

УДК 519.958

О ВЫДЕЛЕНИИ BS-РАСПРЕДЕЛЕНИЯ ИЗ СЕМЕЙСТВА GBS-РАСПРЕДЕЛЕНИЙ

И.Н. Володин, О.А. Джунгурова, С.В. Симушкин

Аннотация

В рамках семейства обобщенных распределений Бирнбаума – Саундерса строится критерий отношения правдоподобия, выделяющий из этого семейства обычное распределение Бирнбаума – Саундерса. Показывается, что предельное распределение статистики критерия есть хи-квадрат распределение с одной степенью свободы; методом статистического моделирования исследуется точность аппроксимации.

Введение

Продолжены исследования, начатые в работах [1, 2], по выделению специальных типов распределений из семейства GBS-распределений (обобщенного распределения Бирнбаума – Саундерса, предложенного в работе [3]). GBS-распределение есть смесь двух распределений и имеет функцию плотности $f(x) = (1 - p)f_1(x) + pf_2(x)$, $x > 0$, где

$$f_1(x) = f_1(x | \theta, \lambda) = \frac{\lambda}{\theta\sqrt{2\pi}} \left(\frac{\theta}{x}\right)^{3/2} \exp \left\{ -\frac{\lambda^2}{2} \left(\sqrt{\frac{\theta}{x}} - \sqrt{\frac{x}{\theta}} \right)^2 \right\}$$

есть обратное гауссовское распределение (IG-распределение),

$$f_2(x) = f_2(x | \theta, \lambda) = \frac{\lambda}{\theta\sqrt{2\pi}} \left(\frac{\theta}{x}\right)^{1/2} \exp \left\{ -\frac{\lambda^2}{2} \left(\sqrt{\frac{\theta}{x}} - \sqrt{\frac{x}{\theta}} \right)^2 \right\}$$

есть смещенное по долговечности (long-biased) IG-распределение (BIG-распределение). Задачи оценки параметров этого распределения решены в работе [4]. В работе [1] построены асимптотически (объем выборки $n \rightarrow \infty$) оптимальные критерии выделения IG- и BIG-распределений из семейства GBS, а также асимптотически (параметр формы $\lambda \rightarrow \infty$) оптимальные критерии приближенной нормальности. В настоящей статье предлагается использовать критерий отношения правдоподобия для выделения BS-распределения (проверки гипотезы $p = 1/2$) из семейства GBS. Хорошо известно (см. [5], где приведена соответствующая библиография), что критерий отношения правдоподобия для проверки гипотез о значении параметра смеси не всегда имеет предельное хи-квадрат распределение, поскольку параметр p принадлежит замкнутому отрезку $[0, 1]$. В данной работе устанавливается, что для проблемы выделения BS-распределения имеет место стандартное предельное хи-квадрат распределение с одной степенью свободы. Методом статистического моделирования исследуется скорость сходимости к предельному закону; предложенный критерий применяется к тестированию BS-распределения на реальных данных.

1. Выделение обратных гауссовских типов из семейства GBS-распределений

Обобщенная модель Бирнбаума–Саундерса содержит как частные случаи три известных семейства распределений, соответствующих значениям $p = 0, 0.5, 1$. Таким образом, перед нами стоит задача проверки гипотезы $p = p_0$ при неизвестных значениях «мешающих» параметров λ и θ . Естественно воспользоваться критерием отношения правдоподобия с критической областью

$$Q = 2 \left[\sup_{p, \theta, \lambda} \mathfrak{L}(\theta, \lambda, p | X^{(n)}) - \sup_{\theta, \lambda} \mathfrak{L}(\theta, \lambda, p_0 | X^{(n)}) \right] > C,$$

где

$$\mathfrak{L}(\theta, \lambda, p | X^{(n)}) = \sum_{i=1}^n \ln f(X_i),$$

$X^{(n)} = (X_1, \dots, X_n)$ – случайная выборка из соответствующего распределения.

