A Poisson Fishing Model[1]

Thomas S. Ferguson, 08/30/94[2]

Abstract: A fishing model of Starr, Wardrop, and Woodroofe is related to the sequential search model of Cozzolino. The latter is generalized to allow an arbitrary joint distribution of capture times and fish sizes. Implications to the foraging models of Oaten and Green and to debugging software are indicated.

0. Introduction.

The central theme of this paper is a result in sequential analysis that has application to a wide variety of problems. These problems have appeared in papers dealing with sequential estimation in statistics, estimation of the number of species, the fishing problem, the proofreading problem, auditing, foraging, search, etc. Authors in different areas are not always aware of each other, and so often recompute the basic results again. This is partly because the basic assumptions required in the different areas necessarily differ in signficant ways. Yet the main result in Section 4 of this paper would be of interest in all these areas. Since the basic assumption of the result is that a certain parameter has a Poisson distribution, it seems appropriate to use the model of Starr, Woodroofe and others as a fishing problem.

1. The Fishing Problem.

One of the first papers to deal with the fishing problem was Starr (1974). The main result of this paper is easy to state. There are m fish in a lake, where m is known. The capture time of fish j if one fishes indefinitely is T_j. We assume $T_1, \ldots, T_m$ are i.i.d. exponential with hazard rate r. Let $K(t)$ denote the number of fish caught by time t, so that

$$K(t) = \sum_{j=1}^{m} \mathrm{I}(T_j \le t), \tag{1}$$

where I denotes the indicator function. There is a constant cost of time, so the payoff for stopping at time t is

$$Y_t = K(t) - ct. \tag{2}$$

[1] Talk given at the XXIV$^{\text{es}}$ Journées de Statistique, Brussels, May 20, 1992.

[2] University of California, Los Angeles, USA.

The problem is to find a stopping rule τ to maximize $\mathrm{E}Y_\tau$. Starr shows that the optimal stopping rule is

$$\tau = \inf\{t \geq 0 : K(t) \geq m - c/r\}. \tag{3}$$

This is a fixed number of captures rule: Fish until there are at most c/r fish left.

This was generalized by Starr and Woodroofe (1974). Again there are m fish and the payoff is given by (2), but now the capture times $T_1, \ldots, T_m$, assumed to be i.i.d. positive random variables with absolutely continuous distribution function $F(t)$. There are two basic results of their paper.

The first is as follows. If F has increasing failure rate (IFR), then it is optimal to stop only at catch times. This allows one to discretize the problem and solve it by backward induction. However, the optimal rule will not generally be a fixed number of captures rule; the optimal decision to stop may also depend on time.

In spite of this, the easy case turns out to be the case when F has decreasing failure rate (DFR). In this case, then it may be optimal to stop between capture times, but the infinitesmal look-ahead rule is optimal. The theory for this rule is developed by Ross (1971). Its application to the present problem produces the optimal rule,

$$\tau = \inf\{t \geq 0 : (m - K(t)) r_F(t) \leq c\}, \tag{4}$$

where $r_F(t)$ is the failure rate of F, namely, $r_F(t) = f(t)/(1 - F(t))$, where $f(t)$ is the density of F.

This result for DFR has been extended by Starr, Wardrop and Woodroofe (1976) to allow the payoff to be of the form

$$Y_t = g(K(t)) - c(t), \tag{5}$$

where g is concave utility function, and c is a convex cost function. The corresponding optimal rule is

$$\tau = \inf\{t \geq 0 : (m - K(t))(g(K(t)+1) - g(K(t))) r_F(t) \leq c'(t)\}. \tag{6}$$

The main application of this result, that motivated generalizing the payoff, is to the statistical problem of estimating the mean of a normal distribution with "delayed" observations. If you start off an experiment with m experimental units and the observations come in sequentially and sporadically, and if you are paying a cost in real time, you might want to stop the experiment early rather than wait for the last observation. If you estimate the mean after $K(t)$ observations, you incur a terminal loss of $\sigma^2/K(t)$. Since $g(k) = -\sigma^2/k$ is concave in k, the rule (6) is optimal.

