Stochastic Modeling for Efficient Use of Public Transportation

Yohanssen Pratama
Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Sitoluama, Indonesia
yohanssen.pratama@del.ac.id

***Abstract*—Recently in Indonesia appeared a new breakthrough in terms of public transport, namely online taxi that can be ordered online via smartphone or mobile telephone. This phenomenon raises some effects to the community, one of the negative effects is the income of conventional public transport drivers (commonly referred as angkot drivers) were drastically reduced. This already caused some clashes between the online taxi drivers and angkot drivers. To maintain the income of angkot drivers we need to build a management system that can provide some information to the drivers. The sufficient time to pick up passengers and how the system will provide many units of vehicles on the specific route that should operate. A stochastic is proposed a model to become a base for an information system that provides some outputs to help the angkot drivers and the transportation agencies make a right decision.**

***Index Terms*—Stochastic; Model; Public Transportation; Online Taxi; Angkot.**

I. Introduction

Online taxi services such as UberTaxi and GrabCar that are based on online mobile applications are increasingly popular. Many success stories from the drivers increase the popularity of the business, More and more people who join the online taxi make the number of drivers increase. The promo fares for the passengers cause the popularity of online taxi to increase from time to time.

Unfortunately, this event causes a reduction in the revenue of conventional public transportation drivers commonly called angkot. Several angkot drivers began to reject the presence of online taxis in several places. Until now the conflict is still present and sometimes followed by violence between two parties. The presence of online taxis cannot be denied along with the advancement of information technology and to overcome this issue, we must use technology approach, not by some of the restrictions.

In this research, a stochastic model is proposed that can be used for the basic development of information system. This information system will help the angkot drivers and the decision makers to obtain information about the sufficient time to pick up the passengers at specific locations and the numbers of vehicles that should operate at that time.

Poisson distribution method and stochastic programming model can be used. The Poisson distribution is used to predict the sufficient time to pick up the passengers, and the stochastic programming model could decide the number of the vehicle that should run on the specific route. The Poisson distribution is a discrete probability distribution that expresses the probability of some events occurring in a fixed period, especially if these events occur with a known average rate and time-independent since the last event occurs [1]. The average number of passengers that can be pick up by angkot in a certain amount of time could be known by monitoring the passengers that get in and out from the angkot. For this research, a single person is assigned as an observer in specific locations to count the number of passengers by entering the data using a smartphone.

Figure 1: Number of passengers sent via smartphone

After collecting the data for the specified period, the number of passengers and its frequency can be obtained. This data is useful to know the sufficient time to pick up the passengers.

On the other hand, the stochastic programming model could be used to optimise a number of angkot that operate at the specific route. Stochastic programming is an approach to modelling optimisation problems that involve uncertainty. We use stochastic programming because the angkot has a random demand of passengers, so the probability distribution (e.g., numbers of passengers) could be estimated from data that have been collected over time. The goal is to find all possible parameter and do optimisation for the expectation function and the random variables.

This paper is organised as follows. In section 2, the Poisson Distribution methods are given. In section 3, the stochastic programming model based on angkot route is presented. The generated model from both methods that proposed will be presented in section 4. Finally, the work of this paper is summarised in the last section

II. Poisson Distribution To See Numbers Of Passengers

Studies and observations were conducted to understand the random variable by using one of the public transport in Bandung city. With a route from Panghegar to Dipatiukur as an object of observation for sampling, the random variable may be known (in Figure 2, Panghegar is represented by ‘A’ and Dipatiukur is represented by ‘B’). Based on observation the object to be sampled was determined, and the number of passengers as a sample was decided.

Figure 2: Route

For the present study, a survey was conducted in the Balubur area. Balubur area is located in between A and B (Panghegar and Dipatiukur). An observer collected the number of passengers from any Panghegar-Dipatiukur angkot that passed the observer per unit of time. The route followed by the angkot is as below:

Figure 3: Panghegar – Dipatiukur angkot route

The random variables to be analysed are the number of passenger per time and its frequency. There are time intervals that have low frequency but there are also time intervals with high frequency. To know the random variables, Poisson Distribution method was used in this research. The main parameter in determining Poisson distribution is the average value (mean value), usually marked by λ. The average value of occurrence of an event is usually at regular intervals (for example: the number of vehicles passing per hour). The probability distribution is given by the following formula:

Table 1
Probability Distribution

x	0	1	2	3	4	…
P(X=x)	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	$\frac{\lambda^3 e^{-\lambda}}{3!}$	$\frac{\lambda^4 e^{-\lambda}}{4!}$	…

In general, if X is a Poisson distribution, then;

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!} \quad (x=0,1,2,....)$$

with X ~ Po(λ)

Poisson Distribution arises when:

- independent
- the probability of an occurrence of two or more appear simultaneously = 0
- random in time and space
- uniform

The survey was conducted by recording public transportation service through one particular place. The numbers of passenger per hour become a random variable probability distribution, but data retrieval which was done every 1 hour will have many possibilities that made a binomial distribution cannot be used here if we consider the possibility of X (the number of passengers per hour) to have no limits. ‘X’ in here (Panghegar – Dipatiukur angkot that passes through one particular point would be calculated based on the number of passengers).

