Notes: 概率论与数理统计

黄宇翔
清华大学计算机系

2023 年 6 月 1 日

1 概率基本概念及事件运算

定义 1.1 (古典概型/等可能概型)**.** 如果每个基本事件都等可能出现，此时某事件的概率是 $P(A)=\frac{\text{事件 }A\text{ 包含的基本事件数}}{\text{全部可能的基本事件数}}$ 或 $P(A)=\frac{\text{事件 }A\text{ 所占区域大小}}{\text{样本空间所占区域大小}}$。

例 1.1 (生日问题)**.** 求 $n(n\leq 365)$ 个人的生日各不相同的概率 p.

只考虑一年 365 天的情况。$p_n=\frac{365\times(365-1)\ldots(365-n+1)}{365^n}$. $p_{10}=0.883, p_{30}=0.2937$.

估值：$p_n=(1-\frac{1}{365})\ldots(1-\frac{n-1}{365}), \ln(1-x)=-x$ 近似 ,$\ln p_n=-\frac{1+\cdots+(n-1)}{365}=-\frac{n(n-1)}{730}, p_n=e^{-\frac{n(n-1)}{730}}$.

例 1.2. 甲乙约定在下午 *4-5* 点见面，并约定等候时间是 *20* 分钟，过时就离开。求二人的会面概率。

有约束条件 $|x-y|\leq 20, 0\leq x,y,\leq 60$（可以画图，其中 $x-20\leq y\leq x+20$），则 $p=\frac{S_A}{S_\Omega}=\frac{5}{9}$。

例 1.3. 组合公式的直观解释举例

$$\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$$

可以看作 n 个小球，染成红、蓝两种颜色，一共有 2^n 种染色方式；也可以看作 n 名同学，挑选几人去完成一项任务，有 2^n 种选派方式。

$$\sum_{k=1}^{n}k\binom{n}{k}=\binom{n}{1}+2\binom{n}{2}+\cdots+n\binom{n}{n}=n2^{n-1}$$

可以看作 n 名同学，随机挑选几人完成任务，再从选出的人中选出一名队长，总共有多少选择方式。等式左边：先选小组，再选组长。等式右边：先选队长（n 个人选），再选队员（另 $n-1$ 人是否参加）。

定义 1.2 (样本和样本空间). 随机事件的一切可能基本结果组成的集合称为样本空间，每一种可能的基本结果叫做样本（点）。

例 1.4. 将一枚硬币抛两次：$\Omega =$ 正正，正反，反正，反反.

定义 1.3 (随机事件). 样本空间的一个子集 $A \subset \Omega, w \in A$ 发生，$w \notin A$ 未发生。

可以使用 Venn 图来刻画随机事件。

例 1.5. 使用 *Venn* 图求证 $(A-B)\bar{C} = A\bar{C} - B$.

定义 1.4 (示性函数). $I_A(w) = 1(w \in A); I_A(w) = 0(w \notin A)$

事件的关系：

- 包含关系：$A \subset B$，事件 B 包含事件 A，$I_A(w) \le I_B(w)$
- 相等关系：$A = B, I_A(w) = I_B(w)$
- 互不相容：$A \cap B = \emptyset$，$I_A(w)I_B(w) = 0$
- 对立事件：$B = \bar{A}, I_A(w) + I_B(w) = 1$

概率的基本性质：

$$P(\emptyset) = 0, P(\bar{A}) = 1 - P(A), P(A-B) = P(A) - P(AB)$$

例 1.6. 袋中有编号为 $1, \ldots, n$ 的 n 个球，从中有放回地取 m 次。求 m 个球中最大编号恰为 k 的概率。

设事件 A_k 表示最大号码恰好为 k，事件 B_k 表示最大号码不超过 k. $A_k = B_k - B_{k-1}, B_{k-1} \subset B_k$，所以 $P(A_k) = P(B_k) - P(B_{k-1})$.

$$P(B_k) = \frac{k^m}{n^m}, P(A_k) = \frac{k^m - (k-1)^m}{n^m}.$$

例 1.7 (匹配问题). n 封信随机放到 n 个信封中，求所有信件都装错了的概率。

设 A_i 表示第 i 封信装对了信封，则所求事件是 $\overline{A_1 \cup \cdots \cup A_n}$.

$$P(A_1 \cup \cdots \cup A_n) = 1 - \frac{1}{2!} + \cdots + (-1)^{n-1}\frac{1}{n!} \to 1 - e^{-1}$$

$$p_0 \to e^{-1} = 0.37$$

例 1.8 (布丰投针问题). 在平面上有一组间距为 d 的平行线，将一根长度为 $l < d$ 的针任意投掷在平面上，求相交的概率。

设针与直线的夹角是 θ，针中点到最近直线的距离是 x, 只有 $0 \le x < \frac{l}{2}\sin\theta$ 时才相交，其中 $0 \le \theta \le \pi, 0 \le x \le \frac{d}{2}$。

那么概率是

$$p = \frac{\int_0^\pi \frac{l}{2}\sin\theta d\theta}{\frac{d}{2}\pi} = \frac{2l}{\pi d}.$$

2 条件概率和独立性

定义 2.1 (条件概率)**.** 设 A, B 是两个事件 b，$P(B) > 0$, 定义

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

命题 2.1 (乘法公式)**.** 设 $P(B) > 0$，则有

$$P(AB) = P(B)P(A|B)$$

$$P(A_1A_2\dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\dots P(A_n|A_{n-1}\dots A_1).$$

证:使用数学归纳法。设 $P(A_1\dots A_k) = P(A_1)\dots(A_k|A_{k-1}\dots A_1)$, 则 $P(A_1\dots A_kA_{k+1}) = P(A_1\dots A_k)P(A_{k+1}|A_1\dots A_k) = P(A_1)\dots P(A_{k+1}|A_1\dots A_k)$.

例 2.1. 设箱子内有 a 个白球和 b 个黑球，在其中不放回地连续取 *3* 次，计算前两次取到黑球而第三次取到白球的概率，应该使用如下算式。

$$P(\bar{A}_1\bar{A}_2A_3) = P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(A_3|\bar{A}_1\bar{A}_2)$$

需要按照取球的顺序展开，这样等式右边的概率更容易计算。

命题 2.2 (全概率公式)**.** 设 $B_1, \dots, B_n$ 是样本空间 Ω 的一个分割，即 $B_1, \dots, B_n$ 互不相容且并集为 Ω，如果 $P(B_i) > 0, \forall i = 1, \dots, n$，那么

$$P(A) = \sum_{i=1}^n P(B_i)P(A|B_i).$$

证：$P(A) = P(A\Omega) = P(A \cap (B_1 \cup \dots \cup B_n)) = P(AB_1 \cup \dots \cup AB_n) = P(AB_1) + \dots + P(AB_n) = \sum_{i=1}^n P(B_i)P(A|B_i)$.

一种特殊情况：$P(A) = P(W)P(A|W) + P(\bar{W})P(A|\bar{W})$.

例 2.2. 设甲箱有 a, b 个白，黑球；乙箱有 c, d 个白，黑球。从甲箱任取一球放入乙箱，然后再从乙箱任取一球，求最后由乙箱取出的是白球的概率。

解: 设事件 A, B 表示最后由乙箱取出的是白球，从甲箱取出了一个白球。那么 $P(A) = P(W)P(A|W) + P(\bar{W})P(A|\bar{W}) = \frac{a}{a+b}\frac{c+1}{a+b+1} + \frac{b}{a+b}\frac{c}{c+d+1} = \frac{ac+bc+a}{(a+b)(c+d+1)}$.

例 2.3. 买彩票，设 n 张彩票中有一张奖券，人们排成一队买票，求第 k 个人买到奖券的概率。

解：$A_k(n): n$ 个人买票，第 k 个人中奖。（如果第一个人没获奖，那么问题规模减小，变成有 $n-1$ 张彩票，$k-1$ 个人买票。）

$$\begin{aligned}P(A_k(n)) &= P(A_1(n))P(A_k(n)|A_1(n)) + P(\overline{A_1(n)})P(A_k(n)|\overline{A_1(n)}) \\ &= 0 + \frac{n-1}{n}P(A_{k-1}(n-1)) = \frac{n-1}{n}\frac{n-2}{n}P(A_{n-1}(n-2)) \\ &\frac{n-1}{n}\cdots\frac{n-k+1}{n-k+1}\frac{1}{n-k+1} = \frac{1}{n}.\end{aligned}$$

命题 2.3 (Bayes 公式)**.** 设 $B_1, \dots, B_n$ 是样本空间 Ω 的一个分割，即 $B_1, \dots, B_n$ 互不相容且并集为 Ω，如果 $P(B_i) > 0, \forall i = 1, \dots, n$，那么

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}.$$

证：$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1)+\cdots+P(B_n)P(A|B_n)}$.

