Determining the Best Trade-Off Between Expected Economic Return and the Risk of Undesirable Events When Managing a Randomly Varying Population

ROY MENDELSSOHN

Southwest Fisheries Center, National Marine Fisheries Service, NOAA, Honolulu, HI 96812, USA

MENDELSSOHN, R. 1979. Determining the best trade-off between expected economic return and the risk of undesirable events when managing a randomly varying population. J. Fish. Res. Board Can. 36: 939–947.

Conditions are given that imply there exist policies that "minimize risk" of undesirable events for stochastic harvesting models. It is shown that for many problems, either such a policy will not exist, or else it is an "extreme" policy that is equally undesirable. Techniques are given to systematically trade-off decreases in the long-run expected return with decreases in the long-run risk. Several numerical examples are given for models of salmon runs, when both population-based risks and harvest-based risks are considered.

Key words: Markov decision processes, risk, salmon management, Pareto optimal policies, trade-off curves, linear programing

MENDELSSOHN, R. 1979. Determining the best trade-off between expected economic return and the risk of undesirable events when managing a randomly varying population. J. Fish. Res. Board Can. 36: 939–947.

L'auteur expose les conditions qui sous-entendent l'existence de politiques aptes à minimiser le risque d'événements indésirables dans l'élaboration de modèles stochastiques de récolte. Il démontre que, pour plusieurs problèmes, une telle politique n'existe pas, ou encore on a une politique «extrême», également indésirable. Il décrit des méthodes selon lesquelles une diminution des gains anticipés à long terme peut être systématiquement échangée contre une diminution du risque à long terme. Il donne plusieurs exemples pour modèles de remontées de saumons, dans lesquels sont considérés risques inhérents à la population et risques inhérents à la récolte.

Received December 28, 1978
Accepted March 26, 1979

Reçu le 28 décembre 1978
Accepté le 26 mars 1979

I. Introduction

AN important consideration in managing ecosystems or wildlife populations is the avoidance of undesirable occurrences, either to the populations themselves, or to the economic agents acting on them. This might mean minimizing the chance that a population becomes too small, or minimizing environmental damage, or perhaps minimizing the chance that the economic return is too small. Such ideas of minimum risk are inherent in such legal concepts as optimal sustainable population (OSP), required for the management of marine mammals in the U.S., or in management under extended jurisdiction, where both the health of the stocks and the health of the industry are vital concerns.

Underlying these concerns is some notion of risk or uncertainty in our management measures. This can arise from two causes. The first cause of risk is risk in the system itself. For example, if oceanographic systems affect fish stocks, we cannot with certainty predict the stock size next year, but only predict with given probability. The second cause of risk is uncertainty about these probabilities. This second cause presumably would diminish as more and more observations are made, and our knowledge about the system increases.

In this paper we are concerned with the first cause of risk — inherent uncertainty in the system. The results and techniques can be extended to include the second case also (for extensions in related areas see Rieder (1975) or van Hee (1978)). Risk enters our decisionmaking process in three ways. Firstly, in constructing the probabilities of the occurrence of given events. Secondly, in the decisionmaker's attitude towards risk, e.g. how much of a gamble will be tolerated to obtain a possibly higher total return. And finally, the policy selected can be analyzed for the degree of risk that it entails.

In this paper, an effort is made to make precise the rather vague decisionmaking goal of "minimizing risk" in a dynamic context. We will see initially that intuitively appealing definitions lead to policies that are not reflective of the true complexity of the problem (section III). As an alternative, acceptable trade-offs between low risk in the long run and high expected economic return must be found. To this end, the concept of a "Pareto optimal solution" is introduced, and

Printed in Canada (J5482)
Imprimé au Canada (J5482)

its uses in evaluating trade-off curves are explored (section IV). Several numerical examples of the techniques of section IV are presented in section V.

II. The Model

We restrict attention to single species Markov models without age structure. A Markov model has transition probabilities (e.g. the probability of a given population size next period) that depend only on the present population size and the present harvesting decision. However, it is possible, by increasing the size of the "state" to include sufficient statistics of the past history, to have Markov models that include time lags and other more complex features. Many of the results extend to models with these complexities, as well as models with age structure or multispecies models; however, the more complex models neither help to further illustrate the limitations of a notion like "minimize risk," nor do they make any clearer the alternative methods proposed. Hence only simple models will be explored. At the beginning of each period t, the population size is "observed" to be x_t. During period t, z_t of the population is removed, leaving a population size of $y_t = x_t - z_t$ at the end of period t. The population size at the beginning of period $t+1$ is assumed to be a random function of y_t, that is

(2.1) $x_{t+1} = D_t s[y_t]; \quad D_t \geq 0$ with probability

where $D_1, D_2, \ldots$ are independent, identically distributed random variables distributed as the generic variable D.

