概率论与数理统计

1. 确定性（必然）；一定发生（不发生）
随机（偶然） 可能发生，可能不发生

1.1 随机事件

试验：观察，测量，实验

随机试验：①在相同条件下可重复
(experiment) ②结果不止一个
③无法预测

事件：每种结果，随机事件 A, B, C
(event) 基本事件：不能再分（不必再分）
（相对于实验目的）
复合事件：基本事件复合
必然事件：Ω
不可能事件：$\varnothing$
} 不随机

样本空间：所有基本事件的集合
(sample space)

样本点：样本空间的元素

事件的集合表示 $A=\{2,4,6\}$ $C=\{1,2\}$ 点数 <3

Ω 必然事件，样本空间

1.1.3 事件间的关系.

① $A\subset B$ A发生导致B发生 $\varnothing\subset A\subset\Omega$

$B\supset A$

$A\subset B$，$B\subset A\Rightarrow A=B$

② $A\cup B$ $A\cup B\supset A$ $A\cup\varnothing=A$ $A\cup A=A$

③ $A\cap B$ $A\cap B\subset A$ $A\cap A=A$ $A\cap\varnothing=\varnothing$ $A\cap\Omega=A$

无限可列个： ① $\mathbb{N}$: $1, 2, 3, \cdots$

② $\mathbb{Z}$: $0, 1, -1, 2, -2, \cdots$

③ $\mathbb{Q}$: $\frac{p}{q}$ $\qquad 0.1\dot{2} = 0.1 + 0.0\dot{2} = \frac{1}{10} + \frac{1}{10} \times \frac{2}{9} = \frac{11}{90}$

$0, \frac{1}{1}, \frac{-1}{1}, \frac{1}{2}, \frac{-1}{2}$

④ $A \setminus B$: A发生而B不发生

⑤ mutually exclusive: A, B 不同时发生 $\qquad A \cap B = \phi$

⑥ complement: A, B mutually exclusive and $A \cup B = \Omega$

$A = B^c \quad B = A^c$

1) $(A^c)^c = A$ $\qquad$ 2) $A \setminus B = A \setminus A \cap B = A \cap B^c$

⑦ 完备事件组 $A_1, A_2, \cdots A_n$ 两两 mutually exclusive and $\bigcup_{i=1}^{n} A_i = \Omega$

运算律 1) $A \cup B = B \cup A \qquad A \cap B = B \cap A$

2) $(A \cup B) \cup C = A \cup (B \cup C)$

3) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

4) $(A \cup B)^c = A^c \cap B^c$

1.2 事件的概率

1.2.1 概率的初等描述

可能性大小、 $P(A) = \frac{1}{2} \quad P(B) = \frac{1}{2}$

$P(\Omega) = 1 \qquad P(\phi) = 0 \qquad 0 \leq P(A) \leq 1$

1.2.2 古典，条件：1) 有限个样本

2) 等可能性

$$P(A) = \frac{\text{A中包含的基本事件数}}{\text{基本事件总数}}$$

排列组合

加法原理：几类方案，加法

乘法原理：分几步（先点包子，再吃米饭）

排列：1) 不重复排列　从n个不同元素中取出m个，排列

$$^{m}P_{n} = n(n-1)(n-2)\cdots(n-m+1) = \frac{n!}{(n-m)!}$$

$$^{10}P_{5} = 10\times 9\times 8\times 7\times 6 = \frac{10!}{5!}$$　　$5^{0} = 5^{1-1} = \frac{5^{1}}{5^{1}} = 1$

$$^{n}P_{n} = n(n-1)\cdots 3\cdot 2\cdot 1 = n!$$　　$0^{0} = 0^{1-1} = \frac{0}{0}$

1) $1! = 1\times 0! \Rightarrow 0! = 1$

2) $^{0}P_{0} = 0! = 1$

3) $\frac{n!}{0!} = n! \Rightarrow 0! = 1$

2) 重复排列　从n个不同元素中取m个排列

$$n\times n\times n\cdots n = n^{m}$$

组合：从n个不同元素中取出m个不同元素

$$^{n}C_{m} = \frac{^{n}P_{m}}{m!} = \frac{n(n-1)\cdots(n-m+1)}{m(m-1)\cdots 2\ 1} = \frac{n!}{m!(n-m)!}$$

$$^{n}C_{n-m} = {^{n}C_{m}} \qquad ^{n}C_{0} = 1,\ ^{n}C_{n} = 1$$

例：5白4黑，任取3球　1) 2白一黑　$\frac{^{5}C_{2}\,^{4}C_{1}}{^{9}C_{3}}$

2) 没有黑　$\frac{^{5}C_{3}}{^{9}C_{3}}$　步数

3) 颜色相同　$\frac{^{5}C_{3}+{^{4}C_{3}}}{^{9}C_{3}}$　情况　$1-\frac{^{5}C_{2}\,^{4}C_{1}}{^{9}C_{3}}-\frac{^{5}C_{1}\,^{4}C_{2}}{^{9}C_{3}}$

例：a白b黑，取一个，白球：$\frac{a}{a+b}$

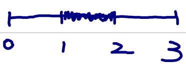
例 a白b黑，从中接连取出m个（$1\le m\le a+b$）第m个是白球

法1) $\frac{a\times(a+b-1)!}{(a+b)!}=\frac{a}{a+b}$　　法2) $\frac{a\times {}^{a+b-1}P_{m-1}}{{}^{a+b}P_{m}}=\frac{a}{a+b}$

法3) 先取m上的球 $\frac{a}{a+b}$

古典性质：1) 非负性 $0\le P(A)\le 1$　2) 规范性：$P(\Omega)=1$ $P(\phi)=0$　3)有限可加：$A_1\dots A_n$ 互不相容

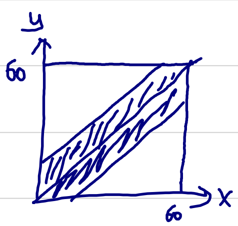
$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$

几何概型 

$P([1,2])=\frac{1}{3}$　$P(A)=\frac{\mu(G)}{\mu(\Omega)}$ 度量

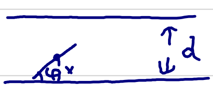
例 甲乙 6点—7点，先到等15min，甲乙一小时内任意时刻可到达

A — 2人见面　X — 甲到达　Y — 乙到达



$|y-x|\le 15$　$\begin{cases} y-x\ge 0, & y-x\le 15 \Rightarrow y\ge x \quad y\le x+15 \\ y-x<0 & x-y\le 15 \end{cases}$

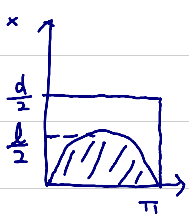
$P=\frac{60\times 60-2\times\frac{1}{2}\times 45\times 45}{60\times 60}=0.4375$

例 

针长度为 l　　设x：针中点离最近线的距离 $0\le x\le \frac{d}{2}$

φ：针与线的夹角　$0\le\varphi\le\pi$

$\Omega=\{(\varphi,x): 0\le x\le\frac{d}{2},\ 0\le\varphi\le\pi\}$



$G=\{(\varphi,x): 0\le\varphi\le\pi,\ 0\le x\le\frac{l}{2}\sin\varphi\}$

斜边 $\frac{x}{\sin\varphi}\le\frac{l}{2}$

$P=\frac{\int_0^{\pi}\frac{l}{2}\sin\varphi\, d\varphi}{\pi\cdot\frac{d}{2}}=\frac{2l}{\pi d}$

关键在于观察到 φ 的对边 $<x$

$P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$　不相容

频率与概率　　n次试验　　A发生了m次　　$\frac{m}{n}$ 频率

性 1) 非负　$0 \le W_n(A) \le 1$

2) 规范　$W_n(\Omega)=1$　　$W_n(\phi)=0$

3) 可加　$A_1 \cdots A_n$ 不相容

$W_n(A_1+\cdots+A_m) = W_n(A_1)+\cdots+W_n(A_m)$

$W_n(A) \longrightarrow P$

公理化　　描述　古典　几何　统计

公理1　　$0 \le P(A) \le 1$

2　　$P(\Omega)=1$

3　　$A_1 A_2 \cdots$ 不相容　　$P(A_1+A_2+\cdots) = P(A_1)+P(A_2)+\cdots$

性质1　$P(\phi)=0$　　$\Omega = \Omega+\phi+\phi+\phi+\cdots$

$P(\Omega) = P(\Omega+\phi+\phi+\cdots) = P(\Omega)+P(\phi)+P(\phi)+\cdots$

$P(\phi)+P(\phi)+\cdots = 0$　　$P(\phi)=0$

2　有限可加　$P(A_1+A_2+\cdots+A_n) = P(A_1)+P(A_2)+\cdots+P(A_n)$

$A_1, A_2, \cdots, A_n, \phi, \phi, \cdots$ 不相容

$P(A_1+\cdots+A_n) = P(A_1)+P(A_2)+\cdots+P(\phi)+\cdots$

3　$P(A^c) = 1-P(A)$　　$P(A)+P(A^c)=1$

$A \cap A^c = \phi$　　$A \cup A^c = \Omega$

$P(\Omega) = P(A+A^c) = P(A)+P(A^c) = 1$

推论：　$A_1 \cdots A_n$ 完备事件组 $\begin{cases} \text{两两不相容} \\ \text{并是}\ \phi \end{cases}$　　$P(A_1)+\cdots+P(A_n)=1$

4.　$P(A \setminus B) = P(A) - P(A \cap B)$　　　$A \supset B$　$P(A-B) = P(A)-P(B)$

$A = (A-B) \cup A \cap B$　　$A-B$ 与 AB 不相容　　and　$P(A) \geq P(B)$

$P(A) = P(A-B)+P(AB)$

$P(A \setminus B) = P(A) - P(A \cap B)$

☆☆☆ 5. 加法 $P(A+B)=P(A)+P(B)-P(A\cap B)$

$A+B=A+(B-A\cap B)$ 创造互不相容

$P(A+B)=P(A)+P(B-A\cap B)$ $\quad p(B)-p(B\cap(A\cup B))$

$=P(A)+p(B)-p(AB)$ $\quad =P(B)-p(AB)$

$p(A+B+C)=p(A)+p(B)+p(C)-p(AB)-p(BC)-p(AC)+p(ABC)$

例1 A B A+B $P(A\bar{B})=0.3$

0.4 0.3 0.6

例 2 $P(A)=P(B)=p(c)=\frac{1}{4}$ $P(AB)=0$ $P(AC)=p(BC)=\frac{1}{16}$

① A B C 至少一个发生 ② ABC 都不发生

$p(A+B+C)=\frac{1}{4}\times 3-0-\frac{1}{16}-\frac{1}{16}-0$ $\quad P(\bar{A}\bar{B}\bar{C})=1-p(A+B+C)=\frac{3}{8}$

$=\frac{5}{8}$

$ABC\subset AB$ $\quad 0\le P(ABC)\le P(AB)=0$

例 3: 4白3黑 取3个 至少2个白. $P(2W)=\frac{{}^{4}C_2\,{}^{3}C_1+{}^{4}C_3}{{}^{7}C_3}$

例 4: ① 不需看 $\to 0.9$ 至少一台需看的概率

② 不需看 $\to 0.8$ $P(A_1+A_2)=0.1+0.2-0.02=0.28$

都要看 $\to 0.02$

例 5: 20件: 一: 6 二: 10 三: 4 取3件 2件相同

$1-\frac{{}^{6}C_1\,{}^{10}C_1\,{}^{4}C_1}{{}^{20}C_3}=\frac{15}{19}$

例 6: n个人至少2人生日相同 $\quad 1-\frac{\overbrace{365\times 364\times\cdots}^{n}}{365^n}$

选座位

条件概率

男:50人 女:50人 100人 Ω

胖 男:30 女:10 吃到胖学生中男生占: $\frac{30}{40}$

定义: Ω样本空间, A B两个事件 $P(B)>0$, 在B已经发生的条件下A发生的概率 $P(A|B)$

① $P(A|B) = \frac{n_{AB}}{n_B}$ ② $P(A|B) = \frac{n_{AB}/n}{n_B/n} = \frac{P(A\cap B)}{P(B)}$

例3: 1~6 个球 B:偶数 (1) $P(A_1)=\frac{1}{6}$ $P(A_1|B)=0$

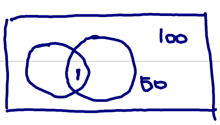
A_1: 1号 A_2: 2号 A_3: >4 (2) $P(A_2)=\frac{1}{6}$ $P(A_2|B)=\frac{1}{3}$

(3) $P(A_3)=\frac{1}{3}$ $P(A_3|B)=\frac{1}{3}$

(1) $P(A|B)\geq 0$ (2) $P(\Omega|B)=1$ (3) $A_1 A_2 \cdots A_n \cdots$ 不相容 $P(\sum_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$

乘法公式 $P(AB)=P(B)P(A|B)$ $P(A)>0$ $P(B)>0$

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



$\frac{1}{100} = \frac{50}{100} \times \frac{1}{50}$

$P(ABC) = P(A)P(B|A)P(C|AB)$ 第一步✓ 第一步✓前提下第2步 ✓

例3: 产品100件, 次品10%, 不放回, 第3次才合格

$P(\bar{A}\bar{B}C) = \frac{10}{100} \times \frac{9}{99} \times \frac{90}{98} = 0.00835$

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

$P(ABC) = P(A)P(B|A)P(C|AB) = \cancel{P(A)}\frac{\cancel{P(AB)}}{\cancel{P(A)}}\frac{P(ABC)}{\cancel{P(AB)}}$

例1 甲占60% 乙占40% 甲合格90% 乙合格80%

$P(\text{甲合格}) = 0.6\times 0.9 = 0.54$ $P(\text{合}|\text{乙}) = 0.4\times 0.8 = 0.32$

例2. 10签 4难 甲乙丙3人抽 1)甲难 2)甲乙难 3)甲易乙难 4)3人难

A B C 表示甲乙丙抽难

1) $P(A)=\frac{4}{10}$ 2) $P(AB)=P(A)P(B|A)=\frac{4}{10}\times\frac{3}{9}=\frac{6}{45}$

3) $P(\bar{A}B)=\frac{6}{10}\times\frac{4}{9}$ 4) $P(ABC)=P(A)P(B|A)P(C|AB)$

$=\frac{4}{10}\times\frac{3}{9}\times\frac{2}{8}$

例4 (传染病) a红 b黑 放入c个颜色相同的球 c=0 放回

三次都是红 A_1 A_2 A_3 表示1,2,3摸红 c=-1 不放回

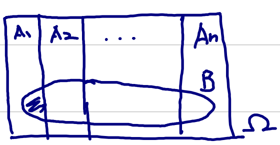
$P(A_1A_2A_3)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)$ c>0

$=\frac{a}{a+b}\times\frac{a+c}{a+b+c}\times\frac{a+2c}{a+b+2c}$

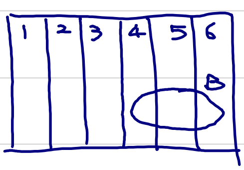
全概率公式

$A_4A_5A_6$ 互不相容 包住B

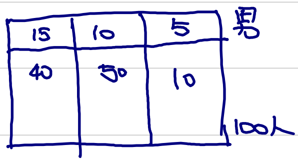
定理1.2 $A_1 A_2 \cdots A_n$ 是完备事件组 $P(A_i)>0$ $P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$



$P(A_1)P(B|A_1)+\cdots+P(A_n)P(B|A_n)$



$P(A_4)P(B|A_4)+\cdots+P(A_6)P(B|A_6)$



男30人

$P(B)=\frac{40}{100}\,\frac{15}{40}+\frac{50}{100}\times\frac{10}{50}+\frac{10}{100}\times\frac{5}{10}$

例1 四生产线

	一	二	三	四
	15%	20%	30%	35%
不合格	0.05	0.04	0.03	0.02

B:不合格

$P(B)=15\%\times0.05+0.2\times0.04+0.3\times0.03+0.35\times0.02=0.0315$

例2： 10台 3台次 卖了2台，再取1台正品

B 第三次正　A_0：两台次　A_1：一次一正　A_2：2正

$$P(B)=P(A_0)P(B|A_0)+P(A_1)P(B|A_1)+P(A_2)P(B|A_2)$$

$$=\frac{{}^{3}C_2}{{}^{10}C_2}\cdot\frac{7}{8}+\frac{{}^{3}C_1\,{}^{7}C_1}{{}^{10}C_2}\cdot\frac{6}{8}+\frac{{}^{7}C_2}{{}^{10}C_2}\cdot\frac{5}{8}=0.7$$

例3： 10产品　次品 $\begin{cases}0\\1\\2\end{cases}$ 件 $\begin{matrix}\frac{1}{3}\\\frac{1}{3}\\\frac{1}{3}\end{matrix}$　正品 $\begin{cases}\nearrow 正\ 0.98\\ \searrow 次\ 0.02\end{cases}$　次品 $\begin{cases}\nearrow 0.05正\\ \searrow 0.95正\end{cases}$

产品通过验证　B：通过　$A_0\,A_1\,A_2$：0, 1, 2　B_1 抽正，$\bar{B_1}$ 抽次

$$P(A_0)=\frac{1}{3}\quad P(A_1)=\frac{1}{3}\quad P(A_2)=\frac{1}{3}$$

$$P(B_1|A_0)=1\quad P(B_1|A_1)=\frac{9}{10}\quad P(B_1|A_2)=\frac{4}{5}$$

$$P(B_1)=P(A_0)P(B_1|A_0)+P(A_1)P(B_1|A_1)+P(A_2)P(B_1|A_2)=0.9\quad P(\bar{B_1})=0.1$$

$$P(B)=P(B_1)P(B|B_1)+P(\bar{B_1})P(B|\bar{B_1})=0.9\times0.98+0.1\times0.05=0.887$$

贝叶斯公式（果→因）

全概率公式　因→果

$$P(A_4|B)\quad \text{refer to last page}$$

$$=\frac{P(A_4B)}{P(B)}=\frac{P(A_4)P(B|A_4)}{P(B)}=0.222$$

	一	二	三	四
1万件	1500	2000	3000	3500
不合格	75	80	90	70

$$P(A_3|B)=\frac{P(A_3B)}{P(B)}=\frac{P(A_3)P(B|A_3)}{P(B)}$$

定理1.3　$A_1\cdots A_n$ 完备　B.　$P(A_i)>0$　$P(B)>0$

$$P(A_k|B)=\frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}=\frac{P(A_kB)}{P(B)}$$

一小块 / 阴影

全概率公式

$P(A_i)$ 先验　$P(A_i|B)$ 后验

例 2：发病率 0.0004

患者 → 有病 99%
患者 → 没病 1%

健人 → 有病 0.1%
健人 → 没病 99.9%

检验有病，真有病概率

A 患病　$\bar{A}$：健人　B 检验有病　$P(B|A)=0.99$

$P(A)=0.0004$　$P(B|\bar{A})=0.001$

$P(\bar{A})=0.9996$

$P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})=0.0004\times0.99+0.9996\times0.001=0.0013956$

