this time: continuous rvs; joint, marginal & conditional distributions

Read: DS ch. 3, 4

AMS 131 9 Aug 17

(doc. com. notes) ①

next time: functions of rvs; expected value

Case Study: the Poisson process (DS p. 294)

You work for a bank that has lately been receiving complaints about long waiting times in the bank teller lines. To quantify the extent of the problem you gather a dataset like the one at left, on Mon, Tue, ..., Sat in a randomly chosen week. This is a new type of dataset for us: it unfolds in time.

time	event
9 AM	bank opens
9.01am	arrival
9.03am	arrival
9.04am	arrival
9.05am	departure
⋮	⋮

There are two equivalent ways to keep track of the arrivals: you could let $N(t) = \#$ arrivals in $[0, t]$, or you could keep track of the inter-arrival times (times between arrivals) $T_1, T_2, \ldots$ ②

Here we'll look at $N(t)$. Definition:

A stochastic process is a collection of random variables indexed by elements of an indexing set, usually $t \in T$, where t represents time. $N(t)$, $t \in [0, t_{max}]$ (9AM; e.g. 5pm)

arrivals in $[0, t]$ is an example of a stochastic process, in which the time index is continuous; $\{T_1, T_2, \ldots\}$ (interarrival times) is an example of a stochastic process with discrete time index. $(T_t, t = 1, 2, \ldots)$

What may reasonably be assumed about the random behavior of $N(t)$?

Assumption 1 The numbers of arrivals in any collection of disjoint (nonoverlapping) time intervals are (mutually) independent (this is reasonable if unrelated customers arrive at the bank haphazardly in time). **Assumption 2** P(1 or more arrivals in short time interval $[s, s+t]$, $t > 0$ small) is proportional to t, for example $\lambda \cdot t$ for a rate parameter $\lambda > 0$ (this is reasonable if the arrivals process is smooth rather than lumpy). **Definition** To say that a function $f(t)$ of t (small) is $o(t)$ (read little-oh of t) is to say that

$$\lim_{t \to 0} \frac{f(t)}{t} = 0$$

In other words, $f(t)$ approaches 0 as $t \to 0$ at a rate faster than t itself.

The mathematically formal way to state Assumption 2 is then

$$P\left(\text{at least one arrival in } [s, s+t] \right) \underset{(t \text{ small})}{=} \underbrace{\lambda t + o(t)}_{\text{independent of } s}$$

Assumption 3	(nearly simultaneous arrivals are rare)

$$P\left(\begin{array}{c}\text{2 or more arrivals} \\ \text{in } [s, s+t) \\ (t \text{ small})\end{array}\right) = o(t)$$

Remarkably, these 3 simple & often plausible assumptions specify the probability behavior of $N(t)$ uniquely (see DS exercise 16, p. 296) for a proof of the following result).

λ constant $\leftrightarrow$ $N(t)$ is a <u>stationary stochastic process</u>

Note that Assumption 2 implies that the rate parameter λ is constant in time; this would be unrealistic in the bank problem over an entire day but would be reasonable during stable subsets of a day (e.g. 10am–11.30am)

Definition For any $\lambda > 0$, a random variable Y has the <u>Poisson distribution</u> with parameter λ (Poisson(λ)) if its pf is

S. Poisson (1800s ~ 1850)

$$f_Y(y \mid \lambda) = P(Y = y \mid \lambda) = \begin{cases} \dfrac{\lambda^y e^{-\lambda}}{y!} & \text{for } y = 0, 1, \ldots \\ 0 & \text{else} \end{cases}$$

(first person to study this was Abraham de Moivre (1711))

Result Under Assumptions 1-3 above, if $Y = \begin{pmatrix} \text{\# of arrivals in any} \\ \text{time interval of length } t \end{pmatrix}$ then $Y \sim \text{Poisson}(\lambda t)$.

Definition

A <u>Poisson process</u> with rate λ per unit time is a stochastic process satisfying

(1) for any $s > 0$ # arrivals in $[s, s+t) \sim \text{Poisson}(\lambda t)$

and (2) # of arrivals in disjoint time intervals are independent

Restatement of Result Under Assumptions 1–3 above, $N(t)$ is a Poisson process with rate parameter λ.

Exploring the Poisson distribution

$$f(y \mid \lambda) = P(Y = y \mid \lambda) = \begin{cases} \dfrac{\lambda^y e^{-\lambda}}{y!} & y = 0, 1, \ldots \\ 0 & \text{else} \end{cases}$$

(for $\lambda > 0$). This is our first example of a discrete rv that takes on a countably infinite number of possible values.

Comment: In reality we don't expect a Poisson rv to be infinite or "nearly infinite"; all the statement ($y = 0, 1, \ldots$) means is that ahead of time we can't place a fixed upper bound on Y.