

概率论与数理统计

第二章: 随机变量 (random variables)

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

1 离散随机变量 (discrete r.v.)

- 引入
- 频率函数
- 分布函数
- 几种重要的离散型随机变量: 两点分布
- 几种重要的离散型随机变量: 二项分布
- 几种重要的离散型随机变量: 几何分布与负二项分布
- 几种重要的离散型随机变量: 泊松流与泊松分布
- 作业

2 连续随机变量 (continuous r.v.)

3 随机变量的函数

古典概型中的几个问题

设 (Ω, \mathcal{A}, P) 为随机试验 E 的概率空间.

问题一

样本空间 Ω 中的元素与试验有关, 从数学角度看, 希望 Ω 是**抽象的集合**.

问题二

非等可能事件 的概率怎么计算?

问题三

在概率论如何使用**微积分理论**?

样本空间 Ω 抽象化

例

抛一枚硬币, 考察正、反面的情况, 则 $\Omega = \{H, T\}$ (有具体含意的空间). 令

$$X(\omega) = \begin{cases} 1, & \omega = H \\ 0, & \omega = T. \end{cases}$$

在上述映射下, 新的“样本空间”为 $\tilde{\Omega} = \{0, 1\}$. 我们有对应关系

$$\{X = 1\} \leftrightarrow \{H\}, \quad \{X = 0\} \leftrightarrow \{T\}.$$

随机变量的定义

一个随机变量是样本空间 Ω 到实数 \mathbb{R} 的映射.

例

将一枚硬币连抛三次, 观察正、反面出现的情况, 则样本空间为

$$\Omega = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}.$$

定义随机变量 $X =$ 正面出现的次数, 则

$$\Omega \xrightarrow{X} \Omega' = \{0, 1, 2, 3\}.$$

我们有事件:

$$\{X = 0\} = \{TTT\},$$

$$\{X = 1\} = \{HTT, THT, TTH\}$$

$$\begin{aligned}\{X \leq 2\} &= \{TTT, HTT, THT, TTH, HHT, HTH, THH\} \\ &= \{X \leq 3\} - \{X = 3\} \\ &= \Omega - \{HHH\}.\end{aligned}$$

随机变量

很多实验产生的结果本身就是随机变量

例

连续型随机变量 (continuous r.v.):

- 某地区的日平均气温 X , 日平均降水量 Y .
- 电子产品的寿命 Y .
- 某城市的日耗电量 W 是一随机变量.

离散型随机变量 (discrete r.v.):

- 一人连续对目标射击 n 次, 击中目标次数 X
- 从一大批产品中随机抽取 n 件进行测试, 其测得的次品数 N .

注

通常用大写字母 X, Y, Z, W 等表示随机变量, 用小写字母 x, y, z, w 表示实数 (非随机).

离散型随机变量

定义

若随机变量仅取有限个或可列个值, 则称 X 为**离散型随机变量**.

例

- 将一枚硬币连抛三次, 观察正、反面出现的情况, 定义 X 为出现正面的次数. X 的取值为 $0, 1, 2, 3$, 故 X 为离散型随机变量.
- 用同一支枪对目标进行射击, 直到击中目标为止, 则射击次数 X 是离散型随机变量.
- 114 查号台一天接到的呼叫次数 X 是离散型随机变量.

问

电子产品的寿命 X 是否是离散型随机变量? **否**.

质量函数

设 X 为离散型随机变量, 设 X 的所有可能取值为 $x_1, x_2, \dots, x_k, \dots$, 且

$$P(X = x_k) = p(x_k) = p_k, \quad k = 1, 2, \dots$$

X 的统计规律完全由数列 $\{x_k\}, \{p_k\}$ 确定.

定义

称 $P(X = x_k) = p(x_k), k = 1, 2, \dots$ 为离散型随机变量 X 的**概率质量函数 (probability mass function)** 或 **频率函数**.

注

离散型随机变量的概率质量函数包括两方面:

- ① 随机变量的所有取值
- ② 随机变量取各个值的概率

例题

将一枚硬币连抛三次, 观察正、反面出现的情况, 记 X 为正面出现的次数. 求 X 的概率质量函数.

解

X 的取值为 $0, 1, 2, 3$, 其样本空间为

$$\Omega = \{TTT\} \cup \{TTH, THT, HTT\} \cup \{THH, HTH, HHT\} \cup \{HHH\}.$$

故 X 的概率密度函数为

$$\begin{aligned} P(X=0) &= \frac{1}{8}, & P(X=1) &= \frac{3}{8} \\ P(X=2) &= \frac{3}{8}, & P(X=3) &= \frac{1}{8}. \end{aligned}$$

问

频率函数有什么特点? **求和为 1.**

频率函数的性质

频率函数的基本性质

① $p(x_k) \geq 0, k = 1, 2, \dots$

② $\sum_{k=1}^{\infty} p(x_k) = 1.$

第二点的证明

$$\sum_{k=1}^{\infty} p(x_k) = \sum_{k=1}^{\infty} P(X = x_k) = P\left(\bigcup_{k=1}^{\infty} \{X = x_k\}\right) = P(\Omega) = 1.$$

两个方向理解基本性质

- 离散型随机变量的频率函数一定满足两个性质
- 满足两个性质的数列 $\{p_k\}$ 一定是某离散型随机变量的频率函数.

频率函数的几种表示方法

解析式法

$$P(X = x_k) = p_k, k = 1, 2, \dots$$

列表法

分布列:

X	x_1	x_2	\cdots	x_k	\cdots
p_k	p_1	p_2	\cdots	p_k	\cdots

矩阵法

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{bmatrix}$$

离散型随机变量的概率分布规律相当于向位于 x_k 处的“盒子”中扔球.

例

一球队要经过四轮比赛才能出线. 设球队每轮被淘汰的概率为 $p = 0.5$. 记 X 表示球队结束比赛时的比赛场数, 求 X 的频率函数.

解

随机变量 X 的可能取值为 $1, 2, 3, 4$. 记 A_k 为球队通过第 k 轮比赛的事件, $k = 1, 2, 3, 4$. 则由独立性,

$$P(X = 1) = P(\bar{A}_1) = p$$

$$P(X = 2) = P(A_1 \bar{A}_2) = P(A_1)P(\bar{A}_2) = (1 - p)p$$

$$P(X = 3) = P(A_1 A_2 \bar{A}_3) = P(A_1)P(A_2)P(\bar{A}_3) = (1 - p)^2 p$$

$$P(X = 4) = P(A_1 A_2 A_3) = (1 - p)^3.$$

代入 $p = 0.5$, 所求的 X 之频率函数为

X	1	2	3	4
p_k	0.5	0.25	0.125	0.125

例

掷两个均匀的骰子, 观测其点数, 令 X 为两骰子点数之和. 求

- a) X 的频率函数
- b) X 为奇数的概率.

解

a)	$X = k$	2	3	4	5	6	7	8	9	10	11	12
	p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b)

$$p_{\text{odd}} = \frac{2 + 4 + 6 + 4 + 2}{36} = \frac{1}{2}.$$

分布函数

思考

随机变量 $X = X(\omega)$ 是样本点 ω 的函数, 是“随机函数”, 不方便应用微积分工具. 怎样将“随机函数”化为“普通函数”?

观察

对于随机变量 X , $\forall x \in (-\infty, \infty)$, 集合

$$\{X \leq x\} = \{\omega \mid X(\omega) \leq x\} \in \mathcal{A}$$

是事件.

定义

称函数 $F(x) = P(X \leq x)$, $-\infty < x < \infty$, 为随机变量 X 的**累积分布函数 (CDF, cumulative distribution function)**, 简称**分布函数**.

注

随机变量的分布函数是关于自变量 x 的普通函数, 它不再是随机的!

例题

设 X 的频率函数为

X	1	2	3
p_k	0.3	0.2	0.5

求 X 的分布函数.

解

由分布函数的定义有

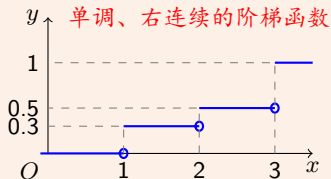
$$F(x) = P(X \leq x) = P(\emptyset) = 0, \quad x < 1$$

$$F(x) = P(X \leq x) = P(X = 1) = 0.3, \quad 1 \leq x < 2$$

$$F(x) = P(X \leq x) = P(X = 1, 2) = 0.5, \quad 2 \leq x < 3$$

$$F(x) = P(X \leq x) = P(\Omega) = 1, \quad 3 \leq x.$$

$$\text{于是, } F(x) = \begin{cases} 0, & x < 1, \\ 0.3, & 1 \leq x < 2, \\ 0.5, & 2 \leq x < 3, \\ 1, & 3 \leq x. \end{cases}$$



分布函数的基本性质

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

分布函数的本质特征

- ① $F(x)$ 是单调不减函数
- ② $0 \leq F(x) \leq 1$ 且

$$F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1.$$

- ③ $F(x)$ 为右连续函数 (或右连左极, càdlàg),

$$F(x+0) := \lim_{t \rightarrow x+0} F(t) = F(x).$$

本质特征的含义

- 随机变量的分布函数必满足这三条性质
- 满足这三条性质的 $F(x)$ 必是某随机变量的分布函数.

