

分布混合分析的深入研究及其在南极磷虾种群结构中的应用*

鲁北伟 王 荣

(中国科学院海洋研究所, 青岛 266071)

提要 介绍分布混合分析的理论, 并运用分布混合分析估计南极磷虾种群体长分布概率密度函数的参数, 首次给出体长取样间隔有效性和方差 σ^2 下限的数学分析。概率密度函数

式为 $Q(x) = \sum_{j=1}^K p_j q(x; \mu_j, \sigma_j)$, 式中 x 是随机变量体长的取值; K 是年龄组数; p_j 是第 A_j

年龄组的强度; $q(x; \mu_j, \sigma_j)$ 是均值为 μ_j ; 方差为 σ_j^2 的正态分布。 $p_j, \mu_j, \sigma_j, j=1, 2, \dots, K$ 是待估计的参数。求解这些参数的最大似然估计值时, 发现最速下降法是一种非常实用的方法, 很容易选取 p_j, μ_j, σ_j 的一组适定的初始近似解, 使得迭代过程收敛。

关键词 体长频数分布 分布混合分析 种群结构 南极磷虾

南极磷虾的年龄鉴定至今是个难题, 因而倍受关注。Siegel (1987) 和 Hosie (1988) 曾将分布混合分析应用于南极近海的大磷虾 (*Euphausia superba*) 的种群分析, 但存在问题。主要是南极磷虾成体后生长越来越缓慢, 体长频数分布的多峰性经常是不明显的, 特别是大龄组的数量急剧减少, 它们的存在不易分辨。本研究在应用该方法时, 不根据多峰性人为选定年龄组数, 而是对所有可能存在的年龄组数都进行计算, 根据计算得出的体长频数分布与实测频数分布的拟合程度及是否符合南极磷虾的生物学规律再作出判断。应用该方法对印度洋区南极磷虾的种群结构进行分析, 取得了满意的结果 (王荣等, 1995)。本文旨在介绍分布混合分析的理论, 重点介绍最速下降法 (steepest descent method), 并给出了南极磷虾种群结构分析的实例。

1 分布混合分析理论

1.1 概率密度函数

一般认为种群内各年龄组的体长 (或其它生长指标) 服从均值为 μ , 方差为 σ^2 的正态分布 $q(x, \mu, \sigma)$ (李淑霞, 1990):

$$q(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

以南极磷虾为例。假设有 M 条同龄虾, 以 $\int_{x-h}^{x+h} Q(x) dx$ 表示体长在 $(x-h, x+h)$

* 国家八五攻关专题, 85-905-02-01 号。鲁北伟, 女, 出生于 1957 年 3 月, 助研。

承李超伦、王文琪和吉鹏同志协助计算, 谨志谢忱。

收稿日期: 1993 年 9 月 13 日, 接受日期: 1995 年 7 月 11 日。

之间的虾的理论频数, 则有 $Q(x) = q(x; \mu, \sigma)$ 。其中 x 是随机变量体长的取值; μ 是体长均值; σ^2 是体长方差, 反映该龄虾体长的差异或生长速率的差异。一般地, 假设有

M 条多龄虾, 包括 A_1, A_2, \dots, A_K 龄虾, 仍然以 $\int_{x-h}^{x+h} Q(x) dx$ 表示体长在 $(x-h,$

$x+h)$ 之间的虾的理论频数, 则有 $Q(x) = \sum_{j=1}^K p_j q(x; \mu_j, \sigma_j)$, 其中 p_j 是 A_j 龄虾的数

目占总虾数的比例或者说是 A_j 龄虾的强度 $\left(\sum_{j=1}^K p_j = 1 \right)$; μ_j 是 A_j 龄虾的体长均值;

σ_j^2 是 A_j 龄虾的体长方差。由已知的样品总数 M 和体长在 (x_i-h, x_i+h) 之间的虾的数目 n_i 求解 $K, p_j, \mu_j, \sigma_j (j=1, 2, \dots, K)$ 称为参数估计。

