Progress report 3/9/23

## Load libraries and the results

```r
library(tidyverse)

# full analysis for 5690 resamples
full_results <- readRDS("full_nc_results.rds")
# subsampling analysis
subsampling_results <- readRDS("subsampling_nc_results.rds")
# full analysis for 5e5 resamples
groundtruth_results <- readRDS("ground_truth_nc_results.rds")
```

Assessing the fit quality of MLE and MoM on all 5e5 resamples

```r
groundtruth_results |>
  filter(ratio < 1000) |>
  pivot_wider(names_from = method, values_from = ratio) |>
  ggplot(aes(x = MLE, y = MoM)) +
  geom_point() +
  geom_abline() +
  facet_wrap(error_type ~ tail, scales = "free") +
  scale_x_log10() +
  scale_y_log10() +
  labs(x = "MLE fit quality on 5e5 resamples",
       y = "MoM fit quality on 5e5 resamples")
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```
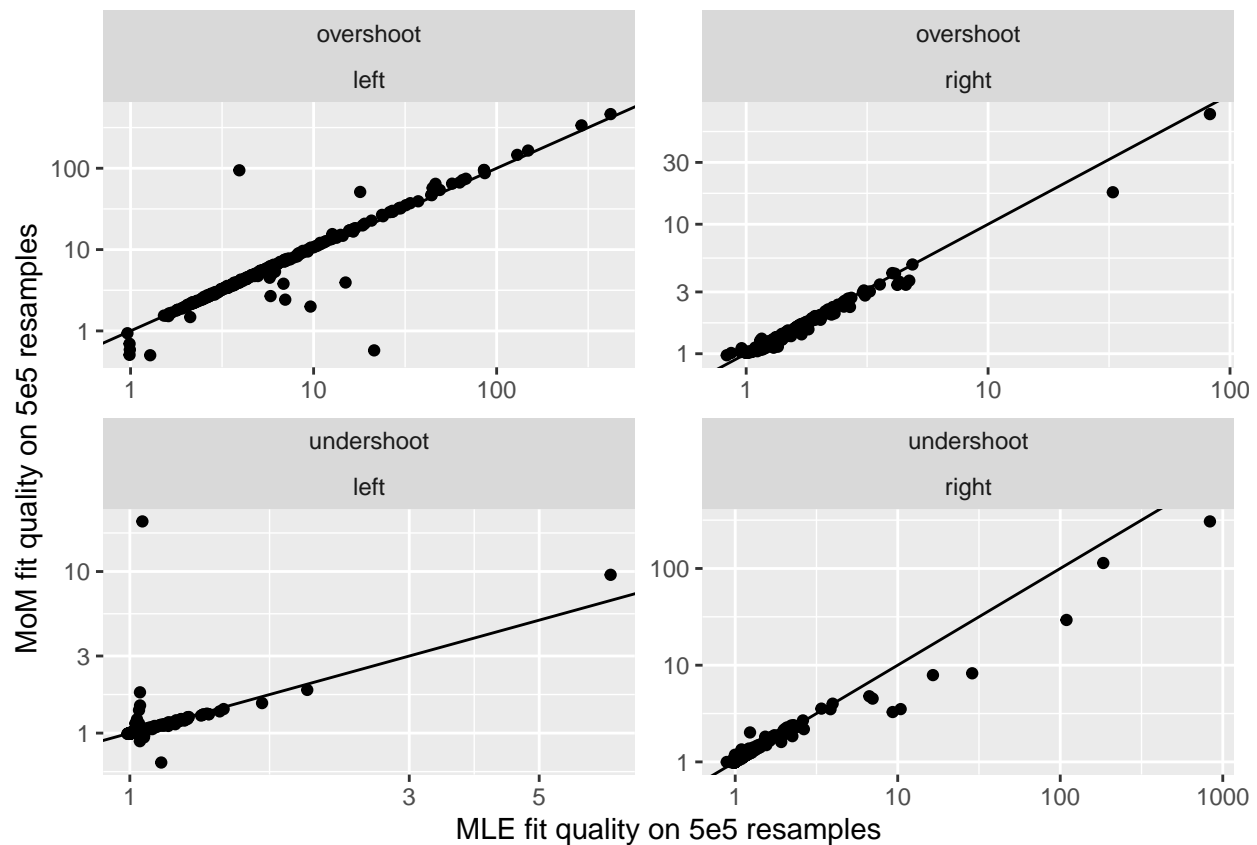
The conclusion here is that MLE and MoM have broadly similar fit quality on 5e5 resamples. Let's take a closer look at the MoM fit quality.

