第九章 工具：期望

9.1 微积分储备知识

泰勒级数：如果 f 是 n 次可微分的 (其中 $f^{(k)}(x)$ 表示 f 在 x 处的 k 阶导数)，那么 f 在 a 点处的 n 阶泰勒级数就是

$$\begin{aligned} T_n(x) &:= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k. \end{aligned}$$

我们把 $f^{(k)}(a)/k!$ 称为 f 关于 a 的第 k 个泰勒系数. 在很多应用中，我们希望得到原点处的泰勒级数，所以 $a=0$ (在一些教材中，这被称作**麦克劳林级数**). 泰勒级数给出了函数及其导数在一点处的性质，由此可以估算出该函数在其他点处的值.

类似的是，掌握大量相关信息后，可以很好地得到近似

9.2 期望值和矩

期望值，矩：设 X 是定义在 $\mathbb{R}$ 上的随机变量，它的概率密度函数是 f_X. 函数 $g(X)$ 的**期望值**是

$$\mathbb{E}[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\mathrm{d}x & \text{若 } X \text{ 是连续的} \\ \sum_n g(x_n) \cdot f_X(x_n) & \text{若 } X \text{ 是离散的.} \end{cases}$$

最重要的情形是 $g(x) = x^r$. 我们把 $\mathbb{E}[X^r]$ 称为 X 的 r 阶**矩**，把 $\mathbb{E}[(X - \mathbb{E}[X])^r]$ 称为 X 的 r 阶**中心矩**.

两个重要的矩 $\begin{cases} \text{期望（一阶矩）} \\ \text{方差（二阶中心矩）} \end{cases}$

9.3 均值和方差

均值和方差： 设 X 是一个连续型或离散型的随机变量，它的概率密度函数是 f_X.

(1) X 的**均值** (即平均值或期望值) 是一阶矩. 我们把它表示为 $\mathbb{E}[X]$ 或 μ_X (当随机变量很明确时, 通常不给出下标 X, 而只写 μ). 具体地说,

$$\mu = \begin{cases} \int_{-\infty}^{\infty} x \cdot f_X(x)\mathrm{d}x & \text{若 } X \text{ 是连续的} \\ \sum_n x_n \cdot f_X(x_n) & \text{若 } X \text{ 是离散的.} \end{cases}$$

(2) X 的**方差** (记作 σ_X^2 或 $\mathrm{Var}(X)$) 是二阶中心距, 也可以说是 $g(X) = (X - \mu_X)^2$ 的期望值. 同样, 当随机变量很明确时, 通常不给出下标 X, 而只写 σ^2. 把它完整地写出来, 就是

$$\sigma_X^2 = \begin{cases} \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\mathrm{d}x & \text{若 } X \text{ 是连续的} \\ \sum_n (x - \mu_X)^2 f_X(x_n) & \text{若 } X \text{ 是离散的.} \end{cases}$$

因为 $\mu_X = \mathbb{E}[X]$, 所以在一系列代数运算后 (参见引理 9.5.3), 我们有

$$\sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

这个式子把方差和 X 的前二阶矩联系起来, 在很多计算中都非常有用. **标准差**是方差的平方根, 即 $\sigma_X = \sqrt{\sigma_X^2}$.

(3) **技术说明：** 为了保证均值存在, 我们希望 $\int_{-\infty}^{\infty} |x| f_X(x)\mathrm{d}x$ (在连续的情形下) 或 $\sum_n |x_n| f_X(x_n)$ (在离散的情形下) 是有限的.

方差与标准差： 与方差相比, 标准差的优势在于它和均值有相同的单位. 因此, 标准差是衡量结果在均值附近波动幅度的自然尺度.

在均值相同的情况下，可以用方差来区分pdf.

关于技术说明：①均值有限才有意义.

② 关于"$|x|$"

①：$\int_{-\infty}^{+\infty} f(x)\,dx$ 是反常积分，当上下限趋∞的方式不一样时，可能会出现无穷的情况，因此仅知道其存在性是不够的.

② 反例：$x \cdot f_1(x) = \frac{\sin x}{x}$.

