OPEN ACCESS

Saudi Journal of Business and Management Studies
Abbreviated Key Title: Saudi J Bus Manag Stud
ISSN 2415-6663 (Print) |ISSN 2415-6671 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: http://scholarsmepub.com/sjbms/

Original Research Article

Statistical Hypothesis Test on the Vessel Arrival Pattern at Hong Kong Port with Peak Time

Dao-zheng Huang*

College of Transport and Communications Shanghai Maritime University No 1550, Haigang Avenue, Pudong New Area, Shanghai, P.R. China

DOI: 10.36348/sjbms.2020.v05i01.003 | **Received:** 01.01.2020 | **Accepted:** 08.01.2020 | **Published:** 16.01.2020

***Corresponding author:** Huang Daozheng

Abstract

The vessel arrival pattern is the basis of research on the port management including berth assignment, quay crane assignment and yard operations. The hypothesis that vessel arrival pattern follows the Poisson distribution (or the inter-arrival time of two consecutive vessel arrivals follows the exponential distribution) is regularly adopted by many researchers. This paper focuses on the vessel arrival pattern at the Hong Kong Port and examines the hypothesis mentioned above based on the real data. The chi-square test method is employed to check the hypothesis under the parameter $\alpha = 0.05$. the result shows the vessel arrival pattern does not follow the Poisson distribution. Researching into the arrival data, we find that there is peak time from 8 to 9 o'clock. Considering the peak time and normal time, respectively, we find that the vessel arrival pattern at both times follows the Poisson distribution. The conclusion is tested in different data set using the chi-square test.

Keywords: Statistics, Hong Kong Port, Vessel Arrival, Poisson distribution.

INTRODUCTION

The vessel arrival pattern is a primary issue for port planning and scheduling. Furthermore, it is also a prerequisite for safety simulation in port water.

Infrastructure of the terminals at Hong Kong Port

There are currently nine container terminals situated at Kwai Chung, Stonecutters Island and Tsing. Additionally, substantial TEU throughput can be handled by the River Trade Terminal at Tuen Mun and by mid-stream.

The Kwai Tsing Container Terminals is located in the north-western part of the harbour, and it has nine container terminals with 24 berths of about 8,500 meters of frontage. The area of container yards and container freight stations covers about 2.7 km². Although 18 existing berths have a handling capacity of over 18 million TEUs, more than 20 million TEUs of container volume make it one of busiest container ports [1].

Current Vessel Traffic Condition of the Hong Kong Port

According to the statistical report in MDHKSAR [2], about 39,020 ocean vessels arrived at the Hong Kong Port and the container vessels takes up 61.7% of all the vessel arrivals and following are general cargo vessel and ferry with the proportion of 16.2% and 8.7%, respectively. The turnaround time for container vessels at Hong Kong Port is estimated at 10 hours [1].

The Hong Kong Port is one of the busiest ports in the world. The deep water and natural shelter provide ideal condition for berthing and vessel handling. In terms of container throughput, it was the busiest port in the world from 1987 to 2004 and then was overtaken by Singapore Port in 2005. Singapore Port, Hong King Port and Shanghai Port are three main container ports in Asia and the competition among them is extremely fierce.

Hong Kong Port was surpassed by Singapore Port in 2005 and was also surpassed by Shanghai Port in terms of container throughput from January 2007 to September 2007. The reasons resulting the present situation originate from two aspects. On one hand, new competitive ports such as Guangzhou Port and Shenzhen Port, which are situated near the Hong Kong and can certainly attract more and more customers via low container handling cost and high-efficient terminal operations, are quickly expanding. On the other hand, Fung *et al*. [3] indicated that high Terminal Handing Charge (THC) had lowered the container throughput at the Hong Kong Port.

