Approximation of Some Hypergeometric Distributions by Binomials

Assume that Y has a hypergeometric distribution, $Y \sim p(y) = \dfrac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$, where $\left\{ \begin{matrix} n << r \\ \text{and} \\ n << N-r \end{matrix} \right\}$.

So we are sampling from a lot of r good and $(N - r)$ bad items, and the number, n, being sampled is much smaller than both the number of good items (r) and the number of bad items ($N - r$) in the lot. Since $0 \le y \le n$ and $0 \le n - y \le n$, both y and $(n - y)$ are also much smaller than r and $N - r$, respectively. Then,

$$p(y) \;=\; \frac{r!}{y!\,(r-y)!} \cdot \frac{(N-r)!}{(n-y)!\,[(N-r)-(n-y)]!} \Bigg/ \frac{N!}{n!\,(N-n)!}$$

$$= \frac{n!}{y!\,(n-y)!} \;\cdot\; \frac{r!}{(r-y)!} \;\cdot\; \frac{(N-r)!}{[(N-r)-(n-y)]!} \Bigg/ \frac{N!}{(N-n)!}$$

$$= \binom{n}{y} \;\cdot\; \frac{r\cdot(r-1)\cdots[r-(y-1)] \;\cdot\; (N-r)\cdot[(N-r)-1]\cdots[(N-r)-(n-y)+1]}{N\cdot(N-1)\cdot[N-(y-1)] \;\cdot\; (N-y)\cdot[(N-y)-1]\cdots[(N-y)-(n-y)+1]}$$

$$= \binom{n}{y} \bullet \underbrace{\left(\frac{r}{N}\right)\cdot\left(\frac{r-1}{N-1}\right)\cdots\left(\frac{r-(y-1)}{N-(y-1)}\right)} \bullet \underbrace{\left(\frac{N-r}{N-y}\right)\cdot\left(\frac{(N-r)-1}{(N-y)-1}\right)\cdots\left(\frac{(N-r)-(n-y)+1}{(N-y)-(n-y)+1}\right)}$$

each of these y fractions is $\approx \dfrac{r}{N} = p$ if $y \le n << r < N$

each of these $(n-y)$ fractions is $\approx \dfrac{N-r}{N} = 1-p$ if $y \le n << N-r < N$

$$\approx \binom{n}{y} \;\bullet\; p^{\mathrm{y}} \;\bullet\; (1-p)^{n-y}.$$

But this last is just the probability function of the binomial distribution with n trials and probability of success $p = \dfrac{r}{N}$ = the proportion of good items originally in the lot.

© John J Currano, 10/08/2004