# 复旦大学计算机科学技术学院

2020-2021 第一学期《概率论与数理统计》期末考试试卷

A 卷

参考答案

课程编号: COMP130006.04 考试形式: 闭卷 考试日期: 2021.1.12

姓名________________ 学号_______________ 专业___________________

提示：请同学们秉持诚实守信宗旨，谨守考试纪律，摒弃考试作弊。学生如有违反学校考试纪律的行为，学校将按《复旦大学学生纪律处分条例》规定予以严肃处理。

| 题号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 总分 |
|---|---|---|---|---|---|---|---|---|
| 分值 | 10 | 10 | 10 | 10 | 20 | 20 | 20 | 100 |
| 得分 | | | | | | | | |

# 1 基础知识

## 1.1 习题，1.2.20

将 3 个球随机地放入 4 个杯子中，试求杯子中球的最大个数 $X$ 的概率分布。

解：只有三种情况，

$$X = \begin{cases} 1 & \text{三个球分别放在三个杯子中} \\ 2 & \text{一个球和另外两个球分别放在两个杯子中} \\ 3 & \text{三个球放在同一个杯子中} \end{cases}$$

由于每个球都是机会均等地放入到 4 个杯子中，用 $a_1, a_2, a_3$ 分别表示三次放球，

$$\begin{aligned}
P(X=1) &= P(a_1\text{任意放}) \cdot P(a_2\text{放入三个空杯}) \cdot P(a_3\text{放入两个空杯}) \\
&= 1 \cdot \frac{3}{4} \cdot \frac{2}{4} \\
&= \frac{3}{8} \\
P(X=3) &= P(a_1\text{任意放}) \cdot P(a_2\text{放入有球杯}) \cdot P(a_3\text{放入有球杯}) \\
&= 1 \cdot \frac{1}{4} \cdot \frac{1}{4} \\
&= \frac{1}{16} \\
P(X=2) &= 1 - P(X=1) - P(X=3) \\
&= 1 - \frac{3}{8} - \frac{1}{16} \\
&= \frac{9}{16}
\end{aligned}$$

或者

$$\begin{aligned}
P(X=2) &= P(a_1\text{任意放}) \, (\cdot (P(a_2\text{放入有球杯}) \cdot P(a_3\text{放入空杯}) + P(a_2\text{放入空杯}) \cdot P(a_3\text{放入有球杯})) \\
&= 1 \cdot \left( \frac{1}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{2}{4} \right) \\
&= \frac{9}{16}
\end{aligned}$$

# 2 随机变量

## 2.1 书本，标准正态分布

服从标准正态分布的随机变量称为标准正态变量，记为 $U$，标准正态分布的概率密度函数 $\phi(u)$ 和分布函数 $\Phi(u)$ 分别为,

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \quad -\infty < u < \infty$$
$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt, \quad -\infty < u < \infty$$

证明:

$$P(|U| < c) = 2\Phi(c) - 1$$

解:

$$\begin{aligned}
P(|U| < c) &= P(-c < U < c) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-\frac{t^2}{2}} dt \\
&= \frac{1}{\sqrt{2\pi}} \left( \int_{-c}^{0} e^{-\frac{t^2}{2}} dt + \int_{0}^{c} e^{-\frac{t^2}{2}} dt \right) \\
&= \frac{1}{\sqrt{2\pi}} \left( \int_{0}^{c} e^{-\frac{t^2}{2}} dt + \int_{0}^{c} e^{-\frac{t^2}{2}} dt \right) \\
&= 2\frac{1}{\sqrt{2\pi}} \int_{0}^{c} e^{-\frac{t^2}{2}} dt \\
&= 2\frac{1}{\sqrt{2\pi}} \left( \int_{-\infty}^{c} e^{-\frac{t^2}{2}} dt - \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt \right) \\
&= 2\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-\frac{t^2}{2}} dt - 2\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt \\
&= 2\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \\
&= 2\Phi(c) - 1
\end{aligned}$$

# 3 多维随机变量

## 3.1 习题，3.2.4

设平面区域 $D$ 由曲线 $y=\frac{1}{x}$ 及直线 $y=0, x=1, x=4$ 所围成，二维随机变量 $(X,Y)$ 在区域 $D$ 上服从均匀分布，试求

