第三章
第一节

给定概率空间(Ω, ***F***, P)，设 X 是从 Ω 到实数 R 的映照，对任何 a<b，区间（a, b]的逆像都属于 ***F***，则称 X 是[定义在概率空间(Ω, ***F***, P)]上的]随机变量。

这与课本上定义 3.1.1 是一致的，但回避了“博雷尔”这几个字。熟知实数 R 自带一个 σ 域，就是开集全体生成的最小 σ 域，叫做 Borel 域 B，因此映照 X 诱导了 B 到 F 的映照，所谓 X 可测就是 B 中任何一个元素(Borel 集)的逆像都属于 ***F***。学完实变函数你会发现要找一个不可测函数还是挺难的，我们平常打交道的函数，你能随口说出的函数，都是可测函数。所以暂时可以不管“可测”这两个字。

给定概率空间(Ω, F, P）和随机变量 X，对于任意实数 x，{ω, X(ω) ≦ x}属于 ***F***，定义 F(x) = P({ω, X(ω) ≦ x})，称函数 F 为随机变量 X 的分布函数。

请注意这里的定义与课本上定义 3.1.2 相差一个等号。因为经常打交道，不会引起歧义的话可以写的省略一些，F(x) = P(ω, X(ω)≦x) = P(X≦x),

分布函数 F 具有如下性质：1.单调上升，2.左端趋向 0，右端趋向 1，3. 左极右连。
严格的证明用到了概率公理假设中的下连续性（1.5.14)。由于我们修改了定义，课本上是左极右连。思考题：如果一个函数满足上述三点，一定是某个随机变量的分布函数吗？

我们根据分布函数对随机变量进行分类：
如果 X 只取可列个数值，分布函数是阶梯形的，称 X 是离散型的，
如果分布函数是可微的，则称 X 是连续型的，此时分布函数 F 的导数 f 叫做 X 的（概率）密度函数（pdf）
当然还有分布函数既不是阶梯形的也不是可微分的，此时称 X 是奇异型的。

显然古典概型诱导的随机变量是离散型的，前面学到的二项分布，伯努利分布，泊松分布，几何分布，帕斯卡分布等等，都是离散型的。
这里有一个小问题，前面在讨论几何概型时，用到了不同概率空间，而这时候不能同时用多个概率空间，更重要的我们希望摆脱实际束缚，找到更多例子，这就要回答前面的思考题。

给定单调函数 F，左极右连。取 Ω=R，F=Borel 域，X 是从 R 到 R 的恒同映照，定义 P((a, b])=F(b)-F(a)，然后把 P 的定义延拓到 Borel 域，这样我们构造了一个概率空间及其上的随机变量 X，其分布函数就是 F.

这个回答很重要，这说明概率空间并不是本质的，可以随时造一个。这样造出来的概率空间叫做典范的(canonical)，在典范概率空间，随机变量和分布函数一一对应，我们感到舒服多了，因此有时候也混为一谈，譬如离散分布实指离散型随机变量的分布函数。

而且任何一个实轴上的非负函数 f，如果积分为 1，则 f 就是某个随机变量的概率密度函数。于是我们就有很多连续型随机变量的例子：

均匀分布 U(a, b)，对应几何概型
正态分布 $N(\mu, \sigma^2)$，又叫做高斯分布，$\mu=0, \sigma=1$ 时， N(0,1)叫做标准正态分布
指数分布 $Exp(\lambda)$， 与离散型的几何分布很像，都具有无记忆性
Gamma 分布 $\Gamma(r, \lambda)$，只要能记得 Γ-积分

大家应当要记得这几个分布的密度函数，没有什么道理，就是我能记得，所以要求大家记得。

在所有分布中，正态分布是最重要的，首先在很多场合出现，譬如误差的分布，其次它具有非常优美的性质，不同的 $N(\mu, \sigma^2)$通过线性变换可以转化为 N(0,1)，因此一张正态分布表就可以了（P133L-3 有误），最重要的它是各种收敛序列的极限分布。

指数分布的无记忆性很重要，今后讲马氏过程，马尔可夫性质，就依赖这个无记忆性。而且指数分布是唯一具有无记忆性的连续型随机变量。这句话里我们不仅把分布和随机变量混为一谈，而且用到了典范概率空间。定义在两个不同概率空间上的两个随机变量，即使它们的分布函数相同，按照定义，它们是不同的。但从使用者角度来看，看不见概率空间，只能感受到分布函数，因此谈论唯一性的时候，限用典范概率空间。

以后我们还可以构造另一个概率空间，其中 Ω=[0,1]，与现在所说的典范概率空间是等价的，也是很常见的。

第二节 随机向量
随机向量是概率空间到 n 维欧氏空间的可测映照，所谓可测，就是要求欧式空间的任意开集的逆像都属于概率空间的 σ 域 ***F***，如前所述我们还真很难碰到不可测的映照。

