Notes on reversible jump Markov chain Monte Carlo

February 2018

1 Reversible jump Markov chain Monte Carlo[1]

1.1 Introduction

Consider the scenario in which we have a data set that we assume has been generated by a particular model. Let $\mathcal{X}$ be the state space of this model, so that each state $x \in \mathcal{X}$ is a realised parameter vector—often, but not always, of real numbers.

Let $\pi(dx)$ denote the target distribution that we are interested in—usually the posterior distribution for the model parameters given the data. We assume π is a probability measure[2] on $\mathcal{X}$, known up to a normalising constant. In standard Markov chain Monte Carlo, we define a conditional probability measure $P(x, dx')$ (a probability kernel) that gives transition probabilities between states in $\mathcal{X}$; if P is irreducible (so all states can be reached from any initial state), aperiodic (there are no states that can only be returned to in multiples of a certain number of steps), and reversible, then samples drawn from a run of the Markov chain defined by P approximate the target distribution π. More specifically, π is the stationary distribution of the Markov chain.

Practically, the most important of the above properties is reversibility. The Markov chain defined by P is said to be reversible with respect to π if it satisfies a property called detailed balance, which requires that for all measurable $A, B \subset \mathcal{X}$,

$$\int_A \int_B \pi(dx) P(x, dx') = \int_B \int_A \pi(dx') P(x', dx) \tag{1}$$

In Metropolis-Hastings MCMC, the transition measure $P(x, dx')$ is made up of a proposal distribution $q(x, A)$ and an acceptance density $\alpha(x, x')$, such

[1] See Peter Green, *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination*, 1995. Much of the current exposition is also based on Rasmus Waagepetersen, *A tutorial on reversible jump MCMC with a view towards applications in QTL mapping*, 2001.

[2] See Section 2 for some introductory measure theory.

that $P(x, dx') = q(x, dx')\alpha(x, x')$, and in this case, there is a choice of α that always leads to detailed balance:

$$\alpha(x, x') = \min\left(1, \frac{\pi(dx')q(x', dx)}{\pi(dx)q(x, dx')}\right). \tag{2}$$

To be clear, we have assumed that $\pi(dx)$ is only known as a density function up to a constant factor, say $\pi(dx) = p(x)d\pi/Z$. Fortunately, this is no real restriction, since $\pi(dx')/\pi(dx)$ can still be computed as

$$\frac{\pi(dx')}{\pi(dx)} = \frac{p(x')/Z}{p(x)/Z} = \frac{p(x')}{p(x)}.$$

Usually, $\pi(dx)$ is the posterior distribution of a model M's parameters $x \equiv \theta$, given data y:

$$\pi(A) = \int_A p(\theta \mid y, M)\, d\theta = \int_A p(y \mid \theta, M)p(\theta \mid M)/p(y \mid M)\, d\theta,$$

in which case $Z = p(y \mid M)$.

1.2 Multiple parameter spaces

In the more general case addressed by reversible jump MCMC, the state space $\mathcal{X}$ may consist of disjoint, disparate subspaces corresponding to a countable set of different models, and the Markov chain may move between them in much the same way as it does within them.

In single-model Metropolis-Hastings simulations, it is quite common to break the transition process into simpler parts by applying a number of different moves, either in sequence or according to a predefined probability vector. One of the notable contributions of reversible jump MCMC is the possibility for state-dependent mixing, in which the set of proposed moves may depend on the current state of the Markov chain.

For example, we might have a countable set of moves indexed by a set M, and a conditional probability $q_m(x, dx')$ of proposing a move of type m to state x', given that the current state of the chain is x. These probabilities need not sum to 1, so that with probability $1 - \sum_m q_m(x, \mathcal{X})$ no move is proposed. Also, it is reasonable to restrict the moves that are available from any given state, so it is quite possible that $q_m(x, \mathcal{X}) = 0$ for arbitrarily many m. As in the previous section, each move is paired with an acceptance probability $\alpha_m(x, x')$, which will be derived abstractly below, and in full for the real case later on. The probability of remaining at the current state x is then:

$$s(x) = 1 - \sum_{m \in M} q_m(x, \mathcal{X}) + \sum_m \int_{\mathcal{X}} q_m(x, dx')(1 - \alpha_m(x, x')),$$

and the transition measure P takes the form:

$$P(x, B) = \sum_m \int_B q_m(x, dx')\alpha_m(x, x') + s(x)[x \in B].$$

Here, B is a measurable subset of $\mathcal{X}$, and $[x \in B]$ is the Iverson bracket, or indicator function.