概率测度的上连续性

对于单调递增的可列集合列 $A_1 \subset A_2 \subset A_3 \subset \cdots$, 有

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} P(A_n)$$

证明

令 $B_n = A_{n+1} \setminus A_n$, 则 B_1, B_2, \dots 两两不相容且

$$\bigcup_{k=1}^{n-1} B_k = A_n, \quad \forall n; \quad \bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k.$$

由有限可加性及可列可加性,

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} P(B_k) = \lim_{n \rightarrow \infty} P(A_n).$$

概率测度的下连续性

对于单调递减的可列集合列 $A_1 \supset A_2 \supset A_3 \supset \cdots$, 有

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} P(A_n).$$

证明

令 $B_n = A_n \setminus A_{n+1}$, $A_{\infty} = \bigcap_{k=1}^{\infty} A_k$. 则 $A_{\infty}, B_1, B_2, \dots$ 两两不相容. 对任意 n , 有

$A_n = A_{\infty} \cup \left(\bigcup_{k=n}^{\infty} B_k\right)$. (若 $x \in A_n$, 则 k 为使得 $x \notin A_{k+1}$ 的最小指标.) 由可列可加性,

$$P(A_n) = P(A_{\infty}) + \sum_{k=n}^{\infty} P(B_k).$$

特别地, 当 $n \rightarrow \infty$ 时, 右边无穷级数收敛, 因此其尾项级数满足 $\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(B_k) = 0$. 在前一恒等式取 $n \rightarrow \infty$ 极限即得结论.

分布函数本质特征证明 I

$F(x)$ 在 $\pm\infty$ 处极限

$$F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1.$$

证明

由概率测度的下连续性,

$$\begin{aligned} F(-\infty) &= \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(\{\omega \mid X(\omega) \leq -n\}) \\ &= P\left(\bigcap_{n=1}^{\infty} \{\omega \mid X(\omega) \leq -n\}\right) = P(\emptyset) = 0. \end{aligned}$$

由概率测度的上连续性,

$$\begin{aligned} F(+\infty) &= \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(\{\omega \mid X(\omega) \leq n\}) \\ &= P\left(\bigcup_{n=1}^{\infty} \{\omega \mid X(\omega) \leq n\}\right) = P(\Omega) = 1. \end{aligned}$$

分布函数本质特征证明 II

 $F(x)$ 右连续性

$$\lim_{t \rightarrow x+} F(t) = F(x).$$

证明

由概率测度的下连续性,

$$\begin{aligned} F(x) &= P(\{\omega \mid X(\omega) \leq x\}) \\ &= P\left(\bigcap_{n=1}^{\infty} \{\omega \mid X(\omega) \leq x + \frac{1}{n}\}\right) \\ &= \lim_{n \rightarrow \infty} F\left(x + \frac{1}{n}\right) = \lim_{t \rightarrow x+} F(t). \end{aligned}$$

区间概率

问

怎样利用分布函数计算概率

$$P(a < X \leq b), \quad a < b?$$

分析

$$\begin{aligned} \{a < X \leq b\} &= \{X \leq b\} - \{X \leq a\} \\ \Rightarrow P(\{a < X \leq b\}) &= P(X \leq b) - P(X \leq a) = F(b) - F(a). \end{aligned}$$

单点概率

问

怎样计算概率 $P(X = c)$ (c 为常数)?

分析

故

$$\begin{aligned} P(X = c) &= P\left(\bigcap_{n=1}^{\infty} \{c - \frac{1}{n} < X \leq c\}\right) = \lim_{n \rightarrow \infty} P(c - 1/n < X \leq c) \\ &= \lim_{n \rightarrow \infty} [F(c) - F(c - 1/n)] \\ &= F(c) - F(c - 0). \end{aligned}$$

注

若 $F(x)$ 在 $x = c$ 处连续, 则 $P(X = c) = 0$. 否则 $P(X = c) > 0$.

单点分布

定义

对常数 c , 如果随机变量 X 的频率函数为

$$P(X = c) = 1,$$

则称随机变量 X 服从单点分布.

注

- 严格说单点分布不具有“随机性”, 视为随机变量是为了理论上的需要.
- 单点分布有时也称为“退化分布”.

定义

某事件发生的概率为 1, 则称该事件“几乎处处” (a.e.) 或“几乎必然” (a.s.) 发生. 如

$$P(X = c) = 1 \quad \Rightarrow \quad X \stackrel{\text{a.e.}}{=} c, \text{ or } X = c, \text{ a.e.}$$

$$P(X = Y) = 1 \quad \Rightarrow \quad X \stackrel{\text{a.e.}}{=} Y, \text{ or } X = Y, \text{ a.e.}$$

两点分布 (伯努利随机变量)

定义

如果随机变量 X 的频率函数为

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

则称随机变量 X 服从**0-1 两点分布** 或 **Bernoulli 分布**, 其中 $0 < p < 1$ 为常数.

注

若一个试验只产生两个结果, 则可以用服从 Bernoulli 分布的随机变量来描述.

例子

- 一门课的考试是“及格”还是“不及格”.
- 刚出生的新生儿是“男”还是“女”.
- 产吕检验的结果是“合格”还是“不合格”.
- 射击结果是“击中目标”还是“没击中目标”.

例: 收入分布

我国 2012 年家庭人均收入 R (千元) 分布如下:

R	1	2	4.5	9	15.9	25.8	34.3
收入低于 R 的家庭比例	0.05	0.1	0.25	0.5	0.75	0.9	0.95

若规定家庭人均收入 2 千元为贫困线, 又令

$$X(\omega) = \begin{cases} 1, & R(\omega) \leq 2, \\ 0, & R(\omega) > 2. \end{cases}$$

则 X 服从参数为 0.15 的两点分布.

注

X 也可以看作某家庭是否是贫困家庭的“示性函数” (indicator function):

$$X(\omega) = \mathbb{1}_{\{R(\omega) \text{ 在贫困以下}\}}(\omega).$$

伯努利试验与二项分布

伯努利试验

只产生两个结果 A, \bar{A} 的试验.

伯努利试验产生两点分布.

n 重伯努利实验

将伯努利试验独立并重复进行 n 次的试验.

例

- 某战士用步枪对目标进行射击, 记 A 为击中目标的事件. 每次射击就是一个伯努利试验. 如果对目标进行 n 次射击, 则是一个 n 重伯努利试验.
- 从一批产品中随机抽取一个产品进行检验, 记 A 为产品合格的事件, 则每检验一个产品就是一个伯努利试验.

伯努利试验分析

在伯努利试验中, 令

$$P(A) = p, \quad P(\bar{A}) = 1 - p$$

注

重复 每次试验中概率 $P(A)$ 保持不变

独立 各次试验的结果互不影响.

记 A_i 为第 i 次试验成功的事件, $i = 1, 2, \dots, n$. 由独立性, 有

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}), \quad 1 \leq i_1 < i_2 < \cdots < i_k \leq n.$$

定义

令 X 为 n 重伯努利试验中事件 A 发生的次数, 则 X 是一个离散型随机变量, 称为**二项分布**.

二项分布的频率函数

记 X 为 n 重伯努利试验中事件 A 发生的次数.

X 的取值为 $0, 1, 2, \dots, n$.

二项分布的频率函数

$\{X = k\}$ 发生意味着 A 发生 k 次, \bar{A} 发生 $(n - k)$ 次, 故由有限可加性

$$\begin{aligned} P(X = k) &= P\left(\bigcup_{1 \leq i_1 < \dots < i_k \leq n} A_{i_1} \cdots A_{i_k} \bar{A}_{j_1} \cdots \bar{A}_{j_{n-k}}\right) \\ &= \sum_{1 \leq i_1 < \dots < i_k \leq n} P\left(A_{i_1} \cdots A_{i_k} \bar{A}_{j_1} \cdots \bar{A}_{j_{n-k}}\right) \\ &= \sum_{1 \leq i_1 < \dots < i_k \leq n} p^k (1 - p)^{n-k} \\ &= \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n. \end{aligned}$$

定义

若随机变量 X 的频率函数为

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n, \quad q = 1 - p,$$

则称 X 服从参数为 (n, p) 的**二项分布**, 记为 $X \sim \text{Bin}(n, p)$. 特别地, 当 $n = 1$ 时, $\text{Bin}(1, p)$ 就是两点分布, 即

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

注

由牛顿二项式定理, 易见

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = [p + (1 - p)]^n = 1.$$

最可能出现次数

问题

令

$$b(k; n, p) := \binom{n}{k} p^k (1-p)^{n-k}.$$

考察当 n, p 固定时, $b(k; n, p)$ 随 k 的变化情况.

