概率论笔记

阴月有晴

国庆

手机:18576427500
邮箱 1：chenay@sustc.edu.cn
邮箱 2：achen@liv.ac.uk

目录

1 Probability

1.1 simple concepts

Ω 代表 sample space,ω 代表基本事件，f 代表 2^{Ω}, 其中有 $2^{|\Omega|}$ 个元素
大写字母 ABC 代表事件，大写字母 XYZ 代表随机变量
ϕ: 例如抛四次硬币，不可能出现反反反。

Operation of Eventl

- Union 并 $\cup$
- Intersection 交 $\cap$
- complement 德摩根律 $A \backslash B$
- contain $A \subset B$
- disjoint A 和 B 不可能同时发生
- **概率测度**：$P : f \to [0, 1]$，不可能事件概率为零，概率为零的事件未必为不可能事件。必然事件亦同

概率空间：(Ω, f, p)

1.2 Computing Probabitity

条件概率，定义，可推广至多个。

Polya's urn schenme 一个例子

School boy 又一个例子

Law of Total Probability: Let$B_1, B_2, B_i, ..$ 为 Ω 的一个划分

由全概率公式可进一步推出贝叶斯公式。

独立性的含义。必然事件和零概率事件与任何事件都独立。分析定义和直观定义。并且也可以进一步推广 (由两个事件推广至 n 个事件)。

2 Random Variables

随机变量的概念：$\Omega \to \mathbb{R}$，将事件的结果与数相关，赋予事件以意义。在数这方面可以进行推广。

Bernouli Random variable，一种很常见的模型。

Binomial Random variable, 是 Bernouli 由一次推广至 n 次的结果，也可以看作是其的应用。

Poisson Random variable Poisson(λ)

Geometric Random variable 超几何分布

反过来，可以用随机变量的结果来代表事件。

随机变量由其核心性质可分为离散型核心变量 (无穷或又穷，可列) 和连续性随机变量。

p.m.f: 相当于根据随机变量的结果在概率空间上的划分。

c.d.f(cumulative dtribution function): 一个非降函数，离散型随机变量会有跳跃点，右连续和左极限。大概是反映了一种顺序吧，类似一种格的结构。

一些离散型随机变量的分布，如泊松分布，伯努利分布等。

3 Continuous Random Varaibles

连续性随机变量：不可列的无穷变量。

类似 p.m.f 可以在连续性随机变量定义 **p.d.f**(概率密度)。(**更好的想法是从 c.d.f 上定义出来，即要求 $F(x)$ 可微，若不可微则为第三类随机变量**)，

而 c.d.f 则可以直接套用在连续性随机变量上，名称依然为 c.d.f 没有改变

p.d.f 在某些点可以取不同的值。这只能从 c.d.f 这方面来理解。

$p(a < x \leq b) = F(b) - F(a)$, 该公式对于所有类型的随机变量均有效。且应该注意左开右闭，对于连续性随机变量则无影响。且 $a > b$ 时右边为零而非反过来

The Exponential Distribution

The Gammma Distribution : 可以看作 Exponential Distribution 的一种更近一步的推广。

Normal Random Variabls: $X \sim N(0,1)$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, 其 c.d.f 又记为

$$\Phi(x) = \int_{-\infty}^{x} f(x)\, dx$$

由于 $f(x)$ 关于 y 轴的对称性，$\Phi(x)$ 有性质：

$$\Phi(x) = 1 - \Phi(-x)$$

其中，可以引出 3σ **原则**

General Normal Random Variables:

$$X \sim N\left(\mu, \sigma^2\right)$$

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

令

$$Y = aX + b$$

则

$$Y \sim N\left(a\mu + b, a^2\sigma^2\right)$$

令

$$Y = \frac{X-\mu}{\sigma}$$

then

$$Y \sim N(0,1)$$

function of a random variable:$R \to R$, 例如：

$$z = x^k, then\, Z(\omega) = X^k\, is\, a\, r.v$$

question if $X \sim N(0,1)$,then what is the distribution of X^2?
Solution : Let $Y = X^2$ and assume the c.d.f of Y is $F_Y(y)$. **answer** :

$$f_Y(y) = \begin{cases} \dfrac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}, \, y > 0 \\ \\ 0, \, y \leq 0 \end{cases}$$

a useful general result:Suppose that $g(x)$ is a strictly monotone differentiable function of x.Then the random variable Y denoted by Y=$g(x)$ has a p.d.f. given by

$$f_Y(y) = \begin{cases} f_X\left[g^-1(y)\right] \cdot \left[\dfrac{d}{dy} g^-1(y)\right], \, if\, y = g(x) \\ \\ 0, \, if\, there\, is\, no\, y = g(x) \end{cases}$$

the cauthy distribution:μ is a real number,if its p.d.f. $f_Y(y)$ is given by:

$$f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{\left[1 + (y-\mu)^2\right]}, (-\infty < y < \infty)$$

when $\mu = 0$ is the standard cauthy distribution.

