

# Large and Moderate Deviations for Projective Systems and Projective Limits

# **Tryfon Daras**

Department of Sciences, Technical University of Crete, Chania, Greece Email: tryfondaras@gmail.com

Received September 11, 2012; revised October 11, 2012; accepted October 18, 2012

# ABSTRACT

One of the active fields in applied probability, the last two decades, is that of *large deviations theory i.e.* the one dealing with the (asymptotic) computation of probabilities of rare events which are exponentially small as a function of some parameter e.g. the amplitude of the noise perturbing a dynamical system. Basic ideas of the theory can be tracked back to Laplace, the first rigorous results are due to Cramer although a clear definition was introduced by Varadhan in 1966. Large deviations estimates have been proved to be the crucial tool in studying problems in Statistics, Physics (Thermodynamics and Statistical Mechanics), Finance (Monte-Carlo methods, option pricing, long term portfolio investment) and in Applied probability (queuing theory). The aim of this work is to describe one of the (recent) methods of proving large deviations results, namely that of *projective systems*. We compare the method with the one of *projective limits* and show the advantages of the first. These advantages are due to the fact that: 1) the arguments are direct and the proofs of the basic results of the theory are much easier and simpler; 2) we are able to extend most of these results using suitable projective systems. We apply the method in the case of a) sequences of i.i.d. r.v.'s and b) sequences of exchangeable r.v.'s. All the results are being proved in a simple "unified" way.

Keywords: Large Deviations; Projective Systems; Projective Limits; Moderate Deviations

# 1. Notation and Basic Results

**Definition 1.1.** Let *E* be a Hausdorff topological space, *F* a  $\sigma$ -algebra of subsets of *E*, and  $\{\mu_d\}_{d\in D}$  (with *D* a directed set) a net of probability measures (p.m.'s) defined on *F*. We say that, the net of p.m.'s  $\{\mu_d\}_{d\in D}$  satisfies the *full large deviations principle* ([1,2]), with normalize-inequarter to  $\{(1,2)\}_{d\in D}$  and  $\{(1,2)\}_{d\in D}$  to  $\{(1,2)\}_{d\in D}$ .

ing constants  $\{r(\mathbf{d})\}_{\mathbf{d}\in D}$   $(r: D \to \mathbb{R}^+$  such that  $\lim_{\mathbf{d}} r(\mathbf{d}) = \infty$ ) and *rate function*  $I: E \to [0, \infty]$ , if I is lower semi-continuous and  $\forall B \in F$ , we have:

1) (upper bound)

$$\limsup_{d} \frac{1}{r(d)} \log \mu_{d}(B) \leq -\inf_{x \in cl(B)} I(x)$$
(1)

2) (lower bound)

$$-\inf_{x \in int(B)} I(x) \le \liminf_{d} \frac{1}{r(d)} \log \mu_{d}(B)$$
(2)

with cl(B) (int(B)) the closure (respectively the interior) of the set B.

If, in addition,  $\forall a \ge 0, L_a = \{x \in E : I(x) \le a\}$  (*level* set) is compact, *I* is called a *good* rate function.

# Remark 1.2.

If the upper bound is valid for all compact sets, while the lower bound is still true for all open sets, we say that the net of p.m.'s  $\{\mu_d\}_{d\in D}$  satisfies the *weak large deviations principle*.

In order to "pass" from a weak LDP to a full LDP we have to find a way of showing that, most of the probability mass (at least on an exponential scale) is concentrated on compact sets. The tool for doing this, is the following.

# **Definition 1.3.**

A net of p.m.'s  $\{\mu_d\}_{d\in D}$  defined on (E,F) is called *exponentially tight*, if  $\forall a > 0$ , there is a compact set  $K_a$  (subset of *E*) such that:

$$\limsup_{\mathbf{d}\in D} \frac{1}{r(\mathbf{d})} \log \mu_{\mathbf{d}} \left( K_a^c \right) \le -a \tag{3}$$

Exponential tightness is applied to the following proposition to strengthen a weak large deviations result. A proof of the proposition can be found in [1].

#### **Proposition 1.4.**

Let  $\{\mu_d\}_{d\in D}$  be a net of p.m.'s defined on (E, F) that is exponentially tight.

Then: a) if the upper bound holds for all compact sets, then it also holds for all closed sets.

b) if the lower bound holds for all open sets, then the rate function is good.

Now, we will characterize families of topological

spaces. This special kind of families will play an important role in proving large deviations results.

**Definition 1.5.** The family  $\{E_{\alpha}, \rho_{\alpha}^{\beta}\}_{\alpha,\beta\in A}$ , with A a directed set, is called a *projective system* if:

1)  $\forall \alpha \in A, E_{\alpha}$  is a Hausdorff topological space

2)  $\forall \alpha, \beta \in A, \alpha < \beta, \rho_{\alpha}^{\beta} : E_{\beta} \to E_{\alpha}$  is a continuous, subjective map such that, if:  $\alpha < \beta < \gamma, \rho_{\alpha}^{\gamma} = \rho_{\alpha}^{\beta} \circ \rho_{\beta}^{\gamma}$ Also,  $\rho_{\alpha}^{\alpha}$  is the identity map on  $E_{\alpha}$ .