Vessel Arrival Data

Data on vessel arrival time is collected from the website of the Maritime Department of Hong Kong, and the data from 1 May 2014 00:00 to 31 May 2014 24:00 are employed. It should be noted that the data given on the website merely includes ocean-going vessels. **Error! Reference source not found.** illustrates the number of vessel arrivals every day and hour. In **Error! Reference source not found.** c), Arrivals in different days of May 2014 are stable at around 70 arrivals per day, while from **Error! Reference source not found.** b), we can see that vessel arrivals in each hour have a peak at 8 o'clock. The number of vessels arrived at 8 o'clock is 232, beyond two times as large as the average number of 93.

The number of vessel arrivals in an hour is counted, and a total of 2230 samples (N) are obtained.

In Fig-1: **The number of vessel arrivals every day and hour**

Table-1, the parameter n_x denotes the number of counting periods (an hour) in which x vessels arrive at the Hong Kong Port.

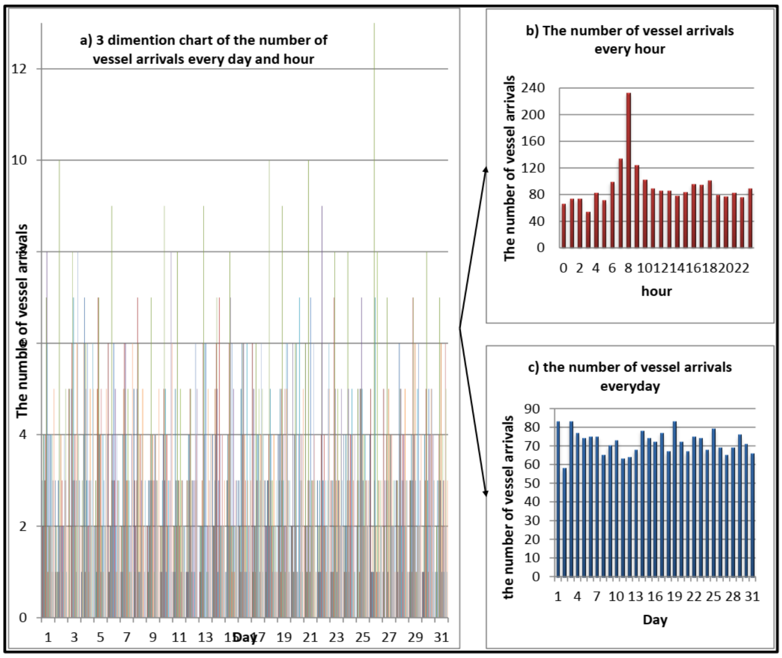

Fig-1: The number of vessel arrivals every day and hour

Table-1: Samples of vessel arrivals in May (t = 1 hour)

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	*Sum*
n_x	39	129	169	162	101	64	42	19	10	5	3	0	0	1	744

Source: MDHKSAR, 2014.

METHODOLOGY

Prerequisites

Poisson Process

It is assumed that the vessel arrival pattern at the Hong Kong Port follows the Poisson Process, and n vessels arrive at the port in counting period t with the probability P, which can be denoted:

$$P\{N(t)=x\}=e^{-\lambda t}\frac{(\lambda t)^x}{x!},\qquad x=0,1,2,\ldots \tag{1}$$

When t = 1 hour, the vessel arrival pattern follows Poisson distribution and the probability mass is shown as follows:

$$P\{N=x\}=e^{-\lambda}\frac{(\lambda)^x}{x!},\qquad x=0,1,2,\ldots \tag{2}$$

There are some assumptions for Poisson Process:

- $N(0)=0;$

- It satisfies the stationary and independent increment properties, h represents a minimal period;
- $P\{N(h)=1\}=\lambda h+o(h)$;
- $P\{N(h)=2\}=o(h)$.

Parameter Estimation

In the formula of Poisson distribution, the parameter λ, which denotes the arrival rate in one hour, controls this distribution. Generally, the parameter λ can be estimated based on the samples which have been collected. Main methods, which can be used to estimate this parameter, have point estimation and interval estimation. Maximum likelihood estimation, one kind of point estimation, is here adopted.