4 Expected Values

4.1 The Expected Value of a Discrete Random Variable

Definition of expexted $E(X)$，表示随机变量的对于概率的加权平均。

$$E(X) = \sum_i p_i \cdot X_i$$

其中若上式为无穷级数，则要求其绝对收敛，否则无意义 (黎曼重排定理)

Binomial Distribution if $X \sim B(n,p)$,then

$$E(X) = np$$

Poisson Distribution if $X \sim Possion(\lambda)$,then

$$E(X) = \lambda$$

The Geometric Distribution $E(X) = \frac{1}{p}$, p 代表一次成功的概率。

notes $E(X)$ does no exist

$$p.m.f\ p_n = \frac{1}{n(n+1)}$$

4.2 The Expected Value of a Continuous Random Variable

Definition of expexted 类比离散型随机变量，

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Standard Normal Distribution 由对称性和**绝对收敛**可知，Standard Normal Distribution 的 $E(X) = 0$

Gamma Distribution 利用 Γ 函数的性质可以演算出：

$$E\left(X\right)=\frac{\alpha}{\lambda}$$

The Cauthy Distribution The cauthy distribution is

$$f\left(x\right)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$$

but

$$\int_{-\infty}^{\infty}\left|x\right|f\left(x\right)\,dx$$

does not convert,so its $E\left(X\right)$ does not exist.

Remark

4.3 Expectation of Function of Random Variables

Problem Suppose X is a r.v. ,the $Y=g\left(X\right)$,how to find the $E\left(Y\right)$?

Notes

- 为什么 $E\left(Y\right)\neq g\left(E\left(X\right)\right)$
- 其中的意义如何理解？
- n 阶矩的意义？

Examples Random Variable Random Variable X:-1,0,1 with pmf:P(X=-1)=0.2 , P(X=0)=0.5 ,P(X=1)=0.3
then

$$E\left(X^2\right)=0.5$$

4.4 二阶矩

Notes

- $E\left(X^2\right)-\left(E\left(X\right)\right)^2$ 是方差 (如何证明？)
- 矩是研究随机变量的一种工具，类似于物理上的矩。

Question

- 如何理解 $E(X^2)$ 的意义？为什么被赋予二阶矩的意义？
- 是否有更快的计算方法？
- 对于连续性变量的方差如何定义？
- 如何理解 $E(X-Y)=E(X)-E(Y)$ 的意义？

5 Joint Distribution

5.1 离散型随机变量

Joint p.m.f 假设 X 和 Y 是定义在同一个样本空间上的离散型随机变量，则

$$p(x_i, y_j) = P(X = x_i, Y = y_i)$$

(其中涉及交运算)，称为 X 和 Y 的 the joint probability mass function ，简写为 Joint p.m.f

Marginal p.m.f $p_X(\cdot)$ 称为 X 和 Y 的 the joint probability mass function 的 X 的边缘分布

M p.m.f 和 Joint p.m.f 的关系 显然 Joint p.m.f 可以决定 M p.m.f, 但是如何决定？而已知 M p.m.f，在什么情况下可以决定 Joint p.m.f?
第一问利用全概率公式即可，类似一种降维的思想
而第二问则需要引入**独立性随机变量**的概念。因为需要知道 X 和 Y 的关系。

独立性随机变量 若对于所有的 x_i 和 y_i 有：

$$P(x_i, y_j) = P(x_i) \cdot P(y_j)$$

, 联想独立事件，可以看作一个独立事件族。

Joint Cumulative Probability Distribution Function 简写为：joint c.d.f，记作 $F_{(X,Y)}(x,y)$, , 有二维递增性和概率连续性：

$$\lim_{x\to-\infty} F(x,y) = 0, \forall y \in R$$

Marginal Cumulative Probability Distribution Function

两者 c.d.f 的关系 类似 c.d.f，但是计算方法有所改变。

5.2 连续性随机变量

类似单变量随机变量，先建立 Joint c.d.f, 它有二维递增性和积分为 1，以及区间性。再由偏导可以进一步定义 Joint p.d.f (Joint Probability Density Function) 以及 M p.d.f

Remark: 独立性和相关系数的关系，相关系数为 0 未必独立，而独立相关系数一定为 0.

5.3 多随机变量的函数的期望

多随机变量的函数，例如 $Z = X + Y$, 是一个复合函数，也是一个 r.v., 仍然以事件为因变量。以正向的思维来看，这个是简单的

5.4 期望的性质

1. $E(C) = C$,**C 为一常数**, 你的初始理解有错误
2. $E(aX) = a \cdot E(X)$,a 为常数，X 为随机变量
3. $E(X \cdot Y) \neq E(X) \cdot E(Y)$, 并不显然
4. $E(X + Y) = E(X) + E(Y)$, 要假设其存在 p.d.f, 然后由积分可证
5. $E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$, 是 2 和 4 的推论，代表了期望的线性性质。

