Elect. Comm. in Probab. **9** (2004), 60–71

**ELECTRONIC COMMUNICATIONS in PROBABILITY**

# LARGE DEVIATIONS FOR MIXTURES

J. D. BIGGINS
*Department of Probability and Statistics, Hicks Building, The University of Sheffield, Sheffield, S3 7RH, U.K.*
email: j.biggins@sheffield.ac.uk

*Submitted* 11 December 2003, *accepted in final form* 29 March 2004

AMS 2000 Subject classification: 60F10
Keywords: Lare deviations, mixture, rate function

*Abstract*
Suppose the probability measures $(\mu^n)$ on $\Theta$ obey a large deviation principle (LDP). Suppose too that $\mu^n$ is concentrated on $\Theta_n$ and that, for $\theta(n) \in \Theta_n$ with $\theta(n) \to \theta \in \Theta$, the probability measures $(P^n_{\theta(n)})$ on $X$ also obey an LDP. The main purpose of this paper is to give conditions which allow an LDP for the mixtures $(P^n)$, given by $P^n(A) = \int P^n_\theta(A)d\mu^n(\theta)$, to be deduced. Chaganty (1997) also considered this question, but under stronger assumptions. The treatment here follows that of Dinwoodie and Zabell (1992) who, motivated by exchangeability, considered the case where $\mu^n$ does not vary with $n$.

## 1 Introduction and motivation

Let $\mu^n$ be a (mixing) probability measure on the Borel $\sigma$-algebra of a topological space $\Theta$, concentrated on (the measurable set) $\Theta_n$. For each $\theta \in \Theta_n$, let $P^n_\theta$ be a probability measure on the Borel $\sigma$-algebra of the topological space $X$ for which the map $\theta \to P^n_\theta(A)$ is measurable on $\Theta_n$ for every measurable $A \subset X$. For definiteness, let $P^n_\theta$ be given by some fixed probability measure on $X$ when $\theta \notin \Theta_n$. Based on these, the joint distribution, $\widetilde{P}^n$, and the marginal distribution, $P^n$, obtained by mixing over $\theta$, have the usual definitions:

$$d\widetilde{P}^n(\theta, x) = dP^n_\theta(x)d\mu^n(\theta) \quad \text{and} \quad dP^n(x) = \int_\Theta dP^n_\theta(x)d\mu^n(\theta) = \int_{\Theta_n} dP^n_\theta(x)d\mu^n(\theta).$$

Throughout, $\Theta$ and $X$ are assumed to be Hausdorff (i.e. distinct points can be separated by disjoint open sets) and $\Theta$ is assumed to be first countable (i.e. for each $\theta$ there is a countable collection of neighbourhoods such that every neighbourhood of $\theta$ contains one of this collection), which implies that convergence in $\Theta$ can be described using sequences. However, $X$ is not assumed to be first countable. Hence $X$ could, for example be a space of measures with the $\tau$-topology, which is the one induced by the bounded measurable functions. In this context, the assumption that $P^n_\theta$ is a measure on the Borel sets of $X$, rather than some smaller $\sigma$-algebra, does rule out some natural examples — see Section 6.2 in Dembo and Zeitouni (1993).

The sequence of probability measures $(P^n)$ (on the Borel $\sigma$-algebra of the topological space $X$) obeys a large deviation principle (LDP) if there is a lower semicontinuous non-negative function $\lambda$ (a rate function) such that for every closed $F$ and open $G$

$$\limsup \frac{\log P^n(F)}{n} \leq -\inf_{y\in F}\lambda(x) \quad \text{and} \quad \liminf \frac{\log P^n(G)}{n} \geq -\inf_{y\in G}\lambda(x).$$

The rate function $\lambda$ is called 'good' (or 'proper') if for every finite $\beta$ the set $\{x : \lambda(x) \leq \beta\}$ is compact. The sequence satisfies a weak LDP if the upper bound holds for compact, rather than closed, $F$. Furthermore, the sequence of probability measures $(P^n)$ is said to be exponentially tight if for every $\alpha > 0$ there is a set $O_\alpha$ whose complement is compact with

$$\limsup \frac{\log P^n(O_\alpha)}{n} < -\alpha. \tag{1}$$

The main idea is to combine large deviation results for $(P_\theta^n)$ and $(\mu^n)$ to give large deviation results for the marginal distributions $(P^n)$. The treatment draws heavily on that in Section 2 of Dinwoodie and Zabell (1992), who consider the case where $\mu^n$ does not depend on $n$. They used their results to consider large deviations for exchangeable sequences in rather general spaces; this motivation led naturally to the assumption that $\mu^n$ was independent of $n$. The basic framework adopted here is used by Chaganty (1997), who also provides a number of statistical applications, but the treatment here is more general in two main ways. Firstly, the use of $\Theta_n$ rather than $\Theta$ is needed to deal with our motivating example, is natural, and produces genuine complication in the argument. Secondly, Chaganty (1997) confines attention to cases where $\Theta$ and $X$ are both Polish, whereas here greater topological generality, in the spirit of Dinwoodie and Zabell (1992), is maintained. A final, arguably less significant, difference is that the focus in Chaganty (1997) is on the LDP for the joint distributions $(\widetilde{P}^n)$, rather than the marginals $(P^n)$. Chaganty's main result will be a consequence of the results here.

