

К ОЦЕНИВАНИЮ ПАРАМЕТРОВ ЗАКОНОВ РАСПРЕДЕЛЕНИЙ И ПРОВЕРКЕ ГИПОТЕЗ ПО ЦЕНЗУРИРОВАННЫМ ВЫБОРКАМ¹

Б.Ю. Лемешко, С.Н. Постовалов, Е.В. Чимитова

Новосибирский государственный технический университет
Новосибирск, Россия. E-mail: headrd@fpm.ami.nstu.ru

Аннотация. Рассматриваются вопросы несмещенного оценивания параметров законов распределений и проверки гипотез о согласии по сильно цензурированным выборкам.

С задачей обработки цензурированных выборок, когда наблюдению оказывается доступной только часть области определения случайной величины, а для выборочных значений, попавших левее и/или правее этой области, фиксируется лишь сам факт этого попадания, приходится сталкиваться в различных приложениях. Особенно часто с цензурированными выборками встречаются в задачах надежности при оценивании продолжительности жизни, когда решают остановить эксперимент, как только часть изделий, подвергнутых испытанию, выходит из строя. Это может быть связано с ограничением на время испытаний (цензурирование I типа), либо с ограничением на наблюдаемое число отказов (цензурирование II типа). В таком случае анализ приходится проводить по выборке, сильно цензурированной справа.

Очевидно, что в такой неполной (цензурированной) выборке содержится меньше информации, чем в полной. Потеря части информации отражается на точности оценивания параметров аппроксимирующего закона распределения. При цензурировании наблюдений снижается способность критериев согласия различать близкие законы распределения. В среде специалистов по надежности, которым наиболее часто приходится сталкиваться с проблемами обработки сильно цензурированных выборок, сложилось даже мнение о бесперспективности различения моделей законов распределений, используемых в задачах надежности и контроля качества, с помощью критериев согласия [1].

В работе [2] и в дальнейшем нами были исследованы потери в информации Фишера в зависимости от степени цензурирования для различных законов распределения. Было показано, что в некоторых случаях, когда доступной наблюдению оказывается лишь незначительная область определения случайной величины, в цензурированной выборке сохраняется достаточно много информации (это же отмечается и в работе [3]). Например, в случае экспоненциального закона распределения при цензурировании слева, когда доступной наблюдению оказывается область справа, вероятность попадания в которую равна 0.05, сохраняется более 52% (!) от информации, имеющейся в полной выборке. Наличие подобных фактов позволяет надеяться, что по таким выборкам можно как находить хорошие оценки параметров, так и достаточно уверенно проверять гипотезы о согласии эмпирического распределения с теоретическим. При цензурировании справа и доступности для наблюдения области слева, вероятность попадания в которую равна также 0.05, в случае того же экспоненциального распределения сохраняется лишь 5% информации.

Оценки максимального правдоподобия (ОМП) параметров распределений по цензурированным наблюдениям являются асимптотически эффективными. Однако при ограниченных объемах выборок и значительной степени цензурирования законы распределения ОМП весьма далеки от асимптотически нормального и, более того, оказываются *асимметричными*, а сами оценки *смещенными*. Например, рис.1 иллюстрирует вид плотностей оценок по полной $\hat{\theta}_n$ и по цензурированным справа выборкам $\hat{\theta}_n^c$ параметра масштаба экспо-

¹ Работа поддержана Российским фондом фундаментальных исследований (проект № 00-01-00913)

нормального закона при различной степени цензурирования для $n=100$. На рисунке, соответственно, в процентах указана вероятность попадания в наблюдаемую область. С уменьшением n и увеличением степени цензурирования увеличивается асимметрия закона распределения $\hat{\theta}_n^c$. Аналогично, на рис. 2 показано, как меняется вид законов распределения ОМП параметра $\hat{\sigma}_n^c$ нормального закона при различной степени цензурирования слева или справа.

