概率论与数理统计

任课教师: 胡家信

2022 年 10 月 31 日

上周内容总结

$$(1).\quad P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}).$$

$$(2).\quad P(B)=\sum_n P(A_n)P(B|A_n).$$

$$(3).\quad P(A_n|B)=\frac{P(A_n)P(B|A_n)}{\sum_k P(A_k)P(B|A_k)}\quad (\textbf{Bayes 定理}).$$

上周内容总结

- **独立性：** 随机变量$X_1, X_2, \cdots, X_n$是**独立的**，当且仅当对任意实数$x_1, x_2, \cdots, x_n$，事件

$$\{X_1 \le x_1\}, \{X_2 \le x_2\}, \cdots, \{X_n \le x_n\}$$

是独立的（此为**一般定义**）。

上周内容总结

但当随机变量$X_1, X_2, \cdots, X_n$取**可数值**时，有下列简化版的定义

- 取**可数值**的随机变量$X_1, X_2, \cdots, X_n$是**独立的**，当且仅当对任意实数$x_1, x_2, \cdots, x_n$，事件

$$\{X_1 = x_1\}, \{X_2 = x_2\}, \cdots, \{X_n = x_n\}$$

是独立的。或者，

- 取**可数值**的随机变量$X_1, X_2, \cdots, X_n$是**独立的**，当且仅当对任意实数$x_1, x_2, \cdots, x_n$，有

$$P(X_1 = x_1, \cdots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

上周内容总结

若$X_1, X_2, \cdots, X_n$是独立的，则

（1）　（分布函数）：$F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n)$。

（2）　（密度函数）：$f(u_1, \cdots, u_n) = f_1(u_1) \cdots f_n(u_n)$。

命题3：设取**可数值**的随机变量$X_1, X_2, \cdots, X_n$是独立的，φ_1，φ_2，$\cdots$，φ_n是定义在$(-\infty, \infty)$上的任意实值函数,则随机变量$\varphi_1(X_1)$，$\varphi_2(X_2)$，$\cdots$，$\varphi_n(X_n)$ 是独立的。

证明：令$Y_1 = \varphi_1(X_1), Y_2 = \varphi_2(X_2), \cdots, Y_n = \varphi_n(X_n)$，则它们取**可数值**。欲证明独立性，只需证明对任意实数$y_1, \cdots, y_n$，

$$P(Y_1 = y_1, \cdots, Y_n = y_n) = P(Y_1 = y_1) \cdots P(Y_n = y_n). \tag{1}$$

令$S_i = \{x : \varphi_i(x) = y_i\}$，则事件

$$\begin{aligned} \{Y_i = y_i\} &= \{\omega \in \Omega : Y_i(\omega) = \varphi_i(X_i(\omega)) = y_i\} \\ &= \{\omega \in \Omega : X_i(\omega) \in S_i\} = \{X_i \in S_i\}. \end{aligned}$$

但事件$\{X_1 \in S_1\}, \cdots, \{X_n \in S_n\}$是独立的（**利用前面命题**），故$\{Y_1 = y_1\}, \cdots, \{Y_n = y_n\}$是独立的，从而（1）成立，证毕。

例：随机从$[0,1]$取两个点，问它们之间的距离小于$1/2$的概率是多少？

解题思路：利用连续的随机变量以及独立性。注意：每个点的取值服从$[0,1]$上的均匀分布。

例子

解： 设X, Y是均匀分布在$[0,1]$的独立随机变量，其密度函数分别为

$$f_1(u)=\begin{cases}1, & u\in[0,1],\\ 0, & 其它,\end{cases}\qquad f_2(v)=\begin{cases}1, & v\in[0,1],\\ 0, & 其它.\end{cases}$$

