Последовательности выпадения орлов

Предположим, что правильная (т.е. выпадение орла и решки равновероятны) монета подбрасывается n раз. Какого количества последовательных выпадений орла можно ожидать? Как покажет последующий анализ, эта величина ведет себя как $\Theta(\lg n)$. Докажем сначала, что математическое ожидание длины наибольшей последовательности орлов представляет собой $\Theta(\lg n)$. Вероятность того, что при очередном подбрасывании выпадет орел, равна $1/2$. Пусть A_{ik} - событие, когда последовательность выпадений орлов длиной не менее k начинается с i-го подбрасывания, или, более строго, A_{ik} - событие, когда при k последовательных подбрасываниях монеты i, $i+1, ..., i+k-1$ (где $1 \le k \le n$ и $1 \le i \le n-k+1$) будут выпадать одни орлы. Поскольку подбрасывания монеты осуществляются независимо, для каждого данного события A_{ik} вероятность того, что во всех k подбрасываниях выпадут одни орлы, определяется следующим образом:

$$Pr\{A_{ik}\} = 1/2^k. \tag{5.8}$$

Для $k = 2\lceil \lg n \rceil$

$$\begin{aligned} Pr\left\{A_{i,2\lceil \lg n \rceil}\right\} &= 1/2^{\lceil \lg n \rceil} \\ &\le 1/2^{\lg n} \\ &= 1/n^2, \end{aligned}$$

так что вероятность того, что последовательность повторных выпадений орлов длиной не менее $2\lceil \lg n \rceil$ начинается с i-го подбрасывания, довольно невелика. Имеется не более $n - 2\lceil \lg n \rceil + 1$ подбрасываний, с которых может начаться указанная последовательность орлов. Таким образом, вероятность того, что последовательность повторных выпадений орлов длиной не менее $2\lceil \lg n \rceil$ начинается при произвольном подбрасывании, равна

$$\begin{aligned} Pr\left\{\bigcup_{i=1}^{n-2\lceil \lg n \rceil+1} A_{i,2\lceil \lg n \rceil}\right\} &\le \sum_{i=1}^{n-2\lceil \lg n \rceil+1} 1/n^2 \\ &\le \sum_{i=1}^{n} 1/n^2 \\ &= 1/n. \end{aligned} \tag{5.9}$$

Справедливость этого соотношения следует из неравенства Буля (B.19), согласно которому вероятность объединения событий не превышает сумму вероятностей отдельных событий. (Заметим, что неравенство Буля выполняется даже для тех событий, которые не являются независимыми.)

Теперь воспользуемся неравенством (5.9) для ограничения длины самой длинной последовательности выпадения орлов. Пусть $L_j (j = 0, 1, 2, ..., n)$ - событие, когда длина самой длинной последовательности выпадения орлов равна j. В соответствии с определением математического ожидания мы имеем

$$E[L] = \sum_{j=0}^{n} jPr\{L_j\}. \tag{5.10}$$

Можно попытаться оценить эту сумму с помощью верхних границ каждой из величин $Pr\{L_j\}$, аналогично тому, как это было сделано в неравенстве (5.9). К сожалению, этот метод не может обеспечить хороших оценок. Однако достаточно точную оценку можно получить с помощью некоторых интуитивных рассуждений, которые вытекают из проведенного выше анализа. Присмотревшись внимательнее, можно заметить, что в сумме (5.10) нет ни одного слагаемого, в котором оба множителя j и $Pr\{L_j\}$ были бы большими. Почему? При $j \geq 2\lceil\lg n\rceil$ величина $Pr\{L_j\}$ очень мала, а при $j < 2\lceil\lg n\rceil$ оказывается невелико само значение j. Выражаясь более формально, можно заметить, что события L_j для $j = 0, 1..., n$ несовместимы, поэтому вероятность того, что непрерывная последовательность выпадения орлов длиной не менее $2\lceil\lg n\rceil$ начинается с любого подбрасывания монеты, равна $\sum_{j=2\lceil\lg n\rceil}^{n} Pr\{L_j\}$. Согласно неравенству имеем $\sum_{j=2\lceil\lg n\rceil}^{n} Pr\{L_j\} < 1/n$. Кроме того, из $\sum_{j=0}^{n} Pr\{L_j\} = 1$ вытекает $\sum_{j=0}^{2\lceil\lg n\rceil-1} Pr\{L_j\} \leq 1$. Таким образом, получаем

$$\begin{aligned}
E[L] &= \sum_{j=0}^{n} jPr\{L_j\} \\
&= \sum_{j=0}^{2\lceil\lg n\rceil-1} jPr\{L_j\} + \sum_{j=2\lceil\lg n\rceil}^{n} jPr\{L_j\} \\
&< \sum_{j=0}^{2\lceil\lg n\rceil-1} (2\lceil\lg n\rceil)Pr\{L_j\} + \sum_{j=2\lceil\lg n\rceil}^{n} nPr\{L_j\} \\
&= 2\lceil\lg n\rceil \sum_{j=0}^{2\lceil\lg n\rceil-1} Pr\{L_j\} + \sum_{j=2\lceil\lg n\rceil}^{n} Pr\{L_j\} \\
&< 2\lceil\lg n\rceil \cdot 1 + n \cdot (1/n) \\
&= O(\lg n).
\end{aligned}$$

Вероятность того, что длина последовательности непрерывных выпадений орла превысит величину $r\lceil \lg n\rceil$, быстро убывает с ростом r. Для $r \geq 1$ вероятность того, что последовательность как минимум $r\lceil \lg n\rceil$ выпадений орлов начнется с i-го подбрасывания, равна

$$\begin{aligned} Pr\left\{A_{i,r\lceil \lg n\rceil}\right\} &= 1/2^{r\lceil \lg n\rceil} \\ &\leq 1/n^r. \end{aligned}$$

Таким образом, вероятность образования непрерывной цепочки из последовательных выпадений орла, имеющей длину не менее $r\lceil \lg n\rceil$, не превышает $n/n^r = 1/n^{r-1}$.Это утверждение эквивалентно утверждению, что длина такой цепочки меньше величины $r\lceil \lg n\rceil$ с вероятностью не менее чем $1 - 1/n^{r-1}$.