Robert Werthman

---

1. *Probability boot camp*

   (a) *Prove Markov's inequality, $Pr[X \geq c] \leq E[X]/c$, with $c > 0$*

   The formula for the probability of a continuous random variable X with probability density function $f(x)$ is

$$Pr[x_1 \leq X \leq x_2] = \int_{x_1}^{x_2} f(x)dx$$

   And the formula for the expected value of a continuous random variable X with probability density function $f(x)$ is

$$E[x_1 \leq X \leq x_2] = \int_{x_1}^{x_2} xf(x)$$

   so for $Pr[X \geq c]$ we have

$$Pr[X \geq c] = \int_{c}^{\infty} f(x)dx$$

   and since X is a nonnegative random variable we have

$$E[X] = \int_{0}^{\infty} xf(x)dx$$

   Notice that $0 < c \leq \infty$. This tells us that the bounds of $E[X]$ are greater than $Pr[X \geq c]$. We can break up the integral formed by $E[X]$ to create an inequality that will begin to look similar to the integral of $Pr[X \geq c]$.

$$\begin{aligned} E[X] &= \int_{0}^{\infty} xf(x)dx \\ &= \int_{0}^{c} xf(x)dx + \int_{c}^{\infty} xf(x)dx \\ &\geq \int_{c}^{\infty} xf(x)dx \end{aligned}$$

We can assume that $x \geq c$ because c is one of the bounds of the integral. This means we can substitute c for x.

$$\int_c^\infty xf(x)dx \geq \int_c^\infty cf(x)dx \geq c\int_c^\infty f(x)dx$$

We now have an equation for $E[X]$ that has $Pr[X \geq c]$.

$$E[X] \geq c\int_c^\infty f(x)dx = cPr[X \geq c]$$

Dividing both sides by c gives us

$$E[X]/c \geq Pr[X \geq c]$$

which is Markov's inequality. We have just shown that $Pr[X \geq c] \leq E[X]/c$, with $c > 0$ is true based on the probability and expected value of the continuous random variable X.

(b) *Prove Chebyshev's inequality* $Pr[|X-\mu| \geq c \cdot \sigma] \leq 1/c^2$

If we let $k = (c \cdot \sigma)$ we have

$$Pr[|X-\mu| \geq c \cdot \sigma] = Pr[|X-\mu| \geq k]$$

One of the properties of $|a|$ is that it can also be represented as $\sqrt{a^2}$ so we can change

$$Pr[|X-\mu| \geq k] = Pr[\sqrt{(X-\mu)^2} \geq k]$$

If we take the square root of both sides of the inequality we get

$$Pr[(X-\mu)^2 \geq k^2]$$

From here we can use Markov's inequality since we know $(X-\mu)^2$ is nonnegative. If we let $(X-\mu)^2 = X$ and $k^2 = c$ and substitute those into Markov's inequality we have

$$Pr[(X-\mu)^2 \geq k^2] \leq E[(X-\mu)^2]/k$$

The variance σ^2 of a continuous random variable X with mean μ is

$$\sigma^2 = E[(X-\mu)^2]$$

So

$$Pr[(X-\mu)^2 \geq k^2] \leq E[(X-\mu)^2]/k$$

Now becomes

$$Pr[(X-\mu)^2 \geq k^2] \leq \sigma^2/k$$

If we now change k back to $c \cdot \sigma$ we have the equation

$$Pr[(X-\mu)^2 \geq (c \cdot \sigma)^2] \leq \sigma^2/c^2 \cdot \sigma^2$$

Reducing the right side of the equation results in

$$Pr[(X-\mu)^2 \geq (c \cdot \sigma)^2] \leq 1/c^2$$

And taking the square root of both sides of the inequality on the left hand side of the equation results in

$$Pr[\sqrt{(X-\mu)^2} \geq c \cdot \sigma] \leq 1/c^2$$

Which can be further reduced to

$$Pr[|X-\mu| \geq c \cdot \sigma] \leq 1/c^2$$

Which is Chebyshev's inequality. We have just shown that Chebyshev's inequality can be proven using Markov's Inequality.

