结论和应用条件，会用泊松分布近似表示二项分布
  - 理解连续型随机变量及其概率密度的概念，掌握均匀分布、正态分布、指数分布及其应用
  - 会求随机变量函数的分布

Q&A

Thank you!

感谢您的聆听和反馈