# 习题 1

## Question 1.

盒子里有 $5$ 个红球， $4$ 个白球和 $3$ 个蓝球，从盒子里随机选一球，记下颜色，然后放回盒子里。求挑选 $6$ 次后出现 $3$ 红、 $2$ 白、 $1$ 蓝的概率。

### Answer 1.

这里使用多项式分布就可以了，根据其分布公式，有

$$P(\vec{x};n,\vec{p}) = n!\prod \frac{p_i^{x_i}}{x_i!} = 6! \times \frac{(5/12)^3}{3!} \times \frac{(4/12)^2}{2!} \times \frac{(3/12)}{1!} = \frac{625}{5184}. \tag{1}$$

## Question 2.

已知某放射源产生的射线中包含 A 、 B 、 C 三种粒子，占比分别为 $1/2$ 、 $1/6$ 、 $1/3$ 。在实验的粒子鉴别过程中， A 粒子被误判为其他粒子的概率为 $10\%$ ， B 、 C 粒子被误判为 A 粒子的概率分别为 $12\%$ 和 $21\%$ 。请计算出被鉴别为 A 粒子的事例来自真实的 A 粒子的概率。

### Answer 2.

设产生的射线中粒子数为 $n$ ，于是真粒子 A 数为 $0.5n$ 、真粒子 B 数为 $\frac{1}{6}n$ ；真粒子 C 数为 $\frac{1}{3}n$ ，被误判的粒子 A 数为 $0.05n$ ，误判为粒子 A 的数目为$0.12 \times \frac{1}{6}n + 0.21 \times \frac{1}{3}n = 0.09n$。

于是总粒子 A 数为

$$n'_A = 0.5n - 0.05n + 0.09n = 0.54n. \tag{2}$$

于是 A 粒子的事例来自真实的 A 粒子的概率为

$$P_A = \frac{0.45n}{0.54n} = \frac{45}{54} = \frac{5}{6}. \tag{3}$$

> **Tip**
>
> 这里说是也可以通过贝叶斯定理计算。
>
> 记 $D_A$ 是被统计为粒子 A 的事件，于是有
>
> $$P(D_A) = P(D_A \mid A)P(A) + P(D_A \mid B)P(B) + P(D_A \mid C)P(C) \tag{4}$$
>
> 于是根据贝叶斯公式，可以得到
>
> $$P(A \mid D_A) = \frac{P(D_A \mid A)P(A)}{P(D_A)} \tag{5}$$
>
> $P(D_A \mid A)$ 是粒子 A 被判定为粒子 A 的概率，这是先验概率，也就是 $P(D_A \mid A) = 0.9$ ； $P(A \mid D_A)$ 是被判定为粒子 A 中真实为粒子 A 的概率。
>
> 这样子不断计算就可以得到最终的结果。

# Question 3.

设随机变量 $X$ , $Y$ 的 pdf 为 $f(x,y)=A\mathrm{e}^{-ax^2+bxy-cy^2}$, $-\infty<x<\infty$, $-\infty<y<\infty$ 问在什么条件下 $X$ 与 $Y$ 相互独立?

## Answer 3

两个联合随机变量 $\{X,Y\}$ 相互独立的条件为

$$f(x,y)=f_X(x)\cdot f_Y(y). \tag{6}$$

考虑通过联合概率密度函数获得边沿概率密度，有

$$\begin{aligned} f_X(x)&=\int_y f(x,y)\mathrm{d}y=A\mathrm{e}^{-ax^2}\int_y \mathrm{e}^{-cy^2+bxy}\mathrm{d}y=A\mathrm{e}^{-ax^2}\sqrt{\frac{\pi}{c}}\mathrm{e}^{\frac{b^2x^2}{4c}},\\ f_Y(y)&=\int_x f(x,y)\mathrm{d}x=A\mathrm{e}^{-cy^2}\int_x \mathrm{e}^{-ax^2+bxy}\mathrm{d}x=A\mathrm{e}^{-cy^2}\sqrt{\frac{\pi}{a}}\mathrm{e}^{\frac{b^2x^2}{4a}}. \end{aligned} \tag{7}$$

考虑我们实际上要求相互独立的条件，所以有如下方程

$$\begin{aligned} &A^2\pi\mathrm{e}^{-ax^2}\mathrm{e}^{\frac{b^2x^2}{4c}}\mathrm{e}^{-cy^2}\mathrm{e}^{\frac{b^2x^2}{4a}}\sqrt{\frac{1}{ac}}=A\mathrm{e}^{-ax^2+bxy-cy^2},\\ \Rightarrow\ &A\pi\mathrm{e}^{-ax^2}\mathrm{e}^{\frac{b^2x^2}{4c}}\mathrm{e}^{-cy^2}\mathrm{e}^{\frac{b^2x^2}{4a}}\sqrt{\frac{1}{ac}}=\mathrm{e}^{-ax^2+bxy-cy^2},\\ \Rightarrow\ &\begin{cases} b=0\\ A=\frac{\sqrt{ac}}{\pi}\end{cases}. \end{aligned} \tag{8}$$

这个就是最终答案。

注：在上面的计算中运用了高斯积分的结果

$$\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}ax^2+bx}\mathrm{d}x=\sqrt{\frac{2\pi}{a}}\mathrm{e}^{\frac{b^2}{2a}}. \tag{9}$$

> **Tip**
>
> 条件概率密度等于边沿概率密度也是独立性的判据。

# Question 4.

随机变量 $X$ 的概率密度函数为:

$$f(x)=\begin{cases} x^2/18 & ,-3<x<3\\ 0 & ,\text{other}\end{cases} \tag{10}$$

求 $Y=(X+1)^2$ 的概率密度函数。

## Answer 4.

可以得到

$$X=\pm\sqrt{Y}-1. \tag{11}$$

显然 $Y$ 不是 $X$ 的双射，也就是说应该有

$$
\begin{aligned}
g(y) &= f(\sqrt{y}-1)\left|\frac{\mathrm{d}(\sqrt{y}-1)}{\mathrm{d}y}\right| + f(-\sqrt{y}-1)\left|\frac{\mathrm{d}(-\sqrt{y}-1)}{\mathrm{d}y}\right|, \\
&= \frac{(\sqrt{y}-1)^2}{18}\cdot\frac{1}{2\sqrt{y}} + \frac{(-\sqrt{y}-1)^2}{18}\cdot\frac{1}{2\sqrt{y}}.
\end{aligned} \tag{12}
$$