A further extension is made in Kramer and Starr (1990). In this paper, the fish are allowed to have different sizes, and the time of capture may be dependent on the size. If we let the size of fish j be denoted by X_j, then the basic assumption of this model is that $(X_1,T_1),\ldots,(X_m,T_m)$ are i.i.d. with absolutely continuous distributions with $T_j > 0$ a.s. and $\mathrm{E}|X| < \infty$. The payoff for stopping at time t is now the total catch size, $R(t)$, minus a cost of time,

$$Y_t = R(t) - c(t) = \sum_{i=1}^{m} X_i \mathrm{I}(T_i \le t) - c(t). \tag{7}$$

The infinitesimal look-ahead rule for this problem is

$$\tau = \inf\{t \ge 0 : (m - K(t))\mathrm{E}(X|T=t)r_F(t) \le c'(t)\} \tag{8}$$

where $r_F(t)$ is the failure rate of the marginal distribution of T. Since it is assumed that $c(t)$ is convex, this rule is optimal provided $\mathrm{E}(X|T=t)r_F(t)$ is nonincreasing in t. In particular, if F has DFR and if $\mathrm{E}(X|T=t)$ is nonincreasing in t, then τ of (8) is optimal. When $\mathrm{E}(X|T=t)$ is nonincreasing in t, bigger fish are easier to catch. This is a natural assumption for Kramer and Starr because they are interested in exploration for oil, where the bigger deposits are easier to find. In fact, because of a nice theorem of theirs from an earlier paper, they restrict attention to joint distributions of X and T for which $\mathrm{P}(T > t|X=x) = (1 - H(t))^x$ for some distribution function H. The nice theorem states that with appropriate regularity conditions, this is a necessary and sufficient condition that sampling becomes proportional to size. That is, at all times t conditional on the sizes $X_1,\ldots,X_m$ of the uncaught fish, the probability that the ith fish is caught next is $X_i/(X_1 + \cdots + X_m)$. (It is assumed that $X_i > 0$ a.s.) Sampling proportional to size is a natural assumption for oil exploration. If H has DFR, then Kramer and Starr show that $\mathrm{E}(X|T=t)r_F(t)$ is nonincreasing so that (8) is optimal.

2. A Search Problem.

Let us go back to an earlier version of this problem found in a paper of Cozzolino (1972). This paper is in the area of optimal allocation of search effort initiated by B. O. Koopman in the late 1950's. In the terminology of the fishing problem, the fish are allowed to have different sizes or values that are dependent on their catch times as in Kramer and Starr. However, the number of fish is allowed to be unknown. This is an important generalization since it is rare in applications that m is known exactly. It is useful to express one's uncertainty of m in a prior probability distribution that may then be updated as information is received. Specifically, Cozzolino assumes that the number of fish, M, has a Poisson distribution with parameter λ, denoted by $\mathcal{P}(\lambda)$. Given $M = m$, the sizes of the fish, $X_1,\ldots,X_m$ are assumed to be i.i.d. with a gamma distribution, and given M and $X_1,\ldots,X_m$, the capture times are assumed to be independent, with T_i having an exponential distribution at rate γX_i. Bigger fish are easier to catch. In symbols,

$$\begin{gathered} M \in \mathcal{P}(\lambda) \\ (X_1,T_1),\ldots,(X_M,T_M)|M \quad \text{i.i.d.} \\ X_j \in \mathcal{G}(\alpha,\beta) \\ T_j|X_j \in \mathcal{G}(1,\gamma X_j) \end{gathered} \tag{9}$$

where $\mathcal{G}(\alpha, \beta)$ denotes the gamma distribution with density proportional to $\exp(-\beta x)x^{\alpha-1}$. The payoff is the sum of the sizes of the fish caught minus a constant cost per unit time, $Y_t = R(t) - ct$. In addressing the problem of when to stop searching, Cozzolino finds that there is an optimal fixed time rule,

$$\tau = \inf\{t \geq 0 : \lambda \mathrm{E}(X|T = t)f(t) \leq c\} = \frac{\beta}{\gamma}\left[\left(\frac{\lambda\gamma\alpha(\alpha+1)}{c\beta^2}\right)^{1/(\alpha+2)} - 1\right]^+ . \tag{10}$$

In contrast to the case with a known number of fish when the optimal rule (3) of Starr stops after a fixed number of catches, the optimal rule with a Poisson number of fish stops at a fixed time.

