Appendix A
Spaces of Measures

A.1 Weak Convergence of Probability Measures

Throughout this section, $\mathscr{X}$ is a Polish space with metric $d(x, y)$. For certain results we will need $\mathscr{X}$ also to be locally compact. In such cases, this assumption will be made explicit. Let $\mathscr{P}(\mathscr{X})$ denote the space of probability measures on $\mathscr{X}$ and let $\mathscr{C}_b(\mathscr{X})$ denote the space of bounded continuous functions mapping $\mathscr{X}$ into $\mathbb{R}$. Consider a sequence $\{\theta_n\}_{n\in\mathbb{N}}$ in $\mathscr{P}(\mathscr{X})$. We say that $\{\theta_n\}$ **converges weakly** to θ, and write $\theta_n \Rightarrow \theta$, if for each $g \in \mathscr{C}_b(\mathscr{X})$,

$$\lim_{n\to\infty} \int_{\mathscr{X}} g\, d\theta_n = \int_{\mathscr{X}} g\, d\theta.$$

Then $\mathscr{P}(\mathscr{X})$ is made into a topological space by taking as the basic open neighborhoods of $\gamma \in \mathscr{P}(\mathscr{X})$ the sets of the form

$$\left\{\theta \in \mathscr{P}(\mathscr{X}) : \left|\int_{\mathscr{X}} g_i\, d\theta - \int_{\mathscr{X}} g_i\, d\gamma\right| < \varepsilon, i = 1, 2, \ldots, k\right\},$$

where $\varepsilon > 0$, k is a positive integer, and $g_1, g_2, \ldots, g_k$ are in $\mathscr{C}_b(\mathscr{X})$. The resulting topology is called the **topology of weak convergence** or simply the **weak topology**.

To introduce a metric on $\mathscr{P}(\mathscr{X})$, for $A \subset \mathscr{X}$ and $\varepsilon > 0$ we define

$$A^{(\varepsilon)} \doteq \{x \in \mathscr{X} : d(x, A) < \varepsilon\}.$$

For γ and θ in $\mathscr{P}(\mathscr{X})$, we then define

$$\mathscr{L}(\gamma, \theta) \doteq \inf\{\varepsilon > 0 : \gamma(F) \leq \theta(F^{(\varepsilon)}) + \varepsilon \text{ for all closed subsets } F \text{ of } \mathscr{X}\}.$$

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

Then $\mathscr{L}(\gamma, \theta)$ defines a metric on $\mathscr{P}(\mathscr{X})$, known as the **Lévy–Prohorov metric** [126, p. 96].

As we state in the next theorem, the Lévy–Prohorov metric is compatible with the weak topology, and with respect to it, $\mathscr{P}(\mathscr{X})$ is Polish.

Theorem A.1 ([126, pp. 101 and 108]) *Let $\{\theta_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathscr{P}(\mathscr{X})$. Then $\theta_n \Rightarrow \theta \in \mathscr{P}(\mathscr{X})$ if and only if $\mathscr{L}(\theta_n, \theta) \to 0$. Furthermore, with respect to the Lévy–Prohorov metric, $\mathscr{P}(\mathscr{X})$ is complete and separable.*

The next result, known as the Portmanteau theorem, gives a number of useful conditions that are equivalent to weak convergence. For $\theta \in \mathscr{P}(\mathscr{X})$, a Borel set A whose boundary ∂A satisfies $\theta(\partial A) = 0$ is called a θ-**continuity set**.

Theorem A.2 ((Portmanteau theorem). [24, p. 11]) *Let $\{\theta_n\}$ and θ be probability measures on $\mathscr{X}$. The following five conditions are equivalent:*

(a) $\theta_n \Rightarrow \theta$.
(b) $\lim_{n\to\infty} \int_{\mathscr{X}} g\, d\theta_n = \int_{\mathscr{X}} g\, d\theta$ *for all bounded uniformly continuous functions g mapping $\mathscr{X}$ into $\mathbb{R}$.*
(c) $\limsup_{n\to\infty} \theta_n(F) \le \theta(F)$ *for all closed subsets F of $\mathscr{X}$.*
(d) $\liminf_{n\to\infty} \theta_n(G) \ge \theta(G)$ *for all open subsets G of $\mathscr{X}$.*
(e) $\lim_{n\to\infty} \theta_n(A) = \theta(A)$ *for all θ-continuity sets A.*

Remark A.3 The standard proof that (b) implies (c) uses a collection of Lipschitz continuous functions. Using this observation, we can augment the Portmanteau theorem with the following additional equivalent condition: $\theta_n \Rightarrow \theta$ if and only if

$$\lim_{n\to\infty} \int_{\mathscr{X}} g\, d\theta_n = \int_{\mathscr{X}} g\, d\theta$$

for all bounded Lipschitz continuous functions g mapping $\mathscr{X}$ into $\mathbb{R}$.

We next state Prohorov's theorem, which characterizes relatively compact subsets of $\mathscr{P}(\mathscr{X})$. It is one of the main results in the theory. A family Φ of probability measures on $\mathscr{X}$ is said to be **tight** if for each $\varepsilon > 0$, there exists a compact set K such that

$$\inf_{\gamma\in\Phi} \gamma(K) \ge 1 - \varepsilon.$$

Theorem A.4 ((Prohorov's theorem). [126, p. 103]) *A family of probability measures on $\mathscr{X}$ is relatively compact with respect to weak convergence if and only if it is tight. In particular, if $\theta_n \Rightarrow \theta$, then $\{\theta_n\}$ is tight.*

Prohorov's theorem yields the following useful fact.

Corollary A.5 *If $\mathscr{X}$ is a compact Polish space, then $\mathscr{P}(\mathscr{X})$ is compact.*

The notion of uniform integrability is useful in proving L^1 convergence.

Definition A.6 A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}^d$ (resp. a sequence $\{\xi_n\}$ of $\mathbb{R}^d$-valued random variables) is said to be **uniformly integrable** if

$$\sup_{n\in\mathbb{N}} \int_{\{x:\|x\|\geq a\}} \|x\| \mu_n(dx) \to 0 \text{ as } a \to \infty,$$

resp.

$$\sup_{n\in\mathbb{N}} E[\|\xi_n\| 1_{\{\|\xi_n\|\geq a\}}] \to 0 \text{ as } a \to \infty.$$

The following definitions are useful for characterizing limits of sequences of probability measures.

Definition A.7 Let $\mathscr{V}$ be a subset of the class of continuous and bounded functions on $\mathscr{X}$. We say that $\mathscr{V}$ is **separating** if for all $\mu, \nu \in \mathscr{P}(\mathscr{X})$, whenever $\int_{\mathscr{X}} g\, d\mu = \int_{\mathscr{X}} g\, d\nu$ for all $g \in \mathscr{V}$, we must have $\mu = \nu$. The class $\mathscr{V}$ is said to be **convergence determining** if for every sequence $\{\theta_n\} \subset \mathscr{P}(\mathscr{X})$ and $\theta \in \mathscr{P}(\mathscr{X})$, $\theta_n \Rightarrow \theta$ if and only if $\int_{\mathscr{X}} g\, d\theta_n \to \int_{\mathscr{X}} g\, d\theta$ for all $g \in \mathscr{V}$.

Note that if $\mathscr{V}$ is convergence determining, then it is separating. One of the basic results in the theory says that there exists a countable convergence determining class $\mathscr{V}$ of bounded uniformly continuous (in fact Lipschitz continuous) functions (cf. [126, Proposition 3.4.4]). Such a class can be given explicitly as follows. Let $\{x_n\}$ be a countable dense set in $\mathscr{X}$. For $i, j \in \mathbb{N}$, let $f_{ij}(x) \doteq 2(1 - jd(x, x_i)) \vee 0, x \in \mathscr{X}$. For a finite subset Λ of $\mathbb{N} \times \mathbb{N}$, let

$$g_\Lambda(x) \doteq \left(\sum_{(i,j)\in\Lambda} f_{ij}(x) \right) \wedge 1, \ x \in \mathscr{X}.$$

Then for each Λ, g_Λ is a bounded Lipschitz continuous function and $\{g_\Lambda : \Lambda \subset \mathbb{N} \times \mathbb{N}\}$ defines a countable convergence determining class. Indeed, suppose $\mu_n, \mu \in \mathscr{P}(\mathscr{X})$ satisfy $\int_{\mathscr{X}} g_\Lambda\, d\mu_n \to \int_{\mathscr{X}} g_\Lambda\, d\mu$ for every Λ, and let $G \subset \mathscr{X}$ be an open set. For $m \in \mathbb{N}$, define

$$\Lambda_m \doteq \{(i, j) : i, j \leq m \text{ and } B(x_i, 1/j) \subset G\}.$$

Then $h_m \doteq g_{\Lambda_m}$ satisfies $h_m \leq 1_G$, and $h_m \uparrow 1_G$ as $m \to \infty$. Thus

$$\liminf_{n\to\infty} \mu_n(G) \geq \lim_{n\to\infty} \int h_m d\mu_n = \int h_m d\mu.$$

Sending $m \to \infty$, we have $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$. Since G is an arbitrary open set, we have $\mu_n \Rightarrow \mu$.

A.2 Skorohod Representation Theorem

For the proof of the following theorem we refer the reader to [167, Theorem 3.30].

Theorem A.8 (SKOROHOD REPRESENTATION THEOREM) *Suppose* $\xi, \xi_1, \xi_2, \ldots$ *are random variables with values in some separable metric space* $(\mathscr{S}, \rho)$ *such that* $\xi_n \Rightarrow \xi$ *as* $n \to \infty$. *Then on some probability space* $(\Omega, \mathscr{F}, P)$, *there exist* $\mathscr{S}$-*valued random variables* $\eta, \eta_1, \eta_2, \ldots$ *such that the distribution of* ξ *is the same as the distribution of* η, *the distribution of* ξ_i *is same as the distribution of* η_i *for every* i, *and* $\eta_i \to \eta$ P-*a.s.*

A.3 Space of Finite Measures

For a Polish space $\mathscr{X}$, $\mathscr{M}(\mathscr{X})$ denotes the set of finite measures on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, i.e., the set of measures γ for which $\gamma(\mathscr{X}) < \infty$. Let $\{\theta_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathscr{M}(\mathscr{X})$. We say that $\{\theta_n\}$ **converges weakly** to θ if for each $g \in \mathscr{C}_b(\mathscr{X})$,

$$\lim_{n\to\infty} \int_{\mathscr{X}} g\, d\theta_n = \int_{\mathscr{X}} g\, d\theta.$$

We next introduce a metric on $\mathscr{M}(\mathscr{X})$ having properties analogous to those of $\mathscr{L}(\cdot,\cdot)$. For γ and θ in $\mathscr{M}(\mathscr{X})$, define

$$m(\gamma, \theta) \doteq \left[\gamma(\mathscr{X}) \wedge \theta(\mathscr{X})\right] \cdot \mathscr{L}\left(\frac{\gamma}{\gamma(\mathscr{X})}, \frac{\theta}{\theta(\mathscr{X})}\right) + |\gamma(\mathscr{X}) - \theta(\mathscr{X})|.$$

The convention is that if γ or θ equals the zero measure on $\mathscr{X}$, then the first term in this definition is 0. Properties of m are given in the next theorem. The straightforward proof is omitted.

Theorem A.9 *The quantity* $m(\gamma, \theta)$ *defines a metric on* $\mathscr{M}(\mathscr{X})$. *The weak convergence on* $\mathscr{M}(\mathscr{X})$ *is equivalent to convergence under this metric. With respect to this metric,* $\mathscr{M}(\mathscr{X})$ *is complete and separable.*

A.4 Space of Locally Finite Measures

For a locally compact Polish space $\mathscr{X}$, we denote by $\Sigma(\mathscr{X})$ the space of all measures ν on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ satisfying $\nu(K) < \infty$ for every compact $K \subset \mathscr{X}$. We endow $\Sigma(\mathscr{X})$ with the weakest topology such that for every $f \in \mathscr{C}_c(\mathscr{X})$ (the space of real continuous functions on $\mathscr{X}$ with compact support), the function $\nu \mapsto \langle f, \nu\rangle =$

$\int_{\mathscr{X}} f(u)\,\nu(du)$, $\nu \in \Sigma(\mathscr{X})$ is continuous. This topology can be metrized such that $\Sigma(\mathscr{X})$ is a Polish space. One metric that is convenient for this purpose is given in the following subsection.

A.4.1 Metric for Measures on a Locally Compact Polish Space

Let $\|f\|_\infty \doteq \sup_{x\in\mathscr{X}} |f(x)|$ and $\|f\|_L \doteq \sup_{x,y\in\mathscr{X}} |f(x) - f(y)|/d(x, y)$, where d denotes the metric on $\mathscr{X}$. According to [220, Theorem 9.5.21], for a locally compact set $\mathscr{X}$ there exists a sequence of open sets $\{O_j\}$ such that $\bar{O}_j \subset O_{j+1}$, each $\bar{O}_j$ is compact, and $\cup_{j=1}^{\infty} O_j = \mathscr{X}$. Let $\phi_j(x) \doteq [1 - d(x, O_j)] \vee 0$. Given any $\mu \in \Sigma(\mathscr{X})$, let $\mu^j \in \Sigma(\mathscr{X})$ be defined by

$$\left[d\mu^j/d\mu\right](x) = \phi_j(x). \tag{A.1}$$

Given $\mu, \nu \in \Sigma(\mathscr{X})$, let

$$\bar{d}(\mu, \nu) \doteq \sum_{j=1}^{\infty} 2^{-j} \left\|\mu^j - \nu^j\right\|_{BL},$$

where $\|\cdot\|_{BL}$ denotes the bounded Lipschitz norm [92, Chap. 11]

$$\left\|\mu^j - \nu^j\right\|_{BL} \doteq \sup_{f:\|f\|_\infty\le 1, \|f\|_L\le 1} \left[\int_{\mathscr{X}} f d\mu^j - \int_{\mathscr{X}} f d\nu^j\right].$$

It is straightforward to check that $\bar{d}(\mu, \nu)$ defines a metric under which $\Sigma(\mathscr{X})$ is a Polish space, and that convergence in this metric is essentially equivalent to weak convergence on each compact subset of $\mathscr{X}$. Specifically, $\bar{d}(\mu_n, \mu) \to 0$ if and only if for each $j \in \mathbb{N}$, $\mu_n^j \to \mu^j$ in the weak topology as finite nonnegative measures, i.e., for all $f \in \mathscr{C}_b(\mathscr{X})$,

$$\int_{\mathscr{X}} f d\mu_n^j \to \int_{\mathscr{X}} f d\mu^j.$$

A.4.2 Determining Convergence from a Countable Class

From the separability of $\mathscr{X}$ it follows that the space $\mathscr{C}_c(\mathscr{X})$ is separable in the uniform metric, from which it follows (see [167, Appendix A.2]) that there is a countable collection $\mathscr{J} \subset \mathscr{C}_c(\mathscr{X})$ such that for $\mu_n, \mu \in \Sigma(\mathscr{X})$, $\bar{d}(\mu_n, \mu) \to 0$ if and only if for every $f \in \mathscr{J}$,

$$\int_{\mathscr{X}} f d\mu_n \to \int_{\mathscr{X}} f d\mu.$$

As an immediate consequence of this fact we have the following result.

Lemma A.10 *Suppose that* $\{N_k\}_{k\in\mathbb{N}}$ *and* N *are* $\Sigma(\mathscr{X})$*-valued random variables defined on a probability space* $(\Omega, \mathscr{F}, P)$ *and that* $E|\langle g, N_k\rangle - \langle g, N\rangle| \to 0$ *for all* $g \in \mathscr{C}_c(\mathscr{X})$*. Then* $N_k \to N$ *in probability.*

A.4.3 Proof of Compactness of S_m^N Defined in Chap. 8

Fix $T < \infty$. Let $\mathscr{X}$ be a locally compact Polish space and let $\mathscr{X}_T = [0, T] \times \mathscr{X}$. Recall the function L_T defined in (8.17):

$$L_T(g) \doteq \int_{\mathscr{X}_T} \ell(g(t, x))\, \nu_T(dt \times dx).$$

Here $\ell : [0, \infty) \to [0, \infty)$ is defined by

$$\ell(r) \doteq r \log r - r + 1,\ \ r \in [0, \infty),$$

with the convention that $0 \log 0 = 0$.

Let $\mathbb{M} \doteq \Sigma(\mathscr{X}_T)$. For $m \in \mathbb{N}$, define

$$S_m^N \doteq \{g : \mathscr{X}_T \to [0, \infty) : L_T(g) \le m\}.$$

A function $g \in S_m^N$ can be identified with a measure $\nu_T^g \in \mathbb{M}$ according to $\nu_T^g(A) = \int_A g(s, x)\, \nu_T(ds \times dx)$, $A \in \mathscr{B}(\mathscr{X}_T)$. The following lemma shows that $\left\{\nu_T^g : g \in S_m^N\right\}$ is a compact subset of $\mathbb{M}$.

Lemma A.11 *For every* $m \in \mathbb{N}$, $\{\nu_T^g : g \in S_m^N\}$ *is a compact subset of* $\mathbb{M}$*.*

Proof We note that the metric $\bar{d}$ on $\mathbb{M}$ introduced in Sect. A.4.1, when $\mathscr{X}$ is replaced with $\mathscr{X}_T$, will be given as follows. There is a sequence of open sets $\left\{O_j,\ j \in \mathbb{N}\right\}$ such that $\bar{O}_j \subset O_{j+1}$, each $\bar{O}_j$ is compact, and $\cup_{j=1}^{\infty} O_j = \mathscr{X}_T$. Also, for $(t, x) \in \mathscr{X}_T$, let $\phi_j(t, x) = \left[1 - d((t, x), O_j)\right] \vee 0$, where d denotes the metric on $\mathscr{X}_T$. Given any $\mu \in \mathbb{M}$, let $\mu^j \in \mathbb{M}$ be defined by $\left[d\mu^j / d\mu\right](t, x) = \phi_j(t, x)$. Then given $\mu, \nu \in \mathbb{M}$, we have

$$\bar{d}(\mu, \nu) \doteq \sum_{j=1}^{\infty} 2^{-j} \left\|\mu^j - \nu^j\right\|_{BL},$$

where $\|\cdot\|_{BL}$ denotes the bounded Lipschitz norm on $\mathscr{M}_F(\mathscr{X}_T)$. Note that for $\mu_n, \mu \in \mathbb{M}$, we have $\bar{d}(\mu_n, \mu) \to 0$ if and only if for each $i \in \mathbb{N}$,

$$\int_{\mathscr{X}_T} f d\mu_n^i \to \int_{\mathscr{X}_T} f d\mu^i$$

for every continuous and bounded function on $\mathscr{X}_T$, where μ_n^i is defined according to $\left[d\mu_n^i/d\mu_n\right](t,x) = \phi_i(t,x)$.

Let $\mu_n \doteq \nu_T^{g_n}$. We first show that $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathbb{M}$ is relatively compact for every sequence $\{g_n\}_{n\in\mathbb{N}} \subset S^N$. For this, using a diagonalization method, it suffices to show that $\{\mu_n^i\} \subset \mathbb{M}$ is relatively compact for every i. Next, since $\{\mu_n^i\}$ are supported on the compact subset K^i of $\mathscr{X}_T$ given by the closure of $\{(t,x) : \phi_i(t,x) \neq 0\}$, to show that $\{\mu_n^i\} \subset \mathbb{M}$ is relatively compact, it suffices to show that $\sup_n \mu_n^i(\mathscr{X}_T) < \infty$. The last property will follow from the fact that $L_T(g_n) \leq m$ for all n and the superlinear growth of ℓ. Specifically, let $c \in (0,\infty)$ be such that $z \leq c(\ell(z)+1)$ for all $z \in [0,\infty)$. Then

$$\begin{aligned}\sup_{n\in\mathbb{N}} \mu_n^i(\mathscr{X}_T) &= \sup_{n\in\mathbb{N}} \int_{\mathscr{X}_T} \phi_i(t,x) g_n(t,x) \nu_T(dt \times dx) \\ &\leq \sup_{n\in\mathbb{N}} \int_{K^i} g_n(t,x)\nu_T(dt \times dx) \leq c(m + \nu_T(K^i)) < \infty.\end{aligned}$$

Next, suppose that along a subsequence (without loss of generality, also denoted by $\{n\}$), $\mu_n \to \mu$. We would like to show that μ is of the form ν_T^g, where $g \in S_m^N$. For this, we will use the lower semicontinuity property of relative entropy. The result holds trivially if $\mu = 0$. Suppose now $\mu \neq 0$. Then there exists $i_0 \in \mathbb{N}$ such that for all $i \geq i_0$, $\inf_{n\in\mathbb{N}} \nu_T^{g_n}(\bar{O}_i) > 0$. Introducing a slight notational inconsistency, let ν_T^i be defined by $\left[d\nu_T^i/d\nu_T\right](t,x) = \phi_i(t,x)$. For $i \geq i_0$, define

$$\begin{aligned} c^i &= \nu_T^i(\mathscr{X}_T), \quad \bar{\nu}_T^i = \nu_T^i/c^i, \\ b_n^i &= \mu_n^i(\mathscr{X}_T), \quad \bar{\mu}_n^i = \mu_n^i/b_n^i, \\ b^i &= \mu^i(\mathscr{X}_T), \quad \bar{\mu}^i = \mu^i/b^i. \end{aligned}$$

Then $\bar{\nu}_T^i$, $\bar{\mu}_n^i$, and $\bar{\mu}^i$ are probability measures, and

$$\begin{aligned} R(\bar{\mu}_n^i \| \bar{\nu}_T^i) &= \frac{1}{b_n^i} \int_{\mathscr{X}_T} \left[\log(g_n(t,x)) + \log\left(\frac{c^i}{b_n^i}\right)\right] g_n(t,x)\phi_i(t,x)\nu_T(dt \times dx) \\ &= \frac{1}{b_n^i} \int_{\mathscr{X}_T} \left[\ell(g_n(t,x)) + g_n(t,x) - 1\right] \phi_i(t,x)\nu_T(dt \times dx) + \log\left(\frac{c^i}{b_n^i}\right) \\ &\leq \frac{1}{b_n^i} m + 1 - \frac{c^i}{b_n^i} + \log\left(\frac{c^i}{b_n^i}\right). \end{aligned}$$

Since $\mu_n^i \to \mu^i$, we have $b_n^i \to b^i$. Hence by the lower semicontinuity property of relative entropy,

$$\begin{aligned} R(\bar{\mu}^i \| \bar{\nu}_T^i) &\le \liminf_{n\to\infty} R(\bar{\mu}_n^i \| \bar{\nu}_T^i) \\ &\le \liminf_{n\to\infty} \left[\frac{1}{b_n^i} m + 1 - \frac{c^i}{b_n^i} + \log\left(\frac{c^i}{b_n^i}\right)\right] \\ &\le \frac{1}{b^i} m + 1 - \frac{c^i}{b^i} + \log\left(\frac{c^i}{b^i}\right) \\ &< \infty. \end{aligned} \tag{A.2}$$

Thus μ^i is absolutely continuous with respect to ν_T^i. Define $g^i = d\mu^i/d\nu_T^i$ and $g = g^i$ on $\bar{O}_i$. It is easily checked that g is defined consistently and that $\mu = \nu_T^g$. Also, by a direct calculation,

$$R(\bar{\mu}^i \| \bar{\nu}_T^i) = \frac{1}{b^i} \int_{\mathscr{X}_T} \ell(g(t,x))\phi_i(t,x)\nu_T(dt \times dx) + 1 - \frac{c^i}{b^i} + \log\left(\frac{c^i}{b^i}\right).$$

Combining the last display with (A.2), we have $\int_{\mathscr{X}_T} \ell(g(t,x))\phi_i(t,x)\nu_T(dt \times dx) \le m$ for all i. Sending $i \to \infty$, we see that $g \in S_m^N$. The result follows. □

Appendix B
Stochastic Kernels

B.1 Regular Conditional Probabilities

Throughout this section, $(\Omega, \mathscr{F}, P)$ is a probability space and $\mathscr{Y}$ is a Polish space. Let Y be a random variable mapping Ω into $\mathscr{Y}$ and $\mathscr{G}$ a sub-σ-field of $\mathscr{F}$. A **regular conditional distribution for** Y **given** $\mathscr{G}$ is a map $(\omega, A) \mapsto \hat{P}(A|\mathscr{G})(\omega)$ from $\Omega \times \mathscr{Y}$ to $[0, 1]$ with the following properties.