The motivating example for developing these results arose in the study of random graphs. The classical random graph is very well understood, but fails to match up to the graphs occurring in many applications. Recently, Cannings and Penman (2003) suggested a model with more flexibility; see also Penman (1998). Suppose a graph is to have $n$ vertices. Then, to produce random graphs with a correlation structure between edge occurrences, Cannings and Penman (2003) proposed that each vertex is independently assigned one of a number of colours, and the probability that an edge arises depends on the colours of its two vertices. The problem posed is to find an LDP for the number of edges, as $n$ becomes large. This falls exactly into the framework proposed. To elucidate, consider the graph with $n$ vertices. Let the proportions of these vertices of the various possible colours be $\theta$; then $\mu^n$ is the distribution $\theta$. Given $n$ and $\theta$ the number of edges is obtained as the sum of independent (but not identically distributed) random variables; this specifies $P_\theta^n$. Note that for finite $n$ the possible values of $\theta$ are confined to those with $n\theta$ containing integers; this defines $\Theta_n$ here. The details of this application are discussed in Biggins and Penman (2003).

The next section contains the statements of the main results. The following two contain their proofs and those of various intermediate results. A brief final section mentions some possible directions for further work.

## 2 The main results

For easier references in the statement and proofs, various assumptions will be labelled. The first two concern the LDP and the exponential tightness for the mixing distributions $(\mu^n)$ on

$\Theta$.
**ldp$\mu$:** *$(\mu^n)$ satisfies an LDP with rate $\psi$.*
**tight$\mu$:** *$(\mu^n)$ is exponentially tight.*
When $\Theta$ is compact **tight$\mu$** holds automatically.
The third assumption is an LDP statement for the conditional distributions. In Dinwoodie and Zabell (1992), this kind of condition is called exponential continuity. To state it, a little more notation is needed. Let $\widetilde{\Theta}$ be the limit set of sequences with the $n$th member from $\Theta_n$; thus

$$\widetilde{\Theta} = \{\theta \in \Theta : \exists\, \theta(n) \in \Theta_n, \theta(n) \to \theta\}.$$

It is easy to check that $\widetilde{\Theta}$ is closed; see Lemma 10 in the next section. Most applications will have $\widetilde{\Theta} = \Theta$.
**exp-cty:** *$\widetilde{\Theta}$ is non-empty and whenever $\theta(n) \in \Theta_n$ and $\theta(n) \to \theta \in \widetilde{\Theta}$, $(P^n_{\theta(n)})$ satisfies an LDP with rate $\lambda_\theta$.*
When $\theta \notin \widetilde{\Theta}$, let $\lambda_\theta(x) = \infty$ for all $x$. Since $\lambda_\theta$ is a rate function it is lower semicontinuous on $X$ for each $\theta$. The fourth assumption is in similar vein.
**lsc:** *The function $\lambda$, defined by*

$$\lambda(x) = \inf\{\lambda_\theta(x) + \psi(\theta) : \theta \in \Theta\}, \tag{2}$$

*is lower semicontinuous on $X$.*
Recall that a topological space is regular if for every open $U$ containing $x$ there is an open $O$ also containing $x$ with its closure contained in $U$.

**Theorem 1** *Suppose* **ldp$\mu$**, **tight$\mu$**, **exp-cty** *and* **lsc** *all hold. Suppose also that $\Theta$ is regular. Then $\{P^n\}$ satisfies an LDP with rate function $\lambda$. When $\Theta$ is compact* **tight$\mu$** *holds automatically. When $\psi$ takes only the value 0 the requirement that $\Theta$ is regular is not needed.*

It turns out that in Theorem 1 it is automatic that $\lambda$ is a good rate function under the extra conditions that $X$ is regular and the rate functions $\psi$ and $\lambda_\theta$ are good. This is the essential content of the next theorem.

**Theorem 2** *Suppose* **ldp$\mu$**, **tight$\mu$** *and* **exp-cty** *all hold and that $\Theta$ and $X$ are regular. Suppose also that the following conditions hold.*
**good$\psi$:** *The rate function $\psi$ in* **ldp$\mu$** *is good.*
**good$\lambda_\theta$:** *For each $\theta \in \widetilde{\Theta}$, the rate function $\lambda_\theta$ in* **exp-cty** *is good.*
*Then $\lambda$, defined at (2), is a good rate function and $(P^n)$ satisfies an LDP with rate function $\lambda$.*

The next result notes that often the rate function being good implies exponential tightness. It shows that the hypothesis **tight$\mu$** in Theorem 2, and later results, is superfluous when **good$\psi$** holds and $\Theta$ is locally compact or Polish. For locally compact spaces the result is contained in Exercise 1.2.19 in Dembo and Zeitouni (1993). For Polish spaces it is Lemma 2.6 in Lynch and Sethuraman (1987) — see also Exercise 4.1.10 in Dembo and Zeitouni (1993).

**Lemma 3** *Suppose $\Theta$ is either locally compact or Polish and* **good$\psi$** *holds. Then* **tight$\mu$** *holds.*

The following condition is a natural extension of the property that $\lambda_\theta$ is lower semicontinuous on $X$ for each $\theta$.