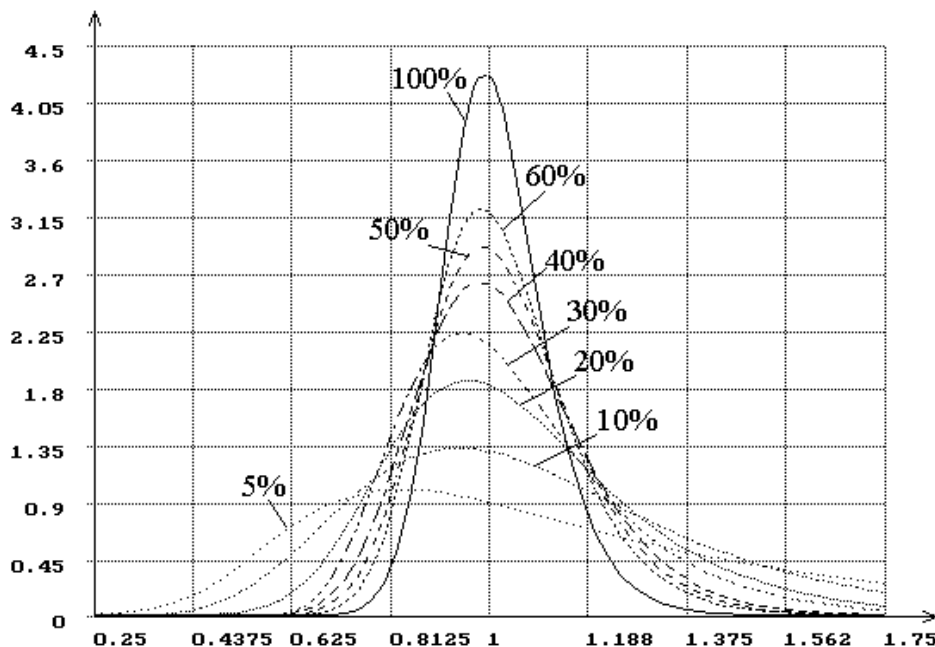


Рис. 1. Плотности распределения оценок масштабного параметра экспоненциального распределения при цензурировании справа при различной величине (%) наблюдаемой области определения случайной величины и полном объеме выборки $n=100$.

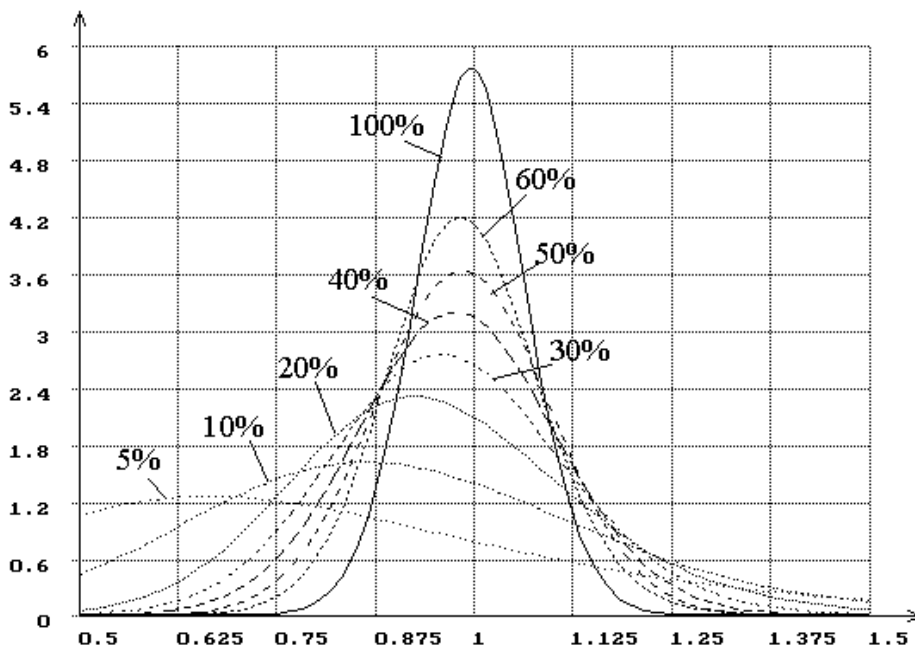


Рис. 2. Плотности распределения оценок масштабного параметра нормального распределения при цензурировании слева (справа) при различной величине (%) наблюдаемой области определения случайной величины и объеме выборки $n=100$.

