GPD vs Skew Normal

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

# should work under low_moi folder
q <- 0.96
samples <- 5e5

# get the file path for GPD and skew normal parameters
subfld <- "figures/power_exploration"
subDir_gpd <- sprintf("%s/tail_prob_%d_resamples_%.2f_percentile", subfld, samples, q)
subDir_sknorm <- sprintf("%s/sknorm_tail_prob_%d_resamples_%.2f_percentile", subfld, samples, q)
param_gpd <- read_csv(sprintf("%s/param_twosides.csv", subDir_gpd))
param_sknorm <- read_csv(sprintf("%s/param_twosides.csv", subDir_sknorm))
gpd_param <- t(param_gpd[,-1])
sknorm_param <- t(param_sknorm[,-1])


# create the boxplot tibble
param <- tibble(method = c(rep("GPD", 660), rep("Sknorm", 660)),
                tail = rep(c(rep("left", 330), rep("right", 330)), 2),
                GoF_statistic = c(gpd_param[, 4], sknorm_param[, 4]),
                GoF_pvalue = c(gpd_param[, 5], sknorm_param[, 5]),
                fit_emp_ratio = c(gpd_param[, 6], sknorm_param[, 6]),
                emp_fit_ratio = c(gpd_param[, 7], sknorm_param[, 7]))

# create the dotplot tibble
param_comparison <- cbind(param[1:660, 2:6], param[661:1320, 3:6])
colnames(param_comparison)[2:9] <- c("GPD_statistic", "GPD_pvalue",
                                     "GPD_EMP_ratio", "EMP_GPD_ratio",
                                     "SKN_statistic", "SKN_pvalue",
                                     "SKN_EMP_ratio", "EMP_SKN_ratio")
```

1. fit_over_emp ratio plot

```r
dot_plot <- param_comparison |>
  filter(GPD_EMP_ratio < 100 & SKN_EMP_ratio < 100) |>
  ggplot(aes_string(x = "GPD_EMP_ratio", y = "SKN_EMP_ratio")) +
  facet_grid(tail ~.) +
  geom_point() +
  geom_abline(linetype = "dashed", color = "red") +
  scale_x_log10() +
  scale_y_log10() +
  labs(title = "fit/emp p-value ratio plot")
```

```
hist_plot <- param |>
  filter(fit_emp_ratio < 100) |>
  ggplot(aes_string(x = "fit_emp_ratio")) +
  facet_grid(tail ~ method) +
  scale_x_log10() +
  geom_histogram() +
  geom_vline(xintercept = 10, linetype = "dashed", color = "red")
plot_grid(dot_plot,
          hist_plot,
          ncol = 2,
          align = "v")
```
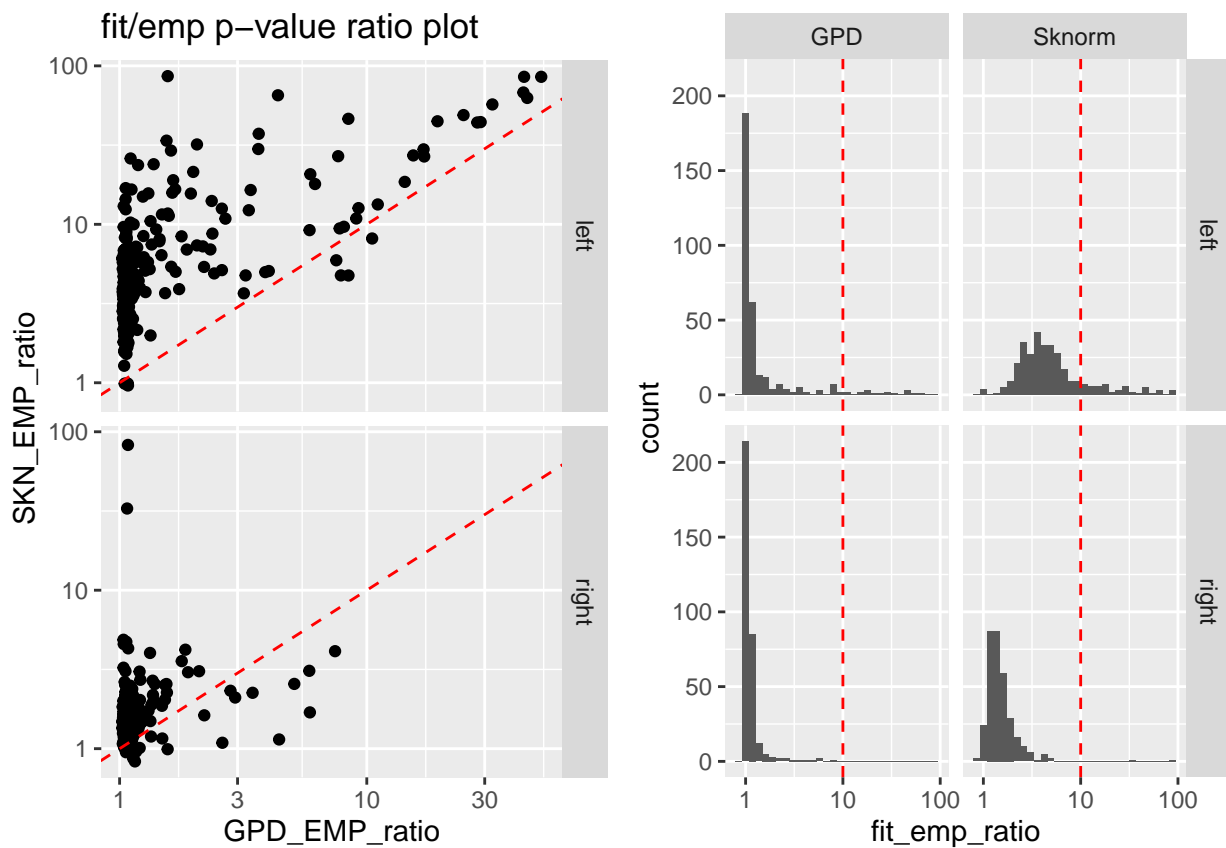

2. emp_over_fit ratio plot

