Математические аспекты проблемы стандартизации уловов на усилие

01 октября 2014

Содержание

Наблюдаемый улов на усилие - случайная функция многих факторов

$$u = U(I(t), x_1, x_2, \ldots, x_d) \tag{1}$$

$$E^{st}_{t,f} = \frac{C_{t,f}}{I_t} \tag{2}$$

$$E^{st}_t = \sum_f E^{st}_{t,f} \tag{3}$$

Метод Галланда

$$I(t) = u_{t,f^*} \tag{4}$$

$$E^{st}_{t,f} = \frac{C_{t,f}}{I(t)} = \frac{C_{t,f}}{u_{t,f^*}} = \frac{C_{t,f}}{C_{t,f^*}} E_{t,f^*} \tag{5}$$

Метод Бивертона -Холта

$$E^{st}_{t,f} = \frac{\overline{u_{t,f}}}{u_{t,f^*}} E_{t,f} = \frac{C_{t,f}}{I(t)} \tag{6}$$

$$I(t) = \frac{C_t}{E_t} \tag{7}$$

Метод Гавариса

$$\ln u = \ln I(t) + \sum_f \ln P_f + \varepsilon \qquad (8)$$

Стандартизация - оптимальная факторизация уловов на усилие

Этапы стандартизации

- Выбор статистической гипотезы о распределении уловов на усилие
- Выбор набора факторов
- Выбор стандартных значений факторов

Линейный предиктор.

Аддитивная факторизация

$$\varphi(E\mathbf{u}) = v = f(x_1) + f(x_2) + \cdots + f(x_d) \tag{9}$$

Линейность непрерывных факторов

$$f(x_j) = \theta_j x_j, j = 1, \ldots, k \tag{10}$$

Категориальные факторы

x_j - пробегает дискретное множество значений $j < k$

Обобщенная линейная модель (GLM) - это статистическая гипотеза о распределения наблюдаемой величины (CPUE)

- Принадлежность к экспоненциальному семейству
- Факторизуемость матожидания

Процедура стандартизация уловов на усилие с использованием GLM есть ни что иное как настройка модели наблюдения.

Построение экспоненциального семейства распределений

Произвольное распределение

$$\rho(u) = \exp(-L(u)) \tag{11}$$

Производящая функция моментов

$$E\exp(\eta u) = \int \exp(u\eta - L(u))du \equiv \exp(G(\eta)) \tag{12}$$

Построение экспоненциального семейства распределений

Экспоненциальное семейство

$$\rho(u,\eta) = \exp(u\eta - L(u) - G(\eta)) \tag{13}$$

Связь с преобразованием Лапласа

$$G(\eta) = \ln \mathcal{L}[\rho](-\eta) \tag{14}$$

Свойства экспоненциального семейства

Производящая функция моментов

$$E \exp(tu) = \exp(G(\eta + \mathbf{t}) - G(\eta)) \tag{15}$$

Для вычисления любого момента функции распределения достаточно соответствующее число раз продифференцировать в нуле производящую функцию

$$M^{(n)} = \partial^{(n)} E \exp(tu) \,|_{t=0} \tag{16}$$

Свойства экспоненциального семейства

Обратная функция связи

$$Eu = \frac{\partial G(\eta)}{\partial \eta} \equiv \mu(\eta) \equiv \varphi^{-1}(\eta) \tag{17}$$

Дисперсия зависит от мат.ожидания

$$D(\mathbf{u}) = \partial^2_{\eta\eta} G(\eta) = \frac{\partial \mu(\eta)}{\partial \eta}(\eta) = \frac{1}{\varphi'(\mu)} \equiv \frac{1}{V(\mu)} \tag{18}$$

Теорема факторизации.

Условное распределение $P(u_1, \ldots, u_N|\eta)$ при $S(u_1, \ldots, u_N) = s$не зависит от η - статистика $S(u_1, \ldots, u_N)$ является достаточной для оценки параметра η . $\Leftrightarrow$ функция правдоподобия представима в виде

$$\rho(u_1, \ldots, u_N; \eta) = \psi(S(u_1, \ldots, u_N), \eta)h(u_1, \ldots, u_N) \qquad (19)$$

