

GPD vs Skew Normal

```
library(eva)
library(tidyverse)
library(katlabutils)
library(cowplot)
library(ggplot2)

# should work under low_moi folder
q <- 0.96
samples <- 5e5
working_dir <- getwd()

# get the file path for gpd parameters
subDir <- sprintf("figures/power_exploration/tail_prob_%d_resamples_%.2f_percentile",
  samples, q)
param <- read.csv(sprintf("%s/param.csv", file.path(working_dir, subDir)))
gpd_param <- t(matrix(as.numeric(unlist(param[, -1])), nrow = 6, ncol = 660))
gpd_param <- as.data.frame(gpd_param)
colnames(gpd_param) <- c("location", "scale", "shape", "statistic", "p-value", "p-value ratio")
p_value_ratio_gpd <- gpd_param$`p-value ratio`

# check if the inf value corresponds to bumpy distribution
# bumpy: 56, 143, 169, 175, 177, 185, 191, 198, 213, 236, 241, 243
# nonbumpy: 327, 325, 324
# both bumpy: 191
sort(which(p_value_ratio_gpd == Inf), decreasing = TRUE)
```

```
## [1] 657 655 654 528 521 515 243 241 236 213 191 177 175 169 143 56
```

```

# get the file path for skew normal parameter
subDir <- sprintf("figures/power_exploration/sknorm_tail_prob_%d_resamples_%.2f_perce
ntile", samples, q)
param <- read.csv(sprintf("%s/param.csv", file.path(working_dir, subDir)))
sknorm_param <- t(matrix(as.numeric(unlist(param[, -1])), nrow = 6, ncol = 660))
