ON A POISSON MODEL FOR THE SIMPLEX ALGORITHM AND THE "SECRETARY PROBLEM"

D. Pfeifer, Aachen

Abstract

It is shown that the simplex algorithm of linear programming (under the assumptions introduced by Ross (1983)) and the "secretary problem" can be described (asymptotically) by the same stochastic model, involving Poisson point processes with either discrete or absolutely continuous mean measure. With these models, a rather precise average-case analysis of the simplex algorithm becomes available, as well as a generalization of the classical "secretary problem".

I. INTRODUCTION AND BASIC RESULTS

A) The simplex algorithm

The (primal) linear programming problem can be written in the form

$$(1.1) \quad \max_x \{c^t x \mid Ax \leq b,\ x \geq 0\}$$

where $c,x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{mxn}$ $(m,n \in \mathbb{N})$.

A well-known tool to solve iteratively problems of this form is the simplex method developed by Dantzig (see Hadley (1962) for references) which in its simplest form moves through the extreme points of the feasible solutions

$$(1.2) \quad \{x \geq 0 \mid Ax \leq b\}$$

in such a way that less or equally good points are successively excluded from further considerations, starting with $x = 0$ in the first step. Although the worst-case behaviour of this algorithm clearly is exponential in time it has proved to be very good "on average". This has been shown rigorously by Borgward (1982 a, 1982 b) in his pioneering papers where it was demonstrated that the average complexity of the simplex method is $O(n^4 m)$ under the assumption of spherically symmetric distributions for the parameters (cf. also the recent bibliography of O'h Eigeartaigh et al. (1985)). A different approach without specifying particular distributions for the parameters was made by Ross (1983) who assumed that the dynamic structure of this algorithm could be modelled by a Markov chain of the following type.

Suppose that the feasible region given by (1.2) contains N extreme points which w.l.o.g. are ranked according to the corresponding value of the gain function (in decreasing order). Let $\{X_n;\ n \in \mathbb{N}\}$ denote the Markov chain of ranks visited during the performance of the algorithm. Since X_1 is the rank of 0, an extreme point of (1.2), and this point could have any rank between 1 and N , Ross suggests that

it is justifiable to assume a uniform distribution for X_1 over the rank set $\{1,2,\ldots,N\}$. If in the last step, rank X_k was visited, then analogously each of the remaining X_k-1 extreme points could be the candidate for the next inspection, hence conditionally on $X_k = j$, rank $1,2,\ldots,j-1$ should be visited with equal probablitiy $\frac{1}{j-1}$ unless $j = 1$ in which case the optimal value is found (i.e. 1 is an absorbing state). (Note that the assumption of uniformity is mainly a working hypothesis which cannot be concluded e.g. from Borgward's assumptions.) Now, if

$$(1.3)\quad I_k = \begin{cases} 1, & \text{if rank k was visited} \\ 0, & \text{otherwise} \end{cases}, \quad 1 \leq k \leq N,$$

then $I_1,\ldots,I_N$ are actually <u>independent</u> random variables with

$$(1.4)\quad P(I_k = 1) = \frac{1}{k}, \quad 1 \leq k \leq N\ !$$

Clearly, $T = \sum_{k=1}^{N} I_k$ represents the number of steps required to find the optimal value, with $E(T) = \sum_{k=1}^{N} \frac{1}{k} \sim \log N$.

Since

$$(1.5)\quad N \leq \binom{n+m}{n} \sim (1 + \frac{m}{n})^n (1 + \frac{n}{m})^m \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

by Stirling's formula, we have asymptotically, for n,m large,

$$(1.6)\quad E(T) \leq 1 + \log N \leq n + m + 2\ .$$

Besides the estimation of E(T), much more can in fact be done here since the distribution of T is of Poisson-Bernoulli type for which precise Poisson approximations have been established recently, in various distance measures (Deheuvels and Pfeifer (1986 a, 1986 b, 1986 c)) which allow for a quite precise estimation of the distribution of T, P^T. With these estimations, it becomes possible to answer questions like these:

a) Given a (small) tolerance probability $\alpha < 1$, what is the maximum number of steps required to finish the procedure with probability $1 - \alpha$?

b) What is the probability of finishing the procedure with at most K steps (K given in advance)?