We also consider a Hausdorff topological space E, F a  $\sigma$ -algebra of subsets of E and  $\forall \alpha \in A, \rho_{\alpha} : E \to E_{\alpha}$  a continuous, surjective map s.t. if  $\alpha < \beta : \rho_{\alpha} = \rho_{\alpha}^{\beta} \circ \rho_{\beta}$  and for  $x, y \in E, x \neq y$  then  $\exists \alpha \in A : \rho_{\alpha}(x) \neq \rho_{\alpha}(y)$ .

The following two theorems give large deviations results in the case of projective systems [3].

**Theorem 1.6.** Let *E* be a Hausdorff topological space, *F* a  $\sigma$ -algebra of subsets of *E* s.t.: a) *F* contains the class of compact sets and b) *F* contains a base *U* for the topology.

Let  $\left\{E_{\alpha}, \rho_{\alpha}^{\beta}\right\}_{\alpha, \beta \in A}$  be a projective system and

 $\rho_{\alpha}, \alpha \in A$  be as above. Assume that  $\rho_{\alpha}, \alpha \in A$  is measurable when *E* is endowed with *F* and  $E_{\alpha}$  with the Borel  $\sigma$ -algebra. Let  $\{\mu_d\}_{d\in D}$  be a net of p.m.'s on *F* and assume that:

i)  $\forall \alpha \in A$ , the net of p.m.'s  $\{\mu_{d} \circ \rho_{\alpha}^{-1}\}_{d \in D}$  satisfies a large deviations principle with normalizing constants  $\{r(d)\}_{d \in D}$  and rate function

$$I_{\alpha}: E_{\alpha} \to [0,\infty]$$

ii) the net of p.m.'s  $\{\mu_d\}_{d\in D}$  is exponentially tight.

Then, the net  $\{\mu_d\}_{d\in D}$  satisfies the large deviations principle with normalizing constants  $\{r(d)\}_{d\in D}$  and good rate function  $I(x) = \sup_{\alpha} I_{\alpha}(\rho_{\alpha}(x))$ .

When *E* is endowed with a specific topology (namely the topology induced by the maps  $\{\rho_{\alpha}\}_{\alpha \in A}$ ), Theorem 1.6 has the following form.

**Theorem 1.7.** Let E,  $\{E_{\alpha}, \rho_{\alpha}^{\beta}\}_{\alpha,\beta\in A}$  be as in theorem 1.6. Endow E with the initial topology induced by the maps  $\{\rho_{\alpha}\}_{\alpha\in A}$  and let F be the  $\sigma$ -algebra of subsets of E such that  $\forall \alpha \in A, \rho_{\alpha}$  is measurable, where  $E_{\alpha}$  is endowed with its Borel  $\sigma$ -algebra  $F_{\alpha}$ . Let  $\{\mu_{d}\}_{d\in D}$  be a net of p.m.'s on F and assume that:

i)  $\forall \alpha \in A$ , the net of p.m.'s  $\{\mu_{d} \circ \rho_{\alpha}^{-1}\}_{d \in D}$  satisfies a large deviation principle with normalizing constants  $\{r(d)\}_{d \in D}$  and rate function  $I_{\alpha}: E_{\alpha} \to [0, \infty]$ 

ii) there is a function  $I: E \to [0,\infty]$  such that

 $\forall b > 0$  the set  $L_b = \{x \in E : I(x) \le b\}$  is compact and  $\forall \alpha \in A, z \in E_{\alpha}$ :

$$I_{\alpha}(z) = \inf \left\{ I(x) : x \in \rho_{\alpha}^{-1}(z) \right\}$$

Then, the net of p.m.'s  $\{\mu_d\}_{d\in D}$  satisfies the large deviations principle with normalizing constants

 $\{r(d)\}_{d\in D}$  and good rate function *I*, and

 $I(x) = \sup_{\alpha} I_{\alpha}(\rho_{\alpha}(x)).$ 

On early days, large deviations results were proved using "large" spaces. One of these spaces is described below.

**Definition 1.8.** Let  $\{E_{\alpha}, \rho_{\alpha}^{\beta}\}_{\alpha,\beta\in A}$  be a projective system. The *projective limit* of this system (denoted by

 $\lim_{\leftarrow} E_{\alpha} \text{ ) is the subset of the product space } Y = \prod_{\alpha \in A} E_{\alpha}$ 

which consists of the elements  $x = (y_{\alpha})_{\alpha \in A}$  for which  $y_{\alpha} = p_{\alpha}^{\beta} (y_{\beta})$ , when  $\alpha < \beta$ , endowed with the topology induced by Y([2]).

The following basic result, analogous to that of Theorem 1.7, allows one to transport a large deviations result on a "smaller" topological space to a "larger" one.

