Two examples of the use of fuzzy set theory in statistics

Glen Meeden
University of Minnesota

http://www.stat.umn.edu/~glen/talks

Fuzzy set theory

Fuzzy set theory was introduced by Zadeh in (1965) as another approach to represent uncertainty.

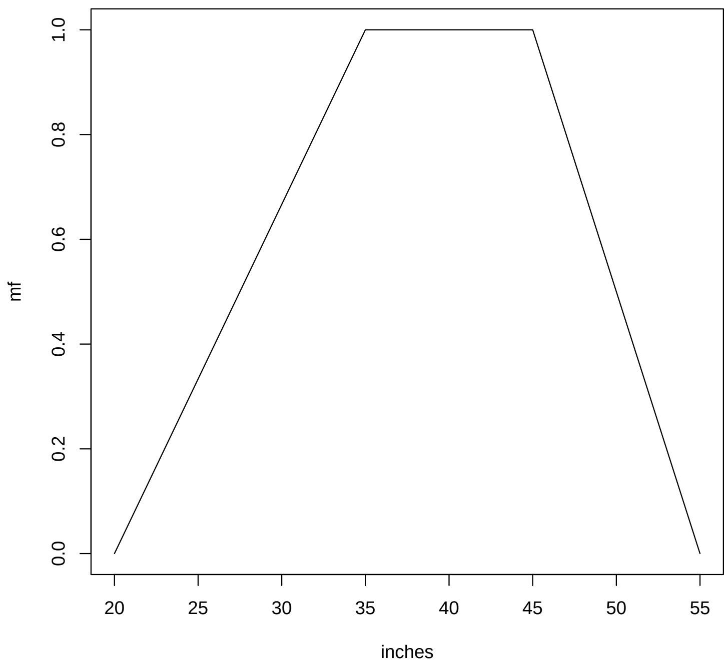
A fuzzy set A is characterized by its membership function.

This is a function whose range is contained in the unit interval.

At a point the value of this function represents the degree of membership of the point in the set A.

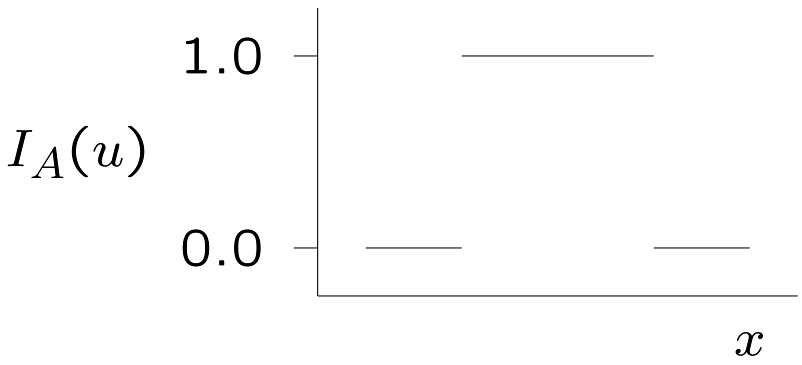
What is the average yearly snowfall in the Twin Cities?

You might answer somewhere between 20 and 50 inches. A better answer could be the following fuzzy membership function.

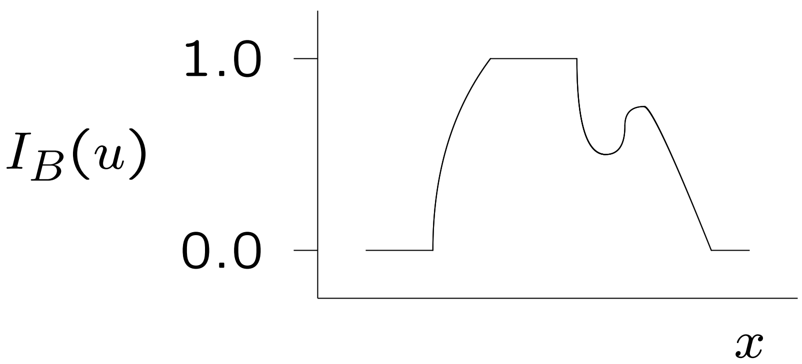


Fuzzy Sets

Indicator function I_A of ordinary set A



Membership function I_B of fuzzy set B



Interpreting a fuzzy membership function

The value $I_B(u)$ is the degree of membership of the point u in the fuzzy set B.