3. The Species Problem.

Another problem related to this is the species problem. The standard objective in the species problem is to estimate the number of unobserved species or the probability of observing a new species. But our interest is in the stopping problem. When should one abandon the search for new species? This problem was investigated in Rasmussen and Starr (1979) "Optimal and adaptive stopping in the search for a new species". Their formulation of the problem is a discrete version of the fishing problem outlined above. Consider an infinite population consisting of m subpopulations, or species. Let p_i denote the proportion of members of the population that belong to species i. Selections are made from the population sequentially at random and the species of the selection is noted. It is assumed that each trial is independent and the probability that the ith species is observed is p_i for each trial. For $i = 1, \ldots, m$ and $n = 0, 1, 2, \ldots$, let $X_i(n)$ denote the number of times species i is observed among the first n observations. Let $K(n) = \sum_1^m \mathrm{I}(X_i(n) > 0)$ denote the number of distinct species observed in the first n trials and let $u(n) = \sum_1^m p_i \mathrm{I}(X_i(n) = 0)$ denote the total probability of the unobserved species. The reward for stopping at stage n is $Y_n = g(K(n)) - nc$, where g is a concave function on the integers and $c > 0$ is a constant. The one-stage look-ahead rule (the 1-sla) is optimal for this problem. It is

$$N = \min\{n \geq 0 : [g(K(n) + 1) - g(K(n))]u(n) \leq c\}. \tag{11}$$

The trouble with this analysis for the species problem is that m and the p_i are assumed known, and which species is associated with p_i is also known. However, since the rule N depends on this knowledge only through $u(n)$, Rasmussen and Starr suggest using Turing's nonparametric estimate of this quantity in (11). This estimate is

$$v(n) = \frac{1}{n}\sum_1^m \mathrm{I}(X_i(n) = 1). \tag{12}$$

Numerical computation shows that the adaptive strategy (11) with $u(n)$ replaced by $v(n)$ for stopping compares well with the original rule.

I would like to suggest another way to deal with this problem. Namely, the Bayes solution in which the prior distribution of $(p_1, \ldots, p_m)$ is taken to be the Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_m)$. One good feature about this approach is that the one-stage look-ahead rule is still optimal for this adaptive problem. It is (11) with $u(n)$ replaced by its Bayes estimate,

$$w(n) = \frac{\sum_1^m \alpha_i \mathrm{I}(X_i(n) = 0)}{n + \sum_1^m \alpha_i}. \tag{13}$$

The drawback of presuming to know which species are associated with which α_i still remains, but at least $w(n) \to 0$ as $n \to \infty$, even if m is incorrectly specified too large. One may also estimate the unknown parameters, (m may even be taken to be infinite), thus providing an adaptive Bayes solution.

Banerjee and Sinha (1985) extend Rasmussen and Starr to sampling in batches of size $k > 1$. They also propose a new estimator of the probability of discovering a new species.

Alsmeyer and Irle (1989) put the problem in continuous time, allow stochastic intensities depending on the past, $\lambda_i(t)$ for species i, and allow the reward, r_i to depend on the species, i. As an example, they take constant intensities, and find as the optimal stopping rule, similar to (11),

$$\tau = \inf\{t \geq 0 : \sum_1^m r_i \lambda_i \mathrm{I}(X_i(t) = 0) \leq c\}. \tag{14}$$

It is assumed that it is known which species are attached to which intensities and rewards. So the fishing model is more appropriate than the species model for this problem.

4. The Basic Fishing Model.

We generalize Cozzolino's formulation as follows. There are a random number, M, of fish. The distribution of M is known and $\mathrm{E}M < \infty$. Given $M = m$, the sizes and times of capture, $(X_1, T_1), \ldots, (X_m, T_m)$, are i.i.d. 2-dimensional random vectors with $\mathrm{E}|X_i| < \infty$ and $T_i > 0$ a.s., with known distribution function, $F(x, t)$. As before, we let for fixed t,

$$\begin{aligned} K(t) &= \textstyle\sum_{i=1}^M \mathrm{I}(T_i \leq t) = \quad \text{\# fish caught by time } t \\ R(t) &= \textstyle\sum_{i=1}^M X_i \mathrm{I}(T_i \leq t) = \quad \text{total value of fish caught} \end{aligned} \tag{15}$$

