From p-Boxes to p-Ellipsoids: Towards an Optimal Representation of Imprecise Probabilities

Konstantin K. Semenov[1] **and Vladik Kreinovich**[2]

[1]Saint-Petersburg State Polytechnical University
29, Polytechnicheskaya str.
Saint-Petersburg, 195251, Russia, semenov.k.k@gmail.com

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

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 2 of 19
Go Back
Full Screen
Close
Quit

1. Probabilistic Information Is Important

- It is very important to take into account information about the probabilities of different possible values.
- This is especially true in many engineering applications, when we have a long history of similar situations.
- There are several mathematically equivalent ways to represent information about a random variable X:
  - *cdf* $F(x) \stackrel{\text{def}}{=} \text{Prob}(x \le X)$;
  - *pdf* $\rho(x) \stackrel{\text{def}}{=} \lim\limits_{\Delta x \to 0} \dfrac{\text{Prob}(x \le X \le x + \Delta x)}{\Delta x}$;
  - *moments* $M_k \stackrel{\text{def}}{=} E[X^k] = \int x^k \cdot \rho(x)\,dx$; instead of M_2, we can describe the *variance* $V = M_2 - M_1^2$;
  - *characteristic function* $E[\exp(\mathrm{i} \cdot \omega \cdot X)] = \int \exp(\mathrm{i} \cdot \omega \cdot x) \cdot \rho(x)\,dx$;
  - expected values $E[u(X)] = \int u(x) \cdot \rho(x)\,dx$ of the utility functions $u(x)$ that describe user preferences.

2. Need to Take Imprecision into Account

- In practice, we rarely have full knowledge of the probability distribution.
- In terms of cdf, this means that we only know the *bounds* uncertainty means that $[\underline{F}(x), \overline{F}(x)]$ (*p-box*).
- Instead of the exact value $\rho(x)$ of the pdf, for each x, we know an interval $[\underline{\rho}(x), \overline{\rho}(x)]$ of possible values.
- Instead of the exact values of the moments M_k, we know intervals $[\underline{M}_k, \overline{M}_k]$ of possible values, etc.
- When we have the exact knowledge of the probabilities, all representations are mathematically equivalent.
- However, in the presence of uncertainty, these representations are no longer equivalent.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

3. Taking Imprecision into Account (cont-d)

- Let us show that in the presence of uncertainty, different representations are no longer equivalent.
- Example: if we know the bounds $\underline{\rho}$ and $\overline{\rho}$ on $\rho(x)$ on $[x^-, x^+]$, we can deduce bounds on $F(x)$:

$$\underline{F}(x) = (x - x^-) \cdot \underline{\rho} \text{ and } \overline{F}(x) = (x - x^-) \cdot \overline{\rho}.$$

- However, these bounds contain a distribution for which:
  - first the cdf $F(x)$ is equal to $\underline{F}(x)$ and
  - then at some point $x_0 \in [x^-, x^+]$, it jumps to $\overline{F}(x)$.
- For this distribution, the probability density $\rho(x)$ is infinite at $x = x_0$, hence $\rho(x_0) = \infty \notin [\underline{\rho}, \overline{\rho}]$.
- So which of these non-equivalent representations of imprecise probability should we use?

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

4. Which Representation Is the Best?

- One of the main objectives of data processing is to make decisions.
- Standard approach: select the action a with the largest expected utility $E[u_a(x)]$.
- In many cases, the utility function $u_a(x)$ is smooth:

$$u_a(x) \approx c_0 + c_1 \cdot (x - x_0) + c_2 \cdot (x - x_0)^2.$$

- So, to compute $E[u_a(x)]$, it's sufficient to know M_k.
- Sometimes, utility function is discontinuous: e.g., there is a fine is pollution is beyond a threshold x_0.
- When $u = u^-$ for $x < x_0$ and $u = u^+ = 1$ for $x \geq x_0$, then $E[u_a(x)] = u^- + (u^+ - u^-) \cdot F(x_0)$.
- So, depending on the application, different representation are optimal: moments M_k or cdf $F(x)$.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

5. Analysis of the Problem

- *Reminder:* we can use several moments M_1, M_2, ..., or several values $F(x_1)$, $F(x_2)$, ..., of cdf $F(x)$.
- In each case, we use several values $v_1, \ldots, v_n$ to describe a distribution.
- In general, all formulas are linear in $\rho(x)$, so relation between different representations is linear:

$$v_i \to v'_i = a_i + \sum_{j=1}^{n} a_{ij} \cdot v_j.$$

- Imprecision is usually represented by bounds $\underline{v}_i$ and $\overline{v}_i$; so, possible values of $v = (v_1, \ldots, v_n)$ form a *box*

$$[\underline{v}_1, \overline{v}_1] \times \ldots \times [\underline{v}_n, \overline{v}_n].$$

- Alas, in general, a linear transformation transforms a box into a parallelepiped – and not into a box.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

6. What Is Needed

- *Reminder:* what was a box in one representation becomes a different objects in another one.
- So, different box representations of imprecise probability are *not* equivalent.
- We therefore need a family F of sets which remains of the same type after a linear transformation T:

$$\text{if } V \in F \text{ then } T(V) \stackrel{\text{def}}{=} \{T(v) : v \in V\} \in F.$$