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B|A)}{P(B)}=\frac{0.0004\times0.99}{0.0013956}=0.284$$

1万人

9996 健人　4人有病

10人有病　4人有病

$\frac{4}{14}=\frac{2}{7}$

事件的独立性：　A的概率不受B发生与否影响　$P(A|B)=P(A)$

定理1.4：$P(A)>0$　$P(B)>0$　A,B 独立 $\Leftrightarrow P(AB)=P(A)P(B)$

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

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

补：$P(A)=0$ or $P(B)=0$

$P(A)=0$　$AB\subset A$　$0\le P(AB)\le P(A)=0$，$P(AB)=0=P(A)P(B)=0$

定义 1.6　$P(AB)=P(A)P(B)\Rightarrow A,B$ 独立

ϕ 与 A 独立 $P(\phi A)=P(\phi)=0$ $P(\phi)P(A)=0$

Ω 与 A 独立 $P(\Omega A)=P(A)$ $P(\Omega)P(A)=P(A)$ $P(A)=0$ 不是 ϕ

定理1.5 (1) AB独立 A与 $\bar{B}$，$\bar{A}$与B，$\bar{A}$与$\bar{B}$独立 $P(A)=1$ 不是 Ω

(2) $P(A)=0$ 或 $P(A)=1$，A与任何事件独立

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

$=P(A)-P(A)P(B)$

$=P(A)(1-P(B))$

$=P(A)P(\bar{B})$

(2) $P(A)=0$ $AB\subset A$ $0\le P(AB)\le P(A)=0$ $P(AB)=0=P(A)P(B)=0$

$P(A)=1$ $P(\bar{A})=0$ $\bar{A}$与B独立 A与B独立

独立：可能性

互不相容：$AB=\phi$ $P(A)>0$ $P(B)>0$ 独立与互不相容不同时成立

① AB独立 $P(AB)=P(A)P(B)>0$

② $AB=\phi$ $P(AB)=P(A)P(B)>0$

A B C 独立：① $P(AB)=P(A)P(B)$ ② $P(BC)=P(B)P(C)$

③ $P(AC)=P(A)P(C)$ ④ $P(ABC)=P(A)P(B)P(C)$

必须指出独立

例2：$P(A+B)=0.9$ $P(A)=0.4$ 求$P(B)$ ①，A B 互不相容 $AB=\phi$ $P(AB)=0$

$P(A+B)=P(A)+P(B)-P(AB)$

$P(B)=0.5$

② A B 独立 $P(AB)=P(A)P(B)$ $P(A+B)=P(A)+P(B)-P(A)P(B)$

$0.9=0.4+P(B)-0.4P(B)$

$P(B)=\frac{5}{6}$

例3 甲乙丙各投一次 投中 0.7，0.8，0.75

1) 恰有一人投中 $P(A\bar{B}\bar{C}+\bar{A}B\bar{C}+\bar{A}\bar{B}C)=P(A)P(\bar{B})P(\bar{C})+P(\bar{A})P(B)P(\bar{C})+P(\bar{A})P(\bar{B})P(C)$

2) 都中 $P(ABC)=P(A)P(B)P(C)$

3) $P(A+B+C)=1-P(\bar{A}\bar{B}\bar{C})=1-P(\bar{A})P(\bar{B})P(\bar{C})$

例4：破译密码，每人译出0.6，以99%译出，至少几个人？

A_i 第i个人能译出，$B=\bigcup_{i=1}^{n}A_i$ $P(B)=1-P(\bigcap_{i=1}^{n}\bar{A_i})=1-\prod_{i=1}^{n}P(\bar{A_i})=1-0.4^n$

$1-0.4^n\geqslant 0.99$

$n\geqslant 5.026$

例5. $0<P(A)<1$ $0<P(B)<1$ $P(A|B)+P(\bar{A}|\bar{B})=1$

$P(A|\bar{B})+P(\bar{A}|\bar{B})=1$

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

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

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

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

伯努利模型

独立实验序列：$E_1,E_2,\cdots,E_n$ 独立

n重独立实验：$E,E,\cdots,E$ 独立 E^n

伯努利实验：结果只有两种

n重伯努利实验：n次独立，结果只有2种

定理：A的概率 $p\,(0<p<1)$ $\bar{A}:1-p$，n重伯努利中发生k次：

$$P_n(k)={}^{k}C_n\,p^k(1-p)^{n-k}$$ 二项概率公式

$$(a+b)^n=(a+b)(a+b)\cdots(a+b)$$

$${}^{n}C_n\,a^n+{}^{n}C_{n-1}\,a^{n-1}b+{}^{n}C_{n-2}\,a^{n-2}b^2+\cdots$$

例1 废品率0.1 合格率0.9，取一个放回去 三次

(1) 恰有1次废品 ${}^{3}C_1\times0.1\times0.9^2$

(2) 恰有2次废品 ${}^{3}C_2\times0.1^2\times0.9$

(3) 三次都废 ${}^{3}C_3\times0.1^3$

(4) 三次取正品 ${}^{3}C_0\times0.9^3$

例2　彩票中奖　十万分之一

买，520次，从未中奖

$$P_{520}(0) = C_{520}^{0} (10^{-5})^{0} (1-10^{-5})^{520}$$
$$= 0.99999^{520}$$
$$\approx 0.9948$$

第2章　随机变量的概念

平常结果确定

1 2 3 4 5 6　　x

正　反　　1 0

白　黑　红　0 1 2

离散型：有限个，无限可列个

$\{w: x(\omega)=a\}$ 事件　$\{x=a\}$　$p\{x=a\}$　$p(x=a)$

例6　公交站　每5分钟一辆　候车时间

$x\in[0,5]$　$p\{1\le x\le 3\}$　$p\{x\ge 0\}=1$　$p\{x>6\}=0$

非离散型：连续型

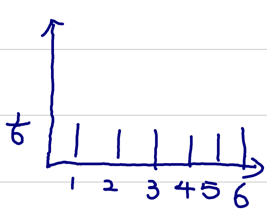
2.2 离散型随机变量及其概率分布

X	1	2	3	4	5	6
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

X	1	0
	$\frac{1}{2}$	$\frac{1}{2}$

X 的所有取值 x_k $(k=1,2,3,\cdots)$ 可以是无限可列个

(变量)　$P\{X=x_k\}=p_k$　概率函数(分布)



① $p_k \ge 0$　② $\sum p_k = 1$

X	2	4	8	16	32	$\cdots$
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\cdots$

例1：5黑3白 每次抽一个 不放回 直到取黑 X为取白数目

$P\{-1<x<0\}=0$ $P\{1<x<3\}=\frac{5}{56}$ $P\{x\le 3\}=1$

$X=0$ $P\{x=0\}=\frac{5}{8}$

$X=1$ $P\{x=1\}=\frac{3}{8}\frac{5}{7}=\frac{15}{56}$

$X=2$ $P\{x=2\}=\frac{3}{8}\frac{2}{7}\frac{5}{6}=\frac{5}{56}$

$X=3$ $P\{x=3\}=\frac{3}{8}\frac{2}{7}\frac{1}{6}\times\frac{5}{5}=\frac{1}{56}$

X	0	1	2	3
P	$\frac{5}{8}$	$\frac{15}{56}$	$\frac{5}{56}$	$\frac{1}{56}$

连续型随机变量及其概率密度函数

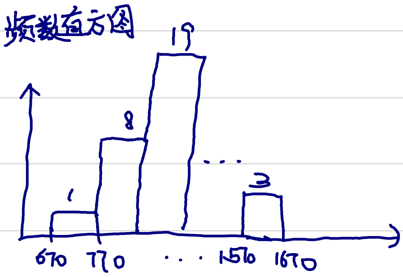
引例：身高 150-200

150~160	10
160~170	20
170~180	40
180~190	15
190~200	15

例：99年的降水量

670~770	1	0.01
770~780	8	0.08
⋮	⋮	⋮
1570~1670	3	0.03
	频数	频率

频数直方图



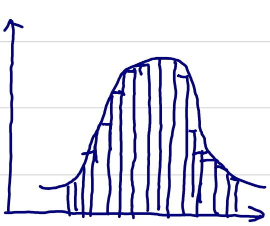
频率密度直方图

频率/组距

① 每个小长方形面积等于该组的概率 $\frac{0.01}{100}\times 100=0.01$

② 所有小长形面积相加等于1

③ 界于 $x=a$ $x=b$ 之间的面积 近似于 $(a,b]$ 的频率



$y=f(x)$

① $f(x)\ge 0$

② $\underbrace{\int_{-\infty}^{+\infty} f(x)=1}_{\text{参数}}$

定义：非负可积 $f(x)$ $f(x)\ge 0$ $a\le b$

$$P\{a<x\le b\}=\int_a^b f(x)dx$$

X：连续： $f(x)$ 概率分布密度函数

③ 连续型随机变量取个别值概率为0

$P(0.1)=0.0001$

取无穷点概率 >1 反证

$0 \le P\{X=x_0\} \le P\{x_0-\Delta x < X \le x_0\} = \int_{x_0-\Delta x}^{x_0} f(x)dx \longrightarrow 0$

连续、端点无所谓 $P\{a\le X\le b\} = P\{a< X\le b\} = P\{a\le X< b\} = P\{a< X< b\}$

$P\{X<a\} = P\{X\le a\}$ $P\{X>a\} = P\{X\ge a\}$

概率为0的事件未必不可能事件

概率为1的事件未必是必然事件 (0,1) 取到0和1

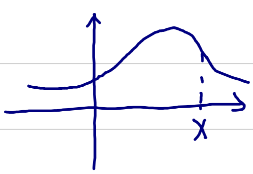
例2: $f(x) = \begin{cases} kx+1 & 0\le x\le 2 \\ 0 & \text{其他} \end{cases}$

求k $\int_{-\infty}^{+\infty} f(x)dx = \int_0^2 (kx+1)dx = 1$

$\Rightarrow k = -\frac{1}{2}$

$P\{X\le 2\} = P\{-\infty < X \le 2\} = \int_0^2 (-\frac{1}{2}x+1)dx = 1$

$P\{1.5 < X < 2.5\} = \int_{1.5}^{2} (-\frac{1}{2}x+1)dx = 0.0625$



X取x附近值的大小

$\lim_{\Delta x\to 0} \frac{P\{x<X<x+\Delta x\}}{\Delta x} = \lim_{\Delta x\to 0}\frac{\int_x^{x+\Delta x} f(x)dx}{\Delta x} = f(x)$

$P\{x<X<x+\Delta x\} \approx f(x)\Delta x$

分布函数：离散，连续

$F(x) = P(X\le x)$ X取值不超过x的概率 $x\in(-\infty,+\infty)$ $F(x)\in[0,1]$

性质：1) $0\le F(x)\le 1$ $x\in(-\infty,+\infty)$

2) F(x) 不减 $x_1 < x_2$ $F(x_1)\le F(x_2)$

$F(1)\le F(1.1)$ $P(X\le 1)$ $P(X\le 1.1)$ $X\le 1 \subset X\le 1.1$

$\lim_{x\to\infty} F(x) = F(+\infty) = 1$ $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} F(-\infty) = 0$ 用于求参数

3) F(x) 右连续 $\begin{cases} \text{离散 右连续} \\ \text{连续} \end{cases}$ 至多可列个间断点

$F(x) = P(X \leq x)$ $\quad P\{x \leq a\} = F(a)$

$P\{x > a\} = 1 - P\{x \leq a\} = 1 - F(a)$

$P\{\underline{a < x \leq b}\} = P\{x \leq b\} - P\{x \underline{\leq} a\} = F(b) - F(a)$

$P\{x = a\} = F(a) - F(a-0)$ 不包含a

对<u>离散</u>,连续都成立 $\quad \lim\limits_{x \to a^+} F(x) = F(a) \qquad \lim\limits_{x \to a^-} F(x) = F(a)$

$\lim\limits_{x \to a} F(x) = F(a)$ <u>极限值,函数值存在</u>，<u>极限值 = 函数值</u>

$P\{\underline{a \leq x < b}\} = F(b) - F(a-0)$ 减掉a左边所有的

$P\{x < a\} = F(a-0)$

$P\{x \geq a\} = 1 - F(a-0)$

例1：

$$F(x) = \begin{cases} a - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} \qquad \lambda > 0 \text{ 求 } a$$

$$\lim_{x \to \infty} \left(a - \frac{1}{e^{\lambda x}}\right) = a = 1$$

例2 ☆

X	-1	2	3
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

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

$x \in (-\infty, +\infty)$

$x = -3$

$F(-3) = P(X \leq -3) = 0$

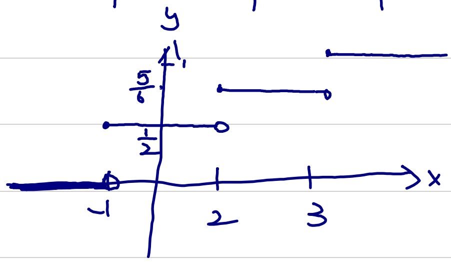
$X < -1, \quad F(x) = P(X \leq x) = 0$

$-1 \leq X < 2 \qquad F(x) = P(x = -1) = \frac{1}{2}$

$2 \leq X < 3 \qquad F(x) = P(X \leq x) = P(x=-1) + P(x=2) = \frac{5}{6}$

$3 \leq x \qquad F(x) = P(X \leq x) = P(x=-1) + P(x=2) + P(x=3) = 1$

$$F(x) = \begin{cases} 0 & x < -1 \\ \frac{1}{2} & -1 \leq x < 2 \\ \frac{5}{6} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$



X	-2	0	1	3
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

4个点、5部分

$X<-2 \quad F(x)=0$

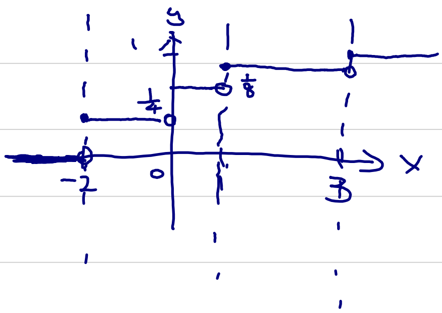
$-2\leq X<0 \quad F(x)=\frac{1}{2}$

$0\leq X<1 \quad F(x)=\frac{1}{2}+\frac{1}{4}$

$1\leq X<3 \quad F(x)=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}$

$3\leq X \quad F(x)=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{8}$

$$F(x)=\begin{cases} 0 & x<-2 \\ \frac{1}{2} & -2\leq x<0 \\ \frac{3}{4} & 0\leq x<1 \\ \frac{7}{8} & 1\leq x<3 \\ 1 & 3\leq x \end{cases}$$



右连续

X取值从小到大排好

分布函数 ⟶ 概率函数

间断点. x_k 是 X 的取值

$P\{X=x_k\}=F(x_k)-F(x_k-0)$

X	-2	0	1	3
P	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$

$\frac{3}{4}-\frac{1}{2}=\frac{1}{2}$ $\quad \frac{7}{8}-\frac{3}{4}=\frac{1}{8}$ $\quad 1-\frac{7}{8}=\frac{1}{8}$

(看高度)

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

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

例3: $f(x)=\frac{1}{\pi(1+x^2)}$ $\quad F(x)=\frac{1}{\pi}\int_{-\infty}^{x}\frac{1}{1+t^2}dt=\frac{1}{\pi}\tan^{-1}x+\frac{1}{2}$

例4 $f(x)=\begin{cases} -\frac{1}{2}x+1 & 0\leq x\leq 2 \\ 0 & 其他 \end{cases}$

$x<0$ 时 $F(x)=\int_{-\infty}^{x}0\,dt=0$

$0\leq x<2$ 时 $F(x)=\int_{-\infty}^{0}f(t)dt+\int_{0}^{x}(-\frac{1}{2}x+1)dt=-\frac{1}{4}x^2+x$

$x\geq 2$ 时 $F(x)=\int_{-\infty}^{x}f(t)dt=\int_{-\infty}^{0}0\,dt+\int_{0}^{2}(-\frac{1}{2}t+1)dt+\int_{2}^{x}0\,dt=1$

例 5 $F(x)=\begin{cases}0 & x<0\\ Ax^2 & 0\leqslant x<1\\ 1 & 1\leq x\end{cases}$ 求A

$\lim\limits_{x\to 0^+}F(x)=0=F(0)=0$ $\lim\limits_{x\to 1^-}Ax^2=A=F(1)=1$

$f(x)=\begin{cases}2x & 0\leqslant x<1\\ 0 & 其他\end{cases}$ $P\{0.3<x<0.7\}=F(0.7)-F(0.3)$

$=\int_{-\infty}^{0.7}f(t)dt-\int_{-\infty}^{0.3}f(t)dt$

$=0.49-0.09=0.4$

常见的分布

0-1分布

X	1	0
P	p	1-p

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

例 1：废品率 10% $x=\begin{cases}1 & 合格\\ 0 & 废\end{cases}$ $P\{x=1\}=0.9$ $P\{x=0\}=0.1$

有两种结果，试验只做一次

几何分布 $P(A)=p$ 第k次首次发生 前k-1次未发生

$P\{x=k\}=(1-p)^{k-1}p$ $x\sim G(p)$

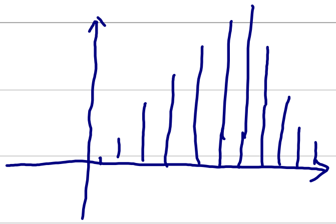
例：射中0.6 X：直到命中目标的次数

$P\{x=k\}=0.4^{k-1}0.6$ $k=1,2,3,\cdots$

二项分布 $P(A)=p$ n次试验发生了k次

$P\{x=k\}={}^{n}C_{k}\,p^k(1-p)^{n-k}$ $k=0,1,2,\cdots,n$ $X\sim B(n,p)$

$n=1$ 时 $P\{x=k\}={}^{n}C_{1}\,p(1-p)^{1-k}$



$n=10 \quad p=0.7$

最可能值

1) $(n+1)p$ 不为整数 $[(n+1)p]$ 达最大值

2) $(n+1)p$ 是整数 $(n+1)p$ $(n+1)p-1$ 最大值

例3：报 0.8，99% 报警

X：台数 n：安装台数 $X \sim B(n, 0.8)$

$0.99 \le P\{X \ge 1\} = 1 - P\{X=0\} = 1 - {}^{n}C_0\, 0.2^n = 1 - 0.2^n$

$1 - 0.2^n \ge 0.99$

$0.01 \ge 0.2^n$

$\ln 0.01 \ge n \ln 0.2$

$n \ge \dfrac{\ln 0.01}{\ln 0.2}$

例4 每台机床维修 $p=0.01$

1) 1个人看 20台 不能及时维修的概率

2) 3人看 80台

1) $n=20 \quad p=0.01 \quad X$ 修的台数

$P(X>1) = 1 - P\{X=0\} - P\{X=1\}$

$= 1 - {}^{20}C_0\, 0.99^{20} - {}^{20}C_1\, 0.01 \times 0.99^{19}$

≈ 0.0169

2) $n=80 \quad p=0.01 \quad P\{X>3\} = 1 - P\{X=0\} - P\{X=1\} - P\{X=2\} - P\{X=3\}$