例 2.4. 考生做题 *4* 选 *1*，会做的概率是 p，不会的时候任选。求回答正确时他确实会做的概率。

解：设事件 A, B 为“答对”,“会做”，那么有

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} = \frac{p \times 1}{p \times 1 + (1-p) \times \frac{1}{4}} = \frac{4p}{1+3p}.$$

例 2.5. 某疾病的发病率是 *0.0004*。现在有一种化验方法，对真正患病的人，其化验结果 *99%* 是阳性，对正常人，*99.9%* 化验结果是阴性。求：*1*）假阳性（占所有阳性），*2*）假阴性（占所有阴性）的概率。

解：1) 设事件 A: 阳性，B：患病。

$$P(\bar{B}|A) = \frac{P(\bar{B})P(A|\bar{B})}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} = 0.716$$

2)

$$P(B|\bar{A}) = \frac{P(B)P(\bar{A}|B)}{P(B)P(\bar{A}|B) + P(\bar{B})P(\bar{A}|\bar{B})} = 4 \times 10^{-6}.$$

定义 2.2 (事件的独立性)**.** 称事件 A, B 独立，当

$$P(A|B) = P(A), P(B|A) = P(B).(P(A) \neq 0, P(B) \neq 0)$$

另一种定义：如果 $P(AB) = P(A)P(B)$ 时，称 A, B 独立。此时可以有 $P(A) = 0$ 或 $P(B) = 0$。否则称 A, B 不独立或相关。

注意：独立和互斥没有任何关系，互斥是 $P(A)P(B)=0$.

定义 2.3 (事件的独立性)**.** *若 A,B,C 三个事件两两独立，且*

$$P(ABC)=P(A)P(B)P(C)$$

则称事件 A,B,C 独立。

注意：三个事件两两独立不能推出三个时间独立。反例：四面体，三个面分别是红黄蓝纯色，另一个面是红黄蓝混色，则掷出三种颜色不独立，尽管两两独立。

命题 2.4. *若事件 A,B 独立，那么 $A,\bar{B}$ 独立，$\bar{A},B$ 独立，$\bar{A},\bar{B}$ 独立。*

证：只证 $A,\bar{B}$ 独立。

$$P(A\bar{B})=P(A-B)=P(A)-P(A)P(B)=P(A)(1-P(B))=P(A)P(\bar{B}).$$

例 2.6. *甲乙轮流掷骰子，计算点数和，甲掷 6 点胜，乙掷 7 点胜。求甲、乙的获胜概率。*

解：掷出 6 点概率为 5/36，7 点概率为 $6/36=1/6$. 记甲第 k 轮掷出 6 点为 A_k，乙第 k 轮掷出 7 点为 B_k.

$$\begin{aligned}P(\text{甲获胜})&=P(A_1\cup\bar{A}_1\bar{B}_1A_2\cup\bar{A}_1\bar{B}_1\bar{A}_2\bar{B}_2A_3\cup\dots)\\&=P(A_1)+P(\bar{A}_1\bar{B}_1A_2)+P(\bar{A}_1\bar{B}_1\bar{A}_2\bar{B}_2A_3)+\dots\\&=P(A_1)+P(\bar{A}_1)P(\bar{B}_1)P(A_2)+P(\bar{A}_1)P(\bar{B}_1)P(\bar{A}_2)P(\bar{B}_2)P(A_3)+\dots\\&=\frac{5}{36}\left(1+\frac{31}{36}\frac{5}{6}+1+\frac{31}{36}\left(\frac{5}{6}\right)^2+\dots\right)=\frac{30}{61}.\end{aligned}$$

定义 2.4 (相关系数)**.**

$$\tau(A,B)=\frac{P(AB)-P(A)P(B)}{\sqrt{P(A)(1-P(A))P(B)(1-P(B))}}$$

相关系数满足：A,B 独立时是 0，A,B 相等时是 1，取值绝对值在 0 到 1 之间。

3 随机变量

3.1 离散型随机变量

定义 3.1 (随机变量)**.** *设 Ω 是样本空间，则定义在样本空间上的函数就成为随机变量*

$$X:\Omega\to\mathbb{R}$$

通常用大写字母 X,Y,Z 表示随机变量，x,y,z 等小写字母表示其取值。$x=X(\omega),\omega\in\Omega$，随机变量 $X(\omega)$ 可以简记为 X.

例 3.1. 扔硬币，正面记 H，背面记 T.

扔一次 $\Omega = \{H,T\}$, $X(H) = 0, X(T) = 4$.

$$P(X = 0) = 1/2, P(X = 4) = 1/2, P(X = x) = 0(x \neq 0, 4)$$

定义 3.2 (随机变量的分布函数)**.** 设 X 是一个随机变量，对任意实数 x，定义

$$F(x) = P(X \leq x)$$

为随机变量 X 的分布函数，且称 X 服从 $F(x)$，记为 $X \sim F(x)$ 或 $F_X(x)$.

容易有

$$\lim_{x\to-\infty} = 0, \lim_{x\to+\infty} = 1;$$
$$F(b) - F(a) = P(a < X \leq b), F(b) - F(a) + P(X = a) = P(a \leq X \leq b).$$

如果随机变量 X 所有可能的取值是有限或可列多个，则其分布可表示为

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots x_n & \cdots \\ p_1 & p_2 & \dots p_n & \dots \end{pmatrix}$$

定义 3.3 (两点分布 (Bernoulli 分布))**.**

$$X \sim \begin{pmatrix} a_0 & a_1 \\ p & q \end{pmatrix}, p + q = 1, p, q \geq 0.$$

定义 3.4 (二项分布)**.** 将 *Bernoulli* 试验独立重复 n 次，比如连续投掷 n 次硬币等，基本结果有 2^n 种，X 为 A（正结果）出现的次数，则 X 的分布为

$$X \sim \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix}, p_k = \binom{n}{k} p^k q^{n-k}, k = 0, 1, \dots, n.$$

此分布称为二项分布，记为 $X \sim b(n, p)$，p_k 为 $(p + q)^n$ 的二项展开系数。

例 3.2. 假设每台自动车床在一段运行时段内需要维修的概率是 $p = 0.01$，不同机床独立，求

1）1 名维修工人看管 20 台机器，不能及时维修的概率；

2）3 名维修工人看管 80 台机器，不能及时维修的概率。

1）$X \sim b(20, 0.01)$，不能及时维修的概率为

$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - 0.99^{20} - 20 \times 0.01 \times 0.99^{19} = 0.017.$

2）$X \sim b(80, 0.01)$，不能及时维修的概率为

$$P(X > 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

$$= 1 - 0.99^{80} - 80 \times 0.01 \times 0.99^{79} - \frac{80 \times 79}{2} \times 0.01^2 \times 0.99^{78} = 0.047$$

定义 3.5 (泊松 (Poisson) 分布)**.**

$$X \sim \begin{pmatrix} 0 & 1 & 2 & \dots \\ p_0 & p_1 & p_2 & \dots \end{pmatrix}, p_k = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, 2, \dots.$$

记作 $X \sim P(\lambda)$. 当二项分布 $b(n,p)$ 中的 p 很小 n 很大时，$b(n,p)$ 与 $P(np)$ 非常接近。

概率接近的含义：$P(X = k) \approx P(Y = k)$.

定理 3.1 (泊松定理)**.** *在 n 重伯努利试验中，记事件 A 在一次试验中发生的概率为 p_n。如果 $n \to \infty$ 有 $np_n \to \lambda$，则*

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

证明：略去，在课本 pp.87-88.