A one-period decision rule in period t is any rule that tells us that if x_t is the observed state at the start of period t, then take action y_t (e.g. harvest z_t), and does so for all possible values of x_t. A policy π is a sequence of decision rules; that is, an n-period policy π_n would be defined as

$$\pi_n = \{\delta_1, \delta_2, \ldots, \delta_n\}.$$

Consider the following two problems:

(2.2a) (i) for fixed ω, "minimize" $Pr\{x_t \leq \omega\}$ for all t

(2.2b) (ii) for fixed η, "minimize" $Pr\{z_t \leq \eta\}$ for all t.

This implies finding some infinite horizon policy π^* such that this probability is always smallest.

To make this more precise, we need to define the notion of *stochastic dominance* (Lehmann 1966; Keilson 1974; Barlow and Proschan 1975; O'Brien 1975). For two random variables x and y defined on a common probability space Ω, let F be the distribution function of x and G the distribution function of y. Then x is said to be stochastically dominant over y, written $x \overset{ST.}{\geq} y$ if and only if

$$F(x(\omega)) \leq G(y(\omega)) \quad \text{for all } \omega \in \Omega$$

that is the random variable y has a distribution function that always has more total accumulated probability at lower values. It is easy to see that if $x \overset{ST.}{\geq} y$, then

$$Pr\{x \leq \omega\} \leq Pr\{y \leq \omega\}$$

for all ω.

In comparing two Markov chains $\{x_n, n \in N_0\}$ and $\{y_n, n \in N_0\}$ defined on a common probability space, $\{x_n\}$ is said to be stochastically dominant over $\{y_n\}$ if

$$x_n \overset{ST.}{\geq} y_n \quad \text{for all } n \in N_0.$$

This implies (2.2a) or (2.2b); as we shall see, stochastic dominance is a useful concept.

III. Policies That are Stochastic Dominant

Our first theorem can be motivated by considering the one-period problem: for a given x_1, choose a y_1 such that

$$D_1 s[y_1] \overset{ST.}{\geq} D_1 s[y] \quad \text{for } 0 \leq y \leq x_1$$

where $s[\cdot]$ is assumed to be unimodal.

Now

$$Pr\{D_1 s[y_1] \leq \omega\} = Pr\left\{D_1 \leq \frac{\omega}{s[y_1]}\right\}.$$

Consider the policy $\delta_1{}^*$ defined by

$$\delta_1{}^* = \text{minimum } \{y_{\max}, x_1\}$$

where $y_{\max}$ is the argument where $s[\cdot]$ achieves its mode. Then, for any other feasible decision rule δ_1

$$s\{\delta_1{}^*(x_1)\} \geq s[\delta_1(x_1)] \quad \text{implies}$$

$$\frac{\omega}{s[\delta_1{}^*(x_1)]} \leq \frac{\omega}{s[\delta_1(x)]} \quad \text{implies}$$

$$Pr\left\{D_1 \leq \frac{\omega}{s[\delta_1{}^*(x_1)]}\right\} \leq Pr\left\{D_1 \leq \frac{\omega}{s[\delta_1(x_1)]}\right\}$$

which is the desired result.

Theorem 3.1 formalizes this result; the proof is given in the appendix, as are all the proofs in this paper.

Theorem 3.1 — In (2.1), assume $s[\cdot]$ is unimodal and obtains its mode at $y_{\max}$. Define the policy $\pi^* = \{\delta_1{}^*, \delta_2{}^*, \ldots\}$ where

$$\delta_t{}^* = \text{minimum } \{x_t, y_{\max}\}.$$

Let $\{x_n{}^{\pi^*}\}$ be the Markov chain that arises from following π^*, and $\{x_n{}^{\pi}\}$ be the Markov chain that arises from following any other feasible policy π. Then

$$x_n{}^{\pi^*} \overset{ST.}{\geq} x_n{}^{\pi} \quad \text{for all } n. \qquad \square$$

The assumption that $s[\cdot]$ is unimodal is weak, as most standard production models satisfy this assumption. The specific form $x_{t+1} = D_t s[y_t]$ is also common; see for example the models of salmon runs in Bristol Bay developed by Mathews (1967).

It is worth comparing the policy that arises from theorem 3.1, with the policy that

$$\text{maximizes} \quad E\left\{\sum_{t=1}^{\infty} \alpha^{t-1} p \cdot (x_t - y_t)\right\}$$
$$\text{s.t.} \quad x_{t+1} = D_t s[y_t]; \quad 0 \leq y_t \leq x_t \tag{3.1}$$

and $s[\cdot]$ restricted as before. It has been shown by Mendelssohn and Sobel (unpublished data) that the solution of (3.1) is of the form

$$\bar{\delta}_t = \text{minimum } \{x_t, y^*\}$$

where $0 \leq y^* \leq y_{\max}$.