In terms of reversibility of the chain, state-dependent mixing imposes no real restrictions, because for detailed balance to hold overall, it need only hold for each move individually. Substituting the above form for $P(x, B)$ into (1) gives:

$$\begin{aligned}\sum_m \int_A \pi(dx) \int_B q_m(x, dx')\alpha_m(x, x') + \int_{A\cap B} \pi(dx)s(x) \\ = \sum_m \int_B \pi(dx') \int_A q_m(x', dx)\alpha_m(x', x) + \int_{B\cap A} \pi(dx')s(x'),\end{aligned}$$

which is satisfied if for each m, A, and B,

$$\begin{aligned}\int_A \pi(dx) \int_B q_m(x, dx')\alpha_m(x, x') \\ = \int_B \pi(dx') \int_A q_m(x', dx)\alpha_m(x', x).\end{aligned} \tag{3}$$

There is an important difference between this equation and (1): in contrast with the previous section, the measures being integrated with respect to here are not necessarily interchangeable. If q_m is a model-switching move, and A and B belong to parameter spaces S_1 and S_2 respectively, then in the first integral we are concerned with a measure on $S_1 \times S_2$ (it may be null elsewhere on $\mathcal{X} \times \mathcal{X}$), whereas the measure in the second integral has to do with $S_2 \times S_1$. For (3) to hold, we must first know that both integrals can at least be expressed with respect to the same measure; we thus make the following assumption: let μ_m be a measure on $\mathcal{X} \times \mathcal{X}$ relative to which πq_m has a finite, computable density on $(S_1 \times S_2) \cup (S_2 \times S_1)$. Also assume that μ_m is symmetric, in the sense that it matters not whether it is applied on $S_1 \times S_2$ or $S_2 \times S_1$, i.e., $\mu_m(dx, dx') = \mu_m(dx', dx)$. (This is all not quite as abstract and circuitous as it may sound—as Section 1.3 will hopefully show.)

With respect to μ_m, assume that $\pi(dx)q_m(x, dx')$ has density $f_1(x, x')$ and $\pi(dx')q_m(x', dx)$ density $f_2(x', x)$. Then detailed balance holds (for move m) if we let

$$\begin{aligned}\alpha_m(x, x') = \min\left(1, \frac{f_2(x', x)}{f_1(x, x')}\right) \\ \text{and} \\ \alpha_m(x', x) = \min\left(1, \frac{f_1(x', x)}{f_2(x, x')}\right),\end{aligned} \tag{4}$$

(we have simply made explicit the fact that α is defined separately on $S_1 \times S_2$ and $S_2 \times S_1$), because this implies that

$$f_1(x, x')\alpha_m(x, x') = f_2(x', x)\alpha_m(x', x),$$

which in turn is all that was required to prove (3):

$$\begin{aligned}\int_A \int_B \pi(dx) q_m(x, dx')\alpha_m(x, x') &= \int_A \int_B \mu_m(dx, dx') f_1(x, x')\alpha_m(x, x') \\ &= \int_B \int_A \mu_m(dx', dx) f_2(x', x)\alpha_m(x', x) \\ &= \int_B \int_A \pi(dx') q_m(x', dx)\alpha_m(x', x).\end{aligned}$$

1.3 Real parameter spaces

Let $\mathcal{X}$ be a state space consisting of real-valued vectors, so $\mathcal{X} = \cup_{i \in I} S_i$, with each $S_i \subseteq \mathbb{R}^{n_i}$, for some countable index set I. We will assume the Borel σ-algebra on $\mathcal{X}$, but won't refer to it directly. Let π again be the target distribution we wish to sample from (a probability measure on $\mathcal{X}$), and assume that for each subspace S_i the measure π has a non-negative, finite density p_i/Z_i with respect to Lebesgue measure λ_{n_i} on S_i.

