# AI 中的数学

## 第五、六讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

1 数学期望

2 随机变量函数的期望

期望 (expectation) 的含义：均值 (mean).

- $X$ 的大量独立观测值 (记为 $a_1, a_2, \cdots, a_n$) 的算术平均:

$$\bar{a} = \frac{1}{n}(a_1 + \cdots + a_n).$$

- $X$ 的所有可能值的加权平均 (总和).
  例，$P(X = x_k) = p_k, k = 1, \cdots, m$.
  记 $n_k = \{m : 1 \leqslant m \leqslant n, a_m = x_k\}$. 那么，根据概率的频率含义，$\frac{n_k}{n} \approx p_k$，于是

$$\bar{a} = \frac{1}{n}\sum_{k=1}^{K} n_k \approx \sum_{k=1}^{K} x_k p_k.$$

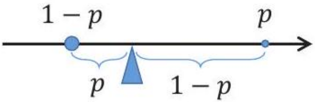
# 1. 离散型随机变量的期望

- 定义 6.1. 假设 $X$ 是离散型，分布列为

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

如果 $\sum_k |x_k| p_k < \infty$，那么，称 $X$ 的期望存在，称 $\sum_k x_k p_k$ 为 $X$ 的数学期望，记为 $EX$.

- $EX$ 是重心．例，(1) 伯努利分布，
$P(X = 1) = p, P(X = 0) = 1 - p$. 则，$EX = p$.




- $E(X)$ 完全由 $X$ 的概率分布确定

(3) 泊松分布.

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

- $\forall k\geqslant 1$,

$$x_k p_k = k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda p_{k-1}.$$

- 因此,

$$EX=\sum_{k=0}^{\infty}kp_k=\sum_{k=1}^{\infty}\lambda p_{k-1}=\lambda\sum_{\ell=0}^{\infty}p_\ell=\lambda.$$

(2) 二项分布.

(2) 二项分布.

$$P(X=k)=C_n^k p^k q^{n-k} =: b(n;k), k=0,1,\cdots,n, (q=1-p).$$

- $\forall 1 \leqslant k \leqslant n$,

$$k\cdot b(n;k) = k\cdot\frac{n!}{k!(n-k)!}p^k q^{n-k} = \frac{n!}{(k-1)!(n-k)!}p^k q^{n-k}$$
$$= \frac{n\cdot(n-1)!}{(k-1)!(n-k)!}p\cdot p^{k-1}q^{n-k} = np\cdot b(n-1,k-1)$$

- 因此,

$$EX=\sum_{k=0}^{n} k\cdot b(n;k) = \sum_{k=1}^{n} np\cdot b(n-1;k-1)$$
$$= np\sum_{\ell=0}^{n-1} b(n-1,\ell) = np$$

(7) 超几何分布.

(7) 超几何分布.

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

- 记 $h(N,D,n;k)=A_1\cdot A_2\cdot A_3=$

$$\frac{D!}{k!(D-k)!}\cdot\frac{(N-D)!}{(n-k)!(N-D-(n-k))!}\cdot\frac{n!(N-n)!}{N!}.$$

- 记 $x'=x-1$. 则, $\forall 1\leqslant k\leqslant n$,

$$k\cdot A_1=\frac{D!}{(k-1)!(D-k)!}=D\times\frac{D'!}{k'!\,(D'-k')!}.$$

- 进一步,

$$A_2=\frac{(N'-D')!}{(n'-k')!\,(N'-D'-(n'-k'))!},$$

$$A_3=\frac{n\cdot n'!\,(N'-n')!}{N\cdot N'!}=\frac{n}{N}\times\frac{n'!\,(N'-n')!}{N'!}.$$

- 记 $x' = x - 1$. 则 $\forall 1 \leqslant k \leqslant n$,

$$k \cdot h(N, D, n; k) = \frac{nD}{N} \times h\left(N', D', n'; k'\right).$$

- 因此,

$$EX = \sum_{k=1}^{n} k \cdot h(N, D, n; k) = \frac{nD}{N} \sum_{k'=0}^{n'} h\left(N', D', n'; k'\right) = \frac{nD}{N}$$

- $D = 1$ 时, 退化为伯努利分布, $EX = p = \frac{D}{N}$.
- $D \geqslant 2$ 时, 不放回抽样, 仍有 $EX = np$.

