

4.13: Kernels and Operators

The goal of this section is to study a type of mathematical object that arises naturally in the context of conditional expected value and parametric distributions, and is of fundamental importance in the study of stochastic processes, particularly Markov processes. In a sense, the main object of study in this section is a generalization of a matrix, and the operations generalizations of matrix operations. If you keep this in mind, this section may seem less abstract.

Basic Theory

Definitions

Recall that a measurable space (S, \mathcal{S}) consists of a set S and a σ -algebra \mathcal{S} of subsets of S . If μ is a positive measure on (S, \mathcal{S}) , then (S, \mathcal{S}, μ) is a measure space. The two most important special cases that we have studied frequently are

1. *Discrete*: S is countable, $\mathcal{S} = \mathcal{P}(S)$ is the collection of all subsets of S , and $\mu = \#$ is counting measure on (S, \mathcal{S}) .
2. *Euclidean*: S is a measurable subset of \mathbb{R}^n for some $n \in \mathbb{N}_+$, \mathcal{S} is the collection of subsets of S that are also measurable, and $\mu = \lambda_n$ is n -dimensional Lebesgue measure on (S, \mathcal{S}) .

More generally, S usually comes with a topology that is locally compact, Hausdorff, with a countable base (LCCB), and \mathcal{S} is the *Borel σ -algebra*, the σ -algebra generated by the topology (the collection of open subsets of S). The measure μ is usually a *Borel measure*, and so satisfies $\mu(C) < \infty$ if $C \subseteq S$ is compact. A discrete measure space is of this type, corresponding to the discrete topology. A Euclidean measure space is also of this type, corresponding to the Euclidean topology, if S is open or closed (which is usually the case). In the discrete case, every function from S to another measurable space is measurable, and every from function from S to another topological space is continuous, so the measure theory is not really necessary.

Recall also that the measure space (S, \mathcal{S}, μ) is σ -finite if there exists a countable collection $\{A_i : i \in I\} \subseteq \mathcal{S}$ such that $\mu(A_i) < \infty$ for $i \in I$ and $S = \bigcup_{i \in I} A_i$. If (S, \mathcal{S}, μ) is a Borel measure space corresponding to an LCCB topology, then it is σ -finite.

If $f : S \rightarrow \mathbb{R}$ is measurable, define $\|f\| = \sup\{|f(x)| : x \in S\}$. Of course we may well have $\|f\| = \infty$. Let $\mathcal{B}(S)$ denote the collection of bounded measurable functions $f : S \rightarrow \mathbb{R}$. Under the usual operations of pointwise addition and scalar multiplication, $\mathcal{B}(S)$ is a vector space, and $\|\cdot\|$ is the natural norm on this space, known as the *supremum norm*. This vector space plays an important role.

In this section, it is sometimes more natural to write integrals with respect to the positive measure μ with the differential *before* the integrand, rather than after. However, rest assured that this is mere notation, the meaning of the integral is the same. So if $f : S \rightarrow \mathbb{R}$ is measurable then we may write the integral of f with respect to μ in operator notation as

$$\mu f = \int_S \mu(dx) f(x) \quad (4.13.1)$$

assuming, as usual, that the integral exists. This will be the case if f is nonnegative, although ∞ is a possible value. More generally, the integral exists in $\mathbb{R} \cup \{-\infty, \infty\}$ if $\mu f^+ < \infty$ or $\mu f^- < \infty$ where f^+ and f^- are the positive and negative parts of f . If both are finite, the integral exists in \mathbb{R} (and f is *integrable* with respect to μ). If μ is a probability measure and we think of (S, \mathcal{S}) as the sample space of a random experiment, then we can think of f as a real-valued random variable, in which case our new notation is not too far from our traditional expected value $\mathbb{E}(f)$. Our main definition comes next.

Suppose that (S, \mathcal{S}) and (T, \mathcal{T}) are measurable spaces. A *kernel* from (S, \mathcal{S}) to (T, \mathcal{T}) is a function $K : S \times \mathcal{T} \rightarrow [0, \infty]$ such that

1. $x \mapsto K(x, A)$ is a measurable function from S into $[0, \infty]$ for each $A \in \mathcal{T}$.
2. $A \mapsto K(x, A)$ is a positive measure on \mathcal{T} for each $x \in S$.

If $(T, \mathcal{T}) = (S, \mathcal{S})$, then K is said to be a *kernel on* (S, \mathcal{S}) .

There are several classes of kernels that deserve special names.

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) . Then

1. K is σ -finite if the measure $K(x, \cdot)$ is σ -finite for every $x \in S$.
2. K is finite if $K(x, T) < \infty$ for every $x \in S$.
3. K is bounded if $K(x, T)$ is bounded in $x \in S$.
4. K is a probability kernel if $K(x, T) = 1$ for every $x \in S$.

Define $\|K\| = \sup\{K(x, T) : x \in S\}$, so that $\|K\| < \infty$ if K is a bounded kernel and $\|K\| = 1$ if K is a probability kernel.

So a probability kernel is bounded, a bounded kernel is finite, and a finite kernel is σ -finite. The terms *stochastic kernel* and *Markov kernel* are also used for probability kernels, and for a probability kernel $\|K\| = 1$ of course. The terms are consistent with terms used for measures: K is a finite kernel if and only if $K(x, \cdot)$ is a finite measure for each $x \in S$, and K is a probability kernel if and only if $K(x, \cdot)$ is a probability measure for each $x \in S$. Note that $\|K\|$ is simply the supremum norm of the function $x \mapsto K(x, T)$.