现在考虑定义域，由于 $x \in (-3,3)$ ，$y \in (0,16)$ 且在 $y \in (0,4)$ 时奇异，最终结果为

$$
g(y) = \begin{cases} \frac{y+1}{18\sqrt{y}}, & y \in (0,4) \\ \frac{(\sqrt{y}-1)^2}{18}\cdot\frac{1}{2\sqrt{y}}, & y \in (4,16) \\ 0, & \text{others} \end{cases} \tag{13}
$$

# Question 5.

已知 $X,\ Y$ 为 $(0,1)$ 上均匀分布的随机变量且相互独立。二维随机变量 $U,\ V$ 和 $X,\ Y$ 之间存在变换关系：

$$
U = \cos(2\pi x)\sqrt{-2\ln y}, \quad V = \sin(2\pi x)\sqrt{-2\ln y}. \tag{14}
$$

证明 $U,\ V \sim N(0,1)$ 。

## Answer 5.

此时有

$$
\begin{aligned}
f(x) &= \begin{cases} 1 & ,\ x \in (0,1) \\ 0 & ,\text{others} \end{cases}, \\
f(y) &= \begin{cases} 1 & ,\ y \in (0,1) \\ 0 & ,\text{others} \end{cases}.
\end{aligned} \tag{15}
$$

由于其相互独立，会有

$$
f(x,y) = \begin{cases} 1 & ,\ x,y \in (0,1) \\ 0 & ,\text{others} \end{cases} \tag{16}
$$

不难计算出反函数为

$$
x = \frac{1}{2\pi}\arctan\frac{v}{u}, \quad y = \mathrm{e}^{-\frac{u^2+v^2}{2}}. \tag{17}
$$

Jacobi 行列式为

$$
J\left(\frac{x,y}{u,v}\right) = \begin{vmatrix} -\frac{v}{2\pi(u^2+v^2)} & -u\mathrm{e}^{-\frac{u^2+v^2}{2}} \\ \frac{u}{2\pi(u^2+v^2)} & -v\mathrm{e}^{-\frac{u^2+v^2}{2}} \end{vmatrix} = \frac{1}{2\pi}\mathrm{e}^{-\frac{u^2+v^2}{2}}. \tag{18}
$$

于是有

$$
g(u,v) = f\left[X(u,v), Y(u,v)\right]\left|J\left(\frac{x,y}{u,v}\right)\right| = \frac{1}{2\pi}\mathrm{e}^{-\frac{u^2+v^2}{2}} \quad ,\ u,v \in (-\infty,\infty) \tag{19}
$$

进而可以求出边沿概率密度

$$
\begin{aligned}
f(u)=\int_v g(u,v)\mathrm{d}v=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{u^2}{2}}\sim N(0,1),\\
f(v)=\int_u g(u,v)\mathrm{d}u=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{v^2}{2}}\sim N(0,1),
\end{aligned}
\tag{20}
$$

证毕。

注：以上计算过程运用了高斯积分公式

$$
\int_{-\infty}^{\infty}\mathrm{e}^{-x^2}\mathrm{d}x=\sqrt{\pi}.
\tag{21}
$$

# Question 6.

随机变量 $X$, $Y$ 都服从 $N(\mu,\sigma^2)$ 且相互独立，令 $U=aX+bY$, $W=aXbY$ ，求 $U$ 与 $W$ 的相关系数 $\rho_{UW}$ 。

## Answer 6.

首先计算协方差，即

$$
\begin{aligned}
cov(u,w)=cov\left(ax+by,abxy\right)=cov\left(ax,abxy\right)+cov\left(by,abxy\right)\\
=a^2b\cdot cov\left(x,xy\right)+ab^2\cdot cov\left(y,xy\right)\\
=a^2b\cdot\left[E(x^2y)-E(x)E(xy)\right]+ab^2\cdot\left[E(xy^2)-E(y)E(xy)\right]\\
=a^2b\cdot\left[\mu E(x^2)-\mu^3\right]+ab^2\cdot\left[\mu E(y^2)-\mu^3\right]\\
=a^2b\mu\cdot\left[E(x^2)-\mu^2\right]+ab^2\mu\cdot\left[E(y^2)-\mu^2\right]\\
=a^2b\mu\sigma^2+ab^2\mu\sigma^2.
\end{aligned}
\tag{22}
$$

现在计算方差

$$
\begin{aligned}
V(u)=V(ax+by)=a^2V(x)+b^2V(y)=a^2\sigma^2+b^2\sigma^2,\\
V(w)=V(abxy)=a^2b^2(\sigma^4+2\mu^2\sigma^2).
\end{aligned}
\tag{23}
$$

于是协方差为

$$
\rho_{uw}=\frac{cov(u,w)}{\sqrt{V(u)V(w)}}=\frac{a\mu+b\mu}{\sqrt{(a^2+b^2)(\sigma^2+2\mu^2)}}.
\tag{24}
$$

> ⚠ **Warning**
>
> 看似完美无缺，但在最后一步需要注意 $\sqrt{a^2b^2}=|ab|$ ！
>
> $$
> \begin{aligned}
> \rho_{uw}=\frac{cov(u,w)}{\sqrt{V(u)V(w)}}=\frac{ab(a+b)\mu\sigma^2}{\sqrt{(a^2\sigma^2+b^2\sigma^2)[a^2b^2\sigma^2(\sigma^2+2\mu^2)]}}=\frac{ab(a+b)\mu\sigma^2}{|ab|\sigma^2\sqrt{(a^2+b^2)(\sigma^2+2\mu^2)}}\\
> =\pm\frac{a\mu+b\mu}{\sqrt{(a^2+b^2)(\sigma^2+2\mu^2)}}.
> \end{aligned}
> \tag{25}
> $$

# Question 7.

离散型随机变量 $X$, $Y$ 的联合概率分布为：

| $X/Y$ | $1$ | $2$ | $3$ |
|---|---|---|---|
| $1$ | $1/6$ | $1/9$ | $1/18$ |
| $2$ | $a$ | $b$ | $1/9$ |