Ordinary sets are special case of fuzzy sets called *crisp sets*.

A non-probabilistic measure of uncertainty.

Think partial credit! Geyer and Meeden (*Statist. Sci.*, 2005)

Buying a used car

Consider the set of cars for sale in your area.

Let A be the fuzzy set of cars that you could consider owning.

For each car you can assign it a value between 0 and 1 which would represent the degree of membership of this particular car in the fuzzy set A.

For a given car this depends on its age, condition, style, price and so forth. Here the fuzzy membership function measures the overall attractiveness of a car to you.

This value cannot be interpreted as a probability!

Can fuzzy set theory be used in statistical inference?

Confidence intervals

Let $X_1, \ldots, X_n$ be iid normal(μ, σ^2).

Let $\bar{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ then

$$\left(\bar{X} - t_{n,0.025} S/\sqrt{n}, \bar{X} + t_{n,0.025} S/\sqrt{n}\right)$$

is a 95% confidence interval for μ.

As a function of μ it has constant coverage probability.

Good CI's come from good tests

In the normal case, given $X_1 = x_1, \ldots, X_n = x_n$, the CI is just the set of μ_o's for which the data "accepts as true" the null hypothesis $H : \mu = \mu_o$.

Optimal properties for CI's comes from the optimal properties of the related tests of hypotheses.

Ordinary Confidence Intervals

OK for continuous data, but a really bad idea for discrete data.

Why? Let X be binomial(n, p)

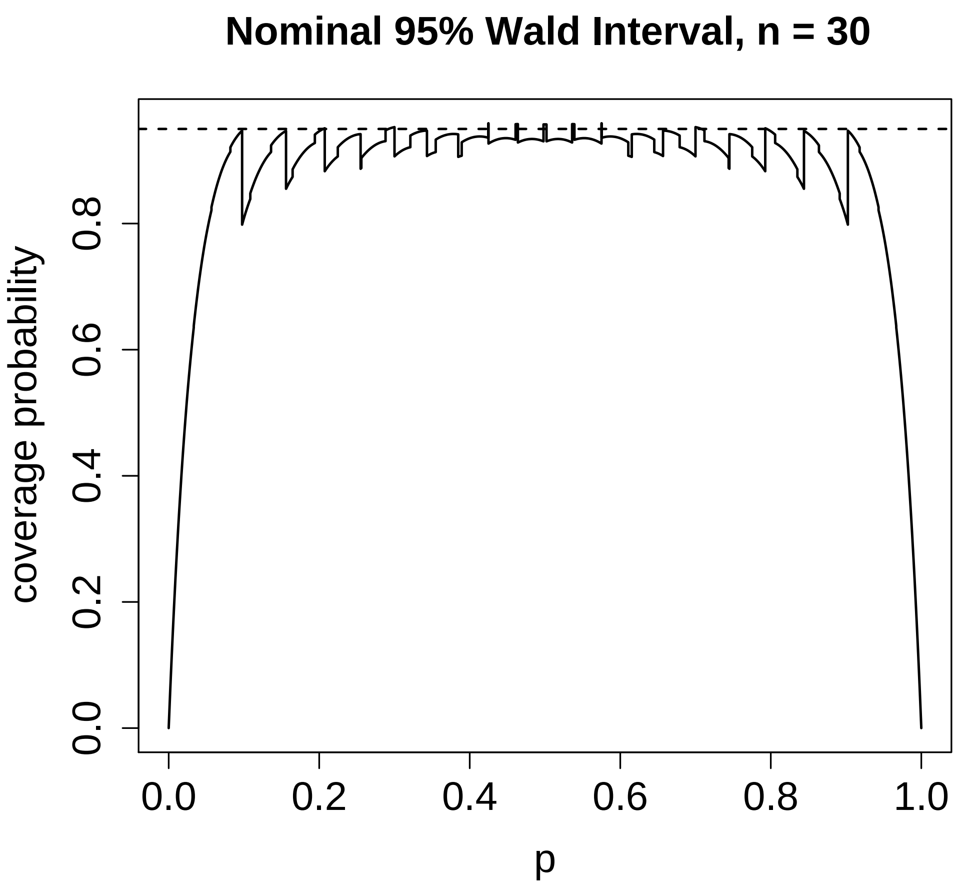
Coverage Probability for a CI $(l(X), u(X))$ is

