FUZZINESS IN FOREST SURVEY DESIGN OPTIMIZATION

George Gertner

Xiangchi Cao

Recommended Citation

FUZZINESS IN FOREST SURVEY DESIGN OPTIMIZATION

by

George Gertner
and
Xiangchi Cao

Department of Forestry
University of Illinois
W503 Turner Hall
Urbana, Illinois 61801
USA
Phone (217) 333-9346

Abstract

When using optimization techniques to optimize a sampling with
partial replacement design, it is often assumed that the
following parameters are known exactly: 1) desired level of
sampling error or total sampling cost for the survey; 2) variable
costs; and 3) population variance and correlation coefficients.
In practice, however, these parameters needed for finding the
optimal design are only educated guesses. The parameters can be
considered to be fuzzy. In this paper, brief consideration is
given to the optimization of a sampling with partial replacement
design using nonlinear programming techniques with fuzzy
parameters. The basis of this method is to obtain the optimal
solution by minimizing the objective function, subject to some
restrictions, when the parameters that appear in both the
objective function and restriction functions are fuzzy. The
method is applied to a two-occasion continuous forest inventory.

Introduction

Continuous forest inventory (CFI) has been used for monitoring forests for centuries. CFI is used to monitor the status of a forest, as well as change over time of a forest. The first approach used for CFI was compete enumeration. Based on compete enumeration, every tree in the population is remeasured periodically. Because of the expense and time, complete enumeration is no longer used. Instead, sampling methods are now widely employed. With sampling techniques, some small area plots are drawn and measured from the population. The traditional sampling method used for CFI is completely repeated sampling. With completely repeated sampling, all sample plots are marked and remeasured periodically through time.

Recently, a more efficient sampling technique, repeated sampling with partial replacement of sample plots (SPR), has been presented by Ware and Cunia (1962) for forest monitoring. In SPR, only part of the sample plots from the previous occasion is remeasured in the next sampling occasion, and some new plots are added. The plots measured on both occasions are refereed to as matched plots, and the plots measured only on one of the occasions are refereed to as unmatched plots.

The high efficiency of the SPR for CFI has been shown analytically and in application by many forest researchers (for example, Ware and Cunia 1962; Bickford, Mayer and Ware 1963; Hazard and Promnitz 1974; deVires 1986). Compared to completely repeated sampling, for a desired level of precision, total survey costs are generally lower with SPR (deVires 1986). In application, Bickford et el. (1963) have shown that the completely repeated sampling method needs more than twice the sample plots needed with SPR when the objective is to attain the same degree of sampling precision and the correlation between the two occasions is high. Based on these results, it can be concluded that SPR can be very efficient.

Two basic methods are used in finding the optimal sampling design when SPR has been employed. They are the Lagrange multipliers (Ware and Cunia 1962) and the convex mathematical programming (Hazard and Promnitz 1974). Although in theory the optimal solution can be obtained with these methods, the optimal solution, as Cunia (1965) describes, is based on three assumptions: (1) variable costs (or cost units) are known exactly; (2) level of the sampling error or total sampling cost

is fixed; and (3) population variance and correlation coefficient are known.

Unfortunately, these assumptions are usually not true in actual practice. Variable costs are generally vague since they are based on uncertain economic situations and forest conditions. Population variance and correlation coefficients are unknown population characteristics that are often only educated guesses prior to the survey and are only estimates after the survey is conducted. Likewise, precision requirements for the survey estimates are also frequently only educated guesses. For instance, Cochran (1977) gives a hypothetical example of a researcher who specifies that he desired the precision of his estimates to be within 5% of true mean, but the researcher would not be bitterly opposed to letting the error be 4% or 6%.

