

Simulation negative binomial distribution

Martin Wohlfender

2024-12-13

```
library(stats)
library(cowplot)

## Warning: Paket 'cowplot' wurde unter R Version 4.2.3 erstellt
library(dplyr)

## Warning: Paket 'dplyr' wurde unter R Version 4.2.3 erstellt
##
## Attache Paket: 'dplyr'
## Die folgenden Objekte sind maskiert von 'package:stats':
##
##     filter, lag
## Die folgenden Objekte sind maskiert von 'package:base':
##
##     intersect, setdiff, setequal, union
library(ggplot2)

## Warning: Paket 'ggplot2' wurde unter R Version 4.2.3 erstellt
library(tibble)

## Warning: Paket 'tibble' wurde unter R Version 4.2.3 erstellt
set.seed(314)

# effective reproduction number
R <- 1.2

# dispersion parameter
k <- 0.3

# mutation probability
mu <- 0.5

# number of samples
N <- 10000

samples_T_1 <- rnbinom(n = N, mu = R, size = k)
samples_V_1 <- unlist(lapply(X = samples_T_1, FUN = function(x) rbinom(n = 1, size = x, prob = 1-mu)))
samples_W_1 <- samples_T_1 - samples_V_1
samples_V_plus_W_1 <- samples_V_1 + samples_W_1

table_samples_T_1 <- tibble(x = sort(unique(samples_T_1)),
```

```

      freq_T_1 = as.numeric(table(samples_T_1)))
table_samples_V_1 <- tibble(x = sort(unique(samples_V_1)),
      freq_V_1 = as.numeric(table(samples_V_1)))
table_samples_W_1 <- tibble(x = sort(unique(samples_W_1)),
      freq_W_1 = as.numeric(table(samples_W_1)))
table_samples_V_plus_W_1 <- tibble(x = sort(unique(samples_V_plus_W_1)),
      freq_V_plus_W_1 = as.numeric(table(samples_V_plus_W_1)))

min_T_V_W_1 <- min(samples_T_1, samples_V_1, samples_W_1, samples_V_plus_W_1)
max_T_V_W_1 <- max(samples_T_1, samples_V_1, samples_W_1, samples_V_plus_W_1)

table_samples_T_V_W_1 <- tibble(x = min_T_V_W_1:max_T_V_W_1) %>%
  left_join(table_samples_T_1, by = "x") %>%
  left_join(table_samples_V_1, by = "x") %>%
  left_join(table_samples_W_1, by = "x") %>%
  left_join(table_samples_V_plus_W_1, by = "x") %>%
  replace(is.na(.), 0)

samples_V_2 <- rnbino(n = N, mu = (1-mu)*R, size = k)
samples_W_2 <- rnbino(n = N, mu = mu*R, size = k)
samples_V_plus_W_2 <- samples_V_2 + samples_W_2

table_samples_V_2 <- tibble(x = sort(unique(samples_V_2)),
      freq_V_2 = as.numeric(table(samples_V_2)))
table_samples_W_2 <- tibble(x = sort(unique(samples_W_2)),
      freq_W_2 = as.numeric(table(samples_W_2)))
table_samples_V_plus_W_2 <- tibble(x = sort(unique(samples_V_plus_W_2)),
      freq_V_plus_W_2 = as.numeric(table(samples_V_plus_W_2)))

min_T_V_W_2 <- min(samples_V_2, samples_W_2, samples_V_plus_W_2)
max_T_V_W_2 <- max(samples_V_2, samples_W_2, samples_V_plus_W_2)

table_samples_T_V_W_2 <- tibble(x = min_T_V_W_2:max_T_V_W_2) %>%
  left_join(table_samples_V_2, by = "x") %>%
  left_join(table_samples_W_2, by = "x") %>%
  left_join(table_samples_V_plus_W_2, by = "x") %>%
  replace(is.na(.), 0)

neg_bin_dist <- tibble(x = 0:max(max_T_V_W_1, max_T_V_W_2)) %>%
  mutate(dist_T = dnbino(x = x, mu = R, size = k),
    dist_V = dnbino(x = x, mu = (1-mu)*R, size = k),
    dist_W = dnbino(x = x, mu = mu*R, size = k))

plot_T_1 <- ggplot() +
  geom_col(data = table_samples_T_V_W_1, aes(x = x, y = freq_T_1 / N), fill = "navy") +
  geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_T), color = "firebrick") +
  scale_x_continuous(limits = c(-0.5, 15.5)) +
  scale_y_continuous(limits = c(0, 1)) +
  xlab(label = "T") +
  ylab(label = NULL) +
  theme_bw()

plot_V_1 <- ggplot() +
  geom_col(data = table_samples_T_V_W_1, aes(x = x, y = freq_V_1 / N), fill = "navy") +