已知 $X$, $Y$ 相互独立，求 $a$ 和 $b$ 的值，并给出 $Y$ 的边沿累积分布。

# Answer 7.

有以下方程

$$\begin{cases} p_x(1)p_y(1) = \frac{1}{6} \\ p_x(1)p_y(2) = \frac{1}{9} \\ p_x(1)p_y(3) = \frac{1}{18} \\ p_x(2)p_y(1) = a \\ p_x(2)p_y(2) = b \\ p_x(2)p_y(3) = \frac{1}{9} \\ p_x(1) + p_x(2) + p_x(3) = p_y(1) + p_y(2) + p_y(3) = 1, \end{cases} \tag{26}$$

观察，假设

$$p_x(1) = \frac{1}{3}, \quad p_x(2) = \frac{2}{3}. \tag{27}$$

可以发现此时应有

$$p_y(1) = \frac{1}{2}, \quad p_y(2) = \frac{1}{3}, \quad p_y(3) = \frac{1}{6}. \tag{28}$$

wow，正好发现了解，于是可以计算出

$$a = \frac{1}{3}, \quad b = \frac{2}{9}. \tag{29}$$

$Y$ 的边沿累积分布为

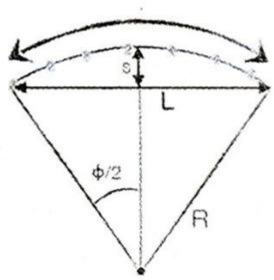
$$\begin{aligned} F_Y(1) &= \frac{1}{2}, \\ F_Y(2) &= \frac{5}{6}, \\ F_Y(3) &= 1. \end{aligned} \tag{30}$$

> ⚠ **Warning**
>
> 注意！这里不能写成上面的样子，因为边沿累积分布毕竟是一个函数，应该写成

> $$F_Y(y) = \begin{cases} 0, & y < 1 \\ \frac{1}{2}, & 1 \leq y < 2 \\ \frac{5}{6}, & 2 \leq y < 3 \\ 1, & y \geq 3 \end{cases} \tag{31}$$

# Question 8.

磁谱仪可以通过测量带电粒子在磁场中的偏转径迹重建粒子的横动量 $p_T = 0.3BR$ 。已知探测器的长度为 $L$ ，通过测量粒子径迹可得偏转距离为 $S$ 误差为 $\sigma_S$ ，如图所示。请计算出粒子动量测量误差的表达式。（ $\sin\frac{\Phi}{2} \approx \frac{\Phi}{2}$，$\cos\frac{\Phi}{2} \approx 1 - \frac{\Phi^2}{8}$ ）




## Answer 8.

$$R - R\cos\frac{\phi}{2} = S \Rightarrow R = \frac{S}{1-\cos\frac{\phi}{2}}. \tag{32}$$

于是有

$$\begin{cases} R\sin\frac{\phi}{2} = \frac{L}{2} \\ R = \frac{S}{1-\cos\frac{\phi}{2}} \end{cases} \Rightarrow \sin\frac{\phi}{2} = L\frac{1-\cos\frac{\phi}{2}}{2S} \to \phi = L\frac{\phi^2}{8S} \Rightarrow \phi = \frac{8S}{L}. \tag{33}$$

继续有

$$R = \frac{L^2}{8S} \Rightarrow p_T = 0.3B\frac{L^2}{8S}. \tag{34}$$

根据不确定度传递公式，应有

$$\begin{gathered} \sigma_{p_T}^2 = (\partial_S p_T)^2 \sigma_s^2 = \left(0.3B\frac{L^2}{8S^2}\right)^2 \sigma_S^2, \\ \sigma_{p_T} = 0.3B\frac{L^2}{8S^2}\sigma_S. \end{gathered} \tag{35}$$

# Question 9.

某实验的内径迹探系统包含像素、硅微条和穿越辐射三种子探测器，三种子探测可以分别独立的测量粒子的横动量，现已知三个子探测器对某个粒子横动量的测量结果分别为：

| detector | pT (GeV) |
| --- | --- |
| pixel detector | $20 \pm 2$ |
| semiconductor tracker | $21 \pm 1$ |

| detector | pT (GeV) |
|---|---|
| transition radiation tracker | $22\pm 4$ |

请综合三个子探测器的数据给出该粒子横动量的测量结果与误差。

## Answer 9.

这里的考点是**联合结果**。

由于是独立测量，应该有

$$\begin{aligned}\mu&=\frac{\sum_i w_i p_T}{\sum_i w_i}=20.9\ \mathrm{GeV},\\ \sigma&=\frac{1}{\sqrt{\sum_i w_i}}=0.9\ \mathrm{GeV}.\end{aligned}\tag{36}$$

所以结果为 $20.9\pm 0.9\ \mathrm{GeV}$ 。

> ⚠ **Warning**
>
> 这里数字的保留也需要特别注意，误差要跟测量精度一致，题目中是个位，联合结果也要保留到个位。

# Question 10.

一条粒子径迹一般需用多层探测器来探测。设一粒子径迹穿过 $n=3$ 层探测器，每层包含 $m=3$ 个子层，当每一层中至少有 $k=2$ 个子层探测器有击中信号，就认为探测到了该粒子径迹。假定各子探测器对于该粒子的探测效率及其误差均为 $\varepsilon\pm\sigma$ ，求该粒子探测系统的粒子径迹探测效率及其误差 $\varepsilon_t\pm\sigma_t$ 。

## Answer 10.

每一层探测器记录的组合有 $4$ 种，也就是**三个子层任意两个有信号**或者**三个子层都有信号**。于是，探测到的概率为

$$p'=C_3^2(1-\varepsilon)\varepsilon^2+C_3^3\varepsilon^3=3(1-\varepsilon)\varepsilon^2+\varepsilon^3=3\varepsilon^2-2\varepsilon^3.\tag{37}$$

由于是每一层都要遵循如上要求，也就是说探测器系统的总效率应该为

$$\varepsilon_t=p'^3=\left(3\varepsilon^2-2\varepsilon^3\right)^3.\tag{38}$$

系统误差应该为

$$\begin{aligned}\sigma_t^2&=(\partial_\varepsilon\varepsilon_t)^2\sigma^2=\left[18\left(3\varepsilon^2-2\varepsilon^3\right)^2\left(\varepsilon-\varepsilon^2\right)\right]^2\sigma^2,\\ \Rightarrow\sigma_t&=18\left(3\varepsilon^2-2\varepsilon^3\right)^2\left(\varepsilon-\varepsilon^2\right)\sigma.\end{aligned}\tag{39}$$