≈ 0.0087

泊松分布 $P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda}$ $k=0,1,2,3,\cdots$

$\lambda>0$ $X\sim P(\lambda)$

电台收呼叫次数，公用设施（等车，收银，挂号）

$\lambda=0.1$ $k=0$ $e^{-0.1}$

$k=1$ $\frac{0.1}{1}e^{-0.1}$

$k=2$ $\frac{0.1^2}{2!}e^{-0.1}$

查表

$k\backslash\lambda$	0.1	0.2	0.3	$\cdots$
0				
1				
2				
3				
$\vdots$				

二项分布 · 0.99^{20}，0.99^{19} 用泊松分布近似当 n 比较大，p 较小

($n\geq100$ $np\leq10$)

例5. 电话台 $X\sim P(3)$ $\lambda=3$

不超过 5次

$P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{3^k}{k!}e^{-\lambda}$ $k=0,1,2,\cdots$

$P\{X\leq5\} = \sum_{k=0}^{5} P\{X=k\} = 0.916$ （查表0 1 2 345）

例6. 1000个帐户 每户10万元，每户提20%概率是0.006

$10\times0.2=20000$元

准备多少现金以95%以上概率

X: 提钱的用户数 $X\sim B(1000,0.006)$

现金x元 $P\{2X\leq x\}\geq0.95$

$P\{X\leq\frac{x}{2}\}\geq0.95$ $\sum_{k=0}^{\frac{x}{2}}\frac{6^k}{k!}e^{-6}\geq0.95$ $\frac{x}{2}\geq10$ $x\geq20$

找到$\lambda=6$，然后全部加，加到≥0.95

例7：发病率 $\frac{1}{1000}$，单位5000人 至少2人得病的概率

X 得病人数 $X\sim B(5000,\frac{1}{1000})$ $X\sim P(5)$

$p\{x\geq 2\}=1-p\{x=0\}-p\{x=1\}$ $p\{x\geq 2\}=1-\sum_{k=0}^{1}\frac{5^k}{k!}e^{-5}$

$=1-0.006738-0.03369$

$=0.959572$

超几何分布 100学生 男60 女40 取10人

X：取10人中男生人数

$$p\{X=k\}=\frac{{}^{60}C_k\,{}^{40}C_{10-k}}{{}^{100}C_{10}}\quad k=0,1,2,\cdots,10$$

N个元素：N_1个属于第一类，N_2个属于第二类

取n个 X：取n个属于第一类的个数

$$P\{x=k\}=\frac{{}^{N_1}C_k\,{}^{N_2}C_{n-k}}{{}^{N}C_n}\quad k=0,1,2,\cdots,\min\{n,M\}$$

例8：20名学生（5女 15男）任取4名 X：4名中女生人数

$$p\{x=k\}=\frac{{}^{5}C_k\,{}^{15}C_{4-k}}{{}^{20}C_4}\quad k=0,1,2,3,4.$$

不放回抽样试验

100000粒 发芽99% 取出10粒 N很大，n相对于N很小

$p=\frac{M}{N}$ 改变小

$$p\{x=k\}=\frac{{}^{M}C_k\,{}^{N-M}C_{n-k}}{{}^{N}C_n}\approx {}^{n}C_k\,p^k(1-p)^{n-k}$$

例9：10000粒 发芽率99% 取200粒 最多一粒不发芽

$10000\times 1\%=100$ 粒 N_1 发芽 9900 粒 N_2

$N=10000$ $N_1=100$ $n=200$

$$p\{x\leq 1\}=p\{x=0\}+p\{x=1\}=\frac{{}^{100}C_0\,{}^{9900}C_{200}}{{}^{10000}C_{200}}+\frac{{}^{100}C_1\,{}^{9900}C_{199}}{{}^{10000}C_{200}}$$

$n=200 \quad p=0.01$

$P\{X\le 1\} = {}^{200}C_0\, 0.01^0\, 0.99^{200} + {}^{200}C_1\, 0.01\times 0.99^{199}$

$n=200 \quad p=0.01 \quad np=2 \quad \lambda=2$

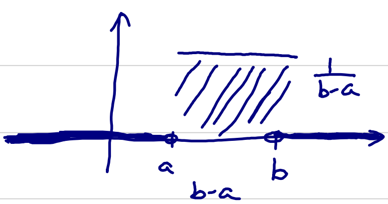
$P\{X\le 1\} = 0.1353 + 0.2707 = 0.406$

二项分布 $n\ge 100$ $np\le 10$ 用泊松分布近似计算 $\lambda=np$

超几何分布 $\xrightarrow{N\text{大}\ \frac{n}{N}\text{小}}$ 二项分布 $\xrightarrow{\text{近似}}$ 泊松

连续型分布

均匀分布 $f(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & \text{else}\end{cases}$ $\quad X\sim U[a,b]$



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

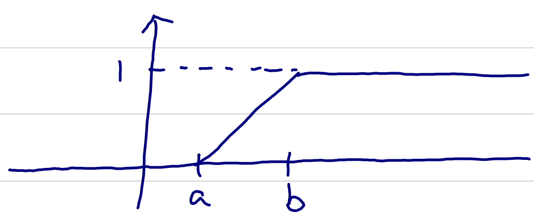
$F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\le x<b\\ 1 & x\ge b\end{cases}$

☆ $\int_{-\infty}^{x} f(t)\,dt$

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

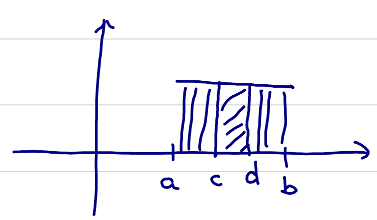
$=\int_{a}^{x}\frac{1}{b-a}\,dt$

$=\frac{1}{b-a}(x-a)$



$X\sim U[a,b] \qquad [c,d]\subset[a,b]$

$P\{c\le X\le d\}=\int_c^d \frac{1}{b-a}\,dt=\frac{d-c}{b-a}$



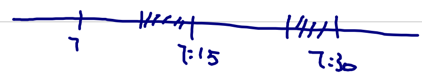
例1. 公共汽车从7点、，每15分开车

乘客7点到7点半到车站是均匀分布

1) 等车不超过5分钟　　2) 等车超10分钟　　$P\{0<X<5\}+P\{15<X<20\}$

过X分　$X \sim U[0,30]$　　　　$=\frac{1}{3}$

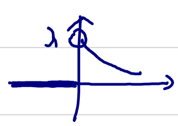
$$f(x)=\begin{cases}\frac{1}{30} & 0\le x\le 30\\ 0 & \text{else}\end{cases}$$



$$P\{10\le X\le 15\}+P\{25\le X\le 30\}=\frac{5}{30}+\frac{5}{30}=\frac{1}{3}$$

指数分布　$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$　$\lambda>0$　$X\sim Exp(\lambda)$

$\frac{\lambda}{e^{\lambda x}}$



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

$$F(x)=P\{X\le x\}$$
$$=\int_{-\infty}^{x} f(t)dt$$
$$=\int_{-\infty}^{0}+\int_{0}^{x}\lambda e^{-\lambda t}dt$$
$$=\int_{0}^{x} e^{-\lambda t}d(-\lambda t)$$
$$=\left[-e^{-\lambda t}\right]_{0}^{x}=1-e^{-\lambda x}$$

服务时间，寿命

例2　$f(x)=\begin{cases}\frac{1}{1000}e^{-\frac{x}{1000}} & x>0\\ 0 & x\le 0\end{cases}$

3个元件　$P\{x>1000\}=\int_{1000}^{+\infty}\frac{1}{1000}e^{-\frac{x}{1000}}dx$

$$=-e^{-\frac{x}{1000}}\Big]_{1000}^{\infty}=e^{-1}$$

$$\left(P(x>1000)\right)^3=e^{-3}$$

例3：X 指数分布 λ　$s>0$　$t>0$　证 $P\{x>s+t : x>s\}=P\{x>t\}$

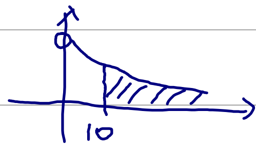
证：$P\{x>s+t : x>s\}$　　$P\{A|B\}=\frac{P(AB)}{P(B)}$　　$P\{x>t\}$

$$=\int_{t}^{\infty}\lambda e^{-\lambda x}dx=e^{-\lambda t}$$

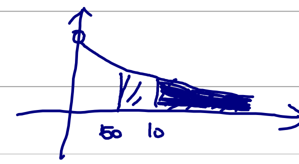
☆　$=\frac{P\{\{x>s+t\}\cap\{x>s\}\}}{P\{x>s\}}$

$$=\frac{P\{x>s+t\}}{P\{x>s\}}=\frac{\int_{s+t}^{\infty}\lambda e^{-\lambda x}dx}{\int_{s}^{\infty}\lambda e^{-\lambda x}dx}=\frac{-e^{-\lambda x}\big|_{s+t}^{\infty}}{-e^{-\lambda x}\big|_{s}^{\infty}}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}$$

$P\{X>10\}$



$P\{X>50+10 : X>50\}$



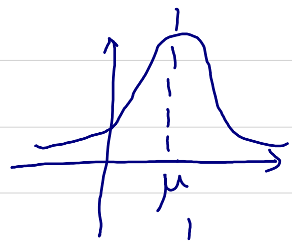
正态分布

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

$$\int_{-\infty}^{\infty} e^{-x^2}dx=\sqrt{\pi}$$

1. $\frac{1}{\sqrt{\pi}\sigma}\int_{-\infty}^{\infty} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2} d\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)=\frac{1}{\sqrt{\pi}}\sqrt{\pi}=1$

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



性 1) $y=\varphi(x)$ 以 $x=\mu$ 为对称轴 bell curve

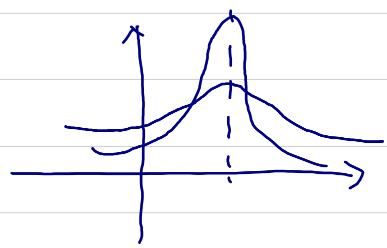
$x=\mu$ 时 $\varphi(x)$ 最大值 $\frac{1}{\sqrt{2\pi}\sigma}$

2) $y=\varphi(x)$ 以 x 轴为渐近线 $x=\mu\pm\sigma$ 拐点.

3) σ 固定 μ 变化 左右移动

μ 固定 σ 变化↓ 最高点↑

σ↑ 最高点↓

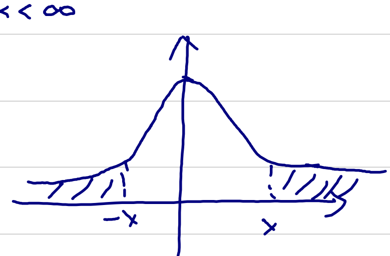


标准正态分布 $\mu=0$ $\sigma=1$ $\phi_0(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ $-\infty<x<\infty$

$$\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

性质:1) 以 y 轴为对称轴 偶函数

$\phi_0(x)=\phi_0(-x)$ ☆ $\Phi_0(-x)=1-\Phi_0(x)$



查表: 给 $0\le x<5$ $x\ge 5$ $\phi_0(x)=0$ $\Phi_0(x)=1$ 面积、

$x\le -5$ $\phi_0(x)=0$ $\Phi_0(x)=0$

$\Phi_0(-4)=1-\Phi_0(4)$

因为 $\varphi_0(x)=\frac{1}{\sqrt{2\pi}}x^{-\frac{x^2}{2}}$

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

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{\infty}^{x}e^{-\left(\frac{t-\mu}{\sigma}\right)^2}d\left(\frac{t-\mu}{\sigma}\right)$$

☆ $\varphi(x)=\frac{1}{\sigma}\varphi_0\left(\frac{x-\mu}{\sigma}\right)$

$\Phi(x)=\Phi_0\left(\frac{x-\mu}{\sigma}\right)$

(4) 求 $\varphi_0(1.65)$ $\quad\varphi_0(-1)=\varphi_0(1)$ $\quad\varphi_0(6.4)=0$ $\quad X\sim N(0,1)$

$\Phi_0(1.65)$ $\quad\Phi_0(-1)=1-\Phi_0(1)$ $\quad\Phi_0(6.4)=1$

$P\{|x|\leq 1.96\}=P\{-1.96\leq X\leq 1.96\}=\Phi_0(1.96)-\Phi_0(-1.96)$

$=\Phi_0(1.96)-1+\Phi_0(1.96)$

$=2\Phi_0(1.96)-1$

(5) $X\sim N(\overset{\mu}{1},\overset{\sigma^2}{4})$ $\quad\mu=1\quad\sigma=2$

$P\{0<X<1.6\}=\Phi(1.6)-\Phi(0)$

$=\Phi_0\left(\frac{1.6-1}{2}\right)-\Phi_0\left(\frac{0-1}{2}\right)$

$=\Phi_0(0.3)-\Phi_0(-0.5)=\Phi_0(0.3)-1+\Phi_0(0.5)$

$P\{|x|\leq 2\}=\Phi(2)-\Phi(-2)$

$=\Phi_0\left(\frac{2-1}{2}\right)-\Phi_0\left(\frac{-2-1}{2}\right)$

$=\Phi_0\left(\frac{1}{2}\right)-\Phi_0\left(-\frac{3}{2}\right)=\Phi_0\left(\frac{1}{2}\right)-1+\Phi_0\left(\frac{3}{2}\right)$

$$P\{|X|\leq 2\}=P\{-2\leq X\leq 2\}=P\left\{\frac{-2-1}{2}\leq\frac{X-1}{2}\leq\frac{2-1}{2}\right\}=\Phi_0(0.5)-\Phi_0(-1.5)$$

(6) 长度 $X\sim N(\overset{\mu}{50},\overset{\sigma^2}{1})$ $\quad\mu=50\quad\sigma=1$ $\quad 50\pm1$ 是合格

$P\{49\leq x\leq 51\}=\Phi(51)-\Phi(49)=\Phi_0\left(\frac{51-50}{1}\right)-\Phi_0\left(\frac{49-50}{1}\right)=\Phi_0(1)-\Phi_0(-1)$

$=2\Phi_0(1)-1$

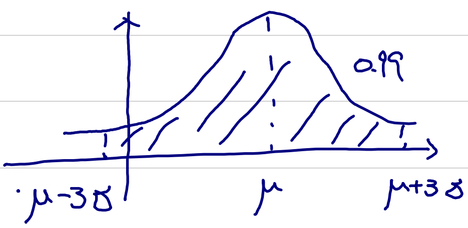
$P\{Y\geq 1\}=1-P\{Y=0\}=1-(1-0.6826)^3\approx 0.968$

7. $X \sim N(\mu, \sigma^2)$ $P\{|x-\mu|<\sigma\}$ $P\{|x-\mu|<2\sigma\}$ $P\{|x-\mu|<3\sigma\}$

$P\{|x-\mu|<\sigma\} = P\{-\sigma<x-\mu<\sigma\} = P\{\mu-\sigma<x<\mu+\sigma\} = \Phi(\mu+\sigma)-\Phi(\mu-\sigma)$

$= \Phi_0\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - \Phi_0\left(\frac{\mu-\sigma-\mu}{\sigma}\right) = \Phi_0(1)-\Phi_0(-1) = 0.6826$

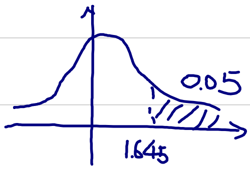
$P\{|x-\mu|<2\sigma\} = 0.9544$ $P\{|x-\mu|<3\sigma\} = 0.9974$



$X \sim N(0,1)$ 给定 α $(0<\alpha<1)$ 找 u_α $P\{x>u_\alpha\}=\alpha$

u_α 叫上 α 分位数

$u_{0.05} = 1.645$ $P\{x>1.645\} = 0.05$



$u_{0.025} = 1.96$ $u_{0.01} = 2.33$

面积 点

随机变量函数的分布

已知 x 是某分布 $Y=3x-5$ 分布？

离散型

1)

X	7	8	9	10
P	0.1	0.3	0.4	0.2

$Y=4x$

Y	28	32	36	40
P	0.1	0.3	0.4	0.2

$Z=x^2$

Z	49	64	81	100
P	0.1	0.3	0.4	0.2

2)

X	-1	0	1	2
P	0.2	0.3	0.4	0.1

$Y=x^2$

Y	1	0	1	4
P	0.2	0.3	0.4	0.1

$\xrightarrow{\text{重复}}$

Y	0	1	4
P	0.3	0.6	0.1

X	-2	-1	0	1	2
P	0.2	0.1	0.5	0.1	0.1

$Y=X^4-1$

Y	15	0	-1	0	15
P	0.2	0.1	0.5	0.1	0.1

Y	-1	0	15
P	0.5	0.2	0.3

连续型

设 X 的 $f_X(x)$ $y=g(x)$ $Y=g(X)$ $F_X(x)=P\{X\le x\}$

① $F_Y(x)\longrightarrow F_X(x)$ 求导 $F_Y(x)=P\{Y\le x\}$

② $f_Y(x)\longleftarrow f_X(x)$

1) X 的密度 $f_X(x)$ $Y=3X+2$

$F_Y(x)=P\{Y\le x\}=P\{3X+2\le x\}$ $f_Y(x)=\frac{1}{3}f_X\left(\frac{x-2}{3}\right)$

$P\{X\le \frac{x-2}{3}\}=F_X\left(\frac{x-2}{3}\right)$

$$f_X(x)=\begin{cases}\frac{1}{4} & 0\le x\le 4\\ 0 & \text{else}\end{cases}\qquad f_Y(x)=\begin{cases}\frac{1}{12} & 2\le x\le 14\\ 0 & \text{else}\end{cases}$$

X 服从 $[a,b]$ 均匀分布 $Y=kx+c$ $(k\ne 0)$ 服从相应区间上的均匀分布

$[ka+c,\ kb+c]$ $k>0$ $[kb+c,\ ka+c]$ $k<0$

$$k>0\quad f_Y(x)=\begin{cases}\frac{1}{kb-ka} & ka+c\le x\le kb+c\\ 0 & \text{else}\end{cases}\qquad k<0\quad f_Y(x)=\begin{cases}\frac{1}{ka-kb} & kb+c\le x\le ka+c\\ 0 & \text{else}\end{cases}$$

2) $X\sim N(\mu,\sigma^2)$ $Y=aX+b$ $a\ne 0$

$a>0$ 时 $F_Y(x)=P\{Y\le x\}=P\{aX+b\le x\}=P\{X\le \frac{x-b}{a}\}=\Phi\left(\frac{x-b}{a}\right)$

$$f_Y(x)=\frac{1}{a}\varphi\left(\frac{x-b}{a}\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(\frac{x-b}{a}-\mu\right)^2}{2\sigma^2}}\cdot\frac{1}{a}$$

$$=\frac{1}{\sqrt{2\pi}a\sigma}e^{-\frac{(x-(b+a\mu))^2}{2\sigma^2a^2}}$$ 对称轴

$N(a\mu+b,\ a^2\sigma^2)$

$a<0$ 时 $f_Y(x) = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}$

线性：$2x+3y$，$x_1+x_2-x_3$，$x+8y+17z+5$

$Y=\frac{x-\mu}{\sigma}$ $N(0,1)$

定理 2.1 X的密度 $f_X(x)$ $Y=kx+b\ (k\neq 0)$ $f_Y(x) = \frac{1}{|k|} f_X(\frac{x-b}{k})$

