Poisson Prediction of the loss of Teachers in High Schools

Li Yingqiu
College of Mathematics and Computing Science, Changsha University of Science and Technology, Changsha, China
liyq-2001@163.com

Sha Yongqiang
College of Mathematics and Computing Science, Changsha University of Science and Technology, Changsha, China
shayongqiang@126.com

Zhao Renke
College of Mathematics and Computing Science, Changsha University of Science and Technology, Changsha, China
zhaorenke2006@126.com

Abstract—It is a key to the sustainable development and stability of colleges and universities education to predict the loss of college teachers. This paper analyzes the historic statistics and applies Chi-square test to do a hypothesis test on the distribution of intervals between two successive teachers’ losses. It also sets up prediction models for the loss based on Poisson Process Theory. After some empirical analysis, it draws a conclusion and provides a new approach to forecast the staff turnover in colleges and universities.

Keywords—Exponential distribution; Poisson process; Prediction model; Data-fitting

I. INTRODUCTION

In today's society, human resources competition is the main competition. The status of human resources in high schools determines its overall comprehensive strength. University teachers are the subject of human resources and their team status determines the level of university teaching and research strength, thus human resource management should focus on construction of teachers. In recent years, many scholars dedicated to human resource management research, such as Markov process with transition probability matrix quantitatively predicting the allocation of teachers staff [1] [2], and this paper attempts to analyze the staff turnover of teacher system in a new perspective. We applied Poisson process theory to predict the loss of university teachers, and then we set up the prediction model in the loss of a system of university teachers' number in a certain period through the statistical analysis of historical data. Following the model, we can obtain the least upper bound of the number of random loss during [0, t), the confidence level is high, which provides a reference that estimates the most actual number of teachers needed to be supplemented for the personnel management department.

II. POISSON PREDICTION MODEL ON THE NUMBER OF SYSTEM STAFF TURNOVER

Teacher wastage process is a counting process, when the time intervals of the loss of two successive staff are independent and exponentially distributed with the same parameters \( \lambda \). In accordance with Poisson process theory we can put the counting process as a Poisson process [3]. Then we established prediction model according to Poisson process for predicting the loss of teachers in college from the current time (denoted as 0 time) to the moment of time \( t (t \geq 0) \).

A. Time interval distribution of system turnover

Statistics conditions of a college teacher turnover system are as follows.

Within the 124 weeks from March 20, 2006 to July 28, 2008, the loss of teachers in a college amounts to 130, the time intervals \( \Delta \) (weeks) of successive staff loss are as follows:

<table>
<thead>
<tr>
<th>( \Delta ) (weeks) fall in the range</th>
<th>( [0,1) )</th>
<th>( (1,2] )</th>
<th>( (2,3] )</th>
<th>( (3,4] )</th>
<th>( (4,5] )</th>
<th>( (5,6] )</th>
<th>( (6,7] )</th>
<th>( (7,\infty) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta ) frequency of occurrence</td>
<td>78</td>
<td>30</td>
<td>12</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Let \( \Delta \) denote time interval between \((l-1)th \) \( (l = 1,2,\ldots) \) staff turnover and \( lth \) \( (l = 1,2,\ldots) \) staff turnover, which is obviously a continuous random variable. Considering the fact that it is independent from each other whether personnel in the system is still within the system in the future, then the time interval \( \{ \Delta_1, |l \geq 1 \} \) can be considered independent of each other.

Next, we apply Chi-square test to do a fit of exponential distribution on the distribution of time intervals, i.e. to test hypothesis \( H_0 \) : the probability density of \( \Delta \) is

\[
 f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0. \end{cases}
\]

Test process is as follows:

The estimated value of maximum likelihood of \( \lambda \) is

\[
 \hat{\lambda} = \frac{1}{\overline{\tau}} = 130/124 = 1.0484 .
\]

If \( H_0 \) is true, the estimate of the distribution function of \( \Delta \) is:

This work was partially supported by the NNSF of China (Grant No.10771021 10471012), the Planned Science and Technology Project of Hunan Province (Grant No.200963098) and Scientific Research Fund of Hunan Provincial Education Department (Grant No.09C113, 09C059, 08C120)

U.S. Government work not protected by U.S. copyright
Thus, table 2 shows the Chi-square test value table of the distribution of $\Delta$ by table 1.

