

RNA-seq course- week1

Serhiy Naumenko

2023-09-30

Contents

1 Overview	1
2 Simulated	4
3 Cumulative probability	4
4 sessionInfo()	5

1 Overview

- Negative binomial distribution
- Від'ємний біноміальний розподіл
- <https://rpubs.com/mpfoley73/458738>
- https://uk.wikipedia.org/wiki/Від'ємний_біноміальний_розподіл
- https://www.youtube.com/watch?v=IJw6Ku_jQkM

```
r <- 3 # n success
p <- 0.20 # success
n <- 7 - r # failures
# exact
dnbinom(x = n, size = r, prob = p)

## [1] 0.049152

# simulated
mean(rnbinom(n = 10000, size = r, prob = p) == n)

## [1] 0.0485
```

```

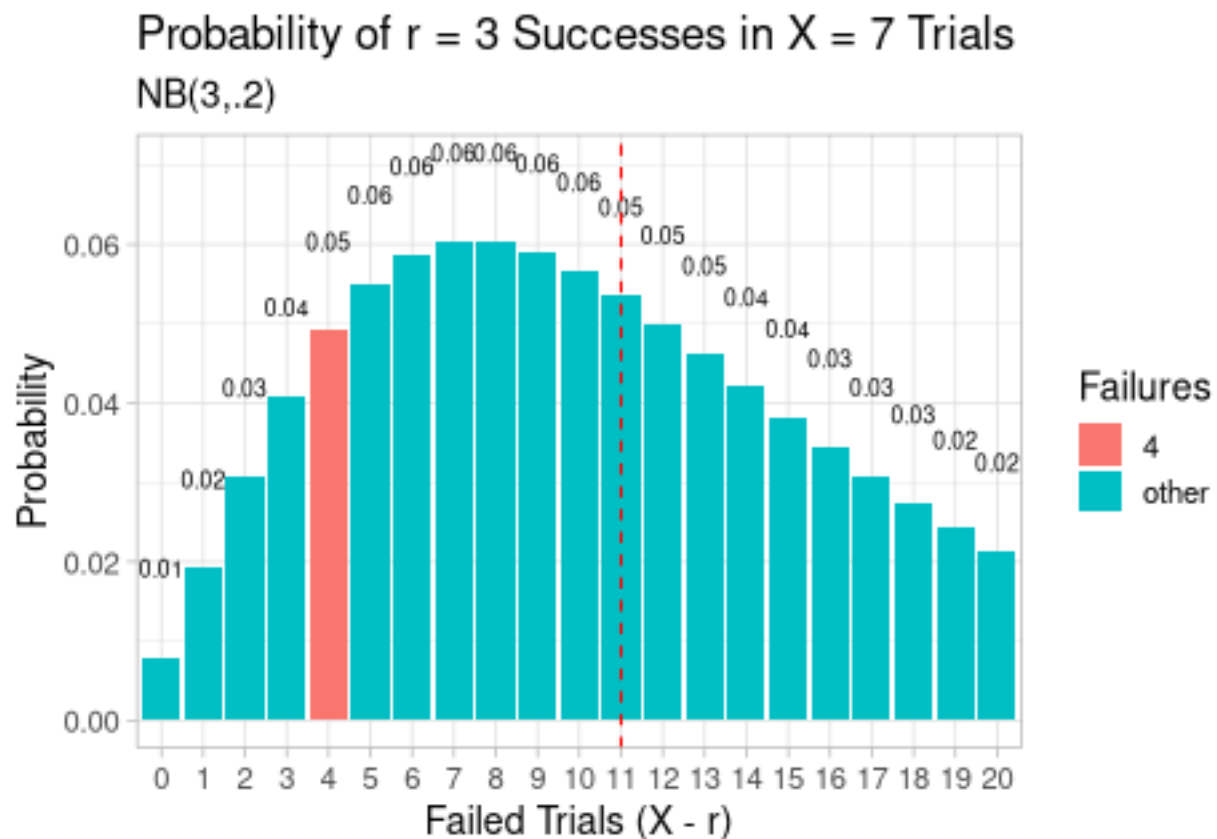
r <- 3
p <- 0.20

sim_nb <- data.frame(x = 0:20, prob = dnbinom(x = 0:20, size = r, prob = p)) %>%
  mutate(Failures = ifelse(x == n, n, "other"))

# NB: factor

sim_nb %>%
  ggplot(aes(x = factor(x), y = prob, fill = Failures)) +
    geom_col() +
    geom_text(
      aes(label = round(prob,2), y = prob + 0.01),
      position = position_dodge(0.9),
      size = 3,
      vjust = 0
    ) +
    labs(title = "Probability of r = 3 Successes in X = 7 Trials",
         subtitle = "NB(3,.2)",
         x = "Failed Trials (X - r)",
         y = "Probability") +
    geom_vline(xintercept = r*(1-p)/p, linetype = "dashed", color = "red")

