NEW RESULTS ON OPTIMAL RULES FOR SELECTING MONOTONE SUBSEQUENCES OF MAXIMAL LENGTH.

F. THOMAS BRUSS and FREDDY DELBAEN

Université Libre de Bruxelles and Eidgenössische Technische Hochschule Zürich

Summary: This article presents new results on the problem of selecting (on-line) a monotone subsequence of maximum expected length from a sequence of i.i.d. random variables. We study the case where the variables are observed sequentially at the occurence times of a Poisson process with known rate. Our approach is a detailed study of the integral equation which determines $v(t)$, the maximum expected number of selected points L^t up to time t. We first show that $v(t), v'(t)$ and $v''(t)$ exist everywhere on $\mathbb{R}^+$. Then, in particular, we prove that $v''(t) < 0$ for all $t \in [0, \infty[$, implying that v is strictly concave on $\mathbb{R}^+$. This settles a conjecture of Gnedin and opens the way to stronger bounds for v and its derivatives. We can show that $v'(t)\sqrt{2t} \sim 1$ and, in particular, that

$$\sqrt{2t} - \log(1 + \sqrt{2t}) + c < v(t) < \sqrt{2t}.$$

Also, using a martingale approach, we show

$$\frac{1}{3}\, v(t) \leq \mathrm{Var}(L^t) \leq \frac{1}{3}\, v(t) + c_1 \log(t) + c_2.$$

Further we obtain several results on the process $(L^t_u)_{0 \leq u \leq t}$, where L^t_u denotes the number of selected points up to time u when applying the optimal rule with respect to time t.

Due to the on-line requirement of selection, we are also interested quick selection rules and their performance, and so we study the class of convenient *graph*-rules. Results by Deuschel and Zeitouni on the concentration measure of record values suggest that the asymptotically best graph rule should be the "diagonal line rule", and we prove this intuition to be correct. Our last short section compares the performance of the optimal rule and the optimal graph rule.

Keywords: Poisson process, on-line selection, patience sorting, Ulam's problem, optimality principle, asymptotic optimality, integral equation, concavity, martingale, squared variation, predictable process, f-record rules, concentration measure, Abelian theorem.

AMS 1991 subject classification: 60G40

1. Introduction. Let $(N(s))_{s\geq 0}$ be a Poisson process with rate 1 on the interval $[0,t]$ and occurrence times $T_1 < T_2 < \cdots$ a.s. Set $T_0 = 0$. Further let $(X_k)_{k=1,2,\cdots}$ be a sequence of i.i.d. uniform random variables on $[0,1]$, independent of the T_j's. We suppose that the bivariate variables $(T_k, X_k)_{k=1,2,\cdots}$ can be observed sequentially. The objective is to select "on-line" a subsequence $(T_{k_1}, X_{k_1}), (T_{k_2}, X_{k_2}), \cdots$ of maximal length, satisfying $X_{k_1} \leq X_{k_2} \leq \cdots$ for $k_1 < k_2 < \cdots$. Here *on-line* means in sequential order, without recall on preceding observations.

This problem is closely related with the problems studied by Samuels and Steele (1981), and Gnedin (1998), of selecting on-line a monotone subsequence of maximal length. Dropping the on-line requirement leads to quite a different problem as we shall shortly outline:

Conditioning on $N(t) = n$, the natural question is now: What is the distribution of the length of the longest subsequence in a random permutation of n different real numbers? This question, seemingly first asked by Ulam (see Critchlow (1988) and Ulam (1972)), has attracted a great deal of scientific attention. A nice way of getting acquainted with this problem is reading the recent survey article by Aldous and Diaconis (1999), which shows the major steps provided by Hammersley (1972), Verȿik and Kerov (1977), Logan and Shepp (1977) and the recent detailed study of Baik, Deift and Johansson (1999) and Baik and Rains(1999), and also draws attention to interesting analogies with sorting problems and other problems.

To contrast the on-line problem of maximizing the expected length of the selected subsequence with the above problem we recall the "prophet" comparison of Samuels and Steele (1981): A prophet with complete foresight of the sequence achieves the maximal length, whereas a sequential decision maker without foreseeing abilities must be satisfied with what an optimal selection strategy can achieve. "Patience sorting" (see Aldous and Diaconis) provides an interesting complementary comparison, because that algorithm *is* on-line. Patience sorting, of n mixed cards, say, requires at each step a comparison of the currently held card with the top cards on the existing piles (from left to right). Interestingly, this algorithm determines (by touching each card once) the length which a prophet is able to achieve (see Aldous and Diaconis, Lemma 1.) However, it does note produce a pile containing this longest subsequence. For our on-line problem in this setting, we are confined to work with two piles: A left pile of discarded cards without specified structure, and a right pile of selected increasing cards which we would like to make as big as possible. The top card of the right pile is now the only legal point of reference, and the goal is to produce the largest possible pile of increasing cards by touching each card exactly once. (In "on-line" life we cannot insert new items into the history, as the patience sorting algorithm does.)

Our Poisson model may be seen as a card game in which we receive cards at the occurrence times of the Poisson process, and the kth card shows X_k, where the $X_1, X_2, \cdots$ are i.i.d. The time is fixed, and the total number of cards is now random. If the X_j's follow some continuous distribution function F, then the transformation $(X'_j) := (F(X_j))$ brings us back to the uniform case, so that solving one of these problems solves the general problem of selecting monotone subsequences from a sample of i.i.d. continuous random variables. The problem of selecting *decreasing* X_{k_j}'s is of course equivalent, since X_{k_j} and $1-X_{k_j}$ follow the same distribution.

We became interested in the problem of selecting monotone subsequences when Gnedin (private communication,1997) asked the following analytical question:

Gnedin's problem: Let $v : \mathbb{R}^+ \to \mathbb{R}^+$ be formally defined by

$$v(0) = 0; \quad v'(t) = \int_0^1 [v(tx) + 1 - v(t)]^+ dx, \tag{1.1}$$

where $y^+ = \max\{0, y\}$. Is $v(t)$ concave on $\mathbb{R}^+$?

We shall show in this paper that the answer to this question is affirmative. As Gnedin noticed before, this question of concavity is different from the question studied by Samuels and Steele (1981) and needs a different approach. Moreover, we shall study other questions concerning (1.1) in detail and point out properties of v which seem important to us. These are, in particular, the existence and the behavior of the derivative $v'(t)$ and tight bounds for the solution v. Apart from mathematical curiosity our motivation herefore is the interpretation of v in the context of the problem: The solution of equation (1.1) determines in fact the value $v(t)$ of the problem, i.e. the maximum expected length of the selected subsequence with increasing X_{k_j}.

Also, equation (1.1) can be adapted to similar problems. For example, if we replace the sequence (X_k) by a i.i.d. random vectors Y_k uniform on the unit m-cube $[0,1]^m$ then the optimal selection equation for increasing (in all components) subsequences Y_{k_j} is the same in t if integration is replaced by multiple integration. This problem was studied by Baryshnikov and Gnedin (1998). Such links with similar problems add to the motivation to get the maximum information from (1.1) itself without having to depend on specific information from the probabilistic model.

The paper is organized as follows.

Section 2. is devoted to the study of $v(t)$, to the form of the optimal selection rule producing L^t, and to the study of fluctuations of L^t. The major steps are the following: v is continuous and increasing (Lemma 2.2), continuously differentiable (Theorem 2.1) and, in particular, v is a strictly concave function on R^+ (Theorem 2.2). New important bounds for v, v' and v'' as well as for an intrinsically connected function (called $\phi(t)$) are provided in Theorems 2.3 through 2.5.

We then turn to the problem of fluctuations of L^t. Theorem 2.6 establishes a uniform lower bound for the variance of L^t. Lemma 2.5 is used to find a close upper bound given by Theorem 2.7.

Section 3 addresses the "on-line" requirement of the problem by trying to find the best *simple* rule. Here a simple rule is defined as the rule to select "greedily" any selectable observation which is below the graph of a deterministic function f (*graph-rule*). The main result (Theorem 3.1) shows that the choice of $f(s) = s/t$ is asymptotically optimal in the class of graph-rules. This confirms the authors' conjecture instigated by results of Deuschel and Zeitouni (1995), on the concentration of measure (see also Goldie and Resnick (1996)). We also show that the asymptotic value of this rule is $\sqrt{t\pi/2}$.

The short Section 4 collects some conclusions from the comparison between graph-rules and asymptotically optimal rules, and draws attention to an Open Problem.

2. Form and Behavior of the Optimal Rule.

Recall that $(N(t))$ is a Poisson process on $\mathbb{R}^+$ of rate 1 with occurrence times $0 = T_0 < T_1 < T_2, \cdots$ a.s., and (X_k) a sequence of i.i.d. r.v.'s uniform on $[0,1]$, independent of the T_k's. Let $(\Omega, \mathcal{A}, P)$ be a probability space large enough to host $(T_k)_{k=1,2,\cdots}$ and $(X_k)_{k=1,2,\cdots}$, and let $\mathcal{A}_k = \sigma(T_1, \cdots, T_k; X_1, \cdots, X_k)$ denote the sigma-field generated by $T_1, T_2, \cdots, T_k$ and $X_1, X_2, \cdots, X_k$ simultaneously. The history of the process at time t is denoted by $\mathcal{F}_t$, i.e. $\mathcal{F}_t$ is the collection of sets which satisfies, for $t \geq 0$, $\mathcal{F}_t \cap \{N(t) = k\} = \mathcal{A}_k \cap \{N(t) = k\}$ for all $k = 0, 1, \cdots$.