The optimal solution is highly dependent on the assumption that these parameters needed for finding the optimal design are known or so called "crisp". If the parameters are not crisp, the solution of the optimal sample design is not truly optimal. A small change of the parameters can cause a great change in the optimal sample numbers. For example, deVires (1986, page 152) gives the following equations for calculating the optimal sample number of plots to minimize sampling costs when estimating mean volume at some desired level of precision:

$$m = \frac{N_1}{r^2} \cdot \sqrt{1-r^2} \cdot \left(\sqrt{\frac{C_u}{C_m}} - \sqrt{1-r^2} \right) \qquad (Eq.1)$$

$$u = \frac{S^2}{V} - \frac{N_1}{r^2} \cdot \left(1 - \sqrt{\frac{C_u}{C_m} \cdot (1-r^2)} \right) \qquad (Eq.2)$$

where m is the matched sample number, u is the unmatched sample number, N_1 is the total sample number at occasion one, ρ^2 is the correlation coefficient between the two occasions, S^2 is the population variance for the second occasion, V is the desired sampling precision, C_m and C_u are respectively the cost units of a matched plot and an unmatched plot in the linear cost function, $C = C_0 + C_m \cdot m + C_u \cdot u$. When the cost of a matched plot equals the cost of an unmatched plot, the optimal matched sample numbers as shown in (Eq.1) can change 50% if the correlation coefficient changes from 0.90 to 0.96. In practice, in order to be sure that the sampling requirements are met, sampling designers usually relax some restrictions when parameters are uncertain adding

"extra sample plots". In such a situation, the "optimal" design based on uncertain parameters is not truly optimal.

Presented in this paper is a brief overview of some aspects of an on going research project where fuzzy methods are being used to account for uncertainties in parameters when optimizing the design of a continuous forest inventory. Since most continuous forest design problems are nonlinear in nature, emphasis in this project has been on the use of nonlinear programming methods with fuzzy parameters.

Nonlinear Programming with Fuzzy Parameters

A continuous forest sampling design problem can generally be written as a nonlinear programming problem,

minimize(or maximize) $Z = f(x, C)$,

subject to $g(x, A) \leq B$,
                          $x \geq 0$,

where $f(x, C)$ is the nonlinear or linear objective function; $g(x, A)$ is a constraint function(s); x is the decision variable; and A, B and C are parameters that are usually assumed to be known constants (crisp). If this is so, then traditional programming methods usually can be used to solve the optimization problem. However, if A, B and C are uncertain due to vagueness, then traditional programming techniques should not be employed.

Techniques in the area of fuzzy mathematics have been developed to solve such programs when parameters (A,B,C) are vague. Fuzzy mathematics provides a framework to account for vagueness in knowledge. Some terms related to vagueness are haziness, cloudiness, unclearness, and indistintiveness. The prime motivation for the development of fuzzy mathematics was the inadequacy of probability theory. Until the development of fuzzy mathematics, the only formal mathematical method for dealing with uncertainty was probability theory.

The theory of fuzzy mathematics is concerned with uncertainties that are not statistical in nature. The relationship of measure theory to probability theory is analogous as fuzzy sets is to possibility theory (Kandel 1986; Kaufmann 1975). Analogous to random variables in the probability domain,

fuzzy numbers (also referred to as fuzzy parameters) are used in the domain of fuzzy sets. There are essential differences between random variables and fuzzy numbers. The uncertainty of a fuzzy number is due to subjectivity and imprecision of human knowledge while the uncertainty of a random variable is because of the occurrence of a random event. Also, a fuzzy number is characterized by a membership degree (or possibility) that is usually subjectively given. The membership degree is the degree of evidence supporting the claim that a specific element of the universe of disclosure belongs to the fuzzy set. A membership degree is analogous to a probability for a random variable. All applications and operations of fuzzy numbers are based on their membership functions that determine membership degrees. Membership functions are analogous to probability functions.

Bellman and Zadeh (1970) were among the first to propose the concept of decision making in a fuzzy environment. The fuzzy optimization has been developed in theory and has been used in practice. Zimmerman (1975, 1977), Hannan (1981), Luhandjula (1983), and Tanaka and Asai (1984) have used the theory of fuzzy sets to formulate and solve fuzzy linear programming problems. Some extensions of the fuzzy linear programming are given by Nakamura (1982). In his work, fuzzy goals and fuzzy constraints are treated as fuzzy sets, and the optimal membership functions of the fuzzy goals and constraints are given by using standard linear programming techniques. In terms of nonlinear programming, Sakawa and Yano (1989) have developed the methodology for converting nonlinear programming with fuzzy parameters to a traditional nonlinear programming problem. The fuzzy solution is then found using multi-objective nonlinear programming techniques.

