BUILDING AN EVALUATION "COMMON GROUND" FOR RESEARCH ON AHP REFINEMENTS

© Robin Rivest, M.Sc. Business Analytics,

Montréal, Qc, Canada robin.rivest@videotron.ca

DECEMBER 3 - DECEMBER 6, 2020 / WEB CONFERENCE
International Symposium on the Analytic Hierarchy Process

Lack of commonality in evaluations

A significant portion of publications that explore the inner workings or theoretical aspects of the AHP* bring attention to potential drawbacks and propose improvements.

When one considers this overall body of work, the disconcerting finding is that **none (or too few of them) share a common basis** to illustrate or highlight the performance of their suggestion.

There is a **compelling need for** an evaluation framework that would enable **informative comparisons** between the various approaches.

* Although introductions to the Analytic Hierarchy Process (AHP) can be found in numerous publications, Brunelli (2014) provides one that is more contemporary as it covers findings of various research areas from the first three decades since the inception of the method.

The purpose of this study is twofold.

First, it aims to raise the **awareness of the absence of a "common ground"** to evaluate and to compare results of experiments.

This is prominently noticeable in the treatment of the following three topics : reduction of the number of comparisons, methods to derive the priority vector and proposals of alternative numerical scales

> « *There are various research papers on methods for dealing with incomplete preferences, but very few investigated the relation between the number of missing comparisons and the stability of the obtained priority vector … It is safe to say that there is need and space for further investigation* ».
>
> – Brunelli, Introduction to AHP (2014, p.40)

The second intention is to **expose some foundational orientations** to circumscribe and resolve this deficiency, which are further detailed in my Master's thesis (Rivest, 2019).

Deficiencies of evaluation frameworks

This section summarizes the literature review found in Rivest (2019) and **makes clear some of these shortcomings**.

The **disparity** is exhibited for **numerous aspects**, such as: the dimensions of matrices (i.e. number of alternatives compared), the extent of cases tested and the use of simulation data. **And, in all cases, no objective limit of precision is stated**.

Disparity of test cases

Evaluations of various suggestions to reduce the number of necessary comparisons are conducted in **dissimilar frameworks**.

This table contains a few examples from this field of research to highlight the **underdevelopment of testing frameworks**.

Source*	Basic idea	Extent of cases tested	Proximity measure	Matrix size	Limit of precision
Shen *et al*, 1992	Evaluate alternatives in multiple subsets, then combine by prorating results (single pivot)	Only one example	None	7	None
Ishizaka, 2012b	Evaluate alternatives in multiple subsets, then combine by prorating results (multiple pivots)	A few examples	None	12	None
Fedrizzi & Giove, 2013	Proceed by iteration until some condition of sufficiency (left to be determined) is attained.	Only one partial example	Incomplete	5, 9	None
Rezaei, 2015-2016	Make only (2n-3) comparisons using best and worst alternatives	46 participants	Rank variation; Total deviation; Euclidian distance	4, 5, 6	None
Pamučar *et al*, 2018	Use only (n-1) comparisons	A few examples	None	4, 5, 8	None
Abastante *et al*, 2019	Proceed with direct estimate of weights which are then calibrated (prorated) with priorities obtained for only a subset of alternatives	98 participants	MSE between non-normalized vectors	10	None

Insufficient coverage

Many are only **considering a few isolated test cases**, while those which make use of more cases **fail to establish why the implied coverage is suitable**. Furthermore, in some instances, **suggestions remain incomplete** or undetermined.

Those evaluations that attempt to assess the validity of resulting priority vectors use a variety of **irreconcilable proximity measures** whose merit remains unstated. Therefore, it is **nearly impossible to juxtapose results** obtained from one study to the next.

Finally, one of the most important deficiency is the **lack of a method to establish an objective threshold** beyond which further gains would be redondant. **This prevents proper gauging** of how far or how near the suggested approaches provide a practical satisfactory approximation.

Similar observations can be made with research on methods to derive the priority vector as well as the exploration of numerical scale alternatives.

Conclusion of review

The review concludes with these remarks:

- Use of simulation data is **neither widespread, nor systematized**.
- Simulated priority vectors are almost always **generated using the *uniform* distribution**, which will we see, provides a test coverage that is **not properly aligned with the value domain**.
- **No characteristics** have been proposed to assert the extent of test cases needed to ensure proper coverage
- There is **no convention for the size** of matrices to be verified
- There is **no convention on the proximity measures** to be used
- The undertaken **limit of precision is either undefined or set arbitrarily**

These findings motivate the identification of solution paths.

Challenges

The main challenges to resolve are:

- To determine **a few key characteristics** that will guide the definition of the **value domain of priority vectors** and in return provide bounds to ensure **proper test coverage**
- To select an **adequate measure of proximity**
- To conceive a way for establishing an **objective limit of precision**

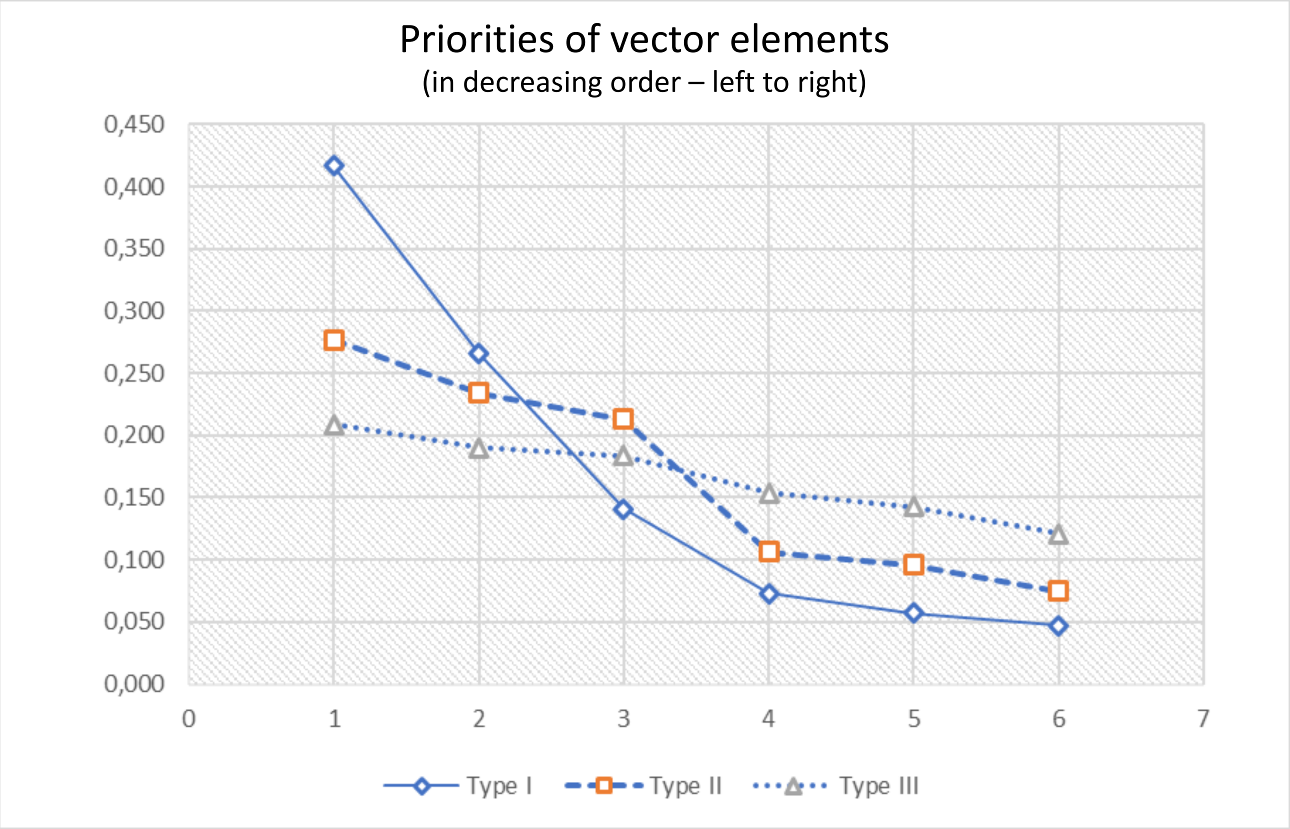
Key characteristics of priority vectors

A self-evident characteristic of a priority vector is its size or **dimension** (n), given by the number of alternatives.

