Accuracy study of parking duration data from patrol survey

Jin Cao, IVT, ETH
Monica Menendez, IVT, ETH

Conference paper STRC 2012

STRC | 12th Swiss Transport Research Conference
Monte Verità / Ascona, May 02-04,2012

Accuracy study of parking duration data from patrol survey

Jin Cao
IVT, ETH
Zürich

Phone: +41/(0) 44 633 27 57
Fax:
email: jin.cao@ivt.baug.ethz.ch

Monica Menendez
IVT, ETH
Zürich

Phone: +41 /(0)44-633-6695
Fax:
email: Monica.menendez@ivt.baug.ethz.ch

September 2007

Abstract

Patrol survey is a widely used method for data collection in parking studies, and one of its goals is to obtain the average parking duration for a parking area. However, the inaccuracy is unavoidable; one main reason is that short stay vehicles are often not observed. To reduce the error from the survey, short observation intervals can be chosen, but that can be very expensive.

In this article, a generalized analytical model is built to analyse this error, several suggestions are made to guarantee the accuracy, and to coordinate it with the survey costs. Through dimensional analysis, the relation between time length of observation interval, estimated parking duration, and survey accuracy is illustrated. In the numerical example, a correction for the value of average duration is introduced to increase the accuracy. We further used simulation-based examples to extend the model and the conclusions to more generalized situations. The final result shows that with this approach, the accuracy of the survey could be estimated and enhanced. It helps the surveyors in choosing a proper length for the survey interval, so one could obtain high quality results from the patrol survey while keeping costs to a minimum.

Keywords

Average parking duration – parking behavior – patrol survey – data collection– parking duration accuracy

1. Introduction

As part of the increasing demand for parking, this topic has gained a lot of attention, as it influences demand, traffic performance and road safety. This study looks at the calculation of parking duration based on data from patrol survey. Patrol survey is also known as "repeat visit survey", "parking beat survey", "periodic check survey" and "fixed period sampling license plate survey". The main idea is that the patrolling observer checks the parking area at fixed time intervals and records the plate number of the car occupying each stall. One of the goals is to obtain the approximate parking duration for each car.

Since the 1950s, patrol surveys have already been used in parking data collection. Procedures for this kind of survey are recorded with detailed examples in the report of "Urban Road Traffic Surveys" (1993). As patrol surveys tend to under sample very short stay parkers, correction factors were developed (Cleveralnd 1963) to account for the bias. But according to Bonsall (1991), the results are not completely reliable. A linear correction model was built upon the parking duration derived from negative exponential distribution (Lautso 1981), but only part of the error was analysed and the improvement in terms of accuracy was not well defined. Richardson (1974) tried to find a common parking duration distribution by studying data from 8 sites in Sydney. The author claimed that it is possible to get more accurate results at a lower survey cost. Though the accuracy is uncertain, patrol surveys are still widely used around the world, even also in the 21st century. In the study of Tong & Wong (2004) on parking in Hong Kong, 72 parking lots were observed by patrol surveyors at 15-mintue intervals. In parking studies from many cities in the U.S., the survey interval is chosen between half an hour to one hour and many of the surveys last for one day.

As one would expect, in most patrol surveys, there is a tradeoff between data collection costs and study accuracy. For example, a survey with long time intervals between observations may have relatively low costs, but the data can be quite inaccurate, and could potentially generate unusable results. Evidently, spending more time and money could increase accuracy. Then the question remains: at what point does the incremental change in accuracy become too expensive?

Although the error which causes the inaccuracy is inherent to the survey, the data can be manipulated afterwards to improve the accuracy. In this article, a generalized analytical model is built to analyse this error. Then several suggestions are made to guarantee the accuracy, and to coordinate it with the survey costs. The relation between time length of observation interval, estimated parking duration, and survey accuracy is illustrated with this model both mathematically and graphically. Using probability theory, the whole process can be simulated based on certain assumptions regarding arriving time and parking duration. Most important, short-stay vehicles which can not be documented during the survey as their arrival and departure times occur within the same observation interval can still be accounted for. Through dimensional analysis, the error between the estimated parking duration and actual parking duration can be estimated. Numerical examples are provided. The final result shows that with this approach, the accuracy of the data could be guaranteed and the reliability of the survey very much enhanced. In this way, one could obtain high quality results from the patrol survey while keeping costs to a minimum.