(a) For each Borel subset B of $\mathscr{Y}$, the function mapping $\omega \in \Omega \mapsto \hat{P}(B|\mathscr{G})(\omega)$ is measurable with respect to $\mathscr{G}$.

(b) For each $\Gamma \in \mathscr{G}$ and Borel subset B of $\mathscr{Y}$,

$$P\{\Gamma \cap \{Y \in B\}\} = \int_{\Gamma} \hat{P}(B|\mathscr{G})(\omega)\, P(d\omega).$$

(c) For each $\omega \in \Omega$, $\hat{P}(dy|\mathscr{G})(\omega)$ is a probability measure on $\mathscr{Y}$.

The first two properties state that $\hat{P}(B|\mathscr{G})(\omega)$ is a version of the conditional probability $P\{Y \in B|\mathscr{G}\}(\omega)$ for every Borel set B. The issue in proving the existence of a regular conditional distribution is to show that a version of the conditional probability can be chosen to be a probability measure on $\mathscr{Y}$ for each fixed value of ω. According to part (a) of the following standard result, this is possible when $\mathscr{Y}$ is a Polish space. Part (b) states the useful property that P-a.s. for $\omega \in \Omega$, conditional expectations can be obtained by integrating with respect to regular conditional distributions. The theorem is proved in Theorems 10.2.2 and 10.2.5 in [92]. If X is a random variable mapping $(\Omega, \mathscr{F})$ into a measurable space $(\mathscr{V}, \mathscr{A})$ and $\mathscr{G}$ denotes the sub-σ-algebra of $\mathscr{F}$ generated by X, then we will write $\hat{P}(dy|X)(\omega)$ for $\hat{P}(dy|\mathscr{G})(\omega)$ and call it a **regular conditional distribution for** Y **given** X.

Theorem B.1 *Let Y be a random variable mapping Ω into $\mathscr{Y}$, $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$, and f a measurable function mapping $\mathscr{Y}$ into $\mathbb{R}$ such that $E\{|f(Y)|\} < \infty$. The following conclusions hold.*

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

(a) *A regular conditional distribution* $\hat{P}(dy|\mathscr{G})(\omega)$ *for* Y *given* $\mathscr{G}$ *exists. It is unique in the sense that if* $\hat{Q}(dy|\mathscr{G})(\omega)$ *also satisfies the definition, then the two distributions* $\hat{P}(dy|\mathscr{G})(\omega)$ *and* $\hat{Q}(dy|\mathscr{G})(\omega)$ *agree* P*-a.s. for* $\omega \in \Omega$.

(b) P*-a.s. for* $\omega \in \Omega$, f *is integrable with respect to* $\hat{P}(dy|\mathscr{G})(\omega)$ *and*

$$E\left[f(Y)|\mathscr{G}\right](\omega) = \int_{\mathscr{Y}} f(y)\, \hat{P}(dy|\mathscr{G})(\omega).$$

Now let $(\mathscr{V}, \mathscr{A})$ be a measurable space and X a random variable mapping Ω into $\mathscr{V}$. A **regular conditional distribution for** Y **given** $X = x$ is defined to be a quantity $\hat{P}(dy|X = x)$ taking values in $[0, 1]$ and having the following properties.

(a) For each Borel subset B of $\mathscr{Y}$, the function mapping $x \in \mathscr{X} \mapsto \hat{P}(B|X = x)$ is measurable.

(b) For each measurable subset A of $(\mathscr{V}, \mathscr{A})$ and Borel subset B of $\mathscr{Y}$,

$$P\{\{X \in A\} \cap \{Y \in B\}\} = \int_A \hat{P}(B|X = x)\, P\{X \in dx\}.$$

(c) For each $x \in \mathscr{X}$, $\hat{P}(dy|X = x)$ is a probability measure on $\mathscr{Y}$.

The first two properties state that $\hat{P}(B|X = x)$ is a version of the conditional probability, in that if $g(x) = \hat{P}(B|X = x)$, then $g(X) = P\{Y \in B|X\}$ a.s. The next result is an immediate consequence of Theorem B.1.

Theorem B.2 *Let* $(\mathscr{V}, \mathscr{A})$ *be a measurable space,* X *a random variable mapping* Ω *into* $\mathscr{V}$*, and* Y *a random variable mapping* Ω *into* $\mathscr{Y}$*. Then a regular conditional distribution* $\hat{P}(dy|X = x)$ *for* Y *given* $X = x$ *exists. It is unique in the sense that if* $\hat{Q}(dy|X = x)$ *also satisfies the definition, then for almost every* x*, the two measures* $\hat{P}(dy|X = x)$ *and* $\hat{Q}(dy|X = x)$ *agree with respect to the distribution of* X.

Proof A regular conditional distribution $\hat{P}(dy|X)$ is measurable with respect to the sub-σ-field generated by X, and so it is a measurable function of X, say $\varphi(X)$. The quantity $\varphi(x)$ is a regular conditional distribution for Y given $X = x$. The uniqueness follows from the uniqueness asserted in part (a) of Theorem B.1. □

B.2 Stochastic Kernels

Throughout this section, $\mathscr{X}$ and $\mathscr{Y}$ are Polish spaces and $(\mathscr{V}, \mathscr{A})$ is a measurable space. Let us recall the definition of a stochastic kernel, which was introduced in Sect. 1.4. Let $\tau(dy|x)$ be a family of probability measures on $\mathscr{Y}$ parametrized by $x \in \mathscr{V}$. We call $\tau(dy|x)$ a **stochastic kernel** on $\mathscr{Y}$ given $\mathscr{V}$ if for every Borel subset E of $\mathscr{Y}$, the function mapping $x \in \mathscr{V} \mapsto \tau(E|x) \in [0, 1]$ is measurable.

In order to establish a useful equivalent condition for a stochastic kernel, we need a preliminary fact given in the next lemma. It is a special case of Proposition 7.25 in [19].

Lemma B.3 *For $E \in \mathscr{B}(\mathscr{Y})$, define $f_E : \mathscr{P}(\mathscr{Y}) \to [0, 1]$ by $f_E(\theta) \doteq \theta(E)$. Then*

$$\mathscr{B}(\mathscr{P}(\mathscr{Y})) = \sigma\left[\bigcup_{E\in\mathscr{B}(\mathscr{Y})} f_E^{-1}(\mathscr{B}(\mathbb{R}))\right].$$

In other words, $\mathscr{B}(\mathscr{P}(\mathscr{Y}))$ is the smallest σ-algebra with respect to which f_E is measurable for every $E \in \mathscr{B}(\mathscr{Y})$.

Proof We write $\mathscr{G} \doteq \sigma[\cup_{E\in\mathscr{B}(\mathscr{Y})} f_E^{-1}(\mathscr{B}(\mathbb{R}))]$. To prove that $\mathscr{G} \subset \mathscr{B}(\mathscr{P}(\mathscr{Y}))$, we show that f_E is $\mathscr{B}(\mathscr{P}(\mathscr{Y}))$-measurable for every $E \in \mathscr{B}(\mathscr{Y})$, so that for every $A \in \mathscr{B}(\mathbb{R})$, we have $f_E^{-1}(A) \in \mathscr{B}(\mathscr{P}(\mathscr{Y}))$. Let

$$\mathscr{D} \doteq \{E \in \mathscr{B}_{\mathscr{Y}} : f_E \text{ is } \mathscr{B}(\mathscr{P}(\mathscr{Y}))\text{-measurable}\}.$$

For every closed set $F \in \mathscr{B}(\mathscr{Y})$ and real number α, the Portmanteau theorem (Theorem A.2) implies that the set $\{\theta \in \mathscr{P}(\mathscr{Y}) : \theta(F) \geq \alpha\}$ is closed. Hence $F \in \mathscr{D}$. It is now straightforward to verify using the Dynkin class theorem [126] that $\mathscr{D}$ equals $\mathscr{B}(\mathscr{Y})$ and thus that $\mathscr{G} \subset \mathscr{B}(\mathscr{P}(\mathscr{Y}))$. The proof that $\mathscr{B}(\mathscr{P}(\mathscr{Y})) \subset \mathscr{G}$ is based on a standard approximation argument. By definition of $\mathscr{G}$, the function

$$\alpha_\varphi : \theta \in \mathscr{P}(\mathscr{Y}) \longmapsto \int_{\mathscr{Y}} \varphi\, d\theta \in \mathbb{R}$$

is $\mathscr{G}$-measurable when $\varphi \doteq 1_E$ for every $E \in \mathscr{B}(\mathscr{Y})$; indeed, in this case, $\alpha_\varphi(\theta) = f_E(\theta)$. Thus α_φ is $\mathscr{G}$-measurable when φ is a $\mathscr{B}(\mathscr{Y})$-simple function. Since when $\varphi \in \mathscr{C}_b(\mathscr{Y})$ there exists a sequence of $\mathscr{B}(\mathscr{Y})$-simple functions $\{\varphi_n\}$ that are uniformly bounded below and satisfy $\varphi_n \uparrow \varphi$, the monotone convergence theorem implies that $\alpha_{\varphi_n} \uparrow \alpha_\varphi$. Thus α_φ is $\mathscr{G}$-measurable for every $\varphi \in \mathscr{C}_b(\mathscr{Y})$. For $\gamma \in \mathscr{P}(\mathscr{Y})$, $\varphi \in \mathscr{C}_b(\mathscr{Y})$, and $\varepsilon > 0$, we define

$$N(\gamma, \varphi, \varepsilon) \doteq \left\{\theta \in \mathscr{P}(\mathscr{Y}) : \left|\int_{\mathscr{Y}} \varphi\, d\theta - \int_{\mathscr{Y}} \varphi\, d\gamma\right| < \varepsilon\right\}.$$

Since $N(\gamma, \varphi, \varepsilon) = \alpha_\varphi^{-1}(\int_{\mathscr{Y}} \varphi\, d\gamma - \varepsilon, \int_{\mathscr{Y}} \varphi\, d\gamma + \varepsilon)$, it follows that $N(\gamma, \varphi, \varepsilon)$ is an element of $\mathscr{G}$, and since the class of sets $\{N(\gamma, \varphi, \varepsilon)\}$ forms an open subbase for $\mathscr{B}(\mathscr{P}(\mathscr{Y}))$, we conclude that $\mathscr{B}(\mathscr{P}(\mathscr{Y})) \subset \mathscr{G}$. This completes the proof of the lemma. □

The following result, taken from Proposition 7.26 in [19], gives a useful equivalent condition for a stochastic kernel. In the latter reference it is assumed that $(\mathscr{V}, \mathscr{A})$ is a Borel space. However, the proof applies without change when $(\mathscr{V}, \mathscr{A})$ is a measurable space.

Theorem B.4 *Let $\tau(dy|x)$ be a family of probability measures on $\mathscr{Y}$ parametrized by $x \in \mathscr{V}$. Then $\tau(dy|x)$ is a stochastic kernel if and only if the function mapping*

$x \in \mathscr{V} \longmapsto \tau(\cdot|x) \in \mathscr{P}(\mathscr{Y})$ is measurable, i.e., if and only if $\tau(\cdot|x)$ is a random variable mapping $\mathscr{V}$ into $\mathscr{P}(\mathscr{Y})$.

Proof We define $g : \mathscr{V} \to \mathscr{P}(\mathscr{Y})$ by $g(x) \doteq \tau(\cdot|x)$, and for $E \in \mathscr{B}(\mathscr{Y})$, we define $h_E : \mathscr{V} \to [0, 1]$ by $h_E(x) \doteq \tau(E|x)$. For $E \in \mathscr{B}(\mathscr{Y})$ we also recall $f_E : \mathscr{P}(\mathscr{Y}) \to [0, 1]$ defined in the previous lemma by $f_E(\theta) \doteq \theta(E)$. These mappings are related by $h_E = f_E \circ g$. The assertion of the theorem is that g is $\mathscr{A}$-measurable if and only if h_E is $\mathscr{A}$-measurable for every $E \in \mathscr{B}(\mathscr{Y})$. Lemma B.3 implies that f_E is $\mathscr{B}(\mathscr{P}(\mathscr{Y}))$-measurable for every $E \in \mathscr{B}(\mathscr{Y})$. Since $h_E = f_E \circ g$, it follows that if g is $\mathscr{A}$-measurable, then h_E is $\mathscr{A}$-measurable for every $E \in \mathscr{B}(\mathscr{Y})$. Conversely, if h_E is $\mathscr{A}$-measurable for every $E \in \mathscr{B}(\mathscr{Y})$, then again by Lemma B.3,

$$\begin{aligned} g^{-1}\left(\mathscr{B}(\mathscr{P}(\mathscr{Y}))\right) &= g^{-1}\left(\sigma\left[\bigcup_{E \in \mathscr{B}(\mathscr{Y})} f_E^{-1}(\mathscr{B}(\mathbb{R}))\right]\right) \\ &= \sigma\left[\bigcup_{E \in \mathscr{B}(\mathscr{Y})} g^{-1}\left(f_E^{-1}(\mathscr{B}(\mathbb{R}))\right)\right] \\ &= \sigma\left[\bigcup_{E \in \mathscr{B}(\mathscr{Y})} h_E^{-1}(\mathscr{B}(\mathbb{R}))\right] \subset \mathscr{A}. \end{aligned}$$

We conclude that g is $\mathscr{A}$-measurable. This completes the proof. □

B.3 A Stochastic Kernel Needed in Sect. 4.8.4

In Sect. 4.8.4, it was required that we find stochastic kernels $\gamma^i, i = 1, 2$, on $\mathbb{R}^d$ given $\mathbb{R}^d \times \mathbb{R}^d$ and on $\mathbb{R}^d$ given $\mathbb{R}^d$, respectively, such that for all $(\xi, \beta^1) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\beta^2 \in \mathbb{R}^d$,

$$R\left(\gamma^1(\cdot|\xi, \beta^1) \,\|\theta(\cdot|\xi)\right) = L(\xi, \beta^1) \quad \text{and} \quad \int_{\mathbb{R}^d} y\gamma^1(dy|\xi, \beta^1) = \beta^1$$

and

$$R\left(\gamma^2(\cdot|\beta^2) \,\|\rho_\sigma(\cdot)\right) = \frac{1}{2\sigma^2}\|\beta^2\|^2 \quad \text{and} \quad \int_{\mathbb{R}^d} y\gamma^2(dy|\beta^2) = \beta^2.$$

While part (g) of Lemma 4.16 can be directly applied to obtain γ^2, it does not directly give γ^1, since we may have $L(\xi, \beta^1) = \infty$ for some (ξ, β^1). Instead, we will mollify and obtain γ^1 as a limit. To simplify notation, we replace (ξ, β^1) by (x, β).

We first note that $L(x, \beta)$ is a lower semicontinuous function of $(x, \beta) \in \mathbb{R}^d \times \mathbb{R}^d$, and thus is measurable on $\mathbb{R}^d \times \mathbb{R}^d$. In particular, $\{(x, \beta) : L(x, \beta) = \infty\}$ is measurable. Fix a sequence $\{\varepsilon_n\}_{n \in \mathbb{N}}$ in $(0, 1)$ that converges to 0, and for $n \in \mathbb{N}$, define

$$H^n(x,\alpha) \doteq \log \int_{\mathbb{R}^d} \exp\langle\alpha, y\rangle\ \theta^n(dy|x),$$
$$\tilde{H}^n(x,\alpha) \doteq \log \int_{\mathbb{R}^d} \exp\langle\alpha, y\rangle\ [\theta(dy|x) + \varepsilon_n \rho_1(dy)],$$

where

$$\theta^n(dy|x) \doteq \frac{1}{1+\varepsilon_n}[\theta(dy|x) + \varepsilon_n \rho_1(dy)]$$

and ρ_1 is the d-dimensional Gaussian distribution with covariance I. For $n \in \mathbb{N}$ and x, α, and β in $\mathbb{R}^d$, we also introduce

$$L^n(x,\beta) \doteq \sup_{\alpha\in\mathbb{R}^d} \left[\langle\alpha,\beta\rangle - H^n(x,\alpha)\right].$$

According to part (g) of Lemma 4.16, for each $n \in \mathbb{N}$, there is a measurable $\alpha^n(x,\beta)$ such that with

$$\bar{\gamma}^n(dy|x,\beta) \doteq e^{\langle\alpha^n(x,\beta),y\rangle - H^n(x,\alpha^n(x,\beta))}\theta^n(dy|x),$$

we have

$$R\left(\bar{\gamma}^n(\cdot|x,\beta)\,\|\theta^n(\cdot|x)\right) = L^n(x,\beta) \quad \text{and} \quad \int_{\mathbb{R}^d} y\bar{\gamma}^n(dy|x,\beta) = \beta. \tag{B.1}$$

Since $H^n(x,\alpha) = \tilde{H}^n(x,\alpha) - \log(1+\varepsilon_n)$ and $\tilde{H}^n(x,\alpha) \geq H(x,\alpha)$, it follows that if $L(x,\beta) < \infty$, then for all $n \in \mathbb{N}$,

$$L^n(x,\beta) = \sup_{\alpha\in\mathbb{R}^d} \left[\langle\alpha,\beta\rangle - \tilde{H}^n(x,\beta)\right] + \log(1+\varepsilon_n) \leq L(x,\beta) + \log 2 < \infty. \tag{B.2}$$

Also, by Condition 4.3, for each x and α in $\mathbb{R}^d$,

$$\sup_{n\in\mathbb{N}} H^n(x,\alpha) < \infty. \tag{B.3}$$

For x and β in $\mathbb{R}^d$, define

$$\gamma^n(dy|x,\beta) \doteq \begin{cases} \bar{\gamma}^n(dy|x,\beta) & \text{if } L(x,\beta) < \infty, \\ \delta_\beta(dy) & \text{if } L(x,\beta) = \infty. \end{cases}$$

Then for each $n \in \mathbb{N}$, γ^n is a stochastic kernel on $\mathbb{R}^d$ given $\mathbb{R}^d \times \mathbb{R}^d$. We will show that for (x,β) such that $L(x,\beta) < \infty$, $\gamma^n(dy|x,\beta)$ converges to $\gamma(dy|x,\beta)$ that satisfies

$$R\left(\gamma(\cdot|x,\beta)\,\|\theta(\cdot|x)\right) = L(x,\beta) \quad \text{and} \quad \int_{\mathbb{R}^d} y\gamma(dy|x,\beta) = \beta, \tag{B.4}$$

which will complete the proof.