例 3.3. *计算 $X \sim b(30000, 0.0001)$ 中的 $P(X = 3)$.*

由于 $b(30000, 0.0001) \approx P(3)$，不妨近似计算。

$$P(X = 3) = e^{-3}\frac{3^3}{3!} = 0.224.$$

例 3.4. *已知某种疾病的发病率是 0.001，某单位有 5000 人，问该单位患有这种疾病不超过 5 人的概率为多少?*

设患病人数为 X，则有 $X \sim b(5000, 0.001)$，所求概率为 $P(X \le 5)$，使用泊松定理近似。

$$P(X \le 5) = \sum_{k=0}^{5} \frac{5^k}{k!} e^{-5} = 0.616.$$

定义 3.6 (几何分布)**.** *考虑伯努利分布 $P(A) = p$，第一次成功的试验次数 $X \sim Ge(p)$*

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

称为几何分布。

几何分布具有无记忆性。

例 3.5 (无记忆性). *进入商场的人有 1/5 的概率进入体育用品商店，如果前 5 个人到商场的人都没有进入这家商店，问到达商场的第 8 个人是第一个进店的概率。*

$$P(X=8|X>5)=\frac{P(X=8)}{P(X>5)}=\frac{0.8^7\times 0.2}{\sum_{k=6}^{+\infty}0.8^{k-1}\times 0.2}=0.128=P(X=3).$$

定理 3.2 (几何分布的无记忆性).

$$P(X>s+t|X>s)=P(X>t)$$

证：$P(X>n)=\sum_{k=n+1}^{+\infty}(1-p)^{k-1}p=p\frac{(1-p)^n}{1-(1-p)}=(1-p)^n$，所以有

$$P(X>s+t|X>s)=\frac{P(X>s+t)}{P(X>s)}=\frac{(1-p)^{s+t}}{(1-p)^s}=(1-p)^t=P(X>t).$$

推论 3.1.

$$P(X=s+t|X>s)=P(X=t).$$

事实 3.1. *无记忆性 $\iff$ 几何分布，也就是如果 X 具有无记忆性，则 X 服从几何分布。*

定义 3.7 (负二项分布). *独立地进行伯努利实验，成功 r 次且最后一次成功的概率是*

$$P(X=k)=\binom{k-1}{r-1}p^r(q-p)^{k-r}.$$

当 $r=1$ 时，$P(X=k)=p(1-p)^{k-1}$，X 服从几何分布。

3.2 连续型随机变量

定义 3.8 (连续型随机变量). *设随机变量 X 的分布函数为 $F(x)$，若存在非负可积函数 $p(x),x\in\mathbb{R}m$，使得 $\forall x\in\mathbb{R}$*

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

则称 X 为连续型随机变量，$p(x)$ 为 X 的概率密度函数，且 $\int_{-\infty}^{\infty}p(x)dx=1$.

Note: 有密度函数的随机变量是连续型随机变量，有分布列的随机变量是离散型随机变量。

定义 3.9 (均匀分布). *$X\sim U(a,b)$,*

$$p(x)=\frac{1}{b-a},x\in[a,b];\ p(x)=0,\ otherwise$$

$$F(x)=0,x<a;\ F(x)=\frac{x-a}{b-a},a\le x<b;\ F(x)=1,x\ge b$$

定义 3.10 (柯西分布 *).

$$p(x)=\frac{1}{\pi}\frac{1}{1+x^2}$$

例 3.6. *某电子元器件的寿命 $X(hour)$ 服从一下分布*

$$p(x)=\frac{1000}{x^2},X>1000;0,otherwise.$$

(1) 任取一只，寿命大于 1500h 的概率

(2) 已知一只元器件使用了 1500h，求再用 500h 的概率

(1)

$$P(X>1500)=\int_{1500}^{+\infty}\frac{1500}{x^2}dx=\frac{2}{3}$$

(2)

$$P(X>2000|X>1500)=\frac{\int_{2000}^{+\infty}\frac{1500}{x^2}dx}{\int_{1500}^{+\infty}\frac{1500}{x^2}dx}=\frac{3}{4}$$

定义 3.11 (指数分布). $X\sim Exp(\lambda),\lambda>0$

$$p(x)=0,x\leq 0;p(x)=\lambda e^{-\lambda x},x>0$$

$$F(x)=0,x\leq 0;F(x)=1-e^{-\lambda x},x>0$$

事实 3.2 (指数分布的无记忆性). *指数分布具有无记忆性，也就是*

$$\forall s>0,t>0,P(Y>s+t|Y>s)=P(Y>t).$$

Note: 指数分布的由来：在任意时间 Δt 内，衰减比率均为 λ, 失效个体 $\lambda\Delta t+o(\Delta t)$. 若 $\lambda_t=\lambda=const.$，则概率分布无记忆性，若 $\lambda_t\neq const.$，则概率分布有记忆性。

例 3.7 (使用无记忆性简化计算). *若某型号的电子元件寿命服从指数分布密度函数*

$$p(x)=\frac{1}{600}e^{\frac{-x}{600}}I_{x\geq 0},$$

若某个元件在使用 1200 小时候仍正常工作，则能再工作 600 小时的概率是?

解:

$$P(X>1800|X>1200)=P(X>600|X>0)$$

$$=P(X>600)=1-F(600)=1-(1-e^{-1})=e^{-1}.$$

定义 3.12 (正态分布). $X\sim N(\mu,\sigma^2)$

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2\right]$$

标准正态分布：$X \sim N(0,1)$,

记号

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

记号

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

.

3σ：在 $\mu - 3\sigma \sim \mu + 3\sigma$ 之间大约有 98% 的数据。

随机变量函数的分布:

$$Y = g(X) \Rightarrow F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

随机模拟生成一般分布:

定理 3.3. 设随机变量 U 服从 $[0,1]$ 上的均匀分布，函数 $F \in \mathcal{C}(\mathbb{R})$ 且满足 F 单调递增，

$$F(-\infty) = 0 \leq F(x) \leq 1 = F(+\infty)$$

则随机变量 $X = F^{-1}(U)$ 的概率分布函数为 $F(x)$.

例 3.8. 利用 $U(0,1)$ 生成服从期望 $1/\lambda$ 的指数分布随机数。

若 $X \sim U(0,1)$，则 $Y = \frac{-\ln X}{\lambda}$ 服从参数为 λ 的指数分布随机数。

当 $x \in (0,1), \frac{-\ln X}{\lambda} > 0$，所以 $F_Y(y) = P(Y \leq y) = 0$.

当 $y > 0, F_Y(y) = P(Y \leq y) = P(\frac{-\ln X}{\lambda} \leq y) = P(\ln X \geq -\lambda y) = P(X \geq e^{-\lambda y}) = 1 - e^{-\lambda y}$.

4 随机变量的数学特征

4.1 期望与方差

定义 4.1 (期望)**.** 随机变量的数学期望:

离散随机变量:

$$E(X) = \sum_{i=1}^{\infty} x_i p(X = x_i)$$

连续随机变量:

$$E(X) = \int_{-\infty}^{\infty} x p(x) dx$$

事实 4.1. 期望存在的条件:

$$\sum_{i=1}^{\infty}|x_i|p(X=x_i)<\infty, \int_{-\infty}^{+\infty}|x|p(x)dx<\infty$$

当级数（或积分）绝对收敛，则收敛，期望一定存在。然而，级数（积分）也可能条件收敛，需要特殊讨论。

事实 4.2 (期望的性质)**.**

$$\begin{aligned}E(c)&=c\\E(cX)&=cE(x)\\E(X+Y)&=E(X)+E(Y)\\E(X_1+\cdots+X_n)&=E(X_1)+\cdots+E(X_n)\end{aligned}$$

例 4.1 (使用期望的性质求二项分布的期望)**.** 设 $X_k\sim b(1,p)$，$X=X_1,\ldots,X_n$，则有

$$E(X)=E(X_1)+E(X_n)=p+\cdots+p=np.$$

定义 4.2 (随机变量函数的期望)**.** 随机变量函数的数学期望:

离散随机变量:

$$E(g(X))=\sum_{i=1}^{\infty}g(x_i)p(X=x_i)$$

连续随机变量:

$$E(g(X))=\int_{-\infty}^{\infty}g(x)p(x)dx$$

例 4.2. 设 $X\sim U(-1,1)$，求 $E(X^2)$

$$E(X^2)=\int_{-1}^{1}x^2p(x)dx=\int_{-1}^{1}\frac{1}{2}x^2dx=\frac{1}{3}$$

事实 4.3 (随机变量函数期望的加法原则)**.**

$$E(g_1(X)+\cdots+g_n(X))=E(g_1(X))+\cdots+E(g_n(X))$$

定义 4.3 (方差)**.** 随机变量偏离期望的程度 $E(X)$（或 $D(X)$）:

$$Var(X)=E((X-E(X))^2)$$

称 $\sigma(X)=\sqrt{Var(X)}$ 为标准差。

可以使用如下方法简单计算方差:

$$\begin{aligned}Var(X)&=E((X-E(X))^2)=E(X^2-2XE(X)+E^2(X))\\&=E(X^2)-2E(X)E(X)+E^2(X)\\&=E(X^2)-E^2(X)\end{aligned}$$

事实 4.4 (期望与方差的关系)**.**

$$Var(X) = E(X^2) - E^2(X)$$

4.2 各种分布的期望与方差

二项分布

二项分布 $X \sim b(n,p)$，$E(X) = np, Var(X) = np(1-p)$.

