条件期望

FXY 2022.10.3

Notations

m : Lebesgue measure

P : Probability measure

$(\Omega, \mathcal{F}, \nu)$: Measure space

Radon-Nikodym导数

好不容易把这导数的名字打了一遍，足见重视!

我们都知道非负Borel函数可以积分:

$$\lambda(A) = \int_A f d\nu \ , \ A \in \mathcal{F}$$

这样λ就是一个集合函数，从$\mathcal{F}$到R，而且是$(\Omega, \mathcal{F})$上的一个测度。顺理成章地，如果$\lambda(\Omega) = 1$，λ就是一个概率测度。很显然有$\nu(A) = 0 \Rightarrow \lambda(A) = 0$. 所以形象地看$\lambda$要比$\nu$更小一点，记作$\lambda << \nu$，学名叫绝对连续。

这是从f, ν生成测度，好像没什么特别的。恐怖的事情是下面这个:

> λ, ν是$(\Omega, \mathcal{F})$上的测度，且ν σ有限，$\lambda << \nu$，则这样的非负f一定存在而且a.e.唯一，记为$d\lambda / d\nu$。

这太恐怖了朋友们！因为它对λ, ν就没什么严格的要求，这样花活就可以很多！如果λ是一个概率测度（ν无所谓），那f称为关于ν的pdf（概率密度函数）；对于一个随机变量X，它的Distribution的定义是$P_X = P \circ X^{-1}$，是一个$(R, \mathcal{B})$上的概率测度，再从$(R, \mathcal{B})$上随便捞一个测度ν，如果$dP_X/d\nu$存在，则称为X关于ν的pdf.

写刚才这段的时候差点砸电脑了，我肯定不适合学概率！

这就很恐怖，你想，比如一个随机变量，在这个测度下没有pdf，没关系，我换个测度说不定又有pdf了，而且导数我只知道存在，它可没告诉我要怎么算....咋办？所以我们就两眼一闭歇菜就好。

什么是连续型随机变量？初等概率论里说存在f使得积分是cdf F之类的。什么积分？嘴上说的是Lebesgue，心里想的大概是黎曼；cdf是怎么定义的？大概是$F(x) = P(X \leq x)$；P是什么？大概是一个概率。我们现在可以来严格定义一下什么是连续型随机变量。首先有一个概率空间$(\Omega, \mathcal{F}, P)$，有个可测函数X，叫做随机变量，它诱导一个$(R, \mathcal{B})$的概率测度P_X：$P_X(B) = P(\{\omega : X(\omega) \in B\}), \forall B \in \mathcal{B}$. 仔细看看，好像没问题。重点来了：如果$dP_X/dm$存在，则$X$是连续型的，且这个导数就是pdf.

> X的cdf定义为：$F(x) = P_X((-\infty, x])$，如果在微积分意义上（我太喜欢这几个字眼了！）有
>
> $$F(x) = \int_{-\infty}^{x} f(t)dt$$
>
> 则这个f就是X关于m的pdf，进而是连续型的。

本节就到这里，已经写晕了，不知道是不是对的....

条件期望

休息一下，不要太着急。

先确认一个事情：P是$(\Omega, \mathcal{F})$上的概率测度，$\mathcal{A} \subset \mathcal{F}$是$\Omega$上的$\sigma$域，则$P$也是$(\Omega, \mathcal{A})$上的概率测度。

好了，开始定义条件期望：设X是$(\Omega,\mathcal{F},P)$上可积随机变量，则$\lambda(A)=\int_A fdP, A\in\mathcal{F}$是$(\Omega,\mathcal{F})$上一个概率测度，当然也是$(\Omega,\mathcal{A})$上的概率测度。同样地，$P$也是$(\Omega,\mathcal{A})$上的概率测度。有了两个概率测度，就可以玩$d\lambda/dP$了，这个a.s. (with respect to P) 唯一的，$(\Omega,\mathcal{A})$上的Borel函数（因而是随机变量）就是$E(X|\mathcal{A})$. 很清晰是吧！

