この壺の中から大きさ r の非復元無作為標本を抽出するとき，標本の中に入っている赤球の個数を x とし，そのときの確率変数を X とする。」

このとき，

$P_r(X=x)=\dfrac{{}_mC_x \cdot {}_nC_{r-x}}{{}_{m+n}C_r}$ であって，

関心の結果空間は

$\{x \mid 0 \leqq x \leqq Min(r, m)$ かつ x は整数$\}$ である。

このようにして定義される確率 model を超幾何分布という。

確率変数 X の平均は，$E(X)=rm/(m+n)$ で，X の分散は，

$V(X)=rmn(m+n-r)/(m+n)^2(m+n-1)$

である。

(2)　最尤推定値

壺の中に入っている球の総数は $t=m+n$ 未知数であるが，赤球の数 m は既知数であるとして，t を標本調査の結果によって定めてみよう。

そこで，$f_x(t)=P_r(X=x)$ とおき，x を固定し，t を動かして $f_x(t)$ を最大とする t の値を求める。

$f_x(t)/f_x(t-1)=(t-m)(t-r)/(t-m-r+x)t$

を調べると，$tx<mr$ のとき，$f_x(t)>f_x(t-1)$ で，$tx>mr$ のとき，$f_x(t)<f_x(t-1)$ となる。

これは，t の増加にともなって系列 $f_x(t)$ は最初増加し，その後減少することを意味する。

したがって，$t=\left[\dfrac{mr}{x}\right]$ のとき $fx(t)$ は最大となる。

i. e. 大きさ r の標本を抽出して標本調査を行った結果，赤球が x 個出現したとすれば，このような結果を最も大きい確率によってもたらす母集団の大きさ t は，$t=\left[\dfrac{mr}{x}\right]$ である。

そこで，$\left[\dfrac{mr}{x}\right]$ によって確定した t の値を $\hat{t}$ で表わし，これを t の最尤推定値という。

(3)　例題

再捕獲の資料から動物の集団の数を推定する問題を考えてみよう。

「湖で獲った1000尾の魚に赤印をつけて放してやるとする。しばらくの期間をおいて，新たに1000尾を捕獲したところ，その中に赤印のものが 100尾いた。この湖の魚の数についてどのような結論が得られるか。」

当然，2回目の魚の捕獲は湖の中のすべての魚からなる母集団からの random な標本と考え，2回目の捕獲までの期間中の魚の数は変わらないものとして，問題を一般化する。

湖の中の魚の数を t（未知数），最初に捕獲した魚の数を m，2回目に捕獲した魚の数を r，2回目に捕獲したときの赤印の魚の数を x とすると，2回目の捕獲にちょうど x 尾の赤印の魚が含まれる確率は $f_x(t)$ となる。

したがって，超幾何分布を適用すれば，t の最尤推定値は $t=\left[\dfrac{1000\times1000}{100}\right]=10000$ となる。

しかし，実際に湖に棲息している魚の総数はちょうど10000尾であるとは限らない。

そこで，t がその中に入っていると合理的に期待し得る範囲を求めるのが普通である。

確率の計算に，次の中心極限定理を応用する。

「$X_1, X_2, \cdot\cdot\cdot, X_n$ が独立な確率変数であるとき，$Y_n=X_1+\cdot\cdot\cdot+X_n$ とおくと

$$P_r\left(\alpha\leqq\frac{Y_n-E(Y_n)}{\sqrt{V(Y_n)}}\leqq\beta\right)\to\frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-\frac{x^2}{2}}dx$$

である。」

ここでは，標本を1尾ずつ抽出するものとして，第 i 回目の抽出によって赤印が選ばれたとき，$X_i=1$，そうでないとき $X_i=0$ とすると，t や m が大きいから，確率変数 $X_1, X_2, \cdot\cdot\cdot, X_n$ を独立であると考えてよいであろう。

$t=8500$, $m=r=1000$ とおき，2回目の標本に赤印の魚が100尾以下である確率を求めると，

$$P_r\left(\frac{Y_{1000}-E(Y_{1000})}{\sqrt{V(Y_{1000})}}\leqq\beta\right)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\beta e^{-\frac{x^2}{2}}dx$$

$E(Y_{1000})=117.647$, $\sqrt{V(Y_{1000})}=9.569$

$E(Y_{1000})+\beta\sqrt{V(Y_{1000})}=100$ より，$\beta=-1.84$

$\therefore\ P_r(Y_{1000}\leqq100)=0.033$

次に，$t=12000$, $m=r=1000$ とおき，2回目の標本に赤印の魚が 100尾以上である確率を求めると，$E(Y_{1000})=83.333$, $\sqrt{V(Y_{1000})}=8.370$ より，

$\beta=1.991$　　$\therefore P_r(Y_{1000}\geqq100)=0.024$

このことは，魚の数 t は $8500<t<12000$ であることを正当化するものであり，そして，この結論に対する信頼度は約94%であることを示す。

4.　おわりに

ここまでの考察から，現象を数学的な考察・処理にのせていくときの教材展開についての留意点を plot すると，次のようになるであろう。

① 現象を理想化する意味について考えさせ，理想化によって現象の記述が可能になることを理解させるようにする。

② 標本点の重さの与え方を，実験によって一般的な方向づけをし，理論的規約として，その重さを導入することの意味が理解できるようにする。

③ 関心の結果空間に確率測度を誘導して，確率 model を求めさせ，その求めた確率 model が考えている偶然変動をうまく記述しているかどうかを考えられるようにする。

④ 確率 model を用いて，一般的傾向を推測し，その推測結果の信頼の程度を求めることができることも知らせるようにする。

次回は，この検討結果にもとづく教材展開についてふれてみたい。