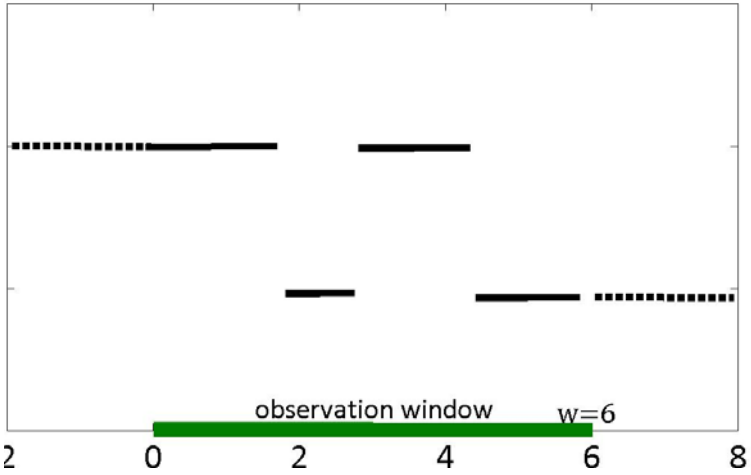
Extreme value methods, tail estimation for window-censored 0-1 processes, and car accidents



Holger Rootzén
Dmitrii Zholud
http://www.math.chalmers.se/~rootzen/

Traffic accidents

- 1.3 miljon 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, … ??

Naturalistic driving studies: cars with drivers like you and me are instrumented with video cameras, radar, GPS, sensors for steering wheel movement, gas- and brake pedal movement, …

Crash Acc.mpg

Generates extremely large data sets – and give completely new opportunities for preventing traffic accidents

Accidents are extreme events – same methods as those for financial risks and for natural catastrophes can be used

New 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, …

Statistical methods used so far:

Odds ratios and logistic regression: Completly 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?

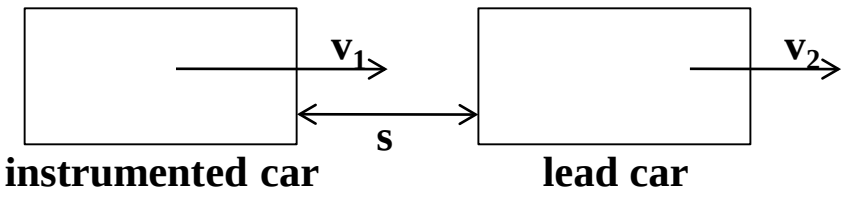
Extreme Value Statistics (almost new): Can near-crashes predict the frequency of real crashes? Do covariates behave in same way for crashes and near-crashes? Requires a continuous crash proximity or crash severity measure.

Underlying philosophy: a traffic accident is a rare and extreme event.

Use near-crashes – require crash proximity measure (EVS uses cont. distributions)

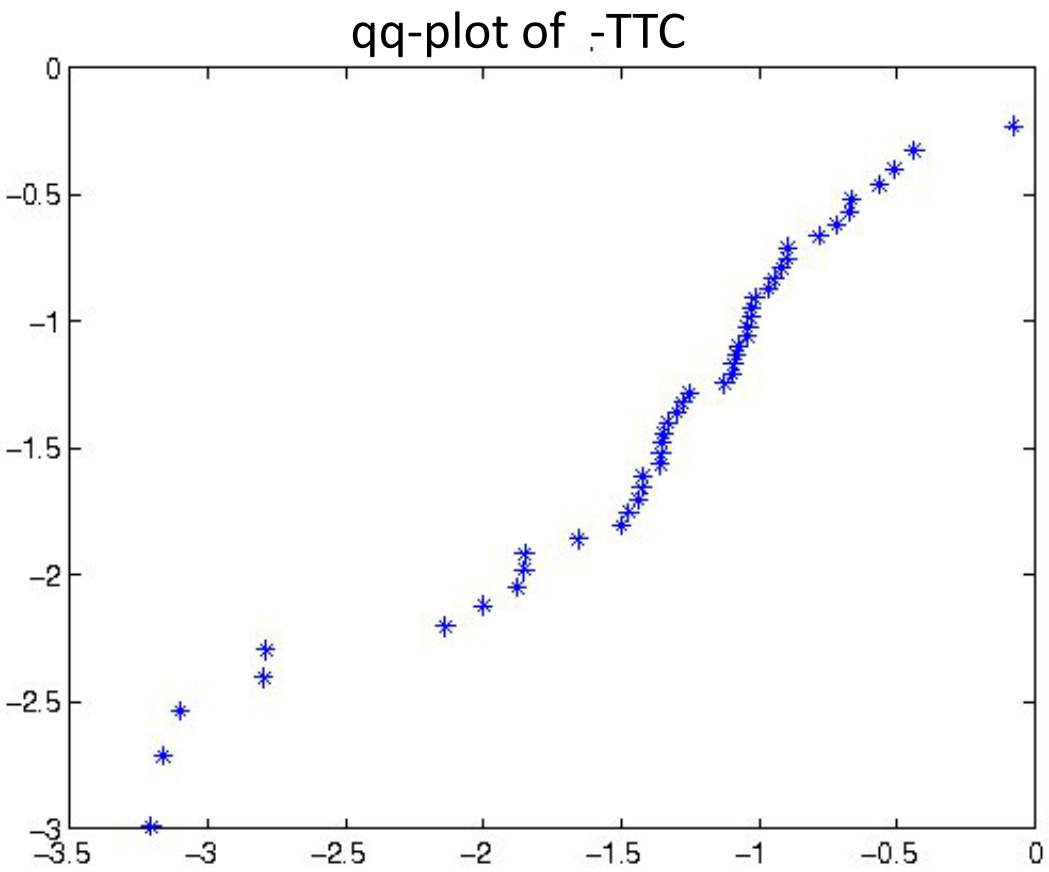
- Measure of how close the near-crash is to a real crash
- Examples: TTEC = Time To edge Crossing, Gap = time between first car leaves conflict area and second car enters conflict area, Time-to-collision (TTC), …
- Here, TTC, the time it takes for the cars to collide when continuing with the same speeds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