(4) 几何分布.

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

- 直接计算:

$$EX=\sum_{k=1}^{\infty}kp_k=\frac{1}{p}$$

## 2. 一般随机变量的期望

- $X$ 为任意随机变量. 做如下近似: $\forall n \in \mathbb{Z}$,

$$当\, n\varepsilon < X \leqslant (n+1)\varepsilon \,时, 令\, X^* = n\varepsilon.$$

- 直观: $X^* \leqslant X < X^* + \varepsilon$, 因此 $EX^* \leqslant EX < EX^* + \varepsilon$.
- 定义 6.2. 若 $EX^*$ 存在且当 $\varepsilon \to 0$ 时有极限, 则称 $X$ 的期望存在, 且称该极限为 $X$ 的期望, 记为 $EX$.
- 对离散型随机变量, 定义 6.1 与定义 6.2 一致.
- 定理 6.1. 对连续型随机变量, 若 $\int_{-\infty}^{\infty} |x| p(x) dx < \infty$, 则

$$EX = \int_{-\infty}^{\infty} x p(x) dx.$$

(2) 指数分布.

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

- $\int_0^\infty x \cdot \lambda e^{-\lambda x} dx = -\int_0^\infty x de^{-\lambda x} = \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}$.
- 一般地, 若 $X$ 为连续型, 且 $X \geqslant 0$. 令尾分布

$$G(x) = P(X > x) = \int_x^\infty p(y) dy$$

则 $G'(x) = -p(x)$. 于是,

$$\int_0^\infty xp(x)dx = \int_0^\infty xdG(x) = \int_0^\infty G(x)dx.$$

(3) 均匀分布.

设随机变量 $X$ 服从区间 $[a, b]$ 上的均匀分布，即 $X$ 有密度分布：

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leqslant x \leqslant b, \\ 0, & \text{其他} \end{cases}$$

由定义知

$$E(X) = \int_b^a x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

(4) 正态分布.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- $X \sim N(0,1)$

$$EX = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

- 同理, $X \sim N(\mu, \sigma^2)$, 则 $p(\mu + x) = p(\mu - x)$, 因此 $EX = \mu$.

(4) 正态分布.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- $X \sim N(0,1)$

$$EX = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

- 同理, $X \sim N\left(\mu, \sigma^2\right)$, 则 $p(\mu+x) = p(\mu-x)$, 因此 $EX = \mu$.
- 例, 柯西分布,

$$p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}.$$

但是, $\int_{-\infty}^{\infty} |x| p(x) dx = \infty$. 因此, $EX$ 不存在!

(5) 伽玛分布.

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

- $\forall x > 0$,

$$xp(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha} e^{-\beta x} = \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} \cdot \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{\alpha} e^{-\beta x} = \frac{\alpha}{\beta} \cdot \hat{p}(x).$$

- 因此,

$$EX = \int_0^{\infty} xp(x)dx = \frac{\alpha}{\beta} \int_0^{\infty} \hat{p}(x)dx = \frac{\alpha}{\beta}$$

## 3. 期望的性质

- 定理 6.2. (1) 若 $X \equiv a$, 则 $EX = a$;
- 定理 6.2. (2) 若 $X \geqslant 0$, 且 $EX$ 存在, 则 $EX \geqslant 0$;
- 定理 6.2. (3)(或, 推论 6.1). 若 $F_X = F_Y$ (或, 若 $X = Y$), 且 $EX$ 存在, 则 $EY$ 存在, 且 $EX = EY$.
- 定理 6.3. (1) & (2), 线性：假设 $EX, EY$ 存在. 则,

$$E(aX) = aEX, \quad E(X+Y) = EX + EY.$$

- 定理 6.3. (3), 单调性：假设 $EX$,$EY$ 存在, 又若 $X \geq Y$, 则 $EX \geq EY$.