We therefore consider such (x,β). Suppose that n indexes a subsequence. It follows from (B.2), (B.1), and (B.3) that the relative entropies $R\left(\bar{\gamma}^n(\cdot|x,\beta)\,\|\theta^n(\cdot|x)\right)$ are uniformly bounded and that the moment-generating functions $H^n(x,\alpha)$ are uniformly bounded for each fixed α and x. The proof of Lemma 3.9 considered an analogous situation but without the n-dependence of the moment-generating function. However, with the uniform bound (B.3), the argument applies with only notational changes, and it shows that $\{\gamma^n(\cdot|x,\beta)\}_{n\in\mathbb{N}}$ is tight and uniformly integrable. Thus by letting n index a convergent subsubsequence with limit $\gamma(\cdot|x,\beta)$, we have

$$\int_{\mathbb{R}^d} y\gamma^n(dy|x,\beta) \to \int_{\mathbb{R}^d} y\gamma(dy|x,\beta).$$

It then follows from the lower semicontinuity of relative entropy (Lemma 2.4) that

$$\begin{aligned}
L(x,\beta) &\le R\left(\gamma(\cdot|x,\beta)\,\|\theta(\cdot|x)\right) \\
&\le \liminf_{n\to\infty} R\left(\bar{\gamma}^n(\cdot|x,\beta)\,\middle\|\,\theta^n(\cdot|x)\right) \\
&= \liminf_{n\to\infty} L^n(x,\beta) \\
&\le L(x,\beta),
\end{aligned}$$

where the last line is due to the fact that $L^n(x,\beta) \le L(x,\beta) - \log(1+\varepsilon_n)$ for all $n \in \mathbb{N}$. According to Lemma 2.4, $R(\cdot\,\|\cdot)$ is strictly convex in the first variable, which shows that $\gamma(\cdot|x,\beta)$ is the unique probability measure that satisfies (B.4). An argument by contradiction then shows that $\gamma^n(\cdot|x,\beta)$ converges to $\gamma(\cdot|x,\beta)$ along the entire sequence $n \in \mathbb{N}$. Since this is true for every (x,β) such that $L(x,\beta) < \infty$, this completes the proof. □

Appendix C
Further Properties of Relative Entropy

C.1 Proof of Part (e) of Lemma 2.4

We denote by Π the class of all finite measurable partitions of the Polish space $\mathscr{X}$. Part (e) of Lemma 2.4 states that for each γ and θ in $\mathscr{P}(\mathscr{X})$,

$$R(\gamma \| \theta) = \sup_{\pi \in \Pi} \sum_{A \in \pi} \gamma(A) \log \frac{\gamma(A)}{\theta(A)}, \tag{C.1}$$

where the summand equals 0 if $\gamma(A) = 0$ and equals ∞ if $\gamma(A) > 0$ and $\theta(A) = 0$. In addition, if A is any Borel subset of $\mathscr{X}$, then

$$R(\gamma \| \theta) \geq \gamma(A) \log \frac{\gamma(A)}{\theta(A)} - 1. \tag{C.2}$$

We first prove that for every finite measurable partition π of $\mathscr{X}$,

$$R(\gamma \| \theta) \geq \sum_{A \in \pi} \gamma(A) \log \frac{\gamma(A)}{\theta(A)}.$$

If $R(\gamma \| \theta) = \infty$, there is nothing to prove, so we assume that $R(\gamma \| \theta) < \infty$. In this case, $\gamma \ll \theta$, and setting $B \doteq \cup_{\{A \in \pi : \gamma(A) = 0\}} A$, we define for $m \in \mathbb{N}$ the bounded measurable function

$$\psi_m(x) \doteq \sum_{\{A \in \pi : \gamma(A) > 0\}} \left(\log \frac{\gamma(A)}{\theta(A)} \right) 1_A(x) - m \, 1_B(x).$$

The Donsker–Varadhan variational formula stated in part (a) of Lemma 2.1 implies that

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

$$R(\gamma\|\theta) \geq \int_{\mathscr{X}} \psi_m \, d\gamma - \log \int_{\mathscr{X}} e^{\psi_m} \, d\theta$$
$$= \sum_{\{A\in\pi:\gamma(A)>0\}} \gamma(A) \log \frac{\gamma(A)}{\theta(A)} - \log(1+e^{-m}\theta(B)).$$

This yields the desired formula, since $\lim_{m\to\infty} \log(1+e^{-m}\theta(B)) = 0$.

In order to complete the proof of equation (C.1), we determine a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ of finite measurable partitions of $\mathscr{X}$ having the property that

$$R(\gamma\|\theta) = \lim_{n\to\infty} \sum_{A\in\pi_n} \gamma(A) \log \frac{\gamma(A)}{\theta(A)}. \tag{C.3}$$

We carry this out via a standard technique, using unpublished notes of Barron [15]. If γ is not absolutely continuous with respect to θ, then the proof is straightforward. Indeed, in this case there exists a Borel subset A of $\mathscr{X}$ having the property that $\theta(A) = 0$ and $\gamma(A) > 0$. We obtain formula (C.3) by setting $\pi_n \doteq \{A, A^c\}$ for each $n \in \mathbb{N}$.

We now suppose that γ is absolutely continuous with respect to θ and let $f \doteq d\gamma/d\theta$. For each $n \in \mathbb{N}$, we then define π_n to be the finite measurable partition of $\mathscr{X}$ consisting of the disjoint Borel sets

$$A_{n,k} \doteq \begin{cases} \{x\in\mathscr{X} : \log f(x) \leq -\sqrt{n}\} & \text{if } k=-n, \\ \left\{x\in\mathscr{X} : \frac{k-1}{\sqrt{n}} < \log f(x) \leq \frac{k}{\sqrt{n}}\right\} & \text{if } k \in \{-n+1, -n+2, \ldots, n-1, n\}, \\ \{x\in\mathscr{X} : \log f(x) > \sqrt{n}\} & \text{if } k = n+1. \end{cases}$$

For $-n+1 \leq k \leq n+1$,

$$\gamma(A_{n,k}) = \int_{A_{n,k}} \exp(\log f) \, d\theta \geq \exp\left[\frac{k-1}{\sqrt{n}}\right] \theta(A_{n,k}). \tag{C.4}$$

The error in the approximation of the relative entropy by the sum over the partition π_n equals

$$R(\gamma\|\theta) - \sum_{A\in\pi_n} \gamma(A) \log \frac{\gamma(A)}{\theta(A)} = \sum_{k=-n}^{n+1} 1_{\{j:\gamma(A_{n,j})>0\}}(k) \int_{A_{n,k}} \log\left(f \, \frac{\theta(A_{n,k})}{\gamma(A_{n,k})}\right) d\gamma.$$

We now bound each term in this sum. For $-n+1 \leq k \leq n$ and $x \in A_{n,k}$, if $\gamma(A_{n,k}) > 0$, then from (C.4), the integrand satisfies

$$\log\left(f(x) \, \frac{\theta(A_{n,k})}{\gamma(A_{n,k})}\right) \leq \frac{k}{\sqrt{n}} - \frac{k-1}{\sqrt{n}} = \frac{1}{\sqrt{n}},$$

which implies that

$$\sum_{k=-n+1}^{n} 1_{\{j:\gamma(A_{n,j})>0\}}(k) \int_{A_{n,k}} \log\left(f \, \frac{\theta(A_{n,k})}{\gamma(A_{n,k})} \right) d\gamma \leq \frac{1}{\sqrt{n}} \sum_{k=-n+1}^{n} \gamma(A_{n,k}) \leq \frac{1}{\sqrt{n}}.$$

For $k = -n$ and $x \in A_{n,-n}$, if $\gamma(A_{n,-n}) > 0$, then the integrand satisfies

$$\log\left(f(x) \, \frac{\theta(A_{n,-n})}{\gamma(A_{n,-n})} \right) \leq \log\left(e^{-\sqrt{n}} \frac{1}{\gamma(A_{n,-n})} \right),$$

and so, since $s \log s \geq -e^{-1}$ for $s \in [0, \infty)$,

$$\int_{A_{n,-n}} \log\left(f \, \frac{\theta(A_{n,-n})}{\gamma(A_{n,-n})} \right) d\gamma \leq -\gamma(A_{n,-n}) \log\left(e^{\sqrt{n}} \gamma(A_{n,-n}) \right) \leq e^{-\sqrt{n}-1}.$$

Finally, for $k = n+1$, if $\gamma(A_{n,n+1}) > 0$, then (C.4) implies that $\theta(A_{n,n+1})/\gamma(A_{n,n+1}) \leq 1$. Thus

$$\int_{A_{n,n+1}} \log\left(f \, \frac{\theta(A_{n,n+1})}{\gamma(A_{n,n+1})} \right) d\gamma \leq \int_{A_{n,n+1}} (\log f) \, d\gamma = \int_{\{\log f > \sqrt{n}\}} (\log f) \, d\gamma.$$

Combining these inequalities yields

$$\begin{aligned} 0 &\leq R(\gamma \| \theta) - \sum_{A \in \pi_n} \gamma(A) \, \log \frac{\gamma(A)}{\theta(A)} \\ &\leq \frac{1}{\sqrt{n}} + e^{-\sqrt{n}-1} + \int_{\{\log f > \sqrt{n}\}} (\log f) \, d\gamma. \end{aligned}$$

If $R(\gamma \| \theta) < \infty$, then the integral in this inequality converges to 0 as $n \to \infty$, and thus

$$\lim_{n \to \infty} \sum_{A \in \pi_n} \gamma(A) \, \log \frac{\gamma(A)}{\theta(A)} = R(\gamma \| \theta).$$

Now assume that γ is absolutely continuous with respect to θ but that $R(\gamma \| \theta) = \infty$. For any Borel set B, if $\theta(B) = 0$, then $\gamma(B) = 0$ and $\gamma(B) \, \log[\gamma(B)/\theta(B)] = 0$, while if $\theta(B) > 0$, then since $s \, \log s \geq s - 1$ for $s \in [0, \infty)$, it follows that

$$\gamma(B) \log \frac{\gamma(B)}{\theta(B)} = \theta(B) \left[\frac{\gamma(B)}{\theta(B)} \, \log \frac{\gamma(B)}{\theta(B)} \right] \geq \theta(B) \left[\frac{\gamma(B)}{\theta(B)} - 1 \right] \geq -1. \tag{C.5}$$

Since $\{A_{n,k}, -n \leq k \leq n\}$ is a finite measurable partition of $\{\log f \leq \sqrt{n}\}$, similar estimates as in the case $R(\gamma \| \theta) < \infty$ yield

$$\int_{\{\log f \leq \sqrt{n}\}} (\log f) \, d\gamma - \sum_{k=-n}^{n} \gamma(A_{n,k}) \, \log \frac{\gamma(A_{n,k})}{\theta(A_{n,k})} \leq \frac{1}{\sqrt{n}} + e^{-\sqrt{n}-1}.$$

Thus

$$\sum_{A\in\pi_n} \gamma(A) \log \frac{\gamma(A)}{\theta(A)} \geq \int_{\{\log f \leq \sqrt{n}\}} (\log f)\, d\gamma - 1 - \frac{1}{\sqrt{n}} - e^{-\sqrt{n}-1}.$$

Since the right-hand side converges to $\infty = R(\gamma\|\theta)$ as $n \to \infty$, we have completed the proof of (C.3) and thus the proof of (C.1).

We now prove formula (C.2). Given A a Borel subset of $\mathscr{X}$, (C.1) yields for the finite measurable partition $\pi \doteq \{A, A^c\}$,

$$R(\gamma\|\theta) \geq \gamma(A) \log \frac{\gamma(A)}{\theta(A)} + \gamma(A^c) \log \frac{\gamma(A^c)}{\theta(A^c)}.$$

If $\theta(A^c) = 0$, then the last term in this display equals either 0 or ∞ depending on whether $\gamma(A^c)$ equals 0 or is positive. In either case, formula (C.2) follows. On the other hand, if $\theta(A^c) > 0$, then by (C.5),

$$R(\gamma\|\theta) \geq \gamma(A) \log \frac{\gamma(A)}{\theta(A)} - 1.$$

This is what we wanted to prove. The proof of part (e) of Lemma 2.4 is complete. □

C.2 Proof of Part (f) of Lemma 2.4

According to part (e) of Lemma 2.4,

$$\begin{aligned} R(\Delta_\psi \nu\|\Delta_\psi \mu) &= \sup_{\pi\in\Pi_{\mathscr{Y}}} \sum_{A\in\pi} \Delta_\psi \nu(A) \log \frac{\Delta_\psi \nu(A)}{\Delta_\psi \mu(A)} \\ &= \sup_{\pi\in\Pi_{\mathscr{Y}}} \sum_{A\in\pi} \nu(\psi^{-1}(A)) \log \frac{\nu(\psi^{-1}(A))}{\mu(\psi^{-1}(A))}, \end{aligned}$$

where $\Pi_{\mathscr{Y}}$ denotes the class of all finite measurable partitions of $\mathscr{Y}$. For each $\pi \in \Pi_{\mathscr{Y}}$, we define $\psi^{-1}(\pi) \doteq \{\psi^{-1}(A) : A \in \pi\}$, which is a finite measurable partition of $\mathscr{X}$. Thus, denoting by $\Pi_{\mathscr{X}}$ the class of all finite measurable partitions of $\mathscr{X}$, we have

$$\begin{aligned} R(\Delta_\psi \nu\|\Delta_\psi \mu) &= \sup_{\pi\in\Pi_{\mathscr{Y}}} \sum_{A\in\pi} \nu(\psi^{-1}(A)) \log \frac{\nu(\psi^{-1}(A))}{\mu(\psi^{-1}(A))} \\ &\leq \sup_{\pi\in\Pi_{\mathscr{X}}} \sum_{A\in\pi} \nu(A) \log \frac{\nu(A)}{\mu(A)} \\ &= R(\nu\|\mu). \end{aligned}$$

This proves the first part of the lemma. Finally, when ψ is one-to-one and ψ^{-1} is measurable, each $\pi_* \in \Pi_{\mathscr{X}}$ has the form $\psi^{-1}(\pi)$ for some $\pi \in \Pi_{\mathscr{Y}}$. In such a case, the inequality in the above display can be replaced by an equality. This completes the proof. □

C.3 Proof of Proposition 2.3

To prove part (a), we note that since k is bounded from below, the right-hand side of equation (2.2) is well defined. Since for $N \in \mathbb{N}$, $k \wedge N$ is bounded and measurable, part (a) of Proposition 2.2 implies that

$$-\log \int_{\mathscr{X}} e^{-(k\wedge N)}\, d\theta = \inf_{\gamma \in \mathscr{P}(\mathscr{X})} \left[R(\gamma \| \theta) + \int_{\mathscr{X}} (k \wedge N)\, d\gamma \right]$$
$$\leq \inf_{\gamma \in \mathscr{P}(\mathscr{X})} \left[R(\gamma \| \theta) + \int_{\mathscr{X}} k\, d\gamma \right].$$

Thus by the dominated convergence theorem,

$$-\log \int_{\mathscr{X}} e^{-k}\, d\theta = \lim_{N\to\infty} \left(-\log \int_{\mathscr{X}} e^{-(k\wedge N)}\, d\theta \right) \leq \inf_{\gamma \in \mathscr{P}(\mathscr{X})} \left[R(\gamma \| \theta) + \int_{\mathscr{X}} k\, d\gamma \right].$$

In order to prove that

$$\inf_{\gamma \in \mathscr{P}(\mathscr{X})} \left[R(\gamma \| \theta) + \int_{\mathscr{X}} k\, d\gamma \right] \leq -\log \int_{\mathscr{X}} e^{-k}\, d\theta,$$

we assume that $-\log \int_{\mathscr{X}} e^{-k}\, d\theta < \infty$, since otherwise, there is nothing to prove. Given $N \in \mathbb{N}$ and $\varepsilon > 0$, there exists a probability measure γ_N on $\mathscr{X}$ such that

$$R(\gamma_N \| \theta) + \int_{\mathscr{X}} (k \wedge N)\, d\gamma_N \leq \inf_{\gamma \in \mathscr{P}(\mathscr{X})} \left[R(\gamma \| \theta) + \int_{\mathscr{X}} (k \wedge N)\, d\gamma \right] + \varepsilon$$
$$= -\log \int_{\mathscr{X}} e^{-(k\wedge N)}\, d\theta + \varepsilon$$
$$\leq -\log \int_{\mathscr{X}} e^{-k}\, d\theta + \varepsilon < \infty.$$

Since k is bounded from below, it follows that $\sup_{N\in\mathbb{N}} R(\gamma_N \| \theta) < \infty$. This implies that the sequence $\{\gamma_N\}_{N\in\mathbb{N}}$ is relatively compact with respect to the weak topology [part (c) of Lemma 2.4]. Moreover, if γ_N converges along a subsequence to $\bar{\gamma}$, then by Lemma 2.5, for every bounded and measurable function ψ,

$$\lim_{N\to\infty}\int_{\mathscr{X}}\psi\,d\gamma_N=\int_{\mathscr{X}}\psi\,d\bar{\gamma}.$$

Thus along the convergent subsequence, we have

$$\begin{aligned}-\log\int_{\mathscr{X}}e^{-k}\,d\theta+\varepsilon &\geq \liminf_{N\to\infty}\left[R(\gamma_N\|\theta)+\int_{\mathscr{X}}(k\wedge N)\,d\gamma_N\right]\\ &\geq \liminf_{M\to\infty}\liminf_{N\to\infty}\left[R(\gamma_N\|\theta)+\int_{\mathscr{X}}(k\wedge N\wedge M)\,d\gamma_N\right]\\ &\geq \liminf_{M\to\infty}\left[R(\bar{\gamma}\|\theta)+\int_{\mathscr{X}}(k\wedge M)\,d\bar{\gamma}\right]\\ &\geq \left[R(\bar{\gamma}\|\theta)+\int_{\mathscr{X}}k\,d\bar{\gamma}\right]\\ &\geq \inf_{\gamma\in\mathscr{P}(\mathscr{X})}\left[R(\gamma\|\theta)+\int_{\mathscr{X}}k\,d\gamma\right],\end{aligned}$$

where the third inequality uses the lower semicontinuity of $\gamma\mapsto R(\gamma\|\theta)$ and Lemma 2.5, and the fourth inequality follows from the monotone convergence theorem. Sending $\varepsilon\to 0$ completes the proof of the variational formula under the assumption that k is bounded from below.

Next consider (b). Since the infimum is restricted to probability measures γ satisfying $R(\gamma\|\theta)<\infty$, the right-hand side of equation (2.3) is well defined. For $N\in\mathbb{N}$, $k\vee(-N)$ is bounded and measurable, and so by part (a) of Proposition 2.2,

$$\begin{aligned}-\log\int_{\mathscr{X}}e^{-[k\vee(-N)]}\,d\theta &= \inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta)+\int_{\mathscr{X}}[k\vee(-N)]\,d\gamma\right]\\ &\geq \inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta)+\int_{\mathscr{X}}k\,d\gamma\right].\end{aligned}$$

The monotone convergence theorem yields

$$\begin{aligned}-\log\int_{\mathscr{X}}e^{-k}\,d\theta &= \lim_{N\to\infty}\left(-\log\int_{\mathscr{X}}e^{-[k\vee(-N)]}\,d\theta\right)\\ &\geq \inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta)+\int_{\mathscr{X}}k\,d\gamma\right].\end{aligned}\tag{C.6}$$

Let $\varepsilon>0$ be given. In order to prove that

$$-\log\int_{\mathscr{X}}e^{-k}\,d\theta\leq\inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta)+\int_{\mathscr{X}}k\,d\gamma\right],$$

we assume that the right-hand side is less than ∞, for otherwise, there is nothing to prove. We choose a probability measure $\tilde{\gamma} \in \Delta(\mathscr{X})$ such that

$$R(\tilde{\gamma}\|\theta) + \int_{\mathscr{X}} k\, d\tilde{\gamma} \le \inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta) + \int_{\mathscr{X}} k\, d\gamma\right] + \varepsilon < \infty.$$

Then

$$\begin{aligned}-\log \int_{\mathscr{X}} e^{-[k\vee(-N)]}\, d\theta &= \inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta) + \int_{\mathscr{X}} [k\vee(-N)]\, d\gamma\right] \\ &\le R(\tilde{\gamma}\|\theta) + \int_{\mathscr{X}} [k\vee(-N)]\, d\tilde{\gamma}.\end{aligned}$$

Since $R(\tilde{\gamma}\|\theta) < \infty$, the monotone convergence theorem yields

$$\begin{aligned}-\log \int_{\mathscr{X}} e^{-k}\, d\theta &= \lim_{N\to\infty}\left(-\log \int_{\mathscr{X}} e^{-[k\vee(-N)]}\, d\theta\right) \\ &\le \lim_{N\to\infty}\left(R(\tilde{\gamma}\|\theta) + \int_{\mathscr{X}} [k\vee(-N)]\, d\tilde{\gamma}\right) \\ &= R(\tilde{\gamma}\|\theta) + \int_{\mathscr{X}} k\, d\tilde{\gamma} \\ &\le \inf_{\gamma\in\Delta(\mathscr{X})}\left[R(\gamma\|\theta) + \int_{\mathscr{X}} k\, d\gamma\right] + \varepsilon.\end{aligned} \tag{C.7}$$

Sending $\varepsilon \to 0$ completes the proof of (2.3) under the assumption that k is bounded from above.

Next we consider (c), where k is not assumed bounded below or above. To simplify notation, we prove the equivalent claim

$$\log \int_{\mathbb{R}^d} e^k\, d\theta = \sup_{\gamma\in\Delta(\mathbb{R}^d)}\left[\int_{\mathbb{R}^d} k\, d\gamma - R(\gamma\|\theta)\right]. \tag{C.8}$$

By assumption, there exists $\zeta > 0$ such that $\int_{\mathbb{R}^d} e^{\zeta\|x\|}\theta(dx) < \infty$. Recalling the inequality

$$ab \le e^a + \ell(b) \text{ for all } a, b \ge 0, \tag{C.9}$$

we have, for every $\gamma \in \Delta(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \|x\|\, d\gamma \le \frac{1}{\zeta}\int_{\mathbb{R}^d} e^{\zeta\|x\|}\theta(dx) + \frac{1}{\zeta}R(\gamma\|\theta) < \infty. \tag{C.10}$$

Thus the right side in (C.8) is well defined.

The first issue is to show that if the left side of (C.8) is ∞, then the right side is also. For $N \in \mathbb{N}$, let $f_N(x) = k(x) \wedge N$. Then f_N is bounded from above, and so the probability measure

$$\gamma_N(dx) = \frac{1}{Z_N} e^{f_N(x)} \theta(dx), \quad Z_N \doteq \int_{\mathbb{R}^d} e^{f_N(x)} \theta(dx)$$

is well defined, and by Fatou's lemma, $Z_N \to \infty$ as $N \to \infty$. With this choice of γ_N, since $f_N(x) \le k(x)$, the right side of (C.8) is bounded below by

$$\begin{aligned}
\left[\int_{\mathbb{R}^d} k \, d\gamma_N - R(\gamma_N \| \theta) \right] &= \left[\int_{\mathbb{R}^d} k \, d\gamma_N - \int_{\mathbb{R}^d} \log \left(\frac{e^{f_N}}{Z_N} \right) d\gamma_N \right] \\
&= \log Z_N + \int_{\mathbb{R}^d} [k - f_N] \, d\gamma_N \\
&\ge \log Z_N.
\end{aligned}$$

Letting $N \to \infty$ shows that the right side in (C.8) also equals ∞. If the left-hand side of (C.8) is finite, then $\log Z_N$ converges to that value, and in this case, sending $N \to \infty$ shows that

$$\log \int_{\mathbb{R}^d} e^k \, d\theta \le \sup_{\gamma \in \Delta(\mathbb{R}^d)} \left[\int_{\mathbb{R}^d} k \, d\gamma - R(\gamma \| \theta) \right].$$

We now argue the reverse inequality. It suffices to show that for all $\gamma \in \Delta(\mathbb{R}^d)$,

$$R(\gamma \| \theta) \ge \int_{\mathbb{R}^d} k \, d\gamma - \log \int_{\mathbb{R}^d} e^k \, d\theta. \tag{C.11}$$

For $M \in \mathbb{N}$, let

$$F_M(x) \doteq k(x) 1_{\{|k(x)| \le M\}} + \frac{M k(x)}{|k(x)|} 1_{\{|k(x)| > M\}}.$$

From part (a) of Lemma 2.4, we have

$$R(\gamma \| \theta) \ge \int_{\mathbb{R}^d} F_M d\gamma - \log \int_{\mathbb{R}^d} e^{F_M} d\theta. \tag{C.12}$$

Next note that by the dominated convergence theorem,

$$\lim_{M \to \infty} \int_{\mathbb{R}^d} F_M d\gamma = \int_{\mathbb{R}^d} k d\gamma$$

and

$$\lim_{M \to \infty} \int_{\{k<0\}} e^{F_M} d\theta = \int_{\{k<0\}} e^k d\theta.$$

Also by the monotone convergence theorem,

$$\lim_{M\to\infty}\int_{\{k\geq 0\}} e^{F_M}\,d\theta = \int_{\{k\geq 0\}} e^{k}\,d\theta.$$

Using the above three convergence properties and sending $M \to \infty$ in (C.12), we have (C.11), completing the proof of the reverse inequality. □

Appendix D
Martingales and Stochastic Integration

We begin with some basic definitions. Fix a finite-time horizon $T \in (0, \infty)$. Let $(\Omega, \mathscr{F}, P)$ be a probability space that is equipped with a **filtration** $\{\mathscr{F}_t\}_{0 \le t \le T}$, which means that $\mathscr{F}_s \subset \mathscr{F}_t \subset \mathscr{F}$ for all $0 \le s \le t \le T$. We will assume throughout that $\mathscr{F}$ is P-complete and the filtration satisfies the **usual conditions**, namely that the filtration is right continuous and for every $t \in [0, T]$, $\mathscr{F}_t$ contains all P-null sets in $\mathscr{F}$.

