

Need for Most Accurate Discrete Approximations Explains Effectiveness of Statistical Methods Based on Heavy-Tailed Distributions

Songsak Sriboonchitta¹, Vladik Kreinovich²

Olga Kosheleva², and Hung T. Nguyen^{1,3}

¹Faculty of Economics, Chiang Mai University, Thailand
songsakecon@gmail.com

²University of Texas at El Paso, USA
olgak@utep.edu, vladik@utep.edu

³Department of Mathematical Sciences, New Mexico State University
Las Cruces, New Mexico 88003, USA, hunguyen@nmsu.edu

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

⏪

⏩

◀

▶

Page 1 of 18

Go Back

Full Screen

Close

Quit

1. Many Effective Statistical Methods Are Based on Light-Tailed Probability Distributions

- In many practical situations, we encounter Gaussian distributions.
- Their ubiquity can be explained by the Central Limit Theorem.
- Gaussian-based methods are also effective when we have a *partial* information about the probabilities.
- The effectiveness of Gaussian methods in such situations comes from the Maximum Entropy approach:
 - if we have only partial information about the probability distribution,
 - i.e., if several different probability density functions $\rho(x)$ are consistent with our knowledge,
 - then it is reasonable to select the distributions with the largest uncertainty $S = - \int \rho(x) \cdot \ln(\rho(x)) dx$.

2. Many Effective Statistical Methods Are Based on Light-Tailed Distributions (cont-d)

- Often, the only information that we have about the distribution is its first two moments:

$$\int \rho(x) dx = 1, \quad \int x \cdot \rho(x) dx = E, \quad \int x^2 \cdot \rho(x) dx = M_2.$$

- For maximizing S under these constraints, the Lagrange multiplier method leads to Gaussian distr.
- For the Gaussian distribution, the probability density has *light tails* in the sense that:
 - as x increases, the probability density tends to 0 fast,
 - so fast that for every k , we have a finite moment $\int x^k \cdot \rho(x) dx$.
- Similarly, many other effective statistical methods are based on light-tailed distributions.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 3 of 18

Go Back

Full Screen

Close

Quit

3. Heavy-Tailed Distributions

- On the other hand, in many empirical situations,
 - it turns out to be efficient to use statistical methods based on *heavy-tailed* distributions,
 - i.e., distributions for which the probability density $\rho(x)$ tends to 0 much slower,
 - so that some moments become infinite.
- A typical case is when asymptotically, $\rho(x) \sim C \cdot x^{-\alpha}$ for some $\alpha > 0$.
- How can we explain the effectiveness of statistical methods based on the heavy-tailed distributions?
- In this paper, we show that:
 - the need for most accurate discrete approximations
 - naturally leads to statistical methods based on heavy-tailed distributions.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 4 of 18

Go Back

Full Screen

Close

Quit

4. Need for Most Accurate Discrete Approximations: The Main Idea

- Many real-life situations – e.g., in econometrics – are well-described by *random walk* models:
 - each component $x(t + 1)$ of the state at the next moment of time $t + 1$
 - is obtained from the state $x(t)$ at the previous moment of time t by adding a random step $r(t)$:

$$x(t + 1) = x(t) + r(t).$$

- Sometimes, the data is well-described by the simplest random walk, when $r(t)$ takes one of the two values.
- Sometimes, we need to use a more complex model, when $r(t)$ takes $n > 2$ possible values with diff. prob.
- When $n \rightarrow \infty$, we get a continuous distribution with pdf $\rho(x)$.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 5 of 18

Go Back

Full Screen

Close

Quit

5. Resulting Need for the Most Accurate Discrete Approximations

- We want to accurately approximate the continuous distribution with an n -point one.
- Since random walk models are widely used, it makes sense to use:
 - the existence of the most accurate discrete approximations
 - as an alternative criterion for selecting a probability distribution
 - in situations when we only have partial information about the probabilities,
 - i.e., in which several different probability distributions are consistent with our knowledge.
- This is our main idea.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 6 of 18

