Variance Estimation for Crash Data in Before-After Studies

2006 Summer TexITE Meeting
College Station
June 24, 2006

Zhirui Ye
Texas A&M University

Background

- Before-After Studies
- Four-step Process (Hauer, 1997)
  - For a single entity:

Estimate λ and π

↓

Estimate $VAR\{\hat{\lambda}\}$ and $VAR\{\hat{\pi}\}$

↓

Estimate δ and θ

↓

Estimate $VAR\{\hat{\delta}\}$ and $VAR\{\hat{\theta}\}$

λ : what safety was

π : what safety would have been

δ : safety effect of a treatment $= \pi - \lambda$

θ : index of effectiveness $= \lambda / \pi$

Background (cont'd)

- Four-step Process
  - For a composite entity
    - Step 1: $\hat{\lambda} = \sum_{i=1}^{n} \lambda_i, \ \hat{\pi} = \sum_{i=1}^{n} \pi_i$
    - Step 2:

Poisson Assumption

For each entity, annual crash data fits a Poisson distribution over a long period.

$\Downarrow$

$VAR\{\hat{\lambda}\} \approx \hat{\lambda}, \ \mathrm{VAR}\{\hat{\pi}\} \approx \hat{\pi}$

Problem Identification

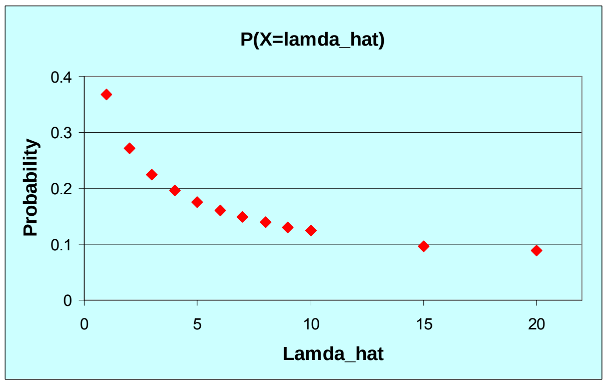
- Question

Poisson Assumption $\xrightarrow{?}$ $VAR\{\hat{\lambda}\} \approx \hat{\lambda}, \ VAR\{\hat{\pi}\} \approx \hat{\pi}$

- Two Possible Situations
  - Situation 1: $\lambda_i, \pi_i \sim Poisson(\,)$
    - Different characteristics (geometries, traffic flow, etc.)
  - Situation 2: $\hat{\lambda}_i = \lambda_i, \ \hat{\pi}_i = \pi_i$

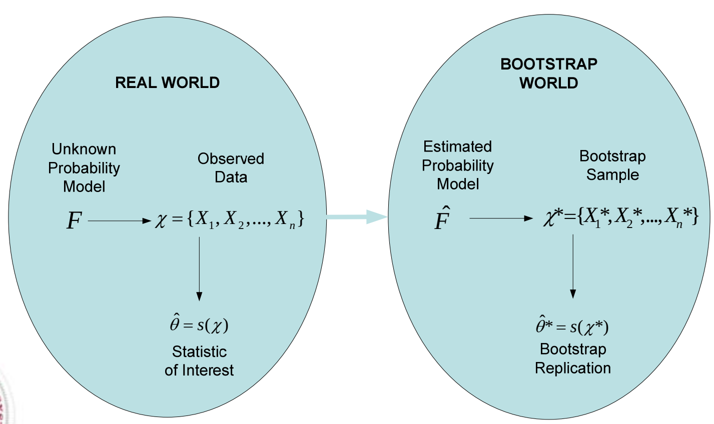
Problem Statement (cont'd)

- Possible Situations
  - Situation 2: $P(X = \lambda_i = \hat{\lambda}_i) = \dfrac{e^{-\hat{\lambda}_i} * \hat{\lambda}_i^{\hat{\lambda}_i}}{\hat{\lambda}_i!}$



Methodology

- Bootstrap Resampling Method



(Efron and Tishirani, 1993)

Methodology (cont'd)

- Algorithm of Variance Estimation

Collect crash data $\chi = \{X_1, X_2, ..., X_n\}$

Bootstrap resamples with replacement

$\chi_1^* = \{X_1^*, X_2^*, ..., X_n^*\} \Longrightarrow S_1^* = sum(\chi_1^*)$

$\chi_2^* = \{X_1^*, X_2^*, ..., X_n^*\} \Longrightarrow S_2^* = sum(\chi_2^*)$

• • • • • •

$\chi_B^* = \{X_1^*, X_2^*, ..., X_n^*\} \Longrightarrow S_B^* = sum(\chi_n^*)$

Variance estimation

$VAR\{S^*\} = VAR\{S_1^*, S_2^*, ..., S_B^*\}$

Simulation Design

Draw a sample from a known distribution (Poisson, Negative Binomial, etc.)
$\chi_i = \{X_1, X_2, ..., X_n\}$