С использованием методики компьютерного моделирования и анализа статистических закономерностей были исследованы величины смещения ОМП параметров наблюдаемого закона в зависимости от объема всей выборки n и величины наблюдаемой её части. Было показано, что для цензурирования I типа оценки параметров экспоненциального и нормального законов распределения оказываются практически несмещенными. Для цензурирования II типа ОМП параметра масштаба экспоненциального закона при цензурировании слева являются практически несмещенными, а при цензурировании справа – существенно смещены вправо (рис.1). Причем смещение растет с уменьшением наблюдаемой части выборки и уменьшением полного объема выборки. В случае нормального закона цензурирование слева приводит к смещению параметра сдвига вправо, а цензурирование справа приводит к смещению параметра сдвига влево. Для параметра масштаба нормального распределения цензурирование слева и справа приводит к смещению оценки влево (рис. 2).

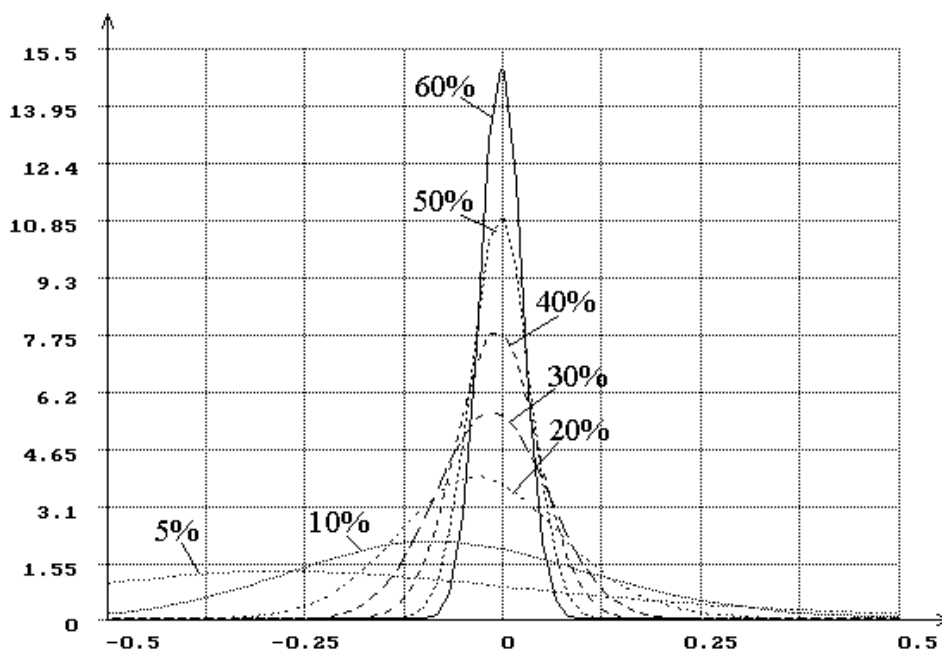


Рис. 3. Плотности распределения величины $(\hat{\mu}_n - \hat{\mu}_n^c) / |\hat{\mu}_n|$ для оценок параметра сдвига нормального закона при цензурировании слева при различной величине (%) наблюдаемой области определения случайной величины и объеме выборки $n=100$.

В целях определения возможной точности оценивания параметров по цензурированным выборкам были исследованы распределения величин вида $(\hat{\theta}_n - \hat{\theta}_n^c) / |\hat{\theta}_n|$. Результаты показали, что эти распределения при малых n и большой степени цензурирования оказываются также существенно асимметричными. Например, на рис. 3 приведены распределения величины $(\hat{\mu}_n - \hat{\mu}_n^c) / |\hat{\mu}_n|$ для параметра сдвига нормального закона при объеме выборки $n=100$. Как видим, в данном случае с уменьшением наблюдаемой части выборки и уменьшением объема выборки n увеличивается смещение оценок $\hat{\mu}_n^c$ и увеличивается дисперсия этих оценок. В данном случае распределения обладают большей симметрией.

