Why Student Distributions? Why Matern's Covariance Model? A Symmetry-Based Explanation

Stephen Schön, Gael Kermarrec, Boris Kargoll, Ingo Neumann, Olga Kosheleva, and Vladik Kreinovich

Abstract In this paper, we show that empirical successes of Student distribution and of Matern's covariance models can be indirectly explained by a natural requirement of scale invariance – that fundamental laws should not depend on the choice of physical units. Namely, while neither the Student distributions nor Matern's covariance models are themselves scale-invariant, they are the only one which can be obtained by applying a scale-invariant combination function to scale-invariant functions.

1 Formulation of the Problem

Scale-invariance: a natural property of the physical world. Scientific laws are described in terms of numerical values of the corresponding quantities, be it physical quantities such as distance, mass, or velocity, or economic quantities such as price or cost.

These numerical values, however, depend on the choice of a measuring unit. If we replace the original unit by a new unit which is λ times smaller, then all the numerical values of the corresponding quantity get multiplied by λ. For example, if instead of meters, we start using centimeters – a 100 smaller unit – to describe distance, then all the distances get multiplied by 100, so that, e.g., 2 m becomes $2 \cdot 100 = 200$ cm.

It is reasonable to require that the fundamental laws describing objects from the physical world – be it material objects or human beings – do not change if we simply

Stephen Schön, Gael Kermarrec, Boris Kargoll, and Ingo Neumann
Leibniz Universität Hannover, Nienburger Strasse 1,
30167 Hannover, Germany, e-mail: schoen@ife.uni-hannover.de, gael.kermarrec@web.de, kargoll@gih.uni-hannover.de, neumann@gih.uni-hannover.de

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, 500 W. University,
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu, vladik@utep.edu

change the measuring unit. In other words, it is reasonable to require that the laws be invariant with respect to *scaling* $x \to \lambda \cdot x$.

Of course, if we change a measuring unit for one quantity, then we may need to also correspondingly change the measuring unit for related quantities as well. For example, in a simple motion, the distance d is equal to the product $v \cdot t$ of velocity v and time t. If we simply change the unit for time without changing the units for distance or velocity, the formula stops being true. However, the formula remains true if we accordingly change the unit for velocity. For example, if we started with seconds and m/sec, then, once we change seconds to hours, we should also change the measuring unit for velocity from m/sec to m/hr.

Thus, scale-invariance means that if we arbitrarily change the units of one or more fundamental quantities, then, after an appropriate re-scaling of related units, we should get, in the new units, the exact same formula as in the old units.

Heavy-tailed distributions: a situation in which we expect scale-invariance. Measurements are rarely absolutely accurate. Usually, the measurement result $\widetilde{x}$ is somewhat different from the actual (unknown) value x of the corresponding quantity. In many cases, we know the upper bound of the measurement error, so that the probability of exceeding this bound is either equal to 0 or very small (practically equal to 0).

In many other practical situations, however, the probability of having reasonably large measurement errors $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ is positive – and does not become negligibly small. In such cases, we talk about *heavy-tailed* distributions.

Such distributions are ubiquitous in physics, in economics, etc., and they have the same shape in different application areas; see, e.g., [8, 10]. This ubiquity seems to indicate that there is a fundamental reason for such distributions. It therefore seems reasonable to expect that for this fundamental law – just like for all other fundamental laws – we have the scale-invariance property. In other words, it is reasonable to expect that, for the corresponding probability density function $\rho(x)$, for every $\lambda > 0$, there exists a value $\mu(\lambda)$ for which

$$\rho(\lambda \cdot x) = \mu(\lambda) \cdot \rho(x). \tag{1}$$

Alas, no scale-invariant pdf is possible. At first glance, the above scale-invariance criterion sounds reasonable, but, alas, it is never satisfied.