This paper describes the causes for biased results and analyses the relation between survey input and accuracy. Section 2 is the analytical model linking the accuracy of the survey to the survey intensity, it includes 3 parts: the model assumptions, and the methodology, a numerical example. Section 3 uses simulation-based examples to extend the basic model to a more general framework with different distributions. Section 4 summarises the finding of this study.

2. Basic Model

2.1 Assumptions

In order to develop a simple analytical model, a few assumptions are initially made. The example below correspond to a parking area (on- or off-street) with no restrictions at all. All the individual parking behaviors are completely spontaneous. The parking area is open 24 hours a day and it has enough parking stalls to satisfy the demand. The cars arrive at different times, but we assume the arrival time is uniformly distributed in the interval [0,C], where C is the length of the cycle (e.g., 1 day). Hence, the probability density function of arrival time is $f(t_a)=\frac{1}{C}$. We are assuming that the survey is only finished when all the cars which arrived within [0,C] leave.

The shortest parking duration across all vehicles is t_s^{min} , and the longest is t_s^{max}. The parking duration is uniformly distributed between these two unknown values, so the real average parking duration is $\overline{T}^{real} = \frac{t_s^{min}+t_s^{max}}{2}$. For a given parked vehicle, denote t_a as the arrival time, t_d as the departure time, and $t_s = t_d - t_a$ as the parking duration. The probability density function of parking duration is $f(t_s) = \frac{1}{t_s^{max} - t_s^{min}}$.

2.2 Model

Denote δ as the survey interval, so the patrolling observer checks the parking area every δ time units and records the parking information for every stall. The percentage of all the parked vehicles that are observed by the patrolling officer throughout the survey is p^{obs}, with $p^{lost} = 1 - p^{obs}$ corresponding to the percentage of the parked vehicles that are not observed (i.e., lost). We define $a = \frac{t_s^{min}}{\delta}$, $b = \frac{t_s^{max}}{\delta}$ and $M = \frac{C}{\delta}$. Then, according to our assumptions, a vehicle can be observed i times, $i \epsilon [\lfloor a \rfloor, \lceil b \rceil]$.

Typically in patrol surveys, one assumes that the parking duration of a vehicle that has been observed i times is $i\delta$. This assumption together with the missing data for short stay vehicles are the two causes for biased data. The estimated average parking duration $\widetilde{T}^{obs}$ can be interpreted as

$$\tilde{T}^{obs} = \delta \cdot \sum_{i=\lfloor a \rfloor}^{\lceil b \rceil} i \cdot \left(\frac{p_i}{p^{obs}}\right) = \frac{a^2(\beta^2 - 1)}{2\beta a - 1 - a^2}\delta \qquad \text{eq. 1}$$

Where

$\beta = \frac{t_s^{max}}{t_s^{min}}$.

p_i is the probability of a car being observed i times.

$$p^{obs} = \sum_{m=1}^{M}\left\{\underbrace{\int_{(m-1)\,\delta}^{m\,\delta - t_s^{min}} f(t_a)\left[\int_{m\,\delta - t_a}^{t_s^{max}} f(t_s)dt_s\right]dt_a}_{\text{Term 1}} + \underbrace{\int_{m\,\delta - t_s^{min}}^{m\,\delta} f(t_a)\left[\int_{t_s^{min}}^{t_s^{max}} f(t_s)dt_s\right]dt_a}_{\text{term 2}}\right\} \qquad \text{eq. 2}$$

In eq.2, “$m\delta - t_s^{min}$” is the moment corresponding to the time of current checking interval minus the minimum parking duration. Term 1 covers not only the probability that vehicles arrive before this moment but also that they stay long enough to be observed. Term 2 covers the probability of vehicles arriving afterwards, they can all be observed as the parking duration is always longer than t_s^{min}.

To further simplify the process, we employ two dimensionless variables, X and Y

$$X = \frac{\tilde{T}^{obs}}{\delta} \qquad \text{eq. 3}$$

$$Y = \frac{\overline{T}^{real}}{\tilde{T}^{obs}} \qquad \text{eq. 4}$$

Recall that $\overline{T}^{real}$ is the real average parking duration, it equals to $\frac{a+b}{2} \cdot \delta$.

