SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 3, pp. 18–33

Robust Bounds on Risk-Sensitive Functionals via Rényi Divergence*

Rami Atar†, Kenny Chowdhary‡, and Paul Dupuis§

Abstract. We extend the duality between exponential integrals and relative entropy to a variational formula for exponential integrals involving the Rényi divergence. This formula characterizes the dependence of risk-sensitive functionals to perturbations in the underlying distribution. It also shows that perturbations of related quantities determined by tail behavior, such as probabilities of rare events, can be bounded in terms of the Rényi divergence. The characterization gives rise to tight upper and lower bounds that are meaningful for all values of a large deviation scaling parameter, allowing one to quantify in explicit terms the robustness of risk-sensitive costs. As applications we consider problems of uncertainty quantification when aspects of the model are not fully known, as well their use in bounding tail properties of an intractable model in terms of a tractable one.

Key words. Rényi divergence, risk-sensitive cost, rare events, large deviation, Laplace principle, robust bounds, risk-sensitive functional comparison bounds, logarithmic probability comparison bounds

AMS subject classifications. 60F10, 60E15, 94A17

DOI. 10.1137/130939730

1. Introduction.

For many models encountered in engineering, the physical sciences, mathematical finance, and elsewhere, rare events play a key role in determining important properties of the system. Given a system model, large deviation theory can often be used to study the impact of rare events, and in particular can provide both qualitative and quantitative information [14, 9, 8, 24]. Of course large deviation theory provides only an asymptotic approximation, and so if nonasymptotic bounds are sought, then one can appeal to other approximations such as Monte Carlo [2, 4, 12]. However, it is well known that the resulting estimates (both asymptotic and nonasymptotic) are sensitive to the underlying assumed distribution, since they are determined by tail properties of the distributions. As a consequence, understanding the impact of modeling errors and model uncertainty becomes especially important. Modeling uncertainty can take many forms. For example, for some parts of the system there may be justification for the use of distributions of a particular form, but with parameters that are not known precisely. For other parts of the system there may be no suitable probabilistic model, and one may instead assume only that parameters belong to some known set.

*Received by the editors October 3, 2013; accepted for publication (in revised form) November 3, 2014; published electronically January 13, 2015.
http://www.siam.org/journals/juq/3/93973.html
†Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (atar@ee.technion.ac.il). This author's research was supported in part by the ISF (grant 1315/12), the US-Israel BSF (grant 2008466), and the Technion fund for promotion of research.
‡Sandia National Laboratories, Livermore, CA 94550 (kennychowdhary@gmail.com).
§Division of Applied Mathematics, Brown University, Providence, RI 02912 (dupuis@dam.brown.edu). This author's research was supported in part by the Department of Energy (DE-SC0002413, DE-SC0010539), the National Science Foundation (DMS-1317199), and the Air Force Office of Scientific Research (FA9550-12-1-0399).

This paper is concerned with probabilities associated with rare events and expected values that are largely determined by rare events. However, the issues just raised regarding model uncertainty and modeling error are also important for ordinary (e.g., order one) probabilities and expected values that are not sensitive to rare events. For such problems, one can obtain tight bounds that hold for a well-defined family of "true" process models by computing certain functionals with respect to a given "nominal" model and then using the duality between exponential integrals and relative entropy. For a detailed discussion we refer the reader to [7]. Following standard terminology in the economics and control literature, we refer to integrals of the form $\int_S e^g d\nu$ as *risk-sensitive* functionals, where $(S, \mathcal{F})$ is a measurable space, $g : S \to \mathbb{R}$ is a measurable function, and ν is a probability measure. Let $\mathcal{P}(S)$ denote the set of probability measures on $(S, \mathcal{F})$. The well-known duality alluded to above is

$$\log \int_S e^g d\nu = \sup_{\theta \in \mathcal{P}(S)} \left[\int_S g d\theta - R(\theta \| \nu) \right], \tag{1.1}$$

where R denotes relative entropy, and g is a bounded function (see (2.2)) [extensions to unbounded g are also valid; see [10, Proposition 4.5.1]]. Based on this identity, the results of [7, 17] give tight bounds on ordinary probabilities and expected values, i.e., quantities of the form $\int_S g d\theta$. The bounds are in terms of a maximum relative entropy distance between the nominal model, ν, and a collection of models which presumably include the true model, plus a risk-sensitive cost with respect to the nominal model. Note that the feasibility of explicit computation, which means computing or approximating exponential integrals, is thus linked to the choice of the nominal model. Robust properties of controls designed on risk-sensitive criteria were first described in [11]. By considering suitable limits such criteria can be linked to other methods for handling model uncertainty, such as H^∞ control [27].

As it turns out, the duality (1.1) is not useful for bounding expectations and analyzing problems with rare events, because the natural scaling properties are such that the probabilities and expected values of interest should themselves be expressed as risk-sensitive functionals (this point will be made precise later on). However, there is a generalization of relative entropy called *Rényi relative entropy* or *Rényi divergence* (introduced in [22]; see section 2), with which risk-sensitive functionals can be expressed in terms of other risk-sensitive functionals. In particular, as we shall prove, the identities

$$\frac{1}{\beta} \log \int_S e^{\beta g} d\nu = \inf_{\theta \in \mathcal{P}(S)} \left[\frac{1}{\gamma} \log \int_S e^{\gamma g} d\theta + \frac{1}{\gamma - \beta} R_{\frac{\gamma}{\gamma-\beta}}(\nu \| \theta) \right] \tag{1.2}$$

and

$$\frac{1}{\gamma} \log \int_S e^{\gamma g} d\nu = \sup_{\theta \in \mathcal{P}(S)} \left[\frac{1}{\beta} \log \int_S e^{\beta g} d\theta - \frac{1}{\gamma - \beta} R_{\frac{\gamma}{\gamma-\beta}}(\theta \| \nu) \right] \tag{1.3}$$

hold for any $\beta, \gamma \in \mathbb{R} \setminus \{0\}$, $\beta < \gamma$, where for $\alpha \in \mathbb{R} \setminus \{0, 1\}$ R_α denotes Rényi divergence of order α, and g is a bounded function (see (2.1) and (2.3)). Moreover, (1.1) is a limit case of (1.3) as $\beta \to 0$, with $\gamma = 1$. Furthermore, as follows from these identities, for fixed probability measures θ and ν, any $0 < \beta < \gamma$, and any event A (a measurable subset of S), one has

$$\frac{1}{\beta}\log\theta(A) \leq \frac{1}{\gamma}\log\nu(A) + \frac{1}{\gamma-\beta}R_{\frac{\gamma}{\gamma-\beta}}(\theta\|\nu), \tag{1.4}$$

$$\frac{1}{\beta}\log\theta(A) \geq \frac{\gamma}{\beta^2}\log\nu(A) - \frac{\gamma}{\beta(\gamma-\beta)}R_{\frac{\gamma}{\gamma-\beta}}(\nu\|\theta). \tag{1.5}$$

Identities (1.2) and (1.3) make it possible to bound risk-sensitive functionals with respect to the true model, θ, in terms of a risk-sensitive functional with respect to the nominal model ν. Similarly, (1.4) and (1.5) provide bounds on probabilities under the true model in terms of probabilities under the nominal model. In this paper we also give elementary examples of how the identities (1.2)–(1.3) and the bounds (1.4)–(1.5) can be used.

