Chapter 1

Introduction

1.1 Sample Space

- 相同条件
- 实验结果不止一个 (可以事先确定所有的可能结果)
- 进行一次实验之前不确定结果

样本空间的子集为事件. 样本空间的元素为样本点. S(样本空间) 必然发生, 为必然事件. $\varnothing$(空集) 不包含任何样本点, 为不可能事件.

1.2 公式

1.2.1 加法

$$P(A \cup B) = P(A) + P(B) - P(A \cdot B) = P(A \cdot \bar{B}) + P(\bar{A} \cdot B) + P(A \cdot B)$$

1.2.2 减法

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

1.2.3 乘法

若 $P(A) > 0$, 则

$$\begin{aligned} P(A \cap B) &= P(A \cdot B) \\ &= P(B \mid A) \cdot P(A) \\ &= P(A \mid B) \cdot P(B) \end{aligned}$$

必须考虑 A 和 B 是否独立, 若不独立不可以概率直接相乘得到积概率.

1.2.4 取反

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

1.2.5 全概率

设有一样本空间 S, 其中 A 为随机事件 $B_1, B_2, B_3 \ldots B_n$ 的划分, 且 $P(B_i) > 0$, 则

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

划分 $\mathbb{B}$ 并起来等于 S, 且两两互不相交.

1.2.6 贝叶斯

$$P(B_j \mid A) = \frac{P(A \mid B_j) \cdot P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)}$$

1.3 条件概率

条件概率本质就是压缩样本空间

$$\begin{aligned} P(B \mid A) &= \frac{P(AB)}{P(A)} \\ &= \frac{P(A \mid B) \cdot P(B)}{P(A)} \\ P(A \mid B) &= \frac{P(AB)}{P(B)} \\ &= \frac{P(B \mid A) \cdot P(A)}{P(B)} \end{aligned}$$

设 S 为样本空间

$$\begin{aligned} P(B) &= \frac{P(B \cdot S)}{P(S)} \\ P(S) &= 1 \end{aligned}$$

1.4 完备事件组

若事件 $B_1, B_2, B_3 \ldots B_n$ 两两互斥, 且 $B_1 \cup B_2 \cup B_3 \cdots \cup B_n = S$($S$ 为样本空间), 则称 $B_1, B_2, B_3 \ldots B_n$ 为一个完备事件组.

1.5 独立事件

若事件 A, B 相互独立, 则.

$$\begin{cases} P(A \mid B) = P(A) \\ P(B \mid A) = P(B) \end{cases}$$

$$P(A \cdot B) = P(A) \cdot P(B)$$

1.6 应用题

1. 甲乙独立对同一目标射击一次, 命中率分别为 0.6 和 0.5, 求两人中至少一人射中的概率.
2. 从甲乙任选一人同一目标射击一次, 命中率分别为 0.6 和 0.5, 求目标被射中的概率.
3. 甲乙独立对同一目标射击一次, 命中率分别为 0.6 和 0.5, 已知目标命中, 求是甲射中的概率.
4. 从甲乙任选一人同一目标射击一次, 命中率分别为 0.6 和 0.5, 已知目标命中, 求是甲射中的概率.

1.7 总结

- 积事件的概率和条件概率
- 独立的积事件和不独立的
- 正面做和反面做 (至多, 至少)
- 样本空间的划分

重点

- 随机事件的概念
- 古典概型的概率计算方法
- 概率的加法公式
- 条件概率和乘法公式的应用
- 全概率公式和贝叶斯公式的应用

Chapter 2

一维随机变量

2.1 Possibility Distribution Functions

2.1.1 PDF

Possibility Density Function. 概率密度函数.

$$f(x) \geq 0$$

被称为非负性

$$\int_{-\infty}^{+\infty} f(x)\ dx = 1$$

被称为归一性

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

哪里求变量哪里求积分

2.1.2 PMF

Possibility Mass Function. 又被称为分布率.
离散型的 PDF

2.1.3 CDF

Cumulative Distribution Function. 累积分布函数.