> ⚠ **Warning**
>
> 最后的结果还是写成这个形式比较好
>
> $$\sigma_t=18\left|\left(3\varepsilon^2-2\varepsilon^3\right)^2\left(\varepsilon-\varepsilon^2\right)\right|\sigma.\tag{40}$$
>
> 这里判错的原因也挺特别的。因为每一层和子层的探测器都是相对独立的，所以应该明确地写成 $\varepsilon_{ij},\ i,j=1,2,3$ 代表第 $i$ 层中 $j$ 子探测器的探测效率。我觉得这个完全属于扯淡。

> 但是**加绝对值保证为正**还是挺重要的。

# 习题 2

## Question 1.

某商店负责某小区 $1000$ 人商品供应，其中某种商品一周内每人需购买一件的概率为 $0.6$ 。假定各人购买商品的行为相互独立，且每人在一周内最多购买一件该商品。问商店一周内应预备多少件这种商品，才能以 $99.7\%$ 以上的概率保证商品不会脱销？（提示：应用极限定理解题）

### Answer 1.

设每周购买人数为 $n$ ，显然这一随机变量服从二项式分布，即

$$P(n)=C_{1000}^{n}\times 0.6^{n}\times 0.4^{1000-n}. \tag{41}$$

其均值以及标准差为

$$\mu=600,\ \sigma=4\sqrt{15}. \tag{42}$$

根据题意，应该有如下表达式成立

$$P(0<n<n_0)=99.7\%. \tag{43}$$

应用德莫佛-拉普拉斯近似，应有

$$a=\frac{0-600}{4\sqrt{15}},\ b=\frac{n_0-600}{4\sqrt{15}}\Rightarrow\Phi(b)-\Phi(a)=0.997 \tag{44}$$

查表可得

$$n_0=643. \tag{45}$$

所以应该预备 $643$ 件商品。

## Question 2.

设容量 $n=10$ 的一组子样观测值为 $(1,\ 4,\ 3,\ 3,\ 4,\ 5,\ 6,\ 4,\ 8)$ ，求子样平均值，子样方差值和子样分布函数的具体形式。

### Answer 2.

子样均值为

$$\mu=\frac{1+2+4+3+3+4+5+6+4+8}{10}=4. \tag{46}$$

子样方差为

$$\sigma^2=\frac{1}{9}\sum_i(x_i-\mu)^2=4. \tag{47}$$

我们将子样按观测数值从小到大排列，认为

$$x_1=1,\ x_2=2,\ x_3=3,\ x_4=4,\ x_5=5,\ x_6=6,\ x_7=8. \tag{48}$$

子样分布函数为

$$F_{10}(x)=\begin{cases}0, & x<x_1\\ 1/10, & x_1\le x<x_2\\ 2/10, & x_2\le x<x_3\\ 4/10, & x_3\le x<x_4\\ 7/10, & x_4\le x<x_5\\ 8/10, & x_5\le x<x_6\\ 9/10, & x_6\le x<x_7\\ 1, & x\ge x_7\end{cases}. \tag{49}$$

# Question 3.

设 $X_1,\cdots,X_{n+m}$ 是总体 $N(0,\sigma^2)$ 的容量 $n+m$ 的子样，求下列统计量服从的概率分布：

(a) $Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2$ (b) $Z=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}}$ (c) $U=\frac{m}{n}\frac{\sum_{i=1}^{n}X_i^2}{\sum_{i=n+1}^{n+m}X_i^2}$

## Answer 3.

- $Y\sim\chi^2(n+m)$；
- $Z=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}}=\frac{\bar{X}\sqrt{n}}{S_m}\sim t(m)$；
- $U=\frac{m}{n}\frac{\sum_{i=1}^{n}X_i^2}{\sum_{i=n+1}^{n+m}X_i^2}=\frac{S_n^2}{S_m^2}\sim F(n,m)$。

> ⚠ **Warning**
>
> 关于 (b) 给出一个详细解释
>
> 已知 $X_i\sim N(0,\sigma^2)$，那么 $X_i/\sigma\sim N(0,1)$，于是有
>
> $$Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2\sim\chi^2(n+m) \tag{50}$$
>
> 以上关系只有 $\sim N(0,1)$ 的变量才成立，不是任意正态分布均成立，所以 $\sigma^2$ 不可忽略。
>
> $$Z=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}} \tag{51}$$
>
> 由于 $\sum_{i=1}^{n}X_i/\sqrt{n}\sigma$ 显然服从标准正态分布，而 $\sum_{i=n+1}^{n+m}X_i^2/\sigma^2\sim\chi^2(m)$，所以根据一个 $t$ 分布的知识点指出
>
> $$\frac{\sqrt{m}\sum_{i=1}^{n}X_i/\sigma}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2/\sigma^2}}=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}}=\sim t(m) \tag{52}$$

# Question 4.

设 $X_1,\cdots,X_n$ 是总体 $N(\mu,\sigma^2)$ 的子样，$\bar{X}$ 和 $S_n^2$ 为**子样平均**和**子样方差**。又 $X_{n+1}\sim N(\mu,\sigma^2)$ 且独立于 $X_1,\cdots,X_n$，求统计量 $Z=\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\bar{X}}{S_n}$ 的概率分布。

## Answer 4.

在这里考虑 $\sigma$ 未知但相等的两个子样的统计量：

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2). \tag{53}$$

现在考虑 $n_1=1,\ n_2=n,\ \mu_1=\mu_2=0,\ S_2^2=S_n^2$，带入结果可以得到

$$\frac{X_{n+1}-\bar{X}}{S_n\sqrt{\frac{n+1}{n}}} = Z \sim t(n-1). \tag{54}$$

解毕。

> **Note**
>
> 注意，以上论述的前提是两个子样的 $\sigma$ 未知但相同。本题显然满足这个前提条件。

# Question 5.

正态总体 $X\sim N(\mu,\sigma^2)$ 的方差为已知，要使均值 $\mu$ 的置信水平为 $1-\alpha$ 的中心置信区间的长度不大于 $d$，问子样容量 $n$ 需多大。