**jnt-lsc:** *$\lambda_\theta(x)$ is jointly lower semicontinuous in $(\theta, x) \in \Theta \times X$.*

Lemma 3.1(i) in Dinwoodie and Zabell (1992) gives some general conditions for **jnt-lsc** to hold; see also Lemma 3.2 in Chaganty (1997). The next result uses this condition to provide one way to check that **lsc** holds.

**Proposition 4** *Suppose* **good**$\psi$ *and* **jnt-lsc** *hold. Then* **lsc** *holds.*

Theorems 1 and 2 approach the LDP for $(P^n)$ directly. The next result, which essentially contains Theorem 2.3 in Chaganty (1997), approaches the question through a weak LDP for the joint distributions $(\widetilde{P}^n)$.

**Theorem 5** *Suppose* **ldp**$\mu$, **exp-cty** *and* **jnt-lsc** *hold. Suppose too that both $\Theta$ and $X$ are regular.*

*(a) Then $(\widetilde{P}^n)$ satisfies a weak LDP with rate function $\lambda_\theta(x) + \psi(\theta)$. Furthermore, when $\Theta$ is locally compact the LDP in* **ldp**$\mu$ *can be replaced by a weak LDP and, similarly, when $X$ is locally compact a weak LDP is enough in* **exp-cty**.

*(b) If in addition $(\widetilde{P}^n)$ is exponentially tight then the (full) LDP holds with a good rate function and $(P^n)$ satisfies an LDP with the good rate function $\lambda$ defined at (2).*

It is desirable to have conditions that ensure that $(\widetilde{P}^n)$ is exponentially tight in order to use the last part of the previous result. The next three Propositions, and Lemma 3, provide a variety of conditions for this. Before giving them one further definition is needed. A family of sequences (of probability measures) is uniformly exponentially tight if, in (1), for every $\alpha > 0$ the same $O_\alpha$ can be used for every sequence.

**Proposition 6** *Suppose that* **tight**$\mu$ *holds. Suppose also that the following condition holds.*

**uni-tight:** *For each $\theta \in \widetilde{\Theta}$, the family of sequences $\{(P^n_{\theta(n)}) : \theta(n) \in \Theta_n,\ \theta(n) \to \theta\}$ is uniformly exponentially tight.*

*Then $(\widetilde{P}^n)$ is exponentially tight.*

Lemma 3.2 in Dinwoodie and Zabell (1992) gives conditions under which **uni-tight** holds when, for each $\theta$, $P^n_\theta$ is the distribution (on a rather general space) of the average of independent identically distributed variables. In Lemma 3 it is noted that **tight**$\mu$ can be replaced by **good**$\psi$ when $\Theta$ is locally compact or Polish. The next Proposition is in a similar spirit.

**Proposition 7** *If $X$ is locally compact then, in Proposition 6,* **uni-tight** *can be replaced by* **good**$\lambda_\theta$.

The final result in this trio is not useful for getting the LDP for the marginal $(P^n)$ from Theorem 5, since the conditions are the same as those in Theorem 2 except for a more restrictive condition on $X$. However, it could be used to strengthen the weak LDP for the joint distributions $(\widetilde{P}^n)$ to a (full) LDP.

**Proposition 8** *Suppose* **ldp**$\mu$, **tight**$\mu$, **exp-cty**, **good**$\psi$ *and* **good**$\lambda_\theta$ *hold, $\Theta$ is regular and $X$ is Polish. Then $(\widetilde{P}^n)$ is exponentially tight.*

# 3 Proofs of Theorems 1 and 2 and Proposition 4

By definition, a function $f$ on $X$ is lower semicontinuous at $x$ if for each $c < f(x)$ there is an open $U$ containing $x$ such that $f(y) > c$ for every $y \in U$.

**Lemma 9** *If $X$ is regular and $f$ is lower semicontinuous on $X$ then for every $x$ and $c < f(x)$ there is a closed set $C_x$ with $x$ in its interior and $f(y) > c$ for all $y \in C_x$.*

**Proof**. Fix $x$ and $c < f(x)$. By the definition of lower semicontinuity, there is an open set $U$ containing $x$ with $f(y) > c$ for $y \in U$. Applying regularity, there is an open set $V_x$ containing $x$ with its closure inside $O_x$. Take $C_x$ to be the closure of $V_x$. □

**Lemma 10** *$\widetilde{\Theta}$ is closed.*

**Proof**. Suppose that $\theta_n^{(k)} \to \theta^{(k)} \to \theta$ with $\theta_n^{(k)} \in \Theta_n$. Take $U_i$ from a countable open neighbourhood base of $\theta$. For some $k(i) > k(i-1)$, $\theta^k \in U_i$ for $k \geq k(i)$. Now $U_i$ is also an open neighbourhood of $\theta^{k(i)}$ and so there is an $n^{k(i)}$ with $\theta_n^{(k(i))} \in U_i$ for all $n \geq n^{k(i)}$. The sequence $\vartheta_n = \theta_n^{(k(i))} \in \Theta_n$ for $n^{k(i)} \leq n < n^{k(i+1)}$ converges to $\theta$, and so $\theta \in \widetilde{\Theta}$. □

