AI 中的数学

第三、四讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

例 1.2. 盒中有 5 个球, 其中有 2 个白球, 3 个黑球.
从中任取 3 个球, 将其中所含的白球的数目记为 X.

- 建模: 将球编号, $1 \sim 3$ 表示黑球, 4,5 表示白球.
- $\omega = (i, j, k)$, 其中 $1 \leqslant i < j < k \leqslant 5. \Omega = C_5^3 = 10.$
- 事件是 Ω 的集合, 如果只关注第一次结果, 所有事件对应的每个集合中可以包括 $2, 3$ 次的所有情况

例 1.2. 盒中有 5 个球, 其中有 2 个白球, 3 个黑球. 从中任取 3 个球, 将其中所含的白球的数目记为 X.

- 建模: 将球编号, $1 \sim 3$ 表示黑球, 4,5 表示白球.
- $\omega = (i,j,k)$, 其中 $1 \leqslant i < j < k \leqslant 5. \Omega = C_5^3 = 10.$
- 事件是 Ω 的集合，如果只关注第一次结果，所有事件对应的每个集合中可以包括 $2,3$ 次的所有情况
- 满足 $X = 0$ 的 ω 有 $C_2^0 C_3^3 = 1$ 个;
  满足 $X = 1$ 的 ω 有 $C_2^1 C_3^2 = 6$ 个;
  满足 $X = 2$ 的 ω 有 $C_2^2 C_3^1 = 3$ 个.
- 事件: $\{X = 1\} = \{\omega : X(\omega) = 1\}$,
  $\{X \leqslant 1\} = \{\omega : X(\omega) \leqslant 1\}$.
- 将 $P(\{X = 1\})$ 简记为 $P(X = 1)$. 例如,

$$P(X = 1) = \frac{6}{10}, P(X \leqslant 1) = \frac{7}{10}.$$

例 1.6. 某公共汽车站每隔 10 min 会有一两某路公交车到达. 某乘客随机在任意时刻到达车站.

- 候车时间 X (单位: min) 为随机变量.
- $0 \leqslant X \leqslant 10$.
- 几何概型 (参阅 1.8): 例,

$$P(X \leqslant 3) = \frac{3}{10}, \quad P(2 \leqslant X \leqslant 6) = \frac{4}{10}.$$

- 直观描述：某变量 X 随机取值，则 X 是随机变量
- 严格描述：对于样本空间 $\Omega = \{\omega\}$，$X = X(\omega)$ 是在 Ω 上有定义的实值函数，而且对任何实数 c，事件 "$\{\omega : X(\omega) \leqslant c\}$" 是有概率的（属于事件域），将 X 称为随机变量

- 定义 2.1. X 是离散型随机变量指: X 取有限个值 $x_1, \cdots, x_n$, 或可列个值 $x_1, x_2, \cdots$. X 的概率分布 (列) 指:

$$p_k = P\left(X = x_k\right), \quad k = 1, \cdots, n \text{ 或} k = 1, 2, \cdots.$$

- 概率分布表:

X	x_1	x_2	$\cdots$	x_k	$\cdots$
p	p_1	p_2	$\cdots$	p_k	$\cdots$

- 非负: $p_k \geqslant 0, \forall k$; 规范: $\sum_{k=1}^{n} p_k = 1$ 或 $\sum_{k=1}^{\infty} p_k = 1$.
- p_k, $k = 1, 2, 3 \ldots$, 是 X 的概率分布, 也称为概率函数或者概率分布律.

1. 两点分布 (伯努利分布), $X \sim B(1,p)$ (参数 $0 \leqslant p \leqslant 1$):

$$P(X=1)=p, \quad P(X=0)=1-p.$$

- 模型：投币，

$$\text{投到} H \text{ 则} X=1; \text{投到} T \text{ 则} X=0.$$

- 示性函数 1_A：事件 A 发生则取 1; A 不发生则取 0 .
- 例 2.1. 100 件产品中有 3 件次品. 从中任取一件.

$$A= \text{“取到合格品”}, X=1_A, p=0.97.$$

2. 二项分布, $X \sim B(n,p)$(参数 $n \geqslant 1, 0 \leqslant p \leqslant 1$) :

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

- 模型: 独立投币 n 次, 正面的总次数.
- 定理 2.1. 分布列的最大值点 k_0 如下:

若 $(n+1)p \notin \mathbb{Z}$, 则 $k_0=[(n+1)p]$;
若 $(n+1)p \in \mathbb{Z}$, 则 $k_0=(n+1)p$ 或 $(n+1)p-1$.

