概率理论基础

随机变量

概率函数

给定一个由任意类型元素构成的集合 E 到 E' 的映射，例如定义在 E 上的单值函数 $u(\xi)$，其值属于 E'。 对于 E' 的每一个子集 A'，我们将与之对应的 E 中的原像（pre-image）记为 $u^{-1}(A')$，它包含所有映射到 A' 中元素的 E 中的元素。令 $\mathcal{E}^{(u)}$ 为E' 的所有子集 A' 组成的系统，其中 A' 的原像属于域 $\mathcal{E}$，则 $\mathcal{E}^{(u)}$ 也是一个域。如果 $\mathcal{E}$ 恰巧是一个波莱尔域，则 $\mathcal{E}^{(u)}$ 也是一个波莱尔域。由此可令：

$$\mathrm{P}^{(u)}(A') = \mathrm{P}\{u^{-1}(A')\} \tag{1}$$

因为定义在 $\mathcal{E}^{(u)}$ 上的集合函数 $\mathrm{P}^{(u)}$ 满足公理 I-VI，所以它可以作为 $\mathcal{E}^{(u)}$ 上的概率函数。在证明以上内容之前，我们需要先给出下面的定义：

定义：给定一个随机事件 ξ 的单值函数 $u(\xi)$，公式 (1) 中定义的函数 $\mathrm{P}^{(u)}(A')$ 称为 u 的概率函数。

备注 1：在研究概率域 $(\mathcal{E}, \mathrm{P})$ 时，我们只简单地把 $\mathrm{P}(A)$ 称为概率函数。但是我们称 $\mathrm{P}^{(u)}(A')$ 为u 的概率函数。当 $u(\xi)=\xi$ 时，$\mathrm{P}^{(u)}(A')$ 恰巧与 $\mathrm{P}(A)$ 相等。

备注 2：事件 $u^{-1}(A')$ 表明 $u(\xi)$ 属于 A'，因此 $\mathrm{P}^{(u)}(A')$ 表示 $u(\xi)$ 属于 A' 的概率。

我们仍需证明以上提到的 $\mathcal{E}^{(u)}$ 和 $\mathrm{P}^{(u)}$ 的性质，不过它们实际上来源于一个简单的事实：

引理：原像集合 $u^{-1}(A')$ 的和、乘积、差，等于对应原始集合 A' 的和、乘积和差的原像。

以上引理的证明过程留给读者。

【补充证明过程】

令 A' 和 B' 为域 $\mathcal{E}^{(u)}$中 的两个集合，它们的原像 A 和 B 则属于 $\mathcal{E}$。由于 $\mathcal{E}$ 构成一个域，则集合 AB、$A+B$ 、$A-B$ 也属于 $\mathcal{E}$；同时这三个集合也分别是集合 $A'B'$、$A'+B'$、$A'-B'$ 的原像（后三者属于域 $\mathcal{E}^{(u)}$）。由此证明 $\mathcal{E}^{(u)}$ 是一个域。同样地，可以证明如果 $\mathcal{E}$ 是一个波莱尔域，则 $\mathcal{E}^{(u)}$ 也是。进一步可得

$$\mathrm{P}^{(u)}(E') = \mathrm{P}\{u^{-1}(E')\} = \mathrm{P}(E) = 1$$

显然 $\mathrm{P}^{(u)}$ 总是非负的。因此接下来只需证明，$\mathrm{P}^{(u)}$ 是完全可加的（参考第二章第 1 节结尾）。

假设所有的集合 A'_n ，以及它们各自的原像 $u^{-1}(A'_n)$ 是互斥的（对于所有的 n， A'_n 两两互斥，原像同理），则有

$$\mathrm{P}^{(u)}(\sum_n A'_n) = \mathrm{P}\{u^{-1}(\sum_n A'_n)\} = \mathrm{P}\{\sum_n u^{-1}(A'_n)\} = \sum_n \mathrm{P}\{u^{-1}(A'_n)\} = \sum_n \mathrm{P}^{(u)}(A'_n)$$

由此证明了$\mathrm{P}^{(u)}$ 的完全可加性。

最后，还有一点需要说明。令 $u_1(\xi)$ 是由 E 映射到 E' 的函数，$u_2(\xi')$ 是由 E' 映射到 E'' 的函数。则乘积函数 $u_2u_1(\xi)$ 从 E 映射到 E''。接下来研究概率函数 $\mathrm{P}^{(u_1)}(A')$ 和 $\mathrm{P}^{(u)}(A'')$，其中 $u(\xi)=u_2u_1(\xi)$。很容易得出，这两个概率函数有以下关联：

$$\mathrm{P}^{(u)}(A'') = \mathrm{P}^{(u_1)}\{u_2^{-1}(A'')\} \tag{2}$$

随机变量和分布函数的定义

定义：定义在基本集合 E 上的实单值函数 $x(\xi)$，当对于任意实数 a，满足 $x<a$ 的所有 ξ 组成的集合属于集合系统 $\mathcal{E}$，则这个实单值函数 $x(\xi)$ 称为随机变量（random variable）。

