How to Take Expert Uncertainty into Account: Economic Approach Illustrated by Pavement Engineering Applications

Edgar Rodriguez[1,2], Carlos M. Chang [1],
Thach Ngoc Nguyen[3], Olga Kosheleva[1], Vladik Kreinovich[1]
[1]University of Texas at El Paso, El Paso, Texas 79968, USA,
edrodriguezvelasquez@miners.utep.edu, cchangalbitres2@utep.edu,
olgak@utep.edu, vladik@utep.edu
[2]Universidad de Piura in Peru (UDEP), edgar.rodriguez@udep.pe
[3]Banking University of Ho Chi Minh City, Vietnam,
thachnn@buh.edu.vn

1. Need for Expert Estimates

- In many practical situations, we use experts to help make decisions.
- In medicine, computer-based systems are not yet able to always provide a correct diagnosis.
- In other case, the corresponding automatic equipment exists, but it is much cheaper to use human experts.
- For example, in pavement engineering, in principle, we can use automatic systems:
  - to gauge the condition of the road surface,
  - to estimate the size of cracks and other faults.
- However, the corresponding equipment is still reasonably expensive to use.
- The use of human grades is explicitly mentioned in the corresponding normative documents.

2. Expert Estimates Come with Uncertainty

- Expert estimates usually come with uncertainty.
- The experts' estimates have, at best, the accuracy of about 10-15%, up to 20%.
- This observed accuracy is in the perfect accordance with the well-known "seven plus-minus two law".
- According to this law, a person normally divides everything into between 5 and 9 categories.
- Thus, a person has the accuracy between $1/9 \approx 10\%$ and $1/5 \approx 20\%$.

3. Traditional Approach to Dealing with This Uncertainty

- Traditionally, we:
  - come up with the most accurate estimate of the desired quantity, and then
  - if needed, we gauge the economic consequences of the resulting estimate.
- The main limitation of the traditional approach is that
  - while our ultimate objective is economic – how to best maintain the pavement within the budget,
  - we do not take this objective into account when computing the numerical estimate.
- In this talk, we show how to take economic factors into account when producing the estimate.
- The resulting formulas are in line with the usual way how decision makers take risk into account.

4. Analysis of the Problem: Main Idea

- An expert describes his/her opinion by a natural-language word or by a number.
- For each such opinion – be it a word or a number –
  - we can find all the cases when this expert expressed this particular opinion, and
  - in all these cases, find the actual value of the estimated quantity q.
- As a result, for each opinion, we get a probability distribution on the set of all possible values of q.
- This distribution can be described:
  - either in terms of the corresponding probability density function (pdf) $\rho(x)$,
  - or in the terms of the cumulative distribution function (cdf) $F(x) \stackrel{\text{def}}{=} \text{Prob}(q \le x)$.

5. Analysis of the Problem (cont-d)

- In many real-life situations, the expert uncertainty is a joint effect of many different small independent factors.
- According to the Central Limit Theorem, such distribution is close to Gaussian (normal).
- Thus, it often makes sense to assume that the corresponding probability distribution is normal.
- For the normal distribution with mean μ and standard deviation σ, we have $F(x) = F_0\left(\dfrac{x-\mu}{\sigma}\right)$.
- Here, $F_0(x)$ is the cdf of the *standard* normal distribution – with mean 0 and standard deviation 1.
- Based on the probability distribution, we describe the most accurate numerical estimate.