Determination of Membership Functions

Determination of membership functions is vital in all applications of fuzzy set theory. Although there is no general technique for the solution of this problem, different techniques appear in the literature. These techniques can be classified into two types: use of prior heuristic knowledge, and use of statistical information. In this ongoing study, both techniques are being used.

The most basic method to define membership functions is the use of prior knowledge. This method is also called the heuristic

method and is usually used for approximate numbers. An approximate number is defined as a fuzzy subset of real numbers. For example, approximately 8, is very close to 5, etc. By this method, membership function is subjectively given. Although the assignment of membership function is based on subjective judgment and prior knowledge, it is not arbitrary. The basis of the heuristic method is to choose an appropriate empirical function as the fuzzy distribution function based on experiences and the properties of the fuzzy number. Shown in Table 1 are six types of the most commonly used fuzzy distribution functions:

Although the heuristic method described above is most often used, a more rigorous method for defining the membership function can be used if there is some statistical information. If a random number can be defined with a known probability distribution, it is possible to convert the random number to a fuzzy random number using a method developed by Civanlar and Trussell (1986). A random fuzzy number is defined as a fuzzy number whose possible values can be obtained from random experiments that are inexact. The basis of this method is to build the membership function by using statistical information provided by the probability distribution function.

Application

An example in which the optimal sample number of plots for the second occasion is given when using SPR for a two occasion continuous inventory. In this application, it is desired to minimizing the total sampling cost subject to a specified sampling error for mean volume per unit area. The example is specific to Allerton Park, a conservation area owned by the University of Illinois. Several continuous forest inventories have been made of the park over the last three decades.

The cost function to be minimized is based on Jessen's (1942) travel cost function and the sampling error formula was taken from Cunia (1965). The optimization problem what written as:

$$\textbf{min } f(m, u) = C = C_0 + C_w \cdot \sqrt{(m+u) \cdot A} + C_m \cdot m + C_u \cdot u + C_{mu} \cdot (m+u)$$

subject to

$$g(m, u) = \frac{S^2 \cdot [N_1 \cdot (1-r^2) + m \cdot r^2]}{N_1 \cdot m - N_1 \cdot u \cdot (1-r^2) + m \cdot u \cdot r^2} - K \leq 0 \qquad (Eq.3)$$

$$m - N_1 \leq 0$$

$$m, u \geq 0$$

where $\sqrt{(m+u) \cdot A}$ is the minimum average travel distance among $(m+u)$ random points with area A; C_0 is fixed cost, C_w is the cost of walking a unit distance; C_m is the cost of establishing and maintaining a matched sample plot; C_u is the cost of establishing an unmatched sample plot; and C_{mu} is the cost of measuring a sample plot. The parameters m, u, S^2, ρ^2 and N_1 are the same as defined for Eqs. 1 and 2. K is the desired level of precision for the survey.

The following are rough estimates of the parameters for Allerton Park: C_0 = 100 man hours (hrs); C_w = 1 hrs; C_m = 2 hrs; C_u = 0.5 hrs; C_{mu} = 2 hrs; ρ^2 = 0.9; S^2 = 6; K= 0.1. The area of the forest, A, was set at 1000 hectares and total number of plots for the first occasion, N_1, was set at 50. Assuming all the parameters are crisp, the traditional nonlinear programming solution would be m=8, u=26, (m+u)=34 and the total cost=215.44 hrs.

The membership functions need to be determined for the parameters. For the variable costs, approximate numbers can be used because their guessed values are usually roughly estimated from the information from the first occasion and current economic situation. The specified error goal, K, is also considered to be an approximate number because it is often obtained in a heuristic manner. Heuristic methods were used to define the membership functions. The form of the membership functions used for both variable costs and specific error goal were logistic membership functions. The correlation coefficient and variance of the population were considered to be fuzzy random numbers. This is because both parameter values were based on previous statistical information from past surveys and on intuition. Civanlar and Trussel's method was used to defined the membership functions. A logarithmic normal membership distribution was used for the

correlation coefficient and normal distribution for the variance
of the population. Specifics of how the membership functions
were defined for Allerton Park will be described in a paper that
is in preparation.

