

Chapter 1

Lecture 2

Lecture 1

重在理解统计的思想。

5°公理化定义: P 是一个实函数: $\mathcal{F} \rightarrow \mathbb{R}$ s.t.

(1) $P(A) \geq 0, \forall A \in \mathcal{F}$

非负性

(2) $P(\Omega) = 1$

归一化

(3) $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \forall i \neq j, A_i \cdot A_j = \emptyset$ 加法公理.

由 σ -algebra, $\sum_{i=1}^{\infty} A_i$ 有意义

称 P 为概率函数, 是给出 P 的一般性质, 不提供赋值方法.

称 $(\Omega, \mathcal{F}, P)$ 为一个概率空间

命题: (1) $P(A) \leq 1, \forall A \in \mathcal{F}$

(2) $P(\emptyset) = 0.$

(3) $P(A) + P(A^c) = 1, \forall A \in \mathcal{F}$

(4) $P(\sum_{i=1}^n A_i) = \sum_{i=1}^n P(A_i), \forall i \neq j, A_i \cdot A_j = \emptyset$

(5) $P(A) \leq P(B), \forall A \subset B$

(6) $P(A+B) = P(A) + P(B) - P(AB) \quad \forall A, B \in \mathcal{F}$

$$P(\sum_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} A_{i_2}) + \sum_{i_1 < i_2 < i_3} P(A_{i_1} A_{i_2} A_{i_3}) - \dots + (-1)^{n+1} \sum_{i_1 < \dots < i_n} P(\prod_{k=1}^n A_{i_k}) + \dots$$

$$= \sum_{r=1}^n (-1)^{r+1} \sum_{i_1 < \dots < i_r} P(\prod_{k=1}^r A_{i_k})$$

e.g. (配对问题) n 个人, 每人一顶中帽子, 参会时摘帽, 会后随机取帽.

1 问无人拿到自己的帽子的概率?

记 A_i = 第 i 个人拿到自己的帽子.

则 $P(A_i) = \frac{1}{n}$

则无人拿到帽子为 $A_1^c \dots A_n^c$

考虑其对立事件: $A_1 + \dots + A_n$ (至少有一人拿到自己的帽子)

而 $P(\sum_{i=1}^n A_i) = \sum_{r=1}^n (-1)^{r+1} \sum_{i_1 < \dots < i_r} P(\prod_{k=1}^r A_{i_k})$

而 $P(A_{i_1} \dots A_{i_r}) = \frac{(n-r)!}{n!}$

$\sum_{i_1 < \dots < i_r} P(A_{i_1} \dots A_{i_r}) = \frac{(n-r)!}{n!} \binom{n}{r} = \frac{1}{r!}$

$P_n = P(\prod_{i=1}^n A_i^c) = 1 - P(\sum_{i=1}^n A_i) = 1 - [1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + (-1)^{n+1} \frac{1}{n!}]$

而 $\lim_{n \rightarrow \infty} P_n = \frac{1}{e} \quad \parallel = \sum_{i=2}^{\infty} \frac{(-1)^i}{i!}$

2 恰有 k 人拿到自己帽子的概率.

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

(前两项相反)

$$e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} = \sum_{n=2}^{\infty} \frac{(-1)^n}{n!}$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

6° 条件概率.

e.g. 掷色子. $A = \{2, 3, 4, 5\}$, $B = \{1, 3, 5\}$

由古典概型: $P(A) = \frac{4}{6}$, $P(B) = \frac{3}{6}$

$$P(A|B) = \frac{2}{3} = \frac{P(AB)}{P(B)}$$

Def. $P(A|B) \triangleq \frac{P(AB)}{P(B)}$ ($P(B) > 0$)

Remark. ① 计算条件概率 { 改变样本空间
依定义计算

② 用于计算乘积的概率: $P(AB) = P(A|B)P(B)$ 乘法法则.

e.g. 取球问题: 8红, 4白. 无放回. 取2球, 均为红色的概率?

$$\frac{C_8^2}{C_{12}^2} = \frac{8 \times 7}{12 \times 11} \quad \left| \begin{array}{l} P(R_1) = \frac{8}{12} \\ \downarrow \\ \text{第一个红球} \end{array} \right. \quad P(R_2|R_1) = \frac{7}{11}$$

$\therefore P(R_2) = P(R_2|R_1) \cdot P(R_1)$ 推荐这种算法, 因为它用到了运算法则. 分析了事件间关系.

推广的乘法法则: $P(A_1 \cdot A_2 \cdots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1, A_2) \cdots P(A_n|A_1 \cdots A_{n-1})$

回顾之前的例子: $P(A_1 \cdots A_r) = \frac{1}{n} \cdot \frac{1}{n-1} \cdots \frac{1}{n-(r-1)} = \frac{(n-r)!}{n!}$ ← Kill掉前 $n-r$ 项, 只留后 r 项.

考虑函数 $P(\cdot|B): \mathcal{F} \rightarrow \mathbb{R}$. (fixed $B, P(B) > 0$)

$A \rightarrow P(A|B)$

命题: $P(\cdot|B)$ 是概率函数

Remark. ① $(\Omega, \mathcal{F}, \mathbb{P})$ 是一个概率空间.

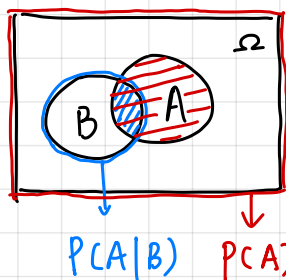
② 区分 $P(A)$ 与 $P(A|B)$ $\hookrightarrow$ 先验 v.s. 后验

e.g. $P(A), P(A|A)$

7° 独立事件:

定义: 若 $P(AB) = P(A)P(B)$, 则称 A, B (相互) 独立.

Remark: 比对 $\forall A, B, P(AB) = P(A|B)P(B)$, 若 $P(A|B)$ 存在, 则 $P(A) = P(A|B)$.



独立性直观上表现为等比缩放

Remark. (1) 从实际角度判断独立性 $\implies$ 可用独立性定义式.

(2)

(3) A 与 B 独立 $\iff A$ 与 B^c 独立.

定义: A, B, C 独立 $\iff \begin{cases} P(ABC) = P(A) \cdot P(B) \cdot P(C) \\ A, B, C \text{ 两两独立} \end{cases}$ 两条件缺一不可.

定义: $A_1, A_2, \dots$ 为有限或可数个事件, 其中任有限个事件, 都有: $P(A_{i_1} \cdots A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$
则称 $A_1, A_2, \dots$ 相互独立

e.g. 彩票每周开奖, 每周开奖, 不累计, 中奖率 10^5 , 十年不中奖的概率 (以 520 周计)

记 A_i = 第 i 周不中奖, 则 $P(A_i) = 1 - 10^{-5}$

$$\text{则 } P(A_1 \cdots A_n) = P(A_1) \cdots P(A_n) = (1 - 10^{-5})^{520} \approx 0.9948$$

定义 (条件独立): A, B 关于事件 E 条件独立 $\iff P(AB|E) = P(A|E) \cdot P(B|E)$

Remark. A, B 关于 E 条件独立与 A, B 独立 互不蕴含

8° Bayes 公式

全概率公式:

给定 Ω 的一个划分 B_i s.t. $\begin{cases} \sum B_i = \Omega \\ \forall i \neq j, B_i \cdot B_j = \emptyset \\ P(B_i) > 0 \end{cases}$ → 指有限或可数个求和

$$\text{则 } P(A) = P(A \sum B_i) = P(\sum AB_i) = \sum P(AB_i) = \sum P(B_i) P(A|B_i)$$

e.g. 敏感问题调查:

$$P(A|F) = \sum P(B_i|F) P(A|B_i, F)$$

$$\frac{k}{n} \approx \frac{1}{2} p + \frac{1}{2} q \implies p \approx \frac{2k}{n} - q$$

e.g. 假阳性悖论

记 B = 患病, A = 阳性, $P(B) = 10^{-4}$, $P(A|B) = 0.99$, $P(A|B^c) = 10^{-3}$
求 $P(B|A)$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B) \cdot P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)} \approx 9\%$$

对 $P(A)$ 进行用 B 诱导的
全概率分解

Bayes 公式: 给定划分 B_j 和事件 B_i .

$$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_j P(B_j) P(A|B_j)} \quad (\text{导致 } A \text{ 的可能性分析})$$

Remark: $P(B_j)$ 先验概率, $P(B_j|A)$ 后验修正.

e.g. 前例续: 重新检测仍为阳性, 且两次检测独立, 同效力. 此时患病概率?
记 A_i = 第 i 次为阳性.

$$P(B|A_1 A_2) = \frac{P(A_1 A_2|B) \cdot P(B)}{P(A_1 A_2)} = \frac{P(A_1|B) \cdot P(A_2|B) \cdot P(B)}{P(A_1 A_2)} \approx 0.99$$

e.g. LA 发生劫案, 目击者看到了 3 嫌犯特征. 几天后, 警方逮捕了一对夫妇, 符合前述特征. 无实证.
一位数学家计算得: 随机选取一对夫妇, 符合特征的概率 $p = 8.3 \times 10^{-8}$, 初审认为有罪
用高法撤消了判决.