A kernel defines two natural integral operators, by operating on the *left* with measures, and by operating on the *right* with functions. As usual, we are often a bit casual with the question of existence. Basically in this section, we assume that any integrals mentioned exist.

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) .

1. If μ is a positive measure on (S, \mathcal{S}) , then μK defined as follows is a positive measure on (T, \mathcal{T}) :

$$\mu K(A) = \int_S \mu(dx) K(x, A), \quad A \in \mathcal{T} \quad (4.13.2)$$

2. If $f : T \rightarrow \mathbb{R}$ is measurable, then $Kf : S \rightarrow \mathbb{R}$ defined as follows is measurable (assuming that the integrals exist in \mathbb{R}):

$$Kf(x) = \int_T K(x, dy) f(y), \quad x \in S \quad (4.13.3)$$

Proof

1. Clearly $\mu K(A) \geq 0$ for $A \in \mathcal{T}$. Suppose that $\{A_j : j \in J\}$ is a countable collection of disjoint sets in \mathcal{T} and $A = \bigcup_{j \in J} A_j$. Then

$$\begin{aligned} \mu K(A) &= \int_S \mu(dx) K(x, A) = \int_S \mu(dx) \left(\sum_{j \in J} K(x, A_j) \right) \\ &= \sum_{j \in J} \int_S \mu(dx) K(x, A_j) = \sum_{j \in J} \mu K(A_j) \end{aligned}$$

The interchange of sum and integral is justified since the terms are nonnegative.

2. The measurability of Kf follows from the measurability of f and of $x \mapsto K(x, A)$ for $A \in \mathcal{T}$, and from basic properties of the integral.

Thus, a kernel transforms measures on (S, \mathcal{S}) into measures on (T, \mathcal{T}) , and transforms certain measurable functions from T to \mathbb{R} into measurable functions from S to \mathbb{R} . Again, part (b) assumes that f is integrable with respect to the measure $K(x, \cdot)$ for every $x \in S$. In particular, the last statement will hold in the following important special case:

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) and that $f \in \mathcal{B}(T)$.

1. If K is finite then Kf is defined and $\|Kf\| = \|K\| \|f\|$.
2. If K is bounded then $Kf \in \mathcal{B}(T)$.

Proof

1. If K is finite then

$$K|f|(x) = \int_T K(x, dy) |f(y)| \leq \int_T K(x, dy) \|f\| = \|f\| K(x, T) < \infty \quad x \in S \quad (4.13.4)$$

Hence f is integrable with respect to $K(x, \cdot)$ for each $x \in S$ so Kf is defined. Continuing with our inequalities, we have $|Kf(x)| \leq K|f|(x) \leq \|f\|K(x, T) \leq \|f\|\|K\|$ so $\|Kf\| \leq \|K\|\|f\|$. Moreover equality holds when $f = \mathbf{1}_T$, the constant function 1 on T .

2. If K is bounded then $\|K\| < \infty$ so from (a), $\|Kf\| < \infty$.

The *identity kernel* I on the measurable space (S, \mathcal{S}) is defined by $I(x, A) = \mathbf{1}(x \in A)$ for $x \in S$ and $A \in \mathcal{S}$.

Thus, $I(x, A) = 1$ if $x \in A$ and $I(x, A) = 0$ if $x \notin A$. So $x \mapsto I(x, A)$ is the *indicator function* of $A \in \mathcal{S}$, while $A \mapsto I(x, A)$ is *point mass* at $x \in S$. Clearly the identity kernel is a probability kernel. If we need to indicate the dependence on the particular space, we will add a subscript. The following result justifies the name.

Let I denote the identity kernel on (S, \mathcal{S}) .

1. If μ is a positive measure on (S, \mathcal{S}) then $\mu I = \mu$.
2. If $f : S \rightarrow \mathbb{R}$ is measurable, then $If = f$.

Constructions

We can create a new kernel from two given kernels, by the usual operations of addition and scalar multiplication.

Suppose that K and L are kernels from (S, \mathcal{S}) to (T, \mathcal{T}) , and that $c \in [0, \infty)$. Then cK and $K + L$ defined below are also kernels from (S, \mathcal{S}) to (T, \mathcal{T}) .

1. $(cK)(x, A) = cK(x, A)$ for $x \in S$ and $A \in \mathcal{T}$.
2. $(K + L)(x, A) = K(x, A) + L(x, A)$ for $x \in S$ and $A \in \mathcal{T}$.

If K and L are σ -finite (finite) (bounded) then cK and $K + L$ are σ -finite (finite) (bounded), respectively.

Proof

These results are simple.

1. Since $x \mapsto K(x, A)$ is measurable for $A \in \mathcal{T}$, so is $x \mapsto cK(x, A)$. Since $A \mapsto K(x, A)$ is a positive measure on (T, \mathcal{T}) for $x \in S$, so is $A \mapsto cK(x, A)$ since $c \geq 0$.
2. Since $x \mapsto K(x, A)$ and $x \mapsto L(x, A)$ are measurable for $A \in \mathcal{T}$, so is $x \mapsto K(x, A) + L(x, A)$. Since $A \mapsto K(x, A)$ and $A \mapsto L(x, A)$ are positive measures on (T, \mathcal{T}) for $x \in S$, so is $A \mapsto K(x, A) + L(x, A)$.