可以使用 0-1 分布来理解二项分布，上面的讨论已经说明了 $E(X) = np$，下面说明方差的求法。

$$Var(X) = Var(X_1 + X_n) = Var(X_1) + Var(X_n) = n(p - p^2) = np(1-p)$$

（注：此处可以使用方差的加法律是因为 $X_1, \dots X_n$ 互相独立。）

泊松分布

泊松分布 $X \sim P(\lambda), \lambda > 0.P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, \dots$

$$E(X) \sim \lambda, Var(X) = \lambda$$

（可以用 $np \to \lambda, p \to 0$ 来理解，此时泊松分布趋近于二项分布，$E(X) = np \to \lambda, Var(X) = np(1-p) \to np \to \lambda.$）

例 4.3 (使用概率意义进行计算)**.** $X \sim b(4, 0.5)$，求 $E(X^2)$。

$$Var(X) = E(X^2) - E(X)^2$$

所以有

$$E(X^2) = Var(X) + E(X)^2 = np(1-p) - (np)^2 = 5$$

几何分布

几何分布 $X \sim Ge(p), p > 0.P(X = k) = p(1-p)^{k-1}, k = 1, 2, \dots$

$$E(X) = \frac{1}{p}, Var(X) = \frac{1-p}{p^2}$$

均匀分布

均匀分布：$X \sim U(a,b)$

$$E(X) = \frac{a+b}{2}, Var(X) = \frac{(b-a)^2}{12}$$

指数分布

指数分布：$X \sim Exp(\lambda), p(x) = \lambda e^{-\lambda x} I_{x>0}$

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

例 4.4 (利用指数分布计算积分)**.** 计算

$$\int_0^{+\infty} 5x^2 e^{-3x} dx$$

考虑 $X \sim Exp(5)$，

$$\begin{aligned}\int_0^{+\infty} 5x^2 e^{-3x} dx &= \frac{5}{3} \int_0^{+\infty} x^2 3e^{-3x} dx \\ &= \frac{5}{3} E(X^2) = \frac{5}{3}(E^2(X) + Var(X)) \\ &= \frac{5}{3}(\frac{1}{\lambda^2} + \frac{1}{\lambda^2}) = \frac{10}{27}\end{aligned}$$

正态分布

$X \sim N(\mu, \sigma)$，其期望和方差为

$$E(X) = \mu, Var(X) = \sigma$$

柯西分布

柯西分布：$X \sim c(\lambda, \mu), p(x) = \frac{\lambda}{\pi(\lambda^2 + (x-\mu)^2)}$

由于

$$\int_{-\infty}^{+\infty} |x| p(x) dx = \int_{-\infty}^{+\infty} |x| \frac{\lambda}{\pi(\lambda^2 + (x-\mu)^2)} dx \to \infty$$

不存在，所以 $E(X), Var(X)$ 均不存在。

4.3 期望与方差的性质

定理 4.1. *X 为随机变量，则 $c = E(X)$ 时，$E((X-c)^2)$ 达到最小。*

证明：$E((X-c)^2) = E(X^2) - 2cE(X) + c^2 = Var(X) + (E(X) - c)^2 \geq Var(X)$ 其中 $Var(X), E(X)$ 是常数，等号成立当且仅当 $c = E(X)$。

定理 4.2 (Chebyshev 不等式)**.**

$$\forall \varepsilon > 0, P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

证明:

$$\begin{aligned}
P(|X-E(X)| \geq \varepsilon) &= \int_{\{x:|x-E(X)| \geq \varepsilon\}} p(x)dx \\
&\leq \int_{\{x:|x-E(X)| \geq \varepsilon\}} \frac{(x-E(X))^2}{\varepsilon^2} p(x)dx \\
&\leq \int_{(-\infty,+\infty)} \frac{(x-E(X))^2}{\varepsilon^2} p(x)dx \\
&= \frac{Var(X)}{\varepsilon^2}
\end{aligned}$$

定义 4.4 (原点矩). 称 $E(X^k)$ 为 k 阶原点矩。

定义 4.5 (中心距). 称 $E((X-E(X))^k)$ 为 k 阶中心矩。

5 多维随机变量

5.1 多维随机变量的基本概念

定义 5.1 (n 维随机变量). 若 $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ 是定义在同一样本空间的 Ω 上的随机变量，则称

$$X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$$

是 n 维随机变量，或随机向量。

定义 5.2 (联合分布函数). 对任意 n 个实数，$x_1, \ldots, x_n$，事件 $X_1 \leq x_1, \ldots, X_n \leq x_n$ 同时发生的概率为

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

称为 n 维随机变量 X 的联合分布函数。

定理 5.1. 二维联合分布函数 $F(x,y)$ 具有以下性质:

1. 单调性。$F(x,y)$ 对 x,y 分别满足单调性

2. 有界性。$0 \leq F(x,y) \leq 1$ 且 $F(-\infty,y) = F(x,-\infty) = 0, F(+\infty,+\infty) = 1$

3. 右连续性。$F(x+0,y) = F(x,y), F(x,y+0) = F(x,y)$

4. 非负性

$$P(a < X < b, c < Y < d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

边缘分布：$P(x = x_j) = \sum_{i=1}^{\infty} P(X = x_j, Y = y_i)$

定义 5.3 (联合密度函数). 存在二元非负函数 $p(x,y)$，使得二维随机变量 (X,Y) 的分布函数 $F(x,y)$ 可以表示为

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(u,v)dudv$$

则称 $p(x,y)$ 为联合密度函数，

$$p(x,y)=\frac{\partial^2}{\partial x\partial y}F(x,y).$$

边缘分布函数：

$$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\left(\int_{-\infty}^{\infty}p(u,v)du\right)dv$$

边缘密度函数：

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dxdy$$

定义 5.4 (随机变量的独立性). 对于 n 元分布函数

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

则称 $X_1,\dots,X_n$ 相互独立，即

$$P(X_1\le x_1,\dots,X_n\le x_n)=\prod_{i=1}^{n}P(X_i\le x_i)$$

离散型等价定义：

$$P(X_1=x_1,\dots,X_n=x_n)=P(X_1=x_1)\dots P(X_n=x_n)$$

连续型等价定义：

$$p(x_1,\dots,x_n)=p(x_1)\dots p(x_n)$$

定理 5.2. 若在 $\mathbb{R}^2$ 上都有如下分解：

$$p(u,v)=f(u)g(v)$$

则 u,v 互相独立。

证：

$$\int_{\mathbb{R}^2}p(u,v)dudv=\int_{\mathbb{R}^2}f(u)g(v)dudv=\int_{\mathbb{R}}f(u)du\int_{\mathbb{R}}g(v)dv=1$$

若 $\int f(u)du=c$，则

$$\frac{1}{c}\int_{\mathbb{R}}f(u)du\times c\int_{\mathbb{R}}g(v)dv=p(u)\times p(v)$$

也就是 $p(u,v)=p(u)p(v)$，所以 u,v 独立。

定义 5.5 (随机变量函数的期望). 对于 n 元随机变量 $(X_1,\dots,X_n)$, 若 $Z=g(X_1,\dots,X_n)$,

离散情型:

$$\sum_{i_1}\cdots\sum_{i_n} g(x_1,\dots,x_n)P(X_1=x_1,\dots,X_n=x_n)$$

连续情型:

$$\int_{\mathbb{R}}\cdots\int_{\mathbb{R}} g(x_1,\dots,x_n)p(x_1,\dots,x_n)dx_1\dots dx_n$$

例 5.1. 设 $X,Y\sim Exp(\lambda)$ 独立同分布, 求 $Z=\max\{X,Y\}$ 的数学期望。

$$\begin{aligned}
E(Z) &= \int\int_{\mathbb{R}} Z(x,y)p(x,y)dxdy = \int\int_{x\geq y} xp(x,y)dxdy + \int\int_{x<y} yp(x,y)dxdy \\
&= \int_0^{+\infty} dx \int_0^x x\lambda e^{-\lambda x}\lambda e^{-\lambda y}dy + \int_0^{+\infty} dy \int_0^y y\lambda e^{-\lambda y}\lambda e^{-\lambda x}dx \\
&= 2\int_0^{+\infty} dx \int_0^x x\lambda e^{-\lambda x}\lambda e^{-\lambda y}dy \\
&= 2\int_0^{+\infty} x\lambda e^{-\lambda x}(1-e^{-\lambda x})dx \\
&= \frac{3}{2\lambda}
\end{aligned}$$

独立随机变量期望和方差的性质:

事实 5.1. 随机变量 X,Y 相互独立, 则

1. $E(XY)=E(X)E(Y), Var(X+Y)=Var(X)+Var(Y)$
2. 无论是否满足独立性, 都有

$$E(c_1X_1+\cdots+c_nX_n)=c_1E(X_1)+\cdots+c_nE(X_n)$$

若 $X_1,\dots,X_n$ 独立时, 有

$$E(X_1\dots X_n)=E(X_1)\dots E(X_n)$$
$$Var(c_1X_1+c_nX_n)=c_1^2Var(X_1)+\cdots+c_n^2Var(X_n)$$

以 0-1 分布求和的角度理解二项分布: $X=X_1+\cdots+X_n$, 其中 $P(1)=p, P(0)=1-p$。

$$E(X)=E(X_1+\dots X_n)=n\cdot p$$
$$Var(X)=Var(X_1+\dots X_n)=Var(X_1)+\cdots+Var(X_n)=n\cdot p(1-p)$$

5.2 协方差、相关系数与条件分布

定义 5.6 (协方差). 设 (X,Y) 是二元随机变量，$E[(X-E(X))(Y-E(Y))]$ 称为 X,Y 的协方差，记为 $Cov(X,Y)$.

