

RNA-seq course- week1

Serhiy Naumenko

2024-08-09

Contents

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

1 Overview

- Negative binomial distribution
- Від'ємний біноміальний розподіл
- <https://rpubs.com/mpfoley73/458738>
- [https://uk.wikipedia.org/wiki/Від\XeTeXglyph\numexpr\XeTeXcharglyph"0027\relax\{ }ємний_біноміальний_розподіл](https://uk.wikipedia.org/wiki/Від\XeTeXglyph\numexpr\XeTeXcharglyph)
- 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
```

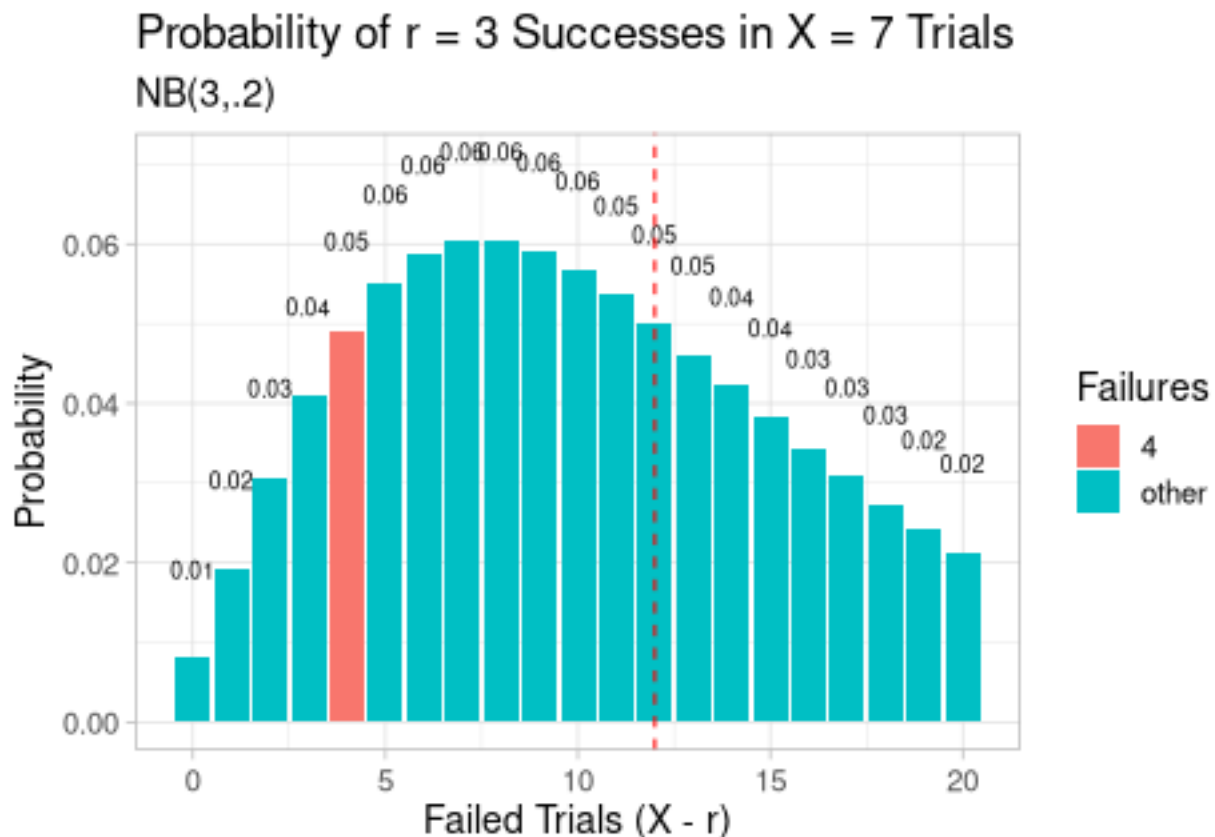
2 Кумулятивна ймовірність

```
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, "other"))

# NB: factor

sim_nb %>%
  ggplot(aes(x = 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

