VARIATIONAL REPRESENTATIONS OF VARADHAN FUNCTIONALS

HAROLD BELL AND WŁODZIMIERZ BRYC

Abstract. Motivated by the theory of large deviations, we introduce a class of non-negative non-linear functionals that have a variational “rate function” representation.

1. Introduction

Let \((X, d)\) be a Polish space with metric \(d()\) and let \(C_b(X)\) denote the space of all bounded continuous functions \(F : X \to \mathbb{R}\). In his work on large deviations of probability measures \(\mu_n\), Varadhan \[12\] introduced a class of non-linear functionals \(L\) defined by

\[
L(F) = \lim_{n \to \infty} \frac{1}{n} \log \int_X \exp(nF(x))d\mu_n
\]

and used the large deviations principle of \(\mu_n\) to prove the variational representation

\[
L(F) = L_0 + \sup_{x \in X} \{F(x) - I(x)\},
\]

where \(I : X \to [0, \infty]\) is the rate function governing the large deviations, and \(L_0 := L(0) = 0\).

Several authors \[1, 3, 4, 9, 10, 11\] abstracted non-probabilistic components from the theory of large deviations. In particular, in \[3\], see also \[10, Theorem 3.1\] we give conditions which imply the rate function representation \((2)\) when the limit \((1)\) exists, and we show that the rate function is determined from the dual formula

\[
I(x) = L(0) + \sup_{F \in C_b(X)} \{F(x) - L(F)\}.
\]

In fact, one can reverse Varadhan’s approach, and show that large deviations of probability measures \(\mu_n\) follow from the variational representation \((4)\) for \((1)\), see \[8, Theorem 1.2.3\]. In this context we have \(\mu_n(X) = 1\) which implies \(L(0) = 0\) in \((3)\) and correspondingly \(L_0 = 0\) in \((2)\).

“Asymptotic values” in \((3)\) are essentially what we call Varadhan Functionals here; the theorems in that paper are not entirely satisfying because the assumptions are in terms of the underlying probability measures. In this paper we present a more satisfying approach which relies on the theory of probability for motivation purposes only.

Received by the editors June 2, 1999. Revised: November 1, 1999.

1991 Mathematics Subject Classification. 60F10.

Key words and phrases. large deviation, Čech-Stone compactification, Varadhan functionals, rate functions.
Definition 1.1. A function $\mathbb{L} : C_b(\mathbb{X}) \to \mathbb{R}$ is a Varadhan Functional if the following conditions are satisfied.

(4) If $F \leq G$ then $\mathbb{L}(F) \leq \mathbb{L}(G)$ for all $F, G \in C_b(\mathbb{X})$

(5) $\mathbb{L}(F + \text{const}) = \mathbb{L}(F) + \text{const}$ for all $F \in C_b(\mathbb{X}), \text{const} \in \mathbb{R}$

Expression (1) provides an example of Varadhan Functional, if the limit exists. Another example is given by variational representation (2).

Definition 1.2. A Varadhan Functional $\mathbb{L}$ is maximal if $\mathbb{L}(\cdot)$ is a lattice homomorphism

(6) $\mathbb{L}(F \lor G) = \mathbb{L}(F) \lor \mathbb{L}(G)$.

It is easy to see that each Varadhan Functional $\mathbb{L}(\cdot)$ satisfies the Lipschitz condition $|\mathbb{L}(F) - \mathbb{L}(G)| \leq \|F - G\|_\infty$, compare (1). Thus $\mathbb{L}$ is a continuous mapping from the Banach space $C_b(\mathbb{X})$ of all bounded continuous functions into the real line. We will need the following stronger continuity assumption, motivated by the definition of the countable additivity of measures.

Definition 1.3. A Varadhan Functional is $\sigma$-continuous if the following condition is satisfied.

(7) If $F_n \downarrow 0$ then $\mathbb{L}(F_n) \to \mathbb{L}(0)$.

Notice that if $\mathbb{X}$ is compact, then by Dini’s theorem and the Lipschitz property, all Varadhan Functionals are $\sigma$-continuous.

Maximal Varadhan Functionals are convex; this follows from the proof of Theorem 2.1, which shows that formula (2) holds true for all Varadhan Functionals when the supremum is extended to all $\mathbb{x}$ in the Čech-Stone compactification of $\mathbb{X}$.

