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

Stephen Schön[1], Gael Kermarrec[1], Boris Kargoll[1]
Ingo Neumann[1], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]Leibniz University Hannover, 30167 Hannover, Germany
schoen@ife.uni-hannover.de, gael.kermarrec@web.de
kargoll@gih.uni-hannover.de, neumann@gih.uni-hannover.de
[5]University of Texas at El Paso, USA, olgak@utep.edu, vladik@utep.edu

1. 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.

2. Scale-Invariance (cont-d)

- 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
  - do not change if we simply change the measuring unit.
- In other words, it is reasonable to require that the laws be invariant with respect to *scaling* $x \to \lambda \cdot x$.

3. Scale-Invariance (cont-d)

- 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 of t without changing the units of d or v, the formula stops working.
- However, the formula remains true if we accordingly change the unit for velocity.

4. Scale-Invariance (cont-d)

- For example:
  - if we started with seconds and m/sec, and we change seconds to hours,
  - then 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.

5. 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.
- In many cases, we know the upper bound of the measurement error.
- Then, the probability of exceeding this bound is either equal to 0 or very small (practically equal to 0).
- Often, however, the probability of large measurement errors $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ is not negligible.
- In such cases, we talk about *heavy-tailed* distributions.
- Such distributions are ubiquitous in physics, in economics, etc.

6. Heavy-Tailed Distributions (cont-d)

- Interestingly, they have the same shape in different application areas.
- 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, we have scale-invariance.
- So, for the corresponding pdf $\rho(x)$, for every $\lambda > 0$, there exists $\mu(\lambda)$ for which

$$\rho(\lambda \cdot x) = \mu(\lambda) \cdot \rho(x).$$

7. 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 be measurable and have $\int \rho(x)\,dx = 1$.
- It is known that every measurable solution of the above equation has the 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 $x \to \infty$.

8. A Simple Explanation of Why Power Laws Are the Only Scale-Invariant Ones

- If we assume that $\rho(x)$ is differentiable, then the power laws $c \cdot x^\alpha$ can be easily derived.
- Indeed, $\mu(\lambda) = \dfrac{\rho(\lambda \cdot x)}{\rho(x)}$ is 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 by λ.
- For $\lambda = 1$, we get $x \cdot \dfrac{d\rho}{dx} = \alpha \cdot \rho$, where $\alpha \stackrel{\text{def}}{=} \dfrac{d\mu}{d\lambda}_{|\lambda=1}$.
- By moving all the terms containing ρ to the left-hand side and all others to the right, we get $\dfrac{d\rho}{\rho} = \alpha \cdot \dfrac{dx}{x}$.
- Integrating both sides, we get $\ln(\rho) = \alpha \cdot \ln(x) + C$.
- Hence for $\rho = \exp(\ln(\rho))$, we get $\rho(x) = c \cdot x^\alpha$.

9. 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

$$\rho(x) = c \cdot (1 + a \cdot x^2)^{-\nu} \text{ for some } c, a, \text{ and } \nu.$$

10. What We Show in This Talk

- In this paper, we "rehabilitate" scale-invariance: 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 Student distributions.
- Thus, indirect scale-invariance explains their empirical success.

11. What We Show in This Talk (cont-d)

- This line of reasoning also provides us with a reasonable next approximation.
- Namely, we should try a scale-invariant combination of three or more scale-invariant functions.
- This approximation is worth trying if we want a more accurate description.

12. 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 $X(\omega)$.
- Namely, to describe, for each frequency ω, the mean value $S(\omega) \stackrel{\text{def}}{=} E[|X(\omega)|^2]$.
- The value $S(\omega)$ is known as the *spectral density*.
- In some cases, this function $S(\omega)$ is mainly concentrated at some frequencies.
- However, often, $S(\omega)$ is not negligible neither for small nor for large ω.

13. Multi-D Case (cont-d)

- 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 such functions, we have $\int S(\omega)\, d\omega = +\infty$.
- However, the integral is equal to the overall energy of the spatial signal and should, therefore, be finite.

14. Multi-D Case (cont-d)

- 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 $S(\omega) = \text{const} \cdot (a_0 + a_1 \cdot \omega^2)^{-\nu}$ is the best.
- In this talk, we show that:
  - while this function is not directly scale-invariant, it is indirectly scale-invariant;
  - namely, it is a result of applying a scale-invariant combination function to two scale-invariant $S(\omega)$.

15. Multi-D Case (cont-d)

- Moreover, 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.