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

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

CDF 单调递增 (累积, 不会减少)

$$\begin{aligned} P\{X = k\} &= P\{X \leq k\} - P\{X < k\} \\ &= F(k)|_{x \geq k\text{的那部分函数}} - F(k)|_{x < k\text{的那部分函数}} \end{aligned}$$

若 k 在边界.

2.1.4 PDF 到 CDF

若有 PDF

$$f(x)=\begin{cases} f_1(x), & a\le x<b\\ f_2(x), & b\le x<c\\ f_3(x), & c\le x<d\\ 0, & \text{elsewhere.}\end{cases}$$

则有 CDF

$$F(x)=\begin{cases} 0, & x<a\\ \int_a^x f_1(x)\,dx, & a\le x<b\\ \int_a^b f_1(x)\,dx+\int_b^x f_2(x)\,dx, & b\le x<c\\ \int_a^b f_1(x)\,dx+\int_b^c f_2(x)\,dx+\int_c^x f_3(x)\,dx, & c\le x<d\\ 1, & x>c\end{cases}$$

2.2 连续性随机变量的函数

若 $Y=g(X)$

2.3 $Y=g(X)$ 单调可导

- 求 $Y=g(X)$ 的值域
- $Y=g(X)\Rightarrow$ 关于 X 的函数 (反函数) $x=h(y)$
- $f_Y(y)=f_X[h(y)]\cdot|\frac{dh(y)}{dy}|$

2.4 $Y=g(X)$ 非单调可导

$$f_X(x)\xrightarrow{\text{find }P\{g(X)\le y\}}F_Y(y)\xrightarrow{\frac{d}{dy}}f_Y(y)$$

Figure 2.1: 步骤图

先求 Y 的 CDF (F_Y) 然后求导得到 f_Y
还是得求 $Y = g(X)$ 的值域

$$
\begin{aligned}
F_Y(y) &= P\{Y \le y\} \\
&= P\{g(X) \le y\} \\
&= P\{X \le h(y)\} && h(y) \text{ 为 } g(X) \text{ 的反函数} \\
&= F_X[h(y)] \\
f_Y(y) &= F'_Y(y) \\
&= \frac{d}{d\,[h(y)]} F_X[h(y)] \cdot \frac{d\,[h(y)]}{dy} \\
&= f_X[h(y)] \cdot \frac{d\,[h(y)]}{dy}
\end{aligned}
$$

例子 下面是很麻烦的例子, 不用看了. 直接套用上面的方法就好了, 没必要分类讨论.

$$
f_X(x) = \begin{cases} 2x, & 0 < x \le 1 \\ 0, & \text{elsewhere} \end{cases}
$$

求 $Y = e^{-X}$ 的 PDF.

假解 根据 $P\{a < X \le b\} = \int_b^a f(x)\,dx = F(b) - F(a)$ 易知

$$
\begin{aligned}
F_Y(y) &= P\{Y \le y\} \\
&= P\{e^{-X} \le y\}
\end{aligned}
$$

开始分类讨论
$y \le 0$ 时, $P\{e^{-X} \le y\} = 0$ (因为 e^{-X} 值域大于 0)
$y > 0$ 时, $P\{e^{-X} \le y\} = P\{X \ge -\ln(y)\}$
就可以转换为 $P\{X \le b\} = \int_b^{+\infty} f(x)\,dx$ 的情况.
积分区间就为 $[-\ln(y), +\infty)$ 与 $(0, 1)$ 的交集.
所以可以根据 $(0, 1)$ 确定 $-\ln(y)$ 的取值.