```
groundtruth_results |>
  filter(method == "MoM",
         ratio < 100) |>
  ggplot(aes(x = ratio)) +
  geom_histogram() +
  geom_vline(xintercept = 1, linetype = "dashed") +
  facet_grid(tail ~ error_type, scales = "free_y") +
  scale_x_log10()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
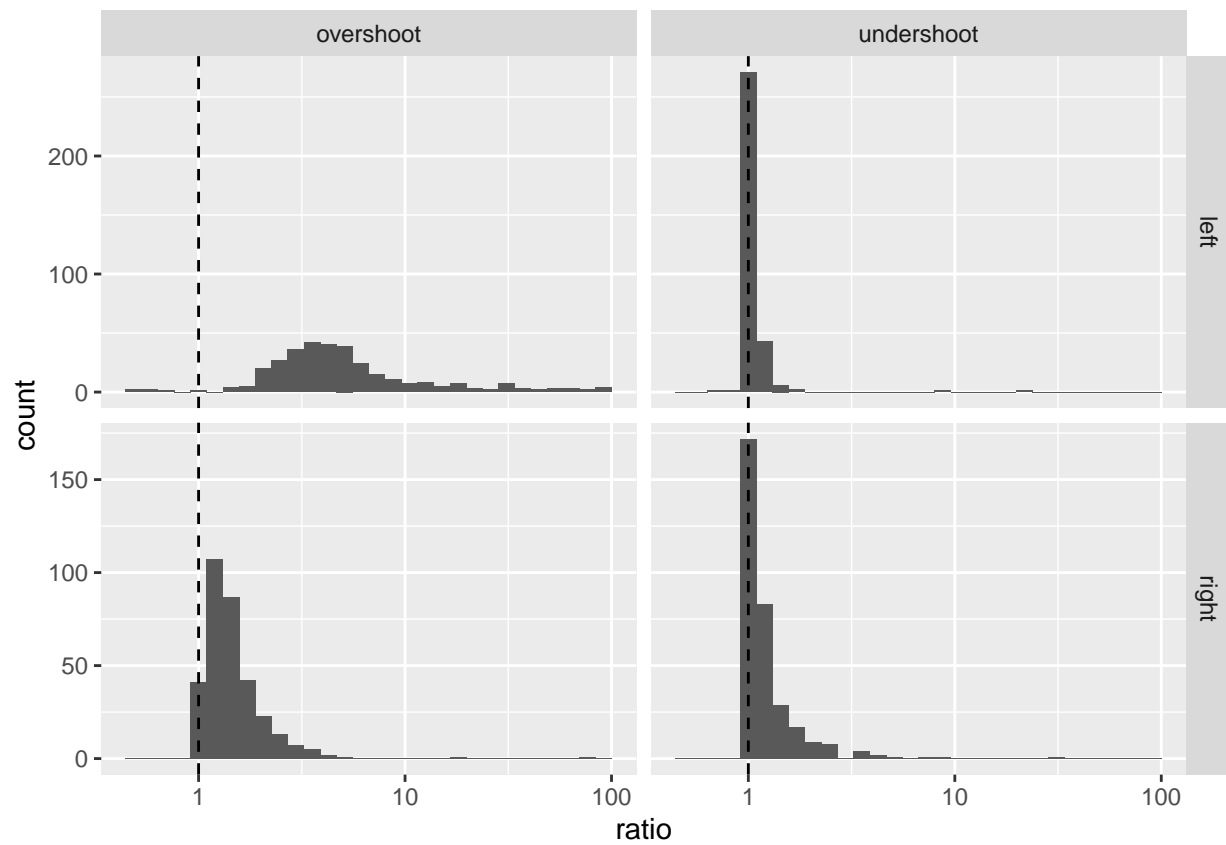
```
groundtruth_results |>
  filter(method == "MoM",
         ratio < 100) |>
  ggplot(aes(x = error_type, y = ratio, fill = tail)) +
