Estimating Highway Speed Distributions From a Moving Vehicle

JON D. FRICKER and HUEL-SHENG TSAY

ABSTRACT

The question of how to develop a distribution of speeds for vehicles passing and being passed by an observer car in the vehicle stream has attracted the interest of several mathematicians, but their published work has not gone beyond the theoretical stage. In practice, the theoretical expressions that have been developed do not permit the construction of an accurate speed distribution from actual data gathered from an observer car. The standard deviation of the distributions they construct are accurate, but their inability to process a few discrete data points causes errors in estimating the mean speed of the traffic stream. An observer car was operated over 817.9 mi of Interstate highway, gathering actual data for use in this study. A simulation program was written to help test an empirical method for constructing a speed distribution. The simulation revealed that the empirical method predicted the mean speed well, but estimated the standard deviation poorly. Furthermore, it illustrated why these inaccuracies resulted. Finally, a practical method was produced by which anyone can determine good estimates of both the mean and standard deviation of the speed distribution without special equipment.

In order to avoid losing certain federal-aid highway funds or to qualify for incentive grants, each state must compile annual speed compliance data. If more than 30 percent of the vehicles on a representative sample of highways in a state exceed 55 mph, that state may have 10 percent of its federal-aid highway funds withheld. If fewer than 20 percent exceed 55 mph, the state can claim an incentive grant (1). Often, as one is driving on an Interstate highway, it is difficult to believe that sufficient compliance is actually taking place. This paper presents a practical method of approximating, not only the mean speed of traffic, but the distribution of speeds in the stream of traffic surrounding a vehicle driven by an individual who is curious about speed compliance levels on a given highway section. This method describes the speed of vehicles throughout the traffic stream, not just at isolated speed monitoring stations.

A THEORETICAL BASIS FOR THE MODEL

A Mathematical Model

The authors are interested in the amount of accurate, useful information an individual in traffic can easily collect. (This person's vehicle shall be called the "observer car" in this paper.) What information must be collected or inferred in order to adequately describe the distribution of vehicle speeds? One way to answer this question is to build a mathematical representation of the traffic stream and the observer car's relationship to other vehicles in it. Employing a mathematical model usually requires the adoption of simplifying assumptions such as

1. Each driver in the traffic stream chooses a speed, v, and drives constantly at this speed.
2. If the observer car approaches a slower vehicle ahead of it, the observer car overtakes it without delay (i.e., there is no interaction between cars).
3. The traffic has reached "steady state."
4. Vehicles entering the highway follow a homogeneous Poisson process.

The following parameter definitions pertain to Equations 1-5:

- q = the flow of vehicles past a point on the highway (vehicles per hr);
- v_o = the speed of the observer vehicle;
- $F(v)$ = the probability that the speed of a vehicle selected at random passing a certain point on the highway is smaller than v. $F(v) = Pr(V \leq v)$. The random variables V are independently and identically distributed;
- $\lambda^+(v_o)$ = the number of vehicles overtaken by the observer vehicle (vehicles per hr); and
- $\lambda^-(v_o)$ = the number of vehicles that overtake the observer vehicle (vehhicles per hr).

From References 2 and 3, the following equations can be assembled:

$$\lambda^+(v_o) = q \int_0^{v_o} [(v_o - v)/v] \, dF(v) \tag{1}$$

$$\lambda^-(v_o) = q \int_{v_o}^{\infty} [(v - v_o)/v] \, dF(v) \tag{2}$$

From (1)

$$\lambda^+(v_o) = qv_o \int_0^{v_o} F(v)dv/v^2 \tag{3}$$

Then

$$F(v_o) = 1/q\{[v_o\, d\lambda^+(v_o)]/dv_o\} - \lambda^+(v_o) \quad (4)$$

Equation 3 shows that $\lambda^+(v_o)$ is an increasing function of v. For specified values of the parameters, Equation 4 estimates the ratio of the number of cars the observer car will pass to the number it will observe. Thus, if a driver travels for sufficiently long periods of time, periodically assumes different (but constant) speeds, and counts the number of cars overtaken (or that overtake him), the speed distribution and the traffic flow can be estimated (2,3). To use Equation 4, the difference between the number of vehicles overtaken by the observer car at speeds $v^1{}_o$ and $v^2{}_o$ needs to be determined. If the number of cars overtaken at speeds $v^1{}_o$ and $v^2{}_o$ are $\lambda^+(v^1{}_o)$, and $\lambda^+(v^2{}_o)$, respectively, then Equation 4 can be used in the following discrete form:

$$F(v^2{}_o) = 1/q\ \{v^2{}_o[\lambda^+(v^2{}_o) - \lambda^+(v^1{}_o)]/(v^2{}_o - v^1{}_o)\} - \lambda^+(v^2{}_o) \quad (5)$$