Following on from the previous section, we can restrict ourselves to moves between a pair of subspaces, say S_1 and S_2. Let $q_1(x, dx')$ and $q_2(x', dx)$ be the proposal measures (technically, these are kernels) on $S_1 \times S_2$ and $S_2 \times S_1$ (i.e., for moves $S_1 \to S_2$ and $S_2 \to S_1$) respectively. Also assume that these moves are only attempted with probabilities $j_1(x)$ and $j_2(x')$, respectively.

The most obvious practical scenario (and the only one we address) is one in which moves can be made with the aid of auxiliary variables, and we have an idea of how these random variables should be generated. In the simplest case, we might have $n_1 < n_2$, and need $n_2 - n_1$ random numbers u to map a state $x \in S_1$ to a proposed state $x' \in S_2$. More generally, u might contain m_1 auxiliary numbers, so that $(x, u) \in \mathbb{R}^{n_1} \times \mathbb{R}^{m_1} = \mathbb{R}^n$ (say).

We assume that the proposed state x' is obtained with the aid of some bijection $h \colon \mathbb{R}^n \to \mathbb{R}^n$ that we have access to, so that $h(x, u) = (x', u')$, where u' is a vector of $m_2 = n - n_2$ auxiliary variables. The so-called 'dimension-matching' assumption of reversible jump MCMC is that we use h^{-1} when we perform moves in the reverse direction; it implies that a vector of auxiliary variables u' will need to be generated when proposing a state in S_1 given $x' \in S_2$ as well. We therefore need to assume that we have access to densities $g_1(u)$ and $g_2(u')$, with respect to λ_{m_1} and λ_{m_2}, that can be used to augment states in S_1 and S_2 respectively.

It is important (for reasons we'll point out below) that the bijection h is a diffeomorphism—that is, both h and h^{-1} are differentiable—and also that the intermediate dimension n is minimal, so that there are no superfluous

auxiliary variables. To be specific, (x, x') must determine (u, u')—otherwise the move $x \to x'$ could not be expressed using $g_1(u)$ for a particular u. We already know that (x, u) determines x' via h and a projection onto S_2; if we set $r_x(u) = x'$, the fact that (x, x') determines (u, u') will imply that r_x^{-1} exists.

Let $A \subseteq S_1$ and $B \subseteq S_2$. We have two measures, on $S_1 \times S_2$ and $S_2 \times S_1$ respectively:

$$\int_A \lambda_{n_1}(dx)\, p_1(x) j_1(x) q_1(x, B)$$
$$\text{and} \tag{5}$$
$$\int_B \lambda_{n_2}(dx')\, p_2(x') j_2(x') q_2(x', A).$$

Consider the first measure, which gives the probability of proposing a move from a state in A to one in B. We wish to define the proposal measure q_1 in such a way that its density (with respect to some new measure on S_2) can be expressed in terms of g_1, which is a density on $\mathbb{R}^{m_1}$. In particular, we require a measure μ_1 on S_2 for which

$$\int_B (g_1 \circ r_x^{-1})\, d\mu_1 = \int_{r_x^{-1}(B)} g_1\, d\lambda_{m_1}, \tag{6}$$

since this implies that the probability of proposing some $x' \in S_2$ is simply $g_1(u')$, where $u = r_x^{-1}(x')$. The change of variables theorem for integrals has just this form: set

$$\mu_1(B) = \lambda_{m_1}(r_x^{-1}(B)) = \lambda_{m_1}(\{\, u : r_x(u) \in B \,\});$$

then μ_1 is a positive measure on S_2, and for any measurable $f \colon S_2 \to \mathbb{R}$:

$$\int_{S_2} f\, d\mu_1 = \int_{\mathbb{R}^{m_1}} (f \circ r_x)\, d\lambda_{m_1}.$$

