Negative Binomial Distribution

Michael Foley

January 16, 2019

R Functions `dnbinom`, `pnbinom`, and `rnbinom`

Random varaible X is distributed $X \sim NB(r, p)$ with mean $\mu = \frac{r}{p}$ and variance $\sigma^2 = \frac{r(1-p)}{p^2}$ if X is the count of independent Bernoulli trials required to achieve the r^{th} successful trial when the probability of success is constant p. The probability of $X = n$ trials is $f(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$.

R function `dnbinom(x, size, prob)` is the probability of `x` *failures prior to the* `r` *th success* (note the difference) when the probability of success is `prob` . R function `pgeom(q, prob, lower.tail)` is the cumulative probability (`lower.tail = TRUE` for left tail, `lower.tail = FALSE` for right tail) of less than or equal to `q` failures prior to success. R function rgeom(n, size, prob) returns `n` random numbers from the geometric distribution `x~geom(prob)` . R function qgeom(p, prob, lower.tail) is the number of failures at the qth percentile (`lower.tail = TRUE`).

Example

An oil company has a p = 0.20 chance of striking oil when drilling a well. What is the probability the company drills x = 7 wells to strike oil r = 3 times?

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

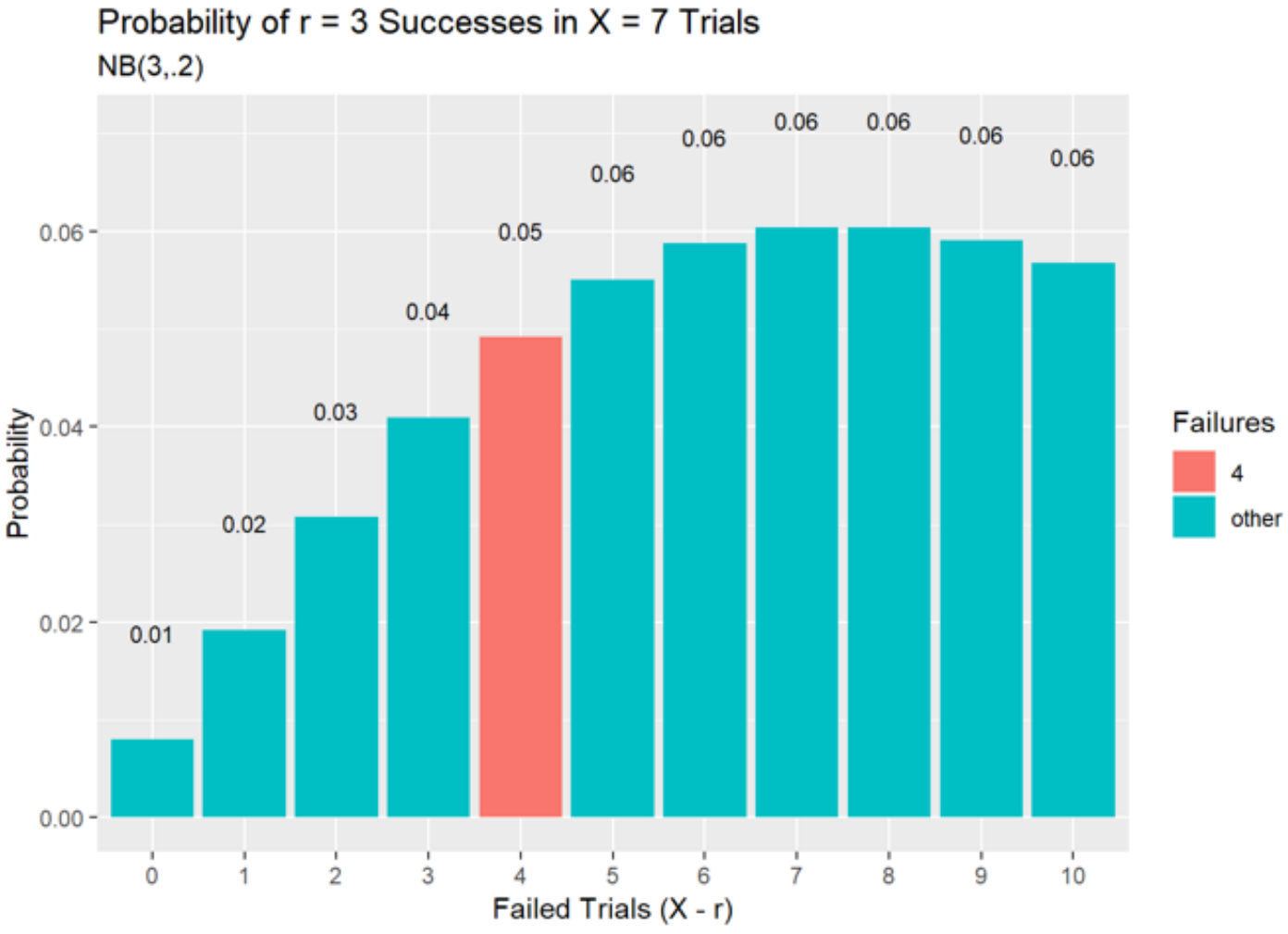
```
## [1] 0.049152
```

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

```
## [1] 0.0463
```

```
library(dplyr)
library(ggplot2)

data.frame(x = 0:10, prob = dnbinom(x = 0:10, size = r, prob = p)) %>%
  mutate(Failures = ifelse(x == n, n, "other")) %>%
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")
```