Table 2
Number of Passengers to the Frequency

Number of passengers per hour (X)	Frequency (f)
0	0.7
1	3.4
2	8.4
3	14
4	17.6
5	17.6
6	14.6
7	10.4
8	6.5
9	3.6
10	1.8
11	0.8
12	0.4
13	0.1
14	0.1
>15	0

Table 2 shows the comparison between the number of passengers for every hour. The random variables calculation (using data in Table 2) is shown in Section 4.

Table 3
Angkot Route Data

Routes / Decision Variables	Panghegar - Dipatiukur	Sadang Serang – Gede Bage	Karang Setra – Cibaduyut
The average result (People/Car)	25	30	20
Gasoline cost (IDR/Car)	150.000	230.000	260.000
Selling price (IDR/People)	2.000	2.000	2.000
Minimum passenger (People)	200	240	180
Car Numbers 50			

III. Stochastic Programming Model To Decide Number Of Vehicles

The idea of this method is to determine how many angkot for each route that should operate in one day. In this research, three routes were used for the examination (Table 3). These are Panghegar-Dipatiukur, Sadang Serang – Gede Bage, and Karang Setra – Cibaduyut routes.

Suppose the data for the angkot management provided by Table 3, then:

Decision Variables:

X1 = Number of Panghegar – Dipatiukur angkot
X2 = Number of Sadang Serang - Gede Bage angkot
X3 = Number of Karang Setra – Cibaduyut angkot
w1 = Number of Passengers who pay in Panghegar – Dipatiukur route
w2 = Number of Passengers who pay in Sadang Serang - Gede Bage route
w3 = Number of Passengers who pay in Karang Setra – Cibaduyut route

This problem can be formulated into a linear programming model (deterministic) by minimizing:

$$150.000x_1 + 230.000\ x_2 + 260.000\ x_3 - 2.000w_1 - 2.000w_2 - 2.000w_3$$

Boundary:

$$x_1 + x_2 + x_3 \le 50$$
$$x_1,\ x_2,\ x_3 \ge 0$$
$$25x_1 \ge 200$$
$$30x_2 \ge 240$$
$$20x_3 \ge 180$$
$$W_1,\ w_2,\ w_3 \ge 0$$

Basically, the model has met the expectation to decide the number of angkot in the route. However, such a model can be valid if there are no other things that affect it, for example, the weather. Here are three weather scenarios that can happen:

1. Good weather: raise 20%
2. Average weather: constant
3. Bad weather: drop 20%

Each scenario has the same opportunity that is 1/3. Following is the model with the scenario:

$$\begin{aligned} &\text{Min } 150.000x_1 + 230.000x_2 + 260.000x_3 \\ &+ \tfrac{1}{3}\,(\text{-}2.000w_{11} - 2.000w_{21} - 2.000w_{31}) \\ &+ \tfrac{1}{3}\,(\text{-}2.000w_{21} - 2.000w_{21} - 2.000w_{31}) \\ &+ \tfrac{1}{3}\,(\text{-}2.000w_{11} - 2.000w_{21} - 2.000w_{31}) \end{aligned}$$

Here is the stochastic programming model with resource:

$$\min\ 150.000x_1 + 230.000x_2 + 260.000x_3 + \textstyle\sum_1^3 P(s)\,(-2.000w_1(s) - 2.000w_2(s) - 2.000w_3(s))$$