Equation 6 follows when $f = g_1 \circ r_x^{-1}$ on B:

$$\begin{aligned}\int_B (g_1 \circ r_x^{-1})\, d\mu_1 &= \int_{S_2} \mathbf{1}_B \cdot (g_1 \circ r_x^{-1})\, d\mu_1 \\ &= \int_{\mathbb{R}^{m_1}} (\mathbf{1}_B \circ r_x) \cdot g_1\, d\lambda_{m_1} = \int_{r_x^{-1}(B)} g_1\, d\lambda_{m_1}.\end{aligned}$$

The first measure in (5) can now be rewritten as

$$\int_A \lambda_{n_1}(dx)\, p_1(x) j_1(x) \int_B \mu_1(dx')\, g_1(r_x^{-1}(x')), \tag{7}$$

and since the product of two measures is a measure on the relevant product space, we have

$$\int_A \int_B \mu(dx, dx')\, p_1(x) j_1(x) g_1(r_x^{-1}(x')),$$

in which the product measure is

$$\mu(A \times B) = \lambda_{n_1}(A)\mu_1(B) = \lambda_n(\{\, (x,u) : x \in A, r_x(u) \in B \,\}).$$

This measure is in fact sufficient for our purposes, for two reasons: one, because it can also be applied to $S_2 \times S_1$, making it symmetric, and two, because both measures in (5) have computable densities with respect to μ.

The density for moves from S_1 to S_2 is clear:

$$f_1(x, x') = p_1(x) j_1(x) g_1(r_x^{-1}(x')).$$

The second density—on $S_2 \times S_1$—must exist by the Lebesgue decomposition and Radon-Nikodym theorems, and can also be given explicitly, because in this case the Radon-Nikodym derivative is the determinant of the Jacobian of $h^{-1} \colon \mathbb{R}^n \to \mathbb{R}^n$.

To see this, note firstly that if we apply all of the above steps to the second expression in (5), then the resulting measure on $S_2 \times S_1$ is

$$\int_B \lambda_{n_2}(dx')\, p_2(x') j_2(x') \int_A \mu_2(dx)\, g_2(r_{x'}^{-1}(x)), \tag{8}$$

mirroring (7). The new functions are straightforward: we have $r_{x'}(u') = x'$, and $\mu_2(A) = \lambda_{m_2}(r_{x'}^{-1}(A))$. Together, λ_{n_2} and μ_2 provide a measure on $S_2 \times S_1$ (as above):

$$\mu^*(B \times A) = \lambda_{n_2}(B)\mu_2(A) = \lambda_n(\{\, (x',u') : x' \in B, r_{x'}(u') \in A \,\}).$$

If $\mu(B \times A) = \mu(A \times B)$ is the Lebesgue measure of the region $R \subseteq \mathbb{R}^n$, then $\mu^*(B \times A)$ is the Lebesgue measure of $h(R)$. In addition, μ^* is absolutely continuous relative to μ, because if R is a null set of μ, one of its elements is determined by the other $n-1$, and since h is one-to-one, the same must be true of $h(R)$. Hence μ^* has a Radon-Nikodym derivative with respect to μ, and (8) implies

$$\int_B \int_A \mu^*(dx', dx)\, p_2(x') j_2(x') g_2(r_{x'}^{-1}(x)) \\ = \int_B \int_A \mu(dx, dx')\, \frac{d\mu^*}{d\mu} p_2(x') j_2(x') g_2(r_{x'}^{-1}(x)).$$

Since $d\mu^*/d\mu$ is the relative rate of change of μ^* with respect to μ, it is the factor by which h scales volume, and is thus given by the absolute value of

the determinant of the Jacobian matrix of h. The Jacobian exists for any differentiable mapping, but its determinant is only non-zero if the function's inverse is also differentiable—hence the diffeomorphism requirement above. In particular then, the density of the joint target-proposal measure on $S_2 \times S_1$ relative to μ is

$$f_2(x', x) = \left| \frac{\partial(x', u')}{\partial(x, u)} \right| p_2(x') j_2(x') g_2(r_{x'}^{-1}(x)),$$

and we are now in a position to prove that detailed balance holds for moves between S_1 and S_2, as long as we make use of the acceptance probability as formulated in (4). We only work out the case in which $\alpha_m(x, x') = f_2(x', x)/f_1(x, x') \leq 1$, the other one being entirely analogous.