$k<0$，$F_Y(x) = P\{Y\le x\} = P\{kX+b\le x\} = P\{X\ge \frac{x-b}{k}\} = 1-P\{X<\frac{x-b}{k}\}$

$= 1-F_X(\frac{x-b}{k})$

$f_Y(x) = -\frac{1}{k} f_X(\frac{x-b}{k})$

$= \frac{1}{|k|} f_X(\frac{x-b}{k})$

$k>0$ 时 $f_Y(x) = \frac{1}{k} f_X(\frac{x-b}{k})$

(3) $X\sim N(0,1)$ $Y=X^2$

$x<0$ 时 $F_Y(x) = P\{Y\le x\} = P\{X^2\le x\} = 0$

$x\ge 0$ 时 $F_Y(x) = P\{Y\le x\} = P\{X^2\le x\} = P\{-\sqrt{x}\le X\le\sqrt{x}\}$

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

$= 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

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

$= \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}} & x>0 \\ 0 & x\le 0 \end{cases}$

$\int_0^{x^2} \frac{1}{t+1} dt = \ln(1+x^2)$

4) X $f(x) = \begin{cases} \frac{1}{x+1} & 0<x<e-1 \\ 0 & \text{else} \end{cases}$ $Y=\sqrt{x}$

$F_Y(x) = \begin{cases} 0 & x<0 \\ \ln(1+x^2) & 0\le x<\sqrt{e-1} \\ 1 & x\ge\sqrt{e-1} \end{cases}$

$x<0$ 时 $F_Y(x) = P\{Y\le x\} = P\{\sqrt{X}\le x\} = 0$

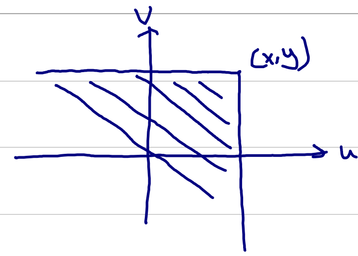
$x\ge 0$ 时 $P\{X\le x^2\} = \int_{-\infty}^{x^2} f(t)\, dt$

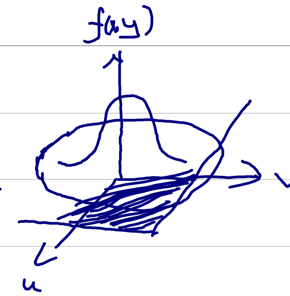
$f_Y(x) = \begin{cases} \frac{2x}{1+x^2} & 0\le x<\sqrt{e-1} \\ 0 & \text{else} \end{cases}$

$= \int_{-\infty}^{0} f(t)\, dt + \int_0^{x^2} f(t)\, dt$ $\begin{cases} 1 & x^2\ge e-1 \\ \ln(1+x^2) & 0\le x^2<e-1 \end{cases}$

第三章 多维随机变量及其分布

身材 身高 体重 三维





3.1 二维随机变量

E试验 Ω空间 X,Y是Ω的两个变量

(X,Y) 变量 向量

分布函数 $F(x,y)=P\{X\le x, Y\le y\}$ 联合分布 $F(x)=P\{X\le x\}$

(1) $0\le F(x,y)\le 1$

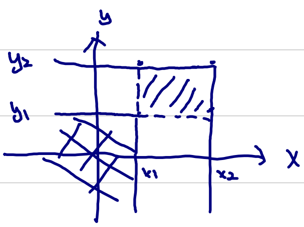
(2) $F(x,y)$ 不减

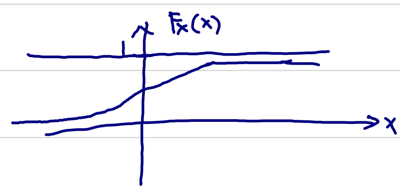
(3) $F(-\infty,y)=0$ $F(x,-\infty)=0$ $F(-\infty,-\infty)=0$ $F(+\infty,+\infty)=1$

(4) $F(x,y)$ 分别关于x和y的右连续

(5) $x_1<x_2$ $y_1<y_2$ $P\{x_1<X\le x_2, y_1<Y\le y_2\}$

$= F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)$ 减多了一次





$F_X(x)=P\{X\le x\}=F(x,+\infty)=P\{X\le x\ Y<+\infty\}$

$F_Y(y)=P\{Y\le y\}=F(+\infty,y)=P\{X<+\infty, Y\le y\}$

二维离散型的联合分布及边缘分布

X\Y	1	2	3
1	0	$\frac{1}{2}$	$\frac{1}{8}$
2	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

分布表

$F(-1,-2)=P\{X\le -1\ Y\le -2\}=0$

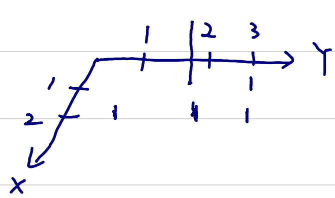
$F(1,2)=P\{X\le 1, Y\le 2\}=\frac{1}{2}$

$F(4,5)=P\{X\le 4, Y\le 5\}=1$

$F(1.5,2.6)=\frac{1}{2}$

(1) $p_{ij}\ge 0$ (2) $\sum_i\sum_j p_{ij}=1$

$F(x,y)=P\{X\le x\ Y\le y\}=\sum_{x_i\le x}\sum_{y_j\le y} p_{ij}$



边缘分布

对行求和，得X边缘分布

对列求和，得Y边缘分布

X\Y	1	2	3
1	0	$\frac{1}{2}$	$\frac{1}{8}$
2	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

X	1	2
P	$\frac{5}{8}$	$\frac{3}{8}$

Y	1	2	3
P	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{4}$

① 联合分布可唯一确定边缘分布

② 边缘分布不能确定联合分布

二维连续的联合密度和边缘密度

$$F(x,y)=P\{X\le x\ \ Y\le y\}=\int_{-\infty}^{x}\int_{-\infty}^{y}f(s,t)\,ds\,dt$$

(1) $f(x,y)\ge 0$ (2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=1$ (3) $\frac{\partial^2 F(x,y)}{\partial x\,\partial y}=f(x,y)$

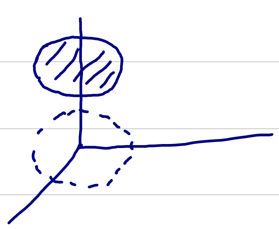
(4) G XY平面的区域 $P\{(X,Y)\in G\}=\iint_G f(x,y)\,dx\,dy$

1) $f(x,y)=\begin{cases} C & (x,y)\in G \\ 0 & \text{else}\end{cases}$ $G: x^2+y^2\le r^2$

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

$$=\iint_G 1\,dx\,dy=C\pi r^2$$

$\Rightarrow C=\frac{1}{\pi r^2}$ 面积倒数

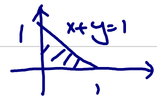


$$f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & \text{else}\end{cases}$$

$$f(x,y)=\begin{cases}\frac{1}{S(G)} & (x,y)\in G\\ 0 & \text{else}\end{cases}$$

(2) $f(x,y) = \begin{cases} e^{-(x+y)} & x>0 \ y>0 \\ 0 & \text{else} \end{cases}$

(1) X, Y 的 $F(x,y)$

(2) 

(3) $F_X(x)$ $F_Y(y)$

(1) $F(x,y) = P\{X \le x \ \ Y \le y\}$ $x>0$ $y>0$ $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, ds\, dt$

$= \int_0^x \int_0^y e^{-s} e^{-t}\, ds\, dt$

x or $y<0$ $F(x,y)=0$

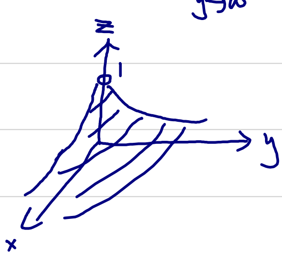
$= \int_0^x e^{-s} ds \cdot \int_0^y e^{-t} dt$

$= (1-e^{-x})(1-e^{-y})$

(2) $P\{(x,y) \in G\} = \iint\limits_G f(x,y)\, dx\, dy = \iint\limits_G e^{-(x+y)}\, dx\, dy$

$= \int_0^1 dx \int_0^{1-x} e^{-(x+y)} dy = 1 - \frac{2}{e}$

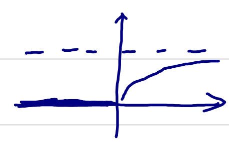
(3) $F_X(x) = \lim\limits_{y \to \infty} F(x,y) = 1 - e^{-x}$ $x>0$ $F_X(x) = 0$ $x \le 0$



$x=0$ $z=e^{-y}$

$y=0$ $z=e^{-x}$

$F_X(x) = \begin{cases} 1-e^{-x} & x>0 \\ 0 & \text{else} \end{cases}$

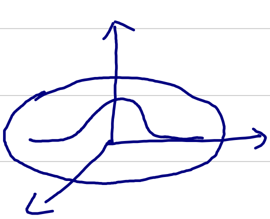


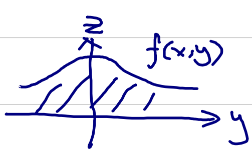
$F_Y(y) = \begin{cases} 1-e^{-y} & y>0 \\ 0 & \text{else} \end{cases}$

边缘密度函数

$F_X(x) = F(x,+\infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(s,t)\, dt\, ds$

$f_X(x) = \int_{-\infty}^{+\infty} f(x,t)\, dt$ $f_Y(y) = \int_{-\infty}^{+\infty} f(s,y)\, ds = \int_{-\infty}^{+\infty} f(x,y)\, dx$





(3) $f(x,y)=\dfrac{1}{\pi^2(1+x^2)(1+y^2)}$ 求 $f_X(x)$ $f_Y(y)$

$$f_X(x)=\int_{-\infty}^{+\infty}\frac{1}{\pi^2(1+x^2)(1+y^2)}\,dy=\frac{1}{\pi^2(1+x^2)}\tan^{-1}y\Big|_{-\infty}^{+\infty}=\frac{1}{\pi(1+x^2)}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}\frac{1}{\pi^2(1+x^2)(1+y^2)}\,dx=\frac{1}{\pi(1+y^2)}$$

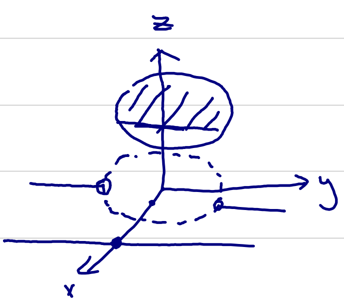
$f(x,y)=f_X(x)\cdot f_Y(y)$ 独立

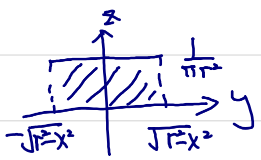
(4) $f(x,y)=\begin{cases}\dfrac{1}{\pi r^2} & x^2+y^2\le r^2\\ 0 & \text{else}\end{cases}$ $f_X(x)$ $f_Y(y)$

$|x|\le r$ 时 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,dy=\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{1}{\pi r^2}\,dy=\frac{2\sqrt{r^2-x^2}}{\pi r^2}$

$|x|>r$ 时 $f_X(x)=0$

$f_X(x)=\begin{cases}\dfrac{2\sqrt{r^2-x^2}}{\pi r^2} & |x|\le r\\ 0 & \text{else}\end{cases}$ $f_Y(y)=\begin{cases}\dfrac{2\sqrt{r^2-y^2}}{\pi r^2} & |y|\le r\\ 0 & \text{else}\end{cases}$





Area $=\dfrac{2\sqrt{r^2-x^2}}{\pi r^2}$

$f(x,y)=\begin{cases}2(x+y-2xy) & 0\le x,y\le 1\\ 0 & \text{else}\end{cases}$

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,dy$$
$$=\int_0^1 2(x+y-2xy)\,dy$$
$$=2\left(xy+\frac{y^2}{2}-xy^2\right)\Big|_0^1=1$$

else $f_X(x)=0$

① 二维正态分度的边缘分布也是正态分度

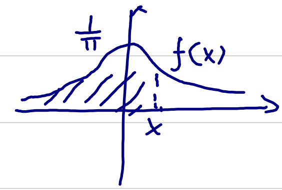
② 两边缘分布是正态，二维并非是正态

3.2.1 条件分布

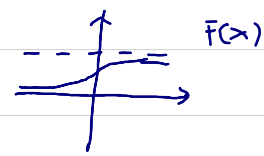
$F(x) = P\{X \le x\}$

$F(x|A) = P\{X \le x \,|\, A\}$

(1) $f(x) = \frac{1}{\pi(1+x^2)}$ 在 $X>1$ 的条件下的条件分布

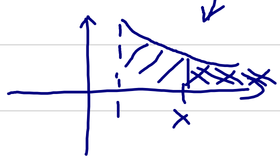


$$F(x|x>1) = P\{X \le x \,|\, X>1\} = \frac{P\{X \le x, X>1\}}{P\{X>1\}}$$



$x \le 1$ 时 $F(x|x>1) = 0$

$x>1$ 时 $P\{1<X\le x\} = \int_1^x \frac{1}{\pi(1+t^2)}\,dt = \frac{1}{\pi}\arctan x - \frac{1}{4}$



$$F(x|x>1) = \begin{cases} 0 & x \le 1 \\ \frac{4}{\pi}\tan^{-1}x - 1 & x>1 \end{cases}$$

$P\{x>1\} = \int_1^{+\infty} \frac{1}{\pi(1+t^2)}\,dt = \frac{1}{4}$

3.2.2 离散型的条件分布

$x_1 \backslash x_2$	0	1	
0	0.1	0.3	0.4
1	0.3	0.3	0.6
	0.4	0.6	

$\begin{cases} x_1=0 & P\{x_2=0 \,|\, x_1=0\} = \frac{0.1}{0.4} = 0.25 \\ x_1=0 & P\{x_2=1 \,|\, x_1=0\} = \frac{0.3}{0.4} = 0.75 \end{cases}$

$\begin{cases} x_1=1 & P\{x_2=0 \,|\, x_1=1\} = \frac{0.3}{0.6} = 0.5 \\ x_1=1 & P\{x_2=1 \,|\, x_1=0\} = \frac{0.3}{0.6} = 0.5 \end{cases}$

x_2	0	1
$P\{x_2 \,|\, x_1=0\}$	0.25	0.75

x_2	0	1
$P\{x_2 \,|\, x_1=1\}$	0.5	0.5

3.2.3 连续型的条件分布

定义3.5 (X,Y) $f(x,y)$ $f_X(x)$ $f_Y(y)$，若 $f_Y(y)>0$

在 $Y=y$ 的条件下 $F(x|y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du$

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}\qquad F(y|x)=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv$$

$$f(y|x)=\frac{f(x,y)}{f_X(x)}$$

(1) $f(x,y)=\frac{1}{\pi^2(1+x^2)(1+y^2)}$ $\quad f_X(x)=\frac{1}{\pi(1+x^2)}$ $\quad f_Y(y)=\frac{1}{\pi(1+y^2)}$

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{1}{\pi(1+x^2)}\qquad f(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\pi(1+y^2)}$$

(2) $f(x,y)=\begin{cases}\frac{1}{\pi r^2} & x^2+y^2\le r\\ 0 & \text{else}\end{cases}$ $\qquad f_X(x)=\begin{cases}\frac{2\sqrt{r^2-x^2}}{\pi r^2} & |x|\le r\\ 0 & \text{else}\end{cases}$

$$f_Y(y)=\begin{cases}\frac{2\sqrt{r^2-y^2}}{\pi r^2} & |y|\le r\\ 0 & \text{else}\end{cases}$$

$|y|<r$ $\quad f(x|y)=\frac{f(x,y)}{f_Y(y)}$

$$=\begin{cases}\frac{1}{2\sqrt{r^2-y^2}} & -\sqrt{r^2-y^2}\le x\le\sqrt{r^2-y^2}\\ 0 & \text{else}\end{cases}$$

$P\{X>0|Y=0\}=\frac{1}{2}$

圆里的一半

$=\int_0^r\frac{1}{2r}dx$

$=\frac{1}{2}$

$\frac{2\sqrt{r^2}}{\pi r^2}=\frac{2}{\pi r}$ $\qquad \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r}}$

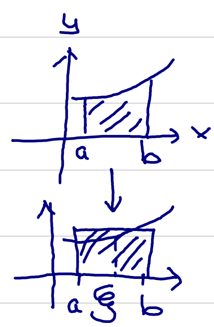
$=\frac{1}{2r}$

$\int_y^{y+\varepsilon}f_Y(v)dv=f_Y(\xi)\varepsilon$

$\int_a^b f(x)dx=f(\xi)(b-a)$

$\frac{1}{\varepsilon}\int_y^{y+\varepsilon}f_Y(v)dv=f_Y(\xi)=f_Y(y)$

$\xi\in(y,y+\varepsilon)$



$x=5$

$5\le X\le 5+\varepsilon$

$$P\{X\le x|Y=y\}=\frac{P\{X\le x\ \ Y=y\}}{P\{Y=y\}}$$

$$=\lim_{\varepsilon\to0}\frac{P\{X\le x\ \ y\le Y\le y+\varepsilon\}}{P\{y\le Y\le y+\varepsilon\}}=\lim_{\varepsilon\to0}\frac{\frac{1}{\varepsilon}\int_{-\infty}^{x}\int_y^{y+\varepsilon}f(u,v)dv\,du}{\frac{1}{\varepsilon}\int_y^{y+\varepsilon}f_Y(v)dv}=\frac{\int_{-\infty}^{x}f(u,v)du}{f_Y(y)}$$

$$=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du$$

3.2.4 随机变量的独立性

扔硬币 第一次第二次 身高 体重

$f(x|y) = f_X(x) = \frac{f(x,y)}{f_Y(y)}$

$f(x,y) = f_X(x) f_Y(y)$ ☆

$F(x,y) = F_X(x) F_Y(y)$

$P\{X\in S_X \ Y\in S_Y\} = P\{X\in S_X\} P\{Y\in S_Y\}$

1) 二维离散的独立性 $P\{X=x_i, Y=y_j\} = P\{X=x_i\} P\{Y=y_j\}$

X \ Y	0	1	
0	0.2	0.2	0.4
1	0.2	0.4	0.6
	0.4	0.6	

$0.4\times 0.4 \neq 0.2$

判断独立

X \ Y	0	1	
0	0.2	0.3	0.5
1	0.2	0.3	0.5
	0.4	0.6	

2) 二维连续性 $f(x,y) = f_X(x) f_Y(y)$

经理 8~12时 秘书7~9时 独立 不超过5分 ($\frac{1}{12}$时)