**Lemma 11** *Suppose* **ldp**$\mu$ *and* **exp-cty** *hold. Let $(\theta, x) \in G^* \subset \Theta \times X$, where $G^*$ is open. Then*

$$\liminf \frac{\log \widetilde{P}^n(G^*)}{n} \geq -(\lambda_\theta(x) + \psi(\theta)).$$

*In particular, for $G$ open in $X$,*

$$\liminf \frac{\log P^n(G)}{n} \geq -\inf\{\lambda_\theta(x) + \psi(\theta) : \theta \in \Theta, x \in G\}.$$

**Proof**. The result is true when $\lambda_\theta(x) = \infty$. Hence attention can focus on $\lambda_\theta(x) < \infty$. There are open sets $O \subset \Theta$ and $U \subset X$ containing $\theta$ and $x$ respectively with $O \times U \subset G^*$. Then

$$\widetilde{P}^n(O \times U) = \int_O P_\vartheta^n(U) d\mu^n(\vartheta).$$

For every $\epsilon > 0$, there exists an open set $O_\epsilon \subset O$ containing $\theta$ and an integer $N_\epsilon$ such that for $n \geq N_\epsilon$ and every $\gamma \in O_\epsilon \cap \Theta_n$

$$P_\gamma^n(U) > \exp(-n[\lambda_\theta(x) + \epsilon]).$$

To demonstrate this, suppose it fails. Then there are $n(i) > n(i-1)$ and $\theta(i) \in \Theta_{n(i)}$ with $\theta(i) \to \theta$ such that

$$P_{\theta(i)}^{n(i)}(U) \leq \exp(-n(i)[\lambda_\theta(x) + \epsilon]),$$

and then

$$\liminf \frac{\log P_{\theta(i)}^{n(i)}(U)}{n(i)} \leq -\lambda_\theta(x) - \epsilon,$$

which contradicts the lower bound in the LDP in **exp-cty**.

Thus, for $n \geq N_\epsilon$,

$$\begin{aligned} \widetilde{P}^n(G^*) \geq \widetilde{P}^n(O_\epsilon \times U) &= \int_{O_\epsilon} P^n_\vartheta(U) d\mu^n(\vartheta) \\ &= \int_{O_\epsilon \cap \Theta_n} P^n_\vartheta(U) d\mu^n(\vartheta) \\ &\geq \exp(-n[\lambda_\theta(x) + \epsilon])\mu^n(O_\epsilon \cap \Theta_n) \\ &= \exp(-n[\lambda_\theta(x) + \epsilon])\mu^n(O_\epsilon) \end{aligned}$$

and so, using **ldp**$\mu$,

$$\liminf \frac{\log \widetilde{P}^n(G^*)}{n} \geq -\lambda_\theta(x) - \epsilon + \liminf \frac{\log \mu^n(O_\epsilon)}{n} \geq -\lambda_\theta(x) - \epsilon - \psi(\theta).$$

The last part comes from taking $G^* = \Theta \times G$, for then $\widetilde{P}^n(G^*) = P^n(G)$. □

**Lemma 12** *Suppose* **ldp**$\mu$ *and* **exp-cty** *hold. Suppose too that* **tight**$\mu$ *holds and* $\Theta$ *is regular. Let* $F \subset X$ *be closed. Then*

$$\limsup \frac{\log P^n(F)}{n} \leq -\inf\{\lambda_\theta(x) + \psi(\theta) : (\theta, x) \in \Theta \times F\}.$$

**Proof**. Fix $F$. Let $c$ and $d$ be such that

$$c < d = \inf\{\lambda_\theta(x) + \psi(\theta) : (x, \theta) \in F \times \Theta\}.$$

Using **tight**$\mu$, let $O$ be such that

$$\limsup \frac{\log \mu^n(O)}{n} < -c$$

and let $S$ be the (compact) complement of $O$. Then

$$P^n(F) = \int_\Theta dP^n_\theta(F) d\mu^n(\theta) \leq \int_S dP^n_\theta(F) d\mu^n(\theta) + \mu^n(O).$$

Let $\Lambda(\theta) = \inf\{\lambda_\theta(x) : x \in F\}$. Let $\epsilon > 0$ with $c < 1/\epsilon$. Now let

$$\Lambda^\epsilon(\theta) = \min\{\Lambda(\theta) - \epsilon, 1/\epsilon\} \quad \text{and} \quad \psi^\epsilon(\theta) = \min\{\psi(\theta) - \epsilon, 1/\epsilon\}.$$

For $\theta \in S$, by **exp-cty**, there is an open set $O_\theta$ containing $\theta$ and an integer $N_\theta$ such that for $n \geq N_\theta$ and every $\gamma \in O_\theta \cap \Theta_n$

$$P^n_\gamma(F) \leq \exp(-n\Lambda^\epsilon(\theta)).$$

To demonstrate this, suppose it fails. Then there are $n(i) > n(i-1)$ and $\theta(i) \in \Theta_{n(i)}$ with $\theta(i) \to \theta$ such that

$$P^{n(i)}_{\theta(i)}(F) > \exp(-n(i)\Lambda^\epsilon(\theta)),$$

and then

$$\limsup \frac{\log P^{n(i)}_{\theta(i)}(F)}{n(i)} \geq -\Lambda^\epsilon(\theta),$$