Indeed, the pdf should have the property that the overall probability to be somewhere should be equal to 1 (i.e., $\int \rho(x)\,dx = 1$), and be measurable, and it is known (see, e.g., [1, 2] that every measurable solution of the equation (1) has the power law form

$$\rho(x) = c \cdot x^\alpha$$

for some c and α. For this function, the integral over the real line is always infinite:

- for $\alpha \geq -1$, it is infinite in the vicinity if 0, while
- for $\alpha \leq -1$, it is infinite for $x \to \infty$.

A simple explanation of why power laws are the only scale-invariant ones. If we additionally assume that the function $\rho(x)$ is differentiable, then the fact that power laws are the only solutions can be easily derived.

Indeed, in this case, the function $\mu(\lambda) = \dfrac{\rho(\lambda \cdot x)}{\rho(x)}$ is also differentiable, as a ratio of two differentiable functions $\rho(\lambda \cdot x)$ and $\rho(x)$. Since both functions $\rho(x)$ and $\mu(\lambda)$ are differentiable, we can differentiate both sides of the equation (1) by λ and take $\lambda = 1$; we then conclude that $x \cdot \dfrac{d\rho}{dx} = \alpha \cdot \rho$, where we denoted $\alpha \stackrel{\text{def}}{=} \dfrac{d\mu}{d\lambda}_{|\lambda=1}$. By moving all the terms containing ρ to the left-hand side and the terms containing x to the right-hand side, we conclude that $\dfrac{d\rho}{\rho} = \alpha \cdot \dfrac{dx}{x}$. Integrating both sides, we get $\ln(\rho) = \alpha \cdot \ln(x) + C$, hence for $\rho = \exp(\ln(\rho))$, we get $\rho(x) = c \cdot x^{\alpha}$, where we denoted $c \stackrel{\text{def}}{=} \exp(C)$.

What is usually done. A usual idea is to abandon scale-invariance completely. For example, one of the most empirically successful ways to describe heavy-tailed distributions is to use non-scale-invariant *Student distributions*, with the probability density $\rho(x) = \text{const} \cdot (1 + a \cdot x^2)^{-\nu}$ for some coefficients const, a, and ν (see, e.g., [7]).

What we show in this paper. In this paper, we "rehabilitate" scale-invariance: namely, we show that while the distribution cannot be "directly" scale-invariant, it can be "indirectly" scale-invariant – namely, it can be described as a scale-invariant combination of two scale-invariant functions.

Interestingly, under a few reasonable additional conditions, we get exactly the empirically successful Student distributions – and thus, indirect scale-invariance explains their empirical success.

This line of reasoning also provides us with a reasonable next approximation (that is worth trying if we want a more accurate description): namely, a scale-invariant combination of three or more scale-invariant functions.

Multi-D case. A similar situation occurs in the multi-D case, e.g., in the analysis of spatial data. Often, spatial data is described as a homogeneous and isotropic process. To describe such processes, it is convenient to use Fourier transforms: namely, to describe, for each frequency ω, the mean value $S(\omega)$ of the square of the absolute value of the ω-Fourier component of the original multi-D data. The value $S(\omega)$ is known as the *spectral density*.

In some cases, this function $S(\omega)$ is mainly concentrated at some frequencies. However, in many other practical situations, the corresponding values do not become negligible neither for small nor for large ω. In many such cases, the shape of the spectral density is approximately the same, so it looks like we have a fundamental law of spatial dependence.

Since it is a fundamental law, it is reasonable to expect it to be scale-invariant, i.e., satisfy the condition $S(\lambda \cdot \omega) = \mu(\lambda) \cdot S(\omega)$.

We already know that every measurable solution to this functional equation has the form $S(\omega) = \text{const} \cdot \omega^{\alpha}$ for some const and α. However, for such functions, we have $\int S(\omega)\,d\omega = +\infty$, while the integral is equal to the overall energy of the spatial signal and should, therefore, be finite.

Similar to the 1-D case, a usual solution is to abandon scale-invariance and to use some non-scale-invariant function for which $\int S(\omega)\,d\omega < +\infty$. It turns out that among all such functions, Matern's function $S(\omega) = \text{const}\cdot(a_0 + a_1\cdot\omega^2)^{-\nu}$ (for some const, a_i, and ν) is, empirically, the best; see, e.g., [3].

