Probability

Probability

[TOC]

1-1: 机率概论

1. 为什么要研究概率
   - 我们对这个世界了解的太少 这世界的运作有很多是未知的
   - 世间事不见得都是必然的(deterministic),有很多事情是有随机性的(random)
2. 机率与统计的差异
   - 概率
     - *概率模型已知*，要学会怎么算某些事件的概率
     - Ex: 已知一颗骰子为 *公平骰*， 看到偶数的概率是多少? 1/2
   - 统计
     - *概率模型未知*， 要学会 怎么从大量的实验结果中去建立概率模型
     - Ex: 不知一骰是否灌铅，欲知各点出现的 概率模型?

1-2: 集合论

集合和概率的关系

- 学生上课不规矩」的机率 = 0. 1
- P(学生上课不规矩) = 0. 1
- P是概率函数, 概率函数的自变量是：**事件，而事件，是一种集合**

annotation:

- 元素 (Element)
- 集合 (Set)
- 子集 (subset)
- 宇集 (Universal Set)
- 空集 (Empty Set)
- 交集 (Intersection)
- 联集 (Union)
- 补集 (Complement) A^c
- 差集 (Difference)
- 不相交 (Disjoint)
- 互斥 (Mutually Exclusive) 若一群集合 $X_1,X_2,\ldots,X_n$中任选两个集合 X_i,X_j都不想交，则我们称这群集合 $X_1,X_2,\ldots,X_n$ 互斥.

De Morgan's Law：

- $(A\cup B)^c = A^c \cap B^c$
- 证明：在数学上，要证明两个集合是同一个集合的独门心法是： 你中有我，我中有你.
- 重要应用:
    - $(A\cup B)^c = A^c \cap B^c$
    - $\Rightarrow P((A\cup B)^c) = P(A^c \cap B^c)$
    - $\Rightarrow 1 - P(A\cup B) = P(A^c \cap B^c)$
    - 集合变补，符号变反，概率求补

1-3: 概率名词

- 实验 (Experiment)
    - 一个机率「实验」包含了： 步骤 (procedures)、模型 (model)、观察(observations)
    - Ex: 丢两公平骰
        - 步骤：「伸手取起桌上二骰，紧握后，手微微开口后向内吹口气。 之后默祷，再将骰丢入碗中，直至停止为止。」
        - 模型：(1, 1)、(1, 2), ... ,(6, 6) 等发生机会均等
        - 观察：每个骰子的点数 (6, 6)
- 结果(Outcome)
    - 是实验中可能的结果
- 样本空间(Sample Space)
    - 机率实验所有可能的结果的集合，通常用 S 来表示
    - Ex: 连丢三次铜板，记录正反面结果
        - S = { HHH, HHT, HTH, HTT, THH, THT, TTH, TTT }
    - Ex: 幸运之轮转一次
        - $S = [0,1)$
    - Ex:幸运之轮转两次
        - $S = [0,1) \times [0,1)$ (2维)
- 事件 (Event)
    - 「事件」指的是对于实验结果的**某种叙述**
        - Ex: 骰子出来的结果是偶数
    - 机率就是在讲实验结果符合某事件叙述的机会多大
    - 在数学上，「事件」可以看成是「结果」的集合，亦即是 「样本空间」 的子集。
        - Ex:台大生的上课出席状况
          -「结果」有哪几种：准时、迟到、旷课
            - 事件1：有出席；E_1={准时、迟到}
            - 事件2：没规矩；E_2={迟到、旷课}
            - 究竟可能会有多少种事件呢? 2^n
            - { {}, {准时} , {迟到} , {旷课} , {准时, 迟到} , {迟到, 旷课} , {准时, 旷课} , {准时, 迟到, 旷课} }
- 事件空间 (Event Space)
    - 「事件空间」是包含所有事件的集合
    - set of sets
    - 若「样本空间」$s = \{ O_1,O_2,...,O_n \}$ 有 n 个「结果」，「事件空间」大小为 2^n
    - 机率是一个函数，其自变量是：事件

- 所以机率可以看成是一个mapping
    - 机率函数是从「事件空间」映射到 [0,1]区间, P:「事件空间」→ [0,1]
    - 它是事件的函数 (你给一个事件，它吐回一个数字给你)

2-1: 机率公理性质

公理 (Axioms)

1. `P(A) ≥ 0`
2. `P(S) = 1`
3. 事件 A_1, A_2, ... 互斥 => $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$
    - 公理 3 搭起了集合运算与机率运算的桥梁!
    - 看到 集合，互斥 就应该想到公理3

公理衍生之机率性质

Ex: 从一副 52 张扑克牌抽中一张，结果 为 Ace 之机率为何?

Ace = {黑桃A,草花A,红心A,方块A}

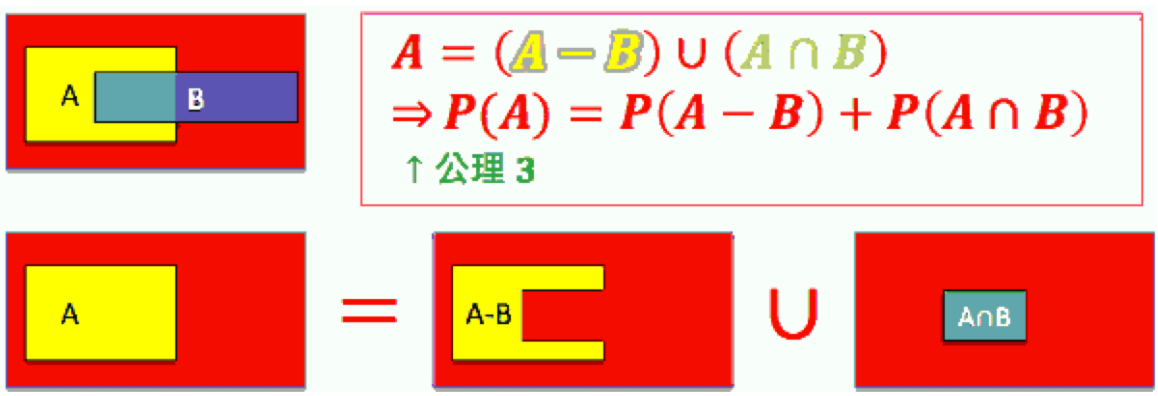
抽中各色A的事件 {黑桃A}, {草花A}, {红心A},{方块A} 互斥，根据公理3，得

P(Ace) = P({黑桃A}) + P({草花A}) + P({红心A}) + P({方块A}) = 1/52 + 1/52 + 1/52 + 1/52 = 1/13