6. Analysis of the Problem: Details

- We want to have an estimate which is as close to the actual values of the quantity q as possible.
- For the same opinion of an expert, we have, in general, different actual values $q_1, \ldots, q_n$.
- These values form a point $(q_1, \ldots, q_n)$ in the corresponding n-dimensional space.
- A natural idea is to select the estimate x_0 for which the point $(x_0, \ldots, x_0)$ is the closest to the point $(q_1, \ldots, q_n)$:

$$d \stackrel{\text{def}}{=} \sqrt{(x_0 - q_1)^2 + \ldots + (x_0 - q_n)^2} \to \min .$$

7. Details (cont-d)

- Minimizing the distance d is equivalent to minimizing

$$d^2 = (x_0 - q_1)^2 + \ldots + (x_0 - q_n)^2.$$

- Differentiating the expression for d^2 with respect to x_0 and equating the derivative to 0, we conclude that

$$2(x_0 - q_1) + \ldots + 2(x_0 - q_n) = 0, \text{ thus}$$

$$x_0 = \mu \stackrel{\text{def}}{=} \frac{q_1 + \ldots + q_n}{n}.$$

- This is equivalent to minimizing the mean square value $\int (x - x_0)^2 \cdot \rho(x)\,dx$, which leads to

$$x_0 = \mu = \int x \cdot \rho(x)\,dx.$$

8. Possible Faults and How Much It Costs to Repair Them

- In pavement engineering, we are interested in estimating the pavement fault index x.
- When the pavement is perfect, this index is 0.
- The presence of any specific fault increases the value of this index.
- Repairing a fault takes money; the larger the index, the more costly it is to repair this road segment.
- Let us denote the cost of repairs for a road segment with index x by $c(x)$.
- We are interested in the case when the road is regularly repaired.
- In this case, the index x cannot grow too much.

9. Possible Faults (cont-d)

- Once there are some faults in the road, these faults are being repaired.
- Thus, the values of the index x remain small.
- So, we can expand the unknown function into Taylor series and keep only the first terms in this expansion.
- For example, we can keep only linear terms: $c(x) \approx c_0 + c_1 \cdot x$.
- When the road segment is perfect, i.e., when $x = 0$, no repairs are needed, so the cost is 0: $c(0) = 0$.
- Thus, $c_0 = 0$, and the cost of repairs linearly depends on the index: $c(x) \approx c_1 \cdot x$.

10. What Is the Cost of Not Repairing a Road?

- If we do not repair a faulty road segment, then:
  - because of the constant traffic load,
  - in the next year, the pavement condition will become worse.
- Let $g(x)$ denote the next-year index corresponding to the situation when this year, the index is x.
- As we have mentioned, it makes sense to consider small values of x.
- So, we can safely expand the function $g(x)$ in Taylor series and keep only linear terms: $g(x) \approx g_0 + g_1 \cdot x$.
- When the pavement is perfect, i.e., when $x = 0$, we usually do not expect it to deteriorate next year.
- So, we should have $g(0) = 0$; thus, $g_0 = 0$, and

$$g(x) \approx g_1 \cdot x.$$

11. Cost of Not Repairing (cont-d)

- Since we did not repair the road segment this year, we have to repair it next year.
- Next year, the index will increase from the original value x to the new value $x' \stackrel{\text{def}}{=} g_1 \cdot x$.
- Thus, the cost of repairs will be $c_1 \cdot x' = c_1 \cdot g_1 \cdot x$.
- This is the cost next year.
- We need to take into account that next year's money is somewhat cheaper than this year's money.
- If the interest rate is r, we can invest a smaller amount $\dfrac{c_1 \cdot q_1 \cdot x}{1 + r}$ now, and get $c_1 \cdot g_1 \cdot x$ next year.
- This formula describes the equivalent this-year cost of not repairing the road segment this year.

Need for Expert Estimates
Expert Estimates . . .
Traditional Approach . . .
Analysis of the . . .
Possible Faults and . . .
What Is the Cost of . . .
Combining The Costs: . . .
Towards Economically . . .
Discussion
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 12 of 19
Go Back
Full Screen
Close
Quit

12. Combining The Costs: What Is the Economic Consequence of Selecting an Estimate

- Once we select an estimate x_0, we perform the repairs corresponding to x_0.
- These repairs costs us the amount $c_1 \cdot x_0$.
- If the actual value x is exactly equal to x_0, this is the ideal situation:
  - the road segment is repaired, and
  - we spend exactly the amount of the money needed to repair it.
- Realistically, the actual x is, in general, somewhat different from x_0.
- As a result, we waste some resources.