Variance Estimation

From the known distribution: $VAR\{S_i\}$

Using bootstrap: $VAR\{S_i^*\}$

Repeat *m=100* times to obtain

$VAR\{S_1\}, VAR\{S_2\}, ..., VAR\{S_m\}$

$VAR\{S_1^*\}, VAR\{S_2^*\}, ..., VAR\{S_m^*\}$

Calculate ratios and variances

$r_1 = VAR\{S_1^*\}/VAR\{S_1\}, ..., r_m = VAR\{S_m^*\}/VAR\{S_m\}$

$E(r) = \sum_{i=1}^{m} r_i, \quad VAR\{r\} = VAR\{r_1, r_2, ..., r_m\}$

Poisson Simulation Results

Sample Size (1)	Mean Value (2)	Theoretical Sum (3)	Sample Sum (4)		Bootstrap Var. (5)		Ratio (6)	
			mean	std.	mean	std.	mean	std.
20	0.5	10	10.16	3.43	9.46	4.32	**0.93**	0.23
	1	20	19.78	4.60	18.44	7.10	**0.93**	0.27
	5	100	100.60	9.20	99.45	33.50	**0.99**	0.31
	10	200	198.22	14.80	189.98	52.00	**0.96**	0.26
	20	400	400.36	21.30	395.50	120.00	**0.99**	0.30
50	0.5	25	25.28	5.00	24.89	6.80	**0.99**	0.20
	1	50	49.60	6.76	48.67	12.20	**0.98**	0.20
	5	250	251.38	15.85	252.39	15.89	**1.01**	0.20
	10	500	501.23	21.35	496.16	87.00	**0.99**	0.16
	20	1000	996.33	33.28	951.16	185.73	**0.96**	0.19
100	0.5	50	49.94	7.50	50.16	9.71	**1.01**	0.13
	1	100	101.47	10.45	102.77	22.20	**1.01**	0.17
	5	500	500.10	22.38	494.83	72.12	**0.99**	0.14
	10	1000	1000.20	33.93	1025.70	169.00	**1.03**	0.16
	20	2000	2003.30	45.60	1996.90	355.60	**1.00**	0.17

Negative Binomial Simulation Results

Sam. Size (1)	Mean (2)	Disp. Para. (3)	Theo. Sum (4)	Theo. Var. (5)	Sample Sum (6)		Mean of Boot. Var. (7)	Ratio (8)	
					mean	Var.		mean	std.
20	1	2	20	30	20.56	29.74	29.59	**0.93**	0.30
	5	1	100	600	100.49	560.80	571.33	**0.91**	0.32
	5	2	100	350	99.19	399.43	326.10	**0.91**	0.31
	20	2	400	4400	395.93	4574.90	4181.30	**0.92**	0.33
50	1	2	50	75	51.15	92.92	76.34	**0.98**	0.21
	5	1	250	1500	256.04	1856.10	1535.50	**0.94**	0.26
	5	2	250	875	253.31	1154.40	876.70	**0.96**	0.20
	20	2	1000	11000	993.62	9492.80	10596.00	**0.96**	0.26
100	1	2	100	150	99.41	177.32	146.51	**0.98**	0.17
	5	1	500	3000	504.72	3287.60	3072.30	**1.00**	0.20
	5	2	500	1750	495.92	2268.50	1732.20	**0.99**	0.17
	20	2	2000	22000	1993.00	22186.00	21477.00	**0.98**	0.19

Case Study

- Treatment: Red-light Cameras
- Data (Charlotte DOT, NC)

Site No.	Crashes Before (π_i)	Crashes After (λ_i)		Site No.	Crashes Before (π_i)	Crashes After (λ_i)
1	4	2		10	23	11
2	29	10		11	24	20
3	27	12		12	61	34
4	18	10		13	32	43
5	44	28		14	25	27
6	26	16		15	27	28
7	9	14		16	33	25
8	43	46		17	18	12
9	25	29		Total	468	367

Results

Statistics	Traditional Method	Bootstrap Method
$\hat{\pi}$	468	468
$\hat{\lambda}$	367	367
$\hat{\delta} = \hat{\pi} - \hat{\lambda}$	101	101
$VAR\{\hat{\pi}\}$	468	2939
$VAR\{\hat{\lambda}\}$	367	2400
$VAR\{\hat{\delta}\}$	835	5339

QUESTIONS