Example

What is the expected number of trials to achieve r = 3 successes when the probability of success is p = 0.2?

```
r = 3
p = 0.20
# mean
# exact
r / p
```

```
## [1] 15
```

```
# simulated
mean(rnbinom(n = 10000, size = 3, prob = p)) + r
```

```
## [1] 15.0766
```

```
# Variance
# exact
r * (1 - p) / p^2
```

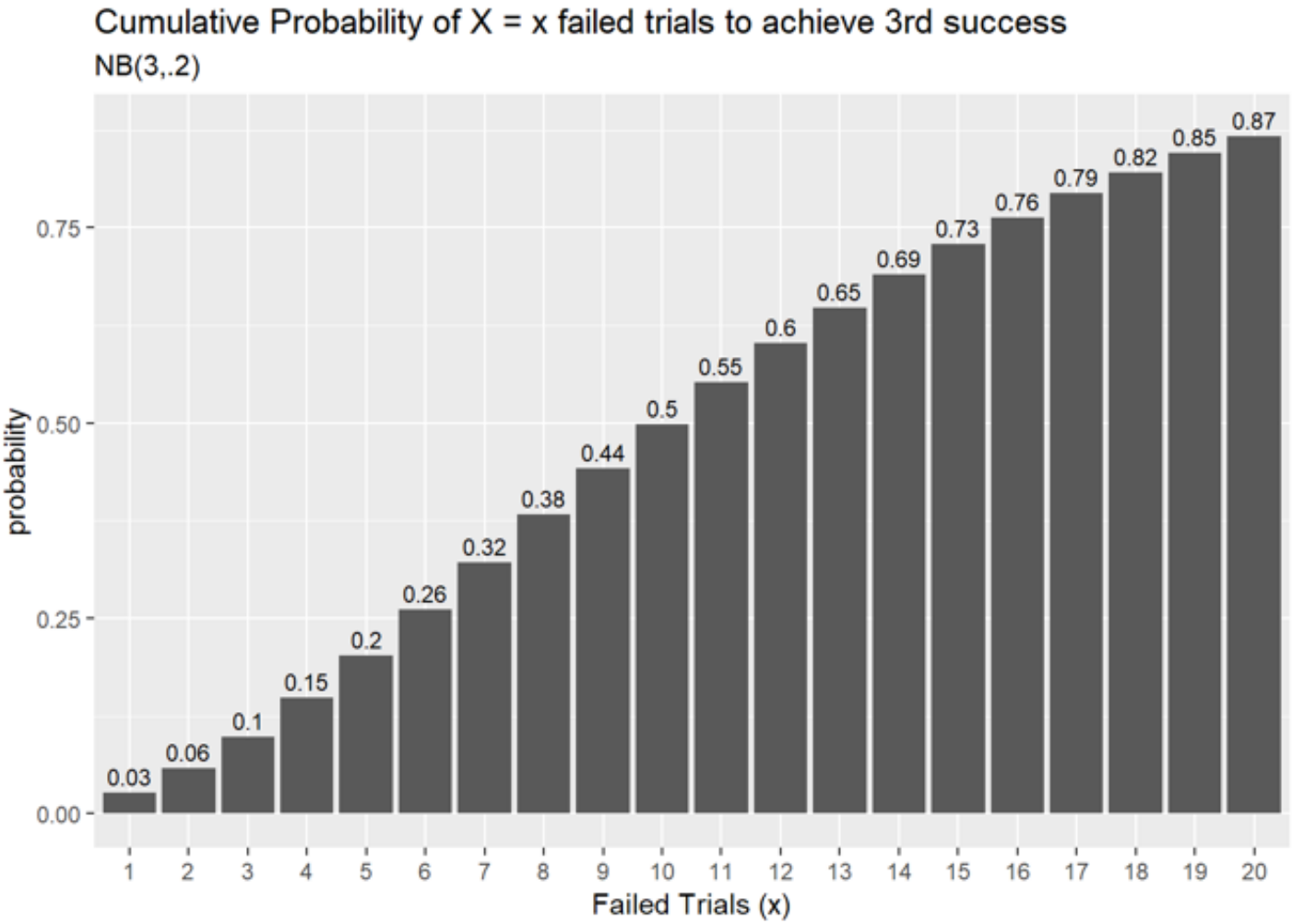
```
## [1] 60
```

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

```
## [1] 59.76289
```

```
library(dplyr)
library(ggplot2)

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 succes
s",
       subtitle = "NB(3,.2)",
       x = "Failed Trials (x)",
       y = "probability")
```


Cumulative Probability of X = x failed trials to achieve 3rd success
NB(3,.2)
probability
0.00
0.25
0.50
0.75
0.03
0.06
0.1
0.15
0.2
0.26
0.32
0.38
0.44
0.5
0.55
0.6
0.65
0.69
0.73
0.76
0.79
0.82
0.85
0.87
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Failed Trials (x)