

P(A>B) Based Analysis

Rain

5/27/2021

This analysis is based on Bayesian result according to $P(A>B) > 95\%$.

Recall that

Case ID	Application	Distribution	Effect
1	Payer Conversion	Bernoulli	False
2	Payer Conversion	Bernoulli	True
3	Total Moves	Poisson	False
4	Total Moves	Poisson	True

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 and 3 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 and 4 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	4.72	4.62
3	6.87	5.03

2. Effect of sample size

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

case	sample_size_per_day	avg_freq_treat	avg_freq_treat_peek	avg_bayes_treat	avg_bayes_treat_peek
1	500	4.95%	22.85%	4.9%	22.75%
1	5000	4.8%	24.45%	5.25%	23.7%
1	50000	4.4%	26.8%	4.5%	23.1%
3	500	3.93%	26.8%	5%	25.07%
3	5000	5.13%	28.27%	5.2%	24.13%
3	50000	4.93%	26.27%	5%	25.93%

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	23.4%	41.88%	32.02%	51.8%
2	5000	59.95%	71.45%	65.65%	78.28%
2	50000	97.9%	99.05%	98.85%	99.48%
4	500	15%	38.63%	21.4%	44.4%
4	5000	52.93%	66.27%	55.93%	67.5%
4	50000	61.83%	73.9%	67.63%	79.3%

3. Effect of prior parameter selections

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

Conclusion: 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	4.93%	25.47%	5.4%	23.13%
1	directional	4.8%	24.07%	5.07%	22.47%
1	neutral	4.53%	26.07%	4.07%	24.33%
1	wrong	4.6%	23.2%	5%	22.8%
2	confident	60.1%	71.1%	65%	76.03%
2	directional	59.97%	70.17%	65.07%	75.87%
2	neutral	60.93%	71.3%	65.9%	77.57%
2	wrong	60.67%	70.6%	66.07%	76.6%
3	confident	4.93%	27.93%	5.13%	24.6%
3	directional	4.4%	26.67%	4.87%	25.33%
3	wrong	4.67%	26.73%	5.2%	25.2%
4	confident	44%	60.47%	49.57%	65.2%
4	directional	43.47%	59.33%	48.1%	63.37%
4	wrong	42.3%	59%	47.3%	62.63%