

Probability and Statistics

Tutorial 4

Siyi Wang

Southern University of Science and Technology

11951002@mail.sustech.edu.cn

October 15, 2020

Outline

- 1 Review
- 2 Homework
- 3 Supplement Exercises
- 4 Further Reading

1. Definition of Random Variables

- A random variable X is a 'measurable' function $X : \Omega \rightarrow T$.
- Here, we often take T to be \mathbf{N} , \mathbf{R} , \mathbf{R}^n , ...

2. Classification of Random Variable

- Discrete r.v.: T finite or countable.
- (Generalized) Continuous r.v.
 - Absolute Continuous r.v. (This is the Continuous r.v. in this course)
 - Singular Continuous r.v.
 - Mixture
- Mixed Type r.v.

3. Discrete r.v. X

- PMF: $P(X = x_k) = p_k$, $k = 1, 2, \dots$
- $p_k \in [0, 1]$ and $\sum_{k=1}^{\infty} p_k = 1$.

频率函数的几种表示方法

解析式法

$$P\{X = x_k\} = p_k \quad (k = 1, 2, \dots)$$

列表法

X	x_1	x_2	\dots	x_k	\dots
p_k	p_1	p_2	\dots	p_k	\dots

矩阵法

$$\begin{pmatrix} x_1 & x_2 & \dots & x_k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}$$

打进第 k 个“盒子”

4. Distribution Function (CDF) $F_X(x)$ of r.v. X

- $F_X(x) = P(X \leq x)$.
- $P(a < X \leq b) = F_X(b) - F_X(a)$. ($P(X = a) = F(a) - F(a-)$).
- $F_X(x)$ is nondecreasing.
- $F_X(x)$ is right continuous.
- $F_X(x) \in [0, 1]$, $F(-\infty) = 0$ and $F(\infty) = 1$.
- If the CDF is continuous, then we call this r.v. is a continuous random variable.

5. Almost Everywhere/Almost Surely.

- $X = Y$ a.e. iff $P(X = Y) = 1$.
- A a.e iff $P(A) = 1$.

6. Bernoulli Distribution $Bern(p)$

- $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

7. Binomial Distribution $Bin(n, p)$

- $P(X = k) = C_n^k p^k (1 - p)^{n-k}$, $k = 0, 1, 2, \dots, n$.
- If $Y_i \sim_{i.i.d.} Bern(p)$, then $X = Y_1 + \dots + Y_n \sim Bin(n, p)$.
- If $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$ and they are independent, then $X = X_1 + X_2 \sim Bin(n_1 + n_2, p)$.

8. Geometric Distribution $Geom(p)$

- $P(X = k) = (1 - p)^{k-1} p$, $k = 1, 2, \dots$
- If $Y_i \sim_{i.i.d.} Bern(p)$, then $X = \inf\{k : Y_k = 1\} \sim Geom(p)$.
- (Memoryless Property) $P(X > n + m | X > n) = P(X > m)$.

9. Negative Binomial Distribution *Negative Binomial*(r, p)

- $P(X = k) = C_{k-1}^{r-1} p^r (1-p)^{k-r}, k = r, r+1, \dots$
- If $Y_i \sim_{i.i.d.} \text{Bern}(p)$, then
 $X = \inf\{k : Y_1 + \dots + Y_k = r\} \sim \text{Negative Binomial}(r, p).$

10. Hypergeometric Distribution $H(r, n, m)$

- $P(X = k) = \frac{C_r^k C_{n-r}^{m-k}}{C_n^m}, k = 0, 1, \dots, m.$
- $\sum_{k=0}^m C_r^k C_{n-r}^{m-k} = C_n^m.$

11. Poisson Distribution $Poisson(\lambda)$

- $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots$
- $e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$
- (Poisson Theorem) If $\lim_{n \rightarrow \infty} np_n = \lambda$ where $\lambda > 0$, then for any integer $k \geq 0$, we have $\lim_{n \rightarrow \infty} C_n^k p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$

12. Poisson Process $N(t)$

- For each $t > 0$, $N(t) \sim Poisson(\lambda t).$
- $N(0) = 0.$
- For complete definition, see Further Reading Part.