- In many situations (e.g., in automatic control), when we need to make decision very fast.
- In general, the more parameters we need to process, the longer our computations.
- It is therefore desirable to select a family F with the smallest possible number of parameters.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

7. Main Result and Its Corollary

Main Result:

- *Let F be a linear-invariant r-parametric family of connected bounded closed domains from $\mathbb{R}^n$.*
- *Then $r \geq \dfrac{n(n+3)}{2}$; and if $r = \dfrac{n(n+3)}{2}$, then:*
  - *either F is the the family of all ellipsoids E,*
  - *or, for some $\lambda \in (0,1)$, F is the family of all sets*

$$E - \lambda \cdot E.$$

Discussion:

- If we restrict ourselves to *convex* sets (or only to simply connected sets), we get ellipsoids only.
- So, to describe imprecision, we should use *p-ellipsoids*: ellipsoid-shaped regions in the space of all cdf f-s $F(x)$.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

8. Towards Auxiliary Result: What Does "Optimal" Mean?

Let $\mathcal{A}$ be a class of families of sets, and let G be a group of transformations defined on $\mathcal{A}$.

- By an *optimality criterion*, we mean a *pre-ordering* (i.e., a transitive reflexive relation) $\preceq$ on the class $\mathcal{A}$.
- An optimality criterion is *G-invariant* if for all $g \in G$, and for all $B, B' \in \mathcal{A}$, $B \preceq B'$ implies $g(B) \preceq g(B')$.
- An optimality criterion is *final* if there exists exactly one $B_{\rm opt} \in \mathcal{A}$ for which $B \preceq B_{\rm opt}$ for all $B \neq B_{\rm opt}$.

Explanation:

- If there are *no* optimal $B_{\rm opt}$, the criterion is useless.
- If there are *several* optimal $B_{\rm opt} \neq B'_{\rm opt}$, we can use this non-uniqueness to optimize something else.
- So, if $B_{\rm opt} \neq B'_{\rm opt}$, the original criterion is *not* final.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

9. Auxiliary Result

Result:

- *Let $\mathcal{A}$ be the family of all r-parametric families of connected bounded closed domains from $\mathbb{R}^n$.*
- *Let $\preceq$ be a linear-invariant final opt. criterion on $\mathcal{A}$.*
- *Then $r \geq \dfrac{n(n+3)}{2}$; and if $r = \dfrac{n(n+3)}{2}$, then:*
  - *either the optimal family $F_{\rm opt}$ is the the family of all ellipsoids E,*
  - *or, for some $\lambda \in (0,1)$, F_{opt} is the family of all sets*

$$E - \lambda \cdot E.$$

Discussion:

- If we restrict ourselves to *convex* sets (or only to simply connected sets), we get ellipsoids only.
- So, to describe imprecision, we should use p-ellipsoids.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

10. Ellipsoids Are Better Than Boxes: Examples

- Several families of sets have been proposed to describe *uncertainty*: ellipsoids, boxes, polytopes, etc.
- Experiments show that in many practical situations with uncertainty, *ellipsoids* lead to the best results.
- Example: *linear programming* – finding min or max of a linear function under linear inequalities.
- The traditional *simplex method* sometimes requires unfeasibly many ($\approx 2^n$) computational steps.
- Ellipsoids lead to polynomial-time algorithms for linear programming (Khachiyan, Karmarkar).
- Ellipsoids are also empirically better in many *pattern recognition* problems.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

11. Ellipsoids Lead to Faster Computations

- In many practical situations, we need to estimate the value of a statistical characteristic $S(v_1, \ldots, v_n)$.
- In the case of imprecision, we only know the range V of possible values of v.
- Different distributions $v \in V$ lead, in general, to different values of $S(v)$.
- It is therefore desirable to compute the *range* $S(V) \stackrel{\text{def}}{=} \{S(v) : v \in V\}$ of possible values of $S(v)$.
- Often, we have a reasonably good knowledge about the probability distribution.
- So, we expand the dependence $S(v)$ around an estimate $\widetilde{v}$ and keep only quadratic terms in $\Delta v \stackrel{\text{def}}{=} v - \widetilde{v}$:

$$S(v) = s_0 + \sum_{i=1}^{n} s_i \cdot v_i + \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij} \cdot v_i \cdot v_j.$$

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

12. Ellipsoids Lead to Faster Computations (cont-d)

- *Reminder:* we need to estimate the range of the following function over the set V describing imprecision:

$$S(v) = s_0 + \sum_{i=1}^{n} s_i \cdot v_i + \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij} \cdot v_i \cdot v_j.$$

- Computing the range of a quadratic function over a *box* is, in general, *NP-hard*.
- This means, crudely speaking, that no feasible algorithm can always solve this range-comp. problem.
- In contrast, Lagrange multipliers lead to *feasible* computation of quadratic $S(v)$ over an *ellipsoid*.
- So, ellipsoids do lead to faster computations.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

