РЕКОРДНЫЕ ВЕЛИЧИНЫ С ОГРАНИЧЕНИЯМИ

В. Б. Невзоров

С.-Петербургский государственный университет,
д-р физ.-мат. наук, профессор, probabil@pisem.net

1. Введение. Рассмотрим случайные величины (с.в.) $X_1, X_2, \dots$ и определим рекордные моменты $L(n)$ и рекордные величины $X(n)$, $n = 1, 2, \dots$, следующим образом:

$$L(1) = 1, \quad X(1) = 1,$$

$$L(n) = \min\{j > L(n-1) : X_j > X(n-1)\}, \\ X(n) = X_{L(n)} = \max\{X_1, X_2, \dots, X_{L(n)}\}, \quad n = 2, 3, \dots$$

Здесь мы имеем дело с так называемыми верхними рекордами. Соответственно, последовательности с.в., определяемые равенствами

$$l(1) = 1, \quad x(1) = X_1, \quad l(n) = \min\{j > l(n-1) : X_j < x(n-1)\}, \\ x(n) = X_{l(n)} = \min\{X_1, X_2, \dots, X_{l(n)}\}, \quad n = 2, 3, \dots,$$

представляют собой нижние рекордные моменты $l(n)$, $n = 1, 2, \dots$, и нижние рекордные величины $x(n)$, $n = 1, 2, \dots$.

Рассмотрение величин $Y_1 = -X_1, Y_2 = -X_2, \dots$ вместо исходных X -ов позволяет перейти от верхних рекордов к нижним. Поэтому достаточно ограничиться исследованием только верхних рекордов. Мы также ограничимся рассмотрением рекордных величин в ситуации, когда исходные с.в. имеют непрерывные функции распределения (ф.р.).

В классической хорошо разработанной (см., например, монографии [1–3]) теории рекордов рассматриваются, как правило, независимые одинаково распределенные с.в. $X_1, X_2, \dots$. Менее исследованными являются ситуации, в которых исходные с.в. могут иметь разные распределения. Здесь в первую очередь упомянем так называемые F^α -схемы [4–7], в которых ф.р. $F_k(x) = P\{X_k < x\}$ связаны соотношениями

$$F_k(x) = F^{\alpha(k)}(x), \quad k = 1, 2, \dots,$$

где $\alpha(1), \alpha(2), \dots$ — произвольные положительные константы, а $F(x)$ — некоторая ф.р.

Отметим также рекордную модель [8], в которой $X_1, X_2, \dots, X_m, X_{m+1}, \dots, X_{2m}, \dots$ представляют собой серии $X_{m(k-1)+1}, X_{m(k-1)+2}, \dots, X_{mk}$, $k = 1, 2, \dots$, независимых случайных величин с ф.р. $F_1, F_2, \dots, F_m$. В этой рекордной схеме предполагается, что

$$F_r(x) = P\{X_{mk+r} < x\}, \quad k = 0, 1, 2, \dots,$$

для каждого $r = 1, 2, \dots, m$.

В приведенных выше схемах уход от «классической» модели был связан с отказом от одинакового распределения исходных X -ов. В то же время появились рекордные схемы, в которых $X_1, X_2, \dots$, как и раньше, были независимыми и одинаково

распределенными, но уже сами рекордные величины определялись «неклассически». Рассмотрим ниже ряд таких схем.

2. Рекорды с превышением. В 1996 г. в работе [9] были введены так называемые рекорды с δ -превышением (δ -exceedance records). В этой схеме для объявления очередного наблюдения рекордным требовалось, чтобы оно превышало предыдущее рекордное достижение более чем на некоторое значение δ . В указанной работе фиксировался некоторый порог превышения $\delta > 0$. Далее для последовательности независимых с.в. $X_1, X_2, \dots$, имеющих одинаковую плотность распределения $f(x)$, рекордные моменты $T(n)$ и рекордные величины $X(n)$ определялись следующим образом:

$$T(0) = 1, X(0) = X_1,$$

$$T(n) = \min\{j > T(n-1) : X_j > X_{T(n-1)} + \delta\}, \quad X(n) = X_{T(n)}, \quad n = 1, 2, \dots$$

Было показано, в частности, что совместная плотность распределения $f(x_0, x_1, \dots, x_n)$ рекордных величин $X(0), X(1), \dots, X(n)$ задается следующими равенствами:

$$f(x_0, x_1, \dots, x_n) = \frac{f(x_0)}{1 - F(x_0 + \delta)} \cdots \frac{f(x_{n-1})}{1 - F(x_{n-1} + \delta)} f(x_n), \quad (1)$$

если $x_j > x_{j-1} + \delta$, $j = 1, 2, \dots, n$, и $f(x_0, x_1, \dots, x_n) = 0$ иначе.