$$TTC = \frac{s}{v_1 - v_2}$$



100-car data, risk of rear-ending, TTC

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



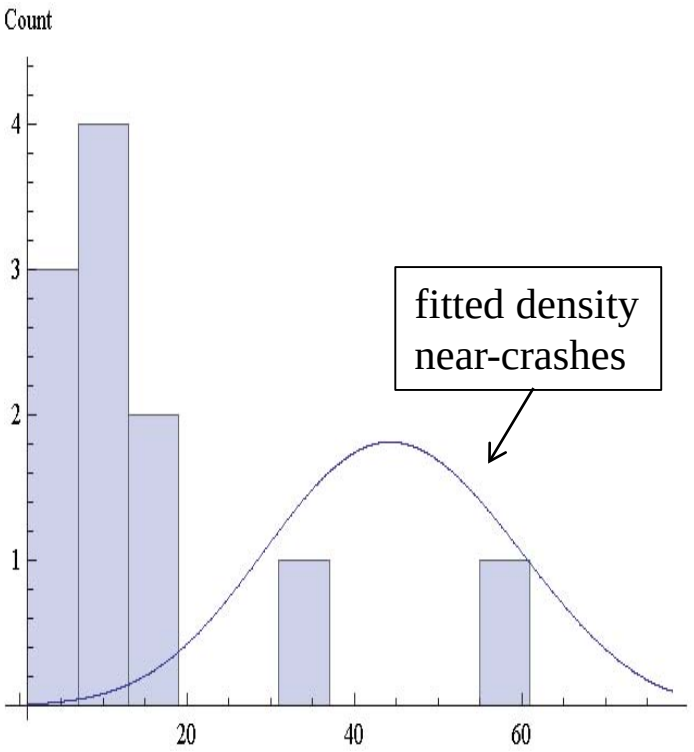
Crash ⇔ TTC < 0

Block maxima 95% confidence interval for expected number of crashes is (0.07, 0,09) (Fitted GEV conditional on –TTC > 0, delta method conf. intervals)