A simple corollary of the last result is that if $a, b \in [0, \infty)$ then $aK + bL$ is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) . In particular, if K, L are probability kernels and $p \in (0, 1)$ then $pK + (1 - p)L$ is a probability kernel. A more interesting and important way to form a new kernel from two given kernels is via a “multiplication” operation.

Suppose that K is a kernel from (R, \mathcal{R}) to (S, \mathcal{S}) and that L is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) . Then KL defined as follows is a kernel from (R, \mathcal{R}) to (T, \mathcal{T}) :

$$KL(x, A) = \int_S K(x, dy)L(y, A), \quad x \in R, A \in \mathcal{T} \quad (4.13.5)$$

1. If K is finite and L is bounded then KL is finite.
2. If K and L are bounded then KL is bounded.
3. If K and L are stochastic then KL is stochastic

Proof

The measurability of $x \mapsto (KL)(x, A)$ for $A \in \mathcal{T}$ follows from basic properties of the integral. For the second property, fix $x \in R$. Clearly $KL(x, A) \geq 0$ for $A \in \mathcal{T}$. Suppose that $\{A_j : j \in J\}$ is a countable collection of disjoint sets in \mathcal{T} and $A = \bigcup_{j \in J} A_j$. Then

$$\begin{aligned}
 KL(x, A) &= \int_S K(x, dy) L(x, A) = \int_S K(x, dy) \left(\sum_{j \in J} L(y, A_j) \right) \\
 &= \sum_{j \in J} \int_S K(x, dy) L(y, A_j) = \sum_{j \in J} KL(x, A_j)
 \end{aligned}$$

The interchange of sum and integral is justified since the terms are nonnegative.

Once again, the identity kernel lives up to its name:

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) . Then

1. $I_S K = K$
2. $K I_T = K$

The next several results show that the operations are *associative* whenever they make sense.

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , μ is a positive measure on \mathcal{S} , $c \in [0, \infty)$, and $f : T \rightarrow \mathbb{R}$ is measurable. Then, assuming that the appropriate integrals exist,

1. $c(\mu K) = (c\mu)K$
2. $c(Kf) = (cK)f$
3. $(\mu K)f = \mu(Kf)$

Proof

These results follow easily from the definitions.

1. The common measure on \mathcal{T} is $c\mu K(A) = c \int_S \mu(dx) K(x, A)$ for $A \in \mathcal{T}$.
2. The common function from S to \mathbb{R} is $cKf(x) = c \int_S K(x, dy) f(y)$ for $x \in S$, assuming that the integral exists for $x \in S$.
3. The common real number is $\mu Kf = \int_S \mu(dx) \int_T K(x, dy) f(y)$, assuming that the integrals exist.

Suppose that K is a kernel from (R, \mathcal{R}) to (S, \mathcal{S}) and L is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) . Suppose also that μ is a positive measure on (R, \mathcal{R}) , $f : T \rightarrow \mathbb{R}$ is measurable, and $c \in [0, \infty)$. Then, assuming that the appropriate integrals exist,

1. $(\mu K)L = \mu(KL)$
2. $K(Lf) = (KL)f$
3. $c(KL) = (cK)L$

Proof

These results follow easily from the definitions.

1. The common measure on (T, \mathcal{T}) is $\mu KL(A) = \int_R \mu(dx) \int_S K(x, dy) L(y, A)$ for $A \in \mathcal{T}$.
2. The common measurable function from R to \mathbb{R} is $KLf(x) = \int_S K(x, dy) \int_T L(y, dz) f(z)$ for $x \in R$, assuming that the integral exists for $x \in S$.
3. The common kernel from (R, \mathcal{R}) to (T, \mathcal{T}) is $cKL(x, A) = c \int_S K(x, dy) L(y, A)$ for $x \in R$ and $A \in \mathcal{T}$.

Suppose that K is a kernel from (R, \mathcal{R}) to (S, \mathcal{S}) , L is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , and M is a kernel from (T, \mathcal{T}) to (U, \mathcal{U}) . Then $(KL)M = K(LM)$.

Proof

This results follow easily from the definitions. The common kernel from (R, \mathcal{R}) to (U, \mathcal{U}) is

$$KLM(x, A) = \int_S K(x, dy) \int_T L(y, dz) M(z, A), \quad x \in R, A \in \mathcal{U} \quad (4.13.6)$$

The next several results show that the *distributive property* holds whenever the operations makes sense.

