

## 第 1 章

# サンプリング法による資源量の推定

生物資源の管理や保全を实践するうえで、対象生物の個体数や分布パターンを把握することは大変重要なステップとなる。野外調査においては、この種の情報を獲得するためにしばしばサンプリング法が適用される。サンプリングの広義の意味は、一部を観測して全体の様子を探ることである。通常はサンプリングに費やせるコストが限られているため、あらかじめ適切なサンプリング計画を立てる必要がある。例えば、分布パターンに関して事前の知識がない場合には対象となる調査範囲をまんべんなく調べる必要があるが、環境等の違いに起因する密度の高低が事前にある程度分かっている場合には、この限りではないであろう。サンプリングの計画は調査や研究の目的に応じて設定されるため、一概にどのサンプリング法が適切かを述べることができない。また、その方法自体も多岐に渡っているため、限られた紙面ですべてを紹介することはできない。そこで本章では、まずサンプリング法の基礎となる単純ランダムサンプリングについて述べ、次に個体の集中的な分布に伴う独特の問題、そして野生生物の資源量推定の代表的な方法の一つであるライントランセクト法について述べることにする。

## 1.1 サンプリング法

### 1.1.1 サンプリング法とは

#### 例題 6.1

ある磯においてアワビの生息数を調べるためにサンプリングを計画したい。調査となる対称海域とサンプリングユニットを図 1.1 のように定める。この海域内に個体が一様にランダムに分布していると想定できるとき、選択するサンプリングユニット数と個体密度の推定精度について、その関係を考察せよ。

サンプリングという言葉の響きは大変わかりやすいが、統計学においてその意味するところは奥深い。たとえば、デザインベーストサンプリング (design-based sampling) とよばれる考え方では、図に示したような各区画の個体数を固定された値とし、どのサンプリ

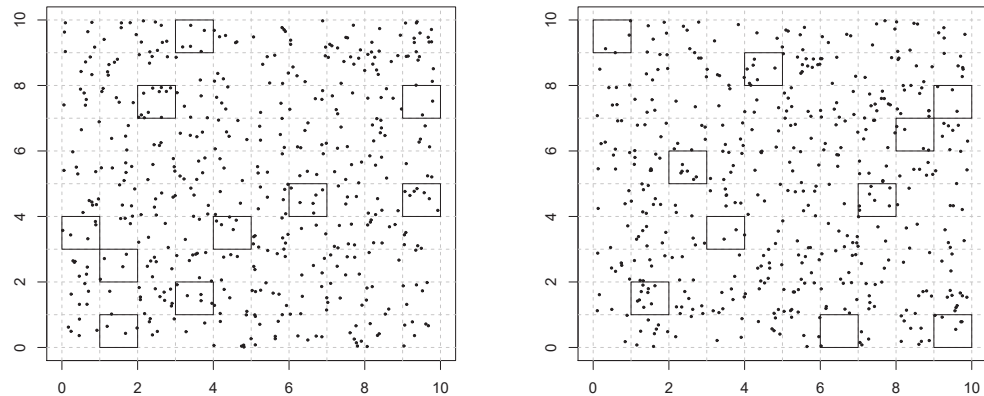


図 1.1: サンプルングのエリアとユニットの例. 真の個体数は 500 とし, 空間内で一様にランダムに分布と仮定.

ングユニットを選ぶかだけにランダムネスがあるという考え方である. 一方で, モデルベーストサンプルング (model-based sampling) という考え方は, サンプルングユニットの選び方に関わらず各サンプルングユニット内の個体数を確率変数と考え, その確率変数の同時分布でランダムネスを表現する. どちらの考え方を採用するかは状況によるが, 例えば例題 6.1 のように調査海域内のユニットを特徴付ける要因がないか, あるいは予備知識がない場合にはデザインベーストサンプルングを仮定するのが自然であろう. その一方で, ユニット間で水温などの生息環境に違いがみられるなど, 事前に補助変数の利用可能性が期待できる場合もある. 例えば, 水温と密度が線形な関係にあると仮定できるときには, デザインベーストサンプルングの一つの方法である比推定量 (ratio estimator) という方法が利用できる. しかし, 非線形性や複数の共変量との関係を表現するには, モデルベーストサンプルングの方が適切なフレームワークとなる. また, ゼロデータが多い場合や, ユニット間の密度に相関がある場合も同様にモデルベースの方が望ましいと考えられる. ただし, そのモデリングにミスペシフィケーションがある場合には必ずしもその恩恵を得られないだろう.

