J Theor Probab (2018) 31:1058–1096
https://doi.org/10.1007/s10959-016-0733-1

Large Deviations of Continuous Regular Conditional Probabilities

W. van Zuijlen[1]

Received: 17 May 2016 / Revised: 6 October 2016 / Published online: 27 December 2016

Abstract We study product regular conditional probabilities under measures of two coordinates with respect to the second coordinate that are weakly continuous on the support of the marginal of the second coordinate. Assuming that there exists a sequence of probability measures on the product space that satisfies a large deviation principle, we present necessary and sufficient conditions for the conditional probabilities under these measures to satisfy a large deviation principle. The arguments of these conditional probabilities are assumed to converge. A way to view regular conditional probabilities as a special case of product regular conditional probabilities is presented. This is used to derive conditions for large deviations of regular conditional probabilities. In addition, we derive a Sanov-type theorem for large deviations of the empirical distribution of the first coordinate conditioned on fixing the empirical distribution of the second coordinate.

Keywords (Product) regular conditional kernel · Weakly continuous · Large deviations

Mathematics Subject Classification (2010) 60A10 · 60F10

1 Introduction and Main Results

In the present paper, we study large deviations of probabilities "of the form"

$$\mathbb{P}(X_n \in A \mid Y_n = y_n), \tag{1.1}$$

✉ W. van Zuijlen
vanzuijlen@wias-berlin.de

[1] Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

where $((X_n, Y_n))_{n\in\mathbb{N}}$ is a sequence of couples of random variables that satisfies a large deviation principle and $y_n \to y$ for some y. As the event $[Y_n = y_n]$ may have probability zero, we make sense of (1.1) in terms of a kernel η_n, so that

$$\eta_n(y_n, A)$$

"represents" (1.1).

Such kernels are called regular conditional probabilities and form an important object in probability theory. The existence of regular conditional probabilities has been studied extensively, for example, by Faden [12] or by Leao et al. [21]. There exist in fact various forms of regular conditional probabilities, namely either with respect to a σ-algebra, with respect to a measurable map or with respect to the projection on one of the coordinates (in case of a product space).

In order to consider large deviations of conditional probabilities, we have to specify which conditional probability we are considering; the conditional probability may not be unique. However, if a (product) regular conditional probability is weakly continuous on the support of the measure composed with the inverse of the measurable map (or projection), it is unique on that domain. For these (product) regular conditional probabilities, it is natural to study their large deviations whenever the argument of the probability is in the domain on which it is unique. In this paper, we study the large deviations in the case when the arguments of these kernels converge, i.e. we study large deviations of $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ for the case that $y_n \to y$. To the best of our knowledge, current literature does not provide a general condition under which such kernels satisfy a large deviation principle.

1.1 Literature

Some examples in this direction are present. For example in Adams et al. [1], the large deviation principle is proved for the empirical distribution that is evolved by independent Brownian motions conditioned on their initial empirical distribution to lie in a ball (see [1, Theorem 1]). They proceed by proving that the large deviation principle rate function converges as the radius of the ball converges to zero. For the purpose of this paper, we have to show that the limit of the radius of the ball and the limit belonging to the large deviation principle can be interchanged. Léonard [22] proves the large deviation principle of the empirical distribution that is evolved by independent Brownian motions conditioned on their initial empirical distribution; those initial empirical distributions are assumed to be converging (see [22, Proposition 2.19]). In both papers, the evolved state is conditioned on the initial state, while there is also interest in large deviations of the initial state conditioned on the evolved state. In this paper, we prove the large deviation principle in this setting for finite state spaces.

There exist various results on quenched large deviations, i.e. large deviations for regular conditional probabilities in the sense that for almost all realisations of the disorder, the conditional probabilities satisfy the large deviation principle with a rate function that does not depend on the disorder. Examples of papers on quenched large deviations are Comets [5] for conditional large deviations of i.i.d. random fields, Greven and den Hollander [14] and Comets et al. [6] for random walks in random

environments, Kosygina–Rezakhanlou–Varadhan [18] for a diffusion with a random drift and Rassoul-Agha et al. [24] for polymers in a random potential.

Biggins [2] obtains the large deviation principle for mixtures of probability measures that satisfy the large deviation principle with kernels that satisfy the large deviation principle as their arguments converge. To some extent, we complement the article in the opposite direction, in the sense that we assume the large deviation principle of the mixture and derive the large deviation principle of the kernels.

Our main motivation to study the above large deviations lies in the theory of Gibbs–non-Gibbs transitions. There is a correspondence between the large deviation rate function of the conditional probability with respect to the evolved coordinate and the evolved state (measure or sequence) being Gibbs (see van Enter et al. [9]). We refer to Sect. 1.4 for further discussions on Gibbs–non-Gibbs transitions.

1.2 Large Deviations

In the literature on large deviations, two dominant definitions of large deviation principles are used. One is in terms of a σ-algebra on the topological space, as is done in the book by Dembo and Zeitouni [7] and in the book by Deuschel and Stroock [8]; the other is in terms of the topology, i.e. in terms of open and closed sets, as is done in the book by den Hollander [16] and in the book by Rassoul-Agha and Seppäläinen [25]. Whenever one considers the Borel σ-algebra on the topological space, the two definitions agree.

We define the large deviation lower bound and the large deviation upper bound separately, as in Sect. 1.3, and in Sect. 6, we describe the necessary and sufficient conditions for each of the bounds separately. Moreover, we define them on a set of subsets of the topological space, which is not required to be a σ-algebra. In Remark 7.4, we motivate the choice for this definition.

Definition 1.1 Let $\mathcal{X}$ be a topological space and $\mathcal{A}$ be a set of subsets of $\mathcal{X}$. Let $I : \mathcal{X} \to [0, \infty]$ be lower semicontinuous. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{A}$. Let $(r_n)_{n\in\mathbb{N}}$ be an increasing sequence in $(0, \infty)$ with $\lim_{n\to\infty} r_n = \infty$. We say that $(\mu_n)_{n\in\mathbb{N}}$ satisfies a large deviation lower bound *on* $\mathcal{A}$ with rate function I and rates $(r_n)_{n\in\mathbb{N}}$ if

$$\liminf_{n\to\infty} \frac{1}{r_n} \log \mu_n(A) \geq -\inf I(A^\circ) \quad (A \in \mathcal{A}). \tag{1.2}$$

We say that $(\mu_n)_{n\in\mathbb{N}}$ satisfies a large deviation upper bound *on* $\mathcal{A}$ with rate function I and rates $(r_n)_{n\in\mathbb{N}}$ if

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \mu_n(A) \leq -\inf I(\overline{A}) \quad (A \in \mathcal{A}). \tag{1.3}$$

In the rest of the paper we only consider the rates $r_n = n$. However, the theory presented is still valid for general rates $(r_n)_{n\in\mathbb{N}}$. We say that $(\mu_n)_{n\in\mathbb{N}}$ satisfies a large deviation principle *on* $\mathcal{A}$ with rate function I whenever it satisfies both the large deviation lower bound and the large deviation upper bound with rate function I.

We omit "on $\mathcal{A}$" whenever $\mathcal{A}$ is the Borel σ-algebra $\mathcal{B}(\mathcal{X})$ on $\mathcal{X}$. In this case, the large deviation lower bound is satisfied if and only if the inequality in (1.2) holds for all open subsets of $\mathcal{X}$ and the large deviation upper bound is satisfied if and only if the inequality in (1.3) holds for all closed subsets of $\mathcal{X}$.

1.3 Main Results

See Sects. 3 and 4 for the definitions of the objects in the statements of the following theorems. In Sects. 6 and 7, we consider a more general situation. Theorem 1.2 is a consequence of Theorem 6.9, and Theorem 1.3 is a consequence of Theorem 7.5.

In this section $\mathcal{X}$ and $\mathcal{Y}$ are metric spaces.

Theorem 1.2 *Let $\pi : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ be given by $\pi(x, y) = y$. Suppose that $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of probability measures on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ that satisfies the large deviation principle with rate function $J : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ that has compact sublevel sets. Suppose that for each $n \in \mathbb{N}$, there exists a product regular conditional probability $\eta_n : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$ under μ_n with respect to π that is weakly continuous on $\operatorname{supp}(\mu_n \circ \pi^{-1})$, which we assume to be non-empty. Let $y \in \mathcal{Y}$ be such that $\inf J(\mathcal{X} \times \{y\}) < \infty$. Define $I : \mathcal{X} \to [0, \infty]$ by*

$$I(x) = J(x, y) - \inf J(\mathcal{X} \times \{y\}). \tag{1.4}$$

I has compact sublevel sets, and, for each $n \in \mathbb{N}$, η_n is unique on $\operatorname{supp}(\mu_n \circ \pi^{-1})$. Moreover,

$$\text{(A1)} \iff \text{(A2) and (B1)} \iff \text{(B2)},$$

where

(A1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n \to y$ and $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ for all n large enough,*[1] *the sequence $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation lower bound with rate function I.*

(A2) *For all $x \in \mathcal{X}$ and $r > 0$, with $U = B(x, r)$,*

$$\sup_{\varepsilon>0} \liminf_{n\to\infty} \inf_{\substack{z\in\mathcal{Y},\delta\in(0,\varepsilon)\\ B(z,\delta)\subset B(y,\varepsilon)}} \frac{1}{n} \log \mu_n\Big(\overline{U} \times \mathcal{Y} \,\Big|\, \mathcal{X} \times B(z, \delta)\Big) \geq -\inf I(U). \tag{1.5}$$

(B1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n \to y$ and $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ for all n large enough, the sequence $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation upper bound with rate function I.*

(B2) *For all $x_1, \ldots, x_k \in \mathcal{X}$ and $r_1, \ldots, r_k > 0$, with $W = \mathcal{X} \setminus [B(x_1, r_1) \cup \cdots \cup B(x_k, r_k)]$,*

[1] Meaning that there exists an $N \in \mathbb{N}$ such that $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ for all $n \geq N$.

$$\inf_{\varepsilon>0} \limsup_{n\to\infty} \sup_{\substack{z\in\mathcal{Y},\delta\in(0,\varepsilon)\\ B(z,\delta)\subset B(y,\varepsilon)}} \frac{1}{n}\log \mu_n\Big(W^\circ\times\mathcal{Y}\,\Big|\,\mathcal{X}\times B(z,\delta)\Big)\le -\inf I(W). \quad (1.6)$$

The next theorem is similar to Theorem 1.2, but considers the large deviation bounds for regular conditional kernels instead of product regular conditional probabilities.

Theorem 1.3 *Let $\tau:\mathcal{X}\to\mathcal{Y}$ be continuous. Suppose that $(\nu_n)_{n\in\mathbb{N}}$ is a sequence of probability measures on $\mathcal{B}(\mathcal{X})$ that satisfies the large deviation principle with rate function $J:\mathcal{X}\to[0,\infty]$ that has compact sublevel sets. Suppose that for each $n\in\mathbb{N}$ there exists a regular conditional probability $\eta_n:\mathcal{Y}\times\mathcal{B}(\mathcal{X})\to[0,1]$ under ν_n with respect to τ that is weakly continuous on $\operatorname{supp}(\nu_n\circ\tau^{-1})$, which is assumed to be non-empty. Let $y\in\mathcal{Y}$ be such that $\inf J(\tau^{-1}(\{y\}))<\infty$. Define $I:\mathcal{X}\to[0,\infty]$ by*

$$I(x)=\begin{cases} J(x)-\inf J(\tau^{-1}(\{y\})) & \tau(x)=y,\\ \infty & \tau(x)\neq y.\end{cases} \quad (1.7)$$

I has compact sublevel sets, and, for each $n\in\mathbb{N}$, η_n is unique on $\operatorname{supp}(\nu_n\circ\tau^{-1})$. Moreover,

$$\text{(A1)} \iff \text{(A2) } \textit{and } \text{(B1)} \iff \text{(B2)},$$

where

(A1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n\to y$ and $y_n\in\operatorname{supp}(\nu_n\circ\pi^{-1})$ for all n large enough, the sequence $(\eta_n(y_n,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation lower bound with rate function I.*
(A2) *For all $x\in\mathcal{X}$ and $r>0$, with $U=B(x,r)$,*

$$\sup_{\varepsilon>0}\liminf_{n\to\infty} \inf_{\substack{z\in\mathcal{Y},\delta\in(0,\varepsilon)\\ B(z,\delta)\subset B(y,\varepsilon)}} \frac{1}{n}\log \nu_n\Big(\overline{U}\,\Big|\,\tau^{-1}(B(z,\delta))\Big)\ge -\inf I(U). \quad (1.8)$$

(B1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n\to y$ and $y_n\in\operatorname{supp}(\nu_n\circ\pi^{-1})$ for all n large enough, the sequence $(\eta_n(y_n,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation upper bound with rate function I.*
(B2) *For all $x_1,\dots,x_k\in\mathcal{X}$ and $r_1,\dots,r_k>0$, with $W=\mathcal{X}\setminus[B(x_1,r_1)\cup\cdots\cup B(x_k,r_k)]$,*

$$\inf_{\varepsilon>0}\limsup_{n\to\infty} \sup_{\substack{z\in\mathcal{Y},\delta\in(0,\varepsilon)\\ B(z,\delta)\subset B(y,\varepsilon)}} \frac{1}{n}\log \nu_n\Big(W^\circ\,\Big|\,\tau^{-1}(B(z,\delta))\Big)\le -\inf I(W). \quad (1.9)$$

1.4 Gibbs–non-Gibbs Transitions and Future Research

In this section we discuss the relation between the large deviation results in this paper and Gibbs–non-Gibbs transitions in more detail. In particular, we discuss possible future directions regarding large deviations of conditional kernels.

The following situation for interacting particle systems occurs in the mean-field context (a similar context holds in the context of lattices). The initial system of so-called spins consists of distributions describing the interaction between spins via a potential V (for each n there is a distribution describing the law of n spins). This initial system is assumed to be Gibbs, which is called sequentially Gibbs in the mean-field context. Allowing the initial state to be transformed, for example, by an evolution of the spins, a question of interest is whether the transformed state is (sequentially) Gibbs. This question has been addressed in the mean-field context by Ermoleav and Külske [11] and by Fernández et al. [13] for $\{-1, +1\}$-valued spins, by den Hollander et al. [17] for $\mathbb{R}$-valued spins and by Külske and Opoku [19] and van Enter et al. [10] for compactly valued spins. In these papers, independent dynamics of the spins are considered (the evolution of each spin is independent of the evolution of the other spins). Independent dynamics simplify the situation. Namely, the evolved measure on either the product space of the initial and the final space, or—in case of an evolution—the space of trajectories, is a tilted measure of the evolved measure when considering $V = 0$. In this case the measure is a product measure, which means that the spins are independent. As a consequence (this will be clarified in a forthcoming paper), the conditional kernel η_n of the initial state on n spins with respect to the final state (for a fixed potential V) is a tilted version of the conditional kernel η_n^0 of the initial state with respect to the final state of independent spins (i.e. $V = 0$). Because of this tilting, by Varadhan's lemma, $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function $V + I_y - \inf(V + I_y)$ if $(\eta_n^0(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I_y. In the forthcoming paper, we will prove that the evolved sequence is sequentially Gibbs if $V + I_\zeta$ has a unique global minimiser.

The large deviation principle of $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ has been mentioned in the case of trajectories in [11, Corollary 2.4] and—as a corollary of that theorem—for the case of the product space of the initial and the final space in [13, Corollary 1.3]. However, no proof was given. Theorem 8.2 provides a rigorous proof of the large deviation principle statement in [13, Corollary 1.3]. In this paper, we do not provide a rigorous proof of [11, Corollary 2.4]. But Theorem 1.3 may be used, as the conditioning on the final state is a regular conditional kernel with respect to the map $\tau : C([0, T], \mathcal{X}) \to \mathcal{X}$, $\tau(f) = f(T)$.

In order to deal with empirical distributions (and not with magnetisations as is done in [17]), in future research we strive to "extend" the statement of Theorem 8.2 to infinite and possibly non-compact state spaces. In the case of non-compact spaces, it may be that topologies on the space of probability measures are considered that are not metrisable.

1.5 Outline

We list some notations, definitions and assumptions in Sect. 2. In Sect. 3, we give and compare the notions of regular conditional kernels, and we show that a regular conditional kernel under a measure ν is in fact a product regular conditional kernel under a measure that is related to ν. In Sect. 4, we introduce and study weakly continuous regular conditional kernels. In Sect. 5, we present some facts about lower semicontinuous functions with compact sublevel sets. Relying on the results of Sects. 4 and 5, in

Sect. 6, we present results on large deviation bounds for product regular conditional probabilities, in particular necessary and sufficient conditions for these bounds to hold. In Sect. 7, we discuss how to obtain large deviation bounds for regular conditional probabilities from the results in Sect. 6. In Sect. 8, we apply the theory to obtain the large deviation principle for the empirical density of the first coordinate given the empirical density of the second coordinate, for independent and identically distributed pairs of random variables. In Sect. 9, we give some examples. We also include an example for which the conditions are not satisfied. For this example we compare the quenched large deviations with large deviations of the weakly continuous regular conditional probabilities and comment on the difference with an example by La Cour and Schieve [20]. In "Appendices 1 and 2" we state some general results considering large deviations bounds that are used in the different sections. In "Appendix 3" we provide the proof of a theorem on which the examples of Sect. 9 rely.

2 Notations and Conventions

$\mathbb{N} = \{1, 2, 3, \dots\}$. For a topological space $\mathcal{X}$, we write $\mathcal{B}(\mathcal{X})$ for the Borel σ-algebra and $\mathcal{P}(\mathcal{X})$ and $\mathcal{M}(\mathcal{X})$ for the spaces of probability and signed measures on $\mathcal{B}(\mathcal{X})$, respectively. For $A \subset \mathcal{X}$ we write A° for the interior of A and $\overline{A}$ for the closure of A. For $x \in \mathcal{X}$ we write δ_x for the element in $\mathcal{P}(\mathcal{X})$ with $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise. For $x \in \mathcal{X}$ we write $\mathcal{N}_x$ for the set of $\mathcal{B}(\mathcal{X})$-measurable neighbourhoods of x. For a $\mu \in \mathcal{M}(\mathcal{X})$ we write $\operatorname{supp} \mu = \{x \in \mathcal{X} : |\mu|(V) > 0 \text{ for all } V \in \mathcal{N}_x\}$ and call this the support of μ. For a function f from a set $\mathcal{X}$ into $\mathbb{R}$ and $c \in \mathbb{R}$ we write $[f \geq c] = \{x \in \mathcal{X} : f(x) \geq c\}$. Similarly, we use the notations $[f > c]$, $[f \leq c]$ and $[f < c]$. Whenever $(x_\iota)_{\iota \in \mathbb{I}}$ is a net, where $\mathbb{I}$ is a directed set by (a direction) $\preceq$, we write $\liminf_{\iota \in \mathbb{I}} x_\iota = \sup_{\iota_0 \in \mathbb{I}} \inf_{\iota \succeq \iota_0, \iota \in \mathbb{I}} x_\iota$ (similarly lim sup). In particular, if $\mathcal{V} \subset \mathcal{N}_x$ and $\bigcap \mathcal{V} = \{x\}$ and $f : \mathcal{V} \to \mathbb{R}$, we write $\liminf_{V \in \mathcal{V}} f(V) = \sup_{V_0 \in \mathcal{V}} \inf_{V \subset V_0, V \in \mathcal{V}} f(V)$ (i.e. we consider $(f(V))_{V \in \mathcal{V}}$ as a net where $\mathcal{V}$ is directed by $\supset$ (as $\preceq$)).