Suppose that K and L are kernels from (R, \mathcal{R}) to (S, \mathcal{S}) and that M and N are kernels from (S, \mathcal{S}) to (T, \mathcal{T}) . Suppose also that μ is a positive measure on (R, \mathcal{R}) and that $f : S \rightarrow \mathbb{R}$ is measurable. Then, assuming that the appropriate integrals exist,

1. $(K + L)M = KM + LM$
2. $K(M + N) = KM + KN$
3. $\mu(K + L) = \mu K + \mu L$
4. $(K + L)f = Kf + Lf$

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , and that μ and ν are positive measures on (S, \mathcal{S}) , and that f and g are measurable functions from T to \mathbb{R} . Then, assuming that the appropriate integrals exist,

1. $(\mu + \nu)K = \mu K + \nu K$
2. $K(f + g) = Kf + Kg$
3. $\mu(f + g) = \mu f + \mu g$
4. $(\mu + \nu)f = \mu f + \nu f$

In particular, note that if K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , then the transformation $\mu \mapsto \mu K$ defined for positive measures on (S, \mathcal{S}) , and the transformation $f \mapsto Kf$ defined for measurable functions $f : T \rightarrow \mathbb{R}$ (for which Kf exists), are both *linear* operators. If μ is a positive measure on (S, \mathcal{S}) , then the integral operator $f \mapsto \mu f$ defined for measurable $f : S \rightarrow \mathbb{R}$ (for which μf exists) is also linear, but of course, we already knew that. Finally, note that the operator $f \mapsto Kf$ is *positive*: if $f \geq 0$ then $Kf \geq 0$. Here is the important summary of our results when the kernel is bounded.

If K is a bounded kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , then $f \mapsto Kf$ is a bounded, linear transformation from $\mathcal{B}(T)$ to $\mathcal{B}(S)$ and $\|K\|$ is the norm of the transformation.

The *commutative property* for the product of kernels fails with a passion. If K and L are kernels, then depending on the measurable spaces, KL may be well defined, but not LK . Even if both products are defined, they may be kernels from or to different measurable spaces. Even if both are defined from and to the same measurable spaces, it may well happen that $KL \neq LK$. Some [examples](#) are given below

If K is a kernel on (S, \mathcal{S}) and $n \in \mathbb{N}$, we let $K^n = KK \cdots K$, the n -fold power of K . By convention, $K^0 = I$, the identity kernel on S .

Fixed points of the operators associated with a kernel turn out to be very important.

Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) .

1. A positive measure μ on (S, \mathcal{S}) such that $\mu K = \mu$ is said to be *invariant* for K .
2. A measurable function $f : T \rightarrow \mathbb{R}$ such that $Kf = f$ is said to be *invariant* for K

So in the language of linear algebra (or functional analysis), an invariant measure is a *left eigenvector* of the kernel, while an invariant function is a *right eigenvector* of the kernel, both corresponding to the eigenvalue 1. By our results above, if μ and ν are invariant measures and $c \in [0, \infty)$, then $\mu + \nu$ and $c\mu$ are also invariant. Similarly, if f and g are invariant functions and $c \in \mathbb{R}$, then $f + g$ and cf are also invariant.

Of course we are particularly interested in *probability* kernels.

Suppose that P is a probability kernel from (R, \mathcal{R}) to (S, \mathcal{S}) and that Q is a probability kernel from (S, \mathcal{S}) to (T, \mathcal{T}) . Suppose also that μ is a probability measure on (R, \mathcal{R}) . Then

1. PQ is a probability kernel from (R, \mathcal{R}) to (T, \mathcal{T}) .
2. μP is a probability measure on (S, \mathcal{S}) .

Proof

1. We know that PQ is a kernel from (R, \mathcal{R}) to (T, \mathcal{T}) . So we just need to note that

$$PQ(T) = \int_S P(x, dy)Q(y, T) = \int_S P(x, dy) = P(x, S) = 1, \quad x \in R \quad (4.13.7)$$

2. We know that μP is a positive measure on (S, \mathcal{S}) . So we just need to note that

$$\mu P(S) = \int_R \mu(dx) P(x, S) = \int_R \mu(dx) = \mu(R) = 1 \quad (4.13.8)$$

As a corollary, it follows that if P is a probability kernel on (S, \mathcal{S}) , then so is P^n for $n \in \mathbb{N}$.

The operators associated with a kernel are of fundamental importance, and we can easily recover the kernel from the operators. Suppose that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , and let $x \in S$ and $A \in \mathcal{T}$. Then trivially, $K\mathbf{1}_A(x) = K(x, A)$ where as usual, $\mathbf{1}_A$ is the indicator function of A . Trivially also $\delta_x K(A) = K(x, A)$ where δ_x is point mass at x .

Kernel Functions

Usually our *measurable* spaces are in fact *measure* spaces, with natural measures associated with the spaces, as in the special cases described in (1). When we start with measure spaces, kernels are usually constructed from *density functions* in much the same way that positive measures are defined from density functions.

Suppose that $(S, \mathcal{S}, \lambda)$ and (T, \mathcal{T}, μ) are measure spaces. As usual, $S \times T$ is given the product σ -algebra $\mathcal{S} \otimes \mathcal{T}$. If $k : S \times T \rightarrow [0, \infty)$ is measurable, then the function K defined as follows is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) :

$$K(x, A) = \int_A k(x, y) \mu(dy), \quad x \in S, A \in \mathcal{T} \quad (4.13.9)$$

Proof

The measurability of $x \mapsto K(x, A) = \int_A k(x, y) \mu(dy)$ for $A \in \mathcal{T}$ follows from a basic property of the integral. The fact that $A \mapsto K(x, A) = \int_A k(x, y) \mu(dy)$ is a positive measure on \mathcal{T} for $x \in S$ also follows from a basic property of the integral. In fact, $y \mapsto k(x, y)$ is the density of this measure with respect to μ .