函数 $x(\xi)$ 将基本集合 E 映射到集合 R^1，即全体实数。如本章第 1 节所言，这个函数确定了一个实数集 R^1 的子集组成的域 $\mathcal{E}^{(x)}$。我们可以用以下方式重新表述随机变量的定义：

> 对于实函数 $x(\xi)$， 当且仅当域 $\mathcal{E}^{(x)}$ 包含形如 $(-\infty;a)$ 的所有区间时，$x(\xi)$ 为一个随机变量。

因为 $\mathcal{E}^{(x)}$ 是一个域，所以连同区间 $(-\infty,a)$ 一起，这个域包含了半开区间 $[a;b)$ 的所有可能的有限和。如果我们的概率域是波莱尔域，则域 $\mathcal{E}$ 和 $\mathcal{E}^{(x)}$ 也是波莱尔域。因此域 $\mathcal{E}^{(x)}$ 包含集合 R^1 的所有波莱尔集。

后面我们会用 $\mathrm{P}^{(x)}(A')$ 来表示随机变量的概率函数。它是定义在域 $\mathcal{E}^{(x)}$ 的所有集合上的。特别地，对于最重要的情形——概率的波莱尔域，$\mathrm{P}^{(x)}$ 定义在实数集 R^1 的所有波莱尔集上。

定义：函数

$$F^{(x)}(a)=\mathrm{P}^{(x)}(-\infty,a)=\mathrm{P}\{x<a\}$$

其中 $-\infty$ 和 $+\infty$ 是 a 的可取值。此时该函数称为随机变量 x 的分布函数（distribution function of the random variable x）。

由定义马上可知

$$\begin{aligned}F^{(x)}(-\infty)&=0\\ F^{(x)}(+\infty)&=1\end{aligned}\tag{1}$$

满足不等式 $a\leq x<b$ 的 x 概率由下式给出

$$\mathrm{P}\{x\subset[a;b)\}=F^{(x)}(b)-F^{(x)}(a)\tag{2}$$

由此可得，对于 $a<b$，有

$$F^{(x)}(a)\leq F^{(x)}(b)$$

上式表明，函数 $F^{(x)}(a)$ 是个非递减函数。现在假设 $a_1<a_2<\cdots<a_n<\cdots<b$，则有

$$\underset{n}{\mathscr{D}}\{x\subset[a_n;b)\}=0$$

因此，根据连续性公理可得，当 $n\rightarrow+\infty$ 时，

$$F^{(x)}(b)-F^{(x)}(a_n)=\mathrm{P}\{x\subset[a_n;b)\}$$

上式趋近于 0。由此可得 函数$F^{(x)}(b)$为左连续函数。

采用类比的方法可以证明：

$$\lim F^{(a)}=F^{(x)}(-\infty)=0,\quad a\rightarrow-\infty\tag{3}$$

$$\lim F^{(a)}=F^{(x)}(+\infty)=0,\quad a\rightarrow+\infty\tag{4}$$

如果概率域 $(\mathcal{E},\mathrm{P})$ 是波莱尔域，则对于所有实数集 R^1 上的波莱尔集合 A，概率函数 $\mathrm{P}^{(x)}(A)$ 的值由分布函数 $F^{(x)}(a)$ 唯一确定（参考第二章第 3 节中的第 III 个例子）。本文重点关注 $\mathrm{P}^{(x)}(A)$ 的值，所以分布函数在后续研究中将发挥极其重要的作用。

如果分布函数 $F^{(x)}(a)$ 可微，则其相对于参数 a 的导数（如下）称为 x 在点 a 处的概率密度（probability density）。

$$f^{(x)}(a)=\frac{d}{da}F^{(x)}(a)$$

如果对于每一个 a，都有

$$\frac{d}{da}F^{(x)}(a)=\int_{-\infty}^{a}f^{(x)}(a)da$$

则任意波莱尔集 A 的概率函数 $\mathrm{P}^{(x)}(a)$ 可以用以下形式表示

$$\mathrm{P}^{(x)}(A)=\int_A f^{(x)}(a)da\tag{5}$$

这种情况下我们称 x 的分布是连续的。上式可以写为另一种更一般的形式

$$\mathrm{P}^{(x)}(A)=\int_A dF^{(x)}(a)\tag{6}$$

所有刚刚介绍到的概念都可以推广到条件概率的情形。

集合函数

$$\mathrm{P}_B^{(x)}(A)=\mathrm{P}_B\{x\subset A\}$$

表示在假设 B 下 x 的条件概率。非递减函数

$$F_B^{(x)}(a)=\mathrm{P}_B\{x<a\}$$

为对应的分布函数。并且，当 $F_B^{(x)}(a)$ 可微时，

$$f_B^{(x)}(a)=\frac{d}{da}F_B^{(x)}(a)$$

是在假设 B 的条件下， x 在 a 点的条件概率密度。

多维分布函数

现在给定 n 个随机变量 $X_1,X_2,\ldots,X_n$。n 维空间 R^n 中的点 $x=(X_1,X_2,\ldots,X_n)$ 是基本事件 ξ 的函数。则根据本章第一节的内容，可得到定义在空间 R^n 上的域 $\mathcal{E}^{(x_1,x_2,\ldots,x_n)}$，这个域包括了空间 R^n 的子集；也可以得到定义在 $\mathcal{E}'$ 上的概率函数 $\mathrm{P}^{(x_1,x_2,\ldots,x_n)}(A')$。这个概率函数称为随机变量 $x_1,x_2,\ldots,x_n$ 的 n 维概率函数。