例題 6.1 ではあらかじめ 100 個のサンプルングのユニットを定めているので, 個体のサンプルングというよりもむしろユニット自体をサンプルングすることになる. したがって, 典型的なデザインベーストサンプルングを想定している. 実際, 図 1.1 に示した 2 つの例では, 1 から 100 までの数字を順にユニットに与え, 10 個のサンプルングユニットを重複を許さずランダムに選んだ結果である. このようなサンプルング法を単純ランダムサンプルング (simple random sampling) という. それでは, この単純ランダムサンプルングの設定をもう少し一般化しよう.

いま, 合計  $M$  個のサンプルングユニットがあり, いずれも同じ面積とする. そこから,

$m(\leq M)$  個のサンプルングユニットをランダムに選ぶ。選ばれたユニット内の個体数 (個体密度) をそれぞれ  $n_1, n_2, \dots, n_m$  とするとき, 単位ユニット当たりの個体密度 ( $D$ ) の推定量は, 単位ユニット当たりの個体数の平均値を用いて

$$\hat{D} = \bar{n} = \frac{1}{m} \sum_{i=1}^m n_i \quad (1.1)$$

のように自然な形で推定できる。これをユニット数で引き延ばせば生息個体数 ( $N$ ) の推定量が

$$\hat{N} = M\hat{D} = \frac{M}{m} \sum_{i=1}^m n_i \quad (1.2)$$

として得られる。この推定量がデザインベーストサンプルングの仮定の下で不偏となることは容易に確かめられる。

ところで, サンプルングユニットはランダムに選ばれるため, サンプルングユニットの選び方によって個体数の推定値も変動する。もちろんサンプルングユニット数を大きくすることによって推定精度も向上する。推定量の分散の計算ではユニット数が有限であることを考慮する。実際, 有限補正と呼ばれる修正を施すことで以下のように求められる。

$$V[\hat{D}] = \frac{M-m}{M} \frac{\sigma^2}{m}, \quad V[\hat{N}] = M^2 V[\hat{D}] = M(M-m) \frac{\sigma^2}{m} \quad (1.3)$$

ただし,  $\sigma^2$  は真の個体密度の分散であり,

$$\sigma^2 = \frac{1}{M-1} \sum_{i=1}^M (n_i - D)^2 \quad (1.4)$$

である。仮にすべてのサイトをサンプルングした全数調査の場合には,  $m = M$  となり推定量の変動はゼロとなる。ところで, 実際の調査では真の分散  $\sigma^2$  は通常未知であるから,  $\sigma^2$  に不偏分散