Homework

1. 假设 X 是离散随机变量, 具有 $P\{X = 0\} = 0.25$, $P\{X = 1\} = 0.125$, $P\{X = 2\} = 0.125$ 和 $P\{X = 3\} = 0.5$. 画出 X 的频率函数和累积分布函数.

Homework

Solution

CDF:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.25, & 0 \leq x < 1 \\ 0.375, & 1 \leq x < 2 \\ 0.5, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases} \quad (1)$$

Homework

15. 两支球队 A 和 B 进行系列赛。如果 A 队赢得比赛的概率为 0.4, 那么对他有利的是 5 局 3 胜制还是 7 局 4 胜制? 假设连续比赛的结果是相互独立的。

Solution

Model 1 (Binomial Distribution) For 5/3, we have

$$P(\text{win}) = C_5^5(0.4)^5 + C_5^4(0.4)^4(0.6) + C_5^3(0.4)^3(0.6)^2 \approx 0.31744$$

For 7/4, we have $P(\text{win}) =$

$$C_7^7(0.4)^7 + C_7^6(0.4)^6(0.6) + C_7^5(0.4)^5(0.6)^2 + C_7^4(0.4)^4(0.6)^3 \approx 0.2898.$$

Hence, 5/3 is better for A.

Model 2 (Supplement Exercises 5) Omit Here.

Remark. The above two model give the same answer.

31. 在某些居住地, 每小时内的被叫电话次数服从参数为 $\lambda = 2$ 的泊松过程.
- a. 如果 Diane 洗浴 10 分钟, 期间电话铃声响起概率是多少?
 - b. 如果她希望没有被叫电话的概率最多为 0.5, 那么她可以洗浴多长时间?

Solution

a. $P(X(\frac{1}{6}) \geq 1) = 1 - P(X = 0) = 1 - e^{-2 * \frac{1}{6}} = 1 - e^{-\frac{1}{3}}$

b. $P(X(t) \geq 1) = 1 - P(X = 0) = 1 - e^{-2 * t} \leq 0.5$. That is, $t \geq 30 \ln 2$.

1. 设随机变量 X 的频率函数为：

$$P(X=x) = c \left(\frac{2}{3} \right)^x, \quad x=1, 2, 3$$

求 c 的值。

Homework

Solution

Since $c(\frac{2}{3}) + c(\frac{2}{3})^2 + c(\frac{2}{3})^3 = 1$, then $c = \frac{27}{38}$.

2. 设随机变量 X 服从泊松分布, 求 k 使 $P(X = k)$ 达到最大。

Solution

Let $h(k) := P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$. Consider $\frac{P(X=k)}{P(X=k-1)} = \frac{\lambda}{k}$. Hence, $h(k)$ is increasing in $\{k \leq [\lambda]\}$ and decreasing in $\{k \geq [\lambda]\}$. Hence, $k_{\max} = [\lambda]$. (If λ is integer, then $k_{\max} = \lambda, \lambda - 1$.)

3. 设在 15 只同类型零件中有 2 只为次品，在其中取 3 次，每次任取 1 只，作不放回抽样，以 X 表示取出的次品个数，求：

(1) X 的分布律；

(2) X 的分布函数并作图；

(3)

$$P\{X \leq \frac{1}{2}\}, P\{1 < X \leq \frac{3}{2}\}, P\{1 \leq X \leq \frac{3}{2}\}, P\{1 < X < 2\}.$$

