Hypegeometric Variance

Suppose for simplicity that $n \geq 2$, $b \geq n$ and $w \geq n$ are all integers [we consider the general case where only $n \leq w+b$ in the Tutorial]. then if:

$$P(X=x) = \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}}, \text{ for } x = 0,1,2,\ldots,n,$$

we have seen already that $E(X) = \frac{nw}{(w+b)}$. So,

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)P(X=x) \quad \Big[= 0(-1)P(X=0) + 1.0.P(X=1) + 2.1\,P(X=2) + \ldots + n(n-1)P(X=n)\Big]$$

$$= \sum_{x=2}^{n} \frac{\cancel{x(x-1)}\,.\,\frac{w(w-1)(w-2)\ldots(w-x+1)}{x\cancel{(x-1)}(x-2)\cdots(1)}\binom{b}{n-x}}{\frac{(w+b)(w+b-1)(w+b-2)\cdots(w+b-n+1)}{n\ (n-1)\ (n-2)\ \ldots\ 1.}}$$

$$= \frac{w(w-1)\,n(n-1)}{(w+b)(w+b-1)} \sum_{x=2}^{n} \frac{\binom{w-2}{x-2}\binom{b}{(n-2)-(x-2)}}{\binom{w+b-2}{n-2}}$$

write $y = x-2$

$$= \frac{w(w-1)\,n(n-1)}{(w+b)(w+b-1)} \underbrace{\sum_{y=0}^{n-2} \frac{\binom{w-2}{y}\binom{b}{(n-2)-y}}{\binom{w+b-2}{n-2}}}$$

Sum of all Hypegeom. probs. when sampling without repl. n-2 times from

$\boxed{\begin{matrix} w-2 \text{ Wh.} \\ b \text{ Bl.} \end{matrix}}$ n-2 times

So then, $\mathrm{Var}(X) = \underbrace{E(X^2)}_{} - [E(X)]^2$

$$= E[X(X-1)] + E(X) - [E(X)]^2$$

$$= \frac{w(w-1)n(n-1)}{(w+b)(w+b-1)} + \frac{nw}{w+b} - \frac{n^2w^2}{(w+b)^2}$$

$$= [\ldots \text{Tutorial Exercise} \ldots]$$

$$= n\left(\frac{w}{w+b}\right)\left(\frac{b}{w+b}\right)\left(\frac{w+b-n}{w+b-1}\right)$$

If we write $p = \frac{w}{w+b}$ for the proportion of White AND $N = w+b$ for the "population size", this becomes

$$\mathrm{Var}(X) = \underbrace{np(1-p)}_{\text{Binomial / with replacement variance}} \left(\frac{N-n}{N-1}\right) \Big\} < 1 \text{ if } n \geq 2,$$

"Finite Population Correction Factor"

Thus the variance is SMALLER for sampling without replacement, compared to sampling with replacement.

<u>Beta-binomial Variance</u> Suppose that for integers $b \geq 1$, $w \geq 1$ and $n \geq 2$,

$$P(X=x) = \frac{\binom{w+x-1}{x}\binom{b+n-x-1}{n-x}}{\binom{w+b+n-1}{n}}, \quad \text{for } x = 0, 1, 2, \ldots, n$$

We have already seen $E(X) = \frac{nw}{w+b}$. ~~Thus~~ We have also then that

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\, P(X=x)$$

$$= \sum_{x=2}^{n} \cancel{x(x-1)} \frac{\frac{w(w+1)(w+2)\cdots(w+x-1)}{\cancel{x(x-1)}(x-2)\cdots 1} \cdot \binom{b+n-x-1}{n-x}}{\frac{(w+b)(w+b+1)(w+b+2)\cdots(w+b+n-1)}{n\,(n-1)\,(n-2)\cdots 1}}$$

$$= \frac{n(n-1)w(w+1)}{(w+b)(w+b+1)} \sum_{x=2}^{n} \frac{\binom{w+x-1}{x-2}\binom{b+(n-2)-(x-2)-1}{(n-2)-(x-2)}}{\binom{\cancel{w+b+n}\ w+b+n-1}{n-2}}$$

$y = x-2$

$$= \frac{n(n-1)w(w+1)}{(w+b)(w+b+1)} \underbrace{\sum_{y=0}^{n-2} \frac{\binom{w+2+y-1}{y}\binom{b+(n-2)-y-1}{(n-2)-y}}{\binom{w+2+b+(n-2)-1}{n-2}}}_{=1}$$

$=1$, its the sum of all beta-binom probs, drawing $n-2$ times using Polya Scheme from $\boxed{\begin{matrix} w+2 \text{ Wh} \\ b \text{ Bl} \end{matrix}}$ → $n-2$ times

Thus

$$Var(X) = E[X(X-1)] + E(X) - [E(X)]^2$$

$$= \frac{n(n-1)w(w+1)}{(w+b)(w+b+1)} + \frac{nw}{(w+b)} - \frac{n^2w^2}{(w+b)^2} = [\cdots \text{Tutorial Exercise}]$$

$$= n\underbrace{\left(\frac{w}{w+b}\right)\left(\frac{b}{w+b}\right)}_{\text{corresponding Binomial variance}}\underbrace{\frac{(w+b+n)}{(w+b+1)}}_{>1 \text{ for } n \geq 2}.$$