Clearly the kernel K depends on the positive measure μ on (T, \mathcal{T}) as well as the function k , while the measure λ on (S, \mathcal{S}) plays no role (and so is not even necessary). But again, our point of view is that the spaces have fixed, natural measures. Appropriately enough, the function k is called a *kernel density function* (with respect to μ), or simply a *kernel function*.

Suppose again that $(S, \mathcal{S}, \lambda)$ and (T, \mathcal{T}, μ) are measure spaces. Suppose also K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) with kernel function k . If $f : T \rightarrow \mathbb{R}$ is measurable, then, assuming that the integrals exists,

$$Kf(x) = \int_S k(x, y) f(y) \mu(dy), \quad x \in S \quad (4.13.10)$$

Proof

This follows since the function $y \mapsto k(x, y)$ is the density of the measure $A \mapsto K(x, A)$ with respect to μ :

$$Kf(x) = \int_S K(x, dy) f(y) = \int_S k(x, y) f(y) \mu(dy), \quad x \in S \quad (4.13.11)$$

A kernel function defines an operator on the left with functions on S in a completely analogous way to the operator on the right above with functions on T .

Suppose again that $(S, \mathcal{S}, \lambda)$ and (T, \mathcal{T}, μ) are measure spaces, and that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) with kernel function k . If $f : S \rightarrow \mathbb{R}$ is measurable, then the function $fK : T \rightarrow \mathbb{R}$ defined as follows is also measurable, assuming that the integrals exists

$$fK(y) = \int_S \lambda(dx) f(x) k(x, y), \quad y \in T \quad (4.13.12)$$

The operator defined above depends on the measure λ on (S, \mathcal{S}) as well as the kernel function k , while the measure μ on (T, \mathcal{T}) plays no role (and so is not even necessary). But again, our point of view is that the spaces have fixed, natural measures. Here is how our new operation on the left with *functions* relates to our old operation on the left with *measures*.

Suppose again that $(S, \mathcal{S}, \lambda)$ and (T, \mathcal{T}, μ) are measure spaces, and that K is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) with kernel function k . Suppose also that $f : S \rightarrow [0, \infty)$ is measurable, and let ρ denote the measure on (S, \mathcal{S}) that has density f with respect to λ . Then fK is the density of the measure ρK with respect to μ .

Proof

The main tool, as usual, is an interchange of integrals. For $B \in \mathcal{T}$,

$$\begin{aligned}\rho K(B) &= \int_S \rho(dx) K(x, B) = \int_S f(x) K(x, B) \lambda(dx) = \int_S f(x) \left[\int_B k(x, y) \mu(dy) \right] \lambda(dx) \\ &= \int_B \left[\int_S f(x) k(x, y) \lambda(dx) \right] \mu(dy) = \int_B fK(y) \mu(dy)\end{aligned}$$

As always, we are particularly interested in stochastic kernels. With a kernel function, we can have *doubly* stochastic kernels.

Suppose again that $(S, \mathcal{S}, \lambda)$ and (T, \mathcal{T}, μ) are measure spaces and that $k : S \times T \rightarrow [0, \infty)$ is measurable. Then k is a *double stochastic kernel function* if

1. $\int_T k(x, y) \mu(dy) = 1$ for $x \in S$
2. $\int_S \lambda(dx) k(x, y) = 1$ for $y \in S$

Of course, condition (a) simply means that the kernel associated with k is a stochastic kernel according to our original definition.

The most common and important special case is when the two spaces are the same. Thus, if $(S, \mathcal{S}, \lambda)$ is a measure space and $k : S \times S \rightarrow [0, \infty)$ is measurable, then we have an operator K that operates on the left and on the right with measurable functions $f : S \rightarrow \mathbb{R}$:

$$\begin{aligned}fK(y) &= \int_S \lambda(dx) f(x) k(x, y), \quad y \in S \\ Kf(x) &= \int_S k(x, y) f(y) \lambda(dy), \quad x \in S\end{aligned}$$

If f is nonnegative and μ is the measure on with density function f , then fK is the density function of the measure μK (both with respect to λ).

Suppose again that $(S, \mathcal{S}, \lambda)$ is a measure space and $k : S \times S \rightarrow [0, \infty)$ is measurable. Then k is *symmetric* if $k(x, y) = k(y, x)$ for all $(x, y) \in S^2$.

Of course, if k is a symmetric, stochastic kernel function on $(S, \mathcal{S}, \lambda)$ then k is doubly stochastic, but the converse is not true.