As mentioned previously, one must evaluate a risk-sensitive functional with respect to a nominal model in order to turn the theoretical results into numerical bounds. This has implications and uses that go beyond assessing model uncertainty. In fact, it suggests an approach for bounding and approximating rare event probabilities when evaluation of this risk-sensitive functional is not possible or convenient for the known true model, by replacing it with the "closest" (in the sense of Rényi divergence) model for which the computation can be carried out and then bounding the Rényi divergence between the nominal and true models. Examples illustrating this use will be given. One can generalize to problems of minimizing risk-sensitive costs with respect to a controlled process and ask for robust bounds (i.e., bounds valid for a family of process models) in terms of the value function and optimal control for the nominal model. This would be analogous to the robust control of order one costs by using controls designed on the basis of risk-sensitive performance criteria [11], and it will be considered elsewhere.

We are aware of two other variational formulas for which the convex duality relation (1.1) is a special case. The first is a duality formula for ϕ-entropy ((2.60) in [19], (20) in [6]), which has played a central role in the study of concentration inequalities [19]. The other is a variational formula for the f-divergence (a notion similar to ϕ-entropy) that has been used to develop f-divergence estimators based on independent and identically distributed (i.i.d.) samples from each of two given distributions. Such estimators are significant in learning problems such as classification, dimensionality reduction, and homogeneity testing (see [20, 23] for the variational formula and its uses). Although Rényi divergence is closely related to f-divergence (in particular, the former is a certain nonlinear transformation of the latter; see [18, 25]), it seems that the representation formulas (1.2) and (1.3) cannot be recovered from these variational characterizations. The issue of robustness for rare events and risk-sensitive functionals has not received a great deal of attention. A paper that does consider the topic is [16], which considers the impact of varying the underlying distributions on the form of the large deviation rate function and related minimizers.

The rest of the paper is organized as follows. In section 2 we recall the definition and some properties of Rényi divergence, state the variational representations based on Rényi divergence, and note some immediate consequences. Section 3 contains elementary applications to functionals of empirical measures of i.i.d. outcomes, queueing, and Brownian motion with drift, and section 4 concludes with the proofs of the representation formulas.

2. Exponential integrals and Rényi divergence.

2.1. Definition and properties of Rényi divergence. Let $(S,\mathcal{F})$ be a measurable space, and recall that $\mathcal{P}(S)$ denotes the set of all probability measures on $(S,\mathcal{F})$. We say that a measure μ on $(S,\mathcal{F})$ dominates $\nu \in \mathcal{P}(S)$ if ν is absolutely continuous with respect to μ and denote this by $\nu \ll \mu$. For two probability measures $\nu,\theta \in \mathcal{P}(S)$, let $\nu' = \frac{d\nu}{d\mu}$ and $\theta' = \frac{d\theta}{d\mu}$ denote the Radon–Nikodym derivatives with respect to a dominating σ-finite measure μ. For $\alpha > 0$, $\alpha \neq 1$, the Rényi divergence of degree α of ν from θ is defined by (cf. [18])

$$R_\alpha(\nu\|\theta) \doteq \begin{cases} \infty & \text{if } \alpha > 1 \text{ and } \nu \not\ll \theta, \\ \dfrac{1}{\alpha(\alpha-1)} \log \displaystyle\int_{\{\nu'\theta'>0\}} \left(\frac{\nu'}{\theta'}\right)^\alpha d\theta & \text{otherwise.} \end{cases} \tag{2.1}$$

We follow [18] in defining R_α with the factor $\frac{1}{\alpha(\alpha-1)}$ rather than $\frac{1}{\alpha-1}$, which is also a common choice [3, 22, 25] (the paper [3] predates Rényi's work but considers only the special case of $\alpha = 1/2$ and for discrete measures). When ν and θ are mutually absolutely continuous, this expression can be written without reference to a dominating measure, namely

$$R_\alpha(\nu\|\theta) = \frac{1}{\alpha(\alpha-1)} \log \int_S \left(\frac{d\nu}{d\theta}\right)^\alpha d\theta = \frac{1}{\alpha(\alpha-1)} \log \int_S \left(\frac{d\theta}{d\nu}\right)^{1-\alpha} d\nu.$$

The definition of R_α is extended to $\alpha = 1$ by letting $R_1 = R$ be the relative entropy, or the Kullback–Liebler divergence, defined by

$$R(\nu\|\theta) \doteq \begin{cases} \infty & \text{if } \nu \not\ll \theta, \\ \displaystyle\int_{\{\nu'\theta'>0\}} \frac{\nu'}{\theta'} \log \frac{\nu'}{\theta'} d\theta & \text{otherwise.} \end{cases} \tag{2.2}$$

The definitions do not depend on the choice of the dominating measure, and since $\nu + \theta$ automatically dominates ν and θ, $R_\alpha(\nu\|\theta)$ is well defined for all pairs $(\nu,\theta) \in \mathcal{P}(S)^2$. For a proof of independence from the dominating measure as well as various properties of R_α, see [15, 18, 25, 26]. To mention a few of these properties, let ν and θ be fixed. Then $\alpha \mapsto \alpha R_\alpha(\nu\|\theta)$ is nondecreasing as a map from $(0,\infty)$ to $[0,\infty]$ and continuous from the left (thus $R = \lim_{\alpha\uparrow 1} R_\alpha$). If ν and θ are mutually singular, then $R_\alpha(\nu\|\theta)$ is infinite everywhere. Otherwise, it is finite and continuous on $(0,\bar{\alpha})$, where $\bar{\alpha} \doteq \sup\{\alpha : R_\alpha(\nu\|\theta) < \infty\} \geq 1$. Moreover, for every $\alpha > 0$, $R_\alpha(\nu\|\theta) = 0$ if and only if $\nu = \theta$.

A further useful property is the identity $R_\alpha(\nu\|\theta) = R_{1-\alpha}(\theta\|\nu)$, which holds for every $\alpha \in (0,1)$. We will use it to extend the definition of R_α to $\alpha \in \mathbb{R} \setminus \{0,1\}$. Namely, we set

$$R_\alpha(\nu\|\theta) \doteq R_{1-\alpha}(\theta\|\nu), \qquad \alpha < 0. \tag{2.3}$$

This definition is consistent with the definition of R_α, $\alpha \in \mathbb{R}$, given in (2.10) of [18], as follows from Remark 2.13 of [18].

2.2. Variational representations for exponential integrals. The variational representation for exponential integrals (1.1) is very closely related to the theory of large deviations and in fact can serve as the natural starting point for the large deviation analysis of any system [10]. It also gives an inequality that allows for robust bounds on ordinary costs with respect to a "true" measure in terms of risk-sensitive costs for a "nominal" model plus relative entropy distance between the two. However, as noted in the introduction, this variational representation does not seem to be useful when bounding risk-sensitive costs. The variational representations in Theorem 2.1 give useful bounds in that respect. Although a variational representation is not stated, some calculations related to a particular case of (2.4) (S a compact metric space) appear in [13] and suggest the statement of Theorem 2.1. However, the discussion in [13] is also restricted to mutually absolutely continuous measures, and because of this one cannot infer the variational formula or the general inequalities that follow. The proof of Theorem 2.1 is given in section 4.

