Abstract

By means of standardized statistical tests applied to empirical traffic data (recorded at the Expressway R1 in Prague, Czech Republic) we verify a hypothesis that vehicular clearances are distributed via Generalized Inverse Gaussian distribution \( g(x) \propto e^{-\beta x} e^{-\lambda x} (x > 0, \beta \geq 0, \lambda > 0) \). We formalize mathematical theory explaining recent results obtained by means of advanced statistical analysis applied to vehicular/pedestrian microstructure. For these purposes we use (and generalize) approaches applied in the theory of counting processes. Quantities standardly analyzed in vehicular headway modeling (headway, multi-headway, interval frequency, and statistical rigidity) are here reformulated into formal mathematical definitions and then analytical predictions for statistical rigidity of particle systems with GIG-distributed headways are compared with empirical behavior. We show that a connection between clearance distribution and rigidity in real-road data is not tight as in theoretical structures. This discrepancy is explained as a consequence of the fact that interaction rules acting in vehicular systems (unlike level processes studied) are not short-ranged, which supports a hypothesis that mutual interactions exist among several succeeding cars (as investigated in Krbálek et al. 2018 Physica A 491, 112).

1. Introduction and motivation

Vehicular headways (gaps or clearances) represent, as is well known, fundamental quantities investigated in empirical/theoretical studies of traffic flow [1–3]. Their statistical distributions characterize how vehicles are distributed on a roadway/freeway [4, 5] or on intersections [6]. Therefore, headway distributions represent a crucial factor for mathematical predictions of freeway/intersection capacities. At present, vehicular headway modeling is an extensively developing discipline which reveals a number of remarkable features in vehicular microstructure. Historical overview and achievements of this discipline are synoptically summarized in the survey [1].

On the other hand, vehicular headways (and clearances) are closely interconnected to the fundamental macroscopic quantities, namely traffic density \( \rho \) and intensity \( I \) (flux). Correct localization of a traffic system in the ID plane strongly depends on accuracy in enumerating the local intensity and local density. Both quantities are, however, calculated with help of random variables like number of cars occurring inside a freeway-segment of a given length, or number of cars crossing a given point during a certain time interval. In fact, a stochastic component of those variables (denoted generally as interval frequencies) plays a principal role in all theoretical branches of the theory. A deeper (and theoretically substantiated) understanding of statistical linkages among usual traffic variables is, therefore, highly desirable.

In this paper we aim to reformulate a standard traffic description into a mathematical language working with associated stochastic variables. Since a standard mathematical theory dealing with events consecutively coming in time is the theory of counting processes [7, 8] we will try to use its approaches to reveal a linkage between clearance distribution and interval frequency of the same particle system. Particularly, we plan to interconnect clearance distribution (time or spatial) with the second central moment of interval frequency. In the theory of
traffic (or in Random Matrix Theory) the latter is usually called a statistical rigidity. Such an advanced quantity (discussed for example in [9–12]) allows to compare a level of synchronization among agents for two different agent systems, which is a noticeable benefit if analyzing empirical data. Besides the transportation science, a similar instrument has been successfully applied also to pedestrian dynamics [13] or public transport [14, 15].

2. Headway modeling—a way from ad-hoc models to formal estimation procedures

Headway modeling (as scientific discipline dealing with analytical predictions/estimations of empirical headway/clearance distributions) has undergone an intense evolution over the last 80 years. First attempt to model of headway statistics in real road flows can be dated to 1930s (see [16]). This pioneering work (working with normal or negative exponential distributions) has been followed by many successors [4–6, 10, 17–23]. However, some of their outputs represent ad-hoc estimations only (which is reasonable at first stages of research). Some of these models, conversely, suggest a statistical description resulting from meaningful mathematical/physical traffic model, which represents a significant progress in this field. Typically, all these approaches use one- or multi-parametric families of distributions (negative exponential, shifted negative exponential, Erlang, Gamma, Pearson type III, semi-Poisson, vehicle-type-specific, log-normal, generalized inverse Gaussian). Unfortunately, respective statistical models have several substantial drawbacks. Firstly, except in a few rare cases (e.g. [24, 25]) almost none of these models is verified by a regular statistical test. Secondly, most of the proposed distributions do not fulfill mathematical criteria for distributions of gaps in one-dimensional particle gases with short- or middle-ranged interaction rules (see below). For these reasons, most of the proposed methods fail for larger samples of data. This is a general feature, so typical for inconsistent statistical estimations. On the other hand, as is well known, consistent statistical estimations generates deviations which are decreasing with length of random vector.

2.1. Data processing & statistical models proposed

According to current knowledge (see [9, 12, 21, 26–28]) we now summarize essential procedures/criteria which should be applied/fulfilled during any estimation procedure made for random variable of the same type as particle clearances.

- **(Segmental data processing)** Variance of real-road clearances is rapidly changing with traffic density [2–4]. It corresponds to the fact that statistical resistivity (measuring a level of resistance to statistical perturbations) depends strongly (as analyzed in [10, 23, 26]) on parameters of actual traffic state (intensity, density, mean velocity). Therefore, necessary part of intended estimation procedure should be a segmentation (i.e. restriction to small sub-areas of ID plane—as discussed in [26, 29] and applied in [23, 26], for example), which prevents a mixing of states with different vigilance of drivers.

- **(Balancing criterion)** Suggested analytical probability density \( g(x) \) should belong to a family \( \mathcal{B} \) of balanced distributions (see appendix A.1). This is a property generally-valid for all one-dimensional particle ensembles, in which number of interacting neighbors is substantially smaller than number of all particles (see a theoretical derivation published in [13]).