B) The "secretary problem"

Here we prefer a "neutral" presentation of the problem in order to avoid "female discrimination", following the excellent review of Freeman(1983), and in order to give an extension which would better fit under the heading "The travelling tourist problem". The following passages are adopted from Freeman's Introduction, describing the general problem.
"A known number N of items is to be presented one by one in random order,all N! possible orders being equally likely. The observer is able at any time to rank the items that have so far been presented in order of desirability. As each item is presented he must either accept it, in which case the process stops, or reject it, when the next item in the sequence is presented and the observer faces the same choice as before. If the last item is presented it must be accepted. The observer's aim is to maximize the probability that the item he chooses is, in fact, the best of the N items available."
A typical case for which this situation applies is a travelling tourist who has exactly one picture left on his film, and who aimes at photographing the most beautiful site during the remainder of his journey. A simple extension of this problem - which will also be considered in the sequel - would be given by the fact that the tourist has exactly K pictures left on his film, and he aims at catching the "best item" (i.e. the most beautiful site) with his reservoir of K.

The usual approach to the solution of the "secretary problem" is dynamic programming, and is - for the simple case above - be given by the following stopping rule:

a) Let $r = \min \{k \in \mathbb{N} \mid \sum_{i=k}^{N-1} \frac{1}{i} < 1\}$.

b) Reject the first r-1 items, and accept the first item which is better than everything before if such exists; otherwise take the last item.

It is easy to see that for N large, r is close to $\frac{N}{e}$, and the corresponding probability q of winning (i.e. of actually obtaining the best item) is

$$(1.7) \quad q = \frac{1}{N}(r-1) \sum_{i=r-1}^{N-1} \frac{1}{i} \sim \frac{1}{e} \log\left(\frac{N-1}{\frac{N}{e} - 1} \right) \sim \frac{1}{e} .$$

An alternative approach to this problem is as follows, paralleling the Markov chain approach of Dynkin (1963). Suppose that $\{Y_k \mid k \in \mathbb{N}\}$ is an iid sequence with continuous cdf F. Define record times $\{U_k \mid k \in \mathbb{N}\}$ by

$$(1.8)\quad U_1 \equiv 1\ ,\ U_{k+1} = \inf\{j \mid Y_j > Y_{U_k}\}\ ,\ k \in \mathbb{N}\ .$$

(This sequence is a.s. well-defined, cf. Shorrock (1972)). The record values then are $\{X_{U_k} \mid k \in \mathbb{N}\}$. The equivalent problem now is to detect the last record in a series of N observations $Y_1,\ldots,Y_N$. Surprisingly, Rényi (1962) has shown that the random variables

$$(1.9)\quad J_k = \begin{cases} 1, \text{ if } Y_k \text{ is a record value (i.e. } U_{J_k} = k) \\ 0, \text{ otherwise} \end{cases}\ ,\ 1 \leq k \leq N$$

again are <u>independent</u>, with

$$(1.10)\quad P(J_k = 1) = \frac{1}{k}\ ,\ 1 \leq k \leq N!$$

The problem hence is to detect the last "one" in the finite sequence $J_1,\ldots,J_N$. If we might-analogously to A)-assume that the counting (point) process

$$(1.11)\quad \xi((1,t]) = \sum_{1<k\leq t} J_k\ ,\ t \geq 1$$

is approximately a Poisson point process ξ with rate $\frac{1}{t}$ (i.e. $E(\xi((a,b])) = \log\frac{b}{a}$, $1 \leq a < b$), then - by the independent increments property - the best strategy would be to wait with a decision until some time s in such a way that the probability q^* of exactly one record (or one "one" in the J-sequence) in the interval $(s,N]$ is maximized. But

$$(1.12)\quad q^* \sim \lambda e^{-\lambda} \text{ with } \lambda = E\{\xi((s,N])\} = \log\frac{N}{s}$$

which is maximal for $\lambda = 1$, hence $\frac{N}{s} = e$ or $s = \frac{N}{e}$, which is the former asymptotic solution.