The $X_1, X_2, ..., X_n$ are the samples of total random variable X, and the probability of events $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$ can be written as:

$$L(\theta) = L(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} P(x_i \theta)$$

In order to fit for the value of samples i.e. we should obtain adequate parameter θ to make the events $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$ most possible and reasonable, we can use:

$\frac{\partial LnL}{\partial \theta} = 0$ and get the $\bar{\theta}$, which is called maximum likelihood value, to maximize $L(\theta)$.

As for Poisson distribution, we have

$$Ln[L(\lambda)] = Ln[L(x_1, x_2, ..., x_n; \lambda)] = Ln\left\{\prod_{i=1}^{n}\left[e^{-\lambda} \cdot \frac{\lambda^{x_i}}{(x_i)!}\right]\right\}$$

$$= -\lambda n + Ln\lambda \cdot \sum_{i=1}^{n} x_i - Ln\left[\prod_{i=1}^{n}(x_i)!\right]$$

$$\frac{\partial Ln[L(\lambda)]}{\partial \lambda} = 0$$

$$\Rightarrow -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\Rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$$

$$Var(\lambda) = \sigma^2 = \lambda = \overline{X} \quad (3)$$

Significance Level and P-Value

Significance level is always denoted with $\alpha \in (0,1)$, which is defined as the probability of making a decision to reject the null hypothesis when the null hypothesis is actually true. The reject range can be obtained based on α and the distribution of statistical variable. The smaller the parameter α is, the assumed value is preferred more. In statistical hypothesis testing, the p-value is the probability of obtaining a result at least as extreme as a given data point, under the null hypothesis. The fact that p-values are based on this assumption is crucial to their correct interpretation. We should reject the null hypothesis if P-value is smaller than α.

Chi-Square Test

We assume vessel arrival pattern at Hong Kong Port follows Poisson distribution with t=1h and parameter λ. and then, hypothesis test should be employed to examine whether the assumption is adequate or not. Chi-Square test is adopted in this research. $X_1, X_2, ..., X_n$ Is the sample of random variable X, which follows Poisson distribution under null hypothesis H_0. Here, H_0 means that vessel arrival pattern follows the Poisson distribution with the parameter λ. To carry out Chi-Square test, the samples should be divided into groups: $A_i\ (i = 1, 2, ..., k)$. According to Pearson Theorem, when n is sufficient large (larger than 50), and under null hypothesis we have $P_i = P(A_i)(i = 1, 2, ..., k)$. The statistical value can be constructed as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \sim \chi^2(k - r - 1) \quad (4)$$

Where f_i is the frequency of group i in historical data, n is the sample size, np_i is the expected frequency of estimated Poisson distribution (np_i should be more than 5, otherwise the group should be emerged into another group), and r is the number of parameters estimated by maximum likelihood estimation. Hence we reject H_0 if $\chi^2 > \chi^2_{(k-r-1;\alpha)}$

(Here, $P\{\chi^2 > \chi^2_\alpha(k-r-1)\} = \int_{\chi^2_\alpha}^{\infty} f(y)dy = \alpha$ is the significant level).

Statistical Hypothesis Test Process

According to the historical knowledge, we assume the vessel arrival pattern in Hong Kong port follows the Poisson distribution (H0). Then, we use chi-square test to test this hypothesis, if the hypothesis is accepted, the process ends. If the hypothesis is rejected, a cluster analysis will be carried out to determine the peak time and normal time according to the data characteristics in section 2. There are lots of cluster analysis methods. Here, we use K-means cluster analysis method to determine the peak time and normal time according to the arrivals in each hour. After dividing the arrival data into two groups: peak time and normal time, the hypothesis is also changed to two hypothesis: H1 (vessel arrival pattern at peak time in Hong Kong port follows the Poisson distribution) and H2 (vessel arrival pattern at normal time in Hong Kong port follows the Poisson distribution). The same test is implemented on hypothesis H1 and H2. If the hypothesizes are rejected, other distributions are suggested to fit the data. If the hypothesizes are accepted, we get the result that the vessel arrival pattern follows Poisson distribution. In order to strengthen our confidence about the conclusion, we use a new data set to test whether the result is right at other time. If the hypothesis is accepted at the new data set, we draw the conclusion that vessel arrival pattern at Hong Kong port follows Poisson distribution.