The payoff we receive if we stop at time t is $Y_t = R(t) - c(t)$, where $c(t)$ is a given increasing function of t. An optimal (finite-valued) stopping time exists if $\mathrm{E} \sup_t Y_t < \infty$. This is true under the assumptions that $\mathrm{E}M < \infty$ and $\mathrm{E}X^+ < \infty$. The infinitesimal look-ahead rule is

$$\tau = \inf\{t \geq 0 : \mathrm{E}(M - K(t)|\mathcal{F}_t)\mathrm{E}(X|T = t) r_F(t) \leq c'(t)\} \tag{16}$$

where $\mathcal{F}_t$ denotes the σ-field generated by the observations up to time t. Moreover, $\mathrm{E}(M - K(t)|\mathcal{F}_t)$ depends only on $K(t)$. In our model, this rule is optimal if the problem

is monotone. The problem is monotone if the validity of the inequality in (16) at $t = t_0$ implies its validity a.s. for all $t > t_0$. For monotonicity, it suffices to have
1. $c'(t)$ to be non-decreasing (i.e. c convex.)
2. $\mathrm{E}(M - K(t)|\mathcal{F}_t)$ to be non-increasing in t a.s.
3. $\mathrm{E}(X|T = t)r_F(t)$ to be non-increasing in t.

The first condition is standard. It was used by Starr, Wardrop and Woodroofe. The third condition breaks into two conditions. It is satisfied if both $\mathrm{E}(X|T = t)$ is non-increasing (bigger fish are easier to catch) and $r_F(t)$ is nonincreasing (T has DFR). However, it is easy to see that it can be satisfied more generally.

The second condition is the critical one. Let us consider some special cases.

(a) *M degenerate at m.* Then $\mathrm{E}(M - K(t)|\mathcal{F}_t) = m - K(t)$ a.s., which is non-increasing a.s. so that condition 2 is satisfied.

(b) *M Poisson, $\mathcal{P}(\lambda)$.* Then $(M - K(t))|\mathcal{F}_t$ has the Poisson distribution, $\mathcal{P}(\lambda \mathrm{P}(T > t))$, so that $\mathrm{E}(M - K(t)|\mathcal{F}_t) = \lambda P(T > t)$. This is a non-random function, non-increasing in t, so again condition 2 is satisfied.

In this case, something more interesting is true. Namely, there is an optimal fixed time rule whether or not the problem is monotone, whether or not bigger fish are easier to catch. This is because at time t, the future is independent of the past. Thus, in Cozzolino's problem, there is an optimal fixed time rule whether or not the assumptions on T and X are satisfied. This means that the optimal rule may be found as a simple maximization problem, namely, find t to maximize

$$\mathrm{E}Y_t = \mathrm{E}\sum_1^M X_i \mathrm{I}(T_i \le t) - c(t) = \lambda \mathrm{E}(X\mathrm{I}(T \le t)) - c(t).$$

The derivative with respect to t is

$$\frac{d}{dt}\mathrm{E}Y_t = \lambda \mathrm{E}(X|T = t)f(t) - c'(t). \tag{17}$$

There exists a unique root of this expression if and only if the problem is monotone. In this case the optimal rule reduces to

$$\tau = \inf\{t \ge 0 : \lambda \mathrm{E}(X|T = t)f(t) \le c'(t)\}. \tag{18}$$

If the problem is not monotone, we must inspect each of the negative-going roots of (17) to find the value.

(c) *M has the binomial distribution, $\mathcal{B}(W, \pi)$.* Then $M - K(t)|\mathcal{F}_t$ has the binomial distribution, $\mathcal{B}(W - K(t), \pi(t))$, where $\pi(t) = \pi \mathrm{P}(T > t)/[1 - \pi + \pi \mathrm{P}(T > t)]$. Hence, $\mathrm{E}(M - K(t)|\mathcal{F}_t) = (W - K(t))\pi(t)$, and the rule (16) becomes

$$\tau = \inf\{t \ge 0 : (W - K(t))\pi \mathrm{E}(X|T = t)f(t) \le c'(t)[1 - \pi + \pi \mathrm{P}(T > t)]\}. \tag{19}$$

(d) *M has the negative binomial distribution,* $\mathcal{NB}(\alpha, 1/(\beta+1))$. This distribution arises when $M|\lambda$ has the Poisson distribution, $\mathcal{P}(\lambda)$ and λ has the gamma distribution, $\mathcal{G}(\alpha,\beta)$. Then $M - K(t)|\mathcal{F}_t \in \mathcal{NB}(K(t)+\alpha, P(T>t)/(\beta+1))$. We have $\mathrm{E}(M-K(t)|\mathcal{F}_t) = (K(t)+\alpha)\mathrm{P}(T>t)/(\beta+\mathrm{P}(T\le t))$. Although $\mathrm{P}(T>t)/(\beta+\mathrm{P}(T\le t))$ is non-increasing, the other term $K(t)+\alpha$ increases in jumps at the time of each observation. Thus the problem is not monotone. The infinitesimal look-ahead rule can be improved.