协方差的性质:

$$
\begin{aligned}
Cov(X,a) &= 0 \\
Cov(X,Y) &= Cov(Y,X) \\
Cov(X+Y,Z) &= Cov(X,Z)+Cov(Y,Z) \\
Cov(X,Y) &= E(XY)-E(X)E(Y) \\
Cov(c_1X+a,c_2Y+b) &= c_1c_2Cov(X,Y) \\
Var(X+Y) &= Var(X)+Var(Y)+2Cov(X,Y) \\
Var(X-Y) &= Var(X)+Var(Y)-2Cov(X,Y) \\
[Cov(X,Y)]^2 &\le Var(X)Var(Y)=\sigma_X^2\sigma_Y^2
\end{aligned}
$$

例 5.2.

$$X_1,X_2,\ldots,X_m \sim b(n,p), Cov(X_1,\frac{X_1+\ldots X_n}{m})=$$

$$
\begin{aligned}
Cov(X_1,\frac{X_1+\ldots X_n}{m}) &= \frac{1}{m}[Cov(X_1,X_1)+\cdots+Cov(X_1,X_n)] \\
&= \frac{1}{m}Cov(X_1,X_1) \\
&= \frac{1}{m}[E(X^2)-E(X)E(X)] \\
&= \frac{np(1-p)}{m}
\end{aligned}
$$

定义 5.7 (相关系数).

$$Corr(X,Y)=Cov\left(\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right)$$

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

若 $Corr(X,Y)>0$，则 X,Y 正相关；若 $Corr(X,Y)<0$，则 X,Y 负相关；若 $Corr(X,Y)=0$，则 X,Y 不相关。

事实 5.2 (相关和独立的关系). 注意有以下事实:

X,Y 独立 $\Rightarrow Cov(X,Y)=E(XY)-E(X)E(Y)=0$ 不相关；

X,Y 不相关 $\not\Rightarrow X,Y$ 独立

例 5.3. $X \sim N(0,1), Y = X^2$，显然 X, Y 不独立，但 X, Y 不相关。

$$Cov(X,Y) = E(X^3) - E(X)E(X^2) = 0$$

不相关指的是不存在线性相关的关系，相关系数并不能充分地表达非线性的相关关系。

定理 5.3. $Corr(X,Y) = \pm 1$ 的充分必要条件是 X, Y 之间几乎处处有线性关系，即存在常数 a, b 使得 $P(YaX+b) = 1$.

证：先证充分性。

$Y = aX+b \Rightarrow Var(Y) = a^2 Var(x) \Rightarrow \sigma_Y = |a|\sigma_X, Cov(X,Y) = aCov(X,X), Corr(X,Y) = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \pm 1.$

再证必要性。

$Var(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = Var(\frac{X}{\sigma_X}) + Var(\frac{Y}{\sigma_Y}) \pm 2Cov(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}) = 2(1 \pm Corr(X,Y))$，当 $Corr(X,Y) = \pm 1 \Rightarrow Var(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = 0 \Rightarrow Y = aX + b$

5.3 常见多元分布

定义 5.8 (二维正态分布)**.** $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

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

$$p(x,y) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

事实 5.3. 二元正态分布的边界密度即为一元正态分布，但逆命题不成立。

若 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则独立 $\iff$ 不相关。

5.4 条件分布

定义 5.9 (离散分布的条件分布)**.** 对一切使 $P(Y = y_j) = p_j = \sum_{i=1}^{\infty} p_{ij} > 0$，称下式为条件分布列。

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

事实 5.4 (泊松分布与二项分布的可加性)**.**

$$X_1 \sim P(\lambda_1), \ldots, X_n \sim P(\lambda_n) \text{ 且彼此独立}, X_1 + X_2 + \cdots + X_n \sim P(\lambda_1 + \cdots + \lambda_n)$$
$$X_1 \sim B(n_1, p), \ldots, X_n \sim B(n_n, p) \text{ 且彼此独立}, X_1 + X_2 + \cdots + X_n \sim B(n_1 + \cdots + n_n, p)$$

定义 5.10 (连续分布的条件分布)**.** 对一切使 $p_Y(y) > 0$ 的 y，给定 $Y = y$ 条件下的条件分布函数与条件密度函数

$$
\begin{aligned}
F(x|y) &= P(X \le x|Y=y) = \lim_{h\to 0} P(X \le x | y \le Y \le y+h) = \lim_{h\to 0} \frac{P(X \le x, y \le Y \le y+h)}{P(y \le Y \le y+h)} \\
&= \lim_{h\to 0} \frac{\int_{-\infty}^{x} du \int_{y}^{y+h} dv}{\int_{y}^{y+h} p_Y(v)dv} = \lim_{h\to 0} \frac{\int_{-\infty}^{x}(p(u, y+c_2h)hdu)}{p_Y(y+c_1h)h} = \frac{\int_{-\infty}^{x} p(u,y)du}{p_Y(y)}
\end{aligned}
$$

$$
p(x|y) = \frac{\partial}{\partial x} \frac{\int_{-\infty}^{x} p(u,y)du}{p_Y(y)} = \frac{p(x,y)}{p(y)}
$$

连续场合下的全概率公式：

$$
p(x,y) = p_X(x)p(y|x) \Rightarrow p_Y(y) = \int_{-\infty}^{+\infty} p_X(x)p(y|x)dx
$$

连续场合下的贝叶斯公式：

$$
p(x|y) = \frac{p_X(x)p(y|x)}{\int_{-\infty}^{+\infty} p_X(x)p(y|x)dx}
$$

6 条件期望与随机变量函数的分布

定义 6.1 (条件期望)**.** 离散随机变量：

$$
E(X|Y=y) = \sum_i x_i P(X = x_i | Y = y)
$$

连续随机变量：

$$
E(X|Y=y) = \int_{-\infty}^{\infty} xp(x|y)dx
$$

例 6.1. $X \sim Ge(\frac{1}{4})$，求 $E(X|X>3)$

$$
\begin{aligned}
E(X|X>3) &= \sum_{k=1}^{\infty}(k+3)P(X=k+3|X>3) \\
&= \sum_{k=1}^{\infty}(k+3)P(X=k) \\
&= E(X+3) = E(X)+3 = 7.
\end{aligned}
$$

注：由几何分布的无记忆性可以推出 $E(X|X>k) = k + E(X) = k + \frac{1}{p}$

定理 6.1 (重期望公式)**.**

$$
E(X) = E(E(X|Y))
$$

证：先证离散型

$$
\begin{aligned}
E(E(X|Y)) &= \sum_{k=1}^{\infty} E(X|Y=y_k)P(Y=y_k) \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} j\frac{P(X=j,Y=y_k)}{P(Y=y_k)}P(Y=y_k) \\
&= \sum_{k=1}^{\infty} jP(X=k) = E(X)
\end{aligned}
$$

再证连续型

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xp_X(x)dx \\
&= \int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} p(x,y)dydx \\
&= \int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} p_Y(y)p(x|y)dydx \\
&= \int_{-\infty}^{+\infty} E(X|Y=y)p_Y(y)dy = E(E(X|Ys))
\end{aligned}
$$

例 6.2 (几何分布的期望). 设随机变量 $X \sim Ge(p), 0<p<1$，求 $E(X), Var(X)$

定义随机变量 $Y = \mathrm{I}_{X=1}$, 由全期望公式 $E(X) = E(E(X|Y))$

$$
\begin{aligned}
E(E(X|Y)) &= P(Y=1)E(X|Y=1) + P(Y=0)E(X|Y=0) \\
&= P(X=1)E(X|X=1) + P(X>1)E(X|X>1) \\
&= p + (1-p)(1+E(X)) = E(X)
\end{aligned}
$$

解得 $E(X) = \frac{1}{p}$.
求 $E(X^2|X>1)$:

$$
\begin{aligned}
E(X^2|X>1) &= \sum_{k=2}^{\infty} k^2\frac{P(X=k|X>1)}{P(X>1)} \\
&= \sum_{k=2}^{\infty} k^2 P(X=k-1) \\
&= \sum_{k=1}^{\infty} (1+k)^2 P(X=k) = E((1+X)^2)
\end{aligned}
$$