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{n!/k!(n-k)!}{n!/(k-1)!(n-k+1)!} \cdot \frac{p}{q} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p - k}{q}.$$

- 当 $k < (n+1)p$ 时, $b(k; n, p)$ 随 k 增加而增加.
- 当 $k = (n+1)p$ 时, $b(k; n, p) = b(k-1; n, p)$.

定义

$m = [(n+1)p]$ ($[x]$ 表示不超过 x 的最大整数) 为**最可能出现次数**, $b(m; n, p)$ 为**中心项**.

例题

一大批电子元件有 10% 已经损坏. 若从这批元件中随机选取 20 只来组成一个线路, 问此线路能正常工作的概率是多少?

解

因为元件的数量很大, 所以取 20 只元件可看作是有放回抽样, 可以用 n 重伯努利试验刻画.

记 X 为 20 只元件中合格品的数量, 则 $X \sim \text{Bin}(20, 0.9)$, 因此

$$P(\text{线路正常}) \approx P(X = 20) = 0.9^{20} \times 0.1^0 = 0.9^{20} \approx 0.1216.$$

例题

已知 100 个产品中有 5 个次品, 现从中有放回地取 3 次, 每次任取 1 个. 求所取的 3 个中恰有 2 个次品的概率.

解

有放回地抽样是条件相同且独立的试验, 它是伯努利试验. 设 X 为所取的 3 个中次品数, 则由题意 $X \sim \text{Bin}(3, 0.05)$, 因此所求概率为

$$P(X = 2) = \binom{3}{2} (0.05)^2 0.95 = 0.007125.$$

注

若本例改为“无放回”, 则此试验不是伯努利试验, 只能用更一般的古典概型求解:

$$P(X = 2) = \frac{\binom{95}{1} \binom{5}{2}}{\binom{100}{3}} \approx 0.00618.$$

例: 加州抢动案 (60 年代)

据报道, 加州某地发生一起抢劫案. 目击嫌疑人有两个: 一个男的理平头黑人, 一个女的黑发梳马尾型. 不久抓到一对具有上述特征的夫妇 (情侣), 能否判断他们有罪?

有罪的论点

数学家通过计算机模拟, 得出一对夫妇具有一述特征的概率为 $p = 8.3 \times 10^{-8}$. 这是一个小概率事件. 陪审团在无其它证据的情况下, 裁决他们有罪.

无罪的论点

加州高院推翻了该裁决. 高院认为犯罪的认定应当具有唯一性. 用上列算式: 设 X 为具有上述特征的夫妇数, 则 $X \sim \text{Bin}(n, p)$. 要计算

$$P(X > 1 | X \geq 1) = \frac{P(X > 1)}{P(X \geq 1)} = \frac{P(X \geq 1) - P(X = 1)}{P(X \geq 1)} = \frac{1 - (1 - p)^n - np(1 - p)^{n-1}}{1 - (1 - p)^n}$$

n 的估计为 8.3×10^6 , 而上式约为 0.2966, 不算小.

保险业的应用

保险业是最早应用概率论的行业之一. 保险公司为了估计企业的利润, 需要计算各种各样的概率.

例题

若某一年某类保险者里面每个人死亡的概率等于 0.005. 现有 10000 个人参加这类人寿保险. 试求在未来一年中在这些保险者里面:

- 1 有 40 个人死亡的概率
- 2 死亡人数不超过 70 个的概率.

解

记 X 为未来一年中在这些人中死亡的人数, 则 $X \sim \text{Bin}(10000, 0.005)$.

- 1 $P(X = 40) = \binom{10000}{40} \times 0.005^{40} \times 0.995^{9960} \approx 0.0214$
- 2 $P(X \leq 70) = \sum_{k=0}^{70} \binom{10000}{k} \times 0.005^k \times 0.995^{10000-k} \approx 0.997.$

例题

设有 80 台同类型的设备, 各台工作是相互独立的, 且发生故障的概率是 0.01, 且一台设备的故障只能由一个人处理. 考虑两种配备维修工作的方法: 第一种, 由 4 个人维护, 每人负责 20 台; 第二种, 由 3 个人共同维护 80 台. 试比较这两种方法在设备发生故障时不能及时维修概率的大小.

解

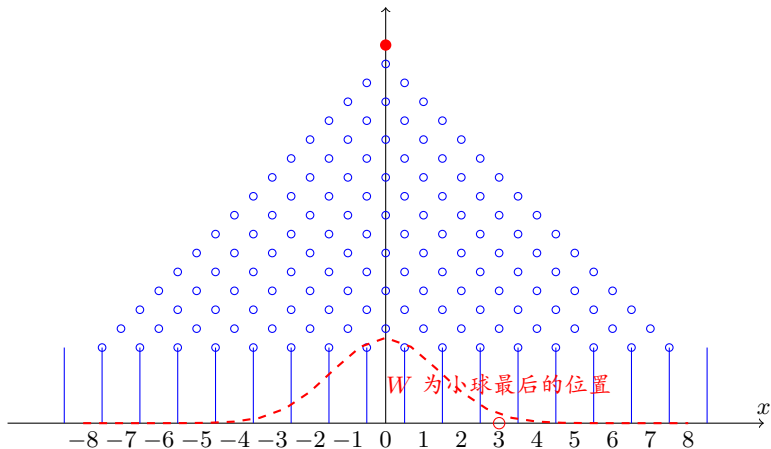
第一种情况 记 X_i 为同一时刻第 i 人维护的 20 台设备中同时发生故障的台数, 则 $X_i \sim \text{Bin}(20, 0.01)$ 且相互独立. 于是 80 台设备中发生故障而不能及时维修的概率为

$$\begin{aligned} P\left(\bigcup_{i=1}^4 \{X_i \geq 2\}\right) &\geq P(X_1 \geq 2) \\ &= 1 - \binom{20}{0} 0.01^0 \times 0.99^{20} - \binom{20}{1} 0.01^1 \times 0.99^{19} \approx 0.0169. \end{aligned}$$

第二种情况 记 X 为 80 台设备中发生故障的台数, 于是 $X \sim \text{Bin}(80, 0.01)$. 则 80 台设备中发生故障而不能及时维修的概率为

$$P(X \geq 4) = 1 - \sum_{k=0}^3 \binom{80}{k} 0.01^k \times 0.99^{80-k} \approx 0.0087.$$

高尔顿钉板试验



共 16 层钉板, X 为小球向右的次数, 则

$$X \sim \text{Bin}(16, 0.5), \quad W = (X - (16 - X))/2 = X - 8.$$

几何分布

定义

考虑不断重复进行成功率为 p 的伯努利试验, 记 X 为第一次成功时所做的试验次数. 称 X 为**几何分布**, 记为 $X \sim \text{Geo}(p)$.

频率函数

$X = k$ 表明第 1 至 $(k-1)$ 次试验失败, 第 k 次试验成功. 由试验的独立性,

$$p(k) = \mathbf{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, \dots$$

归一性

由 $\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$ (几何级数) 知,

$$\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} (1-p)^{k-1}p = \frac{1}{1-(1-p)} \cdot p = 1.$$

负二项分布

定义

考虑不断重复进行成功率为 p 的伯努利试验. 记 X 为第 r 次成功时试验的次数, 称 X 为**负二项分布**.

频率函数

$X = k$ 表示前 k 次试验中, 第 r 次成功发生在第 k 次试验, 其余 $(r-1)$ 次成功发生在前 $(k-1)$ 次试验中. 由独立性, 所有这样的结果序列发生的概率为 $p^r(1-p)^{k-r}$. 由组合计数原理, 这样的样本点一共有 $\binom{k-1}{r-1}$ 个, 因此

$$p(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

注

当 $r = 1$ 时, 负二项分布就是几何分布.

负二项分布归一化

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = 1.$$

负二项公式

$$(1+x)^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} x^k, \quad \binom{-r}{k} = \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!}.$$

证明

$$1 = p^r \frac{1}{[1-(1-p)]^r} = p^r \sum_{m=0}^{\infty} \binom{-r}{m} (-1)^m (1-p)^m = p^r \sum_{k=r}^{\infty} \binom{-r}{k-r} (-1)^{k-r} (1-p)^{k-r}$$

其中

$$(-1)^{k-r} \binom{-r}{k-r} = (-1)^{k-r} \frac{(-r)(-r-1)\cdots(-r-(k-r)+1)}{(k-r)!} = \binom{k-1}{k-r}$$

超几何分布

定义

假设盒中有 n 个球, 其中 r 个黑球, $(n-r)$ 个白球. 从盒中无重复地抽取 m 个球. 令 X 为抽到黑球的个数. 称 X 为参数 r, n, m 的超几何分布随机变量.