A more subtle aspect is hinted at in Ishizaka *et al* (2012, p.4769) with the following sentence: « *A high difference of performances can also be highly discriminating even with a low weight of the criterion.* »

Let's consider 3 **generic types of priority distributions**:

- Type I : Highly discriminating
- Type II : Somewhat discriminating
- Type III : Close to "no difference"

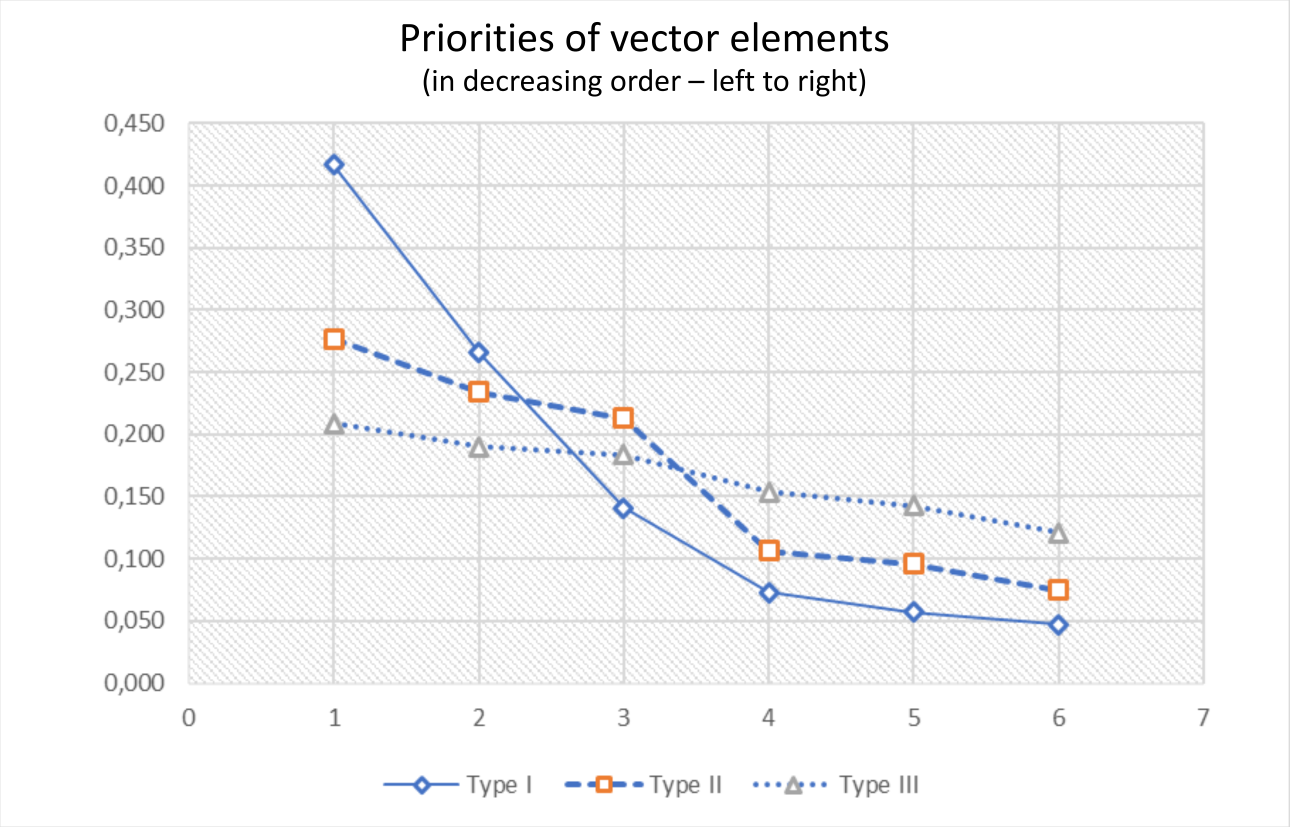


Key characteristics of priority vectors

One can observe two features that make them appear different:

- The relative height of the "best" or "highest" value versus the others
- The overall degree of "compression" or "spacing" between the elements

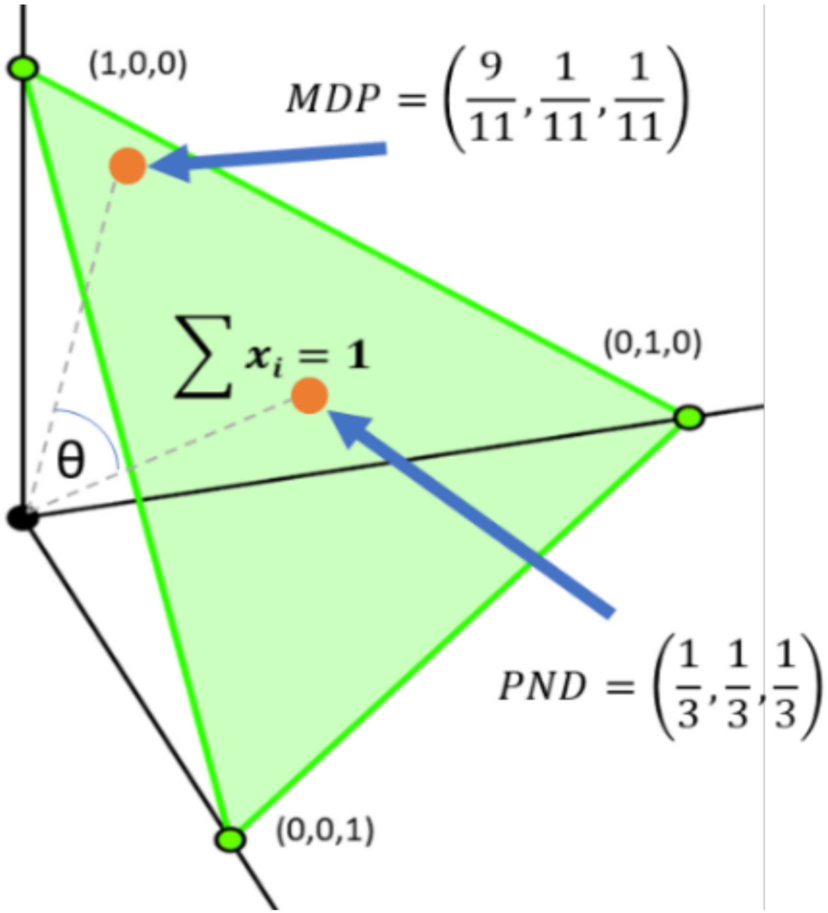
These features can appropriately be captured by a characteristic which is referred to, in what follows, as the **potential for discrimination** and is defined as $\rho = \frac{highest\ priority}{lowest\ priority}$



Measuring the power of discrimination

Based on this analysis of the 3 types, one can consider that the priority vector with the less discriminating power is the one in which all weights are equal. Let us refer to it as the Point-of-No-Difference (PND). We can then entertain the thought that the one which gives the maximum weight to one option and the least weight to all others has the most discriminating power (MDP).

This concept lends itself to the following simple geometric interpretation which can be illustrated with a graph showing both points on a 2-simplex : the **angle (θ) between any vector and the PND can be used as a measure of its power of discrimination.**



Three key characteristics

These characteristics will enable the definition of the value domain for priority vectors. The last step needed to define a practical value domain for priority vectors is to set sensible bounds for each of the three characteristics :

- dimension (n) : comparing from 3 to 18 alternatives should basically encompass all but a few singular cases

- potential for discrimination (ρ) : *de facto,* this one inherits the maximum value of the numerical scale used has an upper bound and 1 as lower bound

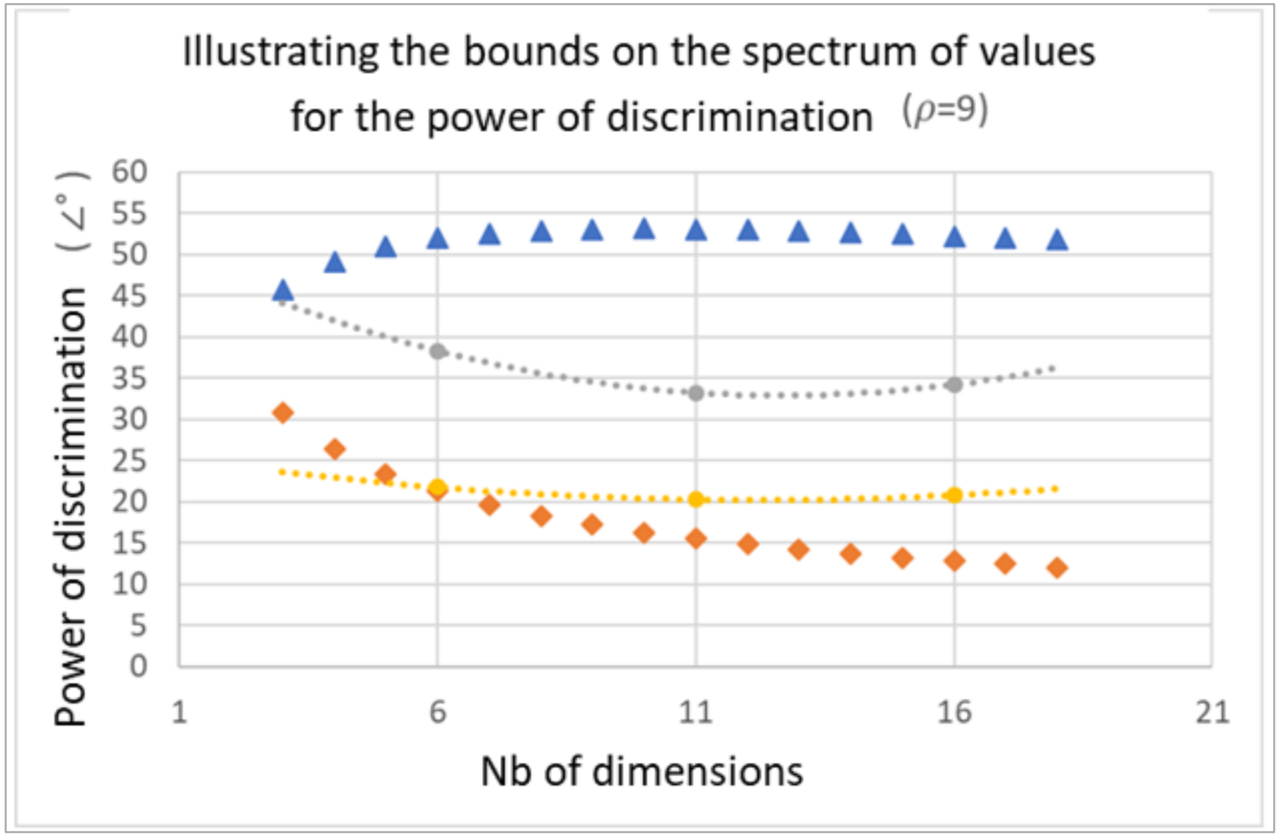
- power of discrimination (°) : for this one, the bounds will be constrained by the potential (ρ)

