Traffic safety research: a challenge for extreme value statistics

Jenny Jonasson, Astra-Zeneca, Gothenburg
Holger Rootzén, Mathematical Sciences, Chalmers
http://www.math.chalmers.se/~rootzen/
Traffic accidents

• 1.3 million deaths/year worldwide, 20-50 million severely injured

• Large economic losses

• Less than 1 death/day in Sweden now. Down from 3 deaths/day a few decades ago – at a time with much less traffic

• First simple measures: seatbelts, helmets, follow traffic rules, drunk driving laws, ..., then more sophisticated ones: rebuild roads, better tires, improve driver education, airbags, ..., then next level of sophistication: more driver training and retraining, ABS, ESP, ..., and ??
New and exciting area for statistics

Selection bias/errors
Risk estimation

Active safety systems for next generation cars. Important for competition with other car makers and for safety (?)

Driver training, traffic laws, ...
Naturalistic Driving Research

- *In Situ* investigation of driver performance
  - Use an instrumented vehicle
  - No experimenter or instructions
  - Data continuously collected for extended period
100-car study

- 100 cars, appr 250 drivers, appr 1 year
- Five video cameras, radar sensors; front, rear (for all 100 cars) and each side (for 20 cars), vision-based lane tracker, glare detectors, GPS, accelerometer
- Still not enough crashes (82) -> try to use near-crashes (761) to learn about crash behavior
Crash

Any contact with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated, and includes other vehicles, roadside barriers, objects on or off of the roadway, pedestrians, cyclists, animals.

Near-crash

Any circumstance requiring a rapid, evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal to avoid a crash. A rapid, evasive maneuver is defined as a steering, braking, accelerating, or any combination of control inputs that approaches the limits of the vehicle capabilities. As a guide: Subject vehicle braking $>0.5$ g or steering input that results in a lateral acceleration $>0.4$ g to avoid a crash constitutes a rapid maneuver.

Selection: “trigger” as above – and then manual selection and annotation
How can information from near-crashes be used to prevent real crashes?

1. Do near-crashes resemble real crashes? Are more extreme near-crashes more like real crashes?

2. Is it possible to find driver behavior or traffic situations which is different in near-crashes than in normal driving? Are these differences even more extreme in real crashes?
Statistical methods used so far:

**Odds ratios and logistic regression:** Completely dominant – but can’t easily extrapolate from less severe events to more severe ones, can’t easily judge extent of selection bias.

**Regression:** Is relative risk the same for crashes and for near-crashes?

**Extreme Value Statistics (almost new):** Can near-crashes predict the frequency of real crashes? Do covariates behave in same way for crashes and near-crashes? Requires a continuous crash proximity or crash severity measure.

*Underlying philosophy: a traffic accident is a rare and extreme event.*
Crash proximity measure

• Measure of how close the near-crash is to a real crash
• Examples: TTEC = Time To edge Crossing, Gap = time between first car leaves conflict area and second car enters conflict area, Time-to-collision (TTC), ... 
• Here, TTC, the time it takes for the cars to collide when continuing with the same speeds – useful for rear-ending

\[
TTC = \frac{s}{v_1 - v_2}
\]
Examples of TTC computed from radar signals

Possible to extract min TTC

Not possible to extract min TTC
100-car data, risk of rear-ending, TTC

384 near-crashes, 29 with good enough radar signals, 14 crashes.

Crash ⇔ TTC < 0

Block maxima 95% confidence interval for expected number of crashes is (0.07, 0.09) (Fitted GEV conditional on –TTC > 0, profile likelihood intervals)

Observed number of crashes = 14

Doesn’t match!
Selection bias!

All but two of the real rear-ending crashes were in start-stop traffic while all the near-crashes with usable TTC were in higher speed situations.

So maybe still: \( \rightarrow \) yes to question 1 (?)
Continuous variables that could influence crash risk:

- Speed
- Absolute value of yaw angle
- Distance to right and left lane markings
- Time the driver looks off-road during last 2 s or 3 s, total length of glances off-road longer than 1.5 s during last 15 s...
- Variance of lateral acceleration
- Variance of longitudinal acceleration
- Length of overlapping glance off road

Do any of these become more and more extreme as TTC gets smaller and smaller?
Fitte logistic bivariate extreme value distribution to min/max of each of these variables and TTC for near-crashes, \( \alpha \in (0, 1] \) dependence parameter, 1 is independence and “0 is complete dependence”

- \( \max(\text{eye off road in 3 s window}) \) \( \alpha = 1.00 \)
- \( \max(\text{speed}) \) \( \alpha = 1.00 \)
- \( \max(\text{variance longitudinal acc}) \) \( \alpha = 1.00 \)
- \( \min(\text{dist left markings}) \) \( \alpha = 1.00 \)
- \( \max(\text{dist right markings}) \) \( \alpha = 0.93 \)

Fitting was not possible for the other variables, however no indication of dependence, except for the last one.
max speed $\rightarrow \alpha = 1$

(12 with overlapping glance, 13 without overlapping glance, 4 without video)
Right now:

