# Math Notes: Third Moment Relationships in Benzyl Cinnamate and Puromycin Diversity

Jeff Cromwell, PhD

*The Mathematical Learning Space Research Portfolio: Math Note 5*

Abstract:

Chemical Properties such as Complexity can have plenty of diversity in the tanimoto functional alignments. Here the identification of the cumulative distribution function and its shape parameters and pearson moments provide a way to compare different properties and chemical similarities for large data sets. Here Benzyl cinnamate CID7652 is antibacterial and antifungal compound and Puromycin CID439530 is an inhibitor of translation, antibiotic protein synthesis inhibitor are considered. For N $\geq$ 1000, maximum likelihood estimation was performed and the parameter relationship with the moment generating function explored for two different chemical species. A statistical test was performed to determine the significance of the difference.

Keywords: Beta Distribution Pearson type I Shape Parameters Moment Generating Functions Third Moment antibacterial antifungal Benzyl cinnamate Puromycin

## 1 Introduction

Molecular properties of species diversity can be summarized with categorification with distribution functions. Consider Beta distributions where the Pearson type I distributions are location-scale transformations. Here the Beta pdf is given by [1]

$$f(x;\alpha,\beta) = \text{constant}\cdot x^{\alpha-1}(1-x)^{\beta-1} \tag{1}$$
$$= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \tag{2}$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1} \tag{3}$$
$$= \frac{1}{Beta(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1} \tag{4}$$

where $\Gamma(z)$ is the gamma function. Consider the location parameter given by

$$\mu = \mathrm{E}[X] = \int_0^1 x f(x;\alpha,\beta)\,dx \tag{5}$$
$$= \int_0^1 x\,\frac{x^{\alpha-1}(1-x)^{\beta-1}}{Beta(\alpha,\beta)}\,dx \tag{6}$$
$$= \frac{\alpha}{\alpha+\beta} \tag{7}$$
$$= \frac{1}{1+\frac{\beta}{\alpha}} \tag{8}$$

when $\alpha = \beta$ $\mu$ is the center of the symmetric distribution. The moment generating function is given by [1]

$$M_X(\alpha;\beta;t) = \mathrm{E}\left[e^{tX}\right] \tag{9}$$
$$= \int_0^1 e^{tx} f(x;\alpha,\beta)\,dx \tag{10}$$
$$= {}_1F_1(\alpha;\alpha+\beta;t) \tag{11}$$
$$= \sum_{n=0}^{\infty}\frac{\alpha^{(n)}}{(\alpha+\beta)^{(n)}}\frac{t^n}{n!} \tag{12}$$
$$= 1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!} \tag{13}$$

Let

$$\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r} \tag{14}$$

multiplying the (exponential series) term $\left(\frac{t^k}{k!}\right)$in the series of the moment generating function [1]

$$\mathrm{E}[X^k] = \frac{\alpha^{(k)}}{(\alpha+\beta)^{(k)}} = \prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r} \tag{15}$$

where $(x)^(k)$ is a Pochhammer symbol representing rising factorial. It can also be written in a recursive form as

$$\mathrm{E}[X^k] = \frac{\alpha+k-1}{\alpha+\beta+k-1}\,\mathrm{E}[X^{k-1}]. \tag{16}$$

Since the moment generating function $M_X(\alpha;\beta;\cdot)$ has a positive radius of convergence, the beta distribution is determined by its moments. [1]

Let k=3 and $\alpha = \beta$ then

$$\mathrm{E}[X^3] = \frac{\alpha+2}{\alpha+\beta+2}\,\mathrm{E}[X^2] \tag{17}$$
$$= \frac{\beta+2}{2\beta+2}\,\mathrm{E}[X^2] \tag{18}$$

and the relationship between the two moments is $\frac{\beta+2}{2\beta+2}$ for the symmetric distribution. If $\alpha \neq \beta$ then (a) $\alpha < \beta$ or (b) $\alpha > \beta$. Here skewness $\tilde{\mu}_3$ is given by

$$\tilde{\mu}_3 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{19}$$
$$= \frac{\mathrm{E}[X^3]-3\mu\,\mathrm{E}[X^2]+3\mu^2\,\mathrm{E}[X]-\mu^3}{\sigma^3} \tag{20}$$
$$= \frac{\mathrm{E}[X^3]-3\mu(\mathrm{E}[X^2]-\mu\,\mathrm{E}[X])-\mu^3}{\sigma^3} \tag{21}$$
$$= \frac{\mathrm{E}[X^3]-3\mu\sigma^2-\mu^3}{\sigma^3}. \tag{22}$$