- 若 E = {o1,o2,...,o_n} , o_i 为单个结果， 则
    - P(E) = P({o1}) + P({o2}) + ... + P({o_n)}
- 证明:
    - E = {o1} ∪ {o2} ∪ ... ∪ {o_n)}
        - 一个事件 可以拆为 它的各个单独结果的事件的合集
    - 因 {o1} , {o2} , ... , {o_n)} 互斥
    - 套用公理3 得证

- P(∅) = 0
- P(A) = 1 - P(A^c)

- P(A) = P(A-B) + P(A∩B)
    - 一个事件的概率，可以表示为 它 与另一个事件的差集 和 交集的 概率之和
- 证明:



- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - 两个事件并集的概率，可以用 集合各自的概率 和 它们交集的 概率来表示
- 证明
  - 

$$\begin{aligned} &P(A \cup B) \\ &= \underbrace{P(A-B)}_{P(A)-P(A\cap B)} + \underbrace{P(A\cap B)+P(B-A)}_{P(B)} \\ &= P(A)+P(B)-P(A\cap B) \end{aligned}$$

- Ex: 在大陆随便碰上一个人，此人爱甜豆花或爱咸豆花 机率为何?
  - P(爱甜 ∪ 爱咸) = P(爱甜) + P(爱咸) − P(爱甜 ∩ 爱咸) = ⋯

- 切面包定理:
  - 若$C_1,C_2,\ldots,C_n$ 互斥，且 $C_1 \cup C_2 \cup \ldots \cup C_n = S$, 则
  - 对任何事件 A: $P(A) = P(A\cap C_1) + P(A\cap C_2) + \ldots + P(A \cap C_n)$
- 证明
  - ![][1]
- Ex: 阿宅心仪某可爱女店员。她的笑容打开了他封 闭的心。阿宅注意到她笑容会受生意的影响，于是 每天忠实记录该店生意与她有无对他笑。店生意有 满、普、惨三态，而她有笑、怒二态。根据记录:
  - $$S = \left\{ \begin{array}{l} (\text{满} \cap \text{笑}) \rightarrow \frac{1}{20}, (\text{满} \cap \text{怒}) \rightarrow \frac{2}{20} \\ (\text{普} \cap \text{笑}) \rightarrow \frac{5}{20}, (\text{普} \cap \text{怒}) \rightarrow \frac{4}{20} \\ (\text{惨} \cap \text{笑}) \rightarrow \frac{5}{20}, (\text{惨} \cap \text{怒}) \rightarrow \frac{3}{20} \end{array} \right\}$$

  - P(笑) = 1/20 + 5/20 + 5/20 = 11/20

- 若 $A \subset B$, 则 $P(A) \le P(B)$

- Boole's 不等式
  - 对任意 n 个事件 $A_1 , A_2 , \ldots , A_n$ 而言, 没有互斥要求
    - $P(\cup^n_{i=1} A_i) \le \Sigma^n_{i=1} P(A_i)$

- Bonferroni's 不等式
  - 对任意n个事件$A_1 , A_2 , \ldots , A_n$ 而言, 没有互斥要求
    - $P(\cap^n_{i=1} A_i) \ge 1-\Sigma^n_{i=1} P(A_i^c)$

条件概率性质:

对任何事件 X 及任何条件事件 Y， 我们有:

- 性质 1：$P(X|Y) = \frac{P(X\cap Y)\geq 0}{P(Y)\geq 0} \geq 0$
- 性质 2：$P(Y|Y) = \frac{P(Y\cap Y)}{P(Y)} = \frac{P(Y)}{P(Y)} = 1$
- 性质 3：A, B 互斥 $\Rightarrow P(A\cup B\,|Y) = \frac{P(A)}{P(Y)} + \frac{P(B)}{P(Y)} = P(A|Y) + P(B|Y)$

Total Probability 定理

- 若$C_1,C_2,\ldots,C_n$ 互斥,且 $C_1 \cup C_2 \cup \ldots \cup C_n = S$, 则对任意事件 A ， 我们有：
  - **$P(A) = P(A|C_1)|P(C_1) + P(A|C_2)|P(C_2) + \ldots + P(A|C_n)|P(C_n)$**
- 其实就是切面包定理的 条件概率版本
  - ![][1]
  - Proof: 切面包定理 => $P(A) = \Sigma^n_{i=1} P(A\cap C_i) = \Sigma^n_{i=1} P(A|C_i)\cdot P(C_i)$
- Ex: 阿宅 vs. 可爱店员：店员对阿宅笑否， 受店的生意影响很大。已知
  - P(满)=1/4, P(普)=1/4, P(惨)=1/2 ; P(笑 | 满)=1/6, P(笑 | 普)=2/6, P(笑 | 惨)=3/6; 问 P(笑) = ?
  - P(笑) = $\Sigma^3_{i=1}P(A|C_i)\cdot P(C_i)$ = 1/6·1/4 + 2/6·1/4 + 3/6·1/2 = 9/24

Bayes' Rule

- 若$C_1,C_2,\ldots,C_n$ 互斥,且 $C_1 \cup C_2 \cup \ldots \cup C_n = S$, 则对任意事件 A ， 我们有：
  - **$P(C_j|A) = P(A|C_j)P(C_j) / \Sigma^n_{i=1} P(A|C_j)\cdot P(C_j)$**
- Proof:
  - $P(C_j|A) = \frac{P(C_j\cap A)}{P(A)} = \frac{P(A\mid C_j)\cdot P(C_j)}{\Sigma^n_{i=1} P(A|C_i)\cdot P(C_i)}$
  - $P(C_j|A) =$
- Bayes' Rule 用在很多时候，我们关心的事件(A) 根条件事件C 互换位置。
  - 其实就是全概率定理 和 条件概率的 联合使用
- Ex: 一日，老板见可爱店员笑， 请问在此情况下，当日生意满座之机率为何?
  - P(满|笑) = P(满∩笑)/P(笑) = P(笑|满)·P(满) / (9/24) = 1/6·1/4 / (9/24) = 1/9
  -

3.1 机率的独立性 (Independence)

- 常见定义:若两事件A,B 之机率 满足
  - `P(A∩B) = P(A)·P(B)`
  - 则 两事件称为机率上的独立事件
- 常见定义:若两事件A,B 之机率 满足
  - P(A|B) = P(A) # <=> P(A∩B)/P(B) = P(A) <=> P(A∩B) = P(A)·P(B)
  - 则 两事件称为机率上的独立事件
- Ex: 已知某生秦始皇作业表现与机率 作业表现相互独立。若秦始皇作业未作机率 为 0.2，机率作业未作机率 为 0.3。 问两科作业同时未作之机率为?
  - 0.2 * 0.3 = 0.06