Bounds for the power of discrimination

The figure (left) shows the upper* (blue) and lower* (orange) bounds for the power of discrimination (°) with $\rho = 9$ for dimensions $n \in [3..18]$

{* see appendix A for bound determination}

Note: The lower bound here is the first step of discretization (see limit of precision)



Incomplete coverage

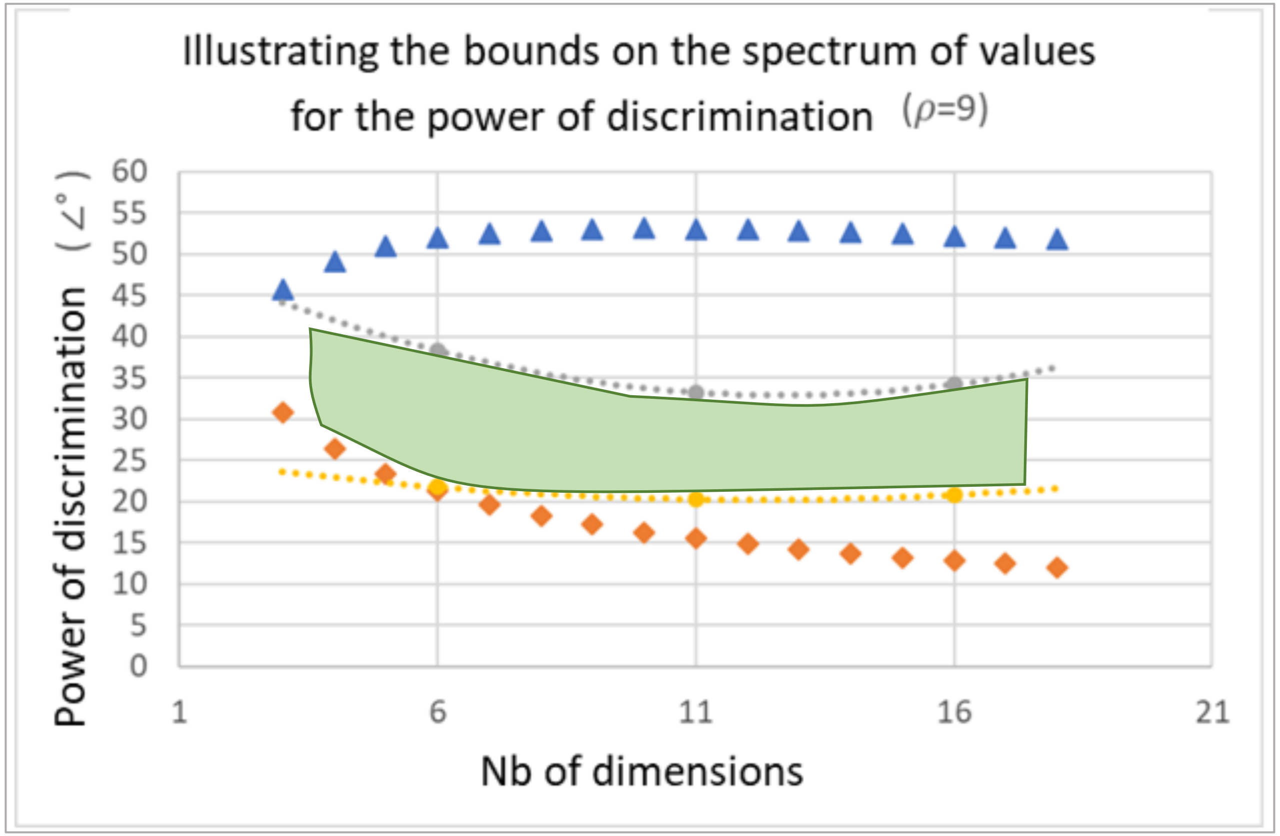
Here, we can observe that generating priorities by drawing from a uniform distribution results in incomplete coverage.

In this figure, the green area between the grey and yellow dotted lines, shows the limited range of generated vectors.

One can observe that, in most dimensions, more than half of the target area is left uncovered by test cases.

Furthermore, a large proportion (40 to 80%)* of vectors that are generated this way actually fall outside the range of values for ρ and therefore are simply not representative of realistic empirical cases.

{* see appendix B for details on proportions}

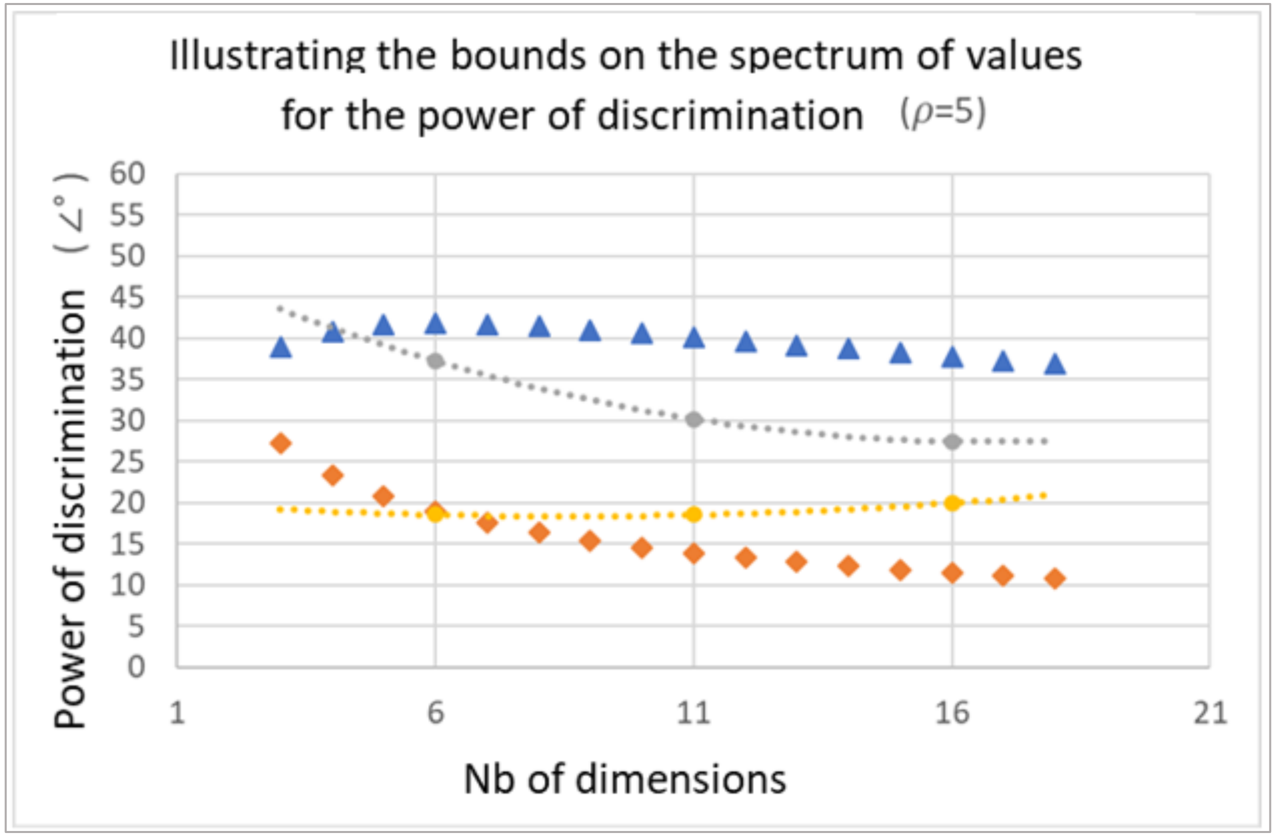


Impact of potential for discrimination

Here the potential (ρ) is set to 5

One can observe that the shape of the values for the upper bounds are reduced significantly from mostly above 50° for $\rho = 9$ to mostly below 40°.

And the upper bound diminishes further as ρ is lowered, thus illustrating that the potential for discrimination is a significant parameter to consider when preparing test cases.