X. $f_X(x) = \begin{cases} \frac{1}{4} & 8<x<12 \\ 0 & \text{else} \end{cases}$ Y $f_Y(y) = \begin{cases} \frac{1}{2} & 7<y<9 \\ 0 & \text{else} \end{cases}$ X Y 独立

$f(x,y) = f_X(x) f_Y(y) = \begin{cases} \frac{1}{8} & 8<x<12 \quad 7<y<9 \\ 0 \end{cases}$

$P\{|X-Y|\leq \frac{1}{12}\}$

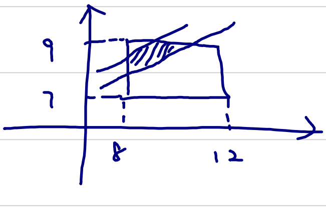
$= \iint_G f(x,y)\,dx\,dy$

$= \frac{1}{8}\iint_G 1\,dx\,dy = \frac{1}{8}\text{Area}G = \frac{1}{48}$

$-\frac{1}{12} \leq y-x \leq \frac{1}{12}$

$-\frac{1}{12}+x \leq y \leq \frac{1}{12}+x$

$y = x-\frac{1}{12}$ $y = x+\frac{1}{12}$



变量独立　构造的函数也独立

定理：X,Y 独立　$g_1(X)$ $g_2(Y)$ 是独立

X,Y 独立　X^2,Y^2 也独立　a_1X+b_1 a_2Y+b_2 独立

3.3 二维随机变量函数的分布

(1) 二维离散　　$Z=XY$

X \ Y	4	4.2
5	0.2	0.4
5.1	0.3	0.1

Z	20	21	20.4	21.42
P	0.2	0.4	0.3	0.1

$Z=X^2-Y$

Z	5^2-4	$5^2-4.2$	5.1^2-4	$5.1^2-4.2$
P	0.2	0.4	0.3	0.1

重复就合并

(2) X_1,X_2 独立　0-1分布　P　X_1+X_2

X_1	0	1
P	1-p	p

X_2	0	1
P	1-p	p

X_1+X_2	0	1	1	2
P	$(1-p)^2$	$(1-p)p$	$p(1-p)$	p^2

X_1+X_2	0	1	2
P	$(1-p)^2$	$2p(1-p)$	p^2

$X_1+X_2\sim B(2,p)$

(3) X,Y 独立　λ_1,λ_2 泊松分布　$Z=X+Y$　$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$

$$\{Z=k\}=\sum_{i=0}^{k}\{X=i,Y=k-i\}$$

$$P\{Z=k\}=\sum_{i=0}^{k}P\{X=i,Y=k-i\}=\sum_{i=0}^{k}P\{X=i\}P\{Y=k-i\}$$

$$=\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}e^{-\lambda_1}\frac{\lambda_2^{k-i}}{(k-i)!}e^{-\lambda_2}=\frac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)}$$

$\lambda_1+\lambda_2$ 泊松分布　(可加性)

3.3.2 二维连续变量函数的分布

(X,Y) $\quad f(x,y) \quad Z=g(X,Y)$

(1) $F_Z(z)=P\{Z\le z\}=P\{g(X,Y)\le z\}=\iint\limits_{D_z} f(x,y)\,dxdy \qquad D_z=\{(x,y)\mid g(x,y)\le z\}$

(2) $f_Z(z)$

(1) $f(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}} \qquad Z=\sqrt{X^2+Y^2}$

$z<0$ 时 $\quad F_Z(z)=P\{Z\le z\}=P\{\sqrt{X^2+Y^2}\le z\}=0$

$z\ge 0$ 时 $\quad P\{X^2+Y^2\le z^2\}$ 圆

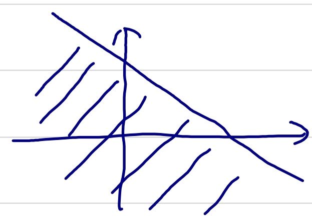
$=\iint\limits_G \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\,dxdy$

$=\int_0^{2\pi}\int_0^z \frac{1}{2\pi}e^{-\frac{r^2}{2}}\,r\,dr\,d\theta \quad =1-e^{-z^2}$

$F_Z(z)=\begin{cases}1-e^{-z^2} & z\ge 0\\ 0 & z<0\end{cases} \qquad f_Z(z)=\begin{cases}2ze^{-z^2} & z\ge 0\\ 0 & z<0\end{cases}$

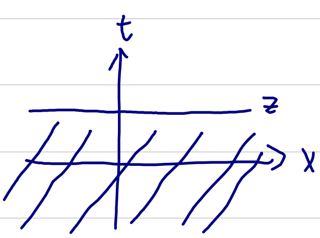
1. $Z=X+Y \quad F_Z(z)=P\{X+Y\le z\}$

$=\iint\limits_{x+y\le z} f(x,y)\,dxdy$



$=\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)\,dy\,dx \qquad t=x+y \quad y\to-\infty \quad t\to-\infty$

$y=t-x \qquad y=z-x \quad t=x+z-x=z$



$=\int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x,t-x)\,dt\,dx \qquad dy=d(t-x)=dt$

$=\int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x,t-x)\,dx\,dt$ 　为了做变上限积分

$f_Z(z)=\int_{-\infty}^{\infty} f(x,z-x)\,dx=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$ 如果独立

卷积公式

① $Z=X+Y$

$f_Z(z)=\int_{-\infty}^{\infty} f(z-y,y)\,dy=\int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\,dy$

② $X\ Y$ 独立

(2) $X\sim N(0,1)$ $Y\sim N(0,1)$ X,Y独立 $Z=X+Y$

$$\phi_Z(z)=\int_{-\infty}^{+\infty}\varphi_X(x)\varphi_Y(z-x)\,dx=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-x)^2}{2}}dx$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-\frac{z^2}{4}}e^{-(x-\frac{z}{2})^2}dx$$

$$=\frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{+\infty}e^{-(x-\frac{z}{2})^2}d(x-\frac{z}{2})=\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{z^2}{2(\sqrt{2})^2}}$$

$$e^{-\frac{x^2}{2}-\frac{(z-x)^2}{2}}$$
$$=e^{\frac{-x^2-z^2+2zx-x^2}{2}}$$
$$=e^{-\frac{2x^2-2zx+z^2}{2}}$$
$$=e^{-\frac{2(x^2-\frac{2z}{2}x+\frac{z^2}{4})+\frac{z^2}{2}}{2}}$$
$$=e^{-\frac{z^2}{4}}e^{-(x-\frac{z}{2})^2}$$

$Z\sim N(0,2)$ $X\sim N(\mu_1,\sigma_1^2)$ $Y\sim N(\mu_2,\sigma_2^2)$ $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

2. $M=\max\{X,Y\}$ $N=\min\{X,Y\}$ X,Y独立

$\{\max\{X,Y\}\le z\}=\{X\le z,Y\le z\}$

$F_M(z)=P\{M\le z\}=P\{X\le z\ Y\le z\}=P\{X\le z\}P\{Y\le z\}=F_X(z)F_Y(z)$

$F_N(z)=P\{N\le z\}=1-P\{N>z\}=1-P\{X>z\ Y>z\}=1-P\{X>z\}P\{Y>z\}$

$=1-(1-P\{X\le z\})(1-P\{Y\le z\})=1-(1-F_X(z))(1-F_Y(z))$

(3) X,Y独立 X: $[0,1]$均匀 Y $\lambda=3$ 指数分布 $M=\max\{X,Y\}$ $N=\min\{X,Y\}$

$$f_X(x)=\begin{cases}1 & 0\le x\le 1\\ 0 & \text{else}\end{cases}\qquad f_Y(y)=\begin{cases}3e^{-3y} & y>0\\ 0 & y\le 0\end{cases}$$

$$F_X(x)=\begin{cases}0 & x<0\\ x & 0\le x<1\\ 1 & 1\le x\end{cases}\qquad F_Y(y)=\begin{cases}1-e^{-3y} & y>0\\ 0 & y\le 0\end{cases}$$

$M=\max\{X,Y\}$ $$F_M(z)=\begin{cases}0 & z<0\\ z(1-e^{-3z}) & 0\le z<1\\ 1-e^{-3z} & 1\le z\end{cases}\qquad f_M(z)=\begin{cases}0 & z<0\\ 1-e^{-3z}+3ze^{-3z} & 0\le z<1\\ 3e^{-3z} & z\ge 1\end{cases}$$

$N=\min\{X,Y\}$ $$F_N(z)=\begin{cases}0 & z<0\\ 1-(1-z)e^{-3z} & 0\le z<1\\ 1 & 1\le z\end{cases}\qquad f_N(z)=\begin{cases}4e^{-3z}-3ze^{-3z} & 0\le z<1\\ 0 & \text{else}\end{cases}$$

4.1 数学期望

平均数 $\frac{180+190+200}{3}$

加权平均数 专业课考90 考100 3分 15

90% 9% 1%

$0.9\times 90+0.09\times 100+0.01\times 15$

4.1.1 离散型的期望

$P(X=x_k)=P_k$ 若 $\sum_{k=1}^{\infty}x_kP_k$ 绝对收敛 Expected $=\sum_{k=1}^{\infty}x_kP_k$

(1) 甲乙 x_1, x_2 次品

x_1	0	1	2	3
P	0.3	0.3	0.2	0.2

x_2	0	1	2	3
P	0.2	0.5	0.3	0

$Ex_1=0\times 0.3+1\times 0.3+2\times 0.2+3\times 0.2$ $Ex_2=1.1$

$=1.3$ 低

(2) 一 二 三 次 60% 20% 10% 10% \$6 \$4.8 \$4 \$0

$E(R)=6\times 0.6+4.8\times 0.2+4\times 0.1+0\times 0.1=4.96$

4.1.2 连续型的期望

x $f(x)$ $\int_{-\infty}^{+\infty}xf(x)dx$ 绝对收敛

$Ex=\int_{-\infty}^{+\infty}xf(x)dx$

(1) $f(x)=\begin{cases}2x & 0<x<1\\ 0 & \text{else}\end{cases}$ $Ex=\int_{-\infty}^{+\infty}xf(x)dx$

$=\int_0^1 2x^2dx=\frac{2}{3}$

(2) 先用后付 寿命 $\lambda=\frac{1}{10}$ 指数分布 $P\{x\le 1\}=\int_0^1\frac{1}{10}e^{-\frac{x}{10}}dx=0.0952$

$X\le 1$年 1500元 $1<X\le 2$ 2000元 $P\{1<x\le 2\}=0.0861$

$2<X\le 3$ 2500元 $X>3$ 3000元 $P\{2<x\le 3\}=0.0779$

$E(R)=1500\times 0.0952+2000\times 0.0861+\cdots=2732.15$元 $P\{X>3\}=0.7408$

4.1.3 随机变量函数的期望 $Y=g(x)$

离散 $\sum x_i p_i$ $\sum g(x_i) p_i$

连续 $\int_{-\infty}^{+\infty} x f(x) dx$ $\int_{-\infty}^{+\infty} g(x) f(x) dx$

(1)

X	0	1	2
P	0.1	0.6	0.3

$Y=4X+1$

Y	1	5	9
P	0.1	0.6	0.3

$EY = 1\times 0.1 + 5\times 0.6 + 9\times 0.3 = 5.8$

$EY = (4\times 0+1)\times 0.1 + (4\times 1+1)\times 0.6 + (4\times 2+1)\times 0.3 = 5.8$

$E(X-EX)^2 = E(X-1.2)^2 = (0-1.2)^2\times 0.1 + (1-1.2)^2\times 0.6 + (2-1.2)^2\times 0.3 = 0.36$

(2) $f(x) = \begin{cases} \frac{1}{2} & 0\le x\le 2 \\ 0 & \text{else} \end{cases}$ $Y=4x+1$

$EY = E(4x+1) = \int_{-\infty}^{+\infty} (4x+1) f(x) dx = \int_0^{2} (4x+1)\frac{1}{2} dx = 5$

(3) X: 需求量 [2000, 4000] 均匀分布

每卖1吨, 3万 卖不出, 1吨损失1万

y 出口量 Y收益 $Y=g(x) = \begin{cases} 3y & x\ge y \\ 3x-(y-x) & x<y \end{cases}$

$f(x) = \begin{cases} \frac{1}{2000} & 2000\le x\le 4000 \\ 0 & \text{else} \end{cases}$

$EY = \int_{-\infty}^{+\infty} g(x) f(x) dx$

$= \int_{2000}^{4000} \frac{g(x)}{2000} dx$

$= \frac{1}{2000}\int_{2000}^{y} (4x-y) dx + \int_y^{4000} 3y\, dx$

$= \frac{1}{1000}(-y^2+7000y-4\times 10^6)$

$y=3500$, 收入最大

二维变量函数 $Z=g(X,Y)$ 的期望

1) 离散 $EZ=\sum_i\sum_j g(x_i,y_j)p_{ij}$

X \ Y	0	1	2
1	0.1	0.1	0.2
2	0.2	0.2	0.2

$Z=X^2Y$ $\quad EZ=(1^2\cdot 0)\times 0.1+(1^2\cdot 1)\times 0.1+(1^2\cdot 2)\times 0.2$

$+(2^2\cdot 0)\times 0.2+(2^2\cdot 1)\times 0.2+(2^2\cdot 2)\times 0.2$

2) 连续型 $EZ=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\,dx\,dy$

例4： $f(x,y)=\begin{cases}x+y & 0\le x\le 1 \quad 0\le y\le 1\\ 0 & \text{else}\end{cases}$

$E(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xy\,f(x,y)\,dx\,dy$

$=\int_0^1\int_0^1 xy(x+y)\,dx\,dy=\frac{1}{3}$

5)： X 进货量 Y：需求量 独立 [10, 20] 均匀

卖一件 1000元 其他商店1件：500元 平均利润

$Z=g(X,Y)=\begin{cases}1000Y & Y\le X\\ 1000X+500(Y-X) & Y>X\end{cases}$

$f_X(x)=\begin{cases}\frac{1}{10} & 10\le x\le 20\\ 0 & \end{cases}$ $\quad f_Y(y)=\begin{cases}\frac{1}{10} & 10\le y\le 20\\ 0 & \text{else}\end{cases}$

$f(x,y)=f_X(x)f_Y(y)=\begin{cases}\frac{1}{100} & 10\le x\le 20 \quad 10\le y\le 20\\ 0 & \text{else}\end{cases}$

$EZ=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\,dx\,dy$

$=\frac{1}{100}\int_{10}^{20}\int_{10}^{20}g(x,y)\,dx\,dy$

$=\frac{1}{100}\int_{10}^{20}dy\left[\int_{10}^{y}500(x+y)dx+\int_{y}^{20}1000\,y\,dx\right]$

$=14166.67$

4.1.4 数学期望的性质

(1) $EC=C$ $\quad C\times 1=C$

(2) $E(X+C)=EX+C$

$$E(X+C)=\int_{-\infty}^{+\infty}(x+c)f(x)\,dx$$
$$=\int_{-\infty}^{\infty}xf(x)dx+c\underbrace{\int_{-\infty}^{+\infty}f(x)dx}_{1}$$

(3) $E(CX)=CE(X)$

(4) $E(kX+b)=kEX+b$

(5) $E(X\pm Y)=EX\pm EY$

$E(\sum C_i x_i)=\sum C_i EX_i$

$E(\frac{1}{n}\sum x_i)=\frac{1}{n}\sum EX_i$

$E(2X_1-3X_2+4X_3)=2EX_1-3EX_2+4EX_3$

$$E(X\pm Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x\pm y)f(x,y)\,dx\,dy$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x f(x,y)\,dxdy\pm\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y f(x,y)\,dxdy$$
$$=\int_{-\infty}^{\infty}x\,dx\int_{-\infty}^{\infty}f(x,y)\,dy\pm\int_{-\infty}^{\infty}y\,dy\int_{-\infty}^{\infty}f(x,y)\,dx$$
$$=\int_{-\infty}^{\infty}x f_X(x)\,dx\pm\int_{-\infty}^{\infty}y f_Y(y)\,dy$$
$$=E(X)\pm E(Y)$$

(6) X,Y 独立 $\quad E(XY)=E(X)E(Y)$

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy\,f_X(x)f_Y(y)\,dxdy$$
$$=\int_{-\infty}^{\infty}x f_X(x)dx\cdot\int_{-\infty}^{\infty}y f_Y(y)dy \quad (2个常数)$$
$$=EXEY$$

例 1 $\quad X,Y$ 独立

X	9	10	11
P	0.3	0.5	0.2

Y	6	7
P	0.4	0.6

$E(X+Y)=EX+EY$

$=9\times0.3+10\times0.5+11\times0.2+6\times0.4+7\times0.6=16.5$

$X+Y$	15	16	16	17	17	18
P	0.12	0.18	0.2	0.3	0.08	0.12

$E(X+Y)=15\times0.12+16\times0.38+17\times0.38+18\times0.12$

$E(XY)=EXEY=9.9\times6.6$

$EY^2=36\times0.4+49\times0.6=43.8$

XY	54	63	60	70	66	77
P	0.12	0.18	0.2	0.3	0.08	0.12

(2)

X \ Y	1	2	3	
0	0.1	0.2	0.3	0.6
1	0.2	0.1	0.1	0.4
	0.3	0.3	0.4	

$E(X-Y) = EX - EY$

$= 0.4 - 0.3 - 2\times 0.3 - 3\times 0.4$

$= -1.7$

X	0	1
P	0.6	0.4

Y	1	2	3
P	0.3	0.3	0.4

$E(XY) \neq EXEY$ （不独立）

$E(XY) = 1\times 0.2 + 2\times 0.1 + 3\times 0.1 = 0.7$

(3) 10个 X点数和 求EX

X_i：第i个点数

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$EX_i = 1\times\frac{1}{6} + 2\times\frac{1}{6} + \cdots + 6\times\frac{1}{6}$

$= \frac{1}{6}(1+2+\cdots+6) = \frac{7}{2}$

$EX = E(X_1+\cdots+X_{10}) = \sum_{i=1}^{10} EX_i = 10\times\frac{7}{2} = 35$

(4) N件产品 M件次品 任取 n件 $n \leq M \leq N$ 求E次品

? 次品 $1 \sim M$ $X_i = \begin{cases} 1 & i\text{件次取} \\ 0 & i\text{件次未取} \end{cases}$ $i = 1, 2, \cdots, M$ X_i 次品数 $X = X_1 + X_2 + \cdots + X_M$

$P\{X_i = 1\} = \frac{n}{N}$ $\qquad \frac{{}^{N-1}C_{n-1}}{{}^{N}C_n}$ $\qquad P\{X_i = 0\} = 1 - \frac{n}{N}$ $\qquad EX_i = \frac{n}{N}$ $\qquad EX = \frac{n}{N}M$