Whenever we write $\mu(A|B)$, we implicitly assume that it is well defined (as $\mu(A \cap B)/\mu(B)$), i.e. that $\mu(B) \neq 0$.

We use the conventions $\log 0 = -\infty$ and $\inf I(\emptyset) = \infty$ whenever I is a function with values in $[0, \infty]$.

All measures in this paper are signed measures, unless mentioned otherwise.

3 Regular Conditional Kernels Being Product Regular Conditional Kernels

In this section we introduce the notion of a (product) regular conditional kernel. For an extensive study on regular conditional kernels, see Bogachev [4, Section 10.4]. The notion of a product regular conditional kernel does not appear in [4], but it does in Faden [12] and in Leao et al. [21]. Besides giving definitions, we make a few observations, of which Theorem 3.6 is used later on to derive statements of regular conditional kernels from statements of product regular conditional kernels.

In this section $(X, \mathcal{A})$, $(Y, \mathcal{B})$ are measurable spaces, ν is a measure on $\mathcal{A}$ and μ is a measure on $\mathcal{A} \otimes \mathcal{B}$, $\tau : X \to Y$ is measurable, and $\pi : X \times Y \to Y$ is given by $\pi(x, y) = y$.

Definition 3.1 A function $\eta : Y \times \mathcal{A} \to \mathbb{R}$ is called a ($\mathcal{B}$-)kernel if $\eta(\cdot, A)$ is ($\mathcal{B}$-)measurable for all $A \in \mathcal{A}$ and $\eta(y, \cdot)$ is a measure for all $y \in Y$. A kernel η is called a probability kernel if $\eta(y, \cdot)$ is a probability measure for all $y \in Y$.

Definition 3.2 Let $\eta : Y \times \mathcal{A} \to \mathbb{R}$ be a (probability) kernel.

(a) η is called a regular conditional kernel (regular conditional probability) under ν with respect to τ if

$$\nu(A \cap \tau^{-1}(B)) = \int_Y \mathbb{1}_B(y)\eta(y, A)\, \mathrm{d}\left[|\nu| \circ \tau^{-1}\right](y) \quad (A \in \mathcal{A}, B \in \mathcal{B}). \tag{3.1}$$

(b) η is called a product regular conditional kernel (product regular conditional probability) under μ with respect to π if

$$\mu(A \times B) = \int_Y \mathbb{1}_B(y)\eta(y, A)\, \mathrm{d}\left[|\mu| \circ \pi^{-1}\right](y) \quad (A \in \mathcal{A}, B \in \mathcal{B}). \tag{3.2}$$

3.3 Suppose that $\mathcal{E}$ is a sub-σ-algebra of $\mathcal{F}$. Let $(Y, \mathcal{B}) = (X, \mathcal{E})$ and $\mathrm{Id} : (X, \mathcal{A}) \to (Y, \mathcal{B})$ be the identity map. In agreement of [4, Definition 10.4.1] a kernel $\eta : Y \times \mathcal{A} \to \mathbb{R}$ is a regular conditional kernel under μ with respect to $\mathcal{E}$ if and only if η is a regular conditional kernel under μ with respect to Id.

3.4 Consider the two kernels $\eta : Y \times \mathcal{A} \to \mathbb{R}$ and $\xi : Y \times (\mathcal{A} \otimes \mathcal{B}) \to \mathbb{R}$, corresponding to each other by the formulas $\xi(y, F) = \int_{\mathcal{X}} \mathbb{1}_F(x, y)\, \mathrm{d}[\eta(y, \cdot)](x)$ and $\eta(y, A) = \xi(y, A \times Y)$. Then ξ is a regular conditional kernel under μ given π if and only if η is a product regular conditional kernel under μ given π.

In general, $X \times Y$ may be equipped with a σ-algebra $\mathcal{F}$ different from $\mathcal{A} \otimes \mathcal{B}$. In this situation, where μ is a measure on $\mathcal{F}$ and π is $\mathcal{F}$-measurable, the above correspondence cannot be used in general to reduce statements about product regular conditional kernels to statements about regular conditional kernels. See also example 4.5.

On the other hand, regular conditional probabilities can be seen as special cases of product regular conditional probabilities; see Theorem 3.6. In the present paper we use this to derive Theorem 1.3 from Theorem 1.2 but also Theorem 7.5 from Theorem 6.9.

Remark 3.5 If $\mathcal{A}$ is generated by a countable set, two regular conditional probabilities under a measure with respect to a σ-algebra (see 3.3) are almost everywhere equal (see Bogachev [4, Theorem 10.4.3]). Similarly one could state an analogous statement for regular conditional kernels with respect to measurable maps and for product regular conditional kernels. In Theorem 4.3 we prove that (product) regular conditional kernels are unique on the domain on which they are weakly continuous, in case the underlying topological space is perfectly normal. For such space the Borel σ-algebra may not be generated by a countable set.[2]

[2] The Sorgenfrey line, the space $\mathbb{R}$ with the right half-open interval topology, is perfectly normal but not second countable (see Steen and Seebach [27, Example 51]).

Theorem 3.6 (a) *There exists a measure $\tilde{\mu}$ on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ for which $\tilde{\mu}(A \times B) = \nu(A \cap \tau^{-1}(B))$.*

(b) *$\eta : Y \times \mathcal{A} \to \mathbb{R}$ is a regular conditional kernel under ν with respect to τ if and only if η is a product conditional kernel under $\tilde{\mu}$ with respect to π.*

Proof (a) We may assume ν to be positive, since $\nu = \nu^+ - \nu^-$. Let $\mathcal{E}$ be the set that consists of $\bigcup_{i=1}^n A_i \times B_i$, where $n \in \mathbb{N}$ and $A_i \in \mathcal{A}, B_i \in \mathcal{B}$ are such that $A_1 \times B_1, \ldots, A_n \times B_n$ are disjoint. Define $\nu^* : \mathcal{E} \to [0, \infty)$ by $\nu^*\left(\bigcup_{i=1}^n A_i \times B_i\right) = \nu\left(\bigcup_{i=1}^n A_i \cap \tau^{-1}(B_i)\right)$ for $A_1, \ldots, A_n \in \mathcal{A}$ and $B_1, \ldots, B_n \in \mathcal{B}$ as above. Checking that $\mathcal{E}$ is a ring of sets and that ν^* is σ-additive is left for the reader. The existence and unicity of the extension $\tilde{\mu}$ follow from the Carathéodory theorem (see Halmos [15, Section 13, Theorem A]).

(b) It follows from by definition of $\tilde{\mu}$ (note that $\nu \circ \tau^{-1} = \tilde{\mu} \circ \pi^{-1}$). □

4 Weakly Continuous Kernels

In this section we introduce the notion of weak continuity for kernels on topological spaces. In Theorem 4.3 we show uniqueness of (product) regular conditional kernels that are weakly continuous. In Theorems 4.6 and 4.8 we describe conditions that imply the existence of weakly continuous regular conditional probabilities. Similarly as is done in the Portmanteau theorem when one considers metric spaces, weak convergence implies lower bounds for open sets and upper bounds for closed sets, as is shown in Theorem 4.10. As described in Lemmas 4.11 and 4.12, these lim inf and lim sup bounds imply bounds for (product) regular conditional probabilities on which the results of Sects. 6 and 7 are based.

In this section $\mathcal{X}$ and $\mathcal{Y}$ are topological spaces, ν is a measure on $\mathcal{B}(\mathcal{X})$, μ is a measure on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$, $\tau : \mathcal{X} \to \mathcal{Y}$ is measurable, and $\pi : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ is given by $\pi(x, y) = y$.

Definition 4.1 We equip the space of measures, $\mathcal{M}(\mathcal{X})$, with the weak topology (generated by $C_b(\mathcal{X})$, which we denote by $\sigma(\mathcal{M}(\mathcal{X}), C_b(\mathcal{X}))$ as in the book of Schaefer [26, Chapter II, Section 5]). In this topology, a net $(\mu_\iota)_{\iota \in \mathbb{I}}$ in $\mathcal{M}(\mathcal{X})$ converges to a μ in $\mathcal{M}(\mathcal{X})$ if and only if $\int_{\mathcal{X}} f \, \mathrm{d}\mu_\iota \to \int_{\mathcal{X}} f \, \mathrm{d}\mu$ for all $f \in C_b(\mathcal{X})$.
Let $D \subset \mathcal{Y}$. A kernel $\eta : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$ is called weakly continuous on D if the map $D \to \mathcal{M}(\mathcal{X})$ given by $y \mapsto \eta(y, \cdot)$ is continuous in the weak topology. η is called *weakly continuous* if η is weakly continuous on $\mathcal{Y}$.

Theorem 4.2 *Let $\mathcal{X}$ be a perfectly normal*[3] *space and $\mu \in \mathcal{M}(\mathcal{X})$. Then*

$$\operatorname{supp} \mu = \left\{ x \in \mathcal{X} : \int_{\mathcal{X}} f \, \mathrm{d}|\mu| > 0 \text{ for all } f \in C(\mathcal{X}, [0, 1]) \text{ with } f(x) > 0 \right\}. \tag{4.1}$$

[3] Perfectly normal means that every open set in $\mathcal{X}$ is equal to $f^{-1}((0, \infty))$ for some $f \in C(\mathcal{X})$. All metric spaces are perfectly normal; Bogachev [4, Proposition 6.3.5].

Moreover, $|\mu|(\mathcal{X}\setminus\operatorname{supp}(\mu)) = 0$.[4] *As a consequence,* $\mu = 0$ *if and only if* $\int_{\mathcal{X}} f \, \mathrm{d}|\mu| = 0$ *for all* $f \in C_b(\mathcal{X})$.

Proof We may assume μ is positive. Let $x \in \operatorname{supp} \mu$. Then $\mu(V) > 0$ for all $V \in \mathcal{N}_x$. Let $f \in C(\mathcal{X}, [0,1])$ be such that $f(x) > 0$. Then $V = f^{-1}(0,\infty)$ has strictly positive measure. Since $\mu(V) = \lim_{n\to\infty} \int_{\mathcal{X}} \min\{nf, 1\} \, \mathrm{d}\mu$, there exists an n such that $\int_{\mathcal{X}} \min\{nf, 1\} \, \mathrm{d}\mu > 0$. Consequently, as $f \geq \frac{1}{n} \min\{nf, 1\}$, we have $\int_{\mathcal{X}} f \, \mathrm{d}\mu > 0$.

Let $x \in \mathcal{X}$ be such that $\int_{\mathcal{X}} f \, \mathrm{d}\mu > 0$ for all $f \in C(\mathcal{X}, [0,1])$ with $f(x) > 0$. Let $V \in \mathcal{N}_x$. As $V = f^{-1}(0,\infty)$ for some $f \in C(\mathcal{X}, [0,1])$, we have $\mu(V) \geq \int_{\mathcal{X}} f \, \mathrm{d}\mu > 0$. □

Theorem 4.3 *Suppose that* $\mathcal{X}$ *is a perfectly normal space.*

(a) *Let* η *and* ζ *be regular conditional kernels under* ν *with respect to* τ *that are weakly continuous on* $\operatorname{supp}(|\nu| \circ \tau^{-1})$. *Then* $\eta(y,\cdot) = \zeta(y,\cdot)$ *for all* $y \in \operatorname{supp}(|\nu| \circ \tau^{-1})$. *If* ν *is a probability measure, then* $\eta(y,\cdot)$ *is a probability measure for all* $y \in \operatorname{supp}(|\nu| \circ \tau^{-1})$.

(b) *Let* η *and* ζ *be product regular conditional kernels under* μ *with respect to* π *that are weakly continuous on* $\operatorname{supp}(|\mu| \circ \pi^{-1})$. *Then* $\eta(y,\cdot) = \zeta(y,\cdot)$ *for all* $y \in \operatorname{supp}(|\mu| \circ \pi^{-1})$. *If* μ *is a probability measure, then* $\eta(y,\cdot)$ *is a probability measure for all* $y \in \operatorname{supp}(|\mu| \circ \pi^{-1})$.

Proof We prove (a), and the proof of (b) is similar (replace "$|\nu| \circ \tau^{-1}$" by "$|\mu| \circ \pi^{-1}$"). To prove $\eta = \zeta$ on $D = \operatorname{supp}(|\nu| \circ \tau^{-1})$, by Theorem 4.2, it is sufficient to prove $\int_{\mathcal{X}} f \, \mathrm{d}\eta(y,\cdot) = \int_{\mathcal{X}} f \, \mathrm{d}\zeta(y,\cdot)$ for all $y \in D$ and all $f \in C_b(\mathcal{X})$. Let $f \in C_b(\mathcal{X})$. Because f is the uniform limit of simple functions, one has for all $B \in \mathcal{B}(\mathcal{Y})$

$$\int_{\mathcal{Y}} \mathbb{1}_B(y) \left[\int_{\mathcal{X}} f \, \mathrm{d}\eta(y,\cdot) \right] \mathrm{d}[|\nu| \circ \tau^{-1}](y) = \int_{\mathcal{Y}} \mathbb{1}_B(y) \left[\int_{\mathcal{X}} f \, \mathrm{d}\zeta(y,\cdot) \right] \mathrm{d}[|\nu| \circ \tau^{-1}](y). \tag{4.2}$$

Therefore there exists a set $Z \in \mathcal{B}(\mathcal{Y})$ with $|\nu| \circ \tau^{-1}(\mathcal{Y} \setminus Z) = 0$ such that

$$\int_{\mathcal{X}} f \, \mathrm{d}\eta(z,\cdot) = \int_{\mathcal{X}} f \, \mathrm{d}\zeta(z,\cdot) \quad (z \in Z). \tag{4.3}$$

Since both $y \mapsto \int_{\mathcal{X}} f \, \mathrm{d}\eta(y,\cdot)$ and $y \mapsto \int_{\mathcal{X}} f \, \mathrm{d}\zeta(y,\cdot)$ are weakly continuous on D, and Z is dense in D by Theorem 4.2, we have $\int_{\mathcal{X}} f \, \mathrm{d}\eta(y,\cdot) = \int_{\mathcal{X}} f \, \mathrm{d}\zeta(y,\cdot)$ for all $y \in D$. The second statement is proved by taking $f = \mathbb{1}_{\mathcal{X}}$. □

4.4 When η is a regular conditional kernel under ν with respect to τ, the value of the function $\eta(\cdot, A)$ on the complement of $\operatorname{supp}(|\nu| \circ \tau^{-1})$ is not determined, in the sense that if $\tilde{\eta}$ is a kernel with $\tilde{\eta}(y,\cdot) = \eta(y,\cdot)$ for all $y \in \operatorname{supp}(|\nu| \circ \tau^{-1})$, then $\tilde{\eta}$ is also a regular conditional kernel under ν with respect to τ.

[4] This is not true in general. For an example, see Bogachev [4, Example 7.1.3].

For example $\tilde{\eta}$ given by $\tilde{\eta}(y,\cdot) = \eta(y,\cdot)$ for $y \in \operatorname{supp}(|\nu| \circ \tau^{-1})$ and $\tilde{\eta}(y,\cdot) = \delta_x$ for $y \in \operatorname{supp}(|\nu| \circ \tau^{-1})^c$ for some chosen $x \in \mathcal{X}$, is such regular conditional kernel.

Whence if ν is a probability measure and there exists a regular conditional kernel under ν with respect to τ that is weakly continuous on $\operatorname{supp}(|\nu| \circ \tau^{-1})$, then we may as well assume this kernel to be a probability kernel. A similar statement is true for product regular conditional kernels.

4.5 By Theorem 3.6, statement (a) of Theorem 4.3 is a consequence of statement (b). In an attempt to reduce statement (b) to statement (a), the following problem occurs to the correspondence between regular conditional kernels and product regular conditional kernels that is mentioned in 3.4.

The Borel σ-algebra of $\mathcal{X} \times \mathcal{Y}$, i.e. $\mathcal{B}(\mathcal{X} \times \mathcal{Y})$, may be strictly larger than $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ (see, e.g. Bogachev [4, Lemma 6.4.1 and Example 6.4.3]). If this is the case, i.e. $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y}) \subsetneq \mathcal{B}(\mathcal{X} \times \mathcal{Y})$, and $\mathcal{B}(\mathcal{X} \times \mathcal{Y})$ equals the Baire-σ-algebra on $\mathcal{X} \times \mathcal{Y}$, i.e. the smallest σ-algebra that makes all continuous function $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ measurable, then there exists a continuous function $f \in C(\mathcal{X} \times \mathcal{Y})$ that is not $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$-measurable. Composing the function f with arctan, we obtain a $g \in C_b(\mathcal{X} \times \mathcal{Y})$ that is not measurable with respect to $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$. So if $\eta : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$ is a product regular conditional kernel under μ with respect to π, and $\xi : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y}) \to \mathbb{R}$ is as in Example 3.4, then g is not integrable with respect to $\xi(y,\cdot)$ for any $y \in \mathcal{Y}$.

$\mathcal{B}(\mathcal{X} \times \mathcal{Y})$ equals the Baire-σ-algebra if $\mathcal{X} \times \mathcal{Y}$ is a metric space (Bogachev [4, Proposition 6.3.4]). Therefore $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{\mathbb{R}}$ equipped with the discrete topology form an example for which the above is the case.

We state two theorems (Theorems 4.6 and 4.8) showing the existence of product regular conditional probabilities that are weakly continuous on $\operatorname{supp}(|\mu| \circ \pi^{-1})$.

Theorem 4.6 *Suppose that $\mathcal{Y}$ is countable and equipped with the discrete topology. Then $\eta : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$ defined by*

$$\eta(y, A) = \begin{cases} \mu(A \times \mathcal{Y} | \mathcal{X} \times \{y\}) & \mu(\mathcal{X} \times \{y\}) \neq 0, \\ 0 & \mu(\mathcal{X} \times \{y\}) = 0, \end{cases} \tag{4.4}$$

is a product regular conditional kernel under μ with respect to π that is weakly continuous on $\operatorname{supp}(|\mu| \circ \pi^{-1})$.