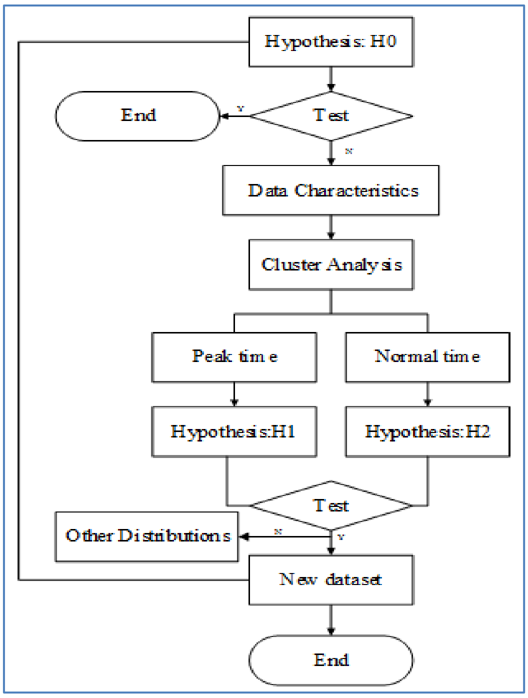

Fig-2: Flowchart of statistical hypothesis test

Note:	H0 means that vessel arrival pattern in Hong Kong port follows the Poisson distribution;
	H1 means that vessel arrival pattern at peak time in Hong Kong port follows the Poisson distribution;
	H2 means that vessel arrival pattern at normal time in Hong Kong port follows the Poisson distribution.

RESULT AND DISCUSSION

The Chi-Square Test of Vessel Arrival Pattern in May

According to the methodology, we can get the average vessel arrivals in May 2014:

$$\overline{X}_5 = \frac{1}{744}(0\times 39 + 1\times 129 + ... + 1\times 13) = 2.9973$$

Form the parameter estimation derived above, $\lambda = \overline{X}_5 = 2.9973$ and n = 744. The test result is shown in table 2.

Table-2: Result of Chi Square Test of Vessel Arrival Pattern in May

$x = i$	f_i	$P_i = \frac{2.9973^i}{i!} e^{-2.9973}$	nP_i
0	39	0.0499	37.1413
1	129	0.1496	111.324
2	169	0.2242	166.8364
3	162	0.2240	166.6869
4	101	0.1679	124.9032
5	64	0.1006	74.8747
6	42	0.0503	37.4038
7	19	0.0215	16.0158
8 and More	19	0.0118	8.8113

Source: compiled based on the data collected from MDHKSAR May, 2014.

The freedom degree is 7. $\chi^2 = \sum_{i=0}^{k} \frac{(f_i - np_i)^2}{np_i} = 22.1155 > \chi^2_{0.05}(7) = 14.07$. We should reject the hypothesis H0 that vessel arrival pattern follows Poisson distribution.

The Chi-Square Test of Vessel Arrival Pattern with Peak Time in May

We use the function *k-means in* Matlab to determine the peak time and normal time according to the total number of arrivals in each hour of May. The cluster analysis shows that 8 o'clock is the peak time and other hours are normal time. Then, the test results of H1 and H2 are got. As for H1, owing to that the observations of 8 to 9 o'clock in May is 31, which is less than 50; we can't use chi-square test directly. The new data set of 19 days data from April is used to improve the observations. So we test the normal time first.

The Chi-Square Test of Vessel Arrival Pattern at Normal Time of May

The same process is carried out on arrival data at normal time of May as shown in table 3.

Table-3: Samples of vessel arrivals at normal time of May (t = 1 hour)

x	0	1	2	3	4	5	6	7	8	*Sum*
n_x	39	129	169	162	101	60	36	12	4	713

Source: MDHKSAR 2014.

