Scenario risk quantification for automated driving

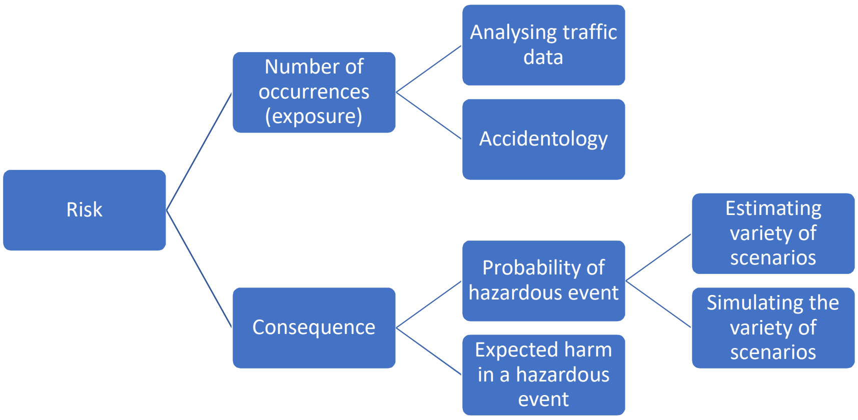
Contents

1. Introduction

- Safety must be validated for all possible traffic scenarios.
- Risk assessment is an essential component of the safety validation.
- TODO: Provide more references and argumentation for relevance.

2. Method

- We want to calculate the risk that a vehicle is exposed to in the scenarios of a specific scenario category $\mathcal{C}$.
- Therefore, everything is under the condition of $\mathcal{C}$, but for the sake of brevity, this is omitted. E.g., $f(\cdot\,|\mathcal{C}) = f(\cdot)$.
- Risk = number of occurrences $\times$ probability of a hazardous event $\times$ expected harm in a hazardous event
- Notation:
  - $P(\cdot)$: Probability of $\cdot$
  - $f(\cdot)$: Probability density of $\cdot$
  - $F(\cdot)$: Cumulative distribution of $\cdot$
  - $\mathrm{E}[\cdot]$: Expectation of $\cdot$



5

2a. Number of occurrences

- Let $P(k)$ denote the probability of k occurrences of scenarios of scenario category $\mathcal{C}$.
- The expected number of occurrences equals:

$$\mathrm{E}[k] = \sum_{k=1}^{\infty} kP(k)$$

- To determine $P(k)$ on real-world traffic data, see <our ITSC2020 paper on scenario mining>
- TODO: summarize steps from ITSC2020 paper
- Remark: In the special case that $P(k)$ follows a Poisson distribution, i.e., $P(k) = \frac{\lambda^k}{k!}e^{-\lambda}$, then we have $\mathrm{E}[k] = \lambda$.

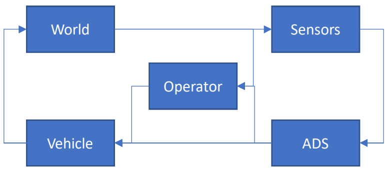
6

2b. Probability of bad outcome

- Let B denote the probability of a bad outcome. We want to estimate $P(B)$.
- To estimate $P(B)$, we need to consider the scenarios of the scenario category $\mathcal{C}$.
- Let the scenarios be parametrized by $\theta \in \Theta$.
- Let $y \in \mathbb{R}$ be the outcome of a test, such that we have a bad outcome if and only if $y \leq 0$. Note that y is a stochastic variable (e.g., stochastic driver model, stochastic sensor noise).
- By definition, we have $P(B|\theta) = \int_{-\infty}^{0} f(y|\theta)\,\mathrm{d}y = F(0|\theta)$.
- $P(B) = \int_{\Theta} F(0|\theta)\,f(\theta)\,\mathrm{d}\theta$, with $f(\theta)$ being the probability density of the scenario parameters θ given the scenario category $\mathcal{C}$.

2b. Probability of bad outcome (cont'd)

- To estimate $F(0|\theta)$, we perform simulations.
- The simulations are based on the scheme shown on the right.
- In the end, we are interested in the quantity $F(0|\theta)\,f(\theta)$.
- Do simulations until $\mathrm{Var}\left[\hat{F}(0|\theta)\right] \cdot \hat{f}(\theta)$ is below a threshold.



2b. Probability of bad outcome (cont'd)

- One idea is to use Kernel Density Estimation (KDE) to estimate $\mathrm{f}(y|\theta)$:
$$\hat{f}(y|\theta) = \sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right),$$
with $K(\cdot)$ denoting the kernel function, n the number of simulations, y_i the result of the i-th simulation and h the bandwidth (estimated using one-leave-out cross validation).
- In that case, Nadaraya (1964) states that
$$\mathrm{Var}\big[\hat{F}(y|\theta)\big] = \frac{F(y|\theta)\big(1 - F(y|\theta)\big)}{n},$$
which we estimate by plugging in $\hat{F}(y|\theta)$ for $F(y|\theta)$.
- $\hat{P}(B|\theta) = \hat{F}(0|\theta)$.

