COMPARISON OF PATTERN RECOGNITION-BASED HOTSPOT IDENTIFICATION METHODS BASED ON SIMULATION(20-02761)

Jiajian Lu, Aditya Medury, Offer Grembek
UC Berkeley Safe Transportation Research & Education Center

INTRODUCTION

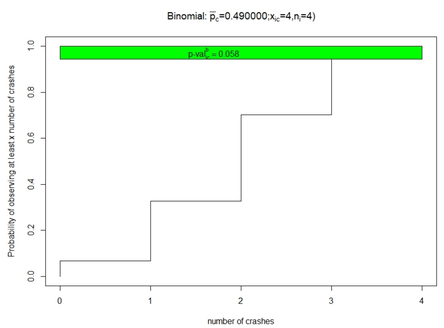
- Hotspot methods rely on SPFs
- Volume and infrastructure can be incomplete or unavailable
- Alternatively is to ascertain if there is a recurring crash pattern.
- We compares Binomial and Beta-Binomial methods for prioritizing hotspots in the absence of volume information and SPFs

METHOD: BINOMIAL TEST

Mean proportion of some crash type c: $\bar{p}_c = \frac{\sum_i x_{ic}}{\sum_i n_i}$

Probability of observing at least x_{ic} crashes of type c of n_i crashes:

$$\text{p-val}_{ic}^{b} = P(x \geq x_{ic}|n_i, \overline{p}) = 1 - P[x \leq (x_i - 1)]$$
$$= 1 - \sum_{x=0}^{x_{ic}-1} \frac{n_i}{x!\,(n_i - x)!} (\overline{p})^x (1 - \overline{p})^{n_i - x}$$



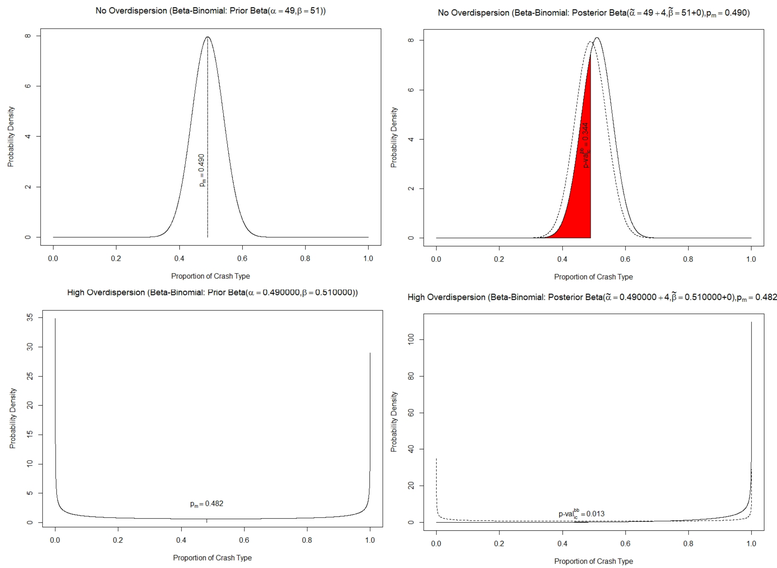
METHOD: BETA BINOMIAL TEST

Binomial distribution: $P(x_{ic}|p, n_i) = \binom{n_i}{x_{ic}} p^{x_{ic}} (1-p)^{n_i - x_{ic}}$

Beta prior: $p \sim Beta(\alpha, \beta).$

Estimate parameters and set threshold proportion: $\int_{p_m}^{1} Beta(p|\hat{\alpha}, \hat{\beta}) dp = 0.5$

Calculate p value: $\text{p-val}_{ic}^{bb} = 1 - P(p_i > p_m|x_{ic}, n_i) = Beta\left(p_m, \tilde{\alpha}, \tilde{\beta}\right)$