$\int_{-\infty}^{+\infty} f(x)dx$是收敛的，但其在正负间不断振荡，不稳定. 而$\int_{-\infty}^{+\infty} \left|\frac{\sin x}{x}\right| dx$是发散的

9.4 联合分布

联合概率密度函数：设 $X_1, X_2, \cdots, X_n$ 都是连续型随机变量，它们的概率密度函数分别是 $f_{X_1}, f_{X_2}, \cdots, f_{X_n}$. 假设每个 X_i 都定义在 $\mathbb{R}$ (实数集) 的一个子集上. 那么, $(X_1, \cdots, X_n)$ 的联合概率密度函数就是一个非负的可积函数 $f_{X_1,\cdots,X_n}$, 满足：对于每一个恰当的集合 $S \subset \mathbb{R}$, 均有

$$\mathrm{Prob}((X_1, \cdots, X_n) \in S) = \int \cdots \int_S f_{X_1,\cdots,X_n}(x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

并且

$$f_{X_i}(x_i) = \int_{x_1 = -\infty}^{\infty} \cdots \int_{x_{i-1} = -\infty}^{\infty} \int_{x_{i+1} = -\infty}^{\infty} \cdots \int_{x_n = -\infty}^{\infty}$$

没有 x_i

$$f_{X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n}(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) \prod_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{d}x_j.$$

我们把 f_{X_i} 称为 X_i 的**边缘概率密度函数**, 可以通过对其他 $n-1$ 个变量求积分来得到.

这 n 个随机变量 $X_1, \cdots, X_n$ 相互独立, 当且仅当

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

对于离散型随机变量, 只需要把积分替换成求和即可.

这是一个极其一般的情况，一般不考虑

联合概率密度函数是体现变量间相互依赖关系的方法之一

给定了n元组的概率

独立：

表 9-2 $(\boldsymbol{X},\boldsymbol{Y})$ 的联合概率密度函数．$\boldsymbol{X}$ 表示独立地抛掷 5 枚均匀硬币，前 3 次掷出正面的次数；$\boldsymbol{Y}$ 表示后 2 次掷出正面的次数

	$Prob(Y=0)$	$Prob(Y=1)$	$Prob(Y=2)$	
Prob$(X=0)$	1/32	2/32	1/32	1/8
Prob$(X=1)$	3/32	6/32	3/32	3/8
Prob$(X=2)$	3/32	6/32	3/32	3/8
Prob$(X=3)$	1/32	2/32	1/32	1/8
	1/4	2/4	1/4	

任取 $X=k_1$, $Y=k_2$.

$Prob(X=k_1, Y=k_2) = Prob(X=k_1)\cdot Prob(X=k_2)$

不独立：

$$f_{U,V}(0,2)=0\neq\frac{1}{8}\cdot\frac{2}{4}=f_U(0)f_V(2).$$

表 9-3 $(\boldsymbol{U},\boldsymbol{V})$ 的联合概率密度函数，其中 $\boldsymbol{U}$ 表示独立地抛掷 5 枚均匀硬币时前 3 次掷出正面的次数，$\boldsymbol{V}$ 表示后 2 次掷出正面的次数

	$Prob(V=0)$	$Prob(V=1)$	$Prob(V=2)$	
Prob$(U=0)$	1/16	1/16	0/16	1/8
Prob$(U=1)$	2/16	3/16	2/16	3/8
Prob$(U=2)$	1/16	3/16	2/16	3/8
Prob$(U=3)$	0/16	1/16	1/16	1/8
	1/4	2/4	1/4	

一个连续型联合概率密度函数：

$$f_X(x,y) = \begin{cases} \frac{1}{\pi} & 0 \le x^2+y^2 \le 1 \\ 0 & \text{其它} \end{cases}$$

X的边缘概率密度函数：

$$f_X(x) = \int_{y=-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy$$

$$= \frac{2}{\pi}\sqrt{1-x^2}$$

9.5 期望的线性性质

定理 9.5.1 (期望的线性性质) 设 $X_1, \cdots, X_n$ 是随机变量，并设 $g_1, \cdots, g_n$ 是满足下列条件的函数：$\mathbb{E}[|g_i(X_i)|]$ 存在且有限. 令 $a_1, \cdots, a_n$ 表示任意实数. 那么

$$\mathbb{E}[a_1 g_1(X_1) + \cdots + a_n g_n(X_n)] = a_1 \mathbb{E}[g_1(X_1)] + \cdots + a_n \mathbb{E}[g_n(X_n)].$$

注意，随机变量不一定是相互独立的. 另外，如果 $g_i(X_i) = \text{c}$ (这里的 c 是固定常数)，那么 $\mathbb{E}[g_i(X_i)] = \text{c}$.

“和的期望等于期望的和”.

懒得抄证明了.