As an example of the difference between these two policies, consider the following model for salmon in the Wood River postulated by Mathews (1967):

$$x_{t+1} = (e^d)4.077 y_t \exp\{-0.800 y_t\}$$

where d is distributed as $N(0, 0.2098)$. Solved on a 51-point grid, $y^* = 0.700$ (Mendelssohn 1978b) when $\alpha = 0.97$. On this same grid, $y_{\max} = 1.26$ which is close to the true $y_{\max}$ of 1.25 (the grid points are in units of 10^6 fish).

Two measures of interest will be examined. The first is the mean per period harvest under the two policies. The second is the long-run probability of being in any state. (This is the vector ϕ, where

$$\phi(x) = \sum_{j \in X} \phi(j) P_{jx}^{\pi}.)$$

Following $\{\delta_t^*\}$ yields a mean per period harvest of 0.916727×10^6 fish, following $\{\bar{\delta}_t\}$ yields a mean per period harvest of 1.188993×10^6 fish.

Figure 1 shows the distribution function of the stationary probabilities arising from the two policies. The policy $\bar{\delta}$ which maximizes the long-run expected harvest has a greater average harvest compared with δ^* of 272 263 fish, and with a lower variance between harvest amounts. $\bar{\delta}$ also has a zero harvest only 3.28% of the years, while the minimum risk policy δ^* does not harvest 19.67% of the time.

Conversely, the minimum risk policy almost halves the probability of being in the low population sizes of 420 000–840 000 fish, from 7.4% of the time to 4.1% of the time.

Clearly these two policies are the two extremes. A natural question is how to find policies, if any exist, that greatly increase the expected total discounted harvest, at only a small increase in risk.

Theorem 3.1 can be generalized slightly by assuming that $x_{t+1} = s[y_t, D_t]$. In this model the random variable need not act multiplicatively as before.

Theorem 3.2 — If, except for the zero state, every $x \in X$ can be reached with positive probability from every other $x \in X$, and (i) if there exists a δ^* such that

$$s[\delta^*(x_1), D_1] \overset{ST.}{\geq} s[\delta(x_1), D_1]$$

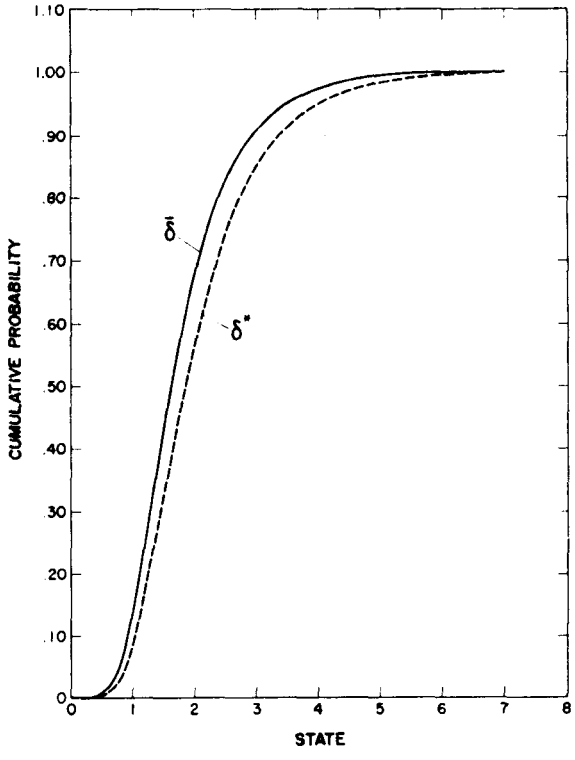

FIG. 1. Cumulative stationary probability distribution ($Pr[x \leq \epsilon]$) when following policies δ^* and $\bar{\delta}$. The minimum risk policy δ^* has a mean harvest of 0.917×10^6, and a variance of 0.894×10^6. The maximal expected total (discounted) harvest policy $\bar{\delta}$ has a mean harvest of 1.189×10^6, and a variance of 0.787×10^6.

for all feasible decision rules δ and all $x_1 \in X$; then

$$\pi^* = (\delta^*, \delta^*, \ldots)$$

is a stochastically dominant policy. (ii) if no such δ^* exists, then there does not exist a policy that is stochastically dominant. □

The advantage of theorem 3.2 is that only a one-period problem need be solved. Either that solution is the long-term solution also, or if no solution exists for this "static" problem, then no solution exists for the larger problem.

