Estimating Variance Under Interval and Fuzzy Uncertainty: Case of Hierarchical Estimation

Gang Xiang and Vladik Kreinovich

Department of Computer Science, University of Texas at El Paso,
El Paso, TX 79968, USA
gxiang@utep.edu, vladik@utep.edu

1 Estimating Variance Under Interval and Fuzzy Uncertainty: Motivations and Known Results

Computing statistics is important. Traditional data processing in science and engineering starts with computing the basic statistical characteristics such as the population mean and population variance:

$$E = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \quad V = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E)^2.$$

Additional problem. Traditional engineering statistical formulas assume that we know the *exact* values x_i of the corresponding quantity. In practice, these values come either from measurements or from expert estimates. In both case, we get only *approximations* $\widetilde{x}_i$ to the actual (unknown) values x_i.

When we use these approximate values $\widetilde{x}_i \neq x_i$ to compute the desired statistical characteristics such as E and V, we only get approximate valued $\widetilde{E}$ and $\widetilde{V}$ for these characteristics. It is desirable to estimate the accuracy of these approximations.

Case of measurement uncertainty. Measurements are never 100% accurate. As a result, the result $\widetilde{x}$ of the measurement is, in general, different from the (unknown) actual value x of the desired quantity. The difference $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ between the measured and the actual values is usually called a *measurement error*.

The manufacturers of a measuring device usually provide us with an upper bound Δ for the (absolute value of) possible errors, i.e., with a bound Δ for which we guarantee that $|\Delta x| \le \Delta$. The need for such a bound comes from the very nature of a measurement process: if no such bound is provided, this means that the difference between the (unknown) actual value x and the observed value $\widetilde{x}$ can be as large as possible.

Since the (absolute value of the) measurement error $\Delta x = \tilde{x} - x$ is bounded by the given bound Δ, we can therefore guarantee that the actual (unknown) value of the desired quantity belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

Traditional probabilistic approach to describing measurement uncertainty. In many practical situations, we not only know the interval $[-\Delta, \Delta]$ of possible

P. Melin et al. (Eds.): IFSA 2007, LNAI 4529, pp. 3–12, 2007.

values of the measurement error; we also know the probability of different values Δx within this interval [7].

In practice, we can determine the desired probabilities of different values of Δx by comparing the results of measuring with this instrument with the results of measuring the same quantity by a standard (much more accurate) measuring instrument. Since the standard measuring instrument is much more accurate than the one use, the difference between these two measurement results is practically equal to the measurement error; thus, the empirical distribution of this difference is close to the desired probability distribution for measurement error.

Interval approach to measurement uncertainty. As we have mentioned, in many practical situations, we do know the probabilities of different values of the measurement error. There are two cases, however, when this determination is not done:

- First is the case of cutting-edge measurements, e.g., measurements in fundamental science. When a Hubble telescope detects the light from a distant galaxy, there is no "standard" (much more accurate) telescope floating nearby that we can use to calibrate the Hubble: the Hubble telescope is the best we have.
- The second case is the case of measurements on the shop floor. In this case, in principle, every sensor can be thoroughly calibrated, but sensor calibration is so costly – usually costing ten times more than the sensor itself – that manufacturers rarely do it.

In both cases, we have no information about the probabilities of Δx; the only information we have is the upper bound on the measurement error.

In this case, after performing a measurement and getting a measurement result $\widetilde{x}$, the only information that we have about the actual value x of the measured quantity is that it belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$. In this situation, for each i, we know the interval $\mathbf{x}_i$ of possible values of x_i, and we need to find the ranges $\mathbf{E}$ and $\mathbf{V}$ of the characteristics E and V over all possible tuples $x_i \in \mathbf{x}_i$.

Case of expert uncertainty. An expert usually describes his/her uncertainty by using words from the natural language, like "most probably, the value of the quantity is between 6 and 7, but it is somewhat possible to have values between 5 and 8". To formalize this knowledge, it is natural to use *fuzzy set theory*, a formalism specifically designed for describing this type of informal ("fuzzy") knowledge [3,6].

As a result, for every value x_i, we have a fuzzy set $\mu_i(x_i)$ which describes the expert's prior knowledge about x_i: the number $\mu_i(x_i)$ describes the expert's degree of certainty that x_i is a possible value of the i-th quantity.

An alternative user-friendly way to represent a fuzzy set is by using its α-cuts $\{x_i \mid \mu_i(x_i) > \alpha\}$ (or $\{x_i \mid \mu_i(x_i) \ge \alpha\}$). For example, the α-cut corresponding to $\alpha = 0$ is the set of all the values which are possible at all, the α-cut corresponding to $\alpha = 0.1$ is the set of all the values which are possible with degree of certainty at least 0.1, etc. In these terms, a fuzzy set can be viewed as a nested family of intervals $[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]$ corresponding to different level α.