A = 至少有一对符合; B = 仅有一对符合; 假设特征独立. 设 $n = 800 \text{ 万} = 8 \times 10^6$

$$P(A^c) = (1-p)^n, \quad P(A) = 1 - (1-p)^n$$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)}{P(A)} = \frac{\binom{n}{1} P(1-p)^{n-1}}{1 - (1-p)^n} = 0.7 \ll 1.$$

Lecture 3

为什么要使用概率:

• 不确定性的来源

1) 被建模系统内部的随机性

2) 观测的不完全性 (e.g. Monty Hall Problem)

3) 建模的不完全性 (e.g. 抛硬币)

• 很多情况下, 简单而不确定的规则比复杂而确定的规则实用

※ { 正确计算概率
正确使用的概率

Chapter 2 随机变量

1° (-维) 随机变量

定义: 样本空间上的实值函数: $X: \Omega \rightarrow \mathbb{R}$

(该定义不严谨)

e.g. 试验 样本空间 随机变量 $\omega \mapsto X(\omega)$ 随机变量的像集 (新样本空间)

随机调查 50 人的意见
支持 or 反对

50 人的所有可能意见
($x_1, x_2, x_3, \dots, x_{50}$)
共 2^{50} 个

X_1
“1 的个数”

$\{0, 1, \dots, 50\}$

随机抽取一个北京
成年市民

所有北京成年市民
构成的集合

X_2
“其年收入”

$(-\infty, +\infty)$

e.g. 事件: $X_1 = 7$; $X_2 > 10$ 万.

Remark. ① 随机变量起概括作用, 是试验的数值摘要.

② 事件 — 静态描述
变量 — 动态描述

分类: ① 离散型: 至多可数多个取值

② 连续型: 区间型取值

③ 其他

定义: $P_X(X \in I) \triangleq P(X^{-1}(I)) \quad \forall I \subseteq \mathbb{R}, I \text{ 可测}$

① X^{-1} 表示 I 的原像集 $\subset \Omega$ (事实上应该要 $\subset \mathcal{F}$)

② 通常记 P_X 为 P

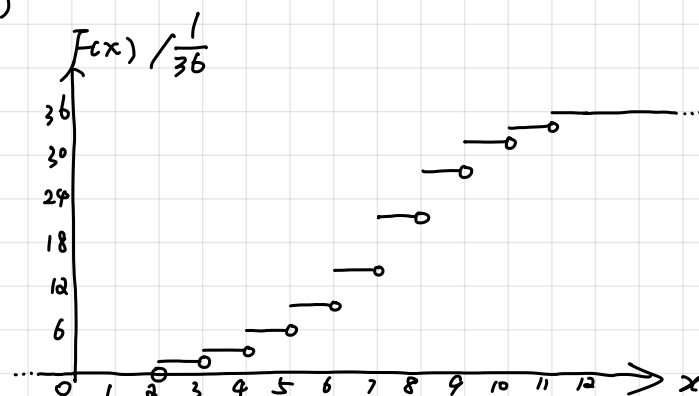
定义: (累计)分布函数: Cumulative Distribution Function, abbr. **CDF**

$$F(x) \triangleq P(X \leq x), \forall x \in \mathbb{R}$$

Remark: $P(a < X \leq b) = F(b) - F(a)$

eg. $X := 2$ 个色子的点数之和

X	2	3	4	5	6	7	8	9	10	11	12
P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Remark:

(1) $F(x)$ 单增 (但不一定严格)

(2) $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$ (单调收敛原理)

(3) $F(x)$ 右连续 (用任意右侧收敛点列, 只需单调点列)

Remark:

(1) 以上三个性质是函数成为概率分布函数的充要条件.

Remark:

(1) 样本空间可以很具体

(2) 随机变量的随机要素来自样本点 ω 的随机选择

(3) 应用中, 随机变量的直观意义往往出现在样本空间的定义之前

eg. $X_i = \begin{cases} 1, & \text{第 } i \text{ 次正面向上} \\ 0, & \text{第 } i \text{ 次反面向上} \end{cases} \quad (i=1, 2)$

考查 3 个随机变量 $X_1 + X_2, X_1, X_2$.

以上三者的定义域都是 $\{\text{正}, \text{反}\} \times \{\text{正}, \text{反}\}$

(4) $aX + bY, XY, \frac{X}{Y} (Y \neq 0), \dots, g(X, Y)$ 在 X, Y 定义域相同时都是随机变量.

(5) 需要 $X^{-1}(I) \in \mathcal{F}$ (否则 $P(X^{-1}(I))$ 缺少定义)

(6) 同分布: 设 CDF of X_i 为 $F_i(x), i=1, 2$.

X_1, X_2 同分布 $\Leftrightarrow \forall I \subset \mathbb{R}$ 可测, 有 $P(X_1 \in I) = P(X_2 \in I)$

$\Leftrightarrow \forall x \in \mathbb{R}$, 有 $F_1(x) = F_2(x)$

(7) 同分布 **不蕴含** 两随机变量相同. **随机变量相同是两个函数的相等**

是映射关系的相等即 $\forall \omega \in \mathcal{F}, X_1(\omega) = X_2(\omega)$

eg. 掷 1 枚硬币: $X = \text{正面向上次数}$
 $Y = \text{反面向上次数}$ } X, Y 同分布, 但含义不同

2° 离散分布

定义: 分布表:
$$\begin{array}{c|c} X & x_1, x_2, x_3, \dots \\ \hline P & p_1, p_2, p_3, \dots \end{array}$$

概率质量函数 (PMF): $f(x) \triangleq P(X=x), \forall x \in \mathbb{R}$.

Remark: ① $f(x_i) = p_i \geq 0, \sum_i f(x_i) = \sum_i p_i = 1 = \sum_{x \in \mathbb{R}} f(x)$ → 因为仅有至多可数个点有意义 满足概率律.

② cdf 为阶梯函数

定义: 期望 $E(X) = \sum_i x_i p_i = \sum_i x_i f(x_i)$

$$\begin{aligned} \text{方差 } \text{Var}(X) &= \sum_i (x_i - E(X))^2 p_i = \sum_i (x_i - E(X))^2 p_i \\ &= E[(X - E(X))^2] \\ &= E(X^2) - E^2(X) \end{aligned}$$

Remark ① 期望即均值: $\frac{1}{n} \sum_{i=1}^n a_i = \sum_j \frac{k_j}{n} x_j$ (相异 x_j)

② 期望存在 $\Leftrightarrow \sum_i x_i f(x_i)$ 绝对收敛 (因为期望要对交换顺序成立)

③ $E(X)$ 刻画了分布特征

$$\textcircled{4} E(g(X)) = \sum_i g(x_i) f(x_i)$$

3° 常见的离散分布

(1) Bernouli 分布 (两点分布)

$$X = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}, p \in (0, 1)$$

记为 $X \sim B(p)$

$$E(X) = p, \text{Var}(X) = p(1-p)$$

e.g. 示性随机变量 $X = \begin{cases} 0, & A \text{ 未发生} \\ 1, & A \text{ 发生} \end{cases}, p = P(A), X \sim B(p)$

(2) 二项分布

X : n 次独立 Bernouli 实验的成功次数

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

记为 $X \sim B(n, p)$

$$E(X) = np, \text{Var}(X) = np(1-p)$$

(3) Poisson 分布

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda > 0 \text{ 为常数}, k \in \mathbb{N}$$

记为 $X \sim P(\lambda)$ 或 $X \sim \text{Poisson}(\lambda)$

$$E(X) = \text{Var}(X) = \lambda$$

引入观察时间 $[0, 1]$ (一定时间内)

e.g. 某路口交通事故数 X

$$I_i = \left[\frac{i-1}{n}, \frac{i}{n} \right] \quad i = 1, \dots, n$$

n 充分大, s.t.

假设: ① I_i 内至多发生一次事故 ($\rightarrow$ 两点分布)

② 恰发生一次的概率 $p \propto \frac{1}{n}$, 记 $p = \frac{\lambda}{n}$ (n 很大, 高阶项略去)

③ 各段独立

$$\text{则 } X \sim B(n, p), p = \frac{\lambda}{n}$$

$$\text{有 } P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \frac{n!}{(n-k)!} \cdot \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$\text{令 } n \rightarrow \infty, \text{ 有 } P(X=k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

注: $X \sim B(n, p) \Rightarrow X \overset{\text{近似}}{\sim} P(\lambda)$, 其中 $\lambda = np$ $\leftarrow$ 一段时间内的平均发生次数

p 很小, n 很大, λ 不太大时, 近似效果较好, 误差最多不超过 $\min\{p, np^2\}$

Poisson 分布用于一定时空内小概率事件发生的次数.

Lecture 4

"All models are wrong, but some are useful." — George Box

Possion 分布 (续)

e.g. 某医院平均每小时出生 λ 个婴儿,
现有一名婴儿出生, 求 接下来 t 小时 中出生的婴儿人数分布.

— 定时空

$$\text{设 } N(t) \sim \text{Possion}(\lambda t)$$

$$\Rightarrow P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad k \in \mathbb{N}.$$

注: 若试验不独立, 但弱相依, 仍为较好近似.

e.g. (配对问题)

$$\left. \begin{aligned} P(A_i) &= \frac{1}{n} \\ P(A_i | A_j) &= \frac{1}{n-1} \quad \forall i \neq j \end{aligned} \right\} A_i, A_j \text{ 弱相依}$$

Σ := 拿到自己帽子的人数

设 $\Sigma \approx \text{Possion}(\lambda)$, $\lambda \approx n \cdot \frac{1}{n} = 1$

$$\therefore P(\Sigma = k) = \frac{1}{k!} e^{-1}$$

下面检验上述近似.

