EXTREMALS OF FUNCTIONS OF RANDOM VARIABLES, OPTIMAL CONTROL THEORY, AND TELETRAFFIC THEORY

Jack M. HOLTZMAN
AT&T Bell Laboratories
Holmdel, New Jersey 07733, USA

The purpose of a previous paper was to see to what extent the apparent success of the equivalent random method was due to constraints placed on blocking probabilities by the mean and variance used by the method. (The mean and variance used by the equivalent random method places stronger constraints on GI/M/N loss system blocking probabilities than the mean and variance of interarrival times). It turned out that the apparent success of the method was not well explained by those constraints but rather by other structure of the overflow processes typically analyzed by the equivalent random method. We return to this problem here by adding a constraint of unimodality to the equivalent random mean and variance. Such a shape constraint might be expected to be quite strong. It is shown that the effect of unimodality is less than expected.

These results are obtained by deriving extrema of functions of random variables subject to constraints. We elaborate on some of these methods, in particular, the Karlin-Isii (KI) Theorem. The KI Theorem is related to optimal control theory, and some other optimization methods are mentioned.

1. INTRODUCTION

The purpose of [1] was to see to what extent the apparent success of the equivalent random method was due to constraints placed on blocking probabilities by the mean and variance used by the method. (The mean and variance used by the equivalent random method places stronger constraints on GI/M/N loss system blocking probabilities than the mean and variance of interarrival times - [2], p.140). It turned out that the apparent success of the method was not well explained by those constraints but rather by other structure of the overflow processes typically analyzed by the equivalent random method. We return to this problem here by adding a constraint of unimodality to the equivalent random mean and variance.

However, our objective is not merely to revisit the equivalent random method but to elaborate on some methodology. Reference 1 used some methods of optimization described in [3], in particular Theorem 2.1 on page 472 (henceforth called the KI Theorem for Karlin and Isii). Given the fact that the KI Theorem may not be familiar to many in teletraffic theory, it seems appropriate to discuss these methods. It will also be instructive to relate the KI Theorem to optimal control theory. Furthermore, there are a host of other relevant optimization methods in the recent book [4] -- we shall mention one aspect of them.

2. THE KI THEOREM

The KI Theorem solves the problem of determining sharp upper and lower bounds of the integral \( \int_T \Omega(t) dF(t) \) subject to the constraints

\[ c_i^0 = \int_T u_i(t) dF(t), \quad i = 0, 1, ..., n, \]  

where \( F \) is a distribution function, \( T \) is a subset of the real line, the \( u_i \) are real-valued Borel measurable functions and \( \Omega \) is continuous. It is also assumed that the \( u_i \) are linearly independent over \( T \). Define

\[ c^0 = \left[ c_0^0, c_1^0, ..., c_n^0 \right]. \]
\[ V(c^0) = \left\{ F \mid \int_T u_i(t) dF(t) = c_i^0, i = 0, \ldots, n \right\} \]  

\[ I_{\text{max}} = \sup_{F \in V(c^0)} \int_T \Omega(t) dF(T) \]  

\[ I_{\text{min}} = \inf_{F \in V(c^0)} \int_T \Omega(t) dF(T) \]  

\[ P_+ = \left\{ u(t) = \sum_{i=0}^{n} a_i u_i(t) \mid u(t) \geq \Omega(t), t \in T \right\} \]  

\[ P_- = \left\{ u(t) = \sum_{i=0}^{n} a_i u_i(t) \mid u(t) \leq \Omega(t), t \in T \right\} \]  

\[ P_+ \text{ and } P_- \text{ are sets of linear combinations of } u_i \text{'s which bound } \Omega \text{ from above and below, respectively.} \]

The KI Theorem (I3, p. 472) is as follows.

**KI Theorem.** Let \( c^0 \) be an interior point of the \((n+1)\)-dimensional cone

\[ M_{n+1} = \left\{ c = (c_0, \ldots, c_n) \mid c_i = \int_T u_i(t) dF(t), i = 0, 1, \ldots, n, F \text{ a dist. fn.} \right\} \]

and let \( \Omega \) be such that \( P_+ \) and \( P_- \) are nonvoid. Then

\[ I_{\text{max}} = \inf_{i=0}^{n} \sum_{i=0}^{n} a_i c_i^0 \]  

\[ I_{\text{min}} = \sup_{i=0}^{n} \sum_{i=0}^{n} a_i c_i^0 \]  

where the inf and sup are extended over all polynomials \( u = \sum_{i=0}^{n} a_i u_i \) contained in \( P_+ \) and \( P_- \), respectively. Moreover, the inf and sup are attained in \((II-9)\) and \((II-10)\).