1.2 参数估计

Harding (1949) 首先使用正态概率纸, 以图解法分析多年龄组种群生长指标的频数分布, 估计参数 p_j, μ_j 和 σ_j , 并突破了以往仅能求解 $K < 3$ 时的 p_j, μ_j 和 σ_j 的局限, 尽管 Cassie (1950, 1954) 对 Harding 的方法做了改善, 但这类方法在很大程度上依赖分析者的人为判断, 计算结果不准确, 分析同一组样本时, 不同的工作人员可能得出不同的结果。后来, Hasselblad (1966) 运用最速下降法和牛顿法, Macdonald 等 (1979) 运用最小二乘法、 χ^2 最小法和修正的 χ^2 最小法, 求解参数的最大似然估计值, 这类数值分析方法, 保证了计算结果的一致性。

1.2.1 最大似然估计 已知 M 条虾的体长分别是 x_1, x_2, \dots, x_M , 体长在 $(x_i-h,$

$x_i+h)$ 之间的虾有 n_i 条 ($i=1, 2, \dots, I; \sum_{i=1}^I n_i = M; I < M$), x 的概率密度函数

$$Q(x) = \sum_{j=1}^K p_j q(x; \mu_j, \sigma_j)。其中 p_K = 1 - \sum_{j=1}^{K-1} p_j, 则$$

$$\begin{aligned} Q(x) &= \sum_{j=1}^K p_j q(x; \mu_j, \sigma_j) = \sum_{j=1}^{K-1} p_j [q(x; \mu_j, \sigma_j) - q(x; \mu_K, \sigma_K)] + q(x; \mu_K, \sigma_K) \\ &= \sum_{j=1}^K p_j [q(x; \mu_j, \sigma_j) - q(x; \mu_K, \sigma_K)] + q(x; \mu_K, \sigma_K) \end{aligned}$$

$$\text{记: } q_{ij} = q(x_i; \mu_j, \sigma_j), Q_i = Q(x_i) = \sum_{j=1}^K p_j q_{ij}, w_{ij} = \frac{n_i q_{ij}}{Q_i}, W_j = \sum_{i=1}^I w_{ij}$$

则: x_1, x_2, \dots, x_M 的似然函数 (李揚碚, 1980):

$$L = \sum_{m=1}^M \ln Q_m = \sum_{i=1}^I n_i \ln Q_i = \sum_{i=1}^I n_i \ln \left[\sum_{j=1}^K p_j (q_{ij} - q_{iK}) + q_{iK} \right]$$

$$\text{由: } \frac{\partial L}{\partial p_j} = \sum_{i=1}^I \frac{n_i}{Q_i} (q_{ij} - q_{iK}) = 0 \quad j=1, 2, \dots, K-1$$

$$\sum_{j=1}^K p_j \left[\sum_{i=1}^I \frac{n_i}{Q_i} (q_{ij} - q_{ik}) \right] = \sum_{i=1}^I \frac{n_i}{Q_i} \sum_{j=1}^K p_j q_{ij} - \sum_{i=1}^I \frac{n_i}{Q_i} q_{ik} \sum_{j=1}^K p_j = M - \sum_{i=1}^I \frac{n_i}{Q_i} q_{ik} = 0$$

$$\sum_{i=1}^I \frac{n_i}{Q_i} q_{ik} = M$$

$$\text{得: } \frac{\partial L}{\partial p_j} = \sum_{i=1}^I \frac{n_i}{Q_i} (q_{ij} - q_{ik}) = W_j - M = \left(\frac{M}{p_j} \right) \left[W_j \frac{p_j}{M} - p_j \right] \quad j=1, 2, \dots, K-1 \quad (1)$$

$$\frac{\partial L}{\partial \mu_j} = \sum_{i=1}^I \frac{n_i q_{ij}}{Q_i} \frac{p_j}{\sigma_j^2} (x_i - \mu_j) = \left(\frac{p_j}{\sigma_j^2} W_j \right) \left[\sum_{i=1}^I \frac{w_{ij} x_i}{W_j} - \mu_j \right] \quad j=1, 2, \dots, K \quad (2)$$