显然随机变量是一维随机向量，也有人把随机变量和随机向量混为一谈。试想若把 n 维欧氏空间换成群，从概率空间到群的可测映照，又叫什么？

随机向量还可以看成是定义在同一概率空间的 n 个随机变量。

随机向量也分为离散型和连续型（还有奇异型），对于离散型随机向量，列表是最简单的，即 P(X=x_i) = p_i
对于连续型随机变量，最常用是其概率密度函数，首先从分布函数说起，为简单计，这里只拿二维随机向量(X,Y)举例，推广到 n 维并非难事。令
F(x,y) = P(X≦x, Y≦y)，注意我们这里修改了等号，对于连续型，不用太在意边界，随机向量落在边界的概率为零。p(x,y) 是 F(x,y) 关于 x 和 y 的二阶偏导数

则 P(X≦x, Y≦y)=F(x,y) =$\iint_{-\infty}^{x,y} p(u,v)dudv$

这一等式可以推广到更一般情形

$$\mathrm{P}\big((\mathrm{X},\mathrm{Y}) \in \mathrm{A}\big) = \iint_A p(u,v)dudv$$

容易验证这等式对于矩形成立，然后可以用若干不相交矩形的并来逼近区域，再次利用概率的连续性。由于这最后一条等式，对于连续型随机变量，通常只用密度函数，很少用分布函数。

任何非负函数，只要积分有限，经过归一化，就可以是一个密度函数，但真正需要记住的只有有限区域上的均匀分布，和高维正态分布(3.2.22)和(4.6.1)，

随机向量的分布函数包含很多信息，可以诱导出边缘（边际）分布，例如

$$F_X(x) = F(x, \infty) = P(X \le x) = P(\{\omega: \ X(\omega) \le x\})$$

对于离散型，就是适当求和。对于连续型，给定密度函数，可以通过积分就出边缘分布的边缘密度函数

$$p_X(\mathrm{x}) = \int_{-\infty}^{\infty} p(x,y)dy$$

对于 3 维情形，还可以有 2 维的边际分布，n 维的分布函数可以诱导出 2^n-2 个边际分布，特别是高维正态分布的边际分布仍是正态的。这些边际分布之间还有相容性，即边际分布的边际分布还是边际分布。但这些边际分布并不能决定原来的分布函数。给定若干（相容）的分布函数，构造一个高维分布函数，以这些分布函数为边际分布？这是一个问题，耦合技术就是来解决这个问题的。

条件分布，对于离散情形，在给定事件{Y=y}= { ω: Y(ω)=y}条件下，P(X≦x | Y=y) = P(X≦x, Y=y) / P(Y=y)，其实更常用的是条件概率 P(X=x | Y=y）= P(X=x, Y=y) / P(Y=y).

对于连续情形，事件{Y=y}是零概率，但零概率事件是会发生的，所以要考察

$$\mathrm{P(X \leqq x \mid Y=y)} = \lim_{\epsilon\to 0} \mathrm{P(X \leqq x \ \ |y-\epsilon \le Y \le y+\epsilon)}$$

这里的等号=实为定义。

$$\mathrm{P(X \leqq x \ \ |y-\epsilon \le Y \le y+\epsilon)} = \frac{\mathrm{P(X \le x, y-\epsilon \le Y \le y+\epsilon)}}{\mathrm{P(y-\epsilon \le Y \le y+\epsilon)}} = \frac{\int_{-\infty}^{x}\int_{y-\epsilon}^{y+\epsilon} p(u,v)dudv}{\int_{-\infty}^{\infty}\int_{y-\epsilon}^{y+\epsilon} p(u,v)dudv}$$

$$\to \frac{\int_{-\infty}^{x} p(u,y)du}{\int_{-\infty}^{\infty} p(u,y)du} = \int_{-\infty}^{x} \frac{p(u,y)}{\int_{-\infty}^{\infty} p(w,y)dw} du$$

因此把 $p(x,y)/\int_{-\infty}^{\infty} p(u,y)du = p(x,y)/p_Y(\mathrm{y})$ 作为事件{Y=y}的条件概率密度函数.

作为微积分的练习题，正态分布的条件分布仍是正态的。看看图 3.2.3 就可以

两个随机变量 X 和 Y，如果对于任何实数 x 和 y，事件$\{\omega: \ X(\omega) \le x\}$和事件$\{\omega: Y(\omega) \le y\}$都独立，等价的是，$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$，则称随机变量 X 和 Y 相互独立。

若把 X 和 Y 看成是随机向量的两个分量，则上述条件可以写为$F(x,y) = \ F_X(x) \ F_Y(y)$，在独立性假设下，边际分布可以决定联合分布。

独立性条件还可以推出更强的形式，对任意可测集 A 和 B，

$$P(X \in A, \quad Y \in B) = P(X \in A)P(Y \in B)$$

学过实变函数或测度论就知道，这样的推广并不困难，使用起来更方便。

不难把两个随机变量的独立性推广到多个随机变量，对任意 x，y,………z

$$P(X \le x, Y \le y, \dots\dots, Z \le z) = P(X \le x)P(Y \le y) \dots\dots P(Z \le z)$$

则称随机变量 X,Y, … Z 相互独立。这里 x，y,………z 可取任意实数，也可以是∞，正因为

如此，这一条等式已经与 2^n 条等式相当。