Moreover,

$$(1.13)\quad q^* \sim \lambda e^{-\lambda} = \frac{1}{e}$$

as expected.

To solve the K-reservoir problem correspondingly, we would have to wait until some time w such that the probability q^* of at most K, but at least one record in the interval (w,N] ist maximized, and to choose all items (i.e. take pictures of sites) after w which are better than all previous ones, if such exist, or to take the last one. Here we have (approximately, for large N)

$$(1.14)\quad q^* \sim \sum_{k=1}^{K} e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{with } \lambda = E\{\xi(w,N])\} = \log \frac{N}{w}$$

which is maximized for $\lambda = \sqrt[K]{K!}$, or

$$(1.15)\quad w = Ne^{-\lambda} = N \exp(-\sqrt[K]{K!}) \sim N \exp(-\frac{K}{e})$$

for large K. By the Central Limit Theorem, applied to (1.14), we would then approximately have, for large K (but $K << N$),

$$(1.16)\quad q^* \sim \phi((e-1)\frac{K}{e})$$

as winning probability where ϕ denotes the cdf of the standard normal distribution. (Note that the above solution is only suboptimal, but rather close to the optimal case; cf. Freeman (1983), 6.1.)

II. THE POINT PROCESS SETTING

From what has been said above it becomes clear that the goodness of the approximate approaches depends heavily on a suitable Poisson point process approximation for an independent sequence $\{I_n \mid n \in \mathbb{N}\}$ of Bernoulli random variables with

$$(2.1)\quad P(I_n = 1) = 1 - P(I_n = 0) = \frac{1}{n}, \quad n \in \mathbb{N}.$$

The corresponding Bernoulli point process ζ then is defined by

$$(2.2)\quad \zeta(A) = \sum_{k \in A} I_k, \quad A \in \mathfrak{S}$$

where $\mathfrak{B}$ denotes the σ-field of all Borel sets in IR. If we especially let - as after (1.4) -

$$(2.3)\quad T = \zeta([1,N]) \quad \text{for } N \in \mathbb{N} ,$$

then it has been proved in Deheuvels and Pfeifer (1986 a, 1986 b, 1986 c) that the (asymptotically, for N large) best Poisson approximation is given by the Poisson distribution $\mathcal{P}(\mu)$ with mean

$$(2.4)\quad \mu = \sum_{k=1}^{N} \frac{1}{k} \sim \gamma + \log N$$

where $\gamma = .5772$ denotes Euler's constant, with a total variation error of

$$(2.5)\quad d(P^T, \mathcal{P}(\mu)) = \sup_{M \subset \mathbb{Z}^+} |P(T \in M) - \mathcal{P}(\mu)(M)|$$

$$\sim \frac{\pi^2}{6\sqrt{2\pi e}} \frac{1}{\log N} = \frac{.398}{\log N} .$$

If we would instead consider the Kolmogorov distance

$$(2.6)\quad d_0(P,Q) = \sup_{x \in \mathbb{R}} |F_P(x) - F_Q(x)|$$

where F_P , F_Q denote the cdf's of the probability measures P, Q, then we would obtain

$$(2.7)\quad d_0(P^T, \mathcal{P}(\mu)) \sim \frac{\pi^2}{12\sqrt{2\pi e}} \frac{1}{\log N} = \frac{.199}{\log N} .$$

This estimation is especially suitable for answering the problems a) and b) at the end of chapter I b).
For the more general point process setting, there are at least three different ways how to proceed.
First consider the coupling approach by Karr and Serfling (1985) which in our case works as follows.
Let $\{Z_n \mid n \geq 2\}$ be an independent sequence of Poisson random variables with means

$$(2.8)\quad \mu_n = E(Z_n) = -\log\left(1 - \frac{1}{n}\right),\ n \geq 2$$

and define the (Poisson) point process ξ by

$$(2.9)\quad \xi(A) = \sum_{\substack{k \in A \\ k \geq 2}} Z_k\ ,\ A \in \mathcal{B}.$$

Let further

$$(2.10)\quad J_n = \begin{cases} Z_n\ , & \text{if } Z_n \leq 1 \\ 1\ , & \text{if } Z_n \geq 2 \end{cases}\ ,\ n \geq 2.$$