```
dot_plot <- param_comparison |>
  filter(EMP_GPD_ratio < 100 & EMP_SKN_ratio < 100) |>
  ggplot(aes_string(x = "EMP_GPD_ratio", y = "EMP_SKN_ratio")) +
  facet_grid(tail ~.) +
  geom_point() +
  geom_abline(linetype = "dashed", color = "red") +
  scale_x_log10() +
  scale_y_log10() +
  labs(title = "emp/fit p-value ratio plot")

hist_plot <- param |>
```
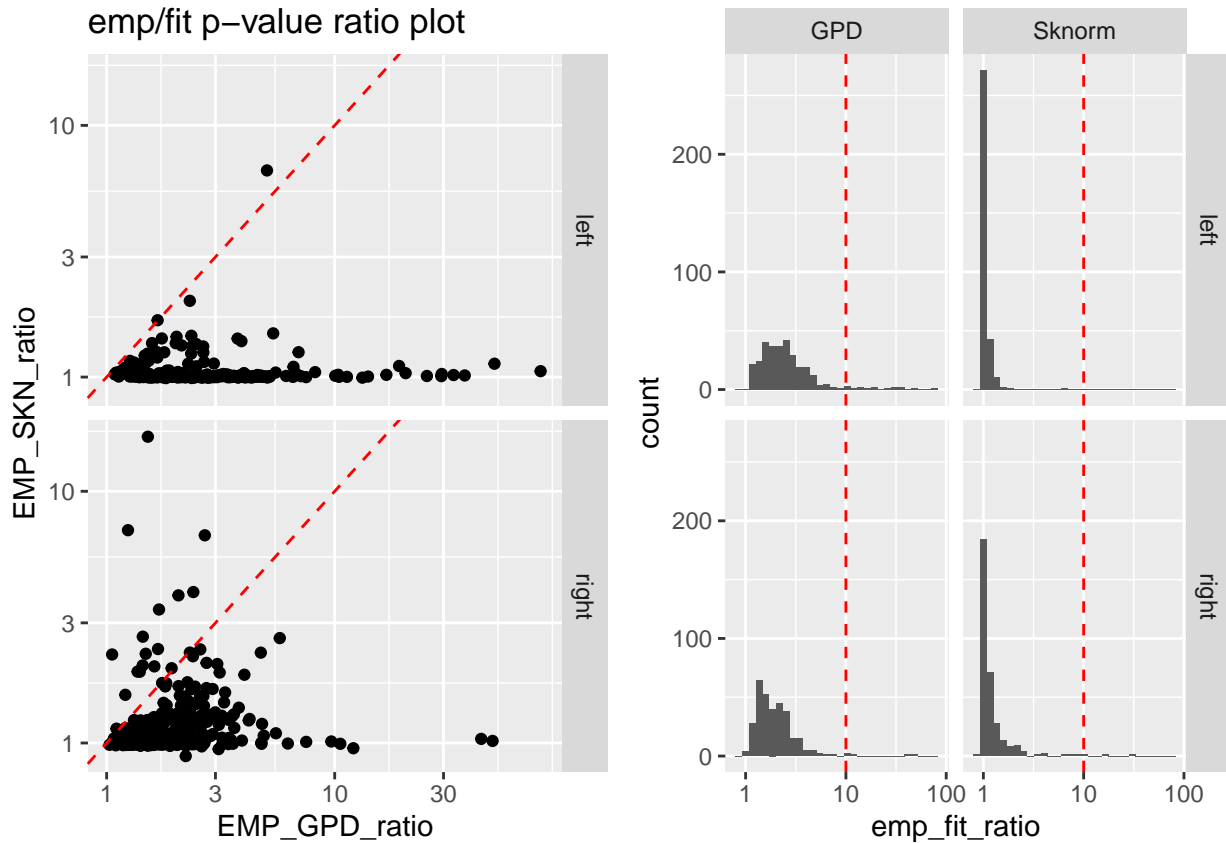
