STOR 435.001 Lecture 8

Random Variables - II

Jan Hannig

UNC Chapel Hill

September 19, 2017

Main properties

- Arise in describing the number of successes in n **independent** trials where each trial can either be a "success" or "failure" and the probability of success is p, constant across all the trials.
- Probability mass function:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

- Moment generating function (writing $q = 1 - p$)

$$\mathbb{E}(e^{tX}) := (q + pe^t)^n, \qquad t \in \mathbb{R}.$$

- $\mathbb{E}(X) = np$, $\mathrm{Var}(X) = np(1-p)$. So standard deviation $SD(X) = \sqrt{np(1-p)}$.

One major area of applications: Redundancy in electronics

Replicating key components so as to decrease the chance of failure. The following two slides taken verbatim from a highly researched article "*When Failure is an Option: Redundancy, reliability and regulation in complex technical systems*" by **John Downer**:

On 24 June 1982, 37,000 feet above the Indian Ocean, a British Airways Boeing 747 began losing power from its four engines. Passengers and crew noticed a numinous blue glow around the engine housings and smoke creeping into the flight deck. Minutes later, one engine shut down completely. Before the crew could fully respond, a second failed, then a third and then, upsettingly, the fourth. The pilots found themselves gliding a powerless Jumbo-Jet, 80 nautical miles from land and 180 from a viable runway.

Unsurprisingly, this caused a degree of consternation among the crew and – despite a memorably quixotic request that they not let circumstances ruin their flight – the passengers, who began writing notes to loved ones while the aircraft earned its mention in *The Guinness Book of Records* for 'longest unpowered flight by a non-purpose-built aircraft'. Finally, just as the crew were preparing for a desperate mid ocean 'landing', the aircraft spluttered back into life: passengers in life-jackets applauded the euphony of restarting engines and the crew landed them safely in Jakarta (Tootell 1985).

Redundancy discussion continued

The cause of the close call was quickly identified by incident investigators. They determined that ash from a nearby volcano had clogged the engines causing them to fail and restart only when the aircraft lost altitude and left the cloud. This unnerved the aviation industry; before the 'Jakarta incident' nobody imagined a commercial aircraft could lose all its propulsion.

A Boeing 747 can fly relatively safely on a single engine so with four it enjoys a comfortable degree of redundancy. This is why Boeing puts them there. Under normal circumstances, engineers consider even a single engine failure highly unlikely, so, until British Airways inadvertently proved otherwise, they considered the likelihood of all four failing during the same flight to be negligible – as near to 'impossible' as to make no difference. In fact, the aviation industry considered a quadruple engine failure so unlikely that they taught pilots to treat indications of it as an instrumentation failure. Nobody had considered the possibility of volcanic ash.

Example 1: Suppose that, in flight, airplane engines will fail with probability $1 - p$, independently from engine to engine. If an airplane needs a majority of its engines operative to complete a successful flight, for what values of p is a 5-engine plane preferable to a 3-engine plane?

☼

Binomial Random variables

Example 2: Jan's dog (referred to as DOG) claims that it can read Jan's mind. This could be bad news. Jan decides to test DOG. Jan takes out DOG's favorite treat and hides it in one of his hands with equal probability. Jan asks DOG to guess which hand. Jan repeats this experiment 10 times. Suppose DOG gets 8 out of 10 correct. What is the probability that DOG would have done at least this well if DOG could not read his mind? You may assume that all the assumptions of a Binomial experiment are satisfied (although in practice one needs to check the assumptions very carefully; e.g. smell resistent bags and then in hands, Jan cannot show any bias to any hands through pressure etc). However if you have a talking DOG, checking assumptions is the least of your worries.

Solution contd.

Definition

A random variable X is called Poisson if it takes values $0, 1, 2, \ldots$ with probabilities

$$p(i) = P(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots$$

for some $\lambda > 0$.

Note: Why $e^{-\lambda}$?

☼

Notation:[1] $X = Pois(\lambda)$

[1] Simeon Denis Poisson (1781-1840), French mathematician and physicist.

Poisson distribution

Motivation: If $Y = B(n,p)$ with large n, small p, and moderate np, then $Y \approx Pois(\lambda)$ with $\lambda = np$.