真解 易得

$$
\begin{aligned}
F_Y(y) &= P\{Y \le y\} \\
&= P\{e^{-X} \le y\} \\
&= P\{X \le -\ln(y)\} \\
&= F_X[-\ln(y)] \\
f_Y(y) &= F'_Y(y) \\
&= \frac{d}{d\,[-\ln(y)]} F_X[-\ln(y)] \cdot \frac{d\,[-\ln(y)]}{dy} \\
&= f_X[-\ln(y)] \cdot \frac{d\,[-\ln(y)]}{dy} \\
&= 2 \cdot (-\ln y) \cdot \frac{-1}{y} \\
&= \frac{2 \ln y}{y}
\end{aligned}
$$

与正确答案差一个负号, 可能是因为定义域映射到值域反转了一下?

Chapter 3

二维随机变量

Table 3.1: 联合分布和边缘分布的例子

$P \backslash Y$ / X	100	90	70	X_{marginal}
100	0.01	0.02	0.07	0.1
80	0.02	0.04	0.14	0.2
40	0.07	0.14	0.49	0.7
Y_{marginal}	0.1	0.2	0.7	1

其中 P 代表 $P\{X = i, Y = j\}$ 的概率 (联合分布).
$X_{\text{marginal}} = P\{X = x_i\}$, $Y_{\text{marginal}} = P\{Y = y_j\}$

3.1 Union Distribution

离散型

Table 3.2: 联合分布

$P \backslash Y$ / X	100	90	70
100	0.01	0.02	0.07
80	0.02	0.04	0.14
40	0.07	0.14	0.49

其中 P 代表 $P\{X = i, Y = j\}$ 的概率 (联合分布).

连续型

$$f(x, y)$$

有着类似的非负性和归一性.

$$f(x,y) \geq 0$$
$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} f(x,y)\, dy = 1$$

哪里求概率哪里求积分

$$P\{(X,Y) \in D\} = \iint_D f(x,y)\, dx\, dy$$

CDF

$$F(x,y) = P\{X \leq x, Y \leq y\}$$

3.2 Marginal Distribution

离散型

$$P\{X = x_i\} = \sum_{i=1}^{\infty} P\{X = x_i, Y = y_j\}$$

可以列出 X 和 Y 的边缘 PMF, 以 $P\{X = 100\}$ 为例.

$P\{X = 100\}$
$= P\{X = 100, Y = 100\} + P\{X = 100, Y = 90\} + P\{X = 100, Y = 70\}$

Table 3.3: X 的边缘分布

X	100	80	40
$P\{X = x_i\}$	0.1	0.2	0.7

Table 3.4: Y 的边缘分布

Y	100	90	70
$P\{Y = y_j\}$	0.1	0.2	0.7

连续型

通过对 y 的积分来求 X 的概率密度, 竖着切
通过对 x 的积分来求 Y 的概率密度, 横着切

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy \qquad \text{最后的结果只有 } x\text{, 积分上下限关于 } x$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\, dx \qquad \text{最后的结果只有 } y\text{, 积分上下限关于 } y$$

CDF

$$F_X(x) = \lim_{x \to +\infty} F(x, y)$$
$$F_Y(y) = \lim_{y \to +\infty} F(x, y)$$

3.3 Conditional Distribution

条件概率联合分布比边缘分布.

$$条件 = \frac{联合}{边缘}$$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \qquad \text{是一个关于 } x \text{ 的函数 (不含有 y) } f_Y(y) \neq 0$$

3.4 连续型随机变量的函数的概率分布

3.4.1 $Z = X + Y$

谁在函数中的表达式较为简单就换谁 (在此以 $Y = Z - X$ 为例)

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

积分上下限必定有 z
相当于求关于 z 的**边缘概率密度**, 故不用求 CDF.
关键是确定 x(被积变量) 的积分区域. 根据题目的变量不等式来列出被积变量的不等式, 一般会有

- 中间是被积变量 x, 两边为关于 z 的不等式.
- 中间是被积变量 x, 两边为常数 (或者只有单边) 的不等式
- 其他限制条件 (不等式)

$$\begin{cases} a < x < b \\ z - a < x < z + b \end{cases}$$

取 z 关于被积变量的交集, 可以得到 z 的范围.