Theorem 2.1. *Let β and γ be members of $\mathbb{R}\setminus\{0\}$, with $\beta<\gamma$. Let $\nu\in\mathcal{P}(S)$. Then for any bounded and measurable $g:S\to\mathbb{R}$, one has*

$$\frac{1}{\beta}\log\int_S e^{\beta g}d\nu=\inf_{\theta\in\mathcal{P}(S)}\left[\frac{1}{\gamma}\log\int_S e^{\gamma g}d\theta+\frac{1}{\gamma-\beta}R_{\frac{\gamma}{\gamma-\beta}}(\nu\|\theta)\right], \tag{2.4}$$

where the infimum is uniquely attained at $d\theta=e^{-(\gamma-\beta)g}d\nu/Z$, $Z=\int_S e^{-(\gamma-\beta)g}d\nu$. In addition,

$$\frac{1}{\gamma}\log\int_S e^{\gamma g}d\nu=\sup_{\theta\in\mathcal{P}(S)}\left[\frac{1}{\beta}\log\int_S e^{\beta g}d\theta-\frac{1}{\gamma-\beta}R_{\frac{\gamma}{\gamma-\beta}}(\theta\|\nu)\right], \tag{2.5}$$

where the supremum is uniquely attained at $d\theta=e^{(\gamma-\beta)g}d\nu/Z$, $Z=\int_S e^{(\gamma-\beta)g}d\nu$.

Remark 2.2. Setting $\beta=\alpha-1$ and $\gamma=\alpha$ gives

$$\frac{1}{\alpha-1}\log\int_S e^{(\alpha-1)g}d\nu=\inf_{\theta\in\mathcal{P}(S)}\left[\frac{1}{\alpha}\log\int_S e^{\alpha g}d\theta+R_\alpha(\nu\|\theta)\right] \tag{2.6}$$

and

$$\frac{1}{\alpha}\log\int_S e^{\alpha g}d\nu=\sup_{\theta\in\mathcal{P}(S)}\left[\frac{1}{\alpha-1}\log\int_S e^{(\alpha-1)g}d\theta-R_\alpha(\theta\|\nu)\right] \tag{2.7}$$

for all $\alpha\neq0$, $\alpha\neq1$. Although (2.6) and (2.7) are special cases of (2.4) and (2.5), the latter can be recovered from the former (in fact with α in the range $\alpha>0$, $\alpha\neq1$), as shown in the proof of Theorem 2.1.

Remark 2.3. By taking the formal limit $\alpha\to1$, we obtain from (2.6) the identity

$$\int_S g d\nu=\inf_{\theta\in\mathcal{P}(S)}\left[\log\int_S e^g d\theta+R(\nu\|\theta)\right]$$

and from (2.7) the well-known convex duality formula (see [7, 10, 11])

$$\log\int_S e^g d\nu=\sup_{\theta\in\mathcal{P}(S)}\left[\int_S g d\theta-R(\theta\|\nu)\right].$$

Note that one can also take $\alpha \to 0$ in (2.6) and (2.7), in which case $\alpha R_\alpha(\nu\|\theta) \to -\log\theta(\nu' > 0)$, recovering the simple fact

$$0 = \inf_{\theta\in\mathcal{P}(S)}[-\log\theta(\nu' > 0)] = \sup_{\theta\in\mathcal{P}} \log\nu(\theta' > 0).$$

The main purpose of this paper is to observe the following inequalities that follow from (2.6) and (2.7) and to discuss how they can be used to study robustness of risk-sensitive functionals. Note that g is not assumed bounded.

Corollary 2.4. *Fix $\theta \in \mathcal{P}(S)$, $\nu \in \mathcal{P}(S)$, and $0 < \beta < \gamma$.*

(i) *Let $g : S \to \mathbb{R}$ be a measurable function. Then*

$$\frac{1}{\beta}\log\int_S e^{\beta g}d\theta \le \frac{1}{\gamma}\log\int_S e^{\gamma g}d\nu + \frac{1}{\gamma-\beta}R_{\frac{\gamma}{\gamma-\beta}}(\theta\|\nu). \tag{2.8}$$

(ii) *Let $A \in \mathcal{F}$ be such that $\theta(A) > 0$ and $\nu(A) > 0$. Then*

$$\frac{1}{\beta}\log\theta(A) \le \frac{1}{\gamma}\log\nu(A) + \frac{1}{\gamma-\beta}R_{\frac{\gamma}{\gamma-\beta}}(\theta\|\nu) \tag{2.9}$$

and

$$\frac{1}{\beta}\log\theta(A) \ge \frac{\gamma}{\beta^2}\log\nu(A) - \frac{\gamma}{\beta(\gamma-\beta)}R_{\frac{\gamma}{\gamma-\beta}}(\nu\|\theta). \tag{2.10}$$

Moreover, given θ and A, (2.9) and (2.10) hold with equality for $\nu = \theta(\cdot|A)$, and given ν and A, (2.9) and (2.10) again hold with equality for $\theta = \nu(\cdot|A)$.

We will refer to (2.8) as the risk-sensitive functional comparison bounds (RSFCBs) and to (2.9) and (2.10) as the logarithmic probability comparison bounds (LPCBs). Note that in the special case where g is bounded, part (i) of the corollary is immediate from (2.4) and (2.5). Moreover, part (ii) follows by considering $g = 0$ on A and $g = -M$ on A^c and then sending $M \to \infty$. The remaining details of the proof appear in section 4.

Similar inequalities can be deduced for other relations between β and γ, but for our present purposes they do not seem to be particularly useful.

3. Elementary applications.

In this section we show how Corollary 2.4 can be used to provide robust bounds of the sort described in the introduction. The examples are intended only to illustrate the main ideas and limited to problems where the driving noises are distributed according to product measure. When assessing probabilities and expected values associated with rare events, it is important to keep in mind that it is usually *relative errors*, and not absolute errors, that are important. Also, it is generally the case that approximations are of an asymptotic nature as some scaling parameter tends to a limit. For light-tailed processes, the scaling is exponential in the parameter. As we will see, this fits in very nicely with the form of RSFCB (2.8).

As described in the introduction, one should have in mind two scenarios. In one case, we think of θ as a probability measure of interest for which the large deviation functional may be hard to compute and of ν as an alternative that is more tractable. In the other case, we are not sure of the model, with the nominal model ν a sort of "best guess" and θ the true model.

3.1. Functionals of the empirical measure. Suppose that $S = \mathbb{R}^n$, where n is a scaling parameter. Let θ^n and ν^n be product probability measures on S, with marginals θ_i^n and ν_i^n. Assume $\nu_i^n = \nu_1$, so that the nominal model corresponds to an i.i.d. sequence. Then (cf. [15])

$$\Delta_\alpha^n \doteq \frac{1}{n} R_\alpha(\theta^n \| \nu^n) = \frac{1}{n} \sum_{i=1}^n R_\alpha(\theta_i^n \| \nu_i^n) = \frac{1}{n} \sum_{i=1}^n R_\alpha(\theta_i^n \| \nu_1).$$

Let X_i, $i = 1, \ldots, n$, denote the canonical process. That is, for $\omega \in S$, $X_i(\omega) = \omega_i$ is the ith coordinate map. If the X_i are also i.i.d. under θ^n with marginal θ_1, then $\Delta_\alpha^n = R_\alpha(\theta_1 \| \nu_1)$ for every n. Consider the empirical measure $L_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ as a random element of the space $\mathcal{P}(\mathbb{R}) = \mathcal{P}(\mathbb{R}, \mathcal{R})$, equipped with the topology of weak convergence, and fix any measurable function $G : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$. Then with $\mathbb{E}_\theta$ and $\mathbb{E}_\nu$ denoting expectation with respect to the indicated distribution, we can take $g(X_n) = nG(L_n)$ in Corollary 2.4 to get, with $0 < \beta < \gamma$ and $\alpha = \gamma/(\gamma - \beta)$,

$$\frac{1}{n}\frac{1}{\beta} \log \mathbb{E}_\theta e^{\beta n G(L_n)} \leq \frac{1}{n}\frac{1}{\gamma} \log \mathbb{E}_\nu e^{\gamma n G(L_n)} + \frac{1}{\gamma - \beta}\frac{1}{n} \sum_{i=1}^n R_\alpha(\theta_i^n \| \nu_1) \tag{3.1}$$

(and also if desired a corresponding lower bound).