*Mathematical Learning Space Research Portfolio

*Email address:* http://mathlearningspace.weebly.com/ (Jeff Cromwell, PhD)

where $\mu = \mathrm{E}[X]$ and the variance of X is equal to the mean of the square of X minus the square of the mean of X is given by

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] && (23)\\ &= \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}[X]^2\right] && (24)\\ &= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2 && (25)\\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 && (26)\end{aligned}$$

Consider (a) $\alpha < \beta$ then $\alpha + z \approx \beta$ with z > 0

$$\begin{aligned}\mathrm{E}[X^3] \approx& \frac{\alpha+2}{2\alpha+z+2}\,\mathrm{E}[X^2] && (27)\\ \approx& \frac{\alpha+2}{2\alpha+z+2}(\sigma+\mu^2) && (28)\end{aligned}$$

where $\frac{\alpha+2}{2\alpha+z+2}$ is the weight on the scale parameter. If $1 < \alpha < \beta$ then $mode \leq median \leq mean$. Expressing the mode (only for $\alpha, \beta > 1$), and the mean in terms of $\alpha$ and $\beta$:

$$\frac{\alpha-1}{\alpha+\beta-2} \leq \text{median} \leq \frac{\alpha}{\alpha+\beta} \qquad (29)$$

# 2 Results

The complexity rating of a compound by the Bertz/Hendrickson/Ihlenfeldt formula has a scaling factor for aromaticity is used with complexity rule that benzene is equal cyclohexane. The scale is from 0 (simple ions) to several thousand (complex natural products). Most often larger compounds are more complex by size, however symmetry or few distinct atom types or elements are downgraded and loosely correlated with synthetic accessibility. [2] [3]

Benzyl cinnamate CID7652 is antibacterial and antifungal compound. [4] Puromycin CID439530 is an inhibitor of translation, antibiotic protein synthesis inhibitor with earlychain termination during translation. Long-term synaptic plasticity (memory processes) requires morphological changes at the protein level as puromycin inhibits protein synthesis in eukaryotic cells and can generate short-term and long-term memory loss. [5]

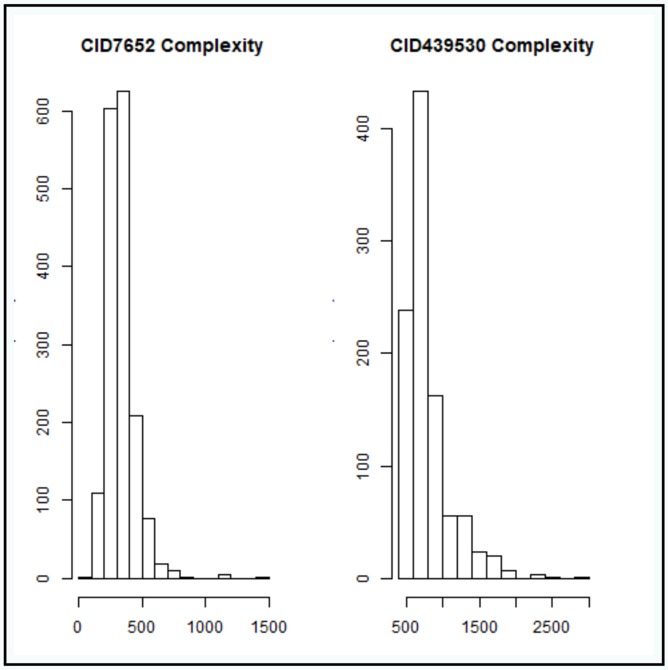


Figure 1: Complexity Frequencies for (a) CID7652 and (b) CID439530 [1001]

For the Two-sample Kolmogorov-Smirnov test D = 0.93171, p-value < 2.2e-16 with alternative hypothesis: two-sided.

