基础数理统计

第三章　　数学期望

期望的定义

定义 1 (期望)

对于随机变量 X，期望的定义为：

$$\begin{aligned}\mu &= \mu_X = E(X) = \int x dF(x) \\ &= \begin{cases} \sum_x xP(X=x) = \sum_x xf(x) & X\text{ 是离散随机变量} \\ \int xf(x)dx & X\text{ 是连续随机变量}. \end{cases}\end{aligned}$$

说明：如果 $\int_x |x| dF_X(x) < \infty$，称 $\mathrm{E}(X)$ 存在。

例 1-几何分布

例 1

抛一枚均匀的硬币一直抛到正面朝上为止，平均抛的次数为:

$$E(X) = \sum_{k=1}^{\infty} \frac{k}{2^k} = 2.$$

计算: 更一般的，对于 $|x| < 1$, 考虑级数

$$\sum_{k=1}^{\infty} kx^{k-1} = (\sum_{k=1}^{\infty} x^k)' = (\frac{x}{1-x})' = \frac{1}{(1-x)^2}.$$

取 $x = 1/2$, 得到

$$\sum_{k=1}^{\infty} \frac{k}{2^k} = (1/2) \sum_{k=1}^{\infty} k(1/2)^{k-1} = \frac{1}{2} * 4 = 2.$$

例 2-正态分布

正态分布 $N(\mu, \sigma^2)$ 的密度函数为：

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

计算期望

$$\begin{aligned} E(X) = \int xf(x)dx &= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}dx \\ &= \int_{-\infty}^{\infty} \frac{\sigma y + \mu}{\sqrt{2\pi}} \exp\{-\frac{y^2}{2}\}dy \\ &= \mu. \end{aligned}$$

例 3-Cauchy 分布

Cauchy 分布的密度函数为：

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

分布函数为：

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\pi}\arctan|_{-\infty}^{x} = \frac{\arctan(x)}{\pi} + \frac{1}{2}.$$

期望不存在，因为

$$E(|X|) = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}dx = \frac{2}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}dx = \infty.$$

如何理解 Cauchy 分布期望不存在

如果对柯西分布多次模拟并取其均值，会发现均值不会稳定，这是因为 Cauchy 分布的尾部较厚，容易出现尾部的观察值。

定理 1 (懒惰统计学家法则)

令 $Y = r(X)$，则

$$E(Y) = E(r(X)) = \int r(x) dF_X(x).$$

例 2

将一根单位长度的棍子从中间某一点折断，令 Y 为较长一段的长度，Y 的均值为多少?

X 为折断点，则 $X \sim U(0,1)$，$Y = \max\{X, 1-X\}$，

$$\mathrm{E}(Y) = \int_0^{1/2} (1-x) dx + \int_{1/2}^1 x dx = \frac{3}{4}.$$

定理 2

如果 k 阶矩存在且 $j < k$，则 j 阶矩存在。

3.2 期望的性质

- 对于任意两个随机变量 X, Y,
$$E(aX + bY) = aE(X) + bE(Y).$$
进一步 $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$.
- 如果两个随机变量 X, Y 是相互独立的，那么
$$E(XY) = E(X) \cdot E(Y).$$
进一步，如果 $X_1, X_2, \ldots, X_n$ 为独立随机变量，则 $\mathrm{E}(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} \mathrm{E}(X_i)$。
- 对于任意的 $x \in \mathbb{R}$, 定义示性函数 $Y = I(X \le x)$,
$$EY = EI(X \le x) = P(X \le x) = F(x),$$
所以概率本身也可以理解为一种期望.

3.3 方差和协方差

方差的定义

定义 2 (方差)

方差刻画了随机变量的离散程度，定义为：

$$\sigma^2 = V(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2.$$

方差的存在性对应 $E(X^2)$ 存在。

方差开根号 σ 或者 $\sqrt{V(X)}$ 记为标准差 *(The standard deviation)*.

对于一个随机变量 X,

$$\arg\min_a E(X-a)^2 = EX = \mu,$$
$$\min_a E(X-a)^2 = E(X-EX)^2 = \mathrm{V}(X).$$

方差的性质

假设方差存在，则它具有如下性质：

1. $\mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2$;
2. $\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$;
3. 如果 $X_1, X_2, \ldots, X_n$ 为独立随机变量，$a_1, a_2, \ldots, a_n$ 为常数，则

$$\mathrm{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{V}(X_i)$$

样本均值和样本方差

如果 $X_1, X_2, \ldots, X_n$ 为 i.i.d 随机变量，则定义样本均值为

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

样本方差为

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

定理 3

$X_1, X_2, \ldots, X_n$ 为 $i.i.d$ 随机变量且 $\mu = E(X_i)$, $\sigma^2 = V(X_i)$，则

$$E(\bar{X}_n) = \mu, \quad V(\bar{X}_n) = \frac{\sigma^2}{n}, \quad E(S_n^2) = \sigma^2.$$