可以求得 $Var(X) = \frac{1-p}{p^2}$

事实 6.1 (非线性的最小二乘最佳预报)**.** 二元随机变量 $(X,Y),Var(X),Var(Y)$ 存在，令 $\phi(X)=E(Y|X)$, 则

$$E[(Y-\phi(X))^2]=\min_{\psi} E[(Y-\psi(X))^2]$$

定理 6.2. 设 n 维随机变量 $X=(X_1,\dots,X_n)$ 的密度函数为 $p_X(x_1,\dots,x_n)$，n 元函数

$$g(x_1,\dots,x_n)=\begin{pmatrix} g_1(x_1,\dots,x_n) \\ g_2(x_1,\dots,x_n) \\ \vdots \\ g_n(x_1,\dots,x_n) \end{pmatrix}$$

且满足条件

(1) 存在唯一的反函数 $x_i(y_1,\dots,y_n)$，即存在方程组 $g_i(x_1\dots,x_n)=y_i$ 的唯一实数解

(2) $x_i(y_1,\dots,y_n)$，$g_i(x_1\dots,x_n)$ 都连续

(3) 存在连续的偏导数 $\frac{\partial x_i}{\partial y_j},\frac{\partial g_i}{\partial x_j}$，记

$$J=\det\left(\frac{\partial x_i}{\partial y_j}\right)$$

则 n 维随机向量 $Y=(Y_1,\dots,Y_n)^T=(g_1(X_1,\dots,X_n),\dots,g_n(X_1,\dots,X_n))^T$ 的密度函数为

$$p_Y(y_1,\dots,y_n)=p_X(x_1(y_1,\dots,y_n),x_n(y_1,\dots,y_n))\cdot|J|\cdot \mathrm{I}_{sta}$$

其中 sta：$(y_1,\dots,y_n)$ 使得 $y=g(x)$ 有解

注：一种形式化的解释方法为，密度函数的映射本质上是 $p_Y(y)dy=p_X(x)dx$.

另注：反函数不容易求解时，可以利用 $|J|=|\frac{\partial y}{\partial x}|^{-1}$ 来计算

再注：若反函数不唯一，即 $g_i(x_1,\dots,x_n)=y_i$ 的解不唯一，为 $x_i^(s)$，此时 y-空间中的点对应于 x-空间的多个点，计算概率时需全部相加

又注：如果 y 的维度小于 x，不妨增补定义 $g_j(x_1,\dots,x_n)=x_i$

7 特征函数与极限定理

7.1 特征函数（不考）

定义 7.1 (特征函数)**.** X 是一个随机变量，称 $\phi(t)=E(e^{itX}),t\in\mathbb{R}$，为 X 的特征函数，特征函数总是存在的。

分布函数与特征函数一一对应。

离散型：

$$\phi(t)=\sum_{k=1}^{\infty}e^{itx_k}P(X=x_k)$$

连续型:

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} p(x) dx$$

定理 7.1. 随机变量的特征函数 $phi(t)$ 在 $\mathbb{R}$ 上一致连续。

注：定义特征函数的动机如下。已知有随机变量 $X, Y, p(x), p(y)$，希望求 $Z = X + Y$ 的概率密度函数，需要使用定理6.2计算，较为麻烦。注意到 $\phi(x)\phi(y) = \phi(z)$，可以先计算 $\phi(x), \phi(y)$，再通过 $\phi(z)$ 求解 $p(z)$ 即可。

几种常见分布的特征函数

$$\begin{aligned}
X \sim b(1,p), &\phi(t) = pe^{it} + 1 - p \\
X \sim Ge(p), &\phi(t) = \frac{pe^{it}}{1 - qe^{it}} \\
X \sim P(\lambda), &\phi(t) = e^{\lambda(e^{it} - 1)} \\
X \sim U(-1,1), &\phi(t) = \frac{\sin t}{t} \\
X \sim N(0,1), &\phi(t) = e^{-\frac{t^2}{2}}
\end{aligned}$$

特征函数的性质

事实 7.1. $Y = aX + b, \phi_Y(t) = e^{ibt}\phi_X(at)$

事实 7.2. 若 X, Y 独立，则 $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$

事实 7.3. 若 $E(X^m)$ 存在，则 $\phi(t) m$ 阶可导。对 $1 \le k \le m$，有 $\phi^{(k)}(0) = i^k E(X^k)$

定义 7.2 (Fourier 变换与逆变换)**.** $\phi(t) = \int_{-\infty}^{+\infty} e^{itx} p(x) dx, p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \phi(t) dt$

7.2 极限定理

事实 7.4 (大数定律)**.** 抽样次数越多，频率越接近于概率，平均值越接近于期望。

设 S_A 为 n 重伯努利试验中事件 A 发生的次数，p 为每次实验中 A 出现的概率，则

$$\forall \varepsilon > 0, \lim_{n \to \infty} P\left\{ \left| \frac{S_A}{n} - p \right| < \varepsilon \right\} = 1$$

证：由切比雪夫不等式知，

$$P\left\{ \left| \frac{S_A}{n} - p \right| < \varepsilon \right\} \ge 1 - \frac{Var(\frac{S_A}{n})}{\varepsilon^2} = 1 - \frac{\frac{p(1-p)}{n}}{\varepsilon^2} \to 1, n \to \infty$$

定理 7.2 (大数定律的一般形式). 设有一个随机变量序列 $\{X_n\}$，若 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right\} = 1$$

则称 $\{X_k\}$ 服从大数定律。

定理 7.3 (切比雪夫大数定律). 设有一个随机变量序列 $\{X_n\}$，所有 X_i, X_j 不相关，则

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right\} = 1$$

定理 7.4 (马尔科夫大数定律). 设有一个随机变量序列 $\{X_n\}$，$\frac{1}{n^2}Var(\sum_{k=1}^{n} X_k) \to 0$，则

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right\} = 1$$

定理 7.5 (辛钦大数定律). 设有一个随机变量序列 $\{X_n\}$ 符合独立同分布，$E(X_1) = \mu$，则

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \varepsilon\right\} = 1$$

定理 7.6 (中心极限定理). 设 $\{X_n\}$ 是独立同分布的随机变量序列。若 $E(X_1) = \mu, Var(X_1) = \sigma^2$，则对于每个固定的 y，有

$$\lim_{n\to\infty} P\left\{\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}} \le y\right\} = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{t^2}{2}} dt$$

推论:

$$(X_1 + \cdots + X_n) \sim N(\sum_{k=1}^{n} E(X_k), \sum_{k=1}^{n} Var(X_k))$$

例 7.1. 将一枚均匀的硬币抛掷 *36* 次，抛出正面的次数不超过 *21* 的概大约是 $\Phi(1.17)$

$$X \sim b(36, 0.5) \dot{\sim} N(18, 3^3)$$

由于 X 是离散变量，需要做 *0.5* 矫正（将 X 正负 *0.5* 以内的概率计算到 X 上），则

$$P(X \le 21) = \Phi(\frac{21.5 - 18}{3}) = \Phi(1.17)$$

例 7.2. 设系统有 *100* 个独立的部件构成，运行时间每个部件损坏的概率是 *0.1*，至少有 *85* 个部件完好时系统才能正常工作，求系统正常工作的概率

解:正常工作部件的数目 $X \sim b(100, 0.9)$, $E(X) = 90, Var(X) = 9, X \dot{\sim} N(90, 9)$, $\frac{X-90}{3} \sim N(0, 1)$。

则 $P(\text{正常工作}) = P(X \ge 85) = 1 - \Phi(-\frac{5}{3}) = \Phi(5/3)$

8 统计学基本概念

定义 8.1. 总体：一个统计问题研究对象的全体。构成总体的每个成员称为个体。

定义 8.2. 样本：从总体中随机抽样的部分个体组成的集合称为样本。

常用统计量：

样本均值

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

定理 8.1. 设 $x_1, \ldots, x_n$ 是来自某个总体的样本，$\bar{x}$ 是样本均值

1) 若总体分布为 $N(\mu, \sigma^2)$，则 $\bar{x} \sim N(\mu, \sigma^2/n)$

2) 若总体分布不是正态分布，$E(x) = \mu, Var(x) = \sigma$，则 $\bar{x} \sim^{\cdot} N(\mu, \sigma^2/n)$