A simple example of convex and maximal but not $\sigma$-continuous Varadhan Functional is $\mathbb{L}(F) = \limsup_{x \to \infty} F(x)$, where $F \in C_b(\mathbb{R})$. This Varadhan Functional cannot be represented by variational formula (2). Indeed, (2) implies that $\mathbb{L}(x) \geq F(x) - \mathbb{L}(F) = F(x)$ for all $F \in C_b(\mathbb{R})$ that vanish at $\infty$; hence $\mathbb{L}(x) = \infty$ for all $x \in \mathbb{R}$ and $\mathbb{L}$ gives $\mathbb{L}(F) = -\infty$ for all $F \in C_b(\mathbb{R})$, a contradiction.

An example of a convex and $\sigma$-continuous but not maximal Varadhan Functional is $\mathbb{L}(F) = \log \int_{\mathbb{X}} \exp F(x) \nu(dx)$, where $\nu$ is a finite non-negative measure.

2. Variational Representations

The main result of this paper is the following.

Theorem 2.1. If a maximal Varadhan Functional $\mathbb{L} : C_b(\mathbb{X}) \to \mathbb{R}$ is $\sigma$-continuous, then there is $L_0 \in \mathbb{R}$ such that variational representation (2) holds true and the rate function $\mathbb{I} : \mathbb{X} \to [0, \infty]$ is given by the dual formula (4). Furthermore, $\mathbb{I}(\cdot)$ is a tight rate function: sets $\mathbb{I}^{-1}([0, a]) \subset \mathbb{X}$ are compact for all $a > 0$.

The next result is closely related to Bryc [3, Theorem T.1.1] and Deuschel & Stroock [4, Theorem 5.1.6]. Denote by $\mathcal{P}(\mathbb{X})$ the metric space (with Prokhorov metric) of all probability measures on a Polish space $\mathbb{X}$ with the Borel $\sigma$-field generated by all open sets.
Theorem 2.2. If a convex Varadhan Functional $L : C_b(X) \to \mathbb{R}$ is $\sigma$-continuous, then there is a lower semicontinuous function $J : \mathcal{P}(X) \to [0, \infty]$ and a constant $L_0$ such that such that

$$(8) \quad L(F) = L_0 + \sup_{\mu \in \mathcal{P}} \left\{ \int F \, d\mu - J(\mu) \right\}$$

for all bounded continuous functions $F$.

A well known example in large deviations is the convex $\sigma$-continuous functional $L(F) := \log \int \exp F(x) \nu(dx)$ with the rate function in (8) given by the relative entropy functional

$J(\mu) = \begin{cases} \int \log \frac{d\mu}{d\nu} \, d\mu & \text{if } \mu \ll \nu \text{ is absolutely continuous} \\ \infty & \text{otherwise}. \end{cases}$

Remark 2.1. Deuschel & Stroock [6, Section 5.1] consider convex functionals $\Phi : C_b(X) \to \mathbb{R}$ such that $\Phi(const) = const$. Such functionals satisfy condition (5).

Indeed, write $F + const$ as a convex combination $F + const = (1 - \theta)F + \theta \left\{ (2 const)^{2} - \Phi(F) \right\}$, where $0 < \theta < 1$. Using convexity and $\Phi(const) = const$ we get $\Phi(F + const) \leq \Phi(F) + const + \theta \left( \Phi(2F) - \Phi(F) \right)$. Since $\theta > 0$ is arbitrary this proves that $\Phi(F + const) \leq \Phi(F) + const$. By routine symmetry considerations (replacing $F \mapsto F - const$, and then $const \mapsto -const$), (5) follows.

3. Proofs

Let $L_0 := L(0)$. Passing to $L'(F) := L(F) - L_0$ if necessary, without losing generality we assume $L(0) = 0$.

Lemma 3.1. Let $\hat{X}$ be a compact Hausdorff space. Suppose $X \subset \hat{X}$ is a separable metric space in the relative topology. If $x_0 \in \hat{X} \setminus X$ then there are bounded continuous functions $F_n : \hat{X} \to \mathbb{R}$ such that

(i) $F_n(x) \downarrow 0$ for all $x \in X$.
(ii) $F_n(x_0) = 1$ for all $n \in \mathbb{N}$.

Proof. Since $\hat{X}$ is Hausdorff, for every $x \in X$ there is an open set $U_x \ni x$ such that its closure $\overline{U}_x$ does not contain $x_0$.

By Lindelöf property for separable metric space $X$, there is a countable subcover $\{U_n\}$ of $\{U_x\}$.

A compact Hausdorff space $\hat{X}$ is normal. So there are continuous functions $\phi_n : \hat{X} \to \mathbb{R}$ such that $\phi_n|_{\overline{U}_n} = 0$ and $\phi_n(x_0) = 1$.