Recall that we are trying to prove that (3) holds, which in the current context means showing that (7) and (8) are equal when paired, respectively, with $\alpha_m(x, x')$ and $\alpha(x', x)$:

$$\begin{aligned}
&\int_A \lambda_{n_1}(dx)\, p_1(x) j_1(x) \int_B \mu_1(dx')\, g_1(r_x^{-1}(x'))\alpha_m(x, x') \\
&\qquad = \int_A \int_B \mu(dx, dx')\, p_1(x) j_1(x) g_1(r_x^{-1}(x'))\alpha_m(x, x') \\
&\qquad = \int_A \int_B \mu(dx, dx')\, \frac{d\mu^*}{d\mu} p_2(x') j_2(x') g_2(r_{x'}^{-1}(x)). \\
&\qquad = \int_B \int_A \mu^*(dx', dx)\, p_2(x') j_2(x') g_2(r_{x'}^{-1}(x)) \\
&\qquad = \int_B \lambda_{n_2}(dx')\, p_2(x') j_2(x') \int_A \mu_2(dx)\, g_2(r_{x'}^{-1}(x))\alpha_m(x', x).
\end{aligned}$$

2 Some measure theory

These definitions and results all come from the *Random*[3] website.

2.1 Algebras and measurable functions

Let S be a set, and $\mathcal{S}$ a non-empty collection of subsets of S.

Definition 1. $\mathcal{S}$ is an *algebra* if it is closed under complements and unions.

In particular, an algebra of sets is closed under a *finite* number of unions, intersections, or complements (or a combination thereof).

Definition 2. $\mathcal{S}$ is a *σ-algebra* if it is closed under complements and countable unions.

[3] `http://www.randomservices.org/random`

It follows that a σ-algebra of sets is closed under countable intersections too. If S is a set and $\mathcal{S}$ a σ-algebra of subsets of S, then the pair $(S, \mathcal{S})$ is called a *measurable space*.

Lemma 1. *The intersection of a (possibly uncountable) collection of σ-algebras of subsets of $\mathcal{S}$ is itself a σ-algebra of subsets of $\mathcal{S}$.*

Let $\mathcal{B}$ be a collection of subsets of S.

Definition 3. The σ-algebra *generated by* $\mathcal{B}$ is the intersection of all σ-algebras of subsets of S that contain $\mathcal{B}$.

The σ-algebra generated by the collection of open subsets of S is called the *Borel* σ-algebra of S.

Suppose that $(S, \mathcal{S})$ and $(T, \mathcal{T})$ are measurable spaces.

Definition 4. A function $f\colon S \to T$ is *measurable* if $f^{-1}(B) \in \mathcal{S}$ for every $B \in \mathcal{T}$.

2.2 Measures

As above, let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be measurable spaces, and $f\colon S \to T$ be a measurable function.

Definition 5. A function $\mu\colon \mathcal{S} \to [0, \infty]$ is a *positive measure* on $(S, \mathcal{S})$ if $\mu(\emptyset) = 0$ and countable additivity holds (i.e., the measure of a countable union of disjoint subsets is the sum of the measures of the subsets).

A positive measure on $(S, \mathcal{S})$ that also satisfies $\mathrm{P}(S) = 1$ is called a *probability measure*. The triple $(S, \mathcal{S}, \mu)$ is called a *measure space*.