Definition 2.1 A strategy is a function $\psi : \mathbb{N} \times \Omega \to \{0,1\}$ such that, for all n, $\psi(n, \cdot)$ is $\mathcal{A}_n$-measurable. We say that ψ is *acceptable* for on-line decisions if $n \leq m$ implies that $X_n \leq X_m$ on the set $\{\psi(n,\omega) = 1\} \cap \{\psi(m,\omega) = 1\}$. The set of acceptable strategies will be denoted by Ψ_a. For ease of notation we write ψ_n for $\psi(n, \cdot)$.

Let

$$Z_t(\psi) = \sum_{T_n \leq t} \psi_n; \quad v(t) = \sup_{\psi \in \Psi_a} E(Z_t(\psi)). \tag{2.1}$$

More generally we define

$$v(t,x) = \sup_{\psi\in\Psi_a(x)} E(Z_t(\psi)), \tag{2.2}$$

where $\Psi_a(x)$ denotes the class of acceptable strategies confined to selections of values $X_{k_j} \geq x$. Note that, by definition of v, $v(t) = v(t,0)$, and by definition of our on-line problem, $\psi_n = \psi_m = 1$ implies $X_m \geq X_n$ for $m \geq n$. Thus, given $X_n = x$ and $\psi_n = 1$, the tail sequence $(\psi_m)_{m\geq n}$ must automatically belong to $\Psi_a(x)$. This will be used throughout.

Lemma 2.1 For all $t \geq 0$ and $0 \leq x \leq 1$, $v(t,x) = v((1-x)t)$.

Proof. Let $X'_n = X_n \mathbf{1}_{\{X_n\geq x\}}$. Remove all T_n with $X'_n = 0$, and renumerate the remaining T_n's by $T'_1, T'_2, \cdots$. The T'_n's are now the occurrence times of a thinned process $(N'(t))_{t\geq 0}$. Since the X_n are i.i.d. on $[0,1]$, this thinning is independent-binomial, so that $(N'(t))_{t\geq 0}$ is again a Poisson process, but now with rate $1-x$. But this determines uniquely the distribution of the process $(N'(t))_{t\geq 0}$, which is the same as the distribution of the process $(N(t(1-x)))_{t\geq 0}$. Since $\Psi_a(x)$ satisfies, by definition, the acceptability property, (2.2) becomes (2.1) with t being replaced by $t(1-x)$, and the Lemma is proved. ■

Lemma 2.2 The optimal value $v(t)$ is increasing and continuous on $\mathbb{R}^+$. (It is Lipschitz with constant 1.)

Proof. Since, for any strategy ψ, all ψ_n are non-negative, and since $N(t)$ is increasing in t, the expected value of the sum in (2.1) is increasing for any $\psi \in \Psi_a$. Hence $v(t)$ is increasing in t. To prove the continuity of $v(t)$, we will show that $|v(t+\delta) - v(t)| \leq |\delta|$. We suppose, w. l. o. g., $\delta \geq 0$. Choose $\epsilon > 0$ and let $\tilde{v}(t)$ denote the expected number of X_j being selected up to time t under a strategy $\psi_\epsilon \in \Psi_a$, which is ϵ-optimal with respect to the time $t+\delta$. Then,

$$\begin{aligned} v(t+\delta) - v(t) &= (v(t+\delta) - \tilde{v}(t+\delta)) + (\tilde{v}(t+\delta) - \tilde{v}(t)) \\ &\quad + (\tilde{v}(t) - v(t)) \\ &\leq \epsilon + |\tilde{v}(t+\delta) - \tilde{v}(t)|, \end{aligned} \tag{2.3}$$

since $\tilde{v}$ is ϵ-optimal with respect to time $t+\delta$, and $\tilde{v}(t) - v(t) \leq 0 \leq v(t+\delta) - v(t)$. Of course, no strategy can select more points in the interval $[t, t+\delta]$, than there are points in this interval. The expected number of such points equals δ. Hence the last term $|\tilde{v}(t+\delta) - \tilde{v}(t)|$ in (2.3) is smaller than δ. This implies the continuity of v with Lipschitz constant 1, since $\epsilon > 0$ is arbitrary. ■

The next result characterizes the optimal rule.

Lemma 2.3. Let ψ be defined by $(\psi_n)_{n=1,2,\cdots}$ satisfying

a) $\psi_n = 1$ if $X_n \geq \max_{k\leq n-1}\{\psi_k X_k\}$ and

$v(t-T_n, X_n) + 1 \geq v(t-T_n, \max_{k\leq n-1}\{\psi_k X_k\})$ for all $T_n \leq t$,

b) $\psi_n = 0$, otherwise.

Then ψ is optimal for time t.

Proof. The condition $X_n \geq \max\{\psi_k X_k\}$ in the definition of ψ assures that $\psi \in \Psi_a$. Now suppose that at some $T_n \leq t$, X_n can be selected, i.e. X_n is larger than any other selected X_j with $j < n$. If ψ selects X_n, then the optimal expected post-T_{n-1} number of selections equals 1 plus the expected number of future selections, which equals $v(t-T_n, X_n)$ for an optimal ψ. If ψ refuses X_n then the optimal post T_n number of selections equals $v(t-T_n, \max\{\psi_k X_k : k \leq n\})$. By the optimality principle, ψ must achieve the maximum of both. Conversely, if ψ does so, then ψ is optimal. ■

Theorem 2.1 $v'(t) = dv(t)/dt$ exists and is continuous on $\mathbb{R}^+$ satisfying equation (1.1) with $v'(0) = 1$. Moreover, $v'(t) > 0$ for all $t \geq 0$.

Proof. We first prove that $v'(t)$ exists everywhere on $\mathbb{R}^+$. This proof is similar to the proof given for a fixed number of variables (see Samuels and Steele (1981), section 2). In our case it is more convenient, however, to condition on what happens on an initial time interval $[0, \delta]$.

We confine our interest to one arrival or no arrival in $[0, \delta]$, because more than one arrival has probability $o(\delta)$. If we have an arrival at time $T_1 \in [0, \delta]$, which happens with probability $\delta + o(\delta)$, then we gain $v(t-T_1)$, if we refuse it and apply the optimal rule thereafter. If we accept it, then we gain $1 + v((1-x)(t-T_1))$, where x is its observed value. By the optimality principle we must opt for whatever is better, i.e. for $\max\{1 + v((1-x)(t-T_1)), v(t-T_1)\}$. If we have no arrival up to time δ, then the optimal gain is $v(t-\delta)$. Hence v must satisfy the optimality equation

$$v(t) = (\delta + o(\delta))\mathrm{E}\Big(\int_0^1 \max\{1 + v((1-x)(t-T_1)), v(t-T_1)\}dx\,\Big|\,0 \leq T_1 \leq \delta\Big)$$

$$+(1-\delta+o(\delta))v(t-\delta),$$

since the value of the observation is uniform on $[0,1]$. Subtracting $v(t-\delta)$ on both sides and dividing by δ yields

$$\frac{v(t)-v(t-\delta)}{\delta} =$$

$$-v(t-\delta)+\mathrm{E}\Big(\int_0^1 \max\{1+v((1-x)(t-T_1)), v(t-T_1)\}dx \,|\, 0\le T_1\le\delta\Big).$$

Here all terms multiplied by $o(\delta)$ sum clearly up to $o(\delta)$, because v is bounded by $v(u)\le u$ for all u. Taking the limit for $\delta\to 0$, and thus $T_1\to 0$, the rhs limit exists, since v is continuous. Hence the lhs limit exists as well and equals $v'(t)$. This yields

$$v'(t) = -v(t)+\int_0^1 \max\{1+v(1-x)t, v(t)\}dx$$

$$= \int_0^1 [v(t(1-x))+1-v(t)]^+dx.$$

Since $v(0)=0$ and since the integration with respect to $1-x$ or x is the same, we obtain

$$v'(t) = \int_0^1 [v(tx)+1-v(t)]^+dx, \tag{2.4}$$

which is equation (1.1) and satisfies $v'(0)=1$. Also, since $v(t)>0$ for all $t>0$, we immediatley get $v'(t)>0$ for all $t\ge 0$.

Finally, the continuity of v' on $\mathbb{R}^+$ follows immediately from the continuity of v on $\mathbb{R}^+$, since it implies the continuity of the integrand of (2.4) on $\mathbb{R}^+$. This proves Theorem 2.1. ■

We will use two other equations involving $v'(t)$ which we now derive. With the change of variables $u:=xt$, equation (2.4) is equivalent to

$$tv'(t) = \int_0^t [1+v(u)-v(t)]^+du, \quad t\ge 0.$$

We define

$$\alpha = \inf\{t\in\mathbb{R}^+ : v(t)=1\}. \tag{2.5}$$

Using definition (2.5) we can drop the positive-part function up to time α so that

$$tv'(t) = \int_0^t (1+v(u)-v(t))du, \quad 0\le t\le\alpha. \tag{2.6}$$

For $t \geq \alpha$ we now define the function ϕ as the unique solution of

$$v(\phi(t)) + 1 = v(t), \quad t \geq \alpha. \tag{2.7}$$

Moreover, we set $\phi(t) = 0$ for all $t < \alpha$. By the implicit function theorem we have for $t > \alpha$

$$v'(\phi(t))\phi'(t) = v'(t). \tag{2.8}$$

We also note that ϕ is in C^1 on $]\alpha, \infty[$, and continuous and strictly increasing on the half-closed interval $[\alpha, \infty[$.