2b. Probability of bad outcome (cont'd)

- To speed up the calculations, we do not sample directly from $f(\theta)$.
- An *importance density* $\mathrm{g}(\theta)$ is chosen.
- $P(B) = \mathrm{E}_f[P(B|\theta)] = \mathrm{E}_g\left[\frac{P(B|\theta)f(\theta)}{g(\theta)}\right]$
- $\mathrm{Var}\,[P(B)] = \mathrm{E}\left[\left(\frac{P(B|\theta)f(\theta)}{g(\theta)} - P(B)\right)^2\right] = \mathrm{E}\left[\left(\frac{P(B|\theta)f(\theta) - P(B)g(\theta)}{g(\theta)}\right)^2\right].$
- Clearly, the optimal importance density (i.e., the one that minimizes the variance) is:
$$g^*(\theta) = \frac{P(B|\theta)f(\theta)}{P(B)}$$

2b. Probability of bad outcome (cont'd)

- The optimal importance density is:
$$g^*(\theta) = \frac{P(B|\theta)f(\theta)}{P(B)}$$
- We do not know $P(B)$ and $P(B|\theta)$. The idea is to use an approximate of $P(B|\theta)$ based on some initial simulations. Suppose that we have $\hat{P}(B|\theta_i)$ with $i \in \{1, \dots, n\}$ and where the parameters θ_i are carefully selected.
- We approximate $P(B|\theta)$ with:
$$\hat{p}(\theta) = \hat{P}\left(B\middle|\theta_j\right) \text{ with } j = \arg\min_{\mathrm{i}\in\{1,\dots,n\}} \|\theta - \theta_i\|,$$
i.e., we use the closest point for which we performed simulations.
- Then $\hat{g}(\theta) \propto \hat{p}(\theta)\hat{f}(\theta)$.
- We can use Markov Chain Monte Carlo to sample from $\hat{g}(\theta)$.

2b. Probability of bad outcome (cont'd)

- There is an important role for the model of the driver.
- Driver models are generally split in two parts:
  - Longitudinal driving behaviour:
    - Very straightforward deterministic models, such as Optimal Velocity Model (Bando et al., 1995), Gipps model (Gipps, 1981), Intelligent Driver Model (IDM) (Treiber et al., 2000), Enhanced IDM (Kesting et al., 2010), IDM+ (Schakel et al., 2010)
    - HDM: any of the aforementioned models, but including "measurement noise" (Treiber et al., 2006)
    - Multi-level framework that includes driver distractions (Van Lint and Calvert, 2018)
  - Lateral driving behaviour:
    - Integrated lane change model seems to be a good option (Schakel et al., 2012)
    - A good overview is given in the survey article of Rahman et al. (2013)

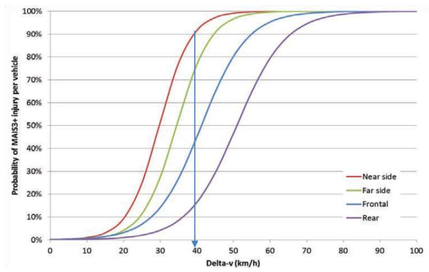
2c. Expected harm with a bad outcome

- Let us assume that the simulations in the previous step provide a vector $\mathrm{z} \in \mathbb{R}^{n_z}$ container metrics that can be used to determine the harm.
- Based on the simulations, we could estimate $f(z|B)$.
- To estimate the harm h, we need a mapping from z to h. This mapping is denoted by $g(\cdot)$.
- Then, the expected harm is $\mathrm{E}[h|B] = \mathrm{E}_{f(z|B)}[g(z)] = \int_{\mathbb{R}^{n_z}} g(z) f(z|B)\,\mathrm{d}z$.