sknorm_param <- as.data.frame(sknorm_param)
colnames(sknorm_param) <- c("location", "scale", "shape", "statistic", "p-value", "p-
value ratio")
p_value_ratio_sknorm <- sknorm_param$`p-value ratio`

# compare left tail case
left_gpd_ratio <- p_value_ratio_gpd[1:330]
left_sknorm_ratio <- p_value_ratio_sknorm[1:330]

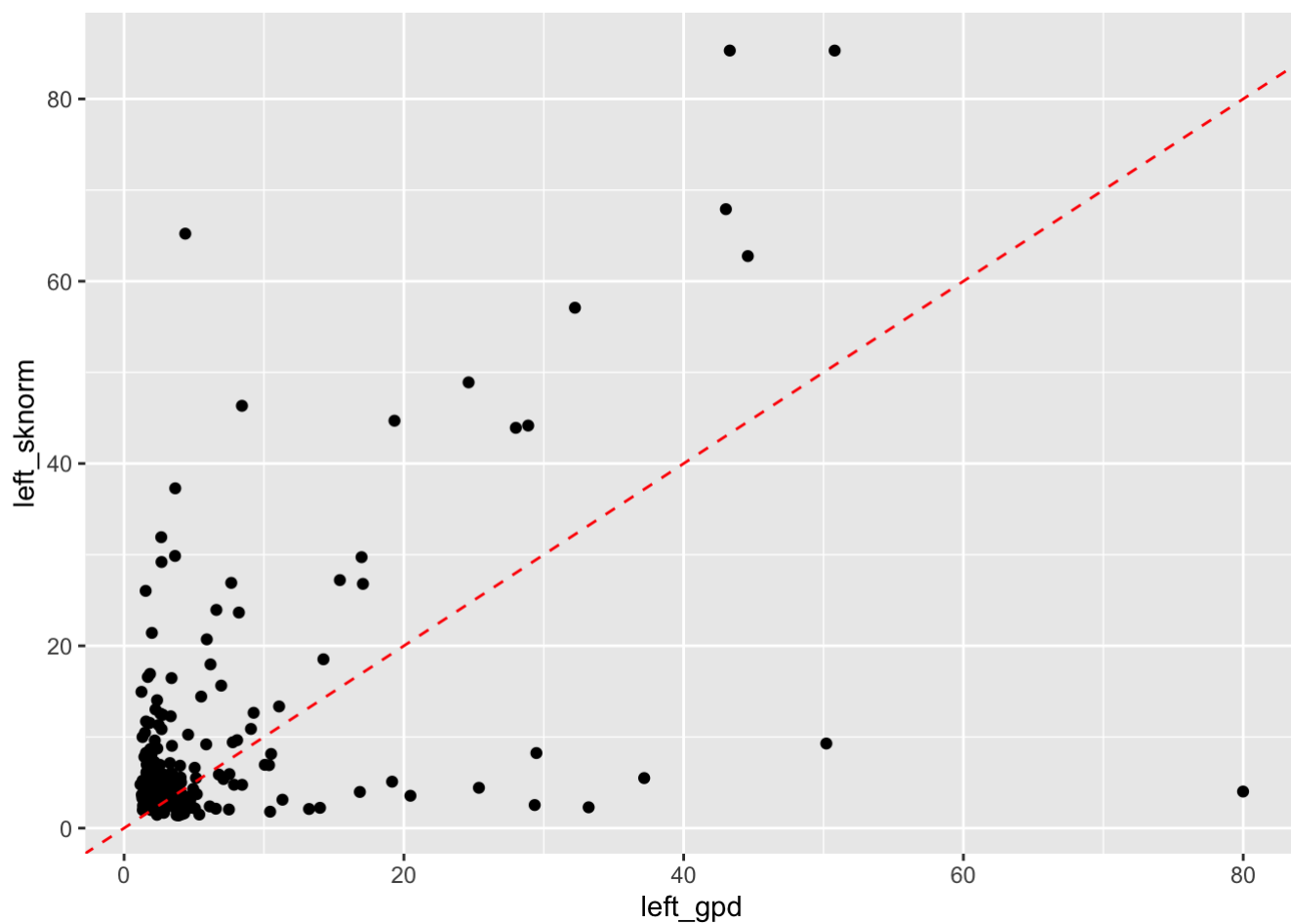
# leave the large value out
large_leftgpd_id <- which(left_gpd_ratio > 100)
large_leftsknorm_id <- which(left_sknorm_ratio > 100)
large_left_id <- c(large_leftgpd_id, large_leftsknorm_id)

# infinite value id
Inf_left_id <- which(left_gpd_ratio == "Inf")

# exclude the infinite value id and large value id
excl_id <- c(Inf_left_id, large_left_id)
left_comparison <- data.frame(left_gpd = left_gpd_ratio[-excl_id],
                             left_sknorm = left_sknorm_ratio[-excl_id])

left_comparison |>
  ggplot(aes(x = left_gpd, y = left_sknorm)) +
  geom_point() +
  geom_abline(linetype = "dashed", color = "red")

```



```
# no of gpd bigger than sknorm for left tail
length(which(left_gpd_ratio > left_sknorm_ratio))
```

```
## [1] 97
```

```
length(which(left_sknorm_ratio > left_gpd_ratio))
```

```
## [1] 233
```

```

# compare right tail case
right_gpd_ratio <- p_value_ratio_gpd[331:660]