Proof It follows from the fact that $\mu(A \times B) = \sum_{y \in B} \mu(A \times \{y\})$ for $A \in \mathcal{B}(\mathcal{X})$, $B \in \mathcal{B}(\mathcal{Y})$. □

4.7 In case $\mathcal{Y}$ is first countable, the notion of open and closed sets and continuity of functions $\mathcal{Y} \to \mathbb{R}$ is characterised by the convergence of sequences. Therefore the following are equivalent for a kernel $\eta : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$

(a) η is weakly continuous in y.
(b) For all $(y_n)_{n \in \mathbb{N}}$ in $\mathcal{Y}$ with $y_n \to y$, one has $\eta(y_n,\cdot) \xrightarrow{w} \eta(y,\cdot)$.

The following theorem is an easy consequence of Lebesgue dominated convergence theorem.

Theorem 4.8 *Let $\mathcal{Y}$ be first countable. Let λ be a probability measure on $\mathcal{B}(\mathcal{X})$. Let $D \subset \mathcal{Y}$. Let $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a bounded $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$-measurable function such that $y \mapsto f(x, y)$ is continuous on D and equal to zero on $\mathcal{Y} \setminus D$ for λ-almost all $x \in \mathcal{X}$. Suppose that $\int_{\mathcal{X}} f(x, y)\ \mathrm{d}\lambda(x) > 0$ for all $y \in D$. If $\eta\colon\ \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$ is given by*

$$\eta(y, A) = \begin{cases} \frac{\int_{\mathcal{X}} \mathbb{1}_A(x) f(x,y)\ \mathrm{d}\lambda(x)}{\int_{\mathcal{X}} f(x,y)\ \mathrm{d}\lambda(x)} & y \in D, \\ 0 & y \notin D. \end{cases} \tag{4.5}$$

then η is weakly continuous on D (even strongly continuous, i.e. $y \mapsto \eta(y, A)$ is continuous for all $A \in \mathcal{B}(\mathcal{X})$). Let κ be a probability measure on $\mathcal{B}(\mathcal{Y})$ and assume $D = \operatorname{supp} \kappa$. Then η is a product regular conditional kernel under

$$\mu : \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y}) \to [0, 1], \quad \mu(A) = \frac{\int_{\mathcal{X}\times\mathcal{Y}} \mathbb{1}_A f\ \mathrm{d}[\lambda \otimes \kappa]}{\int_{\mathcal{X}\times\mathcal{Y}} f\ \mathrm{d}[\lambda \otimes \kappa]} \tag{4.6}$$

with respect to π, that is weakly continuous on $D = \operatorname{supp}(|\mu| \circ \pi^{-1})$.

Remark 4.9 In the above theorem the conditions may be weakened. Instead of assuming f to be bounded and λ, κ to be probability measures, we may as well assume that λ and κ are positive non-zero measures; that for all $y \in D$ there exists a $V \in \mathcal{N}_y$ and a λ-integrable $h : \mathcal{X} \to [0, \infty)$ such that $f(x, z) \le h(x)$ for all $x \in \mathcal{X}$; and all $z \in V \cap D$ and that f is $\lambda \otimes \kappa$-integrable.

In Sect. 6, the condition (b) of Theorem 4.10 is one of the key assumptions. If $\mathcal{X}$ is a metric space, this property follows from weak continuity as in the Portmanteau theorem. We state this in Theorem 4.10.

Theorem 4.10 *Let $\eta : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$ be a probability kernel. Let $D \subset \mathcal{Y}$, $y \in D$ and $\mathcal{V} \subset \mathcal{N}_y$ be such that $\bigcap \mathcal{V} = \{y\}$. Consider the following conditions.*

(a) $D \to \mathcal{M}(\mathcal{X})$, $y \mapsto \eta(y, \cdot)$ *is weakly continuous in y.*
(b) $\liminf_{\iota\in\mathbb{I}} \eta(y_\iota, G) \ge \eta(y, G)$ *for all open $G \subset \mathcal{X}$ and $(y_\iota)_{\iota\in\mathbb{I}}$ in D with $y_\iota \to y$.*
(c) $\limsup_{\iota\in\mathbb{I}} \eta(y_\iota, F) \le \eta(y, F)$ *for all closed $F \subset \mathcal{X}$ and $(y_\iota)_{\iota\in\mathbb{I}}$ in D with $y_\iota \to y$.*
(d) $\sup_{V\in\mathcal{V}} \inf_{v\in V\cap D} \eta(v, G) \ge \eta(y, G)$ *for all open sets $G \subset \mathcal{X}$.*
(e) $\inf_{V\in\mathcal{V}} \sup_{v\in V\cap D} \eta(v, F) \le \eta(y, F)$ *for all closed sets $F \subset \mathcal{X}$.*

(b), (c), (d), (e) *are equivalent. If $\mathcal{X}$ is metrisable, then* (a) *implies* (b). *If $\mathcal{X}$ is metrisable and $\mathcal{Y}$ is first countable, then* (a) *is equivalent to* (b) *and hence to* (c), (d), *and* (e).

Proof We leave it to the reader to check the equivalences between (b), (c), (d), (e). If $\mathcal{X}$ is a metric space, one can follow the lines of the Portmanteau theorem in the book of Billingsley [3, Theorem 2.1] for the implication (a) implies (b); the fact that the measures in the proof are indexed by the natural numbers instead of a general directed set $\mathbb{I}$ does not affect the argument. The proof of (b)$\Longrightarrow$(a) in the book of Billingsley relies on the Lebesgue dominated convergence theorem. But when $\mathcal{Y}$ is first countable, one can restrict to sequences (see 4.7) and obtain the implication (b)$\Longrightarrow$(a) as is done in the book of Billingsley. □

Lemma 4.11 *Assume that μ is a probability measure. Let η be a product regular conditional probability under μ with respect to π. Write $D = \mathrm{supp}(\mu \circ \pi^{-1})$ and let $y \in D$. Then for every $U \in \mathcal{N}_y$, one has $\mu(\mathcal{X} \times U) > 0$ and*

$$\inf_{v \in U \cap D} \eta(v, A) \leq \mu(A \times \mathcal{Y} | \mathcal{X} \times U) \leq \sup_{v \in U \cap D} \eta(v, A) \quad (A \in \mathcal{B}(\mathcal{X})). \tag{4.7}$$

Moreover, if $\mathcal{V} \subset \mathcal{N}_y$ is such that $\bigcap \mathcal{V} = \{y\}$ and η satisfies (b) of Theorem 4.10, then

$$\liminf_{V \in \mathcal{V}} \mu(G \times \mathcal{Y} | \mathcal{X} \times V) \geq \eta(y, G) \text{ for all open } G \subset \mathcal{X}, \tag{4.8}$$

$$\limsup_{V \in \mathcal{V}} \mu(F \times \mathcal{Y} | \mathcal{X} \times V) \leq \eta(y, F) \text{ for all closed } F \subset \mathcal{X}. \tag{4.9}$$

Proof Let $U \in \mathcal{N}_y$. Since $y \in D = \mathrm{supp}(\mu \circ \pi^{-1})$, one has $\mu(\mathcal{X} \times U) > 0$. (4.7) follows from the fact that for all $A \in \mathcal{B}(\mathcal{X})$

$$\begin{aligned} \frac{\mu(A \times U)}{\mu(\mathcal{X} \times U)} &= \frac{\int_{\mathcal{Y}} \mathbb{1}_U(y) \eta(y, A) \, \mathrm{d}[\mu \circ \pi^{-1}](y)}{\int_{\mathcal{Y}} \mathbb{1}_U(y) \, \mathrm{d}[\mu \circ \pi^{-1}](y)} \\ &= \frac{\int_{\mathcal{Y}} \mathbb{1}_{U \cap D}(y) \eta(y, A) \, \mathrm{d}[\mu \circ \pi^{-1}](y)}{\int_{\mathcal{Y}} \mathbb{1}_{U \cap D}(y) \, \mathrm{d}[\mu \circ \pi^{-1}](y)}. \end{aligned} \tag{4.10}$$

For an open $G \subset \mathcal{X}$ we have for $\mathcal{V}$ as above

$$\liminf_{V \in \mathcal{V}} \mu(G \times \mathcal{Y} | \mathcal{X} \times V) \geq \liminf_{V \in \mathcal{V}} \inf_{v \in V \cap D} \eta(v, G) = \sup_{V \in \mathcal{V}} \inf_{v \in V \cap D} \eta(v, G). \tag{4.11}$$

Thus (4.8) follows when assuming (d) of Theorem 4.10. Similarly, one obtains (4.9). □

For a regular conditional probability we have a similar statement; see Lemma 4.12. The proof can be done following the lines of the proof of Lemma 4.11 or as a consequence of Lemma 4.11 using Theorem 3.6.

Lemma 4.12 *Assume that ν is a probability measure. Let η be a regular conditional probability under ν with respect to τ. Write $D = \mathrm{supp}(\nu \circ \tau^{-1})$ and let $y \in D$. Then for every $U \in \mathcal{N}_y$, one has $\nu(\tau^{-1}(U)) > 0$ and*

$$\inf_{v \in U \cap D} \eta(v, A) \leq \nu(A | \tau^{-1}(U)) \leq \sup_{v \in U \cap D} \eta(v, A) \quad (A \in \mathcal{B}(\mathcal{X})). \tag{4.12}$$

Moreover, if $\mathcal{V} \subset \mathcal{N}_y$ is such that $\bigcap \mathcal{V} = \{y\}$ and η satisfies (b) of Theorem 4.10, then

$$\liminf_{V \in \mathcal{V}} \nu(G | \tau^{-1}(V)) \geq \eta(y, G) \text{ for all open } G \subset \mathcal{X}, \tag{4.13}$$

$$\limsup_{V \in \mathcal{V}} \nu(F | \tau^{-1}(V)) \leq \eta(y, F) \text{ for all closed } F \subset \mathcal{X}. \tag{4.14}$$

5 Some Facts About Functions with Compact Sublevel Sets

In this section we present some facts for functions with compact sublevel sets which are used in Sects. 6, 7 and 8.

In this section $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are topological spaces.

Definition 5.1 Let $J : \mathcal{X} \to [0, \infty]$. We call the set $[J \leq \alpha]$ (see Sect. 2) a sublevel set of J for $\alpha \in [0, \infty)$. J is said to be lower semicontinuous if all sublevels of J are closed. J is said to have compact sublevel sets if all sublevels of J are compact.

5.2 Let $J : \mathcal{X} \to [0, \infty]$ be lower semicontinuous. Then

$$J(x) = \sup_{G \in \mathcal{N}_x} \inf J(G). \tag{5.1}$$

Indeed, for all $\alpha < J(x)$ the set $[J > \alpha]$ is open and contains x.

Hence, a function $J : \mathcal{X} \to [0, \infty]$ is lower semicontinuous if and only if

$$\liminf_{\iota \in \mathbb{I}} J(x_\iota) \geq J(x) \tag{5.2}$$

for all $x \in \mathcal{X}$ and all nets $(x_\iota)_{\iota \in \mathbb{I}}$ in $\mathcal{X}$ that converge to x.

Lemma 5.3 *Let $\tau : \mathcal{Z} \to \mathcal{Y}$ be continuous. Let $J : \mathcal{Z} \to [0, \infty]$ have compact sublevel sets. Let $y \in \mathcal{Y}$ and $\mathcal{V} \subset \mathcal{N}_y$, $\bigcap \mathcal{V} = \{y\}$. Let $F \subset \mathcal{Z}$ be closed. Then*

$$\liminf_{V \in \mathcal{V}} \inf J(F \cap \tau^{-1}(\overline{V})) = \inf J(F \cap \tau^{-1}(\{y\})). \tag{5.3}$$

Consequently, if $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, then, for all closed $F \subset \mathcal{X}$ with $\inf J(F \times \{y\}) < \infty$,

$$\liminf_{V \in \mathcal{V}} \inf J(F \times \overline{V}) = \inf J(F \times \{y\}). \tag{5.4}$$

Proof The $\leq$ inequality in (5.3) is immediate. Because $\liminf_{V \in \mathcal{V}} \inf J(F \cap \tau^{-1}(\overline{V})) \geq \liminf_{V \in \mathcal{N}_y} \inf J(F \cap \tau^{-1}(\overline{V}))$, it is sufficient to prove

$$\alpha := \liminf_{V \in \mathcal{N}_y} \inf J(F \cap \tau^{-1}(\overline{V})) \geq \inf J(F \cap \tau^{-1}(\{y\})). \tag{5.5}$$

Note that $\alpha = \sup_{V \in \mathcal{N}_y} \inf J(F \cap \tau^{-1}(\overline{V}))$. If $\alpha = \infty$, there is nothing to prove. Suppose that $\alpha < \infty$. Whence $F \cap \tau^{-1}(\overline{V}) \cap [J \leq \alpha + \varepsilon] \neq \emptyset$ for all $V \in \mathcal{N}_y$ and all $\varepsilon > 0$. Since $[J \leq \alpha + \varepsilon]$ is compact, this implies that $\bigcap_{V \in \mathcal{N}_y} F \cap \tau^{-1}(\overline{V}) \cap [J \leq \alpha + \varepsilon] \neq \emptyset$, i.e. $\inf J(F \cap \tau^{-1}(\{y\})) \leq \alpha + \varepsilon$ for all $\varepsilon > 0$. □

5.4 The assumption that τ be continuous is not redundant, e.g. consider $\mathcal{Y} = \mathcal{Z} = [0, 1]$ and $J = \mathbb{1}_{(\frac{1}{2}, 1]}$ and τ given by $\tau(0) = 0$, $\tau(1) = 1$ and $\tau(x) = 1 - x$ for $x \in (0, 1)$, $F = [0, 1]$ and $y = 1$. Then, for all neighbourhoods V of y, $\tau^{-1}(V)$ contains the interval $(0, \varepsilon)$ for some $\varepsilon > 0$, whence $\inf J(F \cap \tau^{-1}(V)) = 0$ but $\inf J(F \cap \tau^{-1}(\{y\})) = J(1) = 1$.

Lemma 5.5 *Let $\mathcal{X}$ be normal and let $\mathcal{G}$ be a basis for the topology of $\mathcal{X}$. Let $J : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ have compact sublevel sets.*

(a) *For all open $G \subset \mathcal{X}$ and $\varepsilon > 0$, there exists a $U \in \mathcal{G}$ with $U \subset \overline{U} \subset G$ such that*

$$\inf J(G \times \{y\}) + \varepsilon \geq \inf J(U \times \{y\}). \tag{5.6}$$

(b) *For all closed $F \subset \mathcal{X}$ and $\alpha < \inf J(F \times \{y\})$, there exists $U_1, \ldots, U_k \in \mathcal{G}$ such that with $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$ one has $F \subset W^\circ \subset W$ and*

$$\alpha < \inf J(W \times \{y\}) \leq \inf J(W^\circ \times \{y\}) \leq \inf J(F \times \{y\}). \tag{5.7}$$

Proof (a) Let $\varepsilon > 0$. Let $x \in G$ be such that $J(x, y) \leq \inf J(G \times \{y\}) + \varepsilon$. Since $\mathcal{X}$ is a normal topological space, there exists an open set U with $x \in U \subset \overline{U} \subset G$. Because $\mathcal{G}$ is a basis, U may be chosen in $\mathcal{G}$. Then $\inf J(G \times \{y\}) + \varepsilon \geq J(x, y) \geq \inf J(U \times \{y\})$.

(b) Let $\beta > \alpha$ be such that $\beta < \inf J(F \times \{y\})$. The set $K := \{x \in \mathcal{X} : J(x, y) \leq \beta\}$ is a compact set that is disjoint from F. Whence there exists disjoint open $U, V \subset \mathcal{X}$ with $K \subset U$ and $F \subset V$. Since $\mathcal{G}$ is a basis and K is compact, there exists $U_1, \ldots, U_k$ in $\mathcal{G}$ with $K \subset U_1 \cup \cdots \cup U_k \subset U$. Then $\overline{U_1 \cup \cdots \cup U_k} \cap V = \emptyset$. Whence with $W := \mathcal{X} \setminus \overline{U_1 \cup \cdots \cup U_k}$, one has $F \subset W^\circ$ and $W \subset \mathcal{X} \setminus K$, which implies $\inf J(W \times \{y\}) \geq \beta > \alpha$. □

6 Large Deviations for Product Regular Conditional Probabilities

In this section we consider the following situation.

(i) $\mathcal{X}$ and $\mathcal{Y}$ are topological spaces, where $\mathcal{X}$ is normal.
(ii) $\mathcal{G}$ is a basis for the topology of $\mathcal{X}$ and $\mathcal{H}$ is a basis for the topology of $\mathcal{Y}$.
(iii) $\pi : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ is given by $\pi(x, y) = y$.
(iv) $(\mu_n)_{n \in \mathbb{N}}$ is a sequence of probability measures on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ satisfying the large deviation principle on $\{A \times B : A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y})\}$ with a rate function $J : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ that has compact sublevel sets.
(v) For each $n \in \mathbb{N}$ we assume the following: $\operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$,[5] there exists a product regular conditional probability $\eta_n : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$ under μ_n with respect to π, which satisfies the following continuity condition (see Theorem 4.10):

$$\begin{aligned} &\liminf_{\iota \in \mathbb{I}} \eta_n(y_\iota, G) \geq \eta_n(y, G) \text{ for all open } G \subset \mathcal{X} \\ &\text{and } (y_\iota)_{\iota \in \mathbb{I}} \text{ in } \operatorname{supp}(\mu_n \circ \pi^{-1}) \text{with } y_\iota \to y. \end{aligned} \tag{6.1}$$

(vi) Let $y \in \mathcal{Y}$. We assume that $\inf J(\mathcal{X} \times \{y\}) < \infty$ and that there exist $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ with $y_n \to y$. We define $I : \mathcal{X} \to [0, \infty]$ by

$$I(x) = J(x, y) - \inf J(\mathcal{X} \times \{y\}). \tag{6.2}$$

[5] As we are considering large deviation bound for $(\eta_n(y_n, \cdot))_{n \in \mathbb{N}}$ with $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$, we want such y_n to exist. Instead of this condition, one could of course deal with the situation where $\operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$ for some large N and consider sequences $(y_n)_{n \in \mathbb{N}}$ with $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ for $n \geq N$.

In this section we derive necessary and sufficient conditions for the large deviation bounds with rate function I for sequences of the form $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$. We prove this for general topological spaces instead of metric spaces as it does not cost more effort.