In this paper, we show that while this function is not directly scale-invariant, it is indirectly scale-invariant – as a result of applying a scale-invariant combination function to two scale-invariant functions $S(\omega)$. Moreover, it turns out that, under reasonable assumptions, Matern's functions are the only such combinations. Thus, scale invariance explains their empirical success.

We also provide a natural next approximation to Matern's function – a scale-invariant combination of three or more scale-invariant functions.

2 Let Us Describe Scale-Invariant Combination Functions

A combination function: reasonable requirements. By a combination function we would like to mean an operation $a*b$ that transforms two non-negative numbers into a new non-negative number. Intuitively, a combination of a and b should be the same as a combination of b and a, so the operation $*$ should be commutative: $a*b = b*a$.

Similarly, a combination of a, b, and c should not depend on the order in which we combine them, so this operation must be associative: $(a*b)*c = a*(b*c)$. It is also reasonable to require that this operation is continuous (if $a_n \to a$ and $b_n \to b$, then we should have $a_n * b_n \to a*b$) and monotonic (non-decreasing in each of its variables). So, we arrive at the following definition.

Definition 1. *By a* combination function $*$ *we mean a commutative associative continuous non-decreasing function from pairs of non-negative real numbers to non-negative real numbers.*

Scale-invariance. Scale-invariance means that if we have $a*b = c$, then after re-scaling all three values a, b, and c, we conclude that $(\lambda\cdot a)*(\lambda\cdot b) = \lambda\cdot c$. Substituting $c = a*b$ into this formula, we get the following definition.

Definition 2. *We say that a combination function is* scale-invariant *if for all a, b, and λ, we have $(\lambda\cdot a)*(\lambda\cdot b) = \lambda\cdot(a*b)$.*

Proposition. *The only scale-invariant combination functions are $a*b = \min(a,b)$, $a*b = \max(a,b)$, and $a*b = (a^\beta + b^\beta)^{1/\beta}$ for some β.*

For reader's convenience, the proof of this result is given in the Appendix.

3 Resulting Derivation of Student Distribution and Matern's Covariance Model

Derivation of Student distribution. If we use a scale-invariant combination operation to combine two scale-invariant functions $c_i \cdot x^{\alpha_i}$, we get the expressions $\min(c_1 \cdot x^{\alpha_1}, c_2 \cdot x^{\alpha_2})$, $\max(c_1 \cdot x^{\alpha_1}, c_2 \cdot x^{\alpha_2})$, and

$$((c_1 \cdot x^{\alpha_1})^{\beta} + (c_2 \cdot x^{\alpha_2})^{\beta})^{1/\beta} = (C_1 \cdot x^{\gamma_1} + C_2 \cdot x^{\gamma_2})^{\gamma},$$

where $C_i = (c_i)^{\beta}$, $\gamma_i = \beta \cdot \alpha_i$, and $\gamma = 1/\beta$.

It is reasonable to require:

- that the pdf if *analytical* in x – i.e., can be expanded in Taylor series – and
- that it is *monotonically decreasing with* x – since it is reasonable to require that the larger the measurement error, the less probable it is.

Analyticity excludes min and max.

For the sum, if both γ_i are different from 0, the value at 0 is either 0 or infinity. It cannot be infinite – then $\rho(x)$ would be not analytical, and it cannot be 0 – then it will not be able to monotonically decrease to 0. Thus, one of the coefficients γ_i is equal to 0, and we have

$$\rho(x) = C \cdot (1 + c \cdot x^{\gamma_2})^{\gamma}.$$

This expression is analytical when γ_2 is a positive integer. We cannot have $\gamma_2 = 1$, because then we would get $\rho(x) \to +\infty$ either when $x \to +\infty$ or when $x \to -\infty$. Thus, we must have $\gamma_2 \geq 2$.

Out of all possible functions of this type, the *generic* case – when both the 0-th and the second coefficient at Taylor expansion are not 0 – is when $\gamma_2 = 2$. Thus, we get exactly the Student distribution.