## Answer 5.

定义一个新变量

$$Y=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1). \tag{55}$$

也就是有

$$\begin{aligned} &P(-z_{\alpha/2}\le Y\le z_{\alpha/2}) = \int_{-z_{\alpha/2}}^{z_{\alpha/2}} N(y;0,1)\mathrm{d}y = 1-\alpha, \\ \Rightarrow\ &P(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le \mu\le \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha. \end{aligned} \tag{56}$$

题目要求

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le d\ \Rightarrow\ 4z_{\alpha/2}^2\frac{\sigma^2}{d^2}\le n. \tag{57}$$

解毕。

> **Tip**
>
> 想一想在 $\sigma$ 已知的情况下对均值做参数估计的统计量，也只有答案给出的这一个了。

# Question 6.

设总体 $X \sim N(80, 20^2)$，从总体 $X$ 中抽取一个容量 100 的子样，问子样均值与总体均值之差的绝对值大于 5 的概率是多少?

## Answer 6

在这里我们的做法和上面一致，也就是同样构造

$$Y = \frac{\bar{X} - 80}{2}. \tag{58}$$

由于题目要求子样均值与总体均值只差的绝对值大于 5 ，也就是

$$\left|\bar{X} - 80\right| \geq 5 \quad \Rightarrow \quad |Y| \geq \frac{5}{2}. \tag{59}$$

所以有

$$P(|Y| \geq \frac{5}{2}) = 1 - \int_{-5/2}^{5/2} N(y; 0, 1)\mathrm{d}y = 0.0124. \tag{60}$$

解毕。

> **Tip**
>
> 在这里其实是假设检验，并不是参数估计，因为所有参数都已知了。

# Question 7.

设某种油漆的 5 个样品的干燥时间分别为 $10.5,\ 11.0,\ 11.2,\ 12.5,\ 12.8$ 小时。若干燥时间服从正态分布，在置信水平为 $95\%$ 的情况下，求**平均干燥时间**的单侧置信区间上限。

## Answer 7.

子样均值以及方差可以计算为

$$\bar{t} = 11.6,\ S_t^2 = 0.995. \tag{61}$$

由于干燥时间服从高斯分布，所以这样做估计之后应该认为

$$Y = \sqrt{5}\frac{\bar{t} - \mu}{S_t} \sim t(4). \tag{62}$$

所以应该有

$$P\left[Y \leq t_{0.05}(4)\right] = \int_{-\infty}^{t_{0.05}(4)} t(y; 4)\mathrm{d}y = 0.95 = P\left[\mu \geq \bar{t} - t_{0.05}(4)\frac{S}{\sqrt{5}}\right]. \tag{63}$$

但是这里求的变成区间下限了。利用 $t$ 分布的对称性 $t_{0.95}(4) = -t_{0.05}(4)$ ，代入得

$$P\left[Y \geq t_{0.95}(4)\right] = \int_{t_{0.95}(4)}^{\infty} t(y; 4)\mathrm{d}y = 0.95 = P\left[\mu \leq \bar{t} - t_{0.95}(4)\frac{S}{\sqrt{5}}\right] = P\left[\mu \leq \bar{t} + t_{0.05}(4)\frac{S}{\sqrt{5}}\right]. \tag{64}$$

可以解出

$$t_{0.05}(4) = 2.132 \ \Rightarrow \ \mu < 12.55. \tag{65}$$

解毕。

# Question 8.

设总体 $X \sim N(\mu, \sigma^2)$，$\mu, \sigma^2$ 均为未知参数，$X_1, \cdots, X_n$ 为其子样，求:

- $\mu$ 的置信水平 $1-\alpha$ 的**中心置信区间**
- 该区间的长度 $d$
- $d^2$ 的期望值

## Answer 8.

记子样均值与子样方差为 $\bar{X},\ S^2$，构建

$$Y = \sqrt{n}\frac{\bar{X}-\mu}{S} \sim t(n-1). \tag{66}$$

由于 $t$ 分布是关于 $0$ 的中心对称分布，于是有

$$P(-t_{\alpha/2} \leq Y \leq t_{\alpha/2}) = 1-\alpha = P(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}). \tag{67}$$

所以置信度为 $1-\alpha$ 的置信区间就是 $\left[\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$。

区间的长度可以表示为

$$d = 2\frac{S}{\sqrt{n}}t_{\alpha/2}. \tag{68}$$

于是有

$$d^2 = 4\frac{S^2}{n}t_{\alpha/2}^2 \quad\Rightarrow\quad E(d^2) = \frac{4}{n}t_{\alpha/2}^2 E(S^2) = \frac{4}{n}t_{\alpha/2}^2\sigma^2. \tag{69}$$

# Question 9.

某束带电粒子束流通过磁场，利用粒子的偏移径迹可以测量粒子的动量，测得的 $10$ 组数值为:

$$18.87,\ 19.55,\ 19.32,\ 18.70,\ 19.41,\ 19.37,\ 18.84,\ 19.40,\ 18.78,\ 18.76 \tag{70}$$

若束流动量服从正态分布，求方差 $\sigma^2$ 置信水平为 $95\%$ 的置信区间。

## Answer 9.

子样均值与方差分别为

$$\bar{X} = 19.10,\ S^2 = 0.1120. \tag{71}$$

构建如下随机变量

$$\sum_{i=1}^{10}\frac{(X-\bar{X})^2}{\sigma^2} = \frac{9S^2}{\sigma^2} \sim \chi^2(9). \tag{72}$$

所以有

$$\gamma = \int_{\chi^2_{0.975}(9)}^{\chi^2_{0.025}(9)} \chi^2(u,9)\mathrm{d}u = 0.95 \quad \Rightarrow \quad \chi^2_{0.975}(9) = 2.700 \le \frac{9S^2}{\sigma^2} \le \chi^2_{0.025}(9) = 19.023. \tag{73}$$

注意，在这里没有要求中心置信区间，但是为了计算方便还是取中心置信区间，所以取上 $0.975$ 和上 $0.025$ 分位数。于是有

$$\frac{9S^2}{\chi^2_{0.025}(9)} \le \sigma^2 \le \frac{9S^2}{\chi^2_{0.975}(9)} \quad \Rightarrow \quad 0.0530 \le \sigma^2 \le 0.3733 \tag{74}$$