- **(Empirical criterion)** According to recent knowledge all empirical clearance distributions (measured in a congested phase) show the heavy plateau at the origin (see figure 3 in [6] and an explanation in [23]). Such a plateau can be understood as a consequence of strong repulsions among closely-occurring vehicles whose drivers make an effort to prevent a possible crash. Some of older approaches simulate this plateau by means of shifted densities (shifted negative exponential [18], shifted log-normal [19]), which therefore are not smooth. Contemporary models work with smooth functions \( g(x) \) (like GIG distributions [12]) fulfilling the following property

\[
\forall \nu \in \mathbb{N}: \quad \lim_{x \to 0} x^{-\nu} g(x) = 0. \tag{1}
\]

In fact, it corresponds to the mathematical definition of the plateau in \( x = 0 \).

2.2. Origin of empirical traffic data

Induction double-loop data analyzed in this paper have been provided by Road and Motorway Directorate of the Czech Republic. They have been measured at several segments of highway circuit R1 in Prague, Czech Republic, where traffic is saturated for most of the day. For a following data-processing they have been used data from fast lanes only where number of long vehicles is negligible. This procedure guarantees a homogeneity of the data. We have used 12 data-files, each having data from measurements lasting three months approximately.
2.3. Hypothesis of GIG-distributed clearances

Being motivated by the effort to estimate a probability density for clearances gauged at European expressways we suggest (similarly to [4–6, 10–12, 23, 26]) to search for a suitable candidate inside the two-parametric family of Generalized Inverse Gaussian densities [30, 31]

\[ g(x|\beta, D) = A\Theta(x)e^{-\beta x^2/Dx}, \quad (\beta \geq 0, D > 0), \]

which respect all the theoretical criteria discussed in the previous subsection. We note that the function (2) does not represent an ad-hoc assumption only. Quite the contrary, this function has been found (see [3]) to be a steady-state distribution for distances between particles in thermodynamical traffic gas. The parameter \( \beta \) can be understood (see [5, 6, 12, 23]) as statistical resistivity of the gas whose interaction rules are based on a verified premise that a repulsive force (preventing a collision of cars in real-road traffic) is inversely proportional to a square of distance between particles—similarly to the well-known dynamics equations in Intelligent Driver Model [32]. In (2), \( \Theta(x) \) stands for the Heaviside unit-step function and the normalization factor reads

\[ A^{-1} = \begin{cases} 2 \left( \frac{\pi}{2} \right) K_0(2\sqrt{\beta D}) & \beta > 0, \\ D^{-1}, & \beta = 0, \end{cases} \]

where \( K_0(x) \) stand for the Macdonald’s function of the \( \nu \)-th order. Furthermore, the scaling constant \( D \) is related to the first moment (the expectation)

\[ \mu_1(g) \equiv E(x) = \int_{\mathbb{R}} x g(x|\beta, D) dx = \begin{cases} \frac{1}{\sqrt{D\beta}} K_0(2\sqrt{\beta D}) & \beta > 0, \\ D^{-1}, & \beta = 0. \end{cases} \]

In this part of the text we present results of two standard procedures estimating an parametric vector \((\beta, D)\), which are based on a maximization of log-likelihood

\[ \ln L(\beta, D) = \ln \left( \prod_{i=1}^{n} A\Theta(x_i)e^{-\beta x_i^2/Dx_i} \right) = n(\ln(A) - \beta x_i^{-1} - Dx), \]

and a minimization of functional metrics (Kolmogorov distance)

\[ \sigma(\beta, D) = \sup_{x \in \mathbb{R}} |G(x|\beta, D) - H(x)|, \]

respectively. Here \((x_1, x_2, \ldots, x_n)\) is a vector of empirical clearances; \(\bar{x}\) is the associated sample mean, \(\bar{x}\) is the harmonic mean, \(H(x)\) is the empirical distribution function of the sample \((x_1, x_2, \ldots, x_n)\), and \(G(x|\beta, D)\) is the distribution function of the GIG distribution (2). Practically, the proposed procedure is implemented as follows.

Empirical data extracted from congested traffic (and divided into samples of fifty succeeding time clearances) are sorted by a local density (for a definition of a local density please see A.4). Then all the ordered samples are divided into 100-samples sets, each with \(n = 5000\) clearances. Altogether we have \(N = 91\) such data sets with average local densities varying from 30 to 66 veh/km. In fact, such a sorting can be identified as a segmental data processing discussed in section 2.1. For each of the 91 data sets the afore-mentioned estimators are applied. An evolution (induced by changes in density) of point estimates

\[ (\hat{\beta}_1, \hat{D}_1) = \arg\max_{\beta \geq 0, D > 0} \ln L(\beta, D), \]

\[ (\hat{\beta}_2, \hat{D}_2) = \arg\min_{\beta \geq 0, D > 0} \sigma(\beta, D) \]

is summarized in figures 1 and 2.

To decide on the quality of the statistical model proposed we carry out a standardized statistical test \((\chi^2\text{-test in the bootstrap approach}) [33]) with respect to previous considerations we formulate a null-hypothesis in the expected form \(H_0: \text{Empirical time clearances (detected in congested traffic flows) are GIG-distributed}. \) For each of the 91 above described data sets represented by the corresponding value of the local density we repeat the following procedure:

- Denoting \(S = \{x_1, x_2, \ldots, x_n\}\) the set of \(n = 5000\) clearances we randomly pick 300 clearances from \(S\) and test whether \(H_0\) is/\(H_0\) may not be rejected at significance level \(\alpha = 0.05\).
- Such a random choice is repeated \(k = 1000\) times and average rate of rejection \(\hat{\phi}_{\text{MLE}}, \hat{\phi}_{\text{MDE}}\) is then evaluated (according to [33]).

The obtained rejection rates for the assumed values of the local density are drawn in figure 3. In this diagram we see that rejection rates differ only slightly from an ideal value determined by the chosen significance level \(\alpha\). In fact, with exceptions the suggested hypothesis is rejected in less than 15% of all cases. To be more specific, since
the samples were drawn independently the CLT implies that the rejection rates are, for each value of the density, approximately $\mathcal{N}(\alpha, \alpha(1 - \alpha)/k)$ distributed under $H_0$. Taking into account this asymptotic distribution, we plot in figure 3, for each density, the interval where the rejection rate should lie with probability 99% under $H_0$. The dependence on traffic density (visualized by curves) is approximated by quadratic functions $\beta_4(\rho) \approx -0.000 2 \rho^2 + 0.052 8 \rho - 0.577 7$ and $\beta_5(\rho) \approx -0.000 4 \rho^2 + 0.073 1 \rho - 0.971 8$.