Lemma 2. *If μ is a positive measure on $(S, \mathcal{S})$ and $c > 0$, then $c\mu$ is also a positive measure on $(S, \mathcal{S})$.*

Lemma 3. *If $(R, \mathcal{R})$ is a measurable subspace of $(S, \mathcal{S})$—i.e., $(R, \mathcal{R})$ is a measurable space and $\mathcal{R} \subseteq \mathcal{S}$—then μ restricted to $\mathcal{R}$ is a positive measure on $(R, \mathcal{R})$.*

Theorem 1. *The function ν defined by $\nu(B) = \mu(f^{-1}(B))$ for $B \in \mathcal{T}$ is a positive measure on $(T, \mathcal{T})$.*

The construction of a measure according to Theorem 1 is sometimes referred to as a *change of variables*.

2.3 Integrals

Let $(S, \mathcal{S}, \mu)$ be a measure space. Then the integral of a measurable function $f : S \to \mathbb{R}$ is denoted by $\int_S f \, d\mu$, or $\int_S f(x) \, d\mu(x)$, or even $\int_S \mu(dx) f(x)$, and we say that f is *integrable* if the integral exists as a number in the set $\mathbb{R} \cup \{-\infty, \infty\}$.

Definition 6. If $A \in \mathcal{S}$ then $\int_S \mathbf{1}_A \, d\mu = \mu(A)$.

The integral with respect to a measure is defined in three stages: for simple functions, non-negative functions, and then general functions.

Definition 7. A *simple function* is a measurable, real-valued function with finite range.

Lemma 4. *A simple function f can be represented as a sum $\sum_{i \in I} a_i \mathbf{1}_{A_i}$, where I is a finite index set, $a_i \in \mathbb{R}$ for each $i \in I$, and $\{ A_i : i \in I \}$ is a collection of sets in $\mathcal{S}$ that partition S.*

If the a_i are distinct and the A_i non-empty, the representation is called *canonical*.

Let f be a non-negative simple function with representation $\sum_i a_i \mathbf{1}_{A_i}$, where $a_i \geq 0$ for each $i \in I$.

Definition 8. The integral of f with respect to μ is

$$\int_S f \, d\mu = \sum_{i \in I} a_i \mu(A_i).$$

The above definition is consistent, in that the value of the integral is the same for all representations of f.

Next, suppose that $f : S \to [0, \infty)$ is measurable.

Lemma 5. *There exists an increasing sequence $(f_1, f_2, \dots)$ of non-negative simple functions such that $f_n \to f$ on S as $n \to \infty$.*

That is, every non-negative measurable function is the limit of a sequence of non-negative simple functions.

Definition 9. The integral of a non-negative function f is

$$\int_S f \, d\mu = \sup \left\{ \int_S g \, d\mu : g \text{ is simple and } 0 \leq g \leq f \right\}.$$

The above lemma and definition are linked by the *monotone convergence theorem*.

Theorem 2. *If $(f_1, f_2, \dots)$ is an increasing sequence of non-negative simple functions such that $f_n \to f$ as $n \to \infty$, then*

$$\int_S \lim_{n \to \infty} f_n \, d\mu = \lim_{n \to \infty} \int_S f_n \, d\mu.$$

Finally, let $f\colon S \to \mathbb{R}$ be any measurable function. For $x \in \mathbb{R}$, write the positive and negative parts of x as $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$, so that $x = x^+ - x^-$ and $|x| = x^+ + x^-$. This definition extends to functions: f^+ is 0 wherever $f \leq 0$, and f^- is 0 wherever $f \geq 0$.

Definition 10. The integral of a measurable function f is

$$\int_S f\,\mu = \int_S f^+\,d\mu - \int_S f^-\,d\mu.$$

The function f is integrable if and only if $\int_S |f|\,d\mu < \infty$.

Definition 11. The integral of f over a measurable subset $A \in \mathcal{S}$ is

$$\int_A f\,d\mu = \int_S \mathbf{1}_A f\,d\mu.$$

The integral over the union of finitely many disjoint, measurable subsets is the sum of the integrals over the subsets.