Естественно, что при $\delta = 0$ равенство (1) совпадает с соответствующим выражением, полученным для классических рекордных величин. При $\delta > 0$ равенство (1) становится не очень удобным для работы с ним. Однако все существенно упрощается в ситуации, когда исходные с.в. имеют экспоненциальное распределение. В частности, если $X_1, X_2, \dots$ — независимые с.в. с плотностью распределения $f(x) = \exp(-x)$, $x \geq 0$, то основной результат статьи [9] можно переформулировать следующим образом: для любого $n = 1, 2, \dots$ распределение вектора $\{X(0), X(1), \dots, X(n)\}$ совпадает с распределением вектора

$$\{\xi_0, \xi_0 + \xi_1 + \delta, \dots, \xi_0 + \xi_1 + \dots + \xi_n + n\delta\},$$

где $\xi_0, \xi_1, \dots$ независимы и также имеют экспоненциальное распределение с плотностью $f(x) = \exp(-x)$, $x \geq 0$. Отсюда следует, что $X(n)$ при любом $n = 1, 2, \dots$ совпадает по распределению с суммой $\xi_0 + \xi_1 + \dots + \xi_n + n\delta$, т. е. $X(n)$ имеет Гамма-распределение с плотностью

$$f_n(x) = \exp\{-(x - n\delta)\} \frac{(x - n\delta)^n}{n!}, \quad x \geq n\delta. \quad (2)$$

Видим также, что межрекордные величины $X(0), X(1) - X(0), \dots, X(n) - X(n-1)$ в этом случае являются независимыми.

Данная схема предполагает, что не рассматриваются в качестве рекордных небольшие (меньшие δ) превышения предыдущих наблюдений. Такого рода модель выглядит достаточно естественной в ситуации, когда, например, не может быть достигнута высокая точность измерения результатов. Скажем, в марафонском беге (в отличие от спринтерских дистанций) в качестве нового рекордного может быть засчитан лишь результат, который лучше предыдущего рекорда не менее чем на одну секунду (десятыи и тем более сотые доли секунды здесь в расчет не берутся).

3. Рекорды с ограничением. I. В связи с упомянутой выше схемой достаточно естественным дополнением к ней является предлагаемая ниже модель рекордов с ограничениями. В этой рекордной модели игнорируются все наблюдения, которые превышают последнее рекордное значение более чем на некоторую фиксированную положительную константу C . Если вернуться к результатам спортивных соревнований, то там слишком существенные превышения предыдущих рекордов могут просто вызвать опасения, что данный результат достигнут нечестными путями. Например, могли быть использованы нестандартные спортивные снаряды или спортсменами употреблялись запрещенные препараты.

Как и в схеме рекордов с δ -превышением, совместные распределения рекордов с ограничением в общем случае достаточно громоздки, но для экспоненциально распределенных X -ов удается существенно упростить полученные результаты. Итак, рассматриваем последовательность независимых одинаково распределенных с.в. $X_1, X_2, \dots$ и фиксируем некоторую положительную константу C . В качестве первой рекордной величины и первого рекордного момента берем $X(1) = X_1$ и $L(1) = 1$. Последующие рекордные величины $X(n)$ и рекордные моменты $L(n)$ получаем последовательно следующим образом:

$$\begin{aligned} L(n) &= \min\{j > L(n-1) : X(n-1) < X_j \leq X(n-1) + C\}, \\ X(n) &= X_{L(n)}, n = 2, 3, \dots \end{aligned} \quad (3)$$

Предполагаем, что исходные с.в. имеют некоторую плотность распределения $p(x)$. Сравнительно просто получить выражение для условной плотности распределения $p_n(x_n | x_1, x_2, \dots, x_{n-1})$, т. е. для плотности распределения рекордной величины $X(n)$ при условии, что зафиксированы значения

$$X(1) = x_1, \quad X(2) = x_2, \dots, \quad X(n-1) = x_{n-1},$$

где $0 < X_j - X_{j-1} \leq C, j = 2, 3, \dots, n-1$. Несложные рассуждения приводят к равенству

$$p_n(x_n | x_1, x_2, \dots, x_{n-1}) = \frac{p(x_n)}{F(x_{n-1} + C) - F(x_{n-1})}, \quad (4)$$