Поскольку в данном случае не выполняются стандартные условия регулярности, обеспечивающие предельное хи-квадрат распределение статистики Q , то необходимо провести дополнительные асимптотические исследования (см. в связи с этим, например, статью Т. Higori [6]). Основная проблема состоит в том, что оценка максимального правдоподобия (ОМП) $\hat{p}$ гипотетического параметра p может принимать значения на границе параметрического пространства: $\hat{p} = 0$ или 1 , и, что более существенно, производные от логарифмического правдоподобия не обращаются в нуль с положительной вероятностью при значениях $p = \hat{p}$ (равных 0 или 1), $\lambda = \hat{\lambda}$, $\theta = \hat{\theta}$, где $\hat{\lambda}$ и $\hat{\theta}$ – ОМП параметров λ и θ . Тем не менее, в рамках конкретной модели GBS эти производные обращаются в нуль, что позволяет сформулировать

Теорема 1. *Если объем выборки $n \rightarrow \infty$, то распределение статистики Q сходится к распределению хи-квадрат с одной степенью свободы.*

Доказательство. Вывод предельного распределения статистики Q начинается с ее тейлоровского разложения:

$$\begin{aligned} Q &= 2 \left[\sup_{p, \theta, \lambda} \mathfrak{L}(\theta, \lambda, p | X^{(n)}) - \sup_{\theta, \lambda} \mathfrak{L}(\theta, \lambda, p_0 | X^{(n)}) \right] = \\ &= 2 \left[\mathfrak{L}(\hat{\theta}, \hat{\lambda}, \hat{p}) - \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) \right] = \\ &= (\hat{p} - p_0) \frac{\partial}{\partial p} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + 2 (\hat{\lambda} - \hat{\lambda}_0) \frac{\partial}{\partial \lambda} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &+ 2 (\hat{\theta} - \hat{\theta}_0) \frac{\partial}{\partial \theta} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + (\hat{p} - p_0)^2 \frac{\partial^2}{\partial p^2} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &+ (\hat{\lambda} - \hat{\lambda}_0)^2 \frac{\partial^2}{\partial \lambda^2} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + (\hat{\theta} - \hat{\theta}_0)^2 \frac{\partial^2}{\partial \theta^2} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &+ 2 (\hat{p} - p_0) (\hat{\lambda} - \hat{\lambda}_0) \frac{\partial^2}{\partial p \partial \lambda} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &+ 2 (\hat{p} - p_0) (\hat{\theta} - \hat{\theta}_0) \frac{\partial^2}{\partial p \partial \theta} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &+ (\hat{\lambda} - \hat{\lambda}_0) (\hat{\theta} - \hat{\theta}_0) \frac{\partial^2}{\partial \lambda \partial \theta} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + R_n(X^{(n)}), \end{aligned}$$

где $\hat{\lambda}_0, \hat{\theta}_0$ – точки достижения максимума (ОМП) функции правдоподобия $\mathfrak{L}(\theta, \lambda, p_0) = \mathfrak{L}(\theta, \lambda, p_0 | X^{(n)})$ по λ и θ . Остаточный член $R_n(X^{(n)})$ стремится к нулю по вероятности, в чем легко убедиться, решая уравнение правдоподобия для параметра p методом итераций.

Так как для GBS-распределения выполняются стандартные условия регулярности, то первые производные в этом разложении обращаются в нуль, и эти же условия обеспечивают справедливость следующих разложений:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \lambda} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) = \\ &= (p_0 - \hat{p}) \frac{\partial^2}{\partial p \partial \lambda} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + (\hat{\lambda}_0 - \hat{\lambda}) \frac{\partial^2}{\partial \lambda^2} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &\quad + (\hat{\theta}_0 - \hat{\theta}) \frac{\partial^2}{\partial \lambda \partial \theta} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + R_{1,n}(X^{(n)}), \end{aligned}$$

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) = \\ &= (p_0 - \hat{p}) \frac{\partial^2}{\partial \theta \partial p} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + (\hat{\theta}_0 - \hat{\theta}) \frac{\partial^2}{\partial \theta^2} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + \\ &\quad + (\hat{\lambda}_0 - \hat{\lambda}) \frac{\partial^2}{\partial \theta \partial \lambda} \mathfrak{L}(\hat{\theta}_0, \hat{\lambda}_0, p_0) + R_{2,n}(X^{(n)}), \end{aligned}$$

где остаточные члены $R_{1,n}$ и $R_{2,n}$ представляют собой квадратичные многочлены от разности ОМП при нулевой гипотезе и альтернативе с коэффициентами вида третьих производных от функции правдоподобия в средней точке.

