

СРАВНЕНИЕ БУТСТРАП И АНАЛИТИЧЕСКИХ ОЦЕНОК ПОГРЕШНОСТЕЙ ПАРАМЕТРОВ ФОНОВОГО РАСПРЕДЕЛЕНИЯ ЖИТЕЛЕЙ УРАЛЬСКОГО РЕГИОНА

Ю.С. Тимофеев¹

Сегодня при анализе экспериментальных данных в различных областях знаний приобретают популярность методы ресамплинга. В частности, в эконометрике [1], экологии и биологии [2, 3] для построения интервальных оценок и анализа погрешностей активно используется бутстрап (*bootstrap*)-метод [4]. В настоящей работе приводится сравнение результатов аналитических и бутстрап-оценок погрешностей параметров распределения фоновых доз ионизирующего излучения. Результаты, полученные разными методами, хорошо согласуются между собой.

Ключевые слова: статистическая модель; ЭПР измерения; *bootstrap*.

Введение

Ионизирующее излучение индуцирует в эмали зубов человека стабильные CO_2^- радикалы. Метод дозиметрии на основе электронного парамагнитного резонанса (ЭПР) зубной эмали позволяет оценить количество этих радикалов в зубах, удаленных по медицинским показаниям [5]. Таким образом, в результате ЭПР измерения оценивается суммарная поглощенная доза радиации в эмали зубов, накопленная за время жизни донора (до момента экстракции зуба), включая воздействие как антропогенных, так и естественных источников излучения. Естественный радиационный фон представляет собой ионизирующее излучение природных источников космического и земного происхождения. В ЭПР-дозиметрии важной является оценка дозы, полученной в результате изучаемого инцидента, поэтому из общей измеренной дозы необходимо исключить фоновую составляющую.

Средние значения и распределения фоновых ЭПР-доз в разных популяциях могут отличаться [6], так как они зависят от радиационного фона конкретной местности, стиля жизни, типичного для населения региона, а также от индивидуальной вариабельности радиационной чувствительности эмали зубов в популяции [7]. Задача по выделению фонового распределения осложняется тем, что фоновые уровни доз близки к пределу детектирования ЭПР-дозиметрии [8], поэтому погрешность каждого отдельного измерения может быть сопоставима с индивидуальной вариабельностью доз в популяции (зашумленные данные).

В предыдущих исследованиях был предложен подход к выделению фонового распределения доз на основе статистического метода моментов в рамках модели, когда фоновое распределение доз имеет логнормальное распределение, а ошибка измерения – нормальное с зависимой от дозы дисперсией [9]. С его помощью были оценены параметры логнормального распределения фоновых доз сельских жителей Уральского региона: $\text{LogN}(3,9 \pm 0,5; 0,7 \pm 0,1)$, что соответствует среднему значению 61 ± 47 мГр.

С развитием компьютерной техники статистические подходы с использованием бутстрап метода приобретают все большую популярность и в некоторых, особенно в сложных, задачах конкурируют с аналитическими подходами. Целью данной работы является сравнение оценок погрешностей параметров логнормального распределения с помощью бутстрап метода и сравнение их с оценками, полученными аналитически.

Материалы и методы

Используемые в исследованиях образцы были получены от людей, для которых не велась точная история их жизни и перемещений на изучаемой территории. Поэтому нельзя исключить «внешние» в текущей задаче источники радиационного облучения, которыми являются, например, фоновое излучение на других территориях. Некоторые полученные измерения (2–3 %) превышают 500 мГр. Полагая, что к таким результатам привело какое-то внешнее облучение, такие измерения были исключены из анализа.

¹ Тимофеев Юрий Сергеевич – научный сотрудник ООО «Прикладные технологии».
E-mail: ystimofeev@gmail.com

ЭПР-измерения эмали зубов человека – это сложная и многоступенчатая процедура, на каждом шаге которой в получаемый результат привносятся ошибки. Каждая лаборатория использует свои методики измерений, поэтому ошибка измерений также зависит от лаборатории. При измерении фоновых доз получаемые значения сопоставимы с общей погрешностью измерения и большая их часть оказывается ниже предела детектирования используемого ЭПР-метода. Предел детектирования – доза, при которой на высоком уровне значимости статистическими тестами подтверждается нулевая гипотеза «образец был облучен». Были использованы 65 ЭПР-измерений фоновых доз в эмали зубов сельских жителей Уральского региона в возрасте 40–80 лет, проводившихся в Гельмгольц-Центре Мюнхена (HMGU) [10], где более 50 % измерений были выше предела детектирования.