频率函数

所有的抽取都是等可能的, 适用于不考虑次序的古典概型, 因此

$$P(X = k) = \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}, \quad k = 0, 1, \dots, n.$$

归一性

Vandermonde 恒等式:

$$\binom{n}{m} = \sum_{k=0}^m \binom{r}{k} \binom{n-r}{m-k}.$$

超几何分布与二项分布

超几何分布的极限

假设 $r, n \rightarrow \infty, r/n \rightarrow p$, 则

$$\frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}} \rightarrow \binom{m}{k} p^k (1-p)^{m-k}.$$

证明

注意到 m, k 为常数. 因此,

$$\begin{aligned} \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}} &= \frac{\frac{r(r-1)\cdots(r-k+1)}{k!} \cdot \frac{(n-r)(n-r-1)\cdots(n-r-(m-k)+1)}{(m-k)!}}{\frac{n(n-1)\cdots(n-m+1)}{m!}} \\ &= \frac{m!}{k!(m-k)!} \frac{r \cdots (r-k+1) \cdot (n-r) \cdots (n-r-(m-k)+1)}{n(n-1)\cdots(n-m+1)} \\ &\sim \binom{m}{k} \frac{r^k (n-r)^{m-k}}{n^m} \rightarrow \binom{m}{k} p^k (1-p)^{m-k}. \end{aligned}$$

例题

假定在美国 18 家主要计算机公司中, 有 12 家位于加州的硅谷. 如果从这 18 家中随机抽取 3 家, 则至少有 1 家位于硅谷的概率是多少?

解

令 X 为位于硅谷的公司的数量, 则 X 服从参数为 12, 18, 3 的超几何分布. 于是, 所求为

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{12}{0} \binom{6}{3}}{\binom{18}{3}} = 0.9755.$$

泊松流与泊松分布

泊松流

随着时间的推移, 在时间轴上源源不断出现的“随机”粒子流称为泊松流.

典型的泊松流

- 随机服务系统
- 稀疏现象的发生
- 物理学中的现象

构造

可以想像把时间轴按 Δt 分成许多小段, 每一段时间内独立地以某概率 $p = p(\Delta t)$ 出现粒子. 考虑 $\Delta t \rightarrow 0$ 的极限.

泊松分布

定义

设随机变量 X 的取值为 $0, 1, 2, \dots$, 取值概率为

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

其中 $\lambda > 0$ 为参数. 称 X 为服从参数为 λ 的**泊松分布**, 记为 $X \sim \pi(\lambda)$ 或 $X \sim P(\lambda)$.

注

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

泊松分布的性质

- $P(X = k) > 0, k = 0, 1, 2, \dots$
- $\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$

泊松分布与泊松流的关系

泊松流的性质

在泊松流中, 记时间间隔 $(0, t]$ 中出现的粒子数为 X , 则 $X \sim \pi(\lambda t)$, 即

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

其中参数 $\lambda > 0$ 称为**泊松强度**.

泊松定理

定理

设 $\lambda > 0$ 为一常数, n 为正整数, $\lim_{n \rightarrow \infty} np_n = \lambda$. 则对任一非负整数 k , 有

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

应用

当 n 很大时, p 很小时, 有如下估计

$$\binom{n}{k} p^k (1 - p)^{n-k} \approx \frac{(np)^k e^{-np}}{k!}.$$

例题

已知某疾病发病率为 0.001, 某单位共有 5000 人, 问该单位患有这种疾病的人数不超过 5 人的概率.

解

设该单位患有这种疾病的人数为 X , 则有 $X \sim \text{Bin}(5000, 0.001)$. 利用泊松定理近似 $\lambda = np = 5$, 计算得

$$P(X \leq 5) = \sum_{k=0}^5 \binom{5000}{k} 0.001^k 0.999^{5000-k} \approx \sum_{k=0}^5 \frac{5^k}{k!} e^{-5} = 0.616.$$

泊松分布的应用

- 在应用中, 诸如服务系统中对服务的呼叫数, 产品的缺陷 (如布匹上的疵点、玻璃内的气泡等) 数, 一定时期内出现的稀有事件 (如意外事故、自然灾害等) 个数, 放射性发射出的离子数等等, 都以泊松分布为其概率模型.
- 上述的模型都可以看作 n 大 p 小的二项分布, 由泊松分布, 可以近似为服从 $\lambda = np$ 的泊松分布.
- 泊松分布广泛用于社会生活的许多方面, 在运筹学、管理科学中占有突出的地位.

例题

如果某房地产公司每天售出 1.6 套住宅, 且住宅销售量服从泊松分布. 求以下事件的概率

- ① 一天内恰好售出 4 套住宅;
- ② 一天内没有售出住宅;
- ③ 一天售出 5 套以上住宅;
- ④ 每天至少售出 10 套住宅;
- ⑤ 两天内恰好售出 4 套住宅.

解

最后一问: 根据泊松过程的构造, 每两天售出的住宅仍服从泊松分布, 且参数为 $1.6 \times 2 = 3.2$. 因此所求概率为

$$P(\pi(3.2) = 4) = e^{-3.2} \frac{(3.2)^4}{4!} = 0.1781.$$

例题

某人骑了自行车从学校到火车站, 一路上要经过 3 个独立的交通灯. 设各灯工作独立, 且设各灯为红灯的概率为 p , $p \in (0, 1)$. 以 Y 表示一路上遇到红灯的次数.

- 1 求 Y 的分布律;
- 2 求恰好遇到两次红灯的概率.

分析

这是三重伯努利试验, $Y \sim \text{Bin}(3, p)$.

解

- 1 $P(Y = k) = \binom{3}{k} p^k (1-p)^{3-k}$, $k = 0, 1, 2, 3$.
- 2 $P(Y = 2) = 3p^2(1-p)$.

● P46: 1, 15, 31

● 补充题

- (1) 设随机变量 X 的频率函数为 $P(X = x) = c\left(\frac{2}{3}\right)^x$, $x = 1, 2, 3$. 求 c 的值.
- (2) 设随机变量 X 服从参数为 λ 的泊松分布, 求 k 使得 $P(X = k)$ 达到最大.
- (3) 设在 15 只同类型零件中有 2 只为次品. 在其中取 3 次, 每次任取一只, 作不做回抽样, 以 X 表示出取出的次品个数. 求
 - ① X 的分布律;
 - ② X 的分布函数并作图;
 - ③ $P(X \leq 1/2)$, $P(1 < X \leq 3/2)$, $P(1 \leq X \leq 3/2)$; $P(1 < X < 2)$.
- (4) 有 2500 名同一年龄和同社会阶层的人参加了保险公司和人寿保险. 在一年中每个人死亡的概率为 0.002, 每个参加保险的人在 1 月 1 日须交 12 元保险费, 而在死亡时家属可从保险公司领取 2000 元赔偿金. 求
 - (1) 保险公司亏本的概率
 - (2) 保险公司获利分别不少于 10000 元、20000 元的概率.

1 离散随机变量 (discrete r.v.)

2 连续随机变量 (continuous r.v.)

- 定义
- 密度函数
- 几种重要的连续型随机变量 I: 正态分布
- 几种重要的连续型随机变量 II: 均匀分布
- 几种重要的连续型随机变量 III: 指数分布
- 几种重要的连续型随机变量 IV: Gamma 分布与 Beta 分布
- 总结
- 作业

3 随机变量的函数

分布函数回顾

分布函数

随机变量 X 的 (累积) 分布函数

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

分布函数的本质特征

- ① $F(x)$ 单调不减
- ② $F(x) \in [0, 1]$ 且 $F(-\infty) = 0, F(\infty) = 1$.
- ③ $F(x)$ 右连续.

分布函数的应用

区间概率

$$P(X \in (a, b]) = F(b) - F(a).$$

点概率

$$P(X = c) = F(c) - F(c - 0).$$

注

因为分布函数是一个普通的函数, 所以可以应用微积分工具来研究随机现象.

连续型随机变量 (Continuous random variables)

定义

若随机变量 X 的分布函数能够表为

$$F(x) = \int_{-\infty}^x f(t) dt, \quad -\infty < x < \infty, \quad f(x) = F'(x)$$

其中 $f(t) \geq 0$, 则称 X 为**连续型随机变量**. 非负可积函数 $f(t)$ 称为**概率密度函数 (probability density function, pdf)**.

微积分基本定理

假设 $F(x)$ 只有有限个点不可导, 且 $F'(x)$ 可积, 则牛顿-莱布尼茨公式成立:

$$F(b) - F(a) = \int_a^b F'(x) dx.$$

变上限积分

若 $F(x) = \int_a^x f(t) dt$ 且 $f(x)$ 在 $x = c$ 处连续, 则 $F'(c) = f(c)$.

例题

设随机变量 X 的分布函数为

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

求 X 的密度函数.

解

在 $x = 0$ 处 $F(x)$ 不可导, 除此之外我们有

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

因此 X 的密度函数为

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

(补充几个不可导处的值不影响积分值!)