In Theorem 6.3 we consider a fixed sequence $(y_n)_{n\in\mathbb{N}}$ with $y_n \to y$ and describe equivalent conditions for the lower and upper large deviation bound to hold.

We are interested in the question whether for all sequences $(y_n)_{n\in\mathbb{N}}$ with $y_n \to y$ the sequence $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the lower and upper large deviation bound with rate function I. In Theorem 6.9 we give equivalent[6] and sufficient conditions for these bounds in a way that does not depend on sequences $(y_n)_{n\in\mathbb{N}}$ and the sets $(\mathcal{V}_n)_{n\in\mathbb{N}}$ as in Theorem 6.3.

Finally in 6.12 we comment on deriving Theorem 1.2 from Theorem 6.9.

But first we consider specific situations, providing a simple proof of the large deviation bounds with rate function I for sequences of the form $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$. Namely, we consider the case that $\mathcal{Y}$ is a discrete space (Theorem 6.1) and the case where μ_n is a product measure for all $n \in \mathbb{N}$ (Theorem 6.2).

Theorem 6.1 *Suppose that $\mathcal{Y}$ is countable and equipped with the discrete topology. Let $y \in \mathcal{Y}$ be such that* $\inf J(\mathcal{X} \times \{y\}) < \infty$. *For all $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{Y}$ with $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ and $y_n \to y$, the sequence $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I.*

Proof This basically follows from the following inequalities which follow from the large deviation principle and from Theorem 4.6.

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n(G \times \{y\}) \geq -\inf J(G \times \{y\}) \text{ for all open } G \subset \mathcal{X}, \tag{6.3}$$

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n(F \times \{y\}) \leq -\inf J(F \times \{y\}) \text{ for all closed } F \subset \mathcal{X}. \tag{6.4}$$

□

Theorem 6.2 (Independent coordinates) *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are second countable and $\mathcal{Y}$ is regular. Suppose that $\mu_n = \mu_n^1 \otimes \mu_n^2$ for some μ_n^1 on $\mathcal{B}(\mathcal{X})$ and μ_n^2 on $\mathcal{B}(\mathcal{Y})$ for all $n \in \mathbb{N}$. Then $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I for all sequences $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{Y}$. In particular, $\eta_n(y_n, \cdot) = \mu_n^1$ and $I(x) = \inf J(\{x\} \times \mathcal{Y})$.*

Proof I is lower semicontinuous (e.g. by 5.2) and for $c \in \mathbb{R}$ the set $[I \leq c]$ is a subset of the compact set $\{x \in \mathcal{X} : \exists z \in \mathcal{Y}, J(x, z) \leq c - \inf J(\mathcal{X} \times \{y\})\}$.

$[I \leq c] = \pi([J \leq c + \inf J(\mathcal{X} \times \{y\})])$. □

Theorem 6.3 *Let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{Y}$ with $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ that converges to y. For $n \in \mathbb{N}$ let $\mathcal{V}_n \subset \mathcal{N}_{y_n}$ be such that $\bigcap \mathcal{V}_n = \{y_n\}$. Then* (a2) $\iff$ (a3) $\iff$ (a1) *and* (b2) $\iff$ (b3) $\iff$ (b1)

(a1) *For all open $G \subset \mathcal{X}$*

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \eta_n(y_n, G) \geq -\inf I(G). \tag{6.5}$$

[6] Under the condition that $\mathcal{Y}$ is first countable.

(a2) *For all* $U \in \mathcal{G}$[7]

$$\liminf_{n\to\infty} \limsup_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y}|\mathcal{X} \times V) \geq -\inf I(U). \tag{6.6}$$

(a3) *For all open* $U \subset \mathcal{X}$*, one has*

$$\liminf_{n\to\infty} \liminf_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(U \times \mathcal{Y}|\mathcal{X} \times V) \geq -\inf I(U). \tag{6.7}$$

(b1) *For all closed* $F \subset \mathcal{X}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \eta_n(y_n, F) \leq -\inf I(F). \tag{6.8}$$

(b2) *For all* $U_1, \ldots, U_k \in \mathcal{G}$*, one has for* $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$

$$\limsup_{n\to\infty} \liminf_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(W^\circ \times \mathcal{Y}|\mathcal{X} \times V) \leq -\inf I(W). \tag{6.9}$$

(b3) *For all closed* $W \subset \mathcal{X}$

$$\limsup_{n\to\infty} \limsup_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(W \times \mathcal{Y}|\mathcal{X} \times V) \leq -\inf I(W). \tag{6.10}$$

Proof The implications (a3) $\Longrightarrow$ (a2) and (b3) $\Longrightarrow$ (b2) are immediate.

(a1) $\Longrightarrow$ (a3) Let $U \subset \mathcal{X}$ be an open set. By Lemma 4.11, (4.8),

$$\liminf_{n\to\infty} \liminf_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(U \times \mathcal{Y}|\mathcal{X} \times V) \geq \liminf_{n\to\infty} \frac{1}{n} \log \eta_n(y_n, U). \tag{6.11}$$

(b1) $\Longrightarrow$ (b3) Let $W \subset \mathcal{X}$ be a closed set. By Lemma 4.11, (4.9),

$$\limsup_{n\to\infty} \limsup_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(W \times \mathcal{Y}|\mathcal{X} \times V) \leq \limsup_{n\to\infty} \frac{1}{n} \log \eta_n(y_n, W). \tag{6.12}$$

(a2) $\Longrightarrow$ (a1). Let $G \subset \mathcal{X}$ be open. Let $\varepsilon > 0$ and U be as in Lemma 5.5(a). Then we obtain using Lemma 4.11

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \eta_n(y_n, G) &\geq \liminf_{n\to\infty} \frac{1}{n} \log \eta_n(y_n, \overline{U}) \\
&\geq \liminf_{n\to\infty} \limsup_{V\in\mathcal{V}_n} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y}|\mathcal{X} \times V) \\
&\geq -\inf I(U) = -\inf J(U \times \{y\}) + \inf J(\mathcal{X} \times \{y\}) \\
&\geq -\inf J(G \times \{y\}) + \inf J(\mathcal{X} \times \{y\}) - \varepsilon.
\end{aligned} \tag{6.13}$$

As this holds for all $\varepsilon > 0$, we conclude (6.5).

[7] Note that $\mu_n(\mathcal{X} \times V) > 0$ for all $n \in \mathbb{N}$ and $V \in \mathcal{N}_{y_n}$, as $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$.

(b2) $\Longrightarrow$ (b1). Let $\alpha < \inf J(F\times\{y\})$ and $U_1, \dots, U_k$ and W be as in Lemma 5.5(b). Then we obtain using Lemma 4.11

$$\begin{aligned}\limsup_{n\to\infty} \tfrac{1}{n}\log \eta_n(y_n, F) &\le \limsup_{n\to\infty} \tfrac{1}{n}\log \eta_n(y_n, W^\circ)\\ &\le \limsup_{n\to\infty} \liminf_{V\in\mathcal{V}_n} \tfrac{1}{n}\log \mu_n(W^\circ\times\mathcal{Y}|\mathcal{X}\times V)\\ &\le -\inf I(W) \le -\alpha + \inf J(\mathcal{X}\times\{y\}). \end{aligned} \tag{6.14}$$

As this holds for all $\alpha < \inf J(F\times\{y\})$, we conclude (6.8). □

6.4 (**Fixed** y) Note that if $y_n = y$ for all $n\in\mathbb{N}$, one can take $\mathcal{V}_n = \mathcal{V}$ for a $\mathcal{V}\subset\mathcal{N}_y$ with $\bigcap\mathcal{V} = \{y\}$. Then Theorem 6.3 implies that $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I if and only if (a2) and (b2) hold (with $\mathcal{V}_n = \mathcal{V}$).

6.5 Let $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{Y}$ be such that $y_n \in \operatorname{supp}(\mu_n\circ\pi^{-1})$ and $y_n \to y$. From Theorem 6.3 we derive that (a2) holds for some $\mathcal{V}_n\subset\mathcal{N}_{y_n}$ with $\bigcap\mathcal{V}_n = \{y_n\}$ if and only if (a2) holds for all such $\mathcal{V}_n$. Similarly, (b2) holds for some $\mathcal{V}_n\subset\mathcal{N}_{y_n}$ with $\bigcap\mathcal{V}_n = \{y_n\}$ if and only if (b2) holds for all such $\mathcal{V}_n\subset\mathcal{N}_{y_n}$.

In Lemma 6.7, we give a consequence of the large deviation principle of $(\mu_n)_{n\in\mathbb{N}}$. In Theorems 6.9 and 6.10 we use this to formulate sufficient conditions for upper or lower large deviation bounds on sequences $(\eta_n(y_n,\cdot))_{n\in\mathbb{N}}$ with $y_n\to y$ and sequences $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$.

We assumed $\mathcal{X}$ to be normal in this section. For Lemma 6.7 this assumption can be dropped.

6.6 For all neighbourhoods V of y one has by the large deviation principle

$$\liminf_{n\to\infty} \tfrac{1}{n}\log\mu_n(\mathcal{X}\times V) \ge -\inf J(\mathcal{X}\times V^\circ) \ge -\inf J(\mathcal{X}\times\{y\}) > -\infty. \tag{6.15}$$

In particular, there exists an $N\in\mathbb{N}$ such that $\mu_n(\mathcal{X}\times V) > 0$ for all $n\ge N$. Therefore $\mu_n(G\times\mathcal{Y}|\mathcal{X}\times V)$ is well defined for large n.

Lemma 6.7 (a) *For open* $G\subset\mathcal{X}$

$$\liminf_{V\in\mathcal{N}_y} \liminf_{n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0} \tfrac{1}{n}\log\mu_n(G\times\mathcal{Y}|\mathcal{X}\times V) \ge -\inf I(G). \tag{6.16}$$

(b) *For closed* $F\subset\mathcal{X}$

$$\limsup_{V\in\mathcal{N}_y} \limsup_{n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0} \tfrac{1}{n}\log\mu_n(F\times\mathcal{Y}|\mathcal{X}\times V) \le -\inf I(F). \tag{6.17}$$

Proof (a) Let $\varepsilon > 0$. By Lemma 5.3, there exists a $V_0\in\mathcal{N}_y$ such that for all $V\in\mathcal{N}_y$ with $V\subset V_0$

$$\inf J(\mathcal{X}\times\{y\}) \ge \inf J(\mathcal{X}\times\overline{V}) \ge \inf J(\mathcal{X}\times\overline{V}_0) \ge \inf J(\mathcal{X}\times\{y\}) - \varepsilon. \tag{6.18}$$

Let $V \in \mathcal{N}_y$ be such that $V \subset V_0$. As $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\mathcal{X} \times V) > -\infty$ (see 6.6), we can "split the lim inf in two" and we get by the large deviation principle and by (6.18)

$$\begin{aligned} &\liminf_{\substack{n\to\infty \\ n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0}} \frac{1}{n} \log \mu_n(G \times \mathcal{Y} | \mathcal{X} \times V) \\ &\quad = \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G \times V) - \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\mathcal{X} \times V) \\ &\quad \geq -\inf J(G \times \{y\}) + \inf J(\mathcal{X} \times \overline{V}) \geq -\inf I(G) - \varepsilon. \end{aligned} \tag{6.19}$$

(b) Let $\alpha < \inf J(F \times \{y\})$. There exists a neighbourhood V_0 of y such that for all neighbourhoods V of y with $V \subset V_0$

$$\inf J(F \times \{y\}) \geq \inf J(F \times \overline{V}) \geq \inf J(F \times \overline{V}_0) \geq \alpha. \tag{6.20}$$

Let $V \in \mathcal{N}_y$ be such that $y \in V \subset V_0$. Similarly as above, we get

$$\limsup_{\substack{n\to\infty \\ n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0}} \frac{1}{n} \log \mu_n(F \times \mathcal{Y} | \mathcal{X} \times V) \leq -\alpha + \inf J(\mathcal{X} \times \{y\}). \tag{6.21}$$

□

Theorem 6.8 *I has compact sublevel sets.*

Proof $[I \leq c] = \pi([J \leq c + \inf J(\mathcal{X} \times \{y\})])$. □

Theorem 6.9 *We have*

$$(\mathrm{A5}) \Longrightarrow (\mathrm{A4}) \iff (\mathrm{A3}) \Longrightarrow (\mathrm{A2}) \Longrightarrow (\mathrm{A1}),$$

and, if $\mathcal{Y}$ is first countable, then

$$(\mathrm{A1}) \iff (\mathrm{A2}),$$

where

(A1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ and $y_n \to y$, the sequence $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation lower bound with rate function I.*
(A2) *For all $U \in \mathcal{G}$*

$$\sup_{V_0 \in \mathcal{N}_y} \liminf_{n\to\infty} \inf_{\substack{V\in\mathcal{H}, V\subset V_0 \\ V\cap \operatorname{supp}(\mu_n\circ\pi^{-1})\neq\emptyset}} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq -\inf I(U). \tag{6.22}$$

(A3) *For all $U \in \mathcal{G}$*

$$\begin{aligned} &\sup_{V_0 \in \mathcal{N}_y} \liminf_{n\to\infty} \inf_{\substack{V\in\mathcal{H}, V\subset V_0 \\ V\cap \operatorname{supp}(\mu_n\circ\pi^{-1})\neq\emptyset}} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \\ &\quad \geq \liminf_{V\in\mathcal{N}_y} \liminf_{\substack{n\to\infty \\ n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0}} \frac{1}{n} \log \mu_n(U \times \mathcal{Y} | \mathcal{X} \times V). \end{aligned} \tag{6.23}$$

(A4) *For all* $U \in \mathcal{G}$ *we have* $\forall Z_0 \in \mathcal{N}_y\ \forall \varepsilon > 0 \exists V_0 \in \mathcal{N}_y\, \exists Z \in \mathcal{N}_y, Z \subset Z_0\ \forall M\, \exists m \geq M\, \exists N\ \forall n \geq N\ \forall V \in \mathcal{H}, V \subset V_0, V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$*:*

$$\tfrac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq \tfrac{1}{m} \log \mu_m(U \times \mathcal{Y} | \mathcal{X} \times Z) - \varepsilon. \tag{6.24}$$

(A5) *For all* $U \in \mathcal{G}$ *we have* $\forall \varepsilon > 0\ \forall V_0 \in \mathcal{N}_y \exists N \in \mathbb{N}\ \forall n \geq N\ \forall V \in \mathcal{H}, V \subset V_0, V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$*:*

$$\mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq e^{-n\varepsilon} \mu_n(U \times \mathcal{Y} | \mathcal{X} \times V_0). \tag{6.25}$$

Moreover,

$$\text{(B5)} \Longrightarrow \text{(B4)} \iff \text{(B3)} \Longrightarrow \text{(B2)} \Longrightarrow \text{(B1)},$$

and, if $\mathcal{Y}$ *is first countable, then*

$$\text{(B1)} \iff \text{(B2)},$$

where

(B1) *For all* $(y_n)_{n\in\mathbb{N}}$ *with* $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ *and* $y_n \to y$ *the sequence* $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ *satisfies the large deviation upper bound with rate function* I.
(B2) *For all* $U_1, \ldots, U_k \in \mathcal{G}$ *one has for* $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$

$$\inf_{V_0 \in \mathcal{N}_y} \limsup_{n\to\infty} \sup_{\substack{V \in \mathcal{H}, V \subset V_0 \\ V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset}} \tfrac{1}{n} \log \mu_n(W^\circ \times \mathcal{Y} | \mathcal{X} \times V) \leq -\inf I(W). \tag{6.26}$$

(B3) *For all* $U_1, \ldots, U_k \in \mathcal{G}$ *with* $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$

$$\begin{aligned} &\inf_{V_0 \in \mathcal{N}_y} \limsup_{n\to\infty} \sup_{\substack{V \in \mathcal{H}, V \subset V_0 \\ V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset}} \tfrac{1}{n} \log \mu_n(W^\circ \times \mathcal{Y} | \mathcal{X} \times V) \\ &\quad \leq \limsup_{V \in \mathcal{N}_y} \limsup_{\substack{n\to\infty \\ n \in \mathbb{N}: \mu_n(\mathcal{X} \times V) > 0}} \tfrac{1}{n} \log \mu_n(W \times \mathcal{Y} | \mathcal{X} \times V). \end{aligned} \tag{6.27}$$

(B4) *For all* $U_1, \ldots, U_k \in \mathcal{G}$ *with* $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$ *we have* $\forall Z_0 \in \mathcal{N}_y\ \forall \varepsilon > 0\, \exists V_0 \in \mathcal{N}_y\, \exists Z \in \mathcal{N}_y, Z \subset Z_0\ \forall M\, \exists m \geq M\, \exists N\ \forall n \geq N\ \forall V \in \mathcal{H}, V \subset V_0, V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$*:*

$$\tfrac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \leq \tfrac{1}{m} \log \mu_m(U \times \mathcal{Y} | \mathcal{X} \times Z) + \varepsilon. \tag{6.28}$$

(B5) *For all* $U_1, \ldots, U_k \in \mathcal{G}$ *with* $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$ *we have* $\forall \varepsilon > 0\ \forall V_0 \in \mathcal{N}_y\ \exists N \in \mathbb{N}\ \forall n \geq N\ \forall V \in \mathcal{H}, V \subset V_0, V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$*:*

$$\mu_n(W^\circ \times \mathcal{Y} | \mathcal{X} \times V) \leq e^{n\varepsilon} \mu_n(W \times \mathcal{Y} | \mathcal{X} \times V_0) \tag{6.29}$$

Proof The proofs of (B5) $\Longrightarrow$ (B4) $\Longleftrightarrow$ (B3) $\Longrightarrow$ (B2) $\Longrightarrow$ (B1) and of (B1) $\Longrightarrow$ (B2) are similar to the proofs of the following implications.

(A4) $\Longleftrightarrow$ (A3) follows by definition of sup, inf, lim sup and lim inf.

(A5) $\Longrightarrow$ (A3) Let $U \in \mathcal{G}$. Assuming (A5) we obtain $\forall \varepsilon > 0\ \forall V_0 \in \mathcal{N}_y\ \exists N \in \mathbb{N}\ \forall n \geq N$ and one has $\mu_n(\mathcal{X} \times V_0) > 0$ and

$$\inf_{\substack{V \in \mathcal{H}, V \subset V_0 \\ V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset}} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq \frac{1}{n} \log \mu_n(U \times \mathcal{Y} | \mathcal{X} \times V_0) - \varepsilon. \quad (6.30)$$

So $\forall \varepsilon > 0\ \forall V_0 \in \mathcal{N}_y$

$$\sup_{Z \in \mathcal{N}_y} \liminf_{n \to \infty} \inf_{\substack{V \in \mathcal{H}, V \subset Z \\ V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset}} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq \liminf_{\substack{n \to \infty \\ n \in \mathbb{N}: \mu_n(\mathcal{X} \times V_0) > 0}} \frac{1}{n} \log \mu_n(U \times \mathcal{Y} | \mathcal{X} \times V_0) - \varepsilon. \quad (6.31)$$

(A3) $\Longrightarrow$ (A2) Follows by Lemma 6.7.

