Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫
Note by THF

2024 年 9 月 19 日

目录

概述

资源

公众号：狗熊会、大数据文摘，好玩的数学
MOOC：爱课程，Coursera，Edx，网易公开课等

教师要求

教材：概率论与数理统计第二版
参考：The Lady Tasting Tea，程序员数学之概率统计，...
学习目的：自问自答，自言自语
考核及成绩组成：
期中（10）
作业与考勤（10）
期末（70）
MOOC（10）

课程简介

概率：Probability
统计：Statistics
概率论与数理统计：Probability theory and Mathematical statistics

Notation. 第一章重要但不突出

从概率到概率论：新增时间（随机事件、样本空间变化）
从统计到数理统计：统计最开始为记录性质，后来衍生出预测，通过数学模型引入数理统计
类似的还有政府统计、经济统计等

2000-2015 年间，IT 时代逐渐转换为 DT(Data Technology) 时代，大数据逐渐占时代主体

1 第一章

1.1 随机事件

1.1.1 现象

确定性现象：一定条件下必然发生

随机现象强调统计规律性

Notation. 统计规律性：

1. 每次试验前不能预测结果
2. 结果不止一个
3. 大量试验下有一定规律

Example. 星际旅行时宇航员看到的现象不是随机现象：

对星际旅行的人而言，无法完成大量试验

宇航员观测到的结果无规律，只能称为不确定现象 (Uncertain)

Example. 扔一个骰子不能预测结果，但可以知道结果是 $1, 2, 3, 4, 5, 6$ 的一个，因此观察扔骰子是随机现象 (Random)

1.1.2 随机试验

随机试验 (E)：研究随机现象时进行的实验或观察等

Notation. 随机试验的特性：

1. 可以在完全相同的条件下重复进行
2. 试验的可能结果在试验前已知
3. 试验的结果不可预测

1.1.3 样本

在随机试验中，不可再分的最简单结果成为样本点 ω，全体样本点组成样本空间 Ω

Notation. 随机事件是基本事件的集合

Example. 扔骰子存在 6 个基本事件，可以产生 2^6 个随机事件，其中样本空间 $\Omega = \{x | x \in [1,6], x \in \mathbb{R}\}$

Example. 1. 射击时用 ω_i 表示击中 i 环，样本空间为：

$$\Omega = \{\omega_0, \omega_1, \omega_2, \ldots, \omega_{10}\}.$$

2. 微信用户每天收到信息条数的取值范围是 $[0, +\infty)$，样本空间为无限集：

$$\Omega = \{N | N \geq 0, N \in \mathbb{R}\}.$$

3. 电视机的寿命样本空间为 $\Omega = \{t | t > 0\}$，为连续的非负实数集

4. 投掷两枚硬币，样本空间为 $\Omega = \{(x,y) | x, y = 0, 1\}$，其中 $0, 1$ 分别代表正面和背面

Notation. 1. 样本点可以不是数

2. 样本空间可以是无限集

1.1.4 随机事件

1.2 事件关系与运算

1.1. $A \subset B$：A 发生必然 B 发生

1.2. $A = B$ ： $A \subset B, B \subset A$

2. $A \cup B$：A 和 B 至少有一个发生

2.1 $A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_1$

3. $A \cap B$：A 和 B 只发生一个

4.1. A, B 互斥：不能同时发生：$AB = \varnothing$

4.2. A, B 对立：非此即彼：$A \cup B = \Omega$

5. $A - B$：$A\bar{B}$ 或 $A(\Omega - B)$，或 A 发生但 B 不发生

Notation. $A-B=A\bar{B}\subset A$，$B-A=B\bar{A}\subset B$

当 $AB=\varnothing$ 时，$A-B=A, B-A=B$

Notation. $P(\Omega)=1, P(\varnothing)=0$，且 $P(\Omega)+P(\varnothing)=1$ ，即 Ω 与 $\varnothing$ 互斥