16. A Combination Function: Reasonable Requirements

- By a combination function we 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: $a * b = b * a$.
- Also, a combination of a, b, and c should not depend on the order of combination: $(a * b) * c = a * (b * c)$.
- 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).

17. A Combination Function (cont-d)

- **Definition.** *By a* combination f-n $*$ *we mean a commutative associative continuous non-decreasing f-n:*
  - *from pairs of non-negative real numbers*
  - *to non-negative real numbers.*
- 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.** *We say that a combination function is* scale-invariant *if for all a, b, and λ, we have*

$$(\lambda \cdot a) * (\lambda \cdot b) = \lambda \cdot (a * b).$$

18. Main Result

- **Proposition.** *The only scale-invariant combination functions are* $a * b = \min(a, b)$, $a * b = \max(a, b)$, *and*

$$a * b = (a^\beta + b^\beta)^{1/\beta} \text{ for some } \beta.$$

19. 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:

$$\min(c_1 \cdot x^{\alpha_1}, c_2 \cdot x^{\alpha_2}), \quad \max(c_1 \cdot x^{\alpha_1}, c_2 \cdot x^{\alpha_2}), \text{ and}$$

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

- Here, $C_i = (c_i)^\beta$, $\gamma_i = \beta \cdot \alpha_i$, 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.

20. Derivation of Student Distribution (cont-d)

- 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.
- 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 $\gamma_2 \geq 2$.

21. Derivation of Student Distribution (cont-d)

- We want the *generic* case, when both the 0-th and the 2nd coefficient at Taylor expansion are not 0.
- Out of all possible functions of the above type, the *generic* case is only when $\gamma_2 = 2$.
- Thus, we get exactly the Student distribution.
- For dependence of the spectral density on ω, we similarly get exactly Matern's covariance model.

22. What Next?

- Suppose that a scale-invariant combination of *two* scale-invariant functions does not work well.
- Then, we can try a scale-invariant combination of three or more such functions: $f(x) = \left(\sum_{i=1}^{k} C_i \cdot x^{\gamma_i}\right)^{\gamma}$.

23. Alternative Symmetry-Based Explanation

- 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

$$\mathcal{S}(\rho) \stackrel{\text{def}}{=} -\int \rho(x)\,\ln(\rho(x))\,dx.$$

- So, we optimize entropy under the constraints

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

24. Alternative Symmetry-Based Explanation

- The Lagrange multiplier method reduces it to the following unconditional optimization problem: maximize

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

$$\lambda_1\cdot\left(\int x\cdot\rho(x)\,dx\right) + \lambda_2\cdot\left(\int x^2\cdot\rho(x)\,dx - \sigma^2\right).$$

- 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.

25. Entropy Is Scale-Invariant

- Entropy is scale-invariant in the sense that:
  - if we have two distributions $\rho(x)$ and $\rho'(x)$ for which $\mathcal{S}(\rho) = \mathcal{S}(\rho')$, 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 $\mathcal{S}(\rho_\lambda) = \mathcal{S}(\rho'_\lambda)$.
- Entropy is not the only functional with the above scale-invariance properties.
- In addition to entropy, we can also have $\int \ln(\rho(x))\,dx$ and $\int (\rho(x))^q\,dx$ for some q.

26. For Scale-Invariant Generalizations of Entropy, We Get Student Distribution

- Optimizing $\int \ln(\rho(x))\,dx$ under above constraints 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

$$\frac{1}{\rho(x)}-1+\lambda_0+\lambda_1\cdot x+\lambda_2\cdot x^2=0.$$

- Hence $\rho(x)=\dfrac{1}{(1-\lambda_0)-\lambda_1\cdot x-\lambda_2\cdot x^2}$.
- The requirement that the mean is 0 implies $\lambda_1=0$.

27. Scale-Invariant Generalizations (cont-d)

- So we get a particular case of Student distribution.
- Similarly, optimizing $\int (\rho(x))^q\,dx$ under above constraints 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 $a_1 = 0$.
- So we get the general Student distribution.

28. Acknowledgments

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

29. Proof: Case When $1 * 1 = 1$

- We have two possible cases: $1*1 = 1$ and when $1*1 \neq 1$.
- 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.
- 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.

30. Proof: Case When $1 * 1 = 1$ (cont-d)

- So, to fully describe the operation $a * b$, it is sufficient to consider the cases when $a > 0$ and $b > 0$.
- 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, $1 * b^2 = 1 * (b * b^2) = (1 * b) * b^2 = b * b^2 = b^2$.