(A2) $\Longrightarrow$ (A1). Suppose that (A2) holds. Let $U \in \mathcal{G}$ with $\inf J(U \times \{y\}) < \infty$ and let $\varepsilon > 0$. Let $V_0 \in \mathcal{N}_y$ and $N \in \mathbb{N}$ be such that $\frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq -\inf I(U) - \varepsilon$ for all $n \geq N$ and all $V \in \mathcal{H}$ with $V \subset V_0$ and $V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$. Let $(y_n)_{n \in \mathbb{N}}$ be such that $y_n \in \operatorname{supp}(\mu_n \circ \pi^{-1})$ and $y_n \to y$. Let $N_0 \geq N$ be such that $y_n \in V_0$ for all $n \geq N_0$. Then for all $n \geq N_0$ and $V \in \mathcal{N}_{y_n} \cap \mathcal{H}$ with $V \subset V_0$ we have $\frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \geq -\inf I(U) - \varepsilon$. This implies (a2) of Theorem 6.3 (with $\mathcal{V}_n = \mathcal{N}_{y_n} \cap \mathcal{H}$).

(A1) $\Longrightarrow$ (A2) (assuming $\mathcal{Y}$ is first countable). Suppose that (A2) does not hold. Let $(V_m)_{m \in \mathbb{N}}$ be a decreasing sequence in $\mathcal{H}$ with $\bigcap_{m \in \mathbb{N}} V_m = \{y\}$. Then there exists a $U \in \mathcal{G}$ with $\inf J(U \times \{y\}) < \infty$ and an $\alpha > \inf I(U)$ such that for all $M \in \mathbb{N}$ and $N \in \mathbb{N}$ there exist an $n \geq N$ and a $V \in \mathcal{H}$ with $V \subset V_M$ and $V \cap \operatorname{supp}(\mu_n \circ \pi^{-1}) \neq \emptyset$ such that

$$\frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V) \leq -\alpha. \quad (6.32)$$

Let $\beta < \alpha$ be such that $\beta > \inf I(U)$. By Lemma 4.11 we have

$$\inf_{z \in V \cap \operatorname{supp}(\mu_n \circ \pi^{-1})} \frac{1}{n} \log \eta_n(z, U) \leq \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y} | \mathcal{X} \times V). \quad (6.33)$$

For each $m \in \mathbb{N}$ there exist an n_m and a $y_{n_m} \in V_m \cap \operatorname{supp}(\mu_{n_m} \circ \pi^{-1})$ such that

$$\frac{1}{n_m} \log \eta_{n_m}(y_{n_m}, U) \leq -\beta. \quad (6.34)$$

We may choose $n_1 < n_2 < n_3 < \cdots$. With $y_k = y$ for $k \notin \{n_m : m \in \mathbb{N}\}$ we have $y_n \to y$ and

$$\liminf_{n \to \infty} \frac{1}{n} \log \eta_n(y_n, U) \leq \liminf_{m \to \infty} \frac{1}{n_m} \log \eta_{n_m}(y_{n_m}, U) \leq -\beta. \quad (6.35)$$

Therefore (a1) of Theorem 6.3 does not hold, which implies that (A1) does not hold. □

We can also use Lemma 6.7 and Theorem 6.3 (see also 6.4) to obtain sufficient conditions for the lower or upper large deviation bounds for $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$.

Theorem 6.10 *Let $\mathcal{V} \subset \mathcal{N}_y$ be such that $\bigcap \mathcal{V} = \{y\}$.*

(a) *Suppose that for all $U \in \mathcal{G}$ with $\inf J(U \times \{y\}) < \infty$*

$$\liminf_{n\to\infty} \limsup_{V\in\mathcal{V}} \frac{1}{n} \log \mu_n(\overline{U} \times \mathcal{Y}|\mathcal{X} \times V) \geq \liminf_{V\in\mathcal{N}_y} \liminf_{\substack{n\to\infty \\ n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0}} \frac{1}{n} \log \mu_n(U \times \mathcal{Y}|\mathcal{X} \times V). \tag{6.36}$$

Then $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation lower bound with rate function I.

(b) *Suppose that for all $U_1, \ldots, U_k \in \mathcal{G}$ with $W = \mathcal{X} \setminus (U_1 \cup \cdots \cup U_k)$*

$$\limsup_{n\to\infty} \liminf_{V\in\mathcal{V}} \frac{1}{n} \log \mu_n(W^\circ \times \mathcal{Y}|\mathcal{X} \times V) \leq \limsup_{V\in\mathcal{N}_y} \limsup_{\substack{n\to\infty \\ n\in\mathbb{N}:\mu_n(\mathcal{X}\times V)>0}} \frac{1}{n} \log \mu_n(W \times \mathcal{Y}|\mathcal{X} \times V). \tag{6.37}$$

Then $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation upper bound with rate function I.

6.11 (6.36) and (6.37) hold for example when $\forall \varepsilon > 0\ \forall V_0 \in \mathcal{V}\, \exists N \in \mathbb{N}\ \forall n \geq N\ \forall V \in \mathcal{V}, V \subset V_0$:

$$\mu_n(\overline{U} \times \mathcal{Y}|\mathcal{X} \times V) \geq e^{-n\varepsilon} \mu_n(\overline{U} \times \mathcal{Y}|\mathcal{X} \times V_0), \tag{6.38}$$

$$\mu_n(W^\circ \times \mathcal{Y}|\mathcal{X} \times V) \leq e^{n\varepsilon} \mu_n(W \times \mathcal{Y}|\mathcal{X} \times V_0), \tag{6.39}$$

respectively.

6.12 Theorem 1.2 is a consequence of Theorems 4.10, 6.8 and 6.9 with $\mathcal{G} = \{B(x,r) : x \in \mathcal{X}, r > 0\}$ and $\mathcal{H} = \{B(y,\delta) : y \in \mathcal{Y}, \delta > 0\}$.

7 Large Deviations for Regular Conditional Probabilities

In this section $\mathcal{X}$ and $\mathcal{Y}$ are topological spaces, $(\nu_n)_{n\in\mathbb{N}}$ is a sequence of probability measures on $\mathcal{B}(\mathcal{X})$ that satisfies the large deviation principle with rate function $K : \mathcal{X} \to [0,\infty]$, and $\tau : \mathcal{X} \to \mathcal{Y}$ is continuous. For more assumptions, see 7.2.

We derive the analogous statements as in Sect. 6 but for regular conditional kernels instead of product regular conditional kernels (7.3 and Theorem 7.5). First we show that with μ_n the probability measure corresponding on the product space corresponding to ν_n as in Theorem 3.6, the sequence $(\mu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle with a rate function described in terms of K (Theorem 7.1).

If $(\eta_n)_{n\in\mathbb{N}}$ are regular conditional probabilities under $(\nu_n)_{n\in\mathbb{N}}$ given τ, then one could also follow the proofs in Sect. 6 for the product regular conditional probabilities

to obtain similar results for large deviations for sequences of the form $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$. Instead, we make the approach via Theorem 3.6 to translate the results to the setting of regular conditional probabilities.

Theorem 7.1 *For all $n \in \mathbb{N}$ let μ_n be the probability measure on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ for which $\mu_n(A \times B) = \nu_n(A \cap \tau^{-1}(B))$ for $A \in \mathcal{B}(\mathcal{X})$, $B \in \mathcal{B}(\mathcal{Y})$ (as in Theorem 3.6). Then $(\mu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle on $\{A \times B : A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y})\}$ with rate function $J : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ given by*

$$J(x, y) = \begin{cases} K(x) & \tau(x) = y, \\ \infty & \tau(x) \neq y. \end{cases} \tag{7.1}$$

If K has compact sublevel sets, then so does J.

Proof By definition of J we have

$$\inf K(A \cap \tau^{-1}(B)) = \inf J(A \times B) \quad \big(A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y})\big). \tag{7.2}$$

Let $A \in \mathcal{B}(\mathcal{X})$ and $B \in \mathcal{B}(\mathcal{Y})$. Then

$$\begin{aligned} \liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n(A \times B) &= \liminf_{n\to\infty} \tfrac{1}{n} \log \nu_n(A \cap \tau^{-1}(B)) \\ &\geq -\inf K((A \cap \tau^{-1}(B))^\circ). \end{aligned} \tag{7.3}$$

We have $(A \cap \tau^{-1}(B))^\circ = A^\circ \cap \tau^{-1}(B)^\circ$ and $\tau^{-1}(B)^\circ \supset \tau^{-1}(B^\circ)$, whence

$$\begin{aligned} \inf K((A \cap \tau^{-1}(B))^\circ) &\leq \inf K(A^\circ \cap \tau^{-1}(B^\circ)) \\ &= \inf J(A^\circ \times B^\circ) = \inf J((A \times B)^\circ). \end{aligned} \tag{7.4}$$

Similarly

$$\begin{aligned} \limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n(A \times B) &= \limsup_{n\to\infty} \tfrac{1}{n} \log \nu_n(A \cap \tau^{-1}(B)) \\ &\leq -\inf K(\overline{A \cap \tau^{-1}(B)}). \end{aligned} \tag{7.5}$$

We have $\overline{A \cap \tau^{-1}(B)} \subset \overline{A} \cap \overline{\tau^{-1}(B)}$ and $\overline{\tau^{-1}(B)} \subset \tau^{-1}(\overline{B})$, whence

$$\inf K(\overline{A \cap \tau^{-1}(B)}) \geq \inf K(\overline{A} \cap \tau^{-1}(\overline{B})) = \inf J(\overline{A \times B}). \tag{7.6}$$

Suppose that K has compact sublevel sets. Let $c \geq 0$. Then $[J \leq c]$ is contained in the compact set $[K \leq c] \times \tau([K \leq c])$. By Theorem 6.8 I has compact sublevel sets. □

7.2 In the rest of this section $\mathcal{X}$ is normal, and $\mathcal{G}, \mathcal{H}, \pi$ are as in (ii) and (iii) of Sect. 6. Furthermore similarly to (v) and (vi) of Sect. 6, we assume the following.

(v)* For each $n \in \mathbb{N}$ we assume the following: $\operatorname{supp}(\nu_n \circ \tau^{-1}) \neq \emptyset$, there exists a regular conditional probability $\eta_n : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0, 1]$ under ν_n with respect to τ, satisfying the continuity condition (6.1).

(vi)* Let $y \in \mathcal{Y}$. We assume that $\inf K(\tau^{-1}(\{y\})) < \infty$ and that there exists $y_n \in \operatorname{supp}(\nu_n \circ \tau^{-1})$ with $y_n \to y$. Let $I : \mathcal{X} \to [0, \infty]$ be given by

$$\begin{aligned} I(x) &= J(x, y) - \inf J(\mathcal{X} \times \{y\}) \\ &= \begin{cases} K(x) - \inf K(\tau^{-1}(\{y\})) & \tau(x) = y, \\ \infty & \tau(x) \neq y. \end{cases} \end{aligned} \tag{7.7}$$

7.3 As by Theorem 3.6 η_n is the product regular conditional kernel under μ_n with respect to π; by Theorem 7.1 $(\mu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle on $\{A \times B : A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y})\}$ with rate function J, and $\inf J(\mathcal{X} \times \{y\}) = \inf K(\tau^{-1}(y)) < \infty$ and μ_n and η_n are as in Sect. 6 (in the sense that (iv), (v), (vi) hold). Therefore we can translate the results of Sect. 6, but also the results of Sects. 4 and 5, using, for example (7.2), $\nu_n \circ \tau^{-1} = \mu_n \circ \pi^{-1}$ and that for $V \in \mathcal{B}(\mathcal{Y})$ with $\nu_n(\tau^{-1}(V)) > 0$ and for $A \in \mathcal{B}(\mathcal{X})$

$$\nu_n(A|\tau^{-1}(V)) = \mu_n(A \times \mathcal{X}|\mathcal{X} \times V). \tag{7.8}$$

In this sense also Theorem 1.3 follows from Theorem 1.2. We present some of the equivalent statements of Theorem 6.9 in Theorem 7.5.

Remark 7.4 Because of the relation between μ_n and ν_n and between K and J, in Theorem 7.1 we were able to prove the large deviation principle on $\{A \times B : A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y})\}$. Whether it can be extended to the large deviation principle on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ is a priori not clear. However, for the purpose of using the results of Sect. 6, this is not required (as only (iv) of Sect. 6 is required). This is the main reason to define the large deviation bounds as in Definition 1.1.

Theorem 7.5 (A3) $\Longrightarrow$ (A2) $\Longrightarrow$ (A1). *If $\mathcal{Y}$ is first countable, then* (A1) $\Longleftrightarrow$ (A2).

(A1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \operatorname{supp}(\nu_n \circ \tau^{-1})$ and $y_n \to y$ the sequence $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation lower bound with rate function I.*

(A2) *For all $U \in \mathcal{G}$*

$$\sup_{V_0 \in \mathcal{N}_y} \liminf_{n\to\infty} \inf_{\substack{V \in \mathcal{H}, V \subset V_0 \\ V \cap \operatorname{supp}(\nu_n \circ \tau^{-1}) \neq \emptyset}} \frac{1}{n} \log \nu_n(\overline{U}|\tau^{-1}(V)) \geq -\inf I(U). \tag{7.9}$$

(A3) *For all $U \in \mathcal{G}$*

$$\begin{aligned} &\sup_{V_0 \in \mathcal{N}_y} \liminf_{n\to\infty} \inf_{\substack{V \in \mathcal{H}, V \subset V_0 \\ V \cap \operatorname{supp}(\nu_n \circ \tau^{-1}) \neq \emptyset}} \frac{1}{n} \log \nu_n(\overline{U}|\tau^{-1}(V)) \\ &\geq \liminf_{V \in \mathcal{N}_y} \liminf_{\substack{n\to\infty \\ n \in \mathbb{N}: \nu_n(\tau^{-1}(V)) > 0}} \frac{1}{n} \log \nu_n(U|\tau^{-1}(V)). \end{aligned} \tag{7.10}$$

(B3) $\Longrightarrow$ (B2) $\Longrightarrow$ (B1). *If $\mathcal{Y}$ is first countable, then* (B1) $\Longleftrightarrow$ (B2).

(B1) *For all $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \operatorname{supp}(\nu_n \circ \tau^{-1})$ and $y_n \to y$ the sequence $(\eta_n(y_n,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation upper bound with rate function I.*

(B2) *For all $U_1,\dots,U_k \in \mathcal{G}$ one has for $W = \mathcal{X}\setminus(U_1\cup\cdots\cup U_k)$*

$$\inf_{V_0\in\mathcal{N}_y} \limsup_{n\to\infty} \sup_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\operatorname{supp}(\nu_n\circ\tau^{-1})\neq\emptyset}} \frac{1}{n}\log \nu_n(W^\circ|\tau^{-1}(V)) \le -\inf I(W). \tag{7.11}$$

(B3) *For all $U_1,\dots,U_k \in \mathcal{G}$ with $W = \mathcal{X}\setminus(U_1\cup\cdots\cup U_k)$*

$$\begin{aligned}&\inf_{V_0\in\mathcal{N}_y} \limsup_{n\to\infty} \sup_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\operatorname{supp}(\nu_n\circ\tau^{-1})\neq\emptyset}} \frac{1}{n}\log \nu_n(W^\circ|\tau^{-1}(V))\\ &\le \limsup_{V\in\mathcal{N}_y} \limsup_{\substack{n\to\infty\\ n\in\mathbb{N}:\nu_n(\tau^{-1}(V))>0}} \frac{1}{n}\log \nu_n(W|\tau^{-1}(V)).\end{aligned} \tag{7.12}$$

8 An Application to Conditional Probabilities of Empirical Distributions on Finite Sets

In terms of random variables, Sanov's theorem gives us the large deviation principle of empirical densities $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where $X_1, X_2, \dots$ are independent and identically distributed random variables. We consider large deviations of $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ conditioning on $\frac{1}{n}\sum_{i=1}^n \delta_{Y_i} = \psi_n$, where $(X_1,Y_1),(X_2,Y_2),\dots$ are independent and identically distributed couples of random variables, both random variables attaining their values in a finite set. This large deviation principle is formalised in Theorem 8.2.

In this section we consider the following.

- Let $\mathcal{R}$ and $\mathcal{S}$ be finite sets equipped with the discrete topology (discrete metric). Let $\mathcal{P}(\mathcal{R})$, $\mathcal{P}(\mathcal{S})$ and $\mathcal{P}(\mathcal{R}\times\mathcal{S})$ be equipped by the weak topology, and let $\mathfrak{d}$ denote the Prohorov metric (see Billingsley [3, Appendix III]) on each of the spaces.
- Let $\lambda \in \mathcal{P}(\mathcal{R}\times\mathcal{S})$. We assume $\lambda(\mathcal{R}\times\{s\}) > 0$ for all $s\in\mathcal{S}$.
- For $n\in\mathbb{N}$ let $L_n : \mathcal{R}^n \to \mathcal{P}(\mathcal{R})$ be given by $L_n(r) = \frac{1}{n}\sum_{i=1}^n \delta_{r_i}$ for $r = (r_1,\dots,r_n)\in\mathcal{R}^n$.
- Write $\mathcal{P}^n_{emp}(\mathcal{R}) = L_n(\mathcal{R}^n) = \{\frac{1}{n}\sum_{i=1}^n \delta_{r_i} : r_1,\dots,r_n \in \mathcal{R}\}$, similarly $\mathcal{P}^n_{emp}(\mathcal{S}) = L_n(\mathcal{S}^n)$ and $\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S}) = L_n((\mathcal{R}\times\mathcal{S})^n)$.
- Let $\mathfrak{m} : \mathcal{P}(\mathcal{R}\times\mathcal{S}) \to \mathcal{P}(\mathcal{R})\times\mathcal{P}(\mathcal{S})$ be the map that maps a measure in $\mathcal{P}(\mathcal{R}\times\mathcal{S})$ onto the pair of its marginals, i.e. $\mathfrak{m}$ is given by

$$\mathfrak{m}(\xi) = \big(\xi(\cdot\times\mathcal{S}), \xi(\mathcal{R}\times\cdot)\big). \tag{8.1}$$

- Let $\pi : \mathcal{P}(\mathcal{R})\times\mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{S})$ be the map given by $\pi(\xi,\zeta) = \zeta$.

