

随机变量 (Random variables)

1. 随机变量简介.
2. 离散随机变量的概率分布
3. 概率分布函数
4. 一些重要的随机变量

1. 随机变量

简介: 无论一个试验给出的 **定性的** 或者 **定量的** 结果, 统计分析方法都需要所给数据的 **数值** 方面的信息, 例如

- 均值 \bar{x} (mean), 或数学期望, 期望,
- 方差、标准差

因此, 人们需要对样本空间中的结果进行数学化 mathematization

这种**数学化**通常是通过随机变量来实现.

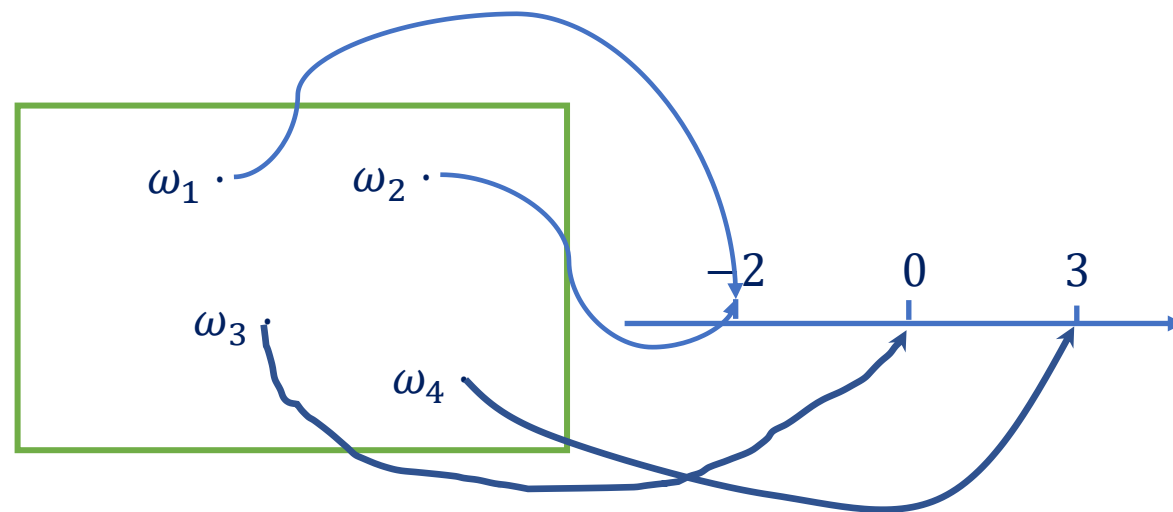
一般地, 试验中的每一个结果都可以与唯一的一个数字相关联

例 1. 抛硬币试验中, 样本空间为
 $\Omega = \{\text{Head}, \text{Tail}\}$.

可定义如下对应法则:

1. “Head” $\rightarrow 1$, “Tail” $\rightarrow 0$;
2. “Head” $\rightarrow \pi$, “Tail” $\rightarrow e$; ...

例2.



这种对应法则（关联法则）称为随机变量——

- 变量: 不同的数值结果都有可能
- 随机: 观测的值依赖于哪一个试验结果出现.

随机变量的定义:

给定某试验的样本空间 Ω , 随机变量是任意一种对应法则: 样本空间中的每一个结果都对应唯一的一个数.

在数学语言中, 随机变量是一个函数 (function)
定义域是样本空间, 值域是由实数构成的某个集合.

随机变量通常由大写字母表示, 例如 X, Y, Z , and etc.

一般形式: $X(\omega) = x$,

x 是样本点 ω 在随机变量 X 下对应的值.

注: 1. 一个随机变量既不是随机的, 也不是一个变量,
它就是简单的一个函数, 定义在样本空间上的函数.

2. 随机变量的取值是随机的, 是一个变量

-- **变量** 不同的可能的值

-- **随机** 不同的值取决于不确定的试验结果.

定义 2.1 设随机试验的样本空间为 Ω , 如果对 Ω 中每一个元素 e , 有一个实数 $X(e)$ 与之对应, 这样就得到一个定义在 Ω 上的实值单值函数 $X = X(e)$, 称之为**随机变量**(random variable).