1. $X$ 的边际密度函数
2. $X$ 的边际分布函数
3. $X$ 的期望
4. $Y$ 的边际密度函数
5. $Y$ 的边际分布函数
6. $Y$ 的期望

解：先计算区域 $D$ 的面积，

$$S_D=\int_1^4\int_0^{\frac{1}{x}} ydydx=\int_1^4 \ln xdx=\ln 4$$

因此，$(X,Y)$ 的联合密度函数是

$$p(x,y)=\frac{1}{\ln 4}=a, \quad 1<x<4, 0<y<\frac{1}{x}$$

1.

$$p(x)=\int_{-\infty}^{\infty} p(x,y)dy=\int_0^{\frac{1}{x}} ady=\frac{a}{x}, \quad 1<x<4$$

2.

$$P(x)=\int_{-\infty}^{x}\frac{a}{t}dt=\int_1^x\frac{a}{t}dt=a\cdot(\ln x-\ln 1)=a\ln x, \quad 1<x<4$$

3.

$$E(X)=\int_{-\infty}^{\infty} xp(x)dx=\int_1^4 x\cdot\frac{a}{x}dx=3a$$

4.

$$\begin{aligned} p(y)&=\int_{-\infty}^{\infty} p(x,y)dy \\ &=\begin{cases} \int_1^4 p(x,y)dx=3a & 0<y<\frac{1}{4} \\ \int_1^{\frac{1}{y}} p(x,y)dx=a\left(\frac{1}{y}-1\right) & \frac{1}{4}<y<1 \end{cases} \end{aligned}$$

5.

$$
\begin{aligned}
P(y) &= \int_{-\infty}^{y} p_y(t)dy \\
&= \begin{cases} \int_0^y 3adt = 3ay & 0 < y < \frac{1}{4} \\ \int_0^{\frac{1}{4}} 3adt + \int_{\frac{1}{4}}^{y} a\left(\frac{1}{t} - 1\right)dt = \frac{3}{4}a + a\left(\ln y - \ln\frac{1}{4}\right) - a(y - \frac{1}{4}) = a + 1 + a\ln y - ay & \frac{1}{4} < y < 1 \end{cases}
\end{aligned}
$$

6.

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} yp(y)dy \\
&= \int_0^{\frac{1}{4}} 3aydy + \int_{\frac{1}{4}}^{1} a\left(\frac{1}{y} - 1\right)ydy \\
&= \int_0^{\frac{1}{4}} 3aydy + \int_{\frac{1}{4}}^{1} ady - \int_{\frac{1}{4}}^{1} aydy \\
&= 3a \cdot \frac{1}{2}y^2|_0^{\frac{1}{4}} + a\left(1 - \frac{1}{4}\right) - a \cdot \frac{1}{2}y^2|_{\frac{1}{4}}^{1} \\
&= \frac{3}{8}a \\
&= \frac{3}{8\ln 4} \\
&= \frac{3}{16\ln 2}
\end{aligned}
$$

# 4 极限性质

## 4.1 习题，4.4.2

某计算机主机的各个终端的使用相互独立，解答下面的问题，

1. 设有 100 个终端，每个终端有 80% 的时间被使用，计算至少 15 个终端空闲的概率。
   (a) 采用二项分布计算
   (b) 采用正态近似计算
2. 设有 100 个终端，要保证至少 15 个终端空闲的概率不小于 95%，每个终端空闲的时间比例不能小于多少?
3. 设每个终端有 80% 的时间被使用，要保证至少 15 个终端空闲的概率不小于 95%，至少需要多少个终端?
4. 谈谈为什么二项分布要采用正态近似?
5. 二项分布近似于正态分布的条件是什么?
6. 采用正态近似计算时，区间修正的原理是什么?