$$s^2 = \frac{1}{m-1} \sum_{i=1}^m (n_i - \bar{n})^2 \quad (1.5)$$

を代入して次式のように推定量の分散を評価する。

$$\hat{V}[\hat{D}] = \frac{M-m}{M} \frac{s^2}{m}, \quad \hat{V}[\hat{N}] = M^2 \text{Var}[\hat{D}] = M(M-m) \frac{s^2}{m}. \quad (1.6)$$

```

1 $図 6.1 の作図と密度 D の推定
2 #サンプルングのエリアとユニットの定義
3 plot(NULL, xlim = c(0,10), ylim = c(0,10), type="n", ann=FALSE)
4 abline(h=seq(0,10), v=seq(0,10), lty=2, col="gray")
5
6 #個体のランダム空間配置
7 Nanimals <- 500
8 X <- runif(Nanimals, 0, 10)
```

```

9  Y <- runif(Nanimals, 0, 10)
10 points(x, y, pch=20, cex=0.5)
11
12 #ユニットの総数とサンプリングユニット（左下から右に順に 1,2,...,100 とラベルを付ける）
13 M <- 100
14 m <- 10
15 Sunits <- sample(1:M, m, replace=FALSE) #サンプリングユニットの選択
16 xs <- Sunits%%10 #サンプリングユニットの左下隅の x 座標（%a は a で割った余り）
17 ys <- floor((Sunits-1)/10) #サンプリングユニットの左下隅の y 座標
18 rec <- cbind(rep(1, m), rep(1, m)) #サンプリングユニットの大きさを支持
19 symbols(xs+0.5, ys+0.5, rectangles=rec, inches=F, bg=rgb(0,1,0,0.3), add=T)
20 #symbols(xs+0.5, ys+0.5, rectangles=rec, inches=F, lwd=2, add=T)
21
22 # サンプリングによる個体数の観測値と密度の推定値
23 nobs <- numeric(m)
24 for(i in 1:m){
25   cond <- (xs[i] < X) & (X <= xs[i]+1) & (ys[i] < Y) & (Y <= ys[i]+1)
26   nobs[i] <- sum(cond)
27 }
28 rbind(Sunits=Sunits, nobs=nobs)
29 Dest <- mean(nobs); Dest
30 s2 <- var(nobs)
31 Dvar <- ((M-m)/M)*(s2/m); Dvar
32 Dcv <- sqrt(Dvar)/Dest; Dcv

```

(Ch6.sampling.map.r)

```

> Dest <- mean(nobs); Dest
[1] 3.9
> s2 <- var(nobs)
> Dvar <- ((M-m)/M)*(s2/m); Dvar
[1] 0.269
> Dcv <- sqrt(Dvar)/Dest; Dcv
[1] 0.1329877

```

(Ch6.SRS.example.r の実行結果)

さて例題 6.1 に戻ろう。個体密度の推定量の分散  $V[\hat{D}]$  は式 (1.3) のように  $\sigma^2$  に依存する。 $\sigma^2$  の経験的な値が利用できる場合や、事前に予備的なサンプリングを実施することが可能であれば、求める精度に応じて式 (1.3) を解くことで必要サンプルサイズを求めることができる。しかし、先述の通り通常は調査を行う前は  $\sigma^2$  が未知である。ところで例題 6.1 では個体が海域内にランダムに分布していると仮定した。正確な計算方法ではないが、総個体数が  $N = M \times D$  であり、単一のユニットに各個体が含まれる確率が  $1/M$  であるから、2 項分布近似を利用すれば  $\sigma^2 = M \times D(1/M) \times (1 - 1/M)$  となり、さらに  $M$  が十分に大きいとしてポアソン近似を利用すれば  $\sigma^2 = D$  となる。個体数推定の際

には，推定量の変動係数を推定精度の目安に用いることが多い．変動係数は，

$$CV[\hat{D}] = \frac{\sqrt{V[\hat{D}]}}{\hat{D}} = \sqrt{\frac{M-m}{MmD}} \quad (1.7)$$

で求められ，図 1.2 のように変動係数の等高線図を描くことで，求められる推定精度に応じて目安となる必要サイト数のある程度見積もることが可能となる．

```

1 #サンプリングユニット数と個体密度に応じた推定量の CV の変化 (図 6.2)
2 M <- 100
3 mvec <- 1:50
4 Dvec <- seq(0.5,10,0.5)
5 CVcomp <- function(M,m,D) sqrt((M-m)/(M*m*D))
6 CV <- outer(mvec, Dvec, FUN=CVcomp, M=M)
7 levels <- seq(0.05, 0.5, 0.05)
8 contour(mvec, Dvec, CV, levels=levels, xlab="Number of sampling units", ylab="Density")
                                     (Ch6.samplesize.r)

