MATH 20, WORKSHEET 5

EDGAR COSTA

DUE FRIDAY OCTOBER 13TH

(1) If X and Y are any two random variables, then the covariance of X and Y is defined by

$$\text{Cov}[X, Y] := E[(X\text{-}E[X]))(Y\text{-}E[Y])].$$

Note that $\text{Cov}(X, X) = V(X)$. Show that, if X and Y are independent, then $\text{Cov}[X, Y] = 0$; and show, by an example, that we can have $\text{Cov}[X, Y] = 0$ and X and Y not independent.

Proof and Example: `https://en.wikipedia.org/wiki/Covariance`

(2) For a sequence of Bernoulli trials, let X_1 be the number of trials until the first success. For $j \geq 2$, let X_j be the number of trials after the (j-1)st success until the jth success. It can be shown that $X_1, X_2, \ldots$ is an independent trials process.

(a) What is the common distribution, expected value, and variance for X_j? You can use the fact that: $x+4x^3+9x^3+16x^4+\cdots = x\dfrac{1+x}{(1-x)^3}$

Answer:: We know that $X_i \simeq \text{Geo}(p)$ and we saw in class $E[\text{Geo}(p)] = \dfrac{1}{p}$.

$$\begin{aligned}
E[X_i^2] &= \sum_{k=1}^{+\infty} k^2 q^{k-1} p \\
&= \frac{p}{q} \sum_{k=1}^{+\infty} k^2 q^k \\
&= \frac{p}{q} q \frac{1+q}{(1-q)^3} && \text{(using the given fromula)} \\
&= \frac{p}{q} q \frac{1+q}{p^3} \\
&= \frac{1+q}{p^2}
\end{aligned}$$

and

$$\begin{aligned} V[X_i] &= E[X_i^2] - E[X_i]^2 \\ &= \frac{1+q}{p^2} - \frac{1}{p^2} \\ &= \frac{q}{p^2} \end{aligned}$$

(b) Let $T_n = X_1 + X_2 + \cdots + X_n$. Then T_n is the time until the nth success. Find $E[T_n]$ and $V[T_n]$.

Answer: We apply the linearity of the expected value:

$$E[T_n] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = n\sum_{i=1}^{n} \frac{1}{p} = \frac{n}{p}.$$

Similarly, since X_i are independent,

$$V[T_n] = V[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} V[X_i] = n\sum_{i=1}^{n} \frac{q}{p^2} = \frac{nq}{p^2}.$$

(c) Find the distribution function of T_n. (Challenge: show that adds up to one.)

Answer: T_n is the number of trials until the nth success, thus $T_n = k$ if before k there were $n-1$ successes and on the kth trial, we also add a success, therefore

$$P(T_n = k) = \binom{k-1}{n-1} p^n q^{k-n}, \text{ for } k \geq n$$

where $\binom{k-1}{n-1}$ is the number of ways to pick the previous $n-1$ successes out of the $k-1$ possibilities.

$$
\begin{aligned}
\sum_{k=n}^{+\infty} P(T_n = k) &= \sum_{k=n}^{+\infty} \binom{k-1}{n-1} p^n q^{k-n} \\
&= \sum_{l=0}^{+\infty} \binom{l+n-1}{n-1} p^n q^l \\
&= \sum_{l=0}^{+\infty} \binom{l+n-1}{l} p^n q^l \\
&= p^n \sum_{l=0}^{+\infty} \binom{l+n-1}{l} q^l
\end{aligned}
$$

The tricky part is to expand $\dfrac{1}{(1-q)^n}$ into a series and make it match the desired series:

$$
\frac{1}{p}^n = \left(\frac{1}{1-q}\right)^n = \cdots = \sum_{l=0}^{+\infty} \binom{l+n-1}{l} q^l.
$$

This is up to you.

(3) The Norwich Beer Company runs a fleet of trucks along the 100 mile road from Hangtown to Dry Gulch. The trucks are old, and are apt to break down at any point along the road with equal probability.

(a) Where should the company locate a garage so as to minimize the expected distance from a typical breakdown to the garage? In other words, if X is a random variable giving the location of the breakdown, measured, say, from Hangtown, and b gives the location of the garage, what choice of b minimizes $E[|X-b|]$?

(b) Now suppose X is not distributed uniformly over $[0, 100]$, but instead has density function $f_X(x) = 2x/10000$. Then what choice of b minimizes $E(|X-b|)$?

Answer: In the case that X is uniformly distributed on $[0, 100]$, one finds that

$$
E[|X-b|] = \frac{1}{200}(b^2 + (100-b)^2),
$$

which is minimized when $b = 50$.

For part (b), we have

$$E[|X\text{-}b|] = \frac{200}{3} - b + \frac{b^3}{15000},$$

which is minimized when $b = 50\sqrt{2}$.