Risk Assessment of Autonomous Vehicles across Diverse Driving Contexts

Akhil Shetty[⋆], Hamidreza Tavafoghi[†], Alex Kurzhanskiy[‡], Kameshwar Poolla[⋆†], and Pravin Varaiya[⋆]

***Abstract*—Traffic crashes are a leading cause of death in the US. The crashes cost more than 36,000 lives and close to $1T in economic loss each year. Autonomous vehicles (AVs) promise a future without crashes. But deploying AVs without adequately assessing their safety might lead to an increase in crashes rather than a reduction. Extensive on-road testing is needed to ensure that AVs bring the intended safety benefits. However, testing AVs across all possible driving contexts is impractical. Moreover, since crashes are rare events, this requires approaches to evaluate AV safety that account for diversity in driving contexts, without testing in all possible scenarios. In this paper, we present a risk assessment framework that uses on-road testing data to provide insights into the safety of AVs across diverse maneuvers and environments. We derive human crash risk baselines to interpret AV safety over the same maneuvers and environments. We also present use cases for our risk assessment framework and suggest how regulators could use it to make decisions about the introduction of AVs in their jurisdictions.**

***Index Terms*—Autonomous vehicles, risk assessment, traffic safety, crashes.**

I. Introduction

Traffic crashes are a leading cause of death in the US [1]. In 2018, more than 36 thousand lives were lost to mishaps on the road [2]. Moreover, 20% of these fatalities were vulnerable road users such as pedestrians and bicyclists. The economic cost of these crashes is estimated to be close to $1T [3].

Autonomous Vehicles (AVs) have recently emerged as a promising solution for eliminating traffic crashes. Since a majority of these crashes result from human error [4], AVs aim to usher in an era of near-zero traffic fatalities by removing humans from the driver's seat. Billions of dollars have been invested across the globe in pursuit of this vision [5]. A report by Intel [6] forecasts that AVs will save 600 thousand lives and $230B in safety costs by 2045.

While AV technology has improved over the last decade, widespread deployment of AVs is still far from being achieved. Designing an AV to be safe on the roads has turned out to be more difficult than originally anticipated. Driving on the road involves navigating complex environments, unpredictable road user behavior, and challenging road and weather conditions. A considerable fraction of crashes are a direct consequence of these intrinsic challenges of driving rather than mistakes made by involved agents [7], [8]. Indeed, several types of crashes are likely to persist in a future with AVs [8]–[10].

Recognizing that some crashes are inevitable, risk assessment is essential before AVs are deployed. The AV industry currently uses three approaches for safety testing: (i) Simulation, (ii) Closed Circuit, and (iii) On-road. Each of these approaches has its own advantages and limitations. Evaluating an AV's safety performance via simulation allows testing over billions of miles, but the simulator is only an approximation of real-world driving. Crashes being rare events, small differences in the simulated and real-world environments can translate into large deviations in the inferences drawn about safety performance. Closed Circuit testing aims to overcome this limitation by replicating real-world conditions and testing AVs in challenging scenarios. However, recreating the immense diversity of real-world driving conditions and road-user behavior is impractical and hence such testing provides an incomplete appraisal of an AV's safety. On-road testing is the gold standard for predicting an AV's performance on future deployment on the roads as there is no gap between the conditions in which the AV is tested and deployed.

Several companies are now testing their AVs in limited operational design domains (ODDs) deemed safe enough for operation. The objectives of on-road testing are: (i) improving AV technology based on the experience gained from driving on the roads, and (ii) demonstrating that AVs are indeed safe enough to be deployed on a large scale. While AV companies are achieving the first objective, the path to the second seems less clear. Several works argue that AVs cannot demonstrate their safety capabilities solely based on on-road testing [11]–[13]. In particular, a study by the RAND corporation [11] suggests that AVs would have to be tested over billions of miles to demonstrate a fatality rate lower than that of human drivers. It would require decades of testing for current AV fleets to cover such a distance. On the other hand, the easier task of demonstrating a lower crash rate would require on the order of millions of miles of driving. This range is well within reach of current AV fleets. Since it is conceivable that lower crash rates would result in lower fatality rates, one might be tempted to draw conclusions about future safety benefits of AV technology based on on-road testing.

This argument breaks down if the diversity of driving scenarios is taken into account. The RAND study [11] uses crashes per mile as the safety metric for comparing AVs and human drivers. Since AVs are currently being tested only in

[⋆]The Department of Electrical Engineering and Computer Science, University of California, Berkeley, USA. Email: {shetty.akhil, poolla, varaiya}@berkeley.edu

[†]The Department of Mechanical Engineering, University of California, Berkeley, USA. Email: tavaf@berkeley.edu

[‡]Partners for Advanced Transportation Technology, University of California, Berkeley, USA. Email: akurzhan@berkeley.edu

limited ODDs, they are not exposed to the wide variety of driving conditions and road user behavior that human drivers come across in billions of miles of cumulative driving every year. Driving a million miles on sparse rural highways is very different from covering the same distance in dense urban settings. Clearly, the context in which these miles are driven is crucial for interpreting the results of on-road testing. Thus, crashes per mile is an insufficient metric for judging AV safety.