The basic idea of the KI Theorem, which may go back to Markov, is as follows. If \( u \in P_+ \), then

\[ \int \Omega(t) \ dF(t) \leq \int u(t) \ dF(t); \]

hence

\[ \int \Omega(t) \ dF(t) \leq \inf_{u \in P_+} \int u(t) \ dF(t) \]

\[ = \inf_{i=0}^{n} \sum_{i=0}^{n} a_i c_i^0 \]  

as in \((II-9)\). The recent contribution of the KI Theorem is the sharpness of the inequality under appropriate conditions. Examples of the application of the KI Theorem are given in chapter XII of [3].

**Remark II.1** It may be noted that the KI Theorem is more general than is needed for many problems. In many cases, the constraint functions and performance criterion function form a complete Tchebycheff system for which available results can be readily applied. See [2] for relevant applications.

We repeat an important remark on the KI Theorem from p. 474 of [3]:

**Remark II.2** If there exists an extremal distribution function \( F^* \) such that \( I_{\text{max}} = \int \Omega dF^* \), then the spectrum of \( F^* \) is confined to the set \( S \subset T \):

\[ S = \left\{ t \mid \sum_{i=0}^{n} a_i u_i(t) = \Omega(t) \right\} \]

This is important in application of the KI Theorem -- in many practical cases \( S \) contains only a finite number of points (sometimes only a few). We shall relate this remark to optimal control theory in the next section.

5.1B.1.2
3. OPTIMAL CONTROL THEORY

The Pontryagin maximum principle ([5]) is relevant to maximizing a performance criterion subject to constraints including constraints defined by differential equations. The maximizing problem of Section II with \( T = [0, T_t] \) can be rewritten as maximizing \( x_{n+1}(T_t) \) subject to

\[
\begin{align*}
\dot{x}_i &= u_i(t)f(t), \quad i = 0, 1, ..., n \\
\dot{x}_{n+1} &= \Omega(t)f(t) \\
x_i(0) &= 0, \quad i = 0, 1, ..., n+1 \\
x_i(T_t) &= c_i^0, \quad i = 0, 1, ..., n \\
u_0(t) &= 1, \quad t \in T_t \\
c_0^0 &= 1
\end{align*}
\]

In this problem formulation, we have assumed that \( F \) has a density function \( f \), which plays the role of the control function. In optimal control theory, the control function is often denoted by \( u \). In Section II, we already adopted the terminology of [2] where the \( u's \) define the constraints.

Remark III.1 The maximum principle ([5], pp. 26-28) is a necessary condition for the control function \( u \) to maximize the performance criterion subject to the constraints. Observe that the maximum principle is only a necessary condition while the KI theorem gives a sharp inequality on the performance criterion. On the other hand, the maximum principle applies to nonlinear systems while the KI theorem is for a linear system (linear in the control). There are specialized optimal control theory results for linear systems (see Chap. 4, [6]) but it is instructive to consider the maximum principle here for these reasons:

(i) Examination of the proof of the KI theorem (with convex cones, supporting hyperplanes) is very reminiscent of the proof of the maximum principle. Apparently, the derivations were done independently. (There is also some commonality with the linear system results of [6]).

(ii) Comparing the forms of the optimality conditions is instructive.

(iii) There may be some problems for which the generality of optimal control theory via the maximum principle may be applicable.

Applying the maximum principle to the problem stated in the first paragraph of this section leads to maximizing a Hamiltonian which is linear in the control \( f \). Since \( f \) is unbounded from above, one is faced with a problem with application of the maximum principle. In fact, the form of the Hamiltonian suggests impulses in the control. We thus turn our direction to an extended maximum principle which allows impulses in the control.