有组合数公式：

$$\frac{p_n(k+1)}{p_n(k)} = \frac{n-k}{k+1} \cdot \frac{p}{1-p}$$

由于 $\frac{n-k}{k+1} \cdot \frac{p}{1-p} > 1$ 等价于 $k < (n+1)p - 1$，于是有：

(a) 当 $k < (n+1)p - 1$ 时，$p_n(k+1) > p_n(k)$

(b) 当 $k > (n+1)p - 1$ 时，$p_n(k+1) < p_n(k)$

(c) 当 $k = (n+1)p - 1$ 时，$p_n(k+1) = p_n(k)$

3. 泊松分布, $X \sim \mathcal{P}(\lambda)$(参数: $\lambda > 0$) :

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

- 模型: 例 2.3. 研究放射性物质在 8 分钟内放射出的粒子数 X.

X 近似服从 $B(n,p)$, 故

$$P(X=k)=C_n^k p^k(1-p)^{n-k}\approx\frac{n!}{k!(n-k)!}p^k(1-p)^n$$
$$\approx\frac{(np)^k}{k!}\left(1-\frac{\lambda}{n}\right)^n=\frac{\lambda^k}{k!}e^{-\lambda}$$

- 上式即为 §1.7 第一近似公式.
- 泊松分布的分布列最大值点 $k_0=[\lambda]$.

泊松分布的分布列最大值点 $k_0 = [\lambda]$.

证明：注意到 $p_{k+1} = \frac{\lambda}{k+1} p_k$，故由分布函数知
若 $k+1 \leqslant \lambda$，则 $p_{k+1} \geqslant p_k$
若 $k+1 \geqslant \lambda$，则 $p_{k+1} \leqslant p_k$
因此当 $k_0 = [\lambda]$ 时，分布列取最大值。

已知某商场一天来的顾客服从参数为 λ 的泊松分布，而每个来商场的顾客购物概率为 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

已知某商场一天来的顾客服从参数为 λ 的泊松分布，而每个来商场的顾客购物概率为 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

解：用 Y 表示商场内一天购物的顾客数，则由全概率公式知，对任意正整数 k 有

$$P(Y=k)=\sum_{i=k}^{\infty}P(X=i)P(Y=k\mid X=i)=\sum_{i=k}^{\infty}\frac{\lambda^i e^{-\lambda}}{i!}C_i^k p^k(1-p)^{i-k}$$

$$=\frac{(\lambda p)^k}{k!}e^{-\lambda}\sum_{i=k}^{\infty}\frac{[\lambda(1-p)]^{i-k}}{(i-k)!}=\frac{(\lambda p)^k}{k!}e^{-\lambda}e^{\lambda(1-p)}=\frac{(\lambda p)^k}{k!}e^{-\lambda p}$$

4. 超几何分布, $X \sim H(N, D, n)$ (参数 N, D, n) :

$$P(X=k)=\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}, \quad k=0,1,\cdots,n.$$

- 模型: N 个产品, D 个次品, 任取 n 个, 抽到的次品数为 X.
- 放回抽样 vs 不放回抽样: 二项分布 vs 超几何分布.
- 定理 2.3. 给定 n. 当 $N \to \infty, \frac{D}{N} \to p$ 时,

$$\frac{C_D^k C_{N-D}^{n-k}}{C_N^n} \to C_n^k p^k (1-p)^{n-k}, \quad k \geqslant 0$$

- 该定理的直观解释是，如果一批产品的总量 N 很大，其中次品占比为 p，则从整批产品随机抽取 n 个，抽到次品的个数 k 近似服从参数为 p，n 的二项分布

证明：由于 $0<p<1$，当 N 充分大时，$n<D<N$，且 n 是固定的，易知

$$\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}=\frac{D!}{k!(D-k)!}\cdot\frac{(N-D)!}{(n-k)!(N-D-n+k)!}\cdot\frac{n!(N-n)!}{N!}$$
$$=\frac{n!}{k!(n-k)!}\cdot\frac{D(D-1)\cdots(D-k+1)}{N^k}$$
$$\cdot\frac{(N-D)(N-D-1)\cdots(N-D-n+k+1)}{N^{n-k}}$$
$$\cdot\frac{N^n}{N(N-1)\cdots(N-n+1)}$$
$$=C_n^k(\prod_{i=1}^{k}\frac{D-i+1}{N})(\prod_{i=1}^{n-k}\frac{N-D-i+1}{N})(\prod_{i=1}^{n}\frac{N}{N-i+1})$$
$$\to C_n^k p^k(1-p)^{n-k}\quad(N\to\infty)$$

5. 几何分布, $X \sim G(p)$, 参数 $0 < p < 1$:

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \cdots .$$

- 模型: 独立重复投币中, 第一次投到 H 时的投币次数.
- $P(X > n) = (1 - p)^n, \forall n \geqslant 0.$
- 无记忆性: $P(X - n = k \mid X > n) = P(X = k).$