**Table 2. The value table of Chi-square test**

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Interval</th>
<th>$f_i$</th>
<th>$p_i$</th>
<th>$n_p_i$</th>
<th>$n_p_i - f_i$</th>
<th>$(n_p_i - f_i)^2 / n_p_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[0,1]</td>
<td>78</td>
<td>0.614879</td>
<td>79.93431</td>
<td>1.934308</td>
<td>0.046807793</td>
</tr>
<tr>
<td>2</td>
<td>(1,2]</td>
<td>30</td>
<td>0.236803</td>
<td>30.78436</td>
<td>0.784357</td>
<td>0.019984702</td>
</tr>
<tr>
<td>3</td>
<td>(2,3]</td>
<td>12</td>
<td>0.091198</td>
<td>11.85569</td>
<td>-0.14431</td>
<td>0.00175649</td>
</tr>
<tr>
<td>4</td>
<td>(3,4]</td>
<td>5</td>
<td>0.035122</td>
<td>4.565873</td>
<td>4.565873</td>
<td>0.019984702</td>
</tr>
<tr>
<td>5</td>
<td>(4,5]</td>
<td>3</td>
<td>0.013526</td>
<td>1.758412</td>
<td>1.758412</td>
<td>0.00175649</td>
</tr>
<tr>
<td>6</td>
<td>(5,6]</td>
<td>1</td>
<td>0.005209</td>
<td>0.677201</td>
<td>0.677201</td>
<td>0.00175649</td>
</tr>
<tr>
<td>7</td>
<td>(6,7]</td>
<td>1</td>
<td>0.002006</td>
<td>0.260804</td>
<td>0.260804</td>
<td>0.00175649</td>
</tr>
<tr>
<td>8</td>
<td>(7,∞)</td>
<td>0</td>
<td>0.001257</td>
<td>0.163351</td>
<td>0.163351</td>
<td>0.00175649</td>
</tr>
<tr>
<td>$\sum$</td>
<td>130</td>
<td>0.961040742</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since $\chi^2_{0.05}(k-r-1) = \chi^2_{0.05}(2) = 5.991 > \chi^2 = 0.9610$ , so it accepts hypothesis $H_0$ . Under the given confidence level of 0.95, it can be considered that the samples are from the exponential distribution with the parameters $\hat{\lambda}$ , that is, the theoretical results are realistic. Therefore, it can be considered that: time intervals series of $\Delta$ is subjected to the exponential distribution with parameters $\hat{\lambda} = 1.0484$ and $\Delta - \hat{f}(x) = \left\{ \begin{array}{ll} 1.0484e^{-1.0484x}, & x > 0; \\ 0, & x \leq 0. \end{array} \right.$

B. The distribution of the number of system turnover

Let $N(t)$ denote the number of system turnover from the current time (denoted as 0 time) to the time $t$ ($t \geq 0$) the loss of teachers in colleges. $N(t)$ is the counting process with time continuous and space state $\{0,1,2,3,\cdots\}$.

By the definition of $\Delta$ , the counting process $N(t)$ corresponds to the distance between points $\{\Delta | I \geq 1\}$. From 2.1 in the discussion results, we found that the distributions of time intervals $\Delta$ of the university distribution with parameters $\hat{\lambda} = 1.0484$ . And the time interval is independent of each other. Then by the Poisson process theory [4], $N(t)$ is a the Poisson process with parameters $\hat{\lambda}$ . Then for any $t \geq 0$ , we have

$$P(N(t) = k) = \frac{(\hat{\lambda}t)^k}{k!} e^{-\hat{\lambda}t}, \quad k = 0,1,2,\cdots, \quad (1)$$

where $\hat{\lambda} = 1.0484$

C. Forecasting Model

Let $0.95 < \beta < 0.99$ , from equation (1), under the given confidence level $\beta$ , the predicted value $s_{i,\beta}$ in the time interval $[0,t]$ denotes the number of system staff turnover in the college system, which is satisfied by

$$s_{i,\beta} = \min\{s : P(N(t) \leq s) = \sum_{r=0}^{\infty}(\alpha t)^r e^{-\beta t} \geq \beta\} \quad (2)$$

where $\hat{\lambda} = 1.0484$.

