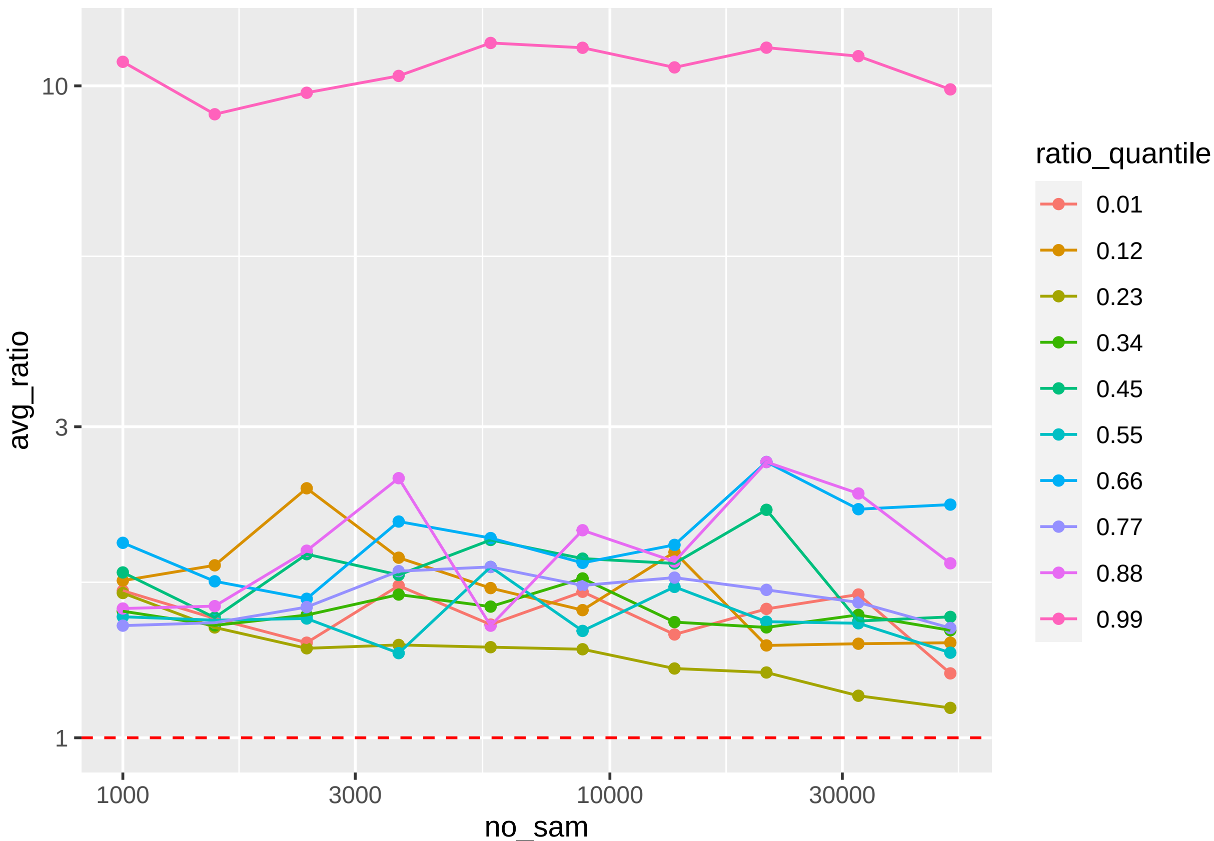
Downsample result analysis

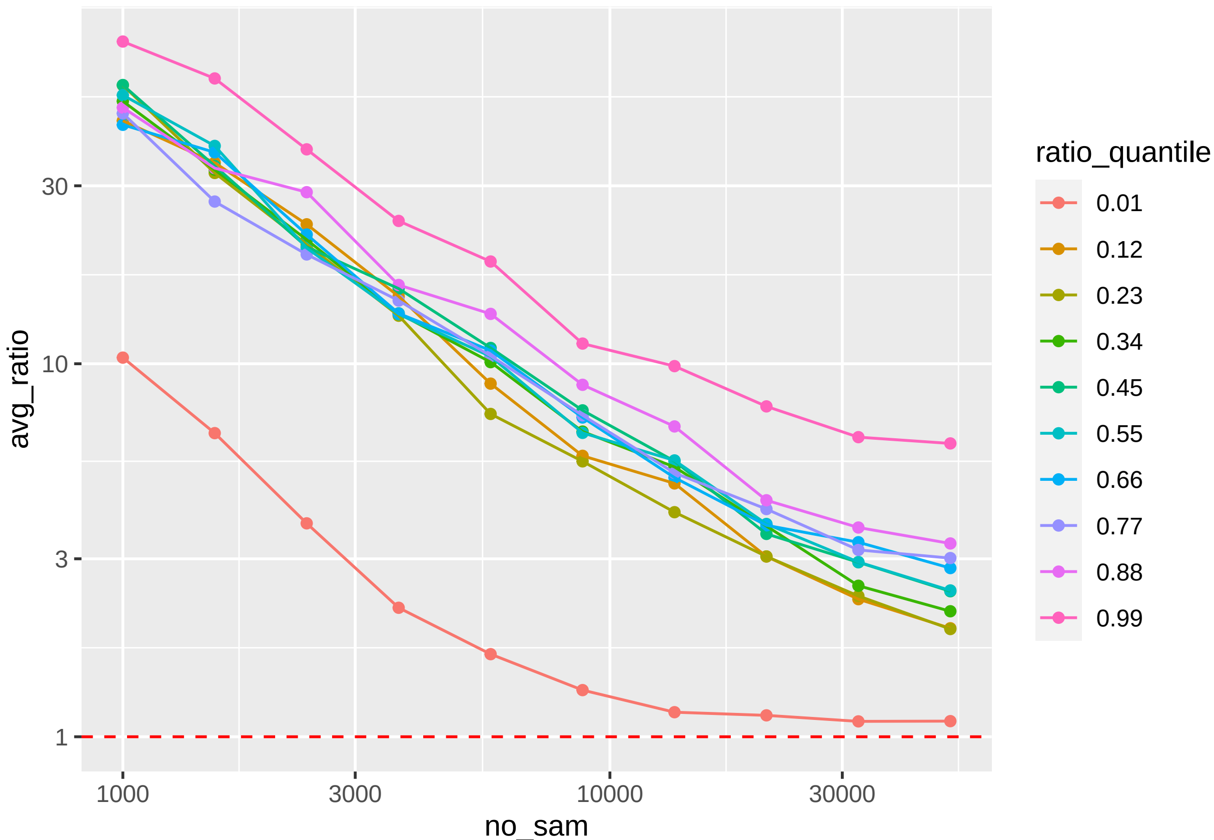
1.Load results

```r
library(dplyr)
library(ggplot2)
library(tidyverse)
undershoot <- read_csv("undershoot_refine_power.csv")[,-1]
overshoot <- read_csv("overshoot_refine_power.csv")[,-1]
quantile_list <- seq(0.01, 0.99, length.out = 10)
no_sam <- round(exp(seq(log(1e3), log(5e4), length.out = 10)))
# rearrange the data frame
B <- 100
undershoot_df <- data.frame(id = rep(1:B, 10*10),
                            ratio_value = 0,
                            no_sam = 0,
                            ratio_quantile = 0)
overshoot_df <- data.frame(id = rep(1:B, 10*10),
                           ratio_value = 0,
                           no_sam = 0,
                           ratio_quantile = 0)

# i: quantile; j: no of sample
for (i in 1:10) {
  for (j in 1:10) {
    start <- (j - 1 + (i-1)*10)*B +1
    end <- (j + (i-1)*10)*B
    undershoot_df[start:end, 2] <- as.vector(undershoot[(((j-1)*B+1) :(j*B)), (i-1)*3+2])[[1]]
    undershoot_df[start:end, 3] <- rep(no_sam[j], B)
    undershoot_df[start:end, 4] <- rep(quantile_list[i], B)
    overshoot_df[start:end, 2] <- as.vector(overshoot[(((j-1)*B+1) :(j*B)), (i-1)*3+2])[[1]]
    overshoot_df[start:end, 3] <- rep(no_sam[j], B)
    overshoot_df[start:end, 4] <- rep(quantile_list[i], B)
  }
}

# plot for undershoot
under_ratio_avg <- undershoot_df |>
  dplyr::group_by_at(c("no_sam", "ratio_quantile")) |>
  summarise(avg_ratio = mean(ratio_value)) |>
  ungroup()
under_ratio_avg$ratio_quantile <- round(under_ratio_avg$ratio_quantile, 2)
under_ratio_avg$ratio_quantile <- as.character(under_ratio_avg$ratio_quantile)

over_ratio_avg <- overshoot_df |>
  dplyr::group_by_at(c("no_sam", "ratio_quantile")) |>