随机变量的类型

连续型随机变量 X 的分布函数可表为

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

注

- ① 连续型随机变量在区间上的取值是“连续的”.
- ② 连续型随机变量的分布函数是连续函数, 且能表示成上述形式.

随机变量的类型

- 离散型随机变量
- 非离散型随机变量
 - 连续型随机变量
 - 奇异型随机变量 (singular random variables)

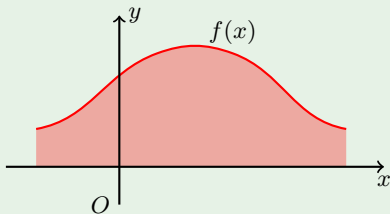
密度函数的性质 I

本质特征

- ① $f(t) \geq 0$;
- ② $\int_{-\infty}^{\infty} f(t) dt = 1$.

几何意义

图形在 x 轴上方, 上方图形面积为 1



密度函数的性质 II

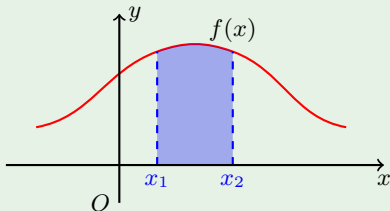
性质

对任意 $x_1 < x_2$, 有

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

几何意义

$P(x_1 < X \leq x_2)$ 等于曲边梯形面积.



密度函数的性质 III

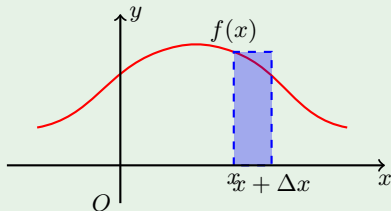
性质

在 $f(x)$ 的连续点处有

$$f(x) = F'(x).$$

几何意义

$P(x < X \leq x + \Delta x)$ 近似于小矩形面积.



例题

设连续型随机变量 X 的密度函数为

$$f(x) = \begin{cases} \frac{k}{x^2}, & x > 100, \\ 0, & x \leq 100. \end{cases}$$

- ① 确定常数 k , 并求 X 的分布函数 $F(x)$.
- ② 计算概率 $P(50 < x \leq 10000)$.

- ① 概率密度函数积分为 1, 因此

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_{100}^{\infty} \frac{k}{x^2} dx = \frac{k}{100} \Rightarrow k = 100.$$

对密度函数积分得到分布函数:

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} \int_{100}^x \frac{100}{t^2} dt, & x > 100, \\ 0, & x \leq 100 \end{cases} = \begin{cases} 1 - \frac{100}{x}, & x > 100, \\ 0, & x \leq 100. \end{cases}$$

② $P(50 < X \leq 1000) = \int_{100}^{1000} \frac{100}{x^2} dx = \frac{9}{10} (= F(1000)).$

单点概率

性质

设 X 为连续型随机变量, c 为任意常数, 则 $P(X = c) = 0$.

分析

注意到连续型随机变量的分布函数是连续函数, 因此

$$P(X = c) = F(c) - F(c - 0) = 0.$$

注

对于连续型随机变量 X , 有

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b).$$

密度函数的意义

问

设 X 为连续型随机变量, c 为任意常数, 那么 $\{X = c\}$ 是否是不可能事件?

答

否. $\{X = c\} \neq \emptyset$, 它只是发生概率为 0 的事件. 在概率论中 “不可能事件” 指的是空集.

密度函数的意义

- 密度函数 $f(x)$ 在某点 c 处的高度, 并不反映 X 取值的概率. 但是, 这个高度越大, 则 X 取 c 附近的值的概率就越大.
- 也可以说, 在某点密度曲线的高度反映了概率集中在该点附近的程度.

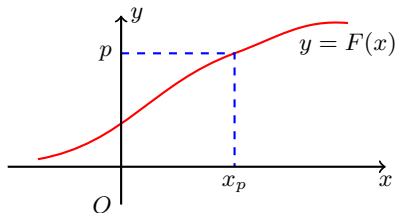
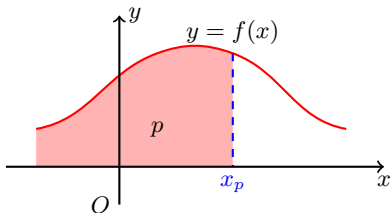
p -分位数

定义

设 $X \sim f(x) dx$. 若对 $p \in (0, 1)$, 存在常数 x_p 满足

$$P(X \leq x_p) = \int_{-\infty}^{x_p} f(x) dx = p, \text{ or } F(x_p) = p,$$

则称 x_p 为密度函数 $f(x)$ 的 **p -分位数 (quantile)**. 特别地, 取 $p = 1/2, 1/4, 3/4$ 时, 称 x_p 为**中位数 (median)**, **下、上四分之一分位数**,



正态分布

定义

如果随机变量 X 的密度函数为

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

其中 $-\infty < \mu < \infty, \sigma > 0$, 则称 X 服从参数为 (μ, σ^2) 的**正态分布**, 记为 $X \sim \mathcal{N}(\mu, \sigma^2)$.

归一性的验证

作 $t = \frac{x - \mu}{\sigma}$ 的换元:

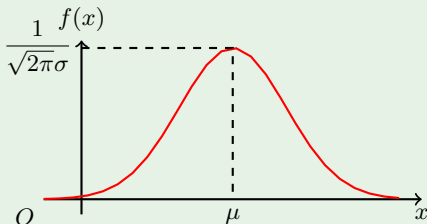
$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \left[\int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{u^2+v^2}{2}} dudv \right]^{1/2} \\ &= \left[\frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}} r dr \right]^{1/2} = 1. \end{aligned}$$

正态分布密度函数的性质

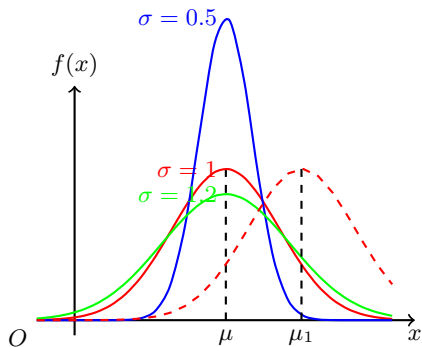
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

性质

- $f(\mu+x) = f(\mu-x)$, 即 $y = f(x)$ 关于 $x = \mu$ 对称.
- 当 $x < \mu$ 时, $f'(x) > 0$, $f(x) \uparrow$; 当 $x > \mu$ 时, $f'(x) < 0$, $f(x) \downarrow$. 因此, $f(x)$ 在 $x = \mu$ 处取最大值 $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$.
- $\lim_{x \rightarrow \pm\infty} f(x) = 0$, 即 $y = f(x)$ 以 x 轴为渐近线.



参数与图像的关系



μ : 位置参数, σ : 刻度参数

- μ 由小变大: 图形向右平移, 形状不变
- μ 由大变小: 图形向左平移, 形状不变
- σ 由小变大: 图形变平坦
- σ 由大变小: 图形变尖锐

正态分布的应用

实际背景

自然界中许多指标都服从或近似服从正态分布

例

- 一个班的某门课程考试成绩
- 海浪的高度
- 一个地区的日耗电量
- 各种测量的误差
- 炮弹弹着点
- ◀ 高尔顿钉板试验

正态分布的发现

为什么叫正态分布?

正态分布密度呈现“中间高，两头低”的形态，它描述了自然界大量存在的随机现象。所以正态分布是自然界的一种“正常状态 (normal)”的分布。

正态分布的发现

正态分布是德国数学家高斯 (Gauss) 在研究误差理论时得到的，故正态分布也称为**高斯分布**。

服从正态分布的指标特点

一般地说，若影响某一数量指标的随机因素很多，而每个因素所起的作用都不太大时，这个指标服从正态分布。

标准正态分布

定义

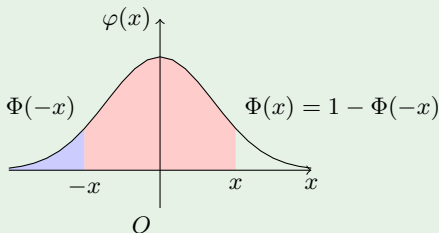
当 $\mu = 0, \sigma = 1$ 时, 称为**标准正态分布**, 记为

$$X \sim \mathcal{N}(0, 1).$$

其概率密度函数和分布函数分别为

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

$\Phi(x)$ 的值通常可查表得到.



$\mathcal{N}(\mu, \sigma^2)$ 与 $\mathcal{N}(0, 1)$ 的关系

引理

若 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则 $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

证明

一个随机变量的分布唯一地由它的分布函数决定. 我们有如下关于 Z 的分布函数的计算:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P\left(\frac{X - \mu}{\sigma} \leq z\right) \\ &= P(X \leq \sigma z + \mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\sigma z + \mu} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du \\ &= \Phi(z). \end{aligned}$$

例题

从某地乘车往火车站有两条路可选. 第一条路线穿过市区, 路程较短但交通拥挤, 所需时间 $X \sim \mathcal{N}(50, 100)$; 第二条路径走环线, 路程较远但意外阻塞较少, 所需时间 $X \sim \mathcal{N}(60, 16)$.