Homework

Solution

$$(1) \text{ PMF: } P(X = 0) = \frac{C_{13}^3}{C_{15}^3} = \frac{22}{35}, \quad P(X = 1) = \frac{C_2^1 C_{13}^2}{C_{15}^3} = \frac{12}{35},$$

$$P(X = 2) = \frac{C_{13}^1}{C_{15}^3} = \frac{1}{35}.$$

$$\text{PMF of another model: } P(X = 0) = \frac{A_{13}^3}{A_{15}^3} = \frac{22}{35}, \quad P(X = 1) = \frac{C_3^1 A_2^1 A_{13}^2}{A_{15}^3} = \frac{12}{35},$$

$$P(X = 2) = \frac{C_3^1 A_{13}^1 A_2^2}{A_{15}^3} = \frac{1}{35}.$$

$$(2) \text{ CDF: } F_X(x) = P(X \leq x).$$

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{22}{35}, & 0 \leq x < 1 \\ \frac{34}{35}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases} \quad (2)$$

$$(3) P(X \leq \frac{1}{2}) = \frac{22}{35}, \quad P(1 < X \leq \frac{3}{2}) = 0, \quad P(1 \leq X \leq \frac{3}{2}) = \frac{12}{35}, \\ P(1 < x < 2) = 0.$$

4. 有 2500 名同一年龄和同社会阶层的人参加了保险公司的人寿保险。在一年中每个人死亡的概率为 0.002，每个参加保险的人在 1 月 1 日须交 12 元保险费，而在死亡时家属可从保险公司领取 2000 元赔偿金。求：
- (1) 保险公司亏本的概率；
 - (2) 保险公司获利分别不少于 10000 元、20000 元的概率。

Homework

Solution

Let $X(n)$ be the number of death out of n people. When $n = 2500$ and $p = 0.002$, $\lambda = 2500 * 0.002 = 5$. Let $Y \sim \text{Poisson}(5)$.

(1)

$$P(2000 * X(2500) > 12 * 2500) = P(X(2500) > 15) \approx P(Y > 15) \approx 0.000$$

$$(2) P(12 * 2500 - 2000 * X(2500) \geq 10000) = P(X(2500) \leq 10) \approx P(Y \leq 10) \approx 0.986$$

$$P(12 * 2500 - 2000 * X(2500) \geq 20000) = P(X(2500) \leq 5) \approx P(Y \leq 5) \approx 0.616$$

Exercise 1

5. 设 10 件产品中有 4 件不合格品, 从中任取两件, 已知其中一件是不合格品, 求另一件也是不合格品的概率.

Supplement Exercises

Solution

解 记事件 A_i 为“第 i 次取出不合格品”, $i = 1, 2$, D 为“有一件是不合格品”, E 为“另一件也是不合格品”. 因为 D 意味着: 第一件是不合格品而第二件是合格品, 或第一件是合格品而第二件是不合格品, 或两件都是不合格品. 而 ED 意味着: 两件都是不合格品. 即 $D = A_1\bar{A}_2 \cup \bar{A}_1A_2 \cup A_1A_2$, $ED = A_1A_2$. 因为

$$P(D) = P(A_1\bar{A}_2) + P(\bar{A}_1A_2) + P(A_1A_2) = \frac{4 \times 6}{10 \times 9} + \frac{6 \times 4}{10 \times 9} + \frac{4 \times 3}{10 \times 9} = \frac{2}{3},$$

$$P(ED) = \frac{4 \times 3}{10 \times 9} = \frac{2}{15},$$

所以根据题意得

$$P(E|D) = \frac{2/15}{2/3} = \frac{1}{3} = 0.33.$$

Exercise 2

12. 一盒晶体管中有 8 只合格品, 2 只不合格品. 从中不返回地一只一只取出, 试求第二次取出合格品的概率.

Solution

$$P(A_2) = P(A_1)P(A_2 | A_1) + P(\bar{A}_1)P(A_2 | \bar{A}_1) = \frac{8}{10} \times \frac{7}{9} + \frac{2}{10} \times \frac{8}{9} = \frac{4}{5}.$$

Exercise 3

21. 将 n 根绳子的 $2n$ 个头任意两两相接, 求恰好结成 n 个圈的概率.