6. 结合律：$(A\cup B)\cup C=A\cup(B\cup C)$
7. 分配律：$(AB)\cup C=(A\cup C)(B\cup C)$ ，$(AUB)C=AC\cup BC$
8. 交换律：$A\cup B=B\cup A, AB=BA$

Notation. 德摩根律：

$$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A_i}.$$

$$\overline{\bigcap_{i=1}^{n}A_i}=\bigcup_{i=1}^{n}\overline{A_i}.$$

Example.

$$\overline{A\cup B}=\bar{A}\bar{B}.$$

$$\overline{(A\cup B)\cup C}=\overline{A\cup B}\bar{C}=\dots.$$

$$\overline{A\cap B}=\bar{A}\cup\bar{B}.$$

1.3 事件的概率

概率分类：

$$\begin{cases}\text{主观概率}\\\text{统计概率}\\\text{古典概型}\\\text{几何概型}\end{cases}.$$

Notation. 德摩根、蒲丰、皮尔逊、维纳均进行过投掷硬币的试验，随着试验次数的增加，出现正面的频率逐渐接近 0.5

大数定律说明，该事件的概率为 0.5

Definition. 统计概率：A 为试验 E 的一个事件，随着重复次数 n 的增加，A 的频率接近于某个常数 p，定义事件 A 的概率为 p，记为 $P(A)=p$

频率的特性：

1. 非负性：$f_n(A) \in [0,1]$
2. 规范性：$f_n(\Omega) = 1$
3. 有限可加性：A_i 两两互斥，则 $f_n\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} f_n(A_i)$

Definition. 主观概率：人对某个事件发生与否的可能性的估计

Definition. 完备事件组：$A_1, A_2, \ldots, A_n$ 两两互斥，且

$$P\left(\sum_{i=1}^{n} A_i\right) = 1.$$

或

$$\sum_{i=1}^{n} A_i = \Omega.$$

则称 $A_1 \to A_n$ 为完备事件组（不重不漏）

Example. $A, \bar{A}$ 是完备事件组

1.3.1 古典概型

古典概型特点：有限等可能性（基本事件数有限，基本事件发生的可能性相等）

Notation. 概率计算：

$$P(A) = \frac{m}{n} = \frac{n(A)}{n(\Omega)}.$$

Example. 某年级有 6 人在 9 月份出生，求 6 个人中没有人同一天过生日的概率

基本事件总数：30^6

目标事件：$30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 = P_{30}^{6}$

概率：

$$P(A) = \frac{P_{30}^{6}}{30^6}.$$

Example. 有 N 个乒乓球中有 M 个白球、$N-M$ 个白球，任取 $n(n<N)$ 个球，分有放回和不放回，求取到 m 个黄球的概率

1. 不放回：

基本事件总数：C_N^n

目标事件：$C_M^m C_{N-M}^{n-m}$

概率：

$$P = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, n = \max\left\{0, n - (N-M)\right\}, \ldots, \min\left\{n, M\right\}.$$

2. 有放回：

$$P = \frac{C_n^m M^m \left(N-M\right)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m}, m \in [0, n].$$

注意到该概率为伯努利分布 $C_n^m B\left(n, \frac{M}{N}\right)$

匹配问题：

Example. 麦克斯韦-玻尔兹曼统计问题：

n 个质点随机落入 $N\left(N > n\right)$ 个盒子，盒子容量不限，设 A 表示指定的 n 个盒子各有一个质点，B 表示恰好有 n 个盒子装一个质点