(From Section 2.11.) Given a measure space $(S, \mathcal{S}, \mu)$, a measurable space $(T, \mathcal{T})$, and a measurable function $f\colon S \to T$, Theorem 1 states that the function $\nu(B) = \mu(f^{-1}(B))$ is a positive measure on $(T, \mathcal{T})$. This change of variables theorem extends to integrals as follows.

Theorem 3. *If $g\colon T \to \mathbb{R}$ is a measurable function, then, assuming the integrals exist,*

$$\int_T g\,d\nu = \int_S (g \circ f)\,d\mu.$$

2.4 Density functions

Let $(S, \mathcal{S}, \mu)$ be a measure space.

Definition 12. A subset $A \in \mathcal{S}$ is a *null* set of μ if $\mu(B) = 0$ for every $B \in \mathcal{S}$ such that $B \subseteq A$.

Definition 13. Let ν be another measure on $(S, \mathcal{S})$. Then ν is *absolutely continuous* with respect to μ—or *dominated* by μ—if every null set of μ is also a null set of ν. On the other hand, μ and ν are *mutually singular* if there exists an $A \in \mathcal{S}$ such that A is null for μ and A^c is null for ν.

Two measures μ and ν on $(S, \mathcal{S})$ are said to be *equivalent* if they are absolutely continuous with respect to one another.

Let $f\colon S \to \mathbb{R}$ be a measurable function whose integral with respect to μ exists.

Definition 14. The function ν defined by

$$\nu(A) = \int_A f \, d\mu$$

is a measure on $(S, \mathcal{S})$ that is absolutely continuous with respect to ν. The measurable function f is called a *density function* of ν with respect to μ.

If f is integrable, then ν is a finite measure; if f is non-negative, ν is a positive measure; and if f is both non-negative and $\int_S f \, d\mu = 1$, then ν is a probability measure. In the final case, f is a *probability density function* of ν relative to μ.

Lemma 6. *If f is a density function of ν with respect to μ, then $g\colon S \to \mathbb{R}$ is a density function of ν relative to μ if and only if $f = g$ almost everywhere on S with respect to μ.*

Theorem 4. *Let ν be a measure on $(S, \mathcal{S})$. Then ν can be decomposed as $\nu = \nu_c + \nu_s$ such that ν_c is absolutely continuous with respect to μ, and ν_s and μ are mutually singular. Furthermore, ν_c has a density function with respect to μ.*

The first and second parts of this theorem are called the *Lebesgue decomposition* and *Radon-Nikodym* theorems respectively. In particular, ν has a density function with respect to μ if and only if μ dominates ν. The derivative in this case is also referred to as the *Radon-Nikodym derivative* of ν with respect to μ, and is sometimes written as $d\nu/d\mu$.

Suppose that ν is absolutely continuous and has density function f relative to μ.

Theorem 5. *If g is a measurable function whose integral with respect to ν exists, then*

$$\int_S g \, d\nu = \int_S gf \, d\mu.$$

This is the *density theorem* for integrals, and is the reason for the notation $f = d\nu/d\mu$.

2.5 The reals

(From Section 0.12.) Consider the set of open intervals on the real line:

$$\mathcal{B} = \{\, (-\infty, a] : a \in \mathbb{R} \,\} \cup \{\, (a, b] : a, b \in \mathbb{R}, a < b \,\} \cup \{\, (b, \infty) : b \in \mathbb{R} \,\}.$$

Lemma 7. *The σ-algebra generated by $\mathcal{B}$ is the σ-algebra of Borel sets of $\mathbb{R}$.*

In $\mathbb{R}^n$, the collection that gives rise to the Borel σ-algebra on $\mathbb{R}^n$ is

$$\mathcal{B}_n = \left\{ \textstyle\prod_{i=1}^n A_i : A_i \in \mathcal{B} \text{ for each } i \in \{1, \ldots, n\} \right\}.$$

(From Section 1.3.)

Definition 15. The standard n-dimensional measure on $\mathbb{R}^n$ is called *Lebesgue measure*, defined by

$$\lambda_n(A) = \int_A 1 \, dx.$$

In particular, λ_1, λ_2, and λ_3 correspond to length, area, and volume, respectively.