**Theorem 1.9. Dawson-Gärtner** (large deviations for projective limits).

Let  $\{\mu_d\}_{d\in D}$  be a net of p.m.'s defined on

$$\begin{split} E &= \lim_{\leftarrow} E_{\alpha} \text{ . Assume that } \forall \alpha \in A \text{ , the net of p.m.'s} \\ & \left\{ \mu_{\mathrm{d}} \circ \rho_{\alpha}^{-1} \right\}_{\mathrm{d} \in D} \text{ satisfies the full large deviations principle} \\ & \text{with constants } \left\{ r(\mathrm{d}) \right\}_{\mathrm{d} \in D} \text{ and good rate function} \\ & I_{\alpha} : E_{\alpha} \to [0,\infty] \text{ . Then, the net of p.m.'s } \left\{ \mu_{\mathrm{d}} \right\}_{\mathrm{d} \in D} \text{ satisfies the full large deviations principle with constants} \\ & \left\{ r(\mathrm{d}) \right\}_{\mathrm{d} \in D} \text{ and good rate function:} \end{split}$$

$$I(x) = \sup_{\alpha} I_{\alpha}(\rho_{\alpha}(x))$$
(4)

**Remark 1.10.** The space *E* of Theorem 1.9 is specific, namely  $E = X = \lim E_{\alpha}$  (in Theorem 1.7 *E* is arbitrary).

Theorem 1.9 is a special case of the Theorem 1.7.

### **Proof.** (of Theorem 1.9)

Define the map  $I(x) = \sup_{\alpha} I_{\alpha}(\rho_{\alpha}(x))$ . It is easy to see (using properties of the projective limits) that  $\forall \alpha \in A$  the map  $I_{\alpha}(y) = \inf \{I(x)/x \in p_{\alpha}^{-1}(y)\}$  *i.e.* 

condition ii) of Theorem 1.7 is satisfied. Then, theorem 1.9 follows from Theorem 1.7.

The motivation for this paper was to find a "unified" way of proving large deviations results. This is done by using the projective systems approach. Using this approach, and not the one of projective limits, the proofs of most of the basic results of the theory are much easier and simpler, the arguments direct. Also, we are able to prove extensions of these results to more abstract spaces, at least in the case of exchangeable sequences of r.v.'s.

# 2. Applications

We now give some of the basic results of the large deviations theory. Extensions of these theorems can be easier proved using projective systems.

# 1) Theorem 2.1. (Cramer)

Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of independent and identically distributed (i.i.d) random variables (r.v.'s), taking values in  $\mathbb{R}^d$  with (common) distribution  $\mu = L(X_1)$ ,

and 
$$S_n = \sum_{j=1}^n X_j$$
.  
1) If  

$$\int_{\mathbb{R}^d} \exp(t ||x||) \mu(dx) < \infty, \text{ for some } t > 0 \quad (5)$$

then: a) (upper bound)  $\forall F \subset \mathbb{R}^d$  closed:

$$\lim_{n \to \infty} \sup \frac{1}{n} \log P\left\{\frac{S_n}{n} \in F\right\} \le -\inf_{x \in F} I(x) \quad (6)$$

with

$$I(x) = \sup_{\xi \in \mathbb{R}^d} \left\{ \left\langle x, \xi \right\rangle - \log \hat{\mu}(\xi) \right\}$$

and

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{\langle x,\xi \rangle} \mu(dx)$$
(7)

b)  $\forall a \ge 0$ , the set  $L_a = \{x/I(x) \le a\}$  is compact.

2) (lower bound)  $\forall G \subset \mathbb{R}^d$  open:

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left\{\frac{S_n}{n} \in G\right\} \ge -\inf_{x \in G} I(x)$$
(8)

Theorem 2.2 generalizes Cramer's theorem in the case of a separable Banach space. The proof is given here using projective systems.

**Theorem 2.2. (Donsker-Varadhan** 1976) (Generalization of Cramer's theorem)

Let *E* be a separable Banach space and *F* its Borel  $\sigma$ -algebra. Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of i.i.d. *E*-valued r.v.'s and  $\forall t > 0$ 

$$\int_{E} \exp(t \|x\|) \mu(\mathrm{d}x) < \infty \quad \text{where} \quad \mu = L(X_1).$$

Then, the sequence of p.m.'s 
$$\left\{L\left(\frac{S_n}{n}\right)\right\}_{n\in\mathbb{N}}$$
 satisfies

the large deviations principle with constants  $\{n\}_{n\in\mathbb{N}}$  and good rate function *I*:  $I(x) = \sup_{\xi\in E^*} \{\langle x,\xi\rangle - \log \hat{\mu}(\xi)\}$ , where  $E^*$  is the dual space of *E* and

$$\hat{\mu}(\xi) = \int \exp(\xi) d\mu, \xi \in E$$

(in other words Theorem 2.1. is true).

Proof.