Solution

解 设事件 A_n 为“恰好结成 n 个圈”, 记 $p_n = P(A_n)$. 又记事件 B 为“第 1 根绳子的两个头相接成圈”, 则由全概率公式得

$$P(A_n) = P(B)P(A_n | B) + P(\bar{B})P(A_n | \bar{B}),$$

容易看出

$$P(B) = \frac{1}{2n-1}, \quad P(A_n | \bar{B}) = 0, \quad P(A_n | B) = P(A_{n-1}) = p_{n-1}.$$

所以得递推公式

$$p_n = \frac{1}{2n-1} p_{n-1}, \quad n = 2, 3, \dots,$$

由此得

Exercise 4

18. 一个人的血型为 A, B, AB, O 型的概率分别为 0.37, 0.21, 0.08, 0.34. 现任意挑选四个人, 试求:

- (1) 此四人的血型全不相同的概率;
- (2) 此四人的血型全部相同的概率.

Solution

解 (1) 若第 1,2,3,4 人血型依次为 A,B,AB,O, 则“四人的血型全不相同”共有 $4! = 24$ 种可能情况, 而每种情况出现的概率都是 $0.37 \times 0.21 \times 0.08 \times 0.34$, 于是所求概率为

$$P(\text{四人的血型全不相同}) = 24 \times 0.37 \times 0.21 \times 0.08 \times 0.34 = 0.0507.$$

(2) 所求概率为

$$\begin{aligned} P(\text{血型全相同}) &= P(\text{全为 A 型}) + P(\text{全为 B 型}) + \\ &\quad P(\text{全为 AB 型}) + P(\text{全为 O 型}) \\ &= 0.37^4 + 0.21^4 + 0.08^4 + 0.34^4 = 0.0341. \end{aligned}$$

Exercise 5

19. 甲、乙两选手进行乒乓球单打比赛,已知在每局中甲胜的概率为 0.6,乙胜的概率为 0.4. 比赛可采用三局二胜制或五局三胜制,问哪一种比赛制度对甲更有利?

Supplement Exercises

Solution

解 (1) 若采用三局二胜制, 则甲在下列两种情况下获胜:

A_1 = “甲胜前两局”,

A_2 = “前两局甲乙各胜一局, 第三局甲胜”,

所以得

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) = 0.6^2 + 2 \times 0.6 \times 0.4 \times 0.6 \\ &= 0.36 + 0.288 = 0.648. \end{aligned}$$

(2) 若采用五局三胜制, 则甲在下列三种情况下获胜:

B_1 = “前三局甲胜”,

B_2 = “前三局中甲胜两局乙胜一局, 第四局甲胜”,

B_3 = “前四局甲乙各胜二局, 第五局甲胜”,

所以得

$$\begin{aligned} P(B_1 \cup B_2 \cup B_3) &= 0.6^3 + \binom{3}{1} \times 0.6^2 \times 0.4 \times 0.6 + \binom{4}{2} \times 0.6^2 \times 0.4^2 \times 0.6 \\ &= 0.216 + 0.259 + 0.207 = 0.682. \end{aligned}$$

所以五局三胜制对甲更有利.

Exercise 6

7. 一批产品的不合格品率为 0.02, 现从中任取 40 件进行检查, 若发现两件或两件以上不合格品就拒收这批产品. 分别用以下方法求拒收的概率: (1) 用二项分布作精确计算; (2) 用泊松分布作近似计算.

Solution

解 记 X 为抽取的 40 件产品中的不合格品数, 则 $X \sim b(40, 0.02)$, 而“拒收”就相当于“ $X \geq 2$ ”.

(1) 拒收的概率为

$$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) = 1 - 0.98^{40} - 40 \times 0.98^{39} \times 0.02 \\ &= 0.1905. \end{aligned}$$

(2) 因为 $\lambda = 40 \times 0.02 = 0.8$, 所以用泊松分布作近似计算, 可得近似值为

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - e^{-0.8} - 0.8 \times e^{-0.8} = 0.1912.$$

可见近似值与精确值相差 0.0007, 近似效果较好.