(e) *M is beta-binomial,* $\mathcal{BB}(W,\alpha,\beta)$. This distribution arises when $M|\pi$ has the binomial distribution, $\mathcal{B}(W,\pi)$ and π has the beta distribution, $\mathcal{B}e(\alpha,\beta)$. To make the family of distributions closed under prior-to-posterior analysis, it is necessary to add another parameter. The 4-parameter beta-binomial $\mathcal{BB}(W,\alpha,\beta,q)$ with $W \ge 0$ integer, $\alpha > 0$, $\beta > 0$, and $q > 0$, is defined as the distribution with mass function proportional to

$$f(m|W,\alpha,\beta,q) \propto \binom{W}{m} B(\alpha+m, \beta+W-m)q^m,$$

where $B(\alpha,\beta)$ represents the beta function. When $q=1$ this is the beta-binomial distribution, and when $\beta = 1$ and $0 < q < 1$, this is the negative binomial distribution truncated at W. If T has a continuous distribution, this problem is never monotone.

5. Proofreading and Testing Computer Software.

In proofreading and in testing computer software, the problem is usually to estimate the number of errors remaining after a debugging process. M represents the initial number of misprints or bugs in the program. I'll mention two papers. In the software debugging paper of Dalal and Mallows (1988), the model is as follows. M has a negative binomial prior distribution, $M \in \mathcal{NB}(\alpha, 1/(1+\beta))$, all bugs are equally valuable to detect, and the times of detection are i.i.d. This is an important problem and Dalal and Mallows suggest a method of solving it in a fairly general setting. When M is Poisson with known mean λ, they note that the optimal rule in their setting is a fixed time rule. When λ is large and there is a large number of observations taken before stopping, one may obtain an adaptive approximately optimal solution by replacing λ in the optimal rule by its estimate, $K(t)/\mathrm{P}(T\le t)$.

In the proofreading paper of Ferguson and Hardwick (1989), the basic model is followed but the setting is discrete and the marginal distribution of T is taken to be a mixture of geometrics. If $\mathrm{P}(T=t) = \mathrm{E}(Q^t(1-Q))$, then $\mathrm{E}(X|T=t)f(t) = \mathrm{E}[Q^t(1-Q)X]$ is decreasing in t, so that in the Poisson case, the 1-sla optimal in general. In the beta-binomial model, the distribution of catch time is discrete, and the 1-sla is optimal in some important cases.

6. Foraging.

Consider an animal that forages for food in spacially separated patches of prey. He feeds at one patch for awhile and then moves on to another. The problem of when to move to a new patch in order to maximize the rate of energy intake is addressed in the paper of Oaten (1977). His results have been extended in a number of ways by Green (1980, 1987).

For example, take the fisherman who moves from waterhole to waterhole catching fish. Here, M represents the number of fish in a waterhole. Given $M = m$, the distribution of the times of catch $T_1, \ldots, T_m$ is known. The expected time to travel from one patch to another is also a known constant, say 1, as is the expected energy required for the trip, a. The problem Oaten and Green consider is to choose a stopping time τ to move to the next waterhole in order to maximize the rate of return, $(\mathrm{E}K(\tau) - a)/(\mathrm{E}\tau + 1)$.

Green in his 1987 paper treats discrete time and considers six cases, when the distribution of M is {degenerate, Poisson, negative binomial}, and when the distribution of catch times is {uniform, exponential}. The reason for considering the uniform case is interesting. Green is very much attuned to applications, and he is especially interested in birds. It seems that certain birds are systematic foragers, that is they are active in their search for prey and avoid covering the same ground twice. For systematic foragers, the time needed to catch a given prey would be approximately uniform over the time needed to cover the whole patch.