Let A = N be the family of finite-dimensional sub-

Copyright © 2012 SciRes.

spaces of  $E^*$ , directed upward by inclusion. For each  $N \in A$ , let  $N^{\perp} = \{x \in E/\langle x, \xi \rangle = 0, \forall \xi \in N\}$  and  $p_N : E \to E/N^{\perp}$  the canonical projection of E onto  $E_N = E/N^{\perp}$ , *i.e.*  $p_N(x) = x + N^{\perp}$ ; for each  $M, N \in A, M \supset N$ , let  $p_N^M : E/M^{\perp} \to E/N^{\perp}$  with  $p_N^M(x+M^{\perp}) = x + N^{\perp}$  be the canonical projection. The family  $\{E_N, p_N^M\}_{N,M \in A}$  is a projective system  $(E_N = E/N^{\perp} \text{ are finite-dimensional normed spaces})$  and  $\{E_N, p_N^M\}_{N,M \in A}, E, \{p_N\}_{N \in A}$  satisfy the assumptions of Theorem 1.7, since:

i) The assumption implies that the sequence of p.m.'s.  $\{\mu_n\}_{n=1}^{+\infty}$  is exponentially tight, since:

$$P\left\{\left\|\frac{S_n}{n}\right\| > r\right\}^{\text{Che by shev}} \leq e^{-ntr} E \exp(t \|S_n\|)$$
$$\leq e^{-ntr} \left(\underbrace{E \exp(t \|X_1\|)}_{a < \infty}\right)^n = e^{-n(tr - \log a)}$$

If t, a are constants, and r such that:  $tr - \log a \ge \varepsilon$ for  $\varepsilon > 0$ , we get  $P\left\{ \left\| \frac{S_n}{n} \right\| > r \right\} \le e^{-n\varepsilon}$ .

For given  $\varepsilon > 0$ , we choose the (compact) set:

$$K_{\varepsilon} = \left\{ x / \|x\| \le \varepsilon \right\}$$

ii) For each  $N \in A$  the sequence of p.m.'s  $\left\{\mu_n \circ p_N^{-1}\right\}_{n=1}^{\infty}$  satisfies the full large deviations principle with good rate function:

$$I_{N}(z) = \sup_{\eta \in N} \left\{ \left\langle z, \eta \right\rangle - \log \widehat{\mu \circ p_{N}^{-1}}(\eta) \right\}.$$

In fact, since:

$$\mu_n = L\left(\frac{S_n}{n}\right) \Longrightarrow \mu_n \circ p_N^{-1} = L\left(\frac{\sum\limits_{i=1}^n p_N(X_i)}{n}\right),$$

If we define the r.v.'s  $Y_i = p_N(X_i), i \in \mathbb{N}$ , they are i.i.d. with common distribution  $L(Y_i) = \mu \circ p_N^{-1}$  and values in the space  $E_N = E/N^{\perp}$ . Also

$$\int \exp\left(t\left\|x+N^{\perp}\right\|\right)\mu\circ p_{N}^{-1}\left(\mathrm{d}x\right)=\int \exp\left(t\left\|x\right\|\right)\mu\left(\mathrm{d}x\right)<\infty$$

from hypothesis, so using Cramer's Theorem 2.1 (for finite dimensional spaces, see e.g. [1,4]), we have that the sequence of p.m.'s:

$$L\left(\frac{\sum_{i=1}^{n} Y_{i}}{n}\right) = L\left(\frac{p_{N}(S_{n})}{n}\right) = \mu_{n} \circ p_{N}^{-1}$$

satisfies the large deviations principle with rate function (using that  $N \cong (E/N^{\perp})^*$ ):

$$\begin{split} I_{N}\left(z\right) &= \sup_{\eta \in \left(E/N^{\perp}\right)^{*}} \left\{ \left\langle z, \eta \right\rangle - \log \widehat{\mu \circ p_{N}^{-1}}\left(\eta\right) \right\} \\ &= \sup_{\eta \in N} \left\{ \left\langle z, \eta \right\rangle - \log \widehat{\mu \circ p_{N}^{-1}}\left(\eta\right) \right\} \end{split}$$

From i) and ii), and Theorem 1.7 we get that, the sequence of p.m.'s.  $\{\mu_n\}_{n=1}^{+\infty}$  satisfies the large deviations principle with good rate function

$$I(x) = \sup_{N \in A} I_N(p_N(x)).$$
  
But:  $\langle p_N(x), \eta \rangle = \langle x, \eta \rangle$  and  $\widehat{\mu \circ p_N^{-1}}(\eta) = \widehat{\mu}(\eta)$ , so  
$$I(x) = \sup_{N \in A} I_N(p_N(x))$$
$$= \sup_{N \in A} \sup_{\eta \in N} \left\{ \langle p_N(x), \eta \rangle - \log \widehat{\mu \circ p_N^{-1}}(\eta) \right\}$$
$$= \sup_{\eta \in N} \sup_{N \in A} \left\{ \langle p_N(x), \eta \rangle - \log \widehat{\mu \circ p_N^{-1}}(\eta) \right\}$$
$$= \sup_{\xi \in E^*} \left\{ \langle x, \xi \rangle - \log \widehat{\mu}(\xi) \right\}$$

When someone deals with the empirical measures of an i.i.d sequence, the following large deviations result is true.