例 4: 一枚硬币连续抛 10 次. 样本空间是什么?

样本空间中多少个元素?

样本空间上定义个随机变量.

Answer: 样本空间 $\Omega = \{s | s \text{ 是由 10 个 H 或 T 构成的序列}\}$

$$|S| = 2^{10}$$

定义随机变量 $X(s) = s$ 中 H 的个数

那么, X 是一个随机变量, 任意一个 s , H 的个数唯一确定的.

i.e, $X(HHTTHTHHHT) = 6$. $R_X = \{0, 1, 2, \dots, 10\}$.

如何用随机变量定义一个事件?

在例4中: $X(s) = 1$ 表示什么事件?

$\{s \in S \mid \text{由10个 H 或 T 构成的序列中, 仅有1 个 H}\}$, 例如

$HTTTTTTTTTT, THTTTTTTTTTT, TTHTTTTTTTTT, \dots$

在例4中: $0 \leq X(s) < 3$ 表示什么意义?

$\{s \in S \mid \text{由10个 H 或 T 构成的序列中, H的个数小于3}\}$

$$0 \leq X(s) < 3 \Leftrightarrow X(s) = 0 \cup X(s) = 1 \cup X(s) = 2$$

$0 \leq X(s) < 3$ 表示一枚硬币连续抛 10 次, 没有正面, 或者出现1次正面, 或者出现2次正面.

$(X = x)$ 表示事件 $\{s \in S | X(s) = x\}$

$(a < X < b)$ 表示事件 $\{s \in S | a < X(s) < b\}$

$(X \leq b)$ 表示事件 $\{s \in S | X(s) \leq b\}$

它的对立事件(逆, 补) 是? $(X > b)$

所以, $P(X > b) = 1 - P(X \leq b)$

三类随机变量: 离散型, 连续型 & 混合型.

定义: 如果随机变量 X 的值域 (取值的集合) 是可数的, 那么 X 称为离散型随机变量. (可数的, 可列的)

一个只含有有限个元素的有限集一定是可数的;

一个无限集的元素如果能够以一个无穷数列的形式 (第1个, 第2个, ... 等形式) 一一列出的话, 就称为可数的(countable).

自然数集合 N , 有理数集合 Z (rational numbers) 都是可数的.
实数集合 R 是不可数的, 是连续的.

定义: 一个随机变量是连续型的, 如果以下两个条件成立:

1. 值域包含了某个区间的所有的点(有限区间, 或无限区间), 或者不相交区间并集的所有的点.(如 $[0, 10] \cup [20, 30]$)
2. 没有任何一个值的概率为正, 即对任意常数 c , $P(X = c) = 0$

离散型随机变量的概率分布.

随机变量 X 的概率分布是说, 整体的概率1是如何分配到 X 的不同的可能取值上.

X	x_1	x_2	x_3	...	x_k	...
p_k	p_1	p_2	p_3	...	p_k	...

1. $p_k \geq 0$, 对所有的 k , 非负性

2. $\sum_{k=1}^{\infty} p_k = 1$ 归一性

例 5: 某统计系实验室为学生预定了6台计算机. 令 X 表示某一天一个特定时间正在使用的计算机数量. X 的概率分布如下:

