#SP

Probability

概率的连续性

为了说明概率的连续性，我们需要定义事件的 Increasing Sequence 和 Decreasing Sequence 。如果一个序列是 Increasing，那么

$$E_n \subset E_{n+1}$$

也就是下一个事件总能包含上一个，Decreasing 定义与其相反。那么，我们定义一个 Increasing Seq 的极限为：

$$\lim_{n\to\infty} E_n = \bigcup_i E_i$$

而一个 Decreasing Seq 的极限为：

$$\lim_{n\to\infty} E_n = \bigcap_i E_i$$

那么，我们有：

Theorem

如果 E_n 是上升序列或者下降序列，那么：

$$\lim P(E_n) = P(\lim E_n)$$

Proof

我们定义一些相互排斥的事件 F_i：

$$F_1 = E_1$$

$$F_n = E_n(\bigcup_i^{n-1} E_i)^c = E_n(E_{n-1})^c$$

那么我们容易注意到：

$$\bigcup_{i=1}^n F_i = \bigcup_{i=1}^n E_i$$

那么

$$\begin{aligned} P(\bigcup^{\infty} E_i) &= P(\bigcup^{\infty} F_i) \\ &= \lim_{n\to\infty} \sum_i^n P(F_i) \\ &= \lim_{n\to\infty} P(\bigcup^n F_i) \\ &= \lim_{n\to\infty} P(E_i) \end{aligned}$$

（主要证明方法：构造互斥序列 F_i，从而将括号里面的并集符号转换成括号外面的求和号，并且在外面取极限）
另外一边证明方式类似，不再赘述。

Lemma : Borel-Catelli Lemma

对于一个序列 E_i，如果

$$\sum_{i}^{\infty} P(E_i) < \infty$$

则记事件 A 为：序列中无穷个事件发生，那么

$$P(A) = 0$$

Proof

我们首先定义序列的上极限：

$$\lim_{i\to\infty} \sup E_i = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i$$

那么，显然可以知道，如果有无穷个 E_i 发生，那么序列的上极限必然发生。那么，由于 $\bigcap_{i=n}^{\infty} E_i$ 是一个递减的序列，我们将无数个递减的序列并起来，那么：

$$\begin{aligned} P(\lim_{i\to\infty} \sup E_i) &= P(\lim_{n\to\infty} \bigcup_{i=1}^{n} E_i) \\ &= \lim_{n\to\infty} P(\bigcup_{i=1}^{n} E_i) \\ &= 0 \end{aligned}$$

其中，最后一步就是 Cauchy 收敛准则

Converse to BC-Lemma

如果 E_i 是**独立事件**组成的序列，且

$$\sum_{n-1}^{\infty} P(E_n) = \infty$$

那么，事件 A： 序列中有无穷个事件 E_i 发生的概率为 1.

Proof

上面已经说过，由于序列的上极限是一系列递减事件的交，那么（注意：这里不能把事件的并的概率放缩成把所有事件的概率加起来，这样会导致一个小于号，然后方向就不对了）

$$\begin{aligned} P(A) &= P(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} E_i) \\ &= \lim_{n\to\infty} P(\bigcup E_i) \\ &= \lim_{n\to\infty} [1 - P(\bigcap E_i^c)] \end{aligned}$$

再注意到：

$$\begin{aligned} P(\bigcap_i E_i^c) &= \prod_i P(E_i^c) \\ &= \prod_i (1 - P(E_i)) \\ &\leq \exp(-\sum_{n}^{\infty} P(E_i)) \\ &= 0 \end{aligned}$$

那么原问题得证。

Random Variables

Mathmatical Expectations

我们介绍一些可能有用的等式。使得 A_i 代表事件，I_i 是一系列的 Indicator Variables。定义：

$$N = \sum_i I_i$$

那么注意到：

$$(1-1)^N = \sum_{i=0}^{N} C_N^i (-1)^i = \sum_{i=0}^{n} C_N^i (-1)^i$$

定义 I，使得在 $N > 0$ 时 $I > 0$，否则 $I = 0$。那么注意到：

$$1 - I = \sum_{i=0}^{n} C_N^i (-1)^i$$

可以改写成：

$$I = \sum_{i=1}^{n} C_N^i (-1)^{i+1}$$

（指数上的 $i+1$ 是从一个负号变过来的）
两边取期望：

$$E[I] = E[N] - E[C_N^2] + \cdots + (-1)^{n+1} E[C_N^n]$$

然后又可以注意到 $E[I]$ 的意义是 A_i 中至少一个事件发生的概率，那么：

$$E[I] = P(N > 0) = P(\bigcup A_i)$$

那么你其实可以得到等式：

$$P(\bigcup A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) \cdots + (-1)^{n+1} P(A_1 A_2 \cdots A_n)$$

其他公式也可以使用类似的方式导出。例如，我们想要求解恰好只有 r 个事件出现的概率，只要定义 indicator variable :

$$C_N^r (1-1)^{N-r} = I_r$$

需要补充

Moment Generating, Characteristic Functions, and Laplace Transforms

随机变量 X 的矩母函数 (Moment Generating Function)定义为：

$$\psi(t) = \int e^{tx} dF(x) = \int e^{tx} f(x) dx = E[e^{tX}]$$

它首要的作用是方便地求解 X 的各阶矩：

$$\psi^{(n)}(t) = E[X^n e^{tX}]$$

接下来计算 $t = 0$ 时的值即可。显然，这个积分可能不收敛，矩母函数未必存在。因此我们定义特征函数 (Characteristic Function)

$$\phi(t) = E[e^{itX}]$$

显然，我们是可以定义联合分布的矩母函数和特征函数。此外，我们定义随机变量的 Laplace 变换为：

$$\tilde{F}(s) = \int_0^\infty e^{-sx} dF(x)$$

其中 $s = a + bi$。

Conditional Expectation

在 $Y = y$ 给定时，X 的条件分布显然为：

$$P(X \le x | Y = y) = \frac{P(X \le x, Y = y)}{P(Y = y)}$$

那么

$$E[X|Y = y] = \int x \, dF(x|Y = y)$$

条件期望的一个性质是：

$$E[X] = E[E[X|Y]] = \int E[X|Y = y] dF(y)$$

这里缺一个 Bayes 推断

Exponential Distribution, Lack of Memory, and Hazard Rate Functions

指数分布的矩母函数为:

$$E[e^{tX}] = \frac{\lambda}{\lambda - t}$$

我们说指数分别是无记忆性的，那么：

$$P[X > s + t | X > t] = P[X > s]$$

或者，我们换个说法：记

$$\bar{F}(x) = P(X \ge x)$$

那么

$$\bar{F}(s + t) = \bar{F}(s)\bar{F}(t)$$

指数分布的失效率函数被记为：

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}$$

意义是在单位时间内失效的零件占目前所剩零件的比例。

Some Probability Inequalities

Lemma : Markov's Inequality

对于 $a > 0$，

$$P[X \geq a] \leq \frac{E[X]}{a}$$

这个容易证明，定义 Indicator Variable $I(x \geq a)$，写出：

$$aI(x \geq a) \leq X$$

两边取期望即可

Theorem : Chernoff Bounds

$$P(X \geq a) \leq e^{-ta} M(t)$$

这个容易证明：

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

Theorem : Jensen's Inequality

如果 f 是 Convex function，那么：

$$E[f(X)] \geq f(E[X])$$

Limit Theorems

Strong Law of Large Numbers

如果 X_i 是 iid 的均值为 μ 的变量，那么

$$P\left(\lim_{n \to \infty} \frac{1}{n} \sum_i X_i = \mu\right) = 1$$

Central Limit Theorem