3.4.2 $Z = XY$

$$Z = XY$$
$$X = \frac{Z}{Y}$$
$$Y = \frac{Z}{X}$$

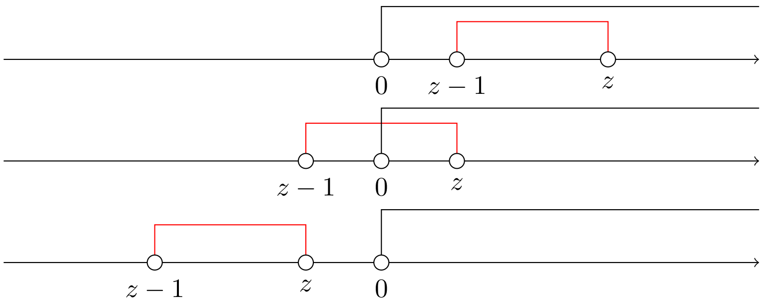

Figure 3.1

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} \cdot f(x, \frac{z}{x})\ dx$$

积分上下限必定有 z
注意前边有个 $\frac{1}{|x|}$.
不等式形式为: x 在中间一个, z 在中间一个.

$$\begin{cases} a < x < b \\ a < z < x \end{cases}$$

3.4.3 $Z = max\{X, Y\}$

- 先求边缘 CDF F_X, F_Y
- $F_{\max} = F_X(z) \cdot F_Y(z)$
- $F_{\min}(z) = 1 - [1 - F_X(z)] \cdot [1 - F_Y(z)]$ 取边缘 CDF 与 1 差值之积, 再取积与 1 差值
- $f_{\max}(z) = \frac{dF_{\max}(z)}{dz}$

若 X, Y 独立同分布

- $F_{\max} = [F_X(z)]^2$
- $f_{\max} = F'_{\max}(z)$

若有 n 个随机变量相互独立且具有相同分布 $F(x)$.

$$F_{\max} = [F(z)]^n \qquad z = \max(X_1, X_2, X_3 \ldots, X_n)\text{的分布函数}$$
$$F_{\min} = 1 - [1 - F(z)]^n \qquad z = \min(X_1, X_2, X_3 \ldots, X_n)\text{的分布函数}$$

3.5 两个随机变量的独立性

联合等于两个边缘相乘则独立.

$$f(x, y) = f_X(x) \cdot f_Y(y) \iff X \text{ 与 } Y \text{ 独立 (充要条件)}$$
$$f_{X|Y}(x \mid y) = f_X(x) \iff X \text{ 与 } Y \text{ 独立}$$

3.5.1 求分布律未知量

若 X, Y 相互独立
先列出边缘分布. 联合分布率等于边缘分布率相乘 (可以任意列).
也可以用归一性, (X 和 Y 的) 边缘分布之和为 1.
当然最好算的是**各行各列成比例**.

3.5.2 连续型的独立性

若 X, Y 相互独立, 则其联合概率密度 (PDF) 中函数值非零的区域必为方形区域 $a < x < b, c < y < b$. (必要不充分)

3.6 补充: 二重积分

3.6.1 竖着切

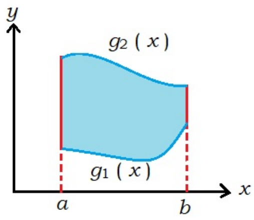

Figure 3.2: 竖着切 (X 型区域)

往哪切哪边就是常数, 提出来放到左边.
画两条竖线 $x = a$, $x = b$, 作为边缘, 当然是常数.
上下是关于 x 的函数 $y = g_1(x)$, $y = g_2(x)$, RHS 不含 y, LHS 就是 y.
从左到右, 从下到上.

