# From Numerical Probabilities to Linguistic Probabilities: A Theoretical Justification of Empirical Granules Used in Risk Management


Beverly Rivera[1], Francisco Zapata[2], and Vladik Kreinovich[1,2]

[1]Computational Science Program

[2]TRACS Center

Department of Computer Science

University of Texas at El Paso

El Paso, TX 79968, USA

barivera@miners.utep.edu, fazg74@gmail.com,

vladik@utep.edu



**Abstract**

In many risk management situations, instead of the exact probability values, specialists use a granule to which this probability belongs. Specifically, they use five granules, corresponding to thresholds 10%, 40%, 60%, and 90%. In this paper, we provide an explanation for such non-uniform granulation.


## 1 Formulation of the Problem

**In risk management, it is often important to group probabilities into granules.** One of the main objectives of risk management is to minimize the expected loss. By definition, the expected loss is equal to $\sum\limits_{i=1}^{n} p_i \cdot x_i$, where $x_i$ is the loss caused by the $i$-th event and $p_i$ is the probability of the $i$-th event.

In many practical situations, we have only crude estimates of potential losses $x_i$ caused by different possible events. In this case, we can only get crude approximations for the expected loss. Thus, it does not make sense to estimate the probabilities $p_i$ with high accuracy. For example, if we know that the loss is between 10 and 100 million dollars, it does not make sense to distinguish between the probabilities of 10% or 11%, it is sufficient to estimate probabilities with a similar low accuracy.

In other words, instead of considering exact probabilities, it makes sense to group possible probability values into granules, so that instead of the actual values of the probability, we use, in our estimates, only the granule containing this value.

**Often, these probability granules have commonsense meaning.** Risk management under uncertainty uses approximate estimates to provide recommendations for managers and other decision makers. Decision makers understand that since these recommendations are based on approximate estimates, they may, in principle, lead to suboptimal decisions. In such situations, it is usually more convincing if a decision not only comes from the corresponding mathematical model, but it is also supported by commonsense reasoning.

For example, in addition to a range of probabilities, it is desirable to provide a natural-language description of this probability, something like "unlikely to occur" or "very unlikely to occur". We have intuitive understanding of such descriptions – in contrast to numbers – and thus, we can feel whether the resulting recommendations sound right.

$7 \pm 2$ **law.** Psychologists have found that in general, we usually divide each quantity into 7 plus plus minus 2 categories – this is the largest number of categories whose meaning we can immediately grasp; see, e.g., [3, 6]. For some people, this "magical number" is $7 + 2 = 9$, for some it is $7 - 2 = 5$. So, to make sure that everyone can grasp the intuitive meaning of the corresponding categories, we need to use a division into no more than 5 granules.

**Five probability granules: empirical description.** Empirically, it was found out that the five probability granules corresponding to actual human decisions are as follows [1, 4]:

- values from 0 to 10% corresponding to "Very unlikely to occur"
- valued from 11% to 40% corresponding to "Unlikely to occur";
- valued from 41% to 60% corresponding to "May occur about half of the time";
- valued from 61% to 90% corresponding to "Likely to occur";
- valued from 91% to 100% corresponding to "Very likely to occur".

In other words, the empirical granules correspond to thresholds 10%, 40%, 60%, and 90%.

**Remaining problem.** If we want to divide the interval from 0 to 100% into five granules, a natural idea is to divide it uniformly, by using equally spaced thresholds 20%, 40%, 60%, and 80%. So why are these thresholds different? How can we explain these particular thresholds?

This is a problem that we solve in this paper.

## 2 Proposed Explanation

**Main idea.** The main idea behind our explanation is that if the two probability values are close to each other, intuitively, we do not feel the difference. For example, there is a clear different between 10% chance of rain or 50% chance of rain, but we do not think that anyone can feel the difference between 50% and 51% chances. So, the discrete scale is formed by probabilities which are distinguishable from each other.

In other words, instead of all possible probability values from the interval $[0, 1]$, we consider only a discrete number of layers, from 0-th to $m$-th, for some small integer $m$.

Let us show how this idea can be formalized.