To end the proof take $F_n(x) = \min_{1 \leq k \leq n} \phi_k(x)$.

The following lemma is contained implicitly in [3, Theorem T.1.2].

Lemma 3.2. Theorem 2.2 holds true for compact $X$.

Proof. Let $I(\cdot)$ be defined by (3). Thus $I(x) \geq F(x) - L(F)$ which implies $L(F) \geq \sup_{x \in X} \{ F(x) - I(x) \}$. To end the proof we need therefore to establish the converse inequality. Fix a bounded continuous function $F \in C_b(X)$ and $\epsilon > 0$. Let $s = \ldots$
sup_{x \in X} \{ F(x) - I(x) \}. Clearly F(x) - I(x) \leq s \leq L(F). By (3) again, for every \( x \in X \), there is \( F_n \in C_b(X) \) such that \( I(x) < F_n(x) - L(F_n) + \epsilon \). Therefore
\[
F(x) \leq s + I(x) < s + \epsilon + F_n(x) - L(F_n)
\]
This means that the sets \( U_n = \{ y \in X : F(y) - F_n(y) < s + \epsilon - L(F_n) \} \) form an open covering of \( X \). Using compactness of \( X \), we choose a finite covering \( U_{n(1)}, \ldots, U_{n(k)} \). Then, writing \( F_i = F_{n(i)} \) we have
\[
F(x) < \max_{1 \leq i \leq k} \{ F_i(x) - L(F_i) \} + s + \epsilon
\]
for all \( x \in X \).

Using (1), (3), and (4) we have
\[
L(F) \leq L \left( \max_{1 \leq i \leq k} \{ F_i - L(F_i) \} + s + \epsilon \right) = L \left( \max_i \{ F_i - L(F_i) \} \right) + s + \epsilon = \max_i \{ L(F_i - L(F_i)) \} + s + \epsilon
\]
Since (4) implies \( L(F_i - L(F_i)) = L(F_i) - L(F_i) = 0 \) this shows that \( s \leq L(F) < s + \epsilon \). Therefore \( L(F) = s \), proving (2).

**Proof of Theorem 2.1.** Let \( X \) be the \( \check{C}ech-Stone \) compactification of \( X \). Since the inclusion \( X \subseteq \hat{X} \) is continuous, we define \( \check{L} : C_b(\hat{X}) \to \mathbb{R} \) by \( \check{L}(\hat{F}) := L(\hat{F}|X) \).

It is clear that \( \check{L} \) is a maximal Varadhan Functional, so by Lemma 3.2 there is \( \hat{L} : \hat{X} \to [0, \infty] \) such that \( \check{L}(\hat{F}) = \sup \{ F(x) - I(x) : x \in X \} \).

Using \( \sigma \)-continuity (5) it is easy to check that \( \check{L}(x) = \infty \) for all \( x \in \hat{X} \setminus X \). Indeed, given \( x_0 \in \hat{X} \setminus X \) by Lemma 3.1 there are \( F_n \in C_b(X) \) such that \( F_n \searrow 0 \) on \( X \), but \( F_n(x_0) = C > 0 \). From (5) we get \( \check{L}(x_0) \geq \check{L}(0) + F_n(x_0) - \check{L}(F_n) \to \check{L}(0) + C \).

Since \( C > 0 \) is arbitrary, \( \check{L}(x_0) = \infty \).

This shows that \( \check{L}(\hat{F}) = \sup \{ \hat{F}(x) - I(x) : x \in X \} \) for all \( \hat{F} \in C_b(\hat{X}) \). It remains to observe that since \( \hat{X} \) is a \( \check{C}ech-Stone \) compactification, every function \( F \in C_b(X) \) is a restriction to \( X \) of some \( \hat{F} \in C_b(\hat{X}) \), see [7, IV.6.22]. Therefore (4) holds true for all \( F \in C_b(X) \).

To prove that the rate function is tight, suppose that there is \( a > 0 \) such that \( \check{L}^{-1}[0, a] \) is not compact. Then there is \( \delta > 0 \) and a sequence \( x_n \in X \) such that \( d(x_m, x_n) > \delta \) for all \( m \neq n \). Since Polish spaces have Lindelöf property, there is a countable number of open balls of radius \( \delta/2 \) which cover \( X \). For \( k = 1, 2, \ldots, \), denote by \( B_k \ni x_k \) one of the balls that contain \( x_k \), and let \( \phi_k \) be a bounded continuous function such that \( \phi_k(x_k) = 2a \) and \( \phi_k = 0 \) on the complement of \( B_k \).