We say that a stochastic process $X = \{X(t)\}_{0 \le t \le T}$ on $(\Omega, \mathscr{F}, P)$ with values in some Polish space $\mathscr{E}$ is **RCLL** (resp. **LCRL**) if for every $\omega \in \Omega$, the map $t \mapsto X(t, \omega)$ from $[0, T]$ to $\mathscr{E}$ is right continuous on $[0, T)$ (resp. left continuous on $(0, T]$) and has left limits on $(0, T]$ (resp. has right limits on $[0, T)$). An $\mathscr{E}$-valued stochastic process X is said to be **$\mathscr{F}_t$-adapted** if for every $t \in [0, T]$, $X(t)$ is $\mathscr{F}_t$-measurable. It is said to be **$\mathscr{F}_t$-progressively measurable** if for every $t \in [0, T]$, the mapping $(s, \omega) \mapsto X(s, \omega)$ from $([0, t] \times \Omega, \mathscr{B}([0, t]) \otimes \mathscr{F}))$ to $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$ is measurable. Denote by $\mathscr{P}\mathscr{F}$ the σ-field on $[0, T] \times \Omega$ generated by the collection of all real $\mathscr{F}_t$-adapted LCRL processes (note that this is the same σ-field as that generated by the elementary functions or simple functions, as used, for example, in Definition 8.2). This σ-field is called the **$\mathscr{F}_t$-predictable σ-field**. For a Polish space $\mathscr{E}$, a $\mathscr{P}\mathscr{F}/\mathscr{B}(\mathscr{E})$-measurable map $X : [0, T] \times \Omega \to \mathscr{E}$ is referred to as an $\mathscr{E}$-valued **$\mathscr{F}_t$-predictable process**.

A $[0, T]$-valued random variable τ on $(\Omega, \mathscr{F})$ is said to be an **$\mathscr{F}_t$-stopping time** if $\{\tau \le t\} \in \mathscr{F}_t$ for every $t \in [0, T]$.

D.1 Martingales

Let $\{X(t)\}_{0 \le t \le T}$ be a real-valued $\mathscr{F}_t$-adapted process such that $E|X(t)| < \infty$ for every $t \in [0, T]$. Such a process is called an **$\mathscr{F}_t$-submartingale** (resp. an **$\mathscr{F}_t$-supermartingale**) if for all $0 \le s \le t \le T$, $E[X(t) \,|\mathscr{F}_s] \ge X(s)$ [resp. $E[X(t) \,|\mathscr{F}_s] \le X(s)$]. A process that is both an $\mathscr{F}_t$-submartingale and an $\mathscr{F}_t$-supermartingale is an **$\mathscr{F}_t$-martingale**. A martingale admits an RCLL modification, which is a

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

martingale with respect to the same filtration, and thus without loss of generality, we use RCLL modifications of martingales. A real-valued stochastic process X is called an $\mathscr{F}_t$**-local martingale** if there is a sequence of $\mathscr{F}_t$-stopping times τ^n increasing to T such that $\{X^{(n)}(t) \doteq X(t \wedge \tau_n)\}_{0 \le t \le T}$ is an $\mathscr{F}_t$-martingale for every n. A **locally square-integrable martingale** is defined in the analogous way. For two square-integrable $\mathscr{F}_t$-martingales X and Y with $X(0) = Y(0) = 0$, their **quadratic covariation**, denoted by $[X, Y]$, is the unique adapted RCLL process A with paths of bounded variation such that $A(0) = 0$, $XY - A$ is a martingale, and $\Delta A = \Delta X \Delta Y$, where for a real RCLL stochastic process Z, $\Delta Z(t) = Z(t) - Z(t-)$, $t \in [0, T]$. For such a process A, there is a unique decomposition $A = M + \tilde{A}$, where $M(0) = \tilde{A}(0) = 0$, M is a martingale, and $\tilde{A}$ is an $\mathscr{F}_t$-predictable process with paths of bounded variation. The process $\tilde{A}$ is called the **predictable quadratic covariation** of X and Y and is denoted by $\langle X, Y \rangle$. These definitions can be extended to local martingales (see [210]). When $X = Y$, these processes are sometimes denoted by $[X]$ and $\langle X \rangle$, and referred to as the **quadratic variation** (resp. **predictable quadratic variation**) of X. When X and Y are continuous, $[X, Y]$ is a continuous adapted process and hence predictable, in which case $[X, Y]$ coincides with $\langle X, Y \rangle$.

The following are some of the martingale inequalities used in this book. Discrete-time analogues of the first three are well known (see, for example, [173, Theorem 11.2] for the first two and [199] for the third). For the continuous time setting, see [172, Theorem 1.3.8] for the first two, [210, Theorem IV.4.48] for the third, and [180, Lemma 2.4] for the last.

Doob's submartingale inequality. For every nonnegative submartingale X, $c \in (0, \infty)$, and $t \in [0, T]$,

$$P\left[\sup_{0 \le s \le t} X(s) \ge c\right] \le \frac{1}{c} E[X(t)]. \tag{D.1}$$

Doob's maximal inequality. For every martingale M and $t \in [0, T]$,

$$E\left[\sup_{0 \le s \le t} |M(s)|^2\right] \le 4E[|M(t)|^2]. \tag{D.2}$$

Burkholder–Davis–Gundy inequality. For every $p \ge 1$, there exist $C_p \in (0, \infty)$ such that for every locally square-integrable martingale M with $M(0) = 0$ and $t \in [0, T]$,

$$E\left[\sup_{0 \le s \le t} |M(s)|^p\right] \le C_p E[[M, M](t)]^{p/2}. \tag{D.3}$$

Lenglart–Lepingle–Pratelli inequality. For $0 < p \le 2$, there exist $C_p \in (0, \infty)$ such that for every locally square-integrable martingale M with $M(0) = 0$ and $t \in [0, T]$,

$$E\left[\sup_{0 \le s \le t} |M(s)|^p\right] \le C_p E[\langle M, M \rangle(t)]^{p/2}. \tag{D.4}$$

The notion of quadratic variation can be extended to vector-valued martingales. Let $M = (M_1, \ldots, M_k)^T$ be an $\mathbb{R}^k$-valued stochastic process such that $\{M_i(t)\}_{0 \le t \le T}$ is an $\{\mathscr{F}_t\}$-martingale for each $i = 1, \ldots, k$. We refer to M as a k-dimensional $\{\mathscr{F}_t\}$-martingale. Let M and N be k-dimensional and r-dimensional $\{\mathscr{F}_t\}$-martingales, respectively. Then $\langle\langle M, N\rangle\rangle$ is the $(k \times r)$-dimensional stochastic process given by

$$(\langle\langle M, N\rangle\rangle(t))_{ij} \doteq \langle M_i, N_j\rangle_t, \ \ 1 \le i \le k, 1 \le j \le r, \ \ t \in [0, T].$$

The martingale inequalities above can be extended to k-dimensional martingales. In particular, the Burkholder–Davis–Gundy inequality for $p = 2$ and a k-dimensional $\{\mathscr{F}_t\}$-martingale M says that

$$E\left[\sup_{0 \le s \le t} \|M(s)\|^2\right] \le C_2 E[\mathrm{tr}(\langle\langle M, M\rangle\rangle(t))]. \tag{D.5}$$

D.2 Stochastic Integration

In this section we summarize the various types of stochastic integrals used in this book. We begin with the setting of d-dimensional Brownian motion.

D.2.1 Brownian Motion in $\mathbb{R}^d$

Let W be a d-dimensional $\mathscr{F}_t$-Brownian motion as introduced in Sect. 3.2. Let $\bar{\mathscr{A}}$ as in Definition 3.12 denote the collection of all $\mathbb{R}^d$-valued $\mathscr{F}_t$-progressively measurable processes $\{v(t)\}_{0 \le t \le T}$ that satisfy $E[\int_0^T \|v(t)\|^2 dt] < \infty$. Then the stochastic integral $M_v(t) = \int_0^t v(s) dW(s)$ (see [172, Chap. 3]) is a square-integrable continuous $\mathscr{F}_t$-martingale, and for $v_1, v_2 \in \bar{\mathscr{A}}$,

$$\langle M_{v_1}, M_{v_2}\rangle(t) = [M_{v_1}, M_{v_2}](t) = \int_0^t \langle v_1(s) v_2(s)\rangle ds.$$

A similar result holds when $d = \infty$. More precisely, let $\{\beta_i\}_{i=1}^{\infty}$ be a sequence of independent one-dimensional $\{\mathscr{F}_t\}$-Brownian motions. Let $f_i \in \bar{\mathscr{A}}$ (with $d = 1$) for each $i \in \mathbb{N}$, and suppose that $\sum_{i=1}^{\infty} E[\int_0^T |f_i(t)|^2 dt] < \infty$. Then

$$M(t) \doteq \sum_{i=1}^{\infty} \int_0^t f_i(s) d\beta_i(s), \ \ t \in [0, T]$$

is a continuous $\{\mathscr{F}_t\}$-martingale and $\langle M\rangle_t = \sum_{i=1}^{\infty} \int_0^t |f_i(s)|^2 ds$ for $t \in [0, T]$.

In considering discontinuous martingales, we will consider integrands v that instead of being progressively measurable, lie in the smaller class of predictable processes. In the setting of Brownian motions, there is not much difference between the two classes, since for every $v \in \bar{\mathscr{A}}$, there is a predictable $\tilde{v}$ such that $v = \tilde{v}$ a.s. $dt \otimes P$, and the stochastic integrals M_v and $M_{\tilde{v}}$ agree a.s.

D.2.2 Point Processes

A reference for the topic of this section is [159]. Let $(\Omega, \mathscr{F}, P)$ and $\{\mathscr{F}_t\}_{0\le t\le T}$ be as at the beginning of this appendix. Let $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{X}_T$, $\mathscr{Y}_T$, ν, $\bar{\nu}$, ν_T, $\bar{\nu}_T$ be as in Sect. 8.2.1. Also let $\bar{N}$ be a Poisson random measure (PRM) with respect to $\{\mathscr{F}_t\}$ on $\mathscr{Y}_T$ with intensity measure $\bar{\nu}_T$ (see Sect. 8.2.1). Let $\bar{\mathscr{A}}$ be as introduced below (8.19), and for $\varphi \in \bar{\mathscr{A}}$, N^{φ} is defined as in Sect. 8.2.1 through (8.16). We will also consider the compensated point processes $\bar{N}_c(ds \times dr) = \bar{N}(ds \times dr) - \bar{\nu}_T(ds \times dr)$ and $N^{\varphi}_c(ds \times dx) = N^{\varphi}(ds \times dx) - \varphi(s,x)\nu_T(ds \times dx)$. Let $\mathscr{PF}$ be the predictable σ-field associated with $\{\mathscr{F}_t\}$. For $t \in [0,T]$ and $\psi : [0,T] \times \Omega \times \mathscr{Y} \to \mathbb{R}$, that is, $(\mathscr{PF} \otimes \mathscr{B}(\mathscr{Y}))/\mathscr{B}(\mathbb{R})$ measurable and satisfying

$$E\int_{\mathscr{Y}_T} |\psi(s,y)|\bar{\nu}_T(ds \times dy) < \infty,$$

the stochastic integral

$$M_{\psi}(t) \doteq \int_{[0,t]\times\mathscr{Y}} \psi(s,y)\bar{N}_c(ds \times dy)$$

is well defined, and the stochastic process M_{ψ} is a martingale. Thus

$$E\int_{\mathscr{Y}_T} \psi(s,y)\bar{N}(ds \times dy) = E\int_{\mathscr{Y}_T} \psi(s,y)\bar{\nu}_T(ds \times dy). \tag{D.6}$$

If in addition

$$E\int_{\mathscr{Y}_T} \psi(s,y)^2\bar{\nu}_T(ds \times dy) < \infty,$$

then M_{ψ} is a square-integrable martingale with quadratic variation

$$[M_{\psi}](t) = \int_{[0,t]\times\mathscr{Y}} \psi(s,y)^2\bar{N}(ds \times dy).$$

For ψ_i, $i = 1, 2$, as above,

$$\langle M_{\psi_1}, M_{\psi_2}\rangle(t) = \int_{[0,t]\times\mathscr{Y}} \psi_1(s,y)\psi_2(s,y)\bar{\nu}_T(ds \times dy).$$

Similarly, if $\psi : [0,T] \times \Omega \times \mathscr{X} \to \mathbb{R}$ is $(\mathscr{PF} \otimes \mathscr{B}(\mathscr{X}))/\mathscr{B}(\mathbb{R})$-measurable and

$$E \int_{\mathscr{X}_T} (|\psi(s,x)| \vee |\psi(s,x)|^2) \varphi(s,x) \nu_T(ds \times dx) < \infty,$$

then the stochastic integral

$$M_\psi(t) \doteq \int_{[0,t]\times\mathscr{X}} \psi(s,x) N_c^\varphi(ds \times dx)$$

is well defined, and the stochastic process M_ψ is a square-integrable martingale with quadratic variation

$$[M_\psi](t) = \int_{[0,t]\times\mathscr{X}} \psi(s,x)^2 N^\varphi(ds \times dx).$$

For two such integrands ψ_1, ψ_2, we have

$$\langle M_{\psi_1}, M_{\psi_2}\rangle(t) = \int_{[0,t]\times\mathscr{X}} \psi_1(s,x)\psi_2(s,x)\varphi(s,x)\nu_T(ds \times dx).$$

D.2.3 Hilbert Space Valued Brownian Motion

Let $(\mathscr{H}, \langle\cdot,\cdot\rangle)$ be a real separable Hilbert space. Let Λ be a strictly positive symmetric trace class operator on $\mathscr{H}$. Let $\{W(t)\}_{0\le t\le T}$ be a Λ-Wiener process with respect to $\{\mathscr{F}_t\}_{0\le t\le T}$ as introduced in Definition 8.1. Also let $(\mathscr{H}_0, \langle\cdot,\cdot\rangle_0)$ be the Hilbert space introduced in Sect. 8.1, i.e., $\mathscr{H}_0 \doteq \Lambda\mathscr{H}$ and $\langle h,k\rangle_0 \doteq \langle \Lambda^{-1/2}h, \Lambda^{-1/2}k\rangle$. Let $\bar{\mathscr{A}}$ be the class of $\mathscr{H}_0$-valued $\mathscr{F}_t$-predictable processes v that satisfy

$$P\left\{\int_0^T \|v(s)\|_0^2 ds < \infty\right\} = 1,$$

as introduced below (8.2). Then for every $\psi \in \bar{\mathscr{A}}$ and $t \in [0,T]$, the stochastic integral $M_t \doteq \int_0^t \langle\psi(s), dW(s)\rangle_0$ is defined as in [69, Sect. 4.2]. Furthermore, M is a continuous $\{\mathscr{F}_t\}$-local martingale, which is a martingale if $E\int_0^T \|\psi(s)\|_0^2 ds < \infty$, in which case $\langle M\rangle_t = \int_0^t \|\psi(s)\|_0^2\, ds$ for $t \in [0,T]$.

D.2.4 Brownian Sheet

Let $\mathscr{O}$ be a bounded open subset of $\mathbb{R}^d$. Let $\{B(t,x),\ (t,x) \in [0,T]\times\mathscr{O}\}$ be a Brownian sheet on $(\Omega, \mathscr{F}, P)$ with respect to the filtration $\{\mathscr{F}_t\}$ as introduced in

Definition 11.5. Let $\bar{\mathscr{A}}$ be, as introduced below Definition 11.7, the class of all $\{\mathscr{F}_t\}$-predictable processes f such that $\int_{[0,T]\times\mathscr{O}} f^2(s,x)dsdx < \infty$ a.s. Then the stochastic integral $M_t(f) \doteq \int_{[0,t]\times\mathscr{O}} f(s,u)B(ds \times du)$, $t \in [0,T]$, is defined as in Chap. 2 of [243]. Furthermore, $\{M_t(f)\}$ is a continuous $\{\mathscr{F}_t\}$-local martingale, which is a martingale if $E\int_{[0,T]\times\mathscr{O}} f^2(s,x)dsdx < \infty$, in which case the quadratic variation is given by $\langle M(f)\rangle_t = \int_{[0,t]\times\mathscr{O}} f^2(s,x)dsdx$.

D.3 Girsanov's Theorem

In this section we summarize some variations of Girsanov's theorem, in addition to those already presented in Chap. 8, that are appealed to in this book. We begin with the classical setting of a finite dimensional Brownian motion. A proof can be found in [172, Sect. 3.5].

Theorem D.1 *Let W be a d-dimensional $\mathscr{F}_t$-Brownian motion and $\{v(t)\}_{0\le t\le T}$ a $\mathbb{R}^d$-valued $\mathscr{F}_t$-progressively measurable process that satisfies $E[\int_0^T \|v(t)\|^2dt] < \infty$. Suppose*

$$E\left[\exp\left\{\int_0^T v(s)\,dW(s) - \frac{1}{2}\int_0^T \|v(s)\|^2 ds\right\}\right] = 1.$$

Then the process

$$\tilde{W}(t) \doteq W(t) - \int_0^t v(s)ds,$$

$t \in [0,T]$, is an $\{\mathscr{F}_t\}$-Brownian motion on $(\Omega, \mathscr{F}, Q)$, where Q is the probability measure defined by

$$\frac{dQ}{dP} = \exp\left\{\int_0^T v(s)\,dW(s) - \frac{1}{2}\int_0^T \|v(s)\|^2 ds\right\}.$$

A similar result holds for $d = \infty$ (see [69, Theorem 10.14]). For that case, W is replaced by a sequence $\{\beta_i\}_{i=1}^\infty$ of independent one-dimensional $\{\mathscr{F}_t\}$-Brownian motions, v with a sequence $f_i \in \bar{\mathscr{A}}$ (with $\bar{\mathscr{A}}$ as in Definition 3.12 and $d = 1$) such that $E[\int_0^T \sum_{i=1}^\infty |f_i(t)|^2 dt] < \infty$, and the integrals $\int_0^T v(s)\,dW(s)$ and $\int_0^T \|v(s)\|^2 ds$ replaced by $\sum_{i=1}^\infty \int_0^T f_i(t)d\beta_i(t)$ and $\sum_{i=1}^\infty \int_0^T |f_i(t)|^2 dt$, respectively. We omit the precise statement.

Girsanov's theorem for a Brownian sheet takes the following form (see [206, Proposition 1.6].

Theorem D.2 *Let $\mathscr{O}$ be a bounded open subset of $\mathbb{R}^d$ and suppose that $\{B(t,x), (t,x) \in [0,T]\times\mathscr{O}\}$ is a Brownian sheet on $(\Omega,\mathscr{F},P)$ with respect to the filtration $\{\mathscr{F}_t\}$. Let f be $\{\mathscr{F}_t\}$-predictable in the sense of Definition 11.7 and satisfy $E\int_{[0,T]\times\mathscr{O}} f^2(s,x)ds\,dx < \infty$. Suppose that*

$$E\left[\exp\left\{\int_{[0,T]\times\mathscr{O}} f(s,u)B(ds\times du) - \frac{1}{2}\int_{[0,T]\times\mathscr{O}} f^2(s,x)dsdx\right\}\right] = 1.$$

Then the random field $\left\{\tilde{B}(t,x),\ (t,x)\in[0,T]\times\mathscr{O}\right\}$ *defined by*

$$\tilde{B}(t,x) \doteq B(t,x) - \int_0^t \int_{(-\infty,x]\cap\mathscr{O}} f(s,y)dyds$$

is a Brownian sheet with respect to $\{\mathscr{F}_t\}$ *on* $(\Omega,\mathscr{F},Q)$*, where* Q *is the probability measure defined by*

$$\frac{dQ}{dP} = \exp\left\{\int_{[0,T]\times\mathscr{O}} f(s,u)B(ds\times du) - \frac{1}{2}\int_{[0,T]\times\mathscr{O}} f^2(s,x)dsdx\right\}.$$

Finally, we present a version of Girsanov's theorem for systems with both Brownian and Poisson noise. We will not aim for maximum generality but rather state the result in the form in which it is used in the book. The result follows from the independence of the Brownian motion and PRM, and their corresponding versions of Girsanov's theorem (cf. [161, Theorem III.3.24]). We consider only a finite dimensional Brownian motion here; extensions to settings with an infinite dimensional Brownian motion can be written similarly.