解毕。这里额外说明一点，其它满足题意的置信区间也是存在的，但是这里给出的是中心置信区间。

# Question 10.

设 $X_1,\cdots,X_n$ 是正态总体 $N(\mu_1,\sigma_1^2)$ 的子样观测值，$Y_1,\cdots,Y_m$ 是正态总体 $N(\mu_2,\sigma_2^2)$ 的子样观测值，求均值 $\mu=\mu_1-\mu_2$ 的置信水平为 $\gamma=1-\alpha$ 的中心置信区间。

- $\sigma_1^2$，$\sigma_2^2$ 已知;
- $\sigma_1^2$，$\sigma_2^2$ 未知，$n$，$m$ 充分大。

## Answer 10.

如果 $\sigma_1^2,\sigma_2^2$ 已知，则有

$$E(\bar{X}-\bar{Y}) = \mu,\ V(\bar{X}-\bar{Y}) = \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m} \Rightarrow \bar{X}-\bar{Y} \sim N(\mu, \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}). \tag{75}$$

此时服从正态分布，所以很显然可以定义一个随机变量

$$\frac{(\bar{X}-\bar{Y})-\mu}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1). \tag{76}$$

所以，这里会有

$$\begin{gathered} P\left(-z_{\alpha/2} \le \frac{(\bar{X}-\bar{Y})-\mu}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \le z_{\alpha/2}\right) = \gamma = 1-\alpha \\ \Rightarrow P\left[-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} + (\bar{X}-\bar{Y}) \le \mu \le z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} + (\bar{X}-\bar{Y})\right] = \gamma = 1-\alpha. \end{gathered} \tag{77}$$

此时的中心置信区间就是

$$\left[-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} + (\bar{X}-\bar{Y}),\quad z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} + (\bar{X}-\bar{Y})\right]. \tag{78}$$

$\sigma_1^2,\sigma_2^2$ 未知，$n,m$ 充分大。此时可以**用样本方差近似为总体方差**，结果为

$$\left[-z_{\alpha/2}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}} + (\bar{X}-\bar{Y}),\quad z_{\alpha/2}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}} + (\bar{X}-\bar{Y})\right]. \tag{79}$$

# 期中真题

## Question 1.

（共 5 分）设总体分布为正态分布 $N(\mu, \sigma^2)$，容量为 $10$ 的一组子样观测值为 $(10，30，50，40，40，20，30，20，40，20)$ 。

- 在总体均值已知为 $30$ 的情况下，求总体方差的有效无偏估计值。

  当总体均值 $\mu = 30$ 已知时：

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = 140 \tag{80}$$

- 在总体的均值未知时，求总体方差的无偏估计值。

  样本均值：

$$\bar{x} = 30 \tag{81}$$

  无偏估计：

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1400}{9} \approx 155.556 \tag{82}$$

- 求子样均值的方差值。

$$V(\bar{x}) = \frac{s^2}{n} = 15.556 \tag{83}$$

## Question 2.

（共 5 分）100 个粒子穿过某探测器，记录到 80 个探测器信号。

- 求探测效率及误差。

  探测效率：

$$p = \frac{80}{100} = 0.8 \tag{84}$$

  标准误差：

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.8 \times 0.2}{100}} = 0.04 \tag{85}$$

  因此：

$$p = 0.8 \pm 0.04 \tag{86}$$

- 若要求探测效率的相对误差达到 $0.05$ ，需要对多少个粒子进行测量？

$$\frac{\sigma_p}{p} = 0.05 \tag{87}$$

  代入具体表达式

$$\sqrt{\frac{0.2}{0.8n}} = 0.05 \tag{88}$$

解得：

$$n = 100 \tag{89}$$

# Question 3.

（共 5 分）某医院开展了一个筛查，测试某种罕见疾病。假设该疾病在总体人群中的患病率为 $0.01$（即 $1\%$ 的人群患有此病）。现有一个测试方法，已知其性能如下：

- 若被测试者患病，测试结果为阳性的概率为 $0.98$；
- 若被测试者不患病，测试结果为阴性的概率 $0.99$ 。

假设某人接受了测试，结果为阳性。请计算此人实际患病的概率。

疾病患病率 $0.01$，灵敏度 $0.98$，特异性 $0.99$。

求阳性时患病概率：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{90}$$

计算：

$$P(B) = 0.0098 + 0.0099 = 0.0197 \tag{91}$$

因此结果：

$$P(A|B) \approx 0.497 \tag{92}$$

# Question 4.

（共 5 分）随机变量 $\{X, Y\}$ 的联合 pdf 为 $f(x, y) = cxy,\ 0 < x \leq 4,\ 1 < y \leq 5$，$c$ 是归一化常数。求

- $c$ 的值。

$$\int_0^4 \int_1^5 cxy \mathrm{d}y \mathrm{d}x = 1 \tag{93}$$

计算得：

$$96c = 1 \quad \Rightarrow \quad c = \frac{1}{96} \tag{94}$$

- $P(1 < X \leq 2,\ 2 < Y \leq 3)$ 。

$$P(1 < X \leq 2,\ 2 < Y \leq 3) = \int_{x=1}^2 \int_{y=2}^3 \frac{1}{96} xy \mathrm{d}y \mathrm{d}x = \frac{5}{128} \tag{95}$$

- $\{X, Y\}$ 的边沿分布 $F_X(x)$ 和 $F_Y(y)$ 。

$$f_X(x) = \int_{y=1}^5 \frac{xy}{96} \mathrm{d}y = \frac{x}{8}, \quad 0 < x \leq 4 \quad \Rightarrow \quad F_x = \int f_X(x) \mathrm{d}x = \frac{x^2}{16}. \tag{96}$$

# Question 5.

（共 5 分）W 玻色子传递弱相互作用。因为粒子物理标准模型对 W 玻色子的质量有严格的限制，所以测量它的质量可以检验标准模型。目前最新的五个实验中对 W 玻色子质量测量的结果如下：
$80415 \pm 50$，$80440 \pm 50$，$80376 \pm 20$，$80370 \pm 20$，$80433 \pm 10\ \mathrm{MeV/c^2}$。试给出以上五个实验对 W 玻色子质量测量的**联合结果**。

