

ON OPTIMAL CHOICE OF REFERENCE PARAMETERS IN THE LIKELIHOOD RATIO METHOD

Reuven Y. Rubinstein

Alexander Shapiro

Faculty of Industrial Engineering & Management,
 Technion, Haifa 32000, ISRAEL

School of Industrial & Systems Engineering
 Georgia Institute of Technology
 Atlanta, Georgia 30332, U.S.A.

ABSTRACT

The likelihood ratio method allows estimation of the expected performance and associated derivatives of a stochastic system for various values of the parameters from a single simulation run. In this method the choice of the reference value v_0 of the parameter vector, which is employed in the generated sample, is crucial. In this paper we discuss how to choose the reference vector v_0 in order to minimize the variance of the considered likelihood ratio estimators. We show that for exponential families of distributions this variance is a convex function of v_0 . We also show that for some exponential families (e.g. Gamma distributions) there is an order relation between the involved parameters and their optimal reference values.

1 INTRODUCTION

Let

$$\ell(v) = \mathbb{E}_v\{L(\mathbf{Y})\} = \int L(\mathbf{y})f(\mathbf{y}, v)d\mathbf{y} \quad (1)$$

be the expected performance of a stochastic system with $L(\mathbf{Y})$ being a sample performance driven by an input vector \mathbf{Y} having a probability distribution function $f(\mathbf{y}, v)$.

Assume that $\ell(v)$ is too complex, and must be evaluated by simulation. To do so, one generates a random sample $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ from $f(\mathbf{y}, v)$ for a particular value v from a permissible parameter set V and then estimates $\ell(v)$ by the sample mean

$$\hat{\ell}_N(v) = N^{-1} \sum_{i=1}^N L(\mathbf{Y}_i). \quad (2)$$

Of course, such a procedure requires generation of a new sample every time $\ell(v)$ should be estimated for a new value of the parameter vector v . The likelihood ratio method on the other hand allows estimation of

$\ell(v)$ for various values of v from a *single* simulation run. That is, let us choose a particular $v_0 \in V$ and consider the likelihood ratio function

$$W(\mathbf{y}, v) = \frac{f(\mathbf{y}, v)}{f(\mathbf{y}, v_0)}.$$

Then it is not difficult to show that

$$\ell(v) = \mathbb{E}_{v_0}\{L(\mathbf{Z})W(\mathbf{Z}, v)\}, \quad (3)$$

where $\mathbf{Z} \sim f(\mathbf{z}, v_0)$. We shall call v_0 the *reference parameter vector*.

By generating now a sample $\mathbf{Z}_1, \dots, \mathbf{Z}_N$ from $f(\mathbf{z}, v_0)$ we can estimate $\ell(v)$ by the corresponding sample mean

$$\bar{\ell}_N(v) = N^{-1} \sum_{i=1}^N L(\mathbf{Z}_i)W(\mathbf{Z}_i, v). \quad (4)$$

Moreover, under suitable regularity conditions ensuring interchangeability of the differentiation and integration operators in the right hand side of (3) we can estimate the gradient $\nabla\ell(v)$ of the expected performance by

$$\nabla\bar{\ell}_N(v) = N^{-1} \sum_{i=1}^N L(\mathbf{Z}_i)\nabla W(\mathbf{Z}_i, v). \quad (5)$$

The above approach can be extended to dynamic systems as follows. Consider a Discrete Event Dynamic System (DEDS) driven by the input sequence of independent identically distributed random vectors $\mathbf{Y}_1, \mathbf{Y}_2, \dots$, with the common pdf $f(\mathbf{y}, v)$. Let $L_t = L_t(\mathbf{Y}_t)$ be a corresponding output process. Here $\mathbf{Y}_t = (\mathbf{Y}_1, \dots, \mathbf{Y}_t)$ represents a history of the input process up to time t and $L_t(\cdot)$ is a sequence of real valued functions. Let τ be a stopping time for the input process $\mathbf{Y}_1, \mathbf{Y}_2, \dots$, and consider

$$\ell(v) = \frac{\mathbb{E}_v\{\sum_{t=1}^{\tau} L_t\}}{\mathbb{E}_v\tau}. \quad (6)$$