Suppose that $(R, \mathcal{R}, \lambda)$, (S, \mathcal{S}, μ) , and (T, \mathcal{T}, ρ) are measure spaces. Suppose also that K is a kernel from (R, \mathcal{R}) to (S, \mathcal{S}) with kernel function k , and that L is a kernel from (S, \mathcal{S}) to (T, \mathcal{T}) with kernel function l . Then the kernel KL from (R, \mathcal{R}) to (T, \mathcal{T}) has density kl given by

$$kl(x, z) = \int_S k(x, y) l(y, z) \mu(dy), \quad (x, z) \in R \times T \quad (4.13.13)$$

Proof

Once again, the main tool is an interchange of integrals via Fubini's theorem. Let $x \in R$ and $B \in \mathcal{T}$. Then

$$\begin{aligned}KL(x, B) &= \int_S K(x, dy) L(y, B) = \int_S k(x, y) L(y, B) \mu(dy) \\ &= \int_S k(x, y) \left[\int_B l(y, z) \rho(dz) \right] \mu(dy) = \int_B \left[\int_S k(x, y) l(y, z) \mu(dy) \right] \rho(dz) = \int_B kl(x, z) \mu(dz)\end{aligned}$$

Examples and Special Cases

The Discrete Case

In this subsection, we assume that the measure spaces are discrete, as described in (1). Since the σ -algebra (all subsets) and the measure (counting measure) are understood, we don't need to reference them. Recall that integrals with respect to counting measure are sums. Suppose now that K is a kernel from the discrete space S to the discrete space T . For $x \in S$ and $y \in T$, let $K(x, y) = K(x, \{y\})$. Then more generally,

$$K(x, A) = \sum_{y \in A} K(x, y), \quad x \in S, A \subseteq T \quad (4.13.14)$$

The function $(x, y) \mapsto K(x, y)$ is simply the kernel function of the kernel K , as defined above, but in this case we usually don't bother with using a different symbol for the function as opposed to the kernel. The function K can be thought of as a *matrix*, with rows indexed by S and columns indexed by T (and so an infinite matrix if S or T is countably infinite). With this interpretation, all of the operations defined above can be thought of as matrix operations. If $f : T \rightarrow \mathbb{R}$ and f is thought of as a column vector indexed by T , then Kf is simply the ordinary product of the matrix K and the vector f ; the product is a column vector indexed by S :

$$Kf(x) = \sum_{y \in S} K(x, y)f(y), \quad x \in S \quad (4.13.15)$$

Similarly, if $f : S \rightarrow \mathbb{R}$ and f is thought of as a row vector indexed by S , then fK is simple the ordinary product of the vector f and the matrix K ; the product is a row vector indexed by T :

$$fK(y) = \sum_{x \in S} f(x)K(x, y), \quad y \in T \quad (4.13.16)$$

If L is another kernel from T to another discrete space U , then as functions, KL is the simply the matrix product of K and L :

$$KL(x, z) = \sum_{y \in T} K(x, y)L(y, z), \quad (x, z) \in S \times U \quad (4.13.17)$$

Let $S = \{1, 2, 3\}$ and $T = \{1, 2, 3, 4\}$. Define the kernel K from S to T by $K(x, y) = x + y$ for $(x, y) \in S \times T$. Define the function f on S by $f(x) = x!$ for $x \in S$, and define the function g on T by $g(y) = y^2$ for $y \in T$. Compute each of the following using matrix algebra:

1. fK
2. Kg

Answer

In matrix form,

$$K = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix}, \quad f = [1 \quad 2 \quad 6], \quad g = \begin{bmatrix} 1 \\ 4 \\ 9 \\ 16 \end{bmatrix} \quad (4.13.18)$$

1. As a row vector indexed by T , the product is $fK = [32 \quad 41 \quad 50 \quad 59]$
2. As a column vector indexed by S ,

$$Kg = \begin{bmatrix} 130 \\ 160 \\ 190 \end{bmatrix} \quad (4.13.19)$$

Let $R = \{0, 1\}$, $S = \{a, b\}$, and $T = \{1, 2, 3\}$. Define the kernel K from R to S , the kernel L from S to S and the kernel M from S to T in matrix form as follows:

$$K = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}, \quad L = \begin{bmatrix} 2 & 2 \\ 1 & 5 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 1 \end{bmatrix} \quad (4.13.20)$$

Compute each of the following kernels, or explain why the operation does not make sense:

1. KL
2. LK
3. K^2
4. L^2
5. KM
6. LM

Proof

Note that these are not just abstract matrices, but rather have rows and columns indexed by the appropriate spaces. So the products make sense only when the spaces match appropriately; it's not just a matter of the number of rows and columns.

1. KL is the kernel from R to S given by

$$KL = \begin{bmatrix} 6 & 22 \\ 7 & 19 \end{bmatrix} \quad (4.13.21)$$

2. LK is not defined since the column space S of L is not the same as the row space R of K .
3. K^2 is not defined since the row space R is not the same as the column space S .
4. L^2 is the kernel from S to S given by

$$L^2 = \begin{bmatrix} 6 & 14 \\ 7 & 27 \end{bmatrix} \quad (4.13.22)$$

5. KM is the kernel from R to T given by

$$KM = \begin{bmatrix} 1 & 12 & 6 \\ 2 & 9 & 7 \end{bmatrix} \quad (4.13.23)$$

6. LM is the kernel from S to T given by

$$LM = \begin{bmatrix} 2 & 6 & 6 \\ 1 & 15 & 7 \end{bmatrix} \quad (4.13.24)$$

Conditional Probability

An important class of probability kernels arises from the distribution of one random variable, conditioned on the value of another random variable. In this subsection, suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and that (S, \mathcal{S}) and (T, \mathcal{T}) are measurable spaces. Further, suppose that X and Y are random variables defined on the probability space, with X taking values in S and that Y taking values in T . Informally, X and Y are random variables defined on the same underlying random experiment.