  geom_boxplot() +
  scale_y_continuous(trans = "log10", breaks = c(1, 2, 5, 10, 100))
```
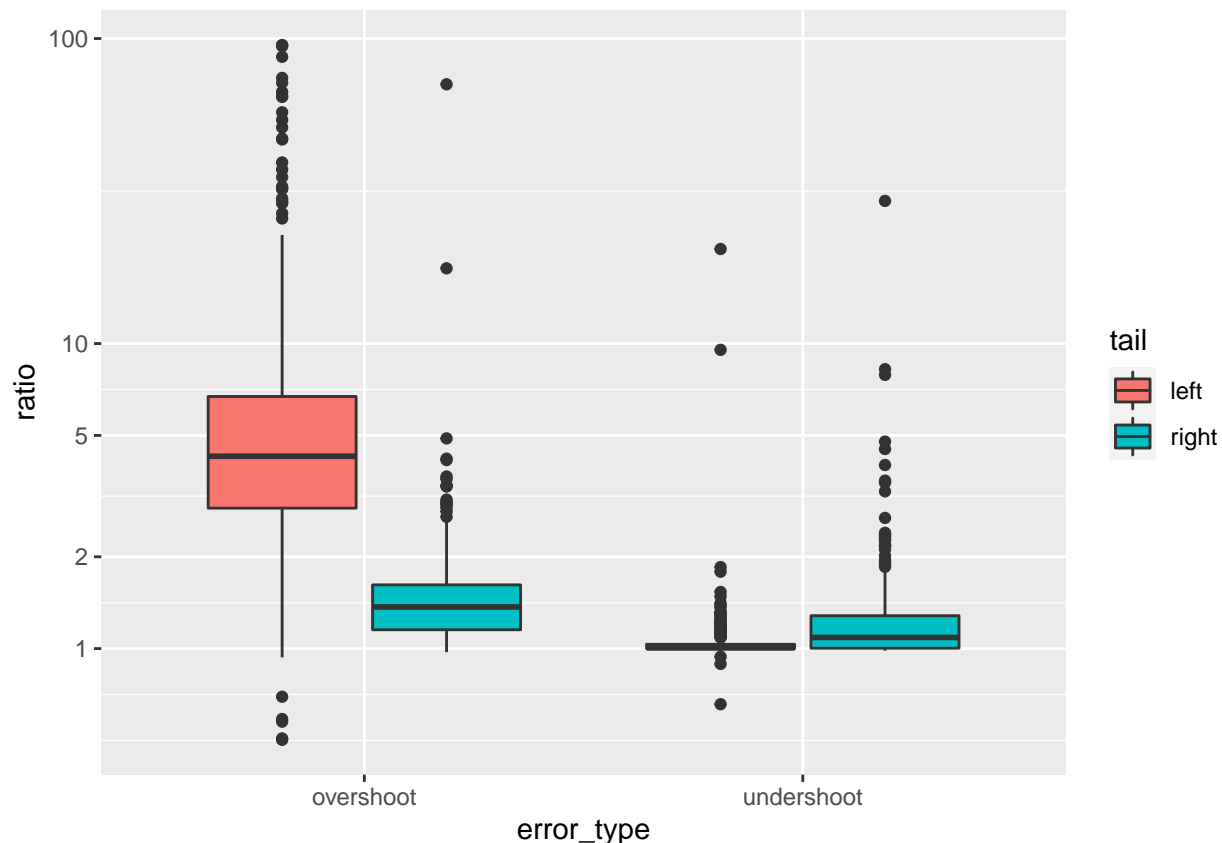
```
groundtruth_results |>
  filter(method == "MoM") |>
  group_by(tail, error_type) |>
  summarise(median = median(ratio), .groups = "drop") |>
  pivot_wider(names_from = tail, values_from = median)
```

```
## # A tibble: 2 x 3
##   error_type  left right
##   <chr>      <dbl> <dbl>
## 1 overshoot   4.34  1.37
## 2 undershoot  1.01  1.09
```