☼

Examples of random variables that usually obey Poisson law: number of misprints on a page of a book; number of people in a community living to 100 years; number of customers entering a post office on a given day; etc.

$X = Pois(\lambda)$: What are EX and $Var(X)$?

Note: Since $X \approx B(n, p)$, expect $EX \approx np = \lambda$ and $Var(X) \approx np(1-p) \approx \lambda$.

Rigorously **using Moment generating function**:

☼

Calculating $\mathbb{E}(X)$ for Poisson without MGF

At home (or see book) how to calculate $\mathbb{E}(X^2)$ without MGF.

Why is the Poisson distribution so important? Example I (World War II)

Figure: Firefighters putting out a blaze in London after an air raid during The Blitz in 1941. Picture By New York Times Paris Bureau Collection [Public domain], via Wikimedia Commons

Why is the Poisson distribution so important? Example I

See: An application of the Poisson distribution by R.D. Clarke [Prudential assurance company].

World War II

- Divided the 144 square km of South London into 576 squares of 1/4 square kilometer each.
- Counted the number of squares where there were $0, 1, 2, \ldots$ bombs
- Total number of bombs: 537. So if we assume that number of bombs in a square is a Poisson random variable with mean λ then λ can be estimated as $\lambda = 537/576 = .93$. Can estimate the number of squares with k bombs by $576 * e^{-\lambda}\lambda^k/k!$ (Poisson distribution formula).

No of bombs per square	Expected number using Poisson	Actual number
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
6 and over	1.57	1

Why is the Poisson distribution so important? Example II

See: Shark attacks and Poisson Approximation by Byron Schmuland

Wayne Gretzky during Edmonton Oiler days

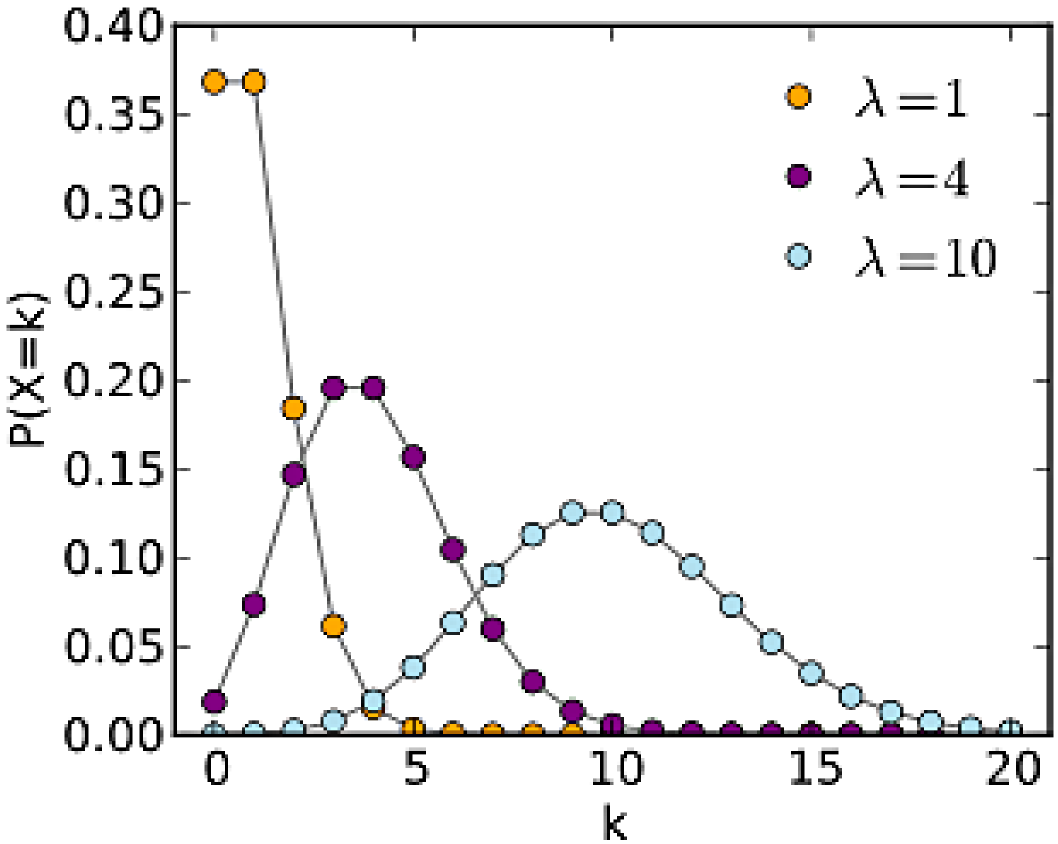
- Scored 1669 points per 696 games.
- Suppose we assume that number of points per game can be approximated by a Poisson random variable
- $\lambda \approx 1669/696$.

