2019 级大气二班

2020~2021 秋季学年概率统计

期末复习资料

整理：蒋斌

前言

《概率论与数理统计》这门课程，是由概率部分与统计部分构成。考查的内容一共涉及八章，分别为前五章的概率论部分和后三章的统计部分。前五章的内容为随机事件及其概率、随机变量及其分布、多维随机变量及其分布、随机变量的数字特征、大数定律及中心极限定理，后三章的内容为样本及抽样分布、参数估计、区间假设。

概率论的部分，涉及概率的计算、积分的运算，重在考查计算能力；统计部分涉及的结论比较多，大家从课本给出的一些表就知道，但是这并不是要求我们去背结论，而是要我们掌握其推导过程，最主要的是把四个抽样定理吃透了，处理统计部分的问题就不会手足无措。总的来说，我们要学好概率统计，最重要的还是要把课本上的基本概念记熟、弄明白——再加上适当的做题，对付期末考试就没有问题了。

本资料将每章的一些典型例题进行了整理，每章附有适量的练习题，希望对大家复习概率统计有所帮助。由于资料缺少审核，难免存在错误之处，请读者能够积极指出（联系邮箱：lesnow_bin@163.com）

祝大家期末考试顺利！

2021 年 1 月 16 日

第一章 概率论的基本概念

➢ *章节知识体系*

- 随机试验
- 样本空间、随机事件
- 频率与概率
- 等可能概型（古典概型）
- 条件概率
- 独立性

➢ *重要术语及主题*

随机试验、样本空间、随机事件、基本事件、德摩根律、频率、概率、古典概型、事件的概率计算、概率加法定理、条件概率、概率乘法公式、全概率公式、贝叶斯公式、事件的独立性、实际推断原理

➢ *典型例题*

1. 设 A、B、C 是三个事件，且$\mathrm{P}(A)=P(B)=P(C)=\frac{1}{4}, P(AB)=P(BC)=0, P(AC)=\frac{1}{8}$，求 A、B、C 三个事件至少有一个发生的概率.

解: 题中“至少有一个发生”的字眼表明求一个和事件的概率，再根据事件计算的公式，有

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$$

从而有$\mathrm{P}(A\cup B\cup C)=\frac{1}{4}+\frac{1}{4}+\frac{1}{4}-0-0-\frac{1}{8}+0=\frac{5}{8}$

注：$P(AB)=P(BC)=0$可以推出$P(ABC)=0$，实际上$P(AB)$、$P(AC)$、$P(BC)$中任意一个为 0，就可以得出$P(ABC)=0$的结论

2. 已知$P(A)=\frac{1}{2},P(B)=\frac{1}{3},P(C)=\frac{1}{4},P(AB)=\frac{1}{10},P(BC)=\frac{1}{20},P(AC)=\frac{1}{15},P(ABC)=\frac{1}{30}$，求$A\cup B,\overline{A}\overline{B},A\cup B\cup C,\overline{A}\overline{B}\overline{C},\overline{A}\overline{B}C,\overline{A}\overline{B}\cup C$的概率.

解： $P(A\cup B)=P(A)+P(B)-P(AB)=\frac{1}{2}+\frac{1}{3}-\frac{1}{10}=\frac{11}{15}$

$P(\overline{A}\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=\frac{4}{15}$

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)=\frac{1}{2}+\frac{1}{3}+\frac{1}{4}-\frac{1}{10}-\frac{1}{15}-\frac{1}{20}+\frac{1}{30}=\frac{17}{20}$$

$P(\overline{A}\overline{B}\overline{C})=P(\overline{A\cup B\cup C})=1-P(A\cup B\cup C)=\frac{3}{20}$

$P(\overline{A}\overline{B}C)=P(\overline{A}\overline{B})-P(\overline{A}\overline{B}\overline{C})=\frac{4}{15}-\frac{3}{20}=\frac{7}{60}$

$P(\overline{A}\overline{B}\cup C)=P(\overline{A}\overline{B})+P(C)-P(\overline{A}\overline{B}C)=\frac{4}{15}+\frac{1}{4}-\frac{7}{60}=\frac{7}{20}$

注：本题较为全面的体现了计算概率用到的一些公式，是一道不错的例题！

3. 一俱乐部有 5 名一年级学生，2 名二年级学生，3 名三年级学生，2 名四年级学生，现从他们当中任选 5 名学生，求四个年级的学生均包含在内的概率.

解： 记事件 A 表示“四个年级的学生均包含在内的概率”.根据题意，五名学生中一定有两名学生来自同一个年级，故所求概率为：

$$P(A)=\frac{C_5^2C_2^1C_3^1C_2^1+C_5^1C_2^2C_3^1c_2^1+l_5^1C_2^1C_3^2C_2^1+C_5^1C_2^1C_3^1C_2^2}{c_{12}^5}=\frac{10}{33}$$

注：另一种解法：

先从四个年及中各选一名学生，再从剩下的八名中任选一名，式子为：

$$P(A)=\frac{C_5^1C_2^1C_3^1C_2^1\cdot C_8^1}{c_{12}^5}\cdot\frac{1}{2}=\frac{10}{33}$$

除以$\frac{1}{2}$的说明：假设二年级的同学 a 和 b，都被选中，这种情况可以是：C_2^1选中 a，C_8^1选中 b；或者C_2^1选中 b，C_8^1选中 a，两者都表示同一个结果，出现了重

复情况，同理可知其他情况也是如此，故需要除以 2.

4. 设袋中有红、白、黑球各 1 个，从中有放回地取球，每次取一个，直到三种颜色都出现即停止摸球，求取球次数恰好为 4 的概率.

解：记事件 A 为“取球次数恰好为 4 的概率”，因为是有放回地取球，故每次摸球都有三个结果，从而样本空间$S = 3^4$.

取球次数恰好为 4，说明最后一次单独一种颜色（假如为红色），有C_3^1种可能；

前三次取剩下的两种颜色（白黑）并且一种颜色出现两次，一种颜色出现一次，有$C_3^2C_2^1$种可能，所以$P(A)=\frac{C_3^1C_3^2C_2^1}{3^4} = \frac{2}{9}$

5. 甲袋中装有 9 个乒乓球，其中 3 个白球，6 个黄球；乙袋中也装有 9 个乒乓球，其中 5 个白球，4 个黄球. 首先从甲袋中任取一个球放入乙袋，再从乙袋中任取一球放入甲袋中，则甲袋中白球数目不发生变化的概率为_______.

解：令事件 A 表示“经过两次交换后，甲袋中白球的数目保持不变”；

事件 B 表示“从甲袋中取出并放入乙袋中的是白球”；

事件 C 表示“从乙袋中取出并放入甲袋中的是白球”；

则事件$A = BC+\overline{B}\overline{C}$，于是$P(A)=P(BC)+P(\overline{B}\overline{C})=P(B)P(C|B)+P(\overline{B})P(\overline{C}|\overline{B})=\frac{3}{9}\times\frac{6}{10}+\frac{6}{9}\times\frac{5}{10}=\frac{8}{15}$

6. 对于二事件 A 和 B，下列说法正确的是______.

A.若$AB\neq\emptyset$，则 A,B 一定独立　　　　B.若$AB\neq\emptyset$，则 A,B 有可能独立

C.若$AB=\emptyset$，则 A,B 一定独立　　　　D.若$AB=\emptyset$，则 A,B 一定不独立

解：选 B。考查独立与互斥的概念。当$AB\neq\emptyset$是，需满足$P(AB)=P(A)P(B)$的条件才能说明 A 和 B 相互独立，从而可知 A 选项错误 B 选项正确。当$AB=\emptyset$时，若 P(A)与 P(B)均大于 0，则 A,B 一定不独立，而当 P(A)与 P(B)至少有一个为零时，A,B 一定是独立的，故 C,D 选项错误。

7. 设两个相互独立的事件 A,B 都不发生的概率为$\frac{1}{9}$,A 发生 B 不发生的概率与 B

发生 A 不发生的概率相等，则 P(A)=______.

解：由题可知：$P(\bar{A}\bar{B})=\frac{1}{9}$,$P(A\bar{B})=P(\bar{A}B)$.又$P(A)=P(AB)+P(A\bar{B})=P(AB)+P(\bar{A}B)=P(B)$.A,B 是相互独立的，所以有$P(\bar{A}\bar{B})=P(\bar{A})P(\bar{B})=(1-P(A))(1-P(B))=(1-P(A))^2=\frac{1}{9}$,从而求得$P(A)=\frac{2}{3}$.

8. 在某城市中发行的三种报纸 A,B,C，经调查，订阅 A 报纸的有 45%，订阅 B 报纸的有 35%，订阅 C 报纸的有 30%，同时订阅 A,B 报纸的有 10%，同时订阅 A,C 报纸的有 8%，同时订阅 B,C 报纸的有 5%，同时订阅 A,B,C 报纸的有 3%.试求下列事件的概率：

(1)只订阅 A 报纸的；(2)只订 A,B 报纸的；(3)只订阅一种报纸的；(4)恰好订阅两种报纸的；(5)至少订阅一种报纸的；(6)不订阅任何报纸的；(7)至多订阅一种报纸的.

解： (1)$P(A\bar{B}\bar{C})=P(A-B-C)=P(A-(B\cup C))=P(A-A(B\cup C))=P(A)-P(A(B\cup C))=P(A)-P(AB)-P(AC)+P(ABC)=0.45-0.1-0.08+0.03=0.30$

(2)$P(AB\bar{C})=P(AB-C)=P(AB-ABC)=P(AB)-P(ABC)=0.1-0.03=0.07$

(3)$P(A\bar{B}\bar{C}\cup\bar{A}B\bar{C}\cup\bar{A}\bar{B}C)=P(A\bar{B}\bar{C})+P(\bar{A}B\bar{C})+P(\bar{A}\bar{B}C)=0.30+P(B-B(A\cup B))+P(C-C(A\cup B))=0.73$

(4)$P(AB\bar{C}\cup A\bar{B}C\cup\bar{A}BC)=P(AB\bar{C})+P(A\bar{B}C)+P(\bar{A}BC)=P(AB)-P(ABC)+P(AC)-P(ABC)+P(BC)-P(ABC)=P(AB)+P(AC)+P(BC)-3P(ABC)=0.1$

$$0+0.08+0.05-3\times0.03=0.14$$

(5)$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$

$$=0.45+0.35+0.30-0.10-0.08-0.05+0.03=0.90$$

(6)$P(\bar{A}\bar{B}\bar{C})=1-P(\cup B\cup C)=0.10$

(7)$P(\bar{A}\bar{B}\bar{C}\cup A\bar{B}\bar{C}\cup\bar{A}B\bar{C}\cup\bar{A}\bar{B}C)=P(\bar{A}\bar{B}\bar{C})+P(A\bar{B}\bar{C})+P(\bar{A}B\bar{C})+P(\bar{A}\bar{B}C)$

$$=0.10+0.73=0.83$$

9. 设有来自三个地区的各 10 名，15 名，25 名考生的报名表，其中女生的报名

表分别有 3 份，7 份和 5 份，随机抽取一个地区的报名表，从中先后抽取两份，求：

(1) 求先抽到一份是女生的概率；

(2) 已知后抽到的一份是男生表，求先抽到一份是女生表的概率.

解：令A_i表示事件“第i次取出的是女生”；B_j表示事件“报名表来自第个地区的考生”（$i=1,2;j=1,2,3$）.根据题意：

$$P(B_1)=P(B_2)=P(B_3)=\frac{1}{3}$$

$$P(A_1|B_1)=\frac{3}{10},P(A_1|B_2)=\frac{7}{15},P(A_1|B_3)=\frac{5}{25}=\frac{1}{5}$$

(1) 由全概率公式：

$$p=P(A_1)=\sum_{i=1}^{3}P(B_i)P(A_1|B_i)=\frac{1}{3}\left(\frac{3}{10}+\frac{7}{15}+\frac{5}{25}\right)=\frac{29}{90}$$

(2) 由条件概率公式得：$q=P(A_1|\bar{A}_2)=\frac{P(A_1\bar{A}_2)}{P(\bar{A}_2)}$，由题意可知：

$$P(\bar{A}_2|B_1)=\frac{7}{10},P(\bar{A}_2|B_2)=\frac{8}{15},P(\bar{A}_2|B_3)=\frac{20}{25}=\frac{4}{5}$$

由全概率公式：$P(\bar{A}_2)=\sum_{i=1}^{3}P(B_i)P(\bar{A}_2|B_i)=\frac{1}{3}\left(\frac{3}{10}+\frac{7}{15}+\frac{5}{25}\right)=\frac{61}{90}$

$$P(A_1\bar{A}_2|B_1)=\frac{C_3^1C_7^1}{C_{10}^2}=\frac{7}{30},P(A_1\bar{A}_2|B_2)=\frac{C_7^1C_8^1}{C_{15}^2}=\frac{8}{30},P(A_1\bar{A}_2|B_3)=\frac{C_5^1C_{20}^1}{C_{25}^2}=\frac{5}{30}$$

$$P(A_1\bar{A}_2)=\sum_{i=1}^{3}P(B_i)P(A_1\bar{A}_2|B_i)=\frac{1}{3}\left(\frac{7}{30}+\frac{8}{30}+\frac{5}{30}\right)=\frac{20}{90}$$

所以$q=P(A_1|\bar{A}_2)=\frac{P(A_1\bar{A}_2)}{P(\bar{A}_2)}=\frac{20}{61}$

- ***第一章巩固练习题***

1. 设 A、B 为随机事件，$P(A)=0.7,P(A-B)=0.3$，则$P(\overline{AB})=$ ______.
2. 某油漆公司发出 17 桶油漆，其中白漆 10 桶，黑漆 4 桶，红漆 3 桶，在搬运过程所有标签脱落，交货人将这些油漆随意发给顾客.问一个订货为 4 桶白漆、3 桶黑漆、2 桶红漆的顾客，能如期按所订颜色得到订货的概率为________.

