Errata for Bayesian Modeling

October 13, 2014

I understand that the last lecture was rushed, so let's try to go through it more slowly. First off, recall that we were trying to predict whether or not Pinky's performance would improve given Pinky's performance. I wrote this as $P(X_{n+1}|X)$ where $X = (X_1, X_2, \dots, X_n)$ is the list of the previous n changes in Pinky's performance. To be more concrete, I should say we are computing $P(X_{n+1} = 1|X)$, which we can then compare to $P(X_{n+1} = -1|X)$ in order to give a prediction.

The problem with computing $P(X_{n+1} = 1|X)$ directly is that there is an unknown variable θ that we have to model. In order to figure θ into our computation, we first need to review some lecture and some basics of probability.

I proposed a Markov assumption regarding the independence of successive trials. The assumption is that Pinky's change in performance only depends on the value of θ and on his change in performance from the day before. In other words, for all i, $P(X_{i+1} = 1|X, \theta) = P(X_{i+1} = 1|X_i, \theta)$.

Now, if we want to compute the value of $P(X|\theta = t)$, we can use the chain rule for conditional probability: Given events, A, B, C, $P(A, B, C) = P(A|B, C)P(B|C)P(C)$. Putting together the Markov assumption and the chain rule, we have

$$\begin{aligned}
P(X|\theta = t) &= P(X_1, X_2, \dots, X_n|\theta = t) \\
&= P(X_n|X_{n-1}, \dots, X_n, \theta = t)P(X_{n-1}|X_{n-2} \dots, X_n, \theta = t) \dots P(X_1|\theta = t) \\
&= P(X_n|X_{n-1}, \theta = t)P(X_{n-1}|X_{n-2}, \theta = t) \dots P(X_1|\theta = t)
\end{aligned}$$

Now to explain the integral business. Recall that when we have events E and F, we know that $P(E) = P(EF) + P(EF^c)$. What if F is a discrete random variable, with possible values 1, 2, or 3. Then $P(E) = P(E, F = 1) + P(E, F = 2) + P(E, F = 3)$. In general, if F can take on any values in the set S, then we can always write

$$P(E) = \sum_{s \in S} P(E, F = s)$$

We want to introduce θ into $P(X_{n+1}|X)$, but θ is not a discrete random variable. In fact, it is continuous random variable, so in order to sum over all of it's possible values, we'll need to take an interval. Now we can write

$$P(X_{n+1} = 1|X) = \int_0^1 P(X_{n+1} = 1, \theta = t|X)dt$$

Putting together everything that we have, we get

$$
\begin{aligned}
P(X_{n+1} = 1|X) &= \int_0^1 P(X_{n+1} = 1, \theta = t|X) dt \\
&= \int_0^1 P(X_{n+1} = 1|\theta = t, X) P(\theta = t|X) dt \\
&= \int_0^1 P(X_{n+1} = 1|X_n, \theta = t) \frac{P(X|\theta = t) P(\theta = t)}{P(X)} dt \\
&= \frac{1}{P(X)} \int_0^1 P(X_{n+1} = 1|X_n, \theta = t) P(X|\theta = t) P(\theta = t) dt
\end{aligned}
$$

and similarly,

$$
P(X_{n+1} = 0|X) = \frac{1}{P(X)} \int_0^1 P(X_{n+1} = 0|X_n, \theta = t) P(X|\theta = t) P(\theta = t) dt
$$

Even though we don't care to compute $P(X)$, we can compute all of the other values, so we can still recover $P(X_{n+1} = 1|X)$ and $P(X_{n+1} = 0|X)$. The same principle applies with $P(\theta = t)$ which we computed up to a normalizing constant as `UPrior`, rather than using the true value `UPrior` divided by `PriorNorm`.