**When probabilities are distinguishable?** Probability of an event is estimated, from observations, as the frequency with which this event occurs. For example, if out of 100 days of observation, rain occurred in 40 of these days, then we estimate the probability of rain as 40%. In general, if out of $n$ observations, the event was observed in $m$ of them, we estimate the probability as the ratio $\frac{m}{n}$.

This ratio is, in general, different from the actual (unknown) probability. For example, if we take a fair coin, for which the probability of head is exactly 50%, and flip it 100 times, we may get 50 heads, but we may also get 47 heads, 52 heads, etc.

It is known (see, e.g., [7]), that the expected value of the frequency is equal to $p$, and that the standard deviation of this frequency is equal to

$$\sigma = \sqrt{\frac{p \cdot (1-p)}{n}}.$$

It is also known that, due to the Central Limit Theorem, for large $n$, the distribution of frequency is very close to the normal distribution (with the corresponding mean $p$ and standard deviation $\sigma$).

For normal distribution, we know that with a high certainty all the values are located within 2-3 standard deviations from the mean, i.e., in our case, within the interval $(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma)$, where $k_0 = 2$ or $k_0 = 3$: for example, for $k_0 = 3$, this is true with confidence 99.9%. We can thus say that the two values of probability $p$ and $p'$ are (definitely) distinguishable if the corresponding intervals of possible values of frequency do not intersect – and thus, we can distinguish between these two probabilities just by observing the corresponding frequencies.

In precise terms, the probabilities $p < p'$ are distinguishable if

$$(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma) \cap (p' - k_0 \cdot \sigma', p + k_0 \cdot \sigma') = \emptyset,$$

where

$$\sigma' \stackrel{\text{def}}{=} \sqrt{\frac{p' \cdot (1-p')}{n}},$$

i.e., if $p' - k_0 \cdot \sigma' \geq p + k_0 \cdot \sigma$. The smaller $p'$, the smaller the difference $p' - k_0 \cdot \sigma'$. Thus, for a given probability $p$, the next distinguishable value $p'$ is the one for which

$$p' - k_0 \cdot \sigma' = p + k_0 \cdot \sigma.$$

When $n$ is large, these value $p$ and $p'$ are close to each other; therefore, $\sigma' \approx \sigma$. Substituting an approximate value $\sigma$ instead of $\sigma'$ into the above equality, we conclude that

$$p' \approx p + 2k_0 \cdot \sigma = p + 2k_0 \cdot \sqrt{\frac{p \cdot (1-p)}{n}}.$$

If the value $p$ corresponds to the $i$-th level, then the next value $p'$ corresponds to the $(i+1)$-st level. Let us denote the value corresponding to the $i$-th level by $p(i)$. In these terms, the above formula takes the form

$$p(i+1) - p(i) = 2k_0 \cdot \sqrt{\frac{p \cdot (1-p)}{n}}.$$

The above notation defines the value $p(i)$ for non-negative integers $i$. We can extrapolate this dependence so that it will be defined for all non-negative real values $i$.

When $n$ is large, the values $p(i+1)$ and $p(i)$ are close, the difference

$$p(i+1) - p(i)$$

is small, and therefore, we can expand the expression $p(i+1)$ in Taylor series and keep only linear terms in this expansion:

$$p(i+1) - p(i) \approx \frac{dp}{di}.$$

Substituting the above expression for $p(i+1) - p(i)$ into this formula, we conclude that

$$\frac{dp}{di} = \text{const} \cdot \sqrt{p \cdot (1-p)}.$$

Moving all the terms containing $p$ into the left-hand side and all the terms containing $i$ into the right-hand side, we get

$$\frac{dp}{\sqrt{p \cdot (1-p)}} = \text{const} \cdot di.$$

Integrating this expression and taking into account that $p = 0$ corresponds to the lowest 0-th level – i.e., that $i(0) = 0$ – we conclude that

$$i(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1-q)}}.$$

This integral can be easily computed if introduce a new variable $t$ for which $q = \sin^2(t)$. In this case,

$$dq = 2 \cdot \sin(t) \cdot \cos(t) \cdot dt,$$

$1 - p = 1 - \sin^2(t) = \cos^2(t)$ and therefore,

$$\sqrt{p \cdot (1-p)} = \sqrt{\sin^2(t) \cdot \cos^2(t)} = \sin(t) \cdot \cos(t).$$

