Normalization-Invariant Fuzzy Logic Operations Explain Empirical Success of Student Distributions in Describing Measurement Uncertainty

Hamza Alkhatib[1], Boris Kargoll[1],
Ingo Neumann[1], and Vladik Kreinovich[2]

[1]Geodätisches Institut, Leibniz Universität Hannover
Nienburger Strasse 1, 30167 Hannover, Germany
alkhatib@gih.uni-hannover.de, kargoll@gih.uni-hannover.de
neumann@gih.uni-hannover.de
[2]Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

1. Traditional Engineering Approach to Measurement Uncertainty

- Traditionally, in engineering applications, it is assumed that the measurement error is normally distributed.
- This assumption makes perfect sense from the practical viewpoint.
- For the majority of measuring instruments, the measurement error is indeed normally distributed.
- It also makes sense from the theoretical viewpoint:
  - the measurement error often comes from a joint effect of many independent small components,
  - so, according to the Central Limit Theorem, the resulting distribution is indeed close to Gaussian.

2. Traditional Engineering Approach (cont-d)

- Another explanation: we only have partial information about the distribution.
- Often, we only know the first and the second moments.
- The first moment – mean – represents a bias.
- If we know the bias, we can always subtract it from the measurement result.
- Thus re-calibrated measuring instrument will have 0 mean.
- Thus, we can always safely assume that the mean is 0.
- Then, the 2nd moment is simply the variance $V = \sigma^2$.

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

3. Traditional Engineering Approach (cont-d)

- There are many distributions w/0 mean and given σ.
- For example, we can have a distribution in which we have σ and $-\sigma$ with probability 1/2 each.
- However, such a distribution creates a false certainty – that no other values of x are possible.
- Out of all such distributions, it makes sense to select the one which maximally preserves the uncertainty.
- Uncertainty can be gauged by average number of binary questions needed to determine x with accuracy ε.
- It is described by *entropy* $S = -\int \rho(x) \cdot \log_2(\rho(x))\,dx$.
- Out of all distributions $\rho(x)$ with mean 0 and given σ, the entropy is the largest for normal $\rho(x)$.

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

4. Need for Heavy-Tailed Distributions

- For the normal distribution,
$$\rho(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{x^2}{2\sigma^2}\right).$$
- The "tails" – values corresponding to large $|x|$ – are very light, practically negligible.
- Often, $\rho(x)$ decreases much slower, as $\rho(x) \sim c \cdot x^{-\alpha}$.
- We cannot have $\rho(x) = c\cdot x^{-\alpha}$, since $\int_0^\infty x^{-\alpha}\,dx = +\infty$, and we want $\int \rho(x)\,dx = 1$.
- Often, the measurement error is well-represented by a Student distribution $\rho_S(x) = (a + b\cdot x^2)^{-\nu}$.
- Our experience is from geodesy, but the Student distributions is effective in other applications as well.
- This distribution is even recommended by the International Organization for Standardization (ISO).

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

5. What We Do

- How to explain the empirical success of Student's distribution $\rho_S(x)$?
- We show that a fuzzy formalization of commonsense requirements leads to $\rho_S(x)$.
- Our idea: uncertainty means that the first value is possible, and the second value is possible, etc.
- Let's select $\rho(x)$ with the largest degree to which all the values are possible.
- It is reasonable to use fuzzy logic to describe degrees of possibility.
- An expert marks his/her degree by selecting a number from the interval $[0, 1]$.

Traditional . . .

Need for Heavy-Tailed . . .

What We Do

Need for Normalization

How to Combine Degrees

Deriving Student . . .

Acknowledgments

6. Need for Normalization

- For "small", we are absolutely sure that 0 is small: $\mu_{\text{small}}(0) = 1$ and $\max\limits_x \mu_{\text{small}}(x) = 1$.
- For "medium", there is no x with $\mu_{\text{med}}(x) = 1$, so $\max\limits_x \mu_{\text{med}}(x) < 1$.
- A usual way to deal with such situations is to *normalize* $\mu(x)$ into $\mu'(x) = \dfrac{\mu(x)}{\max\limits_y \mu(y)}$.
- Normalization is also needed performed when we get additional information.
- Example: we knew that x is small, we learn that $x \geq 5$.
- Then, $\mu_{\text{new}}(x) = \mu_{\text{small}}(x)$ for $x \geq 5$ and $\mu_{\text{new}}(x) = 0$ for $x < 5$, and $\max\limits_x \mu_{\text{new}}(x) < 1$.

7. Need for Normalization (cont-d)

- Normalization is also needed when experts use probabilities to come up with the degrees.
- Indeed, the larger $\rho(x)$, the more probable it is to observe a value close to x.
- Thus, it is reasonable to take the degrees $\mu(x)$ proportional to $\rho(x)$: $\mu(x) = c \cdot \rho(x)$.
- Normalization leads to $\mu(x) = \dfrac{\rho(x)}{\max\limits_y \rho(y)}$.
- Vice versa, if we have the result $\mu(x)$ of normalizing a pdf, we can reconstruct $\rho(x)$ as $\rho(x) = \dfrac{\mu(x)}{\int \mu(y)\,dy}$.