This raises the question: *How should we assess the crash risk of AVs?* We have argued so far that more information than just the total number of crashes and miles driven is needed. Since the promise of AVs is to make our roads safer, safety metrics must reflect the efficacy of AVs across all driving contexts that humans encounter. Our risk assessment framework for AVs should be such that successful testing on sparse rural highways does not imply that they would be safe while navigating dense urban intersections. Rather than having a single performance metric for assessing risk, it is desirable to have a collection of metrics that capture safety performance across diverse contexts. In order to do so, we must first come up with a suitable level of abstraction at which the diversity of driving contexts can be expressed in a tractable manner.

The crash typology presented by NHTSA in [14] is a useful starting point in this endeavour. This typology classifies crashes into 36 types based on the maneuvers of involved vehicles leading up to the crash. For instance, the *Changing Lanes* crash type involves all crashes involving a lane change. This typology is motivated by the observation that there is significant variance in crash risk across driving maneuvers. For example, left turns are twice as likely to result in a crash as compared with right turns. This suggests that analyzing performance across diverse driving maneuvers would be an important first step towards an effective risk assessment framework for AVs.

Another important source of diversity is the environment in which maneuvers are performed. A lane change executed on urban highways with small traffic gaps has a much greater crash risk compared to the same maneuver on sparse rural highways with larger gaps. Foggy weather and slippery roads can substantially increase crash risk as well. Thus, diversity of driving environment is another factor that needs to be captured in our risk assessment framework.

Risk assessment of AVs is of particular interest to regulators who must balance the safety benefits of this promising technology with its potential pitfalls. Failing to impose the required safety standards before deploying AVs can be fatal. For instance, the National Transportation Safety Board concluded that the absence of regulatory oversight contributed to the fatal crash involving an Uber AV in Tempe, Arizona [15]. Currently, states are free to set their own standards for AV testing and deployment. To the best of our knowledge, there are no standardized safety performance metrics used by regulators to govern the deployment of AVs in their jurisdiction. As a result, what kind of testing data should be disclosed by AV companies to regulators for risk assessment

Fig. 1: Top View of Diamond Street and Pacific Coast Highway Intersection (Source: Google Maps)

is an open question.

How should the diversity of driving maneuvers and environment be taken into account during crash risk assessment of AVs? We aim to address this question through our work. Our contributions are as follows:

- We develop a crash risk assessment framework for AVs that provides insights about an AV's safety performance across diverse driving contexts.
- We derive similar crash risk estimates for an average human driver using police crash reports and fine-grained vehicle maneuver data. Since AVs are expected to be significantly better than their human counterparts, these estimates can be used as a baseline to interpret AV safety performance.
- We present several use cases for our risk assessment framework and offer suggestions on how regulators could use it to make decisions about the introduction of AVs in their jurisdictions.

II. AV Crash Risk Assessment

A. The Challenge of Risk Assessment: An Illustrative Example

Consider an AV making an unprotected left turn from the Pacific Coast Highway (PCH) on to Diamond Street, as shown in Figure 1. Suppose we wish to estimate the probability of the AV getting into a crash. Several obstacles could prevent the AV from making this turn safely:

- Through-moving traffic in the opposite direction on PCH with a conflicting trajectory,
- Pedestrians on the crosswalk while the AV makes the left turn,
- Weather conditions that affect visibility and control for both the AV and neighboring vehicles (e.g., foggy weather, smoke, night time)
- Uncertainty in the behavior of neighboring vehicles/pedestrians.

An accurate estimate of the AV's crash probability would have to take all the above factors into account. First, we would need an accurate model of vehicle and pedestrian arrivals at the intersection. Second and arguably more difficult, is the task of probabilistically modeling the response of oncoming traffic to the AV's left turn maneuver. This includes the time taken to detect the AV (reaction time), decision process (to slow or not to slow down) and possibly even evasive maneuvers (e.g., brake, swerve) if the neighboring vehicle was too close to the intersection. Adverse weather conditions would complicate the situation further by modifying the behavior of each road user. Arriving at an accurate estimate of the AV's crash probability in this specific scenario using a mechanistic model of road user behavior seems intractable.

B. Quantifying Crash Risk

This motivates the use of data-driven statistics to describe crash risk. A common statistic used to describe AV safety performance is crashes or disengagements per mile driven.[1] The appeal of this safety metric lies in its simplicity; all that needs to be known is the number of crashes or disengagements encountered by AVs and the total number of miles driven. Moreover, it provides a seemingly straightforward solution for comparing AV safety performance with that of humans. For instance, the crash rate for human driven vehicles in the US is 1.9 per million miles driven [14]. Considering that AVs have already been tested on the roads over millions of miles, it seems reasonable to compare AV crashes per million miles to the above number.