Если из этой системы двух уравнений найти $\hat{\lambda}_0 - \hat{\lambda}$ и $\hat{\theta}_0 - \hat{\theta}$, а результат подставить в разложение статистики Q , то получим представление Q в виде

$$Q = (\hat{p} - p_0)^2 \mathbf{I}_{11}^{-1} + o_p(1),$$

где $\mathbf{I}_{11}^{-1}$ – элемент матрицы, обратной к информационной матрице Фишера. Это представление обеспечивает предельное распределение хи-квадрат с одной степенью свободы для тестовой статистики Q .

Таким образом, этот результат останется справедливым и в нашем случае, если первые логарифмические производные $\mathfrak{L}'_p, \mathfrak{L}'_\lambda, \mathfrak{L}'_\theta$ обращаются в нуль в точке $(\hat{\theta}, \hat{\lambda}, \hat{p})$. Это, очевидно, так для производных по λ и θ при любом значении $\hat{p} \in [0, 1]$. Что же касается производной по p , то достаточно убедиться, что она обращается в нуль в точках $\hat{p} = 0$ и $\hat{p} = 1$.

Действительно, если $\hat{p} = 0$, то уравнение максимального правдоподобия имеет явное решение:

$$\hat{\theta} = \frac{1}{n} T_1, \quad \hat{\lambda} = \left(\frac{T_1 T_2}{n^2} - 1 \right)^{-1}.$$

Нетрудно убедиться в том, что

$$\frac{\partial}{\partial p} \mathfrak{L}(\theta, \lambda, 0) = \sum_{k=1}^n \left[\frac{f_2(X_k)}{f_1(X_k)} - 1 \right],$$

и, подставляя в правую часть приведенные выше оценки λ и θ , получим, что эта производная обращается в нуль. Случай $\hat{p} = 1$ рассматривается аналогичным образом. $\square$

Табл. 1

 $\lambda = 1$

p	n						
	10	50	100	200	300	400	1000
0.01	0.015	0.015	0.020	0.005	0.010	0.010	0.005
0.05	0.075	0.035	0.080	0.040	0.040	0.025	0.060
0.1	0.144	0.060	0.159	0.109	0.085	0.060	0.100
0.2	0.279	0.174	0.259	0.194	0.194	0.174	0.199
0.3	0.517	0.308	0.363	0.299	0.299	0.289	0.274
0.4	0.622	0.388	0.448	0.403	0.373	0.383	0.378
0.5	0.801	0.498	0.532	0.512	0.493	0.488	0.493
0.6	0.925	0.642	0.632	0.597	0.627	0.572	0.587
0.7	0.985	0.791	0.751	0.721	0.741	0.692	0.697
0.8	1.000	0.915	0.871	0.826	0.846	0.771	0.801
0.9	1.000	0.975	0.970	0.915	0.896	0.861	0.891
0.95	1.000	0.995	0.995	0.970	0.965	0.940	0.960
0.99	1.000	1.000	1.000	1.000	0.995	0.985	0.985

Табл. 2

 $\lambda = 3$

p	n						
	10	50	100	200	300	400	1000
0.01	0.000	0.015	0.020	0.015	0.000	0.005	0.015
0.05	0.144	0.090	0.050	0.045	0.035	0.035	0.025
0.1	0.542	0.219	0.080	0.144	0.095	0.085	0.070
0.2	0.990	0.582	0.343	0.368	0.274	0.234	0.204
0.3	1.000	0.881	0.741	0.667	0.527	0.458	0.358
0.4	1.000	0.995	0.930	0.886	0.791	0.706	0.557
0.5	1.000	1.000	0.990	0.990	0.975	0.915	0.761
0.6	1.000	1.000	0.995	1.000	0.995	0.990	0.915
0.7	1.000	1.000	1.000	1.000	1.000	1.000	0.985
0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000

2. Численное исследование точности аппроксимации

Исследование скорости сходимости распределения статистики отношения правдоподобия к предельному хи-квадрат распределению проводилось на данных статистического моделирования. Параметр формы λ выбирался равным 1 и 3, а объем выборки n – 10, 50, 100, 200, 300, 400, 500, 1000. Так как критерий отношения правдоподобия инвариантен относительно масштабных преобразований, то «мешающий» параметр θ полагался равным единице. Ниже приведены данные результатов моделирования на основе 1000 случайных репликаций. В таблицах приводятся частоты попадания статистики отношения правдоподобия в интервалы вида $(-\infty, x_p)$ с правой границей x_p , равной p -квантили хи-квадрат распределения с одной степенью свободы (значения p приводятся в первой колонке таблицы). В соответствии с доказанной теоремой ожидается, что указанные частоты будут близки к p .