所以条件期望满足两个条件（也是一般书上的定义）：

(i) $E(X|\mathcal{A})$是$(\Omega,\mathcal{A})$到$(R,\mathcal{B})$的可测函数；

(ii) $\forall\ A\in\mathcal{A},\ \int_A E(X|\mathcal{A})dP=\int_A XdP$

仔细看看，和上面那段文字说的事情是一样的。

顺理成章地，有条件概率：$P(B|\mathcal{A})=E(I_B|\mathcal{A}), B\in\mathcal{F}$；设$Y$是$(\Omega,\mathcal{F},P)$到$(\Lambda,\mathcal{G})$上的可测函数，则$E(X|Y)=E(X|\sigma(Y))=h\circ Y$，其中$h$是一个定义在$Y$的值域上的可测函数（管他呢，反之就是个函数就行了）。

好端端的搞个h干嘛？可能是为了定义这个：$E(X|Y=y)=h(y)$.

两个技术性证明

好了，你总得说明一下这么抽象的定义和初等概率论里是一样的吧。虽然只是两个小证明，但还是麻烦死了，一切都要怪罪到1933年！

例1: X是$(\Omega,\mathcal{F},P)$上可积随机变量，$A_i\in\mathcal{F}$是Ω的可列的，两两不交的，分割，且概率都大于0. a_i是一列不同的实数，定义

$$Y=\sum_{i=1}^{\infty}a_iI_{A_i}$$

则

$$E(X|Y)=\sum_{i=1}^{\infty}\frac{\int_{A_i}XdP}{P(A_i)}I_{A_i}\ a.s.$$

无力吐槽，反之我们直接验证。先得把$\sigma(Y)$搞出来！考虑$Y^{-1}(B)$，如果你觉得某个A_1里的元素x可以进去，那似乎A_1里的所有元素都可以进去，这样看来诸A们只能打包出售了，所以$\sigma(Y)=\sigma(\{A_1,\ldots,\})$. 看右侧的一堆东西，$\frac{\int_{A_i}XdP}{P(A_i)}$似乎是常数哦，所以也是一堆$I_{A_i}$求和，所以右边关于$(\Omega,\sigma(Y))$可测！好，开始验证积分，对于任意$B\in\mathcal{B},\ Y^{-1}(B)=\cup_{i:\,a_i\in B}A_i$（因为$A_i$两两不交，仔细看看），

$$\begin{aligned}\int_{Y^{-1}(B)}XdP&=\sum_{i:\,a_i\in B}\int_{A_i}XdP\\&=\sum_{i=1}^{\infty}\frac{P(A_i\cap Y^{-1}(B))}{P(A_i)}\int_{A_i}XdP\\&=\sum_{i=1}^{\infty}\frac{\int_{A_i}XdP}{P(A_i)}\int_{Y^{-1}(B)}I(A_i)dP\\&=\int_{Y^{-1}(B)}\sum_{i=1}^{\infty}\frac{\int_{A_i}XdP}{P(A_i)}I(A_i)dP\end{aligned}$$

这就证明了

$$E(X|Y)=\sum_{i=1}^{\infty}\frac{\int_{A_i}XdP}{P(A_i)}I_{A_i}$$

设$A\in\mathcal{F},X=I_A$，则

$$P(A|Y)(\omega)=E(X|Y)(\omega)=\sum_{i=1}^{\infty}\frac{P(A\cap A_i)}{P(A_i)}I_{A_i}(\omega)$$

如果$\omega\in A_i$，则函数值为$\frac{P(A\cap A_i)}{P(A_i)}$.

例2：X是$(R^n,\mathcal{B}^n,\nu)$上的随机向量，Y是$(R^m,\mathcal{B}^m,\lambda)$上的随机向量，这两个测度都$\sigma$有限，$(X,Y)$有关于$\nu\times\lambda$的pdf $f(x,y)$，$g(x,y)$是R^{n+m}上的Borel函数且$E|g(X,Y)|<\infty$，那么