  summarise(avg_ratio = mean(ratio_value)) |>
  ungroup()
```

```r
over_ratio_avg$ratio_quantile <- round(over_ratio_avg$ratio_quantile, 2)
over_ratio_avg$ratio_quantile <- as.character(over_ratio_avg$ratio_quantile)

under_ratio_avg |>
  ggplot(aes_string(x = "no_sam", y = "avg_ratio", colour = "ratio_quantile")) +
  scale_x_log10() +
  scale_y_log10() +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 1, linetype = "dashed", colour = "red")
```




```r
over_ratio_avg |>
  ggplot(aes_string(x = "no_sam", y = "avg_ratio", colour = "ratio_quantile")) +
  scale_x_log10() +
  scale_y_log10() +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 1, linetype = "dashed", colour = "red")
```



2.Quantitative analysis

```
param_nc <- read_csv("figures/power_exploration/sknorm_tail_prob_500000_resamples_0.96_percentile/param_
param_twosides <- t(param_nc[,-1])
overshoot_ratio <- as.numeric(param_twosides[, 6])
undershoot_ratio <- as.numeric(param_twosides[, 7])
quantile_list <- seq(0.01, 0.99, length.out = 10)
overshoot_set <- data.frame(index = numeric(10), ratio = numeric(10))
undershoot_set <- data.frame(index = numeric(10), ratio = numeric(10))

# find distributions based on right tail
for (r in 1:10){
  dist <- abs(overshoot_ratio[331:660] - quantile(overshoot_ratio[331:660], quantile_list[r]))
  overshoot_set[r, 1] <- which(dist == min(dist))
  overshoot_set[r, 2] <- overshoot_ratio[which(dist == min(dist)) + 330]
  dist <- abs(undershoot_ratio[331:660] - quantile(undershoot_ratio[331:660], quantile_list[r]))
  undershoot_set[r, 1] <- which(dist == min(dist))
  undershoot_set[r, 2] <- undershoot_ratio[which(dist == min(dist)) + 330]
}

# accuracy matrix for undershoot matrix
undershoot_acc <- matrix(abs(under_ratio_avg$avg_ratio - rep(undershoot_set$ratio, 10)), 10, 10)