Derivation of Matern's covariant model. For dependence of the spectral density on ω, we similarly get exactly Matern's covariance model.

What next? If the scale-invariant combination of *two* scale-invariant functions does not work well, we can try a scale-invariant combination of three or more such functions: $f(x) = \left(\sum\limits_{i=1}^{k} C_i \cdot x^{\gamma_i}\right)^{\gamma}$.

4 Alternative Symmetry-Based Explanation

How to explain normal distributions: reminder. Many practical applications assume that the distribution is Gaussian (normal). One way to derive the Gaussian distribution is to consider, among all distributions with mean 0 and known standard deviation σ, the distribution with the largest entropy $\mathscr{S}(\rho) \stackrel{\text{def}}{=} -\int \rho(x) \ln(\rho(x))\, dx$ (see, e.g., [4]), i.e., to optimize entropy under the constraints

$$\int \rho(x)\,dx = 1, \ \int x\cdot\rho(x)\,dx = 0, \text{ and } \int x^2\cdot\rho(x)\,dx = \sigma^2. \tag{3}$$

For this constraint optimization problem, the Lagrange multiplier method reduces it to the following unconditional optimization problem

$$-\int \rho(x)\cdot\ln(\rho(x))\,dx + \lambda_0\cdot\left(\int \rho(x)\,dx - 1\right) + \lambda_1\cdot\left(\int x\cdot\rho(x)\,dx\right) +$$

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

Differentiating the objective function with respect to $\rho(x)$ and equating the derivative to 0, we conclude that

$$-\ln(\rho(x)) - 1 + \lambda_0 + \lambda_1\cdot x + \lambda_2\cdot x^2 = 0,$$

hence $\rho(x) = \exp((\lambda_0 - 1) + \lambda_1\cdot x + \lambda_2\cdot x^2)$. The requirement that the mean is 0 implies that $\lambda_1 = 0$, so we get the usual Gaussian distribution.

Entropy is scale-invariant. Entropy is scale-invariant in the sense that:

- if we have two distributions $\rho(x)$ and $\rho'(x)$ for which $\mathscr{S}(\rho) = \mathscr{S}(\rho')$, and
- we re-scale x and thus, transform the original distributions into the re-scaled ones $\rho_\lambda(x)$ and $\rho'_\lambda(x)$, then these re-scaled distributions will also have the same entropy $\mathscr{S}(\rho_\lambda) = \mathscr{S}(\rho'_\lambda)$.

Scale-invariant generalizations of entropy. It turns out that entropy is not the only functional with the above scale-invariance properties. All such scale-invariant functions have been described [5, 6]. In addition to entropy, we can also have $\int \ln(\rho(x))\,dx$ and $\int (\rho(x))^q\,dx$ for some q.

For scale-invariant generalizations of entropy, we get Student distribution. Optimizing $\int \ln(\rho(x))\,dx$ under constraints (3) leads to

$$\int \ln(\rho(x))\,dx + \lambda_0\cdot\left(\int \rho(x)\,dx - 1\right) + \lambda_1\cdot\left(\int x\cdot\rho(x)\,dx\right) +$$

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

Differentiating the objective function with respect to $\rho(x)$ and equating the derivative to 0, we conclude that $\dfrac{1}{\rho(x)} - 1 + \lambda_0 + \lambda_1\cdot x + \lambda_2\cdot x^2 = 0$, hence

$$\rho(x) = \frac{1}{(1-\lambda_0) - \lambda_1\cdot x - \lambda_2\cdot x^2}.$$

The requirement that the mean is 0 implies that $\lambda_1 = 0$, so we indeed get a particular case of the Student distribution.