设 X 是只取自然数的离散随机变量，若 X 的分布具有无记忆性，证明 X 的分布一定为几何分布。

证明：由无记忆性知

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

将 n 换为 $n-1$ 仍有

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

两式相减有

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

设 $P(X=1) = p$，若取 $n = m = 1$ 有

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

若取 $n=2$，$m=1$ 则有

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

若令 $P(X=k)=p(1-p)^{k-1}$，则用数学归纳法得

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

这表明 X 的分布为几何分布。

- 6. 离散均匀分布，

$$P(X = k) = \frac{1}{N}, \quad k = 1, \cdots, N.$$

- 模型：古典概型

- 定义 3.1. 连续型随机变量指：存在 $p(x)$ 使得

$$P(a \leqslant X \leqslant b) = \int_a^b p(x)dx, \quad \forall a < b.$$

称 $p(\cdot)$ 为 X 的概率密度（函数），也记为 $p_X(\cdot)$.

- 非负：$p(x) \geqslant 0$;　　规范：$\int_{-\infty}^{\infty} p(x)dx = 1$.
- $P(X = x) = 0$ 在任意一点选中的概率都为 0.
- $p(\cdot)$ 在 x 连续，则 $P(X \in [x, x + \Delta x]) = p(x)\Delta x + o(\Delta x)$,
- 单独谈论一个点 x 对应的 $p(x)$ 没有意义.

1. 均匀分布, $X \sim U(a,b)$ (参数 $a<b$) :

$$p(x)=\begin{cases}\frac{1}{b-a}; & 若 a \leqslant x \leqslant b;\\ 0, & 否则.\end{cases}$$

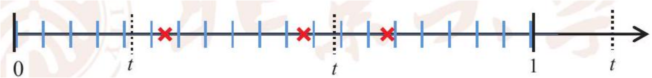
- $a \leqslant b \leqslant x$ 可改为 $a<x<b, a<x \leqslant b, a \leqslant x<b$.
- $p(x)=\frac{1}{b-a} 1_{\{a \leqslant x \leqslant b\}}$.
- $p(x)=\frac{1}{b-a}$, 其中 $a \leqslant x \leqslant b$. 某公共汽车站每隔 10 分钟会

  有一班公交车到达，一位搭乘该车的乘客在任意时刻到达车站是等可能的，则他的候车时间 X 是一个随机变量，且满足 $[0,10]$ 上的均匀分布

2. 指数分布, $X \sim \mathrm{Exp}(\lambda)$(参数 $\lambda > 0$) :

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

- 模型: 例 2.3.X = 第一个粒子的放射时刻. 等待时间、寿命.
- 第一个粒子在第 Y 个微观时间放出, 则



2. 指数分布, $X \sim \mathrm{Exp}(\lambda)$(参数 $\lambda > 0$) :

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

- 定理 3.1. (无记忆性):
  $P(X - s > t \mid X > s) = e^{-\lambda t}, \quad \forall t, s \geqslant 0.$

若 X 服从参数为 λ 的指数分布，则对任何 $0 \leqslant a < b$ 有：

$$P(a < X < b) = \lambda \int_a^b e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}$$

$$P(X > a) = e^{-\lambda a}$$

$$P(X - s > t \mid X > s) = \frac{P(X-s>t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda t} = P(X > t)$$

设 X 是非负的随机变量，$P(X-s>t \mid X>s)=e^{-\lambda t}$ 对 $\forall t,s\geqslant 0$ 恒成立的充分必要条件是 X 服从指数分布

设 X 是非负的随机变量，$P(X-s>t \mid X>s)=e^{-\lambda t}$ 对 $\forall t,s \geqslant 0$ 恒成立的充分必要条件是 X 服从指数分布

必要性：设 X 是非负随机变量满足条件，则

$$P(X>s)>0, \quad P(X>s+t)=P(X>s)P(X>t)$$

令 $f(u)=P(X>u)$，则 $f(s+t)=f(s)f(t)$
于是 $f(1)=f(\frac{1}{n}\times n)=(f(\frac{1}{n}))^n$
从而 $f(\frac{m}{n})=f(\frac{1}{n}\times m)=(f(\frac{1}{n}))^m=(f(1))^{\frac{m}{n}}$
故对任意正有理数 r，有 $f(r)=(f(1))^r$。由于 $0<f(1)<1$ 且 $f(u)$ 是关于 u 的减函数，因此对任意 $u \geqslant 0$，有 $f(u)=(f(1))^u$。
令 $\lambda=-\ln f(1)$，则 $f(u)=e^{-\lambda u}$，即

$$P(X>u)=e^{-\lambda u}=\int_{+\infty}^{u} e^{-\lambda x}dx$$

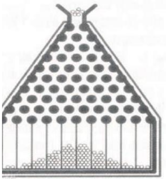
3. 正态分布, $X \sim N\left(\mu, \sigma^2\right)$ (参数 $\mu \in \mathbb{R}, \sigma > 0$) :