For example, if the traffic flow (q) is 750 vehicles per hr, the observer car passes 15 vehicles while moving at 57.6 mph, and passes none at 52.8 mph, Equation 5 becomes

$$F(57.6) = 1/750[57.6(15 - 0)/(57.6 - 52.8)] - 15 = 0.22,$$

where 52.8 is the estimated speed of the slowest vehicle on the highway section of interest. Repeating this equation for four higher speeds (as shown in the following calculations) allows for creation of the plot in Figure 1.

$$F(61.4) = 1/750\ [61.4(29 - 15)/(61.4 - 57.6)] - 29 = 0.263$$

$$F(62.4) = 1/750\ [62.4(38 - 29)/(62.4 - 61.4)] - 38 = 0.69$$

$$F(64.3) = 1/750\ [64.3(56 - 38)/(64.3 - 62.4)] - 56 = 0.74$$

$$F(66.2) = 1/750\ [66.2(79 - 56)/(66.2 - 64.3)] - 79 = 0.96$$

The $\lambda^+(v_o)$ and $\lambda^-(v_o)$ values in Equations 1, 2, and 5 come from counting the number of vehicles that overtake and are overtaken by the observer car.

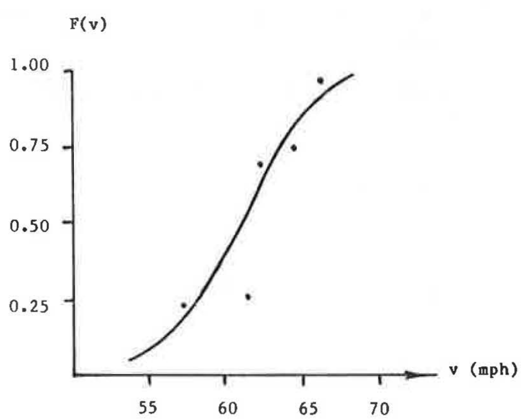


FIGURE 1 CDF of vehicle speeds from mathematical model.

Real traffic flow is not as well-behaved as the simplifying assumptions would indicate. Equations 4 and 5 are intended for use with small differences between $v^1{}_o$ and $v^2{}_o$ and with $\lambda^+(v_o)$ values that increase monotonically with v_o. Neither requirement is likely to be met in practice, however. It would not be practical to collect $\lambda^+(v_o)$ and $\lambda^-(v_o)$ counts for 1 mph increments of v_o. Secondly, there is no guarantee that $\lambda^+(v^2{}_o)$ will be greater than $\lambda^+(v^1{}_o)$, just because $v^2{}_o$ is greater than $v^1{}_o$. Finally, Equation 5 is sensitive to slight variations in $\lambda^+(v_o)$. When v_o = 62.4 mph, $\lambda^+(62.4)$ = 38, which gives F(62.4) = 0.69. For $\lambda^+(62.4)$ = 37, F(62.4) becomes 0.61; for $\lambda^+(62.4)$ = 39, F(62.4) rises to 0.78. These factors explain why the real-life observations that formed the basis for Figure 1 put a point at v_o = 62.4 mph, $F(v_o)$ = 0.263, well away from an otherwise fairly consistent set of points.

The curve in Figure 1 is, of course, a cumulative distribution function (CDF). From this CDF, the proportion of vehicles that are overtaken by the observer car can be estimated for any speed v_o. Furthermore, having a CDF makes it possible to convert it into the corresponding probability density function (PDF) that defines the speed distribution sought by the authors.