The lower bound $q = 0$ corresponds to $t = 0$ and the upper bound $q = p$ corresponds to the value $t_0$ for which $\sin^2(t_0) = p$ – i.e., $\sin(t_0) = \sqrt{p}$ and $t_0 = \arcsin\left(\sqrt{p}\right)$. Therefore,

$$i(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1-q)}} = \text{const} \cdot \int_0^{t_0} \frac{2 \cdot \sin(t) \cdot \cos(t) \cdot dt}{\sin(t) \cdot \cos(t)} =$$

$$\int_0^{t_0} 2 \cdot dt = 2 \cdot \text{const} \cdot t_0.$$

We know how $t_0$ depends on $p$, so we get

$$i(p) = 2 \cdot \text{const} \cdot \arcsin\left(\sqrt{p}\right).$$

We can determine the constant from the condition that the largest possible probability value $p = 1$ should correspond to the largest level $i = m$. From the condition that $i(1) = m$, taking into account that

$$\arcsin\left(\sqrt{1}\right) = \arcsin(1) = \frac{\pi}{2},$$

we conclude that

$$i(p) = \frac{2m}{\pi} \cdot \arcsin\left(\sqrt{p}\right).$$

Thus,

$$\arcsin\left(\sqrt{p}\right) = \frac{\pi \cdot i}{2m},$$

hence

$$\sqrt{p} = \sin\left(\frac{\pi \cdot i}{2m}\right)$$

and thus,

$$p = \sin^2\left(\frac{\pi \cdot i}{2m}\right).$$

*Comment.* Similar computations appeared, in a different context, in [2, 5].

**Resulting values.** For each $m$, the corresponding values divide the interval $[0, 1]$ of possible values of probability into $m$ granules:

- from $p(0) = 0$ to $p(1)$,
- from $p(1)$ to $p(2)$,
- $\ldots$,
- from $p(i)$ to $p(i+1)$,
- $\ldots$,
- from $p(m-1)$ to $p(m) = 1$.

In our case, we want to divide the interval $[0, 1]$ into five granules, so we take $m = 5$. For $m = 5$, we get the following values:

$$p(1) = \sin^2\left(\frac{\pi}{10}\right) \approx 0.095 \approx 10\%, \quad p(2) = \sin^2\left(\frac{2\pi}{10}\right) \approx 0.35 \approx 40\%,$$

$$p(3) = \sin^2\left(\frac{2\pi}{10}\right) \approx 0.55 \approx 60\%, \quad p(4) = \sin^2\left(\frac{2\pi}{10}\right) \approx 0.90 = 90\%.$$

In other words, in the first approximation, we indeed get the empirical values of 10%, 40%, 60%, and 90%.

## Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

## References

[1] P. E. Engert and Z. F. Lannsdowne, *Risk Matrix User's Guide*, Version 2.2, The MITRE Corporation, Document MP99B000029, November 1999.

[2] J. Lorkowski and V. Kreinovich, "Fuzzy Logic Ideas Can Help in Explaining Kahneman and Tversky's Empirical Decision Weights", *Proceedings of the 4th World Conference on Soft Computing*, Berkeley, California, May 25–27, 2014, pp. 285–289.

[3] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *Psychological Review*, 1956, Vol. 63, No. 2, pp. 81–97.

[4] MITRE Corporation, "Definition of Occurrence", In: *Risk Management Toolkit*, Standard Process http://www2.mitre.org/work/sepo/toolkits/risk/StandardProcess/definitions/occurence.html (downloaded on August 19, 2014)

[5] H. T. Nguyen, V. Kreinovich, and B. Lea, "How to combine probabilistic and fuzzy uncertainties in fuzzy control", *Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, December 2–4, 1992, pp. 117–121.

[6] S. K. Reed, *Cognition: Theories and application*, Wadsworth Cengage Learning, Belmont, California, 2010.

[7] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2011.