如何判断事件之间是否机率上相互独立

- Ex: 已知某校生爱乱停车。水源阿伯拖车 时常有车主赶回求情。出现愤宅求情之机率为0.3。 一般而言被人求情阿伯会放行的机率为 0.2 。 水源阿伯，大公无私，天下皆知。 问某日阿伯拖某车时出现愤宅求情且车未放行的机率?
    - 出现愤宅求情 P = 0.3 , 车未放行 P = 1-0.2 = 0.8
    - 出现愤宅求情且车未放行 P = 0.3 * 0.8 = 0.24
- Ex: 某古锥姊有时会在活大停车不当。被拖时若及时赶回求情，古锥姊常一手 摀嘴，一手指着车曰:「啊，那是我的车!」巧笑倩兮。
    - 若以 A 代表古锥姊实时赶回求情之事件，B 代表阿辈放回古锥姊车的事件。根据某愤宅多日观察古 锥姊行为:
    - P(古锥姊未能求情 且 车未放行) = 0.85, P(古锥姊未能求情 且 车放行) = 0.05
    - P(古锥姊及时求情 且 车未放行) = 0.01, P(古锥姊及时求情 且 车放行) = 0.09
    - 愤宅泪眼悲愤控诉阿伯:「你不公平!!!」。 问:水源阿伯清誉，岂容愤宅任意污蔑!愤宅悲愤有理否?吾人该否为其一掬同情之泪?
    - 解:
        - P(古锥姊求情) x P(车放行) =? P(古锥姊及时求情 且 车放行) ?
        - {古锥姊求情} = {古锥姊求情 且 车未放行} ∪ {古锥姊求情 且 车放行}
        - P(古锥姊求情) = 0.01 + 0.09 = 0.1 ; 前页曾述 P(车放行) = 0.2
        - P(古锥姊求情) * P(车放行) = 0.1 * 0.2 = 0.02 ≠ 0.09 不独立 ！！！
        - 实际上， P(车放行|古锥姊求情) = 0.09/0.1 = 0.9 !!!

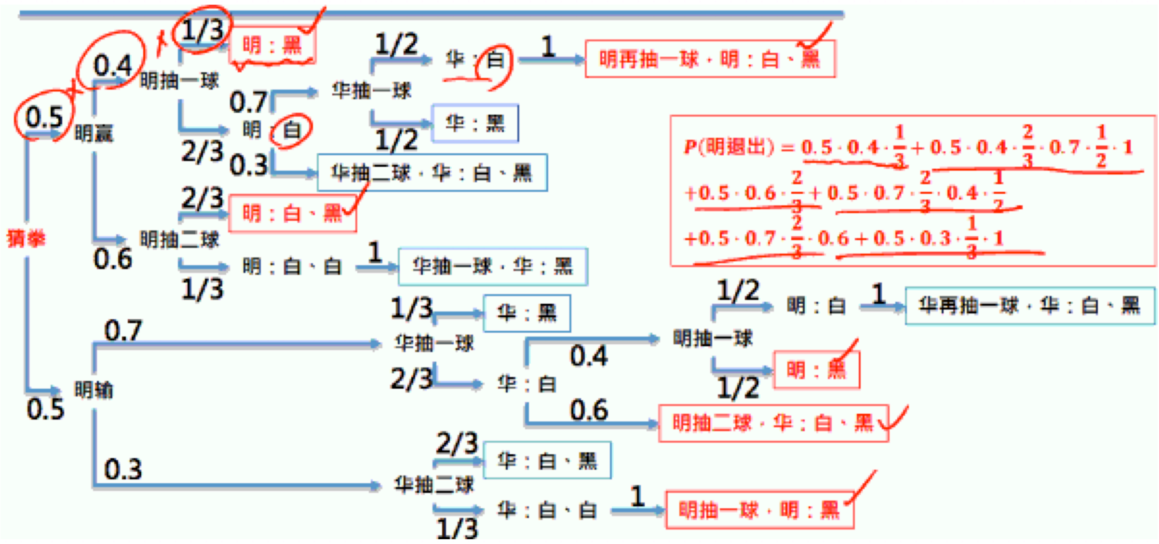
多事件之独立

- 若事件 A_1,A_2,...,A_n 满足下列条件，则称此 n 事件独立(n>2):
    - 从中任选 m 事件 A_{i1},A_{i2},...,A_i_m 均满足
    - P(A_{i1}∩A_{i2}∩...∩A_i_m) = P(A_{i1})P(A_{i2})···P(A_i_m), m = 2,3,... , n.
    - 需要检查所有可能组合之间的独立.

3.2 图解繁复机率

- 当碰到很复杂的机率问题时
    - 先观察这个问题的实验结构
    - 这实验是否能分解成数个子实验?
    - 若可以，则可以利用图解法!

- 范例: 兄弟情
    - 明、华兄弟情笃。故决定一人放弃追求小美以免伤情 谊。于罐中放入两白球、一黑球。游戏规则如下:「猜拳决定谁先，之后轮流罐中取球；每次可取一至二球，直至有人抽 中黑球为止。抽中黑者退出追求。」
    - 已知猜拳输赢机率为 0. 5，每次明取球取一颗之机率为 0. 4，取两颗机 率为 0. 6 。每次华取球取一颗之机率为 0. 7，取两颗机率为 0. 3。
    - 问最后小明退出追求之机率为?



3-3: 数数算机率

– 古典机率常假设每个实验结果（outcome）发生机率相同
– 故计算某事件机率之问题，等同于计算此事件包含多 少实验结果（outcome）。故计算器率等价于数数问题
!

数数基本原则 (Fundamental Principle of Counting)

- 若某种实验有 n 种不同结果，而另一种 实验有 m 种不同结果。若操作此两实验将有n·m 种 不同结果。
    - Ex: 下午茶有 5 种甜点的选择， 10 种饮料的选择。
        - 共有多少种下午茶组合？ 5x10 = 50

- 排列 (Permutation)
- Ex: 小美周末两日惯购物。小美常自明华园 三兄弟找人接送。若两日司机不可重复，问有多少种结果？
    - 3x2 = 6
- 若有 n 异物，从中依序取出 k 物 共有多少种结果？
    - n! / (n-k)!