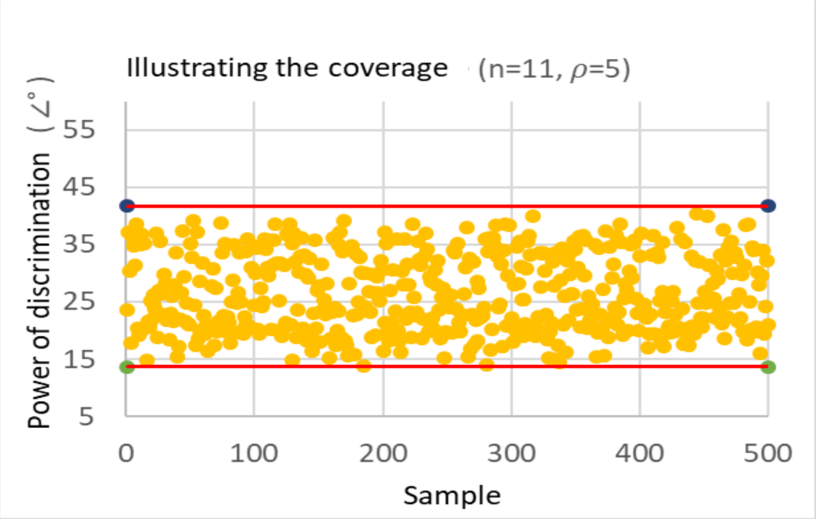


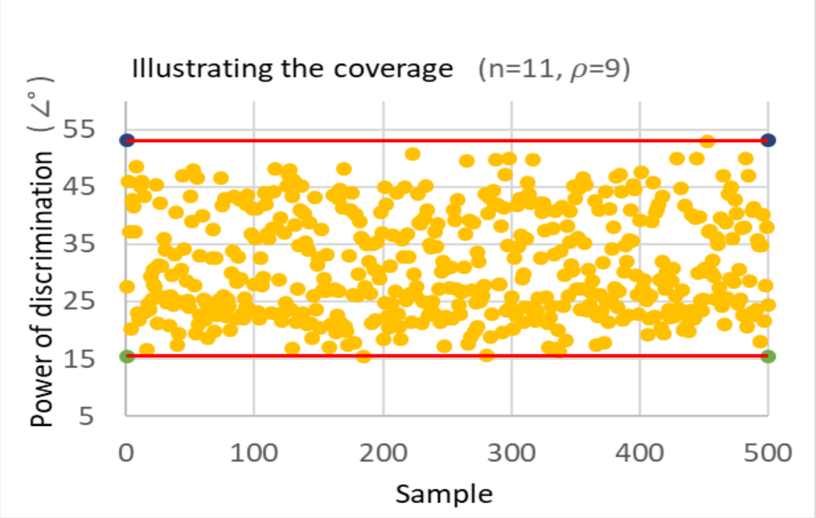
Revised generating process

Here, these two figures (right) illustrate a series of 500 random points generated with a different method that more appropriately covers the spectrum of the power of discrimination for vectors of dimension $n = 11$ and $\rho = 5\ and\ 9$.

Details on the revised generating process* that provides similar results across all combinations of target characteristics can be found in Rivest (2019).

{* see Appendix C for an overview of the revised process}





Proximity measure

In their Encyclopedia of distances, Deza & Deza (2009, p.298) state: « *There are many similarities used in Data Analysis; **the choice** depends on the nature of data and **is not an exact science.*** ».

Measuring the power of discrimination with the cosine distance leads intuitively to the idea of doing the same for the proximity measure. In the original AHP method, vectors are projected on a hyper-plane (normalized to sum 1). In this context, one might argue that the cosine distance is approximately equivalent to the Euclidian distance, which is one of the usual "go to" measures of proximity between vectors. However, certain adaptations use different normalizations. For instance, Schoner et al (1993) explore the use of various normalizations which would make the Euclidian distance less judicious but would not deter the cosine distance.

Thus, the **cosine distance has properties** that make it **more comprehensive** for measuring the proximity of priority vectors, coherent within the evaluation framework, and also easier to interpret.

Purpose of the limit of precision

Why is having an objective limit of precision so pertinent ?

Let's suppose we compare results R1 and R2, which are approximation errors obtained from two different processes. Here, a smaller error is indicative of a better process. With only this information, one can only state whether a process get better results than the other or that they are equivalent. But, it is not possible to assert how satisfactory the results are.

Let's now consider that we have a limit a precision L and that R1 is better than R2. We then have only three possible cases:

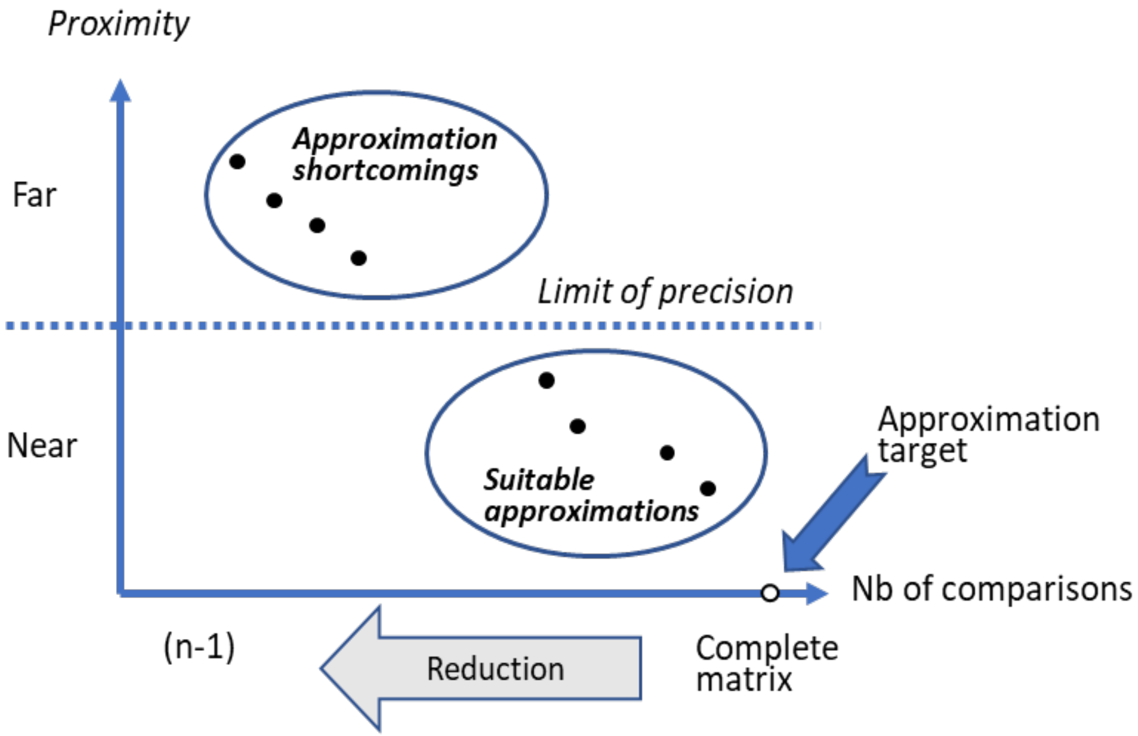
$L < R1 < R2$: Both results are not satisfactory. Further refinements are meaningful.

$R1 < L < R2$: The first result has reached the precision target and the second has not.

$R1 < R2 < L$: Both results are within the precision range. Further improvements are superfluous.

Applying the limit of precision

Applying the "limit of precision" logic to research on the incomplete matrix can provide an upper bound for how far an approximation can deviate while remaining suitable.



This figure (left) illustrates the use of this concept to the reduction in the number of comparisons.

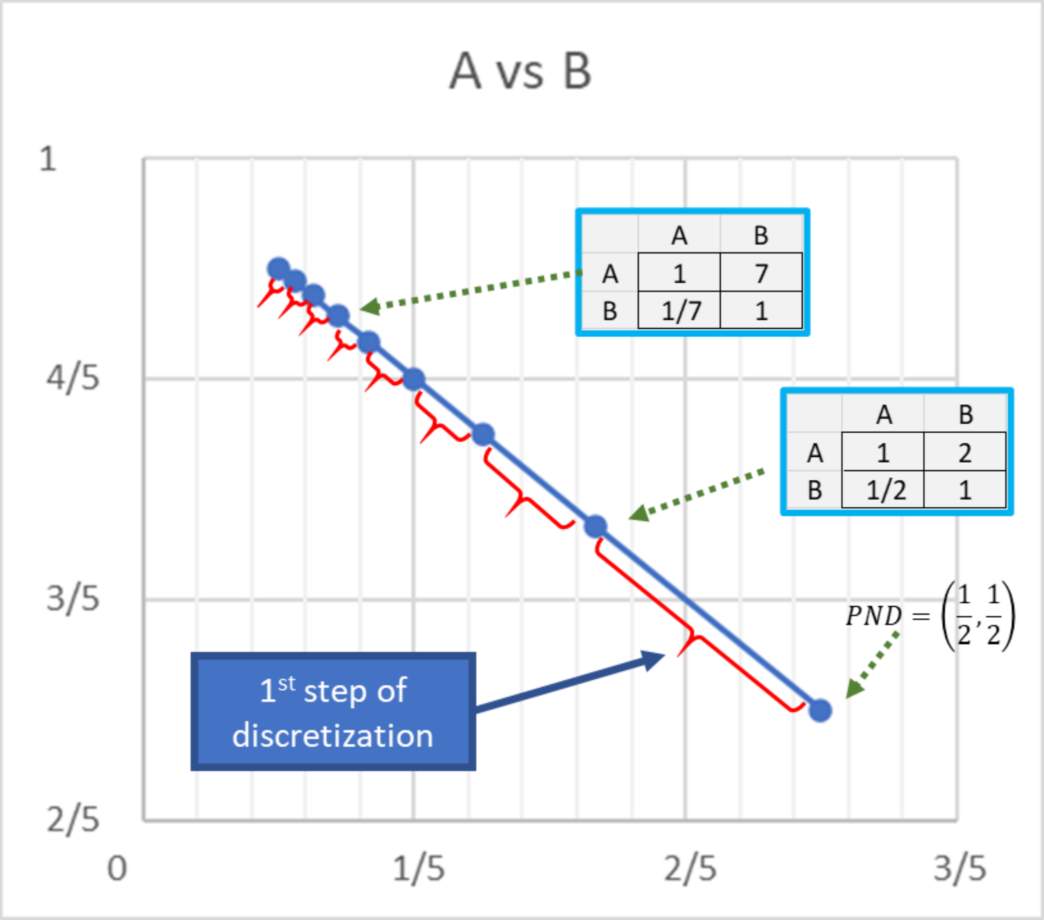
As the number of comparisons is reduced the priority vector obtained may get farther away from the one that would have been obtained if all comparisons had been elicited.