$$\int_a^b dx \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy$$

$$\int_a^b [\int_{g_1(x)}^{g_2(x)} f(x, y)\, dy]\, dx$$

3.6.2 横着切

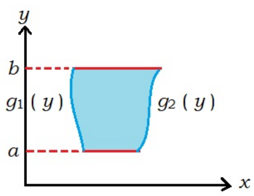

Figure 3.3: 横着切 (Y 型区域)

往哪切哪边就是常数, 提出来放到左边.
画两条横线 $y = a$, $y = b$, 作为边缘, 当然是常数.
上下是关于 y 的函数 $x = h_1(y)$, $x = h_2(y)$, RHS 不含 x, LHS 就是 x.
从下到上, 从左到右.

$$\int_a^b dy \int_{h_1(y)}^{h_2(y)} f(x,y)\,dx$$

Chapter 4

随机变量的数字特征

4.1 常见分布

4.1.1 二项分布

$X \sim \mathcal{B}\left(n,p\right)$

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

$E(X) = np$
$D(X) = np \cdot (1-p)$

两点分布

又被称作 0–1 分布, 就是 $X \sim \mathcal{B}\left(1,p\right)$ 的情况

$$P\{X=k\} = p^k \cdot (1-p)^{1-k} = \begin{cases} p & k=1, \\ (1-p) & k=0. \end{cases}$$

4.1.2 泊松分布

$X \sim \pi(\lambda)$

$$P\{x=k\} = \frac{\lambda^k \cdot e^{-\lambda}}{k\,!}$$

$E(X) = \lambda$
$D(X) = \lambda$

4.1.3 正态分布

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

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

$E(X) = \mu$
$D(X) = \sigma^2$

标准正态分布

服从 $\mu = 0$, $\sigma = 1$ 的正态分布
若 $X \sim \mathcal{N}(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

性质 标准正态分布具有如下性质

- $\mu = 0$, $\Phi(0) = \frac{1}{2}$
- $\Phi(-x) = 1 - \Phi(x)$

若 $X \sim \mathcal{N}(\mu, \sigma^2)$

- $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$
- $P\{X < a\} = \Phi(\frac{a-\mu}{\sigma})$
- $P\{a < X < b\} = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$

正态分布的线性组合

若 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ 且 X 与 Y 相互独立则

$$X \pm Y \sim \mathcal{N}(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$$

若 $U \sim \mathcal{N}(\mu, \sigma^2)$, $V = aU + b$

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

4.1.4 指数分布

$$f(x) = \begin{cases} \frac{1}{\lambda} \cdot e^{-\frac{1}{\lambda} \cdot x} & x > 0, \\ 0 & \text{elsewhere.} \end{cases}$$

$E(X) = \lambda$
$D(X) = \lambda^2$

4.1.5 均匀分布

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

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

4.2 期望值

本质为加权平均.

$$E(X)=\sum_{i=1}^{+\infty}X_i\cdot P\{X=X_i\}$$

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

参数为一维函数情况.

$$E[g(x)]=\sum_{i=1}^{+\infty}g(x_i)\cdot P\{X=x_i\}$$

$$E[g(x)]=\int_{-\infty}^{+\infty}g(x)\cdot f(x)\,dx$$

$$E(x^2)=\int_{-\infty}^{+\infty}x^2\cdot f(x)\,dx \qquad \text{注意 } f(x) \text{ 的参数仍然是 } x$$

参数为二维函数情况 (积事件, 联合分布率)

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

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

$$E[g(x,y)]=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}g(x,y)\cdot f(x,y)\,dy$$

4.3 方差

平方的期望减去期望的平方

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

4.4 期望值与方差的线性组合

4.4.1 期望值

- $E(\text{const})=\text{const}$
- $E(\text{const}\cdot x)=\text{const}\cdot x$
- $E(aX\pm bY)=a\cdot E(X)\pm b\cdot E(Y)$