However, not all miles driven are equal from a traffic safety standpoint. Crash rates differ significantly depending on the type of maneuvers performed while driving. As a case in point, let us compare two maneuvers: (i) maintaining lane, and (ii) making a left turn. Rear-ends are the most common cause of crashes while maintaining lane, and can occur anywhere on the lane. As rear-ends account for 30% of all crashes [14], we can conclude that there is a probability of 6.3×10^{-7} of getting into a rear-end crash for every mile of maintaining lane. On the other hand, left turn crashes can only occur on a short stretch of road leading up to an intersection. A typical left turn involves covering a distance of less than 20 m. It is reasonable to assume that human drivers make less than one left turn per mile on average across all of driving. Considering that left turns are involved in 7% of all crashes, we can conclude that the crash probability per left turn mile must be higher than 1.1×10^{-5}. This is roughly 175 times larger than that for maintaining lane. Clearly, the type of maneuver matters while assessing crash risk.

The total number of miles driven alone cannot inform us about the frequency of different types of maneuvers performed. It would be fair to compare AV performance with the human baseline so long as they are tested across all the diverse maneuvers and environments that humans drive in. However, this is far from the case for AVs at present, which have only been tested in environments deemed safe enough for operation. Driving crash-free over millions of miles of rural highways has very different safety implications compared to doing the same in dense urban settings. Thus, focusing on the crashes per mile safety metric can lull us into a false sense of security about the real safety capabilities of AVs.

[1]California DMV requires AV companies to report how often their vehicles disengaged from autonomous mode during tests because of technology failure or situations requiring the test driver to take manual control of the vehicle to operate safely. Since each AV company has its own rules to determine when a disengagement is warranted, the disengagement rates among companies are not comparable.

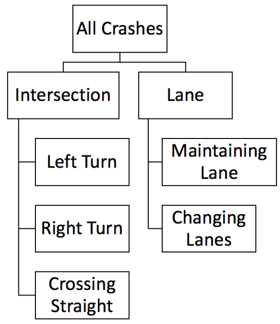

Fig. 2: Decomposition of all crashes based on maneuvers of vehicles involved.

C. Maneuver Level Crash Analysis

We saw in the discussion above how pooling together all of driving in the bucket of total miles driven hides away details about an AV's performance over diverse scenarios that are crucial to assess its crash risk. A first step in incorporating such diversity would be to condition driving based on the type of maneuver. The process of driving can be classified into two broad maneuver groups: (i) intersection, and (ii) lane maneuvers. The former includes such maneuvers as turns and going straight that are related to navigating through an intersection, while the latter consists of lane-related maneuvers such as staying in lane and changing lanes. The frequency of performing these maneuvers depends crucially on the driving environment. Intersection maneuvers are much more common in urban driving as opposed to freeways. Thus, estimating crash risk for specific maneuvers would enable us to put the safety statistics of AVs into context and provide better insights into whether their performance is likely to generalize to regions they haven't been tested in before.

III. Modeling Crash Risk

A. A Simple Maneuver-Level Crash Risk Model

Let us start by developing a simple model to estimate crash probabilities for specific maneuvers. Let us take the example of left-turn crashes. We would like to estimate the probability of a left turn crash at a particular intersection. Suppose each vehicle making a left turn has a probability p of getting into a crash. Let f denote the average rate of left turns at this intersection. Then, the number of left turn crashes C in time T is distributed as

$$C \sim \mathrm{Bin}(fT, p). \tag{1}$$

The maximum likelihood estimate for p in this case is the empirical rear-end crash probability,

$$\hat{p} = \frac{C}{fT}. \tag{2}$$

This provides a simple, intuitive estimate for the probability of a left-turn crash. Note that this estimate is very similar in spirit to the crashes per mile statistic, the only difference being the conditioning on maneuver as opposed to considering all crashes to be equivalent. More generally, such an estimate for maneuver m can be expressed as

$$\hat{p}_m = \frac{C_m}{f_m T}, \tag{3}$$

where C_m denotes the number of crashes involving maneuver m in duration T and f_m represents the average rate of maneuver m.

B. Comparing Human and AV Crash Risk

Waymo's recent safety report [16] described the safety performance of its AVs over 6.1 million miles of driving in Maricopa County, Arizona. Notably, it is the first company to provide a comprehensive breakdown of crashes its AVs were involved in based on maneuver type. Even so, this does not allow us to estimate crash probabilities for maneuvers such as left turns or lane changes since data on the frequency of such maneuvers remains unavailable. However, the crash probability per mile of staying in lane can be computed based on the given data. This enables us to compare AV and human crash risk for this maneuver.