解:

1. 由此知道每个终端空闲的概率为 $p = 0.2$，
   (a) 采用二项式分布，空闲终端数 $X \sim b(100, 0.2)$，因此
   $$\begin{aligned} P(X \geq 15) &= P(X = 15) + P(X = 16) + \cdots + P(X = 100) \\ &= \sum_{i=15}^{1} 00P(X = i) \\ &= \sum_{i=15}^{100} \binom{100}{i} 0.2^i 0.8^{100-i} \end{aligned}$$
   (b) 采用正态分布近似，$X \sim N(100 * 0.2, 100 * 0.2 * 0.8) = N(20, 16)$，即
   $$\frac{X - 20}{\sqrt{16}} \sim N(0, 1)$$
   因此
   $$\begin{aligned} P(X \geq 15) &= 1 - P(X \leq 14) \\ &\approx 1 - P(X < 14.5) \\ &= 1 - P\left(\frac{X - 20}{\sqrt{16}} < \frac{14.5 - 20}{\sqrt{16}}\right) \\ &= 1 - \Phi\left(\frac{-5.5}{4}\right) \\ &= \Phi\left(\frac{5.5}{4}\right) \end{aligned}$$

2. 设单个终端的空闲概率为 $p$，则利用正态近似，要满足

$$P(X \geq 15) \geq 0.95$$

由于

$$\begin{aligned} P(X \geq 15) &= 1 - P(X \leq 14) \\ &\approx 1 - P\left(\frac{X - 100p}{\sqrt{100p(1-p)}} < \frac{14.5 - 100p}{\sqrt{100p(1-p)}}\right) \\ &= 1 - \Phi\left(\frac{14.5 - 100p}{\sqrt{100p(1-p)}}\right) \end{aligned}$$

得到

$$\begin{aligned} 1 - \Phi\left(\frac{14.5 - 100p}{\sqrt{100p(1-p)}}\right) &\geq 0.95 \\ \Phi\left(\frac{14.5 - 100p}{\sqrt{100p(1-p)}}\right) &\leq 0.05 \\ \frac{14.5 - 100p}{\sqrt{100p(1-p)}} &\leq u_{0.05} \end{aligned}$$

解此不等式，得到 $p$ 的最小值。

3. 设终端数为 $n$，则利用正态近似，

$$\begin{aligned} P(X \geq 15) &= 1 - P(X \leq 14) \\ &\approx 1 - P\left(\frac{X - 0.2n}{\sqrt{0.16n}} < \frac{14.5 - 0.2n}{\sqrt{0.16n}}\right) \\ &= 1 - \Phi\left(\frac{14.5 - 0.2n}{\sqrt{0.16n}}\right) \geq 0.95 \end{aligned}$$

因此

$$\begin{aligned} \Phi\left(\frac{14.5 - 0.2n}{\sqrt{0.16n}}\right) &\leq 0.05 \\ \frac{14.5 - 0.2n}{\sqrt{0.16n}} &\leq u_{0.05} \end{aligned}$$

解此不等式，得到 $n$ 的最小值。

4. 略

5. 略

6. 略

# 5 统计量

## 5.1 定理 5.4.1 证明

证明来自正态总体 $N(\mu, \sigma^2)$ 容量为 $n$ 的样本，其样本均值服从 $N(\mu, \sigma^2/n)$。

解：根据连续场合的卷积公式和正态分布的性质，正态分布满足可加性，由于 $x_i \sim N(\mu, \sigma^2)$，因此 $\sum_{i=1}^{n} x_i$ 和 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 都服从正态分布。而且

$$\begin{aligned}
E(\bar{x}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\
&= \frac{1}{n}(\sum_{i=1}^{n} E(x_i)) \\
&= \frac{1}{n} \cdot n \cdot \mu \\
&= \mu \\
Var(\bar{x}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\
&= \left(\frac{1}{n}\right)^2 Var(\sum_{i=1}^{n} x_i) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} Var(x_i) \\
&= \left(\frac{1}{n}\right)^2 \cdot n \cdot \sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

因此，样本均值

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

## 5.2　习题，5.4.2

由正态总体 $N(\mu, \sigma^2)$ 抽取两个独立样本，容量分别是 $m, n$，试求两个样本均值相差不超过 $\epsilon$ 的概率。

解：两个样本均值分别表示为 $\bar{x_m}$ 和 $\bar{x_n}$，易知

$$\bar{x_m} \sim N(\mu, \sigma^2/m), \quad \bar{x_n} \sim N(\mu, \sigma^2/n)$$

则它们的差

$$y = \bar{x_m} - \bar{x_n} \sim N(0, \sigma^2/m + \sigma^2/n)$$

即

$$\frac{y}{\sqrt{\sigma^2/m + \sigma^2/n}} \sim N(0,1)$$

因此，

$$\begin{aligned} P(|y| < \epsilon) &= P\left(\left|\frac{y}{\sqrt{\sigma^2/m + \sigma^2/n}}\right| < \frac{\epsilon}{\sqrt{\sigma^2/m + \sigma^2/n}}\right) \\ &= 2\Phi\left(\frac{\epsilon}{\sqrt{\sigma^2/m + \sigma^2/n}}\right) - 1 \end{aligned}$$

