Expected Loss Based Analysis

Rain

5/27/2021

This analysis is based on Bayesian result according to expected loss < threshold of caring. Threshold of caring is different from probblem to problem.

Recall that

Case ID	Application	Distribution	Effect	Theshold of Caring
1	Payer Conversion	Bernoulli	False	0.001
2	Payer Conversion	Bernoulli	True	0.001
3	Total Moves	Poisson	False	0.00001
4	Total Moves	Poisson	True	0.00001
5	Revenue	Bernoulli-Exponential	False	0.001
6	Revenue	Bernoulli-Exponential	True	0.001

False/True Positive Rate

All the percentages in the following tables are proportion of the simulated experiments that led to the conclusion that the treatment is chosen.

In Case 1, 3, and 5 where the underlying treatment and control have no difference, the percentages mean false positive rates. Therefore, the lower the percentage, the better.

In Case 2, 4, and 6 where the treatment is better than control, the percentages mean true positive rates. Therefore, the higher the percentage, the better.

Areas of Analysis

1. Effect of peeking

We calculate peek_multiplier - how many times more likely we would choose treatment if monitor daily and stop the experiment earlier when we see a result that's positive enough. Note that only sample size of 500 is used because when the sample size is large enough, % of accepting treatment goes to 100% very quickly and would skew the peek_multiplier.

Result: Bayesian suffers from peeking as well.

case	avg_peek_multiplier_freq	avg_peek_multiplier_bayes
1	5.33	4.20
3	5.87	4.33
5	19.55	8.10

2. Effect of sample size

Conclusion: When treatment is not better than control, the false positive rate is controled at 5%.

case	sample_size_per_day	avg_freq_treat	avg_freq_treat_peek	avg_bayes_treat	avg_bayes_treat_peek
1	500	4.45%	23.3%	2.4%	9.65%
1	5000	4.45%	25.2%	7.2%	20.75%
1	50000	5.2%	28.25%	19.4%	43.9%
3	500	5.07%	28.53%	1%	4.27%
3	5000	4.93%	27.47%	3%	9.6%
3	50000	5.33%	28.73%	5.73%	19.6%
5	500	5.75%	79.75%	11.75%	94.75%
5	5000	4.5%	81.75%	29%	99.75%

Conclusion: When treatment is better than control, Bayesian has slightly more power.

case	sample_size_per_day	avg_freq_treat	avg_freq_treat_peek	avg_bayes_treat	avg_bayes_treat_peek
2	500	23.78%	42.98%	26.02%	37.9%
2	5000	60.42%	71.72%	69.23%	78.3%
2	50000	97.8%	99%	99.88%	99.95%
4	500	14.73%	38.13%	6.57%	13.93%
4	5000	52.4%	65.37%	51.97%	57.23%
4	50000	62.57%	74.43%	70.87%	79.53%
6	500	4.33%	79.33%	10%	93.5%
6	5000	6.5%	86.08%	21.92%	98.83%

3. Effect of prior parameter selections

We compare average false/true positive rate by directional, confident, and wrong priors.

Conclusio: Bayesian's prior parameters do not matter much for the sample sizes we have.

case	prior	avg_freq_treat	avg_freq_treat_peek	avg_bayes_treat	avg_bayes_treat_peek
1	confident	4.8%	26.2%	9.8%	24.6%
1	directional	4.6%	25%	9.93%	25%
1	neutral	5.27%	26.8%	10.8%	25.8%
1	wrong	4.13%	24.33%	8.13%	23.67%
2	confident	60.73%	70%	64.63%	70.6%
2	directional	60.47%	71.93%	64.77%	72.57%
2	neutral	60.07%	70.93%	64.57%	72.33%
2	wrong	61.4%	72.07%	66.2%	72.7%
3	confident	5.53%	27.2%	3.2%	11.87%
3	directional	5.6%	28.27%	3.67%	11.33%
3	wrong	4.2%	29.27%	2.87%	10.27%
4	confident	42.6%	59.6%	42.37%	49.93%
4	directional	43.67%	58.73%	43.73%	50.83%
4	wrong	43.43%	59.6%	43.3%	49.93%
5	directional-directional	8.5%	81.5%	22%	98%
5	directional-wrong	2.5%	79.5%	17.5%	98.5%

case	prior	avg_freq_treat	avg_freq_treat_peek	avg_bayes_treat	avg_bayes_treat_peek
5	wrong-directional	6.5%	81%	22%	97.5%
5	wrong-wrong	3%	81%	20%	95%
6	directional-directional	5.17%	83.33%	13.67%	95.67%
6	directional-wrong	5.17%	82.33%	17.5%	96%
6	wrong-directional	5.33%	83.17%	16.17%	96.83%
6	wrong-wrong	6%	82%	16.5%	96.17%