Figure 1. Estimated resistivity as a function of density. We plot values of statistical resistivity $\beta$ estimated by MLE (maximum likelihood estimator (5)) and MDE (minimum distance estimator (6)), respectively. Due to the scaling $g(x) = 1$, the resistivity is dimensionless. The dependence on traffic density (visualized by curves) is approximated by quadratic functions $\beta_4(\rho) \approx -0.000 2 \rho^2 + 0.052 8 \rho - 0.577 7$ and $\beta_5(\rho) \approx -0.000 4 \rho^2 + 0.073 1 \rho - 0.971 8$.

Figure 2. Estimated scaling constant $D$ as a function of density. Crosses and bullets visualize values $D = D(\beta)$ (considered in the GIG distribution (3)) estimated by MLE/MDE.

Figure 3. Rejection rates for a bootstrap method. Crosses and bullets visualize values $\phi_{MLE}$ and $\phi_{MDE}$ describing a percentage of rejected cases. The indices MLE and MDE show which of estimators has been applied/tested. A significance level is visualized by a horizontal line. The grey zone represents 99% probability interval of the asymptotic distribution of rejection rates under $H_0$. A significance level is visualized by a horizontal line. The grey zone represents 99% probability interval of the asymptotic distribution of rejection rates under $H_0$. 

J. Phys. Commun. 3 (2019) 035020 O Kollert et al
For purposes of this article, all types of headways (time headway, distance headway, time clearance, distance clearance) are called level spacings. Terminology used in this paper is inspired by the Random Matrix Theory [34] where the term level spacing represents distance between ordered eigenvalues in the unfolded spectrum of a random matrix.

3. Theory of statistical rigidity

3.1. Fundamental Terms and Properties

The key object we inspect here is a level counting process introduced by the following definition.

**Definition 1.** Let \((S_n)_{n \in \mathbb{N}}\) be a level spacing sequence and \(L_1\) an independent random variable with the density \(f_{L_1} = \lambda (1 - F_{S_1})\) and the intensity parameter \(\lambda = 1/\mathbb{E}(S_1) = 0\). Define \(L_n = L_1 + S_{n-1}\) for \(n \in \mathbb{N}\) where \(S_n = 0\) a.s. and for \(t \in \mathbb{R}_+\), define \(\nu_t = \# \{n \in \mathbb{N} \mid L_n \leq t\}\). The sequence \((L_n)_{n \in \mathbb{N}}\) is then called a level sequence and the process \((\nu_t)_{t \geq 0}\) is said to be a level counting process.

In fact, in terms of Renewal Theory [7] the level counting process is a specific variant of delayed renewal process, where the distribution of \(L_1\) is chosen specifically through the density \(f_{L_1} = \lambda (1 - F_{S_1})\) for the non-zero intensity parameter \(\lambda\). From this similarity one can profit when deriving some beneficial relations and theorems (see Appendix A.2).

3.2. Rigidity and Its Properties

The term rigidity refers to the variance of a level counting process and it represents one of the most powerful tools for describing and comparing various one-dimensional particle-like systems. The formal definition is provided below.

**Definition 2.** Let \((\nu_t)_{t \geq 0}\) be a level counting process. The function defined as \(\Delta(t) = \mathbb{V}(\nu_t)\) for \(t \in \mathbb{R}_+\) is said to be the rigidity.

Analogously, the variance of a renewal process can be defined. The function is then said to be the number variance. Nevertheless, there are numerous disadvantages of this quantity compared to the rigidity. Concerning the theoretical issues, the rigidity is actually much more convenient to handle analytically. It is confirmed by the assertion of the following theorem, which provides the expression of the rigidity through the corresponding renewal function.

**Theorem 1.** Let \(\Delta\) be a rigidity and \(\mathcal{R}\) the corresponding renewal function. Denote \(\lambda\) as the related intensity parameter. Then \(\forall t \geq 0\), it holds

\[
\Delta(t) = \lambda \left( t - 2 \int_0^t \mathcal{R}(s) \, ds \right). \tag{9}
\]

**Proof.** To get the desired relation, it is first necessary to derive the expression for the second moment of the underlying level counting process \((\nu_t)_{t \geq 0}\). Applying the assertion of the theorem 9 then yields

\[
\mathbb{E}(\nu_t)^2 = \lambda \sum_{i=0}^{\left\lfloor \frac{t}{\lambda} \right\rfloor} \binom{2}{i} \mathcal{R}_i(t) * t = \lambda t + 2\lambda \int_0^t \mathcal{R}(s) \, ds \tag{10}
\]

for \(t \geq 0\). After plugging this result into the definition of the variance of the process and using the relation (24), we obtain the desired assertion.
\[
\Delta(t) = \mathbb{E}(\nu_t)^2 - (\lambda t)^2 = \lambda t - 2\lambda \int_0^t \lambda s \, ds + 2\lambda \int_0^t \mathcal{R}(s) \, ds.
\]

Another way of deriving the expression is exploiting the assertion of the theorem 4 and the fact that the increments of a renewal process is a uniformly integrable process. Then, the variance of the increments converge to the variance of the level counting process where the formula for the number variance is

\[
\text{Var}(N_t) = \mathcal{R}(t)(1 - \mathcal{R}(t)) + 2(\mathcal{R} \ast \mathcal{R})(t)
\]

for \( t \geq 0 \). The operator \( \ast \) represents the convolution with respect to generalized functions. The formula (11) is analogous to the equality (9) and its visually even more compact, however, the relation (9) is indeed analytically more simple. The renewal function might be actually quite a complicated one and thus, the expression \( \mathcal{R} \ast \mathcal{R} \) may be extremely hard to get. Moreover, the derivative of the rigidity can be easily obtained, i.e. it is of the form

\[
\Delta'(t) = \lambda - 2\lambda(\lambda t - \mathcal{R}(t))
\]

satisfied for those \( t \geq 0 \) at which the renewal function is continuous.