3. 将 3 只球随机地放入 4 个杯子中去，求杯子中球的最大个数分别为 1,2,3 的概率.
4. 第一只盒子装有 5 个红球，4 直白球；第二只盒子装有 4 只红球，5 只白球. 先从第一只盒子中任取 2 个球放入到第二个盒子中去，然后再从第二个盒子里任取一只球，则取到白球的概率为_______.
5. 分别从 5 双不同的鞋子中任取 4 只，这 4 只鞋子中至少有 2 只配对的概率为_______.
6. 一电子器件工厂从过去经验得知，一位新工人参加培训后能完成生产定额的概率为0.86，而不参加培训完成生产定额的概率只有0.35，假设该工厂中有80%的工人参加过培训.

（1） 一位新工人完成生产定额的概率为多少?

（2） 已知一位新工人完成了生产定额，问他参加过培训的概率是多少?

7. 一口袋中有 6 个球，对其中球的颜色有三种看法：A_1：袋中有 4 只红球、2 只白球；A_2：袋中有 3 只红球、3 只白球；A_3：袋中有 2 只红球、4 只白球. 对这三种看法的某人认为其发生的可能性为：$P(A_1)=\frac{1}{3};P(A_2)=\frac{1}{6};P(A_3)=\frac{1}{2}$. 此人现在从袋中取一个球，取到的是白球，则此人应如何修改自己的看法?
8. 城乡超市销售一批相机共 10 台，其中有 3 台次品，其余均为正品，某顾客去选购时超市已售 2 台，该顾客从剩下的 8 台中任意选购一台，求：

（1） 该顾客购到正品的概率；

（2） 已知顾客购到正品，则已售出两台都是次品的概率.

9. 玻璃杯成箱出售，每箱 20 只，假设各箱含 0,1,2 只残次品的概率分别为 0.8,0.1,0.1. 一顾客欲购买一箱玻璃杯，在购买时，售货员任取一箱，顾客只开箱检查 4 只，若无残次品，则买下该箱玻璃杯，否则退回，问：

（1） 顾客买下这箱玻璃杯的概率；

（2） 在顾客买下的一箱中，确实没有次品的概率.

10. 将两信息分别编码为 A 和 B 传递出去，接收站收到时，A 被误认为 B 的概率为 0.02，而 B 被误认为 A 的概率为 0.01，信息 A 和 B 传递的频繁程度为

2:1，若接收到信息是 A，问原发信息是 A 的概率是多少?

答案:

1. 0.4
2. $\frac{C_{10}^{4}C_{4}^{3}C_{3}^{2}}{C_{17}^{9}}=\frac{252}{2431}$
3. $P(X=1)=\frac{3}{8},P(X=2)=\frac{9}{16},P(X=3)=\frac{1}{16}$
4. $\frac{53}{99}$(第一个盒子摸出两个球的可能情况：红红、白白、红白)
5. $\frac{13}{21}$
6. (1)0.758；(2)$\frac{0.688}{0.758}\approx 0.9077$
7. 提示：利用贝叶斯公式计算得到：$\mathrm{P}(A_1|白)=\frac{4}{19}$；$\mathrm{P}(A_2|白)=\frac{3}{19}$；$\mathrm{P}(A_3|白)$ $=\frac{12}{19}$
8. (1)$\frac{7}{10}$；(2)$\frac{1}{12}$（提示：记事件A_i表示“第i台相机是次品”（$i=1,2,3$））
9. (1)0.94；(2)0.85
10. $\frac{196}{197}$

第二章 随机变量及其分布

➢ *章节知识体系*

- 随机变量
- 离散型随机变量及其分布律
- 随机变量的分布函数
- 连续型随机变量及其概率密度
- 随机变量的函数的分布

➢ *重要术语及主题*

随机变量、分布函数、离散型随机变量及其分布律、连续型随机变量及其概率密度、伯努利试验、两点分布、n 重伯努利试验、二项分布、泊松分布、指数分布、均匀分布、正态分布、随机变量函数的分布

➢ *典型例题*

1. 设随机变量 X 服从参数为(2，p)的二项分布，随机变量 Y 服从参数为(3，p)的二项分布，若 P{X ≥ 1}=$\frac{5}{9}$，则 P{Y ≥ 1}=________.

解： $\frac{5}{9}=P\{X\geq 1\}=1-P\{X<1\}=1-C_2^0p^0(1-p)^2=1-(1-p)^2$,解得$p=\frac{1}{3}$

从而 P{Y ≥ 1}=$1-P\{Y<1\}=1-C_3^0p^0(1-p)^3=1-(1-p)^3=\frac{19}{27}$

2. 有甲、乙两种味道和颜色都极为相似的名酒各 4 杯. 如果从中挑 4 杯，能将甲种酒全部挑出来，算是试验成功一次.

(1) 某人随机地去猜，问他猜对一次的概率为多少？

(2) 某人声称他通过品尝能够区分两种酒. 他连续试验 10 次，成功 3 次，试推断他是猜对的，还是确有区分能力（设各次试验是相互独立的）.

解：(1) 随机试验是从 8 杯酒中任选 4 杯，样本空间为C_8^4，所以猜对一次的概率为$p=\frac{1}{C_8^4}=\frac{1}{70}$；

(2) 假设他是猜对的，则猜对的次数 X 满足二项分布，即 $X\sim B(10,\ \frac{1}{70})$，则猜对的概率为$p\{X=3\}=C_{10}^3\left(\frac{1}{70}\right)^3\left(\frac{69}{70}\right)^7\approx\frac{\left(\frac{1}{7}\right)^3e^{-\frac{1}{7}}}{3!}\approx 3\times10^{-4}$

显然这个概率很小，由实际推断原理，可以认为他是有区分能力的

点评：本题第二问直接计算二项分布显然十分麻烦，实际上对于 n 比较大而 p 比较小的二项分布，常常采用泊松分布来近似替代二项分布，这样可以大大减少计算量

3. 现有同型设备 300 台，各设备的工作是相互独立的，发生故障的概率都是 0.01. 设一台设备的故障可由一名维修工人处理，问至少需要配备多少名维修工人，才能保证设备发生故障但不能及时维修的概率小于 0.01？

解：设需要配备 N 名工人，X 为同一时刻发生故障的设备台数，则 $X\sim B(300,0.01)$. 所以题中的问题转化为确定 N 的最小值，使$P\{X\le N\}\ge 0.99$.

由泊松定理，$\lambda=np=3$，所以$P\{X\le N\}\approx\sum_{k=0}^{N}\frac{3^k}{k!}e^{-3}\ge 0.99$

查表，当 $N\ge 8$ 时上式成立，因此，为达到上述要求，至少配备 8 名维修工人.

4. $f(x)=ce^{-x^2+x}$是随机变量 X 的密度函数，则 $c=$_________.

解：$\int_{-\infty}^{+\infty}f(x)dx=c\cdot\sqrt{\pi}e^{\frac{1}{4}}\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\cdot\frac{\sqrt{2}}{2}}\cdot e-\frac{(x-\frac{1}{2})^2}{2(\frac{\sqrt{2}}{2})^2}=c\cdot\sqrt{\pi}e^{\frac{1}{4}}=1$

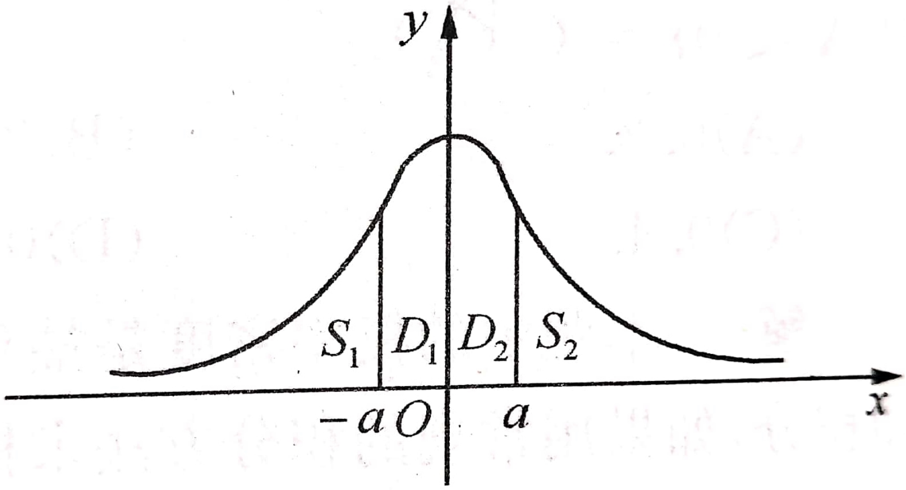
从而解得$c=\frac{1}{\sqrt{\pi}}e^{-\frac{1}{4}}$

5. 设随机变量 X 的密度函数为$\varphi(x)$，且$\varphi(-x)=\varphi(x)$，$F(x)$是 X 的分布函数，则对任意实数 a，有______.

A. $F(-a)=1-\int_0^a\varphi(x)dx$　　B. $F(-a)=\frac{1}{2}-\int_0^a\varphi(x)dx$

C. $F(-a)=F(a)$　　D. $F(-a)=2F(a)-1$

解：因为密度函数为偶函数，可直接采用图像法来求解：



易知 B 选项正确.

6. 设随机变量 X 的概率密度为$f(x)=\begin{cases}\frac{1}{3},若\ x\in[0,1]\\ \frac{2}{9},若 x\in[3,6]\\ 0,\qquad 其他\end{cases}$，若 k 使得$P\{X\geq k\}=\frac{2}{3}$，

则 k 的取值范围是________.

解：由密度函数可以求出分布函数（过程请读者自己完成）：

$$F(x)=\begin{cases}0, & x<0\\ \frac{1}{3}x, & 0\leq x<1\\ \frac{1}{3}, & 1\leq x<3\\ \frac{1}{3}+\frac{2}{9}(x-3), & 3\leq x<6\\ 1, & x\geq 6\end{cases}$$

于是$P\{X\geq k\}=1-P\{X<k\}$，解得$P\{X<k\}=\frac{1}{3}=F(k)$，由分布函数可知，$k$

的取值范围为[1,3]

7. 设$X\sim N(0,1)$. (1) 求$Y=e^X$的概率密度；(2) 求$Y=2X^2+1$的概率密度；(3) 求

$Y=|X|$的概率密度.

解：(1) X 的概率密度为$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},-\infty<x<+\infty$

$F_Y(y)=P\{Y\leq y\}=P\{e^X\leq y\}$，当$y\leq 0$时，显然$F_Y(y)=0$；当$y>0$时，有

$$F_Y(y) = \mathrm{P}\{X \le \ln y\} = \int_{-\infty}^{\ln y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

于是 y 的概率密度为$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}y} \cdot e^{-\frac{(\ln y)^2}{2}}, y > 0 \\ 0, \quad y \le 0 \end{cases}$

(2) 由$Y = 2X^2 + 1$知$Y \ge 1$，故当$y < 1$时$Y \le y$是不可能事件，所以$F_Y(y) = 0$即$f_Y(y) = 0$. 当$y \ge 1$时，有$F_Y(y) = \mathrm{P}\{2X^2 + 1 \le Y\} = P\left\{-\frac{\sqrt{Y-1}}{2} \le X \le \frac{\sqrt{Y-1}}{2}\right\} = \int_{-\frac{\sqrt{y-1}}{2}}^{\frac{\sqrt{y-1}}{2}} f(x)dx = \int_{-\frac{\sqrt{y-1}}{2}}^{\frac{\sqrt{y-1}}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

于是 y 的概率密度为$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{\pi(y-1)}} e^{-\frac{y-1}{4}}, y > 1 \\ 0, \quad y \le 1 \end{cases}$.

(3) 由$Y = |X|$知$Y \ge 0$，故当$y < 0$时$Y \le y$是不可能事件，所以$F_Y(y) = 0$即$f_Y(y) = 0$. 当 $y \ge 0$时，有 $F_Y(y) = \mathrm{P}\{|X| \le Y\} = P\{-Y \le X \le Y\} = \int_{-y}^{y} f(x)dx = \int_{-y}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2\int_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

于是 y 的概率密度为$f_Y(y) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{y^2}{4}}, \quad y > 0 \\ 0, \quad y \le 0 \end{cases}$.

点评：理解和把握分布函数及概率密度的概念，解题时思路就比较清晰.

8. 设随机变量 X 的概率密度为 $f(x) = \begin{cases} \frac{1}{9}x^2, 0 < x < 3 \\ 0, \text{其他} \end{cases}$，令随机变量：

$$Y = \begin{cases} 2, X \le 1 \\ X, 1 < X < 2 \\ 1, X \ge 2 \end{cases}$$

求：(1) Y 的分布函数；(2) 求概率 $P\{X \le Y\}$.

