

THE DECISION PROCESS FOR MAXIMIZING THE PROBABILITY OF OBTAINING THE RANDOM VARIABLE NEAREST TO AN ARBITRARY REAL NUMBER¹

BY E. G. ENNS

University of Calgary

1. Introduction. Let x_i , for $i = 1, 2, \dots, N$ be N independent random variables (rv's) with a common continuous distribution $H(x) = P\{x_i \leq x\}$, $x \in (-\infty, \infty)$. These N rv's will be sampled sequentially beginning with x_1 . Exactly one rv is to be chosen, hence on each presentation the rv sampled is either accepted or forever rejected. Once an rv is accepted, the process ends, as this rv cannot later be rejected. For an arbitrary real number $r \in (-\infty, \infty)$, define the decision function $D(N, r, x)$ as the probability of accepting the first of these N independent rv's if it has the value x . The optimum decision function $D^*(N, r, x)$ will then be defined as that decision function which maximizes the probability of obtaining the rv nearest to r . It will also be assumed that if x_1 is rejected, the problem then becomes that of finding the rv nearest r for the remaining $(N-1)$ rv's, irrespective of the value of x_1 . The special cases when $r = +\infty$ or $-\infty$ then correspond to finding the maximum or minimum rv respectively.

Let the order statistics of the N independent rv's be y_i , $i = 1, 2, \dots, N$, the ordering being defined by $|y_i - r| < |y_{i+1} - r|$, $i = 1, 2, \dots, N-1$. Also define $P_N(i)$ as the probability of obtaining the i th order statistic when following the optimum decision function on a sample of size N . It will be shown that $P_N(i)$ is independent of the distribution $H(x)$ and the real number r .

A special case of this problem, namely when $r = \infty$, has previously been considered by Enns (1969). A somewhat similar problem has also been considered by Karlin (1962). He considers a decision function which maximizes the expected value of the rv chosen. The "optimized" expected value he obtains is, however, a function of the distribution of the rv's.

2. The optimum decision function. For simplicity, let $P_N(1) = P_N$. If the decision $D(N, r, x)$ is made when $x_1 = x$ and the optimum decision function is followed thereafter, then let $P_N(D)$ be the probability of obtaining y_1 in this case. One can then write:

$$(2.1) \quad P_{N+1}(D) = \int_{-\infty}^{\infty} D(N+1, r, x)(f(r, r-x))^N dH(x) \\ + P_N \int_{-\infty}^{\infty} (1 - D(N+1, r, x))(1 - (f(r, r-x))^N) dH(x)$$

for $N = 1, 2, \dots, P_1 = 1$ and

$$(2.2) \quad f(r, u) = 1 + H(r - |u|) - H(r + |u|), \quad r, u \in (-\infty, \infty).$$

Received November 10, 1969.

¹ The author acknowledges the support of the National Research Council of Canada.



The probabilistic interpretation is:

$$(2.3) \quad (f(r, r-x))^N = P\{x_1 = y_1 \mid x_1 = x \text{ and the sample size is } N+1\}.$$

Obviously then:

$$(2.4) \quad \int_{-\infty}^{\infty} (f(r, r-x))^N dH(x) = (N+1)^{-1}.$$

(2.1) can therefore be rewritten as:

$$(2.5) \quad P_{N+1}(D) = (1+P_N) \int_{-\infty}^{\infty} D(N+1, r, x) [(f(r, r-x))^N - (S(N))^N] dH(x) + NP_N/(N+1)$$

where

$$(2.6) \quad S(N) = (P_N/(1+P_N))^{1/N}.$$

Maximizing (2.5) with respect to the decision function D , one obtains the optimum decision function:

$$(2.7) \quad \begin{aligned} D^*(N+1, r, x) &= 1 \quad \text{if } H(r+|r-x|) - H(r-|r-x|) \leq 1 - S(N) \\ &= 0 \quad \text{if } H(r+|r-x|) - H(r-|r-x|) > 1 - S(N) \end{aligned}$$

for $N = 0, 1, 2, \dots$ if one defines $S(0) = 0$.

Using the result (A.4) in the appendix, one can evaluate (2.5) when $D = D^*$. By definition $P_N(D^*) = P_N$, hence one obtains:

$$(2.8) \quad \begin{aligned} P_{N+1} &= (1+P_N)[J(N, N+1, 1) - (S(N))^N J(N, 1, 1)] + NP_N/(N+1) \\ &= (1+NS(N)P_N)/(N+1). \end{aligned}$$

Surprisingly, this result is independent of both the distribution $H(x)$ and the real number r . Using (2.8) one can compute $S(N)$, $N = 1, 2, \dots$, from which the optimum decision function can be obtained. Table 1 presents a tabulation of the optimum decision function for $1 \leq N \leq 1000$.

The asymptotic value of P_N has previously been shown by Enns (1969) to equal $\lim_{N \rightarrow \infty} P_N = 0.4659$.

