# Analysis

Rain

2/19/2021

## Overview

### 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.

### Threshold of Caring

Threshold of caring indicates what level of expected loss you are willing to take by choosing treatment when it's actually not better than control.

A high threshold means you are very willing to choose treatment.

A low threshold means you are very conservative and would only choose treatmet if it's obviously better than control.

Threshold of caring differs from problem to problem. We choose the threshold in each case based on which one is the most comparable to the frequentist test of alpha = 0.05.

In real application, we propose using simulations in this analysis to determine the best threshold, so that it's not abused by bad estimations.

### 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.
2. Effect of sample size. We compare average false/true positive rate by different sample sizes.
3. Effect of prior parameter selections. We compare average false/true positive rate by directional, confident, and wrong priors.

## Case 1

Threshold = 0.001.

Bayesian suffers from peeking as well but not as badly as Frequentist.

| avg_peek_multiplier_freq | avg_peek_multiplier_bayes |
|---|---|
| 5.33 | 4.2 |

Bayesian's false positive rate quickly increases with increasing sample size.

| sample_size_per_day | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| 500 | 4.45% | 23.3% | 2.4% | 9.65% |
| 5000 | 4.45% | 25.2% | 7.2% | 20.75% |
| 50000 | 5.2% | 28.25% | 19.4% | 43.9% |

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

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| confident | 4.8% | 26.2% | 9.8% | 24.6% |
| directional | 4.6% | 25% | 9.93% | 25% |
| neutral | 5.27% | 26.8% | 10.8% | 25.8% |
| wrong | 4.13% | 24.33% | 8.13% | 23.67% |

## Case 2

Threshold = 0.001.

Bayesian suffers from peeking as well but not as badly as Frequentist.

| avg_peek_multiplier_freq | avg_peek_multiplier_bayes |
|---|---|
| 3.06 | 2.02 |

Bayesian's true positive rate quickly increases with increasing sample size.

| sample_size_per_day | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| 500 | 23.78% | 42.98% | 26.02% | 37.9% |
| 5000 | 60.42% | 71.72% | 69.23% | 78.3% |
| 50000 | 97.8% | 99% | 99.88% | 99.95% |

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

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| confident | 60.73% | 70% | 64.63% | 70.6% |
| directional | 60.47% | 71.93% | 64.77% | 72.57% |

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| neutral | 60.07% | 70.93% | 64.57% | 72.33% |
| wrong | 61.4% | 72.07% | 66.2% | 72.7% |

## Case 3

Threshold = 0.00001.

Bayesian suffers from peeking as well but not as badly as Frequentist.

| avg_peek_multiplier_freq | avg_peek_multiplier_bayes |
|---|---|
| 5.87 | 4.33 |

Bayesian's false positive rate quickly increases with increasing sample size.

| sample_size_per_day | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| 500 | 5.07% | 28.53% | 1% | 4.27% |
| 5000 | 4.93% | 27.47% | 3% | 9.6% |
| 50000 | 5.33% | 28.73% | 5.73% | 19.6% |

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

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| confident | 5.53% | 27.2% | 3.2% | 11.87% |
| directional | 5.6% | 28.27% | 3.67% | 11.33% |
| wrong | 4.2% | 29.27% | 2.87% | 10.27% |

## Case 4

Threshold = 0.00001.

Bayesian suffers from peeking as well but not as badly as Frequentist.

| avg_peek_multiplier_freq | avg_peek_multiplier_bayes |
|---|---|
| 3.68 | 2.82 |

Bayesian's true positive rate quickly increases with increasing sample size.

| sample_size_per_day | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| 500 | 14.73% | 38.13% | 6.57% | 13.93% |
| 5000 | 52.4% | 65.37% | 51.97% | 57.23% |
| 50000 | 62.57% | 74.43% | 70.87% | 79.53% |

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

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| confident | 42.6% | 59.6% | 42.37% | 49.93% |
| directional | 43.67% | 58.73% | 43.73% | 50.83% |
| wrong | 43.43% | 59.6% | 43.3% | 49.93% |

## Case 5

Threshold = 0.001.

Bayesian suffers from peeking as well but not as badly as Frequentist.

| avg_peek_multiplier_freq | avg_peek_multiplier_bayes |
|---|---|
| 19.55 | 8.1 |

Bayesian's false positive rate quickly increases with increasing sample size.

| sample_size_per_day | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| 500 | 5.75% | 79.75% | 11.75% | 94.75% |
| 5000 | 4.5% | 81.75% | 29% | 99.75% |

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

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| directional-directional | 8.5% | 81.5% | 22% | 98% |
| directional-wrong | 2.5% | 79.5% | 17.5% | 98.5% |
| wrong-directional | 6.5% | 81% | 22% | 97.5% |
| wrong-wrong | 3% | 81% | 20% | 95% |

## Case 6

Threshold = 0.001.

Bayesian suffers from peeking as well but not as badly as Frequentist.

| avg_peek_multiplier_freq | avg_peek_multiplier_bayes |
|---|---|
| 22 | 9.88 |

Bayesian's true positive rate quickly increases with increasing sample size.

| sample_size_per_day | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| 500 | 4.33% | 79.33% | 10% | 93.5% |
| 5000 | 6.5% | 86.08% | 21.92% | 98.83% |

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

| prior | avg_freq_treat | avg_freq_treat_peek | avg_bayes_treat | avg_bayes_treat_peek |
|---|---|---|---|---|
| directional-directional | 5.17% | 83.33% | 13.67% | 95.67% |
| directional-wrong | 5.17% | 82.33% | 17.5% | 96% |
| wrong-directional | 5.33% | 83.17% | 16.17% | 96.83% |
| wrong-wrong | 6% | 82% | 16.5% | 96.17% |