Four types of crashes can occur while staying in lane: (i) Rear-end, (ii) Other vehicle merging into lane, (iii) Road departure, and (iv) Head-on with vehicle in opposite direction. Together they accounted for 25 of the Waymo's 47 crashes. Since an overwhelming majority of all miles driven involves the staying in lane maneuver, the empirical probability of an AV crashing during this maneuver is $25/(6.1 \times 10^6) = 4.1 \times 10^{-6}$ per mile driven. In order to derive a similar estimate for human drivers, we need to know the number of such crashes and the total vehicle miles travelled in the same region. There were 97,105 crashes in Maricopa County over 37.9 billion miles driven in 2019 [17]. While the distribution of all crashes based on maneuver is not available for Maricopa County, such a breakdown is available for fatal crashes. Observing that this distribution across crash types is similar to that for the entire country, we use the US crash distribution across types to estimate the fraction of crashes corresponding to the staying in lane maneuver [14]. This accounts for 70% of all crashes. Therefore, the human crash probability during the staying in lane maneuver is $(97,105 \times 0.7)/(37.9 \times 10^9) = 1.8 \times 10^{-6}$ per mile. One caveat here is that the human crash statistics are based on police-reports. However, a study by NHTSA [18] found that 30% of all crashes go unreported on average. Accounting for this, the crash probability estimate turns out to be 2.3×10^{-6} per mile. Observe that that the AV crash risk for this maneuver is about twice that for human drivers.

C. Data Required for Estimation

For estimating maneuver-level crash probabilities based on (3), two quantities must be known:

- Number of crashes involving maneuver m,
- Average rate of maneuver m.

At present, these quantities have not been made publicly available by most AV companies. Regulations for on-road testing data disclosure vary based on jurisdiction. While California requires AV companies operating in the state to report the circumstances leading to each disengagement, the state of Arizona has no such requirement. Even in California, companies are required to report only the total number of miles driven, not the frequency of maneuvers their AVs were involved in. Moreover, AVs have been tested in very limited on-road environments. Thus, even AV companies currently do not have data on how their AVs would function across diverse regions and road conditions that humans drive in. In the absence of this data, it is not possible to estimate maneuver-level crash probabilities using (3).

IV. Leveraging Human Driving Data

Even if the required data were available to compute crash risk estimates, there is still a need for baselines to interpret them. The human crash risk for the same maneuvers is an intuitive benchmark for this purpose. Moreover, even though AVs might drive differently compared to humans, their crash risk is likely to be correlated with that of humans due to intrinsic challenges associated with driving on the roads. As a case in point, the left turn example discussed in Section II-A is challenging for both humans and AVs. For maneuvers and road conditions in which AVs do not have suitable driving experience, the corresponding crash risk for humans is a useful indicator of the AV crash risk.

While on-road testing of AVs is at a nascent stage, humans drive billions of miles on the roads every year across a multitude of road conditions. As a result, much more data is available about human driving performance. Two sources of data are of particular interest in the estimation of maneuver-level crash risk: (i) Crash Reports, and (ii) Fine-grained Vehicle Maneuver Data.

A. Crash Reports

Police reports filed in the aftermath of a crash provide detailed information about the events leading to a crash. As a part of the broad trend towards making governmental data openly accessible, many jurisdictions in the US have uploaded their historical crash data online in an easy-to-query format. We use one such data source, the Transportation Injury Mapping System (TIMS) [19], which is a web-based application that presents crash records for the state of California in a map-centric format. This enables an understanding of traffic safety in a particular jurisdiction beyond just knowing the total number of crashes. For instance, we can query the number of crashes involving a vehicle making a left turn colliding with a pedestrian at a particular intersection over a specified time period.

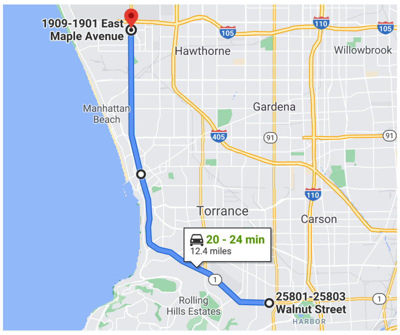

Fig. 3: Torrance, California

B. Fine-grained Vehicle Maneuver Data

As noted in Section III-C, knowledge of crash data alone is not sufficient to estimate maneuver-level crash risk. The frequency of vehicle maneuvers is crucial to put the total number of crashes into context. While vehicle flow and speed are commonly measured for transportation planning purposes, fine-grained maneuver level data such as the number of left turns or lane changes on a particular stretch of road is typically unknown. Recent advancements in technology have led to the emergence of companies collecting GPS traces of vehicles. This has enabled the measurement of previously inaccessible maneuver level data at scale without having to rely on expensive instrumentation of traffic infrastructure. We utilize a trace data set provided by Wejo Ltd., a connected vehicle data collection company, for a 12.4-mile stretch along the Pacific Coast Highway in Torrance, California which includes 29 major signalized intersections (shown in Figure 3). The data set provides information about (location, speed, heading, accelerations/brakes) of each vehicle at 3 s intervals.These traces were collected from General Motors (GM) vehicles that were equipped with an *enhanced* GPS device with higher accuracy ($\pm$ 1.5 m) that enables measurement of location with lane-level precision. We also employ a data set provided by Sensys Networks Inc. containing measurements by in-ground sensors at intersections along the same stretch. This allows us to extract lane-level information about vehicle flow at 29 intersections on the 12.

