No.

Date . .

第七.八章

一. 大数定律

Sample mean 定义: $X_1, \cdots, X_n$ 为 i.i.d, 有 finite mean μ and variance σ^2 则 $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$, 且 mean 为 μ. variance 为 σ^2/n

i.e., $X_{sum} = \sum_{j=1}^{n} X_j$ mean 为 $n\mu$, 方差为 $n\sigma^2$

Theorem: Strong Law of Large Numbers (SLLN)

$n\to\infty$ 时, $\bar{X}_n \to \mu$ 的概率为 1

Weak Law of Large Numbers (WLLN)

$\forall \varepsilon > 0$, $P(|\bar{X}_n - \mu| > \varepsilon) \to 0$, $n \to \infty$.

二. Inequalities.

①Cauchy - Schwarz: Any r.v.s X and Y:

$$|E(XY)| \le \sqrt{E(X^2)E(Y^2)}$$

②Jensen: If f is a convex function, $0 \le \lambda_1, \lambda_2 \le 1$, $\lambda_1 + \lambda_2 = 1$, then for any x_1, x_2:

$$f(\lambda_1 x_1 + \lambda_2 x_2) \le \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

$\Rightarrow$ X be a r.v., g is a convex function. Then:

$E(g(X)) \ge g(E(X))$; if else g is a concave function, then $E(g(X)) \le g(E(X))$.

等号: iff: $\exists$ constant a, b, s.t., $P(g(X) = a + bX) = 1$

加条件: 凸: $E[g(X)|R] \ge g(E[X|R])$.

凹: $E[g(X)|R] \le g(E[X|R])$

KOKUYO

No.

Date

③: Entropy Theory: Let X be a discrete r.v., whose distinct possible values are $a_1, a_2, \cdots, a_n$ with $p_1, p_2, \cdots, p_n$ respectively ($\sum_{i=1}^{n} p_i = 1$)

Entropy of X: $H(X) = \sum_{j=1}^{n} p_j \log_2 (1/p_j)$

Using Jensen: $H(X)$ max when distribution is uniform.

④ Kullback-Leibler Divergence.

Let $\vec{p} = (p_1, \cdots, p_n)$ & $\vec{r} = (r_1, \cdots, r_n)$ be two probability vectors ($\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} r_i = 1$). 把它们看成一个随机变量可能的概率质量函数，其支撑集由n个值组成。则$\vec{p}$与$\vec{r}$之间的 Kullback-Leibler 定义为:

$$D(\vec{p}, \vec{r}) = \sum_{j=1}^{n} p_j \log_2 (1/r_j) - \sum_{j=1}^{n} p_j \log_2 (1/p_j)$$

⑤ 马尔可夫不等式: For any r.v. X and constant $a>0$:

$$P(|X| \geq a) \leq \frac{E|X|}{a}$$

Proof: 令 $Y = \frac{|X|}{a}$, 则欲证: $P(Y \geq 1) \leq E(Y)$.

$\because I(Y \geq 1) \leq Y$ $\therefore$ if $I(Y \geq 1) = 0$, $Y \geq 0$; if $I = 1$, $Y \geq 1$.

⑥ 切比雪夫 (Chebyshev's): X have mean μ & variance σ

Then for any $a>0$:

$$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Proof: Using Markov: $P(|X-\mu| \geq a) = P((X-\mu)^2 \geq a^2)$

$$\leq \frac{E(X-\mu)^2}{a^2} = \frac{\sigma^2}{a^2}$$

⚠补：⑧：坎特利不等式：$P(X-\mu \geq a) \leq \frac{\sigma^2}{a^2+\sigma^2}$ No.

$P(X-\mu \leq -a) \leq \frac{\sigma^2}{\sigma^2+a^2}$ Date

⑦ 切尔诺夫不等式（Chernoff）：

For any r.v. X and constants $a>0$ and $t>0$:

$$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$$

Proof：$g(x)=e^{tx}$，它可逆且严格递增。

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}} \text{ (Markov)}$$

三、条件期望：Given an event

Recall：$P_{X|A}(x) = P(X=x|A) = \frac{P(\{X=x\} \cap A)}{P(A)}$

Bayes' $P_{X|A}(x) = \frac{P(A|X=x)\, P(X=x)}{P(A)}$

LOTP：With a partition $A_1, \cdots, A_n$，$P(A_i)>0$, $i=1,\cdots,n$

$$P(X=x) = \sum_{i=1}^{n} P_{X|A_i}(x)\, P(A_i)$$

上述为条件PMF，下为条件PDF：

$$f_{X|A}(x) = [P(X \leq x|A)]'$$

LOTP：With a partition $A_1, \cdots, A_n$，$P(A_i)>0$, $i=1,\cdots,n$

$$f_X(x) = \sum_{i=1}^{n} P(A_i)\, f_{X|A_i}(x)$$

Baye's：$f_{X|A}(x) = \frac{P(A|X=x)}{P(A)} \cdot f_X(x)$

Def：$E(Y|A) = \sum_y y \cdot P(Y=y|A) = \sum_y y \cdot P_{Y|A}(y)$ (DS)