- ① 若有 70 分钟时间可用, 问应走哪条路线?
- ② 若只有 65 分钟时间可用, 又应走哪条路线?

分析

我们应该保证准时到达的概率尽量大.

解

	路线 I: $\mathcal{N}(50, 100)$	路线 II: $\mathcal{N}(60, 16)$
70 分钟	$\Phi\left(\frac{70-50}{10}\right) = \Phi(2)$	$\Phi\left(\frac{70-60}{4}\right) = \Phi(2.5)$
65 分钟	$\Phi\left(\frac{65-50}{10}\right) = \Phi(1.5)$	$\Phi\left(\frac{65-60}{4}\right) = \Phi(1.25)$

例题

设 $X \sim \mathcal{N}(\mu, \sigma^2)$, 求下列概率值:

$$P(|X - \mu| \leq \sigma), \quad P(|X - \mu| \leq 2\sigma), \quad P(|X - \mu| \leq 3\sigma).$$

分析

由引理知, $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$

解

$$\begin{aligned} P(|X - \mu| \leq \sigma) &= P(|\mathcal{N}(0, 1)| \leq 1) = \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 = 0.6826 \end{aligned}$$

$$\begin{aligned} P(|X - \mu| \leq 2\sigma) &= P(|\mathcal{N}(0, 1)| \leq 2) = \Phi(2) - \Phi(-2) \\ &= 2\Phi(2) - 1 = 0.9544 \end{aligned}$$

$$\begin{aligned} P(|X - \mu| \leq 3\sigma) &= P(|\mathcal{N}(0, 1)| \leq 3) = \Phi(3) - \Phi(-3) \\ &= 2\Phi(3) - 1 = 0.9974 \end{aligned}$$

3 σ -原则

正态分布的值几乎都落在 $(\mu - 3\sigma, \mu + 3\sigma)$ 内.

数据校验与数据“稳健性”

例题

在某体育比赛中, 设裁判给运动员的表演打的分数服从 $\mathcal{N}(\mu, 0.2^2)$ 的正态分布, 其中 μ 为运动员真实的成绩. 已知 4 位裁判的分数分别为 6.8, 6.7, 7.1, 8.6. 请问这些分数是否公正?

分析

我们可以利用 3σ -原则进行判断. 这里参数 μ 未知, 我们考虑用样本均值代替.

解

假设评分是公正的, 那么 $\mu \approx \hat{\mu} = (6.8 + 6.7 + 7.1 + 8.6)/4 = 7.33$. 然而 $|8.6 - \hat{\mu}| = 1.27 > 3\sigma = 0.6$, 根据 3σ -原则这是几乎不可能发生的, 因此认为假设不成立. 所以评分是不公正的.

注

在体育比赛中常常会去掉一个最高分和最低分, 取余下的平均分作为最终的得分.

例题

公共汽车车门的高度是按男子与车门顶头碰头机会在 0.01 以下来设计的. 设男子身高 $X \sim \mathcal{N}(170, 6^2)$ (单位: cm). 问车门高度如何确定?

分析

设车门高度为 h , 则 $P(X \geq h) \leq 0.01$, 也即 $P(X < h) \geq 0.99$. 我们用标准正态分布来求 h .

解

因为 $X \sim \mathcal{N}(170, 6^2)$,

$$P(X < h) = P\left(\frac{X - 170}{6} < \frac{h - 170}{6}\right) = \Phi\left(\frac{h - 170}{6}\right).$$

查表得 $\Phi(2.33) = 0.9901 > 0.9$, 因而 $\frac{h - 170}{6} > 2.33 \Rightarrow h \geq 170 + 13.89 \approx 184$.

均匀分布

定义

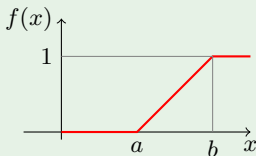
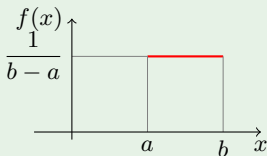
如果随机变量 X 的密度函数为

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

则称 X 服从区间 (a, b) 上的**均匀分布 (Uniform distribution)**, 记为 $X \sim U(a, b)$.

注

- 易见 $f(x) \geq 0$, $\int_a^b \frac{1}{b-a} dx = 1$. 因此 $f(x)$ 满足密度函数的本质特征.
- 均匀分布的密度函数与分布函数图像



均匀分布与几何概型

均匀分布的“等可能性”

对任意 $(c, c+L) \subset (a, b)$ 有

$$P(c < X < c+L) = \int_c^{c+L} \frac{dx}{b-a} = \frac{L}{b-a}$$

即 X 落在 $(c, c+L)$ 的概率只与区间长度有关, 而与位置无关. 这反映了某种“等可能性”, 即随机变量 X 在区间 (a, b) 上“等可能取值”.

注

若 $X \sim U(a, b)$ 为连续型随机变量, c 为常数, 则 $P(X = c) = 0$.

例题

将长度为 2ℓ 的木棒任意截为两段. 求这两段木棒与另一长度为 ℓ 的木棒能构成三角形的概率.

分析

三条边 a, b, c 能组成三角形当且仅当 $a + b > c$, $a + c > b$, $b + c > a$.

解

设两段木棒长度为 X 与 $2\ell - X$, 其中 $X \sim U(0, 2\ell)$. 则组成三角形当且仅当

$$X + \ell > 2\ell - X, \quad 2\ell - X + \ell > X \quad \Leftrightarrow \quad \ell/2 < X < 3\ell/2.$$

$$\text{因此所求概率为 } P(U(0, 2\ell) \in (\ell/2, 3\ell/2)) = \int_{\ell/2}^{3\ell/2} \frac{1}{2\ell} dx = \frac{1}{2}.$$

例题

设随机变量 X 在 $(2, 5)$ 上服从均匀分布. 现对 X 进行 3 次独立观测. 求至少有两次观测值大于 3 的概率.

解

因为随机变量 $X \sim U(2, 5)$, 所以

$$P(X > 3) = \int_3^5 \frac{1}{3} dx = \frac{2}{3}.$$

于是 3 次观测中每次观测到大于 3 的概率为 $2/3$. 易见此为 3 重伯努利试验. 令 Y 为观测值大于 3 的次数, 则 $Y \sim \text{Bin}(3, 2/3)$. 于是所求概率为

$$P(Y \geq 2) = \binom{3}{2} \left(\frac{2}{3}\right)^2 \frac{1}{3} + \binom{3}{0} \left(\frac{2}{3}\right)^3 = \frac{20}{27}.$$

指数分布

定义

如果随机变量 X 的密度函数为

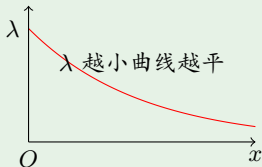
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

则称 X 服从参数为 $\lambda > 0$ 的**指数分布**, 记为 $X \sim \text{Exp}(\lambda)$.

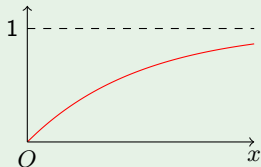
注

- $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1$, 所以 $f(x)$ 确实是密度函数.
- 密度函数与分布函数

$$f(x) = \lambda e^{-\lambda x}$$



$$F(x) = 1 - e^{-\lambda x}$$



指数分布与泊松分布

泊松粒子流回顾

在 $(0, t]$ 中出现的质点数服从参数为 λt 的泊松分布. λ 称为**泊松强度**.

性质

第一个质点出现的时间 T 服从指数分布 $\text{Exp}(\lambda)$.

证明

$$P(T > t) = P(\pi(\lambda t) = 0) = e^{-\lambda t},$$

故 T 的分布函数为 $F(t) = 1 - e^{-\lambda t}$. 这也是 $\text{Exp}(\lambda)$ 的分布函数.

参数 λ 的意义

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}(x).$$

λ 的意义

参数 λ 称为**失效率**, λ^{-1} 表示**平均寿命**.

指数分布的实际背景

指数分布通常用来描述“寿命”的分布. 如:

- 电子元件的寿命,
- 电话的通话时间,
- 机器的修理时间,
- 营业员为顾客提供的服务时间.

指数分布广泛用于可靠性理论和排队论.

指数分布的“无记忆性”

无记忆性

设 $X \sim \text{Exp}(\lambda)$, 则

$$P(X > s + t \mid X > s) = P(X > t).$$

即: 已知寿命长于 s 的情况下, 再活时间 t 的概率与 s 无关.