Theorem D.3 *With notation and processes* W *and* N *that satisfy the conditions of Sect. 8.3, let* $u = (\psi,\varphi)\in\bar{\mathscr{A}}_b$*. Let*

$$\mathscr{E}_1^\varepsilon(t) \doteq \exp\left[\int_{\mathscr{X}_t\times[0,\infty)} 1_{[0,\varepsilon^{-1}\varphi(s,y)]}(r)\log(\tilde{\varphi}(s,y))\bar{N}(ds\times dy\times dr)\right.$$
$$\left. + \int_{\mathscr{X}_t\times[0,\infty)} 1_{[0,\varepsilon^{-1}\varphi(s,y)]}(r)(-\tilde{\varphi}(s,y)+1)\bar{\nu}_T(ds\times dy\times dr)\right],$$

$$\mathscr{E}_2^\varepsilon(t) \doteq \exp\left[-\frac{1}{\sqrt{\varepsilon}}\int_0^t \psi(s)dW(s) - \frac{1}{2\varepsilon}\int_0^t \|\psi(s)\|^2 ds\right],$$

and $\mathscr{E}^\varepsilon(t) \doteq \mathscr{E}_1^\varepsilon(t)\mathscr{E}_2^\varepsilon(t)$*. Then* $\{\mathscr{E}^\varepsilon(t)\}_{0\le t\le T}$ *is an* $\mathscr{F}_t$*-martingale, and consequently,*

$$\bar{Q}^\varepsilon(A) = \int_A \mathscr{E}^\varepsilon(T)dP,\ \ A\in\mathscr{F}$$

defines a probability measure on $(\Omega,\mathscr{F})$*. Furthermore,*

$$\left(W + \frac{1}{\sqrt{\varepsilon}}\int_0^\cdot \psi(s)ds,\ \varepsilon N^{\varphi/\varepsilon}\right)$$

under $\bar{Q}^\varepsilon$ *has the same probability law as* $(W,\varepsilon N^{1/\varepsilon})$ *under* P.

D.4 Criteria for Tightness

The next result, due to Aldous [3], considers tightness of a sequence of random variables $\{X^n\}_{n\in\mathbb{N}}$ with values in $\mathscr{D}([0, T] : \mathscr{E})$. For a proof, see [179, Theorem 2.7]. To simplify notation, it is assumed that all processes are defined on a common probability space $(\Omega, \mathscr{F}, P)$. Recall that τ is an $\mathscr{F}_t$-stopping time if $\{\tau \leq t\} \in \mathscr{F}_t$ for all $t \in [0, T]$.

Theorem D.4 *Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of processes with paths in $\mathscr{D}([0, T] : \mathscr{E})$ and let $\mathscr{F}^n_t$ be the σ-algebra generated by $\{X^n(s), 0 \leq s \leq t\}$. Suppose that $\{X^n(t)\}_{n\in\mathbb{N}}$ is tight for each rational $t \in [0, T]$, and that for every sequence of $\mathscr{F}^n_t$-stopping times $\{\tau_n\}$ such that $\tau_n \leq T$ and every sequence of nonnegative numbers $\{\delta_n\}$ converging to zero as $n \to \infty$,*

$$d(X^n(\tau_n + \delta_n), X^n(\tau_n)) \to 0$$

in probability as $n \to \infty$. Then $\{X^n\}_{n\in\mathbb{N}}$ is tight.

The theorem is also true if $\mathscr{D}([0, T] : \mathscr{E})$ is replaced by $\mathscr{C}([0, T] : \mathscr{E})$.

D.5 Diffeomorphic Properties of Solutions of Itô SDEs

The following is [178, Theorem 4.6.5].

Theorem D.5 *Suppose that the local characteristic (a, b) of a $\mathscr{C}^{k,\nu}$-Brownian motion $\{\Phi(t)\}_{t\geq 0}$ satisfies Condition 12.2 with some $\delta > \nu$. Then the solution of Itô's stochastic differential equation based on the Brownian motion Φ has a modification $\{\phi_{s,t}\}_{0\leq s\leq t\leq T}$ that is a forward stochastic flow of $\mathscr{C}^k$-diffeomorphisms.*

Appendix E
Analysis and Measure Theory

E.1 Measure Theory

The following result is well known. A proof can be found in [167].

Lemma E.1 *Let $\mathscr{X}_1, \mathscr{X}_2$ be Polish spaces and let X be an $\mathscr{X}_1$-valued Borel measurable map defined on some measurable space $(\Omega, \mathscr{F})$. Let $\mathscr{G} = \sigma\{X\}$. Suppose Y is an $\mathscr{X}_2$-valued Borel measurable map given on $(\Omega, \mathscr{F})$ that is $\mathscr{G}$-measurable. Then there is a Borel measurable map $g : \mathscr{X}_1 \to \mathscr{X}_2$ such that $Y = g(X)$.*

E.2 Gronwall's Inequality

Lemma E.2 (Gronwall's lemma) *Let f, g be measurable maps from $[0, \infty)$ to $[0, \infty)$. Suppose that for some $a \in [0, \infty)$,*

$$f(t) \le a + \int_0^t f(s)g(s)ds \quad \text{for all } t \in [0, \infty). \tag{E.1}$$

Also suppose that $\sup_{0 \le s \le t} f(s) < \infty$ for each fixed $t \in [0, \infty)$. Then

$$f(t) \le a e^{\int_0^t g(s)ds} \quad \text{for all } t \in [0, \infty).$$

Proof Fix $t \in [0, \infty)$. We assume without loss of generality that $\int_0^t g(s)ds < \infty$. Iterating (E.1) n times, we get

$$f(t) \le a + a \sum_{k=1}^{n} \int_0^t g(s_1) \int_0^{s_1} g(s_2) \cdots \int_0^{s_{k-1}} g(s_k)\, ds_k \cdots ds_1 + R_n(t), \tag{E.2}$$

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

where

$$R_n(t) = \int_0^t g(s_1) \int_0^{s_1} g(s_2) \cdots \int_0^{s_{n-1}} g(s_n) \int_0^{s_n} f(s_{n+1}) g(s_{n+1})\, ds_{n+1} ds_n \cdots ds_1.$$

Note that

$$R_n(t) \le \left(\int_0^t g(s) f(s) ds \right) \frac{\left(\int_0^t g(s) ds \right)^n}{n!} \le \left(\sup_{0 \le s \le t} f(s) \right) \left(\int_0^t g(s) ds \right) \frac{\left(\int_0^t g(s) ds \right)^n}{n!}.$$

Using the fact that $\int_0^t g(s) ds < \infty$, we see that $R_n(t) \to 0$ as $n \to \infty$. Sending $n \to \infty$ in (E.2), we have

$$\begin{aligned} f(t) &\le a + a \sum_{k=1}^{\infty} \int_0^t g(s_1) \int_0^{s_1} g(s_2) \cdots \int_0^{s_{k-1}} g(s_k) ds_k \cdots ds_1 \\ &= a \sum_{k=0}^{\infty} \frac{1}{k!} \left(\int_0^t g(s) ds \right)^k \\ &= a e^{\int_0^t g(s) ds}. \end{aligned}$$

This completes the proof of the lemma. □

E.3 Measurable Selection and Approximation of Measurable Functions

Let $(\mathscr{X}_2, \rho_2)$ be a complete and separable metric space and let $(\mathscr{X}_1, \rho_1)$ be a metric space. Suppose that for each $x \in \mathscr{X}_1$, $\Gamma_x \subset \mathscr{X}_2$. A **measurable selection** of $\{\Gamma_x\}_{x \in \mathscr{X}_1}$ is a $\mathscr{B}(\mathscr{X}_1) \backslash \mathscr{B}(\mathscr{X}_2)$-measurable function $f : \mathscr{X}_1 \to \mathscr{X}_2$ such that $f(x) \in \Gamma_x$ for every $x \in \mathscr{X}_1$. The following result is proved in Corollary 10.3 in Appendix 10 of [126].

Corollary E.3 *Suppose that if $y_n \in \Gamma_{x_n}$ for $n \in \mathbb{N}$ and $x_n \to x$ as $n \to \infty$, then $\{y_n\}_{n \in \mathbb{N}}$ has a limit point in Γ_x. Then a measurable selection of $\{\Gamma_x\}_{x \in \mathscr{X}_1}$ exists.*

We next state an approximation result (see [90, Theorem V.16a]).

Theorem E.4 *Let $\mathscr{X}$ be a Polish space and suppose $\lambda \in \mathscr{P}(\mathscr{X})$. Let $f : \mathscr{X} \to \mathbb{R}$ be a Borel measurable function. Then there is a sequence of continuous functions $\{f_j\}_{j \in \mathbb{N}}$, $f_j : \mathscr{X} \to \mathbb{R}$, such that*

$$f_j \to f \qquad \lambda\text{-a.e.}$$

as $j \to \infty$. If the function f is bounded in absolute value by B, then all the approximating functions can be taken to be bounded in absolute value by B as well.

E.4 Hilbert Spaces

The definitions in this section are taken from [66, 226].

A real vector space H is called an **inner product space** if for each pair $x, y \in H$ there is a real number $\langle x, y\rangle$ such that the following properties hold for every $x, y, z \in H$ and $\alpha \in \mathbb{R}$: (a) $\langle x, y\rangle = \langle y, x\rangle$, (b) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$, (c) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$, (d) $\langle x, x\rangle \geq 0$, (e) $\langle x, x\rangle = 0$ if and only if $x = 0$. Such a space can be normed by defining $\|x\|^2 \doteq \langle x, x\rangle$. If the resulting metric space is complete, we call it a **Hilbert space**.

A subset H_0 of a Hilbert space H is a said to be an **orthonormal set** if (a) for every $h \in H_0$, $\|h\| = 1$; (b) if $h_1, h_2 \in H_0$ are such that $h_1 \neq h_2$, then $\langle h_1, h_2\rangle = 0$. A maximal orthonormal set is said to be a **complete orthonormal system** (CONS). Every separable Hilbert space has a countable CONS.

For the rest of this section, H will be a separable Hilbert space with inner product $\langle\cdot, \cdot\rangle$ and norm $\|\cdot\|$. A linear mapping $A : H \to H$ is called a **bounded linear operator** on H if

$$\|A\| \doteq \sup_{x\in H:\|x\|\leq 1} \|Ax\| < \infty.$$

In this case, the mapping is continuous, and $\|A\|$ is called the norm of A.

For every bounded linear operator A on H, A^* is the unique bounded linear operator on H, referred to as the **adjoint** of A, with the property that

$$\langle Ax, y\rangle = \langle x, A^* y\rangle \text{ for all } x, y \in H.$$

A bounded linear operator A on H is called **self-adjoint** or **symmetric** if $A = A^*$. It is called **positive** if $\langle Ax, x\rangle \geq 0$ for all $x \in H$, and it is called **strictly positive** if $\langle Ax, x\rangle > 0$ for all nonzero $x \in H$. A positive and self-adjoint operator has a unique positive **square root** S, which is a positive operator satisfying $S^2 = A$.

Let A be a bounded linear operator on H. Let $\{e_i\}_{i\in\mathbb{N}}$ be a CONS in H. Define $\|A\|_2 \doteq [\sum_{i\in\mathbb{N}} \|Ae_i\|^2]^{1/2}$. It can be checked that $\|A\|_2$ thus defined does not depend on the choice of the CONS. We say that A is a **Hilbert–Schmidt operator** if $\|A\|_2 < \infty$, and we refer to $\|A\|_2$ as the **Hilbert–Schmidt norm** of A.

A bounded linear operator A on H is called a **trace class operator** if $A = BC$, where B, C are Hilbert–Schmidt operators. For such an operator, $\sum_{i\in\mathbb{N}} |\langle Ae_i, e_i\rangle| < \infty$ for every CONS $\{e_i\}$, and the sum $\sum_{i\in\mathbb{N}}\langle Ae_i, e_i\rangle$ is independent of the choice of the CONS. This quantity is referred to as the **trace** of the operator A.

Conventions and Standard Notation

Conventions. The following conventions are used throughout the book.

1. The infimum of the empty set is ∞.
2. $0 \log(0/x) = 0$ and $y \log(y/0) = \infty$ for $x \in [0, \infty)$ and $y \in (0, \infty)$.
3. Sigma fields on topological spaces will always be taken to be Borel σ-fields. A set in a Borel σ-field will be referred to as a Borel set. Mappings on a topological space are Borel measurable.
4. Two types of constants are used. Meaningful constants, such as Lipschitz constants, are denoted by uppercase letters, and constants that are used only in the course of a proof are set lowercase; they take values in $(0, \infty)$.

Standard notation, terminology, and abbreviations. The following standard notation is used throughout the book. A list of more specialized notation is given in the list of Specialized Symbols that follows this section.

General

$\mathscr{B}(S)$	the Borel σ-algebra on a Polish space S.
$(\mathscr{H}, \langle\cdot,\cdot\rangle)$	a real separable Hilbert space.
$\mathscr{P}(S)$	the probability measures on the measurable space $(S, \mathscr{F})$.
1_A	the indicator function of the set A.
δ_x	the probability measure with mass 1 at the point x.
$\gamma \ll \theta$	the measure γ is absolutely continuous with respect to θ.
$\frac{d\gamma}{d\theta}$	the Radon–Nikodym derivative of γ with respect to θ when the measure γ is absolutely continuous with respect to θ.
$\theta_n \Rightarrow \theta$	for $\{\theta_n\}_{n \in N} \cup \{\theta\} \subset \mathscr{P}(S)$, with S a metric space, $\int_S f d\theta_n \to \int_S f d\theta$ for all $f \in \mathscr{C}_b(S)$ and called weak convergence; for random variables $\{X_n\}_{n\in\mathbb{N}}$, X, $X_n \Rightarrow X$ means that the induced measures converge weakly, also called convergence in distribution.

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

$d(x, F)$	$\inf\{d(x, y) : y \in F\}$, the distance from the point x to the set F in a metric space with distance $d(\cdot, \cdot)$
$x \vee y$, $x \wedge y$	maximum (resp. minimum) of two real numbers x, y.
x^+, x^-	the positive part (resp. the negative part) of a real number x, equivalently $x \vee 0$ (resp. $(-x) \vee 0$).
$\lfloor x \rfloor$	integer part of x.
$\binom{a}{i}$	$\frac{\prod_{j=0}^{i-1}(a-j)}{i!}$, for $a \in \mathbb{R}$, $a \neq 0$ and $i \in \mathbb{N}$.
$B(x, \delta)$	$\{y : d(y, x) < \delta\}$, the open ball of radius δ centered at x in a metric space with distance $d(\cdot, \cdot)$
$\bar{B}$, B°, ∂B	closure, interior, and boundary of a set B, respectively.
$f_n \uparrow f$	for functions $f_n, f : S \to \mathbb{R}$, $f_n(x)$ increases monotonically $f(x)$ for all $x \in S$.
$f \circ g$	composition of two functions f and g.
$\theta \times \sigma$	for $\theta \in \mathscr{P}(\mathscr{X})$ and $\sigma \in \mathscr{P}(\mathscr{Y})$, $(\mathscr{X}, \mathscr{F})$ and $(\mathscr{Y}, \mathscr{G})$ measurable spaces, the unique probability measure on the product space $\mathscr{X} \times \mathscr{Y}$ that satisfies $[\theta \times \sigma](A \times B) = \theta(A)\sigma(B)$ for all $A \in \mathscr{F}$, $B \in \mathscr{G}$.
$\sigma(dy\|x)$	with $x \in \mathscr{V}$, $y \in \mathscr{Y}$, $\mathscr{V}$ a measurable space and $\mathscr{Y}$ a Polish space, a stochastic kernel on $\mathscr{Y}$ given $\mathscr{V}$: $\sigma(\cdot\|x) \in \mathscr{P}(\mathscr{Y})$ for all $x \in \mathscr{V}$ and $x \mapsto \sigma(A\|x)$ is measurable for every $A \in \mathscr{B}(\mathscr{Y})$.
$\theta \otimes \sigma$	for $\theta \in \mathscr{P}(\mathscr{X})$, $(\mathscr{X}, \mathscr{F})$ a measurable space, and $\sigma(dy\|x)$ a stochastic kernel on $\mathscr{Y}$ given $\mathscr{X}$, $\mathscr{Y}$ a Polish space, the unique probability measure on the product space $\mathscr{X} \times \mathscr{Y}$ obtained from a probability measure θ on $\mathscr{X}$ and a stochastic kernel $\sigma(dy\|x)$ on $\mathscr{Y}$ given $\mathscr{X}$ that satisfies $[\theta \times \sigma](A \times B) = \int_A \theta(dx)\sigma(B\|x)$ for all $A \in \mathscr{F}$, $B \in \mathscr{B}(\mathscr{Y})$.
$[\alpha]_i$, $[\alpha]_{j\|i}$	for $\alpha \in \mathscr{P}(\mathscr{X}_1 \times \cdots \times \mathscr{X}_k)$ with each $\mathscr{X}_i$ a Polish space and the product σ-algebra used, $[\alpha]_i$ is the marginal distribution on $\mathscr{X}_i$, and $[\alpha]_{j\|i}$ is the conditional distribution on $\mathscr{X}_j$ given a point in $\mathscr{X}_i$; $[\alpha]_{i_1,\ldots,i_m}$ and $[\alpha]_{j_i,\ldots,j_l\|i_1,\ldots,i_m}$ are defined in an analogous way.
Distribution of X	The probability measure induced by a random variable X on the space S in which X takes values, also called distribution induced by X.
$\mathscr{F}/\mathscr{G}$-measurable map	for $f : \mathscr{X} \to \mathscr{Y}$, $(\mathscr{X}, \mathscr{F})$ and $(\mathscr{Y}, \mathscr{G})$ measurable spaces, $\{x : f(x) \in B\} \in \mathscr{F}$ for all $B \in \mathscr{G}$.
Level set	for $F : S \to [0, \infty]$, a set of the form $\{x \in S : F(x) \leq M\}$
$\alpha \mapsto f(\alpha)$	the function on space S that maps points $\alpha \in S$ to $f(\alpha)$.
σ^T	the transpose of a vector or a matrix.
$\mathrm{tr}(A)$	$\sum_{i=1}^k a_{ii}$, the trace of a square $k \times k$ matrix $A = (a_{ij})_{i,j=1}^k$.

Spaces of functions (with S a metric space)

$\mathscr{AC}([0,T]:\mathbb{R}^d)$	the space of absolutely continuous functions from $[0,T]$ to $\mathbb{R}^d$, a subspace of $\mathscr{C}([0,T]:\mathbb{R}^d)$.
$\mathscr{AC}_x([0,T]:\mathbb{R}^d)$	the subset of $\mathscr{AC}([0,T]:\mathbb{R}^d)$ with initial condition $\phi(0)=x$.
$\mathscr{C}([0,T]:S)$	the space of continuous functions from $[0,T]$ to S with the supremum norm.
$\mathscr{C}_b(S)$	the space of bounded continuous functions from S to $\mathbb{R}$.
$\mathscr{C}_c(S)$	the space of continuous functions with compact support from S to $\mathbb{R}$.
$\mathscr{D}([0,T]:S)$	the space of functions that are right continuous with limits from the left for all $t\in(0,T]$, with the Skorohod metric.
$\mathscr{L}^1([0,T]:\mathbb{R}_+)$	the space of integrable functions from $[0,T]$ to $\mathbb{R}_+$.
$\mathscr{L}^2([0,T]:\mathbb{R}^d)$	the space of square integrable functions from $[0,T]$ to $\mathbb{R}^d$.
$\mathscr{L}^0([0,T]:\mathbb{R}_+)$	the space of Borel measurable functions from $[0,T]$ to $[0,\infty)$.
$\mathscr{M}_b(S)$	the space of bounded measurable functions from S to $\mathbb{R}$.

Controls and Spaces of Controls

In Chaps. 3 and 8–13, many different spaces of controls are used, and frequently several different spaces are given the same notation. In presenting representations for functionals of a finite dimensional Brownian motion in Sect. 3.2, spaces $\mathscr{A}$ and $\bar{\mathscr{A}}$ are introduced. These denote the collection of all $\mathscr{G}_t$-progressively [resp. $\mathscr{F}_t$-progressively] measurable processes $\{v(t)\}_{0\le t\le T}$ that satisfy the integrability condition $E[\int_0^T \|v(t)\|^2 dt]<\infty$. Here $\mathscr{F}_t$ is a general filtration, and $\mathscr{G}_t$ is the (augmentation of the) filtration generated by the Brownian motion. This section also introduces the subsets of $\mathscr{A}$ denoted by $\mathscr{A}_{b,M}$ and $\mathscr{A}_b$. The first consists of $v\in\mathscr{A}$ such that $\int_0^T \|v(t)\|^2 dt\le M$ a.s. and $\mathscr{A}_b=\cup_{M=1}^\infty \mathscr{A}_{b,M}$.

In Sect. 3.3, in the study of a process, the same notation is used for somewhat different spaces. Specifically, $\mathscr{A}$ is the collection of nonnegative predictable processes, while $\mathscr{A}_{b,M}$ is the subset of $\mathscr{A}$ consisting of φ such that $\int_0^T \ell(\varphi(s))ds\le M$ a.s. and for some $K\in(0,\infty)$ (possibly depending on φ), $K^{-1}\le\varphi\le K$ a.s. Once more, $\mathscr{A}_b=\cup_{M=1}^\infty\mathscr{A}_{b,M}$.

In Chap. 8, we begin with the general theory for continuous time processes. In Sect. 8.1, $\bar{\mathscr{A}}$ denotes the class of $\mathscr{H}_0$-valued $\mathscr{F}_t$-predictable processes v that satisfy

$$P\left\{\int_0^T \|v(s)\|_0^2 ds<\infty\right\}=1,$$

and $\mathscr{A}$ denotes the subset comprising those that are predictable with respect to $\{\mathscr{G}_t\}_{0\le t\le T}$, where $\mathscr{H}_0$ is a Hilbert space with norm $\|\cdot\|_0$ and $\mathscr{F}_t$ and $\mathscr{G}_t$ are similar to their counterparts in Chap. 3. Also, $\mathscr{A}_{b,M}$ consists of $v\in\mathscr{A}$ such that $\int_0^T \|v(s)\|_0^2 ds\le M$ and $\mathscr{A}_b\doteq\cup_{M\in\mathbb{N}}\mathscr{A}_{b,M}$. Also, $\mathscr{A}_s$ denotes the subset of $\mathscr{A}_b$ consisting of all simple processes. The spaces $\bar{\mathscr{A}}_{b,M}$ [resp. $\bar{\mathscr{A}}_b$] are defined exactly like $\mathscr{A}_{b,M}$ [resp. $\mathscr{A}_b$], except that $\{\mathscr{G}_t\}$ is replaced by $\{\mathscr{F}_t\}$.

