APPROXIMATIONS TO THE BIRTHDAY PROBLEM WITH UNEQUAL OCCURRENCE PROBABILITIES AND THEIR APPLICATION TO THE SURNAME PROBLEM IN JAPAN

SHIGERU MASE

Faculty of Integrated Arts and Sciences, Hiroshima University, Naka-ku, Hiroshima 730, Japan

(Received July 4, 1990; revised September 2, 1991)

Abstract. Let $X_1, X_2, \ldots, X_n$ be iid random variables with a discrete distribution $\{p_i\}_{i=1}^m$. We will discuss the coincidence probability R_n, i.e., the probability that there are members of $\{X_i\}$ having the same value. If $m = 365$ and $p_i \equiv 1/365$, this is the famous birthday problem. Also we will give two kinds of approximation to this probability. Finally we will give two applications. The first is the estimation of the coincidence probability of surnames in Japan. For this purpose, we will fit a generalized zeta distribution to a frequency data of surnames in Japan. The second is the *true* birthday problem, that is, we will evaluate the birthday probability in Japan using the actual (non-uniform) distribution of birthdays in Japan.

Key words and phrases: Birthday problem, coincidence probability, non-uniformness, Bell polynomial, approximation, surname.