样本方差

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

定理 8.2. 设总体 X 具有二阶矩，即 $E(X) = \mu, Var(X) = \sigma^2 < +\infty$，$x_1, \ldots, x_n$ 是从 X 得到的样本，$\bar{x}, s^2$ 是均值和方差，则

$$E(\bar{x}) = \mu, Var(\bar{x}) = \frac{\sigma^2}{n}, E(s^2) = \sigma^2$$

对第三个等式的证明：

$$\begin{aligned}
E(s^2) &= E(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2) = \frac{n}{n-1}E((x_i - \bar{x})^2) \\
&= \frac{n}{n-1}[E(x_i^2) - 2E(x_i\bar{x}) + E(\bar{x}^2)] \\
&= \frac{n}{n-1}[E(x_i^2) - 2(\frac{1}{n}E(x_i^2) + \frac{n-1}{n}E^2(x_i)) + E(\bar{x}^2)] \\
&= \frac{n}{n-1}[E(x_i^2) - 2(\frac{1}{n}E(x_i^2) + \frac{n-1}{n}E^2(x_i)) + E[\frac{1}{n^2}(nx_i^2 + (n^2-n)x_ix_j)]] \\
&= \frac{n}{n-1}[\frac{n-1}{n}E(x+i^2) - \frac{2n-2}{n}E^2(x_i) + \frac{1}{n}E(x_i^2) + \frac{n-1}{n}E^2(x_i)] \\
&= \frac{n}{n-1}\frac{n-1}{n}Var(X) = \sigma^2
\end{aligned}$$

三大抽样分布：

定义 8.3 (卡方分布)**.** $x_1, \ldots, x_n$ 是从 $N(0,1)$ 抽样得到的样本，令 $Y = x_1^2 + \cdots + x_n^2$，则 Y 服从自由度为 n 的 χ^2 分布。

若 $Y \sim \chi^2(n), E(Y) = n, Var(Y) = 2n$

定理 8.3. *若 $x_1,\ldots,x_n$ 抽样自 $N(\mu,\sigma^2)$，$\bar{x}$ 是均值，s^2 是均值和方差，则*

1) $\bar{x}, s^2$ 独立

2) $\bar{x} \sim N(\mu, \sigma^2/n)$

3) $(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$

定义 8.4 (t 分布)**.** *若 $X_1 \sim N(0,1), X_2 \sim \chi^2(n)$ 且 X_1, X_2 独立，则*

$$Y = \frac{X_1}{\sqrt{X_2/n}} \sim t(n)$$

定义 8.5 (F 分布)**.** *若 $X_1 \sim \chi^2(n), X_2 \sim \chi^2(m)$，则 $Y = \frac{X_1/m}{X_2/n} \sim F(m,n)$*

分位数

定义 8.6. *对随机变量 X，有概率分布函数 $F(x)$，若 $F(d_\alpha) = \alpha$，也就是 $P(X \le d_\alpha) = \alpha$，则称 d_α 是 α 分位数。*

正态分布的分位数记作 u_α，卡方分布的分位数记作 $\chi^2_\alpha(n)$，t 分布的分位数记作 $t_\alpha(n)$，F 分布的分位数记作 $F_\alpha(m,n)$.

例 8.1. *$X \sim N(0,1)$，求 $P(|X| < u_{0.975})$*

解:

$$P(|X| < u_{0.975}) = 1 - 2P(X \ge u_{0.975}) = 1 - 2(1 - 0.975) = 0.95$$

9 参数点估计的方法与评价

参数点估计：设 $X_1,\ldots,X_n$ 是来自于某个总体的样本，利用样本估计总体分布的参数 θ，构造适当的统计量 $\hat{\theta} = \theta(X_1,\ldots,X_n)$ 称为参数 θ 的点估计。估计的方法不是唯一的，但需要满足一定的合理性。

最常用的两种点估计方法：矩估计和极大似然估计。

9.1 矩估计法

方法：使用替换原理。

矩的理论表达式为参数的函数，用样本矩替换理论上的矩，解方程得到参数的近似。为了计算简单，尽可能用低阶矩。

这里的矩可以是原点矩、中心距，以及样本方差等。估计 m 个参数，使用 m 个方程，也就是

$$E(X^k) = f_k(\theta_1, \theta_m) \Rightarrow f_k(\hat{\theta}_1,\ldots,\hat{\theta}_m) = \frac{x_1^k + \ldots x_n^k}{n}, k = 1,\ldots,m$$

例 9.1. 对于均匀分布总体 $U(a,b)$，估计 a,b

总体期望与方差分别为 $E(X)=\frac{a+b}{2}, Var(X)=\frac{(b-a)^2}{12}$
令 $\frac{\hat{a}+\hat{b}}{2}=\bar{x}, \frac{(\hat{b}-\hat{a})^2}{12}=s^2$，估计得到 $\hat{a}=\bar{x}-\sqrt{3}s, \hat{b}=\bar{x}+\sqrt{3}s$
也可以使用均值和二阶原点矩估计。

9.2 极大似然估计

核心思想：发生的情况概率最大。参数取值应当以最大概率解释样本数据。
样本的联合概率是关于参数 θ 的参数。定义联合概率函数为

$$L(\theta; x_1,\dots,x_n)=\prod_{k=1}^{n} p(x_k;\theta)$$

问题就是求解

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} L(\theta; x_1,\dots,x_n)\\ &= \arg\max_{\theta} \ln L(\theta; x_1,\dots,x_n)\\ &= \arg\max_{\theta} \sum_{k=1}^{n} \ln p(x_k;\theta)\end{aligned}$$

例 9.2. 总体服从于指数分布 $Exp(\lambda)$，设 $x_1,\dots,x_n$ 为观测值，求 λ 的最大似然估计

解：似然函数如下。

$$L(\lambda; x_1,\dots,x_n)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}$$

$$\ln L(\lambda)=n\ln\lambda-\lambda(x_1+\cdots+x_n)$$

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{n}{\lambda}-(x_1+\cdots+x_n)=0$$

所以 $\hat{\lambda}=\frac{1}{\bar{x}}$

例 9.3. 使用极大似然估计和样本 $X_1,\dots,X_n$ 来估计总体 $U(a,b)$ 的参数。

$$L(a,b)=\frac{1}{(b-a)^n}$$

则 b 应当尽可能的小，a 应当尽可能的大。同时，有边界条件

$$a\le\min\{x_1,\dots,x_n\}, b\ge\max\{x_1,\dots,x_n\}$$

所以极大似然估计参数为 $\hat{a}=\min\{x_1,\dots,x_n\}, \hat{b}=\max\{x_1,\dots,x_n\}$

9.3 点估计量的评价标准

定义 9.1 (相合性)**.** 设 $\theta \in \Theta$ 是未知参数，$\hat{\theta}_n$ 是根据 n 个样本得到的估计量。若对 $\forall \varepsilon > 0$，有 $\underline{\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0}$，则称 $\hat{\theta}_n$ 是 θ 的相合估计。

定理 9.1. 设 $\hat{\theta}_n$ 是 θ 的一个估计量，若

$$\lim_{n\to\infty} E(\hat{\theta}_n) = 0, \lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

则 $\hat{\theta}_n$ 是 θ 的相合估计。

定义 9.2 (无偏性)**.** 设 $\theta \in \Theta$ 是未知参数，$\hat{\theta}_n$ 是根据 n 个样本得到的估计量。若对任意 θ 有 $\underline{E(\hat{\theta}) = \theta}$，则称 $\hat{\theta}$ 是 θ 的无偏估计。

定义 9.3 (有效性)**.** 设 $\theta \in \Theta$ 是未知参数，$\hat{\theta}_n$ 是根据 n 个样本得到的估计量。若对任意 θ 有 $Var(\hat{\theta}_1) \le Var(\hat{\theta}_2)$，且存在一个 θ 使得不等式严格成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。

例 9.4. $X_1, \dots, X_n$ 是来自均匀总体 $X \sim U(-a, a)$ 的样本，$\hat{a} = \sqrt{3}s$ 是否为 a 的无偏估计？

需要验证 $E(\sqrt{3}s) =^? a$。注意在非单点取值的时候，都有 $Var(X) = E(X^2) - E^2(X) > 0$，也就是 $a^2 = E(3s^2) > E^2(\sqrt{3s^2}) = E^2(\sqrt{3}s)$

所以有

$$E(\hat{a}) = E(\sqrt{3}s) < \sqrt{E(3s^2)} = a$$

是有偏估计。

例 9.5. 设总体 X 在 $[0, \theta]$ 上服从均匀分布，其中 $\theta > 0$ 未知。

设从总体中抽取了 X_1, X_2, X_3, X_4，则下面都是 θ 的无偏估计量。

$$\begin{aligned}
\hat{\theta}_1 &= \frac{1}{2}(X_1 + \cdots + X_4) \\
\hat{\theta}_2 &= 2X_1 + X_2 - X_3 \\
\hat{\theta}_3 &= X_1 + X_2 + X_3 - X_4
\end{aligned}$$

考察它们的方差：

$$\begin{aligned}
Var(\hat{\theta}_1) &= \frac{1}{4} \times 4Var(X) = Var(X) \\
Var(\hat{\theta}_2) &= 4Var(X) + 2Var(X) = 6Var(X) \\
Var(\hat{\theta}_3) &= 4Var(X)
\end{aligned}$$

