INTERNATIONAL CENTRE OF EXCELLENCE FOR EDUCATION IN MATHEMATICS

Zipf's Law and contiguous topics in probability theory

Andrew Kwok, Department of Mathematics and Statistics, University of Melbourne

My research concerned the study of Zipf's Law in the area of stochastic modeling and probability theory.

Zipf's Law is the observation that the frequency of an event occurring is inversely proportional to its ranking, where the most frequent event has ranking 1 etc. This phenomenological result is somewhat pervasive, applying to areas including word frequency in texts, the popularity of websites, the distribution of income and the population of cities.

For example, if we rank the world's largest city populations in order and plot the logarithm of the population against the logarithm of the rank, we notice an interesting result: the graph we obtain is almost a straight line. This is what Zipf's Law says should happen.

My research project looked at the areas of language and DNA. It can be shown that both proper text and randomly generated text follow a generalised version of Zipf's Law though the gradient of the line may differ in each case. In other words, the text that conveys much information may yield a different gradient to the text that contains little if any information.

DNA also obeys a generalised Zipf's Law. An interesting idea is to compare the gradient of DNA's graph to those of proper and random text. How close the values are to each other might suggest how much information is conveyed by DNA. This is another area I pursued in my project and I hope to continue this research in the future.

Andrew received an ICE-EM Vacation Scholarship in January 2005.
See www.ice-em.org.au/students.html#scholarship.