In Sect. 8.2, we turn to a representation for a Poisson random measure (PRM). In this section, $\mathscr{A}$ is the class of all $(\mathscr{PF} \otimes \mathscr{B}(\mathscr{X}))\backslash\mathscr{B}[0,\infty)$-measurable maps $\varphi : \mathscr{X}_T \times \bar{\mathbb{M}} \to [0,\infty)$. Here $\mathscr{X}$ is the point space associated with the PRM, and $\mathscr{PF}$ is the predictable σ-field. Also,

$$\begin{aligned}\mathscr{A}_{b,M} \doteq \{\varphi \in \mathscr{A} : L_T(\varphi) \le M \text{ a.e. and for some } n \in \mathbb{N},\ n \ge \varphi(t,x,\omega) \ge 1/n \\ \text{and } \varphi(t,x,\omega) = 1 \text{ if } x \in K_n^c, \text{ for all } (t,\omega) \in [0,T] \times \bar{\mathbb{M}}\},\end{aligned} \tag{E.3}$$

where

$$L_T(\varphi)(\omega) \doteq \int_{\mathscr{X}_T} \ell(\varphi(t,x,\omega))\, \nu_T(dt \times dx),\ \ \omega \in \bar{\mathbb{M}}$$

and $\{K_n\}_{n\in\mathbb{N}}$ is an increasing sequence of compact subsets of $\mathscr{X}$ such that $\cup_{n=1}^{\infty} K_n = \mathscr{X}$. As before, $\mathscr{A}_b = \cup_{M=1}^{\infty}\mathscr{A}_{b,M}$. Once more, we let $\bar{\mathscr{A}}_{b,M}$, $\bar{\mathscr{A}}$, and $\bar{\mathscr{A}}_b$ denote the analogous spaces of controls when the canonical filtration $\{\mathscr{G}_t\}$ is replaced by $\{\mathscr{F}_t\}$. In this section, we also consider simple processes. A process $\varphi \in \mathscr{A}_{b,M}$ is in the set $\mathscr{A}_{s,M}$ if the following holds. There exist $n, \ell, n_1, \ldots, n_\ell \in \mathbb{N}$; a partition $0 = t_0 < t_1 < \cdots < t_\ell = T$; for each $i = 1, \ldots, \ell$, a disjoint measurable partition E_{ij} of K_n, $j = 1, \ldots, n_i$; $\mathscr{G}_{t_{i-1}}$-measurable random variables X_{ij}, $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$, such that $1/n \le X_{ij} \le n$; and

$$\varphi(t,x,\bar{m}) = 1_{\{0\}}(t) + \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} 1_{(t_{i-1},t_i]}(t) X_{ij}(\bar{m}) 1_{E_{ij}}(x) + 1_{K_n^c}(x) 1_{(0,T]}(t). \tag{E.4}$$

We define $\mathscr{A}_s \doteq \cup_{M=1}^{\infty}\mathscr{A}_{s,M}$.

In Sect. 8.3, we consider a representation for functionals of both PRM and Brownian motion. This representation involves both types of controls appearing in Sects. 8.1 and 8.2, and therefore we need to modify the notation. We denote by $\bar{\mathscr{A}}^W$ and $\bar{\mathscr{A}}_b^W$ the collections of controls for the Wiener process that were denoted by $\bar{\mathscr{A}}$ and $\bar{\mathscr{A}}_b$ in Sect. 8.1, and by $\bar{\mathscr{A}}^N$, $\bar{\mathscr{A}}_b^N$ the controls for the PRM that were denoted by $\bar{\mathscr{A}}$ and $\bar{\mathscr{A}}_b$ in Sect. 8.2. Similarly, the classes $\bar{\mathscr{A}}_{b,M}^N$ and $\bar{\mathscr{A}}_{b,M}^W$, which give uniform (in ω) bounds, are defined as they were in Sects. 8.1 and 8.2, respectively. Also we let $\bar{\mathscr{A}}_{b,M} \doteq \bar{\mathscr{A}}_{b,M}^W \times \bar{\mathscr{A}}_{b,M}^N$, $\bar{\mathscr{A}}_b \doteq \bar{\mathscr{A}}_b^W \times \bar{\mathscr{A}}_b^N$ and $\bar{\mathscr{A}} \doteq \bar{\mathscr{A}}^W \times \bar{\mathscr{A}}^N$.

The notation in Chaps. 9 and 10 for control spaces is same as that in Sect. 8.3, since we work here with systems that have both types of noise terms. Section 9.2.2, which studies a moderate deviation principle, introduces also a new specialized type of control space $\mathscr{U}_{n,+}^{\varepsilon}$ (see (9.8)) that is the class of controls for both types of noise for which the cost scales proportionally with $a(\varepsilon)^2$.

In Chap. 11, we consider systems with different types of infinite dimensional Brownian motions, and therefore the superscript W in the notation of control spaces is dropped. In considering Brownian sheet-driven systems, we consider the class of control $\bar{\mathscr{A}}$ analogous to those in Sect. 8.1 as the class of all $\{\mathscr{F}_t\}$-predictable processes f such that $\int_{[0,T]\times O} f^2(s,x)ds\,dx < \infty$ a.s. Here predictable processes are functions of (t,x,ω) (see Definition 11.7). Classes $\mathscr{A}_b$, $\mathscr{A}$ and $\bar{\mathscr{A}}_b$ are defined similarly. In this chapter, we also use controls associated with a Hilbert space valued Brownian motion with $\mathscr{H}_0 = l_2$. The spaces of these controls, as in Sect. 8.1, are denoted once more by $\mathscr{A}_b$, $\mathscr{A}$, $\bar{\mathscr{A}}$, and $\bar{\mathscr{A}}_b$. It is made clear at each place they appear which space is intended.

Chapter 12 uses the notation $\bar{\mathscr{A}}$ and $\bar{\mathscr{A}}_b$ (and $\bar{\mathscr{A}}_{b,M}$) for controls as in Sect. 8.1 with $\mathscr{H}_0 = l_2$.

In Chap. 13, we consider systems driven by a PRM, and therefore in denoting spaces of controls, we drop the superscript N. Thus the space $\mathscr{A}$ and its variants are as in Sect. 8.2. In studying a moderate deviation principle, the space $\mathscr{U}^{\varepsilon}_{n,+}$, which was introduced in Sect. 9.2.2, also makes an appearance in Sect. 13.3.2.

Together with the spaces of random controls such as $\mathscr{A}$, we also use many spaces of deterministic controls, employing once again the same notation for several different spaces. In Sect. 3.2, where we consider representations for a k-dimensional Brownian motion, the notation S_M is used for the space

$$\left\{\phi \in \mathscr{L}^2([0,T]:\mathbb{R}^k) : \int_0^T \|\phi(s)\|^2\, ds \le M\right\},$$

while in Sect. 3.3, in the study of a process, we have

$$S_M = \left\{\phi \in \mathscr{L}^0([0,T]:\mathbb{R}_+) : \int_0^T \ell(\phi(s))ds \le M\right\}.$$

In Sect. 8.1, where we consider a Hilbert space valued Brownian motion, we use the notation S_M for the space

$$\left\{u \in \mathscr{L}^2([0,T]:\mathscr{H}_0) : \int_0^T \|u(s)\|_0^2 ds \le M\right\}.$$

In Chap. 9, where we consider systems that have both types of noise terms, we need to distinguish the two types of control spaces. The space S_n from Sect. 8.1 is denoted here by S_n^W, and we define

$$S_n^N \doteq \left\{g : \mathscr{X}_T \to [0,\infty) : L_T^N(g) \le n\right\},$$

where L_T^N is as in (9.1). We define $S_n \doteq S_n^W \times S_n^N$ and $S \doteq \cup_{n\in\mathbb{N}} S_n$. This chapter also uses two other specialized deterministic control spaces. The first corresponds to controls that hit a target ϕ for a given z in the abstract large deviation principle of Sect. 9.2.1, namely

$$S^{\mathscr{G}}_{z,\phi} \doteq \left\{(f,g) \in S : \phi = \mathscr{G}^0(z, \int_0^{\cdot} f(s)ds, \nu_T^g)\right\},$$

while the second is a similar space in the abstract moderate deviation principle of Sect. 9.2.2,

$$S_{z,\eta}^{\mathscr{K}} \doteq \{q = (f_1, f_2) \in \mathscr{L}^2 : \eta = \mathscr{K}^0(z, q)\},$$

where $\mathscr{L}^2$ is introduced below (9.8).

Section 9.2.2 introduces three additional spaces of controls that are needed for the proof of the moderate deviation principle. These are

$$S_{n,+}^{N,\varepsilon} \doteq \{g : \mathscr{X}_T \to \mathbb{R}_+ \text{ such that } L_T^N(g) \leq na^2(\varepsilon)\},$$
$$S_n^{N,\varepsilon} \doteq \{f : \mathscr{X}_T \to \mathbb{R} \text{ such that } f = (g-1)/a(\varepsilon), \text{ with } g \in S_{n,+}^{N,\varepsilon}\},$$

and

$$\hat{S}_n \doteq \{(f_1, f_2) \in \mathscr{L}^2 : \|f_1\|_{W,2}^2 + \|f_2\|_{N,2}^2 \leq n\},$$

where $a(\varepsilon)$ is the scaling sequence in (9.6), and the norms $\|\cdot\|_{W,2}$ and $\|\cdot\|_{N,2}$ are introduced below (9.8).

Chapter 10 uses the same notation as Chap. 9 for the various spaces of deterministic controls.

In Chap. 11, where we consider systems with different types of infinite dimensional Brownian motions, the superscript W in the notation S_n^W is dropped. In particular, either, the space S_n denotes the space in Sect. 8.1 with $\mathscr{H}_0 = l_2$, or it is the space

$$\left\{\phi \in \mathscr{L}^2([0,T] \times O) : \int_{[0,T]\times O} \phi^2(s,r)dsdr \leq n\right\},$$

where O is the open set from Sect. 11.1. The precise space that is being referred to is clear from the context.

In Chap. 12, S_n is the space in Sect. 8.1 with $\mathscr{H}_0 = l_2$.

The spaces of Sect. 9.2.2 appear in Sect. 13.3.2 once more. However, since there is no Brownian motion in the dynamics, the spaces and notations are slightly different. Specifically,

$$\hat{S}_n \doteq \left\{f = \{f_i\}_{i=1}^K : f_i \in \mathscr{L}^2([0,1] \times \mathbb{R}_+) \text{ and } \sum_{i=1}^K \int_{[0,1]\times\mathbb{R}_+} f_i^2(s,y)dyds \leq n\right\}$$

and

$$S_{n,+}^{\varepsilon} \doteq \left\{g = \{g_i\}_{i=1}^K : g_i : [0,1] \times \mathbb{R}_+ \to \mathbb{R}_+ \right.$$
$$\left. \text{and} \sum_{i=1}^K \int_{[0,1]\times\mathbb{R}_+} \ell(g_i(s,y))dyds \leq na^2(\varepsilon)\right\}.$$

Norms and Distances

$\|x\|$	$(\sum_{i=1}^d x_i^2)^{1/2}$ for $x \in \mathbb{R}^d$.
$\|F\|_\infty$	$\sup_{x\in S} \|F(x)\|$ for $F : S \to \mathbb{R}^d$.

$\|F\|_{\infty,T}$	$\sup_{0\le t\le T}\|F(s)\|$ for $F:[0,T]\to\mathbb{R}^d$.
$\|\gamma\|_{\mathrm{TV}}$	for a signed measure γ on a measurable space $(S,\mathscr{F})$, $\sup\left\|\int_S f(x)\,\gamma(dx)\right\|$, where the supremum is over $f\in\mathscr{M}_b(S)$ with $\|f\|_\infty\le 1$, called the total variation norm of γ.
$d_{\mathrm{BL}}(\nu_1,\nu_2)$	for probability measures ν_1 and ν_2 on a Polish space $(\mathscr{X},d)$, $\sup\left\|\int_{\mathscr{X}} f(x)\nu_1(dx)-\int_{\mathscr{X}} f(x)\nu_2(dx)\right\|$, where the supremum is over f with $\|f\|_\infty\le 1$ and $\|f(x)-f(y)\|\le d(x,y)$ for all $x,y\in\mathscr{X}$, called the Dudley metric or bounded-Lipschitz metric.
$\|f\|_1$	for measurable f on a measure space $(S,\mathscr{F},\lambda)$, $\int_S \|f\|\,d\lambda$, called the $\mathscr{L}^1$-norm.
$\|\Lambda\|$	for a bounded linear operator Λ on a Hilbert space $\mathscr{H}$, $\sup_{h\in\mathscr{H}:\|h\|=1}\|\Lambda h\|$, called the operator norm.
$\|\psi\|_\alpha$	for $\alpha\in(0,1)$ and $\psi:S\to\mathbb{R}$, (S,d) a metric space, $\sup\{\|\psi(x)-\psi(y)\|^\alpha/d(x,y),\,x,y\in S\}$, called the α-Hölder norm.

Abbreviations

a.s.	almost surely
CONS	complete orthonormal sequence
DPR	direct probability redistribution
iid	independent and identically distributed
HJB	Hamilton–Jacobi–Bellman
IS	importance sampling
LLN	law of large numbers
LDP	large deviation principle
MDP	moderate deviation principle
PDE	partial differential equation
PRM	Poisson random measure
RCLL	right continuous with left limits
RESTART	repetitive simulation trials after reaching threshold
r.c.p.d.	regular conditional probability distribution
SDE	stochastic differential equation
SPDE	stochastic partial differential equation
w.p.1	with probability 1
WSLQ	weighted serve the longer queue

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

Specialized Symbols

$\mathscr{D}^m$	group of $\mathscr{C}^m$ diffeomorphisms, page 311
$(-\infty, x]$	$\{y : y_i \leq x_i$ for all $i = 1, \ldots, d\}$, page 289
$(\mathscr{H}, \langle\cdot,\cdot\rangle)$	a real separable Hilbert space, page 202
$\bar{V}^\delta$	mollification of $\bar{V}$, page 389
$\bar{\delta}_{Z_i}$	a random measure associated with branching processes, page 427
$\bar{\nu}_T$	$\lambda_T \times \nu \times \lambda_\infty$, page 215
$\bar{L}^n$	controlled empirical measure on $\bar{X}_i^n, i = 1, \ldots, n$, page 47
$\bar{l}_2$	a weighted l_2 space, page 283
$\bar{L}_T(u)$	$L_T^W(\psi) + L_T^N(\varphi)$, page 231
$\bar{M}^n(dw \times dt)$	random measures used in the analysis of a MDP, page 121
$\bar{N}$	an augmented PRM, page 215
$\bar{r}(x, t; v)$	controlled feedback jump rates in WSLQ, page 343
$\bar{U}(x)$	$j\Delta$ if $x \in C_j \backslash C_{j-1}$, page 428
$\bar{U}_k$	$\bar{U}(x)$ if $\sigma(x) = k$, page 428
$\bar{w}^n(t)$	mean of control measure $\bar{\mu}_i^n$ for $t \in [1/n, 1/n + 1/n)$, page 121
$\bar{X}_i^n$	controlled random variables whose conditional distribution is $\bar{\mu}_i^n$, page 46
$\kappa_1(\beta), \bar{\kappa}_1(\beta), \kappa_2(\beta), \kappa_3$	quantities used to describe properties of ℓ, page 238
$\Delta(\mathscr{X})$	$\{\gamma \in \mathscr{P}(\mathscr{X}) : R(\gamma \\| \theta) < \infty\}$, page 31
Δ_ψ	the mapping from $\mathscr{P}(\mathscr{X})$ into $\mathscr{P}(\mathscr{Y})$ defined by $\beta = \Delta_\psi \alpha$ when $\beta(A) = \alpha(x : \psi(x) \in A)$ for Borel sets A, page 34
ℓ	the function $\ell(b) = b \log b - b + 1, b \geq 0$, page 54
$\gamma(dy \mid x, \beta)$	exponentially tilted version of $\theta(dy \mid x)$, page 93
$\hat{\mathscr{W}}_m$	$\mathscr{C}\big([0, T] : \mathscr{D}^m\big)$, page 311
$\hat{A}_b$	a collection of bounded predictable processes, page 219

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

$\kappa(t)$	$\max\{0 \le k \le K : t_k \le t\}$, page 344
Λ	urn type index for occupancy models, page 177
$\lambda_k(l)$	initializing distribution for splitting, page 433
$\langle h, k\rangle_0$	inner product on $\mathscr{H}_0$, page 202
$\\|\alpha\\|_A^2$	$\langle \alpha, A\alpha\rangle$, page 117
$\{\bar{\mu}_i^n\}$	random control probability measures, page 46
$\{v_i(x)\}$	iid random vector fields, page 77
$\bar{\mathbb{M}}$	$\Sigma(\mathscr{Y}_T)$, sample space for an augmented PRM, page 215
$\bar{\mathbb{V}}$	$\mathbb{W} \times \bar{\mathbb{M}}$, page 234
$\mathbb{B}_0^T$	space of all continuous maps from $[0, T] \times \bar{O}$ to $\mathbb{R}$ endowed with the sup–norm, page 293
$\mathbb{B}_\alpha$	Banach space of α-Hölder functions on O, page 293
$\mathbb{B}_\alpha([0, T] \times O)$, $\mathbb{B}_\alpha^T$	Banach space of α-Hölder functions on $[0, T] \times O$, page 293
$\mathbb{H}(p)$	$-H(-p)$, page 382
$\mathbb{H}(x, p)$	$-H(x, -p)$, page 384
$\mathbb{M}$	$\Sigma(\mathscr{X}_T)$, sample space for a PRM, page 215
$\mathbb{R}^\infty$	space of real valued sequences equipped with product topology, page 282
$\mathbb{U}$	range space for the abstract LD and MD results of Chap. 9, page 236
$\mathbb{V}$	$\mathbb{W} \times \mathbb{M}$, page 234
$\mathbb{W}$	$\mathscr{C}([0, T] : \mathscr{H}_0)$, page 234
$\mathscr{D}_a$	feasible domain for an occupancy problem, page 190
$\mathscr{E}^\varphi(t)$	an exponential martingale associated with PRM, page 218
$\mathscr{F}(x, t; \omega, T)$	a collection of probability vectors with certain properties, page 192
$\mathscr{G}^\varepsilon$	measurable maps used in the abstract LDP of Chap. 9, page 236
$\mathscr{H}_0$	$\Lambda^{1/2}\mathscr{H}$, with Λ a symmetric, strictly positive, trace class operator, page 202
$\mathscr{K}^\varepsilon$	measurable maps used in the abstract MDP of Chap. 9, page 238
$\mathscr{L}_{\exp}$	$\cap_{\rho\in(0,\infty)} \mathscr{L}_{\exp}^\rho$, page 251
$\mathscr{L}_{\exp}^\rho$	functions that satisfy an exponential integrability assumption, page 251
$\mathscr{N}$	a collection of simple form absolutely continuous paths, page 341
$\mathscr{P}(\Lambda)$	the probabilities on $0, 1, \ldots, J + 1$, identified with the simplex in $\mathbb{R}^{J+2}$, page 176
$\mathscr{P}\mathscr{F}$	predictable σ-field, page 202
$\mathscr{V}$	set of possible jump vectors for the WSLQ model, page 332
$\mathscr{W}_m$	$\mathscr{C}\big([0, T] : \mathscr{C}^m(\mathbb{R}^d)\big)$, page 311
$\mathscr{X}$	space of types for a PRM, page 214

$\mathscr{X}_T$	$[0, T] \times \mathscr{X}$, page 214
$\mathscr{Y}$	augmented space of types for a controlled PRM, page 215
$\mathscr{Y}_T$	$[0, T] \times \mathscr{Y}$, page 215
$\mathfrak{S}^n(\bar{V})$	second moment of an estimator based on $\bar{V}$, page 388
$\text{Int}(u)$	integrated version of a control u, page 289
ν	measure on the space of types for a PRM, page 214
ν_T	$\lambda_T \times \nu$, page 214
ν_T^g	measure defined by $\int_A g(s, x)\, \nu_T(ds \times dx)$, $A \in \mathscr{B}(\mathscr{X}_T)$, page 235
$\omega(x, \delta)$	modulus of continuity, page 300
$\pi(x)$	indices that maximize the weighted queue length, page 332
ρ_j^δ	weights in the implementation of schemes based on mollified piecewise smooth subsolutions, page 389
$\rho_k(x, t)$	probabilities for ball placement in occupancy models, page 176
$\Sigma(\mathscr{S})$	measures ν on $(\mathscr{S}, \mathscr{B}(\mathscr{S}))$ satisfying $\nu(K) < \infty$ for every compact $K \subset \mathscr{S}$, page 214
$\sigma(x)$	unique integer j such that $x \in C_j \backslash C_{j-1}$, page 428
$\tau(dy\,\lvert x)$	a stochastic kernel in dy given x, page 32
$\theta(\cdot\,\lvert x)$	distribution of iid random vector fields $\{v_i(x)\}$, page 77
$\theta \otimes \sigma(A \times B)$	$\int_{A \times B} \theta(dx)\, \sigma(dy\,\lvert x) = \int_A \sigma(B\,\lvert x)\, \theta(dx)$, page 38
$\lVert f \rVert_{\infty,t}$	$\sup_{0 \le s \le t} \lVert f(s) \rVert$, page 252
$\lVert \cdot \rVert_0$	norm on the Hilbert space $\mathscr{H}_0$, page 202
$\lVert \cdot \rVert_{N,2}$	norm in the Hilbert space $\mathscr{L}^2(\nu_T)$, page 239
$\lVert \cdot \rVert_{W,2}$	norm in $\mathscr{L}^2([0, T] : \mathscr{H}_0)$, page 239
$\{\mathscr{F}_t\}$	a general filtration, page 56
$\{\mathscr{G}_t\}$	a filtration generated by driving noises, page 56
$\{U(t, s)\}$	a two parameter semigroup, page 292
$A(\mu)$	the probability measures on $S \times S$ with both marginals μ, page 149
$a(\varepsilon)$ and $\varkappa(\varepsilon)$	functions used in the statement of the abstract MDP of Chap. 9, page 238
$a(n)$	scaling sequence used in an MDP, page 115
$A_\kappa^{-1}(x)$	matrix obtained by truncating the eigenvalues of $A^{-1}(x)$ at κ^2, page 139
C_j	splitting thresholds, page 425
$C_{x,T}$	collection of paths starting from x and reaching a set B before reaching A, by time T, page 373
C_x	$\cup_{T \in (0,\infty)} C_{x,T}$, page 373
$Db(x)$	matrix of first order partial derivatives, page 119
$G(t, s, r, q)$	kernel of a two parameter semigroup, page 292
$H(\alpha)$	a log moment generating function, page 52
$H(x, \alpha)$	log moment generating function of iid random vector fields $\{v_i(x)\}$, page 78