13. Ellipsoids Are in Good Agreement with Additional Probabilistic Information

- Often, we have a probability distribution on the set V of possible probability distributions.
- There are usually many different reasons for the imprecision with which we know v.
- Due to the Central Limit Theorem, we conclude that the distribution is close to Gaussian.
- Strictly speaking, a Gaussian distribution has positive density $\rho_V(v) > 0$ for all possible vectors $v \in \mathbb{R}^n$.
- In practice, we dismiss v for which the probability is too small $\rho_V(v) < \rho_0$, and keep $V = \{v : \rho_V(v) \geq \rho_0\}$.
- For a Gaussian distribution, the inequality $\rho_V(v) \geq \rho_0$ describes an *ellipsoid*.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

14. How to Extract and p-Ellipsoid from Data

- A p-box can be extracted by using Kolmogorov-Smirnov criterion $\max|F(x) - F_n(x)| \le \Delta$ w/given conf. level,
$$F_n(x) \stackrel{\text{def}}{=} \frac{\#\{i : x_i \le x\}}{n}.$$
- Thus, $F(x) \in [F_n(x) - \Delta, F_n(x) + \Delta]$.
- For p-ellipsoids, we can similarly use Cramer-von Mises ω^2 criterion for goodness of fit:
$$\int (F(x) - F_n(x))^2 \, dF(x) \le \Delta.$$
- In geometric terms, this quadratic inequality describes an ellipsoid, so we get the desired p-ellipsoid.
- In practice, 95% confidence intervals are normally used.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

15. If we Reconstruct a p-Ellipsoid from Data Instead of a p-Box, We Get Better Estimates

We compared the interval of possible values for the mean computed based on p-box and p-ellipsoid.

- We produced a set x_i, $i = 1, 2, \ldots, n$ of random variables of the same bounded distribution (for example, which is uniform on $[a, b]$) for some n.
- We reconstructed a p-box and a p-ellipsoid from this data and estimate confidence intervals I_{KS} and I_{CvM} for mean by solving optimization problems.
  Note: To get correct results we must take into accaunt that all $x_i \in [a, b]$.
- We compared the width w_{KS} of I_{KS} and the width w_{CvM} of I_{CvM} with the width w_t of the classical Student confidence interval.
- We repeated experiment $N = 10^6$ times for different n.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

16. If we Reconstruct a p-Ellipsoid from Data Instead of a p-Box, We Get Better Estimates (cont-d)

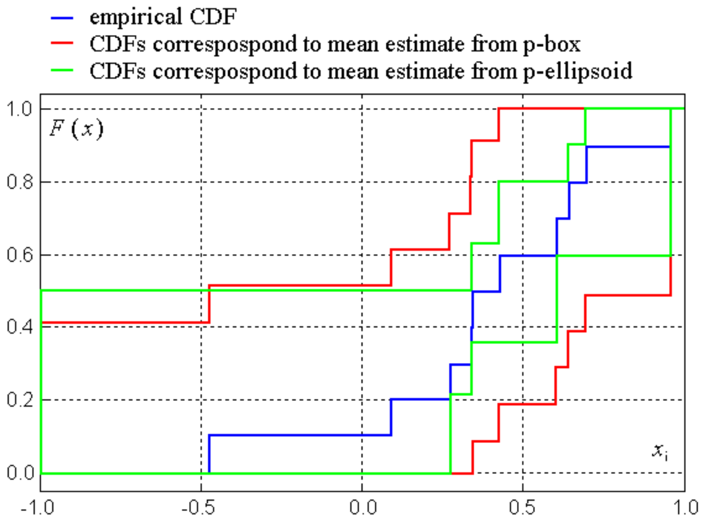
- Results are in the table below.

n	10	25	50	100	200
$\frac{w_{KS}}{w_t}$	1.46	2.05	2.17	2.25	1.88
$\frac{w_{CvM}}{w_t}$	1.21	1.60	1.65	1.67	1.49

- *Conclusion:* estimates w_{CvM} based on p-ellipsoids are narrower.

17. If we Reconstruct a p-Ellipsoid from Data Instead of a p-Box, We Get Better Estimates

- Here is the example of CDFs, which are corresponding to the limit values of I_{KS} and I_{CvM}.



Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .

18. Acknowledgment

This work was supported in part:

- by the National Science Foundation grants HRD-0734825, HRD-1242122, and DUE-0926721,
- by Grant 1 T36 GM078000-01 from the National Institutes of Health, and
- by a grant on F-transforms from the Office of Naval Research.

The authors would like to thank:

- all participants of the 15th GAMM – IMACS Int'l Symposium on Scientific Computing, Computer Arithmetic, and Verified Numerical Computation SCAN'2012 (Novosibirsk, Russia, September 23–29, 2012), especially Sergey Shary, for inspiring discussions,
- the anonymous referees for valuable suggestions.

Probabilistic . . .
Need to Take . . .
Which Representation . . .
Analysis of the Problem
What Is Needed
Main Result and Its . . .
Auxiliary Result
Ellipsoids Are Better . . .
If we Reconstruct a p- . . .