right_sknorm_ratio <- p_value_ratio_sknorm[331:660]

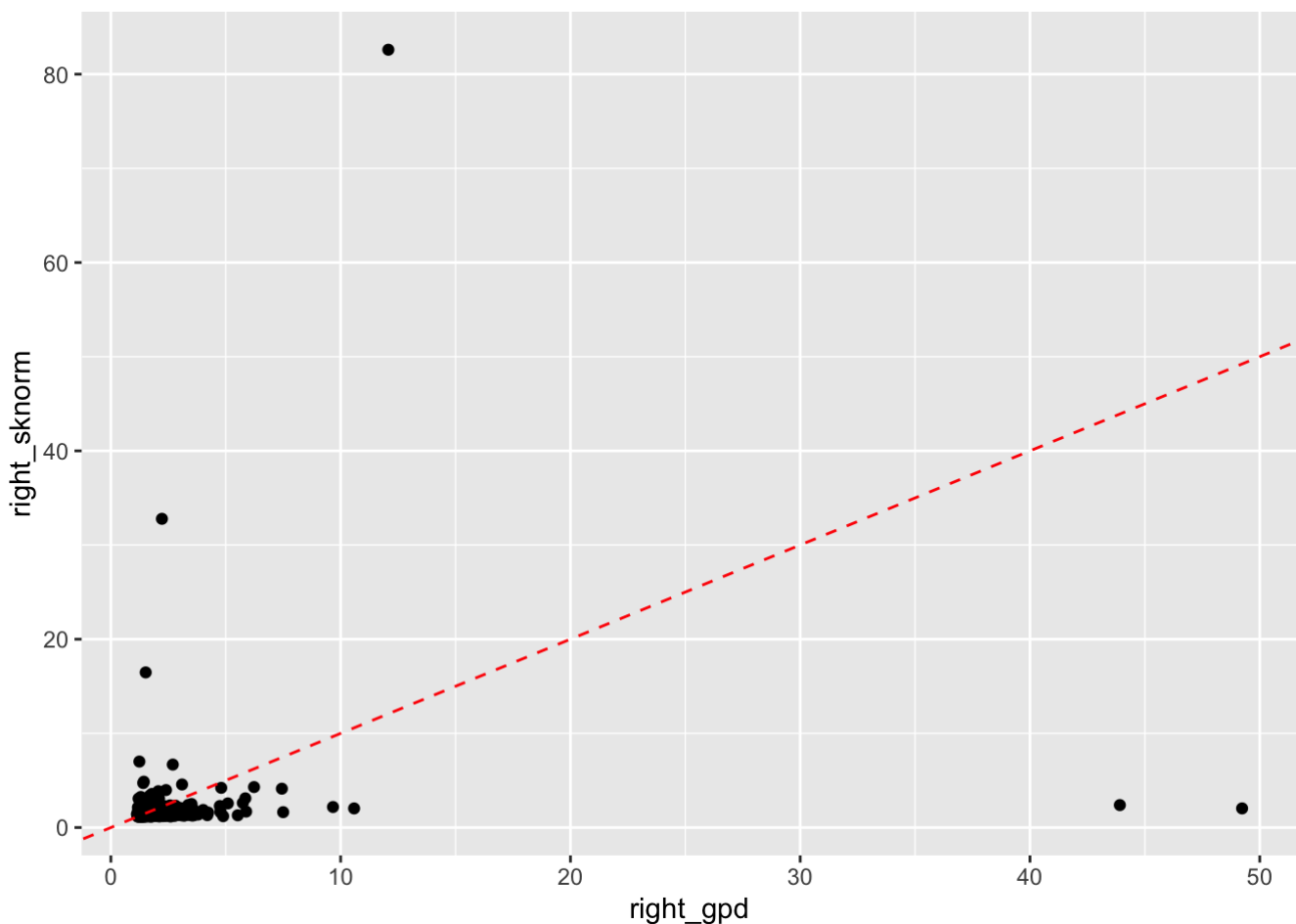
# leave the large value out
large_rightgpd_id <- which(right_gpd_ratio > 100)
large_rightsknorm_id <- which(right_sknorm_ratio > 100)
large_right_id <- c(large_rightgpd_id, large_rightsknorm_id)

# infinite value id
Inf_right_id <- which(right_gpd_ratio == "Inf")

# exclude the infinite value id and large value id
excl_id <- c(Inf_right_id, large_right_id)
right_comparison <- data.frame(right_gpd = right_gpd_ratio[-excl_id],
                               right_sknorm = right_sknorm_ratio[-excl_id])

right_comparison |>
  ggplot(aes(x = right_gpd, y = right_sknorm)) +
  geom_point() +
  geom_abline(linetype = "dashed", color = "red")

```



```

# no of gpd bigger than sknorm for left tail
length(which(right_gpd_ratio > right_sknorm_ratio))

```

```
## [1] 224
```

```
length(which(right_sknorm_ratio > right_gpd_ratio))
```

```
## [1] 106
```

```
# GoF comparison
gpd_gof <- gpd_param$p-value`
sknorm_gof <- sknorm_param$p-value`

# no of significance in terms of gof test for left/right tail.
length(which(gpd_gof[1:330] > sknorm_gof[1:330]))
```

```
## [1] 241
```

```
length(which(gpd_gof[331:660] > sknorm_gof[331:660]))
```

```
## [1] 295
```