Модель ЭПР измерений

Общая измеренная \hat{D} доза состоит из фоновой составляющей D и ошибки измерения E (1):

$$\hat{D} = D + E \quad (1)$$

Распределение фоновых доз принимается логнормальным $D = \text{Log}N[m, s]$ с неизвестными параметрами m и s , а ошибка измерения имеет нормальное распределение с неизвестным средним и известным стандартным отклонением – $N[C, g(D)]$. Форма стандартного отклонения оценивается с помощью компьютерной программы «EPR-dosimetry performance», разработанной авторами [11]. Для изучаемого метода стандартное отклонение описывается функцией (2):

$$g(D) = \begin{cases} 47; & D \leq 100, \\ 30,9 + \frac{15,8}{1 + \exp\left(\frac{D - 304}{42}\right)}; & D > 100. \end{cases} \quad (2)$$

Таким образом, окончательная модель измерений становится (3):

$$\hat{D} = \text{Log}N[m, s] + N[C, g(D)]. \quad (3)$$

Приравнивая теоретические моменты распределения (M_i) правой части уравнения (2) с их выборочными аналогами (\hat{M}_i), полученными из экспериментальных данных, с учетом трёх неизвестных параметров получаем систему из трёх уравнений (4):

$$M_i(m, s, C) = \hat{M}_i, \quad i = 1, 2, 3 \quad (4)$$

Решая эту систему, используя данные единственно удовлетворительного ЭПР метода, находим неизвестные параметры m , s и C , а также погрешности этих оценок. Более подробно метод описан в работе [9]. Полученные оценки параметров $m = 3,9$ и $s = 0,7$, с погрешностями 0,5 и 0,1 соответственно. Среднее значение фоновой дозы при этих оценках будет 61 ± 47 мГр. Учитывая, что возраст доноров на момент извлечения зуба около 60 лет, то ежегодная доза, получаемая от естественных источников радиационного излучения для жителей Уральского региона, находится на уровне 1 мГр/год.

Бутстрап подход

Статистический бутстрап (*bootstrap*) - метод определения статистик вероятностных распределений, основанный на многократной генерации выборок методом Монте-Карло на базе имеющейся выборки. Метод позволяет просто и быстро оценивать различные статистики (доверительные интервалы, дисперсию) для сложных моделей.

Пусть \hat{D}_i – измерения, полученные ЭПР-методом, $i = 1, \dots, 65$. Из имеющегося пула измерений с помощью бутстрап метода сгенерируем j новых выборок \tilde{D}_j (5) такого же размера, что и исходная выборка ($j = 1, \dots, 65$). Каждое значение генерируемой выборки равновероятно выбирается среди значений исходной выборки, при этом выбранное значение не исключается из выбора следующего значения и может быть выбрано повторно.

$$\hat{D}_i \xrightarrow{\text{bootstrap}} \tilde{D}_j^k, \quad i, j = 1, \dots, 65, \quad k = 1, \dots, 1000. \quad (5)$$

Решая систему (4) для каждой бутстрап выборки, находим новые оценки параметров m^k , s^k и их стандартные отклонения – 0,6 и 0,2 соответственно.

Заключение

Использование бутстрап подхода в текущей задаче оценке погрешностей параметров лог-нормального распределения фоновых доз сельского населения Уральского региона дает погрешности 0,6 и 0,2, которые сравнимы с оценками 0,5 и 0,1, полученными аналитически. При этом большие величины, полученные методом бутстрап, вполне ожидаемы, поскольку при выборе с возвращениями, реализуемом в бутстрапе, могут генерироваться нереалистические выборки (например, не исключена возможность получить выборку из повторяющегося одного и того же значения). Таким образом, бутстрап оценка может рассматриваться как консервативное приближение (не занижающее оценки погрешностей и ширины доверительного интервала). Учитывая вышесказанное, а также принимая во внимание, что временные затраты для поиска параметров фонового распределения для каждой сгенерированной выборки несопоставимы с затратами на получение аналитических оценок, применение бутстрап подхода представляется оправданным только в случае невозможности аналитического подхода в решении задачи.