Основная причина использования экспоненциального семейства - существование минимальной достаточной статистики для оценки параметров предиктора

$$\rho(u, v) = \exp(u\eta - L(u) - G(\eta)) \qquad (20)$$

Максимум правдоподобия

$$\sum_j (-L(u_j) + u_j\eta_j + G(\eta_j)) \to \max \tag{21}$$

$$\varepsilon^T X = \sum_j (u_j - \mu(\eta_j)) \frac{\partial \eta_j}{\partial \theta_\alpha} = 0 \tag{22}$$

Итерационный метод Ньютона

$$\theta^{(k+1)} = \beta^{(k)} + (X^T V^{-1}(\theta^{(k)}) X)^{-1} X^T \varepsilon(\theta^{(k)}) \tag{23}$$

Отсутствие внешних факторов

$$\mu(\eta) = \frac{1}{N} \sum u_i \tag{24}$$

$$\eta = \varphi(\frac{1}{N} \sum u_j) = \mu^{-1}(\frac{1}{N} \sum \mu(\eta_j)) \tag{25}$$

Усреднение по Колмогорову

Методы повышения робастности.

- Оценка с использованием функции псевдоправдоподобия
- Байесовские методы
- Медианные и модальные оценки

Функциия псевдоправдоподобия

$$Q(\mu; u) = \int_{u}^{\mu} \frac{u - y}{V(\mu)} dy \qquad (26)$$

Сохраняет связь дисперсии с матожиданием, но позволяет не уточнять вид распределения.

Мода распределения

$$mode\rho(u,\eta) = \arg\max_u \exp(u\eta - L(u) - G(\eta)) \quad (27)$$

$$\eta = \frac{\partial L(u)}{\partial u} \equiv \lambda(u) \quad (28)$$

$$mode\rho(u,\eta) = \lambda^{-1}(\eta) \quad (29)$$

Обратное преобразование Лапласа

$$L(u) = \ln \mathcal{L}^{-1}[\exp(G(-\eta))](u) = \frac{1}{2\pi i}\int_{-a-i\infty}^{-a+i\infty} \exp(-u\eta + G(\eta))d\eta \quad (30)$$

Обобщения.

- Обобщенные аддитивные модели(GAM)
  - отказ от линейности
- Обощенные линейные смешанные модели (GLMM)
  - независимые переменные как коррелированые случайные величины
- Байесовские методы
  - байесовские процедуры оценивания

Выводы.

- Оценка для максимально широкого класса распределений
- Устранение кажушейся гетероскедастичности моделей
- Необходимость учета всех факторов
- Необходимость апостериорной проверки гипотезы о распределении
- Потенциальная неустойчивость параметрических оценок

Боровков А. А. Математическая статистика.— М.: Наука. 1984

McCullagh P., Nelder J. Generalized linear models 2ed., Chapman, 1989

Maunder M. N.,. Punt A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, Volume 70, Issues 2-3, Pages 141-159

McCulagh P., Nelder J. 1989. Generalized Linear Models. Second Edition. Boca Raton: Chapman and Hall / CRC. IBSN 0-412-31760-5.

Venables W.N.,. Dichmont C.M. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research.// Fisheries Research Volume 70, Issues 2-3,Pages 319-337

- Xiao Y., Punt A. E., Millar R. B., Quinn T.J. 2004.Models in fisheries research: GLMs, GAMS and GLMMs.// Fisheries Research, Volume 70, Issues 2-3, Pages 137-139
- Helser T. E, Punt A. E., Methot R. D. 2004 A generalized linear mixed model analysis of a multi-vessel fishery resource survey Original Research Article. // Fisheries Research, Volume 70, Issues 2-3, Pages 251-264
- Колмогоров А. Н. 1985 Математика и механика // Избранные труды / отв. ред. С. М. Никольский, сост. В. М. Тихомиров. - М.: Наука,. - Т. 1. - С. 136-138.
- O'Brien C.M.; L.T. Kell 1997 The use of generalized linear models for the modelling of catch-effort series. I-Theory. Col.Vol.Sci.Pap. ICCAT, 46 (4) : 476-482

- Рябенький В.С. 2000 Введение в вычислительную математику. - 2-е изд., исправл. - М.: Физматлит,.
- Jie Cao, Xinjun Chen, Yong Chen, Bilin Liu, Jin Ma and Siliang Li 2011/ Generalized linear Bayesian models for standardizing CPUE: an application to a squid-jigging fishery in the northwest Pacific Ocean. Scientia Marina 75(4) December 2011, 679-689, Barcelona (Spain) ISSN: 0214-8358
- Хьюбер Дж. П. 1984 Робастность в статистике: Пер. с англ.- М.: Мир,