We can write equation (1.1) in the form

$$tv'(t) = \int_{\phi(t)}^{t} (1 - v(t) + v(u))du = \int_{\phi(t)}^{t} (v(u) - v(\phi(t)))du, \tag{2.9}$$

where we used (2.7) in the second equality. Finally, by Fubini's theorem,

$$tv'(t) = \int_{\phi(t)}^{t} \int_{\phi(t)}^{u} v'(s)dsdu = \int_{\phi(t)}^{t} \int_{s}^{t} v'(s)duds$$

$$= \int_{\phi(t)}^{t} v'(s)(t - s)ds,$$

so that (2.9) is equivalent to

$$tv'(t) = \int_{\phi(t)}^{t} v'(s)(t - s)ds. \tag{2.10}$$

Before we prove the concavity of v let us first see what happens for $t \leq \alpha$.

Lemma 2.4 The (unique) solution $v(t)$ of equation (1.1) on $[0, \alpha]$ is given by

$$v(t) = \int_0^t \frac{1 - e^{-s}}{s} ds, \ 0 \leq t \leq \alpha, \tag{2.11}$$

where $\alpha = 1.34501 \cdots$.

Proof. We use the equivalent equation (2.6). Since v is continuous, (2.6) implies that $tv'(t)$ is continuous on $[0, \alpha]$, and so $v'(t)$ is continuous on $[0, \alpha]$. Deriving (2.6) with respect to t yields then $d(tv'(t))/dt = 1 - \int_0^t v'(t)du = 1 - tv'(t)$. It is easy to verify that this equation has the unique solution (2.11). We then solve $v(t) = 1$ which yields $\alpha = 1.34501 \cdots$. ■

Corollary 2.4 $v''(t) < 0$ for $0 \le t \le \alpha$ (which implies, in particular, that v is strictly concave on $[0,\alpha]$).

Proof. From (2.11) we obtain $v''(t) = (e^{-t}(t+1)-1)/t^2$ and $v''(0) = -1/2$. The numerator of $v''(t)$, $g(t)$ say,satisfies $g(0) = 0$ and $g'(t) < 0$ for $t > 0$, which implies the statement. ■

We have now sufficient material to show that v is concave on $\mathbb{R}^+$.

Theorem 2.2 The function v defined by equation (1.1) is strictly concave on $\mathbb{R}^+$. In fact a stronger property holds: v is twice continuously differentiable on $\mathbb{R}^+$ with $v''(t) < 0$ for all $t > 0$.

Proof. Equation (2.9) tells us that $tv'(t) = \int_{\phi(t)}^{t}(1+v(u)-v(t))du$ and from (2.8) and (2.10) we deduce that $tv'(t)$ is in C^1 on $]\alpha,\infty[$. This means that $v''(t)$ is continuous on $]\alpha,\infty[$. Also, a direct calculation shows that $tv''(t)+v'(t) = 1-(t-\phi(t))v'(t)$, and hence

$$tv''(t) = 1-(1+t-\phi(t))v'(t), \quad t > \alpha. \tag{2.12}$$

This equation also implies that

$$\lim_{t\to\alpha+} tv''(t) = 1-(1+\alpha)v'(\alpha).$$

If we compare this expression with the expression of Lemma 2.4 we see that

$$\lim_{t\to\alpha+} v''(t) = \lim_{t\to\alpha-} v''(t) > -\infty.$$

This implies that the function v' is continuously differentiable on $]0,\infty[$. But then (2.12) also implies that, for $t > \alpha$, the function $tv''(t)$ is continuously differentiable.

We now turn to the proof of $v''(t) < 0$ on $]0,\infty[$. We already know that $v''(t) < 0$ for $t \le \alpha$, and hence, if we define

$$T = \inf\{t : v''(t) = 0\}, \tag{2.13}$$

we necessarily have $\alpha < T \le \infty$. Of course, the aim is to show that $T = \infty$.

Let us suppose on the contrary that $T < \infty$. Since $tv''(t) < 0$ for $t < T$ we must have

$$\frac{d}{dt}(tv''(t))\Big|_{t=T} \ge 0. \tag{2.14}$$

Moreover, since $v''(t) < 0$ for $t < T$ we have also that on $[0, T]$ the function v' is strictly decreasing, and the implicit function theorem shows that $\phi'(t) < 1$ for all $\alpha < t \le T$. Then we get

$$\frac{d(tv''(t))}{dt} = -(1 - \phi'(t))v'(t) + (1 + t - \phi(t))v''(t). \tag{2.15}$$

But for $t = T$ (2.15) implies

$$\frac{d}{dt}(tv''(t))\big|_{t=T} = -(1 - \phi'(T))v'(T) < 0,$$

which contradicts inequality (2.14).

This shows $T = \infty$ and completes the proof of the Theorem. ■

Corollary 2.5

(i) $v'(t) > 0$, for all $t \ge 0$,

(ii) $v''(t) < 0$, for all $t \ge 0$,

(iii) $tv''(t) = 1 - v'(t)(1 + t - \phi(t))$ for $t \ge \alpha$.

(iv) $v'(t)(t - \phi(t)) \le 1 \le v'(t)(1 + t - \phi(t))$ for $t \ge \alpha$.

Proof. Except for property (iv) all other properties were obtained before. To prove property (iv) we observe that the first inequality follows immediately from the concavity of v whereas the second one follows from $v''(t) < 0$, and from property (iii). ■

Theorem 2.3 The functions v and ϕ satisfy

(i) $t - \phi(t) \le \sqrt{2t}$ for $t \ge \alpha$,

(ii) $v'(t) \ge \frac{1}{1+\sqrt{2t}}$ for $t \ge \alpha$,

(iii) $v(t) \ge \sqrt{2t} - \log(1 + \sqrt{(2t)}) + 1 - \sqrt{2\alpha} + \log(1 + \sqrt{2\alpha})$ for $t \ge \alpha$.

Proof. Since the function v is concave, v' is decreasing, and therefore (2.10) implies

$$v'(t)t = \int_{\phi(t)}^{t} v'(s)(t - s)ds \ge v'(t) \int_{\phi(t)}^{t} (t - s)ds = v'(t)\frac{1}{2}(t - \phi(t))^2.$$

Since $v'(t) > 0$ this yields $(t - \phi(t))^2 \leq 2t$, which proves statement (i). But then Corollary 2.5 (iv) implies that $v'(t) \geq 1/(1 + \sqrt{2t})$ for $t \geq \alpha$, proving statement (ii).

Finally, integrating this inequality from α to t yields, using $v(\alpha) = 1$,

$$v(t) = v(\alpha) + \int_\alpha^t v'(s)ds \geq v(\alpha) + \int_\alpha^t \frac{ds}{1 + \sqrt{2s}}$$

$$= \sqrt{2t} - \log(1 + \sqrt{2t}) + 1 - \sqrt{2\alpha} + \log(1 + \sqrt{2\alpha}).$$

This proves statement (iii) and completes the proof. ■

Theorem 2.4 The functions v and ϕ satisfy

(i) $v'(t) \leq \frac{1}{\sqrt{2t}}$ for $t \geq \alpha$,

(ii) $v(t) \leq \sqrt{2t} + 1 - \sqrt{2\alpha} < \sqrt{2t}$ for $t \geq \alpha$,

(iii) $t - \phi(t) \geq \sqrt{2t} - 1$ for $t \geq \alpha$.

(iv) $|v''(t)| \leq \frac{1}{t\sqrt{2t}}$ for $t \geq \alpha$.

Proof. To prove (i) we first note that, from the concavity of v,

$$v'(t)t = \int_{\phi(t)}^t (v(u) - v(\phi(t)))du \leq \int_{\phi(t)}^t (1 + v'(t)(u - t))du,$$

and therefore

$$v'(t)t \leq t - \phi(t) - v'(t)\frac{1}{2}(t - \phi(t))^2.$$

This we rewrite as

$$v'(t) \leq \frac{t - \phi(t)}{t + \frac{1}{2}(t - \phi(t))^2}.$$

But since the function $x \to x/(t + x^2/2)$ is increasing for $x \leq \sqrt{2t}$, we get from this

$$v'(t) \leq \frac{\sqrt{2t}}{t + \frac{1}{2}(\sqrt{2t})^2} = \frac{1}{\sqrt{2t}}.$$

Secondly, direct integration of (i) form a to t yields

$$v(t) \leq v(\alpha) + \int_\alpha^t \frac{ds}{\sqrt{2s}} = 1 + \sqrt{2t} - \sqrt{2\alpha},$$

which proves (ii).