解：(1) 由题可知：$1 \le Y \le 2$，而 $F_Y(y) = P\{Y \le y\}$

所以当 $y<1$ 时，$F_Y(y)=0$，当 $y\geq 2$ 时，$F_Y(y)=1$

当 $1\leq y<2$ 时，$F_Y(y)=P\{Y=1\}+P\{1<Y\leq y\}=P\{X\geq 2\}+P\{1<X\leq y\}$

$$=\int_2^3 \frac{1}{9}x^2dx+\int_1^y \frac{1}{9}x^2dx=\frac{y^3+18}{27}$$

所以 Y 的分布函数为：$F_Y(y)=\begin{cases}0, y<1\\ \dfrac{y^3+18}{27}, 1\leq y<2\\ 1, y\geq 2\end{cases}$

(2) $P\{X\leq Y\}=P\{X<2\}=\int_0^2 \frac{1}{9}x^2dx=\frac{8}{27}$

9. 设随机变量 X 的概率密度为 $f(x)=\begin{cases}\dfrac{1}{3\sqrt[3]{x^2}}, 1\leq x\leq 8\\ 0, \text{其他}\end{cases}$，$F(x)$ 是 X 的分布函数，

求随机变量 $Y=F(X)$ 的分布函数.

解：由题可知，$Y\in[0,1]$

当 $x<1$ 时，$F(x)=0$；当 $x>8$ 时，$F(x)=1$；

当 $x\in[1,8]$ 时，$F(x)=\int_1^3 \frac{1}{3\sqrt[3]{t^2}}dt=\sqrt[3]{x}-1$

令 $G(y)$ 为 $Y=F(X)$ 的分布函数，则有：

当 $y\leq 0$ 时，$G(y)=0$；当 $y\geq 1$ 时，$G(y)=1$；

当 $y\in(0,1)$ 时，$G(y)=P\{Y\leq y\}=P\{F(X)\leq y\}=P\{\sqrt[3]{X}-1\leq y\}$

$$=P\{X\leq (y+1)^3\}=F((y+1)^3)=y$$

所以的分布函数为：$G(y)=\begin{cases}0, y<0\\ y, 0\leq y<1\\ 1, y\geq 1\end{cases}$

点评：本题也可以不用求出$F(x)$，记住该结论：若连续性随机变量X的分布函数为$F(x)$，则不论X服从何种分布，$Y=F(X)$一定服从$(0,1)$上的均匀分布（具体推导过程请读者自行完成）

➢ ***第二章巩固练习题***

1. 设随机变量 X 的密度函数$f(x)$满足$f(1+x)=f(1-x)$，且$\int_0^2 f(x)dx=0.6$，则$P\{X<0\}=$ ______.
2. 设随机变量 X 服从正态分布$N(\mu,\sigma^2)$，随着σ的增大，概率$P\{|X-\mu|<\sigma\}$的变化情况为______.

   A.单调增加　B.单调减少　C.保持不变　D.非单调变化
3. 设X_1,X_2,X_3是随机变量，且$X_1\tilde{}N(0,1),X_2\sim(0,4),X_3\sim(5,9),p_i=\{-2\le X_i\le 2\}$ $(i=1,2,3)$，则______.

   A.$p_1>p_2>p_3$　B.$p_2>p_1>p_3$　C. $p_3>p_2>p_1$　D. $p_1>p_3>p_2$
4. 设$F_1(x)$与$F_2(x)$是两个分布函数,其相应的概率密度$f_1(x)$与$f_2(x)$是连续函数，则必为概率密度的是_____.

   A.$f_1(x)f_2(x)$　B.$2f_2(x)F_1(x)$

   C.$f_1(x)F_2(x)$　D$f_1(x)F_2(x)+f_2(x)F_1(1)$
5. 设$f_1(x)$为标准正态分布的概率密度，$f_2(x)$为$[-1,3]$上的均匀分布的概率密度，若$f(x)=\begin{cases}af_1(x),x\le 0\\ bf_2(x),x>0\end{cases}$ $(a>0,b>0)$为概率密度，则a,b应满足______.

   A. $2a+3b=4$　B. $3a+2b=4$　C. $a+b=1$　D. $a+b=2$

6. 设随机变量X的概率分布为$P\{X=1\}=P\{X=2\}=\dfrac{1}{2}$. 在给定$X=i$的条件下，随机变量$Y$服从均匀分布$U(0,i)(i=1,2)$. 求$Y$的分布函数和概率密度.

答案:

1. 0.2
2. C（转化为标准正分布）
3. A（思路同上）
4. D
5. A(利用归一化条件)
6. $F_Y(y)=\begin{cases}0, y<0\\ \dfrac{3y}{4}, 0\le y<1\\ \dfrac{1}{2}+\dfrac{y}{4}, 1\le y<2\\ 1, y\ge 2\end{cases}, f_Y(y)=\begin{cases}\dfrac{3}{4}, 0<y<1\\ \dfrac{1}{4}, 1\le y<2\\ 0, \quad \text{其他}\end{cases}$

第三章 多维随机变量及其分布

➢ *章节知识体系*

- 二维随机变量
- 边缘分布
- 条件分布
- 相互独立的随机变量 → 联合概率密度等于各随机变量边缘概率密度的乘积
- 两个随机变量的函数分布
$$\begin{cases} Z=X+Y \\ Z=\dfrac{Y}{X}, Z=XY \\ M=\min\{X,Y\}, N=\min\{X,Y\} \end{cases}$$

➢ *重要主题及术语*

二维随机变量 (X,Y) 及其分布函数、离散型随机变量 (X,Y) 的联合分布律与边缘分布律、连续型随机变量 (X,Y) 的联合概率密度与边缘概率密度、条件分布律与条件分布概率密度、两个随机变量的独立性判断、两个随机变量之间的函数分布

➢ *典型例题*

1. 设二维随机变量 (X,Y) 的分布函数为：$F(x,y)=A(B+\arctan\dfrac{x}{2})(C+\arctan\dfrac{y}{3})$，

   求 A,B,C 以及 (X,Y) 的联合概率密度.

解： 由联合分布函数的性质可知：

$$F(+\infty,+\infty)=A(B+\frac{\pi}{2})(C+\frac{\pi}{2})=1$$

$$F(-\infty,+\infty)=A(B-\frac{\pi}{2})(C+\frac{\pi}{2})=0$$

$$F(+\infty,-\infty)=A(B+\frac{\pi}{2})(C-\frac{\pi}{2})=0$$

由此解得 $A=\dfrac{1}{\pi^2}, B=\dfrac{\pi}{2}, C=\dfrac{\pi}{2}$

其联合概率密度为：$f(x,y)=\dfrac{\partial^2 F}{\partial x\partial y}=\dfrac{6}{\pi^2(4+x^2)(9+y^2)}$

2. 设随机变量 $X_i\sim\begin{bmatrix}-1 & 0 & 1\\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4}\end{bmatrix}(i=1,2)$，且满足 $P\{X_1X_2=0\}=1$，则 $P\{X_1=X_2\}=$_______.

解： 由 $P\{X_1X_2=0\}=1$，可知 $P\{X_1X_2\neq 0\}=0$，也即：

$$P\{X_1=-1,X_2=-1\}=P\{X_1=1,X_2=-1\}=P\{X_1=-1,X_2=1\}=P\{X_1=1,X_2=1\}=0$$

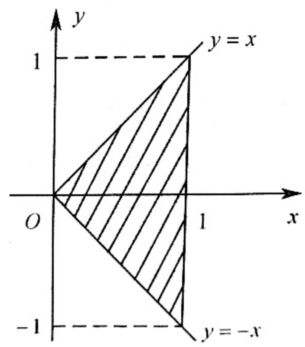
由上述条件，可得 X_1,X_2 的联合分布律如下：

X_1 \ X_2	-1	0	1	$p_{i\cdot}$
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$p_{\cdot j}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

所以 $P\{X_1=X_2\}=0$

3. 设随机变量 (X,Y) 的概率密度为 $f(x,y)=\begin{cases}1, |y|<x, 0<x<1\\ 0, 其他\end{cases}$，分别求条件概率密度 $f_{Y|X}(y|x)$，$f_{X|Y}(x|y)$

解： 概率密度的积分区域如下图所示：



故 $f(x,y)$ 的边缘概率密度为：

$$f_X(x)=\begin{cases}\int_{-x}^{x}1dy,0<x<1\\0,\text{其他}\end{cases}=\begin{cases}2x,0<x<1\\0,\text{其他}\end{cases}$$

$$f_Y(y)=\begin{cases}\int_{|y|}^{x}1dx,-1<y<1\\0,\text{其他}\end{cases}=\begin{cases}1-|y|,-1<y<1\\0,\text{其他}\end{cases}$$

所以当 $0<x<1$ 时， $f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}=\begin{cases}\dfrac{1}{2x},|y|<x\\0,\text{其他}\end{cases}$

当 $|y|<x$ 时， $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(x)}=\begin{cases}\dfrac{1}{1-|y|},|y|<x<1\\0,\text{其他}\end{cases}$

点评：注意连续型条件概率密度的计算公式及自变量的取值范围

4. 随 机 变 量 X_1,X_2,X_3,X_4 独 立 同 分 布 ， $P\{X_i=0\}=0.6,P\{X_i=1\}=0.4(i=1,2,3,4)$，求行列式 $X\sim\begin{vmatrix}X_1 & X_2\\X_3 & X_4\end{vmatrix}$ 的分布律.

解： 记 $Y_1=X_1X_4,Y_2=X_2X_3$， 则 $X=Y_1-Y_2$， 随机变量 Y_1,Y_2 独立同分布

$$P\{Y_1=1\}=P\{Y_2=1\}=P\{X_1=1,X_4=1\}=0.4\times0.4=0.16$$

$$P\{Y_1=0\}=P\{Y_2=0\}=1-0.16=0.84$$

随机变量有三个可能的取值 $-1,0,1$， 易见

$$P\{X=-1\}=P\{Y_1=0,Y_2=1\}=0.84\times0.16=0.1344$$

$$P\{X=1\}=P\{Y_1=1,Y_2=0\}=0.16\times0.84=0.1344$$

$$P\{X=0\}=1-2\times0.1344=0.7312$$

于是行列式的分布律为：

$$X\sim\begin{bmatrix}-1 & 0 & 1\\ 0.1344 & 0.7312 & 0.1344\end{bmatrix}$$

5. 设二维随机变量(X,Y)的概率密度为：$f(x,y)=\begin{cases}1, 0<x<1, 0<y<2x\\ 0, \qquad 其他\end{cases}$，求：

(1) (X,Y)的边缘概率密度；

(2) $Z=2X-Y$的概率密度$f_Z(z)$；

(3) $P\{Y\le\frac{1}{2}\mid X\le\frac{1}{2}\}$

解：(1) 当$0<x<1$时，$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\int_0^{2x}dy=2x$

当$x\le0$或者$x\ge1$时，$f_X(x)=0$，从而$f_X(x)=\begin{cases}2x, 0<x<1\\ 0, \qquad 其他\end{cases}$

当$0<y<2$时，$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx=\int_{\frac{y}{2}}^{1}dx=1-\frac{y}{2}$

当$y\le0$或者$y\ge2$时，$f_Y(y)=0$，从而$f_Y(y)=\begin{cases}1-\frac{y}{2}, 0<y<2\\ 0, \qquad 其他\end{cases}$

(2) （解法一）当$z\le0$时，$F_Z(z)=0$

当$0<z<2$时，$F_Z(z)=P\{2X-Y\le z\}=\iint\limits_{2x-y\le z}f(x,y)dxdy=z-\frac{z^2}{4}$

当$z\ge2$时，$F_Z(z)=1$，从而$f_Z(z)=\begin{cases}1-\frac{z}{2}, 0<z<2\\ 0, \qquad 其他\end{cases}$

（解法二）$f_Z(z)=\int_{-\infty}^{\infty}f(x,2x-z)dx$，

其中 $f(x,2x-z)=\begin{cases}1,0<x<1,0<z<2x\\0, \qquad 其他\end{cases}$，当 $z\le 0$或者$z\ge 2$时，$f_Z(z)=0$

当 $0<z<2$ 时，$f_Z(z)=\int_{\frac{z}{2}}^{1}dx=1-\frac{z}{2}$，所以 $f_Z(z)=\begin{cases}1-\frac{z}{2},0<z<2\\0, \qquad 其他\end{cases}$

(3) $P\{Y\le\frac{1}{2}\mid X\le\frac{1}{2}\}=\dfrac{P\{X\le\frac{1}{2},Y\le\frac{1}{2}\}}{P\{X\le\frac{1}{2}\}}=\dfrac{\frac{3}{16}}{\frac{1}{4}}=\frac{3}{4}$

6. 设 (X,Y) 的概率密度为：$f(x,y)=\begin{cases}\frac{21}{4}x^2y,x^2\le y\le 1\\0, \qquad 其他\end{cases}$

(1) 求条件概率密度 $f_{Y|X}(y\mid x)$，并写出当 $X=\frac{1}{2}$ 时 Y 的条件概率密度；

(2) 求条件概率 $P\{Y\ge\frac{3}{4}\mid X=\frac{1}{2}\}$.