Then \( F_n = \max_{k \geq n} \phi_k \searrow 0 \) pointwise. On the other hand (4) implies \( L(F_n) \geq L_0 + F_n(x_n) - \check{L}(x_n) \geq L_0 + a \), contradicting (4).

**Lemma 3.3.** If \( L(\cdot) \) is a Varadhan Functional then
\[
\inf_{x \in X} \{ F(x) - G(x) \} \leq L(F) - L(G)
\]

**Proof.** Let \( \text{const} = \inf_{x \in X} \{ F(x) - G(x) \} \). Clearly, \( F \geq G + \text{const} \). By positivity condition (4) this implies \( L(F) \geq L(G + \text{const}) = L(G) + \text{const} \).

The next lemma is implicitly contained in the proof of (4, Theorem T.1.1). Let \( \mathcal{P}_a(X) \) denote all regular finitely-additive probability measures on \( X \) with the Borel field.
Lemma 3.4. If \( L(\cdot) \) is a convex Varadhan Functional on \( \mathcal{C}_b(X) \), then there exist a lower semicontinuous function \( \mathcal{J} : \mathcal{P}_a(X) \to [0, \infty] \) such that

\[
L(F) = L(0) + \inf_{\mu} \{ \mu(F) - \mathcal{J}(\mu) : \mu \in \mathcal{P}_a(X) \},
\]

and the supremum is attained.

Proof. Let \( \mathcal{J}(\cdot) \) be defined by

\[
\mathcal{J}(\mu) = L(0) + \inf_{\mu} \{ \mu(F) - L(F) : F \in \mathcal{C}_b(X) \}.
\]

and fix \( F_0 \in \mathcal{C}_b(X) \). Recall that throughout this proof we assume \( L(0) = 0 \).

By the definition of \( \mathcal{J}(\cdot) \), we need to show that

\[
L(F_0) = \sup_{\mu} \inf_{F} \{ \mu(F_0) - \mu(F) + L(F) \},
\]

where the supremum is taken over all \( \mu \in \mathcal{P}_a(X) \) and the infimum is taken over all \( F \in \mathcal{C}_b(X) \). Moreover, since \( \mathcal{J}(\mu) \geq \mu(F_0) - L(F_0) \) for all \( \mu \in \mathcal{P}_a(X) \), therefore \( L(F_0) \geq \sup_{\mu} \inf_{F} \{ \mu(F_0) - \mu(F) + L(F) \} \). Hence to prove \( \mathcal{J}(\cdot) \), it remains to show that there is \( \nu \in \mathcal{P}_a(X) \) such that

\[
L(F_0) \leq \nu(F_0) - \nu(F) + L(F) \quad \text{for all} \quad F \in \mathcal{C}_b(X).
\]

(13) (also, for this \( \nu \), the supremum in (12) will be attained) To find \( \nu \), consider the following sets. Let

\[
\mathcal{M} = \{ F \in \mathcal{C}_b(X) : \inf_{x} \{ F(x) - F_0(x) \} > 0 \}
\]

and let \( \mathcal{N} \) be a set of all finite convex combinations of functions \( g(x) \) of the form \( g(x) = F(x) + L(F_0) - L(F) \), where \( F \in \mathcal{C}_b(X) \).

It is easily seen from the definitions that \( \mathcal{M} \) and \( \mathcal{N} \) are convex; also \( \mathcal{M} \subset \mathcal{C}_b(X) \) is non-empty since \( 1 + F_0 \in \mathcal{M} \), and open since \( \{ F : \inf_{x} \{ F(x) - F_0(x) \} \leq 0 \} \subset \mathcal{C}_b(X) \) is closed. Furthermore, \( \mathcal{M} \) and \( \mathcal{N} \) are disjoint. Indeed, take arbitrary

\[
\mathcal{N} \ni g = \sum \alpha_k F_k + \inf_{x} \{ F_0(x) \} - \sum \alpha_k L(F_k).
\]

Then

\[
\inf_{x} \{ g(x) - F_0(x) \} = \inf \{ \sum \alpha_k F_k(x) - F_0(x) \} - \sum \alpha_k L(F_k) + \inf \{ \sum \alpha_k F_k(x) - F_0(x) \} - \inf \{ \sum \alpha_k F_k(x) - F_0(x) \} + L(F_0) \leq 0,
\]

where the first inequality follows from the convexity of \( L(\cdot) \) and the second one follows from \( (9) \) applied to \( F = \sum \alpha_k F_k(x) \) and \( G = F_0 \).