考虑指定 k 个人拿到自己帽子 (互斥)

E : 指定 k 个人拿自己帽子 F : 其余 $(n-k)$ 人没拿到自己的帽子
恰这指定的 k 个人 没蕴涵 (除非说 "仅")

$$P(EF) = \frac{(n-k)!}{n!} P_{n-k}, \text{ 其中 } P_n = \sum_{k=0}^n \frac{1 \cdot 0^k}{k!}$$

$$P(\Sigma = k) = \binom{n}{k} \cdot P(EF) = \frac{n!}{k!(n-k)!} \cdot \frac{(n-k)!}{n!} P_{n-k} = \frac{1}{k!} \cdot P_{n-k}$$

$$\lim_{n \rightarrow \infty} P(\Sigma = k) = \frac{1}{k!} e^{-1}, \text{ 与 Possion 分布的近似结果一致.}$$

4° 连续分布:

定义: 若存在 $f(x) \geq 0$, 使得对于 $\forall$ 可测的 $I \subset \mathbb{R}$, 都有

$$P(\Sigma \in I) = \int_I f(x) dx$$

则称 Σ 为连续型随机变量,
称 f 为 Σ 的概率密度函数 (PDF)

性质: 1. $\int_{\mathbb{R}} f(x) dx = 1$

$$2. P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) dx = P(a \leq X < b)$$

$$3. \forall a \in \mathbb{R}, P(X=a) = 0.$$

$$4. P(x_0 - \delta < X \leq x_0 + \delta) = \int_{x_0 - \delta}^{x_0 + \delta} f(x) dx \approx 2\delta f(x_0) \quad (\text{若 } f \text{ 于 } x_0 \text{ 处连续})$$

局部概率 $\propto$ 密度 f

$$5. F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

$$F(x) \in C^1(\mathbb{R}),$$

若 f 于 x_0 连续, 则 $F'(x_0) = f(x_0)$

6. 若 pdf 存在, 则 **不** 唯一.

定义: 连续型随机变量的期望 $E(X) = \int_{\mathbb{R}} x f(x) dx$

$$\text{方差 } \text{Var}(X) = \int_{\mathbb{R}} (x - E(X))^2 f(x) dx = E(X^2) - E^2(X)$$

注: (1) 约定: $E(X)$ 存在 $\iff$ 则 $E(X)$ 有限

$E(X)$ 存在 $\iff \int_{-\infty}^{+\infty} x f(x) dx$ 绝对收敛 (不是柯西主值意义下)

$$(2) \text{一般地 } E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

5° 常见连续分布

(1) 均匀分布 (uniform distribution)

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

记为 $X \sim U(a, b)$

$$E(X) = \int_a^b \frac{x dx}{b-a} = \frac{a+b}{2}$$

$$\text{Var}(X) = E(X^2) - E^2(X)$$

$$= \int_a^b \frac{x^2 dx}{b-a} - \frac{(a+b)^2}{4}$$

$$= \frac{1}{3} \frac{b^3 - a^3}{b-a} - \frac{a^2 + 2ab + b^2}{4}$$

$$= \frac{1}{3} (b^2 + ab + a^2) - \frac{1}{4} (a^2 + 2ab + b^2)$$

$$= \frac{1}{12} (4b^2 + 4ab + 4a^2) - \frac{1}{12} (3a^2 + 6ab + 3b^2)$$

$$= \frac{1}{12} (a^2 - 2ab + b^2)$$

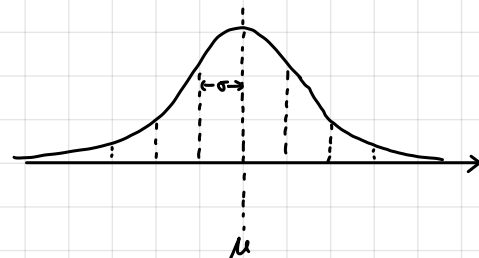
$$= \frac{1}{12} (a-b)^2$$

注: $X \sim U(0, 1)$ 通常称为随机数

(2) 正态分布

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (x \in \mathbb{R})$$

记为 $X \sim N(\mu, \sigma^2)$



注: (1) $E(X) = \mu$, $Var(X) = \sigma^2$

(2) $N(0, 1)$ 称为标准正态

(3) 设 $X \sim N(\mu, \sigma^2) \iff Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$

$$P(y \in [c, d]) = \int_c^d f(y) dy$$

$$\stackrel{\text{变量代换}}{=} \int_{\sigma c + \mu}^{\sigma d + \mu} f\left(\frac{x-\mu}{\sigma}\right) d\frac{x-\mu}{\sigma}$$

$$\stackrel{\text{X, Y 关系定义}}{=} \int_{\sigma c + \mu}^{\sigma d + \mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

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

$$\Rightarrow f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

(3) 指数分布

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

其中 $\lambda > 0$.

记为 $X \sim \text{Exp}(\lambda)$

$$E(X) = \lambda^{-1}, \quad Var(X) = \lambda^{-2}$$

注: ① 有文献认为参数是 $\beta = \lambda^{-1}$

② X 通常用于刻画寿命或等待时间

③ 尾概率 $P(X > x) = 1 - F(x) = e^{-\lambda x} \quad (x > 0)$

e.g. 观察到有一名婴儿出生, 接下来 t 小时有婴儿出生的概率

令 $X :=$ 间隔时间

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

$$\Rightarrow P(X \leq t) = 1 - e^{-\lambda t} \Rightarrow X \sim \text{Exp}(\lambda)$$

指数分布刻画了 Poisson 过程中的间隔时间.

e.g. 失效率 (危险率): 设 X 连续分布, cdf 为 $F(x)$, $F(0) = 0$

$$P(x < X < x+dx \mid X > x) = \frac{P(x < X < x+dx)}{P(X > x)} \approx \frac{F'(x)dx}{1-F(x)} = \frac{F'(x)}{1-F(x)} dx$$

"年龄"为 x 的零件不能继续使用的条件概率密度为 $\frac{F'(x)}{1-F(x)}$

$$\text{令 } \lambda(x) = \frac{F'(x)}{1-F(x)} \Rightarrow F(x) = 1 - e^{-\int_0^x \lambda(t) dt}$$

记 $\lambda(x) \equiv \lambda$ (无老化假设), 有 $F(x) = 1 - e^{-\lambda x}$, $x > 0$ 即 $X \sim \text{Exp}(\lambda)$

这表明指数分布刻画了无老化假设的寿命分布

$$\Rightarrow P(X > t+s | X > s) = \frac{P(X > t+s)}{P(X > s)} = \frac{\lambda e^{-\lambda(t+s)}}{\lambda e^{-\lambda s}} = e^{-\lambda t}$$

这概率与 s 无关, 说明 X 具有无记忆性

注: “无记忆性”这一局限来源于无老化假设

改进: 假设 $\frac{F'(x)}{1-F(x)} = \lambda(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1}$, $x > 0$, $\alpha > 0$, $\beta > 0$

$$\Rightarrow F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

这是 Weibull 分布

$\alpha=1$ 时, 退化成指数分布

6° 随机变量的函数

记 $Y = g(X)$

若 X 离散, 则 Y 离散

若 X 连续, Y 不一定连续, 例如: 设 $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, $Y = \begin{cases} 0, & x \leq t_0 \\ 1, & x > t_0 \end{cases}$, $t_0 > 0$,
则 $P(Y=0) = P(X \leq t_0) = 1 - e^{-\lambda t_0}$
 $P(Y=1) = e^{-\lambda t_0}$

设 $Y = X^2$

$$\forall y > 0: P(Y \leq y) = P(X^2 \leq y)$$

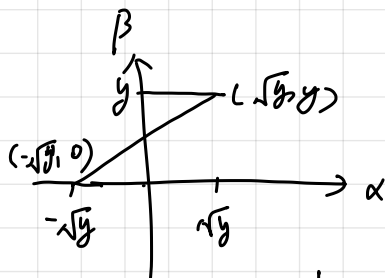
先去掉
无意义部分

$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} f(x) dx$$

$$\stackrel{\text{换元}}{=} \int_0^y l(x) dx$$

$$\Rightarrow l(x) = \begin{cases} \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})) & , y > 0 \\ 0 & , y \leq 0 \end{cases}$$



$$k = \frac{y}{2\sqrt{y}} = \frac{\sqrt{y}}{2}$$

$$\int_0^y \frac{2}{\sqrt{y}} f\left(\frac{2}{\sqrt{y}}\beta - \sqrt{y}\right) d\beta \quad \alpha = \frac{2}{\sqrt{y}}\beta - \sqrt{y} \quad \alpha + \sqrt{y} = \frac{2}{\sqrt{y}}\beta$$

$$\int_0^y f\left(\frac{2}{\sqrt{y}}\beta - \sqrt{y}\right) d\left(\frac{2}{\sqrt{y}}\beta - \sqrt{y}\right) = \frac{\sqrt{y}}{2} (\alpha + \sqrt{y})$$

Lecture 5

Chapter 3 联合分布

1° 随机向量 $(X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$.

↳ 几维随机向量, 每个分量都是同一个 Ω 上的随机变量.

(联合) 累积分布函数 (cdf): $F(x_1, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$
 $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

注: $n=2$ 时, 称为二元分布, 随机向量通常记为 (X, Y) , cdf: $F(x, y)$

2° 离散型随机向量: X_i 均是离散的.

pmf: $f(x_1, \dots, x_n) := P(\forall i \in \{1, \dots, n\}, X_i = x_i)$

$$\begin{cases} f \geq 0 \\ \sum_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) = 1 \end{cases}$$

e.g. (多项分布) 某随机试验有 n 个互斥的可能结果 $A_1, \dots, A_n$,

$\sum_{i=1}^n A_i = \Omega$ (即 $\{A_i\}$ 是 Ω 的一个划分)

A_i 的概率是 p_i , A_i 出的次数是 X_i .