If G is continuous and θ corresponds to an i.i.d. sequence, then in this very simple setting one could use Sanov's theorem to evaluate the limit behavior of the two terms and obtain

$$\frac{1}{\beta} \sup_{\lambda \in \mathcal{P}(\mathbb{R})} [\beta G(\lambda) - R(\lambda \| \theta_1)] \leq \frac{1}{\gamma} \sup_{\lambda \in \mathcal{P}(\mathbb{R})} [\gamma G(\lambda) - R(\lambda \| \nu_1)] + \frac{1}{\gamma - \beta} R_\alpha(\theta_1 \| \nu_1). \tag{3.2}$$

The strength of the general inequalities based on Rényi divergence is that the bound (3.1) holds for all n and, moreover, does not require that θ correspond to an i.i.d. sequence.

We can make (3.1) and (3.2) more concrete by considering, for example, Gaussian distributions $\theta_1 = \mathcal{N}(\mu_1, \sigma_1^2)$ and $\nu_1 = \mathcal{N}(\mu_2, \sigma_2^2)$. In this case, denoting $\sigma_\alpha^2 \doteq \alpha\sigma_2^2 + (1 - \alpha)\sigma_1^2$,

$$R_\alpha(\theta_1 \| \nu_1) = \begin{cases} \frac{1}{\alpha} \log \frac{\sigma_2}{\sigma_1} + \frac{1}{2\alpha(\alpha - 1)} \log \frac{\sigma_2^2}{\sigma_\alpha^2} + \frac{1}{2} \frac{(\mu_1 - \mu_2)^2}{\sigma_\alpha^2} & \text{if } \sigma_\alpha^2 > 0, \\ \infty & \text{otherwise.} \end{cases} \tag{3.3}$$

If $G(L_n) = c\langle 1, L_n \rangle = n^{-1} c(X_1 + \cdots + X_n)$ for some constant c and $\nu_1 = \mathcal{N}(0, 1)$, then $\mathbb{E}_\nu e^{\gamma n G(L_n)} = \mathbb{E}_\nu e^{\gamma c(X_1 + \cdots + X_n)} = e^{\gamma^2 c^2 n/2}$ and (3.1) says that for every θ under which X_n are i.i.d. (but not necessarily Gaussian),

$$\frac{1}{n}\frac{1}{\beta} \log \mathbb{E}_\theta e^{\beta c(X_1 + \cdots + X_n)} \leq \frac{1}{\gamma - \beta} R_\alpha(\theta_1 \| \nu_1) + \frac{\gamma c^2}{2}. \tag{3.4}$$

In (3.4) one obtains equality if θ_1 is $\mathcal{N}(c(\gamma - \beta), 1)$, as can be verified using (3.3) and the relation $\alpha = \gamma/(\gamma - \beta)$. As a result, (3.4) is tight in the following sense. Fix β, γ, and α as above, and fix a constant $d > 0$. Consider the family of θ_1 for which $(\gamma - \beta)^{-1} R_\alpha(\theta_1 \| \nu_1) \leq d$. Then for all θ_1 in this family,

$$\frac{1}{n}\frac{1}{\beta} \log \mathbb{E}_\theta e^{\beta c(X_1 + \cdots + X_n)} \leq d + \frac{\gamma c^2}{2},$$

and one can find c and θ_1 in the family such that this display holds with equality. Indeed, c is chosen so that $\frac{1}{2}(\gamma - \beta)c^2 = d$ (namely, $c = \pm\sqrt{2d/(\gamma - \beta)}$) and $\theta_1 = \mathcal{N}(c(\gamma - \beta), 1)$.

3.2. A sample path large deviation example. We next discuss a well-known example from queueing analysis. Lindley's recursion

$$\begin{cases} Q_n = (Q_{n-1} + X_n - C)^+, & n \geq 1, \\ Q_0 = 0 \end{cases}$$

describes the queue length Q_n in an initially empty queueing system where $X_n \geq 0$ arrivals occur at time $n \geq 1$ and the server is capable of serving C customers at each time slot. Denoting $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$, the solution to this recursion is given by

$$Q_n = \max_{0 \leq i \leq n} [S_n - S_i - Ci].$$

Assume that the system is stable in the sense that $C > \mathbb{E}_\nu[X_1] = 1$. Consider the space-time rescaled processes $\bar{S}^n(t) = n^{-1} S_{[nt]}$ and $\bar{Q}^n(t) = n^{-1} Q_{[nt]}$, $t \geq 0$, and given a constant $b > 0$, let the *buffer overflow* event be given by

$$A_n = \left\{ \max_{t \in [0,1]} \bar{Q}^n(t) > b \right\}.$$

The large deviation asymptotic behavior of this sequence of events has been studied in general; see, for example, [1] and section 11.7 of [24]. Here we will focus on a simple special case. Assume that under ν, X_n are i.i.d. standard Poisson. Let $\mathcal{AC}([0,1] : \mathbb{R})$ [resp., $\mathcal{D}([0,1] : \mathbb{R})$] denote the space of functions that are absolutely continuous [resp., right continuous with left limits] and that map $[0,1]$ to $\mathbb{R}$. Equip $\mathcal{D}([0,1] : \mathbb{R})$ with the Skorohod J_1 topology. The processes $\bar{S}^n$ are known to satisfy a sample-path large deviation principle in $\mathcal{D}([0,1] : \mathbb{R})$ with the rate function I given by

$$I(\varphi) = \begin{cases} \int_0^1 \ell(\dot{\varphi}(t))dt & \text{if } \varphi \in \mathcal{AC}([0,1] : \mathbb{R}), \varphi(0) = 0, \\ \infty & \text{otherwise,} \end{cases}$$

where, with the convention $0 \log 0 = 0$,

$$\ell(x) = \begin{cases} x \log x - x + 1 & \text{if } x \geq 0, \\ \infty & \text{if } x < 0; \end{cases}$$

see [21, Theorem 6.1(b)]. Hence $\lim_{n \to \infty} \frac{1}{n} \log \nu(A_n) = -c$, where

$$c \doteq \inf \left\{ \int_0^1 \ell(\dot{\varphi}(t))dt : \varphi \in \mathcal{AC}, \varphi(0) = 0, \max_{0 \leq s \leq t \leq 1} \varphi(t) - \varphi(s) - C(t-s) \geq b \right\}$$

can be found explicitly. Let m^* and t^* denote the minimum of $t\ell(C + \frac{b}{t})$ over $t > 0$ and the unique minimizer, respectively. Then

$$c = \begin{cases} \ell(C + b) & \text{if } t^* \geq 1, \\ m^* & \text{if } t^* < 1. \end{cases}$$