基本事件总数：N^n

A 考虑顺序，即：

$$P\left(A\right) = \frac{n!}{N^n}.$$

同理：

$$P\left(B\right) = \frac{C_N^n}{N^n}.$$

1.3.2 几何概型

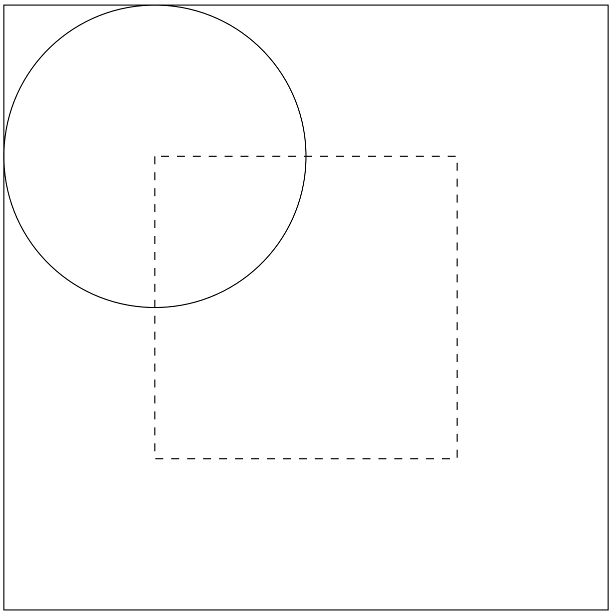
几何概型特点：使用事件所对应的**几何度量**计算

$$P\left(A\right) = \frac{m\left(A\right)}{m\left(\Omega\right)}.$$

Notation. 度量：面积、体积、长度等描述几何量大小的测度方式

Example. 地面铺满 2 dm 的地砖，向地面投掷一个 $r = 0.5$ dm 的光盘，求光盘不与边线相交的概率

如图:

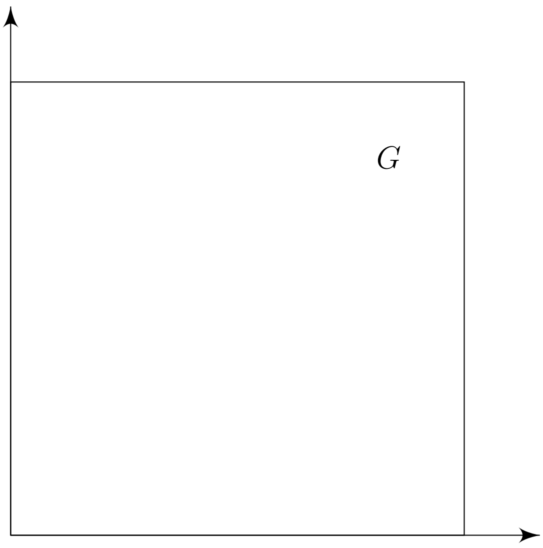


课后习题：A 组 8 题，B 组 3 题

Example. 两人相约 8-9 点间在某地相见，先到的人等待 20 分钟后离去，求二人会面的概率

设 (x, y) 分别表示两人到达的时刻

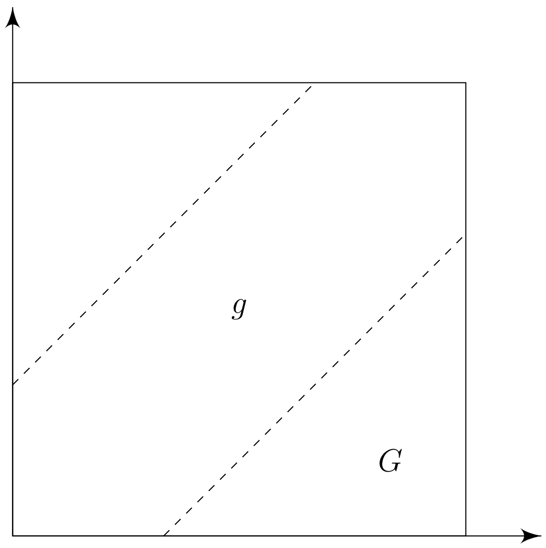
设 G 为样本空间，绘制样本空间:



由题：两人到达的时间之差的绝对值小于 20 分钟（$\frac{1}{3}$ 小时），即：

$$|x-y| \leq \frac{1}{3}.$$

将事件绘制：



$$P(g)=\frac{m(g)}{m(G)}=\frac{S(g)}{S(G)}=\frac{1-\left(\frac{2}{3}\right)^2}{1}=\frac{4}{9}.$$

