Probability Theory for EOR

Some special discrete random variables II

Some **discrete random variables** are associated very special/ubiquitous **distributions (PMFs)**, they get their own names!

Definition

The **probability mass function (PMF)** of a **discrete** r.v., $X : S \mapsto \{x_1, x_2, \cdots\}$, is the function:

$$p_X(x) = P(X = x).$$

Negative Binomial: NBin(r, p)

- If we keep throwing a special coin (with probability p getting head in each flip) until we get r heads, let X denote the number of tails/bottomes/failures.
- $X \sim$NBin(r, p).
- The PMF of X is

$$P(X = n) = \binom{n + r - 1}{r - 1} p^r (1 - p)^n; n = 0, 1, \cdots .$$

- NBin$(1, p)$ is also called Geometric distribution Geom(p) (the number of bottoms before the first head).

Negative Binomial: NBin(r, p)

If $Z \sim$NBin(r, p), what is $E(Z), \mathrm{Var}(Z)$?

- Denote Y_1 the number of bottoms before the first head, Y_2 the number of bottoms after the first head and before the second head, ...
- Y_i follows Geom(p) (i.i.d.); $X = \sum_{i=1}^r Y_i$.
- Then $X \sim$NBin(r, p).
- $EZ = EX = \sum_{i=1}^r EY_i = r \times (1-p)/p$, since

$$EY_i = \sum_{k=0}^{\infty} k(1-p)^k p = p(1-p) \sum_{k=0}^{\infty} k(1-p)^{k-1} = (1-p)/p.$$

- $\mathrm{Var}(Z) = \mathrm{Var}(X) = \sum_{i=1}^r \mathrm{Var}(Y_i) = r \times (1-p)/p^2$, since

$$EY_i^2 = \sum_{k=0}^{\infty} k^2(1-p)^k p = (1-p)(2-p)/p^2.$$

$$\mathrm{Var}(Y_i) = EY_i^2 - (EY_i)^2 = (1-p)/q^2.$$

Negative Hypergeometric: NHGeom(w, b, r)

- Keep marking balls with $*$'s from a urn of w white balls and b **black** balls (each ball is equally likely to be marked) until there are r white balls with marks, and let X denote the number of balls being both **black** and marked with $*$'s.
- $X \sim$NHGeom(w, b, r).
- The PMF of X is

$$P(X = k) = \frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{r+k-1}} \frac{w - r + 1}{w + b - r - k + 1}; k = 0, 1, \cdots$$

Negative Hypergeometric: NHGeom(w, b, r)

If $Z \sim$NHGeom(w, b, r), what is $E(Z)$?

- The underlying sample space contain outcomes such as

$$bbbwwbw \cdots bb$$

which orders all balls.

For each black ball, it can choose $w + 1$ positions with equal likeliness: before the first white ball, between 1st and 2nd white balls, between 2nd and 3rd white balls, $\cdots$, between $(w - 1)$th and wth white balls, and after wth white balls.
- Let's label all black balls from 1 to b. Denote I_j the indicator random varaible whihc equals one if the black ball j chooses any positions before the rth white ball otherwise zero. $EI_j = r/(w + 1)$.
- Let $X = \sum_{i=1}^{b} I_i$, and thus $X \sim$NHGeom(w, b, r).

$$E(Z) = E(X) = \sum_{i=1}^{b} EI_i = br/(w + 1).$$