- 重复选取 (Choose with Replacements)
- Ex: 小美周末两日惯购物。小美常自明华园 三兄弟找人接送。无耻小美竟敢不排除连续凹人两天， 问有多少种结果？
    - 3x3 = 9
- 若有 n 异物，从中选取一物，每 次取完放回。依序选取 k 次，共有多少种 结果？
    - n^k

- 组合 (Combination)
- Ex: 小美爱玩跳棋。小美常自明华园三兄弟 找两人下棋。问有多少种对战组合？
    - 3x2 / 2 = 3
- 若有 n 异物，从中取出 k 物 共有多少种结果？
    - n! / ((n-k)! · k!)

多项组合 (Multinomial)

- Ex: 费雯兄惯于网络八卦版上发废文。第一楼推文常有四类
  - 你妈知道你在发废文吗
  - 见此唉滴必嘘
  - 在五楼…
  - 妈！我在这！

- 问：费雯兄发文 10 次，一楼推文共有多少种组合?
  - 4^{10}

- 有多少组合会看到 4 次「你…」，3 次「见…」，2 次「在五楼…」，1 次「妈…」?
  - $C_4^{10} \cdot C_3^6 \cdot C_2^3 \cdot C_1^1 = 10! / (4! \cdot 3! \cdot 2! \cdot 1!)$

- 若有 m 种异物，每次选物从中选一后放回，依序选 n 次。如此共有 m^n 种实验结果。其中在这 m^n 种 实验结果中，第 1 种异物出现 n_1次且第 2 种异物出现 n_2 次且… 且第
  m 种异物出现 n_m 次，这样的实验结果共有多少种?
  - $n! / (n_1! \cdot n_2! \cdots n_m!)$ <- multinomial coefficient

数数如何应用在算机率上

- 若一事件包含数个实验结果 (outcome)，且每个实验结果发生的机率都一样
  - 先计算任一个实验结果 的机率
  - 再计算该事件共包含多少个实验结果
  - 两者相乘便得到该事件的机率！

- 范例：费雯兄: 继前述费雯兄好发废文之例。若根据统计，费雯兄一楼推文不同 型态之出现机率为:
  - P(「你妈知道你在发废文吗」) = 0.4
  - P(「见此唉滴必嘘」) = 0.2
  - P(「在五楼…」) = 0.1
  - P(「妈！我在这！」) = 0.5

- 问：若费雯兄发文 6 次，会在一楼推文看到 2 次「你…」， 2 次「见…」， 1 次「在五楼…」，1次「妈…」。这样的机率为?
  - P(你你见见五妈) = 0.4*0.4*02*0.2*0.1*0.3 = 0.000192
  - P(2你,2见,1五,1妈) = 0.000192 * 6!/ (2!2!1!1!) = 0.0346
  -

4.1 随机变数 (RANDOM VARIABLE)

- 考虑前面费雯兄的例子, 若根据统计，费雯兄一楼推文 不同型态只有4种, 若
  - P(「你妈知道你在发废文吗」) = 0. 4
  - P(「见此唉滴必嘘」) = 0. 2
  - P(「在五楼…」) = 0. 1
  - 推测 P(「妈!我在这!」) 的概率
  - P(「妈!我在这!」) = 1 − P(「你妈知道你在发废文吗」) − P(「见此 唉滴必嘘」) − P(「在五楼…」) = 0. 3

- 光写字就累翻了!!!
- 若改为:
  - 「你妈知道你在发废文吗」:X = 0
  - 「见此唉滴必嘘」:X = 1
  - 「在五楼...」:X = 2
  - 「妈!我在这!」:X = 3
  - 根据统计: P(X=0)=0.4;P(X=1)=0.2;P(X=2)=0.1;P(X=3)=0.3
    - P(X=3) = 1 - P(X=0) - P(X=1) - P(X=2)
  - 跟前面比起来，你觉得如何呢? 这...这...真是太清爽、太给力了

- 随机变数 (Random Variable, R.V.) 这是一个用来把实验结果 (outcome) *数字化* 的表示方式
- 目的是可以让机率的推导更数学、更简明
- 前面例子中的 X 就是所谓的 *随机变数*
- 随机变数通常都是用 *大写的英文字母* 表示!

探究它的本质!

- 随机变量的本质是什么?
  - 本质是一个 *函数*
  - 「你妈知道你在发废文吗」:X = 0 => X(「你妈知道你在发废文吗」) = 0
- 随机变数 X 其实是一种函数，喂 X 吃一个 outcome ， 就吐出一个对应的数字。数学上的表示法:
  - $X: S \rightarrow \mathbb{R}$

随机变数的种类

- 离散随机变数 (Discrete R. V.)
  - Ex:宅vs.店员:X(微笑)=0,X(不笑)=1 => X = 0, X = 1
  - Ex:小明告白多少次才成功:X(0次)=0,X(1次)=1,X(2次)=2,... => X = 0, X = 1, X = 2, ...
  - 离散 R.V. 的值是有限个，或是「可数的」无穷多个
- 连续随机变数 (Continuous R. V.)
  - 幸运之轮: X 可以是 0 到 1 间内的任意数字
  - 连续 R.V. 的值是有无穷多个，而且是「不可数」的无穷多个

神马叫可数?神马叫不可数?

- 重要性质:0 到 1 之间的所有数字的集合是不可数的!
- 「正整数的集合」 跟 「正偶整数」的集合 相比，哪个集合里面东西比较多? 一样多
- 「长度为一的线段上的点」跟「边长为一的正方形上的点」， 这两个集合，哪一个点的数量比较多? 一样多
- 因为都可以找到一对一对应的方法。

随机变量的函数?

- 阿宅若看到店员微笑，就会点200的套餐。如果店员不笑，他就买15 的饮料。 请问阿宅的消费金额 W是随机变数吗?
  - 店员表情可以由随机变量 X 代表: X(微笑) = 0, X(不笑) = 1
  - W是 X 的函数:W(X(微笑)) = 200, W (X(不笑)) = 15

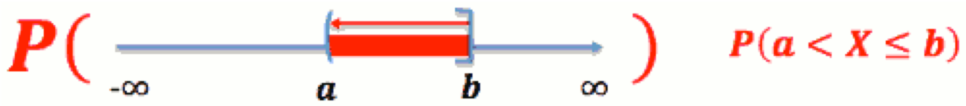
- 所以 W也是喂 outcome 吐数字!因此 W也是一个随机变数!
- 记住:随机变量的函数，也是一个随机变量喔!