Discretization error

One way to establish such a limit is to consider what Triantaphyllou & Mann (1990, p.297) refer to as the *forward error*, which is attributed to the use of a discrete numerical scale to map elicited expressions of relative importance to numerical values.

The figure (right) illustrates the gaps which might be measured to represent the level of imprecision (or *discretization error*) that cannot be overcome with a given numerical scale (e.g. the original linear scale).

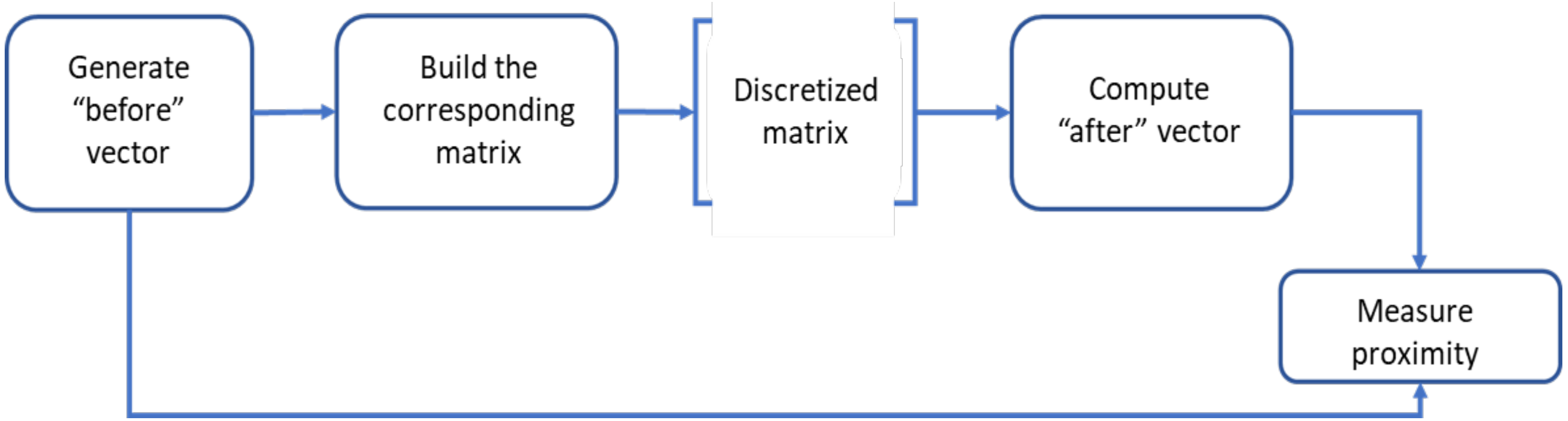
Note: the gap between the PND and the point obtained by using the 2nd lowest value of the numerical scale is referred to as the *1st step of discretization*



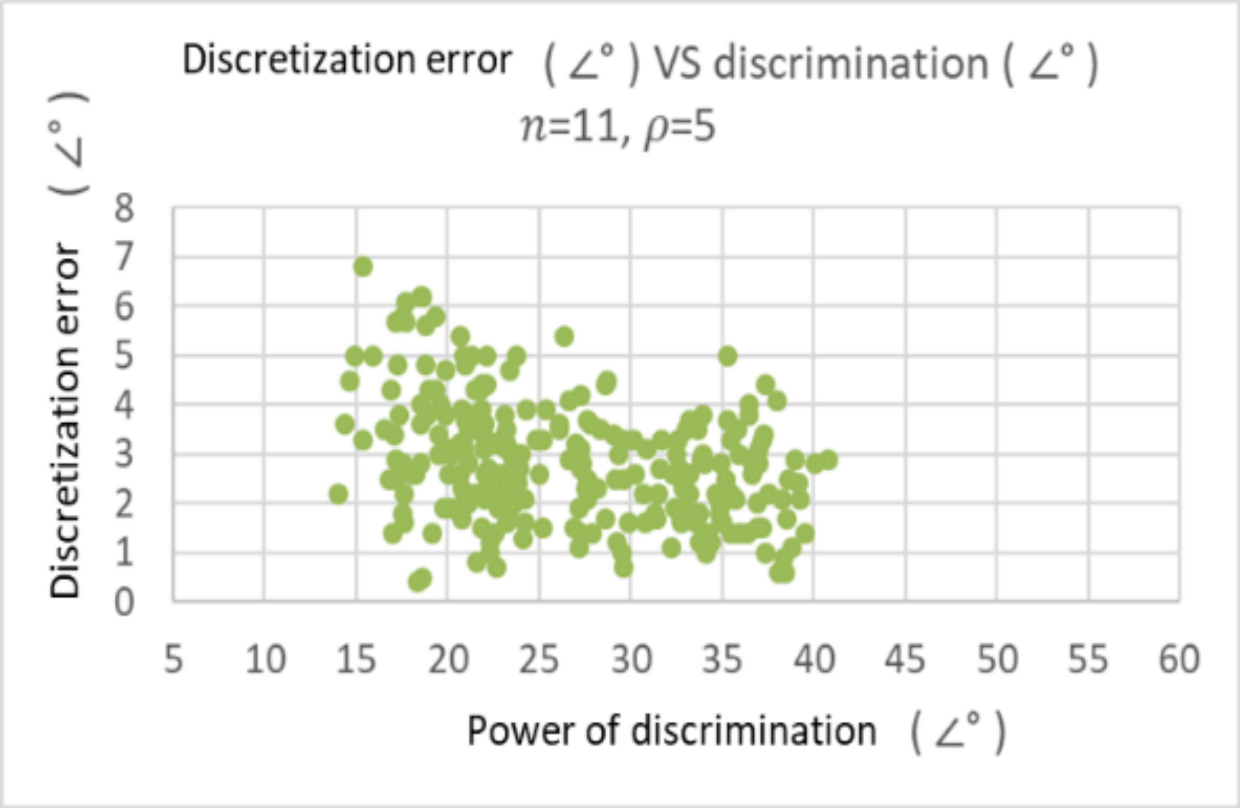
Numerical experiment

Combining the concepts described above and using an appropriate strategy to generate priority vectors, it is possible to establish a justifiable limit of precision in three steps.

First : simulation data have to be produced
Second : discretization must be imparted
Last : consider the results and choose an appropriate value for the intended application



Range for the limit of precision



The figure (left) illustrates the distribution of the discretization error obtained for 250 samples with $n = 11, \rho = 5$. The graph shows that all, but three points (~1%), are less than 6° away. The average angle is 2.85° and the 95th percentile is 5°.

So, one can reasonably make the interpretation that the limit of precision is within the 5° to 6° range.

Revisiting an old debate

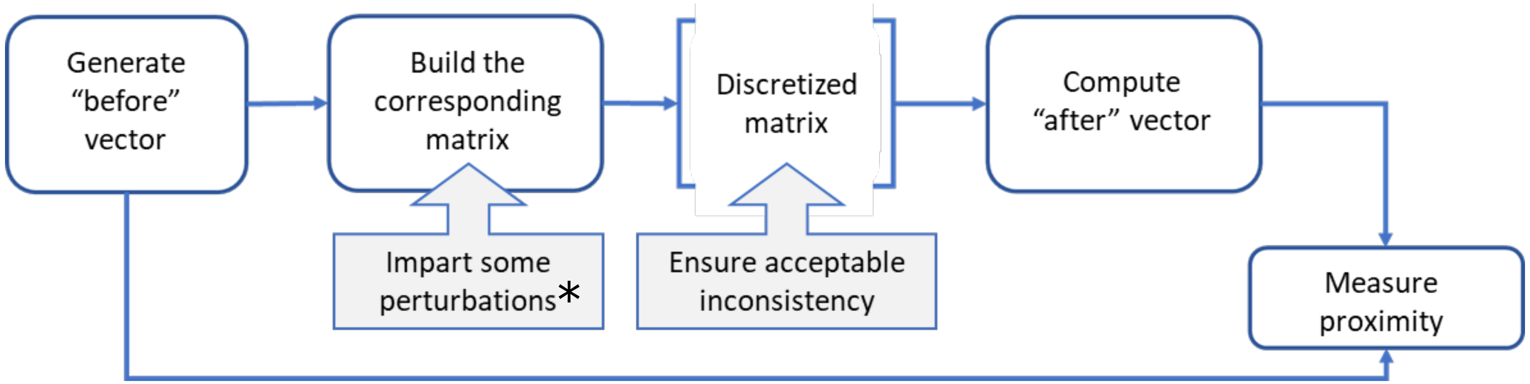
There has been a **long on-going debate** over which method is best suited to derive the priorities from a comparison matrix. Some arguments are made from a purely mathematical perspective while others are made from other standpoints (e.g. practical issues).

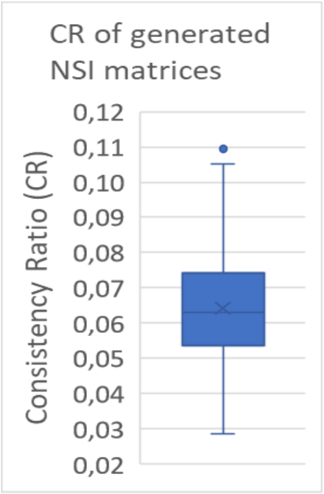
Here it is examined with a different kind of rationale, which can be summed up as: « **What do the actual numbers tell us?** ».

Three methods for obtaining the priority vector are compared: the original right eigenvector (Saaty, 1980), the geometric mean (Williams and Crawford, 1980, p.22) and the cosine maximization (Kou and Lin, 2014).