设 $\sum_{i=1}^n k_i = N$ (总实验次数)

$$P(X_1 = k_1, \dots, X_n = k_n) = \binom{N}{k_1, \dots, k_n} \prod_{i=1}^n p_i^{k_i}$$

其中, 多项式系数 $\binom{N}{k_1, \dots, k_n} = \frac{N!}{\prod_{i=1}^n k_i!}$, 即 $(\sum_{i=1}^n a_i)^N$ 的展开中 $\prod_{i=1}^n a_i^{k_i}$ 的系数

3° 连续分布

定义: 若存在 $f(x_1, \dots, x_n) \geq 0$, 使得对 $\mathbb{R}^n$ 的任何可测子集 Q , 都有

$$P((X_1, \dots, X_n) \in Q) = \int_Q f dx_1 \cdots dx_n$$

则称 $(X_1, \dots, X_n)$ 为连续型, $f(x_1, \dots, x_n)$ 为 pdf.

注: ① $\int_{\mathbb{R}^n} f dx_1 \cdots dx_n = 1$

② $n=2$ 时: $F(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy$

$$\Rightarrow \frac{\partial^2 F}{\partial x \partial y}(a, b) = f(a, b) \quad (\text{a.e.})$$

e.g. 均匀分布 ($n=2$, 矩形分布)

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)}, & (x, y) \in (a, b) \times (c, d) \\ 0, & \text{其他} \end{cases}$$

e.g. 二元正态分布

$$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho), \text{ 其中 } |\rho| < 1$$

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \cdot \frac{1}{\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 - 2\rho \frac{x-\mu_1}{\sigma_1} \cdot \frac{y-\mu_2}{\sigma_2} \right] \right\}$$

$$\text{指数部分} = -\frac{1}{2} \mathbf{Z}^T W \mathbf{Z} = -\frac{1}{2} \mathbf{Z}^T A^T A \mathbf{Z} = -\frac{1}{2} (A\mathbf{Z})^T A\mathbf{Z} = -\frac{1}{2} \|A\mathbf{Z}\|^2$$

$$\text{其中: } \mathbf{Z} = \begin{pmatrix} \frac{x-\mu_1}{\sigma_1} \\ \frac{y-\mu_2}{\sigma_2} \end{pmatrix}, W = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} = \frac{A^T A}{1-\rho^2}, A = \begin{pmatrix} 1 & -\rho \\ 0 & \pm\sqrt{1-\rho^2} \end{pmatrix}$$

其中可能考?

问: ① $f(x, y)$ 的等值线是椭圆

② ρ 有何意义? 直观上看, x 与 y 互相影响的程度

4° 边际分布

考虑 $(X_1, \dots, X_n)$

X_i 的边际cdf: $F_i(x) := P(X_i \leq x) = P(X_i \leq x, \forall j \neq i, X_j \in \mathbb{R})$

e.g. $n=2$ (X, Y) 的cdf $F(x, y)$.

则 X 的边际cdf: $F_X(x) := P(X \leq x)$

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

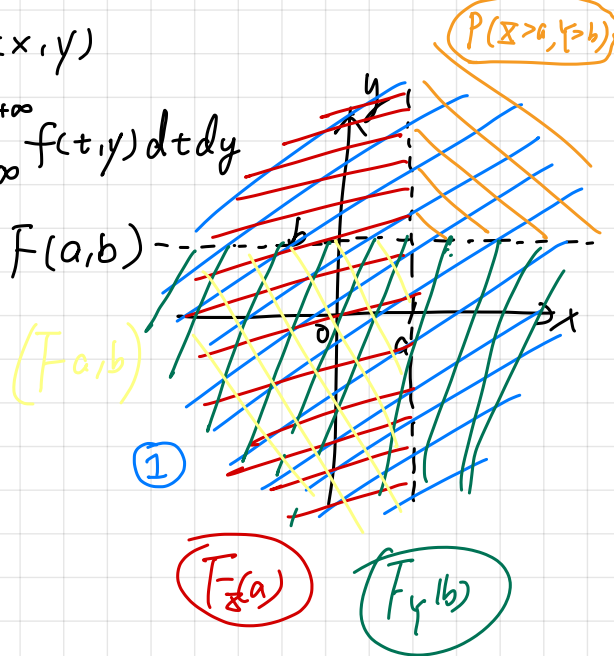
$$= \lim_{y \rightarrow +\infty} P(X \leq x, -\infty < Y \leq y)$$

$$= \lim_{y \rightarrow +\infty} F(x, y)$$

e.g. 连续型, $n=2$ 为例 $F_Z(X) = \lim_{y \rightarrow +\infty} F(x, y)$

$$= \int_{-\infty}^x \int_{-\infty}^{+\infty} f(t, y) dt dy$$

e.g. $P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F(a, b)$



e.g. 离散型, $n=2$ 为例

$$P(Z = z) = \sum_y P(Z = z, Y = y)$$

蕴含所有y都取到
剩余的完全取到即可

e.g. 连续型, $n=2$ 为例 (X, Y) , pdf为 $f(x, y)$

$$F_Z(x) = P(Z \leq x) = \int_{-\infty}^x \int_{\mathbb{R}} f(s, t) dt ds$$

$$Z \text{ 的边沿 pdf } f_Z(x) = \int_{\mathbb{R}} f(x, y) dy$$

e.g. $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$f_Z(x) = \int_{\mathbb{R}} f(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, x \in \mathbb{R}$$

$$\Rightarrow X \sim N(\mu_1, \sigma_1^2)$$

$$\text{同理, } Y \sim (\mu_2, \sigma_2^2)$$

综上, 联合分布确定边缘分布

反之, 边缘分布不能确定联合分布

5° 条件分布

1) $n=2$ 为例 (X, Y)

(i) 离散型

$$P(X=a_i, Y=b_j) = p_{ij}, \quad \sum_{i,j} p_{ij} \equiv 1$$

$$P(X=a_i | Y=b_j) = \frac{P(X=a_i, Y=b_j)}{P(Y=b_j)} = \frac{p_{ij}}{\sum_k p_{kj}}$$

注: ① $\sum_i P(X=a_i | Y=b_j) = 1$

② 高维情形类似

(ii) 连续型, 设 (X, Y) 的 pdf 为 $f_{X,Y}$

$$\begin{aligned} P(X \leq x | y \leq Y \leq y+dy) &= \frac{P(X \leq x, y \leq Y \leq y+dy)}{P(y \leq Y \leq y+dy)} \\ &= \frac{\int_{-\infty}^x \int_y^{y+dy} f(s, t) dt ds}{\int_y^{y+dy} f_Y(s) ds} \end{aligned}$$

X 此时的条件概率密度为

$$f_X(x | y \leq Y \leq y+dy) = \frac{\int_y^{y+dy} f(x, t) dt}{\int_y^{y+dy} f_Y(s) ds} \xrightarrow[\text{如果有}]{dy \rightarrow 0} \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

定义 条件密度函数

$$f_X(x | y) := \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

$$f_Y(y | x) := \frac{f_{X,Y}(x, y)}{f_X(x)}$$

// 这样你就有了 y 取特定值时的 x 的概率分布

注: ① $F(a | y) = P(X \leq a, Y=y)$

$$\textcircled{2} f_{X,Y}(x, y) = f_X(x | y) \cdot f_Y(y) = f_Y(y | x) \cdot f_X(x) \quad \text{乘法法则}$$

$$\Rightarrow f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \int_{\mathbb{R}} f_X(x | y) f_Y(y) dy \quad \text{全概率公式}$$

$$\Rightarrow f_Y(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x|y) \cdot f_Y(y)}{\int_{\mathbb{R}} f_X(x|y) f_Y(y) dy} \quad \text{Bayes 公式}$$

e.g. 二元正态.

$$f_Y(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi}\sigma_2} \cdot \frac{1}{\sqrt{1-\rho^2}} \exp \left\{ -\frac{(y - (\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)))^2}{2(1-\rho^2)\sigma_2^2} \right\}$$

$$\Rightarrow \text{当 } x=x \text{ 时: } Y \sim N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), (1-\rho^2)\sigma_2^2)$$

6° 独立性

定义: (X, Y) cdf 为 $F(x, y)$, 若 $F(x, y) = F_X(x) \cdot F_Y(y)$, 则称 X, Y 独立

注: X, Y 独立 $\Leftrightarrow f(x, y) = f(x) \cdot f(y) \quad \forall x, y \in \mathbb{R}$ (f : pmf/pdf)

Lecture 6

定义: 设 $(X_1, \dots, X_n)$ 的 cdf 为 $F(x_1, \dots, x_n)$ 若

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

则称 $X_1, \dots, X_n$ 相互独立.

注: $X_1, \dots, X_n$ 相互独立 $\Leftrightarrow f(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$

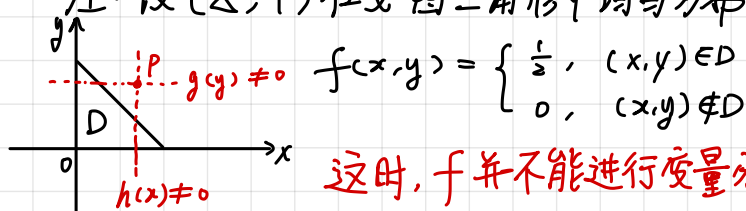
定理(1) 若 $X_1, \dots, X_n$ 独立, $Y_1 = g_1(X_1, \dots, X_m), Y_2 = g_2(X_{m+1}, \dots, X_n)$
则 Y_1, Y_2 独立.

定理(2) 若 pdf: $f(x_1, \dots, x_n) = \prod_{i=1}^n g_i(x_i) \quad \forall x_i \in \mathbb{R}$

则 $X_1, \dots, X_n$ 相互独立, 且 f_{X_i} 与 g_i 仅相差常数因子

这说明相互独立 $\Leftrightarrow$ pdf 变量分离.