# 6 参数估计

## 6.1 习题，6.1.7

设从均值为 $\mu$，方差为 $\sigma^2$ 的总体中，分别抽取容量为 $n_1$ 和 $n_2$ 的两独立样本，$\bar{x_1}$ 和 $\bar{x_2}$ 分别是这两个样本的均值。

1. 证明，对于任意的常数 $a, b$，满足 $a+b=1$，$Y = a\bar{x_1} + b\bar{x_2}$ 都是 $\mu$ 的无偏估计。
2. 确定常数 $a, b$，使 $Var(Y)$ 达到最小。

解:

1.

$$\begin{aligned} E(Y) &= E(a\bar{x_1} + b\bar{x_2}) \\ &= aE(\bar{x_1}) + bE(\bar{x_2}) \\ &= a\mu + b\mu \\ &= (a+b)\mu \\ &= \mu \end{aligned}$$

2.

$$\begin{aligned} Var(Y) &= Var(a\bar{x_1} + b\bar{x_2}) \\ &= a^2 Var(\bar{x_1}) + b^2 Var(\bar{x_2}) \\ &= \frac{a^2}{n_1}\sigma^2 + \frac{b^2}{n_2}\sigma^2 \\ &= \left(\frac{a^2}{n_1} + \frac{b^2}{n_2}\right)\sigma^2 \end{aligned}$$

现在要在 $a+b=1$ 的条件下，求上式的最小值。即求

$$s = \frac{a^2}{n_1} + \frac{(1-a)^2}{n_2}$$

的最小值。对 $a$ 求导数并令其为 $0$，

$$\frac{\partial s}{\partial a} = \frac{2a}{n_1} - \frac{2(1-a)}{n_2} = 0$$

解得，

$$\begin{aligned} a &= \frac{n_1}{n_1+n_2} \\ b &= \frac{n_2}{n_1+n_2} \end{aligned}$$

### 6.2 区间估计

正态总体 $N(\mu,\sigma^2)$，$\sigma^2$ 已知，$x_1,x_2,\ldots,x_n$ 为样本，根据枢轴量法，分别推导 $\mu$ 的置信水平为 $1-\alpha$ 的

1. 同等置信区间
2. 同等单侧置信下限
3. 同等单侧置信上限

解：设枢轴量为

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

1. 根据同等置信区间的定义，即是要确定一对 $c<d$，使得

   $$P(c\le y\le d)=1-\alpha$$

   由于 $u$ 服从标准正态分布，要想区间 $(c,d)$ 最短，则要满足 $c=-d$，因此上式变成

   $$P(-d\le y\le d)=1-\alpha$$

   现在需要求出 $d$。因为

   $$P(-d\le y\le d)=P(|y|\le d)=2\Phi(d)-1$$

   所以

   $$\begin{aligned}2\Phi(d)-1&=1-\alpha\\ \Phi(d)&=1-\frac{\alpha}{2}\\ d&=u_{1-\frac{\alpha}{2}}\end{aligned}$$

   带入得到，

   $$-u_{1-\frac{\alpha}{2}}\le\frac{\bar{x}-\mu}{\sigma}\le u_{1-\frac{\alpha}{2}}$$

   不等式变换得到，

   $$\bar{x}-u_{1-\frac{\alpha}{2}}\sigma\le\mu\le\bar{x}+u_{1-\frac{\alpha}{2}}\sigma$$

   即同等置信区间为

   $$\left[\bar{x}-u_{1-\frac{\alpha}{2}}\sigma,\bar{x}+u_{1-\frac{\alpha}{2}}\sigma\right]$$