对于任意选择的 i 和 a_i（$i=1,2,\ldots,n$），由随机变量的定义可以直接得出 R^n 中满足 $x_i<a_i$ 的所有的点组成的集合。因此 $\mathcal{E}'$ 也包括了以上集合的交集，例如集合 $L_{a_1a_2\ldots a_n}$ 表示 R^n 中满足所有不等式 $x_i<a_i$（$i=1,2,\ldots,n$）的点的集合[1]。

如果把 R^n 空间中满足不等式 $a_i\le x_i<b_i$ 的所有点组成的集合记为 n 维半开区间 $[a_1,a_2,\ldots,a_n;b_1,b_2,\ldots,b_n)$，则可以发现每一个这样的区间都属于域 $\mathcal{E}'$，因为

$$[a_1,a_2,\ldots,a_n;b_1,b_2,\ldots,b_n)=L_{b_1b_2\ldots b_n}-L_{a_1b_2\ldots b_n}-L_{b_1a_2\ldots b_n}-\cdots-L_{b_1b_2\ldots b_{n-1}a_n}$$

所有的 n 维半开区间系统的波莱尔扩展，包含 R^n 中的所有波莱尔集合。由此可得出，在波莱尔概率域下，域 $\mathcal{E}$ 包含了 R^n 空间的所有波莱尔集合。

【关于波莱尔扩展（Borel extension）和波莱尔集合（Borel sets）的区别】

定理：在波莱尔概率域下，每一个定义在有限个随机变量 $x_1,x_2,\ldots,x_n$ 上的波莱尔函数 $x=f(x_1,x_2,\ldots,x_3)$ 也是个随机变量。

要证明以上定理，只需证明 R^n 上满足 $x=f(x_1,x_2,\ldots,x_n)<a$ 的所有的点 $(x_1,x_2,\ldots,x_n)$ 组成的集合是波莱尔集。特别地，所有的对随机变量进行有限的求和和求乘积的操作得到的变量，也是随机变量。

定义：函数

$$F^{(x_1,x_2,\ldots,x_n)}(a_1,a_2,\ldots,a_n)=\mathrm{P}^{(x_1,x_2,\ldots,x_n)}(L_{a_1a_2\ldots a_n})$$

称为随机变量 $x_1,x_2,\ldots,x_n$ 的 n 维分布函数。

与在一维下的情形类似，我们证明证明 n 维分布函数 $F^{(x_1,x_2,\ldots,x_n)}(a_1,a_2,\ldots,a_n)$ 是一个非减函数，并且对于任意一个变量都是左连续的。类比第二节中的公式 (3) 和 (4)，可得：

$$\lim_{a_i\to-\infty}F(a_1,a_2,\ldots,a_n)=F(a_1,a_2,\ldots,a_{i-1},-\infty,a_{i+1},\ldots,a_n)=0 \tag{7}$$

$$\lim_{a_1\to+\infty,a_2\to+\infty,\ldots,a_n\to+\infty}F(a_1,a_2,\ldots,a_n)=F(+\infty,+\infty,\ldots,+\infty)=1 \tag{8}$$

分布函数 $F^{(x_1,x_2,\ldots,x_n)}$ 只针对于特定集合 $L_{a_1a_2\ldots a_n}$ 给出了概率 $\mathrm{P}^{(x_1,x_2,\ldots,x_n)}$ 的值。但如果概率域是波莱尔域，则[2]对于所有 R^n 上的波莱尔集， $\mathrm{P}^{(x_1,x_2,\ldots,x_n)}$ 都可以由分布函数 $F^{(x_1,x_2,\ldots,x_n)}$ 唯一确定（uniquely determined）。

如果存在导数

$$f(a_1,a_2,\ldots,a_n)=\frac{\partial}{\partial a_1\partial a_2\ldots\partial a_n}F^{(x_1,x_2,\ldots,x_n)}(a_1,a_2,\ldots,a_n)$$

则我们称这个导数为随机变量 $x_1,x_2,\ldots,x_n$ 在点 $a_1,a_2,\ldots,a_n$ 处的 n 维概率密度。并且如果对于每个点 $(a_1,a_2,\ldots,a_n)$，都有

$$F^{(x_1,x_2,\ldots,x_n)}(a_1,a_2,\ldots,a_n)=\int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\ldots\int_{-\infty}^{a_n}f(a_1,a_2,\ldots,a_n)da_1da_2\ldots da_n$$