I have not been able to find any policies that are stochastically dominant when the criteria is (2.2b); nor have I been able to prove that such a policy doesn't exist. However, I conjecture the following:

Conjecture 3.1 — There is no policy π^*, such that the Markov chain of the amount harvested following π^*, $\{z_n^{\pi^*}, n \in N_0\}$, compared with the Markov chain $\{z_n^{\pi}, n \in N_0\}$ that arises from following policy π, satisfies

$$z_n^{\pi^*} \overset{ST.}{\geq} z_n^{\pi} \quad \text{for all } n$$

for every possible feasible π. □

It is worth reviewing where we are. We have found a stochastically dominant policy under certain assumptions. This was found to produce a solution that greatly reduces the economic return from the harvest (in fact, if $s[\cdot]$ is nondecreasing, then theorem 3.1 states that the risk is minimized *by not harvesting at all*). Also, for certain other models, it has been either proven or conjectured that *there does not exist a policy that minimizes the risk* of an undesirable event. Finally, a numerical example has led us to speculate that a large gain in economic return can be obtained by incurring only a small increase in the risk.

IV. Pareto Optimal Policies

In most decisionmaking situations, the decisionmaker usually is faced with not one, but rather a myriad of conflicting objectives. However, intuitively we often feel that even in this situation, there is a "best" policy in the sense that this "best" policy most fairly balances the different objectives. To examine all the trade-offs possible could be even more confusing than enlightening for the decisionmaker. It would be desirable to reduce the set of policies that need to be evaluated in examining these trade-offs.

One such set of policies are policies that are "Pareto optimal," or more precisely, whose expected returns are Pareto optimal. If there are two objectives, Pareto optimal policies have an intuitive explanation: A policy is Pareto optimal if there is no other feasible policy that does at least as well in any one of the objectives, and strictly better in the other objective.

To describe this more formally, let Π be the set of all possible policies; let $v^\pi = (v_1{}^\pi, v_2{}^\pi, \ldots, v_k{}^\pi)$ be the vector value of the k objectives $v_1, v_2, \ldots, v_k$. Let V be the set of all possible v^π for $\pi \in \Pi$. Then a policy $\pi^* \in \Pi$ is said to be Pareto optimal if there does not exist some other policy $\pi \in \Pi$ such that

$$v^\pi \geq v^{\pi *}.$$

(When comparing two k-vectors $x = (x_1, \ldots, x_k)$ and $(y = (y_1, \ldots, y_k)$, $x = y$ implies $x_i = y_i$, $i = 1, \ldots, k$; $x \geqq y$ implies $x_i \geq y_i$, $i = 1, \ldots, k$; $x \geq y$ implies $x \geqq y$ but $x \neq y$; and $x > y$ implies $x_i > y_i$, $i = 1, \ldots, k$.)

For the problems of this section, we will restrict Π to the class of stationary policies, that is, policies that follow the same one-period decision rule in each period. There is a body of literature that shows that for our models, when maximizing economic return over a very long planning horizon, this is without loss of generality.

In section III, we were concerned with a class of problem that can be viewed as a multiobjective problem: Any policy is evaluated in terms of the expected (discounted) value of the harvest from that policy; it is also evaluated in terms of the degree of long-run risk of an undesirable event that occurs from following this policy. Our goal now is to find the class of Pareto optimal policies between these two objectives, and to find the trade-off curve on this set of policies, rather than trying to strictly optimize.

To accomplish this, assume problem (3.1) has been discretized on a set of states $X_1 = (x_1, x_2, \ldots, x_k)$ and a set of feasible decisions from each $x \in X$, is given by $Y(x) = \{y | 0 \leq y \leq x; y \in X_1\}$. If it is desired to maximize the long-run average harvest, or more formally

$$E\left\{\lim_{T\to\infty}\left(\frac{1}{T}\right)\left[\sum_{t=1}^{T}(x_t - y_t)\right]\right\}$$

subject to the transition and harvest constraints, then this is equivalent to solving the following linear programing problem (DeGhellinck 1960; Manne 1960):

$$\text{maximize} \sum_{x\in X_1}\sum_{y\in Y(x)} u_x{}^y G_x{}^y$$

(4.1) subject to

$$\sum_{y\in Y(x)} u_x{}^y - \sum_{j\in X_1}\sum_{y\in Y(j)} u_j{}^y P_{jx}{}^y = 0 \quad x \in \{X_1/1\}$$

$$\sum_{x\in X_1}\sum_{y\in Y(x)} u_x{}^y = 1$$

$$u_x{}^y \geq 0 \quad \text{for all } x \in X_1,\ y \in Y(x)$$

where $G_x{}^y$ is the one-period return of observing x and harvesting to y, $P_{jx}{}^y$ is the probability of going from state j to state x when action y is taken, and $\{X_1/1\}$ implies the set of states X_1 less some arbitrary base state. That is, if there are k states, there will be k-1 of these constraints.