$$\frac{\partial L}{\partial \sigma_j} = \sum_{i=1}^I \frac{n_i q_{ij}}{Q_i} p_j \left[\frac{(x_i - \mu_j)^2}{\sigma_j^3} - \frac{1}{\sigma_j} \right] = \left(\frac{p_j}{\sigma_j^3} W_j \right) \left[\sum_{i=1}^I \frac{w_{ij} (x_i - \mu_j)^2}{W_j} - \sigma_j^2 \right]$$

$$j=1, 2, \dots, K \quad (3)$$

使似然函数 L 最大的参数值即为最大似然估计值, 根据微积分的求极值原理, 此时

p_j, σ_j, μ_j 分别是使 $\frac{\partial L}{\partial p_j} = 0, \frac{\partial L}{\partial \mu_j} = 0, \frac{\partial L}{\partial \sigma_j} = 0$ 的解. (1), (2), (3) 是非线性方程组, 需用迭代法求解.

1.3 最速下降法

1.3.1 基本公式 简单迭代法的一般形式是 $z_1 = U(z_0), z_2 = U(z_1), \dots, z_{n+1} = U(z_n)$, z_0 是人为选取的初始近似解, z_{n+1} 是第 $n+1$ 次迭代计算值.

$$\text{令: } r_n = \frac{\partial L}{\partial z_n} = \lim_{\lambda_n r_n \rightarrow 0} \frac{L(z_n + \lambda_n r_n) - L(z_n)}{\lambda_n r_n}$$

$$\text{则: } \lim_{\lambda_n r_n \rightarrow 0} \frac{L(z_n + \lambda_n r_n) - L(z_n)}{\lambda_n} = \lim_{\lambda_n r_n \rightarrow 0} \frac{L(z_n + \lambda_n r_n) - L(z_n)}{\lambda_n r_n} r_n = \frac{\partial L}{\partial z_n} r_n = r_n^2 > 0$$

即: 只要 $\lambda_n > 0$, 恒有 $L(z_n + \lambda_n r_n) > L(z_n)$. 令 $z_{n+1} = z_n + \lambda_n r_n$, 重复这个迭代计算, z_{n+1} 会使 L 逐渐增大, 当 $|z_{n+1} - z_n| < \varepsilon, \lambda_n \gg 0$ 时, 即有 $r_n = \frac{\partial L}{\partial z_n} = 0$, 此时的 z_n 使 L 达到极大.

$$\text{令: } r_n = \frac{\partial L}{\partial z_n} = \left(\frac{1}{\lambda_n} \right) [U(z_n) - z_n]$$

则有 $\lambda_n r_n = U(z_n) - z_n, z_{n+1} = z_n + \lambda_n r_n = z_n + U(z_n) - z_n = U(z_n)$, 与(1), (2)和(3)比较后得到下列公式:

$$\lambda_{p_j} = \frac{p_j}{M} > 0, \lambda_{\mu_j} = \frac{\sigma_j^2}{p_j W_j} > 0, \lambda_{\sigma_j} = \frac{\sigma_j^3}{p_j W_j} > 0$$

$$p_{j,n+1} = \frac{W_{j,n} p_{j,n}}{M} \quad j=1, 2, \dots, K-1 \quad (4)$$

$$\mu_{j,n+1} = \sum_{i=1}^I \frac{w_{ij,n} x_i}{W_{j,n}} \quad j=1, 2, \dots, K \quad (5)$$

$$\sigma_{j,n+1}^2 = \sum_{i=1}^I \frac{w_{ij,n} (x_i - \mu_j)^2}{W_{j,n}} \quad j=1, 2, \dots, K \quad (6)$$

(4), (5), (6)式就是本文使用的迭代方程组。计算时, 首先选择一组大致符合体长频数分布的估计值, 作为初始近似解代入方程组中进行计算。几十组数据的计算结果表明, 这组方程总是收敛的, 只要 ε 足够小, 同一组数据赋予不同的初始近似解时, 计算结果一致。