For example, if $\{L_t\}$ is a discrete time regenerative process with the regenerative cycle of length τ , then it is a well known result in the renewal theory that the expected steady-state performance (long-run average) of L_t is given by the right hand side of (6) (e.g., Ross, 1983). It is possible to show (see Asmussen and Rubinstein, 1992; Rubinstein and Shapiro, 1992) that $\ell_1(\mathbf{v}) = \mathbb{E}_{\mathbf{v}}\{\sum_{t=1}^{\tau} L_t\}$ can be represented in the form

$$\ell_1(\mathbf{v}) = \mathbb{E}_{\mathbf{v}_0}\left\{\sum_{t=1}^{\tau} L_t(\mathbf{Z}_t)\widetilde{W}_t(\mathbf{Z}_t, \mathbf{v})\right\}, \quad (7)$$

where

$$\widetilde{W}_t(\mathbf{z}_t, \mathbf{v}) = \frac{f_t(\mathbf{z}_t, \mathbf{v})}{f_t(\mathbf{z}_t, \mathbf{v}_0)}$$

and

$$f_t(\mathbf{z}_t, \mathbf{v}) = \prod_{i=1}^t f(\mathbf{z}_i, \mathbf{v}).$$

The function $\ell_2(\mathbf{v}) = \mathbb{E}_{\mathbf{v}}\tau$ can be written in a similar way by taking $L_t(\mathbf{z}_t) = 1$, and hence $\ell(\mathbf{v})$ can be represented in the form

$$\ell(\mathbf{v}) = \frac{\mathbb{E}_{\mathbf{v}_0}\left\{\sum_{t=1}^{\tau} L_t(\mathbf{Z}_t)\widetilde{W}_t(\mathbf{Z}_t, \mathbf{v})\right\}}{\mathbb{E}_{\mathbf{v}_0}\left\{\sum_{t=1}^{\tau} \widetilde{W}_t(\mathbf{Z}_t, \mathbf{v})\right\}}. \quad (8)$$

Using representation (8) one can estimate $\ell(\mathbf{v})$ for different values of \mathbf{v} , from a single simulation run, by the ratio $\bar{\ell}_{1N}(\mathbf{v})/\bar{\ell}_{2N}(\mathbf{v})$ with $\bar{\ell}_{1N}(\mathbf{v})$ and $\bar{\ell}_{2N}(\mathbf{v})$ being the corresponding sample mean estimators of $\ell_1(\mathbf{v})$ and $\ell_2(\mathbf{v})$, respectively.

In the above method the choice of the reference value \mathbf{v}_0 of the parameter vector is crucial. A bad choice of \mathbf{v}_0 can result in very large variances of the corresponding estimators and consequently can lead to inaccurate estimation procedures.

Consider, for example, the estimator $\bar{\ell}_N(\mathbf{v})$ defined in (4). In order to improve accuracy of this estimator one can try to minimize its variance. Variance of $\bar{\ell}_N(\mathbf{v})$ is given by N^{-1} times the variance of $L(\mathbf{Z})W(\mathbf{Z}, \mathbf{v})$. Note that $\mathbb{E}_{\mathbf{v}_0}\{L(\mathbf{Z})W(\mathbf{Z}, \mathbf{v})\}$ is equal to $\ell(\mathbf{v})$ and hence is independent of \mathbf{v}_0 . Therefore the problem of minimization of the variance of $\bar{\ell}_N(\mathbf{v})$ is equivalent to minimization of the function

$$\phi(\mathbf{v}, \mathbf{v}_0) = \mathbb{E}_{\mathbf{v}_0}\{[L(\mathbf{Z})W(\mathbf{Z}, \mathbf{v})]^2\} \quad (9)$$

with respect to \mathbf{v}_0 .

In this paper we discuss how to choose the reference vector \mathbf{v}_0 in order to minimize variance of the considered estimators. This problem can be also considered in the context of *importance sampling* scheme (cf. Glynn and Iglehart, 1989). We show that for exponential families of distributions the variance of the likelihood ratio estimators is a *convex* function of \mathbf{v}_0 .

Typically this implies existence of a unique reference value \mathbf{v}_0^* which minimizes the variance. We also show that for some exponential families (e.g. Gamma distributions) there exists an *order relation* between the parameter \mathbf{v} and the corresponding optimal reference value \mathbf{v}_0^* .

2 CONVEXITY RESULTS

In this section we deal with exponential families of distributions in the canonical form. That is, we consider probability density functions of the form

$$f(\mathbf{y}, \mathbf{v}) = c(\mathbf{v}) \exp\left\{\sum_{j=1}^n v_j t_j(\mathbf{y})\right\} h(\mathbf{y}) \quad (10)$$

parameterized by $\mathbf{v} = (v_1, \dots, v_n)$. We show that for such families of distributions, variances of the likelihood ratio estimators are convex functions with respect to \mathbf{v}_0 . Let us start by considering the estimator $\bar{\ell}_N(\mathbf{v})$ and the corresponding function $\phi(\mathbf{v}, \mathbf{v}_0)$ given in (9).

