OXWASP PROBABILITY AND APPROXIMATION MODULE MICRO-PROJECT: ESTIMATING MAXIMUM OCCUPANCY

Suppose there are a large number N of individuals who come to use a facility, such as a public park. People arrive at random times, and stay a random length of time before departing, and everyone is gone before nightfall. We assume individuals' arrival and departure times are independent. We are interested in understanding the maximum occupancy over the course of the day.

One approach is to consider the number of individuals $Y(t)$ in the park at time t as

$$Y(t) = \sum_{i=1}^{N} X_i(t), \text{ where } X_i(t) = \begin{cases} 1 & \text{if individual } i \text{ present at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Assume the functions $Y_i(t)$ are i.i.d. (Does this constrain the model substantially? Explain how. Consider separately the assumptions of independence and identical distribution.) Assume for simplicity that t is on the interval $[0, 1]$.

Possible questions to consider:

(1) Find a limit theorem and tail bound for $Y(t)$ for each t.

(2) How can you turn this into a limit theorem and tail bound for $\max 0 \le t \le 1 Y(t)$? How is it related to the maximum expectated value of $Y_i(t)$?

(3) The random fluctuations are on the order of $N^{1/2}$. If $y(t) := \mathbb{E}[Y_i(t)]$ has its maximum at $t = t_0$, then the $\max Y(t)$ will be at least as large as $Y(t_0)$. How much larger? Try to find the appropriate α such that it is on the order of N^α for large N. (Hint: α is smaller than $\frac{1}{2}$. Near t_0 you have a random Gaussian of one order in N competing with a falling quadratic of another order.)

(4) Taking the above further, you can look up the Functional Central Limit Theorem to understand how $N^{-1/2}(Y(t) - \mathbb{E}Y(t))$ converges to a Brownian motion — a random continuous Gaussian function — and thus derive a more precise limit theorem for the maximum. (For the actual distribution that this leads to, you might have a look at [**?**]. It's mathematically beautiful, but not pretty...)

(5) Test your results with simulations — or use simulations to collect evidence and generate hypotheses if you're stuck. How good are your tail bounds relative to the real errors?

A good overview of the chaining and symmetrisation techniques may be found in [2], available at

`http://www.stat.yale.edu/~pollard/Books/Iowa/Iowa-notes.pdf` .
For more mathematically extensive treatment of these issues see [1].

References

[1] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Springer-Verlag, New York, Heidelberg, Berlin, 1991.

[2] David Pollard. *Empirical Processes: Theory and Applications*, volume 2 of *CBMS-NSF Regional Conference Series in Probability and Statistics*. Institute of Mathematical, Hayward, California, 1991.