если $x_{n-1} < x_n \leq x_{n-1} + C$, и $p_n(x_n | x_1, x_2, \dots, x_{n-1}) = 0$ иначе. Отсюда следует, что плотности распределения соседних рекордных величин $X(n-1)$ и $X(n)$ связаны соотношением

$$p_n(x) = p(x) \int_{x-C}^x \left(\frac{p_{n-1}(u)}{F(u+C) - F(u)} \right) du, \quad n = 2, 3, \dots \quad (5)$$

Последовательно применяя равенство (5) при $n = 2, 3, \dots$, получаем, что совместная плотность распределения $p_n(x_1, x_2, \dots, x_n)$ рекордных величин $X(1), X(2), \dots, X(n)$ имеет следующий вид:

$$p_n(x_1, x_2, \dots, x_n) = p(x_1) \frac{p(x_2)}{F(x_1 + C) - F(x_1)} \dots \frac{p(x_n)}{F(x_{n-1} + C) - F(x_{n-1})}, \quad (6)$$

если $x_1 < x_2 \leq x_1 + C, x_2 < x_3 \leq x_2 + C, \dots, x_{n-1} < x_n \leq x_{n-1} + C$, и $p_n(x_1, x_2, \dots, x_n) = 0$ иначе.

Когда имеем дело со случайными величинами, множество значений которых ограничено снизу некоторым $a > -\infty$, удобнее немного изменить предложенную схему, введя дополнительно $X(0) = a, L(0) = 0$ и определяя $X(1)$ и $L(1)$ равенством

(3), выполняемым при $n = 1$. В этом случае совместная плотность распределения с.в. $X(1), X(2), \dots, X(n)$ примет следующий вид (немного отличающийся от равенства (6)):

$$p_n(x_1, x_2, \dots, x_n) = \frac{p(x_1)}{F(a+C)} \frac{p(x_2)}{F(x_1+C) - F(x_1)} \cdots \frac{p(x_n)}{F(x_{n-1}+C) - F(x_{n-1})} \cdots, \quad (7)$$

если $a < x_1 \leq a + C$, $x_2 < x_3 \leq x_2 + C$, $\dots$, $x_{n-1} < x_n \leq x_{n-1} + C$.

4. Рекорды с ограничением для экспоненциально распределенных величин. Полученные выше общие соотношения существенно упрощаются, если исходные случайные величины имеют экспоненциальное распределение. Рассмотрим этот частный случай. Для удобства введем следующие обозначения. Пусть $Z_0 = 0$ и $Z_1, Z_2, \dots$ — независимые с.в., имеющие экспоненциальное распределение с ф.р. $F(x) = \max\{0, 1 - \exp(-x)\}$. Соответствующие рекордные (с ограничением $C > 0$) величины и рекордные моменты обозначим как

$$0 = Z(0) < Z(1) < Z(2) < \dots \quad \text{и} \quad 0 = T(0) < T(1) < T(2) < \dots$$

Из равенства (7) при $a = 0$ получаем, что для этого частного случая совместная плотность распределения $h(x_1, x_2, \dots, x_n)$ с.в. $Z(1), Z(2), \dots, Z(n)$ определяется соотношением

$$h(x_1, x_2, \dots, x_n) = \frac{\exp\{-x_n\}}{(1 - \exp\{-C\})^n}, \quad (8)$$

если $0 < x_1 \leq C$, $x_1 < x_2 \leq x_1 + C$, $\dots$, $x_{n-1} < x_n \leq x_{n-1} + C$, и $h(x_1, x_2, \dots, x_n) = 0$ иначе. Из равенства (8) получаем, что совместная плотность $g(v_1, v_2, \dots, v_n)$ разностей

$$V(1) = Z(1) - Z(0), \quad V(2) = Z(2) - Z(1), \dots, \quad V(n) = Z(n) - Z(n-1)$$

задается равенством

$$g(v_1, v_2, \dots, v_n) = \frac{\exp\{-(v_1 + v_2 + \dots + v_n)\}}{(1 - \exp\{-C\})^n}, \quad 0 \leq v_1 \leq C, \dots, \quad 0 \leq v_n \leq C. \quad (9)$$

Принимая во внимание соотношение (9), приходим к утверждению, что межрекордные разности $V(1), V(2), \dots, V(n)$ независимы и имеют одинаковую плотность распределения вида

$$g(x) = \frac{\exp\{-x\}}{1 - \exp\{-C\}}, \quad (10)$$

если $0 \leq x \leq C$, и $g(x) = 0$ иначе. Отсюда получаем результат, заключающийся в том, что одинаково распределенными являются случайные векторы

$$\{Z(1), Z(2), \dots, Z(n)\} \quad \text{и} \quad \{W_1, W_1 + W_2, \dots, W_1 + W_2 + \dots + W_n\},$$