# Question 6.

（共 5 分）已知 $x = 10.000 \pm 0.120$，$y = 5.000 \pm 0.050$，问

- 若 $x$ 和 $y$ 不相关，求 $a = x + y$ 的数值。
- 在 $x$ 和 $y$ 相关系数多大的情形下 $b = \frac{x}{y}$ 的数值等于 $2.000 \pm 0.014$ 。

> **Important**
>
> 考点就是
>
> $$V(ax + by) = a^2V(x) + b^2V(y) + 2abcov(x, y) \tag{97}$$
>
> 或者说考的就是误差传递公式。

# 习题 3

## Question 1

设 $x_1, \ldots, x_n$ 是几何分布总体的子样观测值，其分布律为 $P(X=x)=p(1-p)^{x-1}$，$x=1,2,\ldots$。求参数 $p$ 的极大似然估计。

### Answer 1

似然函数为

$$L(x_1,\cdots,x_n \mid p)=\prod_i P(X=x_i)=p^n(1-p)^{x_1+\cdots+x_n-n}. \tag{98}$$

求解如下方程

$$\begin{cases}\partial_p L=\partial_p[p^n(1-p)^{x_1+\cdots+x_n-n}]=0,\\ \partial_{pp} L=\partial_{pp}[p^n(1-p)^{x_1+\cdots+x_n-n}]<0.\end{cases} \tag{99}$$

可以得到

$$\hat{p}=\frac{1}{\bar{x}}. \tag{100}$$

## Question 2

设 $x_1, \ldots, x_n$ 是总体 $X$ 的子样，总体 $X$ 的 pdf 为

$$f(x;\theta)=\begin{cases}(\theta+1)x^\theta, & 0<x_i<1\\ 0, & \text{其它}\end{cases} \tag{101}$$

其中 $\theta>-1$ 是未知参数。用极大似然法求参数 $\theta$ 的估计。

### Answer 2

似然函数为

$$L=\prod_i f(x_i;\theta)=\begin{cases}(\theta+1)^n\left(\prod_i x_i\right)^\theta, & 0<x<1\\ 0. & \text{else}\end{cases} \tag{102}$$

求解如下方程

$$\begin{cases}\partial_\theta L=n(\theta+1)^{n-1}\left(\prod_i x_i\right)^\theta+(\theta+1)^n\left(\prod_i x_i\right)^\theta\ln\left(\prod_i x_i\right)=0,\\ \partial_{\theta\theta}L<0.\end{cases} \tag{103}$$

可以得到

$$\hat{\theta}=-\frac{n}{\ln\left(\prod_i x_i\right)}-1. \tag{104}$$

# Question 3

某实验观测到十组 $e^+e^- \to \mu^+\mu^-$ 过程，相应的 $\cos\theta$（$\theta$ 为散射角度）观测值分别为：

$$-0.5,\ -0.25,\ -0.1,\ -0.05,\ 0.01,\ 0.04,\ 0.11,\ 0.14,\ 0.24,\ 0.6. \tag{105}$$

现假设散射角的分布满足 $f(\cos\theta) \propto 1+\lambda\cos\theta$，请利用极大似然法估计参数 $\lambda$。（可近似取 $\frac{1}{\alpha\lambda+1} \approx 1-\alpha\lambda$）

## Answer 3

似然函数为

$$L \propto \prod_i (1+\lambda\cos\theta_i) \quad \Rightarrow \quad \ln L \propto \sum_i \ln(1+\lambda\cos\theta_i). \tag{106}$$

应满足

$$\partial_\lambda \ln L \propto \partial_\lambda \sum_i \ln(1+\lambda\cos\theta_i) = \sum_i \frac{\cos\theta_i}{1+\lambda\cos\theta_i} = \sum_i \cos\theta_i(1-\lambda\cos\theta_i) = 0. \tag{107}$$

最终可以得到

$$\lambda = 0.309. \tag{108}$$

代入 $\partial_{\lambda\lambda}\ln L$，可以发现确实是极大值。

# Question 4

某厂生产的一种电池，其寿命长期以来服从方差 $\sigma^2 = 5000\ [\text{小时}^2]$ 的正态分布。近期生产一批这种电池，从生产的情况来看不能肯定寿命方差是否改变。现随机抽取 26 个电池，测得寿命的样本方差为 $S^2 = 9200\ [\text{小时}^2]$。问根据这一数据能否推断这批电池寿命方差较以往有显著变化（取 $\alpha = 0.02$）。

## Answer 4

此时原假设和备择假设为

$$H_0: \sigma^2 = 5000,\ H_1: \sigma^2 \neq 5000. \tag{109}$$

构造原假设成立下的检验统计量

$$z = \frac{(n-1)S^2}{\sigma^2} = \frac{25S^2}{5000} \sim \chi^2(25). \tag{110}$$

可以求出双侧临界值

$$\int_{-\infty}^{z_d} \chi^2(z;25)\mathrm{d}z = \int_{z_u}^{\infty} \chi^2(z;25)\mathrm{d}z = 0.01. \tag{111}$$

可以得到

$$z_d = 11.524,\ z_u = 44.314. \tag{112}$$

而观测值为 $z_{\text{obs}} = 46 > z_u$，处在拒绝域中，所以认为有显著变化。

# Question 5

按规定，每 $100\ \mathrm{g}$ 的番茄汁罐头中 VC 的含量不得少于 $21\ \mathrm{mg}$ 。现从某厂生产的一批罐头中抽取 $17$ 个，测得 VC 的含量为：

$$16,\ 22,\ 21,\ 20,\ 23,\ 21,\ 19,\ 15,\ 13,\ 23,\ 17,\ 20,\ 29,\ 18,\ 22,\ 16,\ 25\ \mathrm{mg}. \tag{113}$$

已知 VC 的含量服从正态分布，试以 $0.025$ 的检验水平检验该批罐头的 VC 含量是否合格。

# Answer 5

原假设为

$$H_0 : \mu \geq 21. \tag{114}$$

首先计算出方差

$$S^2 = 15.88. \tag{115}$$

检验统计量为

$$z = \frac{\bar{X} - \mu}{S/\sqrt{17}} \sim t(16). \tag{116}$$

显然 $z$ 是 $\mu$ 的单调递增函数，要使得第二类错误最小，应取 $\mu = 21$ 代入考虑。

显然现在应该求下侧临界值 $t_{0.975} = -2.120$ ，观测值为 $z_{\mathrm{obs}} = -1.035 > t_{0.975}$ ，所以合格。

