Probability Distributions for Categorical Data (1.1-1.2):

- Binomial (Bernoulli)
- Multinomial
- Negative Binomial

Binomial

– Random variable X = # of successes in n trials

Scenario: Binary response – Yes/No → Code 1/0

Parameters: n = # of trials/people
π = probability of a "1" "success"

Assumptions:
1. Trials are independent
2. Probability of success stays constant
3. n fixed
4. Binary outcome

Notation: $X \sim Bin(n, \pi)$

$$P(X = x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{else} \end{cases}$$

name of distn.

R functions: dbinom(), pbinom(), rbinom(), qbinom()

density/mass $P(X = x)$ — cdf $P(X \le x)$ — random # generation — quantile – inverse of cdf

Mean: $E(X) = n\pi$ $Var(X) = n\pi(1-\pi)$

Multinomial – Generalizes binomial

↳ 2 possible outcomes

↳ c possible outcomes

$c = 2 \Rightarrow$ Binomial

Ex: Likert scale $c = 5$: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree

failure

success

Random vector :

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_c \end{pmatrix}$$

↳ one of these is redundant since

$c = 2 \quad \underline{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

$n = X_1 + \dots + X_n$

$\sum \pi_i = 1$

X_i = # of trials w/ i^{th} outcome

R: dmultinom($\underbrace{c(x_1, \dots, x_c)}_{\text{1 outcome you're interested in}}$, n, $c(\pi_1, \dots, \pi_c)$)

$P(X_1 = x_1, \dots, X_c = x_c)$

Negative Binomial – Y = # of failures until r^{th} success

$r = 1 \rightarrow$ Geometric

Alternative parameterization: # of trials until r^{th} success

R uses # failures: $P(Y = y) = \binom{y+r-1}{y} \pi^r (1-\pi)^y$

$\left[\binom{y+r-1}{r-1}\right] \quad y = 0, 1, \dots$

$\underbrace{\_ \; \_ \; \dots \; \_}_{y \text{ failures}} \; S$

r successes

$Y^* = $ # of trials

$\underbrace{y-r \text{ failures} + r \text{ successes}}$

$\underbrace{\_ \; \_ \; \_ \; \dots \; \_}_{y \text{ trials}} \; S$

$\Rightarrow P(Y^* = y) = \binom{y-1}{r-1} \pi^r (1-\pi)^{y-r}$

$\left[\binom{y-1}{y-r}\right] \quad y = r, r+1, \dots$

R: dnbinom( ) , pnbinom( ) , r--- , q---

y, r, π