after N periods). The ... are linear, i.e., for some

(a) Show that the cost-to-go functions $J_k(x_k, y_k)$ are linear, i.e., for some scalars ξ_k, ζ_k,

$$J_k(x_k, y_k) = \xi_k x_k + \zeta_k y_k.$$

(b) Derive an optimal policy $\{\mu_0^*, \dots, \mu_{N-1}^*\}$ under the assumption

$$E\{\gamma_k\} > E\{\delta_k\}$$

and show that this optimal policy can consist of constant functions.

(c) Assume that the proportion of new scientists who become educators at time k is $u_k + \epsilon_k$ (rather than u_k), where ϵ_k are identically distributed independent random variables that are also independent of γ_k, δ_k and take values in the interval $[-\alpha, 1-\beta]$. Derive the form of the cost-to-go functions and the optimal policy.

1.23 (Discounted Cost per Stage)

In the framework of the basic problem, consider the case where the cost is of the form

$$\underset{\substack{w_k \\ k=0,1,\dots,N-1}}{E} \left\{ \alpha^N g_N(x_N) + \sum_{k=0}^{N-1} \alpha^k g_k(x_k, u_k, w_k) \right\},$$

where α is a discount factor with $0 < \alpha < 1$. Show that an alternate form of the DP algorithm is given by

$$V_N(x_N) = g_N(x_N),$$

$$V_k(x_k) = \min_{u_k \in U_k(x_k)} \underset{w_k}{E} \left\{ g_k(x_k, u_k, w_k) + \alpha V_{k+1}\big(f_k(x_k, u_k, w_k)\big) \right\}.$$

1.24 (Exponential Cost Function)

In the framework of the basic problem, consider the case where the cost is of the form

$$\underset{\substack{w_k \\ k=0,1,\dots,N-1}}{E} \left\{ \exp\left(g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k) \right) \right\}.$$

(a) Show that the optimal cost and an optimal policy can be obtained from the DP-like algorithm

$$J_N(x_N) = \exp\big(g_N(x_N)\big),$$

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} \mathop{E}_{w_k} \Big\{ J_{k+1}\big(f_k(x_k, u_k, w_k)\big) \exp\big(g_k(x_k, u_k, w_k)\big) \Big\}.$$

) Define the functions $V_k(x_k) = \ln J_k(x_k)$. Assume also that g_k is a function of x_k and u_k only (and not of w_k). Show that the above algorithm can be rewritten as

$$V_N(x_N) = g_N(x_N),$$

$$V_k(x_k) = \min_{u_k \in U_k(x_k)} \Big\{ g_k(x_k, u_k) + \ln \mathop{E}_{w_k} \Big\{ \exp\Big(V_{k+1}\big(f_k(x_k, u_k, w_k)\big)\Big) \Big\} \Big\}.$$

Note: The exponential is an example of a *risk-sensitive cost function* that can be used to encode a preference for policies with a small variance of the cost $g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$. The associated problems have a lot of interesting properties, which are discussed in several sources, e.g., [DeR79], [Whi90], [FeM94], [JBE94], [BaB95], [Bas00], [Pat01], [Ber16b].

5 (Terminating Process)

ne framework of the basic problem, consider the case where the system evo- n terminates at time i when a given value $\overline{w}_i$ of the disturbance at time i rs, or when a termination decision u_i is made by the controller. If termina- occurs at time i, the resulting cost is

$$T + \sum_{k=0}^{i} g_k(x_k, u_k, w_k),$$

e T is a termination cost. If the process has not terminated up to the final N, the resulting cost is $g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$. Reformulate the lem into the framework of the basic problem. *Hint*: Augment the state space a special termination state.

3.18 (Optimal Termination of Sampling) (www)

This is a classical problem, which when appropriately paraphrased, is known as the job selection, or as the secretary selection, or as the spouse selection problem. A collection of $N \geq 2$ objects is observed randomly and sequentially one at a time. The observer may either select the current object observed, in which case the selection process is terminated, or reject the object and proceed to observe the next. The observer can rank each object relative to those already observed, and the objective is to maximize the probability of selecting the "best" object according to some criterion. It is assumed that no two objects can be judged to be equal. Let r^* be the smallest positive integer r such that

$$\frac{1}{N-1} + \frac{1}{N-2} + \cdots + \frac{1}{r} \leq 1.$$

Show that an optimal policy requires that the first r^* objects be observed. If the r^*th object has rank 1 relative to the others already observed, it should be selected; otherwise, the observation process should be continued until an object of rank 1 relative to those already observed is found. *Hint*: We assume that, if the rth object has rank 1 relative to the previous $(r-1)$ objects, then the probability that it is best is r/N. For $k \geq r^*$, let $J_k(0)$ be the maximal probability of finding the best object assuming k objects have been selected and the kth object is not best relative to the previous $(k-1)$ objects. Show that

$$J_k(0) = \frac{k}{N}\left(\frac{1}{N-1} + \cdots + \frac{1}{k}\right).$$

3.19

A driver is looking for parking on the way to his destination. Each parking place is free with probability p independently of whether other parking places are free or not. The driver cannot observe whether a parking place is free until he reaches it. If he parks k places from his destination, he incurs a cost k. If he reaches the destination without having parked the cost is C.

cost if he is k parking places from his