According to the methodology, we can get the average vessel arrivals at normal time of May, 2014: $\overline{X}_{5nor} = \frac{1}{713}(0\times 39 + 1\times 129 + ... + 8\times 4) = 2.8002$

Form the parameter estimation derived above, $\lambda = \overline{X}_{5nor} = 2.8002$ and n = 713. The result is shown in table 4.

Table-4: Result of Chi-Square Test of Vessel Arrival Pattern in May

$x = i$	f_i	$P_i = \frac{2.8002^i}{i!} e^{-2.8002}$	nP_i
0	39	0.0607	43.2604
1	129	0.1700	121.2262
2	169	0.2382	169.8526
3	162	0.2225	158.6562
4	101	0.1559	111.1483
5	60	0.0874	62.293
6	36	0.0408	29.0933
7	12	0.0163	11.6467
8 and More	5	0.0075	5.3498

Source: compiled based on the data collected from MDHKSAR May 2014.

The freedom degree equals to 9-1-1=7. $\chi^2 = \sum_{i=0}^{k} \frac{(f_i - np_i)^2}{np_i} = 3.6770 < \chi^2_{0.05}(7) = 14.07$. We should accept the hypothesis H2 that vessel arrival pattern at normal time follows Poisson distribution.

The Chi-Square Test of Vessel Arrival Pattern at Peak Time

The same process is carried out on arrival data at peak time of May as shown in table 5.

Table-5: Samples of vessel arrivals at 8 o'clock of May and 18 days of April (t = 1 hour)

x	3	4	5	6	7	8	9	10	11	12	13	*Sum*
n_x	1	1	7	8	11	11	5	4	1	0	1	50

Source: MDHKSAR, 2014.

According to the methodology, we can get the average vessel arrivals at peak time of May, 2014: $\overline{X}_{peak} = \frac{1}{50}(3\times 1 + 4\times 1 + ... + 13\times 1) = 7.28$

Form the parameter estimation derived above, $\lambda = \overline{X}_{peak} = 7.28$ and n = 50. The result is shown in table 6.

Table-6: Result of Chi Square Test of Vessel Arrival Pattern in May

$x = i$	f_i	$P_i = \frac{7.28^i}{i!} e^{-7.28}$	nP_i
3 and 4	2	0.1250	6.2488
5	7	0.1174	5.8719
6	8	0.1425	7.1246
7	11	0.1482	7.4096
8	11	0.1349	6.7427
9	5	0.1091	5.4541
10 and more	6	0.1817	9.0854

Source: compiled based on the data collected from MDHKSAR May 2014.

The freedom degree equals to 7-1-1=5. $\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} = 8.7265 < \chi^2_{0.05}(5) = 11.07$.

We should accept the hypothesis H1 that vessel arrival pattern at peak time follows Poisson distribution.

We get the conclusion that the total vessel arrival pattern at Hong Kong port does not follow Poisson distribution because there is a peak time at 8 to 9 o'clock in the morning. When we consider the vessel arrival pattern at peak time and normal time respectively, the vessel arrival pattern follows the Poisson distribution.

Verification of the Result by Vessel Arrival Data in April

To improve our confidence with this conclusion, we use a new data set from 7th April to 25th April to verify our conclusion.

The Chi-Square Test of Vessel Arrival Pattern in April

The same process is carried out on arrival data at April as shown in table 7.

Table-7: Samples of vessel arrivals at of April (t = 1 hour)

x	0	1	2	3	4	5	6	7	8	9	10	11	*Sum*
n_x	31	72	101	90	77	39	20	12	9	2	1	2	456

Source: MDHKSAR 2014.

According to the methodology, we can get the average vessel arrivals at normal time of 6th April to 24th April, 2014: $\overline{X}_4 = \frac{1}{456}(0\times31+1\times72+...+11\times2) = 3.0110$

Form the parameter estimation derived above, $\lambda = \overline{X}_4 = 3.0110$ and n = 456. The result is shown in table 8.

