Optimal stopping in a cumulative damage model

W. Stadje
Universität Osnabrück, Fachbereich Mathematik/Informatik, Albrechtstrasse 28, W-4500 Osnabück, Federal Republic of Germany

Received December 21, 1989 / Accepted September 20, 1990

Summary. We consider the problem of optimal stopping in a cumulative damage model in which a prescribed level may be surpassed only with small probability, but should be approached as precise as possible. Questions of this kind occur, e.g., in the determination of tolerance levels of medical treatments and in coping with metal fatigue. A control limit policy is proved to be optimal.

Zusammenfassung. Für einen kumulativen Schadensprozeß betrachten wir das Problem, den Prozeß so zu stoppen, daß ein gegebener Schwellewert so exakt wie möglich erreicht, aber nur mit kleiner Wahrscheinlichkeit überschritten wird. Fragen dieser Art treten bei der Festlegung zulässiger Belastungshöchstgrenzen für medizinische Behandlungen auf; ein weiteres Beispiel ist die Bestimmung von Ersatzstrategien für verschleiβende Materialien. Wir zeigen, daß eine Kontrollgrenzenstrategie optimal ist.

Key words: Reliability, cumulative damage, tolerance level

Schlüsselworte: Zuverlässigkeit, additive Belastung, Toleranzgrenze

1. Description of the problem

Imagine a person who is, from time to time, exposed to virtually injurious substances. One may think of a cancer patient whose therapy consists of (or includes) the administration of treatments (e.g., radiations) which have the side-effect of damaging not only the target tissues, but also not attacked parts of the organism. A second example is provided by a worker who is charged with the handling of toxic materials. The amount of damage is assumed to vary in a random manner from person to person and from treatment to treatment, but after having taken place it is observable and can be measured in terms of a non-negative real number. Further we suppose that a certain level of damage can be fixed which can be maximally tolerated. On the other hand, in particular in the medical example, it may be advantageous to approach this threshold level as closely as possible in order to maximize the therapeutic effect of the treatment. In achieving this goal, one runs of course the risk of passing over the threshold. A natural idea is to allow such an exceedance to happen only with a small predetermined probability. A formally similar situation occurs, e.g., in the coping with metal fatigue in some device. Here one is interested in maintaining the device as long as the amount of damage (and therefore the risk of failure) remains below a prescribed tolerance level. If damaging occurs in shocks, the problem is to decide at which distance of the current damage level from the threshold one should withdraw the device from service.

One of the standard mathematical tools to address questions of this kind is the cumulative damage model (Barlow and Proschan 1975, chap. 4.3; Saunders 1982). In this model it is supposed that shocks contributing additively to the amount of damage occur at random points of time. Common assumptions are that the sequence of shock amounts and the process \( N(t) \), say, counting their number are stochastically independent and that the shock amounts \( \xi_1, \xi_2, \ldots \) (where \( \xi_i \) is the increment of damage caused by the \( i \)-th shock) are independent and identically distributed non-negative random variables. The tolerance level is given by some constant \( \beta > 0 \) fixed in advance.

In medical treatments the “shocks” are caused by administered dosages so that the process \( N(t) \) can usually be considered as deterministic. In the application to metal fatigue (and to the situation of working in latent contact with toxic materials) the occurrence times of the shocks should be modeled as random. However, this difference is of no importance for the problem at hand: The objective is to stop the cumulative damage process \( S_n = \xi_1 + \ldots + \xi_n \) before it exceeds the level \( \beta \), but such that \( S_n \) is not too far apart from \( \beta \).

We formalize these requirements as follow. Let \( \mathcal{F}_n \) be the \( \sigma \)-field generated by \( \xi_1, \ldots, \xi_n \), and let \( \mathcal{F}_0 \) be the trivial...
32 W. Stadje: Optimal stopping in a cumulative damage model

a-field. Then consider the set $\mathcal{F}$ of all stopping times relative to $(\mathcal{F}_n)_{n\geq 0}$, i.e., all random variables $\tau$ such that $\{\tau=n\} \in \mathcal{F}_n$ for all $n \geq 0$ and $P(\tau < \infty) = 1$. We fix some $\varepsilon \in (0,1)$ and set $\mathcal{F}_\varepsilon = \{\tau \in \mathcal{F} | P(S_\tau > \beta) \leq \varepsilon\}$. Then the aim is to find a stopping time in $\mathcal{F}_\varepsilon$ which maximizes $E(S_\tau | S_\tau \leq \beta)$.