方差的性质-续

- 对于一列随机变量 $X_1, \cdots, X_n$，如果两两不相关，

$$\mathrm{V}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{V}(X_i).$$

- 如果随机变量 $X_1, \cdots, X_n$ 相互独立，

$$\mathrm{V}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{V}(X_i).$$

协方差和相关系数

定义 3 (协方差和相关系数)

对于两个随机变量 X, Y, 均值分别为 μ_X, μ_Y，标准差分别为 σ_X, σ_Y，定义 X 和 Y 的协方差为

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

相关系数定义为

$$\rho = \rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

协方差和相关系数

定理 4

协方差满足 $Cov(X,Y) = E(XY) - E(X)E(Y)$。相关系数满足 $-1 \le \rho_{X,Y} \le 1$，且

(1) $\rho_{X,Y} = 1$ 当且仅当存在 $a > 0, b \in \mathbb{R}$ 使得 $P(Y = aX + b) = 1$;

(2) $\rho_{X,Y} = -1$ 当且仅当存在 $a < 0, b \in \mathbb{R}$ 使得 $P(Y = aX + b) = 1$;

(3) X, Y 独立，则 $\rho_{X,Y} = 0$，否则通常不成立。

定理 5

对于随机变量 $X_1, X_2, \ldots, X_n$，我们有

$$V(\sum_i a_i X_i) = \sum_i a_i^2 V(X_i) + 2\sum\sum_{i<j} a_i a_j Cov(X_i, X_j).$$

例 3

二项分布 *(Binomial Distribution)* 是刻画 n 个独立事件成功 k 次的概率，

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

可以考虑

$$X=\sum_{i=1}^{n} X_i,$$

其中 $X_1,\cdots,X_n$ 是独立同分布的成功概率为 p 的 *Bernoulli* 分布，

$$E(X)=E(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n} E(X_i)=np,$$

$$V(X)=V(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n} V(X_i)=np(1-p).$$

3.4 一些重要随机变量的期望和方差

1 3.1 随机变量的期望
2 3.2 期望的性质
3 3.3 方差和协方差
4 3.4 一些重要随机变量的期望和方差
5 3.5 条件期望
6 3.6 矩母函数

对于一个向量型随机变量 $X=(X_1,\ldots,X_p)^\top$, 总体协方差矩阵

$$\Sigma = cov(X) = (\text{cov}(X_i, X_j))_{p\times p},$$

是一个非常重要的参数，其刻画了数据的离散程度 (类似于一元随机变量中的方差). 总体协方差矩阵的逆矩阵 $\Omega=\Sigma^{-1}$ 也是一个重要的度量，一般称为精度矩阵 (precision matrix)

引理 1

对于一个多元随机变量 X, 记

$$\boldsymbol{\mu} = EX \in \mathbb{R}^p, \ \Sigma = cov(X) = (cov(X_i, X_j))_{p\times p} \in \mathbb{R}^{p\times p},$$

那么对于其任意的线性组合 $Y = \boldsymbol{\beta}^\top X$, 我们有

$$E(Y) = \boldsymbol{\beta}^\top \boldsymbol{\mu}, \ V(Y) = \boldsymbol{\beta}^\top \Sigma \boldsymbol{\beta}.$$

如果 A 为矩阵，则

$$E(AX) = A\boldsymbol{\mu}, \ \ V(AX) = A\Sigma A^\top.$$

例 4

$X \sim Multinomial(n, p)$，则 $E(X) = n(p_1, p_2, \ldots, p_k)^\top$ 且

$$V(X) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \ldots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \ldots & -np_2p_k \\ \vdots & \vdots & \ldots & \vdots \\ -np_kp_1 & -np_kp_2 & \ldots & np_k(1-p_k) \end{pmatrix}.$$

3.5 条件期望

条件分布

对二元随机变量 (X, Y)，在给定 $X = x$ 时候，可以定义条件分布 $Y|X = x$, 密度函数或者概率函数记为 $f(y|x)$.