证明

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t). \end{aligned}$$

注

指数分布是唯一具有“无记忆性”的连续型随机变量. 几何分布是唯一具有“无记忆性”的离散型随机变量.

例题

假定自动取款机对每位顾客的服务时间 (单位: 分钟) 服从 $\lambda = 1/3$ 的指数分布. 如果有一顾客恰好在你前面走到空闲的取款机, 求:

- 1 你至少等候 3 分钟的概率.
- 2 你等候时间在 3 分钟至 6 分钟之间的概率.

解

以 X 表示你的等待时间, $F(x)$ 为 X 的分布函数, 则

- 1 $P(X \geq 3) = 1 - F(3) = e^{-3\lambda} = e^{-1}.$
- 2 $P(X \in (3, 6)) = F(6) - F(3) = (1 - e^{-6\lambda}) - (1 - e^{-3\lambda}) = e^{-1} - e^{-2}.$

注

若你到达时取款机正在为一名顾客服务, 但你不知道这名顾客已经被服务的时间, 结果又如何?

由无记忆性, 问题的答案不变.

Gamma 分布的导出

问

已知泊松流中第 1 个粒子出现的时间服从指数分布. 那么第 r 个粒子出现的时间服从什么分布?

答

以 X 记第 r 个粒子出现的时间, 则 $\{X > t\}$ 表明 $(0, t]$ 中至多有 $(r-1)$ 个粒子. 因此

$$R(t) = P(X > t) = P(\pi(\lambda t) = 0, 1, \dots, r-1) = e^{-\lambda t} \sum_{i=0}^{r-1} \frac{(\lambda t)^i}{i!}.$$

因此 X 的密度函数为

$$f(t) = -R'(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{r-1}}{(r-1)!}.$$

Gamma 分布

定义

若连续型随机变量 X 的密度函数为

$$f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

其中 $r, \lambda > 0$ 为参数, 则称 X 服从参数为 (r, λ) 的 **Γ -分布**, 记为 $X \sim \Gamma(r, \lambda)$. 特别地, $\Gamma(1, \lambda) = \text{Exp}(\lambda)$.

Gamma 函数

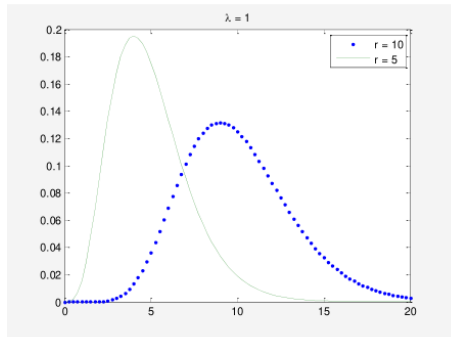
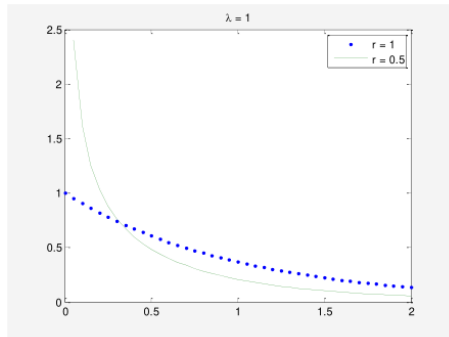
这里

$$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx, \quad r > 0.$$

特别地, 当 r 为自然数时, $\Gamma(r) = (r-1)!$.

Gamma 分布参数的意义

r : 形状参数. λ : 尺度参数.



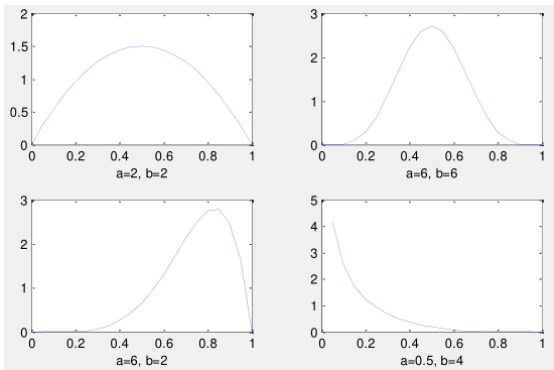
Beta 分布

若连续型随机变量 X 的密度函数为

$$f(u) = \mathbb{1}_{(0,1)}(u) \frac{1}{B(a,b)} u^{a-1} (1-u)^{b-1}$$

其中 $B(a,b) = \int_0^1 u^{a-1} (1-u)^{b-1} du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ 为 Beta 函数, 则称 X 服从参数为 (a,b) 的 **Beta 分布**, 记为 $X \sim \text{Beta}(a,b)$.

特别地, 当 $a = b = 1$ 时, Beta 分布为 $(0,1)$ 上均匀分布.



地震的概率模型

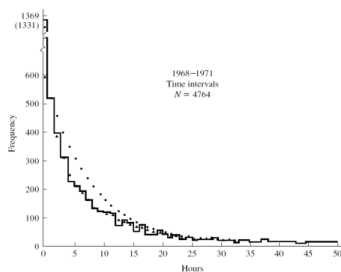


FIGURE 2.12 Fit of gamma density (triangles) and of exponential density (circles) to times between microearthquakes.

指数模型的解释

即使知道上 t 个时间单位内没有地震, 也无法预知下 s 个时间单位内发生地震的概率.

Gamma 模型的解释

对于任意一次地震, 下一次地震紧跟其后的可能性非常大, 并且这种可能性随时间单调下降.

离散型与连续型随机变量的形式统一性

- 设离散型随机变量 X 的频率函数为 $P(X = x_k) = p_k, k = 1, 2, \dots$ 我们改写为

$$p(x) = P(X = x), \quad x = x_1, x_2, \dots$$

- 设连续型随机变量的密度函数为 $f(x)$, 则有 $f(x) dx \approx P(X \in (x, x + dx))$.

对应关系

离散型	连续型
$p(x)$	$f(x) dx$
\sum	\int

例

$$\sum_x p(x) = 1 \longleftrightarrow \int_{-\infty}^{\infty} f(x) dx = 1.$$

概率函数

定义

对于随机变量 X , 它的**概率函数**是指**频率函数** (X 为离散型时) 或**密度函数** (X 为连续型时).

注

- 概率函数表述了离散型随机变量与连续型随机变量的形式统一性.
- 用概率函数来表示, 许多公式既适用于离散型也适用于连续型.

应用中如何确定随机变量的分布?

- ① 确定它是离散型或是连续型
- ② 根据随机变量的来源确定它的分布形式: 正态分布、均匀分布、指数分布、泊松分布、二项分布等等
- ③ 进一步检验某随机变量的分布, 并且给出分布参数 — “分布检验” 与 “参数估计”.

例题

一批钢材 (线材) 长度 X (cm) $\sim \mathcal{N}(\mu, \sigma^2)$.

- ① 若 $\mu = 100$, $\sigma = 2$, 求这批钢材小于 97.8 cm 的概率.
- ② 若 $\mu = 100$, 要使这批钢材的长度至少有 90% 落在区间 (97, 103) 内, 问 σ 至多取何值?

解

① $P(X < 97.8) = \Phi\left(\frac{97.8 - 100}{2}\right) = 1 - \Phi(1.1) = 1 - 0.8643 = 0.1357$.

② 令 $P(97 < X < 103) \geq 0.9$, 则

$$\begin{aligned}\Phi\left(\frac{103 - 100}{\sigma}\right) - \Phi\left(\frac{97 - 100}{\sigma}\right) &= 2\Phi\left(\frac{3}{\sigma}\right) - 1 \geq 0.9 \\ \Rightarrow \Phi\left(\frac{3}{\sigma}\right) &\geq 0.95 \Rightarrow \frac{3}{\sigma} \geq 1.645 \\ \Rightarrow \sigma &\leq 1.8237.\end{aligned}$$

例题

在区间 $(-1, 2)$ 上随机取一数 X . 试写出 X 的概率密度, 并求 $P(X > 0)$ 的值. 若在该区间上随机取 10 个数, 求这 10 个数中恰有两个数大于 0 的概率.

分析

X 服从 $U(-1, 2)$. 大于 0 的个数 $Y \sim \text{Bin}(10, p)$, $p = P(X > 0)$.

解

X 服从 $(-1, 2)$ 上的均匀分布, 所以其密度函数为

$$f(x) = \begin{cases} \frac{1}{3}, & -1 < x < 2, \\ 0, & \text{其它}. \end{cases}$$

易见 $P(X > 0) = \frac{2}{3}$. 因此 $Y \sim \text{Bin}(10, \frac{2}{3})$. 故 $P(Y = 2) = \binom{10}{2} \frac{2^2}{3^{10}}$.

● P48: 33, 40, 45, 52, 53

● 补充题

① 已知随机变量 X 的密度函数为

$$f(x) = Ae^{-|x|}, \quad -\infty < x < \infty.$$

求:

(1) A 的值;

(2) $P(0 < X < 1)$;

(3) $F(x)$.