解： 由 $f(x,y)=\begin{cases}\frac{21}{4}x^2y,x^2\le y\le 1\\0, \qquad 其他\end{cases}$ 可分别求出 X，Y 的边缘概率密度为：

$f_X(x)=\begin{cases}\frac{21}{8}x^2(1-x^4),-1\le x\le 1\\0, \qquad 其他\end{cases}$，$f_Y(y)=\begin{cases}\frac{7}{2}y^{\frac{5}{2}},0\le y\le 1\\0, \qquad 其他\end{cases}$

(1) $f_{Y|X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}=\begin{cases}\frac{2y}{1-x^4},x^2<y<1,-1<x<1\\0, \qquad 其他\end{cases}$

$f_{Y|X}(y\mid x=\frac{1}{2})=\begin{cases}\frac{32}{15}y,\frac{1}{4}<y<1\\0, \qquad 其他\end{cases}$

(2) $P\{Y\ge\frac{3}{4}\mid X=\frac{1}{2}\}=\int_{\frac{3}{4}}^{+\infty}f_{Y|X}(y\mid x=\frac{1}{2})dy=\int_{\frac{3}{4}}^{1}\frac{32}{15}ydy=\frac{7}{15}$

7. (1) 设两个相互独立的随机变量 X和Y 分别服从正态分布 $N(0,1)$和$N(1,1)$，则 ________.

A. $P\{X+Y\le 0\}=\dfrac{1}{2}$　　　　B. $P\{X+Y\le 1\}=\dfrac{1}{2}$

C. $P\{X-Y\le 0\}=\dfrac{1}{2}$　　　　D. $P\{X-Y\le 1\}=\dfrac{1}{2}$

(2) 设随机变量 X_1,X_2,X_3 相互独立，且具有相同的概率分布，$P\{X_i=1\}=p,P\{X_i=0\}=q(i=1,2,3),p+q=1$，考虑随机变量：

$$Y_1=\begin{cases}1, X_1+X_2\text{为奇数}\\0, X_1+X_2\text{为偶数}\end{cases}\qquad Y_2=\begin{cases}1, X_2+X_3\text{为奇数}\\0, X_2+X_3\text{为偶数}\end{cases}$$

则 Y_1Y_2 的分布律为________.

A. $Y_1Y_2\sim\begin{bmatrix}0 & 1\\1-pq & pq\end{bmatrix}$　　　　B. $Y_1Y_2\sim\begin{bmatrix}0 & 1\\pq & 1-pq\end{bmatrix}$

C. $Y_1Y_2\sim\begin{bmatrix}0 & 1\\p & q\end{bmatrix}$　　　　D. $Y_1Y_2\sim\begin{bmatrix}0 & 1\\q & p\end{bmatrix}$

解：(1) B 因为 $X+Y\sim N(1,2),X-Y\sim N(-1,2)$，利用正态分布的几何意义可知：对于 $X\sim N(\mu,\sigma^2)$，有 $P\{X\le\mu\}=\dfrac{1}{2}$，所以 B 项正确.

(2) A 根据 Y_1和Y_2 的取值情况知，Y_1Y_2 只可能取 0,1 两个值，因此只需求出 $P\{Y_1Y_2=0\}$和$P\{Y_1Y_2=1\}$ 即可. 而

$$\begin{aligned}&P\{Y_1Y_2=1\}=P\{Y_1=1,Y_2=1\}=P\{X_1+X_2\text{为奇数},X_2+X_3\text{为奇数}\}\\&=P\{X_1=0,X_2=1,X_3=0\}+P\{X_1=1,X_2=0,X_3=1\}\\&=pq^2+qp^2=pq\end{aligned}$$

所以 $P\{Y_1Y_2=0\}=1-P\{Y_1Y_2=1\}=1-pq$

➢ ***第三章巩固练习题***

1. 设二维随机变量(X,Y)的概率分布为

$X \backslash Y$	0	1
0	0.4	a
1	b	0.1

已知随机事件与相互独立，则________.

A. $a=0.2,b=0.3$　　B. $a=0.4,b=0.1$

C. $a=0.3,b=0.2$　　D. $a=0.1,b=0.4$

2. 设随机变量X在区间$(0,1)$上服从均匀分布，在$X=x(0<x<1)$的条件下，随机变量Y在区间上$(0,x)$服从均匀分布，求：

(1) 随机变量X和Y的联合概率密度；

(2) Y的概率密度；

(3) 概率$P\{X+Y>1\}$.

3. 设随机变量X和Y相互独立，它们的密度函数分变为：

$$f_X(x)=\begin{cases}e^{-x},x>0\\0,x\le 0\end{cases},\quad f_Y(y)=\begin{cases}e^{-y},y>0\\0,y\le 0\end{cases}$$

求：(1) (X,Y)的概率密度；(2) $P\{X\le 1|Y>0\}$.

4. 设随机变量X和Y独立同分布，且的概率密度为：

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

记$U=\min\{X,Y\},V=\max\{X,Y\}$，求$(U,V)$的联合分布律.

5. 设(X,Y)是二维随机变量，X的边缘概率密度为$f_X(x)=\begin{cases}3x^2,0<x<1\\0,\quad 其他\end{cases}$，在给

定 $X=x(0<x<1)$ 的条件下 Y 的条件概率密度为：

$$f_{Y|X}(y|x)=\begin{cases}\dfrac{3y^2}{x^3},0<y<x\\0, \qquad\qquad 其他\end{cases}$$

(1) 求 (X,Y) 的概率密度 $f(x,y)$；

(2) 求 Y 的边缘概率密度 $f_Y(y)$

(3) 求 $P\{X>2Y\}$.

6. 设二维随机变量 (X,Y) 的概率密度为：

$$f(x,y)=\begin{cases}2-x-y,0<x<1,0<y<1\\0, \qquad\qquad 其他\end{cases}$$

(1) 求 $P\{X>2Y\}$;

(2) 求 $Z=X+Y$ 的概率密度 $f_Z(z)$.

7. 设二维随机变量 (X,Y) 的概率密度为 $f(x,y)=\begin{cases}e^{-x},0<y<x\\0,其他\end{cases}$

(1) 求条件概率密度 $f_{Y|X}(y|x)$；

(2) 求条件概率 $P\{X\le 1|Y\le 1\}$.

答案：

1. B

2. (1) $f(x,y)=\begin{cases}\dfrac{1}{x},0<y<x<1\\0, \qquad 其他\end{cases}$；(2) $f_Y(y)=\begin{cases}-\ln y,0<y<1\\0, \qquad 其他\end{cases}$；(3) $1-\ln 2$.

3. (1) $f(x,y)=\begin{cases}e^{-(x+y)},x>0,y>0\\0, \qquad 其他\end{cases}$；(2) $1-e^{-1}$

4. (U,V) 的联合分布律为：

$U \backslash V$	1	2
1	$\frac{4}{9}$	0
2	$\frac{4}{9}$	$\frac{1}{9}$

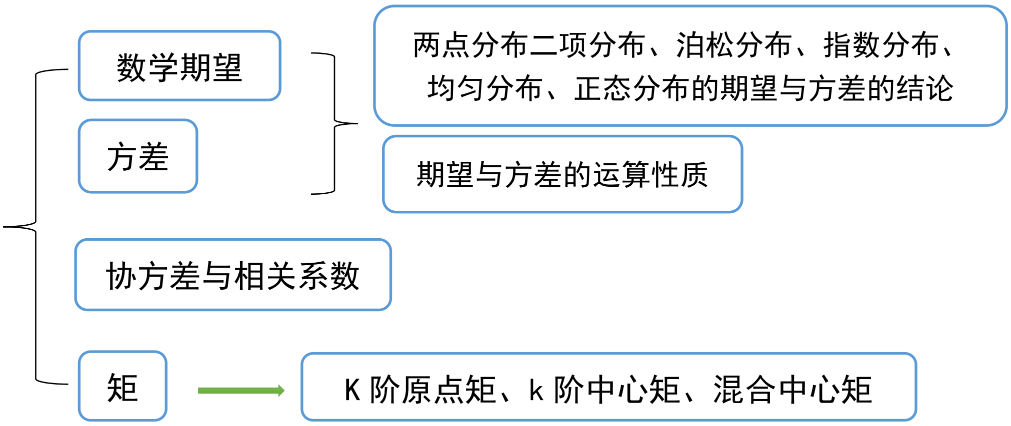
5. (1) $f(x,y)=\begin{cases}\dfrac{9y^2}{x},0<y<x,0<x<1\\ 0, \quad\quad 其他\end{cases}$；(2) $f_Y(y)=\begin{cases}-9y^2\ln y,0<y<1\\ 0, \quad\quad 其他\end{cases}$；(3) $\dfrac{1}{8}$.

6. (1) $\dfrac{7}{24}$；(2) $f_Z(z)=\begin{cases}z(2-z), & 0<z<1\\ (2-z)^2, & 1\le z<2\\ 0, & 其他\end{cases}$

7. (1) $f_{Y|X}(y|x)=\begin{cases}\dfrac{1}{x},0<y<x\\ 0, \quad\quad 其他\end{cases}$；(2) $\dfrac{e-2}{e-1}$

第四章 随机变量的数字特征

➢ ***章节知识体系***



➢ ***重要术语及主题***

期望与方差的性质、离散型与连续型中常见分布的期望与方差、协方差与相关系数、不同类型的矩

➢ ***典型例题***

1. 推导两点分布、二项分布、泊松分布、均匀分布、指数分布、正态分布的期望与方差.

解：(1) 两点分布：

X	0	1
P	$1-p$	p

所以 $E(X)=(1-p)\times 0+1\times p=p, D(X)=E(X^2)-E^2(X)=p-p^2=p(1-p)$

(2) 二项分布：

二项分布可以看成是 $X_1,X_2,\ldots,X_n(n>1)$ 的和，因此有：

$$E(X)=E(X_1+X_2+\cdots+X_n)=np, D(X)=D(X_1+X_2+\cdots+X_n)=np(1-p)$$

(3) 泊松分布：

泊松分布的定义： $P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!}(k=0,1,2,\ldots)$

于是 $E(X)=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k e^{-\lambda}}{k!}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}\cdot e^{\lambda}$

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

所以 $D(X)=E(X^2)-E^2(X)=\lambda^2+\lambda-\lambda^2=\lambda$

(4) **均匀分布**：

均匀分布的概率密度为：$f(x)=\begin{cases}\frac{1}{b-a}, a<x<b \\ 0, \quad 其他\end{cases}$

则 $E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_a^b\frac{x}{b-a}dx=\frac{a+b}{2}$

因为 $E(X^2)=\int_{-\infty}^{\infty}x^2f(x)dx=\int_a^b\frac{x^2}{b-a}dx=\frac{1}{3}(b^2+ab+a^2)$

所以 $D(X)=E(X^2)-E^2(X)=\frac{1}{3}(b^2+ab+a^2)-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}$

(5) **指数分布**：

指数分布的概率密度为：$f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, x>0 \\ 0, \quad 其他\end{cases}$

所以 $E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_0^{\infty}\frac{x}{\theta}e^{-\frac{x}{\theta}}dx=\theta$

又因为 $E(X^2)=\int_{-\infty}^{\infty}x^2f(x)dx=\int_0^{\infty}\frac{x^2}{\theta}e^{-\frac{x}{\theta}}dx=2\theta^2$

所以 $D(X)=E(X^2)-E^2(X)=2\theta^2-\theta^2=\theta^2$

(6) **正态分布**：

正态分布的概率密度为：$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}(-\infty<x<+\infty)$

所以 $E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{\infty}\frac{x}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\mu(令t=\frac{x-\mu}{\sigma})$

对于任意正态分布，均可转化为标准正态分布 $\frac{X-\mu}{\sigma} \sim N(0,1)$

也即 $D(\frac{X-\mu}{\sigma})=1$，由方差的运算性质，得 $D(X)=\sigma^2$

点评：注意公式 $D(X)=E(X^2)-E^2(X)$ 的应用

2. 设二维随机变量 (X,Y) 的联合概率密度为 $f(x,y)=\begin{cases} x+y, 0\le x\le 1, 0\le y\le 1 \\ 0, \quad\quad 其他 \end{cases}$，

求 $E(XY), E(X), E(Y)$.

解： 由公式 $E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$ 得

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy=\int_0^1\int_0^1 xy(x+y)dxdy=\frac{1}{3}$$

$$E(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy=\int_0^1\int_0^1 x(x+y)dxdy=\frac{7}{12}$$

$$E(Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy=\int_0^1\int_0^1 y(x+y)dxdy=\frac{7}{12}$$

点评：本题也可以先求出 X 和 Y 各自的边缘概率密度再利用期望定义求解

3. 设连续型随机变量 X_1, X_2 相互独立且方差均存在，X_1, X_2 的概率密度分别为 $f_1(x), f_2(x)$，随机变量 Y_1 的概率密度为 $f_{Y_1}(y)=\frac{1}{2}[f_1(y)+f_2(y)]$，随机变量 Y_2 为 $Y_2=\frac{1}{2}[X_1+X_2]$，则______.