- Let μ_n be the probability measure on $\mathcal{B}(\mathcal{P}(\mathcal{R})) \otimes \mathcal{B}(\mathcal{P}(\mathcal{S}))$ defined by $\mu_n = \left(\bigotimes_{i=1}^n \lambda\right) \circ L_n^{-1} \circ \mathfrak{m}^{-1}$, so that for $A \in \mathcal{B}(\mathcal{P}(\mathcal{R}))$ and $B \in \mathcal{B}(\mathcal{P}(\mathcal{S}))$

$$\mu_n(A \times B) = \left(\bigotimes_{i=1}^n \lambda\right)(L_n^{-1}(A) \times L_n^{-1}(B)). \tag{8.2}$$

- Define $\theta : \mathcal{S} \times \mathcal{B}(\mathcal{R}) \to [0, 1]$ by $\theta(s, A) = \lambda(A \times \mathcal{S} | \mathcal{R} \times \{s\})$.
- Define $\eta_n : \mathcal{P}(\mathcal{S}) \times \mathcal{B}(\mathcal{P}(\mathcal{R})) \to [0, 1]$ by

$$\eta_n(\xi, A) = \begin{cases} \left[\bigotimes_{i=1}^n \theta(s_i, \cdot)\right] \circ L_n^{-1}(A) & \xi \in \mathcal{P}_{emp}^n(\mathcal{S}), \xi = L_n(s_1, \ldots, s_n) \\ & \text{for } s_1, \ldots, s_n \in \mathcal{S}, \\ 0 & \xi \notin \mathcal{P}_{emp}^n(\mathcal{S}). \end{cases} \tag{8.3}$$

- Let $J : \mathcal{P}(\mathcal{R}) \times \mathcal{P}(\mathcal{S}) \to [0, \infty]$ be given by

$$J(\rho, \sigma) = \inf_{\xi \in \mathfrak{m}^{-1}(\{(\rho,\sigma)\})} H(\xi | \lambda). \tag{8.4}$$

  where $H(\xi|\lambda)$ is the relative entropy of ξ with respect to λ ([7, Definition 2.1.5]).
- Let $\psi \in \mathcal{P}(\mathcal{S})$ be such that

$$\inf_{\xi \in \mathfrak{m}^{-1}(\mathcal{P}(\mathcal{R}) \times \{\psi\})} H(\xi | \lambda) < \infty. \tag{8.5}$$

8.1 We present some facts which follow from the assumptions with little effort; to some facts we give some explanation or references.

(a) $\mathcal{P}_{emp}^n(\mathcal{S})$ is closed in $\mathcal{P}(\mathcal{S})$. Moreover, if ξ_k and ξ in $\mathcal{P}_{emp}^n(\mathcal{S})$ are such that $\xi_k \to \xi$, then there exist s_{ki} and q_i in $\mathcal{S}$ for $i \in \{1, \ldots, n\}$ such that $\xi_k = L_n((s_{k1}, \ldots, s_{kn}))$, $\xi = L_n((q_1, \ldots, q_n))$ and $s_{ki} \to q_i$ for all $i \in \{1, \ldots, n\}$.
(b) $\text{supp}(\mu_n \circ \pi^{-1}) = \mathcal{P}_{emp}^n(\mathcal{S})$.
(c) η_n is a product regular conditional kernel under μ_n with respect to π that is weakly continuous on $\mathcal{P}_{emp}^n(\mathcal{S})$.
(d) $(\bigotimes \lambda^n \circ L_n^{-1})_{n \in \mathbb{N}}$ satisfies the large deviation principle with rate function $H(\cdot|\lambda)$.
(e) $\mathfrak{m}$ is continuous.
(f) $(\mu_n)_{n \in \mathbb{N}}$ satisfies the large deviation principle with rate function J.

(a) follows from the fact that $\mathcal{S}$ is a finite space. (b) follows from (a), from the fact that the complement of $\mathcal{P}_{emp}^n(\mathcal{S})$ has $\mu_n \circ \pi^{-1}$-measure zero and because $\mu_n \circ \pi^{-1}(\{L_n(s)\}) > 0$ for all $s \in \mathcal{S}^n$, which is due to the assumptions on λ. (c) follows by a straightforward calculation, and the continuity follows from (a). For (d) see Sanov's theorem (Dembo and Zeitouni [7, Theorem 6.2.10]). (e) follows from the fact that if $\xi_n \to \xi$ in $\mathcal{P}(\mathcal{R} \times \mathcal{S})$, then the $\mathcal{R}$- and $\mathcal{S}$-marginals of ξ_n converge to the $\mathcal{R}$- and $\mathcal{S}$-marginals of ξ, respectively. Then (f) follows from (a) and (d) by the contraction principle [7, Theorem 4.2.1].

In the rest of this section, we prove the following theorem.

Theorem 8.2 *For all $(\psi_n)_{n\in\mathbb{N}}$ with $\psi_n \in \mathcal{P}^n_{emp}(\mathcal{S})$ and $\psi_n \to \psi$, the sequence $(\eta_n(\psi_n,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function $I : \mathcal{P}(\mathcal{R}) \to [0,\infty]$, given by*

$$I(\phi) = \inf_{\xi\in\mathfrak{m}^{-1}(\{(\phi,\psi)\})} H(\xi|\lambda) - \inf_{\xi\in\mathfrak{m}^{-1}(\mathcal{P}(\mathcal{R})\times\{\psi\})} H(\xi|\lambda). \tag{8.6}$$

I is continuous on $[I<\infty]$.

As $\mathcal{P}(\mathcal{S})$ is first countable, it is sufficient to show that (A2) and (B2) of Theorem 6.9 hold. In 8.4 we use the bounds of Lemma 8.3 to derive other bounds which imply (A2) and (B2). The continuity of I follows by continuity of the map $\nu \mapsto H(\nu|\lambda)$ (Lemma 8.5).

Lemma 8.3 *[7, Lemma 2.1.9] For $\nu \in \mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})$ one has, with $M = (\#\mathcal{R})(\#\mathcal{S})$,*

$$(n+1)^{-M} e^{-nH(\nu|\lambda)} \le \left[\bigotimes_{i=1}^n \lambda\right](L_n^{-1}(\{\nu\})) \le e^{-nH(\nu|\lambda)}. \tag{8.7}$$

8.4 From Lemma 8.3 we obtain the following bounds for $A \in \mathcal{B}(\mathcal{P}(\mathcal{R}))$ and $B \in \mathcal{B}(\mathcal{P}(\mathcal{S}))$.

$$\begin{aligned}\mu_n(A\times B) &\le \#L_n^{-1}(A)\#L_n^{-1}(B) e^{-n\inf_{\nu\in\mathfrak{m}^{-1}(A\times B)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda)} \\ &\le (n+1)^M e^{-n\inf_{\nu\in\mathfrak{m}^{-1}(A\times B)} H(\nu|\lambda)}, \qquad (8.8)\\ \mu_n(A\times B) &\ge (n+1)^{-M} e^{-n\inf_{\nu\in\mathfrak{m}^{-1}(A\times B)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda)}. \qquad (8.9)\end{aligned}$$

Whence

$$\begin{aligned}&\frac{1}{n}\log\left[(n+1)^{-2M}\right] - \left[\inf_{\nu\in\mathfrak{m}^{-1}(A\times B)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda) - \inf_{\xi\in\mathfrak{m}^{-1}(\mathcal{R}\times B)} H(\xi|\lambda)\right] \\ &\quad\le \frac{1}{n}\log\mu_n(A\times\mathcal{S}|\mathcal{R}\times B) \\ &\quad\le \frac{1}{n}\log\left[(n+1)^{2M}\right] - \left[\inf_{\nu\in\mathfrak{m}^{-1}(A\times B)} H(\nu|\lambda) - \inf_{\xi\in\mathfrak{m}^{-1}(\mathcal{R}\times B)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\xi|\lambda)\right].\end{aligned} \tag{8.10}$$

In order to derive (A2) and (B2) of Theorem 6.9, we make the following observation. By (8.10) we have for an open U and a closed W that if for both $A = U$ and $C = \mathcal{R}$

as well as $A = \mathcal{R}$ and $C = W$ we have

$$\inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\mathcal{P}^n_{emp}(\mathcal{S})\neq\emptyset}} \left[\inf_{\nu\in\mathfrak{m}^{-1}(\overline{A}\times V)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda) - \inf_{\xi\in\mathfrak{m}^{-1}(C^\circ\times V)} H(\xi|\lambda)\right]$$
$$\le \inf_{\nu\in\mathfrak{m}^{-1}(A\times\{\psi\})} H(\nu|\lambda) - \inf_{\xi\in\mathfrak{m}^{-1}(C\times\{\psi\})} H(\xi|\lambda), \tag{8.11}$$

then

$$\sup_{V_0\in\mathcal{N}_\psi} \liminf_{n\to\infty} \inf_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\mathrm{supp}(\mu_n\circ\pi^{-1})\neq\emptyset}} \frac{1}{n}\log\mu_n(\overline{U}\times\mathcal{S}|\mathcal{R}\times V) \ge -\inf I(U), \tag{8.12}$$

$$\inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\mathrm{supp}(\mu_n\circ\pi^{-1})\neq\emptyset}} \frac{1}{n}\log\mu_n(W^\circ\times\mathcal{S}|\mathcal{R}\times V) \le -\inf I(W), \tag{8.13}$$

As

$$\inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\mathcal{P}^n_{emp}(\mathcal{S})\neq\emptyset}} \left[\inf_{\nu\in\mathfrak{m}^{-1}(A\times V)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda) - \inf_{\xi\in\mathfrak{m}^{-1}(C\times V)} H(\xi|\lambda)\right]$$
$$\le \inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\mathcal{P}^n_{emp}(\mathcal{S})\neq\emptyset}} \inf_{\nu\in\mathfrak{m}^{-1}(A\times V)\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda)$$
$$- \sup_{V_0\in\mathcal{N}_\psi} \inf_{\substack{V\in\mathcal{H},V\subset V_0\\ V\cap\mathrm{supp}(\mu_n\circ\pi^{-1})\neq\emptyset}} \inf_{\xi\in\mathfrak{m}^{-1}(C\times V)} H(\xi|\lambda)$$
$$\le \inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\zeta\in\mathcal{P}^n_{emp}(\mathcal{S})\cap V_0} \inf_{\nu\in\mathfrak{m}^{-1}(A\times\{\zeta\})\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda)$$
$$- \sup_{V\in\mathcal{N}_\psi} \inf_{\xi\in\mathfrak{m}^{-1}(C\times V)} H(\xi|\lambda), \tag{8.14}$$

(8.11) holds (for both $A = U$ and $C = \mathcal{R}$ as well as for $A = \mathcal{R}$ and $C = W$, where U is open and W is closed) if for all open U and all closed W

$$\inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\zeta\in\mathcal{P}^n_{emp}(\mathcal{S})\cap V_0} \inf_{\nu\in\mathfrak{m}^{-1}(U\times\{\zeta\})\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda)$$
$$\le \inf_{\nu\in\mathfrak{m}^{-1}(U\times\{\psi\})} H(\nu|\lambda), \tag{8.15}$$

$$\sup_{V\in\mathcal{N}_\psi} \inf_{\xi\in\mathfrak{m}^{-1}(W\times V)} H(\xi|\lambda) \ge \inf_{\xi\in\mathfrak{m}^{-1}(W\times\{\psi\})} H(\xi|\lambda). \tag{8.16}$$

(8.16) is a consequence of Lemma 5.3, as $\mathfrak{m}^{-1}(W\times V) = \mathfrak{m}^{-1}(W\times\mathcal{P}(\mathcal{S}))\cap\mathfrak{m}^{-1}(\mathcal{P}(\mathcal{R})\times V)$, the set $F = \mathfrak{m}^{-1}(W\times\mathcal{P}(\mathcal{S}))$ is closed for closed W, $\mathfrak{m}^{-1}(\mathcal{P}(\mathcal{R})\times V) = (\pi\circ\mathfrak{m})^{-1}(V)$, and $\pi\circ\mathfrak{m}$ is continuous. The proof of inequality (8.15) requires a little more attention. First we present some facts which are used to prove this inequality in Lemma 8.8.

Lemma 8.5 [7, Remark below Definition 2.1.5] *The map $\nu \mapsto H(\nu|\lambda)$ is continuous on $[H(\cdot|\lambda) < \infty]$. In particular, for all $\varepsilon > 0$ and $\xi \in \mathcal{P}(\mathcal{R} \times \mathcal{S})$ there exists a $\Theta \in \mathcal{N}_\xi$ such that*

$$H(\nu|\lambda) - \varepsilon \leq H(\xi|\lambda) \quad (\nu \in \Theta \cap [H(\cdot|\lambda) < \infty]). \tag{8.17}$$

Consequently, I as in (8.6) is continuous on $[J < \infty]$.

Lemma 8.6 (a) *Let $k, l \in \mathbb{N}$ and $\zeta \in \mathcal{P}^k_{emp}(\mathcal{S})$. For all $m \geq kl$ there exists a $\nu \in \mathcal{P}^m_{emp}(\mathcal{S})$ such that $\mathfrak{d}(\nu, \zeta) < \frac{1}{l}$.*

(b) *For all open $\Theta \subset \mathcal{P}(\mathcal{S})$ there exists an $N \in \mathbb{N}$ such that $\mathcal{P}^n_{emp}(\mathcal{S}) \cap \Theta \neq \emptyset$ for all $n \geq N$.*

Proof (a) Let $i \in \{1, \ldots, k\}$. Let $\xi \in \mathcal{P}^i_{emp}(\mathcal{S})$. Then the measure $\frac{lk}{lk+i}\zeta + \frac{i}{lk+i}\xi$ is an element of $\mathcal{P}^{lk+i}_{emp}(\mathcal{S})$. For every $A \subset \mathcal{S}$

$$\left|[\tfrac{lk}{lk+i}\zeta + \tfrac{i}{lk+i}\xi](A) - \zeta(A)\right| \leq 2\tfrac{i}{lk+i} \leq 2\tfrac{k}{lk} = \tfrac{2}{l}. \tag{8.18}$$

By definition of the Prohorov metric, this implies $\mathfrak{d}([\frac{lk}{lk+i}\zeta + \frac{i}{lk+i}\xi], \zeta) \leq \frac{2}{l}$.

(b) Let $\xi \in \mathcal{P}(\mathcal{S})$ and $\delta > 0$ be such that $B(\xi, \delta) \subset \Theta$. For each $\xi \in \mathcal{P}(\mathcal{S})$ there is a $k \in \mathbb{N}$ and a $\zeta \in \mathcal{P}^k_{emp}(\mathcal{S})$ such that $\mathfrak{d}(\zeta, \xi) < \frac{\delta}{2}$. Because of this, (b) follows from (a) by letting l be such that $\frac{1}{l} < \frac{\delta}{2}$ and $N = lk$. □

Lemma 8.7 *Let $\xi \in \mathcal{P}(\mathcal{R} \times \mathcal{S})$, $\pi \circ \mathfrak{m}(\xi) = \psi$ and $\xi \ll \lambda$. For all $\delta > 0$ there exist a $\kappa > 0$ and an $N \in \mathbb{N}$ such that for all $n \geq N$ and all $\zeta \in \mathcal{P}^n_{emp}(\mathcal{S})$ with $\mathfrak{d}(\zeta, \psi) < \kappa$ there is a $\nu \in \mathcal{P}^n_{emp}(\mathcal{R} \times \mathcal{S})$ with*

$$\pi \circ \mathfrak{m}(\nu) = \zeta, \quad \nu \ll \lambda, \quad \mathfrak{d}(\nu, \xi) < \delta, \quad \mathfrak{d}\Big(\nu(\cdot \times \mathcal{S}), \xi(\cdot \times \mathcal{S})\Big) < \delta. \tag{8.19}$$

Proof In this proof, for a measure $\xi \in \mathcal{P}(\mathcal{R} \times \mathcal{S})$, we write $\xi_{rs} = \xi(\{(r, s)\})$, so that $\xi = \sum_{rs} \xi_{rs}\delta_{(r,s)}$ where we use the shorthand notation "$\sum_{rs}$" instead of "$\sum_{r\in\mathcal{R},s\in\mathcal{S}}$". Let $M = \#\mathcal{R}\#\mathcal{S}$. Note that

$$\mathfrak{d}(\xi, \nu) \leq M \max_{r\in\mathcal{R},s\in\mathcal{S}} |\xi_{rs} - \nu_{rs}| \quad \big(\xi, \nu \in \mathcal{P}(\mathcal{R} \times \mathcal{S})\big). \tag{8.20}$$

Let $\kappa > 0$ and $n \in \mathbb{N}$. We first give an estimation by which it is clear which κ and N one should choose. By the assumptions on λ for every $s \in \mathcal{S}$ there exists a $r_s \in \mathcal{R}$ with $\lambda_{r_s s} > 0$.

First we show that there exists a $\xi^* \in \mathcal{P}^n_{emp}(\mathcal{X} \times \mathcal{Y})$ with $\xi^* \ll \xi$ and $|\xi^*_{rs} - \xi_{rs}| \leq \frac{2}{n}$ for all $r \in \mathcal{R}$ and $s \in \mathcal{S}$. For each pair $(r, s) \in \mathcal{R} \times \mathcal{S}$ with $\xi_{rs} > 0$ we can choose a $\xi'_{rs} \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ such that $|\xi_{rs} - \xi'_{rs}| < \frac{1}{n}$. By letting $\xi^*_{rs} = 0$ when $\xi_{rs} = 0$ and add or subtract $\frac{1}{n}$ to some of the ξ'_{rs} we obtain a collection of $\xi^*_{rs} \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ with $\sum_{rs} \xi^*_{rs} = 1$ and $|\xi^*_{rs} - \xi_{rs}| \leq \frac{2}{n}$ and $\xi^*_{rs} = 0$ whenever $\xi_{rs} = 0$ for all $r \in \mathcal{R}$ and $s \in \mathcal{S}$.