Figure: Wayne Gretzky. Picture by Kris Krug [CC BY-SA 2.0 (http://creativecommons.org/licenses/by-sa/2.0)], via Wikimedia Commons

Points	Prediction (Poisson)	Actual
0	63.27	69
1	151.71	155
2	181.90	171
3	145.40	143
4	87.17	79
5	41.81	57
6	16.71	14

Poisson distribution

Shape of Poisson distribution:



Example: Brad is in a big dorm with 180 other students. Let X be the number of other students who have the same birthday as Brad. Using Poisson approximation, approximate the probability that (a) there is at least one student with same birthday as Brad? (b) exactly one student with same birthday as Brad? (c) at least two students? Compare this with the exact probability. You may assume that the birthday of each of the other students is equally likely to be any one of the 365 days (no students born on leap years) and independent of each other.

☼ **Approximate probability:**

Problem contd: Exact probabilities

☼

Example

A crime has been committed in a small town containing 1000 people. Blood from every person in the town is taken and matched against the DNA profile found in the crime scene. The probability of a person showing up as a match is $1/500$, independent across individuals (unrealistic assumption in practice since relatives would have close profiles but ok for this made up problem!). Use Poisson approximation to find the chance exactly two people are found as matches?

☼

Poisson distribution

Example: when parameters are given to you

You are planning to take a bus to go home late at night. There are two possible buses you could take to go home. The number of passengers on the first bus (late at night) has a Poisson distribution with $\lambda = 4$. The second bus has a Poisson distribution with $\lambda = 5$. Since you like people, you chose bus one with probability $1/3$ and bus two with probability $2/3$. What is the probability that on your trip, your bus has at least 2 (≥ 2) passengers?

☼

We thus have $Pois(\lambda) \approx B(n, p)$ with $\lambda = np$ (and large n, small p). Binomial distribution involves independent trials. It turns out that this approximation is still good for trials which are "weakly dependent".

Even more generally:

Poisson paradigm

Consider n events, with p_i equal to the probability that event i occurs, $i = 1, \ldots, n$. If all the p_i are "small" and the trials are either independent or at most "weakly dependent," then the number of these events that occur approximately has a Poisson distribution with mean $\sum_{i=1}^{n} p_i$.

Poisson process

Another motivation for Poisson: Suppose “events” occur in time. Assume the following assumptions hold true:

1. The probability that exactly 1 event occurs in a given interval of length h is equal to $\lambda h + o(h)$, where $o(h)$ stands for any function $f(h)$ for which $\lim_{h \to 0} f(h)/h = 0$.
2. The probability that 2 or more events occur in an interval of length h is equal to $o(h)$.
3. For any integers $n, j_1, j_2, \ldots, j_n$ and any set of n nonoverlapping intervals, if we define E_i to be the event that exactly j_i of the events under consideration occur in the ith of these intervals, then events E_1, E_2, ... , E_n are independent.

Let $N(t)$ be the number of events that occur in $[0, t]$.

