

Modeling Human Inventivity as a Random Poisson Process

Ye An

May 11, 2012

This report is both a summary and a critique of the article *Invention and Inventivity Is a Random, Poisson Process: A Potential Guide to Analysis of General Creativity* by John C. Huber from the Institute of Invention and Innovation. The article studies human creativity through measuring the number of patents issued to each individual inventor throughout his/her inventive career. By taking samples from different groups of inventors, the article draws the conclusions that the time pattern of patents for most inventors is random and fit the Poisson Process (Huber, 1998).

1 Motivation

In the field of psychology, the study of creativity has always been a very important aspect. From a theoretical point, it helps us understand humans' cognitive capacities. From a practical point, identifying creative people and promoting people's creative capacity are of particular interest to organizations such as technological firms and universities. According to Huber, most studies of creativity have been focusing on qualitative characteristics such as personality and social background. Thus, the quantitative approach adopted by this article is very innovative and leads the study of inventivity to a new direction for future studies.

2 Basic Models

2.1 Variables

In order to build the model, precise definitions are given to the variables that are to be measured.

2.1.1 Measure of Creativity

Creativity embodies the idea of novelty and innovation, thus by nature it's hard to provide a standard measurement for it. To resolve this difficulty, Huber uses

the number of patents issued to each inventor and the occurrence of each patent in time. This reflects the fact that the issuance of patents follows a standard criterion nationwide, regardless of the field from which the invention comes, and therefore is able to give an objective measurement for creativity.

2.1.2 Duration of Inventive Careers

In a theoretical Poisson Process, the starting point is generally assumed, and the end point can be arbitrarily cut off. However, in reality, since inventors will always start and end their inventive career at some point in their life, it is necessary to specify this boundary of time. For simplicity, Huber makes the first patent as the starting point and the last patent as the end point. Of course such an arrangement does not take into consideration the 'unknown number of years of inventive effort with zero patents before and after the period' (Huber, 1998). However, when there are enough inter-occurrences of patents, a Poisson Process should still be observed.

2.1.3 Inventivity

Let T be a specific time period, then each inventor's inventivity over T is defined as followed:

$$Inventivity(T) = \frac{\text{Number of Patents in } T}{T \text{ (in years)}}$$

Note that when $T=1$, the inventor's inventivity is just the number of patents issued during that year. Similarly, we can define an inventor's inventivity over his/her entire career duration as:

$$Average\ Inventivity = \frac{\text{Total Number of Patents}}{\text{Duration of Inventive Career}}$$

2.2 Structure of the Model

Rather than modeling the time pattern of patents directly as a Poisson Process, Huber splits the process into two layers. At the top layer, patents are separated according to the years they are issued, and the number of patents in each year is counted. At the next level, if indeed the time pattern of patents is a Poisson Process, then the number of patents in each year should be a sample taken from the same Poisson distribution. Two different tests are conducted at each layer to test our hypothesis.

2.2.1 Test of Runs

This text aims at testing whether at the top layer, the number of patents in each year is random and independent of time. We give an easy example to illustrate the point. A vector consisting of the time sequence of patents can be generated after we've calculated the number of patents in each year. For an inventor with

a total of 5 patents in a duration of 10 years, a typical vector will look like (1, 0, 1, 1, 0, 0, 1, 0, 0, 1). By definition, a *run* is a sequence of one or more years with the same number of patents. As in our example, the vector has 7 runs. Next, since the expected number of runs can be calculated for each specific type of vectors, then the vector's deviation in terms of the number of runs from the expected value is in fact a measure of its un-randomness. In our example, for 5 patents over 10 years, the expected number of runs is 6. Therefore, the vector we have been examining has a high degree of randomness.

2.2.2 Goodness-of-fit Test

After confirming that the number of patents are in general random across different years, we can then proceed to conduct a goodness-of-fit test to the Poisson distribution by the chi-square method, since such a statistical test aims at determining whether a sample is taken from a certain distribution.