```
   filter(emp_fit_ratio < 100) |>
   ggplot(aes_string(x = "emp_fit_ratio")) +
   facet_grid(tail ~ method) +
   scale_x_log10() +
   geom_histogram() +
   geom_vline(xintercept = 10, linetype = "dashed", color = "red")
plot_grid(dot_plot,
          hist_plot,
          ncol = 2,
          align = "v")
```
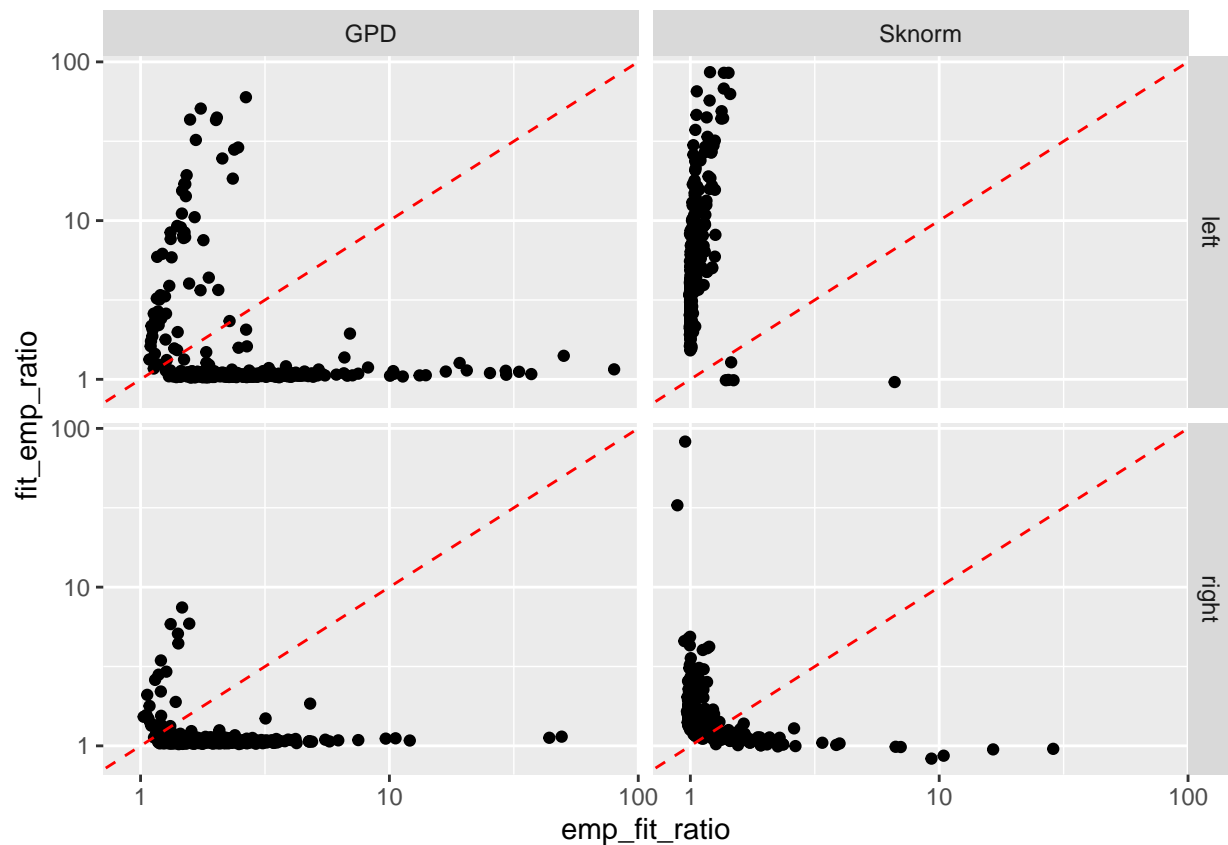

3. ratio compartion plot