31. Proof: Case When $1 * 1 = 1$ (cont-d)

- 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, $1 * a \le 1$.
- 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 implies

$$a' \cdot (1 * r) = (a' \cdot 1) * (a' \cdot r) = a' * b'.$$

32. Proof: Case When $1 * 1 = 1$ (cont-d)

- 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 $a' \geq b'$.
- So, in this case, $a * b = \min(a, b)$ for all a and b.
- Let us now consider the second subcase of the first case, when $0 * 1 > 0$.
- 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 $a = 0$, i.e., $0 * 0 = 0$.
- Let us now prove that in this subcase, $0 * 1 = 1$.

33. Proof: Case When $1 * 1 = 1$ (cont-d)

- 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 $0 * 1 = 1$.
- Let us now prove that for every $a \leq b$, 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 get $0 * b = b$ and $1 * b = b$.
- Due to monotonicity, $0 \leq a \leq b$ implies that $b = 0 * b \leq a * b \leq b * b = b$, thus $a * b = b$.
- The statement is proven.

34. Proof: Case When $1 * 1 \neq 1$

- 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

$$(1 * \ldots * 1) * \ldots * (1 * \ldots * 1).$$

- Here, 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 $v(m \cdot n) = v(m) \cdot v(n)$.
- In particular, this means that for every number p and for every positive integer n, we have $v(p^n) = (v(p))^n$.

35. Proof: Case When $1 * 1 \neq 1$ (cont-d)

- If $v(2) = 1*1 > 1$, then by monotonicity, we get $v(3) = 1*v(2) \geq 1*1 = v(2)$, and, in general, $v(n+1) \geq v(n)$.
- Thus, in this case, the sequence $v(n)$ is (non-strictly) increasing.
- Similarly, if $v(2) = 1*1 < 1$, then we get $v(3) \leq v(2)$ and, in general, $v(n+1) \leq v(n)$.
- In this case, the sequence $v(n)$ is strictly decreasing.
- Let us consider these two cases one by one.

36. Proof: Case When $1 * 1 > 1$

- 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, "if $2^m \le p^n$, then $(v(2))^m \le (v(p))^n$" implies:

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

- Similarly, for all m', n', and p, if $p^{n'} \le 2^{m'}$, then $v(p^{n'}) \le v(2^{m'})$, i.e., $(v(p))^{n'} \le (v(2))^{m'}$.

37. Proof: Case When $1 * 1 > 1$ (cont-d)

- 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, $(v(p))^{n'} \le (v(2))^{m'}$ is equivalent to $\dfrac{\ln(v(p))}{\ln(v(2))} \le \dfrac{m'}{n'}$.
- Thus, "if $p^{n'} \le 2^{m'}$, then $(v(p))^{n'} \le (v(2))^{m'}$" implies:

  for every rational $\dfrac{m'}{n'}$, if $\dfrac{\ln(p)}{\ln(2)} \le \dfrac{m'}{n'}$ then $\dfrac{\ln(v(p))}{\ln(v(2))} \le \dfrac{m'}{n'}$.
- 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 $\alpha - \varepsilon \le \dfrac{m}{n} \le \alpha \le \dfrac{m'}{n'} \le \alpha + \varepsilon$.

38. Proof: Case When $1 * 1 > 1$ (cont-d)

- For these numbers, the above two properties imply that $\dfrac{m}{n} \le \beta$ and $\beta \le \dfrac{m'}{n'}$.
- Thus, $\alpha - \varepsilon \le \beta \le \alpha + \varepsilon$, i.e., $|\alpha - \beta| \le \varepsilon$.
- 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, $\ln(v(p)) = \alpha \cdot \ln(p)$, thus $v(p) = p^\alpha$ for all p.
- We can reach a similar conclusion $v(p) = p^\alpha$ when the sequence $v(n)$ is decreasing.
- By definition of $v(n)$, we have $v(m) * v(m') = v(m + m')$.
- Thus, $m^\alpha * (m')^\alpha = (m + m')^\alpha$.
- By using scale-invariance with $\lambda = n^{-\alpha}$, we get

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

39. Proof: Case When $1 * 1 > 1$ (cont-d)

- 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 $\beta \stackrel{\text{def}}{=} 1/\alpha$.
- Rationals $r = \dfrac{m}{n}$ are everywhere dense among reals.
- Hence the values r^α are also everywhere dense.
- So, 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.