Using crash reports, vehicle trace data, and vehicle detection measurements, we can estimate maneuver-level crash probabilities for an average human driver.

C. Deriving Crash Risk Estimates

Suppose we wish to estimate the probability of a crash while making a left turn in the region illustrated in Figure 3. Let C_{lt} denote the number of left turn crashes in this region between January 2011 and December 2019, i.e., $T = 9$ yrs. Let R_{lt} denote the left turn rate in this region. Using (3), the probability of a left turn crash on the Pacific Coast Highway is estimated as

$$\hat{p}_{\text{lt}} = \frac{C_{\text{lt}}}{R_{\text{lt}} T}. \tag{4}$$

Plugging in $C_{\text{lt}} = 115$ and $R_{\text{lt}} = 1868.4$/hr based on crash reports and vehicle maneuver data for this region, we have $\hat{p}_{\text{lt}} = 7.8 \times 10^{-7}$. Figure 4 illustrates the number of crashes and estimated crash probability across the maneuvers discussed in Section II-C. Observe that there is significant variation in the estimated crash probability across maneuvers. Although most crashes occur while maintaining lane, left turns have the highest crash probability among all maneuvers. This underscores the importance of incorporating the frequency of maneuvers for interpreting crash statistics.

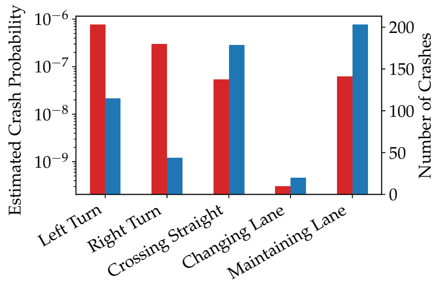

Fig. 4: Estimated crash probability (red) and number of crashes (blue) across diverse maneuvers. While maintaining lane is associated with the most number of crashes, left turns have the highest crash probability.

V. Refined Model

A. Diversity of Environment

In previous sections, we discussed how crash probabilities can vary significantly across maneuvers. From human driving data, we find that there is considerable variance in crash risk across locations even for the same maneuver. Consider the Diamond Street and Hawthorne Boulevard intersections along the Pacific Coast Highway in Torrance, California. The estimated probability of a crash while making a left turn at Diamond Street is 6.7×10^{-6}, whereas it is 1.7×10^{-7} at Hawthorne Boulevard. In other words, the average human driver making a left turn at Diamond Street is more than 20 times as likely to get into a crash as compared with the same maneuver at Hawthorne Boulevard. Thus, it is important that the environment in which a maneuver is performed is taken into account while assessing crash risk. One way to do this is to model crash risk for (maneuver, intersection) pairs rather than the maneuver alone as in (3). Consider a vehicle performing maneuver m at intersection i. Then, the model in (1) changes to

$$C_m^i \sim \text{Bin}(f_m^i T, p_m^i), \tag{5}$$

and the associated maximum likelihood estimate for p_m^i is given by

$$\hat{p}_m^i = \frac{C_m^i}{f_m^i T}. \tag{6}$$

Such crash probabilities for (maneuver, intersection) pairs can be computed at scale for human drivers using the data sources discussed in Section IV. Note that in order to derive meaningful conclusions, these estimates need to be computed over millions of maneuvers at each intersection. While this is

Intersection	Empirical Crash Probability	Left Turn Rate (turns/hr)	95% Lower Confidence Bound	95% Upper Confidence Bound
Diamond Street	6.7×10^{-6}	39.9	4.2×10^{-6}	1.0×10^{-5}
8th Street	5.9×10^{-6}	15.1	2.6×10^{-6}	1.3×10^{-5}
Rolling Hills Way	1.6×10^{-6}	46.2	6.7×10^{-7}	3.8×10^{-6}
Calle Mayor	9.8×10^{-8}	130.1	5.1×10^{-9}	6.3×10^{-7}
Palos Verdes Boulevard	9.3×10^{-8}	136.0	4.9×10^{-9}	6.1×10^{-7}
Prospect Avenue	0	30.2	3.8×10^{-8}	2.0×10^{-6}

TABLE I: Confidence Intervals for left-turn crash probability at selected intersections with high (Diamond, 8th, Rolling Hills) and low (Calle, Palos Verdes, Prospect) empirical crash probabilities. Observe that despite Prospect Avenue having zero crashes, its upper confidence bound is higher than that of Palos Verdes Boulevard since it has a considerably lower left turn rate.

true for human driving data which has been collected across millions of drivers over decades, it would take hundreds of years of on-road testing by AV companies to arrive at statistically valid estimates. A potential solution to tackle this dearth of testing data is to model crash risk for broad classes of intersections rather than modeling each intersection individually. We present an example of such a classification based on human crash risk in Section VI-A.