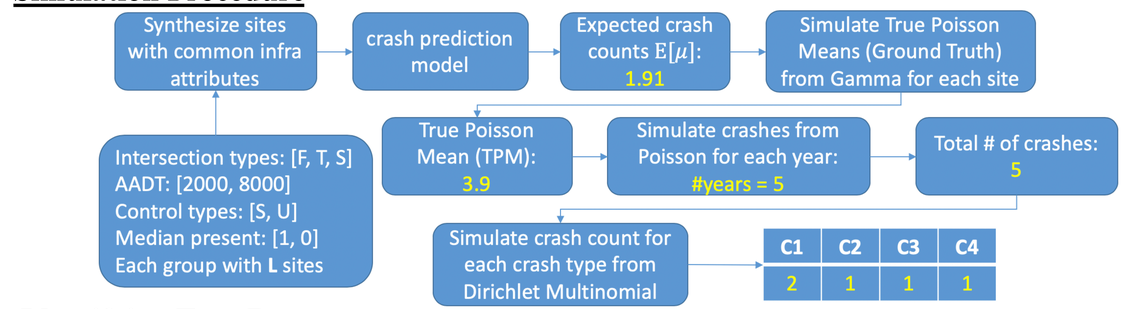


COMPARISON OF TWO METHODS

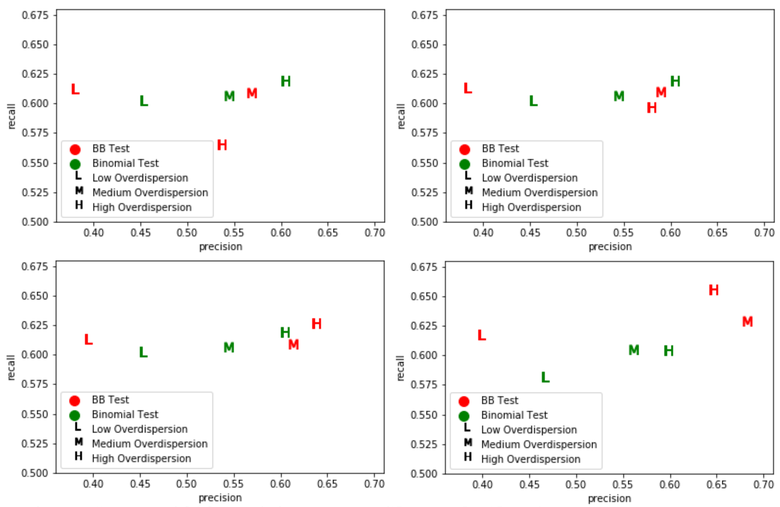
- Binomial distribution does not account for overdispersion in crash proportions
- Binomial test evaluates the exceedance of the observed crashes of a given type observed at a specific site relative to the mean crash proportion in the reference population
- Beta-Binomial test evaluates the median crash proportion in the reference population relative to the EB-adjusted crash proportion of a specific site.

SIMULATION STUDY

Simulation Procedure



Identifying True Patterns



$$\text{precision} = \frac{\text{True Positive}}{\text{Predicted Positive}}$$

$$\text{recall} = \frac{\text{True Positive}}{\text{Actual Positive}}$$

- 75-percentile proportion is better in Beta-Binomial test
- Beta-Binomial test is better in medium and high overdispersion cases
- Beta-Binomial test is better as the number of crashes increases

1) p_{medium} 2) p_{mean}
3) p_{75} 4) 5-time TPM, p_{75}

Ranking HCCLs (Pearson Correlation)

Dispersion	Low	Medium	High
Binomial Test	0.62	0.63	0.66
Beta-Binomial	0.58	0.61	0.62

Two methods perform similar

DISCUSSION

- Beta Binomial performs better in hotspot with motor vehicle crashes as more crashes are included
- Pattern recognition is needed even in the presence of SPFs as the detection of an underlying crash dynamic may aid in the identification of appropriate countermeasures

Berkeley Institute of Transportation Studies
This research was funded by California Department of Transportation
safetrec.berkeley.edu
Berkeley SafeTREC SAFE TRANSPORTATION RESEARCH AND EDUCATION CENTER