X	0	1	2	3	4	5	6
p_k	0.05	0.10	0.15	0.25	0.20	0.15	0.10

计算 $P(X \leq 2)$, $P(X \geq 3)$, $P(2 \leq X \leq 5)$, $P(2 < X < 5)$

$$P(X \leq 2) = 0.05 + 0.10 + 0.15 = 0.30$$

$$P(X \geq 3) = 0.25 + 0.20 + 0.15 + 0.10 = 0.70$$

$$P(2 \leq X \leq 5) = 0.15 + 0.25 + 0.20 + 0.15 = 0.75$$

$$P(2 < X < 5) = 0.25 + 0.20 = 0.45$$

例 6: 假定球员的投篮命中率为 0.8. 现在球员独立投篮两次, 令随机变量 X 为投中次数. 求 X 的概率分布

解: X 可能的取值为 0,1,2 令 A_i 表示第 i 次投篮投中, $i = 1,2$

$$P(X = 0) = P(\bar{A}_1 \bar{A}_2) = 0.2 \times 0.2 = 0.04$$

$$\begin{aligned} P(X = 1) &= P(\bar{A}_1 A_2 \cup A_1 \bar{A}_2) = P(\bar{A}_1 A_2) + P(A_1 \bar{A}_2) \\ &= 0.2 \times 0.8 + 0.8 \times 0.2 = 0.32 \end{aligned}$$

$$P(X = 2) = P(A_1 A_2) = 0.8 \times 0.8 = 0.64$$

X 的概率分布律为,

X	0	1	2
p_k	0.04	0.32	0.64

概率分布函数.

定义： $F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$

$F(x)$ 在 x 处的值等于所有 X 小于或等于 x 的事件概率之和级数的和： X 的所有可能取值可能为无穷多个

也称为累积分布函数 (Cumulative Distribution Function.)

从最小的可能值开始累积, $x_i \leq x$

例 7: 随机变量 X 的概率分布如右所示, 求其概率分布函数.

X	1	2	4	6
p_k	0.05	0.1	0.35	0.5

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.05, & 1 \leq x < 2 \\ 0.15, & 2 \leq x < 4 \\ 0.5, & 4 \leq x < 6 \\ 1, & 6 \leq x \end{cases}$$

注: 用最少的表达式, 表达一个函数

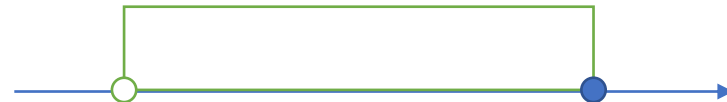
分布函数的性质

定义: $F(x) = P(X \leq x)$.

1. 对任意的 $x_1 < x_2$, $P(x_1 < X \leq x_2)$

$$= P(X \leq x_2) - P(X \leq x_1)$$

$$= F(x_2) - F(x_1)$$



1'.对任意的 $x_1 < x_2$, $P(x_1 \leq X \leq x_2)$

$$= P(X = x_1) + P(x_1 < X \leq x_2).$$

概率分布函数的定义

$$F(x) = P(X \leq x).$$

2. 对任意的 $x_1 < x_2$, $F(x_2) - F(x_1) = P(x_1 < X \leq x_2) \geq 0$

$F(x)$ 是单调不减函数 non-decreasing function

3. $\lim_{x \rightarrow \infty} F(x) = 1$, or $F(\infty) = 1$

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \text{or } F(-\infty) = 0$$

4. $\lim_{x \rightarrow x_0+} F(x) = F(x_0)$, or $F(x)$ 右连续

概率分布函数的性质

$$F(x) = P(X \leq x).$$

$$1. P(X > a) = 1 - P(X \leq a) = 1 - F(a).$$

$$2. P(X < a) = F(a - 0).$$

$$3. P(X = a) = F(a) - F(a - 0).$$

概率分布函数的应用

例 8: Assume that $F(x) = \begin{cases} 1 - e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ is the cumulative

distribution function of a random variable X .

Calculate $P(X \leq 2)$; $P(1 < X \leq 3)$; $P(X = 1)$;

Solution:

$$P(X \leq 2) = F(2) = 1 - e^{-1};$$

$$P(1 < X \leq 3) = F(3) - F(1) = e^{-0.5} - e^{-1.5};$$

$$\begin{aligned} P(X = 1) &= F(1) - F(1 - 0) \\ &= F(1) - \lim_{x \rightarrow 1-} F(x) = 0 \end{aligned}$$

例 9: Assume that $F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1 - e^{-x}, & x \geq 1 \end{cases}$ is the cumulative

distribution function of a random variable X , compute $P(X = 0)$, $P(X = 0.5)$, $P(X = 1)$.