In eq.3, X can be interpreted as the “survey intensity”, so a higher X means a more intensive survey (i.e., with shorter observation intervals). $X \epsilon [1, \infty]$, as the estimated average duration $\tilde{T}^{obs}$ is never smaller than one interval. Evidently X is known from the survey. In practice, surveyors normally evaluate the survey by checking the value of $\frac{1}{X}$, the results are accepted only if this value is below 0.5.

In eq.4, Y can be interpreted as the survey accuracy. The closer Y is to 1, the more accurate the survey is.

When $\delta \in [t_s^{min}, t_s^{max}]$, the relation between X and Y can be calculated as follows:

$$Y = \frac{1+\beta}{2} \cdot a \cdot \frac{1}{X} \qquad \text{eq. 5}$$

Where $a = \frac{t_s^{min}}{\delta} = \frac{1}{\beta - \sqrt{(\beta^2-1)(1-\frac{1}{X})}}$, could be obtained from eq.1 and eq.3.

Figure 1 The correlation between accuracy and survey intensity

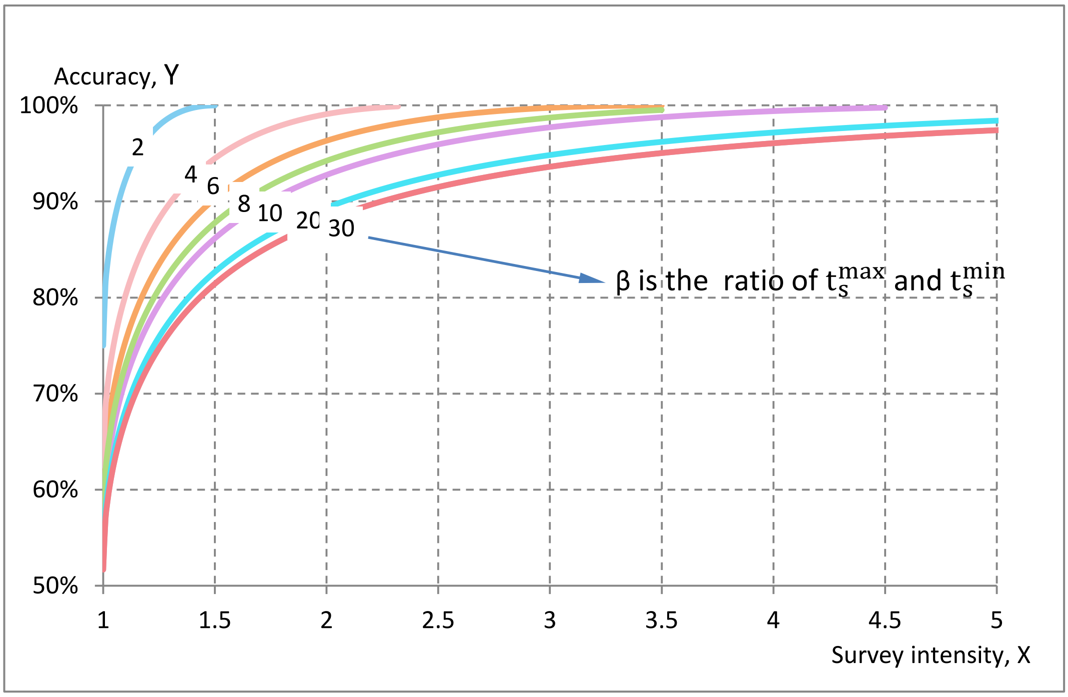


Figure 1 illustrates the results from eq.5 (i.e., the relation between accuracy and survey intensity as a function of β). Below are some conclusions that can be drawn from this graph and the model.

i. As expected, a shorter checking interval or more intensive survey (i.e., higher X) ensures higher accuracy (i.e., higher Y).

ii. For a given β, the value of X varies from 1 to $\frac{1+\beta}{2}$, a higher X corresponds to a lower δ (i.e., X=1 when $\delta = t_s^{max}$ and $X = \frac{1+\beta}{2}$ when $\delta = t_s^{min}$).

iii. The average parking duration is always overestimated, the minimum value of the accuracy is $\frac{1}{2} + \frac{1}{2\beta}$. This means the accuracy will never drop below 50% even if the observation interval, δ, is equal to t_s^{max}.

iv. A basic range of β could be obtained in a survey by approximating the longest and shortest parking duration. For a given survey intensity X, the accuracy decreases as β increases. Hence, one can find the lowest possible accuracy for a given range of β.