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

8. How to Combine Degrees

- For each x, we thus get a degree to which x is possible.
- We want to compute the degree to which x_1 is possible *and* x_2 is possible, etc.
- So, we need to apply an "and"-operation (t-norm) to the corresponding degrees.
- Natural idea: use normalization-invariant t-norms.
- We can compute the normalized degree of confidence in a statement $A \,\&\, B$ in two different ways:
  - we can normalize $f_{\&}(a,b)$ to $\lambda \cdot f_{\&}(a,b)$;
  - or, we can first normalize a and b and then apply an "and"-operation: $f_{\&}(\lambda \cdot a, \lambda \cdot b)$.
- It's reasonable to require that we get the same estimate: $f_{\&}(\lambda \cdot a, \lambda \cdot b) = \lambda \cdot f_{\&}(a,b)$.

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

9. How to Combine Degrees (cont-d)

- It is known that Archimedean t-norms $f_\&(a,b) = f^{-1}(f(a)+f(b))$ are universal approximators.
- So, we can safely assume that $f_\&$ is Archimedean:

$$c = f_\&(a,b) \Leftrightarrow f(c) = f(a) + f(b).$$

- Thus, invariance means that $f(c) = f(a)+f(b)$ implies $f(\lambda \cdot c) = f(\lambda \cdot a) + f(\lambda \cdot b)$.
- So, for every λ, the transformation $T : f(a) \to f(\lambda \cdot a)$ is additive: $T(A+B) = T(A) + T(B)$.
- Known: every monotonic additive function is linear.
- Thus, $f(\lambda \cdot a) = c(\lambda) \cdot f(a)$ for all a and λ.
- For monotonic $f(a)$, this implies $f(a) = C \cdot a^{-\alpha}$.
- So, $f(c) = f(a)+f(b)$ implies $C \cdot c^{-\alpha} = C \cdot a^{-\alpha} + C \cdot b^{-\alpha}$, and $c = f_\&(a,b) = (a^{-\alpha} + b^{-\alpha})^{-1/\alpha}$.

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

10. Deriving Student Distribution

- We want to maximize the degree

  $$f_\&(\mu(x_1), \mu(x_2), \ldots) = ((\mu(x_1))^{-\alpha} + (\mu(x_2))^{-\alpha} + \ldots)^{-1/\alpha}.$$

- The function $f(a)$ is decreasing.
- So, maximizing $f_\&(\mu(x_1), \ldots)$ is equivalent to minimizing the sum $(\mu(x_1))^{-\alpha} + (\mu(x_2))^{-\alpha} + \ldots$
- In the limit, this sum tends to $I \stackrel{\text{def}}{=} \int (\mu(x))^{-\alpha}\, dx$.
- So, we minimize I under constrains $\int x \cdot \rho(x)\, dx = 0$ and $\int x^2 \cdot \rho(x)\, dx = \sigma^2$, where $\rho(x) = \dfrac{\mu(x)}{\int \mu(y)\, dy}$.
- Thus, we minimize $\int (\mu(x))^{-\alpha}\, dx$ under constraints

  $$\int x \cdot \mu(x)\, dx = 0 \text{ and } \int x^2 \cdot \mu(x)\, dx - \sigma^2 \cdot \int \mu(x)\, dx = 0.$$

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

11. Deriving Student Distribution (cont-d)

- Lagrange multiplier method leads to minimizing

$$\int (\mu(x))^{-\alpha}\,dx + \lambda_1 \cdot \int x \cdot \mu(x)\,dx +$$

$$\lambda_2 \cdot \left(\int x^2 \cdot \mu(x)\,dx - \sigma^2 \cdot \int \mu(x)\,dx \right) \to \min .$$

- Equating the derivative w.r.t. $\mu(x)$ to 0, we get:

$$-\alpha \cdot (\mu(x))^{-\alpha-1} + \lambda_1 \cdot x + \lambda_2 \cdot x^2 - \lambda_2 \cdot \sigma^2 = 0.$$

- Thus, $\mu(x) = (a_0 + a_1 \cdot x + a_2 \cdot x^2)^{-\nu}$.
- For $\rho(x) = c \cdot \mu(x)$, we get $\rho(x) = c \cdot (a_0 + a_1 \cdot x + a_2 \cdot x^2)^{-\nu}$.
- So, $\rho(x) = c \cdot (a_2 \cdot (x - x_0)^2 + c_1)^{-\nu}$.
- This $\rho(x)$ is symmetric w.r.t. x_0, so, the mean is x_0.
- We know that the mean is 0, so $x_0 = 0$, and $\rho(x) = \text{const} \cdot (1 + a_2 \cdot x_2)^{-\nu}$: exactly Student's $\rho_S(x)$!

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments

12. Acknowledgments

- This work was performed when Vladik was a visiting researcher with the Geodetic Institute.
- This visit to the Leibniz University of Hannover was supported by the German Science Foundation.
- This work was also supported in part by NSF grant HRD-1242122.

Traditional . . .
Need for Heavy-Tailed . . .
What We Do
Need for Normalization
How to Combine Degrees
Deriving Student . . .
Acknowledgments