注: 设 (X, Y) 在如图三角形中均匀分布



$$f(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in D \\ 0, & (x, y) \notin D \end{cases}$$

这时, f 并不能进行变量分离

$$\Rightarrow h(x)g(y)|_p \neq 0$$

7° 多个随机变量的函数

设 $n > 1$, $Y = g(X_1, \dots, X_n)$

e.g. $X_i \sim B(n_i, p)$, $i=1, 2$, X_1, X_2 独立.

设 $Y = X_1 + X_2$

有 $P(Y=k) = P(X_1 + X_2 = k)$ ($k \in \{0, \dots, n_1 + n_2\}$)

$$\begin{aligned} &= \sum_{k_1=0}^k P(X_1=k_1, X_2=k-k_1) \\ &= \sum_{k_1=0}^k P(X_1=k_1) P(X_2=k-k_1) \quad (\text{独立性}) \\ &= \sum_{k_1=0}^k \binom{n_1}{k_1} p^{k_1} (1-p)^{n_1-k_1} \binom{n_2}{k-k_1} p^{k-k_1} (1-p)^{n_2-k+k_1} \\ &= p^k (1-p)^{n_1+n_2-k} \sum_{k_1=0}^k \binom{n_1}{k_1} \cdot \binom{n_2}{k-k_1} \\ &= \binom{n_1+n_2}{k} p^k (1-p)^{n_1+n_2-k} \end{aligned}$$

$\Rightarrow Y \sim B(n_1+n_2, p)$

e.g. (X_1, X_2) 有 pdf: $f(x_1, x_2)$. $Y = \frac{X_2}{X_1}$, $X_1 > 0$.

$\forall y \in \mathbb{R}, P(Y \leq y) = P(\frac{X_2}{X_1} \leq y)$

$= P(X_2 \leq y X_1)$

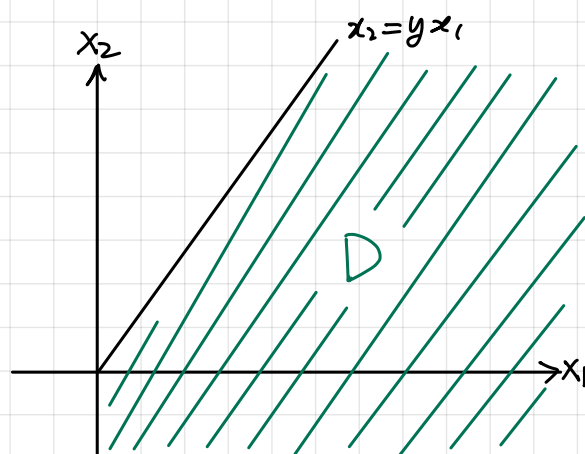
$= \int_D f(x_1, x_2) dx_1 dx_2$

$= \int_0^{+\infty} \left(\int_{-\infty}^{y x_1} f(x_1, x_2) dx_2 \right) dx_1$

跨求导?

$x_2 = t x_1, \int_0^y \left(\int_0^{\infty} x_1 f(x_1, x_1 t) dx_1 \right) dt$

$\Rightarrow y$ 的 pdf: $l(y) = \int_0^{+\infty} x_1 f(x_1, x_1 y) dx_1$



Thm3. 设 $g(t, x), g'_t(t, x) \in C([a, b] \times [c, d])$, $\alpha(t), \beta(t)$ 在 $[a, b]$ 上可导, 且

$c \leq \alpha(t), \beta(t) \leq d, \forall t \in [a, b]$,

则

$$f(t) = \int_{\alpha(t)}^{\beta(t)} g(t, x) dx$$

在区间 $[a, b]$ 上可导, 且

$$\begin{aligned} f'(t) &= \frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} g(t, x) dx \\ &= \int_{\alpha(t)}^{\beta(t)} g'_t(t, x) dx + g(t, \beta(t)) \beta'(t) - g(t, \alpha(t)) \alpha'(t). \end{aligned}$$

令 $\varphi = \int_{\alpha}^{\beta} g(t, x) dx$.

$t \rightarrow \left(\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix} \right) \rightarrow f$

$\frac{df}{dt} = \frac{\partial \varphi}{\partial \alpha} \frac{d\alpha}{dt} + \frac{\partial \varphi}{\partial \beta} \frac{d\beta}{dt} + \frac{\partial \varphi}{\partial t} \frac{dt}{dt}$

积分限
变量换成
积分限

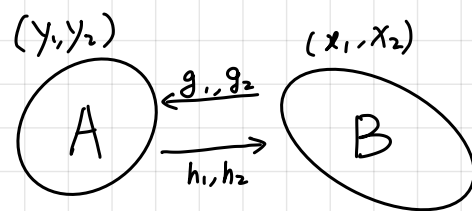
密度函数变换法: ($n=2$ 为例)

设 $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} g_1(X_1, X_2) \\ g_2(X_1, X_2) \end{pmatrix}$ 可逆可微, 即 $\exists h_1, h_2, \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} h_1(Y_1, Y_2) \\ h_2(Y_1, Y_2) \end{pmatrix}$

$$P((Y_1, Y_2) \in A) = P((X_1, X_2) \in B)$$

$$= \int_B f(x_1, x_2) dx_1 dx_2$$

$$= \int_A f(h_1(y_1, y_2), h_2(y_1, y_2)) |\det J| dy_1 dy_2$$



$$\text{其中 } J = \frac{\partial(h_1, h_2)}{\partial(y_1, y_2)} = \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{bmatrix}$$

e.g. (X_1, X_2) pdf 为 $f(x_1, x_2)$, 求 $Y = X_1 + X_2$ 的 pdf (用密度函数变换法)

$$\text{令 } \begin{cases} Y = X_1 + X_2 \\ Z = X_1 \end{cases} \Rightarrow \begin{cases} X_1 = Z \\ X_2 = Y - Z \end{cases} \quad \frac{\partial H}{\partial y} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \quad \det\left(\frac{\partial H}{\partial y}\right) = -1$$

Y, Z 的联合 pdf: $l(y, z) = f(z, y-z) |-1| = f(z, y-z)$

$$l_Y(y) = \int_{\mathbb{R}} l(y, z) dz = \int_{\mathbb{R}} f(z, y-z) dz = \int_{\mathbb{R}} f(y-z, z) dz$$

注: ① 若 X_1, X_2 独立, 则

$$l_Y(y) = \int_{\mathbb{R}} f_1(z) \cdot f_2(y-z) dz = f_1 * f_2(y)$$

注意卷积和积分变换的直观解释.

$$\textcircled{2} (X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$$

③ χ_n^2, t, F 分布

Chapter 4 随机变量的数字特征

1° 期望

$$E(X) = \begin{cases} \sum_{x \in R} x f(x) & , f \text{ 是 pmf} \\ \int_{-\infty}^{+\infty} x f(x) dx & , f \text{ 是 pdf} \end{cases}$$

注: ① 期望 $\Rightarrow f$

② 期望是集中程度的刻画

③ 比一般分布更易计算

性质:

$$\textcircled{1} E(g(X_1, \dots, X_n)) = \begin{cases} \sum_{R^n} g(x_1, \dots, x_n) f(x_1, \dots, x_n) \\ \int_{R^n} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n \end{cases}$$

不必求 pdf,
直接对每个 x 做计算.

$$\textcircled{2} E(\sum_{i=1}^n c_i X_i) = \sum_{i=1}^n c_i E(X_i)$$

$$\textcircled{3} \text{若 } X_1, \dots, X_n \text{ 独立, } E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i) \quad (\text{利用变量分离性, 易证})$$

2° 分位数

定义 $\alpha \in (0, 1)$

包括它自己

若 $P(X \leq a) \geq \alpha$ 且 $P(X \geq a) \geq 1 - \alpha$,

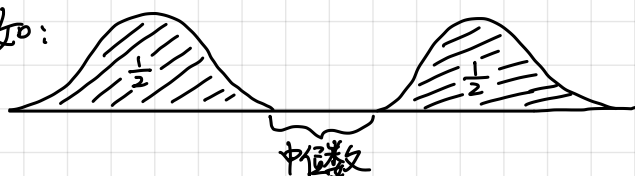
则称 $X=a$ 为 X 的 下侧 α -分位数

① 若 cdf 连续, 则 $F(a) = \alpha$

② $F^{-1}(\alpha)$ 为 α 分位数 $F^{-1}(y) = \inf\{x \mid F(x) \geq y\}$

③ α -分位数不一定唯一

如:



④ $\alpha=0.5 \Leftrightarrow$ 中位数, 是集中趋势的刻画

⑤ e.g.

X	1	2	3	4
P	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{12}$

 中位数为 2 (积累过程中第一个超过 α)

⑥ 众数 (方便定义: f 的最大值点, 不考虑可去间断点)

⑦ 期望 v.s. 中位数 v.s. 众数

↓ ↓
质心 平衡点

3° 方差

定义: $\text{Var}(X) \triangleq E[(X - E(X))^2] = E(X^2) - E(X)^2$

$\text{SD}(X) \triangleq \sqrt{\text{Var}(X)}$

注: 方差类似转动惯量, 是集中程度的刻画

性质: (1) $\text{Var}(C) \equiv 0$

(2) $\text{Var}(X + C) = \text{Var}(X)$

(3) $\text{Var}(cX) = c^2 \text{Var}(X)$

(4) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \overset{2 \text{Cov}(X, Y)}{E[(X - \mu_1)(Y - \mu_2)]}$, 其中 $\mu_1 = E(X)$, $\mu_2 = E(Y)$

$\Rightarrow X, Y$ 独立: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

4° 协方差与相关系数

记 $E(X) = \mu_1$, $E(Y) = \mu_2$, $\text{Var}(X) = \sigma_1^2$, $\text{Var}(Y) = \sigma_2^2$.