To prove (iii) we use Corollary 2.5 (iv). This implies

$$t - \phi(t) + 1 \geq \frac{1}{v'(t)} \geq \sqrt{2t},$$

or equivalently,

$$t - \phi(t) \geq \sqrt{2t} - 1,$$

which proves (iii).

Finally, to prove (iv), we note that

$$\begin{aligned}
|v''(t)| &= \frac{v'(t)(1 + t - \phi(t)) - 1}{t}, \quad \text{(from Corollary 2.5 (iii))} \\
&\leq \frac{1 + t - \phi(t) - \sqrt{2t}}{t\sqrt{2t}}, \quad \text{(from Theorem 2.4 (i))} \\
&\leq \frac{1 + \sqrt{2t} - \sqrt{2t}}{t\sqrt{2t}}, \quad \text{(from Theorem 2.3 (i))} \\
&= \frac{1}{t\sqrt{2t}}.
\end{aligned}$$

This completes the proof. ■

Theorem 2.5 The function ϕ satisfies

(i) $\frac{\phi(t)}{t} \to 1$ as $t \to \infty$,

and

(ii) $\phi'(t) \to 1$ as $t \to \infty$.

Proof. Since $t - \phi(t) \leq \sqrt{2t}$ we immediately get (i) from

$$1 - \frac{\phi(t)}{t} \leq \frac{\sqrt{2t}}{t} \to 0, \text{ as } t \to \infty.$$

The second statement follows from $\phi'(t) = v'(t)/v'(\phi(t))$ since

$$1 \geq \phi'(t) \geq \frac{\sqrt{2\phi(t)}}{\sqrt{2t} + 1} \to 1, \text{ as } t \to \infty.$$ ■

This concludes the study of $v(t) = \mathrm{E}(L^t)$. We now turn to the study of the variance of L^t.

Fluctuations of the number of selected points.

We now address the problem of finding estimates for the random fluctuation of the number of points which are selected by the optimal rule. Our first concern is the variance of this number. Recall that, for fixed $t \geq 0$, L^t denotes the number of points selected up to time t under this rule.

The idea is to write L^t as the final value of a martingale and to use stopping time techniques to estimate $\mathrm{Var}(L^t)$.

Let $L^t(u)$ be the number of selected points up to time u under the rule which is optimal for time t (which we call the "t-optimal" rule). Thus $L^t(t) = L^t$ and $E(L^t) = v(t)$. Further we define

$$Y_u = L^t(u) + v((t-u)(1-M_u)),\ 0 \leq u \leq t, \tag{2.16}$$

where M_u denotes the largest selected *value* up to time u under the t-optimal rule. Clearly

$$Y_0 = v(t) = E(L^t) \text{ a.s.,}$$

and

$$Y_t = L^t \text{ a.s.} \tag{2.17}$$

According to the optimality principle, the process $(Y_u)_{0 \leq u \leq t}$ is a $(\mathcal{F}_u)_{0\leq u \leq t}$ martingale. The next step is to find the characteristics of Y. Intuitively this is done as follows. If an observation X arrives at time u then it will be accepted if and only if

$$X \geq M_{u-}$$

and

$$v((t-u)(1-X)) + 1 \geq v((t-u)(1-M_{u-})) \tag{2.18}$$

In order to describe this we introduce the process $(H_u)_{0\leq u\leq t}$, where $H(u)$ is the maximal acceptable observation at time u, i.e.

$$H_u = \sup\{0 \leq s \leq 1 : v((t-u)(1-s)) + 1 \geq v((t-u)(1-M_{u-}))\}. \tag{2.19}$$

By using the function ϕ, this can be written as

$$(t-u)(1-H_u) = \phi((t-u)(1-M_{u-})). \tag{2.20}$$

In particular, the process $(H_u)_{0\leq u\leq t}$ is predictable. ¿From this it follows that

$$(t-u)(H_u - M_{u-}) = (t-u)(1-M_{u-}) - \phi((t-u)(1-M_{u-})). \tag{2.21}$$

In case there is an acceptable observation X at time u, the jump $\Delta Y_u = Y_u - Y_{u-}$ is given by

$$\Delta Y_u = 1 + v((t-u)(1-X)) - v((t-u)(1-M_{u-})).$$

¿From this it follows that the characteristics (see Jacod and Shiryayev, (1980) p. 76) are given by the measure

$$\lambda_u du = K_u(dx),$$

where $\lambda_u = H_u - M_u$ and $K_u(dx)$ represents the distribution of the random variable

$$1 + v((t-u)(1-X)) - v((t-u)(1-M_{u-})), \tag{2.22}$$

with X being uniformly distributed on $[M_{u-}, H_u]$.

Before we give an upper and a lower bound for $\mathrm{Var}(L^t)$, let us recall the following result from martingale theory (see Protter (1995), squared variation process in Ch. 2, and also Brémaud (19 ??), p. 235, T3):

$$\begin{aligned}\mathrm{Var}(L^t) &= \mathrm{E}\Big(\int_0^t (\Delta_u)^2 dL^t(u)\Big)\\ &= \mathrm{E}\Big(\int_0^t du\lambda_u \int_0^1 y^2 K_u(dy)\Big)\\ &= \mathrm{E}\Big(\int_0^t du(H_u - M_{u-}) \times \qquad (2.23)\end{aligned}$$

$$\int_{M_{u-}}^{H_u} \frac{1}{(H_u - M_{u-})}\big(v((t-u)(1-x)) + 1 - v((t-u)(1-M_{u-})\big)^2 dx\Big).$$

Theorem 2.6

$$\mathrm{Var}(L^t) \geq \frac{v(t)}{3}, \ \forall t \geq 0.$$

Proof. Since v is concave we have that, for each jump time u, the random variable defined in (2.22)

$$1 + v((t-u)(1-X)) - v((t-u)(1-M_{u-}))$$

is stochastically larger than a uniformly distributed random variable on $[0,1]$. Hence, from (2.23),

$$\begin{aligned}\mathrm{Var}(L^t) &\geq \mathrm{E}\Big(\int_0^t du(H_u - M_{u-})\int_0^1 x^2 dx\Big)\\ &\geq \frac{1}{3}\mathrm{E}\,\Big(\int_0^t du(H_u - M_{u-})\Big)\\ &\geq \frac{1}{3}\mathrm{E}\,(\int_0^t dL^t(u)\,)\\ &= \frac{1}{3}v(t).\end{aligned}$$

■

We now turn to the problem of finding an upper bound for $\mathrm{Var}(L^t)$. Clearly, in order to understand the behavior of L^t sufficiently well, our goal is to find an upper bound which is close to the lower bound so that we can obtain strong limit results. Contrarily to the problem of finding a lower bound where we were able to minorize (stochastically) jump sizes, independently of the jump locations, by i.i.d. uniform random variables, a similar approach is now too coarse. Thus we face the additional problem of majorizing, in a suitable way, jump-location-*dependent* jump sizes.

We will need, among other tools, an elementary Lemma which we state and prove first.

Lemma 2.5 Let $a, b \in \mathbb{R}$ with $a < b$ and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a twice differentiable concave function on $[a,b]$ with $\varphi(a) = \varphi(b) = 0$. Suppose that, for some $\gamma \geq 0$,

$$\varphi''(x) \geq -\gamma, \text{ for all } x \in [a,b]. \tag{2.24}$$

Then

$$\int_a^b \varphi(x)dx \leq \frac{\gamma}{12}(b-a)^3.$$

Proof. Define

$$h(x) = \frac{\gamma}{2}(x-a)(b-x) - \varphi(x). \tag{2.25}$$

Then, from (2.24), for $x \in [a,b]$,

$$h''(x) = -\gamma - \varphi''(x) \leq 0,$$

and hence $h(x)$ is concave on $[a,b]$. Also, $\varphi(a)=\varphi(b)=0$, and so from (2.25), $h(a)=h(b)=0$. Concavity of h on $[a,b]$ implies then that h is nonnegative on $[a,b]$. Hence $0\le\varphi(x)\le(\gamma/2)(x-a)(b-x)$, $x\in[a,b]$, and thus

$$\int_a^b \varphi(x)dx \le \int_a^b \frac{\gamma}{2}(x-a)(b-x)dx = \frac{\gamma}{12}(b-a)^3. \qquad [\,]$$

Theorem 2.7

$$\mathrm{Var}(L^t) \le \frac{1}{3}v(t) + c_1\log(t) + c_2.$$

Proof. Let β be defined by $v(\beta)=2$. Since v^{-1} exists on $\mathbb{R}^+$, we have $v^{-1}(1)=\alpha$ and $v^{-1}(2)=\beta$ so that

$$1<\alpha<2<v^{-1}(2)=\beta, \tag{2.26}$$

where the third inequality follows from $v(t)<t$, $t>0$. Now note that, according to Theorem 2.3 (i), $x/\phi(x)\le x/(x-\sqrt{2x})$, and thus from $\beta>2$ in (2.26) we get

$$\frac{x}{\phi(x)} \le \frac{x}{x-\sqrt{2x}} \le \frac{\beta}{\beta-\sqrt{2\beta}},\ x\ge\beta, \tag{2.27}$$

because the function $x/(x-\sqrt{2x})$ is decreasing for $x>2$. We also note that $0<\beta/(\beta-\sqrt{2\beta})<\infty$, since $\beta>2$.