则 $x_1,x_2,\ldots,x_n$ 的分布称为连续的。对于每一个波莱尔集合 $A\subset R^n$，有以下等式

$$\mathrm{P}^{(x_1,x_2,\ldots,x_n)}(A)=\int\int\ldots\int f(a_1,a_2,\ldots,a_n)da_1da_2\ldots da_n \tag{9}$$

在本章最后我们对各种概率函数和分布函数的关系再做一个说明。

给定如下代换

$$S=\begin{pmatrix}1,2,\dots,n\\ i_1,i_2,\dots,i_n\end{pmatrix}$$

令 r_s 表示如下的空间 R^n 到其自身的一个变换：

$$x'_k=x_{ik}\quad(k=1,2,\dots,n)$$

显然可得

$$\mathrm{P}^{(x_{i_1},x_{i_2},\dots,x_{i_n})}(A)=\mathrm{P}^{(x_1,x_2,\dots,x_n)}\{r_s^{-1}(A)\}\tag{10}$$

现在令 $x'=p_k(x)$ 为空间 R^n 在空间 R^k $(k<n)$ 的投影，所以空间 R^n 中的点 $x_1,x_2,\dots,x_n$ 映射到空间 R^k 中为 $x_1,x_2,\dots,x_k$。所以，与第 1 节中的公式 (2) 类似，有

$$\mathrm{P}^{(x_1,x_2,\dots,x_k)}(A)=\mathrm{P}^{(x_1,x_2,\dots,x_n)}(A)\{p_k^{-1}(A)\}\tag{11}$$

对应的分布函数，由公式 (10) 和公式 (11) 可得以下两个方程：

$$F^{(x_{i_1},x_{i_2},\dots,x_{i_n})}(a_{i_1},a_{i_2},\dots,a_{i_n})=F^{(x_1,x_2,\dots,x_n)}(a_1,a_2,\dots,a_n)\tag{12}$$

$$F^{(x_1,x_2,\dots,x_k)}(a_1,a_2,\dots,a_k)=F^{(x_1,x_2,\dots,x_n)}(a_1,a_2,\dots,a_k,+\infty,\dots,+\infty)\tag{13}$$

无限维空间中的概率

在第二章第 3 节我们已经看到如何构造概率论中常见的各种概率域。不过我们仍可以想象某一类问题，在这类问题中基本事件是由无限多个坐标定义的。假定一个由任意基数 m 中的索引 μ 构成的集合 M。我们把实数 x_μ 构成的的系统的总体（其中 μ 可以遍历整个集合 M）

$$\xi=\{x_\mu\}$$

称为空间 R^M（为了定义空间 R^M 中的一个元素 ξ，我们必须将集合 M 中的每一个元素 μ 与一个实数 x_μ 对应，或者等效地给每一个元素 μ 赋予一个定义在 M 上的单值实函数 x_μ）[3]。如果集合 M 中所包含的是 n 个自然数 $1,2,\dots,n$，则 R^M 就是普通的 n 维空间 R^n。如果集合 M 中是所有的实数 R^1，则对应的空间 $R^M=R^{R^1}$ 包含了实变量 μ 的所有的实函数：

$$\xi(\mu)=x_\mu$$

现在我们将集合 R^M（其中 M 为任意集合）作为基本集合 E。令 $\xi=\{x_\mu\}$ 为集合 E 中的一个元素。则我们可以用符号 $p_{\mu_1\mu_2\dots\mu_n}(\xi)$ 表示 n 维空间 R^n 中的点 $(x_{\mu_1},x_{\mu_2},\dots,x_{\mu_n})$。如果 E 的子集 A 可以表示成如下形式，我们称集合 A 为圆柱集合（cylinder set）：

$$A=p^{-1}_{\mu_1\mu_2\dots\mu_n}(\xi)(A')$$

其中 A' 是 R^n 的子集。因此，所有圆柱集组成的类（class），与那些可以按如下形式的关系定义的集合组成的类一致：

$$f(x_{\mu_1},x_{\mu_2},\dots,x_{\mu_n})=0\tag{1}$$

为按照以上关系定义任意圆柱集 $p_{\mu_1\mu_2\dots\mu_n}(A)$，我们只需构造一个函数 f，使它在 A' 处等于 0，并且在除 A' 以外的其他地方等于 1（unity）。

当 A' 是波莱尔集合时，对应的这个圆柱集是波莱尔圆柱集（Borel cylinder set）。空间 R^M 中的所有波莱尔圆柱集构成一个域，记为 $\mathcal{E}^M$ [4]。

我们把域 $\mathcal{E}^M$ 的波莱尔扩展记为 $B\mathcal{E}^M$。$B\mathcal{E}^M$ 中的集合称为空间 R^M 的波莱尔集合。

稍后我们会给出一种在 $\mathcal{E}^M$ 上构造和操作概率函数的方法，并且借由扩展定理（Extension THeorem），该方法同样可用于在 $B\mathcal{E}^M$ 上构造和操作概率函数。由此方式得到的概率域，在当集合 M 为可数的情况下，可以满足所有的（研究）目的。由此我们可以处理涉及到可数随机变量序列的所有问题。但当 M 不可数时，R^M 中的很多简单又有趣的子集就超出 $B\mathcal{E}^M$ 的范围。例如，当集合 M 不可数时，对于所有的索引 μ，那些满足 x_μ 小于某个固定常数的所有元素 ξ 构成的集合，就不属于系统 $B\mathcal{E}^M$。