引理 9.5.2 设 X 是一个随机变量, 它的均值为 μ_X, 方差为 σ_X^2. 如果 a 和 b 是任意两个固定常数, 那么随机变量 $Y = \mathrm{a}X + \mathrm{b}$ 有下列结果

$$\mu_Y = \mathrm{a}\mu_X + \mathrm{b} \quad \text{和} \quad \sigma_Y^2 = \mathrm{a}^2\sigma_Y^2.$$

$$\begin{aligned}
\sigma_Y^2 &= \mathbb{E}[(Y - \mu_Y)^2] \\
&= \mathbb{E}\left[((\mathrm{a}X + \mathrm{b}) - (\mathrm{a}\mu_X + \mathrm{b}))^2\right] \\
&= \mathbb{E}\left[(\mathrm{a}X - \mathrm{a}\mu_X)^2\right] \\
&= \mathbb{E}[\mathrm{a}^2(X - \mu_X)^2] \ = \ \mathrm{a}^2\mathbb{E}[(X - \mu_X)^2] \ = \ \mathrm{a}^2\sigma_X^2.
\end{aligned}$$

引理 9.5.3 设 X 是一个随机变量, 那么

$$\mathrm{Var}(X) \ = \ \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mu_X)^2] \\
&= \mathbb{E}[X^2 - 2\mu_X X + \mu_X^2] \\
&= \mathbb{E}[X^2] - \mathbb{E}[2\mu_X X] + \mathbb{E}[\mu_X^2] \\
&= \mathbb{E}[X^2] - 2\mu_X\mathbb{E}[X] + \mu_X^2 \\
&= \mathbb{E}[X^2] - 2\mu_X \cdot \mu_X + \mu_X^2 \\
&= \mathbb{E}[X^2] - \mu_X^2 \ = \mathbb{E}[X^2] - \mathbb{E}[X]^2,
\end{aligned}$$

9.6 均值和方差的性质

定理 9.6.1 如果 X 和 Y 是相互独立的随机变量，那么

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

一种特别重要的情况是

$$\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[X-\mu_X]\mathbb{E}[Y-\mu_Y] = 0.$$

定理 9.6.2 (随机变量之和的均值与方差) 设 $X_1, \cdots, X_n$ 是 n 个随机变量，它们的均值是 $\mu_{X_1}, \cdots, \mu_{X_n}$，方差是 $\sigma^2_{X_1}, \cdots, \sigma^2_{X_n}$. 如果 $X = X_1 + \cdots + X_n$，那么

$$\mu_X = \mu_{X_1} + \cdots + \mu_{X_n}.$$

如果随机变量是相互独立的，那么还能得到

$$\sigma^2_X = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} \quad 或 \quad \mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

如果这些随机变量是独立同分布的 (因此，每个随机变量的均值都是 μ，方差都是 σ^2)，那么

$$\mu_X = n\mu \quad 且 \quad \sigma^2_X = n\sigma^2.$$

对于方差而言，需要保证事件之间的独立性

反例： X 与 $-X$.

$\mathrm{Var}(X-X) = 0$

But. $\mathrm{Var}(X) + \mathrm{Var}(-X) = 2\,\mathrm{Var}(X)$.

对于随机变量之和的方差计算的证明:

重写代数表达式.

$$
\begin{aligned}
\sigma_X^2 &= \mathbb{E}[(X-\mu_X)^2] \\
&= \mathbb{E}\left[((X_1+X_2)-(\mu_{X_1}+\mu_{X_2}))^2\right] \\
&= \mathbb{E}\left[((X_1-\mu_{X_1})+(X_2-\mu_{X_2}))^2\right] \\
&= \mathbb{E}\left[(X_1-\mu_{X_1})^2+2(X_1-\mu_{X_1})(X_2-\mu_{X_2})+(X_2-\mu_{X_2})^2\right].
\end{aligned}
$$

利用期望的线性性质, 可以把和的期望改写成期望的和, 即

$$
\begin{aligned}
\sigma_X^2 &= \mathbb{E}\left[(X_1-\mu_{X_1})^2\right]+\mathbb{E}\left[2(X_1-\mu_{X_1})(X_2-\mu_{X_2})\right]+\mathbb{E}\left[(X_2-\mu_{X_2})^2\right] \\
&= \sigma_{X_1}^2+2\mathbb{E}\left[(X_1-\mu_{X_1})(X_2-\mu_{X_2})\right]+\sigma_{X_2}^2.
\end{aligned}
$$

因为 X_1 与 X_2 独立, $(X_1-\mu_{X_1})$ 与 $(X_2-\mu_{X_2})$ 也独立.

所以, 有 $\sigma_X^2 = \sigma_{X_1}^2 + 2E(X_1-\mu_{X_1})\cdot E(X_2-\mu_{X_2}) + \sigma_{X_2}^2$

$= \sigma_{X_1}^2 + \sigma_{X_2}^2$

$\Rightarrow$ 利用期望的线性性质

9.7 偏斜度与峰度.

↙ 三阶中心矩　　↓ 四阶中心矩.