Therefore \( \mathcal{M} \) and \( \mathcal{N} \) can be separated, i.e. there is a non-zero linear functional \( f^* \in \mathcal{C}_b^*(X) \) such that for some \( \alpha \in \mathbb{R} \)

\[
f^*(\mathcal{N}) \leq \alpha < f^*(\mathcal{M}),
\]

see e.g. \[13\], V. 2.8]

Claim: \( f^* \) is non-negative.

Indeed, it is easily seen that \( F_0(\cdot) \) belongs to \( \mathcal{N} \), and, as a limit of \( \epsilon + F_0(\cdot) \) as \( \epsilon \to 0 \), \( F_0 \) is also in the closure of \( \mathcal{M} \). Therefore by \( (14) \) we have \( \alpha = f^*(F_0) \). To end the proof take arbitrary \( F \) with \( \inf_{x} F(x) > 0 \). Then \( F + F_0 \in \mathcal{M} \) and by \( (14) \)

\[
f^*(F) = f^*(F + F_0) - f^*(F_0) > \alpha - f^*(F_0) = 0.
\]

This ends the proof of the claim.
Without losing generality, we may assume $f^*(1) = 1$; then it is well known, see e.g. [2, Ch. 2 Section 4 Theorem 1], that $f^*(F) = \nu(F)$ for some $\nu \in \mathcal{P}_a(X)$; for regularity of $\nu$ consult [13, IV.6.2]. It remains to check that $\nu$ satisfies (13). To this end observe that since $F + L(F_0) - L(F) \in \mathcal{N}$, by (14) we have $\nu(F) + L(F_0) - L(F) \leq \alpha = \nu(F_0)$ for all $F \in \mathcal{C}_b(X)$. This ends the proof of (10).

Proof of Theorem 2.2. Lemma 3.4 gives the variational representation (10) with the supremum taken over a too large set. To end the proof we will show that $\mathcal{J}(\mu) = \infty$ on measures $\mu$ that fail to be countably-additive.

Suppose that $\mu$ is additive but not countably additive. Then Daniell-Stone theorem implies that there is $\delta > 0$ and a sequence $F_n \searrow 0$ of bounded continuous functions such that $\int F_n d\mu > \delta > 0$ for all $n$. By (11) and $\sigma$-continuity $\mathcal{J}(\mu) \geq L(0) + C \int F_n d\mu - L(CF_n) \geq L(0) + C\delta - L(CF_n) \to L(0) + C\delta$. Since $C > 0$ is arbitrary, therefore $\mathcal{J}(\mu) = \infty$ for all $\mu$ that are additive but not countably-additive. Thus (10) implies (8). □

References

[1] M. Akian, Densities of idempotent measures and large deviations. Trans. Amer. Math. Soc. 351 (1999), 4515–4543.
[2] H. Bergström, Weak convergence of measures. Acad. Press, New York, 1982.
[3] W. Bryc On the large deviation principle by the asymptotic value method. In: Diffusion Processes and Related Problems in Analysis, Vol. I, ed. M. Pinsky, Birkhäuser, Boston 1990, 447–472.
[4] A. de Acosta, Upper bounds for Large Deviations of Dependent Random Vectors. Zeitsch. Wahrscheinlichk. Theor. Verw. Gebiete 69 (1985), 551–565.
[5] A. Dembo & O. Zeitouni, Large Deviations Techniques and Applications. Jones and Bartlett, Boston, 1993.
[6] J-D. Deuschel & D. W. Stroock, Large Deviations. Pure and Applied Math vol. 137, Academic Press, Boston, 1989.
[7] N. Dunford & J. T. Schwartz, Linear Operators I. Interscience, New York, 1958.
[8] P. Dupuis & R. S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations. Wiley, New York, 1997.
[9] G. L. O’Brien, Sequences of capacities, with connections to large-deviation theory. J. Theoret. Probab. 9 (1996), 19–35.
[10] G. O’Brien & W. Vervaat, Compactness in the theory of large deviations. Stoch. Processes Appl. 57 (1995), 1–10.
[11] A. Puhalskii, Large deviations of Semimartingales: a Maxingale Problem Approach I. Stochastics 61 (1997), 141–243.
[12] S. R. S. Varadhan, Asymptotic probabilities and differential equations. Comm. Pure Appl. Math. 19 (1966), 261–286.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CINCINNATI, PO BOX 210025, CINCINNATI, OH 45221–0025.
E-mail address: bellh@math.uc.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CINCINNATI, PO BOX 210025, CINCINNATI, OH 45221–0025
E-mail address: brycwz@email.uc.edu

URL: http://http://ucms02.csm.uc.edu/preprint/ld-abs