1.3.2 方程组(4), (5), (6)的数学意义 限于篇幅, 仅讨论显而易见的情况。由公

式 $p_{j,n+1} = \frac{W_{j,n} p_{j,n}}{M} = p_{j,n} \sum_{i=1}^I \frac{n_i}{MQ_{i,n}} q_{ij,n}$ 可以得出, 当理论频数等于实测频数 $MQ_i = n_i$, 则 $p_{j,n+1} = p_{j,n}$; 当 $MQ_{i,n} < n_i$, 则 $p_{j,n+1} > p_{j,n}$; 当 $MQ_{i,n} > n_i$, 则 $p_{j,n+1} < p_{j,n}$ 。

$$\text{公式 } \mu_{j,n+1} = \sum_{i=1}^I \frac{w_{ij,n} x_i}{W_{j,n}} = \frac{\sum_{i=1}^I x_i \frac{n_i q_{ij,n}}{MQ_{i,n}}}{\sum_{i=1}^I \frac{n_i q_{ij,n}}{MQ_{i,n}}} \quad \text{中, } w \text{ 相当于权函数, } \mu_j \text{ 是权函数}$$

的曲线重心位置。若 w 是单峰对称曲线, μ_j 是曲线峰值的位置。当 $MQ_{i,n} = n_i$, 则 $\mu_{j,n+1} = \mu_{j,n}$ 。当 μ_j 偏小而使得 $i < i_x$ 时, $\frac{n_i}{MQ_{i,n}} < 1$, $\frac{n_i q_{ij,n}}{MQ_{i,n}} < q_{ij,n}$; $i > i_x$ 时,

$\frac{n_i}{MQ_{i,n}} > 1$, $\frac{n_i q_{ij,n}}{MQ_{i,n}} > q_{ij,n}$; 则 $\mu_{j,n+1} > \mu_{j,n}$ 。当 μ_j 偏大而使得 $i < i_x$ 时, $\frac{n_i}{MQ_{i,n}} > 1$,

$\frac{n_i q_{ij,n}}{MQ_{i,n}} > q_{ij,n}$; $i > i_x$ 时, $\frac{n_i}{MQ_{i,n}} < 1$, $\frac{n_i q_{ij,n}}{MQ_{i,n}} < q_{ij,n}$; 则 $\mu_{j,n+1} < \mu_{j,n}$ 。同理

可以推出 $\sigma_{j,n+1}^2 = \sum_{i=1}^I \frac{w_{ij,n} (x_i - \mu_j)^2}{W_j}$ 对 σ_j^2 的调整。

这说明两点。一是方程组(4), (5), (6)分别对 $p_{j,n}$, $\mu_{j,n}$, $\sigma_{j,n}$ 有调整作用, 使 $p_{j,n}$, $\mu_{j,n}$, $\sigma_{j,n}$ 沿着使理论频数与实测频数吻合的趋势取值。这种调整作用使得人们很容易选取一组适定的 $p_{j,0}$, $\mu_{j,0}$, $\sigma_{j,0}$ 。二是当理论频数与实测频数相等即 $MQ_{i,n} = n_i$ 时, $p_{j,n}$, $\mu_{j,n}$, $\sigma_{j,n}$ 是方程组的解; 当 $MQ_{i,n} \neq n_i$ 时, 不可能同时成立 $p_{j,n+1} = U(p_{j,n})$, $\mu_{j,n+1} = U(\mu_{j,n})$, $\sigma_{j,n+1} = U(\sigma_{j,n})$, 因此方程组(4), (5), (6)的解是唯一的, 是使理论频数与实测频数吻合的解。

2 几点说明

2.1 $q(x; \mu, \sigma)$, $Q(x)$ 离散化后的有效性

$$\text{由公式 } q(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \frac{\partial q}{\partial x} = q(x; \mu, \sigma) \left(-\frac{x-\mu}{\sigma^2}\right),$$

$$\frac{\partial^2 q}{\partial x^2} = \frac{q(x; \mu, \sigma)}{\sigma^2} \left[\frac{(x-\mu)^2}{\sigma^2} - 1 \right]$$

根据微分求极值原理, $\frac{\partial^2 q}{\partial x^2} = 0$ 时, $\frac{\partial q}{\partial x}$ 有极值, 设此时 $x = x_{\max}$, 则 $|x_{\max} - \mu| = \sigma$,

$$\left| \frac{\partial q}{\partial x} \Big|_{x=x_{\max}} \right| = \frac{1}{\sigma^2\sqrt{2\pi}} e^{-0.5} = \frac{0.242}{\sigma^2}. \quad \text{根据微分中值定理, } q(b) - q(a)$$