13. Combining The Costs (cont-d)

- When $x < x_0$, we spend too money on repairs.
- E.g., we bring on heavy and expensive equipment while a simple device would have been sufficient.
- We could spend just $c_1 \cdot x$ and instead, we spend a larger amount $c_1 \cdot x_0$.
- Thus, in comparison with the ideal situation, we waste the amount $c_1 \cdot x_0 - c_1 \cdot x = c_1 \cdot (x_0 - x)$.
- When $x > x_0$, we still have $x - x_0$ faults to repair.
- The cost of these repairs – when translated into this year's costs – is $\dfrac{c_1 \cdot q_1 \cdot (x - x_0)}{1 + r}$.
- Thus, the expected value $W(x_0)$ of the waste is:

$$\int_0^{x_0} c_1 \cdot (x_0 - x) \cdot \rho(x)\,dx + \int_{x_0}^{\infty} \frac{c_1 \cdot q_1 \cdot (x - x_0)}{1 + r} \cdot \rho(x)\,dx.$$

14. Towards Economically Optimal Estimates

- Traditionally, we select the statistically optimal estimate $x_0 = \mu = \int x \cdot \rho(x)\,dx$.
- Let us instead use the estimate that minimizes the waste $W(x_0)$.
- If we equate the derivative of $W(x_0)$ with respect to x_0 to 0, we get: $F(x_0) = \dfrac{q_1}{1 + r + q_1}$.
- So, as the estimate corresponding to the expert's opinion, we should select:
  - *not* the mean of the actual values corresponding to this opinion,
  - but rather a *quantile* corresponding to the level $\dfrac{q_1}{1 + r + q_1}$: $F(x_0) = \dfrac{q_1}{1 + r + q_1}$.
- Here q_1 is the growth rate of the pavement fault, and r is the interest rate.

15. Discussion

- If the fault growth is negligible, i.e., if $q_1 \approx 1$, then $F(x_0) \approx 1/2$.
- So, x_0 should be the median of the corresponding probability distribution.
- For symmetric distributions like normal, median and mean coincide – they both coincide with the center.
- In this case, we can still use the statistically optimal estimate $x_0 = \mu$.
- However, in most real-life situations, when $q_1 \gg 1 + r$, we have $\dfrac{1+r}{q_1} \ll 1$, thus, $1 + \dfrac{1+r}{q_1} \ll 2$ and
$$F(x_0) = \frac{q_1}{1 + r + q_1} = \frac{1}{1 + \dfrac{1+r}{q_1}} \gg 0.5.$$
- So we should select the values larger than the mean.

16. Discussion (cont-d)

- For normal distribution, the above formula takes the form $x_0 = \mu + k \cdot \sigma$, where k is the value for which

$$F_0(k) = \frac{q_1}{1 + r + q_1}.$$

- This is in line with the usual way to taking risk into account when comparing different alternatives:
  - instead of comparing average gains μ,
  - we should compare the values $\mu - k \cdot \sigma$, where k depends on the person's tolerance to risk.

17. Resulting Practical Recommendation

- For each expert opinion, we:
  - collect all the cases in which the expert expressed this opinion, and
  - find, in all these cases, the actual values of the corresponding quantity.
- Based on these actual values, we compute the mean μ and the standard deviation σ.
- Then, as a numerical description of the expert's opinion, we select $\mu + k \cdot \sigma$, where k is such that:

$$F_0(k) = \frac{q_1}{1 + r + q_1}.$$

- This way, we can decrease the losses caused by the expert's uncertainty.

18. Acknowledgments

- This work was supported in part by the US National Science Foundation grant HRD-1242122 (Cyber-ShARE).