2. 根据同等单侧置信下限的定义，即要确定 $d$，使得

   $$P(y\le d)=1-\alpha$$

   因为

   $$P(y\le d)=\Phi(d)$$

所以

$$\begin{aligned}\Phi(d) &= 1-\alpha \\ d &= u_{1-\alpha}\end{aligned}$$

则有

$$\begin{aligned}y &\le u_{1-\alpha} \\ \frac{\bar{x}-\mu}{\sigma} &\le u_{1-\alpha} \\ \mu &\ge \bar{x} - u_{1-\alpha}\sigma\end{aligned}$$

3. 类似地，要确定一个 $c$，使得

$$P(y \ge c) = 1-\alpha$$

因为

$$P(y \ge c) = 1 - P(y < c) = 1 - \Phi(c)$$

因此

$$\begin{aligned}1 - Phi(c) &= 1-\alpha \\ Phi(c) &= \alpha \\ c &= u_\alpha\end{aligned}$$

即

$$\begin{aligned}y &\ge u_a lpha \\ \frac{\bar{x}-\mu}{\sigma} &\ge u_\alpha \\ \mu &\le \bar{x} - u_\alpha\sigma\end{aligned}$$

# 7 假设检验

## 7.1 拒绝域推导

设 $x_1, x_2, \ldots, x_n$ 是来自 $N(\mu, \sigma^2)$ 的样本，$\sigma^2$ 已知，考虑如下关于 $\mu$ 的检验问题:

$$H_0 : \mu = c \quad \texttt{vs} \quad H_1 : \mu \neq c$$

其中 $c$ 为一个常量。

1. 推导这个检验问题在显著性水平为 $\alpha$ 的拒绝域。
2. 当 $\mu = c$ 时，这个检验犯第一类错误和第二类错误的概率分别是多少?

解:

1. 因为 $\sigma^2$ 已知，确定检验统计量为

$$y = \frac{\bar{x} - c}{\sigma/\sqrt{n}}$$

这是一个双侧检验问题，因此拒绝域的形式应该是

$$|y| \geq k$$

已知显著性水平为 $\alpha$，则

$$\begin{aligned} P(|y| \geq k) &= \alpha \\ 1 - P(|y| < k) &= \alpha \\ 1 - (2\Phi(k) - 1) &= \alpha \\ \Phi(k) &= 1 - \frac{\alpha}{2} \\ k &= u_{1-\frac{\alpha}{2}} \end{aligned}$$

因此得到的拒绝域为

$$W = \left\{ (x_1, x_2, \ldots, x_n) | \left| \frac{\bar{x} - c}{\sigma/\sqrt{n}} \right| \geq u_{1-\frac{\alpha}{2}} \right\}$$

2. 第一类错误指的是，此时 $\mu = c$，应该接受 $H_0$，但是获得的样本 $(x_1, \ldots, x_n)$ 落入了拒绝域，导致 $H_0$ 被拒绝，即要计算在 $\mu = c$ 条件下，样本落入拒绝域的概率，而这个概率正是上面的显著性水平 $\alpha$，所以犯第一类错误的概率为 $\alpha$。也可以计算如下，

$$\begin{aligned} P(W|\mu = c) &= P\left( \left| \frac{\bar{x} - c}{\sigma/\sqrt{n}} \right| \geq u_{1-\frac{\alpha}{2}} \right) \\ &= 2 - 2\Phi(u_{1-\frac{\alpha}{2}}) \\ &= 2 - 2(1 - \frac{\alpha}{2}) \\ &= \alpha \end{aligned}$$

第二类错误指的是，当 $\mu \neq c$ 时，样本没有落入拒绝域，而此时 $\mu = c$，因此 $\mu \neq c$ 的前提不存在，因此不会犯第二类错误，即犯第二类错误的概率为 0。

## 7.2 开放题，思维发散

关于区间估计和假设检验的关系，

1. 区间估计中的“置信水平”与假设检验中的显著性水平”有关系吗?
2. 区间估计中的“单侧置信下限”、“单侧置信上限”与假设检验中的“单侧假设”有关系吗?
3. 区间估计中的“置信区间”与假设检验中的“双侧假设”有关系吗?
4. 区间估计问题能不能转化成假设检验问题? 如果能，如何转化?
5. 假设检验问题能不能转化成区间估计问题? 如果能，如何转化?

答：参见同济大学数学系编概率论与数理统计，中国工信出版集团，人民邮电出版社，2017/03，第八章拓展阅读