Theorem 2.1 *Suppose that the random vector \mathbf{Y} is distributed according to an exponential family in the canonical form (10) and let $\phi(\mathbf{v}, \mathbf{v}_0)$ be the corresponding function defined in (9). Then for any permissible \mathbf{v} the function $\phi(\mathbf{v}, \mathbf{v}_0)$ is convex in \mathbf{v}_0 .*

Proof It immediately follows from the definition of $\phi(\mathbf{v}, \mathbf{v}_0)$ and (10) that

$$\phi(\mathbf{v}, \mathbf{v}_0) = c(\mathbf{v})^2 \int L(\mathbf{y})^2 c(\mathbf{v}_0)^{-1} \times \exp\left\{\sum_{j=1}^n (2v_j - v_{0j}) t_j(\mathbf{y})\right\} h(\mathbf{y}) d\mathbf{y}.$$

Also, since $\int f(\mathbf{y}, \mathbf{v}) d\mathbf{y} = 1$, we have that

$$c(\mathbf{v}_0)^{-1} = \int \exp\left\{\sum_{j=1}^n v_{0j} t_j(\mathbf{z})\right\} h(\mathbf{z}) d\mathbf{z}.$$

Putting these two equations together we obtain

$$\begin{aligned} \phi(\mathbf{v}, \mathbf{v}_0) = & c(\mathbf{v})^2 \int \int L(\mathbf{y})^2 \exp\left\{\sum_{j=1}^n [2v_j t_j(\mathbf{y}) \right. \\ & \left. + v_{0j} (t_j(\mathbf{z}) - t_j(\mathbf{y}))]\right\} h(\mathbf{y}) h(\mathbf{z}) d\mathbf{y} d\mathbf{z}. \end{aligned} \quad (11)$$

Now, for any linear function $g(\mathbf{x})$ of $\mathbf{x} \in \mathbb{R}^k$, the function $\exp\{g(\mathbf{x})\}$ is convex in \mathbf{x} . Therefore, since $L(\mathbf{y})^2$ is nonnegative, for any fixed \mathbf{y}, \mathbf{z} and \mathbf{v} the function under the second integral sign in (11) is convex in \mathbf{v}_0 . This implies convexity of $\phi(\mathbf{v}, \cdot)$. \square

The above result can be extended in several directions. Consider the estimator $\nabla \bar{\ell}_N(\mathbf{v})$ defined in (5).

In this case one can be interested in minimization of the sum of variances of the components of $\nabla \bar{\ell}_N(\mathbf{v})$ which is given by the trace of the covariance matrix of $\nabla \bar{\ell}_N(\mathbf{v})$. Since $\mathbb{E}_{\mathbf{v}_0}\{L(\mathbf{Z})\nabla W(\mathbf{Z}, \mathbf{v})\}$ is equal to $\nabla \ell(\mathbf{v})$ we have that this is equivalent to minimization of the function

$$\phi_1(\mathbf{v}, \mathbf{v}_0) = \text{tr} \mathbb{E}_{\mathbf{v}_0}\{L(\mathbf{Z})^2 \nabla W(\mathbf{Z}, \mathbf{v}) \nabla W(\mathbf{Z}, \mathbf{v})'\}. \quad (12)$$

Moreover, we have that $\nabla W(\mathbf{Z}, \mathbf{v}) = W(\mathbf{Z}, \mathbf{v}) \mathbf{S}^{(1)}(\mathbf{Z}, \mathbf{v})$, where

$$\mathbf{S}^{(1)}(\mathbf{y}, \mathbf{v}) = \frac{\nabla f(\mathbf{y}, \mathbf{v})}{f(\mathbf{y}, \mathbf{v})} = \nabla \log f(\mathbf{y}, \mathbf{v})$$

is the score function of the parametric family $f(\mathbf{y}, \mathbf{v})$, and hence

$$\phi_1(\mathbf{v}, \mathbf{v}_0) = \text{tr} \mathbb{E}_{\mathbf{v}_0}\{L(\mathbf{Z})^2 W(\mathbf{Z}, \mathbf{v})^2 \times \mathbf{S}^{(1)}(\mathbf{Z}, \mathbf{v}) \mathbf{S}^{(1)}(\mathbf{Z}, \mathbf{v})'\}.$$

Taking into account that $\text{tr} \mathbf{S}^{(1)}(\mathbf{z}, \mathbf{v}) \mathbf{S}^{(1)}(\mathbf{z}, \mathbf{v})'$ is nonnegative for any \mathbf{z} and \mathbf{v} , we can conclude as in Theorem 2.1 that $\phi_1(\mathbf{v}, \mathbf{v}_0)$ is convex in \mathbf{v}_0 .