Generating process in action

To execute this experiment, the following process (adapted from the one used for the limit of precision) is applied.

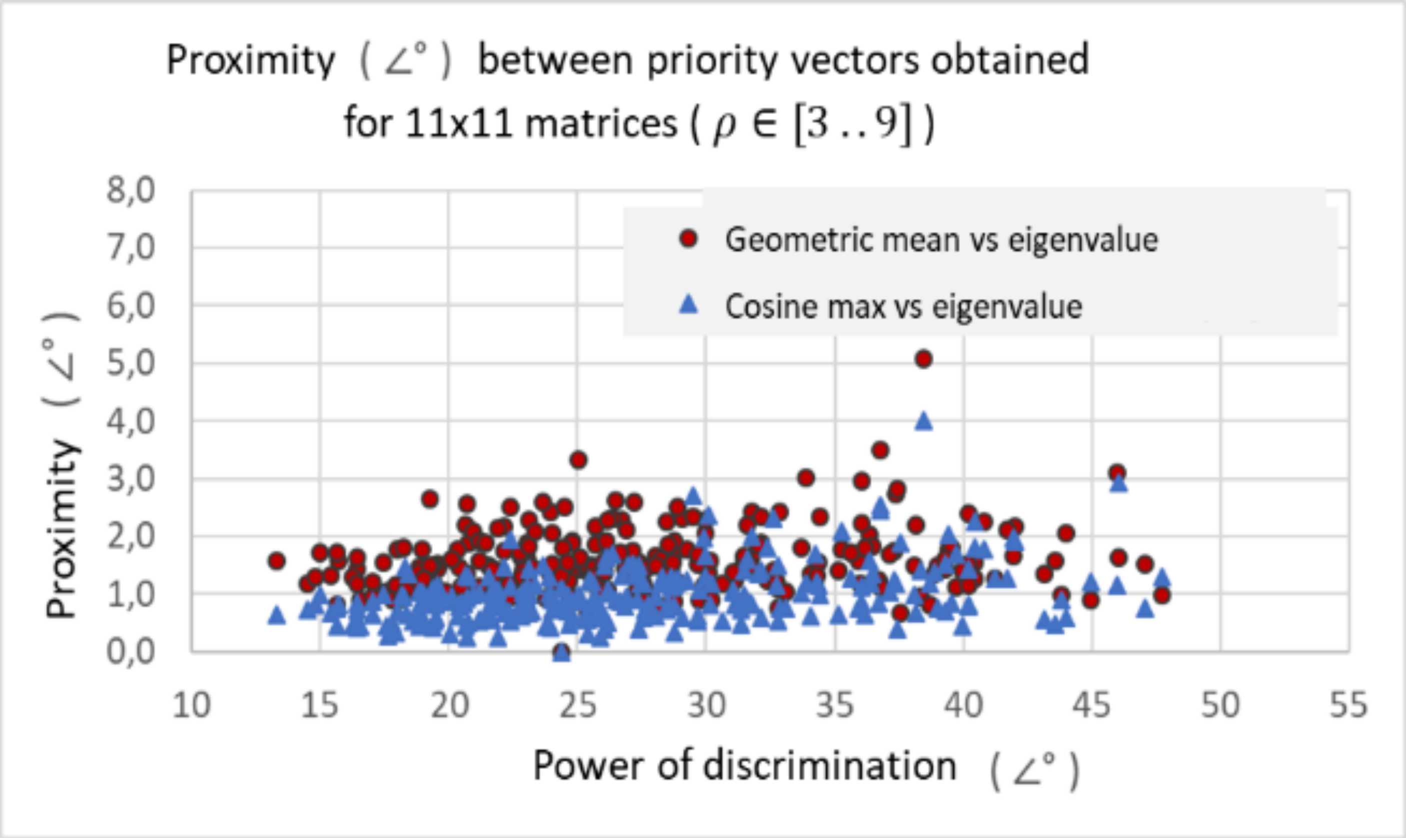




* Herman and Koczkodaj (1996, p.26) describe a method to impart some inconsistency in a generated matrix which they refer to as NSI (or *Not-So-Inconsistent*) matrices. For this application example, the box plot (left) shows the distribution of the consistency ratio (CR) for the set of matrices from which priority vectors were computed.

Within the limit of precision

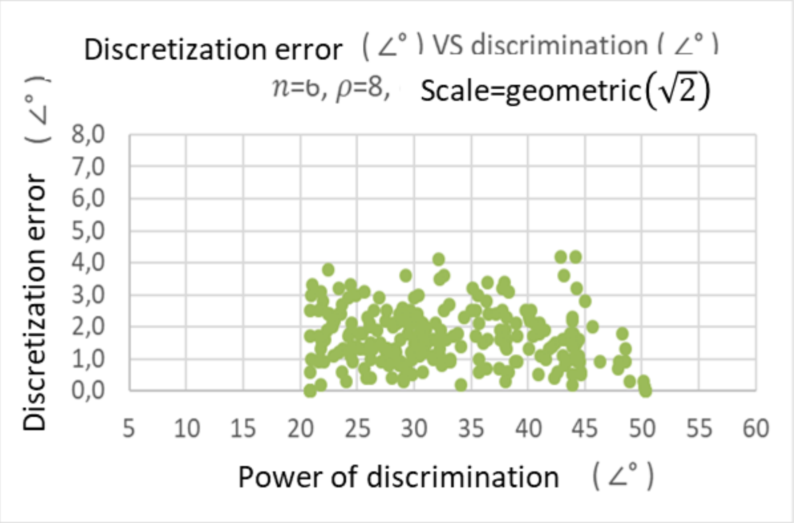
All three methods give results which are well within the limit of precision of each other. Thereby, one can state that, despite the theoretical or conceptual significance of various arguments, in the end, it might not make much of a difference from a practical point of view.

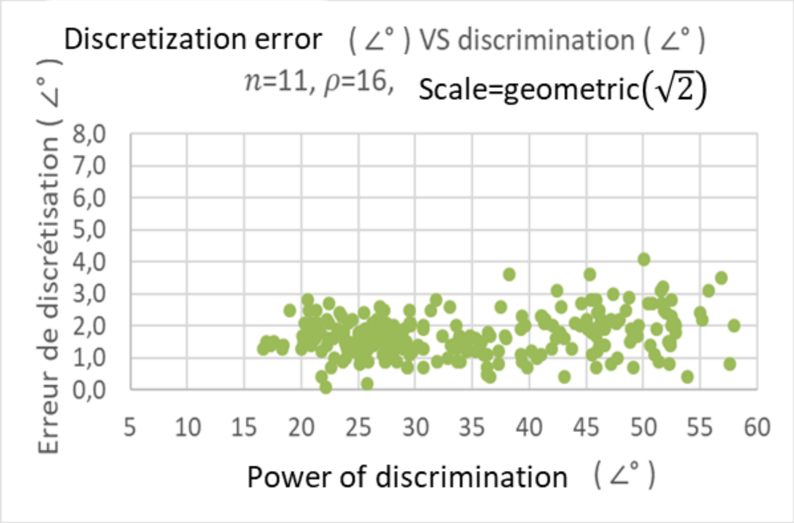


Considerations

The following elements must be taken into consideration before adopting the approach described here to conduct evaluations of potential refinements to the AHP process:

- All simulations and test runs where conducted using the more advanced functions of Microsoft Excel® and require a fair amount of manual interventions. Having a shareable integrated test environment implementing this instrumentation would make adoption easier and more straightforward.
- The limit of precision is dependent on the numerical scale being used, e.g. when using the geometric scale with parameter $\sqrt{2}$, the limit of precision becomes ~ 4° as its discretization gaps are generally less pronounced than those of the linear scale.





Main contributions of this study

- A way to **characterize priority vectors** such that their **value domain can be defined** appropriately in order to provide an **orientation to generate simulation data** that **ensures proper coverage** of test cases

- A **proximity measure with a rational interpretation** to determine the proximity between a priority vector and its approximation obtained via an alternative method (e.g. reduction of comparisons)

- An approach to establish an **objective limit of precision** for priority vectors.

Further information can be found in Rivest (2019).

Key references

Brunelli, M. (2014). Introduction to the analytic hierarchy process. Springer.

Deza, M. M., & Deza, E. (2009). Encyclopedia of distances. In *Encyclopedia of distances* (pp. 1-583). Springer, Berlin, Heidelberg.

Herman, M. W., & Koczkodaj, W. W. (1996). A Monte Carlo Study of Pairwise Comparison. *Inf. Process. Lett., 57(1)*, 25-29.

Ishizaka, A., Pearman, C., & Nemery, P. (2012). AHPSort: an AHP-based method for sorting problems. *International Journal of Production Research*, 50(17), 4767-4784.

Kou, G., & Lin, C. (2014). A cosine maximization method for the priority vector derivation in AHP. *European Journal of Operational Research, 235(1)*, 225-232.

Rivest, R. (2019). Techniques de simulation pour la recherche sur le perfectionnement de la méthode AHP. Mémoire de maîtrise [translation : Simulation techniques for research on AHP refinement, Masters Thesis], *Bibliothèque H.E.C. Montréal*. http://biblios.hec.ca/biblio/memoires/m2019a612764.pdf

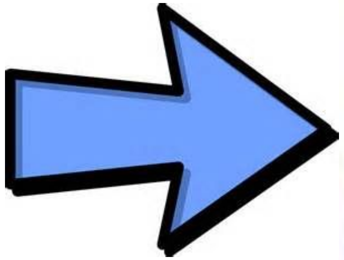
Saaty, T.L. (1980) The Analytic Hierarchy Process. McGraw-Hill, New York.

