

Rated Set Theory

Jason Lind / EVP & Chief Software Architect

jason@lind-i.com 414.788.2820



Definitions

Symbol	Description	Symbol	Description
\mathcal{C}	Count	ρ_i	Probability of i occurrence
\mathbb{U}	Unique Identifier	σ	Standard deviation
\mathbb{Z}	Set of Integers	μ	variance
$x condition$	x such that condition	ϕ_x	Probability density of x
$x \in Y$	x is an element of set Y		

A Rated Set

A rated set is defined as:

$$\mathfrak{R} = \left\{ \left(\begin{array}{l} member = \{ \mathbb{U}, \{characteristics\} = c \} = m, \\ rating = r | (a |_{a \in \mathbb{Z}^+} \leq r \leq b |_{b \in \mathbb{Z}^+, b > a}), \\ rater = \{ \mathbb{U}, c \} = p \end{array} \right) \right\}$$

The unweighted rating for a member is: $R_m = \frac{\sum_{\mathfrak{R}|m} r}{\mathcal{C}_{\mathfrak{R}|m}}$

When comparing unweighted ratings we implicitly consider the following:

$$\rho_r = \frac{\mathcal{C}_{\mathfrak{R}|r}}{\mathcal{C}_{\mathfrak{R}}} \quad \mu = \sum_{r=a}^b \rho_r r \quad \sigma = \sqrt{\sum_{r=a}^b \rho_r (r - \mu)^2}$$

Assuming the rating distribution has a high correlation to a normal distribution in our minds we evaluate the difference between two ratings accordingly:

$$\phi_m = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(R_m - \mu)^2}{2\sigma^2}}$$

$$\phi_{mn} = \iint \phi_m dR_m - \phi_n dR_n = \sqrt{\frac{2}{\pi}} \sigma e^{\frac{\mu}{2\sigma^2}} \left(R_m e^{-\frac{R_n}{2\sigma^2}} - R_n e^{-\frac{R_m}{2\sigma^2}} \right)$$

ϕ_{mn} effectively tells us the magnitude of the difference between two ratings with respect to data.

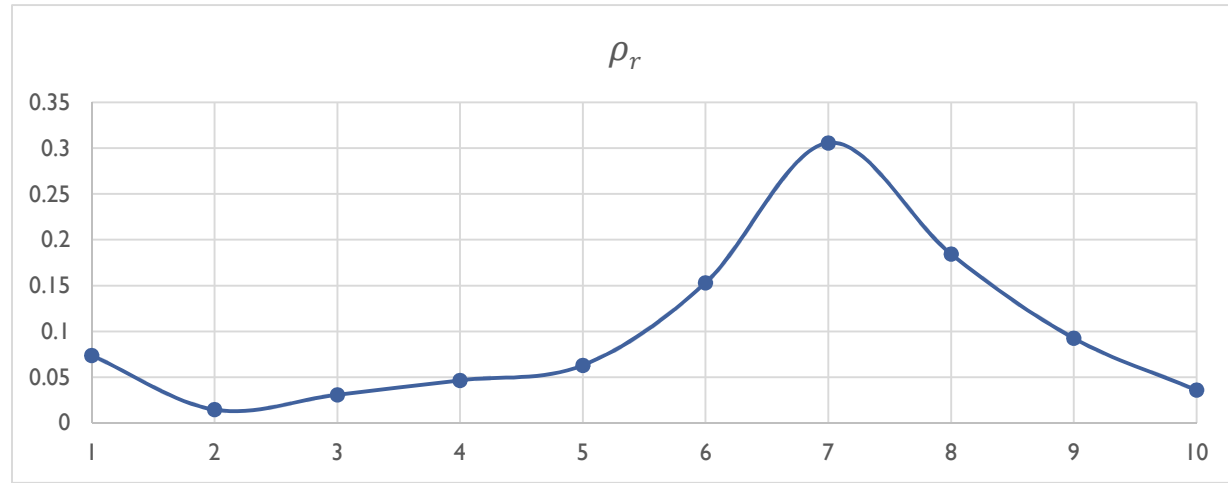
We then normalize this to the terms of the rating system to provide a weighted interpretation of the rating: ω_r .