which contradicts the upper bound in the LDP in **exp-cty**.
Furthermore, using the lower semicontinuity of $\psi$, by taking $O_\theta$ to be smaller if necessary,

$$\psi(\vartheta) > \psi^\epsilon(\theta) \text{ for } \vartheta \in O_\theta,$$

and, using regularity of $\Theta$, there is an open set $V_\theta$ with closure $\overline{V}_\theta$ such that $\theta \in V_\theta$ and $\overline{V}_\theta \subset O_\theta$.
Now $(V_\theta : \theta \in S)$ is an open covering of $S$. Since $S$ is compact a finite subcover $(V_{\theta(i)})_{1\le i\le k}$ exists. Then, for sufficiently large $n$,

$$\begin{aligned}
P^n(F) &\le \mu^n(O) + \sum_{i=1}^k \int_{V_{\theta(i)}} P^n_\vartheta(F)d\mu^n(\vartheta)\\
&= \mu^n(O) + \sum_{i=1}^k \int_{V_{\theta(i)}\cap\Theta_n} P^n_\vartheta(F)d\mu^n(\vartheta)\\
&\le \mu^n(O) + \sum_{i=1}^k \exp(-n\Lambda^\epsilon(\theta(i)))\mu^n\left(\overline{V}_{\theta(i)}\right)\\
&\le \mu^n(O) + \sum_{i=1}^k \exp(-n\Lambda^\epsilon(\theta(i)))\exp(-n\psi^\epsilon(\theta(i))).
\end{aligned}$$

Hence, since $c < 1/\epsilon$,

$$\begin{aligned}
\limsup \frac{\log P^n(F)}{n} &\le -\min\left\{\min_{1\le i\le k}\left\{\Lambda^\epsilon(\theta(i)) + \psi^\epsilon(\theta(i))\right\}, c\right\}\\
&\le -\min\left\{\min_{1\le i\le k}\left\{\Lambda(\theta(i)) + \psi(\theta(i)) - 2\epsilon\right\}, c\right\}\\
&\le -\min\left\{d-2\epsilon, c\right\}.
\end{aligned}$$

Since $c < d$ and $\epsilon > 0$ are arbitrary, the result follows. □

**Lemma 13** *In Lemma 12, if $\psi$ takes only the value 0 then the hypothesis that $\Theta$ is regular is not needed.*

**Proof**. When $\psi$ takes only the value 0 there is no need to introduce $V_\theta$; it suffices to take a finite subcover from $(O_\theta : \theta \in \Theta)$. □

**Proof of Theorem 1**. The last part of Lemma 11 gives the lower bound for open sets, Lemma 12 gives the upper bound for closed sets. Finally, $\lambda$ is lower semicontinuous by assumption. Lemma 13 gives the simplification contained in the final assertion. □

Some further work is needed to deal with the conditions implying that $\lambda$ is good, to produce a proof of Theorem 2.

**Lemma 14** *Suppose* **ldp**$\mu$, **exp-cty** *and* **good**$\psi$ *holds. Fix $c < \infty$. Let $K = \{\theta : \psi(\theta) \le c\}$ and*

$$\overline{\lambda}(y) = \inf\{\lambda_\theta(y) + \psi(\theta) : \theta \in K\}.$$

*Then the sets $\{y : \lambda(y) \le c\}$ and $\{y : \overline{\lambda}(y) \le c\}$ are the same, and $\lambda$ and $\overline{\lambda}$ agree on this set.*

**Proof**. The set $K$ is compact because $\psi$ is good. Then, since $\lambda_\theta$ is non-negative, it is easy to see that

$$\overline{\lambda}(y) \geq \lambda(y) \geq \min\{\overline{\lambda}(y), c\},$$

which gives the result. □

**Lemma 15** *Suppose* **ldp**$\mu$*,* **exp-cty***,* **good**$\psi$ *and* **good**$\lambda_\theta$ *hold. Suppose also that both* $\Theta$ *and* $X$ *are regular. Then* $\lambda$*, given by (2), is a good rate function.*

**Proof**. Take $\epsilon$ and $\alpha$ with $0 \leq \alpha < \alpha + 2\epsilon < \infty$. Let $K$ be $\{\theta : \psi(\theta) \leq c = \alpha + 2\epsilon\}$, which is compact, by **good**$\psi$. Now, by Lemma 14,

$$\{y : \overline{\lambda}(y) \leq \alpha\} = \{y : \lambda(y) \leq \alpha\}.$$

Denote this set by $L^\alpha$ and suppose $\alpha$ was selected so that $L^\alpha$ is not compact. Then there exists a net $\{(\theta(i), x(i)) : i \in I\} \subset K \times X$ such that $\{x(i)\} \subset L^\alpha$, $\{x(i)\}$ has no convergent subnet and