- 推论 6.2. (1) 线性：假设 $EX, EY$ 存在. 则,

$$E(aX + bY) = aEX + bEY.$$

- 推论 6.2. (2) 和的期望：假设 $EX_1, \cdots, EX_n$ 都存在, $\eta = X_1 + \cdots + X_n$. 则 $E\eta$ 存在, 且

$$E\eta = EX_1 + \cdots + EX_n$$

- 例. 超几何分布 $\eta \sim H(N, D, n)$. 若第 $i$ 个产品是次品, 则令 $X_i = 1$; 否则, 令 $X_i = 0$. 则,

$$\eta = X_1 + \cdots + X_n \Rightarrow E\eta = np$$

1 数学期望

2 随机变量函数的期望

# 随机变量函数的期望

- 定理 6.5. $X$ 是离散型，或连续型，且下面的级数或积分绝对收敛，则

$$Ef(X) = \sum_k f(x_k)\, p_k, \qquad \text{或 } Ef(X) = \int_{-\infty}^{\infty} f(x)p(x)dx.$$

- 例 6.1. 设 $X \sim U(0, 2\pi)$，求 $E \sin X$.
- 用公式:

$$E \sin X = \int_{-\infty}^{\infty} \sin x \cdot p(x)dx = \frac{1}{2\pi}\int_0^{2\pi} \sin x dx = 0$$

例：（对应郑书例 6.2）设随机变量 $X$ 服从参数为 $\lambda$ 的指数分布，又 $v_0 > 0$，

$$Y = \begin{cases} X, & X < v_0, \\ v_0, & X \geqslant v_0, \end{cases}$$

求 $E(Y)$。

例：（对应郑书例 6.2）设随机变量 $X$ 服从参数为 $\lambda$ 的指数分布，又 $v_0 > 0$，

$$Y = \begin{cases} X, & X < v_0, \\ v_0, & X \geqslant v_0, \end{cases}$$

求 $E(Y)$。解：设 $f(x) = \min\{x, v_0\}$，则 $Y = f(X)$，由于 $X$ 的分布密度是

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & X \leqslant 0, \end{cases}$$

由式**??**知

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} f(x)p(x)dx = \int_0^{+\infty} f(x)\lambda e^{-\lambda x}dx \\ &= \int_0^{v_0} x\lambda e^{-\lambda x}dx + \int_{v_0}^{+\infty} v_0\lambda e^{-\lambda x}dx \\ &= \frac{1}{\lambda}(1 - e^{-\lambda v_0}) \end{aligned}$$

- 定理 6.4. (马尔可夫不等式). 设 $X \geqslant 0$, 且 $EX$ 存在. 则对任意 $C > 0$, 有

$$P(X \geqslant C) \leqslant \frac{1}{C} EX.$$

- 定理 6.4. (马尔可夫不等式). 设 $X \geqslant 0$, 且 $EX$ 存在. 则对任意 $C > 0$, 有

$$P(X \geqslant C) \leqslant \frac{1}{C} EX.$$

- 证：令 $A = \{X \geqslant C\}$. 则 $1_A \leqslant \frac{X}{C}$. 于是,

$$P(A) = E1_A \leqslant E\frac{X}{C} = \frac{1}{C} EX.$$

- 例, 若 $X \geqslant 0$, 且 $EX = 0$, 则

$$P\left(X \geqslant \frac{1}{n}\right) \leqslant nEX = 0$$

$$\Rightarrow P(X > 0) = \lim_{n\to\infty} P\left(X \geqslant \frac{1}{n}\right) = 0.$$