Then the J-sequence is identically distributed with the original I-sequence since

$$(2.11)\quad P(J_n = 0) = P(Z_n = 0) = e^{-\mu_n} = 1 - \frac{1}{n} = P(I_n = 0)\ ,\ n \geq 2.$$

Obviously, now

$$(2.12)\quad \zeta(A) \overset{\mathcal{D}}{=} \sum_{k \in A} J_k\ ,\ A \in \mathcal{B}\ ,\ 1 \notin A$$

where $\overset{\mathcal{D}}{=}$ means equality in distribution.
Let for simplicity denote $\zeta_B = \zeta(. \cap B)$, $\xi_B = \xi(. \cap B)$, $B \in \mathcal{B}$ denote the point process restricted to the set $B \in \mathcal{B}$, where we also assume that $1 \notin B$ (hence $\zeta_B(A) \overset{\mathcal{D}}{=} \sum_{k \in A \cap B} J_k$, $A \in \mathcal{B}$).

Then with the arguments in Karr and Serfling (1985), we can show that

$$(2.13)\quad d(\zeta_B, \xi_B) \leq \frac{1}{2} \sum_{\substack{k \in B \\ k \geq 2}} \log^2\left(1 - \frac{1}{k}\right) \leq \sum_{\substack{k \in B \\ k \geq 2}} \frac{1}{k^2}$$

which indicates that ξ is a good approximation for ζ if inf(B) is large, which is the case in the "secretary problem" where $B = \left[\frac{N}{e}, \infty\right)$ (in the standard version), hence, for large N,

$$(2.14)\quad d(\zeta_B, \xi_B) \leq \sum_{k \geq \frac{N}{e}} \frac{1}{k^2} \leq \frac{e}{N-e}\ .$$

We could thus have used the Poisson point process ξ with discrete mean measure

$$(2.15)\quad E(\xi(A)) = \sum_{\substack{k\in A\\ k\geq 2}} \mu_k = \sum_{\substack{k\in A\\ k\geq 2}} -\log(1-\frac{1}{k}) \ , \ A \in \mathcal{B}$$

for evaluating the approximate solution to the "secretary problem", which, however, by its discrete nature is a little cumbersome. A more elegant way is to consider the corresponding Poisson point process ξ^* with absolutely continuous mean measure

$$(2.16)\quad E(\xi^*(A)) = \int_{A\cap \mathbb{R}^+} \frac{1}{t}\, dt \ , \ A \in \mathcal{B}.$$

Then obviously, for a set $A = (n-1, n]$, $n \geq 2$, we have

$$(2.17)\quad E(\xi^*(A)) = \int_{n-1}^{n} \frac{1}{t}\, dt = -\log(1-\frac{1}{n}) = \mu_n \ , \ n \geq 2$$

which means that ξ^* "distributes" multiple points of ξ (which may occur at integer times $n \geq 2$) into the adjacent intervals $(n-1,n]$. This means that distributionally the processes ξ and ξ^* coincide for all these sets $A = (n-1,n]$, $n \geq 2$, hence the procedure outlined in Chapter I b) (where only sets of the form $(1,t]$ are considered) is justifiable.
It should be pointed out that the point process ξ^* also arises from an inbedding of the sequence $\{\max_{1\leq k\leq n} Y_k \mid n \in \mathbb{N}\}$ into a so-called extremal process (which is a continuous-time Markov process), providing a continuous analogue of the record times defined in (1.8) (see e.g. Resnick (1973) and Pfeifer (1985)).

Unfortunately, the approach by Karr and Serfling (1985) does not provide the best possible approximation since with

$$(2.18)\quad T^* = 1 + \xi((2,N]) \ , \ N \geq 2$$

we would only get a Poisson approximation for T in (2.3) of the order

$$(2.19)\quad d(P^T, P^{T^*}) = O\left(\frac{1}{\sqrt{\log N}}\right) \ , \ N \to \infty$$