# Question 6

对某个寿命无限长的放射源辐射出的 $\alpha$ 粒子用计数器测 $1$ 分钟计数，共测 $206$ 次，结果如下表：

| **计数/分** $r$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 次数 $n_r$ | 4 | 11 | 31 | 34 | 48 | 26 | 18 | 18 | 10 | 6 |

其中 $n_r$ 是（计数/分）为 $r$ 的出现次数。试用 $\chi^2$ 检验检验计数器 $1$ 分钟计数是否服从泊松分布（$\alpha = 0.05$ ）？

# Answer 6

> **Important**
>
> 如果自由度大于等于 $7$ 的话，是不需要保证每组中的数目大于 $5$ 。

平均值为

$$\mu = 4.19903. \tag{117}$$

原假设为

$$H_0 : r \sim \frac{\mu^r}{r!}\mathrm{e}^{-\mu}. \tag{118}$$

则可以计算出期望概率以及理论频数

| $r$ | $p_{0r}$ | $np_{0r}$ |
|---|---|---|
| 0 | 0.015009 | 3.092 |
| 1 | 0.063031 | 12.984 |
| 2 | 0.132329 | 27.260 |
| 3 | 0.185217 | 38.155 |
| 4 | 0.194433 | 40.053 |
| 5 | 0.163286 | 33.637 |
| 6 | 0.114093 | 23.540 |
| 7 | 0.068548 | 14.121 |
| 8 | 0.035980 | 7.412 |
| 9 | 0.016787 | 3.458 |

（$\chi^2$ 检验）定义统计检验量

$$\chi^2 = \sum_{i=0}^{N} \frac{(n_i - np_{0i})^2}{np_{0i}} \sim \chi^2(10-1-1) = \chi^2(8). \tag{119}$$

上面自由度为 组数 $-\ p_{0i}$ 归一化限制 $-$ 平均值作为原假设未知参数，共计 8 自由度。

可以计算出

$$0.05 = \int_{\chi^2_{0.05}(8)}^{\infty} \chi^2(y,8)\mathrm{d}y \quad \Rightarrow \quad \chi^2_{0.05}(8) = 15.507. \tag{120}$$

而观测值为 $\chi^2_{\text{obs}} = 9.139 < 15.507$，所以认为服从泊松分布。

# Question 7

现对某探测器的探测分辨率进行了模拟，下表展示了某一刚度范围内入射粒子重建结果：

| 区间（重建刚度 - 真实刚度 / GV ） | 中值 | 事例数 |
|---|---|---|
| $-0.20 \sim -0.16$ | $-0.18$ | 25 |
| $-0.16 \sim -0.12$ | $-0.14$ | 52 |
| $-0.12 \sim -0.08$ | $-0.10$ | 96 |
| $-0.08 \sim -0.04$ | $-0.06$ | 163 |
| $-0.04 \sim 0.00$ | $-0.02$ | 215 |
| $0.00 \sim 0.04$ | $0.02$ | 200 |
| $0.04 \sim 0.08$ | $0.06$ | 159 |

| $x$ 区间（重建刚度$^{-1}$ - 真实刚度$^{-1}$ / GV$^{-1}$） | $x_i$ 中值 | $n_i$ 事例数 |
| --- | --- | --- |
| $0.08 \sim 0.12$ | $0.10$ | $92$ |
| $0.12 \sim 0.16$ | $0.14$ | $66$ |
| $0.16 \sim 0.20$ | $0.18$ | $32$ |

1. 假设其满足正态分布，利用极大似然法估计其均值和方差。
2. 在认为**总事例数为泊松变量**且充分大的条件下，试用**似然比检**验方法检验其是否服从（1）中估计的正态分布（$\alpha = 0.01$）。每个区间内的事例数期望值可近似为：

$$v_i = n_t \int_{x_{\min}}^{x_{\max}} f_0(x)\mathrm{d}x \approx n_t f_0(x) \times (x_{\max} - x_{\min}) \tag{121}$$

## Answer 7

1. 似然函数为

$$L(X;\sigma,\mu) = \prod_{i=1}^{10}\left[\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right]^{n_i} \quad \Rightarrow \quad \ln L = 1100\ln\frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^{10} n_i\frac{(x_i-\mu)^2}{2\sigma^2}. \tag{122}$$

求解

$$\begin{cases}\partial_\mu \ln L = 0,\\ \partial_{\sigma\sigma}\ln L < 0.\end{cases} \tag{123}$$

可以得到

$$\mu = \sum_{i=1}^{10}\frac{n_i}{1100}x_i = 0.002, \quad \sigma^2 = \sum_{i=1}^{10}\frac{n_i}{1100}(x_i - \bar{x})^2 = 0.007. \tag{124}$$

2. 则可以计算出理论频数

$$v_i = 1100 \times \int_{x_{\min}}^{x_{\max}} \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\mathrm{d}x \to 1100 \times \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \times 0.04. \tag{125}$$

具体结果如下表

| 区间序号 $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 观测数 $n_i$ | 25 | 52 | 96 | 163 | 215 | 200 | 159 | 92 | 66 | 32 |
| 期望数 $v_i$ | 18.020 | 47.484 | 98.512 | 160.906 | 206.922 | 209.502 | 167.000 | 104.808 | 51.787 | 20.146 |

似然比为

$$\lambda_P = \mathrm{e}^{15}\prod_{i=1}^{10}\left(\frac{\nu_i}{n_i}\right)^{n_i}. \tag{126}$$

检验统计量为

$$\chi_P^2 = -2\ln\lambda_P = -30 - \sum_{i=1}^{10} n_i(\ln v_i - \ln n_i) \sim \chi^2(N-2) = \chi^2(8). \tag{127}$$

这里自由度减去 2 是因为在估计正态分布形式的时候估计了均值和方差。可以计算出

$$0.01=\int_{\chi^2_{0.01}(8)}^{\infty}\chi^2(y,8)\mathrm{d}y \quad\Rightarrow\quad \chi^2_{0.01}(8)=20.090. \tag{128}$$

而观测值为 $\chi^2_{\mathrm{obs}}=15.008$，所以认为服从正态分布。