所以 $\hat{\theta}_1 = \frac{1}{2}(X_1 + \cdots + X_4)$ 是最有效的估计。

10 参数区间估计

定义 10.1 (置信区间、置信水平、置信系数). 对未知参数 θ 估计得到的区间 $S(\theta) = [f(\theta), g(\theta)]$(开闭均可)称作置信区间,$P(\theta \in S(\theta))$ 称作置信水平,置信水平可以取到的最大值称为置信系数。

定义 10.2 (参数区间估计). 设 $x_1, \ldots, x_n$ 是来自总体 $X \sim F(x; \theta)$ 的样本,θ 是未知参数,$I(x_1, \ldots, x_n)$ 是随机区间,由样本值完全确定。称该区间是参数 θ 的一个置信水平为 $1-\alpha(0 < \alpha < 1)$ 的区间估计,是指

$$P_\theta(\theta \in I(x_1, \ldots, x_n)) \geq 1 - \alpha$$

区间估计公式:总体 $N(\mu, \sigma^2)$,期望未知,方差已知,$x_1, \ldots, x_n$ 是样本,求 μ 的 $1-\alpha$ 的置信区间。

解:μ 的点估计 $\bar{x} \sim N(\mu, \sigma^2/n)$,$\bar{x} - \mu$ 的分布与 μ 无关,则

$$Z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0, 1)$$

则有

$$\Phi(d) - \Phi(c) = P\left(c \leq \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \leq d\right) = 1 - \alpha$$

区间估计的构造方法

1. 构造枢轴量 $G(x_1, \ldots, x_n, \theta)$,希望 G 完全由样本值和未知参数确定(概率分布可以直接得到)
2. 选取两个常数 c, d,使得对给定的 α,有 $P(c \leq G \leq d) \geq 1 - \alpha$
3. 求解 $c \leq G(x_1, \ldots, x_n, \theta) \leq d$,得到 $\hat{\theta}_1(x_1, \ldots, x_n) \leq \hat{\theta} \leq \hat{\theta}_2(x_1, \ldots, x_n)$

置信区间的选取

我们总希望置信区间尽可能的小,也就是求解下面的带约束最优化问题。

$$\min\{d - c\}$$
$$s.t\ F(d) - F(c) = 1 - \alpha$$

使用 Lagrange 乘子法,容易得到 $p(c) = p(d)$.

关于卡方分布的注意：设 $x_1,\dots,x_n$ 是来自 $N(\mu,\sigma^2)$ 的样本，则

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$$
$$\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2}\sim\chi^2(n-1)$$

问题：总体 $X\sim N(\mu,\sigma^2)$，σ^2 未知，求 μ 的 $1-\alpha$ 的置信区间

$$\bar{x}\sim N(\mu,\sigma^2/n),\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$
$$\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$$

则有

$$\frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}}=\frac{\sqrt{n}(\bar{x}-\mu)}{s}$$

注意到这实际上是一个 t 分布：

$$X\sim N(0,1),Y\sim\chi^2(n),\frac{X}{\sqrt{Y/n}}\sim t(n)$$

所以

$$\frac{\sqrt{n}(\bar{x}-\mu)}{s}\sim t(n-1)$$

由以上求解过程，可以看到实际上 $n\sim+\infty$ 时，$t(n)\to N(0,1)$

11 假设检验

以女士品茶问题为例，假设检验问题应当如下提出：

设这位女士对一杯饮料能够正确鉴别的概率是 p，原假设（零假设）：女士没有鉴别力 $p\leq 1/2$，备择假设：相反的结果 $p>1/2$.

检验借助一个统计量完成，该统计量成为检验统计量。

11.1 假设检验

检验：在原假设条件下，该统计量的取值是否正常。若异常到一定程度，则拒绝原假设。

显著性水平：异常程度的水平，常用 $\alpha=0.1,0.05,0.01$

假设检验的基本步骤：

- 建立假设（原假设和备择假设）$H_0:\theta\in\Theta_0$ vs. $H_1:\theta\in\Theta_1$
- 选择检验统计量，给出拒绝域 W 的形式（$\bar{W}$ 是接受域）
- 选择显著性水平 $\alpha=\max\{P(拒绝H_0|H_0为真)\}=\max\{P_0(X\in W),\theta\in\Theta_0\}$

例 11.1. *设工厂生产的一种产品，质量超过 10 为优质的标准，服从正态分布 $N(\mu,4)$。现在抽样了 16 个产品，考察接受这批产品为优质的条件。*

$$H_0:\mu\geq 10 \text{ vs. } H_1:\mu<10$$

拒绝域：$\bar{x}<10-u_{1-\alpha}/2$。（$\bar{x}\sim N(10,4/16),2(\bar{x}-10)<u_\alpha,(\bar{x}-10)<-u_{1-\alpha}/2$）

选定 $\alpha=0.05$ 时，有 $\bar{x}<9.1775$ 为拒绝域，则只要 $\bar{x}\geq 9.1775$ 则认为该批产品优质。

11.2 假设检验的两类错误

	原假设成立	原假设不成立
接受	✓	第二类错误（受伪）
拒绝	第一类错误（拒真）	✓

第一类错误的概率：$\alpha(\theta)=P_\theta(X\in W),\theta\in\Theta_0$

第二类错误的概率：$\beta(\theta)=P_\theta(X\in\bar{W}),\theta\in\Theta_1$

例 11.2. *某厂生产长度为 35 的螺钉，实际生产的产品长度服从正态分布 $N(\mu,9)$。做假设检验，样本容量 $n=36,H_0:\mu=35,H_1:\mu\neq 35$，拒绝域为 $W=\{\bar{x}:|\bar{x}-35|>1\}$*

第一类错误：检验统计量 $\bar{X}\sim N(\mu,1/4)$，在假设成立的时候 $\bar{X}\sim N(35,1/4)$，所以第一类错误的概率是

$$\alpha=P(|\bar{X}-35|\geq 1)=P(2|\bar{X}-35|>2)=2(1-\Phi(2))$$

$\mu=36$ 时，犯第二类错误的概率：此时 $\bar{X}\sim N(36,1/4)$，则

$$\beta=P(34<\bar{X}<36)=P(-4<2(\bar{X}-36)<0)=\Phi(0)-\Phi(-4)=\Phi(4)-1/2$$

注意：原假设与备择假设不能交换，以假设检验前的主观意愿作为原假设（保护原假设）。例如，认为厂家生产的产品较为优质，则 H_0: 厂家生产的产品较为优质。在例题11.1中，如果认为产品较为优质（过去长时间的产品较为优质），则得到的拒绝域较小（优质的标准较低）。

样本容量 $n\to+\infty$ 时，两种取法趋于一致。

12 拟合优度检验

P 值：在一个假设检验问题中，利用样本观测值可以做出拒绝原假设的最小显著性水平；即原假设成立条件下，统计量出现在比观测值更异常的范围的概率最大值

例 12.1 (单侧区间)**.** *判断产品是否优质的问题（$n = 16, x \sim N(\mu, 4)$）：$H_0 : \mu \geq 10; H_1 : \mu < 10$*

两次观测结果分别为 $\bar{x} = 9.3, 10.15$ 的 p 值是?

$$p_1 = P(\bar{x} \leq 9.3) = P(2(\bar{x} - 10) \leq 2(9.3 - 10)) = \Phi(-1.4) = 0.081$$

$$p_2 = P(\bar{x} \leq 10.15) = P(2(\bar{x} - 10) \leq 2(10.15 - 10)) = \Phi(0.3) = 0.618$$

双侧区间应当计算（$\bar{x}_0$ 是测量到的值，$\bar{x}$ 是随机变量）

$$p = P(|\bar{x} - \mu| > |\bar{x}_0 - \mu|)$$

检验函数和随机化检验（用来解决离散型随机变量不能将 α 用满的问题）（看 ppt）

χ^2 检验

设总体服从离散分布 $P(X = x_i) = p_i, i = 1, \ldots, k$，进行 n 次独立地观测，k 种取值出现的频次分别是 N_i 则

$$X = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \overset{\cdot}{\sim} \chi^2(k-1)$$

解释：$N_i \sim b(n, p_i) \Rightarrow N_i \overset{\cdot}{\sim} b(np_i, np_i(1-p_i))$，所以有近似 $\frac{N_i - np_i}{\sqrt{np_i(1-p_i)}} \approx \frac{N_i - np_i}{\sqrt{np_i}} \sim N(0,1)$

如果用样本估计 s 个参数，则 $X \sim \chi^2(k - s - 1)$

可以使用 χ^2 检验独立性

列联表检验，见 PPT：

$$Y = \sum_{i=1}^{s} \sum_{j=1}^{t} \frac{(n_{ij} - \frac{c_i d_j}{n})^2}{\frac{c_i d_j}{n}} \sim \chi^2((s-1)(t-1))$$