At an optimal solution, $\bar{u}_x{}^y > 0$ if and only if choosing action y from state x is an optimal decision. Also, the optimal variables $\bar{u}_x{}^y$ have the interpretation of the probability of being in state x and taking decision y. Suppose it is desired to limit the long-run probability that the process is in states contained in some subset X_2 of X_1. This can be included into (4.1) by amending the following constraint:

$$(4.2) \qquad \sum_{x\in X_2}\sum_{y\in Y(x)} u_x{}^y \leq \omega.$$

Similarly, if it is desired to limit the probability that the harvest is less than a given amount, let Z_1 be the set of (x, y) pairs that harvest the fixed amount or less. Then amend the constraint

$$(4.3) \qquad \sum_{(x,y)\in Z_1} u_x{}^y \leq \eta.$$

The advantage of this approach is that linear programing allows for parametric analysis of the right-hand sides of the constraints. Thus, it is simple to start off with ω(resp.η) equal to one, that is, the unconstrained or maximum average return problem, and parameterize it to zero, or the minimum long-run risk problem.

If instead it is desired to maximize the expected

total discounted return, then again, after discretizing the problem, a solution can be found by solving the following linear programing problem (d'Epenoux 1963)

$$\text{maximize} \sum_{x \epsilon X_1} \sum_{y \epsilon Y(x)} u_x^y G_x^y$$

(4.4) subject to

$$\sum_{y \epsilon Y(x)} u_x^y - \sum_{j \epsilon X_1} \sum_{y \epsilon Y(j)} u_j^y \alpha P_{jx}^y = v_x \quad x \epsilon X_1$$

$$u_x^y \geq 0 \quad \text{for all } x \epsilon X_1$$

$$y \epsilon Y(x)$$

where α is the discount factor, and the v_x are chosen so that $v_x / \sum_i v_i$ is the initial probability of being in state x.

Recently Sobel (unpublished data) has shown at an optimal solution to (4.4) that

$$\frac{(1-\alpha)\bar{u}_x^y}{\sum_t v_t} = \text{the normalized discounted fraction of years that state } x \text{ is entered and action } y \text{ is taken.}$$

Note that this is different from the interpretation of the optimal variables in the average return problem, in that here the normalized sum of presence or absence of being in state x and taking action y is discounted, so that near-future behavior is more important than long-run behavior.

Constraints on these probabilities can be amended to (4.4) in a similar manner as probabilistic constraints we amended to (4.3).

Three caveats should be noted carefully. It is always advisable to initially have ω or η identically one and parameterize to zero. This is because at zero the problem may be infeasible, that is, no policy can achieve this constraint. To parameterize an *LP* solution, it is necessary to have an initial feasible solution.

Secondly, a randomized policy may become optimal. A randomized policy is a policy that instead of choosing only one action from each state, chooses one out of several actions from each state by a "lottery" with a given distribution. While this is of some theoretic importance, most policymakers today would not recommend that we manage our resources by tossing dice. However, if only one constraint is amended to (4.2), then Kushner (1971) shows that at most one state will have a randomized decision, and this will involve a "lottery" between at most two actions feasible from that state.

Thirdly, Mendelssohn (1978a) has shown that the size of the grid chosen and the procedure used to discretize the original problem can greatly affect the final values of $\bar{u}_x^y$ for a given problem, as well as the estimates of the stationary distribution $\phi(x)$. Numerical experiments in that paper show that the choice of grid can increase our estimated risk by 3–5 times. Thus it is possible to artificially create a critical trade-off that does not exist in the original, continuous state space problem.

V. Numerical Examples

To see how these techniques work in practice, consider again the model developed by Mathews (1967) for salmon runs in the Wood River:

$$x_{t+1} = (e^d)4.077 y_t \exp\{-0.800 y_t\}$$

where d is distributed as $N(0, 0.6768)$.

This is discretized on the following 16-point grid

$X_1 = (0, 0.467, 0.933, 1.400, 1.867, 2.333, 2.800, 3.267, 4.200, 4.667, 5.133, 5.600, 6.067, 6.533, 7.000)$

in units of 10^6 fish. Numerical experiments suggest that for real stability in terms of caveat three a 51-point grid should be used. However, this 16-point grid produces an *LP* with 136 variables and 17 constraints; the 51-point grid produces an *LP* with 1326 variables and 52 constraints, a very large problem for illustrative purposes. For management purposes, the larger grid size should be used. The return function of interest is (3.1), and (4.4) rather than (4.1) is analyzed. The reason is that while on a 51-point grid, the optimal policy never goes to state zero, it is easy to demonstrate that on the 16-point grid every policy ultimately is absorbed in the zero state. The discounted problem therefore takes into consideration shorter-run behavior, or conversely, the time to absorption.