4.1.5 条件期望：一个变量取某值，另一个变量的期望

(1) 离散 $E(X|Y=y_j) = \sum x_i P(X=x_i | Y=y_j)$

$E(Y|X=x_i) = \sum y_j P(Y=y_j | X=x_i)$

X \ Y	1	2	3	
0	0.1	0.2	0.3	
1	0.2	0.1	0.1	$\to$ 0.4

Y	1	2	3
$P(Y\|X=1)$	0.5	0.25	0.25

$E(Y|X=1) = 0.5 + 2\times 0.25 + 3\times 0.25 = 1.75$

(2) 连续 $E(X|Y=y) = \int_{-\infty}^{\infty} x f(x|y)\,dx$

$E(Y|X=x) = \int_{-\infty}^{\infty} y f(y|x)\,dy$

4.2.1 方差

偏离程度 $|x-Ex|$ $(x-Ex)^2$ 量纲

$Var(x) = E(x-Ex)^2$ $s.d. = \sqrt{Var(x)}$

(1) 离散: $Var(x) = \sum (x_k - Ex)^2 p_k$

(2) 连续: $Var(x) = \int_{-\infty}^{+\infty} (x-Ex)^2 f(x)dx$ $E(x^2 - 2xEx + (Ex)^2)$ Ex是常数

$= Ex^2 - 2(Ex)^2 + (Ex)^2$

$Var(x) = Ex^2 - (Ex)^2$ $= Ex^2 - (Ex)^2$

(1)

x	-2	0	2
p	0.4	0.3	0.3

$Ex = -2\times0.4 + 2\times0.3 = -0.2$

$Var(x) = 4\times0.4 + 4\times0.3 - 0.4 = 2.76$

(2) $f(x) = \begin{cases} 2x & 0<x<1 \\ 0 & else \end{cases}$ $Ex = \int_0^1 2x^2 = \frac{2}{3}$ $Var(x) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$

$Ex^2 = \int_0^1 x^2(2x)dx = \frac{1}{2}$

(3) $f(x) = \begin{cases} ax^2+bx+c & 0<x<1 \\ 0 & else \end{cases}$ $Ex = 0.5$ $DX = 0.15$

$\int_0^1 (ax^2+bx+c)dx = 1$

$Ex = \int_0^1 x(ax^2+bx+c)dx = \frac{1}{2}$

$Ex^2 = \int_0^1 x^2(ax^2+bx+c)dx = \frac{a}{5} + \frac{b}{4} + \frac{c}{3}$

$\begin{cases} \frac{a}{4} + \frac{b}{3} + \frac{c}{2} = 0.5 \\ \frac{a}{5} + \frac{b}{4} + \frac{c}{3} = 0.4 \\ \frac{a}{3} + \frac{b}{2} + c = 1 \end{cases}$

$\begin{bmatrix} 3 & 4 & 6 & 6 \\ 12 & 15 & 20 & 24 \\ 2 & 3 & 6 & 6 \end{bmatrix}$

4.2.2 方差的性质

(1) $Var(c)=0$ $\quad Var(c)=EC^2-(EC)^2=C^2-C^2=0$

(2) $Var(X+C)=Var(X)$ $\quad E(X+C-E(X+C))^2=E(X+C-EX-\underline{EC})^2=E(X-EX)^2$ 一起去台阶

(3) $Var(CX)=C^2 Var(X)$ $\quad E(CX-EC(X))^2=EC^2(X-EX)^2=C^2E(X-EX)^2$

(4) $Var(kX+b)=k^2 Var(X)$

(5) $X\ Y$ 独立 $\quad Var(X\pm Y)=VarX+VarY$

$$E(X\pm Y-E(X\pm Y))^2=E(X\pm Y-EX\mp EY)^2$$
$$=E(X-EX\pm(Y-EY))^2$$
$$=E((X-EX)^2\pm 2(X-EX)(Y-EY)+(Y-EY)^2)$$
$$=E(X-EX)^2+E(Y-EY)^2\pm 2E((X-EX)(Y-EY))$$
$$=VarX+VarY$$

(6) $Var(X)=0 \Leftrightarrow p(X=EX)=1$

标准化 $\quad X^*=\dfrac{X-EX}{s.d.}$ $\quad EX^*=0$ $\quad Var(X^*)=1$

$$E(X^*)=E\left(\frac{X-EX}{s.d.}\right)=\frac{1}{s.d.}(EX-EX)=0$$

$$Var(X^*)=\frac{1}{Var(X)}Var(X-EX)=\frac{Var(X)}{Var(X)}=1$$

4.3.1 常见离散型的期望与方差

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

$EX=p \quad EX^2=p$

$Var(X)=p-p^2=p(1-p)$

X	0	1
P	$1-p$	p

二项分布 $\quad P(X=k)={}^nC_k p^k q^{n-k}$ $\quad k=0,1,\dots,n$

$$EX=\sum_{k=0}^{n} k\,{}^nC_k p^k q^{n-k}=\sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k}=\sum_{k=1}^{n}\frac{n!}{(k-1)!(n-k)!}p^k q^{n-k}$$

$$=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}=np\sum_{k-1=0}^{n-1}{}^{n-1}C_{k-1}p^{k-1}q^{n-k}=np \qquad (p+q)^{n-1}=1$$

$EX^2=npq+n^2p^2$ $\qquad Var(X)=npq$

几何分布 $P(x=k)=(1-p)^{k-1}p \qquad k=1,2,3,\cdots$

积分再求导

$$\sum_{k=1}^{\infty} kx^{k-1} = \left(\sum_{k=1}^{\infty} x^k\right)' = \left(\frac{x}{1-x}\right)' = \frac{1}{1-x^2} \qquad |x|<1$$

$$EX = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p^2}p = \frac{1}{p}$$

$$EX^2 = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = \frac{2-p}{p^2}$$

$$\mathrm{Var}(x) = \frac{1-p}{p^2}$$

$$\sum_{k=1}^{\infty} k^2x^{k-1} = \sum_{k=1}^{\infty} k\cdot kx^{k-1} = \left(\sum_{k=1}^{\infty} k\cdot x^k\right)'$$

$$= \left(x\sum_{k=1}^{\infty} kx^{k-1}\right)' = \left(\frac{x}{(1-x)^2}\right)' = \frac{1+x}{(1-x)^3}$$

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

$$EX = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda} = \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}e^{-\lambda} \quad (m=k-1) \quad = \underline{\lambda}$$

$$EX^2 = \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty}\frac{k\lambda^k}{(k-1)!}e^{-\lambda} = \sum_{k=1}^{\infty}\frac{(k-1)\lambda^k}{(k-1)!}e^{-\lambda} + \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}+\lambda$$

$$=\lambda^2+\lambda$$

$$\mathrm{Var}(x) = EX^2-(EX)^2 = \lambda^2+\lambda-\lambda^2 = \underline{\lambda}$$

4.3.2 常见连续型的期望与方差

均匀分布 $f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & \text{else}\end{cases}$

$$Ex = \int_a^b x\frac{1}{b-a}\,dx = \frac{a+b}{2} \quad \text{中点}$$

$$Ex^2 = \int_a^b x^2\frac{1}{b-a}\,dx = \frac{b^2+ab+a^2}{3}$$

$$\mathrm{Var}(x) = Ex^2-(Ex)^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

指数分布 $f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & \text{else}\end{cases}$

$$EX=-\int_0^{\infty}xe^{-\lambda x}\,d(-\lambda x)=-\int_0^{\infty}x\,de^{-\lambda x}=-xe^{-\lambda x}+\int_0^{+\infty}e^{-\lambda x}dx=\frac{1}{\lambda}$$

$$EX^2=\int_0^{+\infty}x^2e^{-\lambda x}dx=-\int_0^{\infty}x^2de^{-\lambda x}=-x^2e^{-\lambda x}\Big|_0^{\infty}+2\int_0^{\infty}e^{-\lambda x}x\,dx=\frac{2}{\lambda^2}$$

$$Var(x)=EX^2-(Ex)^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}$$

正态分布 $X\sim N(\mu,\sigma^2)$ $Y=\frac{X-\mu}{\sigma}\sim N(0,1)$

EX $Var(X)=\sigma^2$ $EY=0$ $Var(Y)=1$

正态分布最正宗 $\mu\ \sigma^2$ 易知晓

泊松分布最轻松 $\lambda\ \lambda$ 一样的

均匀分布好理解 数学期望是中点

方差是长度$^2\div 12$，为啥是12我也不知道

0-1分布期望 p 二项分布期望是 np

... 方差是 pq ... 方差是 npq

4.4.1 协方差

$$Cov(X,Y)=E((X-EX)(Y-EY))$$
$$=E(XY-XEY-YEX+EXEY)$$
$$=E(XY)-EYEX-EXEY+EXEY$$
$$=E(XY)-EXEY$$

$$Var(X\pm Y)=Var(X)+Var(Y)\pm 2Cov(X,Y)$$

(1)

X \ Y	-1	0	1	
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
0	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	

X	-1	0	1
P	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$

Y	-1	0	1
P	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$

$EX=0 \qquad EY=0 \qquad E(XY)=\frac{1}{8}-\frac{1}{8}-\frac{1}{8}+\frac{1}{8}=0$

$Cov(X,Y)=0-0=0$

(2) $f(x,y)=\begin{cases} x+y & 0\le x\le 1 \quad 0\le y\le 1 \\ 0 & \text{else} \end{cases} \qquad Cov(X,Y)$

$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy \qquad 0\le x\le 1$

$=\int_0^1 (x+y)\,dy = \left[xy+\frac{y^2}{2}\right]_0^1 = x+\frac{1}{2}$

$f_X(x)=\begin{cases} x+\frac{1}{2} & 0\le x\le 1 \\ 0 & \text{else} \end{cases}$

$f_Y(y)=\begin{cases} y+\frac{1}{2} & 0\le y\le 1 \\ 0 & \text{else} \end{cases}$

$EX=\int_0^1 x(x+\frac{1}{2})\,dx=\frac{7}{12} \qquad EY=\frac{7}{12}$

$E(XY)=\int_0^1\int_0^1 xy(x+y)\,dx\,dy=\frac{1}{3}$

$Cov(X,Y)=E(XY)-EXEY=\frac{-1}{144}$

性质：1) $Cov(X,Y)=Cov(Y,X)$

2) $Cov(aX,bY)=ab\,Cov(X,Y)$

$E(aX,bY)-E(aX)E(bY)=abE(XY)-abE(X)E(Y)$

3) $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

4) $Cov(C,X)=0 \qquad E(CX)-EC\,EX=0$

$X^* = \frac{X-EX}{s.d.X}$ $\quad Y^* = \frac{Y-EY}{s.d.Y}$

$Cov(X^*, Y^*) = E(X^*Y^*) - EX^* EY^*$

$= E\left(\frac{X-EX}{s.d.X}, \frac{Y-EY}{s.d.Y}\right) - \cancel{E\left(\frac{X-EX}{s.d.X}\right)} \cancel{E\left(\frac{Y-EY}{s.d.Y}\right)}$ $\quad E(X-EX)=0$

$E(Y-EY)=0$

$= \frac{E((X-EX)(Y-EY))}{s.d.X \ s.d.Y} = \frac{Cov(X^*, Y^*)}{s.d.X \ s.d.Y} = \rho$

4.4.2 相关系数 $\rho = \frac{Cov(X,Y)}{\sqrt{VarX}\sqrt{VarY}}$

$= \frac{E(XY) - EXEY}{s.d.X \ s.d.Y}$ 与 $Cov(X,Y)$ 同正 同负 同 0

例2: $VarX = 4 \quad Var = 9 \quad \rho = 0.5$

$Var(X-Y) = VarX + VarY - 2cov(X,Y) = 7 \quad Cov(X,Y) = \rho\sqrt{VarX}\sqrt{VarY} = 3$

$|\rho| \le 1$ 引理: $[E(XY)]^2 \le EX^2 \cdot EY^2$

$g(t) = E(tX-Y)^2 = E(t^2X^2 - 2tXY + Y^2)$

$= t^2EX^2 - 2tE(XY) + EY^2 \geq 0$

$\Delta = 4[E(XY)]^2 - 4EX^2EY^2 \le 0$

$[E(XY)]^2 \le EX^2EY^2$

$\rho^2 \le 1$ 令 $X_1 = X - EX \quad Y_1 = Y - EY$

$\rho^2 = \frac{(E(X-EX)(Y-EY))^2}{VarX \ VarY} = \frac{(E(X_1 Y_1))^2}{EX_1^2 \ EY_1^2} \le 1$

$DX = EX_1^2 \quad DY = EY_1^2$

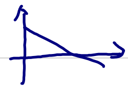
$= E(X-EX)$

$\rho^2 \le 1 \quad -1 \le \rho \le 1$

定理 4.4 $|\rho|=1 \Leftrightarrow$ X与Y 以 $p=1$ 成线性关系 $P(Y=aX+b)=1$

$\rho=-1,1$

① $\rho=1$ X,Y 完全正相关 $Y=2X-3$ 

② $\rho=-1$ X,Y 完全负相关 $Y=-0.5X+1$ 

③ $|\rho|\to 0$ X,Y 线性关系很弱

④ $\rho=0$ X,Y 不存在线性关系

X,Y 不相关 X,Y 独立（没有任何关系）

↓线性

① X,Y 独立，则 X,Y 不相关 $\rho=\frac{E(XY)-EXEY}{\sqrt{DX}\sqrt{DY}}=0 \Rightarrow$ X,Y 不相关

② X,Y 不相关，X,Y 不一定独立 $\rho=0$ $E(XY)=EXEY$

独立 $\Leftrightarrow f(x,y)=f_X(x)f_Y(y)$

$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f(x,y)\,dx\,dy$ 定积分求导是0

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

$xyf(x,y)=xf_X(x)\,yf_Y(y)$ ×

例3： $Y=X^2$ $Cov(X,Y)=0$ $\rho=0$ 不相关

对于二维正态分布 独立与不相关是等价的

4.5 中心矩与原点矩

原点矩：EX^k，期望 EX，一阶原点矩

中心矩：$E(X-EX)^k$ 一阶中心矩 $E(X-EX)=EX-EX=0$

二阶中心矩 $E(X-EX)^2$ 方差

① 离散 $\sum x_i^k P_i$

② 连续 $\int_{-\infty}^{+\infty} x^k f(x)\,dx$ 高于4阶 极少使用

① 离散 $\sum (x_i-EX)^k P_i$

② 连续 $\int_{-\infty}^{+\infty}(x-EX)^k f(x)\,dx$

5.1 大数定律

大量重复实验的平均结果的稳定性

$\frac{DX}{\varepsilon^2}$ → ε越大，外面的概率越小
→ 误差越小 波动越小，落在外面的概率越小

5.1.1 切比雪夫不等式

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

远离 外面 $P(|X-EX| < \varepsilon) \geq 1 - \frac{Var X}{\varepsilon^2}$

EX

Proof X: 连续

$$\int_{|x-EX| \geq \varepsilon} f(x)dx \leq \int_{|x-EX| \geq \varepsilon} \frac{(x-EX)^2}{\varepsilon^2} f(x)dx$$ （强行放缩）

$|x-Ex| \geq \varepsilon^2$

$(x-Ex)^2 \geq \varepsilon^2$

$$\leq \int_{-\infty}^{+\infty} \frac{(x-E(x))^2}{\varepsilon^2} f(x)dx = \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty} (x-Ex)^2 f(x)dx = \frac{Var(X)}{\varepsilon^2}$$

(1) 白细胞平均：7300 s.d.: 700 5200 ~ 7400 $P(|X-EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}$

$Var X = 490000$ $7300 - 5200 = 2100$ $\varepsilon = 2100$

$$P(|X-7300| < 2100) \geq 1 - \frac{490000}{2100^2}$$

(2) $EX = \mu$ $Var X = \sigma^2 > 0$ $P\{|X-\mu| \geq 3\sigma\}$

$= P\{|X-\mu| \geq 3\sigma\} \leq \frac{Var X}{(3\sigma)^2} = \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}$

分布无要求

$X \sim N(\mu, \sigma^2)$ $P\{|X-\mu| \geq 3\sigma\} = 1 - P\{|X-\mu| < 3\sigma\} = 1 - P\{-3\sigma < X-\mu < 3\sigma\}$

$= 1 - P\{-3 < \frac{X-\mu}{\sigma} < 3\}$

$= 1 - (\Phi_0(3) - \Phi_0(-3))$

$= 2 - 2\Phi_0(3) \approx 0.0027$

5.1.2 切比雪夫大数定律

收敛：$a_n \to a$ $\forall \varepsilon > 0$，$\exists N > 0$，$n > N$，$|a_n - a| < \varepsilon$

依概率收敛：$x_n \xrightarrow{P} a$ $\lim_{n \to \infty} P\{|x_n - a| < \varepsilon\} = 1$ 可能有一些点在收敛过程中跳出去

频率 → 概率 $\frac{24}{50}$，$\frac{508}{1000}$，$\frac{4994}{10000}$ $\to \frac{1}{2}$

定理（伯努利大数定律） n重 A发生了m_n次 $\frac{m_n}{n}$ 频率

$$\lim_{n \to \infty} P\{|\frac{m_n}{n} - p| < \varepsilon\} = 1$$

$$\lim_{n \to \infty} P\{|\frac{m_n}{n} - p| \geq \varepsilon\} = 0$$

频率以概率收敛于概率

Proof $m_n \sim B(n,p)$ $\quad Em_n = np$ $\quad Var\, m_n = np(1-p)$

$$E\left(\frac{m_n}{n}\right)=p \qquad Var\left(\frac{m_n}{n}\right)=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}$$

夹逼定理

切比雪夫不等式 $\forall \varepsilon>0$ $\quad 1 \geq P\left\{\left|\frac{m_n}{n}-p\right|<\varepsilon\right\} \geq 1-\frac{\frac{p(1-p)}{n}}{\varepsilon^2} = \underline{\underline{1-\frac{p(1-p)}{n\varepsilon^2}}} \to 1, \quad n\to\infty$

$$\boxed{\frac{m_n}{n} \xrightarrow{P} p \quad n\to\infty}$$

$x_1 \cdots x_n \cdots$ 独立 同分布

$$x_i=\begin{cases}1 & 发生\\ 0 & 未发生\end{cases} \qquad Ex_i=p \qquad Var\,x_i=p(1-p) \qquad m_n=\sum_{i=1}^{n}x_i\ (发生了几次)$$

$$\frac{m_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$p=E\left(\frac{1}{n}\sum x_i\right)=\frac{1}{n}\sum_{i=1}^{n}Ex_i$$

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}Ex_i\right|<\varepsilon\right\}=1$$

切比雪夫大数定理

$x_1 \cdots x_n \cdots$ 不相关的变量

Ex_i 和 $Var\,x_i$ 存在，方差有界 $Var\,x_i \leq M$ $\quad \forall \varepsilon>0$

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}Ex_i\right|<\varepsilon\right\}=1$$

平均数 $\xrightarrow{收敛于}$ 期望均值

Proof $E\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\sum_{i=1}^{n}Ex_i$ $\quad x_1\cdots x_n\cdots$ 不相关 $Cov(x_i,x_j)=0$

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}Dx_i \leq \frac{nM}{n^2}=\frac{M}{n}$$