The function P defined as follows is a probability kernel from (S, \mathcal{S}) to (T, \mathcal{T}) , known as the *conditional probability kernel* of Y given X .

$$P(x, A) = \mathbb{P}(Y \in A \mid X = x), \quad x \in S, A \in \mathcal{T} \quad (4.13.25)$$

Proof

Recall that for $A \in \mathcal{T}$, the conditional probability $\mathbb{P}(Y \in A \mid X)$ is itself a random variable, and is measurable with respect to $\sigma(X)$. That is, $\mathbb{P}(Y \in A \mid X) = P(X, A)$ for some measurable function $x \mapsto P(x, A)$ from S to $[0, 1]$. Then, by definition, $\mathbb{P}(Y \in A \mid X = x) = P(x, A)$. Trivially, of course, $A \mapsto P(x, A)$ is a probability measure on (T, \mathcal{T}) for $x \in S$.

The operators associated with this kernel have natural interpretations.

Let P be the conditional probability kernel of Y given X .

1. If $f : T \rightarrow \mathbb{R}$ is measurable, then $Pf(x) = \mathbb{E}[f(Y) \mid X = x]$ for $x \in S$ (assuming as usual that the expected value exists).
2. If μ is the probability distribution of X then μP is the probability distribution of Y .

Proof

These are basic results that we have already studied, dressed up in new notation.

1. Since $A \mapsto P(x, A)$ is the conditional distribution of Y given $X = x$,

$$\mathbb{E}[f(Y) \mid X = x] = \int_S P(x, dy) f(y) = P f(x) \quad (4.13.26)$$

2. Let $A \in \mathcal{T}$. Conditioning on X gives

$$\mathbb{P}(Y \in A) = \mathbb{E}[\mathbb{P}(Y \in A \mid X)] = \int_S \mu(dx) P(Y \in A \mid X = x) = \int_S \mu(dx) P(x, A) = \mu P(A) \quad (4.13.27)$$

As in the general discussion above, the measurable spaces (S, \mathcal{S}) and (T, \mathcal{T}) are usually *measure* spaces with natural measures attached. So the conditional probability distributions are often given via conditional probability density functions, which then play the role of kernel functions. The next two exercises give examples.

Suppose that X and Y are random variables for an experiment, taking values in \mathbb{R} . For $x \in \mathbb{R}$, the conditional distribution of Y given $X = x$ is normal with mean x and standard deviation 1. Use the notation and operations of this section for the following computations:

1. Give the kernel function for the conditional distribution of Y given X .
2. Find $\mathbb{E}(Y^2 \mid X = x)$.
3. Suppose that X has the standard normal distribution. Find the probability density function of Y .

Answer

1. The kernel function (with respect to Lebesgue measure, of course) is

$$p(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-x)^2}, \quad x, y \in \mathbb{R} \quad (4.13.28)$$

2. Let $g(y) = y^2$ for $y \in \mathbb{R}$. Then $E(Y^2 \mid X = x) = P g(x) = 1 + x^2$ for $x \in \mathbb{R}$

3. The standard normal PDF f is given $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for $x \in \mathbb{R}$. Thus Y has PDF fP .

$$fP(y) \int_{-\infty}^{\infty} f(x) p(x, y) dx = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}y^2}, \quad y \in \mathbb{R} \quad (4.13.29)$$

This is the PDF of the normal distribution with mean 0 and variance 2.

Suppose that X and Y are random variables for an experiment, with X taking values in $\{a, b, c\}$ and Y taking values in $\{1, 2, 3, 4\}$. The kernel function of Y given X is as follows: $P(a, y) = 1/4$, $P(b, y) = y/10$, and $P(c, y) = y^2/30$, each for $y \in \{1, 2, 3, 4\}$.

1. Give the kernel P in matrix form and verify that it is a probability kernel.
2. Find fP where $f(a) = f(b) = f(c) = 1/3$. The result is the density function of Y given that X is uniformly distributed.
3. Find Pg where $g(y) = y$ for $y \in \{1, 2, 3, 4\}$. The resulting function is $\mathbb{E}(Y \mid X = x)$ for $x \in \{a, b, c\}$.

Answer

1. P is given in matrix form below. Note that the row sums are 1.

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{10} & \frac{2}{10} & \frac{3}{10} & \frac{4}{10} \\ \frac{1}{30} & \frac{4}{30} & \frac{9}{30} & \frac{16}{30} \end{bmatrix} \quad (4.13.30)$$

2. In matrix form, $f = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ and $fP = \begin{bmatrix} \frac{23}{180} & \frac{35}{180} & \frac{51}{180} & \frac{71}{180} \end{bmatrix}$.

3. In matrix form,

$$g = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad Pg = \begin{bmatrix} \frac{5}{2} \\ 3 \\ \frac{10}{3} \end{bmatrix} \quad (4.13.31)$$

Parametric Distributions

A parametric probability distribution also defines a probability kernel in a natural way, with the parameter playing the role of the kernel variable, and the distribution playing the role of the measure. Such distributions are usually defined in terms of a parametric density function which then defines a kernel function, again with the parameter playing the role of the first argument and the variable the role of the second argument. If the parameter is thought of as a given value of another random variable, as in Bayesian analysis, then there is considerable overlap with the previous subsection. In most cases, (and in particular in the examples below), the spaces involved are either discrete or Euclidean, as described in (1).