# 2) Theorem 2.3. (Sanov's theorem in $\mathbb{R}$ for independent random variables)

Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of independent and identically distributed r.v.'s, taking values in  $\mathbb{R}$  with (common) distribution  $\mu = L(X_1)$ ,  $P(\mathbb{R})$  the space of probability measures on  $\mathbb{R}$  equipped with the weak topology  $\sigma(P(\mathbb{R}), C_b(\mathbb{R}))$ . Then:

1) a) (upper bound)  $\forall F \subset P(\mathbb{R})$  (weakly) closed:

$$\lim_{n \to \infty} \sup \frac{1}{n} \log P\left\{\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \in F\right\} \leq -\inf_{\substack{\nu \in F \\ \nu \in F}} \lambda_{\mu}(\nu)$$

with  $\delta_x$  Dirac's measure defined on x, and

$$\lambda_{\mu}(\nu) = \begin{cases} \int \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu, \ \nu \ll \mu, \nu \in P(\mathbb{R}) \\ \infty \quad \text{otherwise} \end{cases}$$
(9)

(*Kullback-Leibner information number* or *relative entropy of v with respect to \mu*)

b)  $\forall a \ge 0$ , the set  $L_a = \{ \nu / \lambda_\mu (\nu) \le a \}$  is (weakly) compact.

2) (lower bound)  $\forall G \subset P(\mathbb{R})$  open:

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left\{\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \in G\right\} \ge -\inf_{\substack{\nu \in G \\ \nu \in G}} \lambda_{\mu}(\nu)$$

Remark 2.4. Theorem 2.3 is also true in the case of

r.v.'s taking values in a complete separable topological space *S* and the space of probability measures P(S) is endowed with the weak topology (Donker-Varadhan (1976) and Bahadur-Zabell (1979) [1,5]). We prove now a generalization of Theorem 2.3 (the space P(S) is endowed with the  $\tau$ -topology instead of the weak), using suitable projective systems. Also the r.v.'s are taking values on any set *S* which is endowed with a  $\sigma$ -algebra S (no need for topology on *S*).

Let (S,S) be a *measurable space* (*i.e.* S is any set and S a  $\sigma$ -algebra of subsets of S) and assume that the space P(S) is endowed with the  $\tau$ -topology

$$\left(v_n \xrightarrow{\tau} v \Leftrightarrow \int f \mathrm{d} v_n \xrightarrow{n \to \infty} \int f \mathrm{d} v, \forall f \in B(\mathbf{S})\right)$$

where B(S) the space of the bounded, S measurable maps  $f: S \to \mathbb{R}$ ; convergence of nets of p.m.'s is defined in a similar way). Let also B = B(P(S), B(S)) be the  $\sigma$ -algebra induced on P(S) by B(S).

# Theorem 2.5. (Sanov's theorem for the $\tau$ -topology)

Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of i.i.d. r.v.'s, with (common) distribution  $\mu = L(X_1)$ , and values in the set S and S a  $\sigma$ -algebra of subsets of S. Then:

1) a) (upper bound)  $\forall A \in B$ :

$$\limsup_{n \to \infty} \sup \frac{1}{n} \log P\left\{\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \in A\right\} \leq -\inf_{\substack{\nu \in cl_r A}} \lambda_{\mu}(\nu)$$

b)  $\forall a \ge 0$ , the set  $L_a = \{ v / \lambda_\mu (v) \le a \}$  is  $\tau$ -compact.

2) (lower bound)  $\forall A \in B$ :

$$\liminf_{n \to \infty} \inf \frac{1}{n} \log P\left\{\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \in A\right\} \ge -\inf_{\substack{\nu \in in_r A \\ \nu \in in_r A}} \lambda_{\mu}(\nu)$$

Proof.

Let E = P(S), F = B = B(P(S), B(S)) and A = Nthe family of all finite subsets of B(S), directed upward by inclusion. For  $F \in A, E_F = \mathbb{R}^F, p_F : E \to E_F$  is defined by:  $p_F(v) = \left\{ \int f dv \right\}_{f \in F}$  and for

 $F, G \in A, G \supset F$   $p_F^G : E_G \to E_F$  is the restriction map. It is easy to see that: I) the maps  $\{p_F\}_{F \in A}$  are  $F - B_F$ measurable II) the  $\tau$ -topology on E, is the initial topology induced by the maps  $\{p_F\}_{F \in A}$ , making the family  $\{E_F, p_F^G\}_{F \in G \in A}$  a projective system.