Note that the conclusions above only hold when $\delta \in [t_s^{min}, t_s^{max}]$. Fortunately, it is possible to verify this condition using the same survey data:

- When X=1, then $\delta \geq t_s^{max}$ and clearly the accuracy is low.
- When the percentage of cars being observed only once (i.e., $\frac{p_1}{p^{obs}}$) is much lower than the percentage of cars being observed twice or more times, then $\delta \approx t_s^{min}$ and it's possible to reduce the survey intensity without losing much accuracy.
- When all the observed cars are observed at least twice, then $\delta \leq t_s^{min}$ and the survey intensity is simply higher than needed, survey cost could be reduced by extending the observation interval.

2.3 Numerical example

The following example will be used not only to validate the proposed model, but also to illustrate how it can be used to our advantage.

In the example, a two-day survey will be used to estimate the average parking duration of vehicles, arriving with a uniform distribution between 8am and 6pm. The parking duration of all the vehicles is also known to be uniformly distributed between t_s^{min} and t_s^{max}.

According to the land use and function of the area, we can infer a general range of t_s^{min} and t_s^{max} and in order to decide an appropriate value for the observed interval. Here we conjecture that $t_s^{min} \epsilon [0,2.5]$ and $t_s^{max} \epsilon [7,12]$. We can choose $\delta_1 = 3\text{h}$, then the value of $\delta_2 = 6$ is also contain in the survey, same as $\delta_3 = 9$.

Throughout the 2 days, the patrol observer records 271 vehicles in the survey, 38% of those are observed once, 45% twice, and 17% three times. We can now estimate average duration and survey intensity using eq.1 and eq.3:

$\widetilde{T}_1^{obs} = \delta_1 \cdot \sum_{i=\lfloor a \rfloor}^{\lceil b \rceil} i \cdot \left(\frac{p_i}{p^{obs}}\right)$=3· (1·38%+2·45%+3·17%)=5.37; X_1=1.79.

Similarly, we can find results for the different values of δ, as shown in table 1.

Table 1 Survey results of the numerical example

No.	δ	No. of vehicles observed	Percentage/times			$\tilde{T}^{obs}$	X
			1	2	3		
1	$\delta_1 = 3$	271	38%	45%	17%	5.4	1.79
2	$\delta_2 = 6$	199	92%	8%	-	6.5	1.08
3	$\delta_3 = 9$	164	100%	-	-	9	1

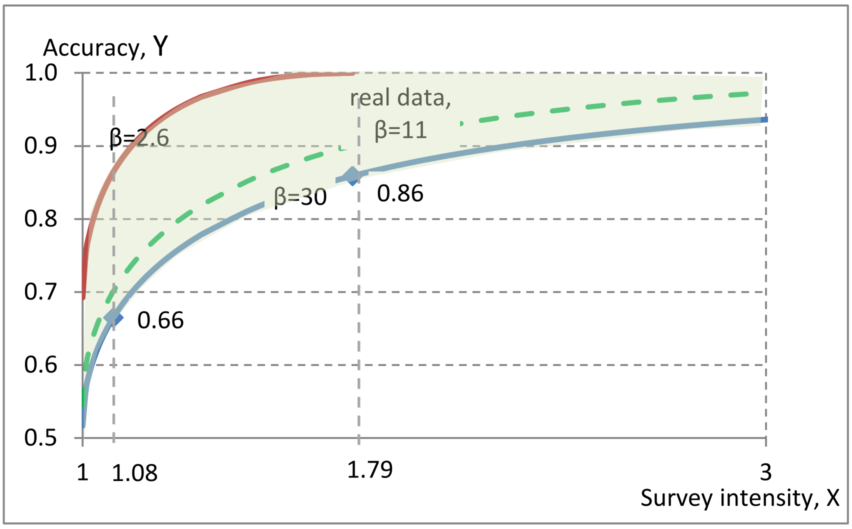
First we need to ensure the results are valid for our model, in other words, to check if $\delta \in [t_s^{min}, t_s^{max}]$ is true for each of the δ value.