A. $E(Y_1)>E(Y_2), D(Y_1)>D(Y_2)$　　　　B. $E(Y_1)=E(Y_2), D(Y_1)=D(Y_2)$

C. $E(Y_1)=E(Y_2), D(Y_1)<D(Y_2)$　　　　D. $E(Y_1)=E(Y_2), D(Y_1)>D(Y_2)$

解： D

$$E(Y_1)=\int_{-\infty}^{\infty} yf_{Y_1}(y)dy=\frac{1}{2}[\int_{-\infty}^{\infty} yf_1(\mathrm{y})dy+\int_{-\infty}^{\infty} yf_2(y)dy]=\frac{1}{2}[E(X_1)+E(X_2)]$$

$$E(Y_2)=\frac{1}{2}[E(X_1)+E(X_2)], 所以 E(Y_1)=E(Y_2)$$

$$E(Y_1^2)=\int_{-\infty}^{\infty} y^2 f_{Y_1}(y)dy=\frac{1}{2}[\int_{-\infty}^{\infty} y^2 f_1(\mathrm{y})dy+\int_{-\infty}^{\infty} y^2 f_2(y)dy]=\frac{1}{2}[E(X_1^2)+E(X_2^2)]$$

又因为$D(Y_1)=E(Y_1^2)-E^2(Y_1), D(Y_2)=E(Y_2^2)-E^2(Y_2)$

而$D(Y_1)-D(Y_2)=E(Y_1^2)-E(Y_2^2)=\frac{1}{2}E(X_1^2)+\frac{1}{2}E(X_2^2)-\frac{1}{4}E[(X_1+X_2)^2]$

$$=\frac{1}{4}E(X_1^2+X_2^2-2X_1X_2)=\frac{1}{4}E[(X_1-X_2)^2]>0$$

所以$D(Y_1)>D(Y_2)$

4. 某箱装有 100 件产品，其中一、二和三等品分别有 80 件，10 件和 10 件，现从中随机中抽取一件，记$X_i=\begin{cases}1, 抽到i等品\\ 0, \quad 其他\end{cases}(i=1,2,3)$，求：

(1) 随机变量X_1, X_2的联合分布；

(2) 随机变量X_1, X_2的相关系数.

解：(1) 设事件A_i表示“抽到i等品”，易知A_1, A_2, A_3相互独立，所以有：

$P(A_1)=0.8, P(A_2)=P(A_3)=0.1$，于是：

$$P\{X_1=0, X_2=0\}=P(A_3)=0.1, P\{X_1=0, X_2=1\}=P(A_2)=0.1$$
$$P\{X_1=1, X_2=0\}=P(A_1)=0.8, P\{X_1=1, X_2=1\}=\varnothing=0$$

所以X_1, X_2的联合分布律为：

X_2 \ X_1	0	1
0	0.1	0.8
1	0.1	0

(2)

$$E(X_1)=0.8, E(X_2)=0.1, D(X_1)=0.16, D(X_2)=0.09$$
$$E(X_1X_2)=0, Cov(X_1,X_2)=E(X_1X_2)-E(X_1)E(X_2)=-0.08$$
$$所以\rho=\frac{Cov(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}=-\frac{2}{3}$$

5. 已知随机变量 $X\sim N(1,3^2),Y\sim N(0,4^2)$，它们的相关系数 $\rho_{XY}=-\frac{1}{2}$，设 $Z=\frac{X}{3}+\frac{Y}{2}$.

(1) 求 Z 的数学期望与方差；

(2) 求 X,Z 的相关系数 ρ_{XZ}.

解： (1) $E(Z)=\frac{1}{3}E(X)+\frac{1}{2}(Y)=\frac{1}{3}$， $D(Z)=\frac{1}{3^2}D(X)+\frac{1}{2^2}D(Y)+2Cov(\frac{X}{3},\frac{Y}{2})=3$

(2)

$$Cov(X,Z)=Cov(X,\frac{X}{3}+\frac{Y}{2})=Cov(X,\frac{X}{3})+Cov(X,\frac{Y}{2})=\frac{1}{3}Cov(X,X)+\frac{1}{2}Cov(X,Y)=0$$

点评：本题较为简单，其中体现了协方差即方差的一些重要运算性质：

$$Cov(X,Y_1+Y_2)=Cov(X,Y_1)+Cov(X,Y_2)$$
$$Cov(aX,bY)=abCov(X,Y)$$
$$D(aX,bY)=a^2D(X)+b^2D(Y)+2abCov(X,Y)$$

➢ ***第四章巩固提高练习***

1. 设随机变量 X 的分布函数为 $f(x)\begin{cases}\frac{1}{\pi\sqrt{1-x^2}}, |x|<1\\ 0, \quad |x|\geq 1\end{cases}$，则数学期望 $E(X)$ 和方差 $D(X)$ 分别为______、_______.

2. 设随机变量 X 在区间 $[-1,2]$ 上服从均匀分布，随机变量 $Y=\begin{cases}1, X>0\\ 0, X=0\\ -1, X<0\end{cases}$，则方差 $D(Y)=$______.

3. (1) 将一枚硬币重复掷 n 次，以 X和Y 分别表示正面向上和反面向上的次数，则 X和Y 的相关系数为______.

(2) 设随机变量 X和Y 的相关系数为 0.9，若 $Z=X-0.4$，则 Y和Z 的相关系数为_______.

(3) 设随机变量 X 在 $[a,b]$ 上服从均匀分布，则 X 的 k 阶原点矩为______，三阶中心矩为______.

4. (1) 设随机变量 $X\sim N(0,1),Y\sim N(1,4)$，且相关系数为 1，则______.

A. $P\{Y=-2X-1\}=1$　　B. $P\{Y=2X-1\}=1$

C. $P\{Y=-2X+1\}=1$　　D. $P\{Y=2X+1\}=1$

(2) 设随机变量 X和Y 独立同分布，记 $U=X-Y,V=X+Y$，则随机变量 U和V 必然_______.

A.不独立　　B.独立　　C.相关系数不为零　　D.相关系数为零

(3) 设随机变量 X和Y 的方差存在且不等于零，则是的______.

A.不相关的充分条件，但不是必要条件

B.独立的必要条件，但不是充分条件

C.不相关的充要条件

D.独立的充要条件

(4) 设随机变量 $X_1,X_2,\ldots,X_n(n>1)$ 独立同分布，且其方差为 $\sigma^2>0$，令 $Y=\frac{1}{n}\sum_{i=1}^{n}X_i$，则_______.

A. $Cov(X_1,Y)=\frac{\sigma^2}{n}$　　B. $Cov(X_1,Y)=\sigma^2$

C. $D(X_1+Y)=\frac{n+2}{n}\sigma^2$　　D. $D(X_1-Y)=\frac{n+1}{n}\sigma^2$

(5) 设 X 是一随机变量，$E(X)=\mu,D(X)=\sigma^2$（$\mu,\sigma>0$ 均为常数)，则对任意常数 c，必有_______.

A. $E[(X-c)^2]=E(X^2)-c^2$　　B. $E[(X-c)^2]=E[(X-\mu)^2]$

C. $E[(X-c)^2]\le E[(X-\mu)^2]$　　D. $E[(X-c)^2]\ge E[(X-\mu)^2]$

(6) 设随机变量 X和Y 相互独立，且 $E(X),E(Y)$ 存在，记 $U=\max\{X,Y\},V=\min\{X,Y\}$，则 $E(UV)=$______.

A. $E(U)\cdot E(V)$　　B. $E(X)\cdot E(Y)$　　C. $E(U)\cdot E(Y)$　　D. $E(X)\cdot E(V)$

5. 随机变量(X,Y)的联合概率密度为$f(x,y)=\begin{cases}\frac{1}{8}(x+y), 0\le x\le 2, 0\le y\le 2\\ 0, \quad\quad 其他\end{cases}$，求

$E(X),E(Y),Cov(X,Y),D(X+Y),\rho_{XY}$.

答案：

1. $E(X)=0$，$D(X)=\dfrac{1}{2}$

2. $\dfrac{8}{9}$

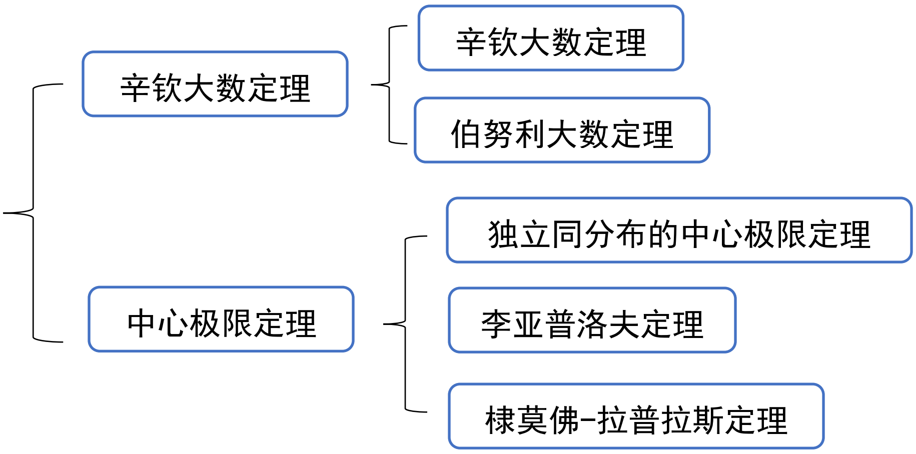
3. (1) -1；(2) 0.9；(3) $E(X^k)=\dfrac{1}{k+1}\cdot\dfrac{b^{k+1}-a^{k+1}}{b-a}, E[(X-EX)^3]=0$

4. (1) D；(2) D；(3) C；(4) B；(5) D；(6) B

5. $E(X)=\dfrac{7}{6}, E(Y)=\dfrac{7}{6}, Cov(X,Y)=-\dfrac{1}{36}, D(X+Y)=\dfrac{5}{9}, \rho_{XY}=-\dfrac{1}{11}$

第五章　大数定律及中心极限定理

➢ *章节知识体系*



➢ *重要主题及术语*

参见上述知识体系导图

➢ *典型例题*

1. 设随机变量 $X_1,\ldots,X_n,\ldots$ 是独立同分布的随机变量，其分布函数为

$F(\mathrm{x})=\mathrm{A}+\dfrac{1}{\pi}\arctan\dfrac{x}{B}$，其中 $B\neq 0$，则辛钦大数定律对此序列________.

A.适用　B.当常数 A,B 取适当值得时候适用　　C.无法判断　D.不适用

解： 辛钦大数定律成立的条件有两个：一是随机变量序列独立同分布；二是数学期望 $E(\mathrm{X}_n)$ ，$n=1,2,\ldots$ 要存在. 本题的第一个条件满足，而第二个条件不满足（请读者自行计算，期望的积分是发散的），所以选 D.

2. 设 $X\sim U[-1,\mathrm{b}]$，若由切比雪夫不等式有 $P\{|\mathrm{X}-1|<\varepsilon\}\geq\dfrac{2}{3}$，则 $b=$_______； ε =________.

解： 因为 $E(\mathrm{X})=\dfrac{b-1}{2},D(\mathrm{X})=\dfrac{(b+1)^2}{12}$，所以 $\dfrac{b-1}{2}=1,1-\dfrac{\dfrac{(b+1)^2}{12}}{\varepsilon^2}=\dfrac{2}{3}$，从而解得

$b=3,\varepsilon=2$

3. 某单位设置一电话总机，共有 200 个电话分机，设每个电话分机有 5%的时间要使用外线通话，假设每个分机是否使用外线通话是相互独立的，问总机多少外线才能以 90%的概率保证每个外机要使用外线时同时使用.（已知 $\phi(1.28)=0.9$）

解： 设同时使用外线的分机台数为 X，则 $X\sim B(n,p)$，其中 $n=200,p=0.05,np=10,\sqrt{np(1-p)}=3.08$，又设该单位安装 N 条外线，依题意，求 $P\{\mathrm{X}\le\mathrm{N}\}\ge0.9$ 的最小 N，由棣莫佛-拉普拉斯中心极限定理：

$$P\{X\le N\}=\mathrm{P}\{\frac{X-np}{\sqrt{np(1-p)}}\le\frac{N-np}{\sqrt{np(1-p)}}\}\approx\Phi(\frac{N-10}{3.08})$$

所以 $\frac{N-10}{3.08}\ge1.28$，即 $N\ge13.94$，所以至少要装 14 条线.

4. 测量某物体的长度，由于存在测量误差，每次测得的长度只能是近似值. 现在进行多次测量，然后选取这些测量值的平均值作为实际长度的估计值，假定 n 个测量值 $X_1,\mathrm{X}_2,\ldots,X_n$ 是独立同分布的的随机变量，具有共同的期望 μ（即实际长度）及方差 σ=1，试问要以 95%的把握可以确信其估计值精确到 ±0.2 以内，必须测量多少次？（已知 $\Phi(1.96)=0.975$）

解： 考虑用中心极限定理来求，则有：

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\le0.2\right\}=P\left\{\left|\frac{\frac{1}{n}\sum_{i=1}^{n}X_i-\mu}{\frac{\sigma}{\sqrt{n}}}\right|\le\frac{0.2\sqrt{n}}{\sigma}\right\}\approx2\Phi(\frac{0.2\sqrt{n}}{\sigma})-1$$

由题意 $2\Phi(0.2\sqrt{n})-1=0.95$，即 $\Phi(0.2\sqrt{n})=0.975$，所以 $0.2\sqrt{n}=1.96$，解得，$n=96.04$，故至少要测量 97 次以上.

5. 在一家保险公司里有 10000 个同龄又同阶层的人参加保险，每人每年付 12 元的保险费。在一年内一个人死亡的概率为 0.006，死亡后家属可向公司领取

1000 元. 求：

(1) 保险公司亏本的概率；

(2) 保险公司一年的利润不少于 60000 元概率.