Reflections on the Mathematical Model

Clearly, the four simplifying assumptions are not strictly correct. The extent of their departure from reality has implications for our study as follows:

1. Drivers maintain constant speeds. Each driver has a speed he or she finds comfortable and would prefer to maintain (4). This speed is a function of that individual's driving attitude, ability, and experience. It is also influenced by the characteristics of the vehicle being driven: type (car, truck, etc.), size, handling, performance at certain speeds, and so forth. These factors support the first assumption, if driving conditions remain constant. But they do not. Furthermore, a driver cannot always drive at his or her "preferred" speed. Physical terrain (hills and curves) may be a limiting factor. So might vehicle densities, if they become great enough. Even at low densities, two or three cars properly positioned may be enough to impede or "push" a driver. Thus, even cruise control does not guarantee keeping to a preferred speed. Besides being boxed in, some drivers tend to adopt the speed of the vehicle just ahead, substituting the preference of the lead driver for their own. The adopted speed may be faster or slower than their own preference, but presumably the difference is not great. And, of course, the ultimate tempering influence on preferred speeds is the speed limit. An associated influence is the driver's perception of the enforcement level. What is the probability that one's speed will be detected? How much above the posted speed limit may one drive without being cited by the authorities? A final complicating influence may be the miscalibration of a vehicle's speedometer. Many speedometers read 2-8 percent high, that is, the speed shown may be 1.02-1.08 times the actual speed. For example, a car equipped with a speedometer that reads 65 mph, but reads 5 percent high, is actually traveling 65/1.05 = 61.9 mph. This 3 mph difference is often not explicitly known by the driver, who instead either (a) guesses at this value or (b) simply assumes a comfortable position in the traffic stream, which includes all the tempering factors mentioned.

2. No interaction between vehicles. Of course, vehicles in close proximity often do influence each other. One can be boxed in, be pushed to higher

speeds by a tailgating car or truck, or follow a lead car in a "platoon" of vehicles. Should this bother us in our study? Not really. If our central question is "How fast are vehicles moving?" it does not matter why they have a certain speed. Neither does it matter to the states' speed monitoring programs and the federal government, unless the speeds are so high that "corrective action" must be taken.

3. and 4. Traffic has reached steady state, in which entering vehicles follow a homogeneous Poisson process. The first half of this statement says that there is no abrupt change in the flow, such as there would be when a road reopens following a blockage. The second half says that Poisson distribution is appropriate for describing discrete random events. When traffic is light and when there is no obvious disturbing factor such as a traffic signal, the behavior of traffic may appear to be random, and the Poisson distribution will give satisfactory results (5). Neither of these assumptions is seriously incorrect. Traffic varies over the course of a day, but the transitions between high and low arrival rates are gradual enough to accept these assumptions for periods of analysis that are not long enough to include a significant change in arrival rates.

While none of the four assumptions are absolutely correct representations of reality, the real question is how well do they approximate the traffic stream being modeled. This can only be answered by gathering data on the phenomenon being described, which is what we describe next.

DATA COLLECTION

A more direct approach to characterizing the speed distribution is to drive at a variety of constant speeds for known distances, and record the number of vehicles that you pass (p') and that pass you (p). This was done with a cruise-control-equipped car on 817.9 mi of Interstate highways in the Midwest. Except where construction zones or impeding traffic prevented it, the car was operated at a constant speed between 52.8 and 69.1 mph for 5-mi segments. (Actually, the speedometer was read to the nearest integer mph, but timing the vehicle between mileposts determined that the observer car's actual speed was 0.960 times the speedometer reading. This is the reason for the unusual speeds used in the Equation 5 example calculations earlier and throughout this paper.) An overtaken vehicle was not added to the p' count if it was in the process of entering or exiting the main travel lanes of the Interstate, either using the on/off ramps or the shoulder. As soon as the observer car was able to resume the desired speed after a construction zone, speed trap, or rest break, data collection resumed. The validity of these data is analyzed later in this section.

Vehicle densities were estimated by counting the number of vehicles between the observer car and a distinctive landmark ahead, then measuring the distance to that point. Typical observed densities ranged from 10 to 40 vehicles per mile for two-lane (in each direction) Interstate highways.

There are several reasons why the observations could not be taken from a strictly homogeneous population of vehicle speeds.

1. Variations in terrain and roadway geometry.
2. Variations in percentage of heavy vehicles in the traffic stream.
3. Variations in vehicle density.
4. The "environment" of the highway: urban versus rural.
5. Enforcement of speed limits: visibility or reputation of the law enforcement agency.