条件期望

定义 4

给定 $Y = y$ 情况下 X 的条件期望为

$$E(X|Y=y) = \begin{cases} \sum x f_{X|Y}(x|y) & \text{离散情形} \\ \int x f_{X|Y}(x|y)\,dx & \text{连续情形} \end{cases}.$$

如果 $r(x, y)$ 为 x 和 y 的函数，则

$$E(r(X,Y)|Y=y) = \begin{cases} \sum r(x,y) f_{X|Y}(x|y) & \text{离散情形} \\ \int r(x,y) f_{X|Y}(x|y)\,dx & \text{连续情形} \end{cases}.$$

定理 6 (迭代期望法则)

对于任意的二元随机变量，假设条件期望存在，那么

$$E\left(E(Y|X)\right) = E(Y), \quad E\left(E(X|Y)\right) = E(X).$$

更一般的,

$$E\left(E(r(X,Y)|X)\right) = E(r(X,Y)).$$

例 5

假设 $X \sim Uniform(0,1)$，假设 $Y|X=x \sim Uniform(x,1)$，求 $E(Y)$。

注意到，$f_{Y|X}(y|x) = 1/(1-x)$，$y \in (x,1)$，故

$$\mathrm{E}(Y|X=x) = \int_x^1 y(1-x)^{-1} dy = \frac{1+x}{2}.$$

故 $\mathrm{E}(Y|X) = \dfrac{1+X}{2}$，$\mathrm{E}(Y) = \mathrm{E}\left(\mathrm{E}(Y|X)\right) = \dfrac{3}{4}$.

条件方差

定义 5

条件方差定义为

$$V(Y|X=x)=\int (y-\mu(x))^2 f_{Y|X}(y|x)dy,$$

其中 $\mu(x)=E(Y|X=x)$.

定理 7

对随机变量 X 和 Y，有

$$V(Y)=E(V(Y|X))+V(E(Y|X)).$$

例 6

从美国任意挑选一个县，在这个县里任意挑选 n 个人，令 X 表示这些人中患有某种疾病的人数。如果 Q 表示该县患有该疾病的人数所占的比例，Q 是一个随机变量 (每个县城的比例可能不同)，假设服从 $Uniform(0,1)$。给定 $Q=q$,X 服从二项分布 $B(n,q)$。则

$$E(X|Q)=nQ, \quad V(X|Q)=nQ(1-Q).$$

从而有

$$\begin{aligned} E(X) &= nE(Q)=\frac{n}{2} \\ V(X) &= n^2V(Q)+nEQ(1-Q)=\frac{n^2}{12}+\frac{n}{6}. \end{aligned}$$

回归函数

我们考虑任意的函数 $g(\cdot)$，最小化 $E(Y-g(X))^2$，记 $Z=E(Y|X)$, 我们有

$$\begin{aligned} E(Y-g(X))^2 &= E(Y-Z+Z-g(X))^2 \\ &= E(Y-Z)^2+E(Z-g(X))^2 \\ &\quad +2E(Y-Z)(Z-g(X)) \\ &= E(Y-Z)^2+E(Z-g(X))^2, \end{aligned}$$

所以在**均方损失**下，最优的回归函数为 $r(x)=E(Y|X=x)$。

作业：13，17，18，21，22

3.6 矩母函数

矩母函数

定义 6 (矩母函数或拉普拉斯变换)

X 的矩母函数 *(MGF)* 或拉普拉斯变换定义为

$$\phi_X(t) = E(\exp^{tX}) = \int e^{tx} dF(x).$$

其中 t 为实数。

说明：假设 MGF 在 $t = 0$ 的某个开区间中的任意 t 都存在，则

$$\phi^{(k)}(0) = \mathrm{E}(X^k).$$

例 7

$X \sim Exp(1)$，则 $\phi_X(t) = \dfrac{1}{1-t}$，$\phi'(0) = 1$，$\phi''(0) = 2$，故 $E(X) = 1, V(X) = 1$。

MGF 的性质

引理 2

MGF 的性质如下:

(1) 如果 $Y = aX + b$，则 $\phi_Y(t) = e^{bt}\phi_X(at)$;

(2) 如果 $X_1, X_2, \ldots, X_n$ 独立且 $Y = \sum_i X_i$，则 $\phi_Y(t) = \prod_i \phi_i(t)$，其中 $\phi_i(t)$ 为 X_i 的 *MGF*。

例 8

$Binomial(1, p)$ 的 *MGF* 为 $\phi(t) = (pe^t + q)$，$q = 1 - p$，故 $Binomial(n, p)$ 的 *MGF* 为 $\phi(t) = (pe^t + q)^n$，其中 $q = 1 - p$。

定理 8

令 X 和 Y 为随机变量，如果对以 *0* 为中心的某个开区间里所有的 t 有 $\phi_X(t) = \phi_Y(t)$，则 X, Y 同分布，记为 $X \overset{d}{=} Y$。

例 9

$X \sim Binomial(n_1, p)$，$Y \sim Binomial(n_2, p), X, Y$ 相互独立，易知 $X + Y \sim Binomial(n_1 + n_2, p)$。

作业：23，24