(c) *Show that for any discrete random variables* $X, X', E[X] = E[E[X|X']]$.

If we let $Y = X'$ we have

$$E[X] = E[E[X|Y]]$$

The expected value of a discrete random variable X is

$$E[X] = \sum_x x \cdot Pr[X = x]$$

The conditional probability for any two discrete random variables X, Y is defined to be

$$Pr[X = x|Y = y] = \frac{Pr[X = x \cap Y = y]}{Pr[Y = y]}$$

The conditional expectation for any two discrete random variable X, Y is defined to be

$$E[X|Y = y] = \sum_x x \cdot Pr[X = x|Y = y]$$

Given the above assertions

$$\begin{aligned}
E[E[X|Y]] &= E[\sum_x x \cdot Pr[X = x|Y]] \\
&= \sum_y [\sum_x x \cdot Pr[X = x|Y = y]] \cdot Pr[Y = y] \\
&= \sum_x x \sum_y Pr[X = x|Y = y] \cdot Pr[Y = y] \\
&= \sum_x x \cdot Pr[X = x] \\
&= E[X]
\end{aligned}$$

Sources I used to complete this problem:

- http://www.maths.qmul.ac.uk/ pettit/MTH5122/notes15.pdf
- https://en.wikipedia.org/wiki/Expected_value
- https://math.stackexchange.com/questions/1353418/expected-value-proof-law-of-total-expectation

(d) *Prove by induction that* $E[X_t] = 0$ *for a martingale.*

We want to show that in a martingale sequence the expected value for a random variable X_{t+1} is the random variable X_t before it.

Base case:

When $t = 0$ with $X_0 = 0$ we have

$$\begin{aligned} E[X_1|X_0] &= \sum_x x \cdot Pr[X_1 = x|X_0 = 0] \\ &= 0 \end{aligned}$$

Inductive step:

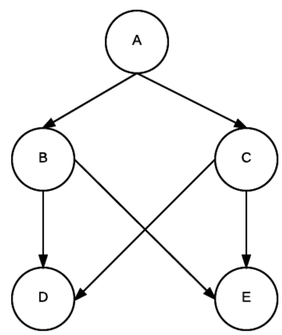
We showed that the base case $E[X_1|X_0] = 0$ is true. Therefore, we can say that $E[X_2|X_1] = E[E[X_1|X0]] = 0$ no matter what X_2 is. This leads to the equation

$$E[X_t|X_0, ..., X_{t-1}] = E[E[...E[X_1|X_0]]] = 0$$

So by induction we have just shown that $E[X_t] = 0$.

2. *Give a graph G with a source vertex s and a set of edges E_π to every vertex in G such that the path from s to v is the shortest path but the set of the edges E_π cannot be produced by running a breadth-first search on G.*

Figure 1: Example of a graph where there exists a set of edges that are the shortest path between vertices but will not be found by a breadth first search.



For this question, the shortest-path from s to v is defined as the minimum number of edges in any path from vertex s to vertex v. If we let the source vertex s be vertex A then the set of edges E_π that cannot be produced by running a breadth first search are

$$\{(A, B), (A, C), (B, D), (C, E)\}$$

This is the case because of how breadth first search discovers vertices. It starts at a source vertex s and then adds every vertex in the adjacency list of s to a queue. In the case above s will be A and it will add B and C to the queue because those are the vertices adjacent to A. An edge from A to B and from A to C will be created. Once it does that it will then pop off the first vertex on the queue which will be either B or C depending on the ordering of the adjacency list of A. It will then add every vertex from the adjacency list of the vertex is pops off to the queue. If either B or C is removed from the queue, D and E are still both added to the queue because there are both in B and C adjacency lists. In either case, edges are created from a single vertex, C or B, to both D and E i.e. $(C, D), (C, E)$. It will never be the case that both C and B will have a only single edge i.e. $(C, D), (B, E)$. It will only be the case that one of them will have two edges so the set of edges E_π above will would never occur with a breadth first search.

3.