Let τ be the the stopping time

$$\tau = \inf\{0\le u\le t : (t-u)(1-M_u)\le\beta\}. \tag{2.29}$$

The time $u=t-\beta$ clearly satisfies the set prescription underlying τ so that

$$\tau \le t-\beta, \tag{2.30}$$

and, since τ is the infimum of this set,

$$(t-u)(1-M_{u-})>\beta, \text{ for } u\in[0,\tau[. \tag{2.31}$$

Now, by the martingale property,

$$\mathrm{E}\Big(Y_\tau(Y_t-Y_\tau)\Big)=0,$$

and hence

$$\mathrm{Var}(L^t) = \mathrm{Var}(Y_t) = \mathrm{Var}(Y_\tau) + \mathrm{Var}(Y_t - Y_\tau).$$

Both summands are estimated in a different way. We start with the easier one.

Let $\mathcal{F}_{\int}$ denote the sigma-field generated by all arrival time points up to time s and their corresponding values. Then

$$\mathrm{Var}(Y_t - Y_\tau) \le \mathrm{E}\Big(\sum_{\tau \le u \le t} (Y_u - Y_{u-})^2 \Big)$$

$$\le \mathrm{E}\big(L^t - L^t(\tau)\big) = \mathrm{E}\big(\mathrm{E}(L^t - L^t(\tau) | \mathcal{F}_\tau)\big).$$

By the martingale property, the latter equals

$$E\big(v((t-\tau)(1-M_\tau)) \big)$$

$$= \mathrm{E}(v(\beta)) = 2,$$

where the second line follows fom the definitions of τ and β (see (2.26) and (2.29), and the continuity of v.

The variance of Y_τ requires more work.

Recall that

$$\mathrm{Var}(Y_\tau) = \mathrm{E}\Big(\int_0^\tau du \int_{M_u}^{H_u} \big(J(s,u)\big)^2 ds \Big),$$

where $J(s,u) = 1 + v((t-u)(1-s)) - v((t-u)(1-M_{u-})$ for $M_{u-} \le s \le H_u$. Since v is concave, we have $\partial^2 J(s,u)/\partial s^2 \le 0$, and thus $J(s,u)$ is a concave function of s. Hence

$$J(s,u) \ge l(s) := \frac{H_u - s}{H_u - M_{u-}},\ M_{u-} \le s \le H_u. \tag{2.32}$$

Also $J^2 - l^2 \le 2(J-l)$, since $0 \le l \le J \le 1$, $M_{u-} \le s \le H_u$, and hence

$$(J(s,u))^2 \le (l(s))^2 + 2(J(s,u) - l(s)),\ M_{u-} \le s \le H_u \tag{2.33}$$

So we get

$$\mathrm{Var}(Y_\tau)$$

$$\le \mathrm{E}\Big(\int_0^\tau du \int_{M_{u-}}^{H_u} (l(s))^2 ds \Big) + 2\,\mathrm{E}\Big(\int_0^\tau du \int_{M_{u-}}^{H_u} (J(s,u) - l(s)) ds \Big). \tag{2.34}$$

For the first term in (2.34) we obtain from (2.32) and (2.30)

$$\mathrm{E}\Big(\int_0^\tau \frac{1}{3}(H_u - M_{u-})du$$

$$\leq \mathrm{E}\Big(\int_0^t \frac{1}{3}(H_u - M_{u-})du \Big) = \frac{1}{3}\mathrm{E}(L^t) = \frac{1}{3}v(t), \tag{2.35}$$

since the intensity of $L^t(u)_{0\leq u\leq t}$ is exactly $(H_u - M_{u-})du$.

To estimate the second term we first note that the function $\varphi(s,u) := J(s,u) - l(s)$ vanishes in $s = M_{u-}$ and in $s = H_u$ and is concave in s on $[M_{u-}, H_u]$. Thus $\varphi(s,u)$ satisfies the hypotheses of Lemma 2.5 with $a = M_{u-}$ and $b = H_u$. Therefore we obtain for the second term of (2.34)

$$2\,\mathrm{E}\Big(\int_0^\tau du \int_{M_{u-}}^{H_u} (J(s,u) - l(s))ds \Big)$$

$$\leq \frac{1}{6}\mathrm{E}\Big(\int_0^\tau du(H_u - M_{u-})^3 \sup_{s\in[M_{u-},H_u]} |v''((t-u)(1-s))|(t-u)^2 \Big). \tag{2.36}$$

Since $(t-u)(1-M_{u-}) > \beta$ (see (2.31)) we obtain from (iv) of Theorem 2.4

$$|v''((t-u)(1-s))| \leq \frac{1}{\sqrt{2}}((t-u)(1-H_u))^{-3/2}.$$

Therefore, from (2.36),

$$\begin{aligned} &\mathrm{Var}(Y_\tau) \\ &\leq \frac{1}{6\sqrt{2}}\,\mathrm{E}\Big(\int_0^\tau (H_u - M_{u-})^3 (t-u)^2\,((t-u)(1-H_u))^{-3/2} du \Big) \\ &= \frac{1}{6\sqrt{2}}\,\mathrm{E}\Big(\int_0^\tau \big((H_u - M_{u-})(t-u)\big)^3 \frac{du}{(t-u)^{5/2}(1-H_u)^{3/2}} \Big), \end{aligned} \tag{2.37}$$

which is, according to (2.21), smaller than

$$\frac{1}{6\sqrt{2}}\,\mathrm{E}\Big(\int_0^\tau \big((t-u)(1-M_{u-}) - \phi((t-u)(1-M_{u-}))\big)^3 \times \frac{du}{(t-u)^{5/2}(1-H_u)^{3/2}} \Big). \tag{2.38}$$

By inequality (2.27) the last expectation is smaller than

$$\frac{1}{6\sqrt{2}}\,\mathrm{E}\Big(\int_0^\tau \frac{du}{t-u} \frac{\beta}{(\beta - \sqrt{2\beta})} \Big)$$

and thus from (2.30)

$$\leq \frac{\beta}{(\beta-\sqrt{2\beta})}\frac{1}{6\sqrt{2}}\int_0^{t-\beta}\frac{du}{t-u}$$

$$\leq \frac{\beta}{(\beta-\sqrt{2\beta})}\,6\sqrt{2}\log\Big(\frac{t}{\beta}\Big).$$

Thus from (2.37) , (2.38) and (2.39),

$$\mathrm{Var}(L^t) \leq \frac{1}{3}v(t) + \frac{\beta}{(\beta-\sqrt{2\beta})6\sqrt{2}}\log\Big(\frac{t}{\beta}\Big) + 2,$$

and the proof is complete. ■

Corollary 2.6

$$\frac{L^t}{v(t)} \to 1 \text{ i.p. and } \frac{L^t}{\sqrt{2t}} \to 1 \text{ i.p. as } t \to \infty.$$

Proof. This is straightforward from Theorems 2.6 and 2.7 and Chebychev's inequality. ■

For t fix, let $(Y^t_u)_{\,0\leq u\leq t}$ be the martingale

$$Y^t_u = L^t_u + v((t-u)(1-M^t_u)).$$

Further let $<\,,\,>$ denote the skew bracket (predictable quadratic variation) for martingales. Then we know already from Theorem 2.6 and Theorem 2.7 that

$$\mathrm{E}\Big((\frac{1}{3}\sqrt{2t})^{-1}\Big(<Y^t,Y^t>_t -(Y^t_0)^2\Big)\Big) \to 1$$

as $t\to\infty$. To prove the convergence in probability, we need the following lemma on the intensity λ^t_u of the process $(L^t_u)_{\,0\leq u\leq t}$.

Lemma 2.6

$$\frac{1}{\sqrt{2t}}\mathrm{Var}\Big(\int_0^t \lambda^t_u du\Big) \to \frac{4}{3}, \quad \text{as } t\to\infty.$$

Proof.

$$\mathrm{E}\Big[\Big(\int_0^t \lambda^t_u du\Big)^2\Big] = 2\mathrm{E}\Big[\int_0^t \Big(\int_s^t \lambda^t_u du\Big)\lambda^t_s ds\Big] = 2\mathrm{E}\Big[\int_0^t (L^t_t - L^t_s)\lambda^t_s ds\Big], \qquad (2.40)$$

where the second equality holds since $(L_s^t - \int_0^s \lambda_u^t du)_{0\le s\le t}$ is a martingale. Since $Y_t^t = L_t^t$ and $(Y_s^t)_s$ is a martingale, the rhs of (2.40) yields

$$2\mathrm{E}\Big[\int_0^t (Y_t^t - L_s^t)\lambda_s^t ds\Big] = 2\mathrm{E}\Big[\int_0^t (Y_s^t - L_s^t)\lambda_s^t ds\Big] = 2\mathrm{E}\Big[\int_0^t (Y_{s-}^t - L_{s-}^t)\lambda_s^t ds\Big],$$

and so, since $(Y_{s-}^t - L_{s-}^t)$ is predictable,

$$= 2\mathrm{E}\Big[\int_0^t (Y_{s-}^t - L_{s-}^t) dL_s^t\Big] = 2\mathrm{E}\Big(\int_0^t Y_{s-}^t dL_s^t\Big) - 2\mathrm{E}\Big(\int_0^t L_{s-}^t dL_s^t\Big) \tag{2.41}$$