Let $\xi \in \mathcal{P}(\mathcal{R}\times\mathcal{S})$. Suppose that $\zeta \in \mathcal{P}^n_{emp}(\mathcal{S})$ is such that $|\zeta_s - \sum_r \xi_{rs}| < \kappa$. Then $|\zeta_s - \sum_r \xi^*_{rs}| < \kappa + \frac{2}{n}M$. We construct a $\nu \in \mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})$ by defining the ν_{rs} by each s separately. Let $s \in S$. If $\zeta_s - \sum_r \xi^*_{rs} < 0$, then we choose $\nu_{rs} \le \xi^*_{rs}$ with $\nu_{rs} \in \{0, \frac{1}{n}, \dots, 1\}$ in such way that $\sum_r \nu_{rs} = \zeta_s$ (note that $|\nu_{rs} - \xi^*_{rs}| \le |\zeta_s - \sum_r \xi^*_{rs}|$). While, if $\zeta_s - \sum_r \xi^*_{rs} \ge 0$, then we let $\nu_{rs} = \xi^*_{rs}$ for all $r \ne r_s$ and we let $\nu_{r_s s} = \xi^*_{r_s s} + \zeta_s - \sum_r \xi^*_{rs}$ (so that $\sum_r \nu_{rs} = \zeta_s$). As $\xi^* \ll \xi$ and $\xi \ll \lambda$, by the construction of ν we have $\nu \ll \lambda$. Moreover, we have $\pi \circ \mathfrak{m}(\nu) = \zeta$ and

$$\begin{aligned}\max_{r\in\mathcal{R},s\in\mathcal{S}} \left|\nu_{rs} - \xi_{rs}\right| &\le \max_{s\in\mathcal{S}} \left|\zeta_s - \sum_r \xi^*_{rs}\right| + \max_{r\in\mathcal{R},s\in\mathcal{S}} |\xi^*_{rs} - \xi_{rs}| \\ &\le \kappa + \tfrac{2}{n}M + \tfrac{2}{n}. \end{aligned} \tag{8.21}$$

which implies by (8.20)

$$\mathfrak{d}(\nu,\xi) \le M\kappa + \tfrac{2}{n}(M^2 + M). \tag{8.22}$$

Moreover, as $|\sum_s \nu_{rs} - \sum_s \xi_{rs}| \le M \max_{s\in\mathcal{S}} |\nu_{rs} - \xi_{rs}|$,

$$\mathfrak{d}\Big(\nu(\cdot\times\mathcal{S}), \xi(\cdot\times\mathcal{S})\Big) \le M \max_{r\in\mathcal{R}} \left|\sum_s \nu_{rs} - \sum_s \xi_{rs}\right| \le M^2\kappa + \tfrac{2}{n}(M^3 + M^2). \tag{8.23}$$

By choosing $\kappa > 0$ and $N \in \mathbb{N}$ such that $M^2\kappa + \frac{2}{n}(M^3 + M^2) < \delta$ the proof is complete. □

Lemma 8.8 *For all open $U \subset \mathcal{R}$*

$$\begin{aligned} 0 &\le \inf_{V_0\in\mathcal{N}_\psi} \limsup_{n\to\infty} \sup_{\zeta\in\mathcal{P}^n_{emp}(\mathcal{S})\cap V_0} \inf_{\nu\in\mathfrak{m}^{-1}(U\times\{\zeta\})\cap\mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})} H(\nu|\lambda) \\ &\le \inf_{\nu\in\mathfrak{m}^{-1}(U\times\{\psi\})} H(\nu|\lambda). \end{aligned} \tag{8.24}$$

Proof We assume $\inf_{\nu\in\mathfrak{m}^{-1}(U\times\{\psi\})} H(\nu|\lambda) < \infty$. Let $\xi \in \mathfrak{m}^{-1}(U\times\{\psi\})$ be such that $H(\xi|\lambda) < \infty$. Let $\varepsilon > 0$. We show there exists a $V_0 \in \mathcal{N}_\psi$ and an $N \in \mathbb{N}$ such that for all $n \ge N$ the set $\mathcal{P}^n_{emp}(\mathcal{S}) \cap V_0$ is not empty and for all $\zeta \in \mathcal{P}^n_{emp}(\mathcal{S}) \cap V_0$ there exists a $\nu \in \mathfrak{m}^{-1}(U\times\{\zeta\}) \cap \mathcal{P}^n_{emp}(\mathcal{R}\times\mathcal{S})$ with

$$H(\nu|\lambda) - \varepsilon \le H(\xi|\lambda). \tag{8.25}$$

Let δ be such that (see Lemma 8.5)

$$B(\xi(\cdot\times\mathcal{S}), \delta) \subset U, \tag{8.26}$$

$$H(\nu|\lambda) - \varepsilon \le H(\xi|\lambda) \quad (\nu \in B(\xi,\delta) \cap [H(\cdot|\lambda) < \infty]). \tag{8.27}$$

Then let $\kappa > 0$ and $N \in \mathbb{N}$ be as in Lemma 8.7. Let $V_0 = B(\psi, \kappa)$. By Lemma 8.6 we may assume that N is large enough such that $\mathcal{P}^n_{emp}(\mathcal{S}) \cap V_0 \neq \emptyset$. Let $n \geq N$ and $\zeta \in \mathcal{P}^n_{emp}(\mathcal{S}) \cap V_0$. By Lemma 8.7 there exists a $\nu \in \mathcal{P}^n_{emp}(\mathcal{R} \times \mathcal{S})$ with $\pi \circ \mathfrak{m}(\nu) = \zeta$, $\nu \ll \lambda$ and $\nu(\cdot \times \mathcal{S}) \in B(\xi(\cdot \times \mathcal{S}), \delta)$, $\nu \in B(\xi, \delta)$, i.e. by (8.26), $\nu \in \mathfrak{m}^{-1}(U \times \{\zeta\})$. $\nu \ll \lambda$ implies $\nu \in [H(\cdot|\lambda) < \infty]$; thus with (8.27) we obtain (8.25). □

9 Examples

In Sect. 8, we showed that the regular conditional kernel η_n as in (8.3) satisfies (A1) and (B1) of Theorem 6.9 by showing that (A2) and (B2) of that theorem hold. This is not always the most optimal approach; in Example 9.1 we show that for a specific example of Gaussian measures the expression of η_n allows us to derive (A1) and (B1) directly.

Furthermore, relying on Theorem 9.2, in Example 9.4, we give an example of a $(\eta_n)_{n\in\mathbb{N}}$ for which (A1) of Theorem 6.9 does not hold. In Remark 9.5 we mention that for the one choice of measures in Example 9.4 a quenched large deviation principle is satisfied, while for the other choice of measures there is no quenched large deviation principle. In Example 9.6 we show that for a choice of measures as in Example 9.4 the conditional regular kernel in a specific chosen point does not satisfy any large deviation principle. In Remark 9.7 we discuss exponential tightness of the regular conditional kernel. In Remark 9.8 we discuss the differences between the present paper and the paper of La Cour and Schieve [20].

Example 9.1 Let $r \neq 0$, $Z_n := \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\frac{n}{2}(x^2 - 2rxy + y^2)} \, \mathrm{d}x \, \mathrm{d}y$ and consider $(\mu_n)_{n\in\mathbb{N}}$ the sequence of probability measures on $\mathcal{B}(\mathbb{R} \times \mathbb{R})$ determined by

$$\mu_n(A \times B) = \frac{1}{Z_n} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{A\times B}(x, y) e^{-\frac{n}{2}(x^2 - 2rxy + y^2)} \, \mathrm{d}x \, \mathrm{d}y \quad (A, B \in \mathcal{B}(\mathbb{R})). \tag{9.1}$$

The sequence satisfies the large deviation principle with rate function $J : \mathbb{R}^2 \to [0, \infty]$ given by $J(x, y) = \frac{1}{2}(x^2 - 2rxy + y^2)$. By Theorem 4.8 η_n given by

$$\eta_n(y, A) = \frac{\int_A e^{-\frac{n}{2}(x^2 - 2rxy)} \, \mathrm{d}x}{\int_{\mathbb{R}} e^{-\frac{n}{2}(x^2 - 2rxy)} \, \mathrm{d}x} = \frac{\int_A e^{-\frac{n}{2}(x - ry)^2} \, \mathrm{d}x}{\int_{\mathbb{R}} e^{-\frac{n}{2}(x - ry)^2} \, \mathrm{d}x}, \tag{9.2}$$

is the weakly continuous product regular conditional probability under μ_n with respect to the projection on the $\mathcal{Y}$-coordinate. If $y_n \to y$, one can show that for $\lambda \in \mathbb{R}$

$$\begin{aligned} \lim_{n\to\infty} \tfrac{1}{n} \log \int_{\mathbb{R}} e^{n\lambda x} \, \mathrm{d}[\eta_n(y_n, \cdot)](x) &= \lim_{n\to\infty} \tfrac{1}{n} \log \int_{\mathbb{R}} e^{n\lambda x} \, \mathrm{d}[\eta_n(y, \cdot)](x) \\ &= \lambda r y + \tfrac{1}{2}\lambda^2. \end{aligned} \tag{9.3}$$

Then by the Gärtner–Ellis theorem (see for example Dembo and Zeitouni [7, Theorem 2.3.6]) we conclude that $(\eta_n(y_n, \cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with the

same rate function as the one of the large deviation principle of $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$, which is $x \mapsto (x-ry)^2$. Note that this equals $J(x,y) - \inf J(\mathbb{R}\times\{y\})$ because of the equality $x^2 - 2rxy + y^2 = (x-ry)^2 + (1-r^2)y^2$.

The proof of the following theorem can be found in "Appendix 3".

Theorem 9.2 *Let $\mathcal{X}$ and $\mathcal{Y}$ be separable metric spaces. Let $(\mu_n^1)_{n\in\mathbb{N}}$ and $(\mu_n^2)_{n\in\mathbb{N}}$ be sequences of probability measures on $\mathcal{B}(\mathcal{X})$. Let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathcal{Y})$ that satisfies the large deviation principle with a rate function $L : \mathcal{Y} \to [0,\infty]$. Suppose that $y \in \mathcal{Y}$ and $W_n \in \mathcal{N}_y$ are such that $\bigcap_{n\in\mathbb{N}} W_n = \{y\}$ and $\alpha_n : \mathcal{Y} \to [0,1]$ is a continuous function with $\alpha_n(y) = 0$ and $\alpha_n = 1$ on $\mathcal{Y}\setminus W_n$ such that*

$$\liminf_{n\to\infty} \tfrac{1}{n} \log\left(\int_{W_n} \alpha_n \, \mathrm{d}\nu_n\right) = 0, \quad \liminf_{n\to\infty} \tfrac{1}{n} \log\left(\int_{W_n} (1-\alpha_n) \, \mathrm{d}\nu_n\right) = 0. \tag{9.4}$$

Assume $(\mu_n^1)_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I. Assume furthermore that for all open $A \subset \mathcal{X}$

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n^1(A) \geq \liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n^2(A), \tag{9.5}$$

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n^1(\mathcal{X}\setminus A) \geq \limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n^2(\mathcal{X}\setminus A). \tag{9.6}$$

Let μ_n be the probability measure on $\mathcal{B}(\mathcal{X})\otimes\mathcal{B}(\mathcal{Y})$ for which for $A \in \mathcal{B}(\mathcal{X})$, $B \in \mathcal{B}(\mathcal{Y})$

$$\mu_n(A\times B) = \mu_n^1(A) \int_{\mathcal{Y}} \mathbb{1}_B \alpha_n \, \mathrm{d}\nu_n + \mu_n^2(A) \int_{\mathcal{Y}} \mathbb{1}_B (1-\alpha_n) \, \mathrm{d}\nu_n. \tag{9.7}$$

Then $(\mu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function $J : \mathcal{X}\times\mathcal{Y} \to [0,\infty]$ given by $J(x,y) = I(x) + L(y)$. $\eta_n : \mathcal{Y}\times\mathcal{B}(\mathcal{X}) \to [0,1]$ defined by

$$\eta_n(y,A) = \alpha_n(y)\mu_n^1(A) + (1-\alpha_n(y))\mu_n^2(A) \tag{9.8}$$

is the weakly continuous product regular conditional probability under μ_n with respect to $\pi : \mathcal{X}\times\mathcal{Y} \to \mathcal{Y}$ given by $\pi(x,y) = y$.

Note that $I(x) = J(x,y) - \inf J(\mathcal{X}\times\{y\})$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

Example 9.3 We give examples of $\mathcal{Y}$, W_n, α_n, ν_n and L such that (9.4) of Theorem 9.2 is satisfied and $(\nu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function L.

(a) Let $\mathcal{Y} = [0,\infty)$, $\alpha_n(y) = \min\{ny, 1\}$ for $y \in \mathcal{Y}$ and let $\nu_n(B) = \int_0^\infty \mathbb{1}_B(y) ne^{-ny} \, \mathrm{d}y$ for $B \in \mathcal{B}([0,\infty))$. Then $\int_0^{\frac{1}{n}} \alpha_n \, \mathrm{d}\nu_n = 1 - 2e^{-1}$ and $\int_0^{\frac{1}{n}} (1-\alpha_n) \, \mathrm{d}\nu_n = e^{-1}$. Therefore with this ν_n, α_n and $W_n = [-\frac{1}{n}, \frac{1}{n}]$ (9.4) is satisfied. Moreover $(\nu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function $L : \mathcal{Y} \to [0,\infty]$, $L(y) = y$ (this follows from example by the Gärtner–Ellis theorem [7, Theorem 2.3.6]).

(b) Let $\mathcal{Y} = \mathbb{R}$ and $\nu_n = \mu_{\mathcal{N}(0,\frac{1}{n})}$ (the Gaussian measure corresponding to a $\mathcal{N}(0, \frac{1}{n})$ distributed random variable). Then there exists a decreasing sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in $(0,\infty)$ with $\varepsilon_n \downarrow 0$, such that with $W_n = [-\varepsilon_n, \varepsilon_n]$ there exist functions α_n as in Theorem 9.2 such that (9.4) is satisfied (see the postscript). With $\nu_n^0 = \frac{1}{2}\delta_0 + \frac{1}{2}\nu_n$ instead of ν_n, (9.4) is also satisfied. Moreover, $(\nu_n)_{n\in\mathbb{N}}$ and $(\nu_n^0)_{n\in\mathbb{N}}$ (use Lemma 9.9) satisfy the large deviation principle with rate function $L : \mathcal{Y} \to [0,\infty]$, $L(y) = \frac{1}{2}y^2$.
Postscript.
Let $\beta = \nu_1([-1,1])$. Let $\kappa_n = \frac{1}{\sqrt{n}}$. Then $\nu_n[-\kappa_n,\kappa_n] = \beta$ for all $n \in \mathbb{N}$. Let $\phi_\varepsilon : \mathbb{R} \to [0,1]$ be defined by $\phi_\varepsilon(z) = \min\{\varepsilon^{-1}|z|, 1\}$. Then $\lim_{\varepsilon\downarrow 0} \int_{[-\kappa_n,\kappa_n]} \phi_\varepsilon \, \mathrm{d}\nu_1 = \beta$, $\lim_{\varepsilon\downarrow 0} \int_{[-\kappa_n,\kappa_n]} 1 - \phi_\varepsilon \, \mathrm{d}\nu_1 = 0$ and

$$\int_{[-\kappa_n,\kappa_n]} \phi_{\kappa_n} \, \mathrm{d}\nu_n < \int_{[-\kappa_n,\kappa_n]} 1 - \phi_{\kappa_n} \, \mathrm{d}\nu_n. \tag{9.9}$$

Therefore, for all $n \in \mathbb{N}$, there exists an $\varepsilon_n \in (0,\kappa_n)$ such that

$$\int_{[-\kappa_n,\kappa_n]} \phi_{\varepsilon_n} \, \mathrm{d}\nu_n = \tfrac{1}{2}\beta = \int_{[-\kappa_n,\kappa_n]} 1 - \phi_{\varepsilon_n} \, \mathrm{d}\nu_n, \tag{9.10}$$

With $\alpha_n = \phi_{\varepsilon_n}$, (9.4) as in Theorem 9.2 is satisfied.

Example 9.4 With $\mathcal{X} = \mathbb{R}$, $\mu_n^1 = \mu_{\mathcal{N}(0,\frac{1}{n})}$, $\mu_n^2 = \delta_{\frac{1}{n}}$ and $I(x) = \frac{1}{2}x^2$ for $x \in \mathbb{R}$ and $\mathcal{Y}$, ν_n (or ν_n^0), α_n, W_n and L as in Examples 9.3 (a) or (b) the conditions of Theorem 9.2 are satisfied (note that $(\delta_{\frac{1}{n}})_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function $H : \mathbb{R} \to [0,\infty]$ given by $H(0) = 0$ and $H(x) = \infty$ for $x \neq 0$).

Then $\eta_n(0,\cdot) = \delta_{\frac{1}{n}}$ and $\eta_n(\varepsilon_n,\cdot) = \mu_{\mathcal{N}(0,\frac{1}{n})}$ for all $n \in \mathbb{N}$. Whence $(\eta_n(0,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function H and $(\eta_n(\varepsilon_n,\cdot))_{n\in\mathbb{N}}$ (and also $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$ for $y > 0$) satisfies the large deviation principle with rate function I. Because $I \leq H$, the sequence $(\eta_n(0,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation upper bound not only with H but also with I instead of H. Therefore (b1) of Theorem 6.3 holds in case $y_n = 0$ for all n. Since $(\eta_n(0,\cdot))_{n\in\mathbb{N}}$ does not satisfy the large deviation principle with rate function I, (a1) of Theorem 6.3 does not hold. Therefore for any decreasing sequence $(V_m)_{m\in\mathbb{N}}$ in $\mathcal{N}_0$ with $\bigcap_{m\in\mathbb{N}} V_m = \{0\}$ there exists an open set U with $\inf I(U) < \infty$ with

$$\liminf_{n\to\infty} \limsup_{m\to\infty} \tfrac{1}{n} \log \mu_n(U \times \mathcal{Y} | \mathcal{X} \times V_m) < -\inf I(U). \tag{9.11}$$

We illustrate this for $\mathcal{Y}$, α_n, W_n, ν_n and L as in Examples 9.3(a): For $V_m = [0, \frac{1}{m})$, $U = (1,\infty)$ we get for $m \geq n$

$$\mu_n(U \times V_m) = \mu_{\mathcal{N}(0,\frac{1}{n})}(U) \int_0^{\frac{1}{m}} ny \cdot ne^{-ny} \, \mathrm{d}y, \tag{9.12}$$

$$\mu_n(\mathcal{X} \times V_m) = \int_0^{\frac{1}{m}} ne^{-ny} \, \mathrm{d}y. \tag{9.13}$$

Since $\int_0^{\frac{1}{m}} ny \cdot ne^{-ny}\, \mathrm{d}y \leq \frac{n}{m} \int_0^{\frac{1}{m}} ne^{-ny}\, \mathrm{d}y$, we get

$$\mu_n(U \times \mathcal{Y} | \mathcal{X} \times V_m) \leq \tfrac{n}{m} \mu_{\mathcal{N}(0,\frac{1}{n})}(U) \tag{9.14}$$

which converges to zero as $m \to \infty$, which implies

$$\limsup_{m\to\infty} \tfrac{1}{n} \log \mu_n(U \times \mathcal{Y} | \mathcal{X} \times V_m) = -\infty < -\tfrac{1}{2} = -\inf I(U). \tag{9.15}$$

Remark 9.5 (Quenched large deviations) Consider the situation as in Example 9.4. For all $n \in \mathbb{N}$ we have the following. If $\zeta_n : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ is a product regular conditional probability under μ_n with respect to π, then $\zeta_n(y,\cdot) = \eta_n(y,\cdot)$ for $[\mu_n \circ \pi^{-1}]$-almost all y (see Remark 3.5).