If  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \mu_n = L(L_n) = P \circ L_n^{-1}$  and for  $F \in A$ 

the probability measure:

ν

$$\mu_n \circ p_F^{-1} = \left(P \circ L_n^{-1}\right) \circ p_F^{-1} = P \circ \left(\frac{S_n}{n}\right)^{-1},$$
  
where  $S_n = \sum_{j=1}^n Z_j, Z_j = \pi_F(X_j),$  with  $\pi_F : S \to \mathbb{R}^F$ 

defined by  $\pi_F(x) = \{f(x)\}_{f \in F}$  and the r.v.'s  $Z_{i}, j = 1, 2, \cdots$  are i.i.d  $\mathbb{R}^{F}$ -valued and

$$\hat{\mu}_{F}(a) = \int \exp\left(\sum_{f \in F} a(f) f(x)\right) d\mu(x) < \infty$$

Using 2.1. (for  $F \in A$ ), we get that the sequence of p.m.'s  $\left\{\mu_n \circ p_F^{-1}\right\}_{n=1}^{\infty}$  satisfies the large deviations principle with rate function:

$$I_{F}(z) = \sup_{a \in \mathbb{R}^{F}} \left\{ \sum_{f \in F} a(f) z(x) - \log \int \exp\left(\sum_{f \in F} a(f) f(x)\right) d\mu(x) \right\}$$

*i.e.* condition i) of Theorem 1.7 is satisfied.

Also using an argument similar to that of Theorem 2.1 in [6], or else Lemma 2.1 [7] implies that  $\forall a \ge 0$ , the set  $L_a = \{ v / \lambda_u (v) \le a \}$  is  $\tau$ -compact. This, using Lemma 2.2 [7], implies that

$$I_F(z) = \inf \left\{ \lambda_\mu(\nu) / \nu \in P(S), p_F(\nu) = z \right\}$$

(condition ii) of Theorem 1.7). So, using Theorem 1.7, the sequence of p.m.'s  $\{\mu_n\}_{n=1}^{\infty}$  satisfies the large deviations principle with rate function  $\lambda_{\mu}(v)$ .

3) Theorem (Sanov's theorem for exchangeable r.v.'s)

Sanov's Theorem 2.5 is still true in the case when the independence, as a dependence relation among the random variables of a stochastic process, is replaced by a weaker one described below.

**Definition 2.6.** Let  $X_1, \dots, X_n, \dots$  be r.v.s defined on the p.s.  $(\Omega, F, P)$  and values in the m.s. (S, S). We say that the r.v.'s are exchangeable or interchangeable [8], if the joint distribution of any  $\kappa$  of them  $(\kappa \in \mathbb{N})$ , depends only on  $\kappa$  and not the specific r.v.'s. (the r.v.'s are identically distributed but not necessarily independent).

The notion of exchangeability is central in Bayesian Statistics and plays a role analogous to that played by i.i.d sequences in classical frequentist theory (in B.S. an exchangeable sequence is one such that future samples behave like earlier samples, meaning that any order of a finite number of samples is equally like). The bivariate normal distribution, the classical Polya's urn model, any convex combination of i.i.d. r.v.'s, are some examples of exchangeable r.v.'s. An i.i.d sequence is (trivially) an exchangeable one and the same is true for a mixture distribution of i.i.d. sequences. A converse proposition (to this) is the well known, powerful result in the case of exchangeable sequences, de Finetti's theorem.

#### **Theorem 2.7. (de Finetti's representation theorem)**

If  $\{X_n\}_{n=1}^{\infty}$  is a sequence of exchangeable r.v.'s, then

there is a probability space  $(\Theta, M, m)$  and transition probability function  $P(\cdot, \cdot): \Theta \times S \rightarrow [0, 1]$ , *i.e.* a function such that:

a)  $\forall \theta \in \Theta, P(\theta, \cdot)$  is a probability measure on S

b) 
$$\forall A \in S, P(\cdot, A)$$
 is a measurable function on  $\Theta$ , and

$$P(\cdot) = \int_{\Theta} P(\theta, \cdot) dm(\theta)$$
(10)

with  $\hat{P}(\theta, \cdot)$  is the product measure on  $(S^{\mathbb{N}}, S^{\mathbb{N}})$  with all its components equal to  $P(\theta, \cdot)$ . We say that, P is a mixture of the p.m.'s  $P_{\theta}(.) = P(\theta, \cdot)$  with mixing measure m.

Theorem 2.8. (Sanov's theorem for exchangeable r.v.s in  $\tau$ -topology)

Let (S,S) be a measurable space, the space P(S)is endowed with the  $\tau$ -topology and B = B(P(S), B(S)). Let also  $\pi_{\theta} = P_{\theta} \circ X_1^{-1}, \theta \in \Theta$  and:

$$\lambda_{\theta}(\nu) = \begin{cases} \int \frac{\mathrm{d}\nu}{\mathrm{d}\pi_{\theta}} \log\left(\frac{\mathrm{d}\nu}{\pi_{\theta}}\right) \mathrm{d}\pi_{\theta}, \ \nu \ll \pi_{\theta}, \nu \in P(S) \\ \infty, \text{ otherwise} \end{cases}$$
(11)

Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of exchangeable r.v.'s taking values in S and suppose that the function

 $\pi: \Theta \to P(S), \pi(\theta) = \pi_{\theta} = P_{\theta} \circ X_1^{-1}$ is  $\tau$ -continuous. Then:

- 1) If the space  $\Theta$  is *compact*
- a) (upper bound)  $\forall A \in B$ :