解

$$P(X = 0) = F(0) - F(0-) = F(0) - \lim_{x \rightarrow 0-} F(x) = \frac{1}{2} - 0 = \frac{1}{2}$$

$$P(X = 0.5) = F(0.5) - F(0.5-) = F(0.5) - \lim_{x \rightarrow 0.5-} F(x) = \frac{1}{2} - \frac{1}{2} = 0$$

$$P(X = 1) = F(1) - F(1-) = F(1) - \lim_{x \rightarrow 1-} F(x) = \frac{1}{2} - e^{-1}$$

该随机变量 X 既不是离散型的，也不是连续型的，是混合型的。

4. 一些重要的随机变量

1. 两点分布

如果随机变量 X 只取两个值，概率分布律为

$$P(X = x_1) = 1 - p, 0 < p < 1; \quad P(X = x_2) = p,$$

称 X 服从参数为 p 的两点分布

特别地, 如果 $x_1 = 0, x_2 = 1$, 两点分布也叫 (0-1) 分布, .

X	0	1	Fail,	Success,
p_k	$1 - p$	p		

2. 二项分布 (Binomial distribution)

如果随机变量 X 的分布律为

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, \dots, n$$

称 X 服从参数为 n, p 的二项分布, 记为 $X \sim b(n, p)$

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n C_n^k p^k (1 - p)^{n-k}$$

$$= [p + (1 - p)]^n = 1$$

$$P(A) = p$$

在 n 重伯努利试验中, 事件 A 发生 k 次的概率为

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, \dots, n,$$

例 10: 某一大批产品的合格率为 98%, 现随机从这批产品中抽样 20 次, 每次抽一个产品. 问抽得的 20 个产品中恰好有 k 个合格品的概率 ($k = 1, 2, \dots, 20$) ?

解 这是不放回抽样. 由于这批产品的总数很大, 而抽出的产品的数量相对于产品总数来说又很小, 那么取出少许几件可以认为并不影响剩下部分的合格品率, 因此可以当做放回抽样来处理, 这样做会有一些误差, 但误差不大. 我们将抽检一个产品看其是否为合格品看成一次试验, 显然, 抽检 20 个产品就相当于做 20 次伯努利试验, 以 X 记 20 个产品中合格品的个数, 那么 $X \sim b(20, 0.98)$, 即

$$P\{X = k\} = C_{20}^k (0.98)^k (0.02)^{20-k}, \quad k = 1, 2, \dots, 20.$$

例 2.2

某大学的校乒乓球队与数学系乒乓球队举行对抗赛. 校队的实力较系队强, 当一个校队运动员与一个系队运动员比赛时, 校队运动员获胜的概率为 0.6. 现在校、系双方商量对抗赛的方式, 提出 3 种方案:

- (1) 双方各出 3 人;
- (2) 双方各出 5 人;
- (3) 双方各出 7 人,

3 种方案中均以比赛中得胜人数多的一方为胜利. 问: 对系队来说, 哪一种方案更有利?

解 设系队得胜人数为 X , 则在上述 3 种方案中, 系队胜利的概率分别为

$$(1) P\{X \geq 2\} = \sum_{k=2}^3 C_3^k (0.4)^k (0.6)^{3-k} = 0.352;$$

$$(2) P\{X \geq 3\} = \sum_{k=3}^5 C_5^k (0.4)^k (0.6)^{5-k} \approx 0.317;$$

$$(3) P\{X \geq 4\} = \sum_{k=4}^7 C_7^k (0.4)^k (0.6)^{7-k} \approx 0.290.$$

因此第一种方案对系队最为有利. 这在直觉上是容易理解的, 因为参赛人数越少, 系队侥幸获胜的可能性也就越大.