Whence, with ν_n as in Examples 9.3 (a) or (b), we have a quenched large deviation principle of the conditional probability with respect to the second coordinate with rate function I; for every product regular conditional probability ζ_n under μ_n with respect to π there exists a $Z \subset \mathcal{Y}$ with $\mu_n \circ \pi^{-1}(Z) = \nu_n(Z) = 1$ such that $(\zeta_n(y,\cdot))_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I for all $y \in Z$.

However, with ν_n^0 as in Examples 9.3 (b) instead of ν_n for such ζ one has $\zeta_n(0,\cdot) = \eta_n(0,\cdot)$ as $\nu_n^0(\{0\}) > 0$. Thus in this case we do not have such a quenched large deviation principle.

Example 9.6 With $\mathcal{X} = \mathbb{N}$, $\mu_n^1 = \sum_{k\in\mathbb{N}} 2^{-k}\delta_k$, $\mu_n^2 = \delta_n$ and $I(x) = 0$ for $x \in \mathbb{N}$ as in Example 9.4, and $\mathcal{Y}$, W_n, α_n, ν_n and L as in Examples 9.3 (a) or (b), the conditions of Theorem 9.2 are satisfied. In this case, $(\eta_n(0,\cdot))_{n\in\mathbb{N}}$ does not satisfy a large deviation principle.

Remark 9.7 (Exponential tightness of the regular conditional kernel) Considering the situation as in Theorem 9.2, we would like to mention that if $(\mu_n^1)_{n\in\mathbb{N}}$ is exponentially tight, then so is $(\mu_n)_{n\in\mathbb{N}}$ since $\mu_n(K_1^c \times K_2^c) = \mu_n^1(K_1^c)\nu_n(K_2^c)$ for large n and (compact) $K_1 \subset \mathcal{X}$, $K_2 \subset \mathcal{Y}$. Similarly $(\eta_n(y,\cdot))_{n\in\mathbb{N}}$ is exponentially tight for all $y > 0$ since $\eta_n(y, K^c) = \mu_n^1(K^c)$ for large n and compact $K \subset \mathcal{X}$. However, as is the case in Example 9.6, $(\eta_n(y_n,\cdot))_{n\in\mathbb{N}}$ need not be exponentially tight for all converging sequences $(y_n)_{n\in\mathbb{N}}$ (e.g. if $(\mu_n^2)_{n\in\mathbb{N}}$ is not exponentially tight, then $(\eta_n(0,\cdot))_{n\in\mathbb{N}}$ is neither).

Remark 9.8 Example 9.4 with ν_n (or ν_n^0) and α_n as in Examples 9.3 (b) fits the assumptions made in Sect. 4 of La Cour and Schieve [20].[8] In Sect. 4 of that paper, it is claimed that the law of the first coordinate conditioned on the second coordinate satisfies the large deviation principle with the rate function I. Their notion of conditioning on y is "condition on an arbitrarily small neighbourhood around y". This approach needs to be justified. Our results are different, as by Example 9.4 the conditioned kernel in 0,

[8] The logarithmic moment generating function (see Dembo and Zeitouni [7, Assumption 2.3.2]) is given by $(x,y) \mapsto \frac{1}{2}x^2 + \frac{1}{2}y^2$, whence the Hessian of it equals the identity matrix and is therefore invertible. In [20], it is mentioned that one cannot proceed the conditioning on all elements, but only those that equal the derivative of $y \mapsto \frac{1}{2}y^2$ at a certain point are considered, of which 0 is an example.

$\eta_n(0,\cdot)$ does not satisfy the large deviation principle with the rate function I (even in the sense of quenched large deviations as discussed in Remark 9.5).

Acknowledgements The author is supported by ERC Advanced Grant VARIS-267356 of Frank den Hollander. The author is grateful to both Frank den Hollander and Frank Redig for valuable suggestions and useful discussions.

Appendix 1: An Elementary Fact About Limsup and Liminf

Lemma 9.9 *Let $k \in \mathbb{N}$ and $a_n^i \in [0,\infty)$ for all $n \in \mathbb{N}$ and $i \in \{1,\dots,k\}$. If there exists an $N \in \mathbb{N}$ such that $\max_{i\in\{1,\dots,k\}} a_n^i > 0$ for all $n \geq N$, then*[9]

$$\lim_{n\to\infty}\left(\frac{1}{n}\log\left(\sum_{i=1}^k a_n^i\right) - \max_{i\in\{1,\dots,k\}}\frac{1}{n}\log a_n^i\right) = 0, \tag{9.16}$$

$$\limsup_{n\to\infty}\frac{1}{n}\log\left(\sum_{i=1}^k a_n^i\right) = \max_{i\in\{1,\dots,k\}}\limsup_{n\to\infty}\frac{1}{n}\log a_n^i, \tag{9.17}$$

$$\liminf_{n\to\infty}\frac{1}{n}\log\left(\sum_{i=1}^k a_n^i\right) = \max_{i\in\{1,\dots,k\}}\liminf_{n\to\infty}\frac{1}{n}\log a_n^i. \tag{9.18}$$

Proof (9.16), (9.17) and (9.18) follow from the inequality

$$\begin{aligned}\max_{i\in\{1,\dots,k\}}\frac{1}{n}\log a_n^i \leq \frac{1}{n}\log\left(\sum_{i=1}^k a_n^i\right) &\leq \frac{1}{n}\log(k\max_{i\in\{1,\dots,k\}} a_n^i)\\ &\leq \frac{1}{n}\log k + \max_{i\in\{1,\dots,k\}}\frac{1}{n}\log(a_n^i).\end{aligned} \tag{9.19}$$

□

Appendix 2: Sufficient Bounds for Large Deviation Bounds

Let $\mathcal{X}$ be a topological space. Let $I : \mathcal{X} \to [0,\infty]$ have compact sublevel sets. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathcal{X})$.

Lemma 9.10 *Let $(F_m)_{m\in\mathbb{N}}$ be a decreasing sequence of closed sets with $F = \bigcap_{m\in\mathbb{N}} F_m$. Then*

$$\sup_{m\in\mathbb{N}}\inf I(F_m) = \inf I(F). \tag{9.20}$$

[9] Equation (9.17) can also be found in Dembo and Zeitouni [7, Theorem 1.2.15].

Proof Let $c := \sup_{m\in\mathbb{N}} \inf I(F_m)$. Note that $c \le \inf I(F)$. If $c = \infty$, there is nothing to prove. Assume that $c < \infty$. Let K be the compact set $[I \le c]$. Then $F_m \cap K \ne \emptyset$ for all $m \in \mathbb{N}$, whence $F \cap K \ne \emptyset$ and thus $\inf I(F) \le c$. □

9.11 For Lemma 9.10 the condition that I has compact sublevel sets is not redundant. For example: Let $I : \mathbb{N}\cup\{0\} \to [0,\infty]$ be given by $I(0) = 1$ and $I(x) = 0$ for $x \in \mathbb{N}$. Then for $F_m = \{0\} \cup \{m, m+1, \dots\}$ and $F = \{0\}$ one has $\sup_{m\in\mathbb{N}} \inf I(F_m) = 0$ and $\inf I(F) = 1$.

Lemma 9.12 (a) *$\mathcal{U}$ be a set of open subsets of $\mathcal{X}$. Suppose that for all $G \in \mathcal{U}$*

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n(G) \ge -\inf I(G). \tag{9.21}$$

Then $G = \bigcup \mathcal{U}$ satisfies (9.21) as well.[10]

(b) *Let $F_1, F_2, \dots$ be closed. Suppose that for all $F \in \{F_m : m \in \mathbb{N}\}$*

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n(F) \le -\inf I(F). \tag{9.22}$$

Then $F = \bigcap_{m\in\mathbb{N}} F_m$ satisfies (9.22) as well.

Proof

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n\Big(\bigcup \mathcal{U}\Big) \ge \sup_{G\in\mathcal{U}} \liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n(G) \ge \sup_{G\in\mathcal{U}} (-\inf I(G)), \tag{9.23}$$

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n\Big(\bigcap_{m\in\mathbb{N}} F_m\Big) \le \inf_{m\in\mathbb{N}} \limsup_{n\to\infty} \tfrac{1}{n} \log \mu_n(F_m) \le \inf_{m\in\mathbb{N}} (-\inf I(F_m)). \tag{9.24}$$

Now apply Lemma 9.10. □

As a consequence of Lemma 9.12 we obtain the following.

Theorem 9.13 *Suppose that $\mathcal{G}$ is a basis for the topology on $\mathcal{X}$, such that (9.21) holds for all $G \in \mathcal{G}$ and (9.22) holds for $F = \mathcal{X} \setminus G$. Suppose that every open G can be written as countable union of elements in $\mathcal{G}$. Then $(\mu_n)_{n\in\mathbb{N}}$ satisfies the large deviation principle with rate function I.*

Appendix 3: Proof of Theorem 9.2

Proof of Theorem 9.2 As $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces, every open subset of $\mathcal{X} \times \mathcal{Y}$ is a countable union of elements of the form $A \times B$ where $A \subset \mathcal{X}$ is open and $B \in \mathcal{H}$, where (with $d_{\mathcal{Y}}$ the metric on $\mathcal{Y}$)

$$\mathcal{H} = \{B(y,\delta) : \delta > 0\} \cup \{B(z,\delta) : z \ne y, 0 < \delta < d_{\mathcal{Y}}(y,z)\}. \tag{9.25}$$

[10] This can also be found in O'Brien [23, Proposition 2.1].

We use Theorem 9.13 to prove the large deviation bounds. Note first that $(\mathcal{X}\times\mathcal{Y})\setminus(A\times B)=(\mathcal{X}\times(\mathcal{Y}\setminus B))\cup((\mathcal{X}\setminus A)\times\mathcal{Y})$, that $\min\{\inf I(\mathcal{X}\setminus A),\inf L(\mathcal{Y}\setminus B)\}=\inf_{(x,y)\in(\mathcal{X}\times\mathcal{Y})\setminus(A\times B)} I(x)+L(y)$ and that by (9.17)

$$\begin{aligned}&\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n((\mathcal{X}\times\mathcal{Y})\setminus(A\times B))\\&\quad\le\max\left\{\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n(\mathcal{X}\times(\mathcal{Y}\setminus B)),\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n((\mathcal{X}\setminus A)\times\mathcal{Y})\right\}.\end{aligned}$$

Using this and Theorem 9.13 it is sufficient to show that for all open sets $A\subset\mathcal{X}$ and $B\subset\mathcal{Y}$

$$\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n(\mathcal{X}\times(\mathcal{Y}\setminus B))\le-\inf L(\mathcal{Y}\setminus B),\tag{9.26}$$

$$\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n((\mathcal{X}\setminus A)\times\mathcal{Y})\le-\inf I(\mathcal{X}\setminus A)\tag{9.27}$$

$$\liminf_{n\to\infty}\tfrac{1}{n}\log\mu_n(A\times B)\ge-\inf I(A)-\inf L(B).\tag{9.28}$$

Let $A\subset\mathcal{X}$ be open and $B\in\mathcal{H}$.

- (9.26) follows from the fact that $\mu_n(\mathcal{X}\times(\mathcal{Y}\setminus B))=\nu_n(\mathcal{Y}\setminus B)$.
- (9.27) follows from the fact that by (9.4), (9.6) and (9.17) we have

$$\begin{aligned}&\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n((\mathcal{X}\setminus A)\times\mathcal{Y})\\&\qquad=\max\left\{\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n^1(\mathcal{X}\setminus A),\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n^2(\mathcal{X}\setminus A)\right\}\\&\qquad=\limsup_{n\to\infty}\tfrac{1}{n}\log\mu_n^1(\mathcal{X}\setminus A)\le-\inf I(\mathcal{X}\setminus A),\end{aligned}\tag{9.29}$$

- (9.28) follows by separating two cases (as either $y\in B$ or $y\notin\overline{B}$): If $y\notin\overline{B}$, then $W_n\cap B=\emptyset$ and so $\mu_n(A\times B)=\mu_n^1(A)\nu_n(B)$ for large n, whence

$$\liminf_{n\to\infty}\tfrac{1}{n}\log\mu_n(A\times B)=\liminf_{n\to\infty}\left(\tfrac{1}{n}\log\mu_n^1(A)+\tfrac{1}{n}\log\nu_n(B)\right).\tag{9.30}$$

Suppose that $y\in B$, i.e. $W_n\subset B$ for large n. By (9.18) we obtain

$$\begin{aligned}&\liminf_{n\to\infty}\tfrac{1}{n}\log\mu_n(A\times B)\\&\quad=\max\left\{\liminf_{n\to\infty}\left(\tfrac{1}{n}\log\mu_n^1(A)+\tfrac{1}{n}\log\left(\int_{W_n}\alpha_n\,\mathrm{d}\nu_n\right)\right),\right.\\&\qquad\liminf_{n\to\infty}\left(\tfrac{1}{n}\log\mu_n^2(A)+\tfrac{1}{n}\log\left(\int_{W_n}(1-\alpha_n)\,\mathrm{d}\nu_n\right)\right),\\&\qquad\left.\liminf_{n\to\infty}\left(\tfrac{1}{n}\log\mu_n^1(A)+\tfrac{1}{n}\log\left(\int_{W_n^c}\mathbb{1}_B\,\mathrm{d}\nu_n\right)\right)\right\}\end{aligned}\tag{9.31}$$

Using that $\liminf_{n\to\infty} \frac{1}{n} \log(\int_{W_n^c} \mathbb{1}_B \, d\nu_n) \leq 0$ together with (9.4) and (9.5), we obtain

$$\begin{aligned} \liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n(A \times B) &= \max \left\{ \liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n^1(A), \liminf_{n\to\infty} \tfrac{1}{n} \log \mu_n^2(A) \right\} \\ &\geq -\inf I(A). \end{aligned} \tag{9.32}$$

Because $\inf L(B) \geq 0$, we conclude (9.28).

We leave it to the reader to check that η_n is the weakly continuous product regular conditional probability under μ_n with respect to π. □

References

1. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. Commun. Math. Phys. **307**(3), 791–815 (2011)
2. Biggins, J.D.: Large deviations for mixtures. Electron. Commun. Probab. **9**, 60–71 (2004). (electronic)
3. Billingsley, P.: Convergence of Probability Measures, Wiley Series in Probability and Statistics: Probability and Statistics, 2nd edn. Wiley, New York (1999)
4. Bogachev, V.: Measure Theory. Springer, Berlin (2007)
5. Comets, F.: Large deviation estimates for a conditional probability distribution. Applications to random interaction Gibbs measures. Probab. Theory Relat. Fields **80**(3), 407–432 (1989)
6. Comets, F., Gantert, N., Zeitouni, O.: Quenched, annealed and functional large deviations for one-dimensional random walk in random environment. Probab. Theory Relat. Fields **118**(1), 65–114 (2000)
7. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, Stochastic Modelling and Applied Probability, vol. 38. Springer, Berlin (2010). [Corrected reprint of the second edition (1998)]
8. Deuschel, J.-D., Stroock, D.W.: Large Deviations, Pure and Applied Mathematics, vol. 137. Academic Press, Boston (1989)
9. van Enter, A.C.D., Fernández, R., den Hollander, F., Redig, F.: A large-deviation view on dynamical Gibbs-non-Gibbs transitions. Mosc. Math. J. **10**(4), 687–711 (2010)
10. van Enter, A.C.D., Külske, C., Opoku, A.A., Ruszel, W.M.: Gibbs-non-Gibbs properties for n-vector lattice and mean-field models. Braz. J. Probab. Stat. **24**(2), 226–255 (2010)
11. Ermolaev, V., Külske, C.: Low-temperature dynamics of the Curie–Weiss model: periodic orbits, multiple histories, and loss of Gibbsianness. J. Stat. Phys. **141**(5), 727–756 (2010)
12. Faden, A.: The existence of regular conditional probabilities: necessary and sufficient conditions. Ann. Probab. **13**(1), 288–298 (1985)
13. Fernández, R., den Hollander, F., Martínez, J.: Variational description of Gibbs-non-Gibbs dynamical transitions for the Curie–Weiss model. Commun. Math. Phys. **319**(3), 703–730 (2013)
14. Greven, A., den Hollander, F.: Large deviations for a random walk in random environment. Ann. Probab. **22**(3), 1381–1428 (1994)
15. Halmos, P.R.: Measure Theory. Springer, Berlin (1974)
16. den Hollander, F.: Large Deviations, Fields Institute Monographs, vol. 14. American Mathematical Society, Providence, RI (2000)
17. den Hollander, F., Redig, R., van Zuijlen, W.: Gibbs-non-Gibbs dynamical transitions for mean-field interacting Brownian motions. Stoch. Process. Appl. **125**(1), 371–400 (2015)
18. Kosygina, E., Rezakhanlou, F., Varadhan, S.R.S.: Stochastic homogenization of Hamilton–Jacobi–Bellman equations. Commun. Pure Appl. Math. **59**(10), 1489–1521 (2006)
19. Külske, C., Opoku, A.A.: Continuous spin mean-field models: limiting kernels and Gibbs properties of local transforms. J. Math. Phys. **49**(12), 125215 (2008)
20. La Cour, B.R., Schieve, W.C.: A general conditional large deviation principle. J. Stat. Phys. **161**(1), 123–130 (2015)
21. Leao Jr., D., Fragoso, M., Ruffino, P.: Regular conditional probability, disintegration of probability and Radon spaces. Proyecciones **23**(1), 15–29 (2004)
22. Léonard, C.: A large deviation approach to optimal transport. arXiv:0710.1461v1 (2007)
23. O'Brien, G.L.: Sequences of capacities, with connections to large-deviation theory. J. Theor. Probab. **9**(1), 19–35 (1996)

24. Rassoul-Agha, F., Seppäläinen, T., Yilmaz, A.: Quenched free energy and large deviations for random walk in random potential. Commun. Pure Appl. Math. **66**(2), 202–244 (2013)
25. Rassoul-Agha, F., Seppäläinen, T.: A Course on Large Deviations with an Introduction to Gibbs Measures, Graduate Studies in Mathematics, vol. 162. American Mathematical Society, Providence, RI (2015)
26. Schaefer, H.H.: Topological Vector Spaces. Springer, New York (1971). (Third printing corrected, Graduate Texts in Mathematics, Vol. 3)
27. Steen, L.A., Seebach Jr., J.A.: Counterexamples in Topology. Holt, Rinehart and Winston, New York (1970)