4.4.2 方差

- $D(\text{const})=0$
- $D(\text{const}\cdot x)=\text{const}^2\cdot x$
- $D(aX\pm bY)=a^2\cdot D(X)+b^2\cdot D(Y)$

4.5 协方差

协方差 (covariance), 乘积的期望减去期望的乘积.

$$cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

4.6 相关系数

协方差比方差开根号之积.

$$\begin{aligned}\rho_(XY) &= \frac{cov(X,Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} \\ &= \frac{E(X \cdot Y) - E(X) \cdot E(Y)}{\sqrt{E(X^2) - E^2(X)} \cdot \sqrt{E(Y^2) - E^2(Y)}}\end{aligned}$$

4.7 中心极限定理

正态分布的普遍存在性
设 $X_1, X_2, X_3 \ldots X_n$ 独立同分布, 当 n 充分大时, $\sum_{i=1}^{n} X_i$ 近似服从正态分布.
(n 越大越接近真实值)

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

μ 为随机变量的数学期望, σ^2 为随机变量的方差
特别的, 设随机变量 Y 服从二项分布时 $Y \sim \mathcal{B}(n,p)$, Y 近似服从如下正态分布.

$$Y \sim \mathcal{N}\left[np, np \cdot (1-p)\right]$$

可以看出 $\mu = np$, $\sigma^2 = np \cdot (1-p)$

4.7.1 解题步骤

- 写出期望值 $E(X)$, 方差 $D(X)$, 知道个数 n
- $\mu = E(X)$, $\sigma^2 = D(X)$, 记得解出标准差 σ (下面用的都是开根号的)

以下拿要比较的随机变量 $X \leq 10$ 作为例子, 实际的随机变量取值不一定.

$$\begin{aligned}P\{X \leq 10\} &= P\{\frac{X - n \cdot \mu}{\sqrt{n} \cdot \sigma} \leq \frac{10 - n \cdot \mu}{\sqrt{n} \cdot \sigma}\} \\ &= \Phi(\frac{10 - n \cdot \mu}{\sqrt{n} \cdot \sigma}) \\ P\{X > 10\} &= 1 - P\{X \leq 10\} \\ &= P\{\frac{X - n \cdot \mu}{\sqrt{n} \cdot \sigma} > \frac{10 - n \cdot \mu}{\sqrt{n} \cdot \sigma}\} \\ &= 1 - P\{\frac{X - n \cdot \mu}{\sqrt{n} \cdot \sigma} \leq \frac{10 - n \cdot \mu}{\sqrt{n} \cdot \sigma}\} \\ &= 1 - \Phi(\frac{10 - n \cdot \mu}{\sqrt{n} \cdot \sigma})\end{aligned}$$

也有可能拿两个概率进行比较来解出 X 的题型.

4.7.2 大数定理

不要求, 不考

当 n 充分大时, 样本的均值 $\bar{x} \underset{\text{无限接近}}{\rightarrow} E(x)$

Chapter 5

数理统计

5.1 抽样分布

5.1.1 样本方差 (Sample Variance)

$$S_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

注意和方差的不同, 是 $\frac{1}{n-1}$ 而非 $\frac{1}{n}$

此章用到的均为样本方差以及样本标准差 (样本方差开个根号)

5.1.2 Chi-Square Distribution

n 个随机变量 $X_1, X_2, X_3 \ldots X_n$ 相互独立且服从标准正态分布. 则随机变量 X 的平方和

$$Z = X_1^2 + X_2^2 + X_3^2 \cdots + X_n^2 = \sum_{i=1}^{n} X_i^2$$

服从自由度为 n 的卡方分布, 记作 $Z \sim \chi^2(n)$

5.1.3 Student's t-Distribution

Normal distribution also called Z distribution.
T 分布是标准正态分布的一种变形, 基于样本量的大小而变化.
随机变量 X, Y 相互独立.X 服从标准正态分布, Z 服从自由度为 n 的卡方分布.
即 $X \sim \mathcal{N}(0, 1)$, $Z \sim \chi^2(n)$. 则随机变量

$$t = \frac{X}{\sqrt{\frac{Z}{n}}}$$

服从自由度为 n 的 t 分布

5.1.4 F Distribution

随机变量 U,V 相互独立. U,V 分别服从自由度为 n_1,n_2 的卡方分布, 即 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 则随机变量

$$F = \frac{\frac{U}{n_1}}{\frac{V}{n_2}}$$

服从自由度为 (n_1,n_2) 的 F 分布.