我们的问题化为求概率$P(|X-Y|<1/2)$。事实上, 利用独立性，二元随机变量(X, Y)的密度函数为

$$f(u,v)=f_1(u)f_2(v)=\begin{cases}1, & u,v\in[0,1],\\ 0, & 其它,\end{cases}$$

从而，我们得

$$P\{(X,Y)\in S\}=\int\int_S f(u,v)dudv=\int\int_S dudv,$$

其中$S=\{(u,v): 0\le u\le 1, 0\le v\le 1, |u-v|<1/2\}$。

故所求概率为

$$P(|X-Y|<1/2)=\int\int_S dudv=1-2\cdot\frac{1}{8}=\frac{3}{4},$$

其中$S=\{(u,v):0\le u\le 1,0\le v\le 1,|u-v|<1/2\}$，如下图

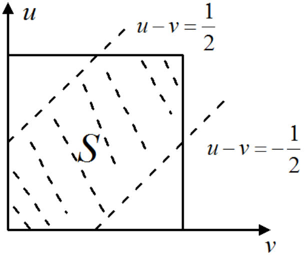


Figure: 区域S

例子

例：设$X_1, X_2, \cdots, X_n$是独立的随机变量，分布函数分别为$F_1, F_2, \cdots, F_n$。求

$$M = \max(X_1, X_2, \cdots, X_n),$$

$$m = \min(X_1, X_2, \cdots, X_n)$$

的分布函数.

分析：M和m是随机变量吗？

解：事实上，有

$$\begin{aligned} F_M(x) &= P(M \le x) = P(X_1 \le x; \cdots; X_n \le x) \\ &= P(X_1 \le x) \cdots P(X_n \le x) \text{（利用独立性）} \\ &= F_1(x) F_2(x) \cdots F_n(x), \end{aligned}$$

该函数为M的分布函数。

下求随机变量m的分布函数。引入函数

$$G_j(x) = P(X_j > x) = 1 - P(X_j \le x) = 1 - F_j(x).$$

则有

$$\begin{aligned} P(m > x) &= P(X_1 > x; \cdots; X_n > x) \\ &= P(X_1 > x) \cdots P(X_n > x) \quad \text{（为什么？）} \\ &= G_1(x) G_2(x) \cdots G_n(x). \end{aligned}$$

从而得

$$\begin{aligned} F_m(x) &= P(m \leq x) = 1 - P(m > x) \\ &= 1 - G_1(x) G_2(x) \cdots G_n(x) \\ &= 1 - (1 - F_1(x))(1 - F_2(x)) \cdots (1 - F_n(x)). \end{aligned}$$

解毕。

解题方法

解概率问题的步骤:

1. 样品空间的描述，即所有可能的结果构成的集合;
2. 计算**单个**事件发生的概率;
3. 计算(**复合**)事件发生的概率，和条件概率等。
   所用知识有: 计算概率的**基本公式**，或者概率的非负性、可加性、正则性等测度性质。

解题方法

经常遇见的情况如下：

1. **点数法**：适用于所有可能的结果**有限**;

2. **序列法**：适用于试验的结果具有**序列性质**;

3. **分割征服法(divide-and-conquer)**：

$$P(B) = \sum_i P(A_i)P(B|A_i),$$

其中$\{A_i\}$是样品空间的不相交的分割.

课堂练习

课堂练习题： 一个班有4名研究生和12名本科生，随机分为4组，每组4人，问每组只有一个研究生的概率？

答案

答案（点数法）：

- 第一步（**确定样品空间的大小**）。16人分成四组，共有

$$\binom{16}{4}\cdot\binom{12}{4}\cdot\binom{8}{4}\cdot\binom{4}{4}=\frac{16!}{(4!)^4}$$

中分法。

- 第二步（**确定事件发生的可能性**）。16人分成四组，每组只有一个研究生的情形共有

$$\binom{4}{1}\binom{12}{3}\cdot\binom{3}{1}\binom{9}{3}\cdot\binom{2}{1}\binom{6}{3}=\frac{4!12!}{(3!)^4}$$

- 第三步（**计算结果**）。所求概率为

$$P(A)=\frac{\frac{4!12!}{(3!)^4}}{\frac{16!}{(4!)^4}}=\frac{12\times 8\times 4}{15\times 14\times 13}.$$ 解毕。

课堂讨论

题目(New York Times, July 21, 1991)： 有3个紧闭的门，一个门后面藏着小轿车，另2个门后面分别藏着一只山羊。现主人让你随机选一个门后（不打开），随机打开其中剩下两个门中的的一个门，结果发现后面是一只山羊。问你换和不换门，那种情况得到小轿车的概率大?