For the first integral in (2.41) we use

$$Y_{s-}^t dL_s^t = d(Y_s^t L_s^t) - L_s^t dY_s^t = d(Y_s^t L_s^t) - (L_{s-}^t + 1)dY_s^t,$$

since $L_s^t = L_{s-}^t + 1$ holds dY-a.s. So we get

$$\mathrm{E}\Big[\int_0^t Y_{s-}^t dL_s^t\Big] = \mathrm{E}\Big[Y_t^t L_t^t\Big] - \mathrm{E}\Big[\int_0^t (L_{s-}^t + 1)dY_s^t\Big].$$

Since the second term equals zero by the martingale property of $(Y_s^t)_{0\le s\le t}$, we get

$$\mathrm{E}\Big[\int_0^t Y_{s-}^t dL_s^t\Big] = \mathrm{E}\Big[\big(L_t^t\big)^2\Big]. \tag{2.42}$$

The second integral in (2.41) can be computed using

$$2L_{s-}^t dL_s^t = d\big(L_s^t)^2\big) - dL_s^t,$$

and hence

$$2\mathrm{E}\Big[\int_0^t L_{s-}^t dL_s^t\Big] = \mathrm{E}\Big[\big(L_t^t\big)^2\Big] - \mathrm{E}(L_t^t).$$

Putting things together ((2.40) through (2.43)) gives

$$\mathrm{E}\Big[\Big(\int_0^t \lambda_u^t du\Big)^2\Big] = \mathrm{E}\Big[\Big(L_t^t\Big)^2\Big] + \mathrm{E}(L_t^t).$$

¿From this we deduce

$$\mathrm{Var}\Big(\int_0^t \lambda_u^t du\Big) = \mathrm{Var}(L_t^t) + \mathrm{E}(L_t^t),$$

and hence, as $t \to \infty$,

$$\frac{1}{\sqrt{2t}}\mathrm{Var}\Big(\int_0^t \lambda_u^t du\Big) \to \frac{1}{3} + 1 = \frac{4}{3},$$

which completes the proof. ∎

Corollary 2.7

$$\frac{1}{\sqrt{2t}}\Big(\int_0^t \lambda_u^t du\Big) \to 1, \qquad \text{i. p. as } t \to \infty.$$

Proof. We know that

$$\mathrm{E}\Big[\int_0^t \lambda_u^t du\Big] = \mathrm{E}\big(L_t^t\big) \sim \sqrt{2t},$$

and from the preceding Lemma,

$$\mathrm{Var}\Big[\frac{1}{\sqrt{2t}}\int_0^t \lambda_u^t du\Big] = \frac{1}{2t}\mathrm{Var}\Big[\int_0^t \lambda_u^t du\Big] \to 0.$$

This proves the statement. ∎

Lemma 2.7

$$\frac{3}{\sqrt{2t}}\Big(< Y^t, Y^t > -(Y_0^t)^2\Big) \to 1 \quad \text{i. p. as } t \to \infty.$$

Proof. The previous section showed that

$$< Y^t, Y^t > -(Y_0^t)^2 = \int_0^t du \lambda_u^t \int_{[0,1]} y^2 \nu_u^t d(y)$$

$$= \frac{1}{3}\int_0^t \lambda_u^t du + \mathcal{O}\big(\log(t)\big).$$

Hence the statement follows from

$$\Big\| \frac{3}{\sqrt{2t}}\Big(< Y^t, Y^t > -(Y_0^t)^2\Big) - 1 \Big\|_2^2$$

$$\leq \frac{1}{2t}\mathrm{Var}\Big(\int_0^t \lambda_u^t du\Big) + C\frac{(\log(t))^2}{2t} \to 0, \text{ as } t \to \infty.$$ ∎

Conjecture. We believe that the preceding results can be strengthened to a functional Central Limit Theorem, namely that $\Big(\frac{L_t^t - \sqrt{2t}}{\sqrt{2t}^{1/2}}, \frac{L_t^t - \int_0^t \lambda_u^t du}{\sqrt{2t}^{1/2}}\Big)$ tends to a two-dimensional normal random variable with mean zero and variance-covariance matrix

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

Indeed, one can see that the preceding results are close to such a result. To prove the above we have repeatedly tried to relate our results to known results (see Billingsley (1995), Jacod and Shiryayev (1980), and others on this topic. However, so far, we must leave this question open.

We now turn to the last objective of this paper, namely to the study of very simple rules.

3. The simplest *graph*-rule.

We recall that the optimal selection rule is of the following form: Select the first observation (T_τ, X_τ) (if any) satisfying $1 + v((1 - X_\tau)(t - T_\tau)) \geq v(t - T_\tau)$, and, recursively, if (T_k, X_k) was the last selected observation, select the first subsequent observation $(T_{\tau(k)}, X_{\tau(k)})$ (if any) satisfying $1+v((1-X_{\tau(k)})(t-T_{\tau(k)})) \geq v(t - T_{\tau(k)})$. The optimal rule is thus a rule defined by v itself. However, the recursive definition of the rule is of limited value. Given that the "on-line" requirement means, in practice, "without delay", this raises the question of finding simpler rules which perform well. We focus our interest on the simplest rules we can imagine, i.e. on drawing a (fixed) graph through $[0, t] \times [0, 1]$ and selecting sequentially all points below the graph which form an increasing sequence. And the goal is, as we should point out, to choose the best *graph*, and *not* an asymptotically optimal rule. (Asymptotically optimal rules are given in Section 4.)

It is interesting to put this goal into a parallel with known results about the concentration of measure of i.i.d. random variables under the condition that they form a record. Deuschel and Zeitouni (1995) have studied limiting curves of concentration of records independently of the question of the existence of good online strategies to select them (for the background see Goldie and Resnick (1996)). The Poisson process model we consider can then easily be compared with the case of t $(t \in \mathbb{N})$ uniform $[0, 1]$ random variables. In this case this curve (on $[0, 1]^2$) is the diagonal line, as Deuschel and Zeitouni have shown. They also establish the link with the monotone subsequence problem (Lemma 8, p. 874), and this shows that a prophet would select the longest subsequence, with expected length in the order of $2\sqrt{t}$, essentially along this line. It is intuitive that a "lazy" decision maker who wants to use a convenient f-record rule cannot do better than just trying to mimic a prophet on-line. And it is nice to see that, as we shall prove, this intuition is indeed true.

The proof consists of several parts which we prepare first.

Definition 3.1 Let $f : [0,t] \to [0,1]$ and let $S_f = \{(s,x) : 0 \le s \le t, 0 \le x \le f(s)$. We say that (T_k, X_k) is a f-record on $[0,t]\times[0,1]$, if X_k is a record among $(T_j, X_j I(X_j \le f(T_j)), j = 1, 2, \cdots, k$, where I denotes the indicator function.

Note that if $Y_j := X_j I(X_j \le f(T_j)) = 0$, then Y_j cannot be a record. Hence f-records are simply records in the planar Poisson process with unit rate confined to the region bounded above by the graph of f.

Definition 3.2 The rule of selecting sequentially all f-records, and only these, will be called *f-record rule.* Further, f^* is called *optimal*, if it maximizes the expected number of f-records for all $f : [0,t] \to [0,1]$. (We do not affirm here yet that such a f^* exists, in which case the definition is understood to be weakened to ϵ-optimality.)

Lemma 3.1 For a given function $f : [0,t] \to [0,1]$ let $r_f(t)$ denote the expected number of f-records. Then

$$r_f(t) = \int_0^t \int_0^{f(s)} e^{-\Lambda_f(s,x)}\, dx\, ds, \tag{3.1}$$

where $\Lambda_f(s,x)$ denotes the Lebesgue measure of the intersection of S_f and the North-West region of the plane with respect to the point (s,x).

Proof. The probability of finding exactly one f-record in a rectangle containing (s,x) with area $ds \times dx$ equals, according to the Poisson hypothesis,

$$(dx\, ds + o(dx\, ds)) \exp\{-\Lambda_f(x,s)\},$$

because it is necessary and sufficient that the rectangle contains at least one observation, and that this observation is not preceded by a f-record with a larger X-value. Further, more than one arrival time in ds has probability $o(ds)$ (more than one X-value in dx has probability $o(dx)$) so that summing over the probabilities, respectively expectations, are here asymptotically equivalent. Letting $ds \to 0$ and $dx \to 0$, the limiting sum exists and yields then the above double integral, because the integrand is continuous in both s and x. ■

Maximizing (3.1) over all f is in general a difficult problem of calculus of variations, but we can reduce the range of functions containing f^*. For sufficiently small t, we have $f^* \equiv 1$ on $[0,t]$, as we can see easily. Indeed, we know from the definition of α (see (2.5)-(2.6)) that $v(t) \le 1$ for $t \le \alpha$, so that it is optimal to select any record if $t < \alpha$. But this simple picture changes quickly, even for moderate t. Further, it will become clear that, as t becomes larger, we can confine our interest to increasing functions f. Therefore we first note that

Corollary 3.1 If f is strictly increasing, then equation (3.1) becomes

$$r_f(t) = \int_0^t \int_0^{f(s)} \exp\Big\{-\int_{f^{-1}(x)}^s (f(u)-x)du\Big\}dx\,ds, \tag{3.2}$$

where $f^{-1}(x) := \inf\{s : f(s) = x\}$.