因此，尽可能地把每一个问题转化为一种形式，使得在这种形式下所有基本事件 ξ 的空间只有一个可数的坐标集（a denumerable set of coordinates），这种方法是非常可取的。

假设概率函数 $\mathrm{P}(A)$ 定义在 $\mathcal{E}^M$ 上，则我们可以把基本事件 ξ 的每一个坐标 x_μ 都当作一个随机变量。因此这些坐标的每个有限群（group）$(x_{\mu_1},x_{\mu_2},\dots,x_{\mu_n})$ 都有一个 n 维的概率函数 $\mathrm{P}_{\mu_1\mu_2\dots\mu_n}(A)$ 和与之对应的分布函数 $F_{\mu_1\mu_2\dots\mu_n}(a_1,a_2,\dots,a_n)$。显然，对每一个波莱尔圆柱集 A：

$$A = p^{-1}_{\mu_1\mu_2...\mu_n}(A')$$

有如下等式成立：

$$\mathrm{P}(A) = \mathrm{P}_{\mu_1\mu_2...\mu_n}(A')$$

其中 A' 是空间 R^n 的一个波莱尔集。这样，空间 $\mathcal{E}^M$ 上所有圆柱集的概率函数 P 可以由空间 R^n 上的所有波莱尔集的优点概率函数 $\mathrm{P}_{\mu_1\mu_2...\mu_n}$ 确定。然而，对于波莱尔集，概率函数 $\mathrm{P}_{\mu_1\mu_2...\mu_n}$ 又是由相应的分布函数唯一确定的。由此我们证明了以下定理：

> 所有有限维分布函数 $F_{\mu_1\mu_2...\mu_n}$ 的集合唯一确定了所有在空间 $\mathcal{E}^M$ 中的集合的概率函数 $\mathrm{P}(A)$。如果 $\mathrm{P}(A)$ 定义在 $\mathcal{E}^M$ 上，则（根据扩展定理）它由分布函数 $F_{\mu_1\mu_2...\mu_n}$ 的值在 $B\mathcal{E}^M$ 上唯一确定。

接下来读者或许想问：在哪种情况下，一个分布函数 $F_{\mu_1\mu_2...\mu_n}$ 系统能够先验地定义一个空间 $\mathcal{E}^M$ 上（或者空间 $B\mathcal{E}^M$）的概率域呢。

首先需要说明，分布函数 $F_{\mu_1\mu_2...\mu_n}$ 必须满足第二章第 3 节的第 III 个例子（无限概率域的例子）。事实上这些条件包含在分布函数的概念当中。此外，作为本章第 2 节方程 (13) 和 (14) 的结果（注，推测应为第 3 节的方程 (12) 和 (13)），我们有如下两个等式关系：

$$F_{\mu_{i_1}\mu_{i_2}...\mu_{i_n}}(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) = F_{\mu_1\mu_2...\mu_n}(a_1, a_2, \ldots, a_n) \tag{2}$$

$$F_{\mu_1\mu_2...\mu_k}(a_1, a_2, \ldots, a_k) = F_{\mu_1\mu_2...\mu_n}(a_1, a_2, \ldots, a_k, +\infty, \ldots, +\infty) \tag{3}$$

其中 $k < n$，并且 $\begin{pmatrix} 1, 2, \ldots, n \\ i_1, i_2, \ldots, i_n \end{pmatrix}$ 是一个任意的排列（permutation）。这些必要条件也被证明为是充分条件。这一点由下面的定理可以得出。

基本定理：每一个满足条件 (2) 和 (3) 的分布函数系统，都定义了一个 $\mathcal{E}^M$ 上的概率函数 $\mathrm{P}(A)$，它满足公理 I - VI。并且这个概率函数 $\mathrm{P}(A)$ 也可以（通过扩展定理）扩展到空间 $B\mathcal{E}^M$ 上。

证明

给定满足第二章第 3 节例子 III 的一般条件，以及满足条件 (2) 和 (3) 的分布函数 $F_{\mu_1\mu_2...\mu_n}$。meige1分布函数都唯一确定了空间 R^n 上的所有波莱尔集的概率函数 $\mathrm{P}_{\mu_1\mu_2...\mu_n}$ （参考本章第 3 节）。后面的讨论中，我们只关注空间 R^n 中的波莱尔集和空间 E 中的波莱尔圆柱集。