$H^{(i)}(\alpha)$	$\mu_i(e^{-\alpha_i}-1)+\sum_{j=1}^d \lambda_j(e^{\alpha_j}-1)$, page 333
$H^*(x,\beta)$	the Legendre-Fenchel transform of $H(x,\alpha)$, also sometimes denoted as $L(x,\beta)$, page 90
H^A	$\max_{i\in A} H^{(i)}$, page 334
$H_c(x,\alpha)$	the centered log cumulant generating function, page 117
$I(A)$	$\inf_{x\in A} I(x)$, page 3
I_M	rate function in a moderate deviation principle, page 119
$L(\beta)$	the Legendre-Fenchel transform of $H(\alpha)$, page 55
$L(x,\beta)$	the Legendre-Fenchel transform of $H(x,\alpha)$, also sometimes denoted as $H^*(x,\beta)$, page 81
$L^{(i)}$	L^A when $A=\{i\}$, page 334
L^A	Legendre transform of H^A, page 334
L^n	the empirical measure on X_i, $i=1,\ldots,n$, page 47
l_2	Hilbert space of square summable sequences, page 283
$L_c(x,\beta)$	the Legendre-Fenchel transform of $H_c(x,\alpha)$, page 119
$L_{i,m}$	support threshold of particle m at time i, page 433
N^φ	Poisson random measure with controlled intensity governed by φ, page 215
N_c^1	the compensated version of N^1, page 218
$Q(j,k)$	random vector defined in terms of $q_l(j,k)$, page 439
$Q(t)$	vector of queue lengths at time t, page 332
$q^{(k)}(x,dy)$	k-step transition probability kernel for $q(x,dy)$, page 150
$q_l(j,k)$	splitting vectors, page 430
$R(\gamma\\|\theta)$	relative entropy of γ with respect to θ, page 29
$r(x,v)$	jump rate for the WSLQ model, page 332
$R^n(\{Y_i^n, w_i^n\}_{i=0,\ldots,Tn-1})$	likelihood ratio in an importance sampling estimator, page 389
R_j	splitting rates, page 425
$s^n(t)$	$\lfloor nt\rfloor/n$, page 124
$W^\psi(t)$	controlled Brownian motion $W(t)+\int_0^t \psi(s)ds$, page 231
$w^n(\delta)$	modulus of continuity, page 85
$w^n(t)$	rescaled mean of control measure $\bar{\mu}_i^n$ for $t\in[1/n,1/n+1/n)$, page 121
$L_T^N(\varphi)$	cost function for a PRM, page 231
$L_T^W(\psi)$	cost function for a Hilbert space valued Brownian motion, page 231

References

1. R.A. Adams, J.J.F. Fournier, *Sobolev Spaces*, 2nd edn. (Academic, New York, 2003)
2. M. Alanyali, B. Hajek, On large deviations of Markov processes with discontinuous statistics. Ann. Appl. Probab. **8**, 45–66 (1998)
3. D. Aldous, Stopping times and tightness. Ann. Prob. **6**, 335–340 (1978)
4. D.F. Anderson, T. Kurtz, *Stochastic Analysis of Biochemical Systems*, Mathematical Biosciences Institute Graduate Lecture (Springer, New York, 2015)
5. G. Ben Arous, F. Castell, Flow decomposition and large deviations. J. Funct. Anal. **140**, 23–67 (1995)
6. S. Asmussen, P.W. Glynn, *Stochastic Simulation: Algorithms and Analysis*. Applications of Mathematics. (Springer Science+Business Media, LLC, Berlin, 2007)
7. R. Azencott, Petits perturbations aléatoires des systems dynamiques: Développments asymptotiques. Bull. des Sci. Math. **109**, 253–308 (1985)
8. R. Azencott, G. Ruget, Mélanges d'équations differentielles et grand écart à la loi des grandes nombres. Z. Wahrsch. Verw. Gebiete **38**, 1–54 (1977)
9. R.R. Bahadur, R. Ranga Rao, On deviations of the sample mean. Ann. Math. Stat. **31**(4), 1015–1027 (1960)
10. P. Baldi, M. Sanz-Solé, Modulus of continuity for stochastic flows, in *Progress in Probability*, vol. 32 (Birkhauser, Basel, 1993)
11. S. Banerjee, A. Budhiraja, M. Perlmutter, Large deviations from the hydrodynamic limit for a system with nearest neighbor interactions. Math. arXiv:1803.09344
12. J. Bao, C. Yuan, Large deviations for neutral functional SDEs with jumps. Stoch. Int. J. Probab. Stoch. Process. **87**(1), 48–70 (2015)
13. V. Barbu, T. Precupanu, *Convexity and Optimization in Banach Spaces* (Springer, Berlin, 2012)
14. M. Bardi, I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations* (Birkhäuser, Basel, 1997)
15. G. Barles, An approach of deterministic control problems with unbounded data. Ann. Inst. Henri Poincaré: Anal. Non linéaire **7**, 235–258 (1990)
16. P. Baxendale, Brownian motions in the diffeomorphism group. I. Compos. Math. **53**(1), 19–50 (1984)
17. P. Baxendale, Asymptotic behaviour of stochastic flows of diffeomorphisms. In *Stochastic Processes and Their Applications* (Springer, Berlin, 1986), pp. 1–19
18. A. Benveniste, M. Metivier, P. Priouret, *Adaptive Algorithms and Stochastic Approximation* (Springer, Berlin, 1990)
19. D. Bertsekas, S. Shreve, *Stochastic Optimal Control: The Discrete Time Case* (Academic, San Diego, 1978)

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0

20. H. Bessaih, A. Millet, Large deviation principle and inviscid shell models. Electron. J. Probab. **14**(89), 2551–2579 (2009)
21. H. Bessaih, A. Millet, Large deviations and the zero viscosity limit for 2D stochastic Navier-Stokes equations with free boundary. SIAM J. Math. Anal. **44**(3), 1861–1893 (2012)
22. H. Bessaih, A. Millet, On stochastic modified 3d Navier-Stokes equations with anisotropic viscosity. J. Math. Anal. Appl. **462**(1), 915–956 (2018)
23. S. Bhamidi, A. Budhiraja, P. Dupuis, R. Wu, Large deviation principle for the exploration process of the configuration model (2017), arXiv:1708.01832
24. P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968)
25. J.M. Bismut, Mécanique Aléatoire. Lecture Notes in Mathematics, vol. 866 (1981)
26. J. Blanchet, H. Lam, State-dependent importance sampling for rare-event simulation: an overview and recent advances. Surv. Oper. Res. Manag. Sci. **17**(1), 38–59 (2012)
27. J.H. Blanchet, P. Glynn, Efficient rare-event simulation for the maximum of heavy-tailed random walks. Ann. Appl. Prob. **18**, 1351–1378 (2008)
28. J.H. Blanchet, P. Glynn, K. Leder, On Lyapunov inequalities and subsolutions for efficient importance sampling. ACM Trans. Model. Comput. Simul. **22**(3), 13:1–13:27 (2012)
29. E. Bolthausen, Markov process large deviations in τ- topology. Stoch. Proc. Appl. **25**, 95–108 (1987)
30. T.E. Booth, J. Hendricks, Importance estimation in forward Monte Carlo calculations. Nucl. Tech./Fusion **6**, 90–100 (1984)
31. C. Borell, Diffusion equations and geometric inequalities. Potential Anal. **12**(1), 49–71 (2000)
32. M. Boué, P. Dupuis, A variational representation for certain functionals of Brownian motion. Ann. Prob. **26**, 1641–1659 (1998)
33. M. Boué, P. Dupuis, R.S. Ellis, Large deviations for small noise diffusions with discontinuous statistics. Prob. Theor. Rel. Fields **116**, 125–148 (2000)
34. L. Breiman, *Probability Theory* (Addison-Wesley, Reading, 1968)
35. P. Brémaud, *Point Processes and Queues: Martingale Dynamics* (Springer, Berlin, 1981)
36. W. Bryc, Large deviations by the asymptotic value method, in *Diffusion Processes and Related Problems in Analysis*. Progress in Probability, vol. I (Evanston, IL, 1989), pp. 447–472
37. Z. Brzeźniak, B. Goldys, T. Jegaraj, Large deviations and transitions between equilibria for stochastic Landau-Lifshitz-Gilbert equation. Arch. Ration. Mech. Anal. **226**(2), 497–558 (2017)
38. A. Budhiraja, J. Chen, P. Dupuis, Large deviations for stochastic partial differential equations driven by a Poisson random measure. Stoch. Proc. Appl. **123**, 523–560 (2013)
39. A. Budhiraja, P. Dupuis, A variational representation for positive functionals of infinite dimensional Brownian motion. Prob. Math. Stat. **20**, 39–61 (2000)
40. A. Budhiraja, P. Dupuis, Large deviations for the empirical measures of reflecting Brownian motion and related constrained processes in $\mathbb{R}_+$. Elec. J. Probab. **8**, 1–46 (2003)
41. A. Budhiraja, P. Dupuis, A. Ganguly, Moderate deviation principles for stochastic differential equations with jumps. Ann. Probab. **44**, 1723–1775 (2016)
42. A. Budhiraja, P. Dupuis, A. Ganguly. Large deviations for small noise diffusions in a fast Markovian environment. Electron. J. Probab. **23** (2018)
43. A. Budhiraja, P. Dupuis, V. Maroulas, Large deviations for infinite dimensional stochastic dynamical systems. Ann. Probab. **36**, 1390–1420 (2008)
44. A. Budhiraja, P. Dupuis, V. Maroulas, Large deviations for stochastic flows of diffeomorphisms. Bernoulli J. **16**, 234–257 (2010)
45. A. Budhiraja, P. Dupuis, V. Maroulas, Variational representations for continuous time processes. Ann. de l'Inst. H. Poincaré **47**, 725–747 (2011)
46. A. Budhiraja, W-T. Fan, R. Wu, Large deviations for Brownian particle systems with killing. J. Theor. Probab. 1–40 (2017)
47. A. Budhiraja, P. Nyquist, Large deviations for multidimensional state-dependent shot-noise processes. J. Appl. Probab. **52**(4), 1097–1114 (2015)

48. A. Budhiraja, R. Wu, Moderate deviation principles for weakly interacting particle systems. Probab. Theory Relat. Fields **168**(3–4), 721–771 (2017)
49. A. Buijsrogge, P. Dupuis, M. Snarski, Splitting algorithms for rare event simulation over long time intervals. To appear in Ann. App. Prob.
50. C. Cardon-Weber, Large deviations for a Burgers'-type SPDE. Stoch. Process. Appl. **84**(1), 53–70 (1999)
51. F. Cerou, P. Del Moral, F. LeGland, P. Lezaud, Genetic genealogical models in rare event analysis. ALEA Lat. Am. J. Probab. Math. Stat. **1**, 181–203 (2006)
52. S. Cerrai, M. Röckner, Large deviations for stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term. Ann. Probab. **32**(1B), 1100–1139 (2004)
53. A. Charalambides, A unified derivation of occupancy and sequential occupancy distributions, in *Advances in Combinatorial Methods and Applications to Probability and Statistics* (1997), pp. 259–273
54. B. Chen, J. Blanchet, C.-H. Rhee, B. Zwart, Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound Poisson processes. Math. arxiv: https://arxiv.org/abs/1706.03981
55. Y. Chen, H. Gao, Well-posedness and large deviations for a class of SPDEs with Lévy noise. J. Differ. Equ. **263**(9), 5216–5252 (2017)
56. F. Chenal, A. Millet, Uniform large deviations for parabolic SPDEs and applications. Stoch. Process. Appl. **72**(2), 161–186 (1997)
57. L. Cheng, R. Li, W. Liu, Moderate deviations for the Langevin equation with strong damping. J. Stat. Phys. **170**(5), 845–861 (2018)
58. S. Chevet, Gaussian measures and large deviations, in *Probability in Banach Spaces IV* (Springer, Berlin, 1983), pp. 30–46
59. T.-S. Chiang, A lower bound of the asymptotic behavior of some Markov processes. Ann. Probab. **10**(4), 955–967 (1982)
60. P.-L. Chow, Large deviation problem for some parabolic Itô equations. Commun. Pure Appl. Math. **45**(1), 97–120 (1992)
61. G. Christensen, Deformable shape models for anatomy. Ph.D. thesis (1994)
62. G.E. Christensen, R.D. Rabbitt, M.I. Miller, Deformable templates using large deformation kinematics. IEEE Trans. Image Process. **5**(10), 1435–1447 (1996)
63. I. Chueshov, A. Millet, Stochastic 2D hydrodynamical type systems: well posedness and large deviations. Appl. Math. Optim. **61**(3), 379–420 (2010)
64. F. Cipriano, T. Costa, A large deviations principle for stochastic flows of viscous fluids. J. Differ. Equ. **264**(8), 5070–5108 (2018)
65. J.F. Collamore, Hitting probabilities and large deviations. Ann. Probab. **24**(4), 2065–2078 (1996)
66. J.B. Conway, *A Course in Functional Analysis*, vol. 96 (Springer Science & Business Media, Berlin, 2013)
67. T.M. Cover, J.A. Thomas, *Elements of Information Theory*, 2nd edn. (Wiley, New York, 2006)
68. H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités. Actualités Scientifiques et Industrielles, 736:2–23 1938; Colloque consacré à la théorie des probabilités, vol. 3 (Hermann, Paris)
69. G. Da Prato, J. Zabczyk, *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, vol. 44 (Cambridge University Press, Cambridge, 1992)
70. H. Dadashi, Large deviation principle for semilinear stochastic evolution equations with Poisson noise. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **20**(02), 1750009 (2017)
71. H. Dadashi-Arani, B.Z. Zangeneh, Large deviation principle for semilinear stochastic evolution equations with monotone nonlinearity and multiplicative noise. Differ. Integral Equ. **23**(7–8), 747–772 (2010)
72. A. de Acosta, Large deviations for empirical measures of Markov chains. J. Theor. Proba. **3** (1990)
73. A. de Acosta, On large deviations of empirical measures in the τ-topology. J. Appl. Proba. **31**, 41–47 (1994)

74. E. De Giorgi, Sulla convergenza di alcune successioni dintegrali del tipo dellarea. Ennio De Giorgi **414** (1975)
75. A. de Oliveira Gomes, Asymptotics for FBSDEs with jumps and connections with partial integral differential equations, in *From Particle Systems to Partial Differential Equations III* (Springer, Berlin, 2016), pp. 99–120
76. T. Dean, P. Dupuis, Splitting for rare event simulation: a large deviations approach to design and analysis. Stoch. Proc. Appl. **119**, 562–587 (2009)
77. T. Dean, P. Dupuis, The design and analysis of a generalized RESTART/DPR algorithm for rare event simulation. Ann. OR **189**, 63–102 (2011)
78. P. Del Moral, J. Garnier, Genealogical particle analysis of rare events. Ann. Appl. Probab. **15**, 2496–2534 (2005)
79. C. Dellacherie, P.A. Meyer, *Probabilities and Potential B: Theory of Martingales*. North-Holland Mathematics Studies, vol. 2 (North-Holland Publishing Company, Amsterdam, 1982)
80. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications* (Springer, New York, 1998)
81. F. den Hollander, *Large Deviations*, Fields Institute Monographs (AMS, Providence, 2000)
82. J.-D. Deuschel, D.W. Stroock, *Large Deviations* (Academic, San Diego, 1989)
83. J. Dieudonné, *Foundations of Modern Analysis*. Number v. 10, pt. 1 in Dieudonné, Jean: Treatise on analysis (Academic, New York, 1960)
84. I.H. Dinwoodie, P. Ney, Occupation measures for Markov chains. J. Theor. Probab. **8**(3), 679–691 (1995)
85. J. Doll, P. Dupuis, On performance measures for infinite swapping Monte Carlo methods. J. Chem. Phys. **142**, 024111 (2015)
86. J. Doll, P. Dupuis, P. Nyquist, A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. Appl. Math. Opt. 1–42 (2017)
87. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, I. Commun. Pure Appl. Math. **28**, 1–47 (1975)
88. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, III. Commun. Pure Appl. Math. **29**, 389–461 (1976)
89. M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, IV. Commun. Pure Appl. Math. **36**, 183–212 (1983)
90. J. Doob, *Stochastic Processes* (Wiley, New York, 1953)
91. J. Duan, A. Millet, Large deviations for the Boussinesq equations under random influences. Stoch. Process. Appl. **119**(6), 2052–2081 (2009)
92. R.M. Dudley, *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics, vol. 74 (Cambridge University Press, Cambridge, 2002). Revised Reprint of the 1989 Original
93. N. Dunford, J.T. Schwartz, *Linear Operators Parts I, II, III* (Interscience Publishers, Geneva, 1963)
94. P. Dupuis, Large deviations analysis of some recursive algorithms with state dependent noise. Ann. Probab. **16**, 1509–1536 (1988)
95. P. Dupuis, R.S. Ellis, Large deviations for Markov processes with discontinuous statistics, II: Random walks. Probab. Theory Rel. Fields **91**, 153–194 (1992)
96. P. Dupuis, R.S. Ellis, The large deviation principle for a general class of queueing systems. I. Trans. Am. Math. Soc. **347**, 2689–2751 (1996)
97. P. Dupuis, R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations* (Wiley, New York, 1997)
98. P. Dupuis, R.S. Ellis, A. Weiss, Large deviations for Markov processes with discontinuous statistics, I: General upper bounds. Ann. Probab. **19**, 1280–1297 (1991)
99. P. Dupuis, U. Grenander, M. Miller, A variational formulation of a problem in image matching. Q. Appl. Math. **56**, 587–600 (1998)
100. P. Dupuis, D. Johnson, Moderate deviations for recursive stochastic algorithms. Stoch. Syst. **1**, 1–33 (2015)
101. P. Dupuis, D. Johnson, Moderate deviations based importance sampling for recursive stochastic equations. J. Appl. Probab. **49**, 981–1010 (2017)

102. P. Dupuis, H.J. Kushner, Stochastic approximation and large deviations: upper bounds and w.p.1 convergence. SIAM J. Control Optim. **27**, 1108–1135 (1989)
103. P. Dupuis, K. Leder, H. Wang, Large deviations and importance sampling for a tandem network with slow-down. QUESTA **57**, 71–83 (2007)
104. P. Dupuis, K. Leder, H. Wang, On the large deviations properties of the weighted-serve-the-longest-queue policy, in *In and Out of Equilibrium 2*, ed. by V. Sidoravicius, M.E. Vares (Birkhauser, New York, 2008)
105. P. Dupuis, K. Leder, H. Wang, Importance sampling for weighted serve-the-longest-queue. Math. Oper. Res. **34**, 642–660 (2009)
106. P. Dupuis, D. Lipshutz, Large deviations for the empirical measure of a diffusion via weak convergence methods. Stoch. Proc. Appl. **128**, 2581–2604 (2018)
107. P. Dupuis, Y. Liu, On the large deviation rate for the empirical measure of reversible pure jump Markov processes. Ann. Probab. **43**, 1121–1156 (2015)
108. P. Dupuis, Y. Liu, N. Plattner, J.D. Doll, On the infinite swapping limit for parallel tempering. SIAM J. Multiscale Model. Simul. **10**, 986–1022 (2012)
109. P. Dupuis, C. Nuzman, P. Whiting, Large deviation asymptotics for occupancy problems. Ann. Probab. **32**, 2765–2818 (2004)
110. P. Dupuis, A. Sezer, H. Wang, Dynamic importance sampling for queueing networks. Ann. Appl. Prob. **17**(4), 1306–1346 (2007)
111. P. Dupuis, K. Spiliopoulos, Large deviations for multiscale diffusions via weak convergence methods. Stoch. Proc. Appl. **122**, 1947–1987 (2012)
112. P. Dupuis, K. Spiliopoulos, H. Wang, Importance sampling for multiscale diffusions. SIAM J. Multiscale Model. Simul. **10**, 1–27 (2012)
113. P. Dupuis, K. Spiliopoulos, X. Zhou, Escaping from an attractor: importance sampling and rest points I. Ann. Appl. Probab. **25**, 2909–2958 (2015)
114. P. Dupuis, H. Wang, Importance sampling, large deviations, and differential games. Stoch. Stoch. Rep. **76**, 481–508 (2004)
115. P. Dupuis, H. Wang, Dynamic importance sampling for uniformly recurrent Markov chains. Ann. Appl. Probab. **15**, 1–38 (2005)
116. P. Dupuis, H. Wang, Subsolutions of an Isaacs equation and efficient schemes for importance sampling. Math. Oper. Res. **32**, 1–35 (2007)
117. P. Dupuis, H. Wang, Importance sampling for Jackson networks. Queueing Syst. **62**, 113–157 (2009)
118. P. Dupuis, O. Zeitouni, A nonstandard form of the rate function for the occupation measure of a Markov chain. Stoch. Proc. Appl. **61**, 249–261 (1996)
119. P. Dupuis, J. Zhang, Explicit solutions for a class of nonlinear PDE that arise in allocation problems. SIAM J. Math. Anal. **39**(5), 1627–1667 (2008)
120. R.S. Ellis, Large deviations for a general class of random vectors. Ann. Probab. **12**, 1–12 (1984)
121. R.S. Ellis, *Entropy, Large Deviations and Statistical Mechanics* (Springer, New York, 1985)
122. R.S. Ellis, Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. Ann. Probab. **16**, 1496–1508 (1988)
123. R.S. Ellis, A.D. Wyner, Uniform large deviation property of the empirical process of a Markov chain. Ann. Probab. **17**, 1147–1151 (1989)
124. K.D. Elworthy, Stochastic dynamical systems and their flows, in *Stochastic Analysis* (Academic Press, New York, 1978), pp. 79–95
125. K.D. Elworthy, Stochastic flows on Riemannian manifolds, in *Diffusion Processes and Related Problems in Analysis*, vol. II (Springer, Berlin, 1992), pp. 37–72
126. S.N. Ethier, T.G. Kurtz, *Markov Processes: Characterization and Convergence* (Wiley, New York, 1986)
127. W.G. Faris, G. Jona-Lasinio, Large fluctuations for a nonlinear heat equation with noise. J. Phys. A **15**(10), 3025–3055 (1982)
128. H. Federer, *Geometric Measure Theory* (Springer, Berlin, 1996)