The optimal policy for (3.1) for this problem is a base stock policy given by

$$y_t = \text{minimum}\{x_t, 0.933\}.$$

The following constraints were added to (4.4), one at a time:

(i) $X_2 = \{x \epsilon X_1 : x \leq 0.467\}$; $v_x = 1$ for all $x \epsilon X$

$$\sum_{x \epsilon X_2} \sum_{y \epsilon Y(x)} \frac{0.03}{16} u_x^y \leq \omega$$

(ii) $X_2 = \{x \epsilon X_1 : x \leq 0.467\}$; $v_0 = 0$; $v_x = \frac{1}{15}$ otherwise

$$\sum_{x \epsilon X_2} \sum_{y \epsilon Y(x)} (0.03)\, u_x^y \leq \omega$$

(iii) $X_2 = \{x \epsilon X_1 : x \leq 0.933\}$; $v_x = 1$ for all $x \epsilon X$

$$\sum_{x \epsilon X_2} \sum_{y \epsilon Y(x)} \frac{0.03}{16} u_x^y \leq \omega$$

(iv) $Z_1 = \{(x, y) : x - y \leq 0.467\}$; $v_0 = 0$, $v_x = \frac{1}{15}$ otherwise

$$\sum_{(x,y) \epsilon Z_1} (0.03)\, u_x^y \leq \eta.$$

The results are summarized in Table 1. For population based probabilistic constraints, an optimal policy moves from a base stock policy of 0.933 to a base stock policy of 1.400, which on the 16-point grid is the optimal minimum risk policy. While any of the

TABLE 1. Bounds on discounted fraction of years $x_t \leq 0.467$, $x_t \leq 0.933$, and $z_t \leq 0.467$.

Total fraction of years, ω	Change in policy: State	Change in policy: Action	Value	$\frac{(1-\alpha)}{\sum_i v(i)}$ (value)
	Bounds on discounted fraction of years $x_t \leq 0.467$			
0.08836	Base stock at	0.933	656.721	1.2314
0.08836	1.867	1.400	656.721	1.2314
0.08735	6.533	1.400	611.634	1.1468
0.08733	3.267	1.400	610.725	1.1451
0.08702	4.667	1.400	596.745	1.1189
0.08696	1.400	1.400	593.696	1.1132
0.08608	4.200	1.400	554.016	1.0401
0.08596	2.333	1.400	548.917	1.0292
0.08518	2.800	1.400	513.729	0.9632
0.08466	7.000	1.400	490.039	0.9188
0.08464	3.733	1.400	489.073	0.9170
0.08444	5.133	1.400	480.126	0.9002
0.08439	6.533	1.400	478.026	0.8963
0.08437	5.600	1.400	477.267	0.8949
0.08434	No change		475.864	
<0.08434	Infeasible			
Initial distribution $v(i) = 1$ for all states i				
	Bounds on discounted fraction of years $x_t \leq 0.933$			
0.02758	Base stock at	0.933	43.7814	
0.02758	1.867	1.400	43.7814	
0.02651	6.067	1.400	40.7756	
0.02349	3.267	1.400	40.7150	
0.02616	4.667	1.400	39.7830	
0.02609	1.400	1.400	39.5797	
0.02515	4.200	1.400	36.9344	
0.02503	2.333	1.400	36.5945	
0.02415	7.000	1.400	34.2486	
0.02417	2.800	1.400	34.1856	
0.02361	3.733	1.400	32.6049	
0.02340	5.133	1.400	32.0084	
0.02335	6.533	1.400	31.8684	
0.02333	5.600	1.400	31.8178	
0.02330	No change		31.7243	
<0.02330	Infeasible			
Initial distribution: $v(0) = 0$, $v(i) = 0.066667$, $i \neq 0$				
	Bounds on discounted fraction of years $x_t \leq 0.933$			
0.15580	Base stock at	0.933	656.721	1.2314
0.15580	1.867	1.400	656.721	1.2314
0.15299	3.267	1.400	611.634	1.1468
0.15212	6.067	1.400	597.667	1.1206
0.15206	4.667	1.400	596.745	1.1189
0.15187	1.400	1.400	593.696	1.1132
0.15041	4.200	1.400	554.016	1.0388
0.14909	2.333	1.400	548.917	1.0292
0.14690	2.800	1.400	513.729	0.9632
0.14543	3.733	1.400	490.039	0.9188
0.14487	5.133	1.400	481.101	0.9021
0.14474	7.000	1.400	479.004	0.8981
0.14468	5.600	1.400	478.026	0.8963
0.14466	6.533	1.400	476.623	0.8937
0.14455	No change		475.864	
<0.14455	Infeasible			
Initial distribution $v(i) \equiv 1$ for all states i				

TABLE 1 (*Concluded*)

