

# R discrete distribution cheat sheet

## Overview

As you'll recall from Lab 04, R includes many built in functions for common distributions. Commands for each follow the same basic format. The abbreviated name of the distribution is preceded by one of four letters:

- “d” for the density function (probability distribution for discrete random variables)
- “p” for the cumulative distribution function (cdf)
- “q” for the quantile function (gives the x value for a given cumulative probability)
- “r” for random (generates a random number based on the distribution)

## Binomial distribution

The **binomial distribution** probability distribution and cdf in R:

- `dbinom(x, size, prob)` – returns  $p(x) = P(X = x)$ , (probability distribution) values for a binomial random variable  $X$ ,
- `pnbinom(x, size, prob)` – returns  $F(x) = P(X \leq x)$ , values for the binomial distribution (cdf)

Note that both have three arguments defined as:

- `x` = the value of the random variable for which a probability is computed
- `size` = number of trials (this is what we usually refer to as  $n$ )
- `prob` = probability of success on a trial (what we usually refer to as  $p$ )

Note that `x` can be a vector, meaning you can compute multiple probabilities with a single command!

## Geometric distribution

**NOTE: In R, the geometric distribution is parameterized as the number of FAILURES until the 1st success.** Whereas, in the book and lecture notes we consider the random variable  $Y$  to be the number of the trial on which the first success occurs. This means generally you will need to convert  $Y$  to the number of failures,  $X$ , as  $X = Y - 1$  before using the built-in R functions.

For example, if you want to find  $P(Y = 3)$  (1st success happens on 3rd trial), this is equivalent to  $P(X = 2)$  (2 failures before 1st success). This is calculated below, assuming a probability of success of 0.75

```
dgeom(2, 0.75)
```

```
## [1] 0.046875
```

We can confirm it by calculating in manually:

```
#P(Y = 3) = p(3) = (1 - p)^(3 - 1)*(p)
(1 - .75)^2*.75
```

```
## [1] 0.046875
```

`pgeom(x, prob)` – returns  $F(x) = P(X \leq x)$ , cumulative probabilities for the geometric distribution, again where  $X$  = number of failures before 1st success.

$P(Y \leq 3) = P(X \leq 2)$

```
pgeom(2, .75)
```

```
## [1] 0.984375
```

```
#check manually
#P(Y <= 3) = p(1) + p(2) + p(3) = sum((1 - p)^(y - 1)*(p)) for y = 1,2,3
y <- c(1,2,3)
p <- 0.75
p_y <- (1 - p)^(y-1)*p
p_y
```

```
## [1] 0.750000 0.187500 0.046875
```

```
sum(p_y)
```

```
## [1] 0.984375
```