问题: 当 n 很大, 而 p 很小的时候, 如何计算

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}?$$

令 $np = \lambda$, 那么 $p = \frac{\lambda}{n}$

$$\begin{aligned} C_n^k p^k (1 - p)^{n-k} &= \frac{P_n^k}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^k \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}(-\lambda)} \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, n \rightarrow \infty \end{aligned}$$

定理 泊松定理 Poisson approximation to the Binomial Theorem:
当 $n \rightarrow \infty$, $p \rightarrow 0$ 且 $np = \lambda$, 有,

$$\lim_{n \rightarrow \infty} C_n^k p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

所以, 当 n 充分大, 而 p 充分小的时候, 有以下近似公式,

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

这里 $\lambda = np$.

k	$C_n^k p^k (1-p)^{n-k}$				$\frac{\lambda^k e^{-\lambda}}{k!}$
	$n = 10$ $p = 0.1$	$n = 20$ $p = 0.05$	$n = 40$ $p = 0.025$	$n = 100$ $p = 0.01$	$\lambda = 1 (= np)$
0	0.349	0.358	0.363	0.366	0.368
1	0.385	0.377	0.372	0.370	0.368
2	0.194	0.189	0.186	0.185	0.184
3	0.057	0.060	0.060	0.061	0.061
4	0.011	0.013	0.014	0.015	0.015
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

在实际应用中，该方法是非常有用的，当 $n \geq 100$ ，且 $np \leq 10$.

例 11: 某十字路口每天有大量汽车通过，假设每辆汽车在此发生事故的的概率为 0.001. 如果每天有 5000 辆汽车通过此处，求发生交通事故的汽车数不少于2的概率。

解: 令 X “发生交通事故的汽车数” $X \sim b(5000,0.001), \lambda = np = 5$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - 0.0404 \approx 0.9596.$$

$$F(x) = P(X \leq x) = \sum_{k=0}^x \frac{\lambda^k e^{-\lambda}}{k!}.$$

x	λ								
	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405

例 12: 某人进行射击, 命中率为 0.02. 独立射击 400 次. 求至少击中 2 次的概率。

解: 令 X “击中的次数” $X \sim b(400, 0.02)$,

$$n = 400, p = 0.02, \lambda = np = 8$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - 0.0030 \approx 0.9970.$$

$$F(x) = P(X \leq x) = \sum_{k=0}^x \frac{\lambda^k e^{-\lambda}}{k!}.$$

x	λ								
	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5
0	0.0041	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001
1	0.0266	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008
2	0.0884	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042
3	0.2017	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149
4	0.3575	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403
5	0.5289	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885

3 泊松分布, (Poisson distribution, 普阿松分布)

若随机变量 X 的概率分布律为

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots$$

其中 $\lambda > 0$ 通常称为单位时间或单位面积上比率. $X \sim P(\lambda)$

泊松分布是独立事件在某时间区间上发生次数的分布

λ 也表示单位时间段内平均发生次数, 例如:

某医院每两小时新生儿数目.

每分钟通过某路口的汽车数,

每10 分钟到达汽车站的人数, ...

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots$$

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

$$\sum_{k=0}^{\infty} P(X = k) = \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) e^{-\lambda} = 1.$$

例 13: 如果每分钟通过某桥的汽车数平均为12辆, 求在某一分钟内, 通过该桥的汽车数大于等于17的概率.