Schoner, B., Wedley, W. C., & Choo, E. U. (1993). A unified approach to AHP with linking pins. *European Journal of Operational Research, 64(3)*, 384-392.

Triantaphyllou, E., & Mann, S. H. (1990). An evaluation of the eigenvalue approach for determining the membership values in fuzzy sets. *Fuzzy Sets and Systems, 35(3)*, 295-301.

Williams, C., & Crawford, G. (1980). Analysis of Subjective Judgment Matrices (No. RAND/R-2572-AF). RAND CORP SANTA MONICA CA.

Questions, comments !



Appendices

Appendix A : Power of discrimination – upper and lower bound determination

The upper θ_U and lower θ_L bounds of the *power of discrimination* (°) for given dimension n and potential ρ are found by solving the following two nonlinear programs:

$$\theta_U(\rho) = max \frac{180}{\pi} \cdot cos^{-1}\left(\frac{v \circ PND}{\sqrt{v \circ v} \cdot \sqrt{PND \circ PND}}\right)$$

s.t. $\sum_{i=1}^{n} v_i = 1$ (1)

$\frac{\max v_i}{\min v_i} = \rho$ (2)

$v \in \mathbb{R}_{+}^{n}$ (3)

$$\theta_L(\rho) = min \frac{180}{\pi} \cdot cos^{-1}\left(\frac{v \circ PND}{\sqrt{v \circ v} \cdot \sqrt{PND \circ PND}}\right)$$

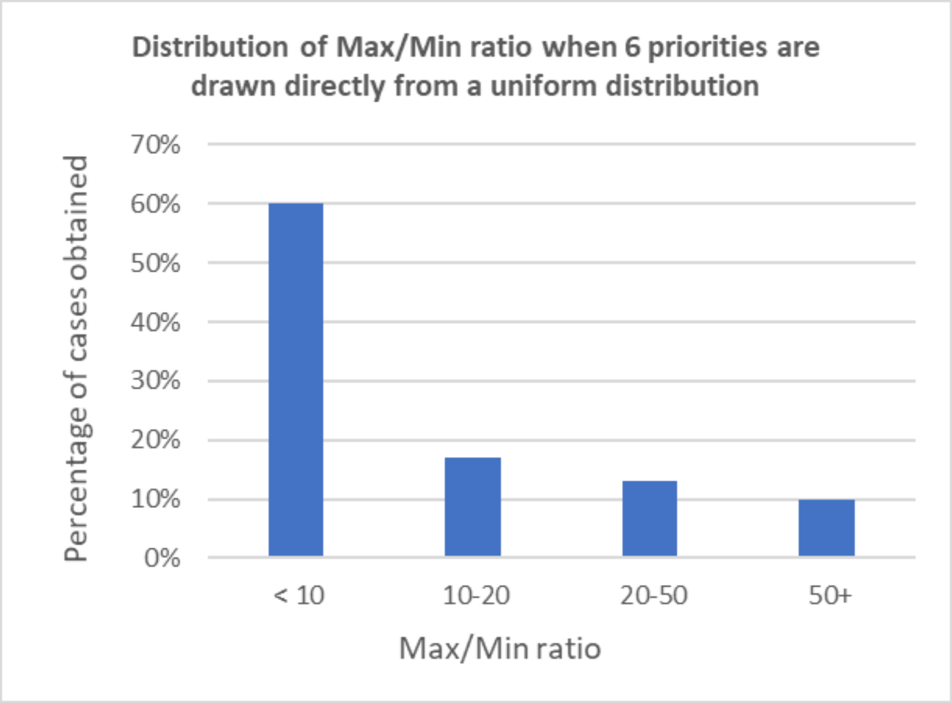
s.t. $\sum_{i=1}^{n} v_i = 1$ (1)

$\frac{\max v_i}{\min v_i} = \rho$ (2)

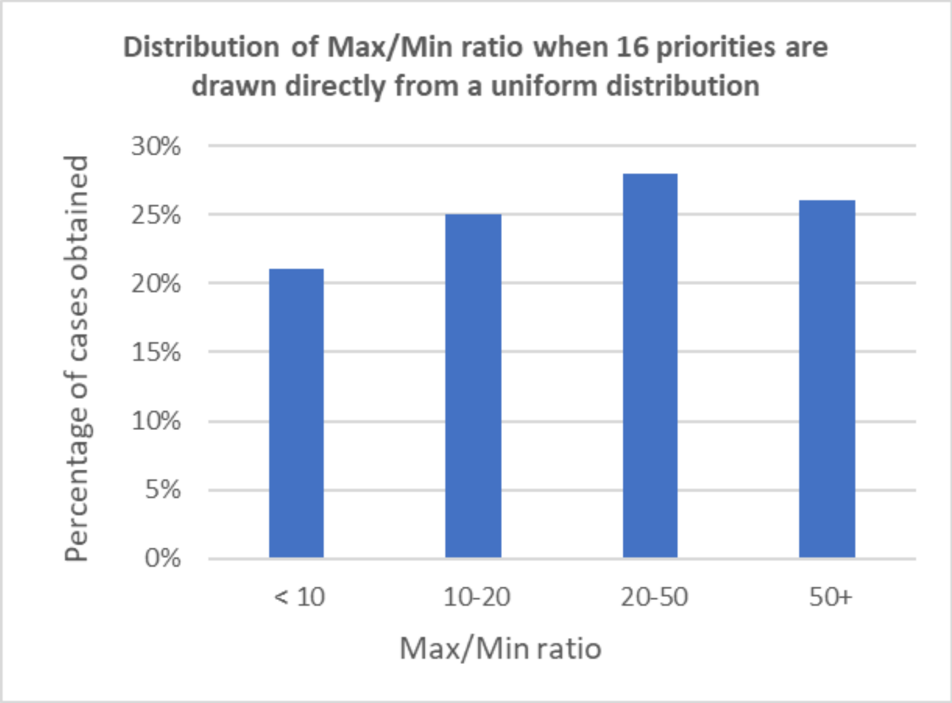
$v \in \mathbb{R}_{+}^{n}$ (3)

Appendix B : Distribution of Max/Min ratios when drawing priorities from uniform

Only 60% of generated cases are below the upper bound of the numerical scale for vectors of dimension 6. Thus, 40% of cases do not represent situations that can be encountered empirically. And it gets much worst as the number of dimensions increases.



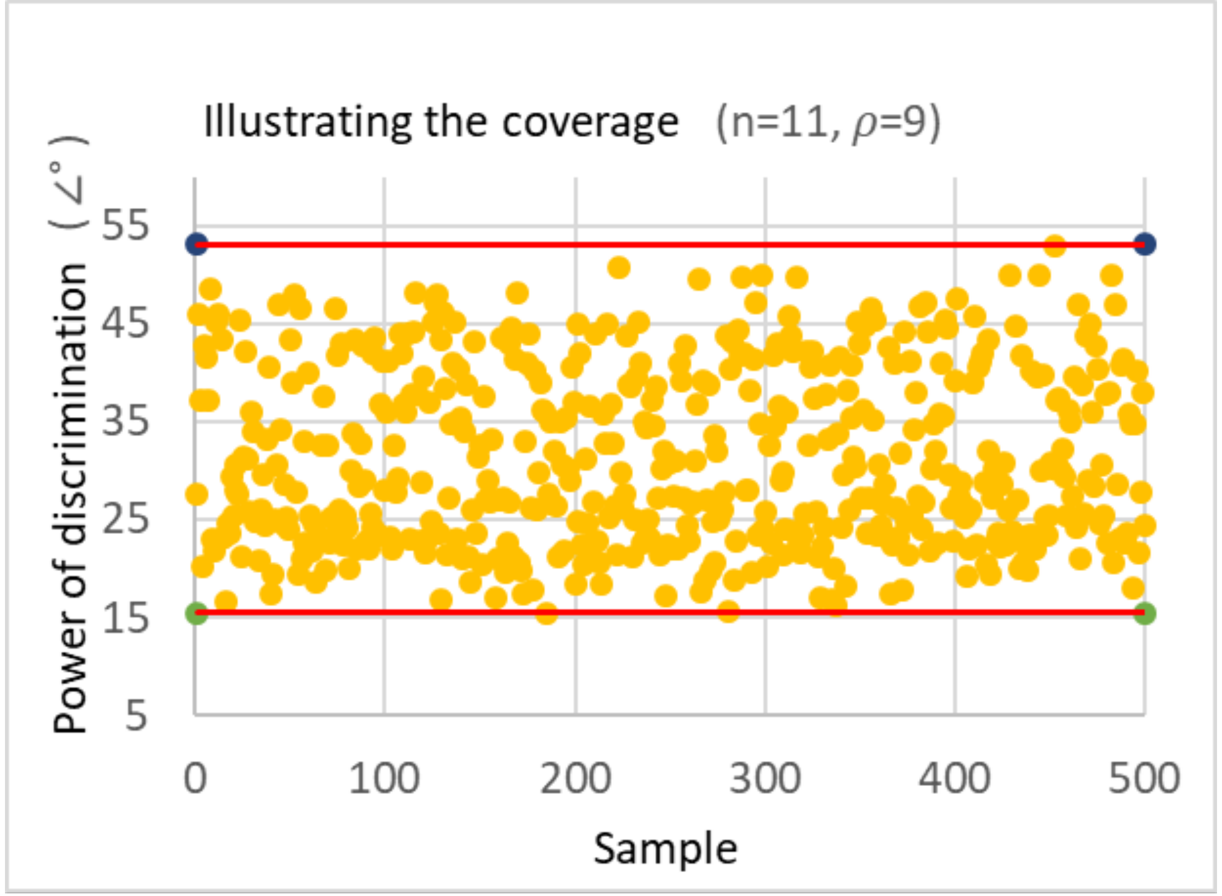
Based on a sample of 1,000 vectors of dimension 6 for which each element was drawn from a uniform distribution



Based on a sample of 1,000 vectors of dimension 16 for which each element was drawn from a uniform distribution

Appendix C : Revised generating process (1 of 4)

The objective for the generation process is to produce a collection of priority vectors of dimension n that properly covers the specific spectrum of power of discrimination from the lower to the upper bound as shown in the figure (below) for a given potential of discrimination ρ.