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- 标准正态分布 $N(0,1)$:

$$-\frac{x^2}{2}$$

高尔顿钉板试验



- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, 将积分的平方写为二重积分:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dxdy.$$

- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, 将积分的平方写为二重积分:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dxdy.$$

- 做极坐标变换:

$$x = r\cos\theta, y = r\sin\theta \Rightarrow \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = r.$$

- 因此,

$$\star\star = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{\infty} e^{-\frac{r^2}{2}} rdr \right) d\theta = \int_0^{\infty} e^{-R} dR = 1$$

- $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$: 令 $y = \frac{x-\mu}{\sigma}$, 则

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

- 函数 Φ：

$$\Phi(x) = \int_{-\infty}^{x} \phi(x)dx.$$

- $\Phi(-x) = 1 = \Phi(x)$.
- 定理 3.2. 令 $x^* = \frac{x-\mu}{\sigma}$, 则

$$P(a < X < b) = \int_a^b \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx = \Phi\left(b^*\right) - \Phi\left(a^*\right).$$

- 推论 3.1. 查表得 $\Phi(3) = 0.9987$, 因此
  $$P(\mu - 3\sigma < X < \mu + 3\sigma) = \Phi(3) - \Phi(-3) = 0.9974$$

4. 伽玛分布, $X \sim \Gamma(\alpha,\beta)$(参数 $\alpha,\beta>0$) :

$$p(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x},\quad x>0.$$

- $\Gamma(\alpha)=\int_0^\infty y^{\alpha-1}e^{-y}dy.$
- $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$

$$\int_0^\infty y^\alpha e^{-y}dy=-\,y^\alpha e^{-y}\Big|_0^\infty+\int_0^\infty \alpha y^{\alpha-1}e^{-y}dy.$$

- $\Gamma(1)=1;\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$

$$\Gamma\left(\frac{1}{2}\right)=\int_0^\infty\frac{1}{\sqrt{y}}e^{-y}dy=\sqrt{2}\int_0^\infty e^{-\frac{x^2}{2}}dx=\sqrt{\pi}.$$

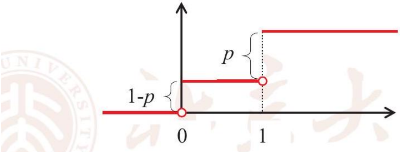
- $\alpha=1$ 时就是指数分布 $\mathrm{Exp}(\beta)$.

1. 随机变量
2. 离散随机变量
3. 连续随机变量
4. 随机变量严格定义

- 定义 4.1. 假设 $(\Omega, \mathcal{F}, P)$ 是概率空间, $X : \Omega \to \mathbb{R}$ 满足:
$$\text{对任意 } x \in \mathbb{R} \text{ 都有 } \{X \leqslant x\} \in \mathcal{F},$$
则称 X 是一个随机变量.
- 定义 4.2. 令 $F(x) = P(X \leqslant x), x \in \mathbb{R}$. 称 F 为随机变量 X 的分布函数, 也记为 F_X.
- 定理 4.2. $F = F_X$ 的三条性质:

(1) 单调性: 若 $x \leqslant y$, 则 $F(x) \leqslant F(y)$.
(2) 规范性: $\lim_{x \to -\infty} F(x) = 0; \lim_{x \to \infty} F(x) = 1$.
(3) 右连续性: $\lim_{y \to x+} F(y) = F(x)$.

- 离散型：$P(X = x_i) = p_i$. x_i 为 F_X 的跳点，p_i 为跳跃幅度.



- 连续型：$F_X(x) = \int_{-\infty}^{x} p(z)dz$，且

$$p(x) = F_X'(x).$$

反过来，若 F_X “几乎”连续可导，则为连续型（定理 4.3, 4.4）.

- 尾分布函数: $G(x) = P(X > x) = 1 - F(x)$.
  连续型: $p(x) = -G'(x)$.
- 例. $X \sim \mathrm{Exp}(\lambda)$.

$$G(x) = e^{-\lambda x}, \quad \forall x > 0,$$
$$\Rightarrow G'(x) = -\lambda G(x). \quad \lambda: \text{速率}.$$

- 由 $F_X(x)$ 可求出 $P(X \in B), \forall B$.
- 若 $F_X = F_Y$, 则称 X 与 Y 同分布, 记为 $X \stackrel{d}{=} Y$.
- $X = Y$, 即 $P(X = Y) = 1$, 则 $F_X = F_Y$. 反之不然.