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}Ex_i\right|<\varepsilon\right\} \geq 1-\frac{D\left(\frac{1}{n}\sum x_i\right)}{\varepsilon^2} \geq 1-\frac{M}{n\varepsilon^2} \to 1$$

推论：$x_1 \cdots x_n \cdots$ 独立同分布 $Ex_i=\mu$ $\quad Var\,x_i=\sigma^2$ $\quad \forall\varepsilon>0$

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}x_i-M\right|<\varepsilon\right\}=1$$

定理（辛钦）

$x_1 \cdots x_n \cdots$ 独立同分布 $Ex_i=\mu$，方差无要求

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|<\varepsilon\right\}=1$$

平均数 $\to$ 期望

5.2 中心极限定理

现象由大量相互独立的因素影响

大量独立同分布的变量和的极限分布是正态分布

定理：$X_1 \cdots X_n \cdots$ 独立同分布 $EX_i=\mu$ $Var(X_i)=\sigma^2$ $0<\sigma^2<\infty$

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\,\sigma} \le x\right) = \Phi_0(x)$$

$Y=\sum_{i=1}^n X_i$ $EY = E\sum_{i=1}^n X_i = n\mu$

$VarY = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) = n\sigma^2$

$\frac{\sum X_i - n\mu}{\sqrt{n}\sigma}$ $N(0,1)$ $\sum_{i=1}^n X_i$ $N(n\mu, n\sigma^2)$

1) 接待100人 $[0,60]$ 均匀分布，独立，月销售额超3500元

X_i 第i人 $EX_i=30$ $Var X_i = 60^2 \div 12 = 300$ $\sum_{i=1}^{100} X_i \ge 3500$

$\frac{\sum X_i - 3000}{100\sqrt{3}} \overset{近似}{\sim} N(0,1)$ $P(\sum_{i=1}^{100} X_i \ge 3500) = 1-P(\sum X_i \le 3500)$

$= 1-P\left(\frac{\sum X_i - 3000}{100\sqrt{3}} \le \frac{3500-3000}{100\sqrt{3}}\right)$

$= 1-\Phi_0(2.887) = 0.002$

2)

X	10	9	8
P	0.5	0.3	0.2

$EX = 10\times0.5+9\times0.3+8\times0.2$ $EX^2 = 100\times0.5+81\times0.3+64\times0.2$

$Var X = EX^2-(EX)^2$ 100次 915 ~ 945

X_i i次射击 $EX_i = 9.3$ $DX_i = 0.61$ $\frac{\sum X_i - 930}{\sqrt{61}} \overset{近似}{\sim} N(0,1)$

$\searrow Y$

$P(915 \le \sum X_i \le 945) = P\left(\frac{915-930}{\sqrt{61}} \le Y \le \frac{945-930}{\sqrt{61}}\right)$

$= P(-1.92 \le Y \le 1.92)$

$= \Phi_0(1.92) - \Phi_0(-1.92) = 2\Phi_0(1.92) - 1$

定理 5.6 Y_n n,p 二项分布 $\xrightarrow{近似}$ 正态分布

$$\lim_{n\to\infty} P\left(\frac{Y_n-np}{\sqrt{np(1-p)}}\le x\right)=\Phi_0(x) \qquad Y_n=\sum_{i=1}^{n}x_i \qquad x_i=\begin{cases}1 & 发生\\ 0 & 不发生\end{cases} \qquad Ex_i=p,\ Var\,x_i=p(1-p)$$

(3) 每人死亡 0.005，1万人，≤70

X 死亡人数 $P(X\le 70)=\sum_{k=0}^{70} C_{10000}^{k}\,0.005^{k}\,0.995^{10000-k}$

$$P(X\le 70)=P\left(\frac{X-np}{\sqrt{np(1-p)}}\le\frac{70-np}{\sqrt{np(1-p)}}\right)$$

$$=\Phi_0(2.84)=0.9977$$

$$P(X=k)=P\left(k-\tfrac{1}{2}<X<k+\tfrac{1}{2}\right)$$

$$=P\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}<\frac{X-np}{\sqrt{np(1-p)}}<\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

$$=\Phi_0\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)-\Phi_0\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

(4) 击中飞机 0.01 500发3发击中 $n=500$ $p=0.01$

$$P(X=5)=P(4.5<X<5.5)=\Phi_0\left(\frac{5.5-5}{\sqrt{4.95}}\right)-\Phi_0\left(\frac{4.5-5}{\sqrt{4.95}}\right)=0.178$$

二项分布
- n大 np适中 → 泊松分布
- n大 np大 → 正态分布

6.1 总体与样本

总体：全体 10万 | 样本： 抽样 $(X_1,\cdots,X_n)$ 确定的数 $(x_1,\cdots,x_n)$

个体 | 变量 观测值

有限总体，无限总体 | 简单随机抽样

X 总体分布 | ① 同分布 ② 独立

样本的分布 $(X_1,\cdots,X_n)$ $F(x_1,\cdots x_n)=F(x_1)\cdots F(x_n)$

$$p(X_1=x_1,\cdots,X_n=x_n)=p(X_1=x_1)\cdots p(X_n=x_n)$$

$$f(x_1,\cdots,x_n)=f(x_1)\cdots f(x_n)$$

(1) X: 0-1 分布 p $(x_1,\cdots,x_n)$ 是样本

$$p(X=x)=p^x(1-p)^{1-x} \quad x=0,1$$

$$p(X_1=x_1,\cdots,X_n=x_n)=p(X_1=x_1)\cdots p(X_n=x_n)=p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_n}$$

$$=p^{x_1+x_2+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$$

6.2.1 统计量定义：不含任何未知参数的样本的函数

$x_1+x_2+\cdots+x_n$ $x_1^2+\cdots+x_n^2$

$\max\{|x_1|,\cdots,|x_n|\}$ $X_1+0X_2+\cdots+0X_n$ $\min\{x_3,x_4\}$

$N(\mu,\sigma^2)$ μ,σ 未知 $(x_1-\mu)^2+\cdots+(x_n-\mu)^2$

$\frac{1}{n\sigma^2}(x_1+\cdots+x_n)$

6.2.2 常用统计量 设样本 $(X_1,X_2,\cdots,X_n)$ 来自 X，即有

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

未修正的样本方差： $S_0^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$ $S^2=\frac{n}{n-1}S_0^2$

样本方差： $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{X})^2$

样本标准差： $S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}$

样本k阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k$ 1阶原点矩是均值

样本k阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$

$EX=\mu$ $Var X=\sigma^2$

(1) $E\bar{X}=E(\frac{1}{n}(x_1+\cdots+x_n))=\frac{1}{n}(Ex_1+\cdots+Ex_n)=\frac{1}{n}n\mu=\mu$

(2) $Var\bar{X}=Var(\frac{1}{n}\sum_{i=1}^{n}x_i)=\frac{1}{n^2}Var(\sum_{i=1}^{n}x_i)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$

(3) $ES^2=\sigma^2$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}[(x_i-\mu)-(\bar{x}-\mu)]^2$$

$$=\sum_{i=1}^{n}[(x_i-\mu)^2-2(x_i-\mu)(\bar{x}-\mu)+(\bar{x}-\mu)^2]$$

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

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

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

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

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

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

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

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

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

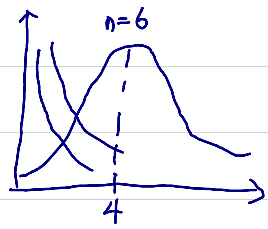
$$=\frac{1}{n-1}(\sum_{i=1}^{n}Dx_i-nD\bar{x})$$

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

$$=\sigma^2$$

6.3 抽样分布　　统计量分布

1. 正态函数　　2. χ^2分布　$\chi^2(n)$



1) $\chi^2(2)$　$\lambda=\frac{1}{2}$ 的指数分布

2) 单峰曲线 $n-2$ 取最大，不对称

n增大，峰向右，越对称

定理6.2：$X_1,\cdots,X_n$ 独立　$N(0,1)$

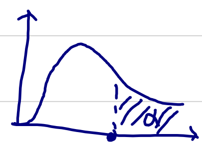
$$\chi^2=\sum_{i=1}^{n}X_i^2\sim\chi^2(n)\qquad EX=n\quad Var\,X=2n$$

由中心极限定理　$X\sim\chi^2(n)$　n充分大　$\frac{X-n}{\sqrt{2n}}\overset{近似}{\sim}N(0,1)$

定理6.3　$X\sim\chi^2(n)$　$Y\sim\chi^2(m)$　$X\ Y$独立　$X+Y\sim\chi^2(m+n)$

推论：$X_i\sim\chi^2(m_i)$ 独立　$\sum_{i=1}^{n}X_i\sim\chi^2(\sum_{i=1}^{n}m_i)$

上α分位数：$P(\chi^2>\chi^2_\alpha(n))=\alpha$



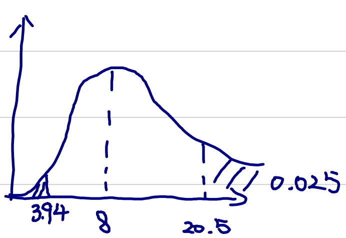
$\chi^2_{0.05}(10)=18.3$　　$\chi^2_{0.1}(25)=34.4$

查表求点，α为概率

(1)　$X\sim\chi^2(10)$　$P(X>a)=0.025$　　$P(X<b)=0.05$

$n=10$　$\alpha=0.025$　$\chi^2_{0.025}(10)=20.5=a$　　　$1-P(X\geq b)=0.05$

$P(X\geq b)=0.95$　　$b=\chi^2_{0.95}(10)=3.94$



(2)　$X_1,\cdots,X_6$　$N(0,2^2)$　　$P(\sum_{i=1}^{6}X_i^2>6.54)$　　$X_i\sim N(0,2^2)$

$\mu\quad\sigma^2$　　　$\frac{X_i-0}{2}\sim N(0,1)$

$\sum_{i=1}^{6}(\frac{X_i}{2})^2\sim\chi^2(6)$　　$P(\frac{\sum_{i=1}^{6}X_i^2}{4}>\frac{6.54}{4})=P(\chi^2(6)>1.635)=0.95$

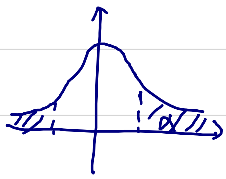
t分布 $X \sim t(n)$

$n \geq 30$ 与正态分布区别很小 n越小，图与正态分布差距大

定理 6.4 $X \sim N(0,1)$ X Y 独立 $\frac{X}{\sqrt{Y/n}} \sim t(n)$

$Y \sim \chi^2(n)$

$P(T > t_\alpha(n)) = \alpha$



$t_{1-\alpha}(n) = -t_\alpha(n)$

(3) $X \sim N(2,1)$ $Y_1, \cdots Y_4$ $N(0,4)$ 独立 $T = \frac{4(X-2)}{\sqrt{\sum_{i=1}^{4} Y_i^2}}$, $P(|T| > t_0) = 0.01$

$\frac{X-2}{1} \sim N(0,1)$ $\frac{Y_i - 0}{2} \sim N(0,1)$ $\sum_{i=1}^{4} \left(\frac{Y_i}{2}\right)^2 \sim \chi^2(4)$

$$\frac{\frac{X-2}{1}}{\sqrt{\sum_{i=1}^{4} \left(\frac{Y_i}{2}\right)^2 / 4}} = \frac{4(X-2)}{\sqrt{\sum_{i=1}^{4} (Y_i)^2}} \sim t(4)$$

$P(T > t_0) = 0.005$ $t_0 = 4.604$

F分布 $F(n_1, n_2)$

定理 6.5 $X \sim \chi^2(n_1)$ $Y \sim \chi^2(n_2)$ X Y 独立 $\frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

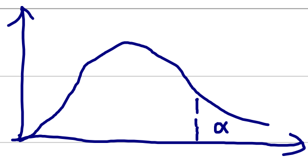
$\frac{Y/n_2}{X/n_1} \sim F(n_2, n_1)$ $F \sim F(n_1, n_2)$ $\frac{1}{F} \sim F(n_2, n_1)$

(4) $X_1, \cdots, X_6$ $N(0, \sigma^2)$ $\frac{2(X_1^2 + X_2^2)}{X_3^2 + X_4^2 + X_5^2 + X_6^2}$ 几个变量 n就是几

$\frac{X_i}{\sigma} \sim N(0,1)$ $\frac{X_1^2}{\sigma^2} + \frac{X_2^2}{\sigma^2} \sim \chi^2(2)$ $\frac{X_3^2}{\sigma^2} + \cdots + \frac{X_6^2}{\sigma^2} \sim \chi^2(4)$

$$\frac{\left(\frac{X_1^2}{\sigma^2} + \frac{X_2^2}{\sigma^2}\right)/2}{\left(\frac{X_3^2}{\sigma^2} + \cdots + \frac{X_6^2}{\sigma^2}\right)/4} = \frac{2(X_1^2 + X_2^2)}{X_3^2 + \cdots + X_6^2} \sim F(2,4)$$

$P(F > F_\alpha(n_1, n_2)) = \alpha$



$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$

$F \sim F(n_1, n_2) \quad \frac{1}{F} \sim F(n_2, n_1)$

$\underline{1-\alpha} = P(F > F_{1-\alpha}(n_1, n_2)) = P\left(\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1,n_2)}\right) = \underline{1 - P\left(\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1,n_2)}\right)}$

$\Rightarrow P\left(\frac{1}{F} > \underline{\frac{1}{F_{1-\alpha}(n_1,n_2)}}\right) = \alpha \qquad F_\alpha(n_2, n_1) = \frac{1}{F_{1-\alpha}(n_1, n_2)}$

上α分位数

(5) $F \sim F(10, 15) \quad \lambda_1, \lambda_2 \quad P(F > \lambda_1) = 0.01 \quad P(F \leq \lambda_2) = 0.01$

$\lambda_1 = F_{0.01}(10, 15) = 3.8 \qquad P(F \leq \lambda_2) = P\left(\frac{1}{F} \geq \frac{1}{\lambda_2}\right) = 0.01$

$\frac{1}{F} \sim F(15, 10) \quad \frac{1}{\lambda_2} = 4.56 \quad \lambda_2 = 0.2193$

6.3.2 正态总体下的抽样分布

总体是正态分布，抽样本，构造统计量的分布

$\boxed{\text{一个正态总体}}$

定理6.6 $X \sim N(\mu, \sigma^2)$ $\{x_1, \dots, x_n\}$ 样本

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

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

$E\bar{X} = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n} n\mu = \mu$

$D\bar{X} = D(\qquad) = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$

(2) $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$

样本均值

(3) $\bar{x}$ 与 S^2 独立

定理6.7 (1) $\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \sim \chi^2(n)$ $\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$

总体期望

(2) $\frac{\bar{x}-\mu}{S}\sqrt{n} \sim t(n-1)$ $\frac{\frac{\bar{x}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/n-1}} \sim t(n-1)$

定理6.8 $X \sim N(\mu_1,\sigma_1^2)$ $Y \sim N(\mu_2,\sigma_2^2)$ $\{x_1,\dots x_{n}\}$ $\{Y_1\dots Y_{n_2}\}$

$\bar{X}$ $\bar{Y}$ S_1^2 S_2^2 (1) $\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

$\bar{X} \sim N(\mu_1, \sigma_1^2/n)$

$\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$ $\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$

(2) $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$ $\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$

$\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/n_1-1}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/n_2-1} \sim F(n_1-1, n_2-1)$

(3) $\sigma_1^2=\sigma_2^2=\sigma$时 $T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$

(1) $X \sim N(\mu^2, \sigma^2)$ $\bar{X}$, S^2 $n=16$ 求k $P(\bar{x}>\mu+kS)=0.95$

$\frac{\bar{x}-\mu}{S}\sqrt{n}=\frac{4(\bar{x}-\mu)}{S} \sim t(16-1)$

$4k=-1.753$

$P(\bar{x}>\mu+kS)=P\left(\frac{4(\bar{x}-\mu)}{S}>\frac{4kS}{S}\right)=0.95$

$k=-0.438$

(2) $X \sim N(\mu, \sigma^2)$ $(x_1, \cdots x_{n+1})$ 样本 $\bar{x}_n$, S_n^2 是 $(x_1, \cdots, x_n)$ 的均值和方差

$\frac{x_{n+1}-\bar{x}}{S_n}\sqrt{\frac{n}{n+1}}$ 的分布 $\sim t(n-1)$

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

$U = \frac{x_{n+1}-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}} \sim N(0,1)$ $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$

$$\frac{\frac{x_{n+1}-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2}\Big/ n-1}} \sim t(n-1)$$

第七章 参数估计

7.1 点估计 区间估计

点估计：185

区间估计：182~187

总体分布	总体参数	样本
$N(\mu, \sigma^2)$	μ, σ^2	$x_1 \cdots x_n$
$P(\lambda)$	λ ←	构造函数
$[a,b]$	a,b	$\hat{\theta} = \hat{\theta}(x_1 \cdots x_n)$

参数空间：取值范围

7.1.1 矩估计法

总体的矩 ← 样本的矩

一阶 EX ← 一阶 $\bar{x} = \frac{1}{n}\sum x_i$

二阶 EX^2 ← 二阶 $A_2 = \frac{1}{n}\sum x_i^2$

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

$= \frac{1}{n}\sum x_i^2 - 2\frac{1}{n}\bar{x}\sum x_i + \frac{1}{n}n\bar{x}^2$

$= \frac{1}{n}\sum x_i^2 - 2\bar{x}^2 + \bar{x}^2$

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

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

$(x_1, x_2, \cdots, x_n)$ 是样本，求 μ, σ^2 的矩估计

$EX = \mu$ $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

总体一阶矩 样本一阶矩

$\hat{\mu} = \bar{x}$

$DX = EX^2 - (EX)^2$ $EX^2 = DX + (EX)^2 = \sigma^2 + \mu^2$ 总体二阶

样本二阶：$A_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

$\hat{\sigma}^2 = A_2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = B_2$

$\Rightarrow \hat{\mu} = \bar{x}$ $\hat{\sigma}^2 = B_2$

(2) $X\sim P(\lambda)$ $(x_1 \cdots x_n)$ λ的矩估计

$EX=\lambda$ $\hat{\lambda}=\bar{X}$ ✓ $\hat{\lambda}=B_2$

(3) $X\sim U[\theta_1,\theta_2]$, $(x_1\cdots x_n)$ θ_1,θ_2的矩估计

$EX=\frac{1}{2}\theta_1+\frac{1}{2}\theta_2$ $DX=EX^2-(EX)^2$

$\frac{(\theta_2-\theta_1)^2}{12}=EX^2-\frac{(\theta_1+\theta_2)^2}{4}$

$EX^2=\frac{(\theta_2-\theta_1)^2}{12}+\frac{(\theta_1+\theta_2)^2}{4}=A_2$

$$\Rightarrow \begin{cases} \frac{1}{2}(\hat{\theta}_1+\hat{\theta}_2)=\bar{X} \\ \frac{(\hat{\theta}_2-\hat{\theta}_1)^2}{12}+\frac{(\hat{\theta}_1+\hat{\theta}_2)^2}{4}=A_2 \end{cases} \Rightarrow \hat{\theta}_1=\bar{X}-\sqrt{3B_2} \quad \hat{\theta}_2=\bar{X}+\sqrt{3B_2}$$