定义: $\text{Cov}(X, Y) = E[(X - \mu_1)(Y - \mu_2)]$

性质: (1) $\text{Cov}(X, X) = \text{Var}(X)$

(2) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

(3) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

(4) $\text{Cov}(aX_1 + bX_2 + c, Y) = a \text{Cov}(X_1, Y) + b \text{Cov}(X_2, Y)$

这表明 Cov 是双线性的、平移不变的

(5) $\text{Cov}(\sum_i X_i, \sum_j Y_j) = \sum_{i,j} \text{Cov}(X_i, Y_j)$ 用于展开 $\text{Var}(\sum_i X_i)$

注: $(X_1, \dots, X_n)$ 的期望记为 $(\mu_1, \dots, \mu_n)$, 方差定义为

$$E \left[\begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_n - \mu_n \end{pmatrix} (X_1 - \mu_1, \dots, X_n - \mu_n) \right] = \left(\text{Cov}(X_i, X_j) \right)_{n \times n}$$

这被称为协方差矩阵

定义: 相关系数

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma_1 \sigma_2} = E\left(\frac{X - \mu_1}{\sigma_1} \cdot \frac{Y - \mu_2}{\sigma_2}\right) := \rho$$

定理 (1) X, Y 独立 $\Rightarrow \text{Cov}(X, Y) = 0 \Rightarrow \text{Corr}(X, Y) = 0$, 此时称 X, Y 不相关

(2) $|\text{Corr}(X, Y)| \leq 1$, 当且仅当存在常数 a, b s.t. $P(Y = aX + b) = 1$

证明(2): Lemma: $E^2(UV) \leq E(U^2)E(V^2)$ (Cauchy-Schwartz 不等式)

当且仅当 $\exists t_0$ s.t. $P(V = t_0 U) = 1$.

Lemma 的证明: $E[(V - t_0 U)^2] \geq 0$, 再由 $\Delta \geq 0$ 即得证.

令 $U = \frac{X - \mu_1}{\sigma_1}$, $V = \frac{Y - \mu_2}{\sigma_2}$, 易证.

注: ① $|\rho| = 1 \Rightarrow a = \pm \frac{\sigma_1}{\sigma_2}$

$$0 = \rho \frac{\sigma_2}{\sigma_1}$$

② $\rho = 0$ 不蕴含独立, 例: 令 $X \sim N(0, 1)$, $Y = X^2$, 有 $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$.

③ ρ 只能表征线性相关性.

全是0

e.g. $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

则 $\text{Corr}(X, Y) = \rho$

思路: $\text{Corr}(X, Y) = E\left(\frac{X - \mu_1}{\sigma_1} \cdot \frac{Y - \mu_2}{\sigma_2}\right)$

$$= \int_{\mathbb{R}^2} \frac{x - \mu_1}{\sigma_1} \cdot \frac{y - \mu_2}{\sigma_2} f(x, y) dx dy$$

然后用 Cholesky 分解进行坐标变换即可.

Lecture 7

5°矩

定义: $E[(X-c)^k]$

称为 X 关于 c 的 k 阶矩

通常 $k \in \mathbb{N}_+$.

其中: $c=0$ 称为原点矩
 $c=E(X)$ 称为中心矩

Remark: ① 期望 $E(X)$ 是 1 阶原点矩, 1 阶中心矩 $\equiv 0$

② 方差 $\text{Var}(X)$ 是 2 阶中心矩

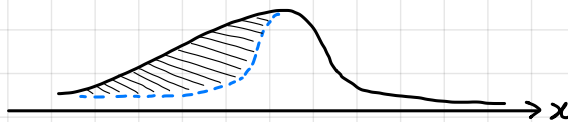
• 偏度系数: $E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] = \frac{E[(x-\mu)^3]}{\sigma^3}$ (3 阶标准矩)

Remark: ① 1 阶标准矩 $\equiv 0$, 2 阶标准矩 $\equiv 1$.

② 相对于 5 阶及以上的奇阶标准矩,
3 阶标准矩易于计算且受噪声影响更小.

③ 不是度量偏度唯一有用的特征数

e.g.



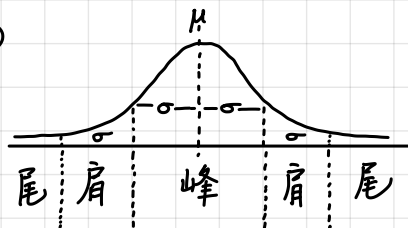
偏度系数 < 0
称为负偏, 左偏

• 峰度系数: $E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right]$

Remark: ① 正态分布的峰度系数 $\equiv 3$.

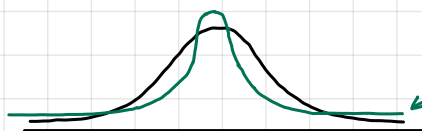
超值(超额)峰度 $\triangleq E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right] - 3$

②



e.g. 金融机构可能更关注尾(小概率情形)

③



峰度 > 3 : 尖峰厚尾

④ 没有唯一的数字特征可以完美地描述尾部行为

⑤ 实际应用中很少涉及 4 阶以上的矩

6° 矩母函数 (Moment Generating Function, MGF)

定义: 记 $M_X(t) \triangleq E(e^{tX})$

若 $M_X(t)$ 在 $t=0$ 的某个邻域内存在, 则称 $M_X(t)$ 是 X 的矩母函数.

e.g. $X \sim \text{Exp}(\lambda)$

$$\text{则 } M_X(t) = E(e^{tX}) = \int_0^{+\infty} e^{tx} \cdot (\lambda e^{-\lambda x}) dx = \frac{\lambda}{\lambda - t}, \quad t < \lambda, \text{ MGF 存在}$$

$X \sim N(0, 1)$

$$\text{则 } M_X(t) = E(e^{tX}) = \int_{\mathbb{R}} e^{tx} \cdot e^{-\frac{x^2}{2}} dx = e^{\frac{t^2}{2}}, \quad t \in \mathbb{R}, \text{ MGF 存在}$$

性质: (1) $M_X(0) = 1$

(2) 设 $Y = aX + b$, 则 $M_Y(t) = E(e^{tY})$

$$= E(e^{t(aX+b)})$$

$$= e^{tb} M_X(at)$$

e.g. 设 $Y \sim N(\mu, \sigma^2)$, 则 $\frac{Y-\mu}{\sigma} = X \sim N(0, 1) \Rightarrow Y = \sigma X + \mu$

$$\text{则 } M_Y(t) = e^{\frac{\sigma^2}{2}t^2 + \mu t}$$

Poisson(λ)

$$M(t) = e^{\lambda(e^t - 1)}$$

• 矩母函数确定矩 // 用矩母函数求高阶原点矩: 求导

定理 $E(X^n) = M_X^{(n)}(0)$ (在一定较好条件下成立)

$$\text{证明 } M_X(t) = \sum_{n=0}^{\infty} M_X^{(n)}(0) \frac{t^n}{n!}$$

$$\text{而 } M_X(t) = E(e^{tX}) = E\left(\sum_{n=0}^{\infty} \frac{X^n t^n}{n!}\right) \stackrel{\text{一定条件下}}{=} \sum_{n=0}^{\infty} E(X^n) \frac{t^n}{n!} \quad \square$$

e.g. 设 $X \sim N(0, 1)$

$$\text{则 } M_X(t) = e^{\frac{t^2}{2}} = \sum_{n=0}^{\infty} \frac{(\frac{t^2}{2})^n}{n!} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^n n!} \cdot \frac{t^{2n}}{(2n)!}$$

$$\Rightarrow \begin{cases} E(X^{2n}) = \frac{(2n)!}{2^n n!} = \frac{(2n)!}{(2n)!!} = \frac{(2n)(2n-2) \cdot 2 \cdot (2n-1)(2n-3) \cdots 1}{(2n)(2n-2) \cdots 2} = (2n-1)!! \\ E(X^{2n+1}) = 0 \end{cases}$$

注: 例 $E(X^2) = 1, E(X^3) = 0, E(X^4) = 3$

• 矩母函数确定分布

定理 若存在 $a > 0$, s.t. $M_X(t) = M_Y(t), \forall t \in (-a, a)$, 则 X, Y 同分布

$$\text{e.g. 设 } M_X(t) = \frac{1}{4}e^{-t} + \frac{1}{2} + \frac{1}{8}e^{4t} + \frac{1}{8}e^{5t}$$

对于离散型 X , 设 $P(X=k) = p_k$, 则 $E(e^{tX}) = \sum e^{tk} p_k$

$$\Rightarrow \begin{array}{c|ccccc} X & -1 & 0 & 4 & 5 \\ \hline P & \frac{1}{4} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} \end{array}$$

Remark: 矩不能确定分布.

e.g. 设 $\ln X \sim N(0,1)$

$$X_1: \text{pdf}: f_1(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln x)^2}{2}}$$

$$X_2: \text{pdf}: f_2(x) = f_1(x) [1 + \sin(2\pi \ln x)]$$

注意到 $\int_0^{+\infty} x^n f_1(x) \sin(2\pi \ln x) dx \equiv 0$.

$$\text{有 } E(X_1^n) = E(X_2^n)$$

$$E(X_1^n) = \int_0^{+\infty} x^n f_1(x) dx$$

$$E(X_2^n) = \int_0^{+\infty} x^n f_2(x) dx$$

$$= \int_0^{+\infty} x^n f_1(x) dx$$

$$+ \int_0^{+\infty} x^n \sin(2\pi \ln x) f_1(x) dx$$

• 独立随机变量和的分布

定理 设 X, Y 独立, $Z = X + Y$, 则 $M_Z(t) = M_X(t) \cdot M_Y(t)$

证明 $M_Z(t) = E(e^{t(X+Y)}) \Rightarrow E(e^{tX}) E(e^{tY}) = M_X(t) M_Y(t)$

$$\cdot Z = \sum_{i=1}^n X_i \Rightarrow M_Z(t) = \prod_{i=1}^n M_{X_i}(t)$$

e.g. 设 $X_1, \dots, X_n$ 是独立的正态, 则 $\sum_{i=1}^n X_i$ 也是正态.