Go Back

Full Screen

Close

Quit

6. Towards Formalizing the Main Idea

- When we approximate a continuous probability distribution by a discrete one, we thus:
 - approximate the actual random variable r which can take, in principle, any value from the real line,
 - by a discrete variable that can only take a finite number of values r_1, \dots, r_n .
- The approximating value r_i is, in general, different from the actual value r : $\delta \stackrel{\text{def}}{=} |r_i - r| \neq 0$.
- A rational decision maker maximizes the expected value of an objective function $u(x)$ called *utility*.
- This function is determined modulo a general linear transformation $u \rightarrow a \cdot u + b$ for some $a > 0$ and b .
- Maximizing utility \Leftrightarrow minimizing *disutility* $U \stackrel{\text{def}}{=} -u$.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 7 of 18

Go Back

Full Screen

Close

Quit

7. Towards Formalizing the Main Idea (cont-d)

- Disutility $U = -u$ is also defined modulo a general linear transformation $U \rightarrow a \cdot U + b$.
- We thus need to describe the disutility $U(\delta)$ caused by inaccuracy δ .
- In most case, the numerical value r of the corresponding quantity depends on the choice of a measuring unit:
 - For a geometric random walk, the value r is different for meters or feet.
 - For a financial random walk, the numerical values are different for dollars or Euros.
- If we choose a k times smaller measuring unit, the numerical values changes: $r \rightarrow k \cdot r$.
- The choice of a measuring unit is usually rather arbitrary.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 8 of 18

Go Back

Full Screen

Close

Quit

8. Towards Formalizing the Main Idea (cont-d)

- The choice of a measuring unit is usually rather arbitrary.
- It is therefore reasonable to require that $U(\delta)$ not depend on the choice of a measuring unit:
 - for every $k > 0$,
 - there exist a value $a(k)$ for which, for every $\delta \geq 0$, we have $U(k \cdot \delta) = a(k) \cdot U(\delta)$.
- For a continuous function, this functional equation implies that $U(\delta) = C \cdot \delta^\beta$ for some $C > 0$ and β .
- The disutility increases with inaccuracy.
- So, we should assign, to each value r , the value r_i which is the closest to r .

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 9 of 18

Go Back

Full Screen

Close

Quit

9. Towards Formalizing the Problem

- For large n :
 - we can describe the selection of the points $r_1 < r_2 < \dots < r_n$
 - by describing the frequency $\rho_0(r)$, i.e., the number of points per unit length.
- The overall number of points is n : $\int \rho_0(r) dr = n$.
- The length of each interval $[r_i, r_{i+1}]$ is approximately equal to $\frac{1}{\rho_0(r)}$.
- The average value of inaccuracy is proportional to the interval width.
- Thus, the average value of the disutility is proportional to the β -th power of this width, i.e., to $U_i \approx \frac{1}{(\rho_0(r_i))^\beta}$.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 10 of 18

Go Back

Full Screen

Close

Quit

10. Towards Formalizing the Problem (cont-d)

- The overall expected value E_U of the disutility is equal to $E_U = \sum_{i=1}^n p_i \cdot U_i$.
- Here, p_i is the probability that the original random variable r occurs in the interval I_i .
- So, $p_i = \int_{I_i} \rho(r) dr$, where $\rho(r)$ is the probability density of the original random variable; thus:

$$E_U = \sum_{i=1}^n \frac{1}{(\rho_0(r_i))^\beta} \cdot \int_{I_i} \rho(r) dr = \sum_{i=1}^n \int_{I_i} \frac{1}{(\rho_0(r_i))^\beta} \cdot \rho(r) dr.$$

- For large n and narrow intervals, $r_i \approx r$, so:

$$E_U \approx \sum_{i=1}^n \int_{I_i} \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr = \int \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr.$$

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 11 of 18

Go Back

Full Screen

Close

Quit

11. Resulting Formalization of the Problem

- So, we arrive at the following precise reformulation of the best discrete optimization problem:
 - we know the probability density $\rho(r)$;
 - we want to find the density $\rho_0(r)$ of the distribution of the discrete points as the one that:
 - minimizes the integral $\int \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr$
 - under the constraint $\int \rho_0(r) dr = n$.
- After that, we will need to select:
 - in each class of probability distributions,
 - a distribution for which this best-case expected disutility has the smallest possible value.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 12 of 18

Go Back

Full Screen

Close

Quit

12. Solving the Resulting Optimization Problem

- For minimizing $\int \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr$ under the constraint $\int \rho_0(r) dr = n$, Lagrange multiplier method implies

$$\int \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr + \lambda \cdot \left(\int \rho_0(r) dr - n \right) \rightarrow \min .$$