As another extension suppose that several sample functions $L_i(\mathbf{Y})$, $i = 1, \dots, k$, are driven by the same input vector \mathbf{Y} . Let $\ell_i(\mathbf{v}) = \mathbb{E}_{\mathbf{v}}\{L_i(\mathbf{Y})\} = \mathbb{E}_{\mathbf{v}_0}\{L_i(\mathbf{Z})W(\mathbf{Z}, \mathbf{v})\}$ be the corresponding expected performances and consider

$$\ell(\mathbf{v}) = g(\ell_1(\mathbf{v}), \dots, \ell_k(\mathbf{v})),$$

with $g(x_1, \dots, x_k)$ being a real valued function of k variables. Take, for example, $g(x_1, x_2) = x_1/x_2$. The corresponding estimator of $\ell(\mathbf{v})$ can be written as

$$\bar{\ell}_N(\mathbf{v}) = g(\bar{\ell}_{1N}(\mathbf{v}), \dots, \bar{\ell}_{kN}(\mathbf{v})),$$

where $\bar{\ell}_{iN}(\mathbf{v})$ are the likelihood ratio estimators of $\ell_i(\mathbf{v})$, $i = 1, \dots, k$. By the Delta Theorem (e.g., Rao, 1973), we have that $N^{1/2}(\bar{\ell}_N(\mathbf{v}) - \ell(\mathbf{v}))$ is distributed asymptotically multivariate normal with mean zero and variance

$$\sigma^2(\mathbf{v}, \mathbf{v}_0) = \text{var}_{\mathbf{v}_0}\{a_1 L_1(\mathbf{Z})W(\mathbf{Z}, \mathbf{v}) + \dots + a_k L_k(\mathbf{Z})W(\mathbf{Z}, \mathbf{v})\},$$

where $a_i = \partial g(\mathbf{x})/\partial x_i$ at $\mathbf{x} = (\ell_1(\mathbf{v}), \dots, \ell_k(\mathbf{v}))$. In this case, similarly to Theorem 2.1 it is possible to prove that $\sigma^2(\mathbf{v}, \mathbf{v}_0)$ is convex in \mathbf{v}_0 .

Let us consider now the expected performance $\ell_1(\mathbf{v})$ represented in the form (7). Variance of the corresponding estimator is given by N^{-1} times the variance of $\sum_{t=1}^{\tau} L_t(\mathbf{Z}_t) \widetilde{W}_t(\mathbf{Z}_t, \mathbf{v})$.

Theorem 2.2 Suppose that the input sequence $\mathbf{Y}_1, \mathbf{Y}_2, \dots$, is distributed according to an exponential family in the canonical form (10). Then the variance function

$$\sigma_1^2(\mathbf{v}, \mathbf{v}_0) = \text{var}_{\mathbf{v}_0}\{\sum_{t=1}^{\tau} L_t(\mathbf{Z}_t) \widetilde{W}_t(\mathbf{Z}_t, \mathbf{v})\} \quad (13)$$

is convex in \mathbf{v}_0 provided the output process $\{L_t\}$ is nonnegative valued.

Proof. This theorem can be easily proved by arguments and derivations similar to those of Theorem 2.1. To do so, consider the random variable

$$X = \sum_{t=1}^{\tau} L_t(\mathbf{Z}_t) \widetilde{W}_t(\mathbf{Z}_t, \mathbf{v}).$$

Note that since $\ell_1(\mathbf{v}) = \mathbb{E}_{\mathbf{v}_0} X$ we have

$$\text{var}_{\mathbf{v}_0} X = \mathbb{E}_{\mathbf{v}_0}\{X^2\} - \ell_1(\mathbf{v})^2,$$

and hence $\text{var}_{\mathbf{v}_0} X$ depends on \mathbf{v}_0 only through the second moment $\mathbb{E}_{\mathbf{v}_0}\{X^2\}$. We show now that for the pdf $f(\mathbf{y}, \mathbf{v})$ in the form (10), $\mathbb{E}_{\mathbf{v}_0}\{X^2\}$ is a convex function of \mathbf{v}_0 provided all $L_t(\mathbf{y}_t)$ are nonnegative valued.