In summary, we see that MoM tends to overshoot more than it undershoots, and that the overshoot problem seems worse in the left tail. The good news is that the extent of the undershoot is fairly low.

Assessing stability of fit quality as a function of number of resamples

```
subsampling_results |>
  mutate(id = as.factor(id)) |>
  filter(evaluation == "ground truth") |>
  group_by(id, error_type, num_samples_fit, tail) |>
  summarise(mean_ratio = mean(ratio), .groups = "drop") |>
  group_by(id, tail, error_type) |>
  filter(min(mean_ratio) < 3) |>
  ungroup() |>
  ggplot(aes(x = num_samples_fit, y = mean_ratio, color = id)) +
  geom_point() +
```

4

```
geom_line() +
facet_wrap(tail ~ error_type, scales = "free") +
scale_x_log10()
```
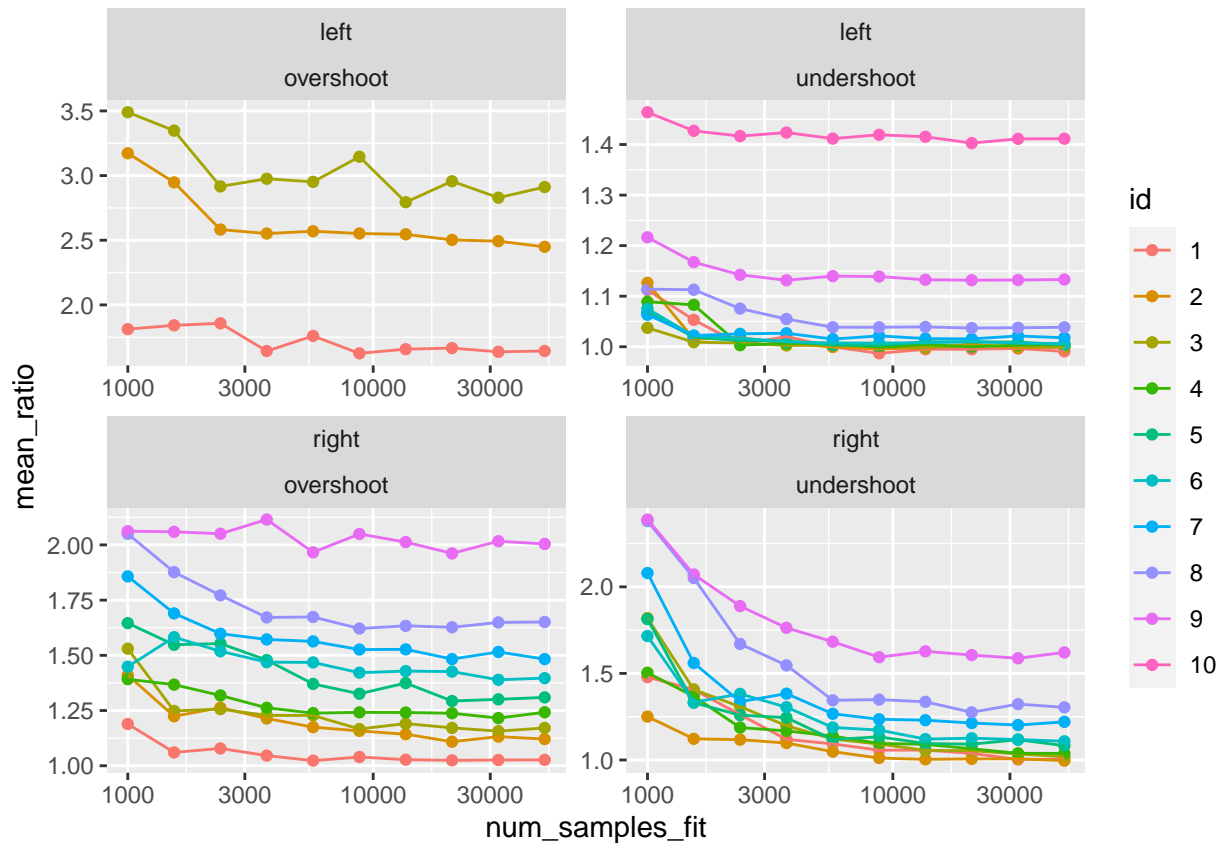

We conclude that the fit quality is relatively stable as a function of the number of samples after roughly 5690.