关键点：确定样品空间。

你如何选择？（随便试一下）

答案

假设你的策略是**换门**，结果如何?

第1步（建立样品空间）：设3个紧闭的门标号分别为$1, 2, 3$，结果用4个字母表示(u, v, w, x)：u表示你最初选定的门牌号码，v表示主人打开的门牌号码，w表示你换的门牌号码，x表示你赢(W)或者输(L)。如$(1, 2, 3, L)$表示你开始选择1号门，主人打开的2号门（发现后面是一只山羊），然后你换成3号门，结果你输，车在1号们后面。**样品空间为**

$$S = \{(1, 2, 3, L), (1, 3, 2, L), (2, 3, 1, W), (3, 2, 1, W)\}.$$

(**注意：这里是假设车在第一号门后面的样品空间**)。从上看出，若你开始选2号或3号门，你一定赢；若你开始选1号，你一定输（以2种不同的方式输）。

答案

第2步（计算概率）：选择1、2、或3号门的概率是一样的，均为1/3。故

$$P(2,3,1,W) = P(3,2,1,W) = 1/3,$$
$$P(1,2,3,L) + P(1,3,2,L) = 1/3.$$

所以你赢得概率为

$$P(\text{赢}) = P(2,3,1,W) + P(3,2,1,W) = 2/3,$$

而你输得概率为

$$P(\text{输}) = P(1,2,3,L) + P(1,3,2,L) = 1/3.$$

答案

不换策略（仍然假设车在第一号门后面）

第1步（建立样品空间）：

$$S = \{(1,2,1,W),(1,3,1,W),(2,3,2,L),(3,2,3,L)\}.$$

第2步（计算概率）：选择1、2、或3号门的概率是一样的，均为1/3。故

$$P(2,3,2,L) = P(3,2,3,L) = 1/3,$$
$$P(1,2,1,W) + P(1,3,1,W) = 1/3.$$

所以你赢的概率为$P(\text{赢}) = P(1,2,1,W) + P(1,3,1,W) = 1/3$，而你输的概率为$P(\text{输}) = P(2,3,2,L) + P(3,2,3,L) = 2/3$。

结论: 换赢的概率为2/3, 不换赢的概率为1/3，故，要换。

回顾**数学期望**：Ω是可数样品空间，$E(X)=\sum_{\omega\in\Omega}X(\omega)P(\omega)$，右边有限值的一个充分条件为$E(|X|)=\sum_{\omega\in\Omega}|X(\omega)|P(\omega)<\infty$。

数学期望的意义：

- 若X是学生的身高，则$E(X)$表示平均身高；
- 若X是工人的工资，则$E(X)$表示平均工资；
- 若X是一天中通过收费站的车辆数，则$E(X)$表示每天平均车流量。

易知，

$$E(I_A) = P(A),$$

这是因为

$$\begin{aligned} E(I_A) &= \sum_{\omega\in\Omega} I_A(\omega)P(\omega) \\ &= \sum_{\omega\in A} 1\cdot P(\omega) = P(A). \end{aligned}$$

数学期望满足下列性质:

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

例子

例题1：彩票中心有100张彩票，只有1张含1万元奖金，其余奖金是0。现我买2张彩票，平均收益收多少？（猜一猜？）

想一想，每张彩票至少卖出100元，否则彩票中心亏本。

例子

解：设每张彩票的收益为随机变量$X(\omega)$，则

$$X = \begin{cases} 10000, & 概率为\dfrac{1}{100}, \\ 0, & 概率为\dfrac{99}{100}. \end{cases}$$

所以，每张彩票的平均收益为

$$E(X) = 10000 \times \frac{1}{100} + 0 \times \frac{99}{100} = 100（元）.$$

两张彩票的平均收益

$$E(X_1 + X_2) = E(X_1) + E(X_2) = 200（元）.$$

解毕。

例子

例题2：现有标号为1到N的兑奖卷，每个兑奖卷抽取后（商家再）添加相同的兑奖卷。你想收集r个不同的兑奖卷，问平均抽多少次?