4.2 累积分布函数 CDF (CUMULATIVE DISTRIBUTION FUNCTION)

- 对任一个随机变数 X ，我们定义 其 CDF 为函数:
  - $F_X(x) = P(X \leq x)$
  - 

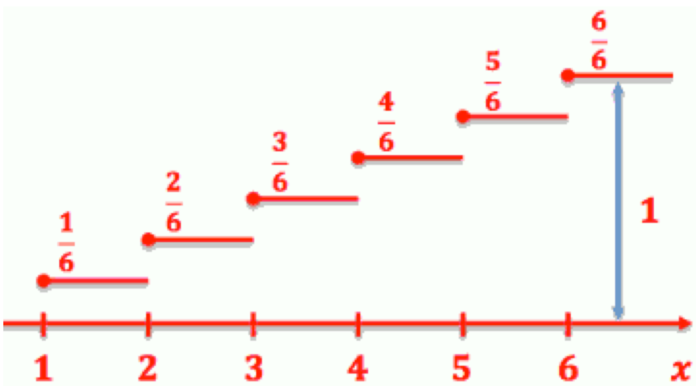
  - 其中 X 是随机变数
- Ex 幸运之轮 $F_X(0.5) = P(X \leq 0.5) = 1/2$

CDF 有什么用?

- 最有用的用途: 计算 X 落在某范围内的机率
  - $P(3 < X \leq 5) = P(X \leq 5) - P(X <= 3)$
  - **$= F_X(5) - F_X(3)$**
    - 

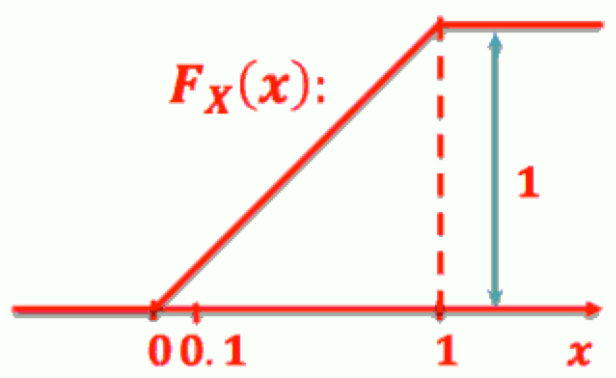
- $P(a < X \leq b) = F_X(b) - F_X(a)$
- $P(a \leq X \leq b) = F_X(b) - F_X(a) + P(X=a)$

离散随机变数的 CDF 长怎样?

- Ex:X为骰子的点数，故P(X=1) =P(X=2) =P(X=3) =P(X=4) =P(X=5) =P(X=6) =1/6
- CDF: $F_X(x) = P(X \leq x)$
  - 

- $P(3 < X \leq 5) = F_X(5) - F_X(3) = 5/6 - 3/6 = 2/6$
- $P(3 < X < 5) = P(3 < X \leq 5^-) = F_X(5^-) - F_X(3) = F_X(5) - P(X=5) - F_X(3) = 1/6$

连续随机变数的 CDF 长怎样?

- Ex: X 为幸运之轮所停下的数字，X∈[0,1)
  - CDF: $F_X(x) = P(X \le x)$
  - 

- $P(0.3< X\le 0.5) = F_X(0.5) - F_X(0.3) = 0.5 - 0.3 = 0.2$
- $P(0.3< X< 0.5) = F_X(0.5^-) - F_X(0.3) = 0.5 - 0.3 = 0.2$

CDF 的性质

- 离散随机变数之CDF:
  - $F_X(x^+) = F_X(x)$
  - $F_X(x^-) = F_X(x) - P(X=x)$
- 连续随机变数之CDF:
  - $F_X(x^+) = F_X(x^-) = F_X(x)$
- 共同性质
  - $F_X(-\infty) = P(X\le-\infty) = 0$
  - $F_X(\infty) = P(X\le \infty) = 1$
  - $0 \le F_X(x) \le 1$

4.3 机率质量函数 PMF (PROBABILITY MASS FUNCTION)

- 只有 离散随机变数 有 PMF
- 对任一个整数值的 *离散随机变数* X , 我们定义其 PMF 为函数:
  - $p_X(x) = P(X=x)$
  - p 大小写无所谓?
- Ex: X 为公平骰子之点数
  - $p_X(3) = P(X=3) = 1/6$

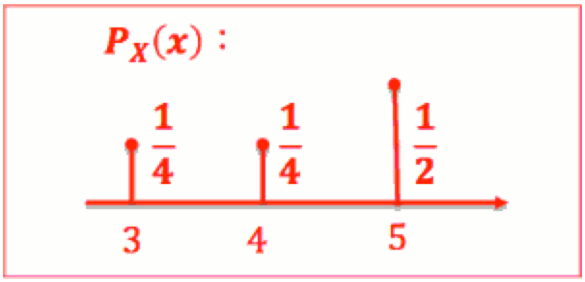
PMF 跟 CDF 的关系?

- PMF -> CDF
  - 对任何 x：

    $$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n)$$
- CDF -> PMF

- $P_X(x) = F_X(x^+) - F_X(x^-)$

机率分布 (Probability Distribution)

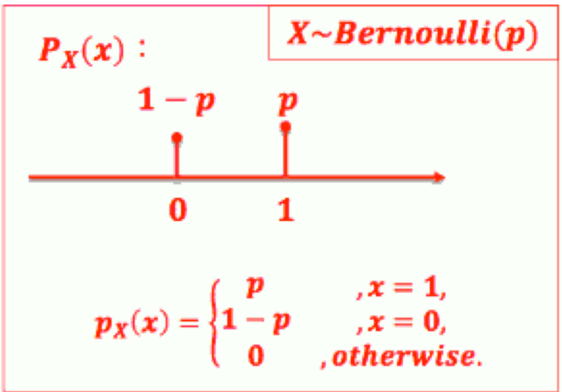
- 任何一个 PMF(或是之后介绍的 PDF)都称作是一种 *机率分布*(将总和为 1 的机率分布在点上之故)
    - 

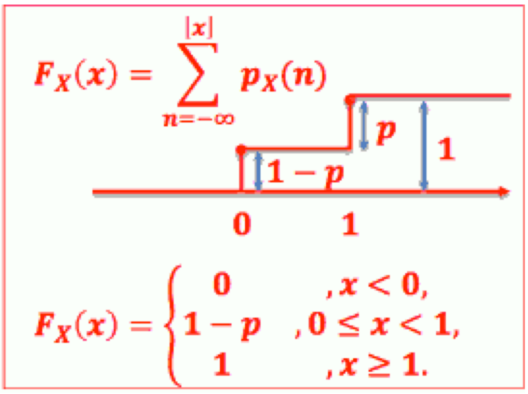
4.4 离散机率分布 I (DISCRETE PROBABILITY DISTRIBUTIONS)