3 Comparison with Experimental Results

In the article, the experimental data indeed is a good fit of the model. According to Huber, the tests of runs show that the probability that each increment is independent exceeds 0.65, across different groups. And the goodness-of-fit tests show that the probability that the time pattern of patents are Poisson distributed is around 0.80. Furthermore, Huber also observes that this model works well when we take into consideration the quality of the inventions. That is, he analyzed a subgroup of inventors that have won several National Awards, and the test results exhibit the same pattern.

4 Analysis of the Model

4.1 Advantages

This model successfully describes the time pattern of patents for individuals with high out-put and long career duration. In particular, the model suggests that there is an innate parameter for every inventor that is constant over time and can be used to measure the level of creativity for that particular individual. From the property of the Poisson Process, it also implies that despite the existence of such a parameter, the time until the next occurrence of a patent is random and independent of previous occurrences, suggesting that creativity cannot be increased due to the demand on time constraints.

On the other hand, for companies hiring inventors, once this parameter is known, several inferences can easily be made. Let an inventor's time pattern of patents be a Poisson Process with parameter λ , and let X be a random variable that represents the inventor's inventivity in each year. Then, X is a Poisson distribution with parameter $\lambda \times 1 = \lambda$. We can then calculate:

1. An inventor's expected inventivity in any year: $E[X] = \lambda$.

2. An inventor's variance of inventivity: $\text{Var}[X]=\lambda$.

This way, although an inventor's annual inventivity is random, the company can still get a good sense of what to expect in the long run.

Furthermore, since the expected value of each patent and the variance of this value can be easily determined, the company will then be able to estimate the expected total value that a certain inventor will bring to the company. That is, let $V(t)$ be the total value of all the patents by some t . And since we know the value of the i th patent, p_i has $E[p_i]=\mu$, and $\text{Var}[p_i]=\sigma^2$, for $i= 1, 2, \dots$, we can then calculate:

1. An inventor's expected value to the company by time t : $E[V(t)]=\mu\lambda t$.
2. An inventor's variance of value by time t : $\text{Var}[V(t)]=\lambda t\sigma^2 + \lambda t\mu^2$

4.2 Drawbacks

This model does have several drawbacks. For example, it is very hard to determine whether someone with only few patents over his/her career exhibits such a pattern. With so few data, the sampling error will be rather large. And yet this is the common scenario, since the number of patents among different inventors exhibits a Pareto distribution, that is, the majority of the inventors has only a few patents, while a small fraction of the inventors accounts for most of the patents in the sample. Therefore, the model is only valid for these top inventors.

To remedy for this, we could do some modifications to this model. Instead of first dividing an inventor's career span into years, we could record the timing of each patent directly by dates and then conduct another goodness-of-fit test to determine whether the inter-arrival times conform with the exponential distribution. We'll use the example in Section 2.2.1 again as an illustration. With only 5 patents in 10 years, it is not possible to test whether a vector like (1, 0, 1, 1, 0, 0, 1, 0, 0, 1) fits into a certain Poisson distribution since our 10 samples consists of five 1's and five 0's. However, if instead we represent this same inventor's time pattern of patents in a vector like (821, 210, 1210, 532), with each number as the inter-arrival time (in days) of a new patent, then we'd be able conduct a test to see whether the 4 numbers fit into a certain exponential distribution. This way, even with a relatively small amount of patents, we would still be able to make some predictions.

5 Other Models

Consequently, while confirming the Poisson Process model for creativity, the experimental results invalidate several other alternative models. For example, as Huber mentioned, one might hope to see more examples of systematic patterns, such as:

1. Increasing output rate over time, indicating learning.
2. Decreasing output rate over time, indicating senescence.
3. Regular, non-random output rate, indicating control and performance to objectives.
4. Sharply peaked bursts of output, indicating breakthroughs.

However, Huber chooses the current model as he believes that the randomness in terms of the timing of the patents can all be accounted for in the Poisson Process and that an inventor's level of creativity is stable over time.

References

John C. Huber (1998): Invention and Inventivity Is a Random, Poisson Process: A Potential Guide to Analysis of General Creativity, *Creativity Research Journal*, 11:3, 231-241g