In the equation (2), predicted values $s_{i,\beta}$ means the minimum upper bound number of random loss during the time interval $[0,t]$ , its confidence is greater than $\beta$ . Or $s_{i,\beta}$ is an indicator to measure stability of teacher system in colleges.

III. MODEL CHECKING

A. Calculation $s_{i,\beta}$ steps:

Since $\sum_{r=0}^{\infty}(\alpha t)^r e^{-\beta t} \geq \beta$ , i.e., $\sum_{r=0}^{\infty}(\alpha t)^r e^{-\beta t} \geq \beta e^{\beta t}$ . Then calculated $s_{i,\beta}$ is as follows:

Step 1: Given $t, \beta, \hat{\lambda}$ , calculated $\mu = \beta e^{\beta t}$ , $\gamma = \hat{\lambda} t$ ;

Step 2: let $s = 0 , f_i = 0$ ;

Step 3: Calculated $f_i = f_{i-1} + \frac{\gamma}{s!} + \sum_{r=0}^{i-1}(\alpha t)^r e^{-\hat{\lambda} t}/k!$ ;

Step 4: If $f_i > \mu$ , let $s_i = s$ , until calculated the end. If $f_i < \mu$ , let $s = s + 1$ , turn to step 3.

B. Empirical analysis

To further illustrate the validity and accuracy of the model (2), the following to verify the model. According to calculated steps, we can obtain table 3 and figure 1 by MATLAB software.

**Table 3. The predicted and actual values ($\beta = 0.96$ )**

<table>
<thead>
<tr>
<th>Time range (0,t(weeks))</th>
<th>Predicted values</th>
<th>Actual loss values</th>
<th>Time range (0,t(weeks))</th>
<th>Predicted values</th>
<th>Actual loss values</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1]</td>
<td>3</td>
<td>2</td>
<td>(0,64]</td>
<td>82</td>
<td>67</td>
</tr>
<tr>
<td>(0,2]</td>
<td>5</td>
<td>4</td>
<td>(0,68]</td>
<td>86</td>
<td>70</td>
</tr>
<tr>
<td>(0,4]</td>
<td>8</td>
<td>9</td>
<td>(0,72]</td>
<td>91</td>
<td>74</td>
</tr>
<tr>
<td>(0,8]</td>
<td>14</td>
<td>12</td>
<td>(0,76]</td>
<td>95</td>
<td>76</td>
</tr>
<tr>
<td>(0,12]</td>
<td>19</td>
<td>16</td>
<td>(0,80]</td>
<td>100</td>
<td>81</td>
</tr>
</tbody>
</table>
Table 3 shows the actual values and prediction values of introducing to the system of teachers in 2007 and 2008. According to the table 3, it shows that the relative error between the number of actual loss and prediction values is less than 5%, generally in correspond with the actual situation, so the model is accurate and effective. By analyzing figure 1, we also find that the predicted values $s_{t, \beta}$ is a linear function of the time $t$, and the probability of the actual value curve below the forecast value is more than 95%, and then compare them, under the given confidence level we can obtain the extent of its stability, and the greater confidence, the better stability.

IV. CONCLUSIONS

According to the system staff turnover, we can establish the prediction model for teachers system using Poisson process theory analysis for the first time, and validate the model accuracy using the school's historical data. The results show that we can set up prediction models for the loss based on Poisson Process Theory and apply the model to predict the loss of staff, and then estimate the actual supplement staff and add the actual number of teachers, which provides a new scientific evaluation criterion for measuring the stability of the system staff turnover.

REFERENCES