$$\lambda_{\theta(i)}(x(i)) + \psi(\theta(i)) \leq \alpha + \epsilon$$

for all $i$. Note that this implies that $\theta(i) \in \widetilde{\Theta}$ because, by definition, $\lambda_\theta(x) = \infty$ for $\theta \notin \widetilde{\Theta}$. Since $K$ is compact and first countable there is a subsequence $(\theta(i_k), x(i_k))$ such that $\theta(i_k) \to \theta$, where $\theta \in \widetilde{\Theta}$ since $\widetilde{\Theta}$ is closed. Furthermore, because $\psi$ is lower semicontinuous, $\liminf \psi(\theta(i)) \geq \psi(\theta)$ and $\psi(\theta) \leq \alpha + \epsilon$.
Take $\beta = \alpha + 3\epsilon$. The level set of $\lambda_\theta$ given by $L^\beta_\theta = \{x : \lambda_\theta(x) \leq \beta - \psi(\theta)\}$ is compact, by **good**$\lambda_\theta$. Hence, for large enough $k_0$, $C_0 = \{x(i_k); k \geq k_0\}$ must be in the complement of $L^\beta_\theta$. Following exactly the argument in Lemma 2.1 in Dinwoodie and Zabell (1992), $C_0$ is closed and so, using the regularity of $X$ and the compactness of $L^\beta_\theta$, there are open sets separating $C_0$ and $L^\beta_\theta$. Hence, there is an open set $U$ containing $C_0$ with closure $C$ in the complement of $L^\beta_\theta$.
Now take $\vartheta(i,n) \in \Theta_n$ with $\vartheta(i,n) \to \theta(i)$. By the LDP lower bound in **exp-cty**,

$$\liminf \frac{\log P^n_{\vartheta(n,i)}(U)}{n} \geq -\inf\{\lambda_{\theta(i)}(x) : x \in U\} \geq \psi(\theta(i)) - \alpha - \epsilon.$$

Hence, selecting suitable subsequences, there is an increasing sequence $n(k)$ and $\vartheta(k) \in \Theta_{n(k)}$ such that $\vartheta(k) \to \theta$ and

$$\frac{\log P^{n(k)}_{\vartheta(k)}(U)}{n(k)} \geq \psi(\theta) - \alpha - 2\epsilon.$$

By the LDP upper bound in **exp-cty**

$$\limsup \frac{\log P^{n(k)}_{\vartheta(k)}(C)}{n(k)} \leq \psi(\theta) - \beta = \psi(\theta) - \alpha - 3\epsilon,$$

because $C$ is in the complement of $L^\beta_\theta$. Since $U \subset C$ this contradicts the previous inequality. Therefore $L^\alpha$ must be compact. It is therefore also closed, since $X$ is Hausdorff, which means $\lambda$ is lower semi-continuous. □

**Proof of Theorem 2**. The last part of Lemma 11 gives the lower bound for open sets, Lemma 12 gives the upper bound for closed sets. Lemma 15 shows that $\lambda$ is good rate function under the stated conditions. □

**Proof of Proposition 4.** It must be shown that for every $x$ and $c < \lambda(x)$ there is a neighbourhood $U$ of $x$ with $\lambda(y) > c$ for every $y \in U$.
Fix $x$, $c < \lambda(x)$ and $\epsilon > 0$. Let $K = \{\theta : \psi(\theta) \le c\}$, which is compact because $\psi$ is good. Then, by Lemma 14, $\{y : \lambda(y) > c\}$ and $\{y : \overline{\lambda}(y) > c\}$ are the same. Thus it will be enough to show that the latter set contains a neighbourhood of $x$.
Let

$$\lambda_\theta^\epsilon(x) = \min\{\lambda_\theta(x) - \epsilon, 1/\epsilon\} \text{ and } \psi^\epsilon(\theta) = \min\{\psi(\theta) - \epsilon, 1/\epsilon\}.$$

For each $\theta$, because $\lambda_\theta(y)$ is jointly lower semicontinuous, and $\psi$ is lower semicontinuous there are open sets $O_\theta \subset \Theta$ and $U_\theta \subset X$ containing $\theta$ and $x$ respectively such that throughout $O_\theta \times U_\theta$

$$\lambda_\vartheta(y) > \lambda_\theta^\epsilon(x) \text{ for } (\vartheta, y) \in O_\theta \times U_\theta$$

and

$$\psi(\vartheta) > \psi^\epsilon(\theta) \text{ for } \vartheta \in O_\theta.$$

The $\{O_\theta : \theta \in C\}$ cover $C$, and so there is a finite subcover, $(O_{\theta(i)})_{1 \le i \le k}$. Let $U = \cap_i U_{\theta(i)}$, which is open and contains $x$. Then for $y \in U$

$$\begin{aligned} \overline{\lambda}(y) &\ge \min_i\{\inf\{\lambda_\theta(y) + \psi(\theta) : \theta \in O_{\theta(i)}\}\} \\ &\ge \min_i\{\lambda^\epsilon_{\theta(i)}(x) + \psi^\epsilon(\theta(i))\} \\ &\ge \min\{\lambda(x) - 2\epsilon, (2\epsilon)^{-1}\} > c, \end{aligned}$$

provided $\epsilon$ is small enough. Then

$$x \in U \subset \{y : \overline{\lambda}(y) > c\}$$

proving the result. □

## 4 Proof of Theorem 5 and associated results

**Proof of Theorem 5.** First, the lower bound for open sets is contained in Lemma 11. Second, by **ldp**$\mu$ and **jnt-lsc**, $\lambda_\theta(x) + \psi(\theta)$ is lower semicontinuous. To prove (a) it remains to consider the upper bound for compact sets.
Fix $F \subset \Theta \times X$, compact. By lower semicontinuity and regularity, for each $(\theta, x)$ there are open sets $O \subset \Theta$ and $U \subset X$ containing $\theta$ and $x$ respectively, with closures $\overline{O}$ and $\overline{U}$, such that

$$\lambda_\vartheta(y) > \lambda_\theta^\epsilon(x) = \min\{\lambda_\theta(x) - \epsilon, 1/\epsilon\} \text{ for } (\vartheta, y) \in \overline{O} \times \overline{U}$$

and

$$\psi(\vartheta) > \psi^\epsilon(\theta) \text{ for } \vartheta \in \overline{O}.$$