```

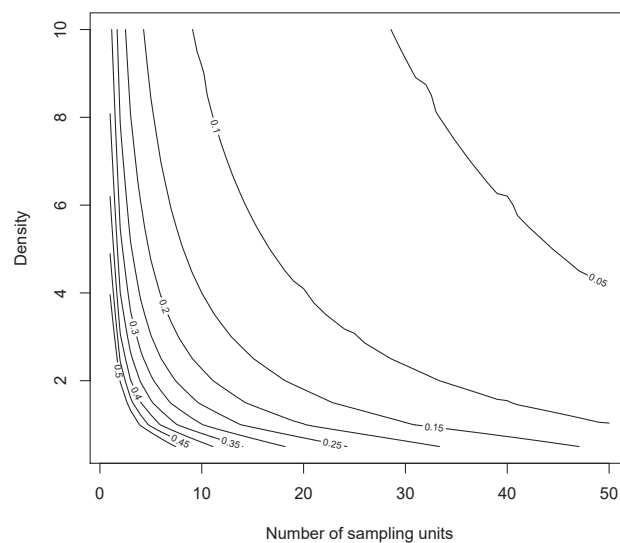


図 1.2: サンプリングユニット数と個体密度に応じた推定量の変動係数の変化．真の分散  $\sigma^2$  にポアソン近似を仮定．

なお，この節で述べた単純ランダムサンプリング以外にも，層化ランダムサンプリング，クラスターサンプリング，多段サンプリングなどデザインベーストサンプリング法にはいくつかの種類がある．層化サンプリングは海域により密度や分布パターンが大きく異なる場合には有効であり，クラスターサンプリングは生物が集中して分布する際に効率の良いサンプリング法となる．また，一定期間に渡って調査を要する場合には，調査日をま

ずランダムにサンプリングし、それからさらにサンプリングの対象をランダムにサンプリングすることも実際的な手法として利用されている。これらのサンプリング法については、Thompson (2002) や北田 (20XX) などを参照されたい。

### 1.1.2 野生生物の空間分布と個体数推定精度の検証

前節ではデザインベーストサンプリングの基礎となる単純ランダムサンプリングについて述べた。もちろん単純ランダムサンプリングはサンプリング法の基礎となるが、野生生物の分布が一様にランダムというのは極めて稀である。したがって、図 1.2 の等高線図では、真の分散が  $\sigma^2 = D$  といういわば計算上理想的な仮定であり、計算される必要サンプル数は実際には過小評価の可能性がある。多様なサンプリング計画に対応するためには、シミュレーションが便利である。そこで、野生生物の分布によって、密度の推定精度の関係を評価するためのシミュレーション実験をしてみよう。

以下では、サンプリングユニットをランダムに選択し個体数を推定する作業は同じであるが、野生生物の分布が

- (A) 一様にランダムな場合
- (B)  $y$  軸方向にのみ  $f(y) = y$  のような正の密度密度勾配がある場合
- (C) 2 変量正規分布  $MVN((3, 3), \mathbf{I}_2)$  と  $MVN((7, 7), \mathbf{I}_2)$  の 50% 混合分布の密度関数に出現頻度が比例する場合

という 3 つの異なる仮定で 10000 回シミュレートし、推定値と分散の挙動を評価してみよう。

空間的に密度勾配のある分布のシミュレートの仕方として、ここでは汎用性のある棄却サンプリング (rejection sampling) を利用する。この方法はベイズ法でも利用されることがあり、ここでは簡単に 1 変数の場合に説明する。いま確率密度関数  $f(x)$  ( $a < x < b$ ) に従う、あるいは確率密度関数が  $f(x)$  に比例する乱数を生成したいが、この乱数を直接生成できないとする。そこで、まず乱数を生成しやすい代替分布として確率密度関数  $g(x)$  から乱数  $x^*$  を生成する。ただし、すべての  $x \in (a, b)$  に対して  $f(x) \leq cg(x)$  となる定数  $c(0 < c < \infty)$  が存在するとする。ここで、 $p(x) = f(x)/cg(x)$  は区間  $(0, 1)$  の値であり、確率  $p(x)$  で  $x^*$  を採択するかどうかを再度ベルヌーイ乱数で決定する。すなわち、一様分布  $U(0, 1)$  に従う乱数  $u$  を生成し、 $u \leq p(x)$  ならば  $x^*$  を採択し、 $u > p(x)$  ならば棄却する。この操作を繰り返し、必要な数だけ乱数を生成する。このような方法で乱数を出すことは  $g(x)$  を条件  $u \leq p(x)$  で制限した条件付分布を考えることと同値である。すなわち、 $U(0, 1)$  の確率密度関数を  $h(u)$  とすると、