$= q'(c) \times (b-a)$, $a < c < b$, 得

$$|q_{i+1} - q_i| < = \frac{0.242}{\sigma^2} (x_{i+1} - x_i) \quad (7)$$

$$\begin{aligned} |Q_{i+1} - Q_i| &= \left| \sum_{j=1}^K p_j [q(x_{i+1}; \mu_j, \sigma_j) - q(x_i; \mu_j, \sigma_j)] \right| \\ &< = \sum_{j=1}^K p_j \left(\frac{0.242}{\sigma_j^2} \right) (x_{i+1} - x_i) \\ &< = \left(\frac{0.242}{\sigma_{\min}^2} \right) (x_{i+1} - x_i) \end{aligned} \quad (8)$$

其中: $\sigma_{\min} = \min(\sigma_1, \sigma_2, \dots, \sigma_K)$

这说明, $x_{i+1} - x_i = 1$ 时, σ 要足够大, $q(x)$, $Q(x)$ 的变化才不大, 使得 $\int_{-\infty}^{\infty} q(x) dx$

$$= 1, \mu = \int_{-\infty}^{\infty} xq(x) dx, \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 q(x) dx, M \int_{x-0.5}^{x+0.5} Q(x) dx \text{ 可以分别近}$$

似为 $\sum_{i=1}^l q(x_i) = 1, \mu = \sum_{i=1}^l x_i q(x_i), \sigma^2 = \sum_{i=1}^l (x_i - \mu)^2 q(x_i), MQ(x_i)$ 。

$|x - \mu| < 3\sigma$ 时, $e^{-\frac{(x-\mu)^2}{2\sigma^2}} > e^{-4.5} > 0.01 \neq 0$, 因此, 须仔细选取 x_1, x_l 值以避免出现 $q(x_1)$ 或 $q(x_l)$ 过大的情况。

2.2 σ_j 的下限

因为 $q(x_i) > 0, \sum_{i=1}^l q(x_i) = 1$, 所以 $q(x_i) < 1$ 应该在任意的 x_i 处成立。由

$$q(x_{\max}) = \frac{e^{-0.5}}{\sigma\sqrt{2\pi}} < 1, \text{ 得 } \sigma_{\min} > 0.242, \text{ 结合(7), (8)式的情况, 当 } x_{i+1} - x_i = 1 \text{ 时,}$$

以 $\sigma_j > 1$ 为宜。

2.3 最佳K值

K 不同则 L 的公式不同, 因而不能依据 L 的值判断最佳 K 值。 χ^2 值可以作为一种有效判据, 它既可以用于判断最佳 K 值, 又可以用于判断理论频数分布与实测频数分布的拟合程度。 χ^2 值越小, 拟合程度越好(陈敦隆, 1982), 此时的 K 应为最佳 K 值。同时根据已经掌握的南极磷虾的生物学资料, 如寿命、性成熟年龄、生长率和各年龄组的体长范围等, 分析拟合的合理性, 确认得出的每个“年龄组”是否确实代表一个年龄组(如 $p_j < 1\%$ 时, 可以认为该“年龄组”不能代表实际年龄组), 进而确定它们代表那一个年龄组。

3 结果与讨论

图 1 和表 1 给出了一个南极磷虾样本的最速下降法分布混合分析实例。样品于 1991 年 2 月 18 日取自普利兹湾外海 ($65^\circ 58' S, 75^\circ 00' W$), 这 1782 尾磷虾的体长频数分布有两个明显的峰, 第一个峰值在 28—29mm, 十分明确它代表 1^+ 龄虾; 第二个峰值在 46—47mm, 代表成体态, 从成体态的体长范围 (19—61mm) 看, 其中可能包括了 2^+ 到 5^+ 乃至 5^+ 以上的所有年龄组。因为 6^+ 和 6^+ 以上的虾在种群中的比例很小 ($< 1\%$), 可以忽略不计, 因此设定最大年龄为 5, 样品的年龄组数为 5, 即 $K = 1, 2, 3, 4, 5$ 。