The number of observations (p + p') made in a segment is a function of the observer car's speed, the segment length, and the density of traffic. As we shall show, if the distribution of vehicle speeds on a highway section is approximately normal with mean μ, p + p' should decrease as $|v_0 - \mu|$ gets smaller. It is obvious that p + p' will increase if segment length increases, density increases, or both. The influence of densities and segment length on the p and p' counts can be eliminated by converting them to ratios as follows

$$\rho = p/(p + p'); \; \rho' = p'/(p + p'); \; \rho + \rho' = 1.00$$

These ratios are summarized in Table 1, with rural Interstate segments shown in a separate column. As expected, ρ' tends to be smaller at most observer car speeds in rural areas than in the overall dataset. The other sources of heterogeneity are not so easily eliminated, and we may not want to. We could decide to take p, p' counts only when certain conditions (e.g., level terrain, low densities) were met, or we could identify and separate our segments into hilly, level, urban, and rural groups. The roads traveled by the observer car were predominantly rural with gentle hills, but even the urban and hillier sections were not drastically different. (Observations were halted during a heavy thunderstorm, however.) In this paper, an attempt is made to estimate the speed distribution on all Interstate sections for which reliable data exist.

TABLE 1 Speed Data from Observer Car

Speed (mph) v_0	Rural ρ'	Total[a] ρ'
54.7	0.00	0.05
56.6	0.02	0.02
57.6	0.20	0.20
58.6	0.50	0.50
59.5	0.27	0.42
60.5	0.20	0.20
61.4	0.45	0.53
62.4	0.69	0.69
63.4	0.71	0.75
64.3	0.72	0.72
65.3	0.82	0.83
66.2	0.95	0.95
68.2	0.92	0.92
69.1	1.00	1.00

[a] Includes rural, urban, and suburban segments.

What are "reliable data"? For instance, it is correct for our purposes to take p, p' counts soon after a construction zone? Table 2 shows the summary of p, p' counts for the 16 segments that followed a construction zone. One would expect a noticeable increase in p + p' in these segments. Vehicles that eventually would have been seen from the observer car would pass or be passed by the observer car sooner, since construction zones tend to cause the formation of platoons. One would not necessarily expect a bias in the r' value, which is the p'/(p + p') value for these particular segments. Faster cars behind the observer car in the construction zone platoon may find some slower vehicles blocking them in the next open segment, but this is typically a 2-lane segment with at least twice the capacity of a 1-lane construction zone. The open segment was also 5 mi long in 14 of the 16 cases. The results of a chi-square test applied to the r' and ρ' columns in Table 2 support the hypothesis that the r' after

TABLE 2 Data Collection After Construction Zones

v_o	p	p'	r'	ρ' (all)	r' − ρ'
56.6	2	0	0.00	0.02	−0.02
57.6	15	2	0.12	0.20	−0.08
59.5	3	2	0.40	0.42	−0.02
61.4	6	18	0.75	0.53	+0.22
62.4	1	9	0.90	0.69	+0.21
64.3	3	3	0.50	0.72	−0.22
65.3	2	16	0.89	0.83	+0.06
66.2	0	13	1.00	0.95	+0.05
Total	32	63			

Note: r' = p'/p + p'. Observation rate = 95 vehicles/76.9 mi = 1.235 vehicles per mile (not a standard density).

construction zones is not biased at the 95 percent confidence level. Furthermore, the expectation that more vehicles will be observed in these segments is also realized: 1.235 per mile versus 0.948 overall.

Should observations made in the presence or aftermath of highway police surveillance be included in our data base? Table 3 is similar in structure to Table 2, but involves only 5 segments over 13.75 mi. A significantly lower r' than ρ' would be expected in this case, but the data are too meager and contradictory to draw such a conclusion. These data were not discarded.

TABLE 3 Data Collection After Police Sighted

v_o	p	p'	r'	ρ' (all)
59.5	5	0	0.00	0.42
62.4	1	7	0.88	0.69
63.4	2	0	0.00	0.75
64.3	0	4	1.00	0.72

Plotting the (v_o,ρ') values from Table 1 results in the points shown in Figure 2. Fitting a smooth curve through these points produces an "empirical" CDF much like that of Figure 1. This CDF can be used to test the fit of any distribution that is proposed as a model of the actual speed distribution along the highway.

The Two Approaches

In both the mathematical model and the empirical method described previously, the counts of vehicles

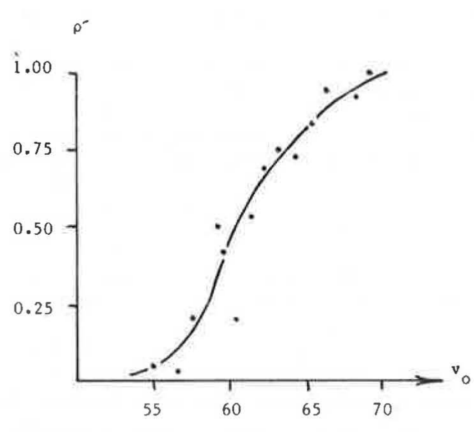


FIGURE 2 CDF of vehicle speeds from observer car data.