注意：此为**放回抽样**。

解：设$X_1, X_2, \cdots$表示得到新奖卷的"依次等时"，则$X_1 = 1$。现X_2是抽取任意不同于X_1(第1次)的等待时间，则$E(X_2) = \frac{N}{N-1}$，这是因为

$$P(X_2 = n) = \left(\frac{1}{N}\right)^{n-1} \cdot \frac{N-1}{N},$$

即前面$n-1$次均抽到X_1（概率为$\left(\frac{1}{N}\right)^{n-1}$），第$n$次抽到$X_2, X_3, \cdots, X_N$中的任何一个（概率为$\frac{N-1}{N}$)。于是，

$$E(X_2) = \sum_{n=1}^{\infty} n \cdot \left(\frac{1}{N}\right)^{n-1} \cdot \frac{N-1}{N} = \frac{N-1}{N} \cdot \frac{1}{(1-1/N)^2} = \frac{N}{N-1},$$

这里利用等式$\sum_{n\geq 1} nx^{n-1} = 1/(1-x)^2, 0 < x < 1$（前面内容）。

例子

同理，

$$E(X_3) = \frac{N}{N-2}.$$

事实上，

$$P(X_3 = n) = \left(\frac{2}{N}\right)^{n-1} \cdot \frac{N-2}{N},$$

从而，我们得

$$\begin{aligned} E(X_3) &= \sum_{n=1}^{\infty} n \cdot \left(\frac{2}{N}\right)^{n-1} \cdot \frac{N-2}{N} \\ &= \frac{N-2}{N} \cdot \frac{1}{(1-2/N)^2} = \frac{N}{N-2}. \end{aligned}$$

例子

同理，$E(X_r) = \frac{N}{N-r+1}$。如此，得到

$$E(X_1 + \cdots + X_r) = \frac{N}{N} + \frac{N}{N-1} + \cdots + \frac{N}{N-r+1}$$
$$= N\left(\frac{1}{N-r+1} + \cdots + \frac{1}{N}\right).$$

特别地，$r = N$，

$$E(X_1 + \cdots + X_N) = N\left(1 + \frac{1}{2} + \cdots + \frac{1}{N}\right) \simeq N\left(\log N + C\right)$$

这里$C = 0.5772...$为欧拉常数，例：$N = 12$ 或$N = 108$，则期望分别为

$12 * (\log 12 + 0.5772) \approx 37$ 或 $108 * (\log 108 + 0.5772) \approx 568$. 解毕。

庞加莱公式

对任意集合$A_1, A_2, \cdots, A_n$, 下列等式成立:

$$P\left(\cup_{j=1}^{n} A_j\right) = \sum_{j} P(A_j) - \sum_{j,k} P(A_j A_k) + \sum_{j,k,l} P(A_j A_k A_l) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

其中求和的指标不同、取值从1到n。特别地，

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

证明： 令$\alpha_j = I_{A_j}$。注意到对任意$A_1, \cdots, A_n$,

$$I_{A_1 A_2 \cdots A_n} = \prod_{j=1}^{n} I_{A_j}.$$

庞加莱公式

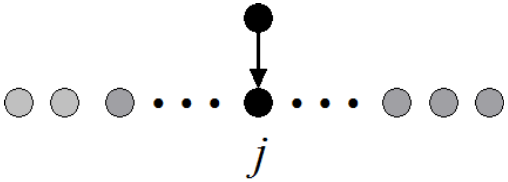
故

$$
\begin{aligned}
I_{A_1\cup\cdots\cup A_n} &= 1 - I_{(A_1\cup\cdots\cup A_n)^c} \\
&= 1 - \prod_{j=1}^{n} I_{A_j^c} = 1 - \prod_{j=1}^{n}(1-\alpha_j) \\
&= \sum_j \alpha_j - \sum_{j,k}\alpha_j\alpha_k + \sum_{j,k,l}\alpha_j\alpha_k\alpha_l \\
&\quad - \cdots + (-1)^{n-1}\alpha_1\cdots\alpha_n.
\end{aligned}
$$