- 观察一下
    - 丢掷铜板:非正面，即反面，正面机率为 0.5
    - 出门天气:非晴天，即雨天，晴天机率为 0.6
- 1 次实验，2 种结果。 在意某结果发生否 Bernoulli 机率分布

Bernoulli 机率分布

- PMF: 若实验成功机率为 p , 作 1 次实验， X 表成功次数
- CDF 见右图





Binomial 机率分布

- 观察一下
    - 阿宅鼓起勇气搭讪 10 人，若每次搭讪成 功机率为 0.6，10 次成功 8 次的机率为?
    - 一周 5 天午餐在晓福买魔石汉堡，若每次制作超时机率为 0.9 5 天中有 3 天制作超时的机率为?
    - 一周有 3 系夜，在活大乱停车 3 次，若每次遭阿伯拖之机率 为 0.8，那这 3 次被拖 2 次之机率为?
- 作 n 次实验，1 个机率，在意 n 次实验出 现某结果 k 次之机率 --> Binomial 机率分布
- 若实验成功几率为 0.6， 做10次实验， X表示成功次数

$$X \sim BIN(10, 0.6)$$

$$p_X(8) = P(X = 8) = \binom{10}{8} 0.6^8 (1-0.6)^{10-8}$$

$\#成功 = 8$

$\#失败 = 10 - 8$

$$F_X(x) = \sum_{m=-\infty}^{\lfloor x \rfloor} p_X(m) = \sum_{m=-\infty}^{\lfloor x \rfloor} \binom{10}{m} \cdot 0.6^m \cdot (1-0.6)^{10-m}$$

- left is PMF, right is CDF

- PMF: 若实验成功机率为 p, 作 n 次实验， X 表成功次数

$$X \sim BIN(n, p)$$

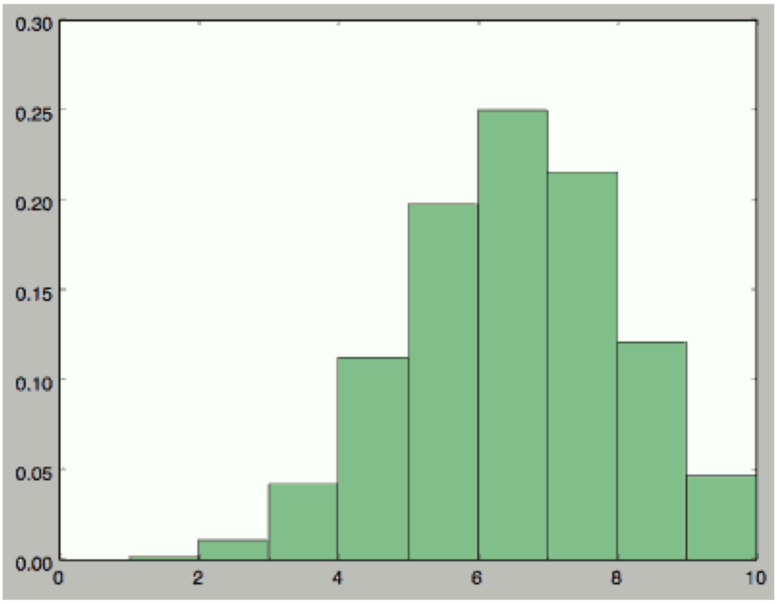
$$p_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$\#成功 = x$

$\#失败 = n - x$

$$F_X(x) = \sum_{m=-\infty}^{\lfloor x \rfloor} p_X(m) = \sum_{m=-\infty}^{\lfloor x \rfloor} \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m}$$

```
import numpy as np
import matplotlib.pyplot as plt
x1 = np.random.binomial( 10, 0.6 , 100000  )
plt.hist(x1, normed=1,  facecolor='green', alpha=0.5)
plt.show()
```

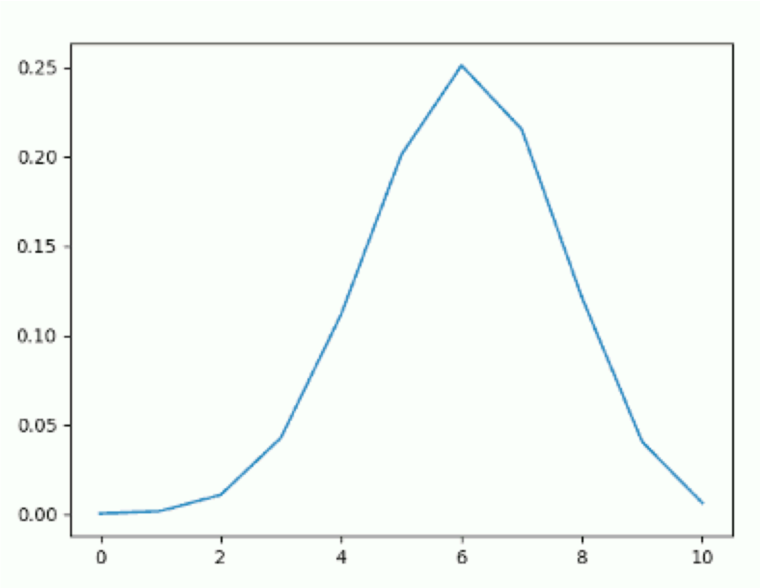


- 为什么这么像 正态分布?

- another method, use scipy

```
import scipy, scipy.stats
```

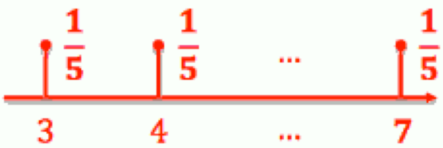
```
x = scipy.linspace(0,10,11)
pmf = scipy.stats.binom.pmf(x,10,0.6)
import pylab
pylab.plot(x,pmf)
pylab.show()
```



Uniform 机率分布

- 观察一下
  - 丢公平骰:1 到 6 各点数出现机会均等
  - 混哥考试:作答 A, B, C, D 机会均等
  - 狡兔三窟:出现在窟 1、窟 2、窟 3 机会均等
- 1 次实验，n 种结果，各结果机率均等。在意某结果发生否 --> Uniform 机率分布
- 如果 X 等于 3,4,...,7 的机率均等