$$\gamma(p) = \text{pr}_p\{l(X) < p < u(X)\}$$
$$= \sum_{x=0}^{n} I_{(l(x),u(x))}(p) \cdot f_p(x)$$

As p moves across the boundary of a possible confidence interval $(l(x), u(x))$, the coverage probability jumps by $f_p(x)$.

Ideally, γ is a constant function equal to the nominal confidence coefficient. But that's not possible.

Performance of $\widehat{p} \pm 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$



Randomized Tests

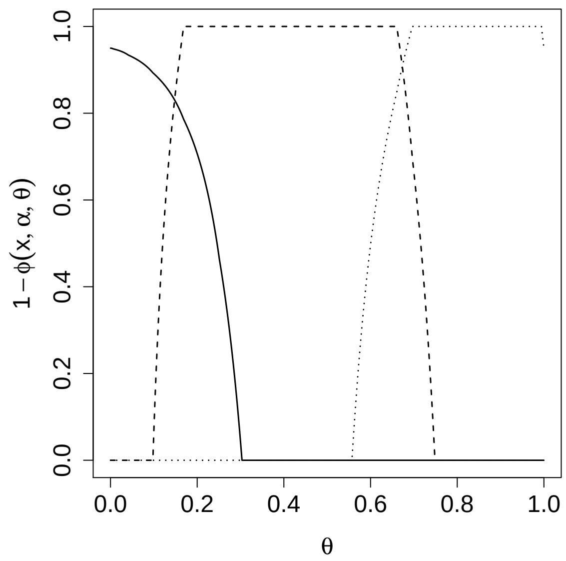
Randomized tests defined by a *critical function* $\phi(x, \alpha, \theta)$. depend on observed data x, the significance level α and the null hypothesis $H_0 : \theta = \theta_0$

Decision is randomized: reject H_0 with probability $\phi(x, \alpha, \theta_0)$.

Since probabilities are between zero and one, so is $\phi(x, \alpha, \theta)$.

Classical uniformly most powerful (UMP) and UMP unbiased (UMPU) tests are randomized when data are discrete and will generate non-crisp CI's.

Binomial Example



Sample size $n = 10$, fuzzy confidence interval associated with UMPU test, confidence level $1 - \alpha = 0.95$.

Data $x = 0$ (solid curve), $x = 4$ (dashed curve) and $x = 9$ (dotted curve)

This is good

If we allow non-crisp fuzzy membership functions to be CI's then our CI's will have exact 95% coverage frequency.

Thinking of CI's as fuzzy membership functions discourages users from believing that a 95% CI contains the unknown parameter with probability equal to 0.95.

The Bayesian approach

X, θ, $f(x \mid \theta)$ and $\pi(\theta)$ the prior distribution.

Inference is based on the posterior

$$\pi(\theta \mid x) = f(x \mid \theta)\pi(\theta) / \int f(x \mid \theta)\pi(\theta)\, d\theta$$

Bayesian credible intervals have the same problem as CI's for discrete data.

Subjective Bayesians will not care since frequentist properties of a procedure are not important to them.