Denoting by I_A the indicator function of an event A , we can write

$$\begin{aligned} X^2 &= \sum_{k=1}^{\infty} \left(\sum_{t=1}^k L_t \widetilde{W}_t \right)^2 I_{[\tau=k]} \\ &= \sum_{k=1}^{\infty} \left(\sum_{s,t \leq k} L_s L_t \widetilde{W}_s \widetilde{W}_t \right) I_{[\tau=k]} \\ &= \sum_{s,t=1}^{\infty} \sum_{k \geq \max(s,t)} L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau=k]} \\ &= \sum_{s,t=1}^{\infty} L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq \max(s,t)]}. \end{aligned} \quad (14)$$

Therefore it will be sufficient to show that

$$\mathbb{E}_{\mathbf{v}_0}\{L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq t]}\}$$

is a convex function of \mathbf{v}_0 for any $s \leq t$. Now for $s \leq t$ we have

$$\begin{aligned} &\mathbb{E}_{\mathbf{v}_0}\{L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq t]}\} = \\ &\int H(\mathbf{z}_t) \widetilde{W}_s(\mathbf{z}_s, \mathbf{v}) \widetilde{W}_t(\mathbf{z}_t, \mathbf{v}) f_t(\mathbf{z}_t, \mathbf{v}_0) d\mathbf{z}_t = \\ &= \int H(\mathbf{z}_t) \frac{f_s(\mathbf{z}_s, \mathbf{v}) f_t(\mathbf{z}_t, \mathbf{v})}{f_s(\mathbf{z}_s, \mathbf{v}_0)} d\mathbf{z}_t, \end{aligned} \quad (15)$$

where

$$H(\underline{z}_t) = L_s(\underline{z}_s)L_t(\underline{z}_t)I_{[\tau \geq t]}(\underline{z}_t)$$

is a nonnegative valued function. We also have that

$$f_s(\underline{z}_s, \mathbf{v}_0)^{-1} = c(\mathbf{v}_0)^{-1} \exp\{\sum_{i=1}^s \sum_{j=1}^n -v_{0j}t_j(\mathbf{z}_i)\}h_s(\underline{z}_s)^{-1},$$

where $h_s(\underline{z}_s) = \prod_{i=1}^s h(\mathbf{z}_i)$ and $c(\mathbf{v}_0)$ is the normalization factor,

$$c(\mathbf{v}_0)^{-1} = \int \exp\{\sum_{i=1}^s \sum_{j=1}^n v_{0j}t_j(\mathbf{x}_i)\}h_s(\underline{\mathbf{x}}_s)d\underline{\mathbf{x}}_s.$$

Putting all this together we obtain

$$\mathbb{E}_{\mathbf{v}_0}\{L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq t]}\} = \int \int A(\underline{z}_t, \underline{\mathbf{x}}_s, \mathbf{v}) \exp\{a(\underline{z}_t, \underline{\mathbf{x}}_s, \mathbf{v}_0)\} d\underline{z}_t d\underline{\mathbf{x}}_s,$$

where $A(\underline{z}_t, \underline{\mathbf{x}}_s, \mathbf{v})$ is a nonnegative valued function independent of \mathbf{v}_0 and $a(\underline{z}_t, \underline{\mathbf{x}}_s, \mathbf{v}_0)$ is a function linear in \mathbf{v}_0 for every fixed \underline{z}_t and $\underline{\mathbf{x}}_s$. It follows that the function inside the above integral is convex in \mathbf{v}_0 . Consequently the expectation $\mathbb{E}_{\mathbf{v}_0}\{L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq t]}\}$ is convex in \mathbf{v}_0 and hence the proof is complete. \square

Note that since $\nabla \widetilde{W}_t(\underline{z}_t, \mathbf{v}) = \widetilde{W}_t(\underline{z}_t, \mathbf{v}) \widetilde{S}_t^{(1)}(\underline{z}_t, \mathbf{v})$, where

$$\widetilde{S}_t^{(1)}(\underline{z}_t, \mathbf{v}) = \sum_{j=1}^t S^{(1)}(\mathbf{z}_j, \mathbf{v}),$$

the above convexity result can be extended to the trace of the covariance matrix of $\nabla \bar{\ell}_{1N}(\mathbf{v})$ as well. Note also that this convexity result holds for the particular case when the stopping time τ is fixed (non-stochastic).