We will now turn our focus onto dealing with the asymptotic behavior of rigidity. Apparently, the derived relation (12) is just shifted and scaled function \( \lambda t - \mathcal{R}(t) \). Combining that result with the well known asymptotic expansion

\[
\mathcal{R}(t) = \lambda t + \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1), \quad t \to \infty
\]

coming from the renewal theory, the derivative of rigidity subsequently converges in infinity to the constant \( \lambda/\sigma^2 \), where \( \sigma^2 = \text{Var}(S) \) and \( \mu = 1/\lambda = \mathbb{E}(S) \). So that this convergence holds, it is necessary to additionally assume the mentioned continuity of the renewal function on the neighborhood of infinity. Below, a more general result which does not require such an assumption is provided.

**Theorem 2.** Let \( \Delta \) be a rigidity and let the corresponding renewal function satisfy the asymptotic expansion (13). Denote \( \lambda \) as the related intensity parameter and \( \sigma^2 \) the variance of the related level spacing. Then \( \forall \ h \geq 0 \), it holds

\[
\lim_{t \to \infty} \Delta(t + h) - \Delta(t) = \lambda h \sigma^2 h.
\]

**Proof.** The proof is based on the relation (9). Using the expansion (13), the limit of the integral part of the investigated difference can be then expressed as

\[
\lim_{t \to \infty} \int_t^{t + h} \lambda s - \mathcal{R}(s) \, ds = \lim_{t \to \infty} \int_0^h \lambda(s + t) - \mathcal{R}(s + t) \, ds = \frac{\mu^2 - \sigma^2}{2\mu^2} h,
\]

where \( h \) is non-negative. The interchange of the limit and the integral is eligible since the integrand is bounded function on \( \mathbb{R}^+ \). In conclusion, the final result is given by

\[
\lim_{t \to \infty} \Delta(t + h) - \Delta(t) = \lambda h + \lim_{t \to \infty} \int_h^{t + h} \lambda s - \mathcal{R}(s) \, ds = \lambda h \sigma^2 h.
\]

The theorem immediately implies that the rigidity must be asymptotically linear once the renewal function has an asymptote.

We will now focus on the derivation of the intercept of the asymptote for the rigidity, which allows us to completely determine the asymptotic behavior of the corresponding rigidity. To accomplish that, the Laplace transform is necessary to be applied. Namely, the mapping defined as

\[
\phi_F(p) = \int_0^\infty e^{-pt} \, dF(t)
\]

for \( p > 0 \) and \( F \) representing some general distribution function will be used. This is very useful tool in the counting process theory since it allows one to determine the asymptotic behavior of a function through the behavior of its Laplace transform on the right neighborhood of zero. Such a procedure applied in the case of rigidity is given in the proof of the following theorem.

**Theorem 3.** Let the rigidity \( \Delta \) has an asymptotic expansion of the first order according to \( (t^{1-n})_{n \in \mathbb{N}} \). For \( k \in \mathbb{N} \), denote \( \mu_k \) as the \( k \)-th moment of the corresponding level spacing and define \( \lambda = 1/\mu_1 \) and \( \sigma^2 = \mu_2 - \mu_1^2 \). If \( \mu_3 < \infty \), then it holds

\[
\Delta(t) = \lambda \sigma^2 t + \frac{3\mu^2_2 - 2\mu_1\mu_3}{6\mu_1^3} + o(1), \quad t \to \infty.
\]
Proof. From the definition of variance and the fact that the expected value of a level counting process is linear, it suffices to perform the asymptotic expansion for the second moment of the process. Denoting \( \varrho_2(t) = \mathbb{E}(n_t)^2 \) for \( t \geq 0 \), it holds
\[
p^2 \varrho_2(p) = p\lambda + 2\lambda p \varrho_R(p),
\]
where the Laplace transform has been applied on the relation (10). Another step requires expressing the Taylor polynomial of the function \( p \varrho_R(p) \) of the second order. It is of the particular form
\[
p \varrho_R(p) = \lambda + \frac{\sigma^2 - \mu_1^2}{2\mu_1}p + \frac{3\mu_2^2 - 2\mu_1\mu_3}{12\mu_1^3}p^2 + o(p^3), \quad p \to 0^+,
\]
which is a very important relation leading finally to the following result. Assuming the existence of the asymptotic expansion of \( \varrho_2 \) to the second order according to (15), the application of the inverse Laplace transform yields
\[
\varrho_2(t) = \lambda^2 t^2 + \lambda^2 \sigma^2 t + \frac{3\mu_2^2 - 2\mu_1\mu_3}{6\mu_1^3}t^3 + o(1), \quad t \to \infty.
\]
Plugging this result into the definition of the variance gives the final asymptotic expansion for the rigidity (15).

4. Simulation experiment

In order to illustrate and check the obtained asymptotic results we carry out the following simulation experiment. We generate 10 000 independent realizations of level counting and renewal processes for \( t \in [0, 7] \) with step equal to 0.1, i.e. the length of every realization is 71 observations. On the basis of these 10 000 realizations we then calculate for each \( t \) the empirical rigidity and number variance, respectively to the process.

The level spacings of the renewal and level counting processes are generated using iid variables coming from the GIG distribution with the density given in (2) and the scaling constant
\[
D = \beta + 3 - e^{-\sqrt{\beta}}.
\]
This particular choice of constant \( D \) implies that the expectation of the GIG distribution is approximately equal to 1 so that the intensity parameter used in the definition of the level counting process (cf. definition 1) is \( \lambda = 1 \) and the corresponding first element \( L_1 \) of the level sequence has the density
\[
f_{L_1}(t) = 1 - A \int_0^t -\frac{2}{\pi}e^{-\sqrt{\beta}x} \, dx.
\]
In the case of renewal process the first level spacing \( S_1 \) has the same GIG distribution as the rest of the spacings, i.e. the one described above.