Given the membership functions for the parameter, the
approach developed by Sakawa and Yano was used to calculate the
optimal sample numbers for different degrees of fuzziness. When
the membership degree is 1, the parameter is crisp, and as the
membership degree approaches 0, the parameter becomes more fuzzy.
Figure 1 shows the optimal designs in terms of the number of
plots and total hours when individually K, S^2, ρ^2, C_m, C_u and C_w
are fuzzy, while the remaining parameters are crisp. From the
figure, it can be seen that optimal design is not very sensitive
to uncertainties ρ^2, but is extremely sensitive to uncertainties
in S^2. When K becomes more uncertain, the total number of hours
to conduct the survey decreases slightly. With increased
uncertainty in K, the number of matched plots decreases while the
number of unmatched plots increases. When C_m and C_u become more
uncertain, the total number of hours to conduct the survey
increases slightly. With increased uncertainty in C_m, the number
of matched plots is decreased while the number of unmatched plots
is increased. In terms of C_u, uncertainty leads to more matched
plots and few unmatched plots. For C_w, the optimal design is not
very sensitive to uncertainty in this parameter.

Conclusion/Summary

The optimal replacement policy of the SPR presented by Cunia
(1965) and deVires (1986) shows that the SPR is an efficient
sampling technique for CFI. As they point out, the numerical
solution of the optimal replacement policy is based on the
assumptions of fixed variable costs, known correlation
coefficients, known population variance, and exact sampling
requirements. However, these assumptions are not always true.
In fact, they can be considered to be fuzzy. Thus, the fuzziness
of parameters in the sampling design leads to the following
questions: (1) what is the optimal replacement policy? and (2)
how does it change with the fuzziness of the fuzzy parameters?
In the ongoing project, an attempt is being made to answer
these questions. It is the ultimate goal of the project to

consider all type of uncertainties, both statistical and nonstatistical, in the design of a survey.

Literature Cited

Bellman, R. and L. Zadeh 1970. Decision-making in a fuzzy environment. Management Sci. Vol. 17, pp. 141-164.

Bickford, C. A., C. E. Mayer and K. D. Ware. 1963. An efficient sampling design for forest inventory: the Northeast forest resurvey. J. For., Vol. 61. pp. 826-833.

Civanlar, M. R., and H. J. Trussell. 1986. Constructing membership functions using statistical data. Fuzzy Sets and Systems, Vol. 18. pp. 1-13.

Cohran, W. G. 1977. Sampling Techniques. John Wiley & Sons, New York. pp. 428.

Cunia, T. 1965. Continuous forestry inventory, partial replacement of samples and multiple regression. Forest Science, Vol. 11, No. 4. pp. 480-502.

Hannan, E. 1981. Linear programming with multiple fuzzy goals. Fuzzy Sets and Systems, Vol. 6. No. 3, pp. 235-248.

Hazard, J. W., and L. C. Promnitz. 1974. Design of successive forest inventories: optimization by convex mathematical programming. Forest Science, Vol. 20, No. 2. pp. 177-127.

Jessen, R. J. 1942. Statistical investigation of a sample survey for obtaining farm facts. Iowa Agr. Exp. Stat. Res. Bull. 304.

Kandel, A. 1986. Fuzzy Mathematical Techniques with Applications, Addison-Wesley Pubalishing Company, Menlo Park, California.

Kaufmann, A. 1975. Introduction to the Theroy of Fuzzy Subsets. Academic Press, New York. pp. 416.

Luhandjula, M. K. 1983. Linear programming under randomness and fuzziness. Fuzzy Sets and Systems, Vol. 10, No. 2, pp. 123-134.

Nakamura, K. 1982. Quantifications of social utilities for multi-aspect decision making. Policy and Information, Vol. 6, pp. 59-69.

Sakawa, M., and H. Yano. 1989. An interactive fuzzy satisfying method for multiobjective nonlinear programming problems with fuzzy parameter. Fuzzy Sets and Systems, Vol. 30. pp. 221-238.