Note that the event A_n depends only on $X_1, \dots, X_n$. Fix $0 < \beta < \gamma$, and, as before, let $\alpha = \gamma/(\gamma - \beta)$. If θ is any probability measure under which X_n are i.i.d. and

$$\frac{1}{\gamma - \beta} R_\alpha(\theta_1 \| \nu_1) \le d_1, \qquad \frac{\gamma}{\beta(\gamma - \beta)} R_\alpha(\nu_1 \| \theta_1) \le d_2,$$

for some constants d_1, d_2, then we obtain from the LPCBs, (2.9) and (2.10), that for all n,

$$\frac{1}{n}\frac{\gamma}{\beta^2} \log \mathbb{P}_\nu(A_n) - d_2 \le \frac{1}{n}\frac{1}{\beta} \log \mathbb{P}_\theta(A_n) \le \frac{1}{n}\frac{1}{\gamma} \log \mathbb{P}_\nu(A_n) + d_1,$$

or

$$\mathbb{P}_\nu(A_n)^{\frac{\gamma}{\beta}} e^{-n\beta d_2} \le \mathbb{P}_\theta(A_n) \le \mathbb{P}_\nu(A_n)^{\frac{\beta}{\gamma}} e^{n\beta d_1}.$$

More generally, the same conclusions hold if under θ, X_n are mutually independent and

$$\frac{1}{n}\frac{1}{\gamma - \beta} \sum_{i=1}^{n} R_\alpha(\theta_i^n \| \nu_1) \le d_1, \qquad \frac{1}{n}\frac{\gamma}{\beta(\gamma - \beta)} \sum_{i=1}^{n} R_\alpha(\nu_1 \| \theta_i^n) \le d_2.$$

3.3. Brownian motion with drift. Let B_t be standard Brownian motion on $0 \le t \le 1$, and let P be the corresponding standard Wiener measure on $\mathcal{C}([0,1] : \mathbb{R})$. Let Q be the measure induced by Brownian motion with constant drift, i.e.,

$$X_t = B_t + \mu t,$$

where $\mu \in \mathbb{R}$. Also, let $\tilde{Q}$ be the measure induced by the paths of the solution X to the stochastic differential equation (SDE)

$$dX_t = m(X_t)dt + dB_t, \qquad X_0 = 0,$$

for measurable m, where, by assumption, weak existence and uniqueness hold. A simple calculation based on Girsanov's theorem yields that the Rényi divergence between Q and P is given by

$$R_\alpha(Q\|P) = R_\alpha(P\|Q) = \frac{\mu^2}{2} \tag{3.5}$$

and that if $|m(x)| \le |\mu|$ for all x, then

$$R_\alpha(\tilde{Q}\|P) \le \frac{\mu^2}{2}, \qquad R_\alpha(P\|\tilde{Q}) \le \frac{\mu^2}{2}. \tag{3.6}$$

Let A be the event that the path exceeds a certain level $K > 0$:

$$A \doteq \left\{ \omega : \sup_{0 \le t \le 1} X_t > K \right\}.$$

The exceedance probability under the measure Q, which represents the probability of Brownian motion with constant drift exceeding K, is given (see [5, section 2.1]) by

$$Q(A) = \frac{1}{2}\operatorname{erfc}\left(\frac{K - \mu}{\sqrt{2}}\right) + \frac{1}{2} e^{2\mu K} \operatorname{erfc}\left(\frac{K + \mu}{\sqrt{2}}\right),$$

where $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-v^2} dv$, and under standard Wiener measure,

$$P(A) = \sqrt{\frac{2}{\pi}} \int_K^\infty e^{-x^2/2} dx = \text{erfc}\left(\frac{K}{\sqrt{2}}\right). \tag{3.7}$$

We would like to identify the bounds on $Q(A)$ and $\tilde{Q}(A)$ that Corollary 2.4 provides. As before, fix $0 < \beta < \gamma$ and let $\alpha = \gamma/(\gamma - \beta)$. Then by the LPCBs, (2.9) and (2.10),

$$\frac{\gamma}{\beta^2} \log P(A) - \frac{\gamma}{\beta(\gamma-\beta)} R_\alpha(P\|Q) \le \frac{1}{\beta} \log Q(A) \le \frac{1}{\gamma} \log P(A) + \frac{1}{\gamma-\beta} R_\alpha(Q\|P).$$

By (3.5) and (3.7) this gives

$$\frac{\gamma}{\beta^2} \log \text{erfc}\left(\frac{K}{\sqrt{2}}\right) - \frac{\gamma}{\beta(\gamma-\beta)} \frac{\mu^2}{2} \le \frac{1}{\beta} \log Q(A) \le \frac{1}{\gamma} \log \text{erfc}\left(\frac{K}{\sqrt{2}}\right) + \frac{1}{\gamma-\beta} \frac{\mu^2}{2},$$

or in probability scale

$$\text{erfc}\left(\frac{K}{\sqrt{2}}\right)^{\frac{\gamma}{\beta}} e^{-\frac{\gamma}{\beta(\gamma-\beta)}\frac{\mu^2}{2}} \le Q(A) \le \text{erfc}\left(\frac{K}{\sqrt{2}}\right)^{\frac{\beta}{\gamma}} e^{\frac{1}{\gamma-\beta}\frac{\mu^2}{2}}.$$

By (3.6), the same conclusion holds for $\tilde{Q}(A)$.

To illustrate these upper and lower bounds, we consider Brownian motion with constant drift with $|\mu| \le .1$ so that $R_\alpha(P\|Q) \le .005$ and $K = 4$. Note that with $K = 4$,

$$P(A) \approx 6.33 \times 10^{-5}.$$

Figures 1 and 2 show the upper and lower bounds in probability scale, plotted as a function of $\gamma \ge \beta$, for $\beta = 10$ and $\beta = 40$.

As another example involving the measures P, Q, and $\tilde{Q}$, consider the random variable

$$H(t) = \inf\left\{ s \in [0,t] : X_s = \sup_{u \in [0,t]} X_u \right\}, \qquad t \ge 0.$$

Let I_0 be the modified Bessel function of the first kind. The Laplace transform of $H(t)$ in the case of the standard Wiener measure is given by

$$\mathbb{E}_P[e^{-vH(t)}] = e^{-vt/2} I_0\left(\frac{vt}{2}\right), \qquad v \in \mathbb{R}.$$

For the case of constant drift,

$$\mathbb{E}_Q[e^{-vH(t)}] = \left(\frac{e^{-vt-\mu^2 t/2}}{\sqrt{\pi t}} + \frac{\mu e^{-vt}}{\sqrt{2}} \text{erfc}\left(-\frac{\mu\sqrt{t}}{\sqrt{2}}\right) \right) * \left(\frac{e^{-\mu^2 t/2}}{\sqrt{\pi t}} - \frac{\mu}{\sqrt{2}} \text{erfc}\left(\frac{\mu\sqrt{t}}{\sqrt{2}}\right) \right),$$

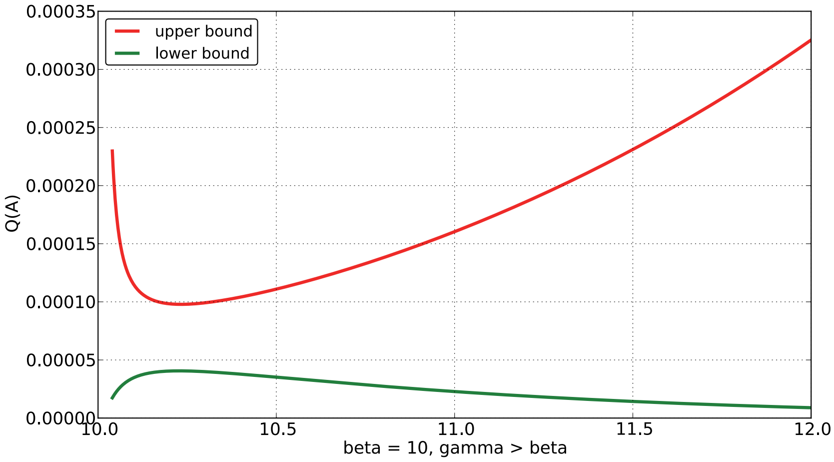

Figure 1. *Upper and lower bounds for* $Q(A)$ *and* $\tilde{Q}(A)$, $\beta = 10$.