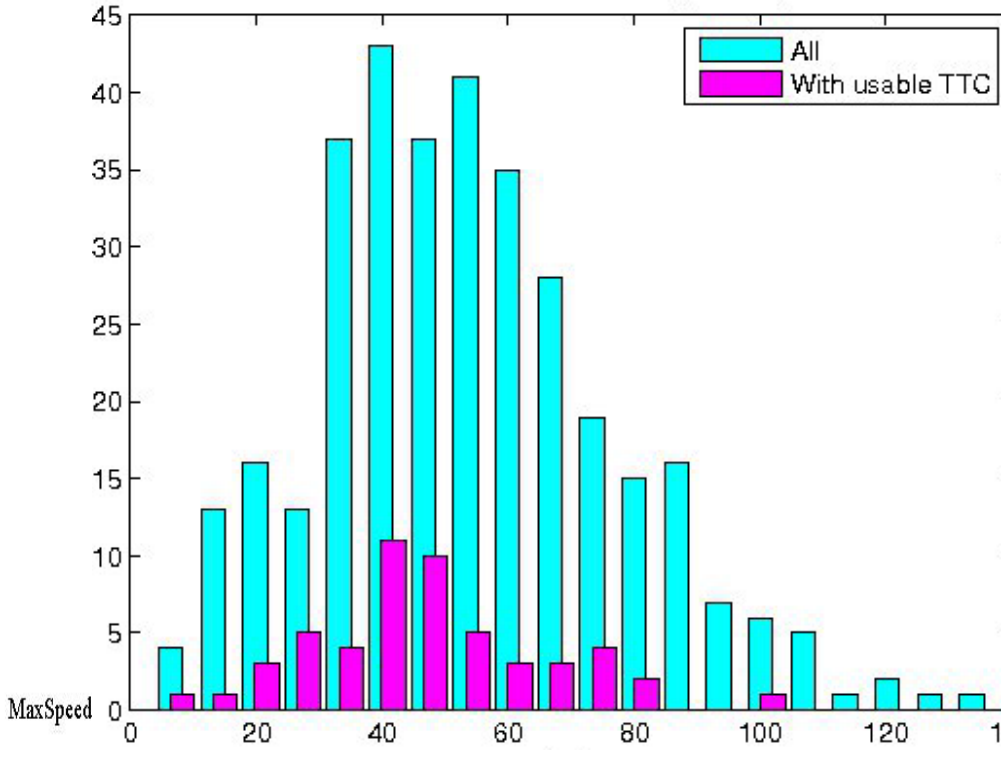
Observed number of crashes = 14

Doesn't match!

Selection bias!



crashes, max speed (km/h)



near-crashes, max speed (km/h)

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

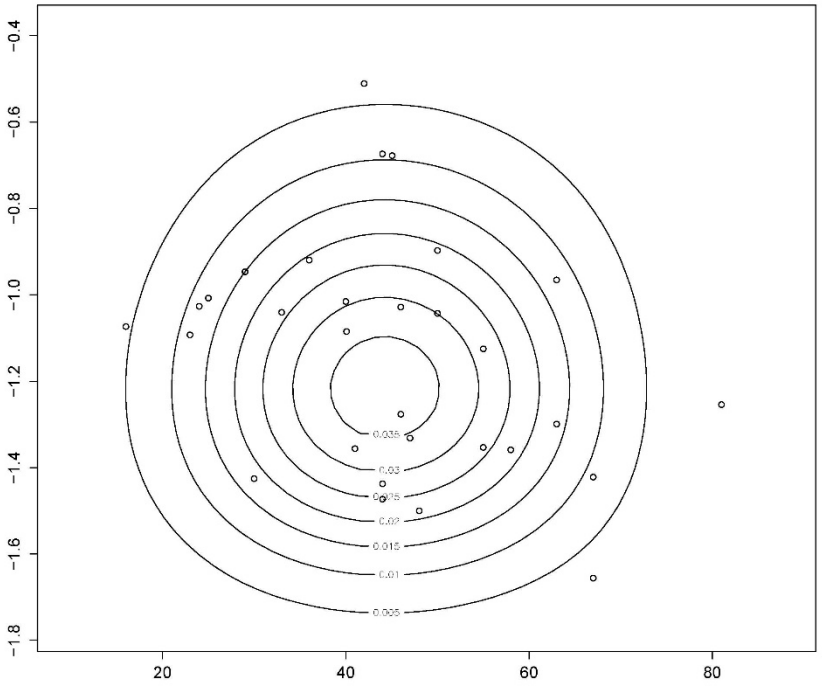
So maybe still: → *yes to question 1 (?)*