$$g(x|u \leq p(x)) = \frac{\int_0^{p(x)} h(u)g(x)du}{\int_a^b \int_0^{p(x)} h(u)g(x)dudx} = \frac{\frac{f(x)}{cg(x)}g(x)}{\int_a^b \frac{f(x)}{cg(x)}g(x)dx} = \frac{f(x)}{\int_a^b f(x)dx} \quad (1.8)$$

となり、棄却サンプリングで目的の乱数を生成できることがわかる。

たとえば  $g(x)$  として一様乱数を仮定し、さらに  $c$  を  $f(x)$  の最大値とすれば、必ずしも効率がよいとは限らないが比較的簡単にシミュレートすることが可能となる。この方法を利用してパターン (A)~(C) に対応し 2 次元の乱数を生成してみよう。

```

1  #棄却サンプリングによる生物の空間分布の生成
2  SpatialDist <- function(Nanimals, FUN){
3    #FUN の最大値を予め 1 としておく
4    X<-Y<-numeric(Nanimals)
5    count <- 1
6    while(count <= Nanimals){
7      xtemp<-runif(1,0,10); ytemp<-runif(1,0,10);
8      prob <- FUN(xtemp,ytemp)
9      if(runif(1,0,1) <= prob) {X[count]<-xtemp; Y[count]<-ytemp; count=count+1}
10   }
11   obj<-NULL; obj$X<-X; obj$Y<-Y; obj
12 }
13
14 #図 6.3
15 #パターン B
16 FUN.B <- function(x,y) y/10
17 SP <- SpatialDist(1000,FUN.B); plot(SP$X,SP$Y, xlab="x", ylab="y", main="Pattern B")
18
19 #パターン C
20 FUN.C <- function(x,y) exp(-(x-3)^2/2-(y-3)^2/2)+exp(-(x-7)^2/2-(y-7)^2/2)
21 SP <- SpatialDist(1000,FUN.C); plot(SP$X,SP$Y, xlab="x", ylab="y", main="Pattern C")
                                     (Ch6.SpatialDist.r)