B. *Uncertainty about Estimate*

While the empirical crash probability estimate in (6) can be used to assess risk, it has two major drawbacks. First, it is uninformative for intersections with zero crashes. Moreover, being a point estimate, it does not quantify the degree of confidence associated with the estimate. For instance, we should be much less confident about our estimate for an intersection with limited crash and flow data. One way to circumvent both drawbacks is to use confidence intervals of p to assess risk in addition to the empirical crash probability. Let $[\hat{p}_{\min}, \hat{p}_{\max}]$ denote the 95%-confidence interval for p based on the available data. Even for intersections having zero crashes, we can be reasonably confident that $\hat{p}_{\max}$ upper bounds the actual crash probability p. We present such confidence intervals for intersections with high and low empirical left turn crash probabilities in Table I. We use the Wilson score with continuity correction method [20] for deriving these confidence bounds. Notice that even though Prospect Avenue has zero crashes, its upper confidence bound is higher than that of Palos Verdes Boulevard since it has a considerably lower left turn rate.

C. *Context of Maneuver*

The crash risk estimates derived so far answer the following question: *What is the apriori probability of a crash while performing a certain maneuver at a particular location?* Here, the use of *apriori* indicates that these estimates do not rely on real-time information such as the locations and behavior of neighboring vehicles, traffic density, weather conditions or time of day. Such factors that inform the context in which a maneuver is performed can have a significance influence on crash risk. For example, the presence of occlusions or the observation of risky behavior by neighboring vehicles is an indicator of heightened risk of a crash. While crash reports contain information about weather conditions and time of day, insights about neighboring vehicle behavior and traffic density are generally absent. On the other hand, AV companies have access to such data from on-road testing. This can be used to compute crash probabilities for specific contexts in which maneuvers are performed.

As an example, consider an AV making an unprotected left turn in the context of its field of view of conflicting traffic being occluded due to a queue of vehicles in the opposing lane. Through on-road testing data, AV companies have access to the number of such left turns made and the resulting number of crashes. This can be used to calculate the empirical probability of a crash in this context.

VI. Use Cases

We now present several use cases for the risk assessment framework developed in previous sections.

A. *Classifying Intersections based on Crash Risk*

As discussed in Section V-A, modeling crash risk for individual intersections is an intractable task for AVs. In such a situation, an intermediate level of abstraction for environment diversity is required that suitably navigates the trade-off between accuracy of crash risk estimates and data scarcity. A potential candidate for such an abstraction is the classification of intersections based on maneuver level crash risk. Let us come up with one such taxonomy for the intersections along the Pacific Coast Highway in Torrance, California. As seen from Figure 4, maintaining lane and left turn are the lane and intersection related maneuvers respectively with the highest crash probability. Let us first consider a taxonomy of intersections based on the number of left turn and maintaining lane crashes, as shown on the left in Figure 5. Observe that Type 4 intersections have a high number of crashes of both types and hence, can be considered as hazardous based on this metric. The crash probabilities for the same intersections are illustrated on the right in Figure 5. This plot paints a slightly different picture. For instance, some intersections that are classified as Type 4 based on number of crashes would be classified as Type 2 based on crash probability. The difference between the two metrics is subtle but nonetheless important. Traffic planners typically rely on the former metric for identifying hazardous intersections as they aim to minimize the total number of crashes in their jurisdiction. On the other hand, crash probabilities are more relevant for AV fleets as they aim to minimize their own crash risk rather than the aggregate crash risk over all road users. Thus, a taxonomy of intersections based on crash probability is more suitable for incorporating environment diversity in AV risk assessment; as discussed in Section V-A, this would enable AVs to aggregate their performance in a context-aware manner based on the above taxonomy. Note that the above example is only one out of the many ways in which environment diversity can be captured based on crash risk.

B. Route Risk

While it is conceivable that AVs will eliminate a significant fraction of crashes with improvements in technology, some types of crashes are likely to persist [7], [10]. As a result, AVs need to be able to mitigate risk while driving on the roads. One way to do this is to consider not just delay but also crash risk in its route planning stage. We now discuss how an AV can go about estimating this risk.

A route can be expressed as a sequence of intersection-maneuver pairs $R = \{r_t\}_{t=0}^{L}$ where $r_t = (i_t, m_t)$ denotes the t^{th} pair in the sequence. Let P_t represent the crash probability estimate at the t^{th} step in the sequence. Then, the probability of a crash occurring over the entire route is given by

$$P(\text{Crash along route } R) = 1 - \Pi_{t=0}^{L}(1 - P_t), \quad (7)$$

$$\approx \sum_t \hat{p}_{m_t}^{i_t}. \quad (8)$$

Let us compute such a route risk estimate for the average human driver along the route shown in Figure 6. This route involves the following sequence of maneuvers:

1) Left turn from Diamond Street on to Pacific Coast Highway,
2) Staying in lane between Diamond Street and Carnelian Street,
3) Going straight through the Carnelian Street intersection,
4) Lane change leading up to Beryl Street intersection,
5) Right turn from Pacific Coast Highway on to Beryl Street.

Plugging in the crash probability estimates for each of the maneuvers, we estimate a crash probability of 6.8×10^{-6} for the given route.