对于每一个圆柱集 $A = p^{-1}_{\mu_1\mu_2...\mu_n}(A')$，我们令

$$\mathrm{P}(A) = \mathrm{P}_{\mu_1\mu_2...\mu_n}(A'') \tag{4}$$

由于可以由不同的集合 A' 经过构造得到相同的圆柱集 A，所以我们首先需要确定公式 (4) 总是得到相同的 $\mathrm{P}(A)$ 。

假设 $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ 为随机变量 x_μ 的有限系统。根据这些随机变量的概率函数 $\mathrm{P}_{\mu_1\mu_2...\mu_n}$，再结合第 3 节提到的规则，我们可以定义每一个子系统 $(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}})$ 的概率函数 $\mathrm{P}_{\mu_{i_1}\mu_{i_2}...\mu_{i_k}}$。由等式 (2) 和 (3) 可得，根据第 3 章内容定义的概率函数与先验给出的函数 $\mathrm{P}_{\mu_{i_1}\mu_{i_2}...\mu_{i_k}}$ 相同。现在我们假设圆柱集 A 是通过下面等式定义的：

$$A = p^{-1}_{\mu_{i_1}\mu_{i_2}...\mu_{i_k}}(A')$$

同时也是通过下面等式定义的：

$$A = p^{-1}_{\mu_{j_1}\mu_{j_2}...\mu_{j_m}}(A'')$$

其中所有的随机变量 x_{μ_i} 和 x_{μ_j} 都属于系统 $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$，这显然不是一个本质上的限制。条件

$$(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'$$

和条件

$$(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset A''$$

是等价的。因此

$$\mathrm{P}_{\mu_{i_1}\mu_{i_2}...\mu_{i_k}}(A') = \mathrm{P}(A) = \mathrm{P}_{\mu_1\mu_2...\mu_n}\{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'\} = \mathrm{P}_{\mu_1\mu_2...\mu_m}\{(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset A''\} = \mathrm{P}_{\mu_{j_1}\mu_{j_2}..}$$

由此证明了 $\mathrm{P}(A)$ 定义的唯一性。

接下来证明概率域 $(\mathcal{E}^M, \mathrm{P})$ 满足所有的公理 I - VI。公理 I 只要求 $\mathcal{E}^M$ 为一个域——这一事实已经在上面的内容中证明。并且，对于任意索引 μ，有如下关系：

$$E = p_\mu^{-1}(R^1)$$
$$\mathrm{P}(E) = \mathrm{P}_\mu(R^1) = 1$$

由此证明了公理 II 和公理 IV 也满足。最后，由 $\mathrm{P}(A)$ 的定义可得 $\mathrm{P}(A)$ 是非负的（满足公理 III）。

证明公理 V 略微复杂一些。为此我们需要考虑两个圆柱集

$$A = p^{-1}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A')$$

和

$$B = p^{-1}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(B')$$

我们假设随机变量 x_{μ_i} 和 x_{μ_j} 属于有限系统（原文为inclusive finite system，inclusive 不知道该怎么翻译）$(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$。如果集合 A 和集合 B 没有交集，则以下两个关系

$$(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A'$$

和（下面的等式，最后的下标 k，或许应该为 m，待确定）

$$(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_k}}) \subset B'$$

是互斥的。因此

$$\mathrm{P}(A+B) = \mathrm{P}_{\mu_1\mu_2\ldots\mu_n}\{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}) \subset A' \quad \mathrm{OR} \quad \{(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}) \subset B'\} = \mathrm{P}_{\mu_1\mu_2\ldots\mu_n}\{(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x$$

由此得证公理 V。

此时只剩下公理 VI。令

$$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$$

为一个圆柱集的递减序列，并且满足以下条件

$$\lim \mathrm{P}(A_n) = L > 0$$

我们需要证明，所有集合 A_n 的乘积为非空。在不会实际限制原有问题的情况下，我们可以假设在前 n 个圆柱集 A_k 的定义中，序列 $x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_{i_n}}, \ldots$ 中只有前 n 个坐标 x_{μ_k} 存在，即

$$A_n = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(B_n)$$

为简写记，记为

$$\mathrm{P}_{\mu_1\mu_2\ldots\mu_n}(B) = \mathrm{P}_n(B)$$

显然有

$$\mathrm{P}_n(B_n) = \mathrm{P}(A_n) \geqq L > 0$$

在每一个集合 B_n 中都有可能找到一个具有闭区间的有界集合 U_n，使得

$$\mathrm{P}_n(B_n - U_n) \leqq \frac{\epsilon}{2^n}$$

根据以上不等式，对于集合 V_n

$$V_n = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(U_n)$$

可得不等式

$$\mathrm{P}(A_n - V_n) \leqq \frac{\epsilon}{2^n} \tag{5}$$

进一步，令

$$W_n = V_1 V_2 \ldots V_n$$

由公式 (5) 可得

$$\mathrm{P}(A_n - W_n) \leqq \epsilon$$

由于 $W_n \subset V_n \subset A_n$，所以有如下关系

$$\mathrm{P}(W_n) \geqq \mathrm{P}(A_n) - \epsilon \geqq L - \epsilon$$

如果 ϵ 足够小，$\mathrm{P}(W_n) > 0$ 并且 W_n 不为空。我们可以在每个集合 W_n 中选择一个坐标为 $x_\mu^{(n)}$ 的点 $\xi^{(n)}$。每个点 $\xi^{(n+p)}$，其中 $p = 0, 1, 2, \ldots$，都属于集合 V_n。因此有

$$(x_{\mu_1}^{(n+p)}, x_{\mu_2}^{(n+p)}, \ldots, x_{\mu_n}^{(n+p)}) = p_{\mu_1\mu_2\ldots\mu_n}^{-1}(\xi^{(n+p)}) \subset U_n$$