Reference 7 treats the problem of choosing a vector control function \( v(t) \) with values in a set \( V \) and a control measure \( \mu \) (a positive Radon measure) so that \( x(t) \) (an \( n \)-vector) is a solution of the equation

\[
\frac{dx}{dt} = f(t, x, v) + g(t, v) \mu
\]

with constraints on initial and terminal states, and the performance criterion

\[
\int_0^{t_1} \int_0^{t_1} f_0(s, x(s), v(s))ds + \int_0^{t_1} g_0(s) \mu(ds)
\]

is minimized. When \( \mu \) is the measure corresponding to the familiar delta function concentrated at \( t, t \in [t_0, t_1] \), we obtain the familiar

\[
\int_0^{t_1} g(s, v(s)) \mu(ds) = g(t, v(t))
\]

The approach of [7] is to convert the above problem into an ordinary problem to which the standard maximum principle applies. Necessary conditions for \( v(t) \) and \( \mu \) to be an optimal control law are given in [7]. We shall focus here on the case where

\[
\begin{align*}
f(t, x, v) &= 0, \\
g(t, v) &= g(t)
\end{align*}
\]
We repeat a remark from (7):

Remark III.2 The necessary conditions imply that

\[
\max_{u \in U} \sum_{j=0}^{n} g_j(t, u) \Phi_j(t) \leq 0
\]  

(almost everywhere with respect to both Lebesgue and \( \mu \) measure, and if \( A \) is a subset of

\[
\left\{ t : \max_{u \in U} \sum_{j=0}^{n} g_j(t, u) \Phi_j(t) < 0 \right\}
\]  

that \( \mu(A) = 0 \). This implies that (III-13) must vanish at each impulse of the measure \( \mu \). If (III-13) as a function of \( t \) on \([t_0, t_1]\) assumes a maximum of zero only a finite number of times, the measure \( \mu \) must consist of a finite number of impulses located at these maximum points. In our case, \( g(t, \nu) = g(t) \) and the functions \( \phi_j(t) \) are constants. Then, this remark is comparable with Remark II.2.

4. OTHER TECHNIQUES

The KI Theorem given in Section II is a fundamental result in determining extrema of functions of random variables. There are, however, a number of other results that should be kept in mind. Ref. 4 presents many of these techniques, along with many references. We shall discuss one of them, based on what is called "convex ordering," a terminology that will become motivated in what follows. We first give a few definitions (pp. 4, 8 of [4]).

Definition IV.1 The random variables \( X \) and \( Y \) satisfy the partial ordering \( \leq_p \) if their respective distribution functions \( F \) and \( G \) satisfy

\[
F(x) \geq G(x) \quad (\text{all real } x)
\]  

(\( X \) is said to be stochastically smaller or smaller in distribution than \( Y \)).

Definition IV.2 The random variables \( X \) and \( Y \) satisfy the ordering \( \leq_c \), called the convex ordering relation, if

\[
E \max (X, Y) \leq E \max (x, Y) \quad (\text{all real } x)
\]  

Convex ordering is much weaker than the stochastic ordering and we shall pursue it some here. A key result is ([4], p.9):

Theorem IV.1 The inequality

\[
\int_{-\infty}^{\infty} f(t) dF_1(t) \leq \int_{-\infty}^{\infty} f(t) dF_2(t)
\]  

holds for all monotonic functions \( f \) (for which the integrals are defined) iff \( F_1 \leq F_2 \). For given \( f \), (IV-3) holds for all \( F_1 \) and \( F_2 \) for which \( F_1 \leq F_2 \) only if \( f \) is non-decreasing and convex. If \( F_1 \leq F_2 \) and their means exist and are equal, then (IV-3) holds for all convex \( f \).

Theorem IV.1, which motivates the notation "convex ordering," is very useful in the following manner. Let \( M \) be a set of distribution functions. It is then useful to find the supremum \( F_{\sup} \) of \( M \) with respect to \( \leq_c \) defined as the smallest \( F \) satisfying

\[
F \leq F_{\sup} \quad \text{all } F \in M
\]  

\((F_{\inf} \) is analogously defined). It is important to note that \( F_{\sup} \) is not necessarily in \( M \). If it is, it is called \( F_{\max} \).

In [8], the KI Theorem is used to provide a nice formalism for determining extremal elements with respect to convex ordering. It may be noted that while the constraints in [8] are only on the moments, the formalism can be extended to more general constraints. Also, unimodality is treated. It should be kept in mind that the inequalities using convex ordering are not necessarily sharp.

5. THE EQUIVALENT RANDOM METHOD AND UNIMODALITY

The equivalent random method first determines the mean \( M \) and variance \( V \) of the number of servers that would be occupied if the traffic were offered to an infinite server group. Then an overflow process with the same \( M \) and \( V \) is offered to the finite server group and its blocking calculated. This blocking is taken as the approximation for the blocking seen by the original traffic.