colnames(undershoot_acc) <- as.character(no_sam)
rownames(undershoot_acc) <- as.character(round(quantile_list, 3))
undershoot_acc
```

```
##              1000        1544      2385      3684      5690      8788
## 0.01   0.712023262  0.555384708  0.4301791  0.7420020  0.5226668  0.7056231
## 0.119  0.752849821  0.849471969  1.4244262  0.8996400  0.7075108  0.5804096
## 0.228  0.672333878  0.481414480  0.3766813  0.3930485  0.3821152  0.3721721
## 0.337  0.556797915  0.480415276  0.5355989  0.6515166  0.5824334  0.7486627
## 0.446  0.753914158  0.489542115  0.8732414  0.7381342  0.9718284  0.8437872
## 0.554  0.454483034  0.436190339  0.4450323  0.2696790  0.7494660  0.3797056
## 0.663  0.840971361  0.587936687  0.4839614  0.9963629  0.8745994  0.7048850
## 0.772  0.233068117  0.248448596  0.3337567  0.5484513  0.5754027  0.4599633
## 0.881  0.008741218  0.004519002  0.3489229  0.9140936  0.1016313  0.4939879
## 0.99  17.834957829 19.676183619 18.9625990 18.3638707 17.0857700 17.2821506
##           13572      20961      32374      50000
## 0.01   0.4712846  0.6073488  0.6897165  0.2863519
## 0.119  0.9368987  0.3962986  0.4046714  0.4097048
## 0.228  0.2822267  0.2643189  0.1650235  0.1163570
## 0.337  0.4977908  0.4698053  0.5365740  0.4547926
## 0.446  0.8119496  1.1979988  0.4712119  0.4934118
## 0.554  0.6250791  0.4280129  0.4196965  0.2716298
## 0.663  0.8264327  1.4987739  1.0920349  1.1285157
## 0.772  0.5076697  0.4330893  0.3600990  0.2204713
## 0.881  0.2760708  1.0608717  0.7825741  0.2645616
## 0.99  18.0483392 17.2775770 17.6159021 18.8454109
```

```r
# accuracy matrix for overshoot matrix
overshoot_acc <- matrix(abs(over_ratio_avg$avg_ratio - rep(overshoot_set$ratio, 10)), 10, 10)
colnames(overshoot_acc) <- as.character(no_sam)