因为集合 U_n 是有界的，所以我们可以从序列 $\{\xi^{(n)}\}$ 中挑选出一个子序列（采用对角线法，by the diagonal method）：

$$\xi^{(n_1)}, \xi^{(n_2)}, \ldots, \xi^{(n_i)}, \cdots$$

在这个子序列中，对于任意的 k，每一个点的坐标 $x_{\mu_k}^{(n_i)}$ 都趋向于一个确定的极限 x_k。最后我们令 ξ 为集合 E 中的一点，它的坐标为

$$x_{\mu_k} = x_k$$
$$x_m = 0, \quad \mu \neq \mu_k \qquad k = 1, 2, 3, \cdots$$

作为序列 $x_1^{(n_i)}, x_2^{(n_i)}, \ldots, x_k^{(n_i)}, \quad i = 1, 2, 3, \ldots$ 的极限，点 $(x_1, x_2, \ldots, x_k)$ 属于集合 U_k。因此对于任意 k，ξ 属于

$$A_k \subset V_k = p_{\mu_1\mu_2\ldots\mu_k}^{-1}(U_k)$$

因此证明

$$A = \underset{k}{\mathscr{D}} A_k$$

等价随机变量及多种收敛

从现在开始，我们将专门讨论概率的波莱尔域（Borel fields of probability）。正如在第 2 节中已经解释过的，这一讨论范围的限定实际上并不会构成对我们研究问题的限制。【有点奇怪】

如果两个随机变量 x 和 y，$x \neq y$ 的概率等于 0，则称这两个随机变量等价（equivalent）。显然，两个等价的随机变量具有相同的概率函数：

$$\mathrm{P}^{(x)}(A) = \mathrm{P}^{(y)}(A)$$

因此，它们对应的分布函数 $F^{(x)}$ 和 $F^{(y)}$ 也相同。在概率理论的许多问题中，我们都可以把某一个随机变量替换为任何其他等价随机变量。

现在，令

$$x_1, x_2, \cdots, x_n, \cdots \tag{1}$$

为一个随机变量序列。现在来研究满足序列 (1) 为收敛序列的所有基本事件 ξ 构成的集合 A。将满足以下不等式的基本事件 ξ 的集合记为 $A_{n\,p}^{(m)}$：

$$|x_{n+k} - x_n| < \frac{1}{m} \qquad k = 1, 2, \cdots, p$$

由此可得

$$A = \underset{m}{\mathscr{D}} \underset{n}{\mathscr{G}} \underset{p}{\mathscr{D}} A_{n\,p}^{(m)} \tag{2}$$

由第 3 节可知，集合 $A_{n\,p}^{(m)}$ 总是属于域 $\mathcal{E}$。等式 (2) 表明，集合 A 也属于 $\mathcal{E}$。因此，我们可以讨论随机变量序列收敛的概率，因为它总是具有明确的意义。

现在令收敛集合 A 的概率 $\mathrm{P}(A)$ 等于 1，则序列 (1) 依概率 1 收敛于一个随机变量 x。其中 x 除了等价关系外是唯一确定的。为确定随机变量 x，令集合 A 中的

$$x = \lim x_n \qquad n \to \infty$$

且集合 A 外有 $x = 0$。我们需要证明，x 是一个随机变量，即满足 $x < a$ 的所有元素 ξ 构成的集合 $A(a)$ 属于 $\mathcal{E}$。但当 $a \leq 0$ 时有

$$A(a) = A \underset{n}{\mathscr{G}} \underset{p}{\mathscr{D}} \{x_{n+p} < a\}$$

相反，当 $a > 0$ 时有

$$A(a) = A \underset{n}{\mathscr{G}} \underset{p}{\mathscr{D}} \{x_{n+p} < a\} + \bar{A}$$

由此，可立即得出**序列 (1) 依概率 1 收敛于一个随机变量** x的结论。

如果收敛序列 (1) 收敛于 x 的概率为 1，则称序列 (1) 几乎必然收敛到 x （converges almost surely to x）。但对概率理论来说，另一种收敛的概念可能更为重要。

定义： 如果对于任意 $\epsilon > 0$，随机变量序列 $x_1, x_2, \ldots, x_n, \ldots$ 依概率收敛于随机变量 x，则当 $n \to \infty$ 时，概率

$$\mathrm{P}\{|x_n - x| > \epsilon\}$$

趋近于 0 [5]。

I. 如果序列 (1) 依概率收敛于 x 和 x'，则 x 与 x' 等价。事实上

$$\mathrm{P}\{|x - x'| > \frac{1}{m}\} \leq \mathrm{P}\{|x_n - x| > \frac{1}{2m}\} + \mathrm{P}\{|x_n - x'| > \frac{1}{2m}\}$$