Consider the parametric family of exponential distributions. Let f denote the identity function on $(0, \infty)$.

1. Give the probability density function as a probability kernel function p on $(0, \infty)$.
2. Find Pf .
3. Find fP .
4. Find p^2 , the kernel function corresponding to the product kernel P^2 .

Answer

1. $p(r, x) = re^{-rx}$ for $r, x \in (0, \infty)$.
2. For $r \in (0, \infty)$,

$$Pf(r) = \int_0^\infty p(r, x)f(x) dx = \int_0^\infty xre^{-rx} dx = \frac{1}{r} \quad (4.13.32)$$

This is the mean of the exponential distribution.

3. For $x \in (0, \infty)$,

$$fP(x) = \int_0^\infty f(r)p(r, x) dr = \int_0^\infty r^2e^{-rx} dr = \frac{2}{x^3} \quad (4.13.33)$$

4. For $r, y \in (0, \infty)$,

$$p^2(r, y) = \int_0^\infty p(r, x)p(x, y) dx = \int_0^\infty \int_0^\infty rxe^{-(r+y)x} dx = \frac{r}{(r+y)^2} \quad (4.13.34)$$

Consider the parametric family of Poisson distributions. Let f be the identity function on \mathbb{N} and let g be the identity function on $(0, \infty)$.

1. Give the probability density function p as a probability kernel function from $(0, \infty)$ to \mathbb{N} .
2. Show that $Pf = g$.
3. Show that $gP = f$.

Answer

1. $p(r, n) = e^{-r} \frac{r^n}{n!}$ for $r \in (0, \infty)$ and $n \in \mathbb{N}$.
2. For $r \in (0, \infty)$, $Pf(r)$ is the mean of the Poisson distribution with parameter r :

$$Pf(r) = \sum_{n=0}^\infty p(r, n)f(n) = \sum_{n=0}^\infty ne^{-r} \frac{r^n}{n!} = r \quad (4.13.35)$$

3. For $n \in \mathbb{N}$,

$$gP(n) = \int_0^\infty g(r)p(r, n) dr = \int_0^\infty e^{-r} \frac{r^{n+1}}{n!} dr = n \quad (4.13.36)$$

Clearly the Poisson distribution has some very special and elegant properties. The next family of distributions also has some very special properties. Compare this exercise with the exercise (30).

Consider the family of normal distributions, parameterized by the mean and with variance 1.

1. Give the probability density function as a probability kernel function p on \mathbb{R} .

2. Show that p is symmetric.
3. Let f be the identity function on \mathbb{R} . Show that $Pf = f$ and $fP = f$.
4. For $n \in \mathbb{N}$, find p^n the kernel function for the operator P^n .

Answer

1. For $\mu, x \in \mathbb{R}$,

$$p(\mu, x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} \quad (4.13.37)$$

That is, $x \mapsto p(x, \mu)$ is the normal probability density function with mean μ and variance 1.

2. Note that $p(\mu, x) = p(x, \mu)$ for $\mu, x \in \mathbb{R}$. So $\mu \mapsto p(\mu, x)$ is the normal probability density function with mean x and variance 1.
3. Since $f(x) = x$ for $x \in \mathbb{R}$, this follows from the previous two parts: $Pf(\mu) = \mu$ for $\mu \in \mathbb{R}$ and $fP(x) = x$ for $x \in \mathbb{R}$.
4. For $\mu, y \in \mathbb{R}$,

$$p^2(\mu, x) = \int_{-\infty}^{\infty} p(\mu, t)p(t, y) dt = \frac{1}{\sqrt{4\pi}} e^{-\frac{1}{4}(x-\mu)^2} \quad (4.13.38)$$

so that $x \mapsto p^2(\mu, x)$ is the normal PDF with mean μ and variance 2. By induction,

$$p^n(\mu, x) = \frac{1}{\sqrt{2\pi n}} e^{-\frac{1}{2n}(x-\mu)^2} \quad (4.13.39)$$

for $n \in \mathbb{N}_+$ and $\mu, x \in \mathbb{R}$. Thus $x \mapsto p^n(\mu, x)$ is the normal PDF with mean μ and variance n .

For each of the following special distributions, express the probability density function as a probability kernel function. Be sure to specify the parameter spaces.

1. The general normal distribution on \mathbb{R} .
2. The beta distribution on $(0, 1)$.
3. The negative binomial distribution on \mathbb{N} .

Answer

1. The normal distribution with mean μ and standard deviation σ defines a kernel function p from $\mathbb{R} \times (0, \infty)$ to \mathbb{R} given by

$$p[(\mu, \sigma), x] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (4.13.40)$$

2. The beta distribution with left parameter a and right parameter b defines a kernel function p from $(0, \infty)^2$ to $(0, 1)$ given by

$$p[(a, b), x] = \frac{1}{B(a, b)} x^{a-1} y^{b-1} \quad (4.13.41)$$

where B is the beta function.

3. The negative binomial distribution with stopping parameter k and success parameter α defines a kernel function p from $(0, \infty) \times (0, 1)$ to \mathbb{N} given by

$$p[(n, \alpha), k] = \binom{n+k-1}{n} \alpha^k (1-\alpha)^n \quad (4.13.42)$$

This page titled [4.13: Kernels and Operators](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.