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left\{\frac{1}{n} \sum_{i=1}^{n} \delta_{X_{i}} \in A\right\}$$
  
$$\leq -\inf_{\substack{\nu \in cl_{r}A}} \lambda(\nu) \quad \text{with } \lambda(\nu) = \inf_{\substack{\theta \in s(m)}} \lambda_{\theta}(\nu)$$

β)  $\forall a \ge 0$ , the set  $L_a = \{ \nu / \lambda(\nu) \le a \}$  is *τ*-compact. 2) (lower bound)  $\forall A \in B$ :

$$\liminf_{n \to \infty} \frac{1}{n} \log P\left\{\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \in A\right\} \ge -\inf_{\substack{\nu \in in_{\tau}A}} \lambda(\nu)$$

#### Proof.

Using Theorems 2.1 and 2.2 [9], it is enough to prove: whenever  $\theta_n \xrightarrow[n \to \infty]{} \theta, \theta_n, \theta \in \Theta$ , the sequence of p.m.'s  $\mu_n = P_{\theta_n} \circ L_n^{-1}$  satisfies the large deviations principle with rate function  $\lambda_{\theta}(v)$ .

We define the projective system  $\{E_F, p_F^G\}_{F,G,A}$  where A = N the family of all finite subsets of B(S), directed upward by inclusion,  $E_F = \mathbb{R}^F$  for  $F \in A$ , the map  $p_F: E \to E_F$  is defined by  $p_F(v) = \left\{ \int f \mathrm{d} v \right\}_{e \in E}$ and for  $F, G \in A, G \supset F$   $p_F^G : E_G \to E_F$  is the restriction map.

Finally E = P(S), E = B = B(P(S), B(S)). Then:

I) For  $F \in A$ : the p.m.

$$\mu_n \circ p_F^{-1} = \left(P_{\theta_n} \circ L_n^{-1}\right) \circ p_F^{-1} = P_{\theta_n} \circ \left(\frac{S_n}{n}\right)^{-1}$$

where  $S_n = \sum_{j=1}^n Z_j, Z_j = \pi_F(Y_j), \pi_F : S \to \mathbb{R}^F$ 

 $\pi_F(x) = \{f(x)\}_{f \in F}$  and the r.v.s  $Z_j, j = 1, 2, \cdots$  are i.i.d (with respect to the p.m.  $P_{\theta_n}$ ) with values in  $\mathbb{R}^F$ . The map:

$$\varphi_{F}(\theta, z) = \sup_{a \in \mathbb{R}^{F}} \left\{ \sum_{f \in F} \left( a(f) z(x) \right) -\log \int \exp \left( \sum_{f \in F} a(f) f(x) \right) d\pi_{\theta}(x) \right\}$$

is jointly lower-semi-continuous, so using Theorem 3.1. [9] (or directly using Gartner-Ellis theorem), we get that the sequence of p.m  $\{\mu_n \circ p_F^{-1}\}_{n=1}^{\infty}$  satisfies the large deviations principle with rate function:

$$I_{F}^{\theta}(z) = \sup_{a \in \mathbb{R}^{F}} \left\{ \sum_{f \in F} a(f) z(x) - \log \int \exp \left( \sum_{f \in F} a(f) f(x) \right) d\pi_{\theta}(x) \right\}$$

II) It can be proved (in a way analogous to Theorem 2.1 Daras [6], see also the proof of Theorem 2.5) that

$$I_{F}^{\theta}(z) = \inf \left\{ \lambda_{\theta}(v) / v \in P(S), p_{F}(v) = z \right\}$$

Finally, the result follows using I) and II) and Theorem 1.7.

Remark 2.9.

a) Sanov's theorem is true in a more general setting, namely when the p.m. P is a mixture of p.m.'s [6]. Then, Theorem 2.8 follows, as a corollary, using de Finetti's theorem.

b) Theorem 2.8 extends a result of Dinwoodie and Zabell [9]. They prove their statement for a sequence  $\{X_n\}_{n=1}^{\infty}$  of r.v.'s taking values in a Polish space S (no need here for topology on S) and the space P(S) is endowed with the weak topology (stronger than the  $\tau$ -topology).

#### 4) Moderate deviations

Let  $\{b_n\}_{n=1}^{\infty}$  be a positive real sequence such that:

$$\frac{b_n}{n^{\frac{1}{2}}} \xrightarrow[n \to \infty]{} \infty, \frac{b_n}{n} \xrightarrow[n \to \infty]{} 0, \qquad (12)$$

and  $\{X_n\}_{n=1}^{\infty}$  a sequence of exchangeable r.v.s with distribution  $\mu = P \circ X_1^{-1}$  and for  $n \in \mathbb{N}$ :

$$M_{n} = \frac{n}{b_{n}} (L_{n} - \mu) = \frac{1}{b_{n}} \sum_{i=1}^{n} (\delta_{X_{i}} - \mu)$$
(13)