由于对于足够大的 n，最后这些概率可以任意小，因此可以得出

$$\mathrm{P}\{|x - x'| > \frac{1}{m}\} = 0$$

并且可立即得出

$$\mathrm{P}\{x \neq x'\} \leq \sum_m \mathrm{P}\{|x - x'| > \frac{1}{m}\} = 0$$

II. 如果序列 (1) 几乎必然收敛于 x，则它也依概率收敛于 x。令 A 为序列 (1) 的收敛集合，则

$$1 = \mathrm{P}(A) \leq \lim_{n \to \infty} \{|x_{n+p} - x| < \epsilon, p = 0, 1, 2, \ldots\} \leq \lim_{n \to \infty} \mathrm{P}\{|x_n - x| < \epsilon\}$$

由此可以推得依概率收敛。

III. 对于序列 (1) 依概率收敛的情况，以下条件既是充分的，也是必要的：对任意 $\xi > 0$，总存在一个 n，使得对于任意 $p > 0$，以下等式成立【这里有问题，因为都没有 p】：

$$\mathrm{P}\{|x_{n+p} - x_n| > \epsilon\} < \epsilon$$

令 $F_1(a), F_2(a), \ldots, F_n(a), \ldots, F(a)$ 为随机变量 $x_1, x_2, \ldots, x_n, \ldots, x$ 的分布函数。如果序列 x_n 依概率收敛于 x，分布函数 $F(a)$ 由 $F_n(a)$ 唯一确定。事实上有以下定理

定理：如果序列 $x_1, x_2, \ldots, x_n, \ldots$ 依概率收敛于 x，则对应的分布函数 $F_n(a)$ 的序列在 $F(a)$ 的每一个连续点处收敛于 x 的分布函数 $F(a)$。

$F(a)$ 是左连续的单调函数，由连续点处的值唯一确定 [6]。由此可得，$F(a)$ 由 $F_n(a)$ 确定。为证明这一定理，假设 F 在点 a 连续。令 $a' < a$，则当 $x < a'$、$x_n \geq a$ 时，必须有 $|x_n - x| > a - a'$。因此有

$$\begin{gathered}
\lim \mathrm{P}(x < a', x_n \geq a) = 0, \\
F(a') = \mathrm{P}(x < a') \leq \mathrm{P}(x_n < a) + \mathrm{P}(x < a', x_n \geq a) = F_n(a) + \mathrm{P}(x < a', x_n \geq a), \\
F(a') \leq \liminf F_n(a) + \lim \mathrm{P}(x < a', x_n \geq a), \\
F(a') \leq \liminf F_n(a)
\end{gathered} \tag{3}$$

注：$\liminf$ 表示下极限（limit inferior），即序列 $F_n(a)$ 在所有子序列极限中的最小极限值。下文中的 $\limsup$ 表示上极限。【感慨一下，翻译到这里时，deepseek已经声名鹊起】

由以上结果类比，当 $a'' > a$ 时有

$$F(a'') \geq \limsup F_x(a) \tag{4}$$

因为当 $a' \to a$，且 $a'' \to a$时，$F(a')$ 与 $F(a'')$ 都收敛于 $F(a)$，所以由公式 (3) 和 (4) 可得

$$\lim F_n(a) = F(a)$$

由此证明以上定理。

$$\begin{gathered}
f(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n}) = 0 \\
g(x_{\lambda_1}, x_{\lambda_2}, \ldots, x_{\lambda_m}) = 0
\end{gathered}$$

由此可根据以下关系定义集合 $A + B$，AB，以及 $A - B$：

$$f \cdot g = 0$$
$$f^2 + g^2 = 0$$
$$f^2 + \omega(g) = 0$$

其中当 $x \neq 0$ 时 $\omega(x) = 0$，并且 $\omega(0) = 1$。如果 f 和 g 都是波莱尔函数，则 $f \cdot g$、$f^2 + g^2$ 和 $f^2 + \omega(g)$ 也是波莱尔函数。因此 $A + B$，AB，以及 $A - B$ 都是波莱尔圆柱集。由此我们证明集合 $\mathcal{E}^M$ 的系统为一个域。

---

1. a_i 也可以取无限值 $\pm\infty$。↩
2. 参见第 IV 章第 3 节。↩
3. 参考HAUSDORFF, Mengenlehre, 1927, p. 23。↩
4. 由以上内容可知，波莱尔圆柱集是可以通过公式 (1) 中的关系定义的波莱尔集合。现假定集合 A 和 B 是两个由如下关系定义的波莱尔圆柱集：↩
5. 这一概念来源于 Bernoulli，它的完整的、一般性的分析由 E. E. Slutsky 提出（见参考文献 [1]）。↩
6. 事实上 $F_n(a)$ 至多只有可数个间断点（见 LEBESGUE, Leçons sur l'intégration, 1928, p. 50）。因此连续点是处处稠密的（everywhere dense），并且函数 $F(a)$ 在间断点的值由它左连续点的函数值的极限决定。↩