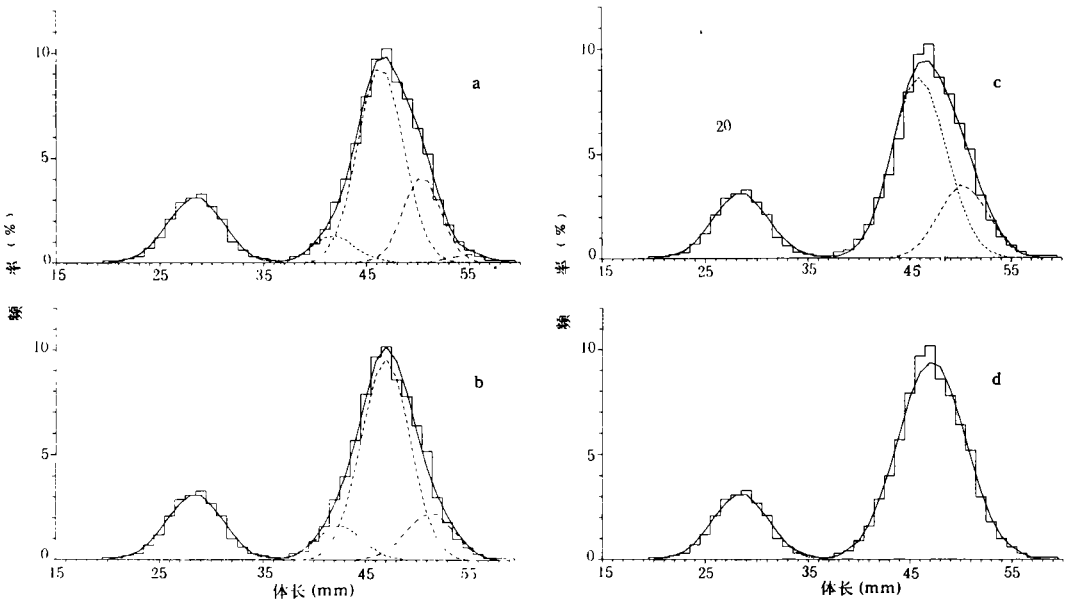


图 1 一个南极磷虾样本的分布混合分析结果 (不同 K 值的比较)

Fig. 1 Results of distribution mixture analysis of a krill (*Euphausia superba*) sample (comparison of different K)

比较 4 种计算结果 (表 1), 可以看出, $K = 5$ 时拟合效果最好, 同时也符合南极磷虾的生长规律。拟合的 5 个年龄组的平均体长和标准差与已有资料 (自然条件下和实验室培养条件下) 的 1^+ 到 5^+ 龄 5 个年龄组的平均体长和标准差基本一致。其它计算结果

不仅拟合效果不如 $K=5$ 的好，而且很难从已知的有关南极磷虾的生物学规律对拟合结果作出解释。如 $K=3$ ，年龄组 1 代表 1^+ 龄，年龄组 2 从平均体长 (46mm) 看只能是 3^+ 龄组，这样就不存在 2^+ 龄组了，这几乎是不可能的。 1^+ 龄 (分布区限制) 和 5^+ 龄 (数量少) 有可能在某个样本中不出现，但 2^+ ， 3^+ ， 4^+ 龄虾在某个样本中完全不出现的记录迄今还未有过。同样， $K=2$ 时，意味着成体态中只有一个 3^+ 龄组，这更是不可能的，首先 3^+ 没有这样大的体长跨度，其次 2^+ 和 4^+ 可能较少但不可能完全没有。

从理论上讲，因为实测结果是客观存在，最佳拟合结果应当最符合生物学规律。本文所计算的大多数样品中，最佳拟合结果最符合生物学规律。但由于取样的代表性、测量误差和种群组成的复杂性 (不同来源种群的混杂等) 有时可能出现最佳拟合结果不是最符合生物学规律的。在拟合结果 (χ^2 值) 相差不大时，应更多考虑生物学规律，相差较大时可以同时给出两种结果。

表 1 一个南极磷虾样本的分布混合分析结果 (不同 K 值的比较)