Let $F$ be the common distribution function of the $X_i$ and assume that $F$ is continuous and $F(\beta) > 0$. Our main result is that for $\varepsilon \in (1 - F(\beta), 1)$ a control limit policy is optimal, i.e., there is an $s = s(\varepsilon) \in (0, \beta)$ such that $a = \inf\{n > 1 | S_n \leq s(\varepsilon)\}$ is optimal. The control limit $s(\varepsilon)$ can be determined as solution of the equation

$$1 - F(\beta) + \int_0^s (1 - F(\beta - t)) \, dU(t) = \varepsilon,$$

where $U$ is the renewal measure associated to $F$. This is proved in Sect. 3. As examples, exponential and uniform shock amount variables are treated.

The problem of approaching a goal value $\beta$ as closely as possible can also be formalized in another way. If any exceedance of $\beta$ is to be avoided, one can reward the reached degree of closeness of $S_n$ to $\beta$ from below by a function $f(S_n)$ as long as $S_n < \beta$, and impose a penalty $\alpha$, say, in the case $S_n > \beta$. We shall assume that $f: [0, \beta] \rightarrow [0, \infty)$ is a concave, non-decreasing function and $\alpha \in \mathbb{R}$ is a constant satisfying $\alpha < f(\beta)$. This is carried out in Sect. 2. Incidentally, the case $f(S_n) = S_n$ will be the starting point for the solution of our main problem formulated above.

If it is only the closeness of $S_n$ to $\beta$ which matters, one may measure the distance to the goal by some “loss function” $g(S_n - \beta)$, where $g: [0, \infty) \rightarrow [0, \infty)$ is non-decreasing and satisfies $g(0) = 0$. If $g$ is additionally assumed to be convex, an optimal stopping time for the corresponding minimization problem is found in Sect. 2.

2. Two problems of optimal stopping

We retain the notation already introduced in Sect. 1. Additionally, let $S_0 = 0$ and denote by $1_A$ the indicator function of the event $A$. In this section the following two optimal stopping problems are solved:

(a) Maximize $E(f(S_\tau) 1_{[S_n \leq \beta]} + \alpha 1_{[S_n > \beta]})$ with respect to all stopping times $\tau \in \mathcal{F}$. Here $f: [0, \beta] \rightarrow [0, \infty)$ is assumed to be a concave, non-decreasing function and $\alpha \in \mathbb{R}$ is a constant satisfying $\alpha < f(\beta)$.

(b) Minimize $E(g(S_\tau - \beta))$ with respect to all stopping times $\tau \in \mathcal{F}$. Here $g: [0, \infty) \rightarrow [0, \infty)$ is assumed to be a convex function for which $g(0) = 0$ and $g(S_n)$ is integrable for all $n \geq 1$.

Observe that $g$ attains its minimum at 0 and is monotone non-decreasing. For $s \in \mathbb{R}$ we define $\sigma(s) \in \mathcal{F}$ by $\sigma(s) = \inf\{n \geq 1 | S_n \geq s\}$.

Theorem 1. (a) If

$$f(0) > \int_0^\beta f(x) \, dF(x) + \alpha(1 - F(\beta)),$$

the stopping time $\sigma = 0$ is optimal for problem (a). If (2.1) does not hold, there is an $s \in (0, \beta)$ satisfying

$$f(s) = \int_s^\beta f(s + x) \, dF(x) + \alpha(1 - F(\beta - s)).$$

(b) If

$$g(\beta) \leq \int_0^\infty g(x - \beta) \, dF(x) + \int_0^\beta g(\beta - x) \, dF(x),$$

$\sigma = 0$ is optimal. Otherwise there is an $u \in (0, \beta)$ such that

$$g(u) = \int_u^\infty g(x - u) \, dF(x) + \int_0^u g(u - x) \, dF(x),$$

and $\sigma(\beta - u)$ is optimal.

Proof. Let $X_n = f(S_n) 1_{[S_n \leq \beta]} + \alpha 1_{[S_n > \beta]}$. We shall show that, for any $n \geq 0$,

$$\{X_n \geq E(X_{n+1} | \mathcal{F}_n)\} \subseteq \{X_{n+1} \geq E(X_{n+1} | \mathcal{F}_{n+1})\}. \quad (2.5)$$

To prove (2.5), note that on $\{S_{n+1} > \beta\}$ we have $X_{n+1} = X_{n+2} = \ldots = \alpha$, so that $\{S_{n+1} > \beta\} \cap \{X_{n+1} \geq E(X_{n+1} | \mathcal{F}_{n+1})\}$. Thus it suffices to prove that

$$\{S_{n+1} \leq \beta, X_{n+1} \geq E(X_{n+1} | \mathcal{F}_n)\}$$

$$\subseteq \{S_{n+1} \leq \beta, X_{n+1} \geq E(X_{n+1} | \mathcal{F}_{n+1})\}. \quad (2.6)$$