Total fraction of years, ω	Change in policy		Value	$\frac{(1-\alpha)}{\sum_i v(i)}$ (value)
	State	Action		
	Bounds on discounted fraction of years $z_t \leq 0.467$			
0.30498	Base stock at	0.933	43.78143	1.3134
0.25000	1.400	Randomized between 0.467, 0.933	42.98377	1.2900
0.16647	1.400 3.267	0.467 Randomized: 1.400, 0.933	41.7717	1.2532
0.16549	3.267 6.067	1.400 Randomized: 1.400, 0.933	41.0265	41.2308
0.16542	6.067 4.667	1.400 Randomized: 1.400, 0.933	40.9748	1.2292
0.16521	4.667 2.333	1.400 Randomized: 1.400, 0.933	40.8160	1.2245
0.16251	4.667 4.200	1.400 Randomized: 1.400, 0.933	38.7655	1.1630
0.16216	4.200 2.800	1.400 Randomized: 1.400, 0.933	38.4950	1.1549
0.16043	2.800 7.000	1.400 Randomized: 1.400, 0.933	37.1798	1.1154
0.16036	7.000 3.733	1.400 Randomized: 1.400, 0.933	37.1263	1.1138
0.15974	3.733 5.133	1.400 Randomized: 1.400, 0.933	36.6561	1.0997
0.15960	5.133 6.533	1.400 Randomized: 1.400, 0.933	36.5452	1.0964
0.15954	6.533 5.600	1.400 Randomized: 1.400, 0.933	36.5010	1.0950
0.15944	5.600	1.400	36.42505	1.0928
<0.15944	Infeasible			
Initial distribution: $v(0) = 0$, $v(i) = 0.066667$				

intermediate policies could be implemented, the shift between the two base stock sizes does not occur in any monotone fashion, that is, first all the larger (or smaller) stock sizes switch their policy, and then the other subset of states are switched over. This feature would make an intermediate policy easier to implement, but does not occur.

The fourth column in Table 1 $(1-\alpha)/\sum_i v_i \cdot$ (value) is in a sense the "discounted mean" value of the harvest. If $\bar{u}_x{}^y$ is the discounted fraction of years that state x is observed and action y is taken, and if $G_x{}^y$ is the one-period return, then the fourth column is equivalent to

$$\sum_{x \epsilon X} \sum_{y \epsilon Y(x)} \bar{u}_x{}^y G_x{}^y$$

which is the one-period returns averaged over the discounted fraction of years it occurs. Hence, it is a "discounted mean" harvest (or value). Columns 1 and 4 of Table 1 provide the information needed by the

decisionmaker to determine the best trade-off between risk and return.

For the harvest-based probability constraints, an optimal policy is randomized for all alternative policies except for the original base stock policy of 0.933 and the final policy. The final policy is of some interest. Summarized it is

States	Harvest to	Harvest amount
0	0	0
0.467	0.467	0
0.933	0.933	0
1.400	0.467	0.933
1.867	0.933	0.933
>1.867	1.400	$x - 1.400$

This policy is very similar to an optimal policy that arises when "smoothing costs" (costs on the amount of year to year fluctuations in the allowed harvest) are added to the problem, with a positive cost only for decreasing the harvest (Mendelssohn 1978b). This suggests that smoothing costs act like a probabilistic constraint, most likely constraining the variance (or discounted variance) of the fraction of periods a state–action combination occurs.

Also, since randomized policies do not seem to occur in population-based constraints but do occur in harvest-based constraints, this leads to the following conjecture: If for every state that has at least one action included in a probabilistic constraint, all actions are also included in the probabilistic constraint, then a nonrandomized policy will be optimal. Otherwise, a randomized policy will be optimal whenever the probabilistic constraint is binding, except perhaps at the two extreme points.

A proof of the conjecture has not been found. But trying to find such rules are important. They lend the analyst insight into how to set up a problem both to obtain the desired trade-offs *and* to obtain policies that can be implemented. And ultimately, the purpose of all model building is the added insight they give to *people* who make the final decisions.

BARLOW, R. E., AND F. PROSCHAN. 1975. Statistical theory of reliability and life testing. Probability models. Holt, Rhinehart, and Winston, Inc., New York, N.Y. 290 p.

DEGHELLINCK, G. 1960. Les problèmes de décisions séquentielles. Cahiers Centre Etudes Recherche Opèr. 2: 161–179.

D'EPENOUX, G. 1963. A probabilistic production and inventory problem. Manage. Sci. 10: 98–108.

KEILSON, J. 1974. Markov chain models — rarity and exponentiality. Center for System Science, Rep CSS 74-01, University of Rochester, New York, N.Y. 182 p.

KUSHNER, H. 1971. Introduction to stochastic control. Holt, Rhinehart, and Winston, Inc., New York, N.Y. 390 p.

LEHMANN, E. L. 1966. Some concepts of dependence. Annals of Mathematic Statistics 37: 1137–1153.

MANNE, A. 1960. Linear programming and sequential decisions. Manage. Sci. 6: 259–276.