(p, p') over a range of speeds have been converted into proportions or ratios, which, in turn, have been plotted in CDF--Pr ($V \leq v_o$)--form. The resulting S-shaped curves (Figures 1 and 2) are "similar," based on goodness-of-fit tests at the 95 percent confidence level. Thus, our data collection effort seems to validate the mathematical model's ability to describe the distribution of speeds, despite our critique of its simplifying assumptions, and subject to the limitations on the differences in v_o already discussed. The model's F(v_o) values closely match the ρ'(v_o) values generated for Table 1, except at v_o = 61.4 mph. What remains is to devise a procedure to convert either of the CDFs to the PDF of vehicle speeds and then test the results for accuracy.

TESTING THE METHOD

How do we test the model's ability to estimate the actual distribution of speeds in the highway section of interest? Ideally, radar could be used to record the exact speed of a vehicle at the time of observation, but this involves more expense or effort than is desired. Besides, there are complicating factors to consider. If a radar-derived speed is assigned to each vehicle, that speed ignores the variation of speeds a vehicle assumes as it reacts to different roadway geometry, traffic conditions, and driver behavior. A stationary radar site removes the variable of geometry, but at the moving observation post, vehicles are being viewed under a full range of conditions, which gives (theoretically) a better sample of overall vehicle speeds in the total traffic stream. However, the authors propose to limit the data collection to the simple question "When a vehicle is observed, is it passing or being passed?" Nothing can precisely be said about the other vehicle's speed, v, except $v > v_o$ or $v < v_o$. The difference is speeds may be 1, 5, or 20 mph. The observer can only guess, but need not (if the authors' method succeeds) even try.

Can a speed distribution be effectively reconstructed from this limited information? Equally important, against what can the estimates be compared? Speed monitoring data are usually gathered at a few fixed locations. Although care is taken to disguise the sites and to establish them in places unaffected by merging maneuvers and other factors, they are constrained to level highway sections where access to the recorders is easy and where protection of the equipment (e.g., by guardrails) is possible. Each location used probably has some bias as a result of prevailing vehicle mix, variations in vehicle densities, weather conditions, and other factors. If one is interested in what speeds exist in the entire traffic stream during a particular trip, only a considerable number of perfectly disguised recording stations along the route could give the desired distribution. For reasons of economy and reproducibility of results, a simulation model was developed.

The Simulation Model

The authors wished to sample speeds of vehicles in the traffic stream by observing whether the driver is passing them ($v_o > v$) or they are passing the driver ($v > v_o$). Then this sample and a constructed PDF of speeds are compared with the "actual" speed distribution for the traffic stream in which the authors were traveling. For reasons already mentioned, the actual distribution is not possible to accurately describe. But, for these testing purposes, it is not really necessary to know the exact distribution of an actual traffic stream. If one could create an exact distribution, pretend not be know its shape, then sample from it to generate an estimated distribution, one would have the basis for testing the authors' method. A simulation model can create such distributions.

Steps in Simulation Program

1. Reading input values that include the following;
   a. Parameters of distribution describing vehicle speeds;
   b. Traffic density (vehicles per mile);
   c. Speed of observer car (v_o);
   d. Termination criteria as follows;
      • Length of observation section, Δx (mi),
      • Length of time that v_o will be maintained (min), and
      • Minimum number of observations k^* to be made at v_o mph;
   e. Speed for random number generator.
2. Increment count of vehicles observed, $k = k + 1$.
3. Generate the location x_k of the k^{th} vehicle with respect to the observer car ($x_o = 0$) at the start of the observation period, sampling from a uniform distribution.
4. Generate the k^{th} vehicle's speed, v_k, using the hypothesized distribution.
5. If $x_k > x_o$, go to step 6. If $x_k \leq x_o$, calculate whether the k^{th} vehicle will overtake the observer car before the observation period terminates. If so, $p = p + 1$. Go to step 7.
6. Calculate whether the observer car will overtake the k^{th} vehicle before the observation period terminates. If so, $p' = p' + 1$.
7. If the observation period has not ended, return to Step 2.
8. Stop simulation. Print summary of observations made.