$$E(Y|A) = \int_{-\infty}^{\infty} y \cdot f_{Y|A}(y)\,dy$$

$$E(g(Y)|A) = \sum_y g(y) \cdot P_{Y|A}(y)$$

$$E(g(Y)|A) = \int_{-\infty}^{\infty} g(y) \cdot f_{Y|A}(y)\,dy$$

KOKUYO

No.

Date

Eg: $X \sim Expo(\lambda)$, find $E(X|X>1)$ & $Var(X|X>1)$.

"$A = X>1$" $P(A) = P(X>1) = \int_1^\infty \lambda e^{-\lambda x} dx = e^{-\lambda}$.

$E(X|X>1) = \int_1^\infty x \cdot f_{X|A}(x)^* dx = \int_1^\infty x \cdot \lambda e^{-\lambda(x-1)} dx = 1+\frac{1}{\lambda}$

*: $f_{X|A}(x) = \frac{f(x)}{P(A)} = \lambda e^{-\lambda(x-1)}$ $(x>1)$

$E(X^2|X>1) = \int_1^\infty x^2 f_{X|A}(x) dx = \int_1^\infty x^2 \lambda e^{-\lambda(x-1)} dx = \frac{\lambda^2+2\lambda+2}{\lambda^2}$

$\therefore Var(X|X>1) = E(X^2|X>1) - [E(X|X>1)]^2 = \frac{1}{\lambda^2}$

Law of Total Expectation

Let $A_1, \cdots, A_n$ be a partition of a sample space with $P(A_i) > 0$. Y be a r.v. on this sample space:

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

四. 条件期望: Given an r.v

Let $g(x) = E(Y|X=x)$. 则 $E(Y|X)$ 定义为 $g(X)$

Eg: 木棍长为1，随机选在X点折断，$X=x$，再在$[0, x]$上随机选点Y. $E(E(Y|X))$ & variance.

Solution: $X \sim Unif(0,1)$ $Y \sim Unif(0,x)$.

$E(Y|X=x) = \frac{x}{2}$ 用X换x: $E(Y|X) = X/2$

$\therefore E(E(Y|X)) = E(X/2) = 1/4$

$Var(E(Y|X)) = Var(X/2) = 1/48$

*: X具象为x，$E(Y|X)$具象为$g(x)$

Campus

No.

Date . .

Theorem: 1. 若 X、Y 独立，$E(Y|X)=E(Y)$.

2. For any function h:

$$E(h(X)Y|X) = h(X)E(Y|X).$$

条件期望性质：$E(Y_1+Y_2|X) = E(Y_1|X)+E(Y_2|X)$.

Adam's Law. For any r.v.s X & Y:

$$E(E(Y|X)) = E(Y)$$

类似于LOTP

Proof: 令 $E(Y|X)=g(X)$

$$E(g(X)) = \sum_x g(x)P(X=x) = \sum_x \left(\sum_y y P(Y=y, X=x)\right) P(X=x)$$

$$= \sum_x \sum_y y P(X=x) P(Y=y|X=x) = \sum_y y \sum_x P(X=x, Y=y)$$

$$= \sum_y y P(Y=y) = E(Y).$$

With Conditioning: $E(E(Y|X,Z)|Z) = E(Y|Z)$

$$E(E(X|Z,Y)|Y) = E(X|Y).$$

条件方差：$Var(Y|X) = E((Y-E(Y|X))^2|X)$.

$$\Leftrightarrow Var(Y|X) = E(Y^2|X) - [E(Y|X)]^2$$

Eve's law: $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$.

Proof: $g(X) = E(Y|X)$, $E(g(X)) = E(Y)$

$$E(Var(Y|X)) = E(E(Y^2|X) - g(X)^2) = E(Y^2) - E(g(X)^2)$$

$$Var(E(Y|X)) = E(g(X)^2) - [E(g(X))]^2 = E(g(X)^2) - (EY)^2.$$