解：(1) 设参加保险的 10000 人中一年死亡的人数为 X，则有：

$$X \sim B(10000, 0.006), E(X)=60, D(X) \approx 7.72^2$$

公司亏本，则转化为求 $P\{1000X - 120000 > 0\} = P\{X > 120\}$ 根据中心极限定理，有 $X \sim N(60, 7.72^2)$，从而公司亏本的概率为：

$$P\{X > 120\} = 1 - P\left\{\frac{X-60}{7.72} \le \frac{120-60}{7.72}\right\} = 1 - P\left\{\frac{X-60}{7.72} \le 7.77\right\} \approx 1 - \Phi(7.77) \approx 0$$

(2) 由题意所求概率为

$$P\{0 \le X \le 60\} = P\left\{\frac{0-60}{7.72} \le \frac{X-60}{7.72} \le \frac{60-60}{7.72}\right\} = P\left\{-7.77 \le \frac{X-60}{7.72} \le 0\right\} \approx \Phi(0) - \Phi(-0.7.77) \approx 0.5$$

6. 对一个学生而言，来参加家长会的人数是随机变量，设一个学生的无家长、1 名家长、2 名家长来参加会议的概率分别是 0.05、0.8、0.15. 若学校共有 400 名学生，已知各学生参加会的家长人数相互独立，且服从同一分布.

(1) 求参加会议的家长人数 X 超过 450 的概率；

(2) 求有 1 名家长来参加会议的学生人数不超过 340 的概率.

解：(1) 以 $X_k(k=1,2,...,400)$ 记第k个学生来参加家长会的家长人数，则 X_k 的分布律如下：

X_k	0	1	2
p_k	0.05	0.8	0.15

易知 $E(X_k)=1.1, D(X_k)=0.19, k=1,2,...,400.$ 而 $X=\sum_{k=1}^{400} X_k$.由中心极限定理，

随机变量 $\dfrac{\sum_{k=1}^{400} X_k - 400\times 1.1}{\sqrt{400}\sqrt{0.19}} = \dfrac{X - 400\times 1.1}{\sqrt{400}\sqrt{0.19}}$ 近似服从正态分布 $N(0,1)$，于是

$$P\{X>450\}=P\left\{\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}>\frac{450-400\times1.1}{\sqrt{400}\sqrt{0.19}}\right\}=1-P\left\{\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}\le1.147\right\}\approx1-\Phi(1.147)=0.1251$$

(2) 以 Y 记有 1 名家长参加会议的学生人数，则 $Y\sim B(400,0.8)$，由棣莫佛-拉普拉斯定理，有

$$P\{Y\le340\}=P\left\{\frac{Y-400\times0.8}{\sqrt{400\times0.8\times0.2}}\le\frac{340-400\times0.8}{\sqrt{400\times0.8\times0.2}}\right\}=P\left\{\frac{Y-400\times0.8}{\sqrt{400\times0.8\times0.2}}\le2.5\right\}\approx\Phi(2.5)=0.9938$$

➢ *第五章巩固练习题*

1. 设随机变量 X 和 Y 的数学期望分别为-2 和 2，方差分别为 1 和 4，而相关系数为-0.5，则根据切比雪夫不等式$P\{|X+Y|\ge6\}\le$__________.
2. 在每次试验中事件 A 发生的概率为 0.5，利用切比雪夫不等式，则在 1000 次独立重复试验中事件 A 发生的次数在 450~550 之间的概率为_______.
3. 设随机变量的 X 的概率密度为$\begin{cases}\frac{1}{2}x^2e^{-x}, & x>0\\ 0, & x\le0\end{cases}$，试利用切比雪夫不等式估计概率$P\{1<X<5\}>$_________.
4. 设$\Phi(x)$为标准正态分布函数，$X_i=\begin{cases}0, & A\text{不发生}\\ 1, & A\text{发生}\end{cases}(i=1,2,\dots,100)$，且$P(A)=0.8$，$X_1,X_2,\dots,X_{100}$相互独立．令$\mathrm{Y}=\sum_{\mathrm{i}=1}^{100}\mathrm{X_i}$，则由中心极限定理知$Y$的的分布函数$F(y)$近似于_______.

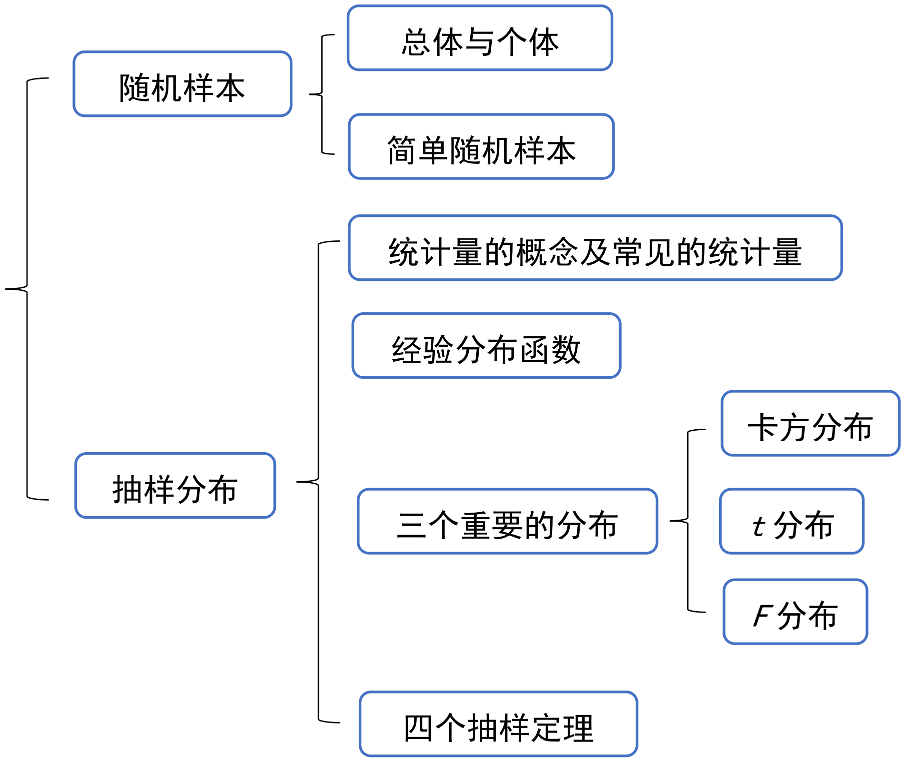
   A. $\Phi(\mathrm{y})$　　B. $\Phi(\frac{y-80}{4})$　　C. $\Phi(16y+8)$　　D. $\Phi(4y+80)$
5. 一生产线生产的产品成箱包装，每箱的重量是随机的，假设每箱平均重 50 千克，标准差为 5 千克，若用最大载重为 5 吨的汽车承运，使用中心极限定理说明每辆车最多能装多少箱，才能保证不超载的概率大于 0.977.（已知$\Phi(x)$为标准正态分布函数，且$\Phi(2)=0.977$）

答案：

1.$\frac{1}{12}$　　2. 0.9　　3. $\frac{1}{4}$　　4.B　　5.98 箱

第六章　样本及抽样分布

➢ ***章节知识体系***



➢ ***重要术语及主题***

总体、简单随机样本、统计量、常见统计量（样本均值、样本方差、样本 k 阶矩、样本 k 阶中心矩）、χ^2 分布和 t 分布及 F 分布的定义和某些特殊的性质、上 α 分位点、抽样定理中的重要结论

➢ ***典型例题***

1. 设 $X_1,X_2,\dots,X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本，$\bar{X}$ 为样本均值，记

$$S_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,\qquad S_2^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,\qquad S_3^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2,$$

$S_4^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$，则服从自由度为 $n-1$ 的 t 分布的随机变量是________.

A. $t=\dfrac{\bar{X}-\mu}{\frac{S_1}{\sqrt{n-1}}}$　　B. $t=\dfrac{\bar{X}-\mu}{\frac{S_2}{\sqrt{n-1}}}$　　C. $t=\dfrac{\bar{X}-\mu}{\frac{S_3}{\sqrt{n}}}$　　D. $t=\dfrac{\bar{X}-\mu}{\frac{S_4}{\sqrt{n}}}$

解： B　因为 $X_1,X_2,...,X_n$ 服从 $N(\mu,\sigma^2)$ 分布，所以有 $t=\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$，

$\sum\limits_{i=1}^{n}\dfrac{(X_i-\bar{X})^2}{\sigma^2}\sim\chi^2(n-1)$，所以 $\dfrac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2}}}=\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}}\sim t(n-1)$，

所以 $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}}=\dfrac{(\bar{X}-\mu)}{\sqrt{\frac{1}{n}\cdot\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}}=\dfrac{\bar{X}-\mu}{\frac{S_2}{\sqrt{n-1}}}$，B 正确.

2. 设 $X_1,X_2,...,X_n(n\geq 2)$ 是来自总体 $N(\mu,1)$ 的简单随机样本，记 $\bar{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$，

则下列结论不正确的是________.

A. $\sum\limits_{i=1}^{n}(X_i-\mu)^2$ 服从 χ^2 分布　　B. $2(X_n-X_1)^2$ 服从 χ^2 分布

C. $\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 服从 χ^2 分布　　D. $n(\bar{X}-\mu)^2$ 服从 χ^2 分布

解： B　对 A 项，$X_i-\mu\sim(0,1)$，所以 $\sum\limits_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$，正确；C 项，

$\sum\limits_{i=1}^{n}(X_i-\bar{X})^2=(n-1)S\sim\chi^2(n-1)$，正确；D 项，$\bar{X}\sim N(\mu,\dfrac{1}{n})$，标准化后得

$n(\bar{X}-\mu)^2\sim\chi^2(1)$，

正确；对于 B 项，因为 $X_n-X_1\sim N(0,2)$,所以 $\dfrac{X_n-X_1}{\sqrt{2}}\sim N(0,1)$，则

$\dfrac{(X_n-X_1)^2}{2}\sim\chi^2(1)$，所以 B 错误.

3. (1) 设 $X_1, X_2, ..., X_n$ 是来自总体 $N(0,\sigma^2)$ 的简单随机样本，记 $\bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i, 1 \le k \le n$，则 $Cov(\bar{X}_k, \bar{X}_{k+1}) =$ ________.

(2) 设随机变量，$X \sim t(n), Y \sim F(1,n)$, 给定 $\alpha(0<\alpha<0.5)$，常数 c 满足 $P\{X>c\}=\alpha$，则 $P\{Y>c^2\}=$ _______.

解：(1)本题考查协方差的性质.

$$Cov(\bar{X}_k, \bar{X}_{k+1}) = \frac{1}{k(k+1)} Cov(\sum_{i=1}^{k} X_i, \sum_{i=1}^{k} X_i + X_{k+1}) = \frac{1}{k(k+1)} Cov(\sum_{i=1}^{k} X_i, \sum_{i=1}^{k} X_i) = \frac{1}{k(k+1)} D(\sum_{i=1}^{k} X_i) = \frac{\sigma^2}{k+1}$$

(2) 因为 $X \sim t(n), Y \sim F(1,n)$，则 $X^2 \sim F(1,n)$ 与 Y 同分布，所以：

$$P\{Y>c^2\} = P\{X^2>c^2\} = P\{X>c\} + P\{X<-c\} = 2P\{X>c\} = 2\alpha$$

4. 设 $X_1, X_2, ..., X_9$ 是总体的一个简单的随机样本，X 服从正态分布 $N(\mu,\sigma^2)$，

$$Y_1 = \frac{1}{6}(X_1 + \ldots + X_6), Y_2 = \frac{1}{3}(X_7 + X_8 + X_9), S^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - Y_2)^2, T = \frac{\sqrt{2}(Y_1 - Y_2)}{S},$$

证明：$T \sim t(2)$.

证明：因为 $X \sim N(\mu,\sigma^2)$，$X_i \sim N(\mu,\sigma^2)$ 所以，$Y_1 \sim N(\mu, \frac{\sigma^2}{6}), Y_2 \sim N(\mu, \frac{\sigma^2}{3})$，

故 $Y_1 - Y_2 \sim N(0, \frac{\sigma^2}{2})$，因此有 $\frac{Y_1 - Y_2}{\sigma/\sqrt{2}} \sim N(0,1)$

又由于 $S^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - Y_2)^2 = \frac{1}{3-1}\sum_{i=7}^{9}(X_i - Y_2)^2$，而 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

所以 $\frac{2S^2}{\sigma^2} \sim \chi^2(2)$. 因为 Y_2 和 S^2 相互独立，而且 Y_1 与 Y_2，Y_1 与 S^2 也相互独立，

所以 $Y_1 - Y_2$ 与 S^2 相互独立，则 $\frac{\sqrt{2}(Y_1 - Y_2)}{\sigma}$ 与 $\frac{2S^2}{\sigma^2}$ 也相互独立.

于是，$T=\dfrac{\sqrt{2}(Y_1-Y_2)}{S}=\dfrac{\dfrac{\sqrt{2}(Y_1-Y_2)}{\sigma}}{\sqrt{\dfrac{2S^2}{\sigma^2}}}\sim t(2)$，证毕

5. 设 X_1,X_2,X_3,X_4 是来自正态总体 $N(0,2^2)$ 的简单随机样本，某统计量满足 $X=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$,则当 $a=$_______,$b=$_______时，统计量 X 服从 χ^2 分布，其自由度为_______.