$r_f(t)$ in (3.2) can be computed conveniently for $f(s) = s/t$, and this will give us immediately a lower bound for $\sup_f r_f(t)$. If $f(s) = s/t$ then we denote the corresponding expected number of f-records by $r_d(t)$ (d being mnemonic for "diagonal line").

Lemma 3.2

$$\sup_f r_f(t) \geq r_d(t) = \sqrt{\frac{\pi}{2}t} - 1 + o\Big(\frac{1}{t}\Big). \tag{3.3}$$

Proof: The inequality in (3.3) is obvious. Further, for $f(s) = s/t$ we have $f^{-1}(s) = ts$ and so from (3.2)

$$r_d(t) = \int_0^t \int_0^{s/t} \exp\Big\{-\int_{tx}^s \Big(\frac{u}{t} - x\Big)du\Big\}dxds. \tag{3.4}$$

With the change of variable $s := s/t$, (3.4) becomes

$$r_d(t)/t = \int_0^1 \int_0^s \exp\Big\{-(t/2)(s-x)^2\Big\}dxds. \tag{3.5}$$

We now use

$$\int_0^s \exp\Big\{-(t/2)(s-x)^2\Big\}dx = \int_0^s \exp\Big\{-(t/2)x^2\Big\}dx = \frac{1}{\sqrt{t}}\int_0^{\sqrt{t}s} e^{-w^2/2}dw.$$

Plugging this into (3.5) we see that

$$r_d(t) = \sqrt{t}\int_0^1 (G(\sqrt{t}s) - G(0))ds, \tag{3.6}$$

where G denotes the Gaussian error-integral. (3.6) can be straightforwardly integrated which yields the rhs of (3.3). ■

Asymptotically optimal f-record rules. To tackle the question of an asymptotically optimal performance, we transform the time by $s := s/t$, so that we have now a planar Poisson process with rate t on $[0,1]^2$. Consequently, we may and will confine our interest to those functions f which satisfy $0 \leq f(s) \leq 1$ for all $s \in [0,1]$.

We first need two preliminary easy estimates.

Corollary 3.2 Let $0 < m < \infty$ and let $f(s) = ms$ on $[0, t_1]$, and let $r_f(t_1, t)$ denote the expected number of f-records up to time t_1. Then there exists a constant $c_1 > 1$ such that

$$c_1\sqrt{t_1 f(t_1)t} \leq r_f(t_1,t) \leq 2\sqrt{t_1 f(t_1)t},\ 0 < t_1 \leq \max\{1, 1/m\} \tag{3.7}$$

for all t sufficiently large.

Proof. The first inequality in (3.7) is obvious from the proof of Lemma 3.2 (by time scale transformation). The second inequality holds since $r_f(t_1,t) \leq v(t_1 f(t_1)t) \sim \sqrt{2t_1 f(t_1)t} < 2\sqrt{t_1 f(t_1)t}$. ■

Lemma 3.3 Let $0 < a \leq 1$, and let $f(s) \equiv a$ on $[0, t_1]$ for some $0 < t_1 \leq 1$. Then, for all constants $c_2 > 1$,

$$r_a(t_1, t) < c_2 \log(t) \tag{3.8}$$

for all t sufficiently large.

Proof. It is easy to check (and well-known) that the expected number of records from n i.i.d. continuous random variables equals $\mu(n) = 1 + 1/2 + \cdots + 1/(n-1) + 1/n$ (see e.g. Arnold et al. (1998), p. 23). Since $\mu(n) \leq 1 + \log(n)$ we have $r_a(t_1, t) \leq E(1 + \log(N))$, where N is a Poisson-distributed random variable with mean at_1t). Since log is a concave function, Jensen's inequality yields $E(\log(N)) \leq \log(E(N))$, and thus

$$r_a(t_1,t) \leq 1 + \log(E(N)) = 1 + \log(at_1) + \log(t) \leq 1 + \log(t) < c_2 \log(t),$$

where the last inequality holds for all t sufficiently large. ■

We are now ready for the essential Lemma of this section.

Lemma 3.4 To find an asymptotically optimal function f on $[0,1]$ as $t \to \infty$, it suffices to study the class of functions f which are increasing and satisfy $f(0) = 0$ and $f(1) = 1$.

Proof (i) We first show that we can confine to the class of increasing functions f. Recall (3.1), which becomes (in time scale $s := s/t$)

$$r_f(t) = \int_0^1 \int_0^{f(s)} te^{-t\,\Lambda_f(s,x)}\,dx\,ds. \tag{3.9}$$

Suppose now that f is not increasing on $[0,1]$. We will see then that the choice $\tilde{f}(s) := \max_{\{0\le u\le s\}} f(s)$ is asymptotically at least as good as f. Clearly, $\tilde{f}$ is well-defined since f is continuous. Since f is not increasing on $[0,1]$, there exists a largest point, $s_1 < 1$ say, such that f is increasing on $[0,s_1]$ (where $s_1 = 0$ is defined to signify that f decreases in a neigborhood of 0). Again from continuity of f, we can find an $\epsilon > 0$, a $\delta := \delta(\epsilon) > 0$ and a largest time $s_2 := s_2(\epsilon)$ (truncated at 1) such that $f(u) \le f(s_1) - \epsilon$ for $s_1 + \delta \le u \le s_2$. But then $\Lambda_f(u,x)$ is positive and bounded away from 0 on the set $\mathcal{U} = \{(u,x) : s_1 + \delta \le u \le s, 0 \le x \le f(u)\}$. Therefore $t\exp\{\tau\,\Lambda_f(u,x)\} \to 0$ on U as $t \to \infty$, and thus

$$\int_{s_1+\delta}^{s_2} \int_0^{f(s)} te^{-t\,\Lambda_f(s,x)}\,dx\,ds \to 0 \text{ as } t \to \infty \tag{3.10}$$

This is true for $0 < \epsilon' < \epsilon$ as well, and then, having chosen a corresponding $\delta > 0$, also for all δ' with $0 < \delta' < \delta$, since s_1 is maximal. Extending the outer integral in (3.10) to $\int_{s_1}^{s_2}$ adds at most δ to the value of the integral, and hence, as $t \to \infty$, the limiting contribution of the time interval $[s_1, s_2]$ to the expected number of f records is zero. The same argument holds for all subsequent intervals (if any) on which f drops below $f(s_1)$. Thus replacing f by $\tilde{f}$ for $s \ge s_1$ cannot but improve on f, since the integrand is positive. (Note also that $N_{\tilde{f}}(s,x)$ does not stay bounded away from zero.) This implies that we can focus on increasing functions f.

(ii). We now show that an asymptotically optimal f must satisfy $f(0) = 0$. According to (i), we suppose f is increasing. Suppose $f(0) = a > 0$. Let, as before, $r_f(\epsilon, t)$ be the expected number of f-records on $[0,\epsilon]$. The number of f-records which are not a-records is the number of those f-records which lie in the region bounded by the graphs of f and a. Since the Poisson arrival rate is constant everywhere, it has the same distribution as the number of $f-a$-records (where $f - a$ is shorthand for $g(s) := f(s) - a$). Hence its expectation is the same as $r_{f-a}(t,\epsilon)$. Counting f-records and a-records separately yields the upper bound

$$r_f(t,\epsilon) \le r_a(t,\epsilon) + r_{f-a}(t,\epsilon), \tag{3.11}$$

because all those a-records which are preceded by f-records do not contribute to the number of f-records.

Now note that $r_{f-a}(t,\epsilon) \leq v(\epsilon(f(\epsilon)-a)t)$. Indeed, $v(\epsilon f(\epsilon))$ yields the expected number of selectable points on the box $[0,\epsilon]\times[0,f(\epsilon)-a)]$ under the (overall) optimal selection rule, and the class of all selection rules contains the class of f-record rules. Also recall that $v(t)\sim\sqrt{2t}$. Using this and Lemma 3.3 yields then from (3.11)

$$r_f(t,\epsilon)/\sqrt{\epsilon t} \leq 2\sqrt{f(\epsilon)-a}+\eta_t, \tag{3.12}$$

where $0<\eta_t<c_2\log(t)/\sqrt{\epsilon t}\to 0$ as $t\to\infty$. Since f is continuous, the rhs tends to zero as ϵ tends to zero. Replace f on $[0,\epsilon]$ (only) by the line joining the origin and the point $(\epsilon,f(\epsilon))$, i.e. by $g(s)=sf(\epsilon)/\epsilon>0$. Since $f(0)=a>0$ and f is increasing, this line has a finite strictly positive slope for all $\epsilon>0$. Thus from Corollary 3.2, for t sufficiently large,

$$\frac{r_g(t,\epsilon)}{\sqrt{\epsilon t}} \geq \frac{c_1\sqrt{t\epsilon f(\epsilon)}}{\sqrt{\epsilon t}} = c_1\sqrt{f(\epsilon)}.$$

The rhs stays now bounded away from 0, since $f(0)=a>0$. Thus there exists $\epsilon>0$ such that $r_f(t,\epsilon)<r_g(t,\epsilon)$ for t sufficiently large, which proves that f is asymptotically suboptimal unless $f(0)=0$.