Notation. 几何概型的特点：

1. 非负性：
$$P(A)\in[0,1].$$
2. 规范性：
$$P(\Omega)=1.$$
3. 可列可加性：
$$P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i).$$

1.4 公理化

$$(\Omega,\mathscr{F},p).$$

Definition. Ω：随机试验所产生的所有样本点的集合

$\mathscr{F}$：集合内所有子集为元素的集合

$P\left(X\right)$：概率函数

Axiom. 非负性：

$$P\left(A\right)\geq 0, A\in\mathscr{F}.$$

Axiom. 规范性：

$$P\left(\Omega\right)=1.$$

Axiom. 可列可加性：对两两互斥的事件 $A_1, A_2, \ldots,$

$$P\left(\sum_{i=1}^{+\infty}A_i\right)=\sum_{i=1}^{+\infty}P\left(A_i\right).$$

从三条公理得出的性质：

Notation. 1. $P\left(\varnothing\right)=0$

2. 有限可加性：

$$\sum_{i=1}^{n}P\left(A_i\right)=P\left(\sum_{i=1}^{n}A_i\right).$$

3. $P\left(\bar{A}\right)=1-P\left(A\right)$

4. $A\subset B \implies P\left(B-A\right)=P\left(B\right)-P\left(A\right)$

5. $A\subset B \implies P\left(A\right)\leq P\left(B\right)$

6. $P\left(A\cup B\right)=P\left(A\right)+P\left(B\right)-P\left(AB\right)$

Notation. 6.1.

$$P\left(A\cup B\cup C\right)=P\left(A\right)+P\left(B\right)+P\left(C\right)-P\left(AB\right)-P\left(BC\right)$$

$$-P\left(AC\right)+P\left(ABC\right).$$

6.2.

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P\left(A_i\right)-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P\left(A_iA_j\right)+\sum_{i=1}^{n-2}\sum_{j=i+i}^{n-1}\sum_{k=j+1}^{n}P\left(A_iA_jA_k\right)$$

$$\ldots+\left(-1\right)^{n-1}P\left(\prod_{i=1}^{n}A_i\right).$$

Example. 从有号码 $1,2,...,n$ 的 n 个球中有放回地取 m 个球，求取出的 m 个球中最大号码为 k 的概率

$$P\left\{k=1\right\}=\left(\frac{1}{n}\right)^m.$$

逐个列举计算较复杂，记事件 B_k 为取出的 m 个球最大号码不超过 k，只需保证每次摸出的球都不超过 k 即可：

$$P\left(B_k\right)=\frac{k^m}{n^m}.$$

又有 $P\left(A_k\right)=P\left(B_k\right)-P\left(B_{k-1}\right)$，且 $B_{k-1}\subset B_k$

所以：

$$P\left(A_k\right)=\frac{k^m}{n^m}-\frac{\left(k-1\right)^m}{n^m}.$$

Example. 匹配问题：n 个学生各带有一个礼品，随机分配礼品，设第 i 个人抽到自己的礼品称为一个配对，求至少有一个配对的概率

设 A_i 是第 i 个人抽到自己的礼品，A 为目标事件，则：

$$A=\bigcup_{i=1}^{n}A_i.$$

$$P\left(A_i\right)=\frac{(n-1)!}{n!}=\frac{1}{n}.$$

$$P\left(A_iA_j\right)=\frac{(n-2)!}{n!}=\frac{1}{P_n^2}.$$

$$P\left(A_iA_jA_k\right)=\frac{1}{P_n^3}.$$

$$\cdots\cdots\cdots\cdot$$

$$P\left(\prod_{i=1}^{n}A_i\right)=\frac{1}{n!}.$$

$$P\left(A\right)=P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P\left(A_i\right)-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P\left(A_iA_j\right)+\ldots.$$