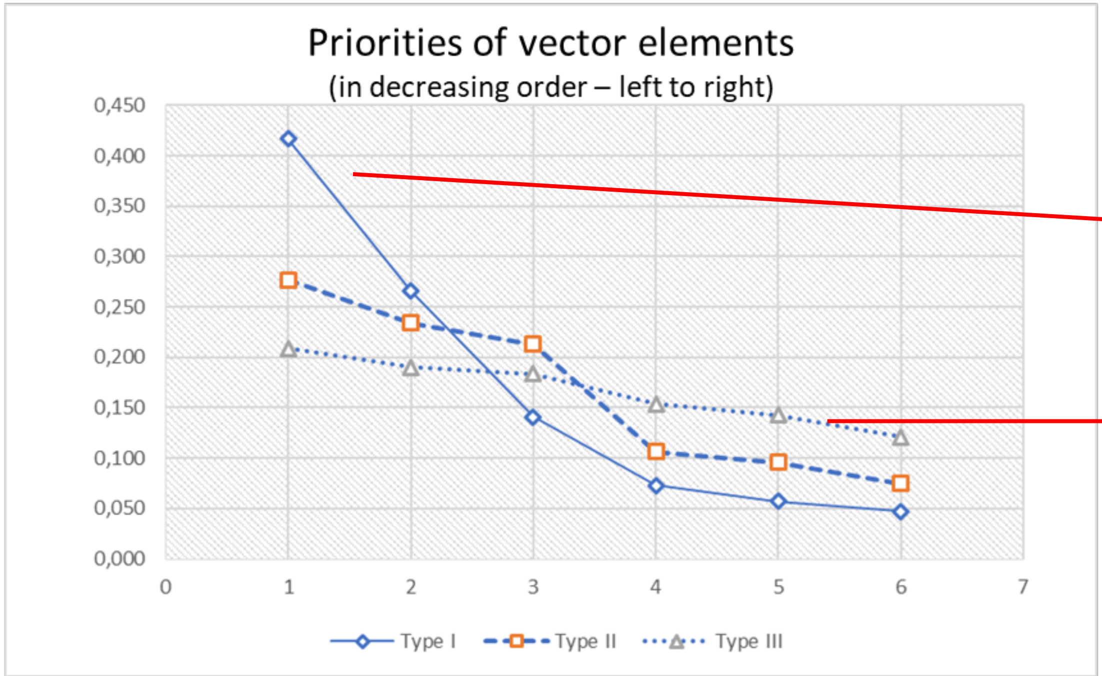


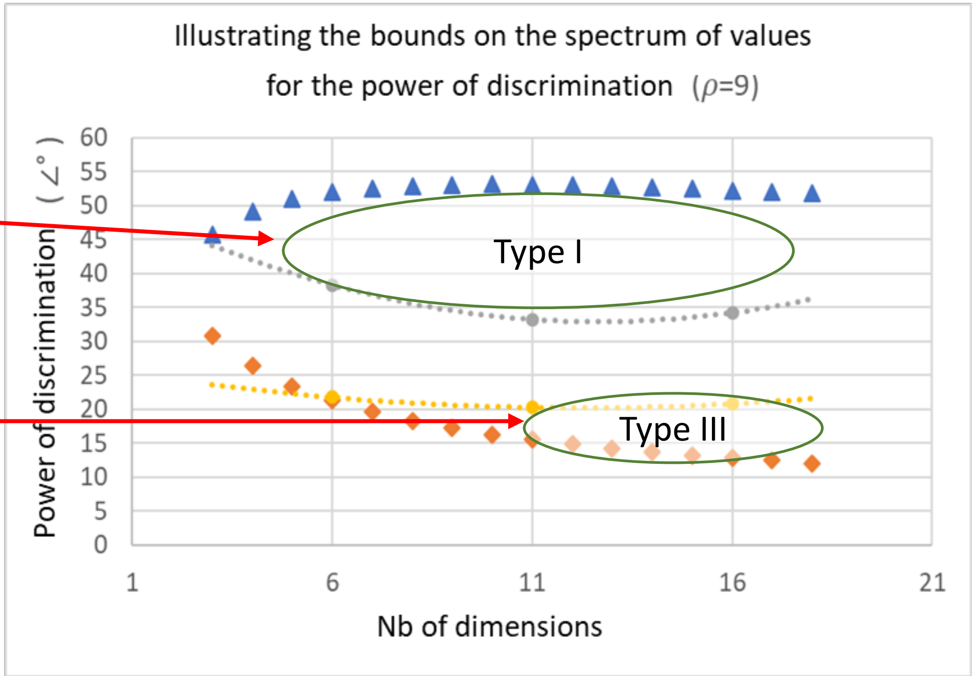
Revised generating process (2 of 4)

Let's first consider the three generic types of priority distributions (below – left):

- Type I : Highly discriminating
- Type II : Somewhat discriminating
- Type III : Close to "no difference"

We can observe (below – right) that types I and III correspond to the zones left vacant by the usual generating process (i.e. drawing priorities directly from a *uniform* distribution)





Revised generating process (3 of 4)

So, in order to enhance the generating process, the following logic facilitates the addition of priority vectors of types I and III

- First, set the highest element to ρ and the lowest to 1, thus securing the constraint $\frac{max\ v_i}{min\ v_i} = \rho$

- Second, draw the other $n-2$ elements from a probability distribution over $[1..\rho]$ that will favor values either closer to ρ (for type I vectors) or closer to 1 (for type III vectors)

- Finally, divide each element by the sum of all elements

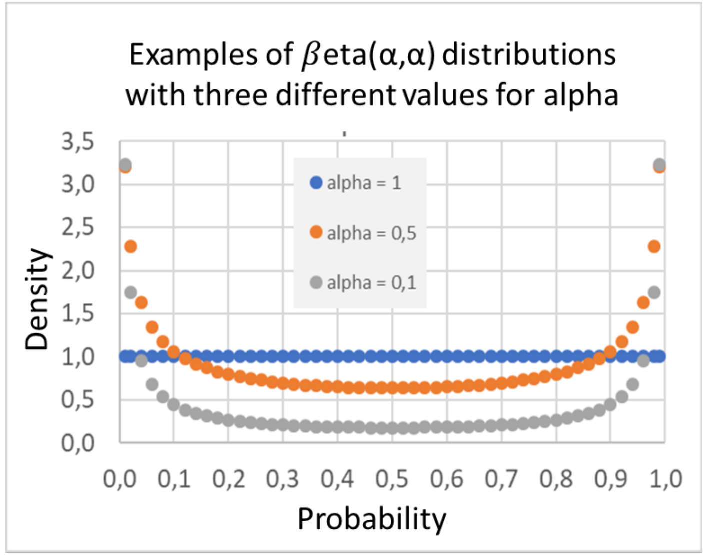
Revised generating process (4 of 4)

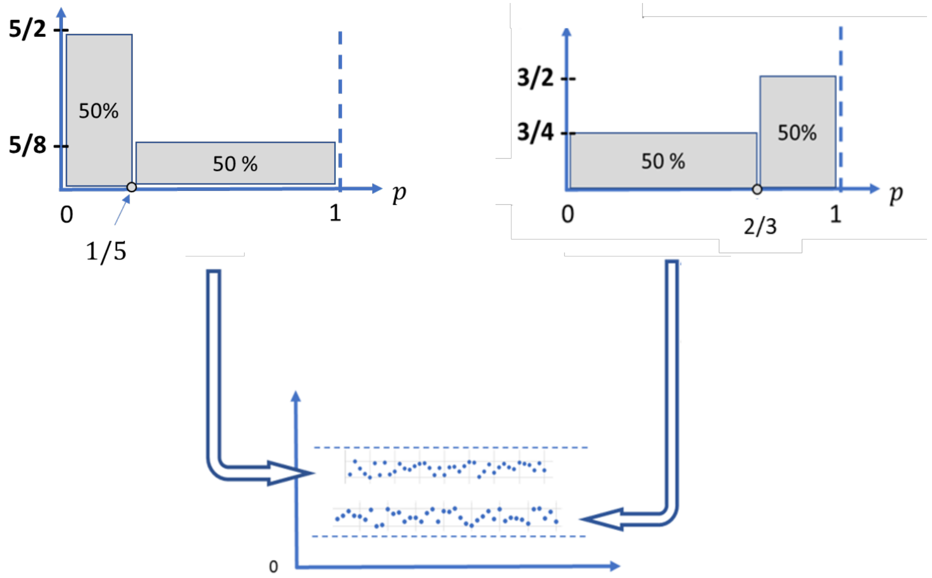
There are multiple ways to accomplish this.
The following three steps describe one such way.

- First, draw from a properly calibrated βeta distribution that favors values closer to the bounds.

- Second, use that value to build a mixture of two uniform distributions which will favor either low or high values for the power of discrimination.

- Last, draw values from the mixture.





© Robin Rivest