$n=2$ 时, 设 $X_i \sim N(\mu_i, \sigma_i^2), i \in \{1, 2\}$

$$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t) = e^{\frac{\sigma_1^2}{2}t^2 + \mu_1 t + \frac{\sigma_2^2}{2}t^2 + \mu_2 t} = \exp\left\{\frac{\sigma_1^2 + \sigma_2^2}{2}t^2 + (\mu_1 + \mu_2)t\right\}$$

而这是 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 的 MGF, MGF 确定分布.

$$\Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

归纳可知结论正确.

Remark

① 联合分布 $(X_1, \dots, X_n)$ 的矩母函数

$$M_{X_1, \dots, X_n}(t_1, \dots, t_n) = E(e^{t_1 X_1 + \dots + t_n X_n})$$

• 若 $\exists p > 0, s.t. M_{X_1, \dots, X_n} = M_{Y_1, \dots, Y_n} \forall x \in B(0, p)$, 则 $(X_1, \dots, X_n)$ 与 $(Y_1, \dots, Y_n)$ 同分布

② 概率母函数 (PGF), X 为非负整数取值, 设 $p_k = P(X=k)$

$$\text{PGF: } E(t^X) = \sum_{k=0}^{\infty} p_k t^k, t \in [-1, 1]$$

$$\text{当 } t \in (0, 1] \text{ 时 } E(t^X) = E(e^{X \ln t}) = M_X(\ln t)$$

③ 特征函数: $E(e^{itX})$ 其中 $i^2 = -1$.

优势: 没有存在性问题

④ 矩母、特征分别对应于 Laplace 变换, Fourier 变换

7° 条件期望

$$E(Y|X \in A) = \begin{cases} \sum y_i P(Y=y_i|X \in A), & Y \text{ 为离散型} \\ \int_{\mathbb{R}} y f_Y(y|X \in A) dy, & Y \text{ 为连续型} \end{cases}$$

$$E(Y|x) := E(Y|X=x) = \begin{cases} \sum y_i P(Y=y_i|X=x), & Y \text{ 为离散型} \\ \int_{\mathbb{R}} y f_Y(y|x) dy, & Y \text{ 为连续型} \end{cases}$$

$E(Y|X)$ 是 X 的函数, 一个新的随机变量, 称为 Y 对 X 的回归函数

e.g. $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$\text{则 } E(Y|x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1)$$

e.g. 甲乙是两种同类产品, 平均寿命分别是 10 年, 15 年, 市场占有率分别是 60%, 40%. 随机买一个, 寿命的期望是?

naïve: $12 = 0.6 \times 10 + 0.4 \times 15$

记 $X = \begin{cases} 1, & \text{甲} \\ 2, & \text{乙} \end{cases}$, Y 是买到产品的寿命.

$$\begin{aligned} E(Y) &= E(Y|X=1)P(X=1) + E(Y|X=2)P(X=2) \\ &= E(E(Y|X)) \end{aligned}$$

• 定理 (全期望公式) $E(Y) = E(E(Y|X))$

证明 以连续型为例

$$E(Y|x) = \int_{\mathbb{R}} y f_Y(y|x) dy = \int_{\mathbb{R}} y \cdot \frac{f_{X,Y}(x,y)}{f_X(x)} dy$$

$$\Rightarrow E(Y|x) f_X(x) = \int_{\mathbb{R}} y f_{X,Y}(x,y) dy$$

$$\Rightarrow E(E(Y|X)) = \int_{\mathbb{R}} E(Y|x) f_X(x) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y f_{X,Y}(x,y) dy \right) dx = E(Y)$$

Remark. 一般地, $E(g(X, Y)) = E[E(g(X, Y)|X)]$

多维类似

多元情形可对某个变量做全期望

• 定理 $E[(Y - g(X))^2] \geq E[(Y - E(Y|X))^2]$

理解 注意到 $\forall c \in \mathbb{R}$.

$$E[(Y - c)^2] \geq E[(Y - E(Y))^2]$$

$$\text{故 } E[(Y - g(X))^2|X] \geq E[(Y - E(Y|X))^2|X]$$

两边取期望即得上式.

Remark. $E(Y|X)$ 依赖 (X, Y) 联合分布, 通常难以获得

转而求最优线性预测 (i.e. $\min_{a,b} E[(Y - (aX + b))^2]$)

这是最小二乘问题, a, b 与 $E(X), E(Y), \text{Var}(X), \text{Var}(Y), \text{Corr}(X, Y)$ 这些低阶数字特征有关

$\hat{Y} \triangleq E(Y|X)$, 估计误差: $\tilde{Y} = \hat{Y} - Y$. 有

① $E(\tilde{Y}) = 0$, 无偏性

② $E(\tilde{Y}\hat{Y}) = 0 \Rightarrow \text{Cor}(\hat{Y}, \tilde{Y}) = E(\hat{Y}\tilde{Y}) - E(\hat{Y})E(\tilde{Y}) = 0 - 0 = 0$
 $\Rightarrow \text{Var}(Y) = \text{Var}(\hat{Y}) + \text{Var}(\tilde{Y})$

Lecture 8

Chapter 5

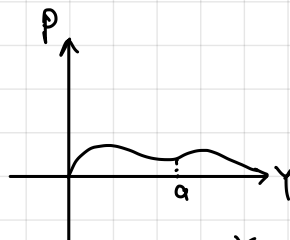
1° 概率不等式

• Markov 不等式: $Y \geq 0$, 则 $\forall a > 0$,

$$P(Y \geq a) \leq \frac{E(Y)}{a}$$

证明: 令 $I = \begin{cases} 1, & Y \geq a \\ 0, & Y < a \end{cases} \Rightarrow I \leq \frac{Y}{a} \Rightarrow E(I) \leq E\left(\frac{Y}{a}\right) \Rightarrow P(Y \geq a) \leq \frac{E(Y)}{a}$

intuition: $Y \geq a \rightarrow$ 缩到 a
 $Y < a \rightarrow$ 缩到 0



Remark. $P(|X| \geq a) \leq \frac{E(|X|)}{a}$

eg. 设 $X \geq 0$, $P(X \geq kE(X)) \leq \frac{1}{k}$

• Chebyshev 不等式: 设 $\text{Var}(Y)$ 存在, 则对 $\forall a > 0$,

$$P(|Y - E(Y)| \geq a) \leq \frac{\text{Var}(Y)}{a^2}$$

证明: $P(|Y - E(Y)| \geq a)$

$$= P((Y - E(Y))^2 \geq a^2)$$

$$\stackrel{\text{Markov}}{\leq} \frac{E((Y - E(Y))^2)}{a^2}$$

$$= \frac{\text{Var}(Y)}{a^2}$$

Remark. $\text{Var}(Y) = 0 \Rightarrow P(Y = E(Y)) = 1$

• Chernoff 不等式: $\forall a > 0, \forall t > 0$, 有

$$P(Y \geq a) \leq \frac{E(e^{tY})}{e^{ta}}$$

证明: 对 $\forall t > 0$,

$$P(Y \geq a) = P(e^{tY} \geq e^{ta})$$

$$\stackrel{\text{Markov}}{\leq} \frac{E(e^{tY})}{e^{ta}}$$

e.g. 设 $X \sim N(0,1)$

$$\text{考查 } P(|X| \geq 3) : \begin{cases} \text{Markov: } \leq \frac{E(|X|)}{3} = \frac{1}{3} \sqrt{\frac{2}{\pi}} \approx 0.27 \\ \text{Chebyshev: } \leq \frac{1}{9} \approx 0.11 \\ \text{Chernoff: } = 2P(X \geq 3) \leq 2 \cdot \frac{e^{-\frac{t^2}{2}}}{e^{3t}} = 2e^{\frac{t^2}{2} - 3t} \leq 2 \cdot e^{-\frac{9}{2}} \approx 0.022 \end{cases}$$

而正态经验法则易计算得 $P(|X| \geq 3) \approx 0.003$

2° 大数定律 (Law of Large Numbers, LLN)

$X_1, \dots, X_n$, iid (independent identical distribution), $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2 > 0$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, E(\bar{X}) = \mu, \text{Var}(\bar{X}) = \frac{1}{n} \sigma^2 \quad (\text{Var}(\bar{X}) \rightarrow 0, n \rightarrow +\infty)$$

随 n 增大, $\bar{X}$ 越来越趋于接近 μ

• 定理 (Khinchin 弱大数定律)

// 荐读: Khinchin, 数学分析讲义

设 iid $X_1, \dots, X_n$, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$
则对 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P(|\bar{X} - \mu| \geq \varepsilon) = 0$$

也即

$$\lim_{n \rightarrow +\infty} P(|\bar{X} - \mu| < \varepsilon) = 1$$

$$\text{证明: } P(|\bar{X} - \mu| \geq \varepsilon) \stackrel{\text{Chebyshev}}{\leq} \frac{\text{Var}(\bar{X})}{\varepsilon^2} \rightarrow 0$$