Fuzzy and Bayes

Typically Bayesians are not interested in the fuzzy notion of uncertainty. An exception are those interested in imprecise probability. See Gert de Cooman (*Fuzzy sets and systems*, 2005)

Didier Dubois has also written about the relationship between a fuzzy membership function and families of priors.

How can a Bayesian change their prior or posterior into a fuzzy set membership function which gives a good representation of their uncertainty and can be interpreted like a CI.

A Bayesian decision problem

$\theta \in \Theta$ an interval of real numbers.

$\pi(\theta)$ a continuous prior density

$A \in \mathcal{A}$, the class of measurable membership functions on Θ.

The loss function depends on four known parameters specified by the statistician: $a_1 \geq 0$, $a_2 \geq 0$, $b_1 \geq 0$ and $b_2 \geq 0$ where at least one of the a_i's and one of the b_i's must be > 0.

$$L(A,\theta) = a_1\{1 - I_A(\theta)\} + \frac{a_2}{2}\{1 - I_A(\theta)\}^2 + \int_\Theta \left\{ b_1 I_A(\theta) + \frac{b_2}{2}(I_A(\theta))^2 \right\} d\theta$$

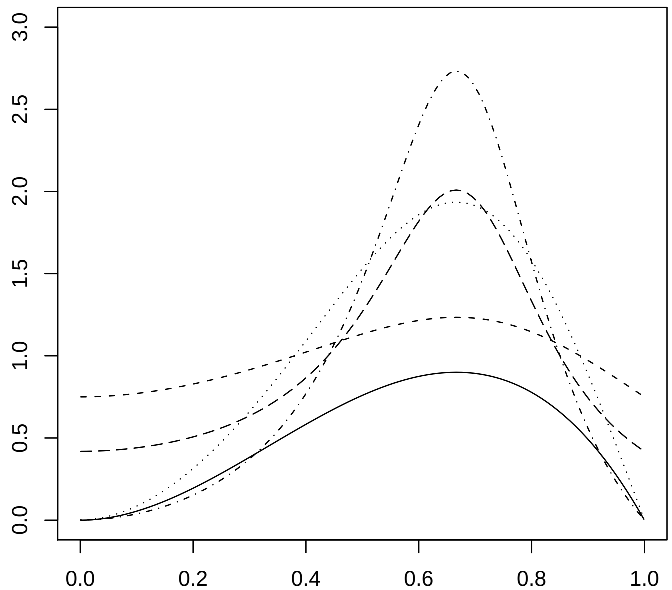
The solution

The fuzzy set membership A which satisfies

$$\int_{\Theta} L(A,\theta)\pi(\theta)\,d\theta = \inf_{A' \in \mathcal{A}} \int_{\Theta} L(A',\theta)\pi(\theta)\,d\theta$$

is given by

$$I_A(\theta) = \begin{cases} 0 & \text{for} \quad 0 \le \pi(\theta) < b_1/(a_1+a_2) \\ \frac{(a_1+a_2)\pi(\theta)-b_1}{a_2\pi(\theta)+b_2} & \text{for} \quad b_1/(a_1+a_2) \le \pi(\theta) \le (b_1+b_2)/a_1 \\ 1 & \text{for} \quad \pi(\theta) > (b_1+b_2)/a_1 \end{cases}$$



A fuzzy membership function (solid line) and for 4 different loss functions the 4 priors whose "solution" is the fmf.

Final comments

We have seen how to convert a prior or posterior density into a fuzzy membership function. Formally this increases a Bayesian's uncertainty but in some cases it could be a good thing. Why?

For a Bayesian the posterior distribution summarizes their information about the parameter given the data. Bayesian credible intervals are a way to summarize the information in the posterior. These intervals have little formal justification in the Bayesian paradigm and seem to be popular because they mimic frequentist CI's. For discrete problems it may more useful to convert the posterior to a fuzzy membership function.

Perhaps frequentists and Bayesians should be more interested in fuzzy membership functions as a way to represent uncertainty about a parameter value.