где независимые с.в. $W_1, W_2, \dots, W_n$ имеют плотность распределения (10). Таким образом, рекордная величина $Z(n)$ представима в виде суммы n независимых одинаково распределенных случайных величин. Отметим, что математическое ожидание и дисперсия величин W_k , задаются равенствами

$$a(c) = EW_k = \frac{1 - (1 + C) \exp\{-C\}}{1 - \exp\{-C\}}$$

и

$$\sigma^2(c) = DV_k = \frac{1 - (2 + C^2) \exp\{-C\} + \exp\{-2C\}}{(1 - \exp\{-C\})^2}, \quad k = 1, 2, \dots$$

Учитывая приведенные факты, можно получить для $Z(n)$ ряд предельных теорем, справедливых для сумм независимых случайных величин. В частности, сразу получаем асимптотическую нормальность с.в. $(Z(n) - na(c))/\sigma(c)n^{1/2}$.

Естественным образом результаты, сформулированные выше для случайных величин Z_1, Z_2 , имеющих стандартное экспоненциальное распределение, переносятся на случай, когда эти величины имеют более общие экспоненциальные плотности распределения вида $\lambda \exp\{-\lambda(x - \alpha)\}$, $x \geq \alpha$, где $\lambda > 0$ и $-\infty < \alpha < \infty$.

5. Рекорды с ограничением. II. Близкой к описанной выше рекордной модели является схема, в которой также зафиксировано некоторое значение $C > 0$, определяющее допустимый порог превышения предыдущего рекорда. Но в этом случае наблюдение, превышающее более чем на C существующий рекорд $X(n)$, не теряется, и очередной рекордной величиной объявляется $X(n+1) = X(n) + C$. Если же новое наблюдение попадает в интервал $(X(n), X(n) + C]$, то оно и становится рекордной величиной $X(n+1)$. В частности, если мы вернемся к исходным независимым экспоненциальным (с плотностью распределения $p(x) = \exp\{-x\}$, $x \geq 0$) с.в. $Z_1, Z_2, \dots$ и оставим обозначения $Z(1) < Z(2) < \dots$ для видоизмененных таким образом рекордных величин, то рассуждения, аналогичные приведенным выше, позволяют установить, что разности $V_k = Z(k) - Z(k-1)$, $k = 1, 2, \dots$, как и в предыдущей ситуации, независимы, но только в этом случае

$$P\{V_k < x\} = 1 - \exp\{-x\}, \quad \text{если } 0 \leq x \leq C, \quad \text{и } P\{V_k < x\} = 1, \quad \text{если } x > C. \quad (11)$$

Это означает, в частности, что

$$P\{V_k = C\} = \exp\{-C\}, \quad k = 1, 2, \dots$$

Видим, что для данной рекордной схемы и любого $n = 1, 2, \dots$ распределение вектора $\{Z(1), Z(2), \dots, Z(n)\}$ совпадает с распределением вектора $\{v_1, v_1 + v_2, \dots, v_1 + v_2 + \dots + v_n\}$, где $v_1, v_2, \dots$ — независимые с.в. с функцией распределения (11).

Литература

1. *Ahsanullah M.* Record values. New York: Nova Science Publishers, 1995.
2. *Arnold B. C., Balakrishnan N., Nagaraja H. N.* Records. New York: John Wiley and Sons, 1998. 312 p.
3. *Невзоров В. Б.* Рекорды. Математическая теория. М.: Фазис, 2000. 244 с.
4. *Yang M. C. K.* On the distribution of the inter-record times in an increasing population // Journal of Applied Probability. 1975. Т. 12. Р. 148–154.
5. *Деовельс П., Невзоров В. Б.* Рекорды в F^α -схеме. I. Мартингалльные свойства // Записки научных семинаров ПОМИ. 1993. Т. 207. С. 19–36.
6. *Деовельс П., Невзоров В. Б.* Рекорды в F^α -схеме. II. Предельные теоремы // Записки научных семинаров ПОМИ. 1994. Т. 216. С. 42–51.
7. *Shakil M., Ahsanullah M.* Review on order statistics and record values from F^α -distributions. Pakistan Journal of Statistics and Operation Research. 2008. Vol. VIII, N 1. P. 101–120.
8. *Невзоров В. Б.* О среднем числе рекордов в последовательностях неодинаково распределенных случайных величин // Вестн. С.-Петербург. ун-та, 2012. Т. 45, № 4. С. 28–32.
9. *Balakrishnan N., Balasubramanian K., Panchapakesan S.* δ -exceedance records // Journal of Applied Statistical Science. 1996. Vol. 4. P. 123–132.

Статья поступила в редакцию 28 марта 2013 г.