rownames(overshoot_acc) <- as.character(round(quantile_list, 3))
overshoot_acc
```

```
##           1000      1544      2385      3684       5690     8788     13572
## 0.01   9.429248  5.558956  2.776257  1.259675  0.7074014 0.376081 0.2059259
## 0.119 43.700717 33.471509 22.528478 14.043679  7.7242960 4.544698 3.6565718
## 0.228 54.621585 31.283491 19.707961 12.263640  6.1472134 4.286252 2.8120461
## 0.337 49.300583 31.862873 20.239954 12.382024  8.8332068 5.312701 4.0214665
## 0.446 54.521899 32.533712 19.211461 14.538990  9.6642239 6.138706 4.1191257
## 0.554 51.046158 36.917162 18.972348 12.120031  8.9999074 5.071928 4.0499043
## 0.663 42.140249 35.298523 20.633435 12.071566  9.2767664 5.605511 3.3793264
## 0.772 45.198670 25.508186 17.926587 13.060465  8.7888847 5.564574 3.3875622
## 0.881 46.487514 31.331926 26.715365 14.136036 11.4820449 6.657766 4.6679910
## 0.99  68.364169 53.475688 32.849880 19.428023 14.0748175 6.603144 5.1319048
##          20961     32374     50000
## 0.01  0.1835201 0.1419676 0.1435876
## 0.119 1.9215794 1.2170980 0.8329757
## 0.228 1.8556045 1.1962786 0.7561891
## 0.337 2.4059755 1.2741013 0.9038144
## 0.446 2.1368151 1.5750519 1.0942652
## 0.554 2.2634439 1.4849114 1.0114871
## 0.663 2.1161133 1.7509255 1.2610081
## 0.772 2.3689944 1.4601335 1.3041824
## 0.881 2.1824488 1.5156658 1.1744338
## 0.99  2.9627714 1.6337128 1.3922274
```

From the above figure and table, we can clearly see the distance between the estimated overshoot ratio and the true ratio decrease substantially with the increase of the number of resamples. But for undershoot case, such decrease is not obvious possibly due to the fact that the skew normal fit does not undershoot a lot and the maximum undershoot ratio are most arround 1.