$$\varphi_r = \sum_{\lambda=b}^a \rho_r \phi_{r\lambda} \quad \omega_r = \begin{cases} r + \frac{1}{\varphi_r} & \text{except} \\ r + \varphi_r & | r = \frac{b - a + 1}{2} \end{cases}$$

Example Data Set

For this example we will use my rating data from songs played on RadioParadise. http://www.radioparadise.com/rp_2.php?#name=Members&file=userinfo&u=6814

r	Occurrence	ρ	μ	σ
10	57	0.035872	0.358716	0.459483
9	147	0.092511	0.832599	0.615304
8	293	0.184393	1.475142	0.459724
7	486	0.305853	2.140969	0.102527
6	243	0.152926	0.917558	0.027107
5	100	0.062933	0.314663	0.12708
4	74	0.04657	0.186281	0.272963
3	49	0.030837	0.092511	0.360897
2	23	0.014475	0.028949	0.28291
1	117	0.073631	0.073631	2.163834
	1589		6.42102	2.207222



As you can see I favor a 7 rating, which is because I tend to like what RadioParadise plays. I also heavily weight 1's with respect to the set in general, so the distribution isn't exactly normal, however it leads to a much higher μ , σ than a standard normal distribution. I contend that people tend to assume a standard distribution when interpreting raw ratings with the absence of an understanding of the distribution.

$\phi_{r\lambda}$	10	9	8	7	6	5	4	3	2	1	$\rho_r \phi_{r\lambda}$	10	9	8	7	6	5	4	3	2	1	ω_r
10	0.0	-2.5	-5.2	-8.1	-11.1	-14.3	-17.7	-21.4	-25.3	-29.5	10	0.00	-0.09	-0.19	-0.29	-0.40	-0.51	-0.63	-0.77	-0.91	-1.06	9.79
9	2.5	0.0	-2.7	-5.5	-8.4	-11.6	-14.9	-18.5	-22.2	-26.3	9	0.09	0.00	-0.25	-0.51	-0.78	-1.07	-1.38	-1.71	-2.06	-2.43	8.90
8	5.2	2.7	0.0	-2.8	-5.7	-8.8	-12.1	-15.5	-19.2	-23.1	8	0.19	0.49	0.00	-0.51	-1.06	-1.63	-2.23	-2.86	-3.54	-4.26	7.94
7	8.1	5.5	2.8	0.0	-2.9	-6.0	-9.2	-12.5	-16.1	-19.8	7	0.29	1.67	0.85	0.00	-0.89	-1.82	-2.80	-3.83	-4.92	-6.07	6.94
6	11.1	8.4	5.7	2.9	0.0	-3.0	-6.2	-9.5	-13.0	-16.6	6	0.40	1.29	0.88	0.45	0.00	-0.46	-0.95	-1.45	-1.98	-2.54	5.77
5	14.3	11.6	8.8	6.0	3.0	0.0	-3.1	-6.4	-9.8	-13.3	5	0.51	0.73	0.55	0.38	0.19	0.00	-0.20	-0.40	-0.62	-0.84	5.31
4	17.7	14.9	12.1	9.2	6.2	3.1	0.0	-3.2	-6.6	-10.0	4	0.63	0.69	0.56	0.43	0.29	0.15	0.00	-0.15	-0.31	-0.47	4.55
3	21.4	18.5	15.5	12.5	9.5	6.4	3.2	0.0	-3.3	-6.7	3	0.77	0.57	0.48	0.39	0.29	0.20	0.10	0.00	-0.10	-0.21	3.40
2	25.3	22.2	19.2	16.1	13.0	9.8	6.6	3.3	0.0	-3.4	2	0.91	0.32	0.28	0.23	0.19	0.14	0.10	0.05	0.00	-0.05	2.46
1	29.5	26.3	23.1	19.8	16.6	13.3	10.0	6.7	3.4	0.0	1	1.06	1.94	1.70	1.46	1.22	0.98	0.74	0.49	0.25	0.00	1.10