Tanaka, H. and K. Asai 1984. Fuzzy linear programming probelms
    with fuzzy numbers. Fuzzy Sets and Systems, Vol. 13, No. 1,
    pp.1-10.
de Vries, P. G. 1986. <u>Sampling Theory for Forest Inventory</u>.
    Springer-Verlag, New York.pp. 398.
Ware, K. D., and T. Cunia. 1962. Continuous forest inventory
    with partial replacement of samples. Forest Science,
    Monograph 3.
Zadeh, L. A. 1978. Fuzzy sets as a basis for a theory of
    possibility. Fuzzy Sets and Systems, Vol. 1. pp. 3-28.
Zimmermann, H. J. 1975. Description and optimization of fuzzy
    systems. Int. J. General Syst., Vol. 2. No. 4, pp. 209-215.
Zimmermann, H. J. 1977. Fuzzy programming and linear programming
   with several objective functions. Fuzzy Sets and Systems.
    Vol. 1. pp. 44-55.

Table 1. Commonly used fuzzy membership functions.	
Membership Function	Curve
i) Step function $\mu(x) = \begin{cases} 0 & x < -a \\ 1 & -a \leq x \leq a \\ 0 & a < x \end{cases}$	
ii) Exponential function $\mu(x) = \begin{cases} e^{kx} & x \leq 0 \\ e^{-kx} & 0 \leq x \end{cases}$	
iii) Normal function $\mu(x) = e^{-\left(\frac{x-a}{b}\right)^2}$	
iv) Linear function $\mu(x) = \begin{cases} 0 & x \leq -a \\ 1 + \dfrac{x}{a} & -a \leq x \leq 0 \\ 1 - \dfrac{x}{a} & 0 \leq x \leq a \\ 0 & a \leq x \end{cases}$	
v) Power function $\mu(x) = \begin{cases} 0 & x \leq -(a)^{-\frac{1}{b}} \\ 1 - a(-x)^b & -(a)^{-\frac{1}{b}} \leq x \leq 0 \\ 1 - a(x)^b & 0 \leq x \leq (a)^{-\frac{1}{b}} \\ 0 & (a)^{-\frac{1}{b}} \leq x \end{cases}$	