The most interesting product of the simulation program is the plot (Figure 3) of the distribution of speeds for the vehicles that were seen from the observer car. These speeds are not known to the observer, except to the extent that they make up the p and p' counts, but their distribution is at once fascinating and important to the estimation method. Figure 3 illustrates how the actual distribution of vehicle speeds (dashed curve) is distorted when viewed from a moving vehicle. The reason for the dip in curve (b) near v_o is easy to understand.

• For a vehicle with speed v_k not much different from v_o, its initial location must be relatively close to x_o to expect meeting the observer car during the study period.
• For a given v_o, the probability of the k^{th} vehicle meeting the observer car during the study period increases as $|v_k - v_o|$ increases, especially as v_k moves in the direction of the mode of the speed distribution.
• However, as v_k begins to take on extreme values (in the tails of the speed distribution) that are less likely to occur, this decreases the probability of a vehicle with such speed even existing to meet the observer car.

Note that both curves in Figure 3 are drawn as PDFs. Each encloses an area of 1.00, based on their respective definitions of an event. However, the speed distribution of the population (the dashed curve) represents a greater number of vehicles in the study section, since a majority of these vehicles will never meet (be sampled or observed from) the observer car. The standard deviation for curve (b) is greater than that for (a) because there are fewer speeds observed near the mean for curve (b).

HOW TO USE THE METHOD

The task before us is to take a few p, and p' counts and convert them into an estimate of the speed distribution's parameters. The authors are only concerned with putting each observed vehicle into one of two categories: $v < v_o$ or $v > v_o$.

At each v_o, the data collected will be used to locate a point on a CDF such as in Figure 2. The accuracy of that point's location $p'(v_o)$ with respect to its correct location should improve with an increase in the length of time at v_o, since $\rho'(v_o)$ will be based on a larger sample size, p + p'. However, each time the v_o is changed, a new "view" of the distribution is obtained, which the authors are trying to reproduce. Because each view is so limited (either $v_o > v_k$ or $v_k > v_o$), the greater the number of v_o's that are adopted, the better the CDF's shape

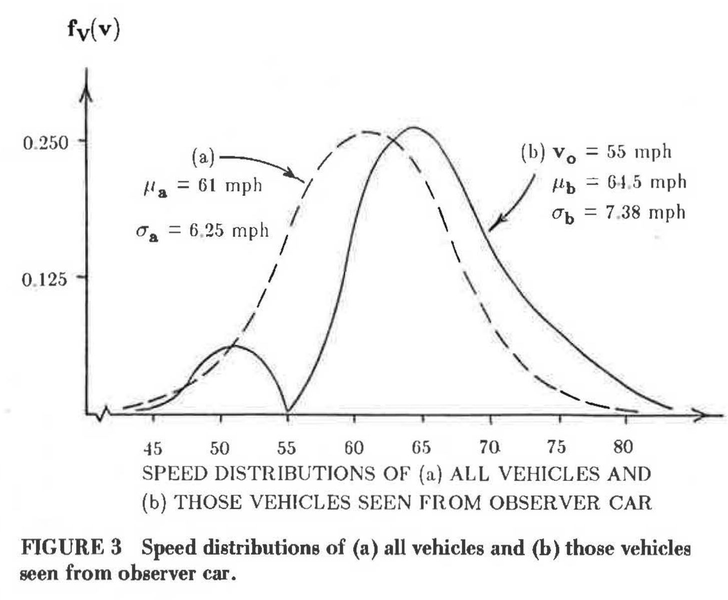


FIGURE 3 Speed distributions of (a) all vehicles and (b) those vehicles seen from observer car.

is likely to be approximated. For a trip of moderate length, say 100 mi, a trade-off occurs. One can travel at two different speeds for 50 mi each, or at 10 different speeds for 10 mi each, or any other combination of distance (Δx) and number of different speeds (n) whose product does not exceed 100 mi. The first strategy would seem to provide two points of high quality, while the second approach gives ten fewer reliable points through which to fit a CDF.