For δ_1=3, the percentage of vehicles being observed once is still quite high(38%) compare to the percentage of cars being observed twice (45%) or more times, so $t_s^{min} < \delta$ (i.e., $t_s^{min} < 3$). In addition, some vehicles are observed more than 2 times, but no vehicles is observed 4 times, indicating that $2\delta < t_s^{max} < 4\delta$ (i.e., $6 < t_s^{max} < 12$). In summary, $\delta_1 \in [t_s^{min}, t_s^{max}]$. Using the same analyze, we can see that the same is true for δ_2; However, for δ_3 the result is not available for the model because X_3=1 which means δ_3 is not less than t_s^{max}, so $t_s^{max} \leq 9$.

Since the results from δ_1 and δ_2 are valid for the model, $t_s^{max} \in [6,9]$. For the minimum parking duration in this example, we will assume based on real life parking experiences that a person only parks the car when he/she needs to stay for more than 0.3 hours. In other words, a parking maneuver is worth to be conducted only when the parking duration is longer than 0.3 hours, then we have $t_s^{min} \in [0.3,3]$.

As $X \epsilon \left[1, \frac{1+\beta}{2}\right]$, a lower bound of β =2.58 could be calculated by $X_1 < \frac{1+\beta}{2}$, and the upper bound of β=30 can be calculated by $\frac{t_s^{max}}{t_s^{min}}$. So β ϵ[2.58,30], then we can find both the lower and upper bounds for the accuracy of the survey as a function of X.

Figure 2 Survey accuracy of the numerical example

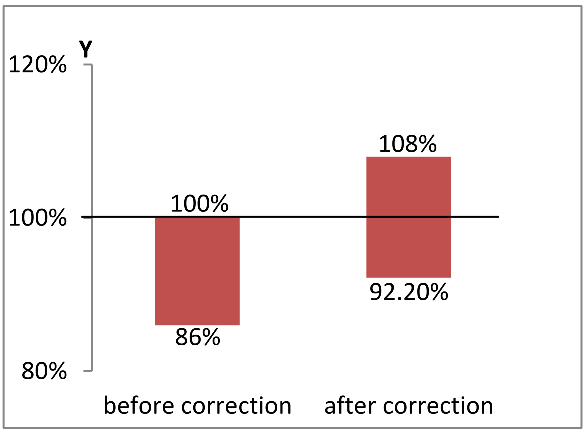


Seen in figure 2, when patrolling observers check the parking stalls at 3-hour intervals, X=1.79 and the accuracy is above 86%. As the estimated average duration is 5.4 hours, the real average duration would be higher than 4.64 hours (86% · 5.4), so $\widetilde{T}^{obs}$ =5.4 and $\overline{T}^{real}$ ∈[4.64, 5.4]. Denote μ as relative error, the range of μ is between 0 to 16.4% according to eq.6.

$$\mu = \frac{\left|\widetilde{T}^{obs} - \overline{T}^{real}\right|}{\overline{T}^{real}} = \left|\frac{1}{Y} - 1\right| \qquad \text{eq. 6}$$

We can adjust the average duration to the interval of 5.00 hours. By this correction, μ is controlled below 7.8% instead of 16.4%. Seen in figure 3, the range of a possible accuracy (i.e., $Y=\frac{\overline{T}^{real}}{\widetilde{T}^{obs}}$) is between 92.2% and 108%, the chance of a low accuracy (i.e., 86% to 92%) is eliminated. Since the real underlying data of $\overline{T}^{real}$ is 4.8h in this example, the accuracy of the survey increased by 7% (from 89% to 96%) with this approach.

Figure 3 The range of accuracy before and after the correction of $\widetilde{T}^{obs}$



(The underlying data for this example was: t_s^{min}=0.8h, t_s^{max}=8.8h, $\overline{T}^{real}$=4.8h. The survey accuracy for δ=3 was 89%. During the survey, the real amount of vehicles that parked in the area was 300.)