- Fitted 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

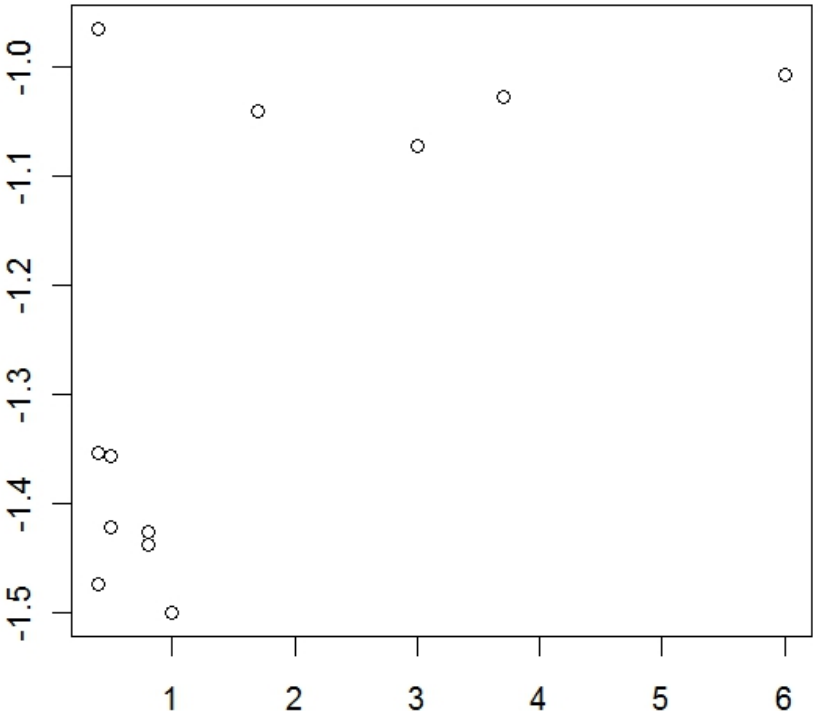
max(eye off road in 3 s window)	α=1.00
max (speed)	α=1.00
max (variance longitudinal acc)	α=1.00
min (dist left markings)	α=1.00
max (dist right markings)	α=0.93

Fitting was not possible for the other variables, however no indication of dependence, except for the last one (length overlapping glance off road)

max -TTC



max speed $\rightarrow \alpha = 1$



length overlapping glance

(12 with overlapping glance, 13 without overlapping glance, 4 vithout video)

Visual behavior/censoring

How much do you look off road while driving?

- ❑ 5% of the time
- ❑ 10% of the time
- ☒ 15% of the time
- ❑ 20% of the time

There is a 1 in 1000 chance that the lengths of an off road glances is longer than

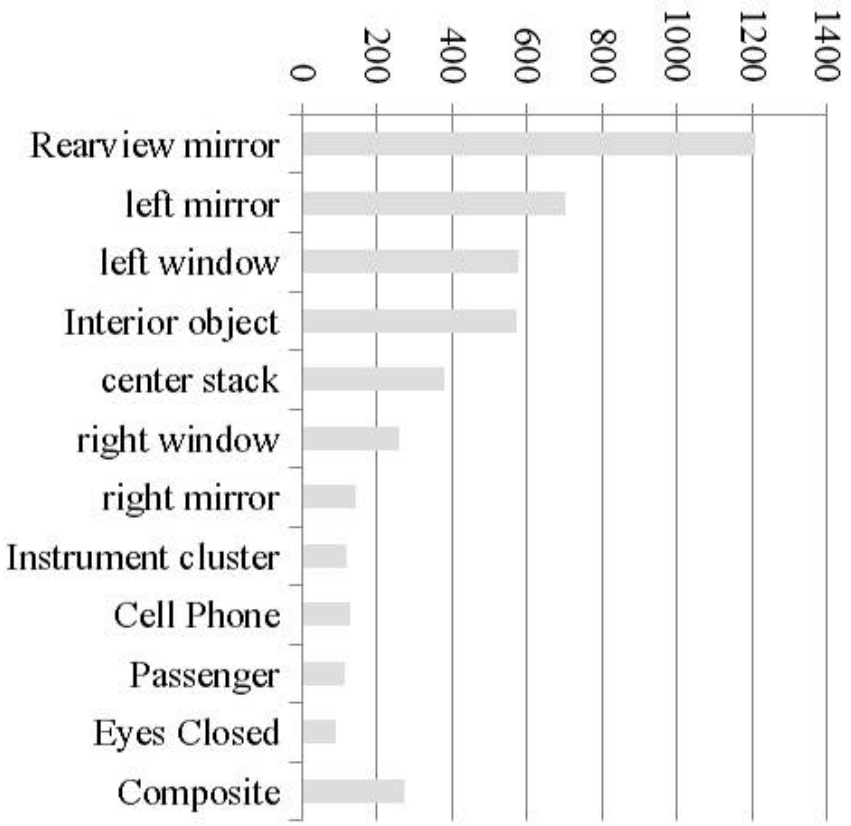
- ❑ 1 second
- ❑ 2 seconds
- ❑ 3 seconds
- ❑ 4 seconds
- ☒ 5 seconds
- ❑ 10 seconds

Is glance behavior different in different circumstances?