For the assumed level counting process the asymptotic expansion of rigidity given in (15) takes on the form
\[
\Delta(t) = \chi t + \varkappa + o(1), \quad t \to \infty,
\]
where
\[
\chi = \chi(\beta, D) = \sqrt{\frac{D}{\beta}} \frac{K_4 K_1^4 - K_3^2 K_1}{K_3^2}, \quad \varkappa = \varkappa(\beta, D) = \frac{K_4^2 (3K_4^2 - 2K_3 K_1)}{6K_3^4},
\]
is a so-called compressibility (see also [12, 13]),
\[
K_i = K_i(2\sqrt{\beta D}).
\]
The results obtained for rigidity and number variance are presented in figure 4. From this figure we can observe the following. First, the asymptotic approximation of rigidity works well and convergence is quick for small values of \( \beta \) and it is getting slower with increasing value of the parameter \( \beta \). Second, there is a clear visible difference between the behavior of rigidity and number variance even though the corresponding processes differ only in the first level spacing. It seems, when considering a fixed value of \( \beta \), that the asymptote for number variance has the same slope as the one for rigidity (as predicted by the theory) but the intercepts are different. Again the difference increases with growing value of the parameter \( \beta \).
So in practical applications, when applying the concept of statistical rigidity, one should be aware of the process of collecting the data. For example, if we are studying the number of vehicles in time intervals of length $x$ it is important to know how the first time interval is placed on the time axis: if it is positioned randomly or if it is placed in such a way that there is a vehicle in its left border.

5. Rigidity in empirical traffic samples

In this section we present a numerical analysis of the statistical rigidity extracted (similarly to [10]) from empirical traffic data described in section 2.2. Following the same segmentation procedure as that applied in the section 2.3, we calculate the quantity $\Delta(t)$ in all the samples being sorted by a local density. Then a course of associated rigidities is plotted in figure 5 where it is clearly evident that an asymptotical trend is linear for all density bands, i.e.

\[
\Delta(t) \approx \chi(\rho) \cdot t + \mathcal{A}(\rho) \quad (t \rightarrow +\infty).
\]  

This fully corresponds to the theory discussed. However, the compressibility $\hat{\chi}(\rho)$—determined from a linear tail of an empirical dependency $\Delta(t)$ by a linear regression—shows a substantial deviation from the theory presented. To be specific, a regression coefficient $\hat{\chi}(\rho)$ and analytical value (17) enumerated for $\beta$, $D$ (plotted in figures 1 and 2) differ significantly, as visualized in figure 6. The direct consequence of such inconsistency is the fact that the assumptions of the above-discussed theorems must necessarily be violated in real traffic flows. It means that clearances of neighboring vehicles (as random variables) are not independent which implies that traffic flow does certainly not represent a system with short-ranged interactions (according to [12]). Hence we assert that for a given driver the number of actively-followed vehicles is greater than one. It supports a contemporary opinion (published in [12]) that vehicular systems are driven by potentials/forces/rules that are definitely middle-ranged. It means that driver’s decision routines include more vehicles moving in front of a reference car.

Figure 4. Rigidity $\Delta(t)$ and number variance calculated from simulated data for different values of the parameter $\beta$. The solid line represents the theoretic asymptote given in (16).

So in practical applications, when applying the concept of statistical rigidity, one should be aware of the process of collecting the data. For example, if we are studying the number of vehicles in time intervals of length $x$ it is important to know how the first time interval is placed on the time axis: if it is positioned randomly or if it is placed in such a way that there is a vehicle in its left border.

5. Rigidity in empirical traffic samples

In this section we present a numerical analysis of the statistical rigidity extracted (similarly to [10]) from empirical traffic data described in section 2.2. Following the same segmentation procedure as that applied in the section 2.3, we calculate the quantity $\Delta(t)$ in all the samples being sorted by a local density. Then a course of associated rigidities is plotted in figure 5 where it is clearly evident that an asymptotical trend is linear for all density bands, i.e.

\[
\Delta(t) \approx \chi(\rho) \cdot t + \mathcal{A}(\rho) \quad (t \rightarrow +\infty).
\]  

This fully corresponds to the theory discussed. However, the compressibility $\hat{\chi}(\rho)$—determined from a linear tail of an empirical dependency $\Delta(t)$ by a linear regression—shows a substantial deviation from the theory presented. To be specific, a regression coefficient $\hat{\chi}(\rho)$ and analytical value (17) enumerated for $\beta$, $D$ (plotted in figures 1 and 2) differ significantly, as visualized in figure 6. The direct consequence of such inconsistency is the fact that the assumptions of the above-discussed theorems must necessarily be violated in real traffic flows. It means that clearances of neighboring vehicles (as random variables) are not independent which implies that traffic flow does certainly not represent a system with short-ranged interactions (according to [12]). Hence we assert that for a given driver the number of actively-followed vehicles is greater than one. It supports a contemporary opinion (published in [12]) that vehicular systems are driven by potentials/forces/rules that are definitely middle-ranged. It means that driver’s decision routines include more vehicles moving in front of a reference car.
6. Summary and conclusions

In the first part of the article we discuss three substantial attributes that should be taken into account if one aims to estimate vehicular clearance distributions. We explain that a necessary part of such estimations is a data segmentation preventing an undesirable mixing of states with different stochastic resistivity or compressibility. Also, we assert that any proposed statistical model (distribution family) should belong to the class of balanced distributions since otherwise the model does not fulfill a necessary condition which is generally valid in all one-dimensional particle systems with short- or middle-ranged interactions. The third essential attribute of a suggested estimation procedure is a requirement that probability density \( g(x) \) should generate a plateau in \( x = x_0 \), which implies (for smooth probability density functions) that \( g'(0) = g''(0) = g'''(0) = \ldots = 0 \). This requirement is an output of earlier studies of traffic data.