```

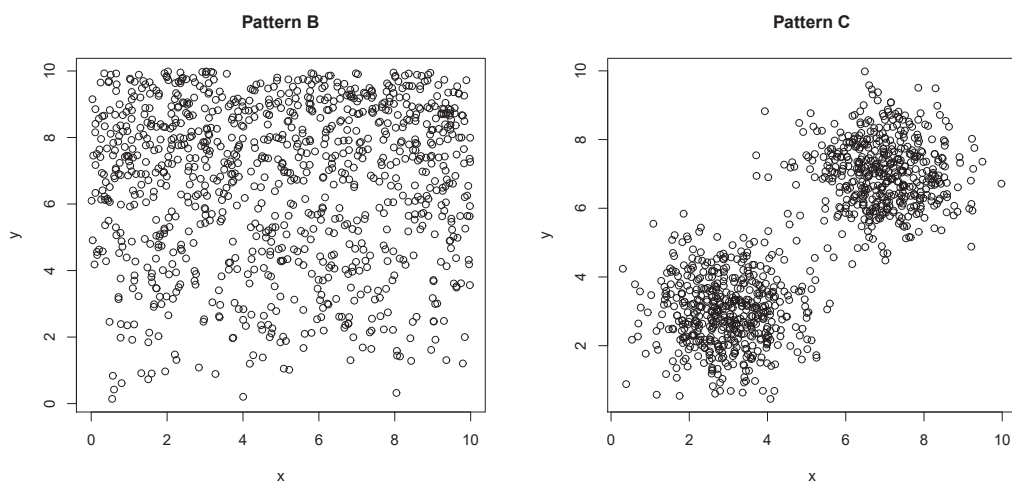


図 1.3: パターン B と C の空間分布の例



```

1  # SRS (Simple Random Sampling)
2  Sim.SRS <- function(Nit, M=100, m, Density, Pattern){
3
4      Dest <- numeric(Nit)    #個体密度の推定値を保存するベクトル (Nit はシミュレーション回数)
5      Dcvest <- numeric(Nit)  #変動係数の推定値を保存するベクトル
6      Nanimals <- M*Density
7
8      if(Pattern=="A") FUN <- function(x,y) 1
9      if(Pattern=="B") FUN <- FUN.B
10     if(Pattern=="C") FUN <- FUN.C
11
12     for(it in 1:Nit){
13
14         SP <- SpatialDist(Nanimals, FUN)
15         X <- SP$X; Y <- SP$Y
16
17         nobs <- numeric(m)
18         Sunits <- sample(1:M, m, replace=FALSE)
19         xs <- Sunits%%10
20         ys <- floor((Sunits-1)/10)
21         for(i in 1:m){
22             nobs[i] <- sum( (xs[i] < X) & (X <= xs[i]+1) & (ys[i] < Y) & (Y <= ys[i]+1) )
23         }
24         s2 <- var(nobs)
25         Dest[it] <- mean(nobs)
26         Dcvest[it] <- sqrt(((M-m)/M)*(s2/m))/Dest[it]
27     }
28
29     obj <- NULL
30     obj$Dest <- Dest
31     obj$Dest.mean <- mean(Dest)
32     obj$Dest.cv <- sd(Dest)/mean(Dest)
33     obj$Dcvest <- Dcvest
34     obj$Dcvest.mean <- mean(Dcvest)
35     return(obj)
36 }
37
38 Density.true<-5
39 mvec <- c(5,10,20)
40 Res.A <- lapply(mvec, Sim.SRS, Nit=1000, M=100, Density=Density.true, Pattern="A")
41 Res.B <- lapply(mvec, Sim.SRS, Nit=1000, M=100, Density=Density.true, Pattern="B")
42 Res.C <- lapply(mvec, Sim.SRS, Nit=1000, M=100, Density=Density.true, Pattern="C")
43
44 par(mfrow=c(3,3), cex=0.7)
45 Res <- Res.A
46 #Res <- Res.B
47 #Res <- Res.C
48 for(i in 1:length(mvec)){
49     hist(Res[[i]]$Dest, xlim=c(0, 2*Density.true), col="gray80", ann=FALSE)
50     title(main=paste("(C) #Samling Units=", mvec[i]), xlab="Estimate of Density")
51     points(Density.true, 0, pch=4, col="red")
52     points(Res[[i]]$Dest.mean, 0, pch=4, col="blue")

```



53 }

(Ch6.Sim.SRS.r)

図 1.4 には、真の個体数密度が 5 で、サンプリングユニット数を 5, 10, 20 と変化させたときの結果をヒストグラムで示している。想定する生物の空間分布が何であろうと単純ランダムサンプリングによる推定量は不変であるが、その推定精度は大きく異なることがわかるであろう。実際、前節で計算したポアソン近似の変動係数値とパターン A のそれはほぼ同じであるが、パターン B, C の変動係数はかなり大きな値をとっている。したがって、ある程度分布の形状が想定できる場合には、シミュレーションなどにより推定精度を事前に検証することが必要であろう。