```

```

geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_V), color = "firebrick") +
scale_x_continuous(limits = c(-0.5, 15.5)) +
scale_y_continuous(limits = c(0, 1)) +
xlab(label = "V") +
ylab(label = NULL) +

theme_bw()

plot_W_1 <- ggplot() +
geom_col(data = table_samples_T_V_W_1, aes(x = x, y = freq_W_1 / N), fill = "navy") +
geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_W), color = "firebrick") +
scale_x_continuous(limits = c(-0.5, 15.5)) +
scale_y_continuous(limits = c(0, 1)) +
xlab(label = "W") +
ylab(label = NULL) +
theme_bw()

plot_V_plus_W_1 <- ggplot() +
geom_col(data = table_samples_T_V_W_1, aes(x = x, y = freq_V_plus_W_1 / N), fill = "navy") +
geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_T), color = "firebrick") +
scale_x_continuous(limits = c(-0.5, 15.5)) +
scale_y_continuous(limits = c(0, 1)) +
xlab(label = "V+W") +
ylab(label = NULL) +
theme_bw()

plot_V_2 <- ggplot() +
geom_col(data = table_samples_T_V_W_2, aes(x = x, y = freq_V_2 / N), fill = "navy") +
geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_V), color = "firebrick") +
scale_x_continuous(limits = c(-0.5, 15.5)) +
scale_y_continuous(limits = c(0, 1)) +
xlab(label = "V") +
ylab(label = NULL) +
theme_bw()

plot_W_2 <- ggplot() +
geom_col(data = table_samples_T_V_W_2, aes(x = x, y = freq_W_2 / N), fill = "navy") +
geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_W), color = "firebrick") +
scale_x_continuous(limits = c(-0.5, 15.5)) +
scale_y_continuous(limits = c(0, 1)) +
xlab(label = "W") +
ylab(label = NULL) +
theme_bw()

plot_V_plus_W_2 <- ggplot() +
geom_col(data = table_samples_T_V_W_2, aes(x = x, y = freq_V_plus_W_2 / N), fill = "navy") +
geom_point(data = neg_bin_dist, mapping = aes(x = x, y = dist_T), color = "firebrick") +
scale_x_continuous(limits = c(-0.5, 15.5)) +
scale_y_continuous(limits = c(0, 1)) +
xlab(label = "V + W") +
ylab(label = NULL) +
theme_bw()

```

```

plot_grid_V_W_1 <- plot_grid(plot_V_1 + theme(plot.margin = margin(0, 5, 5, 20)),
                             plot_W_1 + theme(plot.margin = margin(0, 5, 5, 20)),
                             labels = c('B', 'C'), label_size = 12)

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_col()`).

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_col()`).

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

plot_grid_1 <- plot_grid(plot_T_1 + theme(plot.margin = margin(0, 5, 5, 20)),
                        plot_grid_V_W_1 ,
                        plot_V_plus_W_1 + theme(plot.margin = margin(0, 5, 5, 20)),
                        labels = c('A', '', 'D'), label_size = 12, nrow = 3)

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_col()`).
## Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_col()`).

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

plot_grid_V_W_2 <- plot_grid(plot_V_2 + theme(plot.margin = margin(0, 5, 5, 20)),
                             plot_W_2 + theme(plot.margin = margin(0, 5, 5, 20)),
                             labels = c('E', 'F'), label_size = 12)

## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_col()`).
## Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_col()`).

## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

plot_grid_2 <- plot_grid(plot_grid_V_W_2,
                        plot_V_plus_W_2 + theme(plot.margin = margin(0, 5, 5, 20)),
                        labels = c('', 'G'), label_size = 12, nrow = 2)

## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_col()`).
## Removed 17 rows containing missing values or values outside the scale range
## (`geom_point()`).

plot_grid <- plot_grid(plot_grid_1,
                      plot_grid_2,
                      labels = c('', ''), label_size = 12, ncol = 2)

```

```

ggsave(filename = "plots/simulation/varia/plot_grid_sim_neg_binom.png",
        plot = plot_grid,
        width = 7, height = 7, units = "in")

ggsave(filename = "plots/simulation/varia/plot_grid_sim_neg_binom.pdf",
        plot = plot_grid,
        width = 7, height = 7, units = "in")

```

plot_grid