Exercise 7

9. 已知某商场一天来的顾客数 X 服从参数为 λ 的泊松分布, 而每个来到商场的顾客购物的概率为 p , 证明: 此商场一天内购物的顾客数服从参数为 λp 的泊松分布.

Solution

证 用 Y 表示商场一天内购物的顾客数, 则由全概率公式知, 对任意正整数 k 有

$$\begin{aligned} P(Y = k) &= \sum_{i=k}^{+\infty} P(X = i) P(Y = k | X = i) = \sum_{i=k}^{+\infty} \frac{\lambda^i e^{-\lambda}}{i!} \binom{i}{k} p^k (1-p)^{i-k} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{i=k}^{+\infty} \frac{[\lambda(1-p)]^{i-k}}{(i-k)!} = \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p}. \end{aligned}$$

这表明: Y 服从参数为 λp 的泊松分布.

Exercise 8

23. 一本 500 页的书共有 500 个错误,若每个错误等可能地出现在每一页上 (每一页上至少有 500 个印刷符号). 试求指定的一页上至少有三个错误的概率.

Supplement Exercises

Solution

解 设 X 为指定一页上错误的个数, 则 $X \sim b(500, p)$, 且 $p = 1/500$. 所求的概率为

$$P(X \geq 3) = 1 - P(X < 3) = 1 - \sum_{i=0}^2 \binom{500}{i} \left(\frac{1}{500}\right)^i \left(\frac{499}{500}\right)^{500-i}.$$

利用二项分布的泊松近似, 取 $\lambda = np = 500 \times 1/500 = 1$, 于是上述概率的近似值为

$$P(X \geq 3) \approx 1 - \sum_{i=0}^2 \frac{e^{-1}}{i!} = 1 - e^{-1} - e^{-1} - \frac{e^{-1}}{2} = 0.0803.$$

Exercise 9

25. 设 X 是只取自然数为值的离散随机变量, 若 X 的分布具有无记忆性, 即对任意自然数 n 与 m , 都有

$$P(X > n + m \mid X > m) = P(X > n),$$

则 X 的分布一定是几何分布.

Supplement Exercises

Solution

证 由无记忆性知

$$P(X > n + m | X > m) = \frac{P(X > n + m)}{P(X > m)} = P(X > n),$$

或

$$P(X > n + m) = P(X > n)P(X > m).$$

若把 n 换成 $n - 1$ 仍有

$$P(X > n + m - 1) = P(X > n - 1)P(X > m).$$

上两式相减可得

$$P(X = n + m) = P(X = n)P(X > m).$$

若取 $n = m = 1$, 并设 $P(X = 1) = p$, 则有

$$P(X = 2) = p(1 - p).$$

若取 $n = 2, m = 1$, 可得

$$P(X = 3) = P(X = 2)P(X > 1) = p(1 - p)^2.$$

若令 $P(X = k) = p(1 - p)^{k-1}$, 则用数学归纳法可推得

$$P(X = k + 1) = P(X = k)P(X > 1) = p(1 - p)^k, \quad k = 0, 1, \dots$$

Further Reading

1. Poisson Process

- A **counting process** $N(t)$ represents the total number of "events" that have occurred up to time t .
 - $N(t) \geq 0$;
 - $N(t)$ is integer valued;
 - $N(t)$ is non-decreasing.
 - For $s < t$, $N(t) - N(s)$ represents \dots (?)
- **Independent increments**: The numbers of events that occur in disjoint time intervals are independent.
- **Stationary increments**: The distribution of $N(t) - N(s)$ depends only on $t - s$.

• **Definition 1:** The counting process $N(t)$ is a **Poisson process** with rate λ if

- (i) $N(0) = 0$;
- (ii) It has independent increments;
- (iii) It has stationary increments and the stationary distribution of $N(t + s) - N(s)$ is **Poisson**(λt), i.e.,

$$P(N(t + s) - N(s) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Reference. [Ross] Stochastic Processes

Thank you!