另一方面，$E(I_A) = P(A)$。在上式两边取数学期望，得到所证。证毕。

例子

例3：标号从1到n的两副牌随机配对，问至少有1个配对的概率是多少?（如下图，第j张牌正好配对。）



例子

解：令A_j表示标号为j的牌配对，不管其它，则

$$P(A_j) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

同理，若标号为j和k的牌均配对，不管其它，则

$$P(A_jA_k) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, \quad j \neq k.$$

如此，

$$P(A_jA_kA_l) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}.$$

等等（目的：欲计算$P\left(\cup_{j=1}^{n} A_j\right)$）。

例子

利用上面公式，计算公式右边和，得

$$\binom{n}{1}\frac{1}{n}-\binom{n}{2}\frac{1}{n(n-1)}+\binom{n}{3}\frac{1}{n(n-1)(n-2)}-\cdots+(-1)^{n-1}\frac{1}{n!}$$
$$=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n-1}\frac{1}{n!}=\sum_{k=1}^{n}(-1)^{k-1}\frac{1}{k!}.$$

故所求概率为

$$P\left(\cup_{j=1}^{n}A_j\right)=\sum_{k=1}^{n}(-1)^{k-1}\frac{1}{k!}.$$

注意，当$n\to\infty$时，有

$$\sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k!}=1-e^{-1}\simeq 0.63.$$

解毕。

例子

随机配对数为

$$N := I_{A_1} + \cdots + I_{A_n},$$

它是一个随机变量，其数学期望(即平均配对数)为

$$E(N) = \sum_{j=1}^{n} E(I_{A_j}) = \sum_{j=1}^{n} P(A_j) = n \cdot \frac{1}{n} = 1.$$

题目： 已知$\Omega = \{\omega_1, \omega_2, \omega_3\}, P(\omega_1) = P(\omega_2) = P(\omega_3) = \frac{1}{3}$，定义

$$X(\omega_1) = 1, X(\omega_2) = 2, X(\omega_3) = 3,$$
$$Y(\omega_1) = 2, Y(\omega_2) = 3, Y(\omega_3) = 1,$$
$$Z(\omega_1) = 3, Z(\omega_2) = 1, Z(\omega_3) = 2.$$

证明X, Y, Z**有相同的概率分布**，并求$X + Y, Y + Z, Z + X$**的概率分布**。

我们的问题如下：

1. 证明：$P(X = 1) = P(Y = 1) = P(Z = 1)$,
   $P(X = 2) = P(Y = 2) = P(Z = 2)$,
   $P(X = 3) = P(Y = 3) = P(Z = 3)$?
2. 求三个随机变量$X + Y, Y + Z, Z + X$**的概率分布**

解答：

1. 这里只证明：$P(X=1)=P(Y=1)=P(Z=1)$，其他类似（略）。事实上，

$$P(X=1)=P(\{\omega\in\Omega:X(\omega)=1\})=P(\{\omega_1\})=P(\omega_1)=\frac{1}{3},$$

$$P(Y=1)=P(\{\omega\in\Omega:Y(\omega)=1\})=P(\omega_3)=\frac{1}{3},$$

$$P(Z=1)=P(\{\omega\in\Omega:Z(\omega)=1\})=P(\omega_2)=\frac{1}{3}.$$

2. 这里只求随机变量$W:=X+Y$的**概率**分布，其他两个随机变量一样处理。按钟书随机变量的**加法定义**：$W(\omega)=X(\omega)+Y(\omega)$有如下取值：$W(\omega_1)=X(\omega_1)+Y(\omega_1)=1+2=3$, $W(\omega_2)=5$，$W(\omega_3)=4$, 故$P(W=3)=P(W=4)=P(W=5)=\frac{1}{3}$，即为所求，解毕。

题目：一个箱子里有n张标号从1到n的彩票，(不放回)抽两张，标号小的彩票记为X，大的记为Y。求(X, Y)的**联合分布**、**边际分布**、以及随机变量$Y - X$的分布？

解答：(X, Y)**的联合分布**

$$P((X, Y) = (p, q)) = P(X = p, Y = q) = \frac{1}{\binom{n}{2}}, \ 1 \le p < q \le n,$$