В результате исследования методами статистического моделирования были получены оценки математических ожиданий смещений ОМП как эмпирические функции от объема выборки и степени цензурирования. Использование полученных величин в качестве поправок, ликвидирующих смещение при оценивании параметров экспоненциального и нормального законов, показывает, что такие оценки с поправками оказываются несмещенными (выборочное математическое ожидание новых оценок практически совпадает со значением параметра, с которым проводилось моделирование). При этом дисперсия несмещенных оценок

меньше, чем дисперсия оценок, полученных без применения поправок на смещение. Результат применения поправок иллюстрирует рис. 4, на котором приведены функции распределения ОМП параметра сдвига нормального закона по полной выборке $\hat{\mu}_n$ (на рис. обозначено 1), ОМП $\hat{\mu}_n^c$ по цензурированной справа выборке при величине наблюдаемой области, вероятность попадания в которую равна 0.1 (2) и ОМП по цензурированной выборке с поправкой на смещение (3).

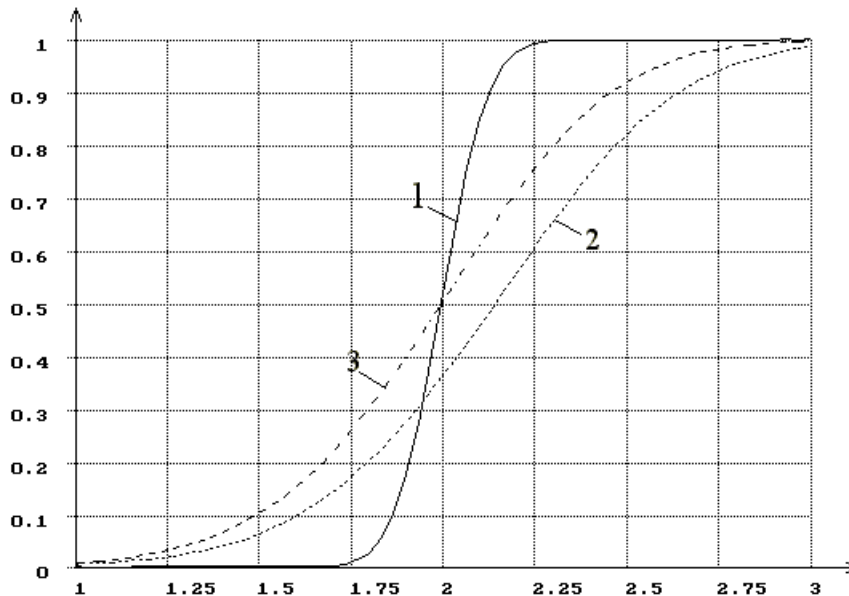


Рис. 4. Функции распределения ОМП по полным выборкам (1), по цензурированным (2) и по цензурированным с поправками (3)

Для проверки согласия при цензурированных наблюдениях и простых гипотезах могут использоваться критерии типа Реньи [4], которые в этой ситуации являются “свободными от распределения”. Очевидно, что при проверке сложных гипотез они теряют это свойство. Поэтому необходимы соответствующие исследования распределений этих статистик при проверке различных сложных гипотез с целью построения моделей предельных распределений. Проверка сложных гипотез тесно взаимосвязана с проблемой оценивания параметров. Наибольшая неприятность связана со смещенностью ОМП параметров, вычисляемых по цензурированным выборкам. В случае использования смещенных оценок задача проверки сложных гипотез с использованием критериев согласия теряет смысл. Возможность построения несмещенных оценок параметров по цензурированным выборкам, позволяет исследовать распределения статистик критериев согласия при проверке соответствующих сложных гипотез, что, в свою очередь, позволит корректно проверять сложные гипотезы относительно законов распределения по цензурированным наблюдениям.

ЛИТЕРАТУРА

1. Демидович О.Н. Особенности проверки соответствия опытного распределения теоретическому в задачах надежности // Методы менеджмента качества. – 1999. – № 11. – С. 29-33.
2. Б.Ю. Лемешко, С.Я. Гильдебрант, С.Н. Постовалов. К оцениванию параметров законов распределений по неполным выборкам // Тр. IV межд. конференции “Актуальные проблемы электронного приборостроения” АПЭП-98. Т. 3, Новосибирск, 1998. - С. 17-23.
3. Тихов М.С. Оценивание показателей качества по неполным выборкам // Надежность и контроль качества. 1996. № 11. - С. 16-24.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. -М.:Наука, 1983. - 416с.