3. The order statistic accepted. If the sample is of size N and the optimum decision function (2.7) is followed, then the probability of accepting the order statistic y_i can be written as:

$$(3.1) \quad \begin{aligned} P_N(i) &= \int_{-\infty}^{\infty} D^*(N, r, x) P\{x_1 = y_i \mid x_1 = x\} dH(x) \\ &\quad + P_{N-1}(i-1) \int_{-\infty}^{\infty} (1 - D^*(N, r, x)) \sum_{k=1}^{i-1} P\{x_1 = y_k \mid x_1 = x\} dH(x) \\ &\quad + P_{N-1}(i) \int_{-\infty}^{\infty} (1 - D^*(N, r, x)) \sum_{k=i+1}^N P\{x_1 = y_k \mid x_1 = x\} dH(x) \end{aligned}$$

where

$$(3.2) \quad P\{x_1 = y_i \mid x_1 = x\} = \binom{N-1}{i-1} (1-f(r, r-x))^{i-1} (f(r, r-x))^{N-i}.$$

Then using

$$\int_{-\infty}^{\infty} P\{x_1 = y_i \mid x_1 = x\} dH(x) = P\{x_1 = y_i\} = 1/N$$

and the result (A.4) in the appendix, (3.1) can be rewritten as:

$$(3.3) \quad P_N(i) = J(N-1, N, i) - P_{N-1}(i-1) [\sum_{k=1}^{i-1} J(N-1, N, k) - (i-1)/N] \\ - P_{N-1}(i) [\sum_{k=i+1}^N J(N-1, N, k) - (N-i)/N].$$

Let $nJ(N, n, i) = 1 - B(N, n, i)$ where $B(N, n, i)$ has the binomial form:

$$(3.4) \quad B(N, n, i) = \sum_{k=0}^{i-1} \binom{n}{k} (1 - S(N))^k (S(N))^{n-k}.$$

Employing the identity:

$$(3.5) \quad \sum_{k=1}^i B(N, n, k) = iB(N, n, i+1) - n(1 - S(N))B(N, n-1, i),$$

(3.3) can be finally written as:

$$(3.6) \quad NP_N(i) = (1 - B(N-1, N, i)) \\ + P_{N-1}(i-1) [(i-1)B(N-1, N, i) - N(1 - S(N-1)) \\ \cdot B(N-1, N-1, i-1)] \\ + P_{N-1}(i) [NS(N-1) - iB(N-1, N, i) + N(1 - S(N-1)) \\ \cdot B(N-1, N-1, i-1)].$$

It is readily shown that $\sum_{i=1}^N P_N(i) = 1$ and that $P_N(i)$ is monotonically decreasing for $i = 1, 2, \dots, N$.

4. A moment of interest. If x is the value of the rv chosen when following the optimum decision function on a sample of size N , then let:

$$(4.1) \quad y(N, r, x) = H(r + |r-x|) - H(r - |r-x|) \geq 0.$$

$y(N, r, x)$ is therefore the probability that an arbitrary rv selected from the distribution $H(x)$ will be closer to the real number r than the rv x chosen by the sequential decision process.

The expected value of $y(N, r, x)$ is therefore:

$$(4.2) \quad E(y(N, r, x)) = \int_{-\infty}^{\infty} D^*(N, r, x) y(N, r, x) dH(x) \\ + E(y(N-1, r, x)) \int_{-\infty}^{\infty} (1 - D^*(N, r, x)) dH(x) \\ = J(N-1, 2, 2) + E(y(N-1, r, x)) [1 - J(N-1, 1, 1)] \\ = S(N-1) E(y(N-1, r, x)) + \frac{1}{2} (1 - S(N-1))^2$$

for $N \geq 2$ and $E(y(1, r, x)) = \frac{1}{2}$.

This recursive relation is easily solved; however it is more informative in the above form.

APPENDIX

When following the optimum decision function, the following integral often arises.

$$(A.1) \quad J(N, n, i) = \binom{n-1}{i-1} \int_{-\infty}^{\infty} D^*(N+1, r, x) (1-f(r, r-x))^{i-1} (f(r, r-x))^{n-i} dH(x)$$

where

$$\begin{aligned} D^*(N+1, r, x) &= 1 \quad \text{if } f(r, r-x) \geq S(N) \\ &= 0 \quad \text{if } f(r, r-x) < S(N) \end{aligned}$$

and $f(r, x) = f(r, -x) = 1 + H(r - |x|) - H(r + |x|)$ is a monotonically decreasing function of $x \geq 0$. (A.1) can therefore be written as:

$$(A.2) \quad J(N, n, i) = \binom{n-1}{i-1} \int_a^{2r-a} (1-f(r, r-x))^{i-1} (f(r, r-x))^{n-i} dH(x)$$

with $a < r$ and $f(r, r-a) = S(N)$. Thus

$$(A.3) \quad J(N, n, i) = -\binom{n-1}{i-1} \int_0^{r-a} (1-f(r, u))^{i-1} (f(r, u))^{n-i} df(r, u).$$

Successive integration by parts yields:

$$\begin{aligned} (A.4) \quad J(N, n, i) &= J(N, n, i+1) + \binom{n}{i} (1-S(N))^i (S(N))^{n-i} / n \\ &= \sum_{k=i}^n \binom{n}{k} (1-S(N))^k (S(N))^{n-k} / n. \end{aligned}$$

REFERENCES

- [1] ENNS, E. G. (1970). The optimum strategy for choosing the maximum of independent random variables. *Unternehmensforschung* 14 89–96.
- [2] KARLIN, S. (1962). Stochastic models and optimal policy for selling an asset. *Studies in Applied Probability and Management Science*. Stanford Univ. Press.