Taking into account all these attributes we suggest the GIG distribution as a suitable candidate for a respective estimation procedure. In the next part of the text we prove (by means of standardized statistical tests and the bootstrap approach) that a hypothesis on GIG-distributed vehicular clearances is justified in all sub-areas of ID-plane (a rejection rate is significantly low) and that stochastic resistivity \( \beta \) (which is a free parameter in GIG family) increases almost linearly with traffic density.

In the next part of the text we derive a direct theoretical connection between a clearance distribution and a statistical rigidity (property being observed in several investigations of traffic microstructure). We show that the statistical rigidity has a perspicuous linear asymptote whose slope \( \chi \) (called a compressibility) is (in systems where spacing are scaled to the unit average) equal to the variance \( \sigma^2 \) of spacings. All these results have been thoroughly tested and verified with help of simulation experiments.
Furthermore, we present results of statistical analysis working with data measured on Czech freeways. We show that empirical rigidity generates (in accordance with the theory) a linear tail whose slope depends on traffic density. Inconsistently with the theory, however, the empirical compressibility do not correspond to the variance $\sigma^2$ detected from clearance distributions. To be specific, in all sub-areas of ID-plane theory underestimates a real value of the compressibility. This discrepancy represents an interesting open problem in a field of vehicular headway modeling. The elementary explanation (if we disregard dynamical segregation effects such as a clustering or occurrence of dramatically faster vehicles behind a slower one) is, as understandable, hidden in the fact that vehicular spacings do not fulfill the fundamental prerequisite in the definition of a level process. Indeed, vehicular headways/clearances/gaps do not represent an iid sequence of random variables and analytical predictions therefore deviates from reality. Such a deviation generates, however, a following very strong conclusion. Vehicular systems (in a congested phase) do not represent a short-ranged system. It means, beyond all doubt, that driver reacts to more front-running vehicles, which results in a middle-ranged decision making. This knowledge is in a consonance with previously-published findings [12].

To conclude, we show that in short-ranged one-dimensional systems (where spacings are iid) the spacing distribution unambiguously determines a statistical rigidity so that a respective compressibility $\chi$ is equal to the variance $\sigma^2$ of spacings. In contrast, a significant deviation $\chi \approx \sigma^2$ detected in empirical measurements supports a recent opinion that the number of actively-followed vehicles (in all typical vehicular systems) is greater than one. Unfortunately, these findings are omitted in most traffic models commonly used.

**Acknowledgments**

The research presented in this work was supported by Grant 15-15049S of Czech Science Foundation (GA ČR) and by the Grant SGS18/188/OHK4/3T/14 provided by the Ministry of Education, Youth, and Sports of the Czech Republic (MSMT ČR).

**Appendix**

A.1. Class of balanced densities

A function $g(x): \mathbb{R} \rightarrow \mathbb{R}$ is called a **balanced density** (balanced density function) if it fulfills the following axioms:

1. **(piecewise continuity):** $g(x) \in \mathcal{Pc}(\mathbb{R})$, i.e. $g(x)$ is piecewise continuous on $\mathbb{R}$;
2. **(non-negativity of the range):** $\text{Ran}(g) \subset (0, +\infty)$, i.e. $g(x)$ is non-negative;
3. **(non-negativity of the support):** $\text{supp}(g) \subset (0, +\infty)$, i.e. $g(x) = \Theta(x)g(x)$, where $\Theta(x)$ is the Heaviside unit-step function;
4. **(Lebesgue integrability):** $g(x) \in L(\mathbb{R})$, i.e. $g(x)$ is Lebesgue integrable;
5. **(balancing criterion):** $\exists \omega > 0$ so that

$$\forall \alpha > \omega: \lim_{x \rightarrow +\infty} g(x)e^{\alpha x} = +\infty,$$  \hspace{1cm} (19)

and

$$\forall \alpha < \omega: \lim_{x \rightarrow +\infty} g(x)e^{\alpha x} = 0.$$  \hspace{1cm} (20)

The parameter $\omega$ (unambiguously determined by the balancing criterion) is then called a **balancing index** and denoted by $\mathcal{B}$. The class of balanced densities is denoted by $\mathcal{B}$. The balanced density $g(x)$ will be referred to as the **normalized (balanced) density** if $\int_{\mathbb{R}} g(x)dx = 1$. Therefore, every normalized density can be understood as the probability density function (pdf) of a certain continuous random variable.

A.2. Level counting process theory

The essential object we inspect here is a level counting process (see some important results published in [7, 8, 35–37]). To completely understand its nature, it is first necessary to deal extensively with a renewal process. However, providing the results of the renewal theory is not the purpose of this paper. Additionally, the theory is quite well known and understood [7]. Nevertheless, we will at least provide the definition of the process since there is a newly used term needed for further investigations.
Definition 3. Let \((R_i)_{i \in \mathbb{N}}\) be an iid sequence of a.s. positive random variables. Define \(S_n = \sum_{i=1}^{n} R_i\) for \(n \in \mathbb{N}\) and \(N_t = \# \{ n \in \mathbb{N} \mid S_n \leq t \}\) for \(t \in \mathbb{R}_+^\ast\). The sequence \((S_n)_{n \in \mathbb{N}}\) is then called the level spacing sequence and the process \((N_t)_{t \geq 0}\) is said to be a renewal process.