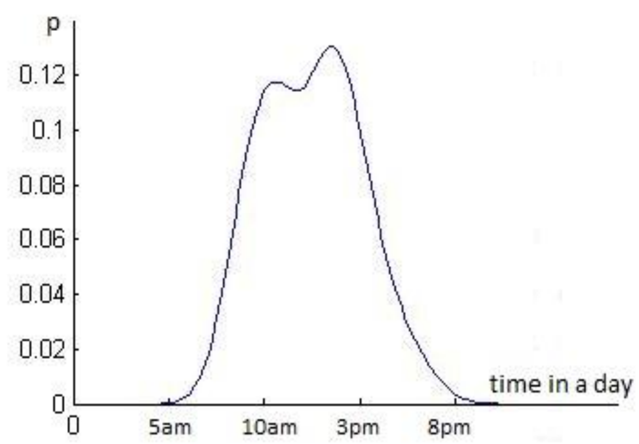
3. Extended Models

In the previous model, we assumed a uniform distribution both for the arrivals and for the parking duration in order to obtain analytical answers. However in real life we expect the parking duration to follow different distributions. In this section we will test if some of the results from the previous section can be extended to other distributions. Below we will show simulation-based examples using gamma distribution.

Gamma distribution is denoted by $G(k, \theta)$, the probability density function (PDF) is $f(x;\ k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$.

Evidently the profile of arrival times could be very different across parking areas considering the location, the function of the neighborhood and the purpose of parking. Here, we assume 500 vehicles arrive the parking area during a day following a continuous double peak distribution. The probability density function of the arrival time is showing in figure 4. This distribution can be generated by combining two gamma distributions, one is based on 6am and the other is based on 10am (k =7, θ =0.6). When the parking duration obeys a Gamma distribution, it can be proved that the mean of the parking duration, $\overline{T}^{real}$, is equal to the value of $k \cdot \theta$. We assume this value is between 10 min to 8 hours. This assumption is typically suitable for most parking areas except long-term parking facilities (e.g., facilities supplied for residences/ airports, etc.).

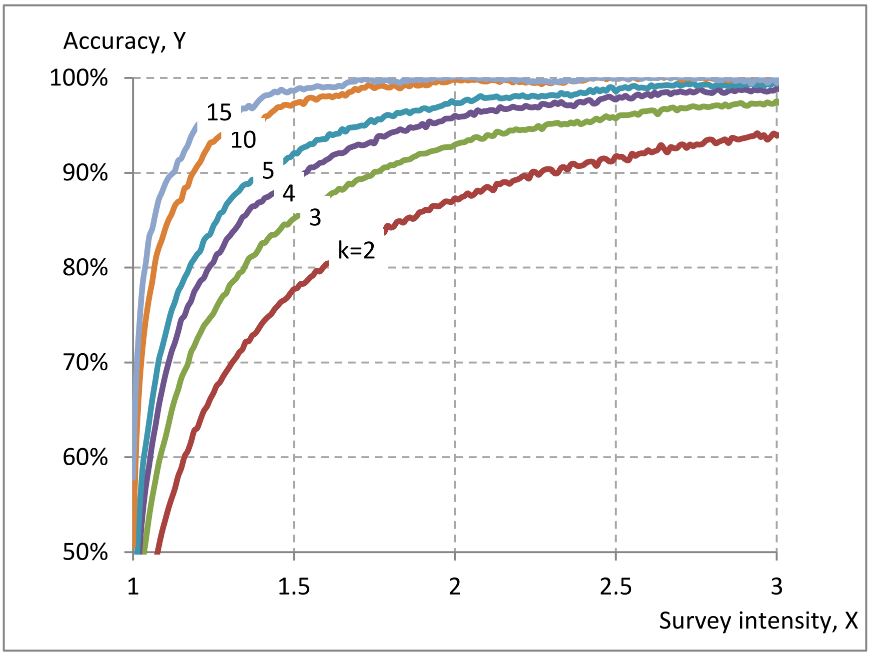
Figure 4 The probability density function of the arrival time



For a given k, θ is defined by $\frac{\overline{T}^{real}}{k}$ and $\overline{T}^{real}$ is chosen based on the vector [10, 30, 120, 210, 300, 390, 480] (unit: minutes) . For each value of k among 2, 3, 4, 5, 10, 15, 3500 simulations have been ran.