Converting a CDF to a PDF

Say, for example, that one chose to travel at a v_o of 55 mph for 50 mi, and that the simulation program reported that $\rho'(55) = .104$. This means that $Pr(v_k < 55 \text{ mph}) = .104$. Because the CDFs from both the mathematical model and the empirical method (Figures 1 and 2) resemble a CDF derived from a normal distribution, one is justified in assuming a normal distribution. Branston (6) supports this choice of distribution. Thus, we seek the standard normal deviate

$$z = (v_o - \mu)/\sigma$$

that causes $\Phi(z) = .104$. This is represented by $\Phi^{-1}(.104)$, the solution to which can be found in a CDF table of the standard normal distribution. Here,

$$\Phi^{-1}(.104) = -1.26 = (v_o - \mu)/\sigma = (55 - \mu)/\sigma \qquad (6)$$

with μ and σ being the two parameter values to be solved for. Equation 6 can be rearranged to be

$$\hat{\mu} - 1.26\,\hat{\sigma} = 55 \qquad (6')$$

Driving at a different v_o, say, 65 mph, gives $\rho'(65) = 0.787$. This becomes $\Phi^{-1}(.787) = 0.795$ and, in a few steps,

$$\hat{\mu} - 0.795\,\hat{\sigma} = 65 \qquad (7)$$

Solving Equations 6' and 7 simultaneously gives $\hat{\mu} = 61.1$ mph and $\hat{\sigma} = 4.87$ mph. Trying other v_o values gives an expanded number of pairs of values with which to find $\hat{\mu}$ and $\hat{\sigma}$. Table 4 shows how these estimates will vary with choices of v_o pairs. Insuring that the number (n) of observer speeds (v_o) exceeds two produces $n!/2!(n-2)!$ solutions for $\hat{\mu}$, $\hat{\sigma}$ to be examined for trends and clues, such as those solutions that should be discarded.

TABLE 4 Solutions for $\hat{\mu}$, $\hat{\sigma}$

v_o pairs	$\hat{\mu}$	$\hat{\sigma}$
55 and 60	60.89	4.67
60 and 65	61.0	5.08
65 and 70	61.6	4.16
55 and 65	61.1	4.87
60 and 70	60.85	4.55
55 and 70	60.77	4.59
Avg. values	61.04	4.65

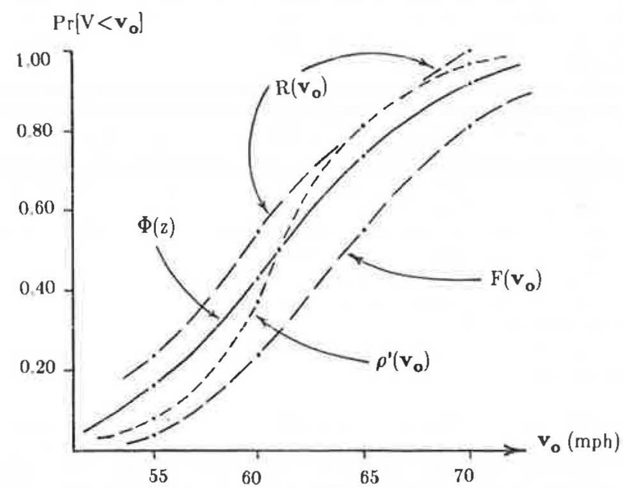
Note: $\Delta x_o = 50$ mi, $\mu = 61$ mph, and $\sigma = 6.25$ mph.

At this point, the authors could select the "most central" of the six solutions in Table 4, say, $\hat{\mu} = 60.89$ with $\hat{\sigma} = 4.67$, adopt the average of the parameter solution values ($\hat{\mu} = 61.04$, $\hat{\sigma} = 4.65$), or use any legitimate variation. In this example, the estimate of $\hat{\mu}$ would be quite accurate, but $\hat{\sigma}$ is consistently about 25 percent below the 6.25 value entered into the simulation program. Table 5 and Figure 4 illustrate why this is so.

The solid curve labeled "$\Phi(z)$" in Figure is the CDF for the normal distribution the authors are attempting to duplicate from the observed data. The curve with short dashes--"$\rho'(v_o)$"--is the CDF constructed from the proposed method, with (Δx = 1,000 mi) to ensure statistical stability. Note how it starts below the $\Phi(z)$ curve, crosses it at approximately 0.50, and finishes above it. The $\rho'(v_o)$ curve is "steeper" than the $\Phi(z)$ curve; therefore, it will have a narrower PDF than $\Phi(z)$'s target PDF.

TABLE 5 Comparison of CDF Values

v_o:	55	60	65	70
$\Phi(z)$	0.1685	0.4364	0.7389	0.9251
$\rho'(v_o)$	0.080	0.382	0.821	0.976
$F(v_o)$	0.044	0.244	0.554	0.816
$R(v_o)$	0.244	0.554	0.816	1.00



CUMULATIVE SPEED DISTRIBUTIONS: THREE METHODS

FIGURE 4 Cumulative speed distributions: three methods.