The term level spacing comes from the theory of random matrices [34] where the level spacings refer to the spacings between consecutive random eigenvalues. In this paper, the term will eventually represent the spacings between two consecutive vehicles on a road. Definition of the level counting process \((\nu_t)_{t \geq 0}\) is introduced in the main part of the paper (see definition 1). The level sequence, which determines the level counting process, is entirely defined through the level spacing sequence which determines the renewal process. It is then not a surprise that the two counting processes are very alike in terms of their structures. This close relation can be described through the equation

\[
P(\nu_t = n \mid L_1 = s) = P(N_{t-s} = n - 1)
\]

where \(n \in \mathbb{N}\) and \(t, s \geq 0\). It actually says that after the realization of the first level, the level counting process becomes the renewal one. The property (21) directly follows from the definition 1 since for \(n, k \in \mathbb{N}\), it holds

\[
L_{n+k} - L_n = L_1 + S_{n+k-1} - L_1 = S_{n+k-1} - S_{n-1} \overset{D}{=} S_k,
\]

where \(\overset{D}{=}\) means equality in distribution, i.e. \(L_{n+k} - L_n\) and \(S_k\) have the same distribution functions. In other words, the spacings between two consecutive random variables in the sequence of levels have the distributions determined by the sequence of level spacings. Additionally, that distribution depends only on the difference between indexes of the levels. Such property is apparently satisfied regardless of the distribution of \(L_1\). If the distribution of this first random variable is general, the corresponding counting process is then truly called a delayed renewal process which is also quite a well known kind of process (see [7]). Nevertheless, in the definition of a level counting process, the distribution of \(L_1\) is chosen specifically through the density \(f_{L_1} = \lambda(1 - F_{S_1})\) for the non-zero intensity parameter \(\lambda\). This particular choice puts the two counting processes into another close and important relation which is provided in the following theorem (proved in [7]).

Theorem 4. Let \((N_t)_{t \geq 0}\) be a renewal process with the non-zero intensity parameter and let \((\nu_t)_{t \geq 0}\) be an independent level counting process both defined through the same level spacing sequence. Then for \(t \to \infty\) and \(\forall y \geq 0\), it is satisfied

\[
N_{t+y} - N_t \overset{D}{=} \nu_y.
\]

The proposition above says that the random process consisting of the increments of a renewal process converge in distribution to the level counting process. To put it in another way, the properties of a level counting process are actually given by the limiting properties of the corresponding renewal process. Before we prove some crucial results, we first need the formula providing the relation between the distribution of levels and the corresponding level spacings.

Theorem 5. Let \((L_n)_{n \in \mathbb{N}}\) be a level sequence and \((S_n)_{n \in \mathbb{N}}\) the corresponding level spacing sequence. Define \(S_0 = 0\) a.s. and put \(\lambda = 1/E(S_1) \neq 0\). Then \(\forall n \in \mathbb{N}\), it holds

\[
f_{L_n} = \lambda(F_{S_{n-1}} - F_{S_n}).
\]

Proof. In the case where \(n = 1\), the proposition contains the random variable \(S_0\). Because \(S_0\) is defined to be zero almost surely, the respective distribution function takes only the value one on the non-negative part of the real line. According to the distribution of \(L_1\) introduced in the definition 1, the relation (23) is satisfied.

Suppose now that the natural \(n\) is greater than one. Using the definition of \(L_n\) as the sum of two independent random variables \(L_1\) and \(S_{n-1}\), it is possible to write

\[
f_{L_n}(t) = (f_{L_1} \ast F_{S_{n-1}})(t) = \int_0^t \lambda(1 - F_{S_1}(t - s))dF_{S_{n-1}}(s)
\]

\[
= \lambda \int_0^t dF_{S_{n-1}}(s) - \lambda \int_0^t F_{S_1}(t - s)dF_{S_{n-1}}(s),
\]

where \(t \geq 0\). That is already the desired expression considering the fact that \(S_n\) represents the sum of \(n\) iid random variables all having the same distribution as that of \(S_1\).

Equipped with the theorem 5, we can now proceed to the most important result of this section. Namely, it will be shown that the expected value of a level counting process is a linear function.
Theorem 6. Let \((\nu_t)_{t \geq 0}\) be a level counting process and \(\lambda\) the respective non-zero intensity parameter. Then \(\forall \, t \geq 0\), it holds
\[
E(\nu_t) = \lambda t. \tag{24}
\]

Proof. Here, we will use the well known result from the renewal theory stating that the expected value of a renewal process can be written as the series of all the distribution functions of the corresponding level spacings. Mathematically written, it holds
\[
E(N_t) = \sum_{n=0}^{\infty} nP(N_t = n) = \sum_{n=0}^{\infty} n(F_{S_n}(t) - F_{S_{n+1}}(t)) = \sum_{n=1}^{\infty} F_{S_n}(t) \tag{25}
\]
for \(t \geq 0\). In fact, this proposition can be actually generalized to a counting process of any kind. Its application on a level counting process along with the relation (23) yields
\[
E(\nu_t) = \lambda \int_0^t \sum_{n=1}^{\infty} F_{S_{n-1}}(s) - F_{S_n}(s) \, ds = \lambda \int_0^t F_0(s) \, ds = \lambda t. \]

From the results above, we can see that the expected value of arbitrarily chosen increments of some fixed length are still the same for a level counting process. As a matter of fact, there is a much stronger proposition saying that the increments even have an identical distribution. Such property is called homogeneity and the respective theorem (proved in [7]) is given below.

Theorem 7. A level counting process satisfies the properties of a homogenous process, i.e. \(\forall \, t, \, h \geq 0\), it holds
\[
\nu_{t+h} - \nu_t \overset{D}{=} \nu_h.
\]

The essential results of the paper are based on the expression of the moments of a level counting process through the moments of a renewal process. To conveniently perform the derivation of such a relation, we will first establish the corresponding terminology and notation.

Definition 4. Let \((N_t)_{t \geq 0}\) be a renewal process. Then for \(k \in \mathbb{N}_0\), the function \(R_k(t) := E(N_t^k)\) defined for \(t \in \mathbb{R}_0^+\) is said to be the \(k\)-th renewal function. Usually, \(R_0(t) := \mathcal{R}(t)\) is said to the renewal function.

The definition of the renewal functions makes sense since for all \(k \in \mathbb{N}\) and \(t \in \mathbb{R}_0^+\), the moments \(E(N_t^k)\) are finite. Such property holds even for the moments of a general counting process as discussed in [7]. With this kind of notation, the mentioned expression for the moments of a level counting process have more elegant and slick form. However, before providing its proof, the following auxiliary proposition needs to be proved.