The relation between average accuracy and survey intensity based on the data obtained is shown in figure 5.

Figure 5 Survey accuracy for simulation-based examples



The results are quite similar with the ones from the basic model:

- The real average duration is typically overestimated. (i.e., $\overline{T}^{real} < \widetilde{T}^{obs}$)
- For a given X, the accuracy is very much influenced by k, a parameter from the gamma distribution.

In the previous section, $\beta = \frac{t_s^{max}}{t_s^{min}}$ was the decisive factor of the accuracy (besides X). By analyzing the PDF of uniform and gamma distributions ($f(t_s) = \frac{1}{t_s^{max} - t_s^{min}} = \frac{1}{t_s^{min}(\beta-1)}$ and $f(x;\ k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$), we can see the similarity: k and β are both "shape parameters" while t_s^{min} and θ are used to decide the scale or location of the PDF in the coordinate system.

In our model, X stands for the dimensionless value "survey intensity" instead of observation interval. Hence, the exact values of parking duration are not as decisive to the relation between Y and X as the "shape parameter". For instance, for a given X, it does not matter to the accuracy if the value of t_s^{min} is 10 min or one hour when β stays the same. A similar result is found with the Gamma distribution.

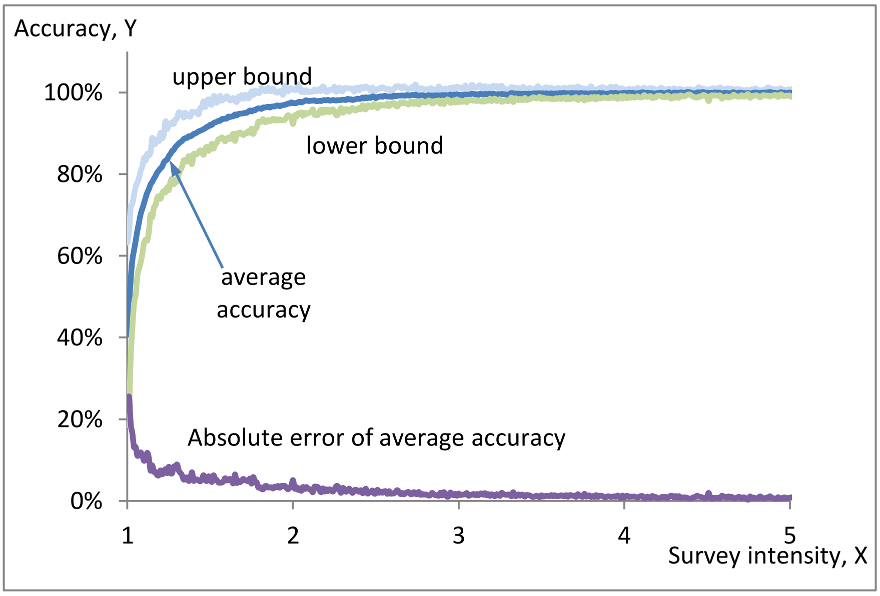
Other things to notice from figure 5:

- For a given X, a higher accuracy correspondences to a larger k, but in the basic model it is the opposite: a higher accuracy correspondences to a smaller β, even though they are both "shape parameters". The main reason is the different shape of the PDF between a uniform distribution and a Gamma distribution.
- With a higher value of k (e.g., k=15), the parking duration tends to be longer. However the survey lasts only for 1 day. This causes the fluctuations in the accuracy shown in the curve.

Evidently, the accuracy obtained is not always a same value as variations typically exist among both simulations and real surveys. In other words, when survey intensity X is known in practice even for a given k, the accuracy could be any value among a certain range instead of a precise value. As an example, we take k=5 to show this range of variation. The absolute error of average accuracy is up to 30% in the simulation, but it becomes quite small (below 8%) when X is above 1.5. We have tested this range for each k in the simulation, and the results are consistent across the different k values. In other words, for all tested k values, the absolute error of average accuracy is below 8% for X>1.5.