Let us consider now the ratio estimator $\bar{\ell}_{1N}(\mathbf{v})/\bar{\ell}_{2N}(\mathbf{v})$ of $\ell(\mathbf{v})$ given in (8). By the Delta Theorem its asymptotic variance is N^{-1} times the variance

$$\begin{aligned} \sigma^2(\mathbf{v}, \mathbf{v}_0) &= \text{var}_{\mathbf{v}_0}\{a \sum_{t=1}^{\tau} L_t(\underline{\mathbf{Z}}_t) \widetilde{W}_t(\underline{\mathbf{Z}}_t, \mathbf{v}) \\ &\quad - b \sum_{t=1}^{\tau} \widetilde{W}_t(\underline{\mathbf{Z}}_t, \mathbf{v})\} \\ &= a^2 \text{var}_{\mathbf{v}_0}\{\sum_{t=1}^{\tau} M_t(\underline{\mathbf{Z}}_t) \widetilde{W}_t(\underline{\mathbf{Z}}_t, \mathbf{v})\}, \end{aligned} \tag{16}$$

where $a = 1/\ell_2(\mathbf{v})$, $b = \ell_1(\mathbf{v})/\ell_2^2(\mathbf{v})$ and $M_t = L_t - \ell_1(\mathbf{v})/\ell_2(\mathbf{v})$.

Unfortunately, Theorem 2.2 can not be directly extended to $\sigma^2(\mathbf{v}, \mathbf{v}_0)$ given in the form (16), because M_t are not nonnegative valued. Therefore, convexity of $\sigma^2(\mathbf{v}, \cdot)$ in the form (16) still remains an open problem.

3 OPTIMAL CHOICE OF THE REFERENCE PARAMETERS

The results of the previous section imply that the problem of minimization of the variance of the likelihood ratio estimators is a convex problem, provided $f(\mathbf{y}, \mathbf{v})$ comes from an exponential family given in the canonical form and the permissible parameter set V is convex. In this section we show that for some exponential families there is an order relation between v and the corresponding optimal reference value v_0^* .

Consider Gamma distributions with the corresponding density function

$$f(y, \lambda, \beta) = \frac{\lambda^\beta y^{\beta-1} e^{-\lambda y}}{\Gamma(\beta)}, \quad y > 0.$$

Note that this is an exponential family in the canonical form with two parameters $\lambda > 0$ and $\beta > 0$. We assume further that the parameter β is fixed and consider this family with respect to the parameter λ only.

Proposition 3.1 Consider an output process $L_t = L_t(\underline{Y}_t)$ driven by the input sequence Y_1, Y_2, \dots , of iid random variables distributed $\text{Gamma}(\lambda, \beta)$. Suppose that $L_t(\underline{y}_t)$, $t = 1, 2, \dots$, are nonnegative valued and monotonically increasing in every component y_i , $i = 1, \dots, t$, of \underline{y}_t , and that the indicator function $I_{[\tau \geq t]}(\underline{y}_t)$ is nondecreasing in every component of \underline{y}_t . Then there exists a unique minimizer λ_0^* of the function

$$\varphi(\lambda_0) = \mathbb{E}_{\lambda_0}\{[\sum_{t=1}^{\tau} L_t(\underline{Z}_t) \widetilde{W}_t(\underline{Z}_t, \lambda)]^2\} \tag{17}$$

which satisfies $\lambda_0^* < \lambda$.

Proof Let us first observe that it follows from the results of section 2 that the function $\varphi(\lambda_0)$ is convex. We show now that its derivative $d\varphi(\lambda_0)/d\lambda_0$ at $\lambda_0 = \lambda$ is positive. Since $\varphi(\lambda_0)$ is convex and tends to infinity as λ_0 approaches zero, this will imply that the minimizer λ_0^* exists and is less than λ . It is not difficult to verify here that in fact $\varphi(\lambda_0)$ is strictly convex and hence uniqueness of λ_0^* will follow.

By (15) we have that

$$\mathbb{E}_{\lambda_0}\{L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq t]}\} = \int H(\underline{z}_t) \frac{f_s(\underline{z}_s, \lambda) f_t(\underline{z}_t, \lambda)}{f_s(\underline{z}_s, \lambda_0)} d\underline{z}_t,$$

where

$$H(\underline{z}_t) = L_s(\underline{z}_s)L_t(\underline{z}_t)I_{[\tau \geq t]}(\underline{z}_t)$$

is monotonically increasing in every component of \underline{z}_t . It follows that

$$\begin{aligned} \frac{d}{d\lambda_0}[\mathbb{E}_{\lambda_0}\{L_s L_t \widetilde{W}_s \widetilde{W}_t I_{[\tau \geq t]}\}]|_{\lambda_0=\lambda} &= \\ &= \int H(\underline{z}_t) \widetilde{S}_s^{(1)}(\underline{z}_s, \lambda) f_t(\underline{z}_t, \lambda) d\underline{z}_t. \end{aligned} \tag{18}$$