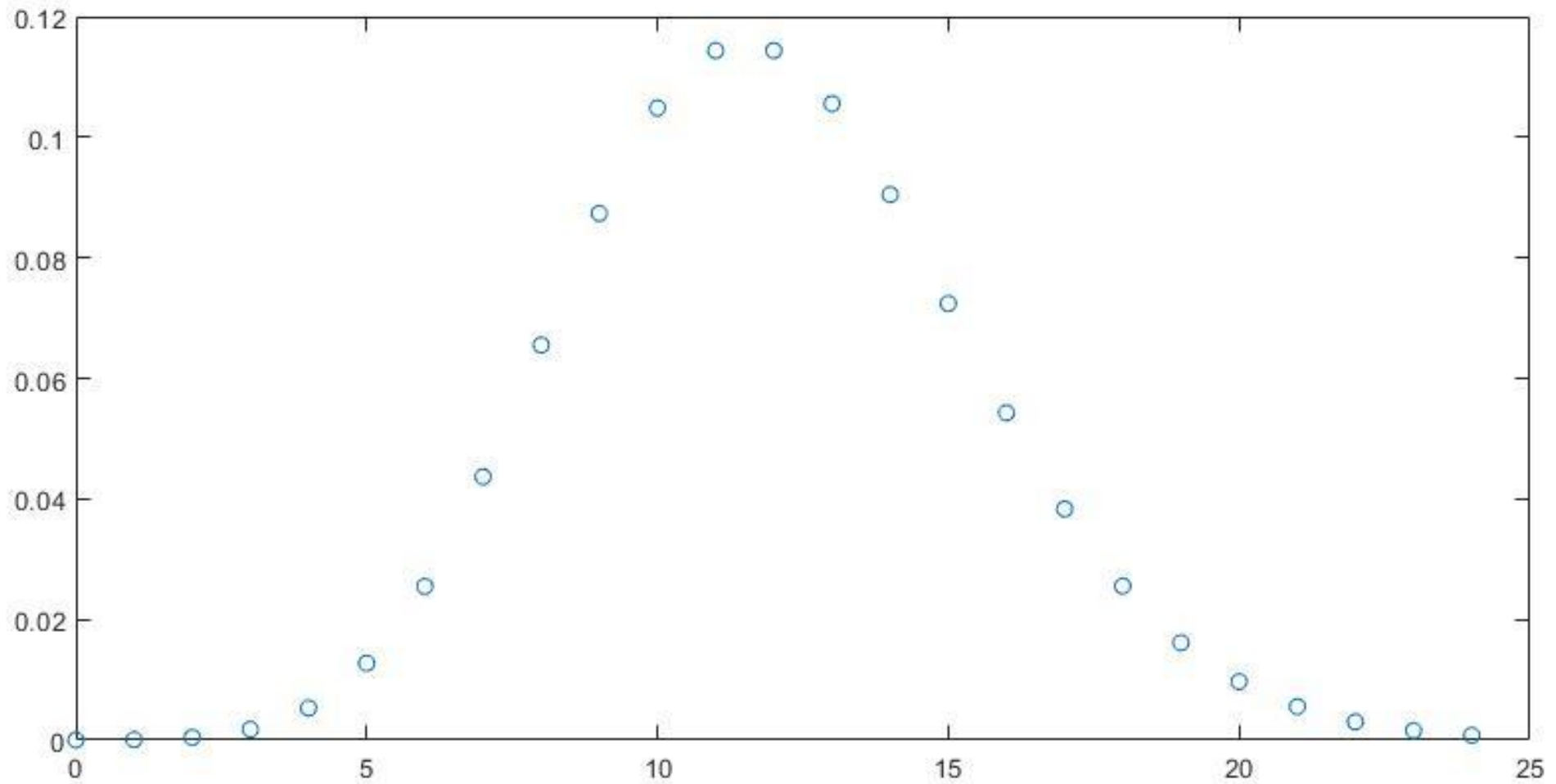
解: $\lambda = 12$. $P(X \geq 17) = 1 - P(X < 17) = 1 - P(X \leq 16)$

