STAT 380 Lecture 30
Spring 2018
19 March 2018

- I did 3 problems from the text.
- I am now done Poisson Processes.
- On Wednesday I will start continuous time Markov Chains.

Page 348, number 61: Some system has N flaws; N has a Poisson(λ) distribution. The ith flaw is discovered at time T_i. Given N the T_i are iid with cdf G. Let $D(t)$ be the number of flaws discovered by time t and $R(t) = N - D(t)$ be the number undiscovered. Then we are asked to find the distributions of $D(t)$ and $R(t)$ and whether or not these two are independent.

Given $N = n$ the random variable $D(t)$ has a Binomial distribution with parameters n and $p = G(t)$. So by the example I did in class:

1. The number discovered, $D(t)$ has a Poisson($\lambda G(t)$) distribution;
2. The number discovered, $R(t)$ has a Poisson($\lambda \{1 - G(t)\}$) distribution;
3. and $D(t)$ and $R(t)$ are independent.

In class I said D is an inhomogeneous Poisson process but did not prove it. Fix

$$s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k.$$

Let N_i be the number discovered in the interval $(s_i, t_i]$. Each point is now being labelled by one of $k+1$ labels: labels 1 to k indicate discovery. Label $k+1$ is just not discovered during those intervals. Points get label i with probability $p_i = G(t_i) - G(s_i)$ for $i = 1, \ldots, k$ and $p_{k+1} = 1 - \sum_1^k p_i$. So $N_1, \ldots, N_{k+1}$ are independent Poissons with parameters $\lambda p_i, i = 1, \ldots, k+1$. This meets the NHPP definition with

$$\Lambda(t) = \lambda G(t).$$

I also suggested D and R are independent as processes (rather than for a fixed t). But R is a decreasing process and if I tell you $R(0)$ you can figure out $D(\infty)$ because both are equal to N. So what I said is wrong. It is just for a single fixed t that $D(t)$ and $R(t)$ are independent.

Page 346, number 52: Two players (1 and 2) are playing a game. Player i scores points at times of a Poisson Process N_i with rate λ_i. The processes are independent. The player who is first to have scored k more points than the other is the winner. What is the probability that Player 1 wins?

First define $N = N_1 + N_2$. Then N is a Poisson Process with rate $\lambda_1 + \lambda_2$. Think about the arrival times $S_1 < S_2 < \cdots$ of N. At time 0 you are waiting for the first player to score. That time is the minimum of two exponential random

variables and the one for player 1 is smaller than the one for player two with probability

$$p = \frac{\lambda_1}{\lambda_1 + \lambda_2}/$$

Now imagine that n points have been scored; the time is S_n. Then the system starts over (by the memoryless property) so the probability that player 1 scores next is p, regardless of what has happened till now.

So: if you define X_n to be the number of points for player 1 - the number for player 2 at time S_n then X_n is a Markov Chain. The transition probabilities are given by

$$P(X_{n+1} = j | X_{=} i = \begin{cases} p & j = i+1 \\ 1-p & j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

We are to compute

$$P(X_n \text{ hits } k \text{ before it hits } -k).$$

This is actually the Gambler's Ruin problem. See Markov Chains slides number 59 to 65.

If you define $Y_n = X_n + k$ then Y_n is also a Markov Chain. It starts at k and we want to compute the chance that Y_n reaches $2k$ before it hits 0. I didn't do the algebra in class but here I will. Let

$$P_j = P(Y_n \text{ hits } k \text{ before it hits } 0).$$

The answer we need is P_k. Then

$$P_j = pP_{j+1} + (1-p)P_{j-1}$$

by first step analysis. Also $P_{2k} = 1$ and $P_0 = 0$. Re-arrange to get

$$p(P_{j+1} - P_j) = (1-p)(P_j - P_{j-1}).$$

Define $\theta = (1-p)/p = \lambda_2/\lambda_2$. Start at $j = 1$ and remember $P_0 = 0$ to get

$$P_2 - P_1 = \theta P_1$$

Then

$$P_3 - P_2 = \theta(P_2 - P_1) = \theta^2 P_1$$

and so on. In general

$$P_{j+1} - P_j = \theta^j P_1$$

Now sum this equation from $j = 0$ to $j = 2k-1$ to get

$$1 = P_{2k} - P_0 = \left(1 + \theta + \cdots + \theta^{2k-1}\right) P_1$$

This makes

$$P_1 = \frac{\theta - 1}{\theta^{2k} - 1}$$

for $\theta \neq 1$. For $theta = 1$ the geometric series adds up to $2k$ so

$$P_1 = \frac{1}{2k}.$$

We need P_k. If you stop summing at k not at $2k$ you get

$$P_k - P_0 = \theta^{k-1} P_1 = \frac{\theta^k - \theta^{k-1}}{(\theta^{2k} - 1)}.$$

or $P_k = 1/2$ for $\theta = \lambda_2/\lambda_1$.

Page 348, number 62: A book has N typos; this number has a Poisson(λ) distribution. Two proofreaders study the book to find typos. Reader i finds a typo with probability p_i; the proofreaders behave independently and typo discoveries are independent and all have the same probability. The problem asks you to develop estimators of λ, p_1 and p_2. In class I discussed the method of moments. Here is Maximum Likelihood.

Define N_{++} to be the number of typos which are found by both proofreaders. Let N_{+-}, N_{-+}, and N_{--} be the numbers found only by proofreader 1, only by proofreader 2 and not by either, respectively. The observed data is

$$N_{+-}, N_{-+}, N_{++}.$$

As in previous problems each typo is assigned to 1 of 4 categories randomly and independently so $N_{++}, N_{+-}, N_{-+}, N_{--}$ are independent Poissons with means $\lambda p_1 p_2$, $\lambda p_1(1-p_2)$, $\lambda(1-p_1)p_2$, and $\lambda(1-p_1)(1-p_2)$ respectively. Define $\mu_{++} = \lambda p_1 p_2$ to be the mean of N_{++} and so on. Then the log-likelihood is

$$-\mu_{++} + N_{++}\log(\mu_{++}) - \mu_{+-} + N_{+-}\log(\mu_{+-}) - \mu_{-+} + N_{-+}\log(\mu_{-+}).$$

Maximize this over μ_{++}, μ_{+-}, and μ_{-+} (by setting derivatives equal to 0) to find

$$\hat{\mu}_{++} = N_{++}$$

and so on. Then set $\hat{\mu}_{++} = \hat{\lambda}\hat{p}_1\hat{p}_2$ and so on and solve to find

$$\hat{p}_2 = \frac{N_{++}}{N_{++} + N_{+-}}$$

which is the fraction of the errors found by reader 1 which were also found by reader 2. Similarly

$$\hat{p}_1 = \frac{N_{++}}{N_{++} + N_{-+}}$$

and

$$\hat{\lambda} = \frac{(N_{++} + N_{+-})(N_{++} + N_{-+})}{N_{++}}.$$

Finally to forecast the number of undiscovered errors we use

$$\hat{\lambda}(1-\hat{p}_1)(1-\hat{p}_2) = \frac{N_{+-}N_{-+}}{N_{++}}.$$

Page 346, number 49: I did not do this one in class but I used to use this one for homework.