vi) Logistic function $\mu(x) = \begin{cases} 1 & x \leq C \\ \dfrac{1}{1 + (a(x-c))^b} & x > C \end{cases}$	
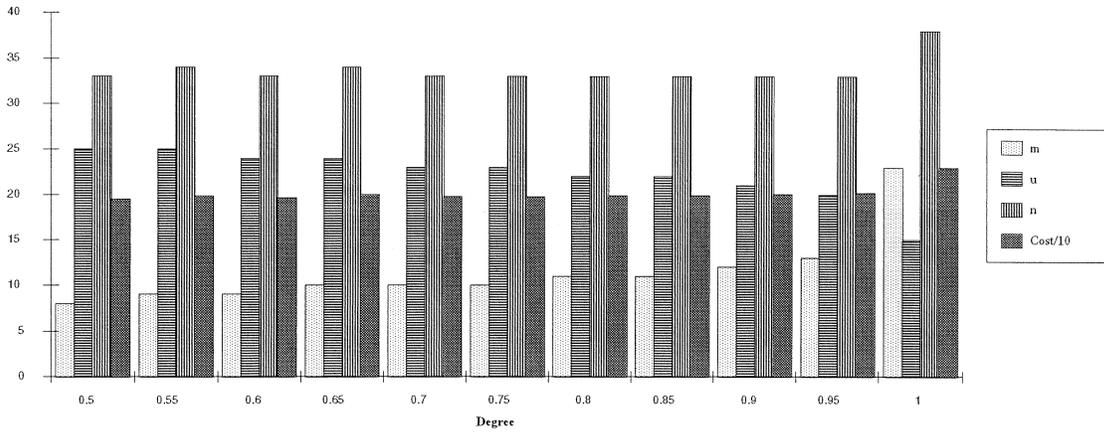
Sampling Design with Fuzzy Error Goal K for
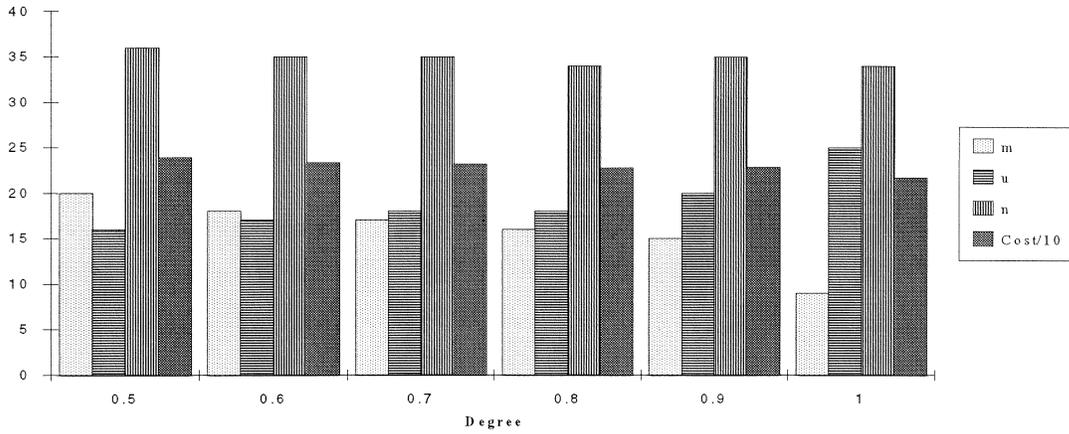Estimating the Mean.
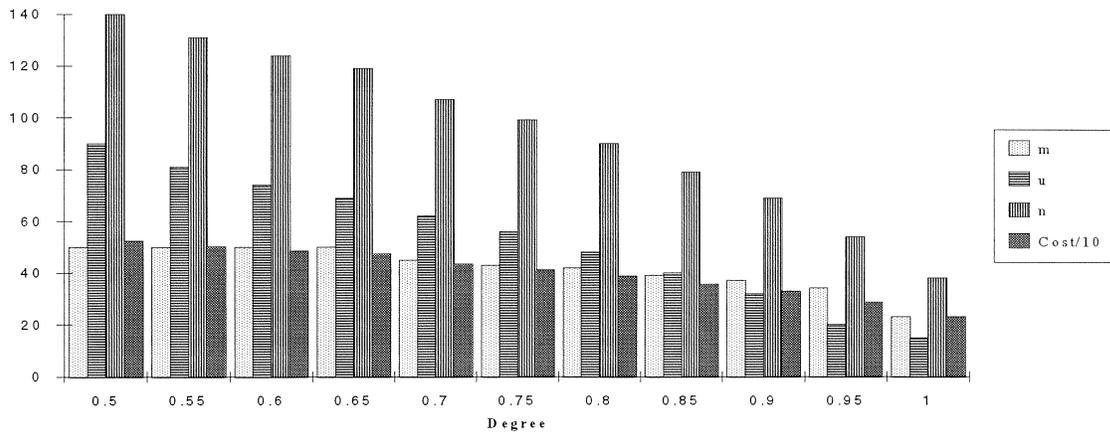
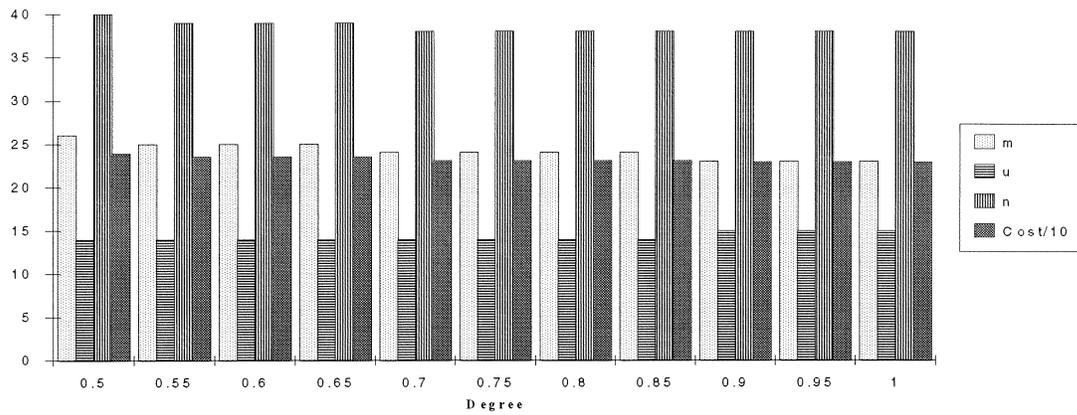Sampling Design with Fuzzy Variance for Estimating
the Mean.
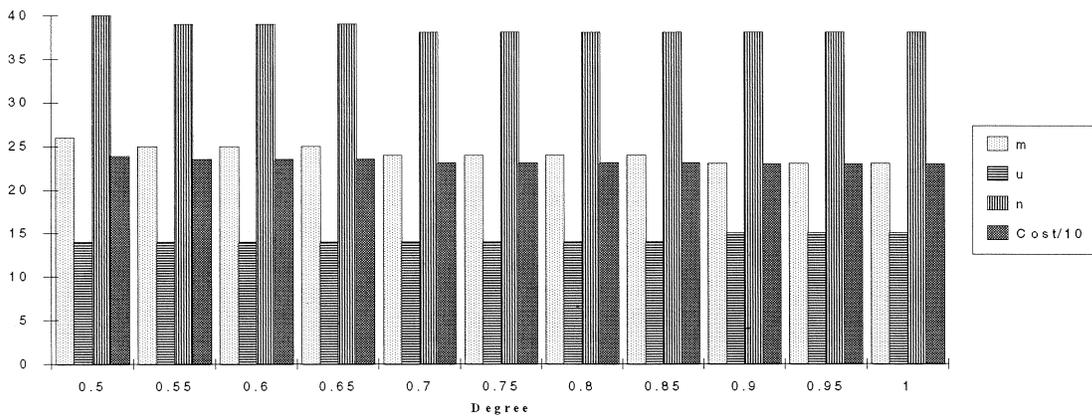
Sampling Design with Fuzzy Correlation Coefficient
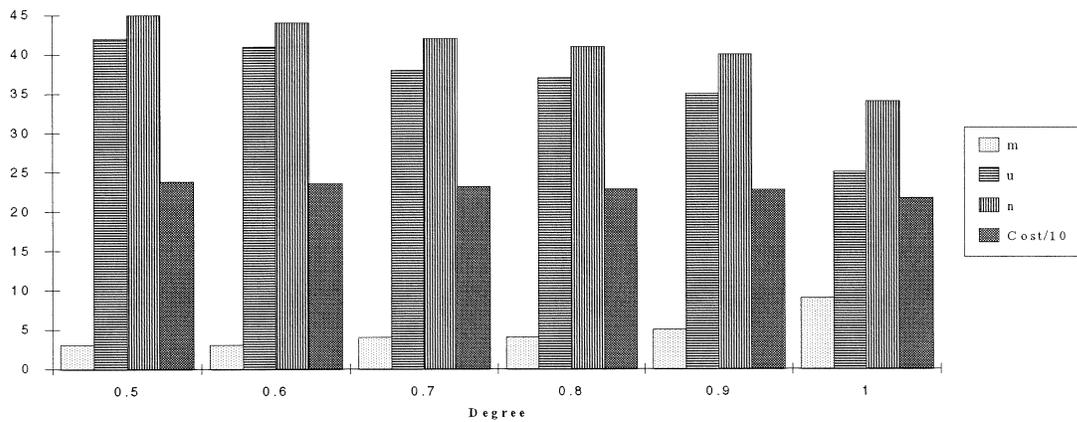for Estimating the Mean.

Figure 1.

Sampling Design with Fuzzy Variable Cost Cm
for Estimating the Mean.

Sampling Design with Fuzzy Variable Cost Cu for
Estimating the Mean.

Sampling Design with Fuzzy Variable Cost Cw
for Estimating the Mean.

Figure 1. Continued.