(*): Assigned to weekly problem set.

Conditional Variance

1. (*) Let X_1, X_2, and Y be random variables (not necessarily independent), such that Y has finite variance. Define

$$A = \mathbb{E}[Y|X_1] \qquad \text{and} \qquad B = \mathbb{E}[Y|X_1, X_2]$$

   (a) Show that $\mathrm{Var}(A) \leq \mathrm{Var}(B)$. *Hint: start by using Eve's law on B.*

   (b) Check that the result in (a) makes sense in the extreme case where Y is independent of X_1.

   (c) Check that the result in (a) makes sense in the extreme case where $Y = h(X_2)$ for some function h.

2. Suppose you see a total of $N \sim \mathrm{Geom}(s)$ movies in your lifetime. Suppose that in each movie, you have a probability p of liking the movie, independently of other movies and of N. Let T be the number of movies you like in your lifetime.

   (a) Find the mean of T.

   (b) Find the variance of T.

3. Emails arrive one at a time in an inbox. Let T_j be the time at which the j-th email arrives. Suppose the waiting times between emails are iid $\mathrm{Exp}(\lambda)$ random variables (i.e. $T_1, T_2 - T_1, T_3 - T_2, \ldots$ are are iid $\mathrm{Exp}(\lambda)$).

   Each email is non-spam with probability p and spam with probability $1 - p$, independently of other emails and of the waiting times. Let X be the time at which the first non-spam email arrives (so X is a continuous random variable). For example, $X = T_1$ is the first email is non-spam, $X = T_2$ if the first email is spam and the second email is non-spam, etc.

   Hint for both parts: Let N be the number of emails until the first non-spam (including that one), and given N, write X as a sum of N terms.

   (a) Find the mean of X.

   (b) Find the variance of X.