3 Simulated

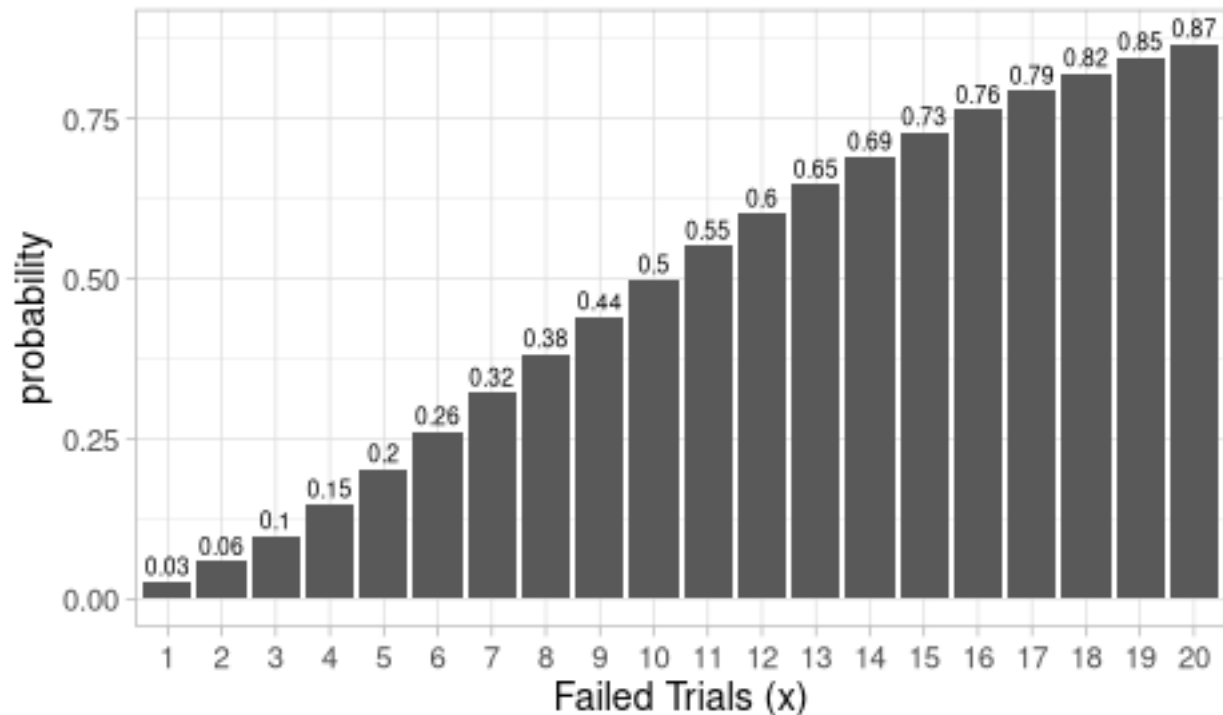
```
var(rnbinom(n = 100000, size = r, prob = p))
```

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

4 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)



5 sessionInfo()

```
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14)
## Platform: x86_64-redhat-linux-gnu
## Running under: Fedora Linux 40 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: FlexiBLAS OPENBLAS-OPENMP; LAPACK version 3.11.0
##
## 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
##
## time zone: America/Toronto
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
```

```

## other attached packages:
## [1] knitr_1.48          mosaic_1.9.1      mosaicData_0.20.4 ggformula_0.12.0
## [5] Matrix_1.7-0        lattice_0.22-6    lubridate_1.9.3   forcats_1.0.0
## [9] stringr_1.5.1       dplyr_1.1.4       purrr_1.0.2       readr_2.1.5
## [13] tidyr_1.3.1         tibble_3.2.1      ggplot2_3.5.1     tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] utf8_1.2.4          generics_0.1.3     stringi_1.8.4      hms_1.1.3
## [5] digest_0.6.36       magrittr_2.0.3     evaluate_0.24.0     grid_4.4.1
## [9] timechange_0.3.0    fastmap_1.2.0      fansi_1.0.6         scales_1.3.0
## [13] cli_3.6.3           labelled_2.13.0    rlang_1.1.4         munsell_0.5.1
## [17] withr_3.0.0         yaml_2.3.9         tools_4.4.1         tzdb_0.4.0
## [21] colorspace_2.1-0    mosaicCore_0.9.4.0 vctrs_0.6.5         R6_2.5.1
## [25] ggribges_0.5.6      lifecycle_1.0.4    MASS_7.3-60.2       pkgconfig_2.0.3
## [29] pillar_1.9.0        gtable_0.3.5       glue_1.7.0          highr_0.11
## [33] haven_2.5.4         xfun_0.45          tidyselect_1.2.1    rstudioapi_0.16.0
## [37] farver_2.1.2        htmltools_0.5.8.1 labeling_0.4.3      rmarkdown_2.27
## [41] compiler_4.4.1

```