Let $X \sim \mathrm{Beta}(\alpha, \beta)$ then $\frac{X}{1-X} \sim \beta'(\alpha, \beta)$ where The beta prime distribution is a special case of the type 6 Pearson distribution obtained by the maximum likelihood estimation of (b) with parameters a=2.643474 b=3.904749 location 471.4477 and scale 366.5962.

For (a) the results are Pearson type 4 with the four parameters of m=3.311518, v=-3.990226, $\lambda$=190.8067 and $\alpha$=161.3513. Here Pearson 4 is given by 3 components

$$\begin{aligned}p(x) =& \frac{\left|\frac{\Gamma\left(m+\frac{\nu}{2}i\right)}{\Gamma(m)}\right|^2}{\alpha\,\mathrm{Beta}\left(m-\frac{1}{2},\frac{1}{2}\right)} && (30)\\ & \left[1+\left(\frac{x-\lambda}{\alpha}\right)^2\right]^{-m} && (31)\\ & \exp\left[-\nu\arctan\left(\frac{x-\lambda}{\alpha}\right)\right]. && (32)\end{aligned}$$

The Cauchy distribution is the limiting case of a Pearson Distribution type 4. For the Cauchy the parameters are location=319.39311 and scale= 40.05748 with

$$\begin{aligned}f(x; x_0, \gamma) =& \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]} && (33)\\ =& \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-x_0)^2+\gamma^2}\right] && (34)\end{aligned}$$

where $x_0$ is the location parameter and $\gamma$ is the scale. The Sen weighted mean statistic $\mathcal{S}_{n,k}$ is a robust estimator of the mean of a distribution

$$\mathcal{S}_{n,k} = \binom{n}{2k+1}^{-1}\Sigma_{i=1}^{n}\binom{i-1}{k}\binom{n-i}{k}x_{i:n}, \qquad (35)$$

where $x_{i:n}$ are the sample order statistics and k is a weighting or trimming parameter. If k = 2, then the $\mathcal{S}_{n,2}$ is the first symmetrical TL-moment (trim = 1). Since the first moment has three distinct measures of central tendency for (a) and (b) the Sen weighted mean statistic can be used. Consider that

$\mathcal{S}_{n,0} = \mu = \overline{X}_n$ or the arithmetic mean and $\mathcal{S}_{n,k}$ is the sample median if either n is even and k = (n/2) - 1 or n is odd and k = (n-1)/2. Here for (a) 316.8072 and (b) 730.4129 and the value of $0 < k < (n-1)/2$ jas k=2. The The Gini mean difference statistic $\mathcal{G}$ is a robust estimator of distribution scale and is closely related to the second L-moment $\lambda_2 = \mathcal{G}/2$.

$$\mathcal{G} = \frac{2}{n(n-1)} \Sigma_{i=1}^{n}(2i-n-1)x_{i:n}, \tag{36}$$

where $x_{i:n}$ are the sample order statistics.[1002] For (a) gini=116.6499 and L2=58.32497 and (b) gini=292.8128 L2=146.4064. The sample L moments for (a) lambdas 330.548523 58.324973 11.155413 11.724774 5.121723 and ratios NA 0.17644905 0.19126306 0.20102494 0.08781356 and (b) lambdas 801.60700 146.40642 54.45218 31.56046 15.38780 ratios NA 0.1826411 0.3719248 0.2155674 0.1051033.[1002][6]

The equivalent cumulative distribution function of two distributions has the nonexceedance probability of a given quantile from a linear weighted combination of two quantile functions. The left-tail parameter is the distribution and determines the left tail; the right-tail parameter governs the right tail. The quantile function algebra is

$$Q(F) = (1-F) \times Q_{\triangleleft}(F) + F \times Q_{\triangleright}(F), \tag{37}$$

where Q(F) is the equivalent quantile for nonexceedance probability F computed by the tail weigthing. $Q_{\triangleleft}(F)$ is the left-tail quantile function; $Q_{\triangleright}(F)$ is the right-tail quantile function. Here the combined median is (a) 320.67 and (b) 757.79. [6]

For the linear regression of complexity by rotatable bonds for CID7652, the results are presented in Table 1.

| Parameter | Estimate | Std. Error | t value | |
|---|---|---|---|---|
| (Intercept) | 199.0786 | 5.2318 | 38.05 | 2e-16 *** |
| rotbonds | 18.6242 | 0.6622 | 28.12 | 2e-16 *** |

Figure 2 has the 4 diagnostics for the regression.