Литература

1. Орлов, А.И. Эконометрика. Учебник / А.И. Орлов. – М.: Издательство «Экзамен», 2002. – 576 с.
2. Van Dongen, S. One and two-sample tests for single locus inbreeding coefficients using the bootstrap / S. Van Dongen, T. Backeljau // *Heredity*. – 1995. – Vol. 74. – P. 129–135.
3. Manly, B. F. J. Randomization, Bootstrap and Monte Carlo Methods in Biology. 2nd edition. / B. F. J. Manly. – Chapman and Hall, London, 1997. – 300 p.
4. Davison, A.C. Bootstrap Methods and their Application / A.C. Davison, D.V. Hinkley. – Cambridge University Press, Cambridge, 1997. – 582 p.
5. Ikeya, M. ESR dosimetry for atomic bomb survivors using shell buttons and tooth enamel / M. Ikeya, J. Miyajima, S. Okajima // *Japanese J. of Applied Physics*. – 1984. – Vol. 23. – P. 697–699.
6. Individual biodosimetry at the natural radiation background level / A.A. Romanyukha, V. Nagy, M.F. Desrosiers *et al.* // *Health Phys.* – 2001. – Vol. 80. – P. 71–73.
7. Variability of the radiation sensitivity for tooth enamel of the Ural residents / E. Shishkina, E. Tolstykh, M. Degteva *et al.* // *ANRI (Instruments and Methods of Radiat. Meas.)*. – 2012. – Vol. 69. – P. 41–50.
8. Assessment of performance parameters for EPRdosimetry with tooth enamel / A. Wieser, P. Fattibene, E.A. Shishkina *et al.* // *Radiation Measurements*. – 2008. – Vol. 43. – P. 731–736.
9. Заляпин, В.И. Статистическая реконструкция распределения фоновых доз облучения по результатам ЭПР измерений / В.И. Заляпин, Ю.С. Тимофеев, Е.А. Шишкина // *Вестник ЮУрГУ. Серия «Математика. Механика. Физика»*. – 2014. – Т. 6, № 1. – С. 22–27.
10. Shishkina, E.A. EPR dosimetry of radiation background in the Urals region / E.A. Shishkina, P. Fattibene, A. Wieser *et al.* // *Proceedings of Second European IRPA congress on radiation protection - Radiation protection: from knowledge to action*. – 2006. – CD, № TA-33. – P. 12.
11. Shishkina, E.A. Software for evaluation of EPR-dosimetry performance / E.A. Shishkina, Y.S. Timofeev, D.V. Ivanov // *Radiation Protection Dosimetry*. – 2014. – Vol. 159, no. 1–4. – P. 188–193.

Поступила в редакцию 11 декабря 2014 г.

COMPARISON OF BOOTSTRAP AND ANALYTICAL ERRORS OF ESTIMATED PARAMETERS OF BACKGROUND DISTRIBUTION OF THE POPULATION OF THE URAL REGIONYu.S. Timofeev¹

Data analysis methods based on data resampling are becoming popular today. Particularly, the bootstrap method is used in econometrics [1], ecology and biology [2, 3] to estimate such statistical values as standard deviation, confidence interval, etc [4]. The comparison of analytical and bootstrap errors of parameters of background distribution is presented in this work. Bootstrap results are in good agreement with the analytical ones.

Keywords: statistical model; EPR measurements; bootstrap.

References

1. Orlov A.I. *Ekonometrika. Uchebnik* (Econometrics. Textbook). Moscow: Izdatel'stvo "Ekzamen" Publ., 2002. 576 p. (in Russ.).
2. Van Dongen S., Backeljau T. One and two-sample tests for single locus inbreeding coefficients using the bootstrap. *Heredity*. 1995. Vol. 74. pp. 129–135. doi:10.1038/hdy.1995.19
3. Manly B.F.J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edition. Chapman and Hall, London, 1997. 300 p.
4. Davison A.C., Hinkley D.V. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, 1997. 582 p. doi: 10.1017/CBO9780511802843
5. Ikeya M., Miyajima J., Okajima S. ESR dosimetry for atomic bomb survivors using shell buttons and tooth enamel. *Japanese J. of Applied Physics*. 1984. Vol. 23. pp. 697–699.
6. Romanyukha A.A., Nagy V., Desrosiers M.F., Jiang J., Heiss A. Individual biodosimetry at the natural radiation background level. *Health Phys*. 2001. Vol. 80. pp. 71–73.
7. Shishkina E., Tolstykh E., Degteva M., Ivanov D., Aladova E. Variability of the radiation sensitivity for tooth enamel of the Ural residents. *ANRI (Instruments and Methods of Radiat. Meas.)*. 2012. Vol. 69. pp. 41–50.
8. Wieser A., Fattibene P., Shishkina E.A., Ivanov D.V., De Coste V., Güttler A., Onori S. Assessment of performance parameters for EPRdosimetry with tooth enamel. *Radiation Measurements*. 2008. Vol. 43. pp. 731–736.
9. Zalyapin V.I., Timofeev Yu.S., Shishkina E.A. Statistical Reconstruction of the Distribution of Background Doses Based on the Results of the EPR Measurements. *Bulletin of South Ural State University. Series of "Mathematics. Mechanics. Physics"*. 2014. Vol. 6, no. 1. pp. 22–27.
10. Shishkina E.A., Fattibene P., Wieser A., Degteva M.O., Onori S., Ivanov D.V., Shved V.A., Bayankin S.N., Knyazev V.A., Vasilenko E.K., Gorelov M. EPR dosimetry of radiation background in the Urals region. *Proceedings of Second European IRPA congress on radiation protection – Radiation protection: from knowledge to action*. 2006. CD, № TA-33. p. 12.
11. Shishkina E.A., Timofeev Y.S., Ivanov D.V. Software for evaluation of EPR-dosimetry performance. *Radiation Protection Dosimetry*. 2014. Vol. 159, no. 1–4. pp. 188–193. doi:10.1093/rpd/ncu167

Received 11 December 2014

¹ Timofeev Yury Sergeevich is Researcher, «Applied Technologies» Ltd.
E-mail: ystimofeev@gmail.com