5.1B.1.4
Consider a nonlattice renewal process, with distribution function \( F \) for the interarrival times, offered to a group of \( N \) trunks. The holding times are mutually independent exponentially distributed random variables with unity mean (or the mean is the time unit). Blocked calls are cleared and the system is in equilibrium.

Define

\[
m = \int_0^\infty t dF(t), \tag{V-1}
\]

\[
\phi(s) = \int_0^\infty e^{-st} dF(t). \tag{V-2}
\]

Then it is known that the blocking probability is

\[
B = \left\{ 1 + \left[ \sum_{i=1}^{N} \frac{1 - \phi(1)}{\phi(1)} \right] + \ldots + \left[ \frac{1 - \phi(1)}{\phi(1)} \frac{1 - \phi(2)}{\phi(2)} \ldots \frac{1 - \phi(N)}{\phi(N)} \right] \right\}^{-1}, \tag{V-3}
\]

(see, e.g., [9], Chapter 4). Observe that \( B \) depends on \( N \) values of \( \phi(i) \), \( i = 1, \ldots, N \), and that it is an increasing function of these \( \phi(i) \). We have the following relationships:

\[
M = m^{-1}, \tag{V-4}
\]

\[
V = M \left[ \frac{1}{1 - \phi(1)} - M \right]. \tag{V-5}
\]

Thus, \((M, V)\) uniquely determines \((m, \phi(1))\) and vice versa. Specifically,

\[
\phi(1) = \frac{V/M - 1 + M}{V/M + M}. \tag{V-6}
\]

Hence, the equivalent random method fixes \( \phi(1) \) which is particularly important in (V-3). Furthermore, there are additional implicit restraints on the other \( \phi(i) \), as shown in [11.

In [1] it was shown that the equivalent random mean and variance placed only weak constraints on the blocking in a \( GI/M/N \) loss system (but still much stronger than the mean alone and stronger than the mean and variance of interarrival times). It was suggested that additional structure should be taken into account to explain the apparent success of the equivalent random method in practice. We examine the effect of adding the constraint of unimodality. Such a shape constraint might be expected be quite strong.

\( F \) is now assumed to be absolutely continuous with a unimodal density function. Assume that \( F \) has a unimodal density function \( f \) with derivative satisfying

\[
f(t) \geq 0 \quad t < t_m \tag{V-7}
\]

\[
f(t) \leq 0 \quad t > t_m \tag{V-8}
\]

Consider the following optimization problem. Find the

\[
\sup \text{ (or inf) } \text{ of } \int_0^\infty e^{-st} f(t) dt \quad (s > 1)
\]

subject to

\[
\int_0^\infty f(t) dt = 1 \tag{V-9}
\]

\[
\int_0^\infty tf(t) dt = m \tag{V-10}
\]

\[
\int_0^\infty e^{-t} f(t) dt = \phi(1) \tag{V-11}
\]

\[
f(t) = \int_0^1 u(y) dy \tag{V-12}
\]

\[
u(t) \geq 0, \quad t < t_m \tag{V-13}
\]

\[
u(t) \leq 0, \quad t > t_m \tag{V-14}
\]

The control function is the derivative of the density function. Restating our problem in the formulation of Section III, we have:
\[ \dot{x}_0 = e^{-s} x_1 \quad (s > 1) \]  
\[ \dot{x}_1 = \text{sign}(t - t_0) \mu \]  
\[ \dot{x}_2 = x_1 \]  
\[ \dot{x}_3 = t x_1 \]  
\[ \dot{x}_4 = e^{-s} x_1 \]  
\[ x_0(0) = x_1(0) = x_2(0) = x_3(0) = x_4(0) = 0 \]  
\[ x_2(\infty) = 1, x_3(\infty) = m, x_4(\infty) = \phi(1) \]  
\[ \text{performance criterion} = \int_0^\infty e^{-s} x_1(t) dt \]

We apply the extended maximum principle of [7].

Remarks:

(i) As mentioned, the maximum principle is a necessary condition. We do not try to verify sufficiency here. Nor do we try to rigorously fill in some technical details (e.g., \( g(t, u) \) is assumed differentiable with respect to \( t \) in [7] and ours is not; a finite time interval is assumed in [7]). We use the maximum principle to suggest candidate optimal functions. With the candidate optimal functions to be displayed, we are able to make our point about how strong is the unimodality constraint.