$P_X(x)$:　　$X \sim UNIF(3,7)$

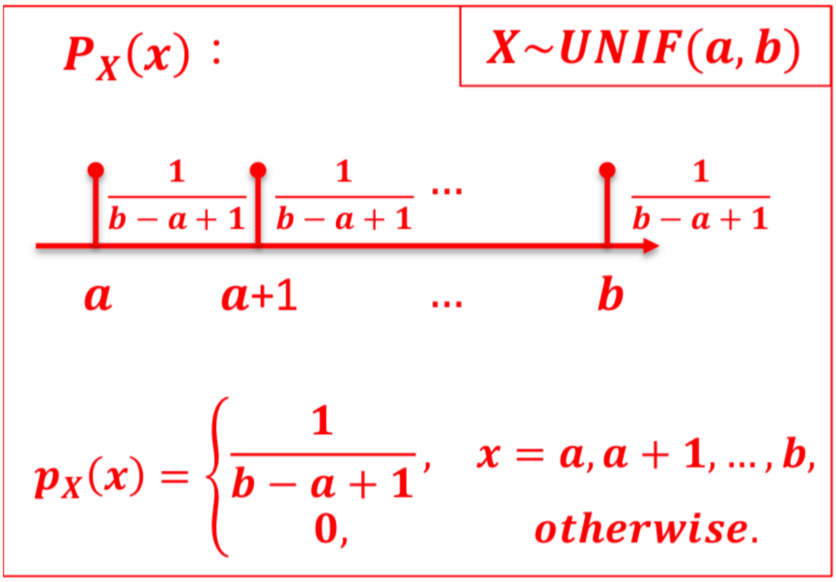


$$p_X(x) = \begin{cases} \frac{1}{7-3+1} = \frac{1}{5}, & x = 3,4,\ldots,7, \\ 0, & otherwise. \end{cases}$$

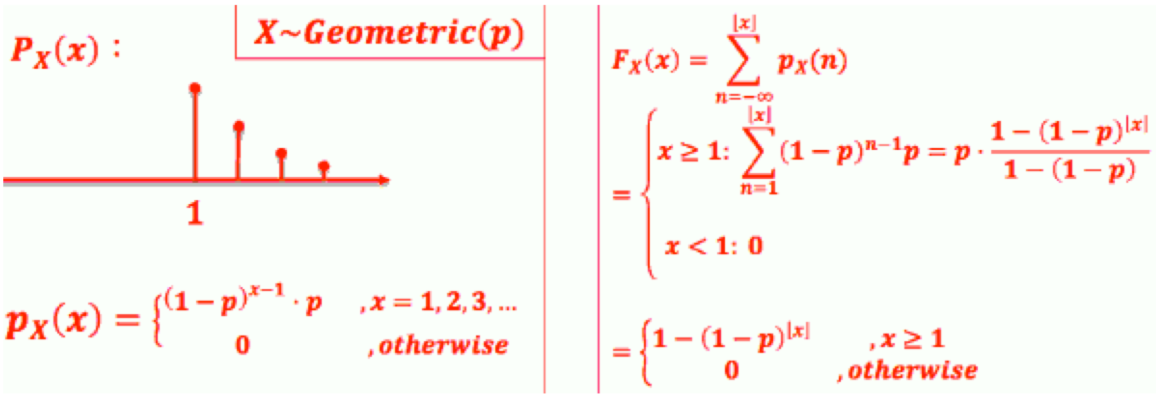
$$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n) = \begin{cases} 0 & ,x < 3, \\ \frac{\lfloor x \rfloor - 3 + 1}{5} & ,3 \le x \le 7, \\ 1 & ,x > 7. \end{cases}$$

)

- 如果X等于a,a+1...,b 的机率均等

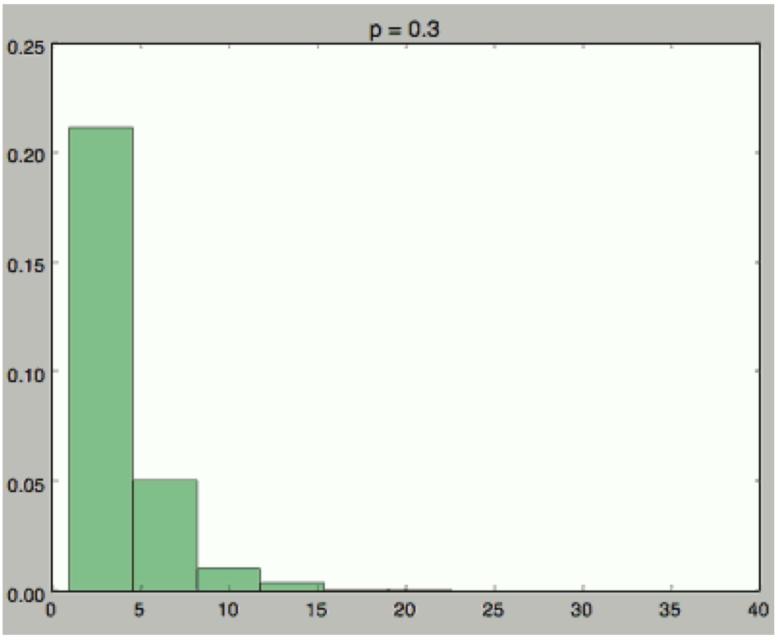
- 

Geometric 机率分布

- 观察一下
  - 阿宅告白:成功机率为 0.3，不成功誓不休。 问到第 5 次才告白成功之机率?
  - 孙文革命:成功机率为 0.1，不成功誓不休。问到第 11 次才 成功之机率?
  - 六脉神剑:那纠缠狂妈宝废物段誉每次要打出六脉神剑，打 的出来的机率为 0. 1。他在 10 次才打出六脉神剑的机率?
- 实验中出现某结果机率已知，重复操作实验至该结果出现为止。 在意某结果是在第几次实验才首次出现 --> Geometric 机率分布
- 六脉神剑:那妈宝废物段誉每次要打 六脉神剑，打的出来的机率为 0.1。他在第 10 次 才打出六脉神剑的机率?
  - 败败败败败败败败败成 => 机率 = 0.9^9 x 0.1
- 六脉神剑:那妈宝废物段誉每次要打 六脉神剑，打的出来的机率为 p 。他在第 X 次 尝试才成功打出六脉神剑。 X = x 的机率?
  - 机率 = $(1\text{-}p)^{X\text{-}1} \cdot p$
- 若实验成功机率为 p，尝 试到成功为止，作了 X 次尝试
  - 