C. Economic Cost of Crashes

Although the terms crash probability and risk have been used interchangeably in the preceding sections, crash probability is not the only risk metric of interest. In particular, the business case for AV fleets being used for ride-hailing and delivery services requires an understanding of the economic costs of crashes. A maneuver level analysis is critical for such an understanding. For instance, although rear-ends are the most common type of crashes, crossing path crashes have the highest economic cost since they cause about twice the damage as rear-ends on average [14]. The crash probability estimate derived in (3) can be used to estimate the corresponding economic cost. Let e_m denote the average economic cost of a crash involving maneuver m. Then, the expected economic cost of a crash involving maneuver m is given by

$$\hat{E}_m = \hat{p}_m e_m. \quad (9)$$

Let us compare the expected economic cost of left and right turns along the stretch of Pacific Coast Highway shown in Figure 3. Using (9) and estimates for economic cost of crashes from [14], we find that the expected economic cost over 1000 left turns is \$100.8, while that for a right turn is \$14.2. The huge difference in the economic costs of these two seemingly similar maneuvers suggests that careful route planning can provide substantial safety benefits. This also explains why fleet operators such as UPS plan routes that avoid left turns [21]. Estimates for the expected economic cost for other maneuvers are presented in Figure 7. Although the crossing straight maneuver has the highest economic cost per crash, left turns have the highest expected economic cost per maneuver since their crash probability is significantly higher.

VII. Conclusion

Autonomous vehicles hold great promise for leading us to a future without crashes. At the same time, deploying them without adequate assessment of their safety capabilities might lead to increase in crashes rather than a reduction. Extensive on-road testing is crucial to ensure that AVs bring about the intended safety benefits.

However, testing AVs in all possible driving contexts faced by humans is impractical. Coupled with the fact that crashes are rare events requires approaches for evaluating AV safety that account for diversity in driving contexts, yet do not require testing over all possible scenarios. In this paper, we presented one such risk assessment framework that uses on-road testing data to provide insights into the safety performance of AVs across diverse maneuvers and environments. We derived human crash risk baselines that can be used to interpret an AV's safety performance over the same maneuvers and environments. Several applications for this framework were also presented.

The above risk assessment framework can be used to inform regulations governing the testing and deployment of AVs. In particular, it suggests that regulators should require AV companies to disclose data not only regarding crashes but also the frequency of the various maneuvers their AVs performed as well as the context in which they were performed. Our framework is just one of out of the many different ways in which diversity of driving contexts can be accounted for in crash risk assessment. We hope that this work motivates further research along this thread.

Acknowledgement

This research was supported by National Science Foundation EAGER award 1839843, Berkeley Deep Drive, and California Department of Transportation. We are grateful to Jo Birch of Wejo, Ltd. and Christopher Flores of Sensys Networks, Inc. for data.

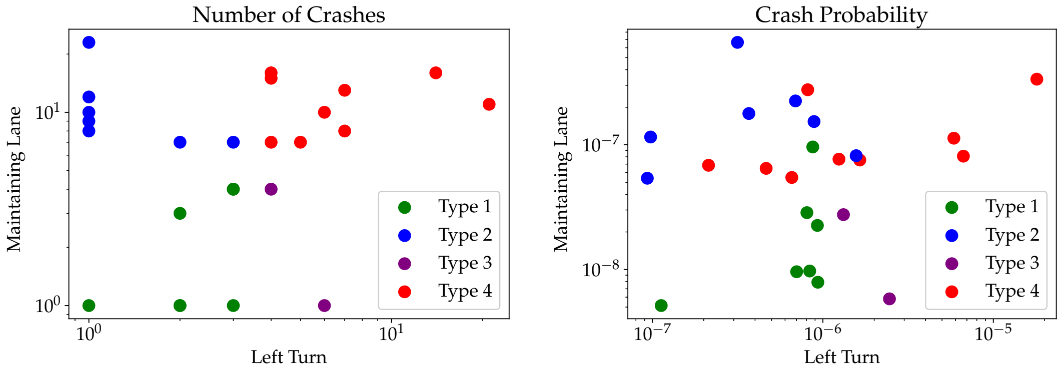

Fig. 5: Number of crashes and estimated crash probability for intersections along Pacific Coast Highway.

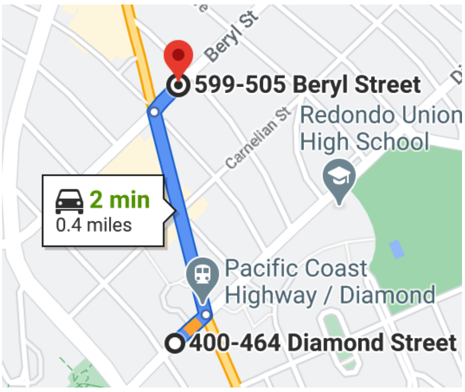

Fig. 6: Route risk example: Diamond Street to Beryl Street along Pacific Coast Highway.