1.5 条件概率与乘法公式

Definition.

$$P(A) > 0, P(B|A) = \frac{P(AB)}{P(A)}.$$

即：在 **A 发生的条件下**，B 发生的概率

Definition. 乘法公式：

$$P(AB) = P(A)P(B|A) = P(B)P(A|B).$$

Notation. A, B 独立：$P(AB) = P(A)P(B)$

结合乘法公式：

$$P(B) = P(B|A).$$

$$P(A) = P(A|B).$$

1.6 全概率公式

Corollary. 事件 $A_1, A_2, \ldots, A_n$ 为完备事件组，事件 $B \subset \Omega = \bigcup_{i=1}^n A_i$，则：

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i).$$

Notation. 此时完备事件组的情况应该已知，通过完备事件组 A 的辅助可以求得较复杂事件 B 的概率

1.7 贝叶斯公式

Corollary.

$$P(A_k|B) = \frac{P(A_kB)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}.$$

贝叶斯公式被称为“逆概率公式/后验公式”，其中事件 B 更可能是事件的结果，将事件组 A 看作结果出现的原因，则贝叶斯公式是一个从“结果”推“原因”的可能性的公式

Notation. 对比一般公式：事件 A 导致 B，求 B 发生的概率

贝叶斯公式：事件 A 导致 B，A 中的一个事件 A_i 导致 B 发生的概率

Axiom. 条件概率的公理：

1. 非负性：$P(A) \in [0,1]$
2. 规范性：$P(\Omega|A) = 1$
3. 可列可加性：

$$P\left(\sum_{i=1}^{\infty} B_i | A\right) = \sum_{i=1}^{\infty} P(B_i|A).$$

Corollary.

$$P\left(\bar{B}|A\right) = P(\Omega - B|A) = P(\Omega|A) - P(B|A).$$

Corollary.

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1B_2).$$

$$\implies P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) - P(B_1B_2|A).$$

Corollary. 乘法公式：

$$P(ABC) = P(A(BC)) = P(A)P(BC|A) = P(A)P(B|A)P(C|AB).$$

Example. 8 个红球 2 个白球，求前三次结果是“红红白”的概率：

1. 不放回取 3 个（和一次取三个球相同）

所有可能性：$10 \times 9 \times 8$

目标事件：$8 \times 7 \times 2$

或使用乘法公式：设 A_i 为第 i 次取到红球，目标事件可表示为 $A_1A_2\bar{A}_3$

概率：

$$P\left(A_1A_2\bar{A}_3\right) = P(A_1)P(A_2|A_1)P\left(\bar{A}_3|A_1A_2\right) = \frac{8}{10} \times \frac{7}{9} \times \frac{2}{8} = \frac{7}{45}.$$

2. 每次取后放回，并加入两个同色的球，取 3 次（不能使用古典概型）

概率：

$$P\left(A_1A_2\bar{A}_3\right) = \frac{8}{10} \times \frac{8}{12} \times \frac{2}{14} = \frac{8}{105}.$$

Example. 某疾病的发病率为 0.0004，患病检测呈阳性的概率为 0.99，误诊为阴性的概率为 0.01，误诊为阳性的概率为 0.05，不患病检测呈阴性概率为 0.95，一个人检测呈阳性，求其患病的概率

设阳性为 A，患病为 B

则：

$$P\left(A|B\right)=0.99,P\left(A|\bar{B}\right)=0.05,P\left(B\right)=0.0004.$$

要求：$P\left(B|A\right)$

使用贝叶斯公式：

$$P\left(B|A\right)=\frac{P\left(AB\right)}{P\left(A\right)}=\frac{P\left(B\right)P\left(A|B\right)}{P\left(AB\right)+P\left(A\bar{B}\right)}.$$

$$=\frac{P\left(B\right)P\left(A|B\right)}{P\left(B\right)P\left(A|B\right)+P\left(A|\bar{B}\right)P\left(\bar{B}\right)}=0.0079.$$