- Differentiating with respect to $\rho_0(r)$ and equating the derivative to 0, we get $-\beta \cdot \frac{\rho(r)}{(\rho_0(r))^{\beta+1}} + \lambda = 0$.
- So, $\rho_0(r) = c \cdot (\rho(r))^{1/(\beta+1)}$.
- The constraint $\int \rho_0(r) dr = n$ implies

$$\rho_0(r) = n \cdot \frac{(\rho(r))^{1/(\beta+1)}}{\int (\rho(s))^{1/(\beta+1)} ds} .$$

- We know that $E_U = \int \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr$.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 13 of 18

Go Back

Full Screen

Close

Quit

13. Solving the Resulting Optimization Problem (cont-d)

- We know that $E_U = \int \frac{1}{(\rho_0(r))^\beta} \cdot \rho(r) dr$.
- Substituting the above expression for $\rho_0(r)$ into the formula for E_U , we get

$$E_U = \int \frac{\rho(r)}{(\rho_0(r))^{1/(\beta+1)}} dr = \frac{\int (\rho(r))^{1/(\beta+1)} dr}{\left(\int (\rho(r))^{1/(\beta+1)} dr\right)^\beta} = \frac{1}{\left(\int (\rho(r))^{1/(\beta+1)} dr\right)^{\beta-1}}.$$

- Thus, depending on whether $\beta < 1$ or $\beta > 1$, minimizing E_U is equivalent to minimizing or maximizing

$$J \stackrel{\text{def}}{=} \int (\rho(r))^{1/(\beta+1)} dr.$$

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 14 of 18

Go Back

Full Screen

Close

Quit

14. Solving the Resulting Optimization Problem: Final Step

- We have a class of probability density functions which are consistent with our knowledge.
- In this class, we need to find the ones for which the above expression J is optimal.
- Let's consider the case when our knowledge consists of a moment-related constraint $\int |r|^k \cdot \rho(r) dr = M_k$.
- Then, the Lagrange multiplier method leads to optimizing the following expression:

$$\int (\rho(r))^{1/(\beta+1)} dr + \lambda_0 \cdot \left(\int \rho(r) dr - 1 \right) + \lambda_k \cdot \left(\int |r|^k \cdot \rho(r) dr - M_k \right).$$

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 15 of 18

Go Back

Full Screen

Close

Quit

15. Solving the Resulting Optimization Problem: Final Step (cont-d)

- Differentiating with respect to unknowns $\rho(r)$ and equating the derivative to 0, we get

$$\frac{1}{\beta + 1} \cdot (\rho(r))^{-\beta/(\beta+1)} = -\lambda_0 + \lambda_k \cdot |r|^k.$$

- So, $\rho(r) = \frac{1}{(C_0 + C_1 \cdot |r|^k)^{1+1/\beta}}$, for const. C_0, C_1 .
- Asymptotically, when $|r|$ increases, we have $\rho(r) \sim |r|^{-\alpha}$ for $\alpha = k \cdot \left(1 + \frac{1}{\beta}\right)$.
- Thus, *we have an explanation for the effectiveness of statistical methods based on heavy-tailed distributions.*

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 16 of 18

Go Back

Full Screen

Close

Quit

16. Possible Future Work

- We concentrated on the use of heavy-tail-based statistical methods in econometric (and similar) applications.
- However, such methods are also effectively used in many other application areas.
- For example, they are used in the analysis of graphs and networks.
- It would be interesting to analyze:
 - to what extent similar ideas
 - can explain the effectiveness of these methods in network studies and in other application areas.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

Home Page

Title Page



Page 17 of 18

Go Back

Full Screen

Close

Quit

17. Acknowledgments

- We acknowledge the support of the Center of Excellence in Econometrics, Chiang Mai Univ., Thailand.
- This work was also supported in part:
 - by the National Science Foundation grants
 - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
 - DUE-0926721, and
 - by an award from Prudential Foundation.

Many Effective ...

Heavy-Tailed ...

Need for Most ...

Resulting Need for the ...

Towards Formalizing ...

Towards Formalizing ...

Resulting ...

Solving the Resulting ...

Possible Future Work

[Home Page](#)

[Title Page](#)



Page 18 of 18

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)