```
dot_plot <- param |>
  filter(emp_fit_ratio < 100 & fit_emp_ratio < 100) |>
  ggplot(aes_string(x = "emp_fit_ratio", y = "fit_emp_ratio")) +
  facet_grid(tail ~ method) +
  scale_x_log10() +
  scale_y_log10() +
  geom_point() +
  geom_abline(linetype = "dashed", color = "red")
print(dot_plot)
```
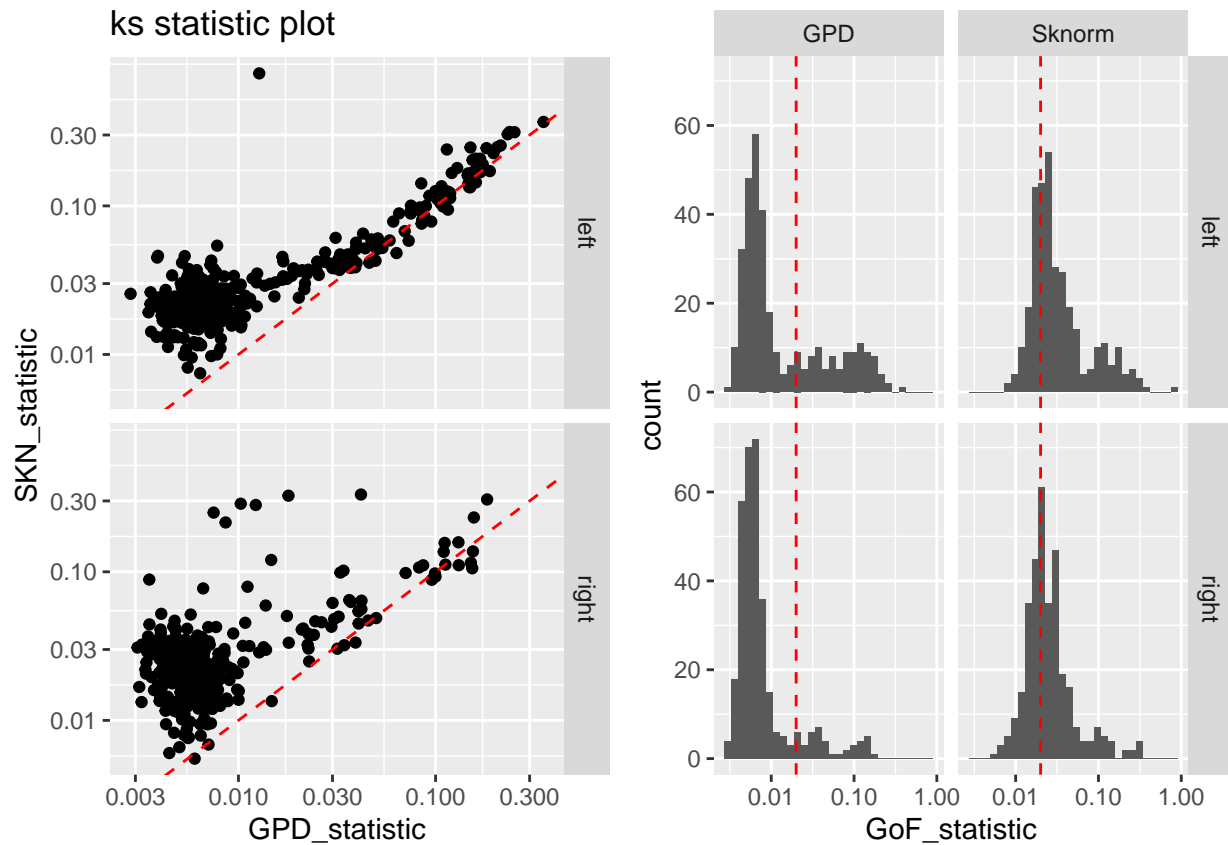
4. GoF statistic

```
dot_plot <- param_comparison |>
  ggplot(aes_string(x = "GPD_statistic", y = "SKN_statistic")) +
  facet_grid(tail ~.) +
  geom_point() +
  scale_x_log10() +
  scale_y_log10() +
  geom_abline(linetype = "dashed", color = "red") +
  labs(title = "ks statistic plot")

hist_plot <- param |>
  ggplot(aes_string(x = "GoF_statistic")) +
  facet_grid(tail ~ method) +
  scale_x_log10() +
  geom_histogram() +
  geom_vline(xintercept = 0.02, linetype = "dashed", color = "red")

plot_grid(dot_plot,
          hist_plot,
          ncol = 2,
          align = "v")
```

5. p-value of GoF test