By taking $O$ to be smaller, if necessary, there is an integer $N$ such that for $n \ge N$ and $\gamma \in O \cap \Theta_n$

$$P_\gamma^n(\overline{U}) \le \exp(-n\lambda^\epsilon(x, \theta)) \text{ and } \mu^n(\overline{O}) \le \exp(-n\psi^\epsilon(\theta)).$$

Thus, for $n \ge N$,

$$\begin{aligned} \widetilde{P}^n(O \times U) &\le \int_{\overline{O} \cap \Theta_n} P_\vartheta^n(\overline{U}) d\mu^n(\vartheta) \\ &\le \exp(-n\lambda_\theta^\epsilon(x))\mu^n(\overline{O}) \\ &\le \exp(-n\lambda_\theta^\epsilon(x)) \exp(-n\psi^\epsilon(\theta)). \end{aligned}$$

Hence

$$\limsup_n \frac{\log \widetilde{P}^n(U \times O)}{n} \leq -\lambda_\theta^\epsilon(x) - \psi^\epsilon(\theta).$$

As $(x, \theta)$ varies over $F$ the corresponding sets $U \times O$ cover $F$. Taking a finite subcover, using it to get an upper bound on $\widetilde{P}(F)$ and then letting $\epsilon$ go to zero completes the proof. In the locally compact cases, $O$ and $U$ can be taken so that $\overline{U}$ and $\overline{O}$ are compact and so a weak LDP is enough to bound the corresponding terms. This completes the proof of (a).
Part (b) follows immediately from Lemma 1.2.18 in Dembo and Zeitouni (1993), which gives the LDP for $(\widetilde{P}^n)$, and the contraction principle (given in Theorem 4.2.1 of Dembo and Zeitouni (1993)) applied to the projection from $\Theta \times X$ to $X$, which gives the LDP for $(P^n)$. □

**Proof of Proposition 6**. Fix $\alpha$. Using **tight**$\mu$, let $O$ be such that

$$\limsup \frac{\log \mu^n\{O\}}{n} < -\alpha$$

and let $S$ be the (compact) complement of $O$. For $\theta \in \widetilde{\Theta}$, let $U_\theta \subset X$ be a set with compact complement such that for any $\theta(n) \in \Theta_n$ with $\theta(n) \to \theta$

$$\limsup \frac{\log P^n_{\theta(n)}\{U_\theta\}}{n} < -\alpha.$$

The existence of $U_\theta$ is guaranteed by **uni-tight**. For $\theta \notin \widetilde{\Theta}$, let $U_\theta = X$. Then there is an open set $V_\theta$, containing $\theta$, and an integer $N_\theta$ such that for $n \geq N_\theta$ and $\gamma \in V_\theta \cap \Theta_n$

$$P^n_\gamma(U_\theta) < \exp(-n\alpha).$$

Otherwise a suitable subsequence contradicts **uni-tight**.
The collection $\{V_\theta : \theta \in S\}$ covers $S$. Take a finite cover $(V_{\theta(i)} : 1 \leq i \leq k)$ of $S$; then let $U$ be the set $\cap_i U_{\theta(i)}$ and $K$ be its complement, which, as the union of a finite number of compact sets is itself compact. Then $(O \times X) \cup (S \times U)$ has the complement $S \times K$, which is compact, and, for $n$ large enough

$$\begin{aligned} \widetilde{P}^n\left((O \times X) \cup (S \times U)\right) &\leq \mu^n(O) + \sum_{i=1}^k \int_{V_{\theta(i)} \cap \Theta_n} P_\vartheta(U_{\theta(i)}) d\mu^n(\vartheta) \\ &< (k+1)\exp(-n\alpha). \end{aligned}$$

Hence

$$\limsup \frac{\log \widetilde{P}^n(O \times X) \cup (S \times U))}{n} \leq -\alpha,$$

which suffices, since $\alpha$ was arbitrary. □

**Lemma 16** *Suppose $X$ is locally compact and* **exp-cty** *holds. Then* **uni-tight** *holds when* **good**$\lambda_\theta$ *holds.*

