混獲率の推定と推定に必要な標本数

The sampling size for the estimation of fish populations in the tagging method.

清水利厚

Toshiatsu SIMIZU

We have a total fishing catch containing N members(unknown) which is unknown to contain S marked members and drawn a single representative sample of n members containing s marked.

In current text-books on statistics, the probability that s is hyper-geometric distribution, and is approximately binominal distribution when N is large.

The confidence limits of the estimate of the tag-ratio in the population were obtained as follows

$$\left[\hat{p}-1.96\sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}}\ ,\ \hat{p}+1.96\sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}}\right]$$

where $\hat{p}=s/n$.

The problem what values n when standerd error is smaller than k. When k is equal to $\alpha\,p$ ($0<\alpha<1$, free), the sampling size were obtained by the formula

$$n=\frac{1}{\alpha^2}\left(\frac{1}{p}-1\right).$$

はじめに

標識放流個体の混獲率を求めて資源量を推定したり放流種苗の再捕数を推定することができる*1。市場調査等で全漁獲物中に標識個体がどれだけ混ざっているかを知りたいとき，全数調査では時間と経費がかさむので一般には確率的に抽出された標本によって全漁獲物中の標識個体の割合が推定される。ここではその際に必要な標本の数について検討する。

混獲率の推定

総漁獲数N個体中に標識個体がS個体存在するとき

$$p=\frac{S}{N}$$

は母集団混獲率である。(p は未知)

今，母集団からランダムに抽出したn個の標本中にs個の標識個体を得る確率は超幾何分布

*1 推定資源量$N_S=T/\hat{p}$ (Tは標識放流数)，推定再捕数 $\hat{S}=\hat{p}N$

$$\mathrm{P}\ [s]=\frac{\binom{S}{s}\binom{N-S}{n-s}}{\binom{N}{n}}$$

で与えられる。

平均と分散は

$$\mathrm{E}\ [s]=\frac{nS}{N}=np$$

$$\mathrm{V}\ [s]=\frac{N-n}{N-1}\ \frac{nS}{N}\ (1-\frac{S}{N})$$

$$=\frac{N-n}{N-1}\,np\,(1-p)$$

となり，標本中の標識個体の割合$s/n=\hat{p}$ の標本平均と標本分散は

$$\mathrm{E}\ [s]=\frac{np}{n}=p$$

$$\mathrm{V}\ [s]=\frac{N-n}{N-1}\ \frac{1}{n^2}\ np\,(1-p)$$

$$=\frac{N-n}{N-1}\ \frac{p\ (1-p)}{n}$$

となる。

標本平均s/nはnを大きくしていくとき，中心極限定理により

$$Z=\frac{\frac{s}{n}-p}{\sqrt{\frac{N-n}{N-1}\ \frac{p\ (1-p)}{n}}}$$

はN（0，1^2）とみなせる。

ここで

$$\Pr\ \{-1.96\leqq Z\leqq 1.96\}=0.95$$

なので

$$\frac{s}{n}-1.96\sqrt{\frac{N-n}{N-1}\ \frac{p\ (1-p)}{n}}\leqq p,$$

$$\frac{s}{n}+1.96\sqrt{\frac{N-n}{N-1}\ \frac{p\ (1-p)}{n}}\geqq p$$

であり，pの95％信頼区間は

$$[\hat{p}-1.96\sqrt{\frac{N-n}{N-1}\ \frac{p\ (1-p)}{n}},\ \hat{p}+1.96\sqrt{\frac{N-n}{N-1}\ \frac{p\ (1-p)}{n}}]$$

となる。$\sqrt{\{(N-n)/(N-1)\}\ \{p(1-p)/n\}}$は標準誤差である。

pが未知であるから計算にあたってはnが十分大きいときはpの代りに$\hat{p}$を用いて近似する。

またNも未知なのでNが十分大きいときに限り二項分布で近似できることを用いて$(N-n)/(N-1)=1$とみなして近似的にpの95％信頼区間を

$$[\hat{p}-1.96\sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}},\ \hat{p}+1.96\sqrt{\frac{\hat{p}\ (1-\hat{p})}{n}}]$$