Similarly, optimizing $\int(\rho(x))^q\,dx$ under constraints (3) leads to

$$\int(\rho(x))^q\,dx+\lambda_0\cdot\left(\int\rho(x)\,dx-1\right)+\lambda_1\cdot\left(\int x\cdot\rho(x)\,dx\right)+$$

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

Differentiating the objective function with respect to $\rho(x)$ and equating the derivative to 0, we conclude that $q\cdot(\rho(x))^{q-1}+\lambda_0+\lambda_1\cdot x+\lambda_2\cdot x^2=0$, hence $\rho(x)=(a_0+a_1\cdot x+a_2\cdot x^2)^{1/(q-1)}$. The requirement that the mean is 0 implies that $a_1=0$, so we indeed get (the generic case of) the Student distribution.

Acknowledgments

This work was performed when Olga Kosheleva and Vladik Kreinovich were visiting researchers with the Geodetic Institute of the Leibniz University of Hannover, a visit supported by the German Science Foundation. This work was also supported in part by NSF grant HRD-1242122.

References

1. J. Aczél, *Lectures on Functional Equations and Their Applications*, Dover, New York, 2006.
2. J. Aczél and H. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, Cambridge, UK, 1989.
3. N. Cressie, *Statistics for Spatial Data*, Wiley, New York, 2015.
4. E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.
5. O. Kosheleva, "Symmetry-group justification of maximum entropy method and generalized maximum entropy methods in image processing", In: G. J. Erickson, J. T. Rychert, and C. R. Smith (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1998, pp. 101–113.
6. V. Kreinovich, O. Kosheleva, H. T. Nguyen, and S. Sriboonchitta, "Why some families of probability distributions are practically efficient: a symmetry-based explanation", In: Van Nam Huynh, V. Kreinovich, and S. Sriboonchitta (eds.), *Causal Inference in Econometrics*, Springer Verlag, Cham, Switzerland, 2016, pp. 133–152.
7. K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t-distribution", *Journal of the American Statistical Association*, 1989, Vol. 84, No. 408, pp. 881–896.
8. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, California, 1983.
9. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.
10. S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Varlag, New York, 2007.
11. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

5 Proof

1°. Depending on whether the value $1*1$ is equal to 1 or not, we have two possible cases: $1*1=1$ and when $1*1\neq 1$. Let us consider these two cases one by one.

2°. Let us first consider the case when $1*1=1$. In this case, the value $0*1$ can be either equal to 0 or different from 0. Let us consider both subcases.

2.1°. Let us first consider the first subcase, when $0*1=0$.

In this case, for every $b>0$, scale invariance with $\lambda=b$ implies that

$$(b\cdot 0)*(b\cdot 1)=(b\cdot 0),$$

i.e., that $0*b=0$. By taking $b\to 0$ and using continuity, we also get $0*0=0$. Thus, $0*b=0$ for all b.

By commutativity, we have $a*0=0$ for all a. So, to fully describe the operation $a*b$, it is sufficient to consider the cases when $a>0$ and $b>0$.

$2.1.1^\circ$. Let us prove, by contradiction, that in this subcase, we have $1*a\le 1$ for all a.

Indeed, let us assume that for some a, we have $b\stackrel{\text{def}}{=}1*a>1$. Then, due to associativity and $1*1=1$, we have $1*b=1*(1*a)=(1*1)*a=1*a=b$.

Due to scale-invariance with $\lambda=b$, the equality $1*b=b$ implies that $b*b^2=b^2$. Thus, $1*b^2=1*(b*b^2)=(1*b)*b^2=b*b^2=b^2$.

Similarly, from $1*b^2=b^2$, we conclude that for $b^4=(b^2)^2$, we have $1*b^4=b^4$, and, in general, that $1*b^{2^n}=b^{2^n}$ for every n.

Scale invariance with $\lambda=b^{-2^n}$ implies that $b^{-2^n}*1=1$. In the limit $n\to\infty$, we get $0*1=1$, which contradicts to our assumption that $0*1=0$. This contradiction shows that indeed, $1*a\le 1$.

$2.1.2^\circ$. For $a\ge 1$, monotonicity implies $1=1*1\le 1*a$, so $1*a\le 1$ implies that $1*a=1$.