Taking (14) into account, it is readily seen that in order to prove the positivity of $d\varphi(\lambda_0)/d\lambda_0$ at $\lambda_0 = \lambda$ it is sufficient to show that for $s \leq t$ the expression given in the right hand side of (18) is positive. To proceed, we write the right hand side of (18) in the form

$$\begin{aligned} & \Gamma(\beta)^{-t} \int (z_1 + \dots + z_s - s\beta\lambda^{-1}) H(\underline{z}_t) \times \\ & \lambda^{t\beta} e^{-(\lambda_1 z_1 + \dots + \lambda_t z_t)} \prod_{j=1}^t z_j^{\beta-1} dz_t = \\ & \lambda^{t\beta} \Gamma(\beta)^{-t} \sum_{i=1}^s \int (z_i - \beta\lambda^{-1}) \times \\ & H(\underline{z}_t) \prod_{j=1}^t (e^{-\lambda z_j} z_j^{\beta-1}) dz_t. \end{aligned}$$

Therefore it will be sufficient to show that each integral in the above sum is positive for $i = 1, \dots, s$. We prove this for $i = 1$. The proof for the remaining values of i is similar. We have

$$\begin{aligned} & \int (z_1 - \beta\lambda^{-1}) H(\underline{z}_t) \prod_{j=1}^t (e^{-\lambda z_j} z_j^{\beta-1}) dz_t = \\ & \int \prod_{j=2}^t (e^{-\lambda z_j} z_j^{\beta-1}) \left(\int_0^\infty (z_1 - \beta\lambda^{-1}) \times \right. \\ & \left. H(\underline{z}_t) e^{-\lambda z_1} z_1^{\beta-1} dz_1 \right) dz_2 \dots dz_t. \end{aligned}$$

Now, for fixed z_2, \dots, z_t , integration by parts gives us the inner integral in the form

$$\lambda^{-1} \int_0^\infty z_1^\beta e^{-\lambda z_1} dH(z_1, z_2, \dots, z_t).$$

Since for fixed z_2, \dots, z_t , the function $H(z_1, z_2, \dots, z_t)$ is monotonically increasing in z_1 , this integral is positive. Thus, the corresponding multiple integral is also positive and the proof is complete. \square

Note that the result of Proposition 3.1 holds in the particular case when τ is fixed (nonstochastic). In this case the indicator functions $I_{[\tau \geq t]}(\cdot)$ are constants and the condition that $I_{[\tau \geq t]}(\cdot)$ must be nondecreasing functions holds automatically.

Let, for example, L_t be the waiting time process in a $GI/G/c$ queue with c servers. The input sequence is given here by the two-dimensional random vectors $\mathbf{Y}_i = (Y_{1i}, Y_{2i})$, $i = 1, 2, \dots$, with independent components Y_{1i} and Y_{2i} , Y_{1i} being the service time of the i^{th} customer, Y_{2i} being the interarrival time between i^{th} and $(i-1)^{\text{th}}$ customers and τ being the number of customers served during the busy period. Suppose, for example, that Y_{1i} are distributed exponentially with $f(y, \lambda) = \lambda e^{-\lambda y}$, i.e. consider a $G/M/c$ queue with service rate λ . (Note that the exponential distribution is a particular case of Gamma distributions with $\beta = 1$.) Then conditions of Proposition 3.1 are satisfied and hence we obtain here that the optimal value λ_0^* of the reference parameter λ_0 , which minimizes function $\varphi(\lambda_0)$ defined in (17), is less than λ .

Note that if we assume in Proposition 3.1 that $L_t(\underline{y}_t)$ are nonnegative and monotonically decreasing

in every component of \underline{y}_t and $I_{[\tau \geq t]}(\underline{y}_t)$ are nonincreasing in every component of \underline{y}_t , then similar arguments show that the corresponding minimizer λ_0^* of $\varphi(\lambda_0)$ is greater than λ . Consider, for example, again the waiting time process in a $GI/G/c$ queue and suppose now that the interarrival times Y_{2i} are distributed exponentially, i.e. consider an $M/G/c$ queue with interarrival rate λ . Then the above assumptions hold and the optimal value λ_0^* of the reference parameter will be greater than λ .

As another example suppose now that the service times Y_{1i} and the interarrival times Y_{2i} are independent, each distributed Gamma with respective parameters λ_1, β_1 and λ_2, β_2 . Let β_1 and β_2 be fixed and consider the corresponding variance of $\bar{\ell}_{1N}(\mathbf{v})$, $\mathbf{v} = \boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, as a function of the reference value $\boldsymbol{\lambda}_0 = (\lambda_{01}, \lambda_{02})$ of the parameter vector $\boldsymbol{\lambda}$. We obtain then that this variance is a convex function of $\boldsymbol{\lambda}_0$ and the corresponding partial derivatives with respect to λ_{01} at $\lambda_{01} = \lambda_1$ are positive while the partial derivatives with respect to λ_{02} at $\lambda_{02} = \lambda_2$ are negative. This implies that $\lambda_{01}^* < \lambda_1$ and $\lambda_{02}^* > \lambda_2$.