とする。

誤差の設定と必要標本数

ここで標準誤差をk以下にしたいとき，必要な最小の標本数nはいくつかという問題を考える。Nが十分大きいとき

$$\sqrt{\frac{p\ (1-p)}{n}}\leqq k$$

から

$$n=\frac{p\ (1-p)}{k^2}$$

となる。

ここでpに対する信頼幅を均衡させるために誤差の許容限界をpの一定割合と考えることにする。すなわち$k=\alpha p$　$(0<\alpha<1)$ とすると

$$n=\frac{p\,(1-p)}{\alpha^2 p^2}$$

から

$$n=\frac{1}{\alpha^2}\ (\frac{1}{p}-1)\ \cdots\cdots\cdots(1)$$

となる[*2]。

αを変化させたときのpとnの関係を図１に示す。

ところで一般的には誤差の許容限界を絶対値として設定する。しかしその設定のしかたではpのいろいろな値に対して信頼幅に不均衡を生じ，pが小さい値から0.5に近づくに従い，必要標本数が大きくなるという不合理が生じる。

すなわち

$$n=\frac{p\ (1-p)}{k^2}$$

で

$$p\ (1-p)=\{\frac{1}{4}-(p-\frac{1}{2})^2\}\leqq\frac{1}{4}$$

[*2] Nが比較的小さい値のときは $n=\dfrac{(1-p)\,N}{\alpha^2 p(N-1)+(1-p)}$，$((N-n)/(N-1)\neq 1)$ となる。

実用上，$\alpha>0.05$とするならば，$N\geqq 5{,}000$で(1)式に近似できる。

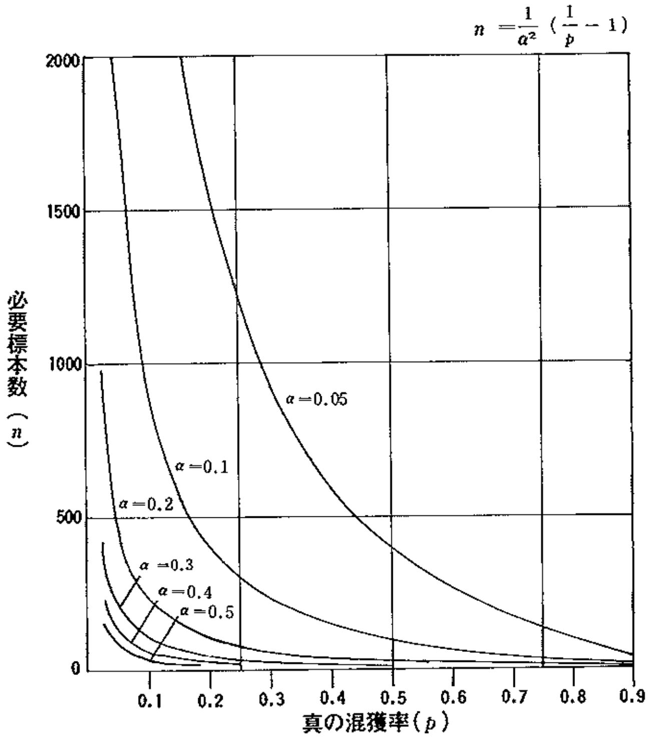

図１　誤差の許容限界をαpとしたときの真の混獲率pに対する必要標本数n

から$p=½$のときnは最大となる。例えば$k=0.01$とした場合，$p=0.2$のときの必要標本数は1,600個体であるが$p=0.5$のときは$n=25,000$個体が必要になってしまう。したがって$k=\alpha p$とするのが合理的である。

例えばpが経験的に0.2とわかっている場合で誤差の割合を母集団比率の５％以内にしたいときに必要な標本数nは1,600個体となる。10％以内では400個体である。pが大きければ必要標本数は小さくてよいがpが小さくなるに従い必要標本数は急激に大きくなる。標本数があまり大きいと実用的でない。標本数を小さくするためにはpが大きくなるように標識数を計画するとよい。$p=0.05$が一つの目安であろう。

この結果は一般的に母比率を推定する場合に適用できる。

要　約

混獲率$\hat{p}$の推定に必要な標本の数nは真の混獲率をp，誤差の限界をαp（$0<\alpha<1$）とすると

$$n=\frac{1}{\alpha^2}\left(\frac{1}{p}-1\right)$$

で示される。

αの各水準についてpとnの関係を図示した。