Fact

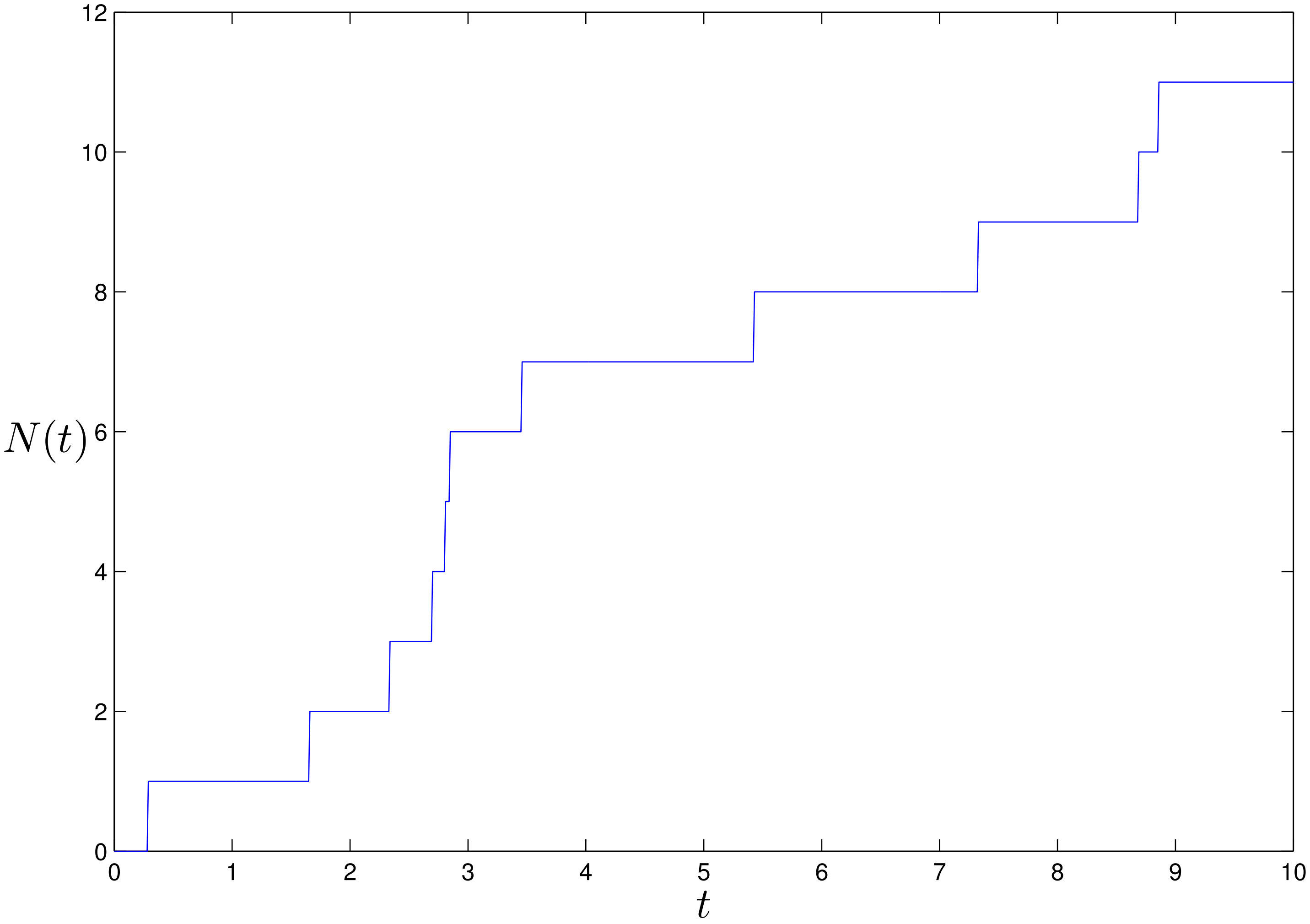
Under conditions 1, 2 and 3, $N(t) = Pois(\lambda t)$.

Note: $\lambda =$ rate per unit time at which events occur.

☼

Examples: number of earthquakes occurring in a period of time; number of deaths of policyholders of insurance company in a period of time; number of shark attacks in a period of time (see Extras); number of buses showing up at a bus stop etc.

Poisson process (with $\lambda = 1$):



Example In a big university, lights are kept on day and night and they burn at the rate $\lambda = 7.2$ per day (i.e. 1 day is the unit of time and lights burning out occur in accordance with the three assumptions at a rate of 7.2 per day). You may also assume for simplicity that they are replaced as soon as they burn out. Find the probability of:

(a) More than two burnouts between noon and 2 p.m. tomorrow.

(b) Exactly 50 burnouts next week (i.e. in the next 7 days).

☼

Solution continued