Lecture 19

David Aldous

12 October 2015

Today's topics

- 2 examples using "marked PPPs" theory from last class.
- Examples of informal Poisson approximation for approximately independent events.

[From last class]
Suppose we have a rate-λ PPP of times of events $0 < W_1 < W_2 < \ldots$. Suppose that associated with the i'th event is a $\mathbb{R}$-valued random variable Y_i, where $(Y_1, Y_2, \ldots)$ are IID with density $g(y)$, independent of (W_i). Then we can regard the points (W_1, Y_1), (W_2, Y_2), $\ldots$ as a point process on $[0, \infty) \times (-\infty, \infty)$.

Theorem (PK Theorem 5.8)

The points (W_1, Y_1), (W_2, Y_2), $\ldots$ form a Poisson PP with rate $\lambda(t, y) = \lambda g(y)$.

- -

Question; what is the distribution of $Z = \min(W_1 + Y_1, W_2 + Y_2, \ldots)$?
Answer [board]

$$\mathbb{P}(Z > z) = \exp\left(-\lambda \int_0^z G(s)ds\right), \quad G(\cdot) = \text{ distribution function of } Y.$$

[PK] page 370 gives a "crack failure" story for this question.

Question: Take a rate-λ PPP on the plane $\mathbb{R}^2$. For each point of the PPP, draw a circle of radius R, where $R > 0$ is random with density function $f(r)$, independent for different points. What is the distribution of $Q =$ number of circles covering the origin?

Answer [board]
Q has Poisson distribution with mean $\lambda \int_0^\infty \pi r^2 f(r) dr$.

This is [PK] problem 5.5.7.

Some models of coincidences.

Almost every textbook and popular science account of probability discusses the *birthday problem*, and the conclusion

> *with 23 people in a room, there is roughly a 50% chance that some two will have the same birthday.*

And it's easy to check this prediction with real data, for instance from MLB active rosters, which conveniently have 25 players and their birth dates.

[show]

The predicted chance of a birthday coincidence is about 57%. With 30 MLB teams one expects around 17 teams to have the coincidence – can check in freshman seminar course.

Mathematicians have put great ingenuity into finding exact formulas, but it's simpler and more broadly useful to use approximate ones, based on the informal Poisson approximation. If events $A_1, A_2, \ldots$ are roughly independent, and each has small probability, then the random number that occur has mean (exactly) $\mu = \sum_i \mathbb{P}(A_i)$ and distribution (approximately) Poisson(μ), so

$$\mathbb{P}(\text{none of the events occur}) \approx \exp\left(-\sum_i P(A_i)\right). \qquad (1)$$

So if we list all possible coincidences in our model as $A_1, A_2, \ldots$ then

$$\mathbb{P}(\text{at least one coincidence occurs}) \approx 1 - \exp\left(-\sum_i \mathbb{P}(A_i)\right).$$

For the usual birthday problem, people often ask whether the fact that birthdays are not distributed exactly uniformly over the year makes any difference. So let's consider k people and non-uniform distribution

$$p_i = \mathbb{P}(\text{born of day } i \text{ of the year}).$$

For each *pair* of people, the chance they have the same birthday is $\sum_i p_i^2$, and there are $\binom{k}{2}$ pairs, so from (1)

$$\mathbb{P}(\text{no birthday coincidence}) \approx \exp\left(-\binom{k}{2}\sum_i p_i^2\right).$$

Write *median-k* for the value of k that makes this probability close to 1/2 (and therefore makes the chance there *is* a coincidence close to 1/2). We calculate [board]

$$\text{median-}k \approx \tfrac{1}{2} + \frac{1.18}{\sqrt{\sum_i p_i^2}}.$$

For the uniform distribution over N categories this becomes

$$\text{median-}k \approx \tfrac{1}{2} + 1.18\sqrt{N}$$

which for $N = 365$ gives the familiar answer 23.

$$\text{median-}k \approx \frac{1}{2} + \frac{1.18}{\sqrt{\sum_i p_i^2}}.$$

- -

To illustrate the non-uniform case, imagine hypothetically that there were twice as many births per day in one half of the year as in the other half, so $p_i = \frac{4}{3N}$ or $\frac{2}{3N}$. The approximation becomes $\frac{1}{2} + 1.12\sqrt{N}$ which for $N = 365$ becomes 22.

The smallness of the change ("robustness to non-uniformity") is in fact **not** typical of combinatorial problems in general. In the coupon collector's problem, for instance, the change would be much more noticeable.

Here are two variants.
(**1.**) If we ask for the coincidence of *three* people having the same birthday, then we can repeat the argument above to get

$$\mathbb{P}(\text{no three-person birthday coincidence}) \approx \exp\left(-\binom{k}{3}\sum_i p_i^3\right)$$

and then in the uniform case,

$$\text{median-}k \approx 1 + 1.61 N^{2/3}$$

which for $N = 365$ gives the less familiar answer 83.

(**2.**) If instead of calendar days we have k events at independent uniform times during a year, and regard a coincidence as seeing two of these events within 24 hours (not necessarily the same calendar day), then the chance that a particular two events are within 24 hours is $2/N$ for $N = 365$, and we can repeat the calculation for the birthday problem to get

$$\text{median-}k \approx \tfrac{1}{2} + 1.18\sqrt{N/2} \approx 16.$$