Estimating fit quality given only the resamples we have

```
subsampling_results |>
  mutate(id = as.factor(id)) |>
  pivot_wider(names_from = evaluation, values_from = ratio) |>
  mutate(estimate_to_truth = estimate / `ground truth`) |>
  group_by(id, error_type, num_samples_fit, tail) |>
  summarise(mean_estimate_to_truth = mean(estimate_to_truth), .groups = "drop") |>
  ggplot(aes(x = num_samples_fit, y = mean_estimate_to_truth, color = id)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 1, linetype = "dashed") +
  facet_wrap(tail ~ error_type, scales = "free") +
  scale_x_log10()
```
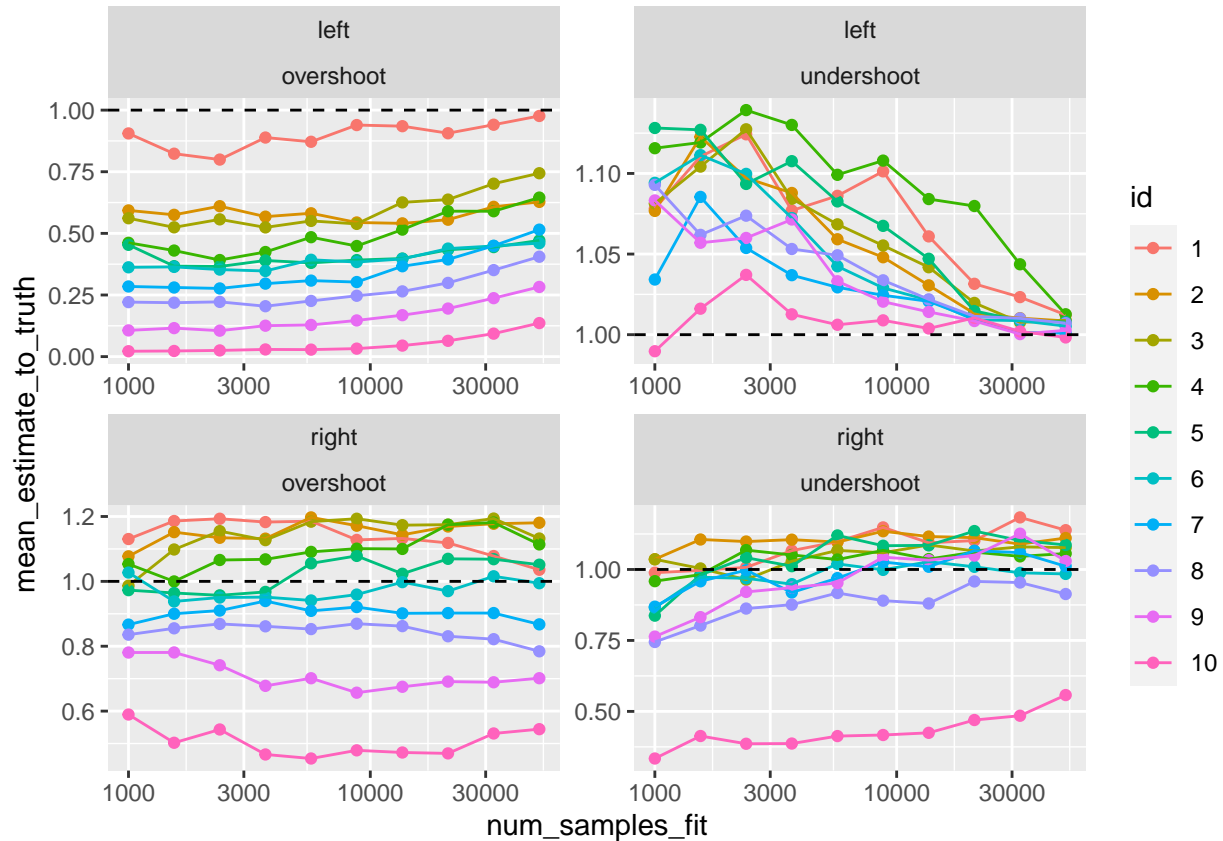
This set of plots reveal the following information.\ (1) The top left plot indicates that most (all) of the downsample estimate tends to underestimate the p-value ratio compared with ground truth at least under the average sense.

(2) The bottom left plot indicates similar phenomenon for right tail in terms of overshoot although less severe underestimate.