13

2c. Expected harm with a bad outcome (cont'd)

- The mapping g could be based on literature.
- For example, Jurewicz et al. (2016) provide probability for MAIS3+ injuries based on impact speed (see figure on the right).
- Bahouth et al. (2014) provide tables with *High-Severity Injury Risk Thresholds.*



14

2d. Computing the risk

- To compute the risk, the results of the previous steps are combined.
- Thus, the risk is:

$$\mathrm{E}[k] \cdot P(B) \cdot \mathrm{E}[h|B]$$

3. Relation with ISO 26262

- An important component of ISO 26262 is the Hazard Analysis and Risk Assessment (HARA).
- Within the HARA, the risk of an *hazardous event* is determined using an *Automotive Safety Integrity Level (ASIL)*. The ASIL is based on three aspects:
  - exposure,
  - controllability, and
  - severity.
- ASIL D corresponds to the highest risk.
- The ASIL risk rating is *qualitative* and based on "expert opinions" (thus *subjective*)

Table 4 — ASIL determination

Severity class	Exposure class	Controllability class		
		C1	C2	C3
S1	E1	QM	QM	QM
	E2	QM	QM	QM
	E3	QM	QM	A
	E4	QM	A	B
S2	E1	QM	QM	QM
	E2	QM	QM	A
	E3	QM	A	B
	E4	A	B	C

3. Relation with ISO 26262 (cont'd)

- "The two distinct short-comings of the current ISO 26262-2011 standard are guided by the subjective nature of the experts' mental models leading to unreliable ratings and the ability to identify a hazard (including the black swan events)" (Khastgir et al., 2017)
- "the classification of the ASIL for a vehicle function depends very strong on the included engineers which causes to very different results defining these parameters." (Teuchter, 2012)

3. Relation with ISO 26262 (cont'd)

- Our proposed method can be used to determine an ASIL-like indicator in a *quantitative* and *objective* manner.
- The method can be used to determine the contributions to the three aspects separately:
  - Exposure: follows directly from our exposure.
  - Controllability: "the likelihood that, when subject to a hazardous situation, the vehicle driver avoids an accident". This can be determine by computing $P(B)$ while including a hazardous situation (i.e., a component failure) in the simulations.
  - Severity: follows directly from our expected harm.

4. Case study

- In this case study, three different scenario categories are considered:
  - Lead vehicle braking
  - Cut in
  - Slower vehicle in front while other vehicle overtaking
- TODO: Obtain and describe the results
  - Simulation of lead vehicle braking is finished.
  - Also obtaining $P(B|\theta)$ (and its variance) is finished.
  - The importance sampling (in order to obtain $P(B)$ is still work to do.
  - For the cut-in scenario, the estimation of the parameter distribution is done.
  - For the third scenario (slower vehicle in front), still the scenarios need to be obtained from the data set.

5. Discussion

- TODO: Describe discussion items

6. Conclusions

- TODO: write conclusions

A. Driver model

- For modeling the longitudinal driver behavior, the Human Driver (meta-)Model is used from Treiber et al. (2006) in combination with the Intelligent Driver Model plus (IDM+) from Schakel et al. (2010).
- With HDM, it is assumed that the acceleration is a function of the
  - gap between the ego vehicle and the leading vehicle (s),
  - the speed difference between the ego vehicle and the leading vehicle (Δv), and
  - the ego vehicle speed (v), i.e.

$$\dot{v}(t) = m\big(s(t), \Delta v(t), v(t)\big)$$

- First, HDM introduces a finite reaction time T'.
- Next, imperfect measurements are assumed for $s(t)$ and $\Delta v(t)$

A. Driver model (cont'd)

- The imperfect measurements are modelled using a Wiener process:
  - $s^{\text{est}}(t) = s(t)\exp\{V_s w_s(t)\}$
  - $(\Delta v)^{\text{est}}(t) = \Delta v(t) + s(t) r_c w_{\Delta v}(t)$
  - $w_s(t)$ and $w_{\Delta v}(t)$ come from a Wiener process with $w(t+\Delta t) = \exp\left\{\frac{-\Delta t}{t}\right\} w(t) + \sqrt{\frac{2\Delta t}{t}}\eta(t)$, where Δt is the sample time and $\eta(t)$ is sampled from a i.i.d. normal distribution with mean zeros and standard deviation 1.
  - V_s and r_c are parameters that determine the estimation errors.
- It is assumed that the driver anticipates its measurements of the gap and its speed while it cannot anticipate on the speed difference:
  - $s'(t) = [s^{\text{est}} - T'(\Delta v)^{\text{est}}]_{t-T'}$
  - $v'(t) = [v^{\text{est}} + T'\dot{v}]_{t-T'}$
  - $\Delta v'(t) = (\Delta v)^{\text{est}}(t - T')$

A. Driver model (cont'd)

- With IDM+, the following model is used:

$$m\big(s(t), \Delta v(t), v(t)\big) = a \min\left(1 - \left(\frac{v(t)}{v_0}\right)^{\delta}, 1 - \left(\frac{s^*\big(v(t), \Delta v(t)\big)}{s(t)}\right)^2\right),$$

with $s^*(v, \Delta v) = s_0 + Tv + \frac{v\Delta v}{2\sqrt{ab}}$

where a, b, s_0, δ, T, v_0 are parameters.