129. W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1 (Wiley, New York, 1968)
130. W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2 (Wiley, New York, 1971)
131. J. Feng, Large deviation for diffusions and Hamilton-Jacobi equation in Hilbert spaces. Ann. Probab. **34**(1), 321–385 (2006)
132. J. Feng, T.G. Kurtz, *Large Deviations for Stochastic Processes*. Mathematical Surveys and Monographs, vol. 131 (American Mathematical Society, Providence, 2006)
133. W.H. Fleming, Exit probabilities and optimal stochastic control. Appl. Math. Optim. **4**, 329–346 (1978)
134. W.H. Fleming, H.M. Soner, Asymptotic expansions for Markov processes with Lévy generators. Appl. Math. Optim. **19**, 203–223 (1989)
135. W.H. Fleming, H.M. Soner, *Controlled Markov Processes and Viscosity Solutions* (Springer, New York, 1992)
136. W.H. Fleming, P.E. Souganidis, Asymptotic series and the method of vanishing viscosity. Indiana Univ. Math. J. **35**, 425–447 (1987)
137. R. Foley, D. McDonald, Join the shortest queue: stability and exact asymptotics. Ann. Appl. Probab. **11**, 569–607 (2001)
138. M.I. Freidlin, The averaging principle and theorems on large deviations. Russian Math. Surv. **33**, 117–176 (1978)
139. M.I. Freidlin, Random perturbations of reaction-diffusion equations: the quasideterministic approximation. Trans. Am. Math. Soc. **305**(2), 665–697 (1988)
140. M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1984)
141. A.J. Ganesh, N. O'Connell, D.J. Wischik, *Big Queues* (Springer, Berlin, 2004)
142. H. Gao, C. Sun, Well-posedness and large deviations for the stochastic primitive equations in two space dimensions. Commun. Math. Sci. **10**(2), 575–593 (2012)
143. P. Gao, The stochastic Swift-Hohenberg equation. Nonlinearity **30**(9), 3516 (2017)
144. J. Gärtner, On large deviations from the invariant measure. Theory Probab. Appl. **22**, 24–39 (1977)
145. M.J.J. Garvels, The splitting method in rare event simulation. Ph.D. thesis, University of Twente, The Netherlands (2000)
146. M. Ghosh, Probabilities of moderate deviations under m-dependence. Can. J. Stat. **2**(1–2), 157–168 (1974)
147. D. Gilbarg, N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd edn. (Springer, Berlin, 1983)
148. P. Glasserman, P. Heidelberger, P. Shahabuddin, T. Zajic, A large deviations perspective on the efficiency of multilevel splitting. IEEE Trans. Automat. Control **43**, 1666–1679 (1998)
149. P. Glasserman, S. Kou, Analysis of an importance sampling estimator for tandem queues. ACM Trans. Model. Comput. Simul. **4**, 22–42 (1995)
150. P. Glasserman, Y. Wang, Counter examples in importance sampling for large deviations probabilities. Ann. Appl. Probab. **7**, 731–746 (1997)
151. U. Grenander, M.I. Miller, Representations of knowledge in complex systems. J. R. Stat. Soc. B **56**(3), 549–603 (1994)
152. Z. Haraszti, J.K. Townsend, The theory of direct probability redistribution and its application to rare event simulation, in *Proceedings of the IEEE International Conference on Communications* (1998), pp. 1443–1450
153. Z. Haraszti, J.K. Townsend, The theory of direct probability redistribution and its application to rare event simulation. ACM Trans. Model. Comput. Simul. **9**, 105–140 (1999)
154. L. Holst, On the coupon collectors and other urn problems. Int. Stat. Rev. **54**(1), 15–27 (1986)
155. R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, New York, 1991)
156. Y. Hu, D. Nualart, T. Zhang, Large deviations for stochastic heat equation with rough dependence in space. Bernoulli **24**(1), 354–385 (2018)

157. I.A. Ibragimov, Conditions for smoothness of trajectories of random functions. Teor. Veroyatnost. i Primenen. **28**(2), 229–250 (1983)
158. I. Ignatiouk-Robert, Large deviations for processes with discontinuous statistics. Ann. Probab. **33**, 1479–1508 (2005)
159. N. Ikeda, S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library, vol. 24 (North-Holland Publishing Co., Amsterdam; 2nd edn., Kodansha, Ltd., Tokyo, 1989)
160. V.M. Imaĭkin, A.I. Komech, Large deviations of solutions of nonlinear stochastic equations. Trudy Sem. Petrovsk. 258(13), 177–196 (1988)
161. J. Jacod, A.N. Shiryaev, *Limit Theorems for Stochastic Processes* (Springer, Berlin, 1987)
162. N. Jain, Large deviation lower bounds for additive functionals of Markov processes. Ann. Probab. **18**, 1071–1098 (1990)
163. Y. Le Jan, Flots de diffusions dans $\mathbb{R}^d$. C.R. Acad. Sci., Paris, Ser. I **294**, 687–689 (1982)
164. Y. Le Jan, S. Watanabe, Stochastic flows of diffeomorphisms, in *North-Holland Mathematical Library*, vol. 32 (Elsevier, Amsterdam, 1984), pp. 307–332
165. N.L. Johnson, S. Kotz, *Urn Models and Their Applications* (Wiley, New York, 1977)
166. H. Kahn, T.E. Harris, Estimation of particle transmission by random sampling. Natl. Bur. Stand. Appl. Math. Ser. **12**, 27–30 (1951)
167. O. Kallenberg, *Foundations of Modern Probability*. Probability and its Applications (New York), 2nd edn. (Springer, New York, 2002)
168. G. Kallianpur, *Stochastic Filtering Theory*, vol. 13 (Springer Science & Business Media, Berlin, 2013)
169. G. Kallianpur, J. Xiong, *Stochastic Differential Equations in Infinite-Dimensional Spaces*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 26 (Institute of Mathematical Statistics, Hayward, 1995)
170. G. Kallianpur, J. Xiong, Large deviations for a class of stochastic partial differential equations. Ann. Probab. **24**(1), 320–345 (1996)
171. H.-W. Kang, T.G. Kurtz, Separation of time-scales and model reduction for stochastic reaction networks. Ann. Appl. Probab. **23**(2), 529–583 (2013)
172. I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer, New York, 1988)
173. A. Klenke, *Probability Theory: A Comprehensive Course* (Springer Science & Business Media, Berlin, 2013)
174. P. Kotelenez, Existence, uniqueness and smoothness for a class of function valued stochastic partial differential equations. Stoch.: Int. J. Probab. Stoch. Process. **41**(3), 177–199 (1992)
175. P. Kotelenez, *Stochastic Ordinary and Stochastic Partial Differential Equations: Transition from Microscopic to Macroscopic Equations*, vol. 58 (Springer Science & Business Media, Berlin, 2007)
176. S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959)
177. S. Kullback, R.A. Leibler, On information and sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951)
178. H. Kunita, *Stochastic Flows and Stochastic Differential Equations* (Cambridge University Press, Cambridge, 1990)
179. T.G. Kurtz, *Approximation of Population Processes*, CBMS-NSF Regional Conference, vol. 36 (Series in Applied Mathematics) (SIAM, Philadelphia, 1981)
180. T.G. Kurtz, P.E. Protter, Weak convergence of stochastic integrals and differential equations II: Infinite dimensional case, in *Probabilistic Models for Nonlinear Partial Differential Equations* (Springer, Berlin, 1996), pp. 197–285
181. H.J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations* (Academic, New York, 1977)
182. H.J. Kushner, G.G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Stochastic Modelling and Applied Probability (Springer, New York, 2003)
183. P. L'Ecuyer, J.H. Blanchet, B. Tuffin, P.W. Glynn, Asymptotic robustness of estimators in rare-event simulation. ACM Trans. Model. Comput. Simul. **20**(1), 1–41 (2010)

184. J. Lehec, Representation formula for the entropy and functional inequalities. Ann. Inst. H. Poincar Probab. Stat. **49**(3), 885–899 (2013)
185. C. Léonard, Large deviations for Poisson random measures and processes with independent increments. Stoch. Process. Appl. **85**(1), 93–121 (2000)
186. Y. Li, R. Wang, N. Yao, S. Zhang, A moderate deviation principle for stochastic Volterra equation. Stat. Probab. Lett. **122**, 79–85 (2017)
187. Y. Li, R. Wang, N. Yao, S. Zhang, Moderate deviations for a fractional stochastic heat equation with spatially correlated noise. Stoch. Dyn. **17**(04), 1750025 (2017)
188. Y. Li, S. Zhang, Moderate deviations and central limit theorem for positive diffusions. J. Inequalities Appl. **2016**(1), 87 (2016)
189. D. Lipshutz, Exit time asymptotics for small noise stochastic delay differential equations (2017). arXiv:1710.09771
190. J.S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2004)
191. W. Liu, Large deviations for stochastic evolution equations with small multiplicative noise. Appl. Math. Optim. **61**(1), 27–56 (2010)
192. W. Liu, M. Röckner, X.-C. Zhu, Large deviation principles for the stochastic quasi-geostrophic equations. Stoch. Process. Appl. **123**(8), 3299–3327 (2013)
193. L. Ljung, T. Söderström, *Theory and Practice of Recursive Identification*. Series in Signal Processing, Optimization, and Control (MIT Press, Cambridge, 1985)
194. X. Ma, F. Xi, Moderate deviations for neutral stochastic differential delay equations with jumps. Stat. Probab. Lett. **126**, 97–107 (2017)
195. K. Majewski, Large deviations of the steady-state distribution of reflected processes with applications to queueing systems. Queueing Syst. **29**, 351–381 (1998)
196. U. Manna, S.S. Sritharan, P. Sundar, Large deviations for the stochastic shell model of turbulence. NoDEA Nonlinear Differ. Equ. Appl. **16**(4), 493–521 (2009)
197. M. Métivier, *Semimartingales: A Course on Stochastic Processes*, vol. 2 (Walter de Gruyter, Berlin, 1982)
198. M. Métivier, J. Pellaumail, *Stochastic Integration* (Academic, New York, 2014)
199. Paul-André Meyer, Les inégalités de Burkholder en théorie des martingales, d'après Gundy. Séminaire de probabilités de Strasbourg **3**, 163–174 (1969)
200. R. Michel, Results on probabilities of moderate deviations. Ann. Probab. **2**(2), 349–353 (1974)
201. M.I. Miller, G.E. Christensen, Y. Amit, U. Grenander, Mathematical textbook of deformable neuroanatomies. Proc. Natl. Acad. Sci. **90**(24), 11944–11948 (1993)
202. A. Millet, D. Nualart, M. Sanz-Solé, Large deviations for a class of anticipating stochastic differential equations. Ann. Probab. **20**(4), 1902–1931 (1992)
203. C. Mo, J. Luo, Large deviations for stochastic differential delay equations. Nonlinear Anal. **80**, 202–210 (2013)
204. P. Ney, E. Nummelin, Markov additive processes I: eigenvalue properties and limit theorems. Ann. Prob. **16**, 561–592 (1987)
205. P. Ney, E. Nummelin, Markov additive processes II: large deviations. Ann. Prob. **15**, 593–609 (1987)
206. D. Nualart, E. Pardoux, Markov field properties of solutions of white noise driven quasi-linear parabolic PDEs. Stoch.: Int. J. Probab. Stoch. Process. **48**(1–2), 17–44 (1994)
207. V. Ortiz-López, M. Sanz-Solé, A Laplace principle for a stochastic wave equation in spatial dimension three, in *Stochastic Analysis 2010* (Springer, Berlin, 2011), pp. 31–49
208. V. Demers, P. L'Ecuyer, B. Tuffin, Rare events, splitting, and quasi-Monte Carlo. ACM Trans. Model. Comput. Simul. **17**(2) (2007)
209. S. Peszat, Large deviation principle for stochastic evolution equations. Probab. Theory Related Fields **98**(1), 113–136 (1994)
210. P.E. Protter, *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability, vol. 21 (Springer, Berlin, 2005). 2nd edn. Version 2.1, Corrected Third Printing
211. F. Rassoul-Agha, T. Seppäläinen, *A Course on Large Deviations with an Introduction to Gibbs Measures*, vol. 162 (American Mathematical Society, Providence, 2015)

212. P.A. Razafimandimby, Viscosity limit and deviations principles for a grade-two fluid driven by multiplicative noise. Annali di Matematica Pura ed Applicata **197**(5), 1547–1583 (2018)
213. J. Ren, S. Xu, X. Zhang, Large deviations for multivalued stochastic differential equations. J. Theoret. Probab. **23**(4), 1142–1156 (2010)
214. J. Ren, X. Zhang, Freidlin-Wentzell's large deviations for homeomorphism flows of non-Lipschitz SDEs. Bull. Sci. Math. **129**(8), 643–655 (2005)
215. J. Ren, X. Zhang, Schilder theorem for the Brownian motion on the diffeomorphism group of the circle. J. Funct. Anal. **224**(1), 107–133 (2005)
216. J. Ren, X. Zhang, Freidlin-Wentzell's large deviations for stochastic evolution equations. J. Funct. Anal. **254**(12), 3148–3172 (2008)
217. R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)
218. M. Röckner, T. Zhang, X. Zhang, Large deviations for stochastic tamed 3D Navier-Stokes equations. Appl. Math. Optim. **61**(2), 267–285 (2010)
219. S. Ross, *Introduction to Probability Models*, 8th edn. (Academic, New York, 2002)
220. H.L. Royden, P. Fitzpatrick, *Real Analysis*, vol. 198 (Macmillan, New York, 1988)
221. B.L. Rozovskii, *Stochastic Evolution Systems: Linear Theory and Applications to Non-Linear Filtering*, vol. 35 (Springer Science & Business Media, Berlin, 2012)
222. H. Rubin, J. Sethuraman, Probabilities of moderate deviations. Sankhyā Ser. A **27**, 325–346 (1965)
223. G. Rubino, B. Tuffin (eds.), *Rare Event Simulation Using Monte Carlo Methods* (Wiley, New York, 2009)
224. R.Y. Rubinstein, D.P. Kroese, *Simulation and the Monte Carlo Method*, 3rd edn. (Wiley, New York, 2016)
225. R.Y. Rubinstein, A. Ridder, R. Vaisman, *Fast Sequential Monte Carlo Methods for Counting and Optimization*, 1st edn. (Wiley Publishing, New York, 2013)
226. W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1991)
227. M. Salins, A. Budhiraja, P. Dupuis, Uniform large deviation principles for Banach space valued stochastic differential equations (2017), arXiv:1803.00648 (To appear in Trans. Am. Math. Soc.)
228. I.N. Sanov, On the probability of large deviations of random variables. Mat. Sbornik **42**(84), 11–44 (1957)
229. M. Schilder, Some asymptotic formulas for Wiener integrals. Trans. Am. Math. Soc. **125**(1), 63–85 (1966)
230. Z. Schuss, *Theory and Applications of Stochastic Differential Equations* (Wiley, New York, 1988)
231. A. Shwartz, A. Weiss, *Large Deviations for Performance Analysis: Queues, Communication and Computing* (Chapman and Hall, New York, 1995)
232. D. Siegmund, Importance sampling in the Monte Carlo study of sequential tests. Ann. Stat. **4**, 673–684 (1976)
233. M. Sion, On general minimax theorems. Pac. J. Math. **8**, 171–176 (1958)
234. H. Soner, Optimal control with state-space constraint I. SIAM J. Control. Optim. **24**, 05 (1986)
235. R.B. Sowers, Large deviations for a reaction-diffusion equation with non-Gaussian perturbations. Ann. Probab. **20**(1), 504–537 (1992)
236. S.S. Sritharan, P. Sundar, Large deviations for the two-dimensional Navier-Stokes equations with multiplicative noise. Stoch. Process. Appl. **116**(11), 1636–1659 (2006)
237. D.W. Stroock, *An Introduction to the Theory of Large Deviations* (Springer, New York, 1984)
238. S.R.S. Varadhan, Asymptotic probabilities and differential equations. Comm. Pure Appl. Math. **19**(3), 261–286 (1966)
239. S.R.S. Varadhan, *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Mathematics. (SIAM, Philadelphia, 1984)
240. C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, 2009)
241. M. Villen-Altamirano, J. Villen-Altamirano, RESTART: A method for accelerating rare event simulations, in *Proceedings of the 13th International Teletraffic Congress, Queueing, Performance and Control in ATM* (Elsevier, Amsterdam, 1991), pp. 71–76

242. M. Villen-Altamirano, J. Villen-Altamirano, RESTART: A straightforward method for fast simulation of rare events, in *Proceedings of the 1994 Winter Simulation Conference* (1994), pp. 282–289
243. J.B. Walsh, An introduction to stochastic partial differential equations, in *École d'été de probabilités de Saint-Flour, XIV—1984*. Lecture Notes in Mathematics, vol. 1180 (Springer, Berlin, 1986), pp. 265–439
244. W. Wang, J. Duan, Reductions and deviations for stochastic partial differential equations under fast dynamical boundary conditions. Stoch. Anal. Appl. **27**(3), 431–459 (2009)
245. A.D. Wentzell, Rough limit theorems on large deviations for Markov stochastic processes. I. Theory Probab. Appl. **21**, 227–242 (1976)
246. A.D. Wentzell, Rough limit theorems on large deviations for Markov stochastic processes. II. Theory Probab. Appl. **21**, 499–512 (1976)
247. A.D. Wentzell, Rough limit theorems on large deviations for Markov stochastic processes. III. Theory Probab. Appl. **24**, 675–692 (1979)
248. A.D. Wentzell, Rough limit theorems on large deviations for Markov stochastic processes. IV. Theory Probab. Appl. **27**, 215–234 (1982)
249. R.J. Williams, Recurrence classification and invariant measure for reflected Brownian motion in a wedge. Ann. Probab. **13**, 758–778 (1985)
250. M. Winter, L. Xu, J. Zhai, T. Zhang, The dynamics of the stochastic shadow Gierer-Meinhardt system. J. Differ. Equ. **260**(1), 84–114 (2016)
251. J. Wu, Uniform large deviations for multivalued stochastic differential equations with Poisson jumps. Kyoto J. Math. **51**(3), 535–559 (2011)
252. J. Xiong, Large deviations for diffusion processes in duals of nuclear spaces. Appl. Math. Optim. **34**(1), 1–27 (1996)
253. J. Xiong, J. Zhai, Large deviations for locally monotone stochastic partial differential equations driven by Lévy noise. Bernoulli **24**(4A), 2842–2874 (2018)
254. T. Xu, T. Zhang, White noise driven SPDEs with reflection: existence, uniqueness and large deviation principles. Stoch. Process. Appl. **119**(10), 3453–3470 (2009)
255. D. Yang, J. Duan, Large deviations for the stochastic quasigeostrophic equation with multiplicative noise. J. Math. Phys. **51**(5), 053301 (2010)
256. D. Yang, Z. Hou, Large deviations for the stochastic derivative Ginzburg-Landau equation with multiplicative noise. Phys. D **237**(1), 82–91 (2008)
257. J. Yang, J. Zhai, Asymptotics of stochastic 2d hydrodynamical type systems in unbounded domains. Infin. Dimens. Anal., Quantum Probab. Relat. Top. **20**(03), 1750017 (2017)
258. X. Yang, J. Zhai, T. Zhang, Large deviations for SPDEs of jump type. Stoch. Dyn. **15**(04), 1550026 (2015)
259. G. Yin, C. Zhu, *Hybrid Switching Siffusions: Properties and Applications*, vol. 63 (Springer, New York, 2010)
260. L. Ying, R. Srikant, A. Eryilmaz, G.E. Dullerud, A large deviation analysis of scheduling in wireless networks. IEEE Trans. Inf. Theory **52**(11), 5088–5098 (2006)
261. J. Zabczyk, On large deviations for stochastic evolution equations, in *Stochastic Systems and Optimization (Warsaw, 1988)*. Lecture Notes in Control and Information Sciences, vol. 136 (Springer, Berlin, 1989), pp. 240–253
262. J. Zhai, T. Zhang, Large deviations for 2-d stochastic Navier-Stokes equations driven by multiplicative Lévy noises. Bernoulli **21**(4), 2351–2392 (2015)
263. J. Zhai, T. Zhang, Large deviations for stochastic models of two-dimensional second grade fluids. Appl. Math. Optim. **75**(3), 471–498 (2017)
264. J. Zhai, T. Zhang, W. Zheng, Large deviations for stochastic models of two-dimensional second grade fluids driven by Lévy noise (2017). arXiv preprint arXiv:1706.08862
265. J. Zhai, T. Zhang, W. Zheng, Moderate deviations for stochastic models of two-dimensional second grade fluids. Stoch. Dyn. **18**(03), 1850026 (2018)
266. J. Zhang, P. Dupuis, Large deviation principle for general occupancy models. Comb., Probab., Comput. **17**, 437–470 (2008)

267. R. Zhang, G. Zhou, Large deviations for nematic liquid crystals driven by pure jump noise. Math. Methods Appl. Sci. **41**(14), 5552–5581 (2018)
268. X. Zhang, Euler schemes and large deviations for stochastic Volterra equations with singular kernels. J. Differ. Equ. **244**(9), 2226–2250 (2008)
269. X. Zhang, Clark-Ocone formula and variational representation for Poisson functionals. Ann. Probab. **37**(2), 506–529 (2009)
270. X. Zhang, Stochastic Volterra equations in Banach spaces and stochastic partial differential equation. J. Funct. Anal. **258**(4), 1361–1425 (2010)
271. X. Zhang, Well-posedness and large deviation for degenerate SDEs with Sobolev coefficients. Rev. Mat. Iberoam. **29**(1), 25–52 (2013)
272. H. Zhao, S. Xu, Freidlin-Wentzells large deviations for stochastic evolution equations with Poisson jumps. Adv. Pure Math. **6**(10), 676 (2016)
273. W. Zheng, J. Zhai, T. Zhang, Moderate deviations for stochastic models of two-dimensional second grade fluids driven by Lévy noise (2018). arXiv preprint arXiv:1801.08429

Index

A. Budhiraja and P. Dupuis, *Analysis and Approximation of Rare Events*, Probability Theory and Stochastic Modelling 94,
https://doi.org/10.1007/978-1-4939-9579-0