Lecture 05: Discrete Distributions

Mathematical Statistics I, MATH 60061/70061

Tuesday September 14, 2021

Reference: Casella & Berger, 3.1-3.2

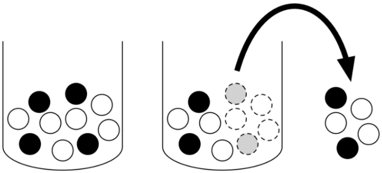
Drawing balls

Consider an urn with w white balls and b black balls, and draw n balls out of the urn:

- When sampling *with replacement*, the # of white balls follows a *Binomial distribution*

$$X \sim \mathrm{Bin}(n, w/(w+b))$$

- When sampling *without replacement*, the # of white balls follows a *Hypergeometric distribution*



Structure of the Hypergeometric story

Items in a population are classified using two sets of *tags*:

- Each ball is either white or black (the first set)
- Each ball is either sampled or not sampled (the second set)

At least one of these sets of tags is assigned completely at random.

Then, the number of twice-tagged items (e.g., balls are both white and sampled) follows a Hypergeometric distribution

$$X \sim \mathrm{HGeom}(w, b, n).$$

Identical Hypergeometric distributions

If $X \sim \mathrm{HGeom}(w, b, n)$ and $Y \sim \mathrm{HGeom}(n, w + b - n, w)$, then X and Y have the same distribution.

Identical Hypergeometric distributions

If $X \sim \mathrm{HGeom}(w, b, n)$ and $Y \sim \mathrm{HGeom}(n, w + b - n, w)$, then X and Y have the same distribution.

The story of the Hypergeometric with w white and b black balls, and a sample of size n made without replacement:

- $X \sim \mathrm{HGeom}(w, b, n)$: the # of white balls in the sample
  - First set of tags: white/black
  - Second set of tags: sampled/not sampled

Identical Hypergeometric distributions

If $X \sim \mathrm{HGeom}(w, b, n)$ and $Y \sim \mathrm{HGeom}(n, w + b - n, w)$, then X and Y have the same distribution.

The story of the Hypergeometric with w white and b black balls, and a sample of size n made without replacement:

- $X \sim \mathrm{HGeom}(w, b, n)$: the # of white balls in the sample
  - First set of tags: white/black
  - Second set of tags: sampled/not sampled
- $Y \sim \mathrm{HGeom}(n, w + b - n, w)$: the # of sampled balls among the white balls
  - First set of tags: sampled/not sampled
  - Second set of tags: white/black

Identical Hypergeometric distributions

If $X \sim \mathrm{HGeom}(w, b, n)$ and $Y \sim \mathrm{HGeom}(n, w+b-n, w)$, then X and Y have the same distribution.

The story of the Hypergeometric with w white and b black balls, and a sample of size n made without replacement:

- $X \sim \mathrm{HGeom}(w, b, n)$: the # of white balls in the sample
  - First set of tags: white/black
  - Second set of tags: sampled/not sampled
- $Y \sim \mathrm{HGeom}(n, w+b-n, w)$: the # of sampled balls among the white balls
  - First set of tags: sampled/not sampled
  - Second set of tags: white/black

Both X and Y count the # of white sampled balls, so they have the same distribution.

Binomial vs. Hypergeometric

Consider an urn with w white balls and b black balls. Consider also drawing n balls as performing n Bernoulli trials:

- When sampling *with replacement*, the # of white balls follows a *Binomial distribution*

  $$X \sim \mathrm{Bin}(n, w/(w+b)).$$

  The Bernoulli trials involved are *independent*.

- When sampling *without replacement*, the # of white balls follows a *Hypergeometric distribution*

  $$X \sim \mathrm{HGeom}(w, b, n).$$

  The Bernoulli trials involved are *dependent*.

What if $N = w + b \to \infty$?

The Binomial is a limiting case of the Hypergeometric

If $X \sim \mathrm{HGeom}(w, b, n)$ and $N = w + b \to \infty$ such that $p = \frac{w}{w+b}$ remains fixed, the PMF of X converges to the $\mathrm{Bin}(n, p)$ PMF.

The Binomial is a limiting case of the Hypergeometric

If $X \sim \text{HGeom}(w, b, n)$ and $N = w + b \to \infty$ such that $p = \frac{w}{w+b}$ remains fixed, the PMF of X converges to the $\text{Bin}(n, p)$ PMF.

$$\begin{aligned}
P(X = x) &= \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} \\
&= \frac{\binom{n}{x}\binom{w+b-n}{w-x}}{\binom{w+b}{w}} \quad [\text{HGeom}(w, b, n) = \text{HGeom}(n, w+b-n, w)] \\
&= \binom{n}{x} \frac{w!}{(w-x)!} \frac{b!}{(b-n+x)!} \frac{(w+b-n)!}{(w+b)!} \\
&= \binom{n}{x} \frac{w(w-1)\dots(w-x+1)b(b-1)\dots(b-n+x+1)}{(w+b)(w+b-1)\dots(w+b-n+1)} \\
&= \binom{n}{x} \frac{p(p-\frac{1}{N})\dots(p-\frac{x-1}{N})q(q-\frac{1}{N})\dots(q-\frac{n-x-1}{N})}{(1-\frac{1}{N})(1-\frac{2}{N})\dots(1-\frac{n-1}{N})}
\end{aligned}$$

As $N \to \infty$, $P(X = x) \to \binom{n}{x} p^x q^{n-x}$.