Stat 400, section 3.6a Poisson & 3.5c Geometric Random Variables

notes prepared by Tim Pilachowski

In probability, statistics, and related fields, a Poisson point process or Poisson process is a kind of random mathematical object which consists of points randomly located on a mathematical space. For our purposes, we will define the Poisson process on the Real number line where it can be used to model random events in time, or in the Real number plane where it can represent the locations of scattered objects in space. There are two important characteristics shared in general by Poisson point processes:

the number of points in each finite interval has a Poisson distribution, and
the number of points in disjoint intervals are independent random variables.

Poisson distributions model (some) discrete random variables, usually $X = 0, 1, 2, \ldots$ Typically, a Poisson random variable is a count of the number of events that occur in a certain time interval or spatial area. For example, the number of cars passing a fixed point in a 5 minute interval, or the number of calls received by a switchboard during a given period of time. A discrete random variable X is said to follow a Poisson distribution with parameter μ, if it has probability distribution

$$P(X = x) = p(x; \mu) = \frac{e^{-\mu} * \mu^x}{x!} \text{ where } x = 0, 1, 2, 3, \ldots$$

The following requirements must be met:

the length of the observation period is fixed in advance;
the events occur at a constant average rate;
the number of events occurring in non-connected intervals are statistically independent;
the random variable X has no upper limit (although large values may be extremely unlikely).

The Poisson distribution has expected value $E(X) = \mu$ and $V(X) = \mu$.

Example A: Between 1900 through 2000, 168 hurricanes made landfall in the United States. Statistical evidence suggests a Poisson distribution. Find the probabilities that a) 3, b) fewer than 3, and c) more than 3 hurricanes will reach the U.S. during a particular year. Also identify d) the variance of this probability distribution.
answers: ≈ 0.1450, ≈ 0.7677, ≈ 0.0873, 1.66, 1.66

In Example A, we came in through the backdoor to an application of Poisson distributions outlined in your text on the pages titles "The Poisson Process". Time is the variable, represented by the letter t, and the expected number of events during a specified interval of time t is $\mu = \alpha t$. The Poisson probability function in this scenario is written

$$P_k(t) = \frac{e^{-\alpha t} * (\alpha t)^k}{k!}.$$

Another handy use of Poisson probability distributions: They can be used to model binomial probability distributions under limited conditions. As a rule of thumb, if $n > 50$ and $np < 5$, then $b(x;\, n,\, p) \to p(x;\, np)$.

3.4 Example C revisited. From prior experience and testing, Shockingly Good, Inc. has determined that the probability of a spark plug being defective is 0.01. In a production run of 360 spark plugs, what is the probability that the number of defectives is at most 5? *answer*: ≈ 0.8411

At the very end of section 3.5, your text makes a passing reference to geometric probability distributions, which had already been introduced in section 3.2 (text Example 3.12). In my Lecture, Example B revisited was geometric probability.

Geometric distributions model (some) other discrete random variables. Typically, a geometric random variable is the number of trials required before one obtains the first failure, for example, the number of tosses of a coin before the first 'tail' is obtained, or a process where components from a production line are tested, in turn, until the first defective item is found. The trials must meet the following requirements:

the total number of trials is potentially infinite;
there are just two outcomes of each trial, defined success and failure;
the outcomes of all the trials are statistically independent;
all the trials have the same probability of success.

Example B. Suppose one die is rolled over and over until a 2 is rolled. If we define "success" as getting something other than a 2, what are the probabilities that a 2 is rolled
a) on the first roll? b) on the 2nd roll? c) on the 3rd roll? d) after at most 4 rolls? e) from the 3rd to the 6th rolls? f) after at least 5 rolls? *answers*: ≈ 0.1667, ≈ 0.1389, ≈ 0.1157, ≈ 0.5177, ≈ 0.3595, ≈ 0.4823

Using the above results, we can generalize for any geometric probability distribution. For discrete random variable X = number of successes before failure, and p = probability of a success, $P(X = x) = p^x(1 - p)$. Note that x must be an integer greater than or equal to 0 ($x = 0, 1, 2, 3, \ldots$) and $0 \le p \le 1$.

A geometric probability distribution has expected value $E(X) = \frac{p}{1-p}$ and variance $V(X) = \frac{p}{(1-p)^2}$.

Considering the similarity in the formulae for geometric distributions and negative binomial distributions, it is not surprising that the mean and variance share similar characteristics.

Considering that a geometric probability distribution is exponential in nature (although not base *e* like an exponential probability distribution), the formulae for geometric and exponential expected value and variance bear a striking resemblance. (We'll get to exponential probability distributions in chapter 4.)

Example B revisited. Suppose one die is rolled over and over until a 2 is rolled. If we define "success" as getting something other than a 2, what is the expected number of successes before we roll a 2? What are the variance and standard deviation? *answers*: 5, 30, ≈ 5.4772