偏斜度衡量了概率分布的对称性.

如：正态分布的偏斜度是0.完全对称.

偏斜度>0，则右侧尾巴比左侧更厚、长

〃 <0，则左侧尾巴更厚、长.

峰度衡量的是概率密度函数到达峰值的方式.

峰度越小，到达峰值越平稳，反之有一个尖的点和陡峭的落点.

9.8 协方差

协方差: 设 X 和 Y 是两个随机变量. X 和 Y 的协方差记作 σ_{XY} 或者 $\mathrm{Cov}(X,Y)$, 它的表达式为

$$\sigma_{XY} = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)].$$

注意, $\mathrm{Cov}(X,X)$ 等于 X 的方差. 另外, 如果 $X_1,\cdots,X_n$ 都是随机变量, 并且 $X = X_1+\cdots+X_n$, 那么

$$\mathrm{Var}(X) = \sum_{i=1}^{n}\mathrm{Var}(X_i) + 2\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i,X_j).$$

① X_1, X_2 的协方差为 0 并不代表两变量相互独立.

② 协方差衡量了两个变量的线性相关程度.

相关系数:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

是对协方差的标准化. $\rho \in [-1, 1]$

绝对值越大越相关.

对于任意两变量 X，Y 有

$$Cov(X,Y) = E(X,Y) - \mu_X \mu_Y$$

证明：$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

$$= E(XY - X\cdot\mu_X - Y\cdot\mu_Y + \mu_X\mu_Y)$$

$$= E(XY) - \mu_X\mu_Y$$

练习补充：

① 标准正态分布：

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

奇数阶中心矩 = 0

四阶中心矩 $\mu^4 = 3\sigma^2$

所有完全对称的pdf都满足奇矩=0

计算n阶矩时注意可能遇到奇函数对称区间上的积分

② 协方差 $cov(X_1, X_2)$ 的取值范围

$$\left[-\sqrt{Var(X_1)Var(X_2)}, \sqrt{Var(X_1)Var(X_2)}\right]$$

③ 不要求严格独立的：

期望的线性性 .（ $E(X+Y) = E(X) + E(Y)$ ）

要求严格独立的：

$E(XY) = E(X) \cdot E(Y)$

$Var(X+Y) = Var(X) + Var(Y)$

④ 联合概率密度函数的判定定理

$$\begin{cases} \text{非负性} \\ \text{归一性} \left(\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1 \right) \end{cases}$$

9.10.21

$X \sim B(n,p)$，计算二阶矩/三阶矩

$$Var(X) = E(X^2) - E^2(X)$$

$$Var(X) = n\cdot(1-p)\cdot p$$

$$E(X) = np$$

$$\Rightarrow E(X^2) = np\cdot(1-p) + n^2p^2$$

三阶矩

考虑二项分布的矩生成函数

$$M_X(t) = E(e^{tX})$$

$$= E\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k\right) \quad (泰勒展开)$$

$$= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$$

$$\frac{d^n}{dt^n} M_X(t) = \sum_{k=0}^{\infty} \frac{k!}{(k-n)!} \cdot \frac{t^{k-n}}{k!} \cdot E(X^k)$$

$$= \sum_{k=0}^{\infty} \frac{t^{k-n}}{(k-n)!} E(X^k)$$

令 $n=3$，$t=0$，有 $E(X^3) = \frac{d^3 M_X(t)}{dt^3}\Big|_{t=0}$

只有 $k=3$ 这一项会保留 $E(X^3)$

$$E(e^{tX}) = \sum_{k=0}^{n} e^{tk} \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} (e^t \cdot p)^k \cdot (1-p)^{n-k} \binom{n}{k}$$

$$= (e^t \cdot p + 1 - p)^n$$

$$E(X^3) = \frac{d^3}{dt^3} E(e^{tX}) \Big|_{t=0}.$$

$$= \cdots$$

公式化：求二项分布n阶矩

$$E(X^n) = \frac{d^n}{dt^n} E(e^{tX}) \Big|_{t=0}.$$

$$= \frac{d^n}{dt^n} (e^t p + 1 - p) \Big|_{t=0}$$

边缘累积分布函数：

考虑Y取所有可能值的情况下

$$F_X(x) = Pr(X \le x)$$

计算方法1：

$$F_X(x) = \lim_{Y \to +\infty} F_{X,Y}(x, y)$$

方法2：　　★

$$f_X(x) = \int_{Y=-\infty}^{+\infty} f_{X,Y}(X, Y)\, dY$$

$$F_X(X) = \int_{-\infty}^{X} f_X(t)\, dt.$$

由联合pdf计算边缘cdf的经典方法.