The "$F(v_o)$" curve is based on the use in Equation 5 of the same data Δx = 1,000 mi) that went into the $\rho'(v_o)$ curve. Note how $F(v_o)$ is consistently lower than the desired $\Phi(z)$ value. However, the $F(v_o)$ curve has one very desirable characteristic: its slope is much like that of the $\Phi(z)$ curve. On the theory that the $F(v_o)$ curve is too low because Equation 5 does not handle large changes in v_o well, the authors rearranged that equation to get

$$R(v^1{}_o) = 1/q\ \{v^1{}_o[\lambda^+(v^1{}_o) - \lambda^+(v^2{}_o)] \div (v^1{}_o - v^2{}_o) - \lambda^+(v^1{}_o)\} \qquad (5')$$

Because this is just Equation 5 with $v^1{}_o$ and $v^2{}_o$ reversed (and with $v^2{}_o$ still greater than $v^1{}_o$), this is called the Reverse $F(v_o)$ Curve, denoted by $R(v_o)$. The $R(v_o)$ curve is consistently above $\Phi(z)$, but with a similar slope. The "average" of the $F(v_o)$ and $R(v_o)$ curves would approximate the $\Phi(z)$ curve quite closely. In addition, the common structure of Equations 5 and 5' gives rise to a repetition of solution values (see .244, .554, and .816 in Table 5). This allows for a check on the accuracy of calculations or the elimination of the three redundant ones.

By using .5 $[F(v_O) + R(v_O)]$ as input to Equation 6, the same procedure that built Table 4 provides the entries for Table 6. A highway section of 1,000 mi is too long to be realistic, but it is used here to guard against statistical aberrations in the p and p' counts that can occur with small samples (short sections). The μ estimates are slightly high [since $F(v_O)$ and $R(v_O)$ are not quite symmetric about the $\Phi(z)$ curve], but $\hat{\sigma}$ is close to the desired 6.25 value of σ, thanks to the similarity in the slopes of the three CDFs involved.

TABLE 6 Solutions for $\hat{\mu}$, $\hat{\sigma}$ Using $F(v_O)$ and $R(v_O)$ Values

v_O Pairs	$\hat{\mu}$	$\hat{\sigma}$
55 and 60	61.58	6.21
60 and 65	61.73	6.80
65 and 70	62.18	5.88
55 and 65	61.88	6.49
60 and 70	61.60	6.31
55 and 70	61.65	6.27
Avg. values	61.77	6.33

Note: Δx_O = 1,000 mi, μ = 61 mph, and σ = 6.25 mph.

SUMMARY

The method developed in this paper offers some advances in the practical aspects of estimating the highway speed distribution from a moving vehicle. Mathematical models of the sort described by Equation 4 call for a fully specified CDF in order to produce the PDF. This specification is seldom possible. What is possible is a count of vehicles overtaken by (and overtaking) an observer car. Our first proposed method estimated μ well, but consistently underestimated σ. The mathematical model's CDF was consistently low, but had the correct slope. Use of this CDF and its "reverse" counterpart in a step preliminary to our method of solving simultaneous equations yielded good estimates of μ and σ.

The result is a description of highway speeds which, in some ways, is superior to the stationary speed monitoring summaries in that a more complete picture is obtained of the traffic stream under a variety of driving conditions.

ACKNOWLEDGMENTS

The authors thank Charles Oliva, who wrote the simulation program, and Rita Wolf, who performed most of the word processing chores.

REFERENCES

1. Speed Monitoring Program Procedural Manual for the National Maximum Speed Limit. FHWA, U.S. Department of Transportation, May 1980.
2. E. Hauer. Accidents, Overtaking, and Speed Control. In Accident Analysis & Prevention, Vol. 3, 1971, pp. 1-13.
3. A. Renyi. On Two Mathematical Models of the Traffic on a Divided Highway. In Journal of Applied Probability, Vol. 1, 1964, pp. 311-320.
4. E. Hauer, F.J. Ahlin, and J.S. Bower. Speed Enforcement and Speed Choice. In Accident Analysis & Prevention, Vol. 14, No. 4, 1982, pp. 267-278.
5. D.L. Gerlough and M.H. Huber. Traffic Flow Theory. In Special Report 165, TRB, National Research Council, Washington, D.C., 1975, 222 pp.
6. D. Branston. A Method of Estimating the Free Speed Distribution for a Road. In Transportation Science, Vol. 13, No. 2, May 1979.

Publication of this paper sponsored by Committee on Simulation and Measurement of Driving.