$$E(g(X,Y)|Y)=\frac{\int g(x,Y)f(x,Y)d\nu(x)}{\int f(x,Y)d\nu(x)}\ a.s.$$

虽然我不知道它是什么，但反正看着像是$\int g(x,Y)\frac{f(x,Y)}{\int f(x,Y)d\nu(x)}d\nu(x)$，还挺靠谱。

证明：

右边是Y的函数，记为$h(Y)$，由Fubini，它是$(\Omega,\sigma(Y))$上的可测函数（反之我也不会细看，Fubini是块砖哪里需要往哪搬）；还是Fubini，$f_Y(y)=\int f(x,y)d\nu(x)$是$Y$关于$\lambda$的pdf（反之和初等概率论一样的东西就看得很开心，人生苦短我搬Fubini），任意$B\in\mathcal{B}^m$，

$$
\begin{aligned}
\int_{Y^{-1}(B)} h(Y)dP &= \int_B h(y)dP_Y \text{ (万恶的积分变换!)} \\
&= \int_B \frac{\int g(x,y)f(x,y)d\nu(x)}{\int f(x,Y)d\nu(x)} f_Y(y)d\lambda(y) \\
&= \int_B \int g(x,y)f(x,y)d\nu(x)d\lambda(y) \\
&= \int_{B\times R^n} g(x,y)dP_{(x,y)} = \int_{Y^{-1}(B)} g(X,Y)dP
\end{aligned}
$$

好了，我已经晕了，反之就是连续情况下初等概率论算条件期望的公式还能用，然后条件分布也还是一样的，这样就行了！

条件期望的性质

X,Y,X_i都在$(\Omega,\mathcal{F},P)$上可积，$\mathcal{A}$是$\mathcal{F}$的子σ域。

(i) If $X=c\ a.s.$, then $E(X|\mathcal{A})=c\ a.s.$

(ii) $E(E(X|\mathcal{A}))=EX$

(iii) $E(E(X|\mathcal{A})|\mathcal{A}_0)=E(X|\mathcal{A}_0)$，$\mathcal{A}_0$是$\mathcal{A}$的子$\sigma$域

(iv) $\sigma(Y)\subset\mathcal{A},\ E|XY|<\infty$, then $E(XY|\mathcal{A})=YE(X|\mathcal{A})$

(v) Fatou, Holder, Minkowski, 控制收敛等都成立

> 条件期望是一个随机变量，所以取liminf等都是函数列的收敛

独立性

只有Y_2与X和Y_1都独立时，才有$E(X|Y_1,Y_2)=E(X|Y_1)\ a.s.$成立。

学了还是不会，大眼瞪小眼

发现没有，关于Radon-Nikodym导数和条件期望，我们都是在验证而没有直接计算。所以当我平时看论文看到下面这玩意的时候，还是只能大眼瞪小眼！

Lemma 3. *Let $\Theta \in \mathbb{R}^{n\times n}$ such that $\Theta_{ij}=\rho$ for all $1\le i,j\le n$, where $0<\rho<1/2$. Let A be an adjacency matrix of an inhomogeneous Bernoulli network with independent edges such that $\mathbb{E}(A)=\Theta$. For any $v_b, v_c \in [-\sqrt{\rho},\sqrt{\rho}]^n$, let B and C be adjacency matrices of inhomogeneous Bernoulli networks with independent edges such that $\mathbb{E}(B)=v_bv_b^\top+\Theta$ and $\mathbb{E}(C)=v_cv_c^\top+\Theta$. Let P_A, P_B, P_C be the distributions of A, B and C. Then*

$$\mathbb{E}_{P_A}\left(\frac{dP_B}{dP_A}\frac{dP_C}{dP_A}\right)\le \exp\left(\frac{(v_b^\top v_c)^2}{\rho(1-\rho)}\right).$$

Let $A'=A-\mathrm{diag}(A)$, $B'=B-\mathrm{diag}(B)$ and $C'=C-\mathrm{diag}(C)$. Then

$$\mathbb{E}_{P_{A'}}\left(\frac{dP_{B'}}{dP_{A'}}\frac{dP_{C'}}{dP_{A'}}\right)\le \exp\left(\frac{(v_b^\top v_c)^2}{\rho(1-\rho)}\right).$$

致谢和参考教材

感谢我亲爱的导师，也碰巧为本学期的数理统计授课，是他灌输的自由散漫让我获益无穷，也是他仁慈的作业量让我有空在这边码些无用的东西！

参考教材：数理统计（Shaojun）