5.2 矩估计和最大似然检验

5.2.1 矩估计

$$E(X) = \bar{x}$$

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

$f(x)$ 由题目给出, 得到 $\hat{\theta} = \bar{x}$

5.2.2 最大似然估计

$$L(\theta) = \prod_{i=1}^{n} f(x_i\ ;\ \theta)$$

$f(x_i\ ;\ \theta)$ 由题目给出.

- 估计量 X
- 估计值 x
- 估计 (大小写均可)

1. 写 $L(\theta)$ 化简 (将 θ 提出来)
2. 写 $\ln[L(\theta)]$ 化简 (取对数, 使得连乘化为连加)
3. $\frac{d\ \ln[L(\theta)]}{d\theta} = 0$ 对 θ 求导
4. 由上式解出 θ, 得到 $\hat{\theta} = \theta$

5.3 区间估计和假设检验

5.3.1 区间估计

一般显著性水平 $\alpha = 0.05$ 即 "显著".

Table 5.1: 区间估计表

待估参数		置信区间	单侧置信限
μ	σ^2 已知	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$	$\bar{\mu} = \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha}$ $\underline{\mu} = \bar{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha}$
	σ^2 未知	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot t_{\frac{\alpha}{2}} \cdot (n-1)$	$\bar{\mu} = \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot t_{\alpha} \cdot (n-1)$ $\underline{\mu} = \bar{X} - \frac{\sigma}{\sqrt{n}} \cdot t_{\alpha} \cdot (n-1)$
σ^2	μ 未知	$(\frac{(n-1)\cdot S^2}{\chi^2_{\frac{\alpha}{2}}\cdot(n-1)}, \frac{(n-1)\cdot S^2}{\chi^2_{1-\frac{\alpha}{2}}\cdot(n-1)})$	$\bar{\sigma}^2 = \frac{(n-1)\cdot S^2}{\chi^2_{1-\alpha}\cdot(n-1)}$ $\underline{\sigma^2} = \frac{(n-1)\cdot S^2}{\chi^2_{\alpha}\cdot(n-1)}$

5.3.2 Confidence Interval

The mean for the population lies between (a, b), instead of equaling.
Its centre is still sample mean μ.

$$\begin{aligned} P\{-\mu_{\frac{\alpha}{2}} < X < \mu_{\frac{\alpha}{2}}\} &= 1-\alpha \\ \Phi(z_\alpha) &= 1-\alpha \end{aligned}$$

Standard Error

$$\begin{aligned} \text{Standard Error} &= \sqrt{\frac{\sigma^2}{n}} \\ &= \sqrt{\frac{\text{Deviation}}{n}} \\ &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{\text{Stardard Deviation}}{\sqrt{n}} \end{aligned}$$

- Variation σ^2
- Sample Size n

Confidence Interval (deviation known)

$$\text{Confidence Interval} = \underbrace{\bar{X}}_{\text{Sample Mean}} \pm \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{Standard Error}} \cdot \underbrace{z_\alpha}_{\text{Z Score}}$$

5.3.3 What is Z score

What is Critical Value?

$$\begin{aligned} z_{0.025} &= 1.96 \\ \Phi(1.96) &= 1-0.025 = 0.975 \end{aligned}$$

大概 Z score, T score, Chi-square score 题目都会给的.
Normal distribution also called Z distribution.