(ii) Although \( x_1(0) \) is assumed zero, \( x_1(t) \) can have a jump at \( 0^+ \) (it is right continuous).

(iii) The constancy condition of [7] is trivially satisfied.

Thus we use the extended maximum principle manner to suggest candidate optimal density functions, which are shown in Figure 1. The problem reduces to finding the sup or inf of

\[ \frac{c(1-e^{-at})}{s} + \frac{(c+d)(e^{-at} - e^{-t(a+b)})}{s} \]

subject to

\[ ac + (c+d)b = 1, \]
\[ \frac{a^2}{2} + \frac{(c+d)(2ab + b^2)}{2} = m, \]
\[ c(1-e^{-at}) + (c+d)(e^{-at} - e^{-t(a+b)}) = \phi(1) \]
\[ a \geq 0, b \geq 0, c \geq 0, c + d \geq 0. \]

Figures 2 and 3 show some numerical results for \( z = 2 \) and \( 4 \) for \( N = 10 \) servers. It is seen that while the range of blocking probabilities is narrowed from the results of [1] (particularly the lower bound) there is still a rather wide range. Thus, the practical success of the equivalent random method depends on further structure. Observe that adding continuity to unimodality does not help because there are approximating sequences of continuous functions to the density functions of Figure 1. It was shown in Reference 10 how log-convexity, a stronger constraint than unimodality, narrows the range of delays for delay systems. It would be interesting to see how bounds on the derivative of the density function would affect the results.

6. CONCLUDING REMARKS

The value of analyses such as in Section V is that they help to improve intuition about why approximations work well or not. Reference 1 showed that the equivalent random mean and variance were not as constraining as might be thought by showing what "weird" distribution functions can satisfy the mean/variance conditions. Here, unimodality, which might be thought to introduce an additional strong constraint of the shape of the density functions, does narrow the range of blocking probabilities but still leaves a rather wide range. Unimodality still allows a fair degree of freedom with respect to the step functions satisfying the mean/variance conditions.

Optimal control theory is an overkill for the types of problems treated by the KI Theorem and yields less; e.g., the result quoted was only a necessary condition. However, it was felt to be instructive to show the relationship. Furthermore, the optimal control result was used to suggest candidates optimal functions for the unimodal case in Section V (although it is felt that a more direct approach should be possible)\(^8\). And finally,
we already remarked that there may be cases in which the additional generality of the optimal control formulation may be applicable (for extremely wide ranging generality, see [11]).

ACKNOWLEDGEMENT

I thank H. Heffes and W. Whitt for their comments.

FOOTNOTES

1. The KI Theorem actually applies to more general functions than distribution functions. With our focus on distribution functions, the first constraint will have \( c^0 = 1, u(t) = 1 \).
2. See Sect. 4.5 of [6] for a perspective on those linear control results.
3. In [7], the control is \( u \).
4. It is a partial ordering on the set of all distribution functions with finite means.
5. If \( F < G \) and \( E \max(0,Y) < \infty \), then \( F \preceq G \).
6. The peakedness \( z \) is the ratio of \( V \) to \( M \) from (V-4) and (V-5).
7. The upper and lower bounds with unimodality displayed in Figures 2 and 3 are not exact since they were obtained via a numerical procedure with a stopping condition.
8. In [2], Chap. XII, Sect. 4, some unimodal problems are treated; also see [8] for the use of unimodality with convex ordering.

REFERENCES


![Figure 1 - Candidate Density Functions](image-url)
$B_{er}$ = EQUIVALENT RANDOM BLOCKING PROBABILITY
$B_u, B_l$: UPPER AND LOWER BOUNDS FROM [1]
$B_u^{ii}, B_l^{ii}$: UPPER AND LOWER BOUNDS WITH UNIMODALITY

FIGURE 2 - BOUNDS ON BLOCKING PROBABILITIES
FOR $z=2, N=10$

$B_{er}$ = EQUIVALENT RANDOM BLOCKING PROBABILITY
$B_u, B_l$: UPPER AND LOWER BOUNDS FROM [1]
$B_u^{ii}, B_l^{ii}$: UPPER AND LOWER BOUNDS WITH UNIMODALITY

FIGURE 3 - BOUNDS ON BLOCKING PROBABILITIES
FOR $z=4, N=10$