A supplementary material for the lecture

Tomonari SEI

Apr 13, 2017

1 A problem mentioned in the lecture and its solution

Problem 1. There are N students in the class. Each student tossed a coin repeatedly until the tail appears. For the i-th student ($1 \leq i \leq N$), let T_i be the number of consecutive heads. In other words,

$$T_i = \max\{t \mid X_i(s) = 1,\ 1 \leq s \leq t\},$$

where $X_i(s)$ denotes the result of the coin toss: $X_i(s) = 1$ if the s-th outcome was head and 0 if it was tail. Let $N(t)$ be the number of students still alive at time t, that is,

$$N(t) = \sharp\{i \mid T_i \geq t\}.$$

Define the random variable R by

$$R = \max\{t \mid N(t) > 0\}.$$

Then find the probability mass function of R.

Answer. By definition, we have

$$R = \max(T_1, \ldots, T_N). \tag{1}$$

Indeed, $R = r$ for $r \in \mathbb{Z}_+$ means $N(r) > 0$ and $N(r+1) = 0$, which is equivalent to $\max(T_1, \ldots, T_N) = r$. Then, by independence of T_i's, we have

$$\begin{aligned}
P(R \leq r) &= P(T_1 \leq r, \ldots, T_N \leq r) \\
&= \prod_{i=1}^{N} P(T_i \leq r) \\
&= \left(1 - (1/2)^{r+1}\right)^N,
\end{aligned}$$

where the last equality follows from

$$\begin{aligned} P(T_i \le r) &= 1 - P(T_i \ge r+1) \\ &= 1 - P(X_i(1) = \cdots = X_i(r+1) = 1) \\ &= 1 - (1/2)^{r+1} \quad \text{(geometric distribution)}. \end{aligned}$$

We finally obtain

$$\begin{aligned} P(R = r) &= P(R \le r) - P(R \le r-1) \\ &= \{1 - (1/2)^{r+1}\}^N - \{1 - (1/2)^r\}^N. \end{aligned}$$

Note that the probability of $R = 0$ is $(1/2)^N$ as you can easily check. □

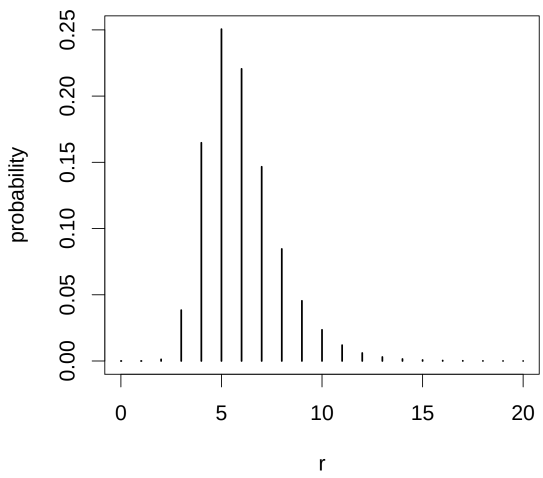
The problem above was essentially to obtain the distribution of maximum of random variables. See Eq. (1).

In the lecture, the result[*1] of the experiment was $R = 6$ with $N = 50$. The upper probability is

$$\begin{aligned} P(R \ge 6) &= 1 - (1 - 1/64)^{50} \\ &= 0.5450. \end{aligned}$$

One may deduce that $R = 6$ is not a surprising result.

The following figure shows the probability mass function (histogram) of R.



It is natural to ask the following problems.

[*1] I cannot remember exactly. Was it $R = 7$?

Problem 2. The coin is now changed so that $P(X_i = 1) = p$ and $P(X_i = 0) = q = 1 - p$. Then find the probability mass function of R.

Answer. The mass function is similarly obtained as Problem 1. The result is

$$P(R = r) = (1 - p^{r+1})^N - (1 - p^r)^N.$$

since $P(T_i \geq r + 1) = p^{r+1}$. $\square$

Problem 3. What happens if $N \to \infty$?

Answer. For simplicity, consider the case $p = 1/2$ again. We focus on

$$P(R < r) = (1 - (1/2)^r)^N.$$

It will be of interest to find r such that $(1 - (1/2)^r)^N$ does not degenerate as $N \to \infty$. This will be the case if $(1/2)^r = O(1/N)$. So we put $r = (x + \log N)/\log 2$, which is equivalent to $(1/2)^r = e^{-x}/N$. Then we have

$$\begin{aligned} P(R < (x + \log N)/\log 2) &= \left(1 - \frac{e^{-x}}{N}\right)^N \\ &\simeq e^{-e^{-x}}. \end{aligned}$$

The function $F(x) = e^{-e^{-x}}$ is actually a cumulative distribution since

$$\frac{dF(x)}{dx} = e^{-x}e^{-e^{-x}} > 0,$$

$F(-\infty) = e^{-\infty} = 0$, and $F(\infty) = e^0 = 1$. Here we allowed x takes any real numbers although it should take discrete values since $x = (\log 2)r - \log N$. The distribution $F(x)$ is known as *the Gumbel distribution*, one of the generalized extreme value distributions. $\square$