Figure 6 shows the upper, lower and average accuracy values for k=5. We focus more on the lower bound as low accuracy is what we are trying to prevent. Shown in the figure 6, with a smaller X, the accuracy is lower and it can further drop to a much lower accuracy caused by natural variations.

Figure 6 Accuracy fluctuation caused by natural variations



4. Conclusion

Patrol surveys are supported by new technology nowadays (Bonsall,1997), but the idea hasn't change much and the bias of the data still exists. The data collected are often used to learn about the parking demand in an area, or to support local government for policy developing. It is also used to find the proportion of overtime parkers. Moreover, it can supply data support for new construction of parking facilities or land use planning, for example to combine parking areas into a centralized one with better management or to split one to balance the demand. Even for a city which has very limited public parking area, patrol surveys could be used. In Zurich for example, the city wondered how many parking lots are dispensable. By analyzing the average parking duration and parking inventory, one can tell the demand and parking purposes and that could support other decisions.

However, only high quality survey data could reflect the real situation. Moreover, the accuracy and the survey costs (related to survey intensity) should be balanced. In this paper, we analyzed the relation between survey accuracy of average parking duration and the survey intensity.

Below are the findings:

First, the accuracy of the survey could be estimated given the survey intensity. Then a method could be used to adjust the result to obtain on average a higher accuracy.

Second, we have found that the most influential factor to the accuracy is the shape parameter of the PDF of parking duration. Former survey experiences or results can be used to support fitting the distribution and estimating this value.

Third, the coordination between survey costs and accuracy could be done by choosing a proper value of survey intensity. Higher survey intensity means a more expensive survey while it also gives better accuracy. Based on the model, we could find a proper value to guarantee that both are acceptable, or the minimum possible cost for a desired level of accuracy.

Fourth, the length of the survey affects the accuracy although not significantly. However, when the parameter is within a certain range (e.g., a high value of k for gamma distribution), or when the duration obtained is comparatively long, the survey should be prolonged correspondingly.

Fifth, through our study, it is proved that when survey intensity is below 1.5, the survey could be misleading, not only because a comparatively low accuracy, but also because the data recorded have with large natural variations.

clearly the model and the results presented here are limited by the parking duration and arrival distributions we used. Also, in the model we have counted all the vehicles that arrived within the day, while in practice, the vehicles with no information on arrival interval or departure interval are eliminated or modified. The errors caused by this data manipulation are not considered. Further research is necessary to extend and generalize the findings to even more distributions and other parking scenarios.

5. Acknowledgements

The authors are very grateful to Dr. Michael MacNicholas for his valuable information on patrol surveys based on his experience in the field of parking data collection.

6. References

Richardson, A.J. (1974) An improved parking duration study. Proc. Australian Road Research Board Conference, Vol. 7

Central City Parking Analysis (2008) Office of Transportation, city of Portland, Oregon, pp. 23–24.

Cleveland DE (1963) Accuracy of the periodic check parking study. Traffic Engineering. 33(12): 14–17

Tong, CO, SC Wong and BS Leung (2004) Estimation of parking accumulation profiles from survey data. Transportation, 31, pp. 183–202

Lautso, K.(1981) Mathematical relationships among parking characteristics and revising and reduction methods of parking field survey information. Transportation Res. B 15B:73–83.

Marler, N.W. and F.O. Montgomery (1993). Urban Road Traffic Surveys. Overseas Road Note 11, Transport Research Laboratory and Overseas Development Administration. TRL, Crowthorne, UK.

Bonsall, P. (1991) The changing face of parking-related data collection and analysis: the role of new technologies. Transportation, 18, pp. 94–100

Bonsall, P. (1997) Transport Planning and Traffic Engineering(chapter 13: Observational Traffic Surveys)

TJKM-Transportation Consultants (2007) Downtown Parking Study For the City of Grass Valley, pp. 12–14

Associated Engineering Consultants, Inc. (2011) Parking study of city of Medicine Hat, pp.16–18

BSC group (2002) Municipal parking study town of Belmont, pp.17–19

Carl Walker, Inc.(2009) Downtown Lincoln comprehensive parking study (2009), pp. 37–39