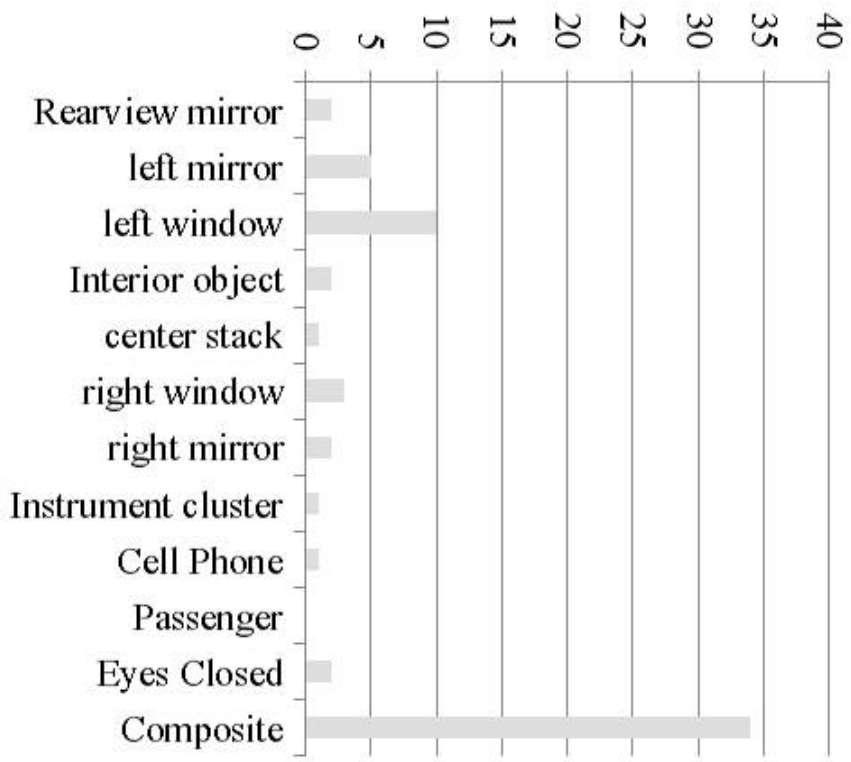
Not well understood

Glance behavior in the 100-car study

Raw data: 19,616 annotated 6-second intervals from 100-car study: 4582 with 1 or more off road glances



Glances shorter than 3 seconds



Glances longer than 3 seconds

M. Barnes*, A. Blankespoor, D. Blower, T. Gordon, P. Green, L. Kostyniuk, D. LeBlanc, S. Bogard., B. R. Cannon, and S.B. McLaughlin (2010). Development of Analysis Methods Using Recent Data: A Multivariate Analysis of Crash and Naturalistic Event Data in Relation to Highway Factors Using the GIS Framework. *Final Report* ***SHRP S01,*** *University of Michigan Transportation Research Institute*

Dozza, M. and Trent, V. (2012). Inattention – risk function (lead vehicle crashes). *A SHRP2 S08 analysis report.*

P. Tarko, and P. Songchitruksa (2006). Estimating frequency of chrashes as extreme traffic events. *Report, Purdue University*

J. Jonasson and H. Rootzén (2012). Internal validation of near-crashes in naturalistic driving studies: a continuous and multivariate approach. *Accident Analysis and Prevention* ***62,*** 102-109

K.-F. Wu and P.P. Jovanis. Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. *Accident Analysis and Prevention*, **45**, 507–516, 2012.

Alvarez, A. Smoothed nonparametric estimation in window censored semi-Markov processes. *J. Statist. Planning inf.* **131**, 209-229 (2005)

Einmahl, J.H.J., Fils-Villetard, A, & Guillou, A. Statistics of extremes under random

Censoring. *Bernoulli* **14**, 207–227 (2008)

Gill, R.D. & Keiding, N. Product-limit estimators of the gap time distribution of a renewal process under different sampling patterns. Lifetime Data Anal **16,** 571–579 (2010)

Gomes, Y. and Neves, M. Estimation of the Extreme Value Index for Randomly Censored Data. 2010

H. Rootzén and D. Zholud (2015): Tail estimation for window censored processes. *Technometrics,* online

Vardi, Y. Emirical distributions in selection bias models. *Ann. Statist.* **13**, 178-203 (1985)

Zhao, Y & Nagaraja, H.N. Fisher information in window censored renewal process

data and its applications. *Ann Inst Stat Math* **63**, 791–825 (2011)

Zhu, Y, Yashchin, E. & Hosking, J. R.M. Parametric Estimation for Window Censored

Recurrence Data, *Technometrics online*, DOI:10.1080/00401706.2013.804442 (2013)

Zuo, J., Meeker, J.W.Q. & Wu, W.Q. H. Analysis of Window-Observation Recurrence Data. *Technometrics* **50**, 128-143 (2008)