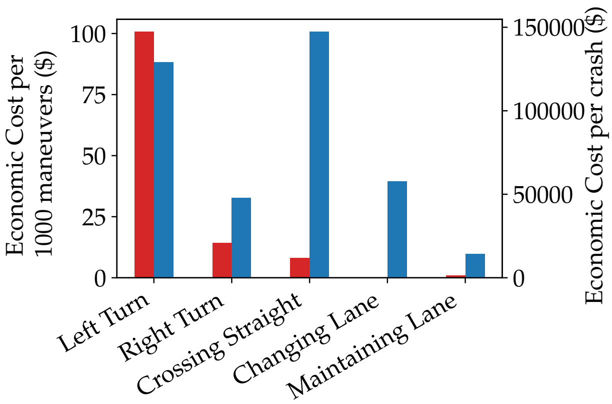

Fig. 7: Economic cost of crashes per 1000 maneuvers (red) and per crash (blue). Note that the crossing straight maneuver hs the highest economic cost per crash, but left turns have the highest expected economic cost per maneuver as their crash probability is significantly higher.

References

[1] Centers for Disease Control and Prevention, "Web-based Injury Statistics Query and Reporting System (WISQARS)[Online].(2020)," *National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (producer). Available from: URL: www. cdc. gov/ncipc/wisqars*, 2020.

[2] National Center for Statistics and Analysis, "2018 fatal motor vehicle crashes: Overview. (Traffic Safety Facts Research Note. Report No. DOT HS 812 826)," United States National Highway Traffic Safety Administration, Tech. Rep., 2019.

[3] L. Blincoe, T. R. Miller, E. Zaloshnja, and B. A. Lawrence, "The economic and social impact of motor vehicle crashes, 2010 (Revised) (Report No. DOT HS 812 013)," United States National Highway Traffic Safety Administration, Tech. Rep., 2015.

[4] National Highway Traffic Safety Administration, "National Motor Vehicle Crash Causation Survey: Report to Congress," 2008.

[5] C. F. Kerry and J. Karsten, "Gauging investments in self-driving cars," https://www.brookings.edu/research/gauging-investment-in-self-driving-cars, 2017.

[6] R. Lanctot *et al.*, "Accelerating the future: The economic impact of the emerging passenger economy," *Strategy Analytics*, vol. 5, 2017.

[7] P. Koopman, "A Reality Check on the 94 Percent Human Error Statistic for Self-Driving Cars," *Safe Autonomy Blog*, 2018, http://safeautonomy.blogspot.com/2018/06/a-reality-check-on-94-percent-human.html.

[8] A. S. Mueller, J. B. Cicchino, and D. S. Zuby, "What humanlike errors do autonomous vehicles need to avoid to maximize safety?" *Journal of Safety Research*, vol. 75, pp. 310–318, 2020.

[9] L. Lubbe, H. Jeppsson, A. Ranjbar, J. Fredriksson, J. Bärgman, and M. Östling, "Predicted road traffic fatalities in Germany: The potential and limitations of vehicle safety technologies from passive safety to highly automated driving," in *Proceedings of IRCOBI conference. Athena, Greece*, 2018.

[10] A. Shetty, M. Yu, A. Kurzhanskiy, O. Grembek, H. Tavafoghi, and P. Varaiya, "Safety challenges for autonomous vehicles in the absence of connectivity," *arXiv preprint arXiv:2006.03987*, 2020.

[11] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.

[12] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

[13] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," *arXiv preprint arXiv:1708.06374*, 2017.

[14] E. D. Swanson, F. Foderaro, M. Yanagisawa, W. G. Najm, P. Azeredo *et al.*, "Statistics of Light-Vehicle Pre-Crash Scenarios Based on 2011–2015 National Crash Data," Washington, DC: National Highway Traffic Safety, Tech. Rep., 2019.

[15] N. Board, "Collision between vehicle controlled by developmental automated driving system and pedestrian," *Nat. Transpot. Saf. Board, Washington, DC, USA, Tech. Rep. HAR-19-03*, 2019.

[16] M. Schwall, T. Daniel, T. Victor, F. Favaro, and H. Hohnhold, "Waymo public road safety performance data," *arXiv preprint arXiv:2011.00038*, 2020.
[17] Maricopa Association of Governments, "Mag fast facts - transportation," https://azmag.gov/About-Us/Divisions/Transportation-Division/MAG-Fast-Facts-Transportation, 2018.
[18] M. Davis *et al.*, "National telephone survey of reported and unreported motor vehicle crashes," in *Findings report no. DOT HS-812-183*. National Highway Traffic Safety Administration Washington, DC, 2015.
[19] Safe Transportation Research and Education Center, "Transportation Injury Mapping System (TIMS)," 2020. [Online]. Available: https://tims.berkeley.edu/
[20] R. G. Newcombe, "Two-sided confidence intervals for the single proportion: comparison of seven methods," *Statistics in medicine*, vol. 17, no. 8, pp. 857–872, 1998.
[21] C. Holland, J. Levis, R. Nuggehalli, B. Santilli, and J. Winters, "UPS optimizes delivery routes," *Interfaces*, vol. 47, no. 1, pp. 8–23, 2017.