```



Expected number of trials

```

r <- 3 # success
p <- 0.20

```

```
# mean
# exact
r / p
```

```
## [1] 15
• Mean = 12`
• Var = 60
```

2 Simulated

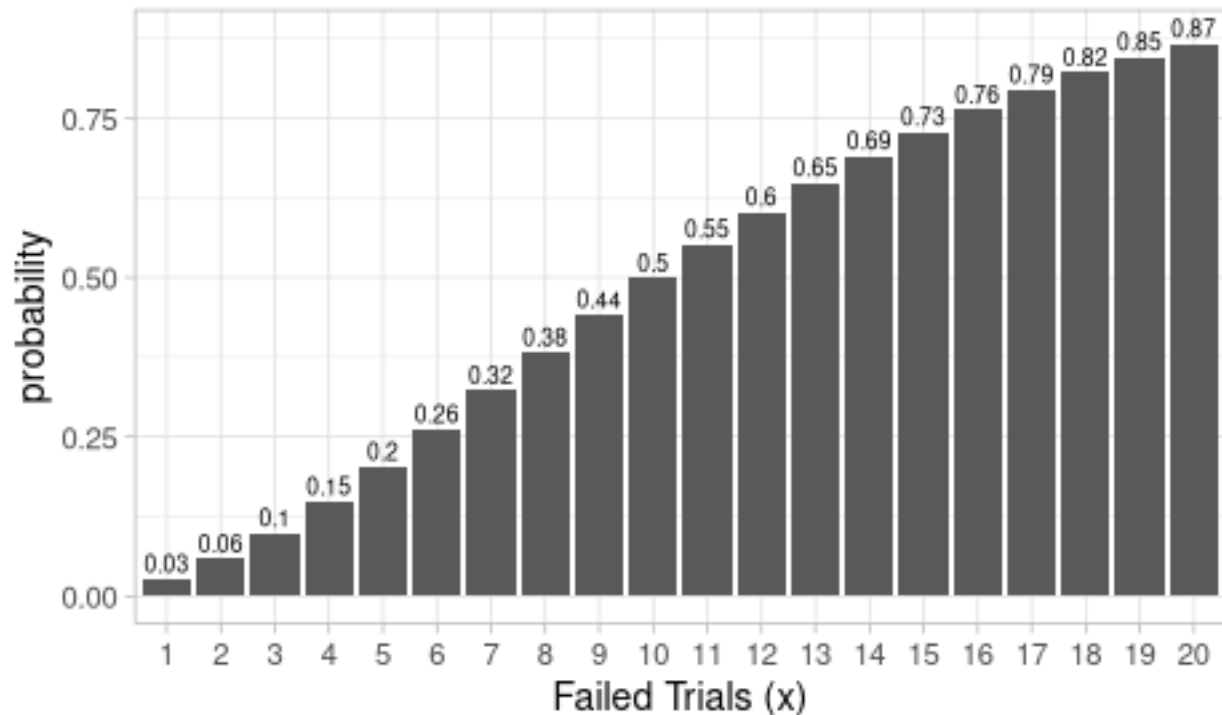
```
var(rnbinom(n = 100000, size = r, prob = p))
```

```
## [1] 59.03792
```

3 Cumulative probability

```
data.frame(x = 1:20,
            pmf = dnbinom(x = 1:20, size = r, prob = p),
            cdf = pnbinom(q = 1:20, size = r, prob = p, lower.tail = TRUE)) %>%
ggplot(aes(x = factor(x), y = cdf)) +
  geom_col() +
  geom_text(
    aes(label = round(cdf,2), y = cdf + 0.01),
    position = position_dodge(0.9),
    size = 3,
    vjust = 0
  ) +
  labs(title = "Cumulative Probability of X = x failed trials to achieve 3rd success",
        subtitle = "NB(3,.2)",
        x = "Failed Trials (x)",
        y = "probability")
```

Cumulative Probability of $X = x$ failed trials to achieve 3rd NB(3,.2)



4 sessionInfo()

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora Linux 37 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libflexiblas.so.3.3
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] knitr_1.41      mosaic_1.8.4.2  mosaicData_0.20.3 ggformula_0.10.4
##  [5] Matrix_1.6-1    lattice_0.20-45 forcats_0.5.2    stringr_1.5.0
```

```

## [9] dplyr_1.0.10      purrr_1.0.0      readr_2.1.3      tidyr_1.2.1
## [13] tibble_3.1.8      ggplot2_3.4.0    tidyverse_1.3.2
##
## loaded via a namespace (and not attached):
## [1] httr_1.4.4          jsonlite_1.8.4      modelr_0.1.10
## [4] assertthat_0.2.1    highr_0.10           googlesheets4_1.0.1
## [7] ggstance_0.3.6      cellranger_1.1.0     yaml_2.3.6
## [10] pillar_1.8.1        backports_1.4.1      glue_1.6.2
## [13] digest_0.6.31        polyclip_1.10-4      rvest_1.0.3
## [16] colorspace_2.0-3     htmltools_0.5.4      pkgconfig_2.0.3
## [19] broom_1.0.2          labelled_2.12.0      haven_2.5.1
## [22] scales_1.2.1         tweenr_2.0.2         tzdb_0.3.0
## [25] ggforce_0.4.1        timechange_0.1.1     googledrive_2.0.0
## [28] generics_0.1.3       farver_2.1.1         ellipsis_0.3.2
## [31] withr_2.5.0          cli_3.5.0            magrittr_2.0.3
## [34] crayon_1.5.2         readxl_1.4.1         evaluate_0.19
## [37] fs_1.5.2             fansi_1.0.3          MASS_7.3-58.1
## [40] xml2_1.3.3           tools_4.2.2          hms_1.1.2
## [43] gargle_1.2.0         lifecycle_1.0.3      munsell_0.5.0
## [46] reprex_2.0.2         compiler_4.2.2       rlang_1.0.6
## [49] grid_4.2.2           ggribes_0.5.4        rstudioapi_0.14
## [52] mosaicCore_0.9.2.1   labeling_0.4.2       rmarkdown_2.19
## [55] gtable_0.3.1         DBI_1.1.3            R6_2.5.1
## [58] lubridate_1.9.0      fastmap_1.1.0        utf8_1.2.2
## [61] stringi_1.7.8        Rcpp_1.0.9           vctrs_0.5.1
## [64] dbplyr_2.2.1         tidyselect_1.2.0     xfun_0.36

```