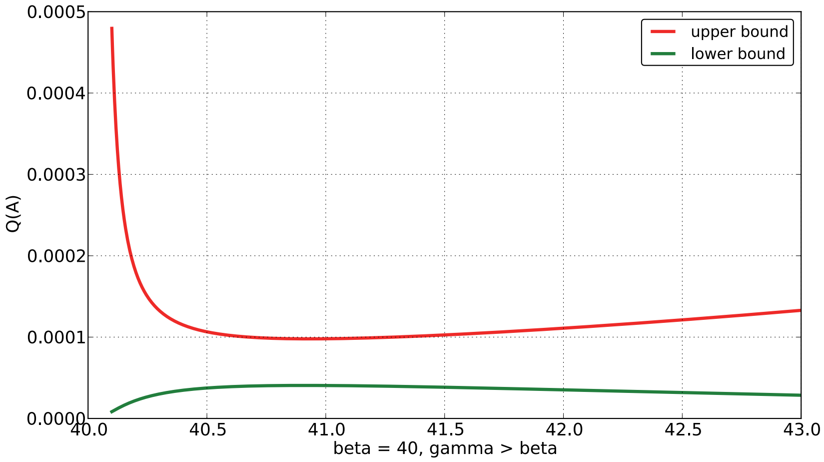

Figure 2. *Upper and lower bounds for* $Q(A)$ *and* $\tilde{Q}(A)$, $\beta = 40$.

where $f(t) * g(t)$ denotes the convolution of f and g evaluated at t (see [5]). There is no explicit expression for the case of an SDE. To obtain bounds on the behavior under Q and $\tilde{Q}$ we apply Corollary 2.4, which gives

$$\begin{aligned}\frac{\gamma}{\beta^2}\left[-\frac{\beta^2 vt}{2\gamma} + \log I_0\left(\frac{\beta^2 vt}{2\gamma}\right)\right] - \frac{\gamma}{\beta(\gamma-\beta)}\frac{\mu^2 t}{2} \\ \leq \frac{1}{\beta}\log \mathbb{E}_Q[e^{-\beta v H(t)}] \\ \leq \frac{1}{\gamma}\left[-\frac{\gamma vt}{2} + \log I_0\left(\frac{\gamma vt}{2}\right)\right] + \frac{1}{\gamma-\beta}\frac{\mu^2 t}{2},\end{aligned}$$

where the upper bound uses (2.8) directly and the lower bound is based on (2.8) where θ is switched with ν and (β, γ) with $(\beta^2/\gamma, \beta)$. As before, the same upper and lower bounds are valid for $\tilde{Q}$ as well.

4. Proofs of Theorem 2.1 and Corollary 2.4.

Proof of Theorem 2.1. The main part of the proof will be to show the validity of (2.6) and (2.7) for all $\alpha > 0$, $\alpha \neq 1$. Before proving these identities, let us show that they imply (2.4) and (2.5). First, note that (2.6) and (2.7) for $\alpha > 0$, $\alpha \neq 1$ imply (2.6) and (2.7) for all $\alpha \in \mathbb{R} \setminus \{0, 1\}$. Indeed, if $\alpha < 0$, then (2.6) with $\bar{\alpha} = 1 - \alpha > 1$ and $\bar{g} = -g$ reads as

$$\frac{1}{\bar{\alpha}-1} \log \int e^{(\bar{\alpha}-1)\bar{g}} d\theta = \inf_{\theta \in \mathcal{P}(S)} \left[\frac{1}{\bar{\alpha}} \log \int e^{\bar{\alpha}\bar{g}} d\theta + R_{\bar{\alpha}}(\nu \| \theta) \right].$$

Expressed in terms of α and g,

$$-\frac{1}{\alpha} \log \int e^{\alpha g} d\theta = \inf_{\theta \in \mathcal{P}(S)} \left[-\frac{1}{\alpha - 1} \log \int e^{(\alpha-1)g} d\theta + R_{\alpha}(\theta \| \nu) \right],$$

where we used (2.3). Multiplying by (-1) establishes the validity of (2.7) for $\alpha < 0$. In a similar way, the validity of (2.6) for $\alpha < 0$ follows from that of (2.7) for $\bar{\alpha} > 1$.

Next, to show that (2.6) and (2.7) with $\alpha \in \mathbb{R} \setminus \{0, 1\}$ imply (2.4) and (2.5), fix β and γ in $\mathbb{R} \setminus \{0\}$, $\beta < \gamma$. Apply (2.6) with $\alpha = \frac{\gamma}{\gamma - \beta}$ and $g = (\gamma - \beta) f$ (note that $\alpha \notin \{0, 1\}$) and divide by $\gamma - \beta$ to get (2.4) (with f in place of g). In a similar way, (2.5) follows from (2.7).

We turn to proving (2.6) for $\alpha > 0$, $\alpha \neq 1$. Fix ν, and consider first the case $\alpha > 1$. Given any θ, let $\mu = \mu(\theta)$ be a measure dominating both ν and θ, and denote by ν' and θ' the corresponding densities. Define $\lambda \in \mathcal{P}(S)$ by $d\lambda = e^{-g} d\nu / Z$, where $Z = \int_S e^{-g} d\nu$, and let λ' be the density of λ with respect to μ. Then $\lambda'/\nu' = e^{-g}/Z$, and so

$$\begin{aligned}
\log \int_S e^{\alpha g} d\theta &\geq \log \int_{\{\nu' > 0\}} e^{\alpha g} d\theta \\
&= \log \int_{\{\nu' > 0\}} \frac{1}{Z} \frac{\nu'}{\lambda'} e^{(\alpha-1)g} d\theta \\
&= \log \int_S \frac{1}{Z} \frac{\nu' \theta'}{\lambda'} e^{(\alpha-1)g} d\mu \\
&= \log \int_S \frac{1}{Z} \frac{\theta'}{\lambda'} e^{(\alpha-1)g} d\nu.
\end{aligned} \tag{4.1}$$