With the observations we have made about the Poisson fishing model, we may extend the model of Oaten/Green to allow size of catch to depend on time. Therefore, our model is given by M with a known distribution, $(X_1, T_1), \ldots, (X_M, T_M)|M$ i.i.d. with a known distribution independent of M. We are to maximize $(\mathrm{E}R(\tau) - a)/(\mathrm{E}\tau + 1)$. This problem may be related to the problem of finding a stopping rule to maximize the return $\mathrm{E}(R(\tau) - a - \phi\tau - \phi)$ and then to adjust ϕ so that the optimal return is zero. The resulting ϕ is the optimal rate of return and the optimal rules for the two problems are the same.

In the Poisson case, since the optimal rule is a fixed time rule, we need only compute

$$\mathrm{E}(R(t) - a - \phi t - \phi) = \mathrm{E}M\,\mathrm{E}(X\mathrm{I}(T \le t)) - a - \phi t - \phi$$

Setting this to zero gives one equation, and setting the derivative to zero gives a second equation, to be solved jointly for t and ϕ. Eliminating ϕ from these two equations gives

$$\mathrm{E}M\,\mathrm{E}(X\mathrm{I}(T \le t)) = a + (t+1)\mathrm{E}M\,\mathrm{E}(X|T = t)f(t). \tag{20}$$

As an example, suppose T has the inverse power distribution with density, $f(t) = \theta(1+t)^{-(\theta+1)}$, and suppose that the expectation of X given $T = t$ is $\mathrm{E}(X|T = t) = \alpha(1+t)^{\gamma}$ where $\gamma < \theta$. (This arises, for example, when the distribution of Z given $T = t$ is the gamma, $\mathcal{G}(\alpha, (1+t)^{-\gamma})$.) If $\gamma < 0$, bigger fish are easier to catch, and if $\gamma > 0$, smaller fish are easier to catch. In this example, (20) can be solved explicitly for t to give

$$t = (A^{1/(\theta-\gamma)} - 1)^{+}$$

as the optimal time to stop, where

$$A = \frac{(\theta - \gamma + 1)\lambda\alpha\theta}{\lambda\alpha\theta - a(\theta - \gamma)}.$$

References.

G. Alsmeyer and A. Irle (1989) "Optimal strategies for discovering new species in continuous time" *J. Appl. Prob.* **26**, 695-706.

P. K. Banerjee and B. K. Sinha (1985) "Optimal and adaptive strategies in discovering new species" *Sequential Anal.* **4***(1&2)*, 111-122.

J. M. Cozzolino (1972) "Sequential search for an unknown number of objects of nonuniform size" *Op. Res.* **20**, 293-308.

S. R. Dalal and C. Mallows (1988) "When should one stop testing software?" *J. Amer. Statist. Assoc.* **83**, 772-779.

T. S. Ferguson and J. P. Hardwick (1989) "Stopping rules for proofreading" *J. Appl. Prob.* **26**, 304-313.

R. F. Green (1980) "Bayesian birds: A simple example of Oaten's stochastic model of optimal foraging" *Theor. Pop. Biol.* **18** 244-256.

R. F. Green (1987) "Stochastic models of optimal foraging" (in) *Foraging Behavior*, A. C. Kamil, J. R. Krebs, H. R. Pullman, Eds. Plenum Press, New York, 273-302.

Kramer and N. Starr (1990) "Optimal stopping in a size dependent search" *Sequential Anal.* **9**, 59-80.

A. Oaten (1977) "Optimal foraging in patches: A case for stochasticity" *Theor. Population Biol.* **12**, 263-285.

S. Rasmussen and N. Starr (1979) "Optimal and adaptive search for a new species" *J. Amer. Statist. Assoc.* **74**, 661-667.

S. M. Ross (1971) Infinitesimal look-ahead stopping rules *Ann. Math. Statist.* **42**, 297-303.

N. Starr (1974) Optimal and adaptive stopping based on capture times *J. Appl. Prob.* **11**, 294-301.

N. Starr and M. Woodroofe (1974) "Gone fishin': Optimal stopping based on catch times" U. Mich. Tech. Report #33, Dept. of Statistics.

N. Starr, R. Wardrop and M. Woodroofe (1976) "Estimating a mean from delayed observations" *Z. fur Wahr.* **35**, 103-113.