Remark: ① $\forall \varepsilon > 0, \forall \alpha > 0 \exists N$ s.t. $n \geq N$ 时

精度 置信度

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \alpha \text{ 即 } P(|\bar{X} - \mu| < \varepsilon) \geq 1 - \alpha$$

② μ (未知) 很大概率意义上 $\approx \bar{X}$

③ 设 $X_i \sim B(p) \iff$ Bernoulli 大数定律

④ 关于方差的条件不是必要的

$\text{Var}(X_i)$ 不存在时, 定理也成立

⑤ 该定理可推广

如 Chebyshev: X_i, X_j 两两不相关, $\text{Var}(X_i)$ 一致有界

Markov: $\text{Var}(\bar{X}) \rightarrow 0$

Remark: 依概率收敛

$$Y_n \xrightarrow{P} Y \iff \lim_{n \rightarrow \infty} P(|Y_n - Y| \geq \varepsilon) = 0, \forall \varepsilon > 0$$

不要收敛
偏差值的
概率测度

• 定理 (Kolmogorov 强大数定律)

设 iid $X_1, \dots, X_n$, $E(X_i) = \mu$, 则

$$P(\lim_{n \rightarrow \infty} \bar{X} = \mu) = 1$$

// 在 Ω 中几乎处处满足 $\lim_{n \rightarrow \infty} \bar{X} = \mu$

即 $\forall \varepsilon > 0$

$$P(\lim_{n \rightarrow \infty} |\bar{X} - \mu| < \varepsilon) = 1$$

Remark. 设 $X_i \sim B(p)$

该定理说明概率的频率解释是合理

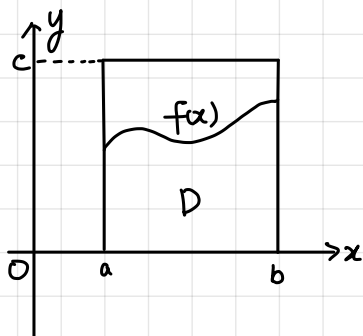
Remark. ① 以概率 1 收敛: 几乎必然收敛 (almost surely converge)

$$Y_n \xrightarrow{a.s.} Y \iff P(\lim_{n \rightarrow \infty} Y_n = Y) = 1 \iff \forall \varepsilon > 0, P(\lim_{n \rightarrow \infty} |Y_n - Y| < \varepsilon) = 1$$

② 强大数定律: $\bar{X} \xrightarrow{a.s.} \mu$

a.s. 收敛要求先有收敛
这是在用概率的测度进行评估

e.g. (Monte Carlo 积分)



设 (X_i, Y_i) iid. $\sim U([a, b] \times [0, c])$

$$\text{令 } I_i = \begin{cases} 1, & (X_i, Y_i) \in D \\ 0, & \text{其他} \end{cases} \Rightarrow I_i \sim B(p) \text{ 而 } p = \frac{\int_a^b f(x) dx}{c(b-a)}$$

$$\Rightarrow \int_a^b f(x) dx \approx c(b-a) \cdot \left(\frac{1}{n} \sum_{i=1}^n I_i \right)$$

e.g. 两种收敛的区别

设 $\Omega = [0, 1]$, 均匀分布 (从而有 $\Omega, \mathcal{F}, P$)

设 $s \in \Omega$

$$Y_1(s) = s + I_{[0, 1/2]}(s)$$

$$Y_2(s) = s + I_{[0, 1/2]}(s)$$

$$Y_3(s) = s + I_{[1/2, 1]}(s)$$

$$Y_4(s) = s + I_{[0, 1/3]}(s)$$

$$Y_5(s) = s + I_{[1/3, 2/3]}(s)$$

$$Y_6(s) = s + I_{[2/3, 1]}(s)$$

⋮

$$Y(s) = s$$

$$\text{令 } Z_n(s) = s^n$$

$$Y_n \xrightarrow{P} Y, \text{ 但 } Y \not\xrightarrow{a.s.} Y_n, \text{ 因为 } \lim_{n \rightarrow \infty} Y_n \text{ 不存在}$$

3° 中心极限定理 (CLT)

• 定理 设 $X_1, \dots, X_n$ iid, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2 > 0$, 则

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x)$$

其中 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, 是 $N(0,1)$ 的 cdf.

// 趋向过程中, 在 $n\mu$ 附近怎样分布

证明 只证 $M_{X_i}(t)$ 存在的情形

$$\text{令 } M(t) = E(e^{tX_i})$$

(未看)

WLOG, 令 $\mu=0, \sigma^2=1$, 则 $M(0)=1, M'(0)=0, M''(0)=1$

$$E\left[\exp\left\{t \frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right\}\right] \xrightarrow{\text{独立性}} M^n\left(\frac{t}{\sqrt{n}}\right)$$

$$\lim_{n \rightarrow \infty} n \ln M\left(\frac{t}{\sqrt{n}}\right) \xrightarrow{y=\frac{t}{\sqrt{n}}} \lim_{y \rightarrow 0} \frac{\ln M(yt)}{y^2} \xrightarrow{\text{L'Hospital}} \frac{t^2}{2}$$

$$M_{\frac{\sum X_i}{\sqrt{n}\sigma}} \rightarrow e^{\frac{t^2}{2}} \quad (n \rightarrow \infty)$$

Remark. ① 上述定理常被称为 Lindeberg CLT. 可在不同分布时推广

② 依分布收敛: 只衡量随机变量的 cdf 到

$$Y_n \xrightarrow{d} Y \iff \text{cdf. } F_{Y_n}(x) \rightarrow F_Y(x) \quad (n \rightarrow \infty)$$

③ 设 $Z_n := \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ (对样本均值做标准化)

$$\text{CLT} \implies Z_n \xrightarrow{d} Z, Z \sim N(0,1), \text{也记为 } Z_n \rightarrow N(0,1)$$

$$\text{④ CLT} \implies \begin{aligned} \sum_{i=1}^n X_i &\rightarrow N(n\mu, n\sigma^2) \\ \bar{X} &\rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \end{aligned}$$

⑤ $X_i \sim B(p)$, 则 $\sum_{i=1}^n X_i \sim B(n, p) \rightarrow \text{Normal}$

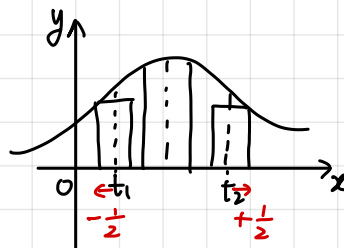
此时 CLT 称为 De Moivre - Laplace CLT

$$P(t_1 \leq \sum_{i=1}^n X_i \leq t_2) \approx \Phi(y_2) - \Phi(y_1)$$

$$y_1 = \frac{t_1 - np - \frac{1}{2}}{\sqrt{np(1-p)}}$$

$$y_2 = \frac{t_2 - np + \frac{1}{2}}{\sqrt{np(1-p)}}$$

连续性修正



io 在相应的 $N(np, np(1-p))$ 中, $(t_2 + \frac{1}{2})$ 处于什么位置.

e.g. (选举问题) 设 p 为选民支持率(未知) $// p_n \rightarrow N(\mu, \frac{\sigma^2}{n})$
 随机调查 n 个人, 支持比例 $p_n = \frac{1}{n} \sum_{i=1}^n \delta_i$, $\delta_i \sim B(p)$, iid δ_i (因为 $n \ll N$)
 若 $\varepsilon = 0.03$, $1 - \alpha = 0.95$, $n = ?$

$$\text{即 } P(|p_n - p| \geq \varepsilon) \leq \alpha$$

$$// -\varepsilon \leq p_n - p \leq \varepsilon$$

$$// \mu = 0$$

$$// \sigma^2 = \frac{p(1-p)}{n}$$

$$\text{用 CLT: } P(|p_n - p| \geq \varepsilon) \stackrel{\text{CLT}}{\approx} 2(1 - \phi(\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}})) \leq \alpha$$

$$\Leftrightarrow \phi(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}) \geq 1 - \frac{\alpha}{2}$$

为使所有 p 成立, 左边取最小值令 $p = \frac{1}{2}$,

即有 p 是未知数

$$\phi(2\varepsilon\sqrt{n}) \geq 1 - \frac{\alpha}{2}$$

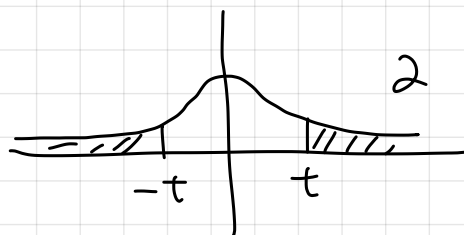
$$\text{注意到 } \phi(1.96) \approx 0.975 = 1 - \frac{\alpha}{2}$$

取 $n \geq 1068$ 即可 (这与 N 无关)

$$P(|p_n - p| \geq \varepsilon) \leq 1 - \alpha$$

$$P\left(\left|\frac{p_n - p}{\sqrt{\frac{p(1-p)}{n}}}\right| \geq \frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) \leq \alpha$$

$\downarrow$ $\downarrow$
 $N(0,1)$ t



$$2(1 - \phi(t)) \leq \alpha$$

$$2 - 2\phi(t) \leq \alpha$$

$$2\phi(t) \geq 2 - \alpha$$

$$\phi(t) \geq 1 - \frac{\alpha}{2}$$

$$\phi(\sqrt{4n}\varepsilon) \geq 1 - \frac{\alpha}{2}$$

$$\phi\left(\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) \geq 1 - \frac{\alpha}{2}$$