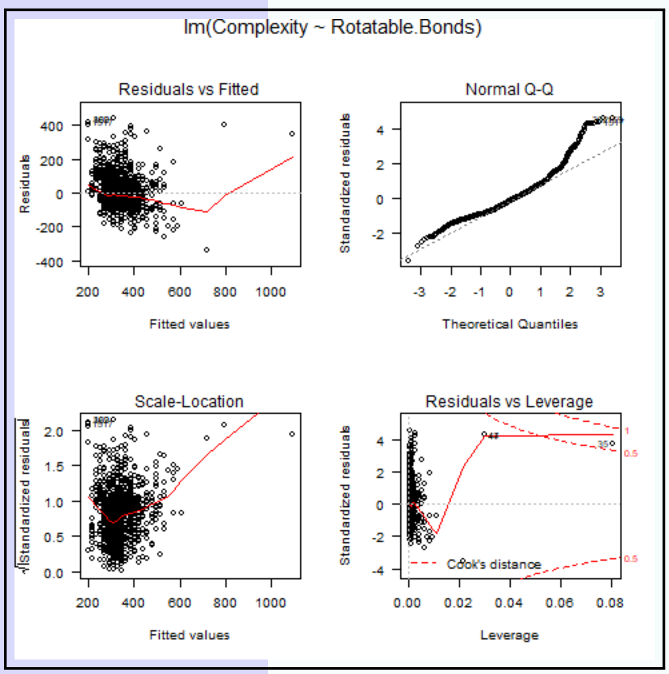


Figure 2: Diagnostics for the regression. [1001]

## 3 Conclusion

Chemical Properties such as Complexity are diverse in tanimoto alignments. Here the identification of the cumulative distribution function and its shape parameters and moments provide a way to compare different properties and chemical similarities for large data sets. Maximum likelihood estimation was performed and the parameter relationship with the moment generating function explored for two different chemical species. A statistical test was performed to determine the significance of the difference. Since these properties often demonstrate one of three distinct types with respect to the third parameter and its numerical magnitude of negative, 0 and positive categories, the idea that small values are more likely than large values suggest that nature prefers less complexity to more with respect to the volume of diversity for control systems to provide flexibility, adaptability and error handling.

## 4 References


[1] Wikipedia contributors. "Beta distribution." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 8 Aug. 2021. Web. 18 Aug. 2021.

[2] S.H. Bertz, The first general index of molecular complexity J. Am. Chem. Soc., 1981, 103 (12), pp 3599–3601

[3] Hendrickson et al., Molecular complexity: a simplified formula adapted to individual atoms J. Chem. Inf. Comput. Sci., 1987, 27 (2), pp 63–67.

[4] Wikipedia contributors. "Benzyl cinnamate." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 14 Jun. 2021. Web. 19 Aug. 2021.

[5] Wikipedia contributors. "Puromycin." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 7 Jun. 2021. Web. 19 Aug. 2021.

[6] Wikipedia contributors. "L-moment." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 30 Apr. 2021. Web. 19 Aug. 2021.

[7] Wikipedia contributors. "Skewness." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 7 May. 2021. Web. 20 Aug. 2021.

[8] Wikipedia contributors. "Pearson distribution." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 6 May. 2021. Web. 20 Aug. 2021.

[9] Wikipedia contributors. "Cauchy distribution." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 6 Jul. 2021. Web. 20 Aug. 2021.

[1000] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[1001] Osorio, D., Rondon-Villarreal, P. and Torres, R. Peptides: A package for data mining of antimicrobial peptides. The R Journal. 7(1), 4-14 (2015).

[1002] Asquith, W.H., 2017, lmomco—L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.2.7, Texas Tech University, Lubbock, Texas.