这是因为从n张彩票抽两张有$\binom{n}{2}$种可能，而该两张彩票中较小号码为p、同时较大号码为q的情况只有1种可能。X**的边际分布**

$$\begin{aligned} P(X = p) &= P(X = p, -\infty < Y < \infty) \\ &= \sum_{q=p+1}^{n} P(X = p, Y = q) = \frac{n-p}{\binom{n}{2}}, \ 1 \le p \le n. \end{aligned}$$

Y的边际分布

$$P(Y=q)=\sum_{p=1}^{q-1}P(X=p,Y=q)=\frac{q-1}{\binom{n}{2}},\ 1\leq q\leq n.$$

随机变量$Y-X$的分布

$$\begin{aligned}P(Y-X&=k)=\sum_{p=1}^{n-k}P(X=p,Y=k+p)\\&=\frac{n-k}{\binom{n}{2}},\ 1\leq k\leq n-1.\end{aligned}$$

解毕。

习题讲解（钟开来书，第158页3题）

There are two kinds of tubes in an electronic gadget. It will cease to function if and only if one of each kind is defective. The probability that there is a defective tube of the first kind is .1; the probability that there is a defective tube of the second kind is .2. It is known that two tubes are defective, what is the probability that the gadget still works?

电子元件（如电视机）由两种电子管构成，第一种电子管次品率为0.1, 第二种电子管次品率为0.2。当且仅当每种电子管均有一个为次品时，电子元件不工作（发生故障）。现有两个电子管为次品，问电子元件仍能工作的概率为多少？

求条件概率。

解：设$A=\{$电子元件正常工作$\}$，$S=\{$两个电子管为次品$\}$。欲求$P(A|S)$.

- 两个次品均为第一种电子管，概率为$0.1*0.1=0.01$;
- 两个次品均为第二种电子管，概率为$0.2*0.2=0.04$;
- 第一个次品为第一种电子管、同时第二个次品为第二种电子管，概率为$0.1*0.2=0.02$;
- 第一个次品为第二种电子管、同时第二个次品为第一种电子管，概率为$0.2*0.1=0.02$。

所以，$P(S)=0.01+0.04+0.02*2=0.09$,

$$P(AS)=0.01+0.04=0.05,\quad P(A|S)=\frac{P(AS)}{P(S)}=\frac{0.05}{0.09}=\frac{5}{9}.$$

解毕。

One hundred trout are caught in a little lake and returned after they are tagged. Later another 100 are caught and found to contain 7 tagged ones. What is the probability of this if the lake contains n trout? [What is your best guess as to the true value of n? The latter is the kind of question asked in statistics.]

湖里有n条红鳟鱼，其中100条有标签。现捕鱼**100**条，发现**7**条有标签，问此事件发生的概率为多少？

答案：所求概率为

$$p_n = \begin{cases} \dfrac{\binom{100}{7}\binom{n-100}{93}}{\binom{n}{100}}, & \text{if } n \geq 193, \\ 0, & \text{if } n < 193. \end{cases}$$

注：湖里至少有193条鱼，或$\binom{n-100}{93} = \binom{n-100}{n-193}$有意义。

问：猜想n的真值是多少，即求n为何值，概率p_n达到最大？

答案：注意p_n关于n递增，当且仅当

$$p_n < p_{n+1} \Leftrightarrow n < \frac{99^2+192}{7} \approx 1427.6,$$

故当$n=1428$时，p_n达到最大。实际上，计算如下：

$$\begin{aligned} p_{1427} &= \frac{\binom{100}{7}\binom{1427-100}{93}}{\binom{1427}{100}} \approx 0.160197\mathbf{6066}, \\ p_{1428} &= \frac{\binom{100}{7}\binom{1428-100}{93}}{\binom{1428}{100}} \approx 0.160197\mathbf{9700}, \\ p_{1429} &= \frac{\binom{100}{7}\binom{1429-100}{93}}{\binom{1429}{100}} \approx 0.160197\mathbf{6979}. \end{aligned}$$

附注: 利用平均值近似相等：$\frac{7}{100} \approx \frac{100}{n} \Rightarrow n = \frac{10000}{7} = 1428.6 \approx 1429$，也对。

作业

第6次作业(钟开来书)：

P. 195-196：第2，3，4，5，6，7，11，12，15题.

预习内容：方差、生成函数