Chapter 4: Extra notes

ST 554 | Fall 2017

Negative Binomial as Sum of Geometric Rvs

If $X_i \overset{iid}{\sim} Geom(p)$, then $Y = \sum_{i=1}^{r} X_i \sim NBin(r, p)$

- Let $r = 1$. $P(Y = k) = p(1-p)^k$ for $k = 0, 1, \ldots$ is Geometric
- Let $r = 2$. The number of failures before 1st success is $Geom(p)$. Because the Bernoulli trials are iid, the number of *additional* failures before another success is also $Geom(p)$.
- $E[Y] = \sum_{i=1}^{r} E[X_i] = r(1-p)/p$

4.7 Poisson

- Poisson often used to model the number of events
- If $N \sim Pois(\lambda)$, then we say N has a Poisson distribution with parameter $\lambda > 0$.
  - λ is the expected number of events ($\mathbb{E}[N] = \lambda$)
- PMF

$$P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{for } k = 0, 1, \ldots$$

Computer Attacks

Under normal operating conditions, suppose packets arrive at a network server according to a *Poisson process* with a rate of $0.5/sec$. However during an malicious attack packets will arrive with a rate of $1.25/sec$. Suppose you monitor the network for $t = 8$ seconds and observe $x = 7$ packet arrivals.

- Poisson Process: $N(t) \sim Pois(\lambda t)$
- Let Y be the number of packets in $t = 8$ seconds
- Let A be the event that system is under attack
- $Y|A \sim Pois(\lambda = 8(.5) = 4)$, $Y|A^{\mathsf{c}} \sim Pois(\lambda = 8(1.25) = 10)$

1. What is the probability of observing 7 packets if there is *not* an attack?
   - First, notice that $Y|A^{\mathsf{c}} \sim Pois(4)$. Therefore,

$$P(Y = 7|A^{\mathsf{c}}) = \frac{e^{-4}\,4^7}{7!} = 0.06$$

2. What is the probability of observing 7 packets if there is an attack?
   - $Y|A \sim Pois(10)$. Therefore,

$$P(Y = 7|A) = \frac{e^{-10}\,10^7}{7!} = 0.09$$

3. If the prior probability of an attack is $P(A) = .25$, what is the probability of an attack given the 7 arrivals in 8 seconds? (i.e. find $P(A|Y = 7)$)
   - Using Bayes Theorem:

$$P(A|Y = 7) = \frac{P(Y = 7|A)P(A)}{P(Y = 7)}$$

   - We need to find $P(Y = 7)$: By Law of Total Probability,

$$\begin{aligned}P(Y=7) &= P(Y=7|A')P(A') + P(Y=7|A)P(A)\\ &= (.06)(.75) + (.09)(.25) = 0.0675\end{aligned}$$

- Thus,
$$P(A|Y=7) = \frac{.0225}{.0675} = .33$$

Survival Functions

- $G(k) = 1 - F(k) = P(X > k)$ is called the *survival function*.
- We can find the survival function for the first success distribution.
- Let $X \sim FS(p)$
  - If $X > k$, then there must be 0 successes in first k trials. So $P(X > k) = (1-p)^k$
- We can come to the same conclusing by summing the PMF

$$\begin{aligned}G_X(k) &= P(X > k)\\ &= \sum_{j=k+1}^{\infty} p(1-p)^{j-1}\\ &= p\sum_{j=0}^{\infty}(1-p)^{j+k}\\ &= p(1-p)^k\sum_{j=0}^{\infty}(1-p)^j\\ &= p(1-p)^k(1/p) \qquad \text{[Geometric Series]}\\ &= (1-p)^k\end{aligned}$$

Expected Value from Survival Functions

In certain situations, the survival function can be used to find the expected value. Let X be a discrete RV with support of non-negative integers ($X \in 0, 1, \ldots$). Then $E[X] = \sum_{k=0}^{\infty} G(k) = \sum_{k=0}^{\infty} P(X > k)$

- Proof: (also some details on pg 156)
  - First note: $G(k) = P(X > k) = \sum_{j=k+1}^{\infty} P(X = j)$

$$
\begin{aligned}
E[X] &= \sum_{k=0}^{\infty} G(k) \\
&= \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} P(X=j) \\
&= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \mathbb{1}(j>k) P(X=j) && \text{[use indicator for sum region]} \\
&= \sum_{j=0}^{\infty} P(X=j) \sum_{k=0}^{\infty} \mathbb{1}(j>k) && \text{interchange sums} \\
&= \sum_{j=0}^{\infty} P(X=j) \left(\sum_{k=0}^{j-1} 1 \right) \\
&= \sum_{j=0}^{\infty} P(X=j)\,(j) \\
&= \sum_{j=0}^{\infty} j \cdot P(X=j)
\end{aligned}
$$

Variance of Sample Mean

For *independent* RVs $V[\sum_i X_i] = \sum_i V[X_i]$

- Suppose we have $E[X_i] = \mu$ and $V[X_i] = \sigma^2$ for all $i = 1, \dots, n$
- Sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.
  - $E[\bar{X}] = \mu$
  - $V[\bar{X}] = \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] = V[X_i]/n$
  - What happens to variance when n gets big?
  - Ans: $\bar{X}$ becomes close to the constant μ