1.8 独立性

Definition. A,B 独立，则：$P\left(A|B\right)=P\left(A\right)$

Notation. 证明独立性：

1. $P\left(A\right)P\left(B\right)=P\left(AB\right)$

Notation. 独立事件的特点：

1. A,B 独立有：A,B 所有的组合（包含补集）均独立
2. A,B 独立的充要条件：$P\left(A|B\right)=P\left(A\right)$ or $P\left(B|A\right)=P\left(B\right)$
3. $\varnothing$ 与任何随机事件独立，Ω 与任何随机事件独立

对于三个事件相互独立：

$$\begin{cases}P\left(AB\right)=P\left(A\right)P\left(B\right)\\P\left(AC\right)=P\left(A\right)P\left(C\right)\\P\left(BC\right)=P\left(C\right)P\left(C\right)\\P\left(ABC\right)=P\left(A\right)P\left(B\right)P\left(C\right)\end{cases}.$$

对比乘法公式：$P\left(ABC\right)=P\left(A\right)P\left(B|A\right)P\left(C|AB\right)$

Definition. 相互独立：

有 $A_1, A_2, \ldots, A_n$ 事件组，对 $\forall s \in [2, n]$ 个事件 $A_{k_1}, A_{k_2}, \ldots, A_{k_s}$ 均有：

$$P\left(\prod_{n=1}^{s} A_{k_n}\right) = \prod_{n=1}^{s} P\left(A_{k_n}\right).$$

称事件 $A_1, A_2, \ldots, A_n$ 相互独立

Definition. 两两独立：对事件 $A_1, A_2, \ldots, A_n$ ，若任意两个事件独立，则称为两两独立

Notation. 相互独立一定两两独立，反之不一定

Notation. 相互独立事件组的性质：

1. 事件 $A_1, A_2, \ldots, A_n$ 相互独立，将其中任意部分改为对立事件，事件组仍为相互独立

2. 事件相互独立，将事件组任意分为两组（或多组），对组内事件进行“并、交、差、补”操作后，事件间依然相互独立

独立重复实验

Definition. E_1, E_2 中一个试验的任何结果和另一个试验的任何结果相互独立，则试验相互独立；若 n 个独立试验相互独立且试验相同，称 $E_1, E_2, \ldots, E_n$ 为 n 次独立重复实验，或 n 重独立试验

Example. 扔硬币和掷骰子为独立试验，其中扔硬币为伯努利试验（只有两个结果）

Definition. n 重独立试验 E 中，每次试验都是伯努利试验（可能结果只有两个），称 E 为 n 重伯努利试验

1. 二项概率公式：成功 k 次的概率记为 $P_n(k)$，假定前 k 次成功，后 $n-k$ 次失败，则

$$P_i = p^k (1-p)^{n-k}.$$

指定事件 A 发生的位置有 C_n^k 种，则：

$$P_n(k) = C_n^k p^k (1-p)^{n-k}.$$

称为二项概率公式

2. 几何概率公式：首次成功恰好发生在第 k 次的概率记为 $G(k)$，设前 $k-1$ 次失败，则：

$$G(k) = q^{k-1}p.$$

可以验证：$\sum G(k) = 1$

3. 负二项概率：需要成功 r 次，第 r 次成功恰好发生在第 k 次的概率记为 $G_r(k)$，设前 $k-1$ 次试验有 $r-1$ 次成功，则：

$$G_r(k) = C_{k-1}^{n-1}p^r q^{k-r}.$$

同样有：$\sum G_r(k) = 1$

2 一维随机变量及其分布

2.1 随机变量及其分布函数

随机变量

$$\text{随机变量}\begin{cases}\text{函数}\\ \text{连续}\end{cases}.$$

Example. 下一个进入教室的同学可能是男是女，分别记为 1,2，则有映射：

$$\{\text{男}, \text{女}\} \to \{1, 2\}.$$

将离散的结果映射为坐标轴上离散的数值，所有的数值性的观测结果无需改变，如：下一个进入教室的同学身高为 ω，则有映射：

$$X(\omega) = \omega.$$

Definition. 实值变量（无分布函数）使用小写字母，随机变量（有分布函数）使用大写字母

对 $\forall x \in \mathbb{R}$，$\{X \le x\} = \{\omega | X(\omega) \le x, \omega \le \Omega\} \in \mathscr{F}$，则 X 称为**概率空间的随机变量**