- Left and right censoring of baselines
- Models for “nonsymmetric dependence”
- “Mix Block Maxima and PoT”
Stable mixture models for Block Maxima

\[ P(G \leq x) = \exp \{-e^{-x/\sigma}\} \quad \text{Gumbel distribution} \]
\[ E(e^{-tS}) = e^{-t^\alpha}; \quad 0 < \alpha < 1 \quad \text{positive stable distribution} \]
\[ X = G + \sigma \log S, \]
\[ P(X \leq x) = E(P(X \leq x \mid S)) = E(\exp\{-e^{-x/\sigma+\log S}\}) \]
\[ = \exp\{-e^{-x/\sigma/\alpha}\} \quad \text{Gumbel distribution!} \]

\[ X_1 = G_1 + \sigma_1 \log(S) \quad \text{bivariate logistic distribution with} \]
\[ X_2 = G_2 + \sigma_2 \log(S) \quad \text{Gumbel marginals – symmetric if same marginals} \]
Stable mixture models for Block Maxima

\[ X_1 = G_1 + \sigma_1 \log(S + a_{1,1}S_{1,1}) \quad (2, 1) \text{ stable mixture distribution with } \]
\[ X_2 = G_2 + \sigma_2 \log(S) \quad \text{Gumbel marginals (not symmetric)} \]

\[ X_1 = G_1 + \sigma_1 \log(S) \quad (1, 2) \text{ stable mixture distribution with } \]
\[ X_2 = G_2 + \sigma_2 \log(S + a_{2,1}S_{2,1}) \quad \text{Gumbel marginals} \]

\[ X_1 = G_1 + \sigma_1 \log(S + a_{1,1}S_{1,1}) \quad (2, 2) \text{ stable mixture distribution with } \]
\[ X_2 = G_2 + \sigma_2 \log(S + a_{2,1}S_{2,1}) \quad \text{Gumbel marginals} \]

And so on ...
UMTRI (Gordon et al (2010)) “do near-crashes give similar risk estimates as crashes?”

Seemingly Unrelated Regression $\rightarrow$ yes to question 1 (?)

**EVS:** TTEC $\rightarrow$ road departure $\rightarrow$ road way departure crash

2.3 mile segment of US-23 with 117 traversals by 43 different drivers in instrumented cars.

EV distribution fit to minimum TTEC values for the 117 traversals $\rightarrow$ predicts 12 road departures/year

On the average there were 1.8 road way departure crashes/year

$\rightarrow$ yes to question 1 (?)
four year actual counts of daytime right angle collisions

EVS estimate of crash frequency from gap measurements

-- 8-hour observations of crossing gaps. Signalized intersections in the Lafayette area. Summer 2003

-- Error bars show 95% Poisson confidence intervals based on observed counts

→ yes to question 1 (??)
Example Victor & Dozza (2012)

Eyes Off Road Over Time (100ms bins)

 précipitating Event
  e.g. braking lead vehicle

Victor & Dozza (2012)
(Crashes & Near Crashes)

Eyes Off the Road [%]

Events
Baselines

Precipitating Event
e.g. braking lead vehicle

Time [s]
### Inattention & Risk example

*Victor and Dozza, 2012*

#### The overlapping glance gives the highest OR-s

<table>
<thead>
<tr>
<th>Time [s]</th>
<th>Events</th>
<th>Baselines</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5s</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.5-1.0 s</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>1.0-1.5 s</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>1.5-2.0 s</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>More than 2.0 s</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

#### Odds Ratio

<table>
<thead>
<tr>
<th>Glance Duration</th>
<th>Total Glance Time case control (Klauer 2006)</th>
<th>Total Glance Time case crossover (Liang et al 2011)</th>
<th>Total Glance Time case control (present paper)</th>
<th>Last Single Glance (present paper)</th>
<th>Glance History (previous glances) (present paper)</th>
<th>LG5s glances at t=5s (present paper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 0.5s</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5-1.0 s</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>1.0-1.5 s</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1.5-2.0 s</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>More than 2.0 s</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

*SHRP2 Inattention & Risk*
Example

Victor & Dozza (2011)

Events (Crashes and Near Crashes)
Baselines
Talking on the phone

Eyes Off the Road [%]

Time [s]

Off Road

0.3
0.2
0.1

0 1 2 3 4 5 6

0.17
0.12

***

**SAFER**
One conclusion

different kinds of nearcrashes and crashes; naturalistics studies; vehicles; drivers, all lead to different kinds of

• Selection bias
• Crash proximity measures
• Driver behavior – and ”covariates”

All require separate careful analysis

No omnibus answer to ”is there selection bias in choice of near-crashes”
The future

• Use near-crashes to investigate how (and if) attention measures and other driving and traffic characteristics influence crash risk → high-dimensional variable selection → new research questions

• Develop statistical predictors of crash risk → optimal choice of predictors → new research questions

• Investigate the relation of risk estimates obtained in different in naturalistic driving studies (Semifot, 100-car, SHRP 2, ...)

• Study the normal driving – near-crash/crash relation in naturalistic driving experiments

More and better data crucial
SHRP 2

• 2000 cars
• 3 years
• Much better instrumentation (?)
• Started a year ago


Crash Acc.mpg