Tab.1 Results of distribution mixture analysis of a krill (*Euphausia superba*) sample (comparison of different K)

K	A_j	μ_j (mm)	σ_j (mm)	P_j (%)	M	χ^2	自由度
5	1	28.40	2.66	21.06	1 782	5.18	33
	2	41.81	2.07	6.64			
	3	46.40	2.20	51.73			
	4	50.39	1.82	18.74			
	5	54.85	1.90	1.84			
4	1	28.41	2.66	21.07	1 782	9.91	36
	2	42.47	2.22	9.05			
	3	46.94	2.37	56.76			
	4	51.31	2.32	13.12			
3	1	28.44	2.71	21.14	1 782	18.48	39
	2	46.01	2.69	57.57			
	3	50.30	2.44	21.29			
2	1	28.40	2.67	21.04	1 782	18.67	42
	2	47.15	3.33	78.96			

4 结语

为根据体长研究南极磷虾的种群结构，本文详细讨论了分布混合分析的理论 and 具体计算方法，给出取样间隔的有效性和方差 σ_j^2 的下限。估计体长分布概率密度函数的参数时，发现最速下降法非常实用，其公式的物理意义明晰，计算过程简单，很容易选取一组适定的初始近似解。

参 考 文 献

- 王荣、鲁北伟, 1995, 海洋与湖沼, 26(6): 598—605.
- 李淑霞, 1990, 生物学中常用数学方法, 青岛海洋大学出版社(青岛), 1—371.
- 李惕碛, 1980, 实验的数学处理, 科学出版社(北京), 1—370.
- 陈敦隆, 1982, 海洋科学研究中的概率统计方法, 海洋出版社(北京), 1—541.
- 谢如彪、蒋培庆, 1984, 非线性数值分析, 上海交通大学出版社(上海), 1—187.
- Cassie, R.M., 1950, *N.Z. Sci. Rev.*, **8**: 89—91.
- Cassie, R.M., 1954, *Aust. J. Mar. Freshwater Res.*, **5**: 513—522.
- Harding, J.P., 1949, *J. Mar. Biol. Assoc. U.K.*, **28**: 141—153.
- Hasselblad, V., 1966, *Technometrics*, **8**: 431—444.
- Hosie, G.W. et al., 1988, *Polar Biol.*, **8**: 213—224.
- Macdonald, P.D.M. and Pitcher, T.J., 1979, *J. Fish. Res. Board Can.*, **36**: 987—1001.
- Siegel, V., 1987, *Mar. Biol.*, **96**: 483—495.

FURTHER STUDY ON DISTRIBUTION MIXTURE ANALYSIS AND ITS APPLICATION TO AGE-GROUPS ANALYSIS OF ANTARCTIC KRILL (*EUPHAUSIA SUPERBA*)

Lu Beiwei, Wang Rong

(Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071)

Abstract This paper presents the principle and algorithm of distribution mixture analysis, and its application to age-groups analysis of Antarctic krill (*Euphausia superba*), and presents, for the first time, the analytically deduced effective range of sampling interval and the lower limit of variance σ_j^2 . The probability density function is $Q(x) = \sum_{j=1}^K p_j q(x; \mu_j, \sigma_j)$, where x is the observed value of random variable, body length; K is number of age-group; p_j is strength of the A_j age-group, $q(x; \mu_j, \sigma_j)$ is normal probability distribution function with mean μ_j and variance σ_j^2 . $p_j, \mu_j, \sigma_j, j=1, 2, \dots, K$ are parameters to be estimated. The steepest descent method was used to estimate the parameters. The result shows that the steepest descent method is practical. Its deduction process is simple and easy to understand, its calculation is simple, and selection of a set of proper iterative initial values of the parameters is easy also.

A age-groups analysis of Antarctic krill by means of the steepest descent method is given in this paper. Using this method, better understanding of the population structure of Antarctic krill is obtained based on the length-frequency data (Wang et al., 1995).

Key words Length-frequency distribution Distribution mixture analysis Population structure Antarctic krill

Key words Length-frequency distribution Distribution mixture analysis Population structure Antarctic krill