Now, for any a' and b' for which $0<a'\le b'$, if we denote $r\stackrel{\text{def}}{=}\dfrac{b'}{a'}\ge 1$, then scale-invariance with $\lambda=a'$ implies that $a'\cdot(1*r)=(a'\cdot 1)*(a'\cdot r)=a'*b'$. Here, $1*r=1$, thus $a'*b'=a'\cdot 1=a'$, i.e., $a'*b'=\min(a',b')$. Due to commutativity, the same formula also holds when $a'\ge b'$. So, in this case, $a*b=\min(a,b)$ for all a and b.

2.2°. Let us now consider the second subcase of the first case, when $0*1>0$.

$2.2.1^\circ$. Let us first show that in this subcase, we have $0*0=0$.

Indeed, scale-invariance with $\lambda=2$ implies that from $0*0=a$, we can conclude that

$$(2\cdot 0)*(2\cdot 0)=0*0=2\cdot a.$$

Thus $a=2\cdot a$, hence $a=0$. The statement is proven.

$2.2.2^\circ$. Let us now prove that in this subcase, $0*1=1$.

Indeed, in this case, for $a\stackrel{\text{def}}{=}0*1$, we have, due to $0*0=0$ and associativity, that

$$0 * a = 0 * (0 * 1) = (0 * 0) * 1 = 0 * 1 = a.$$

Here, $a > 0$, so by applying scale invariance with $\lambda = a^{-1}$, we conclude that

$$0 * 1 = 1.$$

2.2.3°. Let us now prove that for every $a \le b$, we have $a * b = b$. So, due to commutativity, we have $a * b = \max(a,b)$ for all a and b.

Indeed, from $1 * 1 = 1$ and $0 * 1 = 1$, due to scale invariance with $\lambda = b$, we conclude that $0 * b = b$ and $1 * b = b$. Due to monotonicity, $0 \le a \le b$ implies that $b = 0 * b \le a * b \le b * b = b$, thus $a * b = b$. The statement is proven.

3°. Let us now consider the remaining case when $1 * 1 \ne 1$.

3.1°. Let us denote $v(k) \stackrel{\text{def}}{=} 1 * \ldots * 1$ (k times). Then, for every m and n, the value $v(m \cdot n) = 1 * \ldots * 1$ ($m \cdot n$ times) can be represented as

$$(1 * \ldots * 1) * \ldots * (1 * \ldots * 1),$$

where we divide the 1s into m groups with n 1s in each. For each group, we have $1 * \ldots * 1 = v(n)$. Thus, $v(m \cdot n) = v(n) * \ldots * v(n)$ (m times).

We know that $1 * \ldots * 1$ (m times) $= v(m)$. Thus, by using scale-invariance with $\lambda = v(n)$, we conclude that $v(m \cdot n) = v(m) \cdot v(n)$, i.e., that that function $v(n)$ is multiplicative. In particular, this means that for every number p and for every positive integer n, we have $v(p^n) = (v(p))^n$.

3.2°. If $v(2) = 1 * 1 > 1$, then by monotonicity, we get $v(3) = 1 * v(2) \ge 1 * 1 = v(2)$, and, in general, $v(n+1) \ge v(n)$. Thus, in this case, the sequence $v(n)$ is (non-strictly) increasing.

Similarly, if $v(2) = 1 * 1 < 1$, then we get $v(3) \le v(2)$ and, in general, $v(n+1) \le v(n)$, i.e., in this case, the sequence $v(n)$ is strictly decreasing.

Let us consider these two cases one by one.

3.2.1°. Let us first consider the case when the sequence $v(n)$ is increasing. In this case, for every three integers m, n, and p, if $2^m \le p^n$, then $v(2^m) \le v(p^n)$, i.e., $(v(2))^m \le (v(p))^n$.