(cf. also Deheuvels and Pfeifer (1986 a, 1986 b, 1986 c)). A third method of treating the problems outlined in Chapter I could thus be obtained by a suitable "thinning" of the point process ξ^* in the following way:
Let $\{W_n \mid n \in \mathbb{N}\}$ be an independent sequence of Bernoulli random variables with

$$(2.20)\quad P(W_n = 1) = 1 - P(W_n = 0) = n(1 - e^{-1/n}) < 1,\ n \in \mathbb{N}\ .$$

Then the sequence $\{W_n J_n \mid n \geq 2\}$ also is independent Bernoulli with

$$(2.21)\quad P(W_n J_n = 1) = P(W_n = 1)\, P(J_n = 1) = 1 - e^{-1/n},\ n \in \mathbb{N}.$$

If we now consider the Karr-Serfling approximation ξ^{**} , say for the latter sequence, then we would in fact have a Poisson point process with

$$(2.22)\quad E(\xi^{**}(\{n\})) = \frac{1}{n}\ ,\ n \in \mathbb{N}\ ,$$

for which

$$(2.23)\quad T^{**} = \xi^{**}([1,N]),\ n \geq 1$$

would realize the (asymptotically) optimal Poisson approximation for T. Note that here the point process ξ^* would yield

$$(2.24)\quad E(\xi^*([1,N])) = \int_1^N \frac{1}{t}\, dt = \log N,\ N \geq 1$$

which differs only (asymptotically) by Euler's constant γ from the true mean E(T). Nevertheless, an approximation with a $\mathcal{P}(\log N)$-distribution for T would still give an error of

$$(2.25)\quad d(P^T, \mathcal{P}(\log N)) \sim \frac{\gamma}{\sqrt{2\pi \log N}} \qquad (N \to \infty)$$

(cf. Deheuvels and Pfeifer (1986 b)).

REFERENCES

Borgward, K.H. (1982 a): Some distribution-independent results about the asymptotic order of the average number of pivot steps of the simplex method. Math. Oper. Res. 7, 441 - 462.

Borgward, K.H. (1982 b): The average number of pivot steps required by the simplex-method is polynomial. Z. Oper. Res. 26, 157 - 177.

Deheuvels, P. and Pfeifer, D. (1986 a): A semigroup approach to Poisson approximation. Ann. Prob. 14, 663 - 676.

Deheuvels, P. and Pfeifer, D. (1986 b): Semigroups and Poisson approximation. In: Perspectives And New Directions In Theoretical And Applied Statistics, M. Puri, J.P. Villaplana, W. Wertz, eds.; Wiley, N.Y.

Deheuvels, P. and Pfeifer, D. (1986 c): Operator semigroups and Poisson convergence in selected metrics. To appear in: Semigroup Forum.

Dynkin, E.B. (1963): The optimal choice of the stopping moment for a Markov process. Dokl. Akad. Nauk. SSSR 150, 238 - 240.

Freeman, P.R. (1983): The secretary problem and ist extensions: A review. Int. Stat. Review 51, 189 - 206.

Hadley, G. (1962): Linear Programming. Addison-Wesley, Reading, Mars.

Karr, A. and Serfling, R.J. (1985): Poisson approximation of Bernoulli point processes and their superpositions, via coupling. Johns Hopkins University Baltimore, USA, Tech. Rept. 458.

O'h Eigeartaigh, M., Lenstra, J.K., and Rinnooy Kan, A.M.G. (1985): Combinatorial Optimization - Annotated Bibliographies. Wiley, N.Y.

Pfeifer, D. (1985): Extremal processes, record times and strong approximation. To appear in: Publ. Inst. Stat. Paris.

Rênyi, A. (1962): Thêorie des êlêments saillants d'une suite d'observations. Coll. on Combinatorial Methods in Prob. Th., 104 - 115 , Mathematisk Institut, Aarhus Universitet, Denmark.

Resnick, S.I. (1973): Extremal processes and record value times. J. Appl. Prob. 10, 863 - 868.

Ross, S.M. (1983): Stochastic Processes. Wiley, N.Y. (Chapter IV).

Shorrock, R.W. (1972): On record values and record times. J. Appl. Prob. 9, 316 - 326.

Priv.-Doz. Dr. Dietmar Pfeifer
Institut für Statistik und Wirtschaftsmathematik
RWTH Aachen, Wüllnerstr. 3
5100 Aachen