(3)The right set of plot shows more promising results. Except for the case where id = 10, most of other curves are around 1, implying dowsample estimates are not underestimating a lot in terms of undershoot p-value ratio.

This figure adds the quantile error bar to supplement last plot.

```r
full_results |>
  filter(ratio < 1e3) |>
  pivot_wider(names_from = evaluation, values_from = ratio) |>
  ggplot(aes(x = `ground truth`, y = estimate)) +
  geom_point() +
  geom_abline() +
  facet_wrap(tail ~ error_type, scales = "free") +
  scale_x_log10() +
  scale_y_log10()
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```
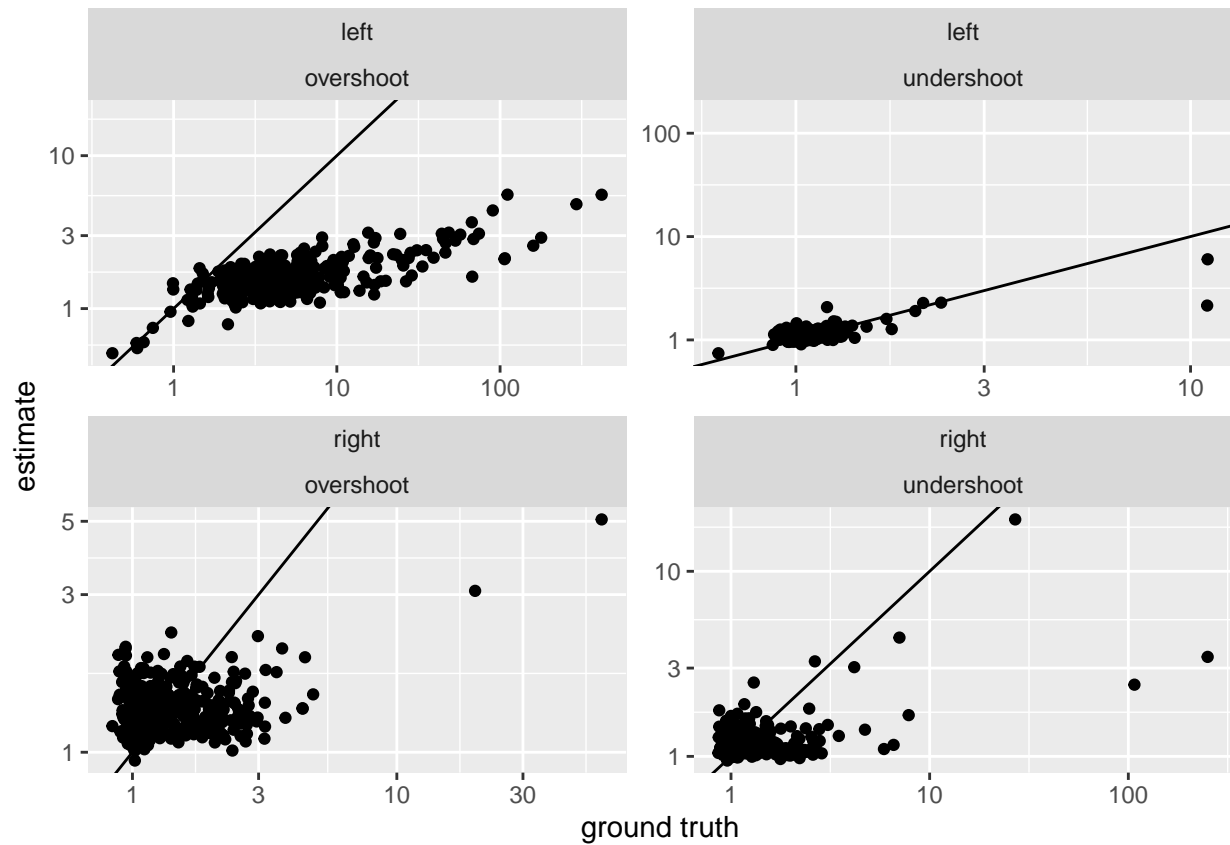
This set of plot marks the alignment between estimate and ground truth p-value ratio. We can see the most aligned result is undershoot left. The worst is overshoot left case. Some refined analysis is needed to right tail case.