6. $E\left(\sum_{n=1}^{\infty} Y_n\right) = \sum_{n=1}^{\infty} E(Y_n)$, 其中要保证无穷项符号要一致
   控制收敛定理：若 $\sum_{n=1}^{\infty} E(|Y_n|)$ 有限，则前式也成立

7. 若 X 和 Y 两者独立（上面的定理不需要这个条件），那么 $E(XY) = E(X) \cdot E(Y)$, 证明过程中只需要利用 $f(x,y) = f_X(x) \cdot f_Y(y)$ 套积分公式即可
   进一步推广，
   $$E(g(X) \cdot h(Y)) = E(g(X)) \cdot E(h(Y))$$

Remark 数学归纳法，只对有限数有用，不能过渡到无穷。(超限归纳法)

5.5 方差

因为 $E(X)$ 无法反映随机变量的所有性质，无法反映其“spread”
$E|X - E(X)|$ 可以用来测定“spread”, 但不好计算
为了在数学上更好的处理，故用 $E\left((X - E(X))^2\right)$ 来定义方差

$$Var(X) = E\left(X^2\right) - (E(X))^2$$

这个公式的证明体现了这个体系的优越性. 均方差为方差的开方。

方差的性质

1. 方差大于零

2. $Var(C) = 0$，C 为常数

3. $Var(a \cdot X) = a^2 \cdot Var(X)$, 不是线性性质，但利用了期望的线性性质

4. 如果 X 和 Y 互相独立，有：
   $$Var(X+Y) = Var(X) + Var(Y)$$
   反映了方差的线性性质，这里利用了期望的第三条性质，虽然形式上跟第二条性质相似

5. a,b 为常数，则
   $$Var(a+bX) = b^2 \cdot Var(X)$$

5.6 协方差

Covariance:

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

它的性质蛮多的。不过均是可以由条件期望的三条性质推导出来的。它可以取任意实数.，代表一种相关性

可以引出一个 Varaince Formula, 蛮有用的

5.7 相关系数

由于 Covariance 可能太大，故引进 Correlation Cofficient: $\rho\left(X,Y\right)$

其中两个方差严格大于零.

可以构造证明

$$-1\leqq\rho\leqq1$$

如果 $\rho=1$, 则两者呈现正线性关系。

不过对于这个线性关系怎样理解？

另外，可以与线性代数来使 n 个随机变量的 Covariance 形式更漂亮

5.8 Moment Generating Function

矩发生函数,MGF

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

是关于 t 的实函数。

显然，

$$M_X\left(t\right)=E\left(e^{tX}\right)=\int_{-\infty}^{\infty}e^{tx}f\left(x\right)dx$$

由于绝对收敛性，所以可能对某些 t 不成立

矩发生函数的性质：

1. $M_X\left(0\right)=1$
2. If $Y=aX+b$,$M_Y\left(t\right)=e^{bt}M_X\left(at\right)$, 利用期望的线性性质即可。

3. If Z=X+Y ,and X,Y are independent then $M_Z(t) = M_X(t) \cdot M_Y(t)$, 可拓展至 n 个独立随机变量

4. mgf 与其随机变量一一对应。拉氏变换的唯一性

Remark 你对几何分布的认识有问题。
矩发生函数的应用：

1. 用来求随机变量的矩：

$$M_X'(t) = \frac{d}{dt}E\left(e^{tX}\right) = E\left(\frac{d}{dt}e^{tX}\right) = E\left(e^{tX}X\right)$$

则：

$$E\left(X^n\right) = M_X^{(n)}(0)$$

称为原点矩，另外还有均值矩 $E\left[(X - E(X))^n\right]$, **也是可以求出来的**

2. 用来决定分布利用性质三和性质四可以求组合独立随机变量的分布

Remark $Y \sim \Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$, 其中 $Y \sim \chi^2(n)$

6 条件概率

6.1 引入概念

首先从离散型随机变量引入。并且与 Joint p.m.f 和 Marginal p.m.f

Conditional p.m.f 是 x 的函数，y 在其中起参数的作用，

$$p_{X|Y}(x|y)$$

Conditional c.d.f 由 Conditional p.m.f 求和而得，仍然是 x 的函数

$$F_{X|Y}(x|y)$$

条件期望

$$E(X|Y = y) = \sum_x x \cdot p_{X|Y}(x|y)$$

其中要求 $p_Y y > 0$, 这也是难以向连续性随机变量拓展的困难之处。

an example 如果 X 和 Y 是独立的泊松分布，求

$$E\left(X|X+Y=n\right)$$

answer is $n\dfrac{\lambda_1}{\lambda_2+\lambda_1}$

连续性随机变量的条件期望的引入 取极限就是

$$f_{X|Y}\left(x|y\right)=\frac{f\left(x,y\right)}{f_Y\left(y\right)}\quad f_Y\left(y\right)>0$$

$$E\left(X|Y=y\right)=\int_{-\infty}^{\infty}x\cdot f_{X|Y}\left(x|y\right)/,dx$$

是 y 的函数