Представленные данные моделирования убедительно свидетельствуют о крайне низкой скорости сходимости распределения тестовой статистики к предельному закону, особенно при больших λ . На наш взгляд, это можно объяснить двумя причинами. Во-первых, хорошо известно, что оценки максимального правдоподобия $\hat{p}$

параметра смеси p имеют крайне низкую скорость сходимости. Например, в работе [7] показывается, что скорость сходимости $\hat{p}$ к p варьируется от $\sqrt{\ln n}$ до $\sqrt[4]{n}$.

Во-вторых, для нашей задачи плохая сходимость обуславливается асимптотической ($\lambda \rightarrow \infty$) нормальностью обоих распределений смеси с очень близкими значениями параметров: $IG \rightsquigarrow N(\lambda, \lambda)$, $BIG \rightsquigarrow N(\lambda + 1, \lambda + 2)$ (см. [1]). Так, статистический анализ реальных данных, приведенных в статье [8], показывает, что любая из гипотез, касающаяся значения параметра p ($= 0, 0.5, 1$), а также гипотеза нормальности распределения подтверждается с высоким критическим уровнем значимости. Для этих данных оценки максимального правдоподобия параметра λ принимают очень большие значения (от 50 до 130).

Summary

I.N. Volodin, O.A. Dzungurova, S.V. Simushkin. About distinguishing of BS-distribution from the family of GBS-distributions.

A likelihood-ratio test is constructed which distinguishes the Birnbaum–Saunders distribution from the family of the generalized Birnbaum–Saunders distributions. It is established that the asymptotical distribution of the test statistics is chi-square distribution with one degree of freedom. An accuracy of approximation is investigated by Monte–Carlo method.

Литература

1. *Dzungurova O.A., Volodin I.N.* On limit distributions emerging in the generalized Birnbaum–Saunders model // J. Math. Sci. – 2000. – V. 99, No 3. – P. 1348–1366.
2. *Дзунгурова О.А.* Критерий отношения правдоподобия для обобщенной модели Бирнбаума–Саундерса // Тез. докл. «Обозрение прикл. и пром. математики». Третий всеросс. симпозиум по прикладной и промышленной математике, Сочи, 1–6 окт. 2002 г. – 2002. – Т. 9, № 2. – С. 365.
3. *Jorgensen B., Seshadri V., Whitmore G.A.* On the mixture of the inverse Gaussian distribution with its complimentary reciprocal // Scand. J. Statist. – 1991. – V. 18, No 1. – P. 77–79.
4. *Gupta R.C., Akman H.O.* On the reliability studies of a weighted inverse Gaussian model // J. Statist. Plann. Inference. – 1995. – V. 48, No 1. – P. 69–83.
5. *Vu H.T.V, Zhou S.* Generalization of likelihood ratio tests under nonstandard conditions // Ann. Statist. – 1997. – V. 25, No 2. – P. 897–916.
6. *Hironori T.* On the likelihood ratio test for a single model against the mixture of two known densities // Commun. Stat. Theory Meth. – 2001. – V. 30, No 5. – P. 931–942.
7. *Chen J.* Optimal rate of convergence for finite mixture models // Ann. Statist. – 1995. – V. 23, No 1. – P. 221–233.
8. *Birnbaum Z.W., Saunders S.C.* Estimation for a family of life distribution with application to fatigue // J. App. Probab. – 1969. – V. 6, No 2. – P. 328–337.

Поступила в редакцию
20.06.06

Володин Игорь Николаевич – доктор физико-математических наук, профессор, заведующий кафедрой математической статистики Казанского государственного университета.

E-mail: *Igor.Volodin@ksu.ru*

Джунгурова Ольга Александровна – кандидат физико-математических наук, старший преподаватель кафедры математической статистики Казанского государственного университета.

E-mail: *Olga.Dzhungurova@ksu.ru*

Симушкин Сергей Владимирович – кандидат физико-математических наук, доцент кафедры математической статистики Казанского государственного университета.

E-mail: *Sergey.Simushkin@ksu.ru*