Boundary

$$x_1 + x_2 + x_3 \le 50$$
$$x_1,\ x_2,\ x_3 \ge 0$$
$$\varepsilon_1(s)x_1 \ge 200$$
$$\varepsilon_2(s)x_2 \ge 240$$
$$\varepsilon_3(s)x_3 \ge 180$$
$$w_{1,2,3}(s) \ge 0$$

s-scenario

$$P(s) = \tfrac{1}{3},\ s = 1,2,3$$

$$\begin{bmatrix} \varepsilon_1(1) & \varepsilon_2(2) & \varepsilon_3(1) \\ \varepsilon_1(2) & \varepsilon_2(2) & \varepsilon_3(2) \\ \varepsilon_1(3) & \varepsilon_2(3) & \varepsilon_3(3) \end{bmatrix} = \begin{bmatrix} 30 & 36 & 24 \\ 25 & 30 & 20 \\ 20 & 24 & 16 \end{bmatrix} = random\ matrix$$

therefore

$$\underbrace{\min\ 150.000x_1 + 230.000x_2 + 260.000x_3\ +}_{\text{deterministic part (phase I)}}$$

$$\underbrace{\textstyle\sum_1^3 P(s)(-2.000w_1(s) - 2.000w_2(s) - 2.000w_3(s))}_{\text{stochastic part (phase II)}}$$

Boundary

$$\left.\begin{aligned} &x_1 + x_2 + x_3 \le 50 \\ &x_1,\ x_2,\ x_3 \ge 0 \end{aligned}\right\}\ \text{deterministic boundary (phase I)}$$

$$\left.\begin{aligned} &\varepsilon_1(s)x_1 \ge 200 \\ &\varepsilon_2(s)x_2 \ge 240 \\ &\varepsilon_3(s)x_3 \ge 180 \\ &w_{1,2,3}(s) \ge 0 \end{aligned}\right\}\ \text{stochastic boundary (phase II)}$$

From this model the resource function can be written as:

$$Q(x_1,\ x_2,\ x_3,\ s) = -\ 2.000w_1(s) - 2.000w_2(s) - 2.000w_3(s)$$

Boundary

$$\varepsilon_1(s)x_1 \ge 200$$
$$\varepsilon_2(s)x_2 \ge 240$$
$$\varepsilon_3(s)x_3 \ge 180$$
$$w_{1,2,3}(s) \ge 0$$

The expectation value from the resource function:

$$Q(x) = E_3Q(x,\ \varepsilon) = \textstyle\sum_1^3 P(s)\ Q(x_1,\ x_2,\ x_3,\ s)$$

Therefore, the model resource becomes:

$$\min\ 150.000x_1 + 230.000x_2 + 260.000x_3 + E_3Q(x,\ \varepsilon)$$

Boundary:

$$x_1 + x_2 + x_3 \le 50$$
$$x_1,\ x_2,\ x_3 \ge 0$$

In general, two phases resource model could be written:

$$\min c^T x + Q(x)$$

Boundary:

$$Ax = b,\ x \geq 0$$

More generally this resource model could be in the form

$$\min f_1(x) + E\varepsilon[Q(x, \varepsilon)]$$

Boundary:

$$Ax = b,\ x \geq 0$$

is repeated until the overall expected cost becomes optimal.

IV. EXPERIMENTAL RESULT

In this section, the random variable calculation from both models will be performed. For the Poisson distribution process, the data from Table 3 can be presented as shown in Figure 4.

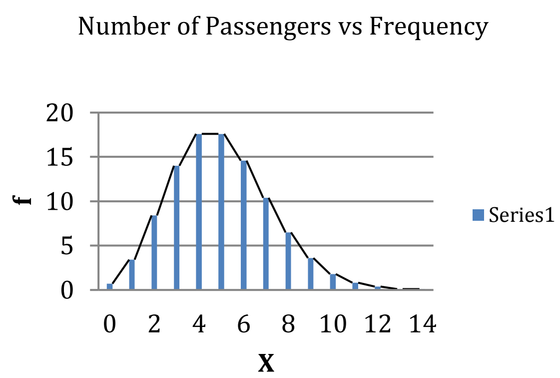

Figure 4: Graph number of passengers vs frequency

From the graph we could do a calculation for:

1. The mean value of the data:

$$\bar{x} = \frac{0 \times 0.7 + 1 \times 3.4 + \ldots + 1.4 \times 1}{100} \approx 4.997$$

2. While the value of variance is:

$$s^2 = \frac{0^2 \times 0.7 + 1^2 \times 3.4 + \ldots + 14^2 \times 0.1}{100} - 4.997^2 \approx 5.013$$

3. By allowing a small error value at the sampling process, we can see that the mean value and the variance value are almost identic in distribution above. The relationship between each frequency above can be seen by performing division between one frequency with frequency afterwards respectively:

$\frac{3.4}{0.7} \approx \frac{5}{1}$	$\frac{8.4}{3.4} \approx \frac{5}{2}$	$\frac{14}{8.4} = \frac{5}{3}$	$\frac{17.6}{14} \approx \frac{5}{4}$
$\frac{17.6}{17.6} \approx \frac{5}{5}$	$\frac{14.6}{17.6} \approx \frac{5}{6}$	$\frac{10.4}{14.6} = \frac{5}{7}$	$\frac{6.5}{10.4} \approx \frac{5}{8}$
$\frac{3.6}{6.5} \approx \frac{5}{9}$	$\frac{1.8}{3.6} \approx \frac{5}{10}$		