Suppose that $\nu \ll \theta$. Then

$$\begin{aligned}
R_{\alpha}(\nu \| \theta) &= \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu'\theta' > 0\}} \left(\frac{\nu'}{\theta'} \right)^{\alpha} d\theta \\
&= \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu'\theta' > 0\}} \left(\frac{\nu'}{\theta'} \right)^{\alpha-1} d\nu \\
&= \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu' > 0\}} \left(\frac{\nu'}{\theta'} \right)^{\alpha-1} d\nu \\
&= \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu' > 0\}} Z^{\alpha-1} \left(\frac{\lambda'}{\theta'} \right)^{\alpha-1} e^{(\alpha-1)g} d\nu,
\end{aligned}$$

where changing the domain of integration in the third equality uses the fact that $\nu\{\theta'=0\}=0$, which follows from $\theta\{\theta'=0\}=0$ and the assumption $\nu \ll \theta$. Thus with $d\tilde{\nu}=e^{(\alpha-1)g}d\nu$, and since the terms involving Z cancel,

$$\frac{1}{\alpha}\log\int_S e^{\alpha g}d\theta + R_\alpha(\nu\|\theta) \geq \frac{1}{\alpha}\log\int_S \frac{\theta'}{\lambda'}d\tilde{\nu} + \frac{1}{\alpha(\alpha-1)}\log\int_{\{\nu'>0\}}\left(\frac{\lambda'}{\theta'}\right)^{\alpha-1}d\tilde{\nu}. \tag{4.2}$$

On the set $\{\lambda'\theta'>0\}$, define

$$\varphi=\left(\frac{\lambda'}{\theta'}\right)^{\frac{\alpha-1}{\alpha}}, \qquad \psi=\left(\frac{\theta'}{\lambda'}\right)^{\frac{\alpha-1}{\alpha}},$$

so that $\varphi\psi=1$ on $\{\lambda'\theta'>0\}$. By Hölder's inequality with $1/p=1/\alpha$ and $1/q=(\alpha-1)/\alpha$, and with p attached to φ and q attached to ψ, and again using $\nu \ll \theta$, we have

$$\int_S d\tilde{\nu} = \int_{\{\lambda'\theta'>0\}} d\tilde{\nu} \leq \left(\int_{\{\lambda'\theta'>0\}}\left(\frac{\lambda'}{\theta'}\right)^{\alpha-1}d\tilde{\nu}\right)^{\frac{1}{\alpha}}\left(\int_S \frac{\theta'}{\lambda'}d\tilde{\nu}\right)^{\frac{\alpha-1}{\alpha}}. \tag{4.3}$$

Since $d\lambda/d\nu = e^{-g}/Z$ and g is bounded, $\{\lambda'\theta'>0\}\subset\{\nu'>0\}$. Taking logs, dividing by $\alpha-1$, and using (4.2) gives that for any $\theta\in\mathcal{P}(S)$ with $\theta \gg \nu$,

$$\frac{1}{\alpha}\log\int_S e^{\alpha g}d\theta + R_\alpha(\nu\|\theta) \geq \frac{1}{\alpha-1}\log\int_S d\tilde{\nu} = \frac{1}{\alpha-1}\log\int_S e^{(\alpha-1)g}d\nu. \tag{4.4}$$

If $\nu \not\ll \theta$, then $R_\alpha(\nu\|\theta)=\infty$, and again the inequality holds.

Taking the infimum over all $\theta\in\mathcal{P}(S)$ shows that the right-hand side of (2.6) is bounded below by the left-hand side. Note that since g is bounded, $\lambda\{\nu'>0\}=\lambda\{S\}=1$. Thus the choice $\theta=\lambda$ gives equality in both (4.1) and (4.3), and also $\{\lambda'\theta'>0\}=\{\nu'>0\}$. Hence (4.4) holds with equality, and therefore λ is a minimizer.

Finally, we show that the minimizer is unique. Assume that $\theta \gg \nu$ attains the infimum over $\mathcal{P}(S)$. Then both (4.1) and (4.3) must hold with equality. For (4.1) to hold with equality requires $\theta\{\nu'=0\}=0$, and so $\theta\sim\nu$ must be true. Recall that Hölder's inequality will give an equality if and only if θ'/λ' is constant on $\{\theta'\lambda'>0\}$. Thus for any measurable set A,

$$\theta(A)=\int_A \frac{d\theta}{d\lambda}d\lambda = \int_{A\cap\{\theta'\lambda'>0\}}\frac{\theta'}{\lambda'}d\lambda = c\lambda(A),$$

and so the only probability measure that satisfies these conditions is $\theta=\lambda$. This shows that λ attains the infimum uniquely.

Next we consider (2.6) for the same ν, but for $\alpha\in(0,1)$. In this case, we can no longer assume $\theta \gg \nu$. To show that the left-hand side of (2.6) is a lower bound for the right-hand side, consider any $\theta\in\mathcal{P}(S)$. As with the case $\alpha>1$, let μ be a measure dominating both ν and θ, and define ν',θ', and λ' with respect to this measure, where $d\lambda=e^{-g}d\nu/Z$. Starting with the right-hand side of (2.6),

$$\begin{aligned}\log\int_S e^{\alpha g}d\theta &\geq \log\int_{\{\nu'\theta'>0\}} e^{\alpha g}d\theta \\ &= \log\int_{\{\nu'\theta'>0\}}\frac{1}{Z}\frac{\theta'}{\lambda'}e^{(\alpha-1)g}d\nu,\end{aligned} \tag{4.5}$$

and

$$R_\alpha(\nu\|\theta) = \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu'\theta'>0\}} \left(\frac{\nu'}{\theta'}\right)^\alpha d\theta$$
$$= \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu'\theta'>0\}} Z^{\alpha-1} \left(\frac{\lambda'}{\theta'}\right)^{\alpha-1} e^{(\alpha-1)g} d\nu.$$

With $\tilde{\nu}$ again defined by $d\tilde{\nu} = e^{(\alpha-1)g} d\nu$,

$$\frac{1}{\alpha} \log \int_S e^{\alpha g} d\theta + R_\alpha(\nu\|\theta) \geq \frac{1}{\alpha} \log \int_{\{\nu'\theta'>0\}} \frac{\theta'}{\lambda'} d\tilde{\nu} + \frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu'\theta'>0\}} \left(\frac{\lambda'}{\theta'}\right)^{\alpha-1} d\tilde{\nu}. \tag{4.6}$$

Define $\varphi = 1$ and $\psi = (\lambda'/\theta')^{\alpha-1}$ on the set $\{\nu'\theta' > 0\}$. Using Hölder's inequality with $p = 1/\alpha$ attached to φ and $q = 1/(1-\alpha)$ attached to ψ gives

$$\int_{\{\nu'\theta'>0\}} \left(\frac{\lambda'}{\theta'}\right)^{\alpha-1} d\tilde{\nu} \leq \left(\int_{\{\nu'\theta'>0\}} d\tilde{\nu}\right)^\alpha \left(\int_{\{\nu'\theta'>0\}} \frac{\theta'}{\lambda'} d\tilde{\nu}\right)^{1-\alpha}.$$

Taking logs and dividing by $\alpha(\alpha-1) < 0$ gives

$$\frac{1}{\alpha(\alpha-1)} \log \int_{\{\nu'\theta'>0\}} \left(\frac{\lambda'}{\theta'}\right)^{\alpha-1} d\tilde{\nu} \geq \frac{1}{\alpha-1} \log \int_{\{\nu'\theta'>0\}} d\tilde{\nu} - \frac{1}{\alpha} \log \int_{\{\nu'\theta'>0\}} \frac{\theta'}{\lambda'} d\tilde{\nu}.$$

Using (4.6) and recalling $\alpha \in (0,1)$ gives

$$\frac{1}{\alpha} \log \int_S e^{\alpha g} d\theta + R_\alpha(\nu\|\theta) \geq \frac{1}{\alpha-1} \log \int_{\{\nu'\theta'>0\}} d\tilde{\nu}$$
$$= \frac{1}{\alpha-1} \log \int_{\{\nu'\theta'>0\}} e^{(\alpha-1)g} d\nu$$
$$\geq \frac{1}{\alpha-1} \log \int_S e^{(\alpha-1)g} d\nu, \tag{4.7}$$

showing that (2.6) holds as an inequality. To show equality, substitute λ for θ and note that all the inequalities hold as equalities.