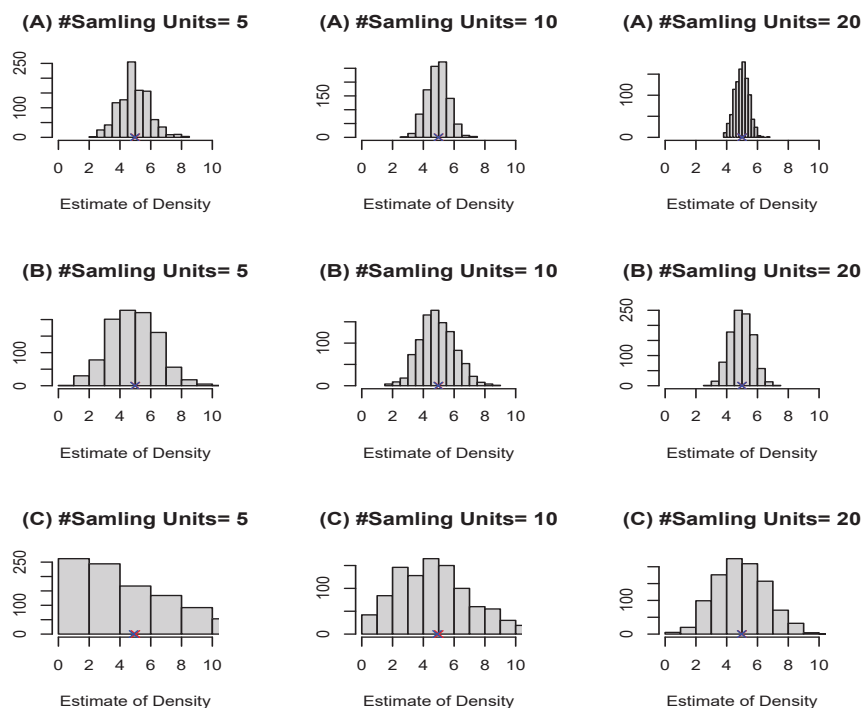


図 1.4: 真の空間分布を変化させたときの密度の推定値。真の個体数密度は 5 を仮定。

```
> CV[10,10]
[1] 0.1341641
> Res.A[[2]]$Dest.cv
[1] 0.1375979
> Res.B[[2]]$Dest.cv
[1] 0.226557
> Res.C[[2]]$Dest.cv
[1] 0.5246171
```

(Ch6.SRS.comparison.r の実行結果)

なお，空間分布の仮定の仕方は上記に限らず，いくつかの共変量を組み合わせることでより实际的となる．このような生物の空間的な分布の表現法には，一般化加法モデルや，点仮定モデルなど様々な方法が考えられている．これらは，第 x x 章「空間分布の推定」で詳しく述べる．

## 1.2 ライントランセクト法 計算，コード未チェック

### 例題 6.3

図 1.5 はある年に北海道日本海側実施されたトド資源に対するライントランセクト調査デザインである．このように事前に定められたコース上を船で走行し目視観察を実施した結果，表 1.1 のように各トラックライン の努力量 (走行距離，単位は海里)，発見群数，そして発見した群れ毎にトラックラインとの垂直距離 (横距離とよぶ，単位は海里)，群れの大きさ，そして観測環境が記録された．このデータを利用して，資源量推定値を求めよ．

表 1.1: 北海道日本海側トド資源量調査結果 (抜粋)

Region.Label	Area	Sample.Label	Pre.Effort	Act.Effort	Sea.state	Size	PD
Area1	15962	2010_SS12	90.8	90.8	2	1	0.077
Area1	15962	2010_SS13	102.6	94.3	2	1	0.074
Area1	15962	2010_SS13	102.6	94.3	2	1	0.199
...							
Area1	15962	2010_SS17	140.6	140.6	2	1	0.093
Area1	15962	2010_SS17	140.6	140.6	2	2	0.129
Area1	15962	2010_SS18	140.7	139.5	2	1	0.135
...							
Area3	7385	2010_KS14	19.2	19.2			
Area3	7385	2010_KS15	22.6	22.6			
Area3	7385	2010_KS16	33.6	33.6	1	1	0.222
Area3	7385	2010_KS16	33.6	33.6	1	1	0.02
Area3	7385	2010_KS17	43.8	41.5	1	1	0.039

この節で紹介するライントランセクト法は生物資源量の調査手法の一つである．もちろんこの方法は陸上の生物にも適用可能であるが，ここでは水産資源を対象し，調査船による目視調査を想定し，ライントランセクト法の基礎となるストリップランセクト法から話を始め，次に群れの発見関数を考慮した実際のモデルへと展開することとする．途