Theorem 8. Let \(\mathcal{R}_k\) represent the \(k\)-th renewal function for \(k \in \mathbb{N}\) and let \(\lambda\) be the corresponding non-zero intensity parameter. Then \(\forall \, k \in \mathbb{N}_0\), it holds
\[
\sum_{n=1}^{\infty} n^k f_{L_n} = \lambda \sum_{i=0}^{k} \binom{k}{i} R_i. \tag{26}
\]

Proof. The proof is based on expressing the relation (23) through the probability distribution of the corresponding renewal process \((N_t)_{t \geq 0}\). The essential equality \(P(N_t = n) = F_{S_n}(t) - F_{S_{n+1}}(t)\) used also in (25) then allows one to write
\[
\sum_{n=1}^{\infty} n^k f_{L_n}(t) = \lambda \sum_{n=0}^{\infty} (n+1)^k P(N_t = n) = \lambda \sum_{n=0}^{\infty} \sum_{i=0}^{k} \binom{k}{i} n^i P(N_t = n) = \lambda \sum_{i=0}^{k} \binom{k}{i} R_i(t)
\]
for \(t \in \mathbb{R}_0^+\) and \(k \in \mathbb{N}_0\). To get the last equality, the sums need to be swapped and consequently, the definition of the function \(\mathcal{R}_i\) as the \(i\)-th moment of a renewal process is necessary to be applied.

Just proven relation involves already obtained result related to the linearity of the expected value of a level counting process. Namely, since \(\mathcal{R}_0 = 1\) on the non-negative part of the real line, the equality (26) for \(k = 0\) represents the derivative of \(E(\nu_t)\) as discussed in the proof of the theorem 6. The generalization of the relation (24) is also included in the following proposition.

Theorem 9. Let \((\nu_t)_{t \geq 0}\) be a level counting process with the intensity parameter \(\lambda\) and for \(k \in \mathbb{N}\), let \(\mathcal{R}_k\) be the corresponding \(k\)-th renewal function. Then \(\forall \, t \in \mathbb{R}_0^+\) and \(\forall \, k \in \mathbb{N}\), it holds
\[
\sum_{n=1}^{\infty} n^k f_{L_n}(t) = \lambda \sum_{i=0}^{k} \binom{k}{i} R_i(t).
\]
Proof. In the proofs of the theorems 6 and 8, we used the expression of the distribution of a renewal process as the difference of the distribution functions of the corresponding level spacings. Again, such a property holds generally for any counting process. Applying that result for a level counting process then yields the equality

\[ E(\nu_t)^k = \lambda \int_0^t \sum_{i=0}^{k-1} \binom{k}{i} R_i(s) ds. \]  

(27)

A.3. Testing the alternative hypothesis

In this part of Appendix we test the hypothesis \( \tilde{H}_0 \): Empirical time clearances (detected in congested traffic flows) are Gamma-distributed, which represents a natural alternative to the main hypothesis \( H_0 \) studied in section 2.3.

Applying the bootstrap approach with same setting parameters one can see (compare figures 3 and A1) that rejection rates for \( \tilde{H}_0 \) are several times higher than in case of \( H_0 \). This confirm the well-known fact \([6]\) that Gamma distribution is not a suitable candidate for vehicular clearance distributions.

A.4. Calculation of the local density in a data sample

Consider one sample \( \{y_1, y_2, \ldots, y_n\} \) of \( n \) succeeding time headways (brutto time intervals between vehicles) and a set \( \{v_1, v_2, \ldots, v_n\} \) of corresponding velocities. Then, using the fluid-dynamic equation \( \rho = \rho v \) as discussed in [2] and applied e.g. in [23]), one can calculate a local density (density associated to the selected sample) by means of

\[ \theta = \frac{n^2}{\sum_{k=1}^{n} y_k \cdot \sum_{k=1}^{n} v_k}. \]
[3] Treiber M and Kesting A 2013 *Traffic Flow Dynamics* (Berlin: Springer)

[4] Krbálek M and Helbing D 2004 *Physica A* 333 370

[5] Krbálek M 2007 *J. Phys. A: Math. Theor.* 40 5813

[6] Krbálek M and Sleis J 2015 *J. Phys. A: Math. Theor.* 48 015101

[7] Cox D R 1962 *Renewal Theory* (London, New York: Methuen & Co.)

[8] Cox D R and Lewis P A W 1978 *The statistical Analysis of Series of Events* (England: Chapman & Hall)

[9] Nishinari K, Helbing D and Treiber M 2003 *Phys. Rev. E* 68 067101

[10] Krbálek M and Šeba P 2009 *J. Phys. A: Math. Theor.* 42 345001

[11] Krbálek M 2010 *Kybernetika* 46 1108

[12] Krbálek M, Apeltauer J, Apeltauer T and Szabová Z 2018 *Physica A* 491 112

[13] Krbálek M, Hrabák P and Bukáček M 2018 *J. Phys. A: Math. Gen.* 33 L229

[14] Krbálek M 2010 *Kybernetika* 46 1108

[15] Krbálek M, Apeltauer J, Apeltauer T and Szabová Z 2018 *Physica A* 491 112

[16] Krbálek M and Hobza T 2016 *Phys. Lett. A* 380 1839

[17] Ha D, Aron M and Cohen S 2012 *Transportation Research Part C* 25 181

[18] Rossi R and Gaspard P 2012 *Proceedings—Social and Behavioral Science* 54 865

[19] Chen X, Li Z and Li L 2014 *Transportation Research Part C* 30 820

[20] Johnson N L, Kotz S and Balakrishnan N 1994 *Continuous Univariate Distributions* 1 2nd edn (New York: John Wiley & Sons)

[21] Treiber M, Hennecke A and Helbing D 2000 *Phys. Rev. E* 62 1805

[22] Mehta M L 2004 *Random Matrices* (New York: Academic)

[23] Gut A 2009 *Stopped Random Walks: Limit Theorems and Applications* (New York: Springer)

[24] Mikosh T 2004 *An introduction with stochastic processes* (Berlin: Springer)

[25] Bohigas O and Giannoni M J 1984 *Chaotic motion and random matrix theories* (Berlin: Springer)