解： 令 $Y_1=X_1-2X_2$，则 $\dfrac{Y_1}{\sqrt{20}}\sim N(0,1)$；同样令 $Y_2=3X_3-4X_4$，则 $\dfrac{Y_2}{10}\sim N(0,1)$，

此时 $\dfrac{Y_1^2}{20}+\dfrac{Y_2^2}{100}\sim\chi^2(2)$，对比可知 $a=\dfrac{1}{20},b=\dfrac{1}{100},n=2$.

6. 在总体 $N(12,4)$ 中随机抽一容量为 5 的样本 X_1,X_2,X_3,X_4,X_5.

(1) 求样本均值与总体均值之差绝对值大于 1 的概率.

(2) 求概率 $P\{\max(X_1,X_2,X_3,X_4,X_5)>15\}$.

(3) 求概率 $P\{\min(X_1,X_2,X_3,X_4,X_5)<10\}$.

解： (1) 由题意 $\bar{X}\sim N(12,\dfrac{2}{5})$，从而

$$P\{|X-12|>1\}=\mathrm{P}\left\{\left|\frac{X-12}{\frac{2}{\sqrt{5}}}\right|>\frac{\sqrt{5}}{2}\right\}=2-2\Phi(1.12)=0.2628$$

(2) $P\{\max(X_1,X_2,X_3,X_4,X_5)>15\}$

$=1-P\{X_1\le 15,X_2\le 15,X_3\le 15,X_4\le 15,X_5\le 15\}$

$=1-\prod_{i=1}^{5}P\{X_i\le 15\}=1-\prod_{i=1}^{5}P\{\dfrac{X_i-12}{2}\le\dfrac{15-12}{2}\}$

$=1-[\Phi(1.5)]^5=1-0.9332^5=0.2923$

(3) $P\{\min(X_1,X_2,X_3,X_4,X_5)<10\}$

$$=1-P\{X_1\ge 10, X_2\ge 10, X_3\ge 10, X_4\ge 10, X_5\ge 10\}$$

$$=1-\prod_{i=1}^{5}P\{X_i\ge 10\}=1-\prod_{i=1}^{5}P\{\frac{X_i-10}{2}\ge\frac{15-10}{2}\}$$

$$=1-[1-\Phi(-1)]^5=1-0.8413^5=0.5785$$

7. 设总体 X 服从正态分布 $N(\mu,\sigma^2)$. 从该整体中抽取简单的随机样本 $X_1,X_2,\ldots,X_{2n}(n\ge 2)$，其样本均值为 $\bar{X}=\frac{1}{2n}\sum_{i=1}^{2n}X_i$，试求统计量

$$Y=\sum_{i=1}^{n}(X_i+X_{i+1}\text{-}2\bar{X})$$

的数学期望 $E(Y)$.

解： 由已知条件 $X_1,X_2,\ldots,X_{2n}$ 均服从 $N(\mu,\sigma^2)$ 且相互独立，所以

$(X_1+X_{n+1}),(X_2+X_{n+2}),\ldots,(X_n+X_{2n})$ 均相互独立且服从 $N(2\mu,2\sigma^2)$

其样本均值为 $\frac{1}{n}\sum_{i=1}^{n}(X_i+X_{n+i})=\frac{1}{n}\sum_{i=1}^{2n}X_i=2\bar{X}$

样本方差为 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i+X_{n+i}-2\bar{X})^2=\frac{1}{n-1}Y$

因为 $E(S^2)=\sigma^2$，所以 $E(\frac{1}{n-1}Y)=2\sigma^2$，解得 $E(Y)=2(n-1)\sigma^2$

点评: 本题用的方法十分巧妙，需要读者细心观察。但是这种方法不易想到，也可以利用直接展开的方法计算，请读者自行完成.

➢ *第六章巩固练习题*

1. 设随机变量 X 和 Y 都服从标椎正态分布，则______.

A. $X+Y$ 服从正态分布　　B. X^2+Y^2 服从 χ^2 分布

C. X^2 和 Y^2 都服从 χ^2 分布　　D. $\dfrac{X^2}{Y^2}$ 服从 F 分布

2. 设总体 X 服从正态分布 $N(0,2^2)$，而 $X_1,X_2,...,X_{15}$ 是来自总体 X 的简单随机样本，则随机变量 $Y=\dfrac{X_1^2+\cdots+X_{10}^2}{2(X_{11}^2+\cdots+X_{15}^2)}$ 服从_____分布，参数为_______.

3. X_1,X_2,X_3,X_4 为来自总体 $X\sim N(1,\sigma^2)$ 的简单随机样本，则统计量 $\dfrac{X_1-X_2}{|X_3+X_4-2|}$ 的分布为_______.

A. $N(0,1)$　　B. $t(1)$　　C. $\chi^2(1)$　　D. $F(1,1)$

4. 设 $X_1,X_2,...,X_5$ 是取自正态分布 $N(0,\sigma^2)$ 的一个简单随机样本，若 $\dfrac{a(X_1+X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$ 服从 t 分布，则 $a=$________.

5. 设在总体 $N(\mu,\sigma^2)$ 中抽取一个容量为 16 的样本，则 $P\{\dfrac{S^2}{\sigma^2}\le 1.664\}=$______.

（已知 $\chi^2_{0.05}(15)=24.996$）

6. 设 $X_1,X_2,...,X_m$ 是来自二项分布总体 $B(n,p)$ 的简单随机样本，$\bar{X}$ 和 S^2 分别为样本均值和样本方差，记统计量 $T=\bar{X}-S^2$，则 $E(T)=$________.

7. 设总体 X 服从参数为 $\lambda(\lambda>0)$ 的泊松分布，$X_1,X_2,\ldots,X_n(n\ge 2)$ 为来自总体的简单随机样本，则对于统计量 $T_1=\dfrac{1}{n}\sum_{i=1}^{n}X_i, T_2=\dfrac{1}{n-1}\sum_{i=1}^{n-1}X_i+\dfrac{1}{n}X_n$，有_______.

A. $E(T_1)>E(T_2),D(T_1)>D(T_2)$　　B. $E(T_1)>E(T_2),D(T_1)<D(T_2)$

C. $E(T_1)<E(T_2),D(T_1)>D(T_2)$　　D. $E(T_1)<E(T_2),D(T_1)<D(T_2)$

8. 设 $X_1,X_2,\ldots,X_n(n\ge 2)$ 为来自总体的 $N(0,1)$ 的简单随机样本，$\bar{X}$ 为样本均值，

S^2 为样本的方差，则________.

A. $n\overline{X} \sim N(0,1)$　　B. $nS^2 \sim \chi^2(n)$

C. $\dfrac{(n-1)\overline{X}}{S} \sim t(\mathrm{n}-1)$　　D. $\dfrac{(n-1)X_1^2}{\sum_{i=2}^{n} X_i^2} \sim t(\mathrm{n}-1)$

9. 设总体 X 服从正态分布 $N(\mu_1,\sigma^2)$，总体 Y 服从正态分布 $N(\mu_2,\sigma^2)$，$X_1,X_2,\ldots,X_{n_1}$ 和 $Y_1,Y_2,\ldots,Y_{n_2}$ 分别是来自总体 X 和 Y 的简单随机样本，则

$$E\left[\frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2}{n_1+n_2-2}\right]=\underline{\qquad\qquad}.$$

答案：

1.C（没有说明 X 和 Y 是否相互独立）

2. F；（10,5）

3.B（去绝对值）

4. $\sqrt{\dfrac{3}{2}}$（直接按定义求）

5.0.95（利用抽样定理二的结论求解）

6. np^2

7.D（$E(T_1)=\lambda, E(T_2)=\lambda+\dfrac{\lambda}{n}; \mathrm{D}(T_1)=\dfrac{\lambda}{n}, D(T_2)=\dfrac{\lambda}{n-1}+\dfrac{\lambda}{n^2}$）

8.D（按照定义逐一验证可得正确答案）

9. σ^2（利用 $E(S^2)=\sigma^2$）

第七、八章 参数估计与假设检验

➢ *章节知识体系*

- **参数估计**
  - 点估计
    - 矩估计 → 关键：根据未知参数个数写出样本的的 k 阶矩
    - 最大似然估计 → 关键：写出似然函数
  - 估计量的评选标准 → 无偏性、有效性、相合性
  - 区间估计
    - 正态总体均值与方差区间估计
    - 单侧置信区间
    - → 关键：利用第六章的抽样定理找到适合的枢轴量
- **假设检验**
  - 假设检验的有关概念
    - 显著性水平
    - 原假设与备择假设
    - 两类错误与显著性检验
    - 统计量、拒绝域与临界点
    - 双边假设检验与单边假设检验
  - 假设检验的步骤
    - 根据实际问题提出原假设 H_0 与备择假设 H_1
    - 给定显著性水平 α 及样本数据
    - 确定统计量和拒绝域形式，并求出拒绝域
    - 根据样本观测值给出决策（如果未落在拒绝域内则接受原假设）
  - 正态总体均值与方差的假设检验

➢ ***重要术语及主题***

矩估计量、最大似然估计量、估计量的无偏性、枢轴量、参数θ的置信水平为$1-\alpha$的置信区间以及单侧置信（上或下）限、单个正态总体的方差和均值在不同情形下的置信区间与单侧置信限、两个正态总体的方差和均值在不同情形下的置信区间与单侧置信限

原假设、备择假设、检验统计量、单边与双边检验、一个正态总体的参数的检验、两个正态总体均值差与方差比的检验、成对数据的检验

➢ ***典型例题***

1. 设总体 X 的密度函数为 $f(x;\theta)=\begin{cases}\theta x^{\theta-1}, 0<x<1\\ 0 \quad\quad ,\text{其他}\end{cases}(\theta>0)$，则未知参数$\theta$的矩估计量为______.

解： $E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_0^1 x\theta x^{\theta-1}dx=\dfrac{\theta}{\theta+1}$， 令 $E(X)=\bar{X}$

解得 $\hat{\theta}=\dfrac{\bar{X}}{1-\bar{X}}(\bar{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i)$，则$\hat{\theta}$即为参数$\theta$的矩估计量.

点评：求矩估计量的一般步骤如下：

(1) 根据未知参数的个数，求出总体的各阶矩（分为离散型和连续型）；

(2) 总体矩中含有未知参数，故可构造方程或者方程组；

(3) 解方程或方程组，即把未知参数用总体矩表示出来；

(4) 再用样本矩代替总体矩，得到矩估计量.

2. 设总体 X 的密度函数为 $f(x;\lambda)=\begin{cases}\lambda\alpha x^{\alpha-1}e^{-\lambda x^{\alpha}} & \text{若}x>0\\ 0 & \text{若}x\le 0\end{cases}$，其中$\lambda>0$为未知参数，$\alpha>0$是已知常数，$X_1,X_2,\dots,X_n$是来自总体的简单随机样本，求$\lambda$的最大似然估计量$\hat{\lambda}$.

解：由题意可得似然函数为：$L(x_1,x_2,...,x_n;\lambda)=\prod_{i=1}^{n}f(x_i;\lambda)=(\lambda\alpha)^n e^{-\lambda\sum_{i=1}^{n}x_i^\alpha}\prod_{i=1}^{n}x_i^{\alpha-1}$

当 $x_i>0$ 时， $L>0$，有 $\ln L=n\ln(\lambda\alpha)+\ln\prod_{i=1}^{n}x_i^{\alpha-1}-\lambda\sum_{i=1}^{n}x_i^\alpha$，对 λ 求导得：

$\frac{d\ln L}{d\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i^\alpha$，令 $\frac{d\ln L}{d\lambda}=0$，解得 $\hat{\lambda}=\frac{n}{\sum_{i=1}^{n}x_i^\alpha}$

点评：求最大似然估计的步骤如下：

(1) 根据总体的分布写出似然函数 L（这一步是关键）；

(2) （一般情况下）求对数似然函数 $\ln L(\theta)$（因为涉及连乘）；

(3) 对 $\ln L(\theta)$ 求导或者求偏导，令其为零；

(4) 解方程或者方程组，得到未知参数的极大似然估计量.

3. 设某种元件的使用寿命 X 的概率密度为 $f(x;\theta)=\begin{cases}2e^{-2(x-\theta)}, x>\theta\\ 0 \quad\quad x\le\theta\end{cases}$，其中 $\theta>0$ 为未知参数，又设 $x_1,x_2,...,x_n$ 是 X 的一组样本观测值，求参数 θ 的极大似然估计值.

解：由题意知，似然函数为：$L(x_1,x_2,...,x_n;\theta)=\begin{cases}2^n e^{-2\sum_{i=1}^{n}(x_i-\theta)}, x_i>\theta\,(\mathrm{i}=1,2,...,n)\\ 0, \quad\quad x_i\le\theta\end{cases}$

当 $x_i>0$ 时， $L>0$，有 $\ln L=n\ln 2-2\sum_{i=1}^{n}(x_i-\theta)$，而 $\frac{d\ln L}{d\theta}=2n>0$

即似然函数单调增加无极值，但是题目给出了 $\theta<x_i(\mathrm{i}=1,2,...,n)$，所以当 θ 取

$x_1,x_2,...,x_n$ 中的最小值时，似然函数有最大值，所以参数 θ 的极大似然估计值

为 $\hat{\theta}=\min\{x_1,x_2,...,x_n\}$

点评：注意题目中是求极大似然估计值还是估计量.

4. 设 $X\sim N(\mu,\sigma^2)$， $x_1,x_2,...,x_n$ 为一组样本值，求参数 μ,σ^2 的最大似然估计.