(iii) Finally, suppose that $f(1)=b<1$. Replace f on $[1-\epsilon,1]$ by the line-segment joining $(1-\epsilon,b)$ and $(1,1)$, which has again a strictly positive finite slope for each $0<\epsilon\leq 1$. Similarly to the proof of (ii) we see then that, for some $\epsilon>0$ this replacement improves on f as t becomes large.

This completes the proof. ■

We are now ready for the main result.

Theorem 3.1 The f-record rule using the diagonal line $f(s)=s$ is asymptotically optimal in the class of differentiable functions $f:[0,1]\to[0,1]$.

Proof. According to Lemma 3.4 we may and do suppose that $f(0)=0, f(1)=1$ and $f'(s)\geq 0$ on $[0,1]$. The latter implies that $\Lambda_f(s,f(s))=0$ and that each point (s,x) with $0\leq x<f(s)$ has a single connected North-West region below f attached to it. For $0\leq\beta\leq 1$ let

$$B(\beta)=\{(s,x)\in[0,1]^2 : 0\leq x\leq f(s), \Lambda_f(x,s)\leq\beta\}. \tag{3.13}$$

Note that $B(\beta)$ is bounded above by $B(0)$, i.e. the graph of f, and below by the curve $c(s,\beta):=c_f(s,\beta)$ defined by $\Lambda_f(s,x)=\beta$, both being well defined and continuous in β since f is increasing and continuous. Therefore the Lebesgue measure of $B(\beta)$ denoted by $\mathcal{L}(B(\beta))$ is (right)-continuous in $\beta=0$.

Recall the intergral in (3.9). As t increases, $t\exp\{-t\,\Lambda_f(s,c(s,\beta))\}\to 0$ for each $\beta>0$, so that integration on the complement of $B(\beta)$ is asymptotically negligible for any fixed $\beta>0$. This means that the asymptotic behavior of the double integral (3.9) is solely determined by the behavior of the integrand on $B(\beta)$ as $\beta\to 0+$. Now interpret (3.9) as t times the Laplace transform of Λ_f. This opens the way to applying an Abelian Theorem (Feller, Vol. II, Chapter XIII, section 5, Theorem 4). This shows that maximizing (3.9) over the set of f is equivalent to maximizing $t^\rho\mathcal{L}(B(\beta))$ for some factor $\rho>0$. Moreover, we see that ρ must be equal to $1/2$, because $\sup_{\{f\}} r_f(t)\le v(t)\sim\sqrt{2t}$, and, from Lemma 3.2 $\sup_{\{f\}} r_f(t)\ge c\sqrt{t}$, for each $c<\sqrt{2\pi}$ and t sufficiently large.

Now, to study the limiting behavior of

$$\sqrt{t}\mathcal{L}(B(\beta))=\sqrt{t}\int_0^1 (f(s)-c(s,\beta))ds, \tag{3.14}$$

as $\beta\to o+$ we use the differentiability of f. Note that

$$\begin{aligned}\Lambda_f(s,c(s,\beta))&=\frac{1}{2}(s-f^{-1}(c(s,\beta))(f(s)-c(s,\beta))+o(\beta)\\&=\frac{1}{2}\frac{(f(s)-c(s,\beta))^2}{f's)}+o(\beta).\end{aligned}$$

Since $\Lambda_f(s,c(s,\beta))=\beta$, this implies

$$f(s)-c(s,\beta)=\sqrt{2\beta}\sqrt{f'(s)}+o(\beta). \tag{3.15}$$

Therefore

$$\mathcal{L}(B(\beta))=\int_0^1 (f(s)-c(s,\beta))ds=\int_0^1\sqrt{f'(s)}ds+o(\beta).$$

The Cauchy-Schwarz inequality implies, however, that

$$\int_0^1\sqrt{f'(s)}ds\le\sqrt{\int_0^1 ds}\sqrt{\int_0^1 f'(s)ds}=\sqrt{f(1)-f(0)}=1,$$

and that equality holds if and only if $f'(s)$ is proportonal to 1 on $[0,1]$. With the border conditions this implies that only $f(s)=s$ is asymptotically optimal, and the proof is complete. ■

4. Conclusions and comparisons. One practical solution of the problem is thus to graph the line $f(s)=s/t$ through the ribbon $[0,t]\times[0,1]$ and to select greedily all f-records. This is the "laziest" f-record rule. Note that any monotone increasing function f whose horizontal projection into $[0,1]^2$ is the main diagonal is an asymptotically optimal f-record rule as t tends to infinity.

For finite (even large) t the projection argument is for on-line strategies of limited value, of course, and there are strategies other than the optimal one which perform better. According to the optimality conditions implied by Lemma 2.1, for instance, we can do better by replacing after each selection the old diagonal line by a new one passing through the presently selected point and the point $(t,1)$. A little reflection shows that this rule is indeed always strictly better.

However, this rule of drawing new lines after each selection is more complicated, so that a convenient approximation of the optimal rule by "close-to-opimal" rules become comparable with respect to on-line speed, and, due to their asymptotic optimality, clearly preferable. (Here we mean by an *asymptotically optimal* rule a rule whose value $w(t)$ satisfies $w(t)/v(t) \to 1$ as $t \to \infty$.)

Such close-to-optimal rules are easy to obtain. According to Theorem 2.3 (iii) and Theorem 2.4 (ii) we have close lower and upper bounds for v. Using for instance an approximation $\hat{v}$ of v in Lemma 2.3 between these bounds ($\hat{v}(t) = \sqrt{2t} - \frac{1}{4}\log(t)$, say) will yield close-to-optimal asymptotically optimal results, because both $v(t)$ and $\hat{v}(t)$ are continuous.

Open problems. An analytic challenge for this problem is to know whether the logarithmic gap between the lower and the upper bound for $v(t)$ can be tightened to a gap of lower order. The second open problem is to know whther the conjecture of the functional CLT for $(L_u^t)_{0\le u\le t}$, stated at the end of Section 3, is true.

Acknowledgement. We thank the Referee and the Associate Editor for critical comments and suggestions.

References

Aldous D.J. and Diaconis, P. (1999). Longest Increasing Subsequences: From Patience Sorting to the Baik-Deift-Johansson Theorem. *Bulletin Amer. Math. Society*, Vol. 36, 413-432.

Arnold, B.C., Balakrishnan N. and Nagaraja H.N. (1998) Records. *John Wiley & Sons*, New-York, Chichester, Weinheim, Brisbane, Singapore, Toronto.

Baik, J. , Deift, P.A. and Johansson, K. (1999). On the Distribution of the Length of the Longest Increasing Subsequence of Random Permutations. *Technical Report math.Co/99810105.* To appear in *J. Amer. Math. Society.*

Baryshnikov Y.M. and Gnedin A.V. (1998). Sequential Selection of an Increasing Sequence from a Multidimensional Random sample. Preprint.

Brémaud, J. (1981). Point Processes and Queues. Springer Verlag New York - Heidelberg - Berlin.

Critchlow, D. (1988). Ulam's Metric. In *Encyclopedia of Statistical Sciences*, Vol. 9, 379-380.

Deuschel J.-D. and O. Zeitouni (1995). Limiting Curves for i.i.d. Records. *Ann. Probab.* **23**, 852-878.

Feller, W. (1971) An Introduction to Probability Theory and its Applications, Volume II. Second edition. *John Wiley & Sons*, New York, London, Sidney, Toronto.

Gnedin, A.V.(1998), Sequential Selection of an Increasing Subsequence from a Sample of Random Size. To appear in *IMS Lecture Notes.*

Goldie, C.M. and Resnick, S.I. (1996) Ordered Independent Scattering. *Commun. Statist.-Stochastic Models*, **12** (4), 523-528.

Hammersley, J.M. (1972). A Few Seedlings of Research. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **1**, 345-394.

Jacod J. and Shiryavev A.N. (1980). Limit Theorems for Stochastic Processes. Springer Verlag Berlin, Heidelberg, New York, London Paris, Tokyo.

Logan, B.F. and L.A. Shepp (1977). A Variational Problem for Random Young Tableaux. *Advances in Math.*, **26**, 206 -222.

Protter, P. (1995). Stochastic Integration and Differential Equations. *Springer Verlag*, Berlin, Heidelberg, New York.

Samuels, S.M. and J.M. Steele (1981). Optimal Sequential Selection of a Monotone Sequence from a Random Sample. *Ann. Probab.* **9**, 937-947.

Ulam, S.M. (1972). Some Ideas and Prospects in Biomathematics. *Ann. Rev. Biophys. Bioengin.*, Vol.1, 277-292.

Vershik A.M. and S.V. Kerov (1977). Asymptotics of the Plancherel Measure of the Symmetric Group and the Limiting Form of the Young Tables. *Soviet Math. Dokl.* **18**, 527 - 531.

Authors' addresses

F. Thomas Bruss,
Université Libre de Bruxelles,
Département de Mathématique et ISRO,
Campus Plaine, CP 210, B-1050 Brussels.
(Email: tbruss@ulb.ac.be)

Freddy Delbaen,
Eidgenössische Technische Hochschule,
ETH-Zentrum, Departement für Mathematik,
CH-8092 Zürich.
(Email: delbaen@math.ethz.ch)