MATHEWS, S. B. 1967. The economic consequences of forecasting sockeye salmon runs to Bristol Bay, Alaska: A computer simulation study of the potential benefits to a salmon canning industry from accurate forecasts of the runs. Ph.D. dissertation, Univ. Washington, Seattle, Wash.

MENDELSSOHN, R. 1978a. The effect of grid size and approximation technique on the solution of a Markov decision process. Southwest Fisheries Center Administrative Report 20H, 14 p., National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

1978b. Using Markov decision models and related techniques for purposes other than simple optimization I: Analyzing the consequences of policy alternatives on the management of salmon runs. Southwest Fisheries Center Administrative Report 27H. National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

O'BRIEN, G. L. 1975. The comparison method for stochastic processes. Annals of Probability 3(1): 80–88.

RIEDER, U. 1975. Bayesian dynamic programming. Advances in Applied Probability 7: 330–348.

VAN HEE, K. M. 1978. Bayesian control of Markov chains. Mathematical Centre Tract No. 98, 193 p., Amsterdam.

Appendix

A lemma is needed before the theorems can be proven. Lemma 1 is a specialized version of theorem 3 in O'Brien (1975). The proof can be found in O'Brien (1975).

Lemma 1. Let $\{x_n, n \in N_0\}$ and $\{y_n, n \in N_0\}$ be two real-valued, discrete time Markov processes. If

ST.

(i) $x_0 \leq y_0$

(ii) $Pr\{x_{n+1} \geq t | x_n = x\} \leq Pr\{y_{n+1} \geq t | y_n = y\}$
for all t, n, and $x \leq y$

then there exists two new processes $\{\bar{x}_n\}$ and $\{\bar{y}_n\}$ such that on a common probability space Ω

(a) $\bar{x}_n$ is distributed as x_n for all n

(b) $\bar{y}_n$ is distributed as y_n for all n

(c) $\bar{x}_n(\omega) \leq y_n(\omega)$, for all n, and all $\omega \in \Omega$. □

Proof of theorem 3.1: For any x_1, the policy

$$y_1 = \min\{y_{\max}, x_1\}$$

can be shown to be stochastically dominant as follows. Let δ^* be this decision rule, and δ any other one-period decision rule. Then

$$Pr\{x_2 \leq \omega | x_1, \delta\} = Pr\{Ds[\delta(x_1)] \leq \omega\} = Pr\left\{D \leq \frac{\omega}{s[\delta(x_1)]}\right\}.$$

Since on the set $\{y : 0 \leq y \leq x_1\}$, $s[\delta^*(x_1)] \geq s[\delta(x_1)]$, then

$$\frac{\omega}{s[\delta^*(x_1)]} \leq \frac{\omega}{s[\delta(x_1)]}$$

or that δ^* is stochastically dominant over δ. This implies part (i) of lemma 1.

Conditioned on x_t, the policy

$$y_t = \min\{y_{\max}, x_t\} = \delta^*$$

compared with any other one-period decision rule δ can be shown to satisfy

$$Pr\{x_{t+1} \geq \omega | x_t, \delta^*\} \geq Pr\{x_{t+1} \geq \omega | x_t, \delta\}.$$

by the same argument as above. This is part (ii) of lemma 1, and hence the theorem is proven. □

Proof of theorem 3.2. Part (i): By assumption, π^* is stochastically dominant for x_2 given any x_1. This is condition (i) of lemma 1.

Suppose for some $x_t^1 \leq x_t^2$, x_t^1, $x_t^2 \in X$, there exists a decision rule δ part of policy π, such that

$$Pr\{s[\delta(x_t^1), D_t] \geq \omega\} > Pr\{s[\delta^*(x_t^2), D_t] \geq \omega\} \quad \text{for some } \omega \in \Omega.$$

Since the D_t's are i.i.d. random variables, this would imply

$$Pr\{s[\delta(x_1^1), D_1] \geq \omega\} > Pr\{s[\delta^*(x_1^2), D_1] \geq \omega\} \quad \text{for some } \omega \in \Omega.$$

However, this contradicts the assumption of the theorem that δ^* is a stochastically dominant decision rule for x_2 given x_1. This contradiction implies condition (ii) of lemma 1.

Part (ii): The proof is by contradiction. Suppose a stochastically dominant policy exists, say $\pi^* = \{\delta_1^*, \delta_2^*, \ldots\}$, but no one-period decision rule δ^* exists that is stochastically dominant. By assumption, any state except perhaps the absorbing state is reached with positive probability. Then, conditioned on x_t, there is no policy that is stochastically dominant for x_{t+1}, since by assumption δ^* does not exist, and the transition probabilities are stationary. This contradicts condition (ii) of lemma 1; from the proof of theorem 3 in O'Brien (1975) it is evident that this is sufficient to disprove that the chain resulting from π^* is stochastically dominant.