For all m, n, and p, the inequality $2^m \le p^n$ is equivalent to $m \cdot \ln(2) \le n \cdot \ln(p)$, i.e., to $\dfrac{m}{n} \le \dfrac{\ln(p)}{\ln(2)}$. Similarly, the inequality $(v(2))^m \ge (v(p))^n$ is equivalent to $\dfrac{m}{n} \le \dfrac{\ln(v(p))}{\ln(v(2))}$. Thus, the above conclusion

$$\text{if } 2^m \le p^n, \text{ then } (v(2))^m \le (v(p))^n$$

takes the following form:

$$\text{for every rational number } \frac{m}{n}, \text{ if } \frac{m}{n} \le \frac{\ln(p)}{\ln(2)} \text{ then } \frac{m}{n} \le \frac{\ln(v(p))}{\ln(v(2))}.$$

Similarly, for all m', n', and p, if $p^{n'} \le 2^{m'}$, then $v(p^{n'}) \le v(2^{m'})$, i.e., $(v(p))^{n'} \le (v(2))^{m'}$. The inequality $p^{n'} \le 2^{m'}$ is equivalent to $n' \cdot \ln(p) \le m' \cdot \ln(2)$, i.e., to $\dfrac{\ln(p)}{\ln(2)} \le \dfrac{m'}{n'}$. Also, the inequality $(v(p))^{n'} \le (v(2))^{m'}$ is equivalent to $\dfrac{\ln(v(p))}{\ln(v(2))} \le \dfrac{m'}{n'}$. Thus, the above conclusion

$$\text{if } p^{n'} \le 2^{m'}, \text{ then } (v(p))^{n'} \le (v(2))^{m'}$$

takes the following form:

$$\text{for every rational number } \frac{m'}{n'}, \text{ if } \frac{\ln(p)}{\ln(2)} \le \frac{m'}{n'} \text{ then } \frac{\ln(v(p))}{\ln(v(2))} \le \frac{m'}{n'}.$$

Let us denote $\alpha \stackrel{\text{def}}{=} \dfrac{\ln(v(2))}{\ln(2)}$ and $\beta \stackrel{\text{def}}{=} \dfrac{\ln(v(p))}{\ln(p)}$. For every $\varepsilon > 0$, there exist rational numbers $\dfrac{m}{n}$ and $\dfrac{m'}{n'}$ for which $\alpha - \varepsilon \le \dfrac{m}{n} \le \alpha \le \dfrac{m'}{n'} \le \alpha + \varepsilon$. For these numbers, the above two properties imply that $\dfrac{m}{n} \le \beta$ and $\beta \le \dfrac{m'}{n'}$ and thus, that $\alpha - \varepsilon \le \beta \le \alpha + \varepsilon$, i.e., that $|\alpha - \beta| \le \varepsilon$. This is true for all $\varepsilon > 0$, so we conclude that $\beta = \alpha$, i.e., that $\dfrac{\ln(v(p))}{\ln(v(2))} = \alpha$. Hence, $\ln(v(p)) = \alpha \cdot \ln(p)$ and thus, $v(p) = p^\alpha$ for all integers p.

3.2.2°. We can reach a similar conclusion $v(p) = p^\alpha$ when the sequence $v(n)$ is decreasing.

3.3°. By definition of $v(n)$, we have $v(m) * v(m') = v(m+m')$. Thus, we have

$$m^\alpha * (m')^\alpha = (m+m')^\alpha.$$

By using scale-invariance with $\lambda = n^{-\alpha}$, we get

$$\frac{m^\alpha}{n^\alpha} * \frac{(m')^\alpha}{n^\alpha} = \frac{(m+m')^\alpha}{n^\alpha}.$$

Thus, for $a = \dfrac{m^\alpha}{n^\alpha}$ and $b = \dfrac{(m')^\alpha}{n^\alpha}$, we get $a * b = (a^\beta + b^\beta)^{1/\beta}$, where $\beta \stackrel{\text{def}}{=} 1/\alpha$.

Rational numbers $r = \dfrac{m}{n}$ are everywhere dense on the real line, hence the values r^α are also everywhere dense, i.e., every real number can be approximated, with any given accuracy, by such numbers. Thus, continuity implies that $a * b = (a^\beta + b^\beta)^{1/\beta}$ for every two real numbers a and b.

The proposition is proven.