To show that λ is the unique minimizer, note that any $\theta \in \mathcal{P}(S)$ satisfying all inequalities as equalities must, in particular, give equality in (4.5), for which it is necessary that $\theta \ll \nu$. Equality in (4.7) implies $\nu \ll \theta$. For Hölder's inequality to hold with equality, ψ must be a constant, and the only probability measure satisfying these conditions is λ. This completes the proof of (2.6).

Toward proving (2.7), note that (2.6) implies

$$\frac{1}{\alpha-1} \log \int_S e^{(\alpha-1)g} d\nu \leq \frac{1}{\alpha} \log \int_S e^{\alpha g} d\theta + R_\alpha(\nu\|\theta), \qquad \nu, \theta \in \mathcal{P}(S),$$

which is equivalent to

$$\frac{1}{\alpha} \log \int_S e^{\alpha g} d\nu \geq \frac{1}{\alpha-1} \log \int_S e^{(\alpha-1)g} d\theta - R_\alpha(\theta\|\nu), \qquad \nu, \theta \in \mathcal{P}(S).$$

Thus to prove part (2.7), it suffices to show that the measure $d\theta = e^g d\nu/Z$, and only this measure, gives equality in the above display. The proof is similar to that of (2.6), and therefore the details are omitted.

Proof of Corollary 2.4. We first prove inequality (2.8). If g is bounded, the result follows from (2.5) in Theorem 2.1. Otherwise, since the claim holds trivially if the right-hand side is infinite, assume it is finite. Let $g^{M,N} = (g \vee -M) \wedge N$ for $M, N \geq 0$. Then

$$\frac{1}{\beta} \log \int_S e^{\beta g^{M,N}} d\theta \leq \frac{1}{\gamma} \log \int_S e^{\gamma g^{M,N}} d\nu + \frac{1}{\gamma - \beta} R_{\frac{\gamma}{\gamma-\beta}}(\theta \| \nu). \tag{4.8}$$

We first take $M \to \infty$ and use bounded convergence on both sides of the inequality. To this end note that for fixed N and all $M \geq 0$, $e^{\gamma g^{M,N}(\cdot)} \leq e^{\gamma N}$, and so, as $M \to \infty$, the first term on the right-hand side of (4.8) converges to $\gamma^{-1} \log \int_S e^{\gamma g^{\infty,N}} d\nu$. A similar remark holds for the left-hand side of (4.8), and as a result, (4.8) holds with $g^{\infty,N}$ on both sides. Now we take $N \to \infty$ and use monotone convergence (recall that $0 < \beta < \gamma$). This gives inequality (2.8).

As for part (ii), consider (2.8) with $g = 0$ on A and $g = -M$ on A^c. Then taking $M \to \infty$ gives (2.9) by bounded convergence. Next, (2.10) is obtained from (2.9) by switching θ with ν, and (β, γ) with $(\beta^2/\gamma, \beta)$ (note that $0 < \beta^2/\gamma < \beta$). The final assertion about equality follows by straightforward calculation.

Acknowledgments. We are grateful to Ramon van Handel for bringing to our attention the dualities in [6, 19], and also thank two referees for suggestions that improved the paper.

REFERENCES

[1] V. Anantharam, *How large delays build up in a $GI/G/1$ queue*, Queueing Systems Theory Appl., 5 (1989), pp. 345–367.

[2] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Stoch. Model. Appl. Probab. 57, Springer, New York, 2007.

[3] A. Bhattacharyya, *On some analogues of the amount of information and their use in statistical estimation*, Sankhyā, 8 (1946), pp. 1–14.

[4] J. H. Blanchet and P. Glynn, *Efficient rare-event simulation for the maximum of heavy-tailed random walks*, Ann. Appl. Probab., 18 (2008), pp. 1351–1378.

[5] A. N. Borodin and P. Salminen, *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed., Probab. Appl., Birkhäuser Verlag, Basel, 2002.

[6] D. Chafaï, *Entropies, convexity, and functional inequalities: On Φ-entropies and Φ-Sobolev inequalities*, J. Math. Kyoto Univ., 44 (2004), pp. 325–363.

[7] K. Chowdhary and P. Dupuis, *Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification*, ESAIM Math. Model. Numer. Anal., 47 (2013), pp. 635–662.

[8] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, 1993.

[9] P. Dupuis and R. S. Ellis, *Large deviations for Markov processes with discontinuous statistics,* II: *Random walks*, Probab. Theory Related Fields, 91 (1992), pp. 153–194.

[10] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, New York, 1997.

[11] P. Dupuis, M. R. James, and I. R. Petersen, *Robust properties of risk–sensitive control*, Math. Control Signals Systems, 13 (2000), pp. 318–332.

[12] P. Dupuis and H. Wang, *Subsolutions of an Isaacs equation and efficient schemes for importance sampling*, Math. Oper. Res., 32 (2007), pp. 1–35.

[13] K. Dvijotham and E. Todorov, *A unified theory of linearly solvable optimal control*, Artificial Intelligence (UAI), 2011.

[14] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1984.

[15] L. Golshani, E. Pasha, and G. Yari, *Some properties of Rényi entropy and Rényi entropy rate*, Inform. Sci., 179 (2009), pp. 2426–2433.

[16] H. J. Kushner, *Robustness and approximation of escape times and large deviations estimates for systems with small noise effects*, SIAM J. Appl. Math., 44 (1984), pp. 160–182.

[17] J. Li and D. Xiu, *Computation of failure probability subject to epistemic uncertainty*, SIAM J. Sci. Comput., 34 (2012), pp. A2946–A2964.

[18] F. Liese and I. Vajda, *Convex Statistical Distances*, Teubner-Texte Math. 95, Teubner, Leipzig, 1987.

[19] P. Massart, *Concentration Inequalities and Model Selection*, Lecture Notes in Math. 1896, Springer, Berlin, 2007.

[20] X. Nguyen, H. J. Wainwright, and M. I. Jordan, *Estimating divergence functionals and the likelihood ratio by convex risk minimization*, IEEE Trans. Inform. Theory, 56 (2010), pp. 5847–5861.

[21] A. A. Puhalskii and W. Whitt, *Functional large deviation principles for first-passage-time processes*, Ann. Appl. Probab., 7 (1997), pp. 362–381.

[22] A. Rényi, *On measures of entropy and information*, in Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, University of California Press, Berkeley, CA, 1961, pp. 547–561.

[23] A. Ruderman, M. Reid, D. García-García, and J. Petterson, *Tighter Variational Representations of f-Divergences via Restriction to Probability Measures*, preprint, arXiv:1206.4464, 2012.

[24] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis: Queues, Communication and Computing*, Chapman and Hall, New York, 1995.

[25] I. Vajda, *Distances and discrimination rates for stochastic processes*, Stochastic Process. Appl., 35 (1990), pp. 47–57.

[26] T. van Erven and P. Harremoës, *Rényi divergence and majorization*, in Proceedings of the IEEE International Symposium on Information Theory (ISIT), 2010, pp. 1335–1339.

[27] P. Whittle, *Risk-sensitive linear/quadratic/Gaussian control*, Adv. in Appl. Probab., 13 (1981), pp. 764–777.