(4)

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

总体 样本$\{1,2,1\}$

$EX=\theta^2+4\theta(1-\theta)+3(1-\theta)^2=$ $\bar{X}=\frac{1+2+1}{3}=\frac{4}{3}$

$\hat{\theta}=\frac{5}{6}$

7.1.2 极大似然估计 ☆

100个球 黑99 白1 求黑球,白球的概率 0.99, 0.01

100个球 $\underbrace{\text{黑、}\theta=99\text{个或1个}\quad \text{白}\quad \text{1个或99个}}_{\text{总体}}$ 摸一个黑的 $\theta=99$

学生: $\underbrace{8:2 \quad \text{会}\ 8\ \ 2 \quad \text{不会}\ 2\ \ 8}_{\text{总体}}$ $\underbrace{\text{提问3个题}}_{\text{样本}}$,不会8 会2

概率大的事件比概率小的事件更易发生

将使A发生的P最大的参数值做为估计值

1) 写出概率/密度函数

2) 写似然函数 $L(\lambda)$ 参数

3) 两边取 $\ln$

4) 对 λ 求导（偏导）

5) 导数 $=0$

例 2 总体 $X \sim P(\lambda)$ $(x_1,\dots,x_n)$ 样本 λ 的极大似然估计

总体概率函数 $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\ (k=0,1,2,\cdots)$

则 L 的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{x_1+\cdots+x_n}}{\prod_{i=1}^{n}x_i!}e^{-n\lambda}$$

$$\ln(L(\lambda))=-\ln\prod_{i=1}^{n}x_i!+(x_1+\cdots+x_n)\ln\lambda-n\lambda$$

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

$$\hat{\lambda}=\frac{x_1+x_2+\cdots+x_n}{n}=\bar{x}$$

例 3 总体 $X\sim Exp(\lambda)$

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

$$L(\lambda)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}=\lambda^n e^{-\lambda(x_1+\cdots+x_n)}$$

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

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{n}{\lambda}-(x_1+\cdots+x_n)=0 \Rightarrow \hat{\lambda}=\frac{n}{x_1+\cdots+x_n}=\frac{1}{\bar{x}}$$

例 4 $X\sim N(\mu,\sigma^2)$

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

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=\left(\frac{1}{\sqrt{2\pi}}\right)^n\left(\frac{1}{\sigma}\right)^n e^{-\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}}$$

$$\ln L(\mu,\sigma^2)=n\ln\frac{1}{\sqrt{2\pi}}-\frac{n}{2}\ln\sigma^2-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}$$

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=+\frac{2(x_1-\mu)+\cdots+2(x_n-\mu)}{2\sigma^2}=\frac{x_1+\cdots+x_n-n\mu}{\sigma^2}=0\Rightarrow\hat{\mu}=\bar{x}$$

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2}\left(\frac{1}{\sigma^2}\right)+\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^4}=0$$

$$\hat{\sigma}^2=\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{n}$$

$$=B_2$$

5) $X \sim U[\theta_1, \theta_2]$

$$f(x,\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2-\theta_1} & x\in[\theta_1,\theta_2] \\ 0 & \text{else} \end{cases}$$

$$L(\theta_1,\theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2-\theta_1} = \frac{1}{(\theta_2-\theta_1)^n}$$

$$\ln L(\theta_1,\theta_2) = -n\ln(\theta_2-\theta_1)$$

$$\left.\begin{array}{l} \frac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_1} = \frac{n}{\theta_2-\theta_1} = 0 \\ \frac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_2} = \frac{-n}{\theta_2-\theta_1} = 0 \end{array}\right\} \text{无解}$$

$\frac{1}{(\theta_2-\theta_1)^n}$ 最大，θ_1,θ_2越近越好

$\hat{\theta}_1 = \min\{x_1,\cdots,x_n\}$ $\hat{\theta}_2 = \max\{x_1,\cdots,x_n\}$

7.2 点估计的优良性准则

无偏性 $E\hat{\theta} = \theta$

(1) 总体X $EX=\mu$ $DX=\sigma^2$ $(x_1 \cdot\cdot x_n)$

① $\bar{X}$是μ的无偏估计 $E\bar{X}=\mu$

② 样本方差 S^2是σ^2的无偏估计 $ES^2=\sigma^2$

③ 未修正方差是σ^2的有偏估计

$\hat{\theta}$是θ的无偏 $g(\hat{\theta})$不一定是$g(\theta)$的无偏

(2) S^2是σ^2的无偏 $\sqrt{S^2}$不是$\sqrt{\sigma^2}$的无偏

$$DS = ES^2 - (ES)^2$$
$$= \sigma^2 - (ES)^2$$
$$ES = \sqrt{\sigma^2 - DS} \leq \sigma$$

(3) $\mu = EX$ $(x_1, \cdots, x_n)$

$\hat{\mu} = C_1 x_1 + \cdots + C_n x_n$ $C_1 + C_2 + \cdots + C_n = 1$

$E\hat{\mu} = C_1\mu + \cdots + C_n\mu = \mu$

7.2.2 有效性 $D(\hat{\theta}_1) \leqslant D(\hat{\theta}_2)$

(1) 总体 X $EX = \mu$ $DX = \sigma^2$ $\mu \begin{cases} x_1 & EX_1 = \mu & DX_1 = \sigma^2 \\ \bar{x} \quad 有效 & & D\bar{x} = \frac{\sigma^2}{n} \end{cases}$

(2) $\begin{cases} a_1 x_1 + \cdots + a_n x_n & a_1 + \cdots + a_n = 1 \\ \bar{x} \end{cases}$

$D\hat{\theta} = a_1^2 Dx_1 + \cdots + a_n^2 Dx_n = \sigma^2(a_1^2 + \cdots + a_n^2) \geqslant \frac{\sigma^2}{n}$ $a_1^2 + \cdots + a_n^2 \geqslant \frac{1}{n}$

$1 = (a_1 + \cdots + a_n)^2 \quad \cdots \quad \leqslant n(a_1^2 + \cdots + a_n^2)$

$(a_1 + a_2 + a_3)^2 = a_1^2 + a_2^2 + a_3^2 + 2a_1a_2 + 2a_1a_3 + 2a_2a_3 \leqslant a_1^2 + a_2^2 + a_3^2 + a_1^2 + a_2^2 + a_1^2 + a_3^2 + a_2^2 + a_3^2$

$= 3(a_1^2 + a_2^2 + a_3^2)$

7.2.3 相合性（一致性）

$\lim_{n\to\infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$

7.3.1 置信区间 区间估计 180~185 越小越好

① 区间长度 ②概率落入区间 $P(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2) = 1 - \alpha$ 置信度

$[\hat{\theta}_1, \hat{\theta}_2]$ 区间 ← 求

$[\hat{\theta}_1, \hat{\theta}_2]$ 能套住 θ 的概率 ✓ | θ 落在区间的概率是 90%

θ 未知但是确定 90%

取样 100 次，100 个区间，90 次套中 θ，10 次未套中

枢轴变量法

① 位置θ [ ]，90%

拉女生男生（已知），三人位置好确定，扔二人，剩自己

② 求θ的区间 $-1<\frac{\theta+2}{4}<1$

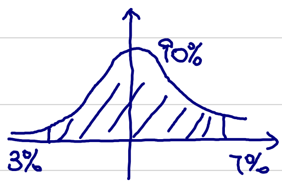
定义：1) $I=I(T,\theta)$ ⟶ 枢轴变量

已知　未知参数

分布F已知且与θ无关　　（不一定是F分布）

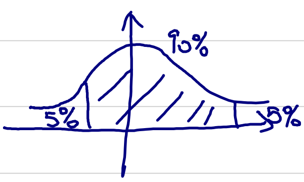
2) 给定 $1-\alpha$，确定F的上 $\frac{\alpha}{2}$ 分位数，上 $(1-\frac{\alpha}{2})$ 分位数

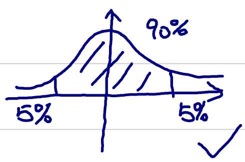
$V_{\frac{\alpha}{2}}$　　$V_{1-\frac{\alpha}{2}}$



$P(V_{1-\frac{\alpha}{2}}\le I(T,\theta)\le V_{\frac{\alpha}{2}})=1-\alpha$

$\hat{\theta}_1\le\theta\le\hat{\theta}_2$





区间长度越小越好

7.3.2 一个正态总体　均值和方差的区间估计

① σ^2 已知，估计 μ　　$U=\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$

给定 $1-\alpha$，令 $P(U>u_{\frac{\alpha}{2}})=\frac{\alpha}{2}$　　$\Phi_0(u_{\frac{\alpha}{2}})=1-\frac{\alpha}{2}$

右端点

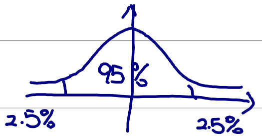
$P\left(-u_{\frac{\alpha}{2}}\le\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\le u_{\frac{\alpha}{2}}\right)=1-\alpha$　　区间长度与 $1-\alpha$ 矛盾

$-\bar{X}-\frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}\le-\mu\le\frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}-\bar{X}$

$\bar{X}-\frac{\sigma u_{\frac{\alpha}{2}}}{\sqrt{n}}\le\mu\le\bar{X}+\frac{u_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}$

(1) 5个灯泡 1650 1700 1680 1820 1800 $X \sim N(\mu, 9)$ $\alpha = 0.05$

$$\bar{x} = \frac{1650 + \cdots + 1800}{5} = 1730 \qquad \sigma^2 = 9 \quad \sigma = 3$$



$u_{0.025} = 1.96$

$$-1.96 \le \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \le 1.96$$

$$-1.96 \le \frac{\sqrt{5}(1730-\mu)}{3} \le 1.96$$

$$1727.37 \le \mu \le 1732.63$$

② σ^2 未知，估计 μ $T = \frac{\sqrt{n}(\bar{x}-\mu)}{S}$ 样本标准差 $\sim t(n-1)$

(不可用)

给定 $1-\alpha$ 上 $\frac{\alpha}{2}$ 分位数 $t_{\frac{\alpha}{2}}(n-1)$

$$P\left(-t_{\frac{\alpha}{2}}(n-1) \le \frac{\sqrt{n}(\bar{x}-\mu)}{S} \le t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

$$\bar{x} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \le \mu \le \bar{x} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)$$

方差的区间估计

① μ 已知，对 σ^2 的区间估计

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

给定 $1-\alpha$ $\chi^2_{1-\frac{\alpha}{2}}(n)$ $\chi^2_{\frac{\alpha}{2}}(n)$ 2点都要查表

$$\chi^2_{1-\frac{\alpha}{2}}(n) \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \le \chi^2_{\frac{\alpha}{2}}(n)$$

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)} \le \sigma^2 \le \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}$$

② μ 未知，估计 σ^2 $\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

给定 $1-\alpha$ $\chi^2_{1-\frac{\alpha}{2}}(n-1)$ $\chi^2_{\frac{\alpha}{2}}(n-1)$

$$\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right]$$

估计

μ	σ^2已知	$\frac{\bar{x}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$	$[\bar{x}-\frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}}, \bar{x}+\frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}}]$
μ	σ^2未知	$\frac{\bar{x}-\mu}{S}\sqrt{n}\sim t(n-1)$	$[\bar{x}-\frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1), \bar{x}+\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)]$ 对称,
σ^2	μ已知	$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\sim\chi^2(n)$	$\left[\frac{\sum(x_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \frac{\sum(x_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}\right]$
σ^2	μ未知	$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right]$ 不对称

第八章 假设检验

8.1 基本概念

$\mu=100$ $\quad$ $\mu\neq 100$

H_0：原假设 $\quad$ H_1：对立假设 $\quad$ 接受H_0拒绝H_1 or 接受H_1拒绝H_0

(不能轻易否定)

小概率 α 越小，发生后怀疑 H_0 理由越显著 ⟶ α 称为显著性水平

拒绝域 $W=\{(x_1,\cdots,x_n)\mid t\in I\}$

if $(x_1,\cdots,x_n)\in W$，拒绝H_0接受H_1

两类错误：① 弃真错误 $P(\text{reject } H_0 \mid H_0 \text{ is True})=\alpha$

② 纳伪错误 $P(\text{accept } H_0 \mid H_0 \text{ is False})=\beta$

8.2 正态总体的参数假设检验

① σ^2 已知，检验 μ

(1) H_0: $\mu=\mu_0$ $\quad$ H_1: $\mu\neq\mu_0$

$U=\frac{\bar{X}-\mu_0}{\sigma_0}\sqrt{n}\sim N(0,1)$ 观测值：$u=\frac{\bar{x}-\mu_0}{\sigma_0}\sqrt{n}$

$A=(|U|>C)$

$W=\{(x_1,\cdots,x_n)\mid |u|>c\}$

For given α $\quad$ $P(|U|>C)=\alpha$ $\quad$ $C=u_{\frac{\alpha}{2}}$

$\Rightarrow W=\{(x_1,\dots,x_n): \left|\frac{\bar{x}-\mu_0}{\sigma_0}\sqrt{n}\right|>u_{\frac{\alpha}{2}}\}$

(2) $H_0: \mu=\mu_0$ $H_1: \mu>\mu_0$ $P(U>u_\alpha)=\alpha$

$W=\{(x_1,\dots,x_n): \frac{\bar{x}-\mu}{\sigma}\sqrt{n}>u_\alpha\}$

(3) $H_0: \mu=\mu_0$ $H_1: \mu<\mu_0$ $P(U<-u_\alpha)=\alpha$

$W=\{(x_1,\dots,x_n): \frac{\bar{x}-\mu}{\sigma}\sqrt{n}<-u_\alpha\}$

1. $\mu=10kg$ 甲: 10.1, 10, 9.8, 9.9, 9.9 $\alpha=0.05$ $\sigma=0.1$

$H_0: \mu=10$ $H_1: \mu\neq 10$ $\bar{x}=9.94$

$W=\{(x_1,\dots,x_5): \left|\frac{\bar{x}-\mu}{\sigma}\sqrt{n}\right|>u_{0.025}\}$ $u_{0.025}=1.96$

$u=\frac{\bar{x}-10}{0.1}\sqrt{5}=-1.34$ $|u|<1.96$ $\Rightarrow$ 甲正常

③ σ^2 未知，检验 μ

H_0成立: $T=\frac{\bar{x}-\mu_0}{S}\sqrt{n}\sim t(n-1)$

$W=\{(x_1,\dots,x_n): |t|>c\}$

算法同上

	H_0	H_1		
U检验	$\mu=\mu_0$	$\mu\neq\mu_0$	$U=\frac{\bar{x}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$	$\{\|u\|>u_{\frac{\alpha}{2}}\}$
	$\mu\le\mu_0$	$\mu>\mu_0$		$\{u>u_\alpha\}$
	$\mu\ge\mu_0$	$\mu<\mu_0$		$\{u<-u_\alpha\}$
T检验	$\mu=\mu_0$	$\mu\neq\mu_0$	$T=\frac{\bar{x}-\mu}{S}\sqrt{n}\sim t(n-1)$	$\{\|t\|>t_{\frac{\alpha}{2}}(n-1)\}$
	$\mu\le\mu_0$	$\mu>\mu_0$		$\{t>t_\alpha(n-1)\}$
	$\mu\ge\mu_0$	$\mu<\mu_0$		$\{t<-t_\alpha(n-1)\}$

③ μ 未知 检验 σ^2

$H_0: \sigma^2=\sigma_0^2 \quad H_1: \sigma^2\neq\sigma_0^2$

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}=\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\sim\chi^2(n-1)$$

事件 $A=(\chi^2<C_1)\cup(\chi^2>C_2)$

$W=\{(x_1,\dots,x_n):\chi^2<C_1 \text{ or } \chi^2>C_2\}$

$P(\chi^2<C_1)=P(\chi^2>C_2)=\frac{\alpha}{2} \Rightarrow C_1=\chi^2_{1-\frac{\alpha}{2}}(n-1) \quad C_2=\chi^2_{\frac{\alpha}{2}}(n-1)$

④ μ 已知 检验 σ^2 $\quad X\sim N(\mu,\sigma^2)$

$H_0: \sigma^2=\sigma_0^2 \quad H_1: \sigma^2\neq\sigma_0^2$

$$\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\sim\chi^2(n)$$

$W=\{(x_1,\dots,x_n):\chi^2<\chi^2_{1-\frac{\alpha}{2}}(n) \text{ or } \chi^2>\chi^2_{\frac{\alpha}{2}}(n)\}$

	H_0	H_1	H_0成立，统计量	
χ^2检验 μ 已知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\sim\chi^2(n)$	$\{\chi^2<\chi^2_{1-\frac{\alpha}{2}}(n) \text{ or } \chi^2>\chi^2_{\frac{\alpha}{2}}(n)\}$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\{\chi^2>\chi^2_{\alpha}(n)\}$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\{\chi^2<\chi^2_{1-\frac{\alpha}{2}}(n)\}$
χ^2检验 μ 未知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\{\chi^2<\chi^2_{1-\frac{\alpha}{2}}(n-1) \text{ or } \chi^2>\chi^2_{\frac{\alpha}{2}}(n-1)\}$
	$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\{\chi^2>\chi^2_{\alpha}(n-1)\}$
	$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\{\chi^2<\chi^2_{\alpha}(n-1)\}$

8.3 两个总体的假设检验

① σ_1^2, σ_2^2 已知，检验 $\mu_1-\mu_2$

$$U=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$|u|>u_{\frac{\alpha}{2}}$ ，$u>u_\alpha$ ，$u<-u_\alpha$

② $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知，检验 $\mu_1-\mu_2$

$$T=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

③ μ_1, μ_2 未知，检验 $\frac{\sigma_1^2}{\sigma_2^2}$

$$F=\frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$$

8.4 拟合优度检验

提出假设 H_0：总体 X 的分布函数为 $F(x,\theta)$

$\theta=(\theta_1,\theta_2,\cdots,\theta_r)$

若未知，则估计 $\hat{\theta}$ ，得 $F(x,\hat{\theta})$

将区间分成 k 个小区间

$A_1=(-\infty,a_1]$, $A_2=(a_1,a_2]$, $\cdots$, $A_k=(A_{k-1},\infty)$

f_i 落入区间 A_i 的频数 $\sum_{i=1}^{k} f_i=n$

根据 $F(x,\hat{\theta})$ ，算出 $p_i\geq 0$, $i=1,\cdots,n$ $\sum_{i=1}^{k} p_i=1$

np_i：理论频数

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-np_i)^2}{np_i} \sim \chi^2(k-r-1)$$

要求 $n\geq 30$ $np_i\geq 5$ or $np_i\geq 10$

$W=\{\chi^2>\chi^2_\alpha(k-r-1)\}$