Example. 对于 $\{M,F\} \to \{1,2\}$ ，取 $x=1$ ，写出定义式：

$$\{X \leq 1\} = \{M\}.$$

同时由于 $x \in \mathbb{R}$, 取 $x = 1.5$ 时, $\{X \leq 1.5\} = \{M\}$ 取 $x = 4$ 时, $\{X \leq 4\} = \{M, F\}$

由于 $x \in \mathbb{R}$ ，则可以引入其他分布函数辅助，继而引用微积分理论

对于 $(\Omega, \mathscr{F}, P)$：

$$P : \Omega \to [0,1].$$

Notation. X 具有随机性（样本点具有随机性），是定义在 Ω 上的函数

X 是随机变量时 $\{a \leq X \leq b\}, a < b, a, b \in \mathbb{R}$ 均为随机事件

X 是随机变量，$g(x)$ 是非单点的实值函数，则 $Y = g(X)$ 也是随机变量：

$$Y(\omega) = g(X(\omega)).$$

Example. 对灯泡做寿命试验，用 X 表示测得灯泡的寿命，样本空间 $\Omega = [0, +\infty)$，则：

$A =$ “测得灯泡寿命大于 500 h” $= \{X > 500\}$

$B =$ “测得灯泡寿命小于 5000 h” $= \{X \leq 5000\}$

分布函数

Definition. 分布函数：记

$$F(x) = P\{X \leq x\}, x \in \mathbb{R}.$$

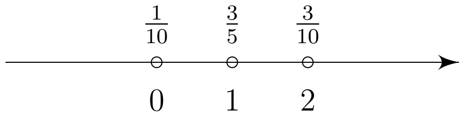
为 X 的分布函数

Example. 3 白 2 黑，不放回取三次球，求取到的黑球个数 X 的分布函数

X 可以取到：0,1,2

$$P\{X=0\} = \frac{C_3^3}{C_5^3}, P\{X=1\} = \frac{C_2^1 C_3^2}{C_5^3}, P\{X=2\} = \frac{C_2^2 C_3^1}{C_5^3}.$$

概率在坐标轴上体现：



$$F(x) = P\{X \le x\} = \begin{cases} 0, x < 0 \\ \frac{1}{10}, x \in [0,1) \\ \frac{1}{10} + \frac{3}{5} = \frac{7}{10}, x \in [1,2) \\ 1, x \ge 2 \end{cases}.$$

图像：P43

Notation. 分布函数的特性：

1. 非负性：$P \in [0,1]$
2. 单调不减性
3. 右连续性：

$$F(x) = \lim_{t \to x+0^+} F(t).$$

3.1 不满足左连续，例：$P(0) - P(0^-) \neq 0$

4. 规范性：

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$$

关于 X 的事件都可以使用分布函数表示：

$$\begin{cases} P\{X = a\} = \lim_{\varepsilon \to 0^+} P\{a - \varepsilon < X \le a\} = F(a) - F(a - 0^+) \\ P\{a \le X < b\} = F(b - 0^+) - F(a - 0^+) \\ \ldots\ldots \end{cases}.$$

Example. 在 $[a,b]$ 内随机取一个数 X ，求 X 的分布函数

关键区域：$x \in [a,b)$,

$$\{X \le x\} = \{a \le X \le x\}.$$

$$F(x) = P\{X \le x\} = \frac{x-a}{b-a}.$$

作业：预习第 2,3 节