4. Initial probability, that is P(X=0) = 0.007, can be used to calculate another probability:

- $P(X=1) = \frac{5}{1} P(X=0)$
- $P(X=2) = \frac{5}{2} P(X=1) = \frac{5^2}{2 \times 1} P(X=0)$
- $P(X=3) = \frac{5}{3} P(X=2) = \frac{5^3}{3 \times 2 \times 1} P(X=0)$
- $P(X=4) = \frac{5}{4} P(X=3) = \frac{5^4}{4 \times 3 \times 2 \times 1} P(X=0)$
- Etc.

5. Thus, the probability distribution can be written as follows (in the form of *generating function*):

$$P(X=n) = \frac{5^n}{n(n-1)\ldots.2 \times 1} P(X=0)$$

$$P(X=n) = \frac{5^n}{n!} P(X=0)$$

Because of the sum of all probabilities = 1, then:

$$1 = p + 5p + \frac{5^2 p}{2!} + \frac{5^3 p}{3!} + \frac{5^4 p}{4!} + \ldots..$$

$$1 = p\left(1 + 5 + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \ldots..\right)$$

$$1 = pe^5,\ dengan\ \ e^5 = 1 + 5 + \frac{5^2}{2!} + \ldots.$$

$$p = e^{-5}$$

with e = 2.71828..

Exponential function when translated into the form of series can be written as follows:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Based on the calculations above, x = 5. So daily gross income of angkot driver on average travelling distance and maximum number of passenger that appearing in a place is five people. This is shown in Table 4.

Table 4
Distance to Income

Average Travel Distance (km)	Daily Gross Income (rupiah)
1	109.200 ± 5.856
2	68.586 ± 5.710
3	56.450 ± 3.998
4	44.780 ± 3.184
5	39.603 ± 2.300
6	35.146 ± 1.340
7	30.683 ± 1.033
8	29.997 ± 1.606
9	26.320 ± 763
10	25.789 ± 1.094

V. CONCLUSION

From the probability distribution calculation above by using the Poisson distribution, the random variable for the number of passengers within a specified time interval could be known.

For the stochastic programming model, the expected value from the objective value of the second phase is a resource. In the first stage, a decision was made based on any data available at the time. In the second phase, every possible implementation from a random variable ε (a new decision) taken depends on the decision in the single phase. Expectations cost in the second phase is calculated, and after that, the decision in the first stage could be revised to achieve a balance for overall costs that is better between phase 1 and phase 2. This process is repeated until the expected overall cost is optimal. Then, we could decide the number of angkot that should be operating in a day.

From these two methods, a model that could be used to predict the sufficient time to pick up the passengers and the number of vehicles (angkot) that should be run on the specific route can be made. For the future works, this model will be adapted to build an information system that provides some recommendation to help the angkot drivers and the transportation agencies make right decisions. If they could make the right decisions, then the income of angkot drivers could become stable.

ACKNOWLEDGEMENTS

This work supported by Institut Teknologi Del and Blackberry Innovation Center ITB.

REFERENCES

[1] M. Casey, S. Sen. The scenario generation algorithm for multi-stage stochastic linear programming. *Mathematics of Operations Research*, 2005. Forthcoming

[2] M. Koivu. Variance reduction in sample approximation of stochastics programs. *Mathematical Programming* 103(3):463-485, 2005.

[3] R. J.-B. Wets. Stochastic programs with fixed recourse: the equivalent deterministic program. *SIAM J. on Applied Mathematics* 17:638-663, 1969..

[4] Shapiro. Lec*tures on stochastic programming modeling and theory*. Rutgers University, New Jersey.

[5] D. L. Gerlough. *Use of Poisson Distribution in Highway Traffic. Eno Foundation for Highway Traffic Control*, 1955. The University of California

[6] I. J. Good. Some statistical applications of poisson's work. *Statistical Science*, 1(2):157, 1986.

[7] J. R. Birge, F. V. Louveaux. *Introduction to stochastic programming*. Springer, 1997.

[8] J. L. Higle, S. Sen. Statistical approximations for stochastic linear programming problems. *Annals of operations research* 85:173-192, 1999.

[9] S. Ahmed, A. J. King, G. Parija. A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. *Journal of Global Optimization* 26:3-24, 2003.

[10] A. J. Kleywegt, A. Shapiro, T.H. De Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* 12(2):479-502, 2002.