```r
dot_plot <- param_comparison |>
  ggplot(aes_string(x = "GPD_pvalue", y = "SKN_pvalue")) +
  facet_grid(tail ~.) +
  geom_point() +
  scale_x_sqrt() +
  scale_y_sqrt() +
  geom_abline(linetype = "dashed", color = "red") +
  labs(title = "ks p-value plot")

hist_plot <- param |>
  ggplot(aes_string(x = "GoF_pvalue")) +
  facet_grid(tail ~ method) +
  geom_histogram() +
  scale_x_log10() +
  geom_vline(xintercept = 0.05, linetype = "dashed", color = "red")

plot_grid(dot_plot,
          hist_plot,
          ncol = 2,
          align = "v")
```

ks p−value plot
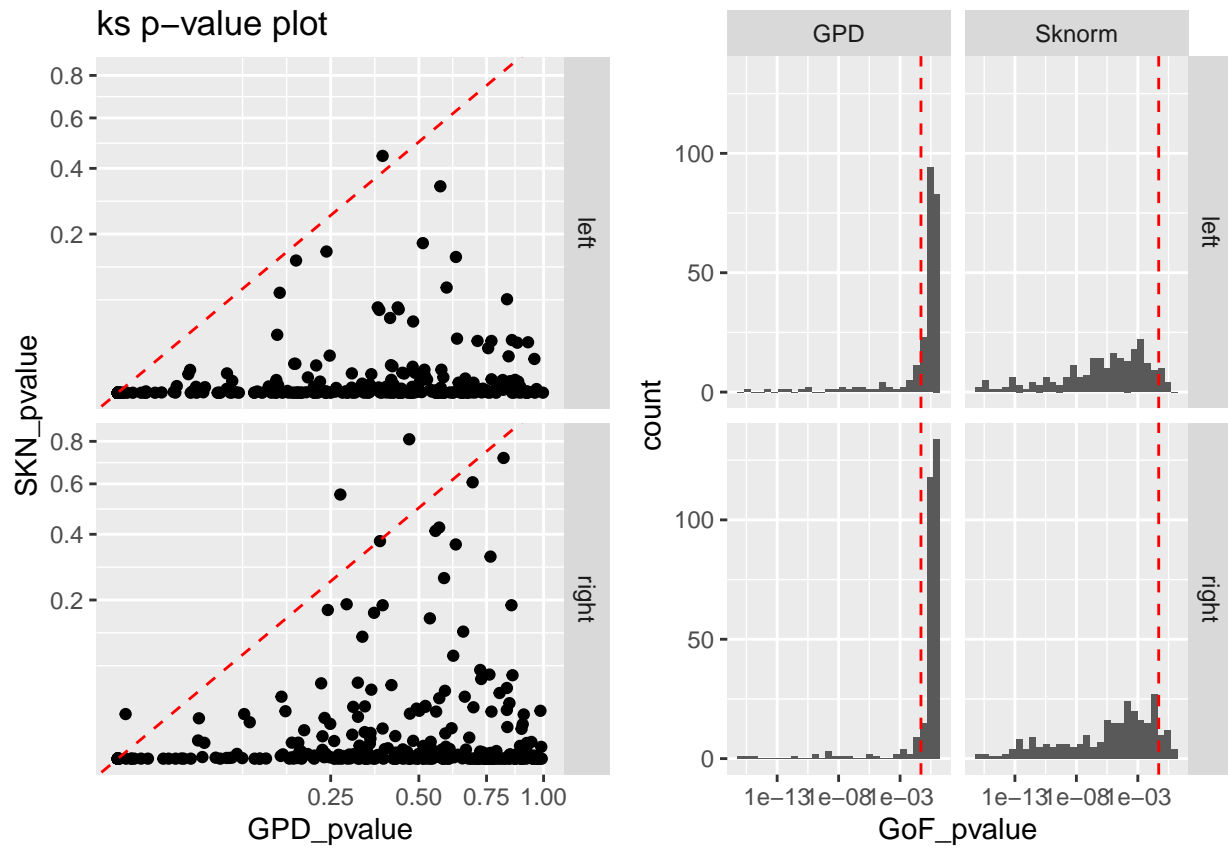
```
# check if the inf value corresponds to bumpy distribution
# bumpy: 56, 143, 169, 175, 177, 185, 191, 198, 213, 236, 241, 243
# nonbumpy: 327, 325, 324
# both bumpy: 191
```