**Proof**. Locally compact means that for every $x \in X$ there is $U_x$ open and $C_x$ compact with $x \in U_x \subset C_x$. Fix $\theta \in \widetilde{\Theta}$. Take $\beta < \alpha < \infty$. Since $\lambda_\theta$ is good,

$$K = \{x : \lambda_\theta(x) \leq \alpha\}$$

is compact. Let $\{U_{x(i)} : i = 1, 2, \ldots, k\}$ be a finite subcover of $K$ taken from $\{U_x : x \in K\}$. Now let $O$ be the complement of the compact set $\cup_i C_{x(i)}$, and let $F$ be the complement of the open set $\cup_i U_{x(i)}$. Then $F \cap K = \emptyset$.
Consider $\{P^n_{\theta(n)}\}$ where $\theta(n) \in \Theta_n$, and $\theta(n) \to \theta$. Then, by **exp-cty**,

$$\limsup \frac{\log P^n_{\theta(n)}(O)}{n} \leq \limsup \frac{\log P^n_{\theta(n)}(F)}{n} \leq -\inf_{y \in F} \lambda_\theta(y) \leq -\alpha < -\beta.$$

Since the set $O$ is independent of the particular sequence $(\theta(n))$ the result is proved. □

**Proof of Proposition 7**. This is an immediate consequence of Lemma 16. □

**Proof of Proposition 8**. The argument is borrowed from the last part of the proof of Theorem 2.3 in Chaganty (1997). Fix $\alpha$. Using **tight**$\mu$, let $O \subset \Theta$, with a compact complement $S$, be such that

$$\limsup \frac{\log \mu^n\{O\}}{n} < -\alpha/2.$$

By Theorem 2, $(P^n)$ satisfies an LDP with the good rate function $\lambda$. Then, by Lemma 3, $(P^n)$ is exponentially tight and so there is an open set $U \subset X$ with a compact complement $K$ such that

$$\limsup \frac{\log P^n\{U\}}{n} < -\alpha/2.$$

Then $(O \times X) \cup (S \times U)$ has the complement $S \times K$, which is compact, and

$$\limsup \frac{\log \widetilde{P}^n\{(O \times X) \cup (S \times U)\}}{n} < -\alpha/2 - \alpha/2 = -\alpha.$$

□

# 5 Possible extensions and refinements

This is a brief note of things that have not been attempted but seem to have some interest.
Clearly, it would be desirable to have some variant of Lemma 16 for Polish spaces. However, the proof that a good rate function implies exponential tightness in a Polish space seems to work only for a given sequence — see Lemma 2.6 in Lynch and Sethuraman (1987). Hence, it does not produce the uniformity needed in **uni-tight**.
This note aims to generalise Theorem 2.3 in Dinwoodie and Zabell (1992). In that Theorem, the mixing LDP, **ldp**$\mu$, and the associated exponential tightness, **tight**$\mu$, hold automatically, while exponential continuity, **exp-cty**, and joint lower semicontinuity of $\lambda_\theta(x)$, **jnt-lsc**, are taken as hypotheses. In a further study, Dinwoodie and Zabell (1993), they give results that relax these assumptions and also their assumption that $\Theta$ is compact, which is, in a sense, analogous to **tight**$\mu$ here. Their ideas could be taken up in this context.
As hinted in the introduction, the possibility that $X$ is a space of measures with the $\tau$-topology raises the question of whether the arguments can be adapted to allow $P^n_\theta$ to be defined on a smaller $\sigma$-algebra than the Borel one.
Finally, the approach to large deviations described in Puhalskii (2001) could be explored. Theorem 1.8.9 and Lemma 1.8.12 there are relevant. Roughly translated into the language here, they give conditions on $\psi(\theta)$ and $\lambda_\theta(x)$ which make $\lambda_\theta(x) + \psi(\theta)$ a rate function on $\Theta \times X$.

## Acknowledgements

David Penman brought Dinwoodie and Zabell (1992) to my attention, worked with me on the motivating problem, reported in Biggins and Penman (2003), and commented on the manuscript. Sandy Zabell looked at an earlier version and brought Chaganty (1997) to my attention. The referee pointed out the desirability of allowing $P_\theta^n$ to be defined on a smaller $\sigma$-algebra than the Borel one.

## References

[1] Biggins, J.D., Penman, D. B. Large deviations for randomly coloured random graphs. (2003) Preprint available at `http://www.shef.ac.uk/~st1jdb/ldircrg.html`

[2] Cannings, C., Penman, D. B. Models of random graphs and their applications *Handbook of Statistics 21. Stochastic Processes: Modeling and Simulation.* Eds: D.N. Shanbhag, and C.R. Rao. Elsevier (2003) 51–91.

[3] Chaganty, N. R. Large deviations for joint distributions and statistical applications. *Sankhyā. Series A* **59** (1997) 147–166.

[4] Dembo, A., Zeitouni, O. *Large deviations techniques and applications*, (1st Edition), Jones and Bartlett, Boston (1993).

[5] Dinwoodie, I. H., Zabell, S. L. Large deviations for exchangeable random vectors. *Ann. Prob.* **20** (1992) 1147–1166.

[6] Dinwoodie, I. H., Zabell, S. L. Large deviations for sequences of mixtures. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift*, eds: J.K. Ghosh, S.K. Mitra, K.R. Parthasarathy and B.L.S. Prakasa Rao. Wiley Eastern, New Delhi, (1993) 171–190.

[7] Lynch, J., Sethuraman, J. Large deviations for processes with independent increments. *Ann. Prob.* **15** (1987) 610–627.

[8] Puhalskii, A. *Large Deviations and Idempotent Probability.* Chapman & Hall/CRC, Boca Raton, 2001.