解：似然函数为： $L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$

两边取对数，有： $\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln 2\pi\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

似然方程组为： $\begin{cases}\frac{\partial}{\partial\mu}\ln L=\frac{1}{\sigma^2}(\sum_{i=1}^{n}x_i-n\mu)=0\\ \frac{\partial}{\partial\sigma^2}\ln L=-\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$

解得： $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i,\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

因此 μ,σ^2 的最大似然估计量为： $\hat{\mu}=\bar{X},\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

5. 设 $X_1,X_2,\ldots,X_n$ 为取自总体的简单随机样本，为了估计统计量，我们利用统计量 $\hat{\sigma}^2=K\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$ ，则 $K=$ ________ 时， $\hat{\sigma}^2$ 是 σ^2 的无偏估计.

解：由题意得， $E(\hat{\sigma}^2)=\sigma^2$，

又因为 $E[(X_{i+1}-X_i)^2]=D(X_{i+1}-X_i)+E^2(X_{i+1}-X_i)=2\sigma^2$

所以 $E(\hat{\sigma}^2)=K\sum_{i=1}^{n-1}E(X_{i+1}-X_i)^2=K(n-1)2\sigma^2$，从而 $K=\frac{1}{2(n-1)}$

6. 设总体 X 的概率密度为 $f(x;\theta)=\begin{cases}\dfrac{1}{2\theta}, & 0<x<\theta \\ \dfrac{1}{2(1-\theta)}, & \theta\le x<1 \\ 0, & \text{其他}\end{cases}$，其中 $\theta(0<\theta<1)$ 是未知参数，$X_1,X_2,\dots,X_n$ 为来自总体 X 的简单随机样本，均值为 $\overline{X}$

(1) 求参数 θ 的矩估计量 $\hat{\theta}$；

(2) 判断 $4\overline{X}^2$ 是否为 θ^2 的无偏估计量，并说明理由.

解： (1) $E(X)=\int_{-\infty}^{+\infty}xf(x;\theta)dx=\int_0^\theta\frac{x}{2\theta}dx+\int_\theta^1\frac{x}{2(1-\theta)}dx=\frac{1}{4}+\frac{\theta}{2}$，令 $\overline{X}=E(X)$

得 θ 的矩估计量为：$\hat{\theta}=2\overline{X}-\frac{1}{2}$

(2) 因为
$$\begin{aligned}E(4\overline{X}^2)&=4E(\overline{X}^2)=4[D(\overline{X})+E^2(\overline{X})]\\&=4[\frac{1}{n}D(X)+(\frac{1}{4}+\frac{1}{2}\theta)^2]=\frac{4}{n}D(X)+\frac{1}{4}+\theta+\theta^2\end{aligned}$$

其中 $D(X)\ge 0,\theta>0$，所以 $E(4\overline{X}^2)>\theta^2$，座椅 $4\overline{X}^2$ 不是为 θ^2 的无偏估计量

7. 从总体 $X\sim N(\mu_1,25)$ 中抽取出一容量为 $n_1=10$ 的样本，其样本均值为 $\overline{X}_1=19.8$；从总体 $Y\sim N(\mu_2,36)$ 中抽取出一容量为 $n_2=12$ 的样本，其样本均值为 $\overline{Y}=24.0$，已知两个样本之间相互独立，则 $\mu_1-\mu_2$ 的 0.90 的置信区间为________.（$z_{0.05}=1.645$）

解： 这是 $\sigma_1{}^2,\sigma_2{}^2$ 都为已知时，求均值差的区间估计问题.

$\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$，从而有 $P\left\{\left|\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\right|\le z_{\frac{\alpha}{2}}\right\}=1-\alpha$

由于 $1-\alpha=0.90$，所以 $\frac{\alpha}{2}=0.05$，解上述不等式有：

$$\bar{X}-\bar{Y}-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le \mu_1-\mu_2\le \bar{X}-\bar{Y}+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

其中 $\bar{X}-\bar{Y}=-4.2$， $z_{0.05}=1.645$， $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}=2.345$，代入上式，解得

$$-8.06\le \mu_1-\mu_2\le -0.34$$

8. 设总体 $X\sim N(\mu,\sigma^2)$， σ^2 已知，则样本容量n至少为_____时，才能保证μ的置信度$1-\alpha$的置信区间长度不大于d.

解： 在 σ^2 已知的情况下，易知μ的置信度$1-\alpha$的置信区间为 $[\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$，所以$\mu$的置信度$1-\alpha$的置信区间长度为$2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$，依题意， $2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le d$，解得 $n\ge\left(\frac{2z_{\frac{\alpha}{2}}\sigma}{d}\right)^2$

点评：关于求区间估计的步骤：

(1) 根据题目条件，寻找枢轴量 W（枢轴量：它是一个样本和未知参数的函数，并且它的分布不依赖于未知参数）；

(2) 对于给定的置信区间为 $1-\alpha$，定出两个常数 a,b，使它们满足 $P\{a<W<b\}=1-\alpha$；

(3) **由(2)中确定的不等式，解出未知参数的范围，即为置信区间**

枢轴量的寻找是求区间估计的关键，下面给出不同情况下应如何选取枢轴量：

单个正态总体:

✧ σ^2已知，求μ 的置信区间，枢轴量为: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

✧ σ^2未知，求μ 的置信区间，枢轴量为: $\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

✧ μ已知，求σ^2的置信区间，枢轴量为：$\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 \sim \chi^2(n)$

✧ μ未知，求σ^2的置信区间，枢轴量为：$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

两个正态总体：

✧ $\sigma_1^{\,2},\sigma_2^{\,2}$已知，求$\mu_1-\mu_2$的置信区间，枢轴量为：$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

✧ $\sigma_1^{\,2}=\sigma_2^{\,2}=\sigma^2$未知，求$\mu_1-\mu_2$的置信区间，

枢轴量为：$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$ $\quad (S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2})$

✧ μ_1,μ_2未知，求$\frac{\sigma_1^{\,2}}{\sigma_2^{\,2}}$的置信区间，枢轴量为：$\frac{S_1^2/S_2^2}{\sigma_1^{\,2}/\sigma_2^{\,2}} \sim F(n_1-1,n_2-1)$

9. 设某次考试的考生成绩服从正态分布，从中随机抽取 36 位考生的成绩，算得平均分为 66.5 分，标准差为 15 分. 问在显著性水平 0.05 下，能否认为此次考试全体考生的平均成绩为 70 分？并给出检验过程($t_{0.025}(35)=2.0301$).

解： 设这次考试考生的成绩为 X，则 $X \sim N(\mu,\sigma^2)$，且σ^2未知，由题意建立假设：

$$H_0:\mu=70;H_1:\mu\neq 70$$

选取检验统计量$T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$，则拒绝域为：$\left|\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\right| \geq t_{\frac{\alpha}{2}}(n)=t_{0.025}(35)=2.0301$

当H_0成立时，有$\bar{X}=66.5,S=15$，代入上式，求得 $t=\left|\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\right|=1.4<2.0301$

t 的值未落在拒绝域内，所以接受原假设，即可以认为此次考试全体考生的平均成绩为 70 分

点评：假设检验的步骤：

(1) 根据实际问题的需求，提出原假设 H_0 和备择假设 H_1；

(2) 给定显著性水平 α 和样本容量 n;

(3) 根据题目所给条件，确定适合的检验统计量，并求出拒绝域;

(4) 在原假设成立的条件下，把样本的统计量代入验证;

(5) 如果计算结果未落在拒绝域内，则接受原假设，否则，拒绝原假设.

10. 已知总体 $X \sim N(\mu,\sigma^2)$，其中 μ 是未知参数，$X_1,X_2,\ldots,X_{16}$ 是其样本，$\bar{X}$ 为样本平均值，如果对检验 $H_0:\mu_1=\mu_0$，取拒绝域 $\{|\bar{X}-\mu_0|>k\}$，则 $k=$______.（$\alpha=0.05$）

解： $P\{|\bar{X}-\mu_0|>k\}=0.05$，也即 $P\{|\dfrac{\bar{X}-\mu_0|}{\sigma/\sqrt{n}}>\dfrac{k}{\sigma/\sqrt{n}}\}=0.05$，从而有：

$$\frac{k}{\sigma/\sqrt{n}}=z_{0.025}=1.96，解得 k=0.49\sigma$$

11. 设总体 $X \sim N(\mu,\sigma^2)$，现对 μ 进行假设检验，如在显著性水平 α =0.05 下接受了 $H_0:\mu=\mu_0$，则在显著性水平 α =0.01 下_______.

A.接受 H_0　　　　B.拒绝 H_0

C.可能接受，可能拒绝 H_0　　　　D.第一类错误的概率变大

解： A. 无论 σ^2 已知还是未知，当 α 变小时，拒绝域更小，原显著性水平条件下能接受 H_0，现在也能接受.

➢ 第七八章巩固提高练习

1. 设总体X的概率密度为$f(x;\theta)=\begin{cases}\theta & ,0<x<1\\ 1-\theta, & 1\le x<2\\ 0 & ,其他\end{cases}$，其中$\theta(0<\theta<1)$是未知参数，$X_1,X_2,\ldots,X_n$为来自总体 X 的简单随机样本，记 N 为样本值$x_1,x_2,\ldots,x_n$中小于 1 的个数．求：

(1) θ的矩估计；

(2) θ的最大似然估计.

2. 已知总体 X 的期望为$E(X)=0$，方差为$D(X)=\sigma^2$， $X_1,X_2,\ldots,X_n$为其简单随机样本，均值为$\overline{X}$，方差为S^2，则σ^2的无偏估计量为______.

A. $n\overline{X}^2+S^2$　　B. $\frac{1}{2}n\overline{X}^2+\frac{1}{2}S^2$

C. $\frac{1}{3}n\overline{X}^2+S^2$　　D. $\frac{1}{4}n\overline{X}^2+\frac{1}{4}S^2$

3. 设总体 X 的概率密度为$f(x;\theta)=\begin{cases}\frac{\theta^2}{x^3}e^{-\frac{\theta}{x}}, & x>0\\ 0, & 其他\end{cases}$,其中$\theta(\theta>0)$是未知参数，$X_1,X_2,\ldots,X_n$为来自总体 X 的简单随机样本.

(1) θ的矩估计量；

(2) θ的最大似然估计量.

4. 设总体 X 的概率密度为$f(x;\lambda)=\begin{cases}\lambda^2xe^{-\lambda x}, & x>0\\ 0, & 其他\end{cases}$,其中$\lambda(\lambda>0)$是未知参数，$X_1,X_2,\ldots,X_n$为来自总体 X 的简单随机样本.

(1) λ的矩估计量；

(2) λ的最大似然估计量.

5. 设总体 X 的概率密度为 $f(x;\theta)=\begin{cases}\dfrac{2x}{3\theta^2},0<x<2\theta\\0, \qquad\qquad 其他\end{cases}$，其中 θ 是未知参数，

$X_1,X_2,\ldots,X_n$ 为来自总体 X 的简单随机样本，若 $c\sum_{i=1}^{n}X_i^2$ 是 θ^2 的无偏估计，则.

$c=$______.

6. 研究两种固体燃料火箭推进器的燃烧率，设两者都服从正态分布，并且已知燃烧率的标注准差均近似地为 $0.05\,\mathrm{cm/s}$，取样本容量为 $n_1=n_2=20$，得燃烧率的样本均值分别为 $\bar{x}_1=18cm/s,\bar{x}_2=24cm/s$，则两燃烧率总体均值差 $\mu_1-\mu_2$ 的置信度为 0.99 的置信区间为_______.

7. 用甲、乙两种方法生产同一种药品，其成品的得率的方差分别为. 现测得甲方法生产的药品得率的 25 个数据，$\bar{X}=3.81$；乙方法生产的药品得率的 30 个数据，$\bar{Y}=3.56$. 设得率服从正太分布. 问在给定显著性水平 $\alpha=0.05$ 条件下，甲乙两种方法的平均得率是否有显著的差异？（$z_{0.025}=1.96$）

8. 某种导线，要求其电阻的标准差不得超过 0.005Ω. 今在生产的一批导线中，取样品 9 根，测得标准差为 0.007Ω. 设总体为正态分布，参数均未知，问在显著性水平 0.05 下能否认为这批导线的标准差显著地偏大？（$\chi^2_{0.05}(9)=16.919,\chi^2_{0.05}(8)=15.507$）

答案

1. (1) 矩估计：$\hat{\theta}=\dfrac{3}{2}-\bar{X}$；(2)最大似然估计： $\hat{\theta}=\dfrac{N}{n}$

2. B(直接按无偏估计求解，只有 B 项的均值为 σ^2)

3.(1) 矩估计：$\hat{\theta}=\overline{X}(\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i)$；(2)最大似然估计： $\hat{\theta}=\frac{2n}{\sum_{i=1}^{n}\frac{1}{X_i}}$

4. 矩估计：$\hat{\lambda}=\frac{2}{\overline{X}}$；(2)最大似然估计： $\hat{\lambda}=\frac{2}{\overline{X}}$

5. $c=\frac{2}{5n}$ 6. $(-6.04,-5.86)$

7. 原假设：$H_0:\mu_1=\mu_2$;备择假设$H_1:\mu_1\neq\mu_2$

$$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1),|u|=1.426<1.96$$，没有显著差异

8. 原假设：$H_0:\sigma\leq 0.005$;备择假设$H_1:\sigma>0.005$，$\chi^2=15.68>15.507$，落在拒绝域内，所以有显著差异