$$P(X \leq 16) = \sum_{k=0}^{16} \frac{\lambda^k e^{-\lambda}}{k!} = 0.8987.$$

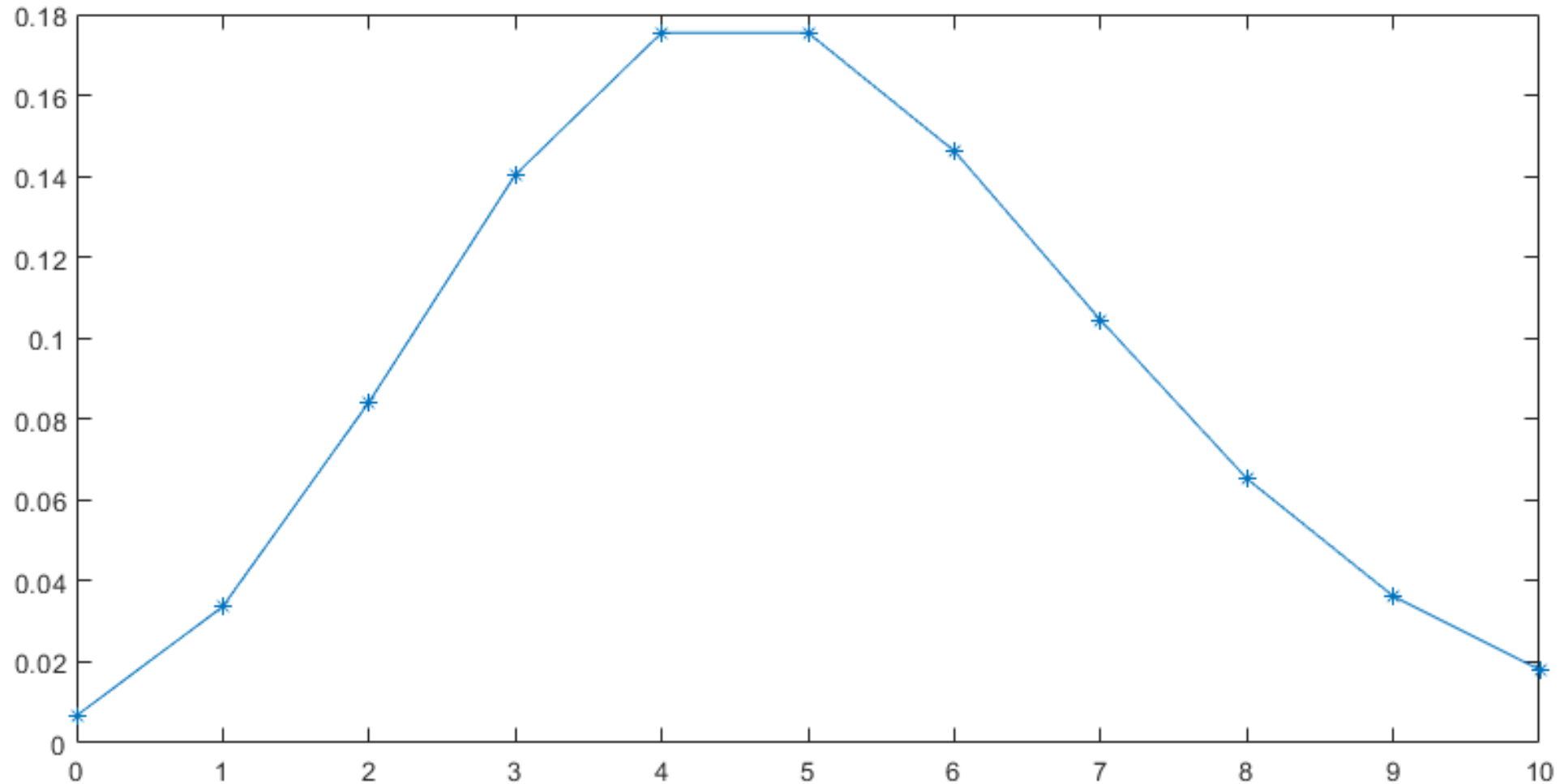
$$\text{故 } P(X \geq 17) = 1 - 0.8987 = 0.1013$$

x	λ								
	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0
0	0.0000	0.0000	0.0000						
1	0.0005	0.0002	0.0001	0.0000	0.0000				
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000		
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3750
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = 12, k = 0, 1, 2, \dots$$



$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = 5, k = 0, 1, 2, \dots$$



练习: 某人进行射击, 有5发子弹, 击中则终止 (子弹用完当然也终止). 设每次击中的概率为 p , X 为射击次数, 求 X 的分布律。

$$P(X = 1) = p \qquad P(X = 2) = p(1 - p)$$

$$P(X = 3) = p(1 - p)^2 \qquad P(X = 4) = p(1 - p)^3$$

特别地, 当 $X=5$ 的时候, 包含两种情况, 射中射不中, 两个事件是互斥的, 和的概率等于概率的和

$$P(X = 5) = p(1 - p)^4 + (1 - p)^5 = (1 - p)^4$$

X	1	2	3	4	5
p_k	p	$p(1 - p)$...		