It is possible to show, in a way similar to the proof of Proposition 3.1, that if the input sequence Y_1, Y_2, \dots , is distributed normal $N(\mu, \sigma^2)$ and the monotonicity assumptions of Proposition 3.1 hold, then the optimal values μ_0^*, σ_0^* of the reference parameters satisfy $\mu_0^* < \mu$ and $\sigma_0^* > \sigma$.

4 CONCLUSIONS

In the above likelihood ratio method (importance sampling scheme) choice of the reference value \mathbf{v}_0 of the parameter vector is very important. We showed that for a large class of distribution families the problem of minimization of the variance of the likelihood estimators is convex with respect to \mathbf{v}_0 . We also showed that in some cases there are order relations between the parameters and their optimal reference values. In those cases it is possible to give clear recommendations how the reference values should be chosen in relation to the considered values of the parameters.

Let us finally mention that the problem of minimization of the variance of the likelihood ratio estimators can be approached numerically by the likelihood ratio method. To do so consider for example the function $\phi(\mathbf{v}, \mathbf{v}_0)$ in (9). For a fixed \mathbf{v} its gradient with respect to \mathbf{v}_0 can be written in the form

$$\begin{aligned} & \nabla_{\mathbf{v}_0} \phi(\mathbf{v}, \mathbf{v}_0) = \\ & -\mathbb{E}_{\mathbf{v}_0} \{ L(\mathbf{Z})^2 W(\mathbf{Z}, \mathbf{v})^2 \mathbf{S}^{(1)}(\mathbf{Z}, \mathbf{v}_0) \}, \end{aligned}$$

provided the corresponding derivatives can be taken inside the expected value. In order to find a mini-

mizer \mathbf{v}_0^* of the function $\phi(\mathbf{v}, \cdot)$ it is sufficient to solve the associated system of equations $\nabla_{\mathbf{v}_0} \phi(\mathbf{v}, \mathbf{v}_0) = \mathbf{0}$. To estimate now \mathbf{v}_0^* from simulation one can consider the following stochastic equivalent (stochastic counterpart) of $\nabla_{\mathbf{v}_0} \phi(\mathbf{v}, \mathbf{v}_0) = \mathbf{0}$,

$$\sum_{i=1}^N L(\mathbf{Z}_i)^2 W(\mathbf{Z}_i, \mathbf{v})^2 \frac{\nabla f(\mathbf{Z}_i, \mathbf{v}_0)}{f(\mathbf{Z}_i, \bar{\mathbf{v}})} = \mathbf{0}. \quad (19)$$

Here $\mathbf{Z}_1, \mathbf{Z}_2, \dots$, is a sample from $f(\mathbf{z}, \bar{\mathbf{v}})$, where $\bar{\mathbf{v}}$ is a chosen value of the parameter vector. The solution of the system of equations (19) will provide an estimator of \mathbf{v}_0^* .

REFERENCES

- Asmussen, S. and Rubinstein, R.Y. 1992. The efficiency and heavy traffic properties of the score function method in sensitivity analysis of queueing models. *Advances of Applied Probability*, 24:172-201.
- Glynn, P.W., and Iglehart, D.L. 1989. Importance sampling for stochastic simulations. *Management Sci.* 35:1367-1392.
- Rao, C.R. 1973. *Linear Statistical Inference and its Applications*. New York: John Wiley and Sons.
- Ross, S.M. 1983. *Stochastic Processes*. New York: John Wiley and Sons.
- Rubinstein, R.Y. and Shapiro, A. 1992. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. New York: John Wiley and Sons. To appear.

AUTHOR BIOGRAPHIES

REUVEN Y. RUBINSTEIN is a Professor in the Faculty of Industrial Engineering and Management at the Technion-Israel Institute of Technology. He received a Ph.D. degree in Operations Research from Rigas Politechnical Institute (USSR). His research interests are focused on stochastic modeling and optimization, applied probability and discrete event systems.

ALEXANDER SHAPIRO is an Associate Professor in the Department of Industrial and Systems Engineering at Georgia Institute of Technology. He received a Ph.D. degree in applied mathematics from the Ben-Gurion University (Israel) in 1981. His research interests are focused on stochastic programming, sensitivity analysis of mathematical programs and multivariate statistical analysis.