② 设顾客在某银行的窗口等待服务的时间 X (以分钟计) 服从指数分布 $\text{Exp}(1/5)$. 某顾客在窗口等待服务, 若超过 10 分钟他就离开. 他一个月要来银行 5 次, 以 Y 表示一个月内他未等到服务而离开窗口的次数. 试写出 Y 的分布律, 并求 $P(Y \geq 1)$.

① 离散随机变量 (discrete r.v.)

② 连续随机变量 (continuous r.v.)

③ 随机变量的函数

- 引言
- 离散型随机变量函数的频率函数
- 连续型随机变量函数的分布

实际例子

- 若要得到一个圆的面积 Y , 总是测量其半径, 半径的测量值可看作随机变量 X . 若 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则 $Y = \pi X^2$ 的分布是什么?
- 若已知体重 W (kg) 服从正态分布. 在身高 L (m) 确定的情形下, 体质指数 $\text{BMI} = W/L^2$ 服从什么分布?

抽象化的问题

已知随机变量 X 的分布, 考虑一个新的随机变量 $Y = g(X)$, 其中函数 $g(\cdot)$ 已知. 求 Y 的分布.

例题

已知 X 有概率分布

X	-1	0	1
p_i	0.2	0.5	0.3

令 $Y = X^2$. 求 Y 的概率分布.

分析

Y 的所有可能取值为 0, 1, 因此是一个离散型随机变量. 我们只需要求出 $Y = 0$ 或 1 的概率就能得到分布列.

解

$$P(Y = 0) = P(X = 0) = 0.5,$$

$$P(Y = 1) = P(\{X = 1\} \cup \{X = -1\}) = P(X = 1) + P(X = -1) = 0.2 + 0.3 = 0.5.$$

离散型 → 离散型

引理

设随机变量 X 的频率函数为

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

则 $Y = g(X)$ 的频率函数为

Y	$g(x_1)$	$g(x_2)$	\cdots	$g(x_n)$	\cdots
p_k	p_1	p_2	\cdots	p_n	\cdots

(若有些 $g(x_k)$ 值相同, 则把相应的概率值合并相加!)

连续型 \rightarrow 离散型

例子

设随机变量 $X \sim U(0, 1)$, 定义

$$Y = \begin{cases} 0, & 0 < X \leq 0.25, \\ 1, & 0.25 < X \leq 0.75, \\ 2, & 0.75 < X < 1. \end{cases}$$

求随机变量 Y 的频率函数.

分析

Y 的取值为 $0, 1, 2$, 是离散型随机变量. 我们只需要求出 $P(Y = 0)$, $P(Y = 1)$, $P(Y = 2)$ 的概率即可.

解

由 $X \sim U(0, 1)$ 及 Y 的定义知

$$P(Y = 0) = P(0 < X \leq 0.25) = \int_0^{0.25} 1 \cdot dx = 0.25$$

$$P(Y = 1) = P(0.25 < X \leq 0.75) = 0.50$$

$$P(Y = 2) = P(0.75 < X < 1) = 0.25.$$

因此, Y 的频率函数为

Y	0	1	2
p_k	0.25	0.50	0.25

例题 (儿童智商)

设儿童智商 $X \sim \mathcal{N}(100, 100)$. 将儿童按智商分为 3 类, 类标号 Y 规定如下:

$$Y = \begin{cases} 1, & X > 110, \\ 0, & 90 < X \leq 110, \\ -1, & X \leq 90. \end{cases}$$

求 Y 的频率函数.

解

Y	-1	0	1
p_k	0.16	0.68	0.16

小结

$Y = g(X)$ 的概率分布

若 Y 为离散型, 则先写出 Y 的可能取值 $y_1, y_2, \dots, y_j, \dots$, 再找出 $\{Y = y_j\}$ 的等价事件 $\{X \in I\}$. 最后用 $P(Y = y_j) = P(X \in I)$ 求解频率函数.

关键

找出“等价事件”!

例题

设随机变量 X 的密度函数为

$$f_X(x) = \frac{x}{8} \mathbb{1}_{(0,4)}(x).$$

求随机变量 $Y = 2X + 8$ 的密度函数.

分析

我们利用“等价事件”求得 Y 的分布函数, 再对其求导得到密度函数.

解

Y 的分布函数为

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(2X + 8 \leq y) \\ &= P(X \leq (y - 8)/2) = F_X\left(\frac{y - 8}{2}\right) := F_X(g^{-1}(y)), \end{aligned}$$

其中 $g^{-1}(y) = (y - 8)/2$. 因此

$$f_Y(y) = f_X(g^{-1}(y)) (g^{-1}(y))' = \begin{cases} \frac{1}{8} \cdot \frac{y - 8}{2} \cdot \frac{1}{2}, & 0 < \frac{y - 8}{2} < 4 \\ 0, & \text{其它} \end{cases} = \frac{y - 8}{32} \mathbb{1}_{(8,16)}(y).$$

线性变换

问题

设随机变量 X 的概率密度函数为 $f_X(x)$, 随机变量 $Y = a + bX$. 求 Y 的概率密度函数.

分析

$F_Y(y) = P(Y \leq y) = P(a + bX \leq y)$. 这里等价事件的处理需要区分 $b > 0$ 与 $b < 0$ 的情况.

解

$$b > 0 \quad F_Y(y) = P\left(X \leq \frac{y-a}{b}\right) = \int_{-\infty}^{\frac{y-a}{b}} f_X(x) dx, \text{ 因此}$$

$$f_Y(y) = F'_Y(y) = \frac{1}{b} \cdot f_X\left(\frac{y-a}{b}\right).$$

$$b < 0 \quad F_Y(y) = P\left(X \geq \frac{y-a}{b}\right) = 1 - \int_{-\infty}^{\frac{y-a}{b}} f_X(x) dx. \text{ 因此}$$

$$f_Y(y) = F'_Y(y) = -\frac{1}{b} f_X\left(\frac{y-a}{b}\right).$$

流程总结

如何求 $Y = g(X)$ 的概率密度函数

- ① 求随机变量 Y 的分布函数 $F_Y(y) = P(Y \leq y)$.
- ② 转化为关于随机变量 X 的概率计算问题.
需用到函数 $y = g(x)$ 的性质!
- ③ 求导 $f_Y(y) = F'_Y(y)$.

$g(x)$ 单调增情形

问题

设随机变量 X 的概率密度函数为 $f_X(x)$, $y = g(x)$ 单调增且处处可导, 求随机变量 $Y = g(x)$ 的概率密度函数.

解

$$F_Y(y) = P(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x) dx.$$

由链式法则求导得

$$f_Y(y) = F'_Y(y) = f_X(g^{-1}(y)) \cdot [g^{-1}(y)]'.$$

讨论

- 若随机变量 X 有取值范围 (a, b) , 则 $f_Y(y)$ 有定义域 $g(a) < y < g(b)$.
- 要求 $y = g(x)$ 严格递增, 才能保证 $g^{-1}(y)$ 存在.
- 若 $y = g(x)$ 严格递减, 有什么结论?

$$f_Y(y) = -f_X(g^{-1}(y)) [g^{-1}(y)]' = f_X(g^{-1}(y)) |[g^{-1}(y)]'|.$$

定理

设随机变量 X 的密度函数为 $f(x)$, 又设 $y = g(x)$ 为严格单调函数 (严格增或严格减), 其反函数 $h(y) = g^{-1}(y)$ 连续可导. 则 $Y = g(X)$ 的密度函数为

$$f_Y(y) = \begin{cases} |h'(y)| \cdot f(h(y)), & h(y) \text{ 有意义} \\ 0, & \text{其它.} \end{cases}$$

定理的应用

例子

设 $X \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. 求 $Y = \tan x$ 的密度函数.

解

记 $y = \tan x$, 则 $h(y) = \arctan y$, $h'(y) = \frac{1}{1+y^2}$, $-\infty < y < \infty$. 于是, Y 的密度函数为

$$f_Y(y) = |h'(y)| \cdot f_X(h(y)) = \frac{1}{\pi(1+y^2)}, \quad -\infty < y < \infty.$$

注

上述分布称为 Cauchy 分布.

正态分布的线性变换

例题

设 $X \sim \mathcal{N}(\mu, \sigma^2)$. 求 $Y = aX + b$ ($a \neq 0$) 的密度函数.

解

记 $y = ax + b$, 则 $h(y) = \frac{y-b}{a}$, $h'(y) = \frac{1}{a}$, $-\infty < y < \infty$. 于是 Y 的密度函数为

$$f_Y(y) = |h'(y)| \cdot f_X(h(y)) = \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{((y-b)/a - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{[y - (a\mu + b)]^2}{2(a\sigma)^2}}.$$

即 $aX + b \sim \mathcal{N}(a\mu + b, (a\sigma)^2)$.

注

正态分布对线性变换封闭.