Table-8: Result of Chi Square Test of Vessel Arrival Pattern by April data

$x = i$	f_i	$P_i = \frac{2.8002^i}{i!} e^{-2.8002}$	nP_i
0	31	0.0492	22.4553
1	72	0.1483	67.6122
2	101	0.2232	101.789
3	90	0.2240	102.161
4	77	0.1686	76.9008
5	39	0.1016	46.3091
6	20	0.0510	23.2392
7	12	0.0219	9.9961
8 and More	14	0.0121	5.5036

Source: compiled based on the data collected from MDHKSAR May 2014.

The freedom degree equals to 9-1-1=7. $\chi^2 = \sum_{i=0}^{k} \frac{(f_i - np_i)^2}{np_i} = 20.1136 > \chi^2_{0.05}(7) = 14.07$. We should reject the hypothesis H0 that vessel arrival pattern follows Poisson distribution.

The Chi-Square Test of Vessel Arrival Pattern at Normal Time in April

Table-9: Samples of vessel arrivals at normal time of April (t = 1 hour)

x	0	1	2	3	4	5	6	7	8	9	10	11	*Sum*
n_x	31	72	101	89	76	36	18	8	4	1	0	1	437

Source: MDHKSAR 2014.

According to the methodology, we can get the average vessel arrivals at normal time of April, 2014: $\overline{X}_{4nor} = \frac{1}{437}(0\times31+1\times72+...+11\times1) = 2.8398$

Form the parameter estimation derived above, $\lambda = \overline{X}_{4nor} = 2.8398$ and n = 437. The result is shown in table 10.

Table-10: Result of Chi Square Test of Vessel Arrival Pattern at normal time in April

$x = i$	f_i	$P_i = \frac{2.8398^i}{i!} e^{-2.8398}$	nP_i
0	31	0.0584	25.5367
1	72	0.1659	72.5195
2	101	0.2356	102.9711
3	89	0.2231	97.473
4	76	0.1584	69.2014
5	36	0.0899	39.3038
6	18	0.0426	18.6026
7 and more	14	0.0260	11.3732

Source: compiled based on the data collected from MDHKSAR April, 2014.

The freedom degree equals to 8-1-1=6. $\chi^2 = \sum_{i=0}^{k} \frac{(f_i - np_i)^2}{np_i} = 3.5187 < \chi^2_{0.05}(6) = 12.59$.

We should accept the hypothesis H2 that vessel arrival pattern at normal time follows Poisson distribution.

CONCLUSION

The vessel arrival pattern at Hong Kong Port is examined based on the real data which are collected from the website of Maritime Department of Hong Kong Port. Under the null hypothesis that arrival pattern follows the Poisson distribution, chi-square test is adopted to test the hypothesis. It is observed that there is a peak time of vessel arrival at 8 to 9 o'clock. Moreover, it is concluded that even though the total arrival pattern of vessels at Hong Kong Port does not follow the Poisson distribution under the significance level it follows Poisson distribution at peak time and normal time, respectively.

ACKNOWLEDGMENTS

The authors sincerely thank the editor and anonymous reviewers for their valuable comments that help improve this article. This research was funded by MOE (Ministry of Education in China) Liberal arts and Social Sciences Foundation (Grant no. 19YJCGJW003).

REFERENCE

1. Wikipedia. (2007). "Port of Hong Kong" <en.wikipedia.org/wiki/Port_of_Hong_Kong>. (Accessed 22 Nov 2007).
2. Maritime Department of the Hong Kong Special Administrative Region (MDHKSAR). (2007). "Publications and Press Releases" <www.mardep.gov.hk/en/publication/portstat.html > (Accessed 22 Nov 2007).
3. Fung, M.K., Cheng, L.K., Qiu, L.D. (2003). The impact of terminal handling charges on overall shipping charges: an empirical study, Transportation Research Part A, 37: 703-716.