Let  $\Delta(S)$  be the subspace of B(S) consisting of all those maps g, such that  $\int_{S} g(x) d\pi_{\theta}(x) = 0, \forall \theta \in \Theta$ . Endow the space M(S) of finite signed measures on S with the topology  $\tau_{\Delta}$  generated by  $\Delta(S)$ , *i.e.* the smallest topology making the maps of the form:

 $v \xrightarrow{\nu \in M(S)} \int g d\nu, g \in \Delta(S)$  continuous and let

 $B_{\Delta} = B(M(S), \Delta(S))$  the  $\sigma$ -algebra induced on M(S)

by 
$$\Delta(S)$$
. Then if  $\widehat{\mu_n} = P \circ M_n^{-1}$  and

$$\hat{I}_{\theta}(\nu) = \begin{cases} \frac{1}{2} \int \left(\frac{\mathrm{d}\nu}{\mathrm{d}\pi_{\theta}}\right)^2 \mathrm{d}\pi_{\theta}\nu \ll \pi_{\theta} \\ \infty \quad \text{otherwise} \end{cases}$$
(14)

the following large deviations principle is true [6].

Theorem 2.10. (moderate deviations for empirical measures)

Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of exchangeable r.v.'s taking values in S. Assume that the map

 $\pi: \Theta \to M(S), \pi(\theta) = \pi_{\theta} = P_{\theta} \circ X_1^{-1}$  is  $\tau$ -continuous. Then:

1) If the space  $\Theta$  is compact, then

a) (upper bound)  $\forall B \in B_{\Delta}$ :

$$\limsup_{n \to \infty} \sup \frac{n}{b_n^2} \log \widehat{\mu_n}(B) \le -\inf_{x \in cl_{\tau_\Delta} B} \widehat{I}(v) \quad \text{with}$$

$$\widehat{I}(v) = \inf_{x \in cl_{\tau_\Delta} B} \widehat{I}(v)$$

$$I(v) = \inf_{\theta \in s(m)} I_{\theta}(v)$$

b)  $\forall a \ge 0$ , the level set  $L_a = \{\nu / \lambda(\nu) \le a\}$  is  $\tau$ -compact.

2) (lower bound)  $\forall B \in B_{\Delta}$ :

$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log \widehat{\mu_n}(B) \ge -\inf_{x\in \operatorname{int}_{\tau_\Delta} B} \hat{I}(v)$$

#### Remark 2.11.

a) Large deviations with normalizing constants of the form (12) are being called *moderate deviations* [6,10].

b) Theorem 2.10 generalizes Theorem 3.1. in [11]. There, the sequence  $\widehat{\mu_n}$  is based on a sequence of r.v.'s taking values in a m.s. (S,S) and the space M(S) is endowed with the  $\tau$ -topology.

c) Theorem 2.10 is true in general, namely when the p.m. P is a mixture of p.m.'s [6]. Then, Theorem 2.10 follows using de Finetti's theorem.

#### REFERENCES

- A. Dembo and O. Zeitouni, "Large Deviations Techniques and Applications," Jones and Bartlett, Boston, 1993.
- [2] D. W. Strook, "An Introduction to the Theory of Large

Deviations," Springer-Verlag, New York, 1984. doi:10.1007/978-1-4613-8514-1

- [3] A. de Acosta, "Exponential Tightness and Projective Systems in Large Deviation Theory," In: D. Pollard, E. Togersen and G. Yang, Eds., *Festschrift for Lucien Le Cam*, Springer, New York, 1997, pp. 143-156,
- [4] J. A. Bucklew, "Large Deviation Techniques in Decision, Simulation and Estimation," John Wiley & Sons, New York, 1990.
- [5] P. Dupuis and R. Ellis, "A Weak Convergence Approach to the Large Deviations," Wiley Series in Probability, New York, 1997.
- [6] T. Daras, "Large and Moderate Deviations for the Empirical Measures of an Exchangeable Sequence," *Statistics & Probability Letters*, Vol. 36, No. 1, 1997, pp. 91-100. doi:10.1016/S0167-7152(97)00052-7

- [7] A. de Acosta, "On Large Deviations of Empirical Measures in the T-Topology," *Journal of Applied Probability*, Vol. 31, 1994, pp. 41-47.
- [8] D. J. Aldous, "Exchangeability and Related Topics. Ecole d' Ete de Probabilites de Saint-Flour XIII 1983," Lecture Notes in Mathematics 117, Springer, New York, 1985.
- [9] I. H. Dinwoodie and S. L. Zabell, "Large Deviations for Exchangeable Random Vectors," *The Annals of Probability*, Vol. 20, No. 3, 1992, pp. 1147-1166. <u>doi:10.1214/aop/1176989683</u>
- [10] T. Daras, "Trajectories of Exchangeable Sequences: Large and Moderate Deviations Results," *Statistics & Probability Letters*, Vol. 39, No. 4, 1998, pp. 289-304.
- [11] A. de Acosta, "Projective Systems in Large Deviation Theory II: Some Applications," *Probability in Banach Spaces* 9, Vol. 35, 1994, pp. 241-250.