- 有失忆性！ 离散分布中唯一的失忆性分布

```python
import numpy as np
import matplotlib.pyplot as plt
x1 = np.random.geometric( 0.3 , 100000  )
plt.hist(x1, normed=1, facecolor='green', alpha=0.5)
plt.title('p = 0.3')
plt.show()
```



- 近似 指数分布！

Pascal 机率分布

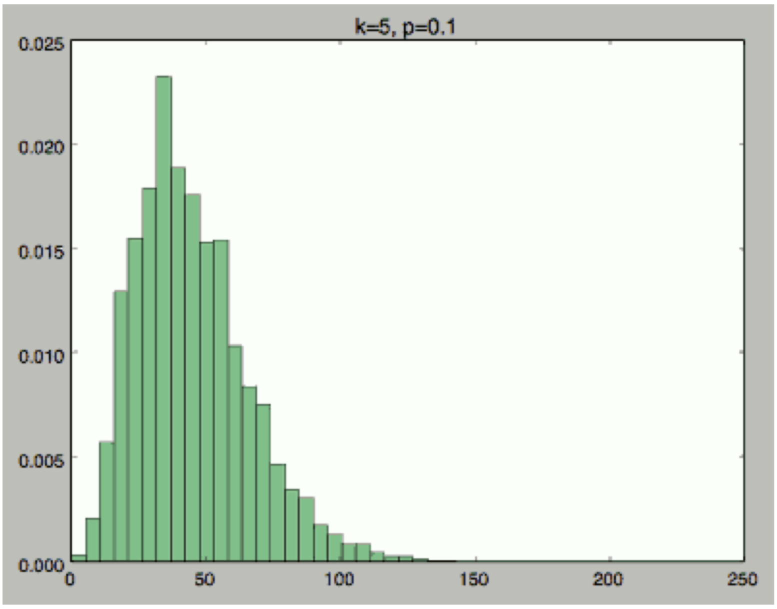
- 观察一下
    - 自尊阿宅:阿宅邀约店员失败机率为 0.9， 若邀约失败达 4 次，阿宅便会自尊有损而放弃追求。问在阿 宅第 7 次邀约时决定放弃追求之机率?
    - 六脉神剑:妈宝废物段誉每次打成功 5 次六脉神剑便功力耗 尽。若每次打的出来的机率为 0. 1。请问他在第 9 次时刚好 功力耗尽的机率?
- 实验中出现某结果机率已知，重复操作实验至该结果出现第 k 次 为止。在意到底在第几次实验才结束 --> Pascal 机率分布
- 六脉神剑:那妈宝废物段誉每次要打 六脉神剑，打的出来的机率为 0.1。成功 5 次便功 力耗尽。请问他在第 9 次时刚好功力耗尽的机率?
    - 可能情况之一:败 成 败 成 败 成 成 败 成
    - 此情况机率 = 0.9^4 x 0.1^5
    - 刚好第9次才成功第5次的情况有几种? C(8,4)·C(1,1) = C(8,4)
    - 所求机率 = C(8,4) x 0.9^4 x 0.1^5
- 六脉神剑:那妈宝废物段誉每次要打六脉神 剑，打的出来的机率为 p 。成功 k 次便功力耗尽。 他在第 X 次尝试才成功打出 k 次六脉神剑。 X = x 的机率?
    - C(x-1, k-1) x $(1-p)^{x-k}$ x p^k
- 若实验成功机率为 p , 到第 k 次成功为止共作了 X 次
    - $\boxed{X \sim Pascal(k,p)}$

      $$p_X(x) = \begin{cases} \binom{x-1}{k-1}(1-p)^{x-k}p^k & , x = k, k+1, \dots \\ 0 & , otherwise \end{cases}$$

      $$\begin{aligned} F_X(x) &= P(X \le x) \\ &= P(\text{第 } k \text{ 次成功在第 } x \text{ 次以前发生}) \\ &= P(\text{在 } x \text{ 次实验中} \ge k \text{ 次成功}) \\ &= P(Y \ge k), Y \sim BIN(x,p) \end{aligned}$$

      ※故 *Pascal* 又称作 *Negative Binomial*

```python
import numpy as np
import matplotlib.pyplot as plt
x1 = np.random.negative_binomial( 5, 0.1 , 100000  )
plt.hist(x1, normed=1, bins=40, facecolor='green', alpha=0.5)
plt.title('k=5, p=0.1')
plt.show()
```



Poisson 机率分布

- 观察一下
    - 转角夜宵:在晚上 平均每小时会有 10 人 来 跟转角哥买夜宵。 问摆摊 5 小时 有 60 人光顾之机率?
    - 费雯被嘘:费雯兄 po 文后， 平均每分钟会有 5 人嘘之 。 问发文后 二十分钟 变成 XX (100 嘘) 之机率?
- 某结果出现之平均速率(rate: 次数/时间)已知。问持续观察某 时间长度后，看到该结果出现 k 次之机率? --> Poisson 机率分布
- 已知某事发生速率为每单位时间 λ 次，观察时间为 T 时间单位。 X 为该观察时间 内发生该事的总次数。则:
    - PMF:
    $$\boxed{X \sim POI(\lambda T)}$$
    $$p_X(x) = P(X = x) = e^{-\lambda T} \cdot \frac{(\lambda T)^x}{x!}$$
    $$※ \boxed{\mu = \lambda T}, X \sim POI(\mu) \Rightarrow P_X(x) = e^{-\mu} \cdot \frac{\mu^x}{x!}$$
    - CDF:
    $$\boxed{X \sim POI(\lambda T)}$$
    $$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(x) = \begin{cases} \sum_{n=-\infty}^{\lfloor x \rfloor} e^{-\mu} \cdot \frac{\mu^n}{n!}, & x = 0, 1, 2 \dots \\ 0, & otherwise \end{cases}$$
- 费雯被嘘:费雯兄 po 文后，平均 每分钟会有 5 人嘘之。问发文后 20 分钟变 成 XX (100 嘘) 之机率?
    - λ = 5 嘘/分，若定义随机变量 X 为 20 分钟内的嘘数

- => X ~ POI(λT) = POI(100) = $e^{-100} \cdot 100^{100}$ / 100!
- 若条件是 每小时 300人嘘之，答案一样

- 理解泊松分布的特性:
  - 它常用来描述大量随机试验中稀有事件出现的次数
  - 比如 抽卡抽到 詹姆斯卡次数？

```
import numpy as np
import matplotlib.pyplot as plt
x1 = np.random.poisson( 5 , 100000  )
plt.hist(x1, normed=1, bins=40, facecolor='green', alpha=0.5 )
plt.title('lambda = 5 ')
plt.show()
```