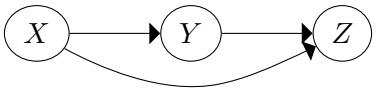
Topic 15: Graphical Representations of Random Variables

02-680: Essentials of Mathematics and Statistics

November 13, 2024

1 Bayesian Networks

Sometimes we need a more complicated conditional description. One way to do this is using a network (graph).



In general for a network such as this we define the probabilities as:

$$p\Big(\langle Q_1, Q_2, ..., Q_d\rangle\Big) = \prod_{i=0}^{d} p\left(Q_i \middle| \bigwedge_{j \in N^{in}(Q_i)} Q_j\right)$$

remember that $N^{in}(v)$ is the set of in-neighbors of v in a graph.

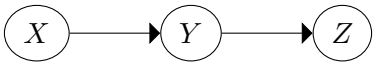
So for the Bayesian Network above, we may have something like:

x	$p(X = x)$
0	0.7
1	0.3

	$p(Y \mid X)$	
x	$y = 1$	$y = 0$
0	0.5	0.5
1	0.9	0.1

		$p(Z \mid Y, X)$	
x	y	$z = 1$	$z = 0$
0	0	0.3	0.7
0	1	0.1	0.9
1	0	0.7	0.3
1	1	0.4	0.6

But if we alter the network to the following (i.e. remove the dependence of Z on X),



we end up with

x	$p(X = x)$
0	0.7
1	0.3

	$p(Y \mid X)$	
x	$y = 1$	$y = 0$
0	0.5	0.5
1	0.9	0.1

	$p(Z \mid Y, X)$	
y	$z = 1$	$z = 0$
0	0.2	0.8
1	0.7	0.3

2 Markov Chains

Markov chains are usually used to model sequences of items. We saw when talking about conditional probabilities that:

$$\begin{aligned} p(X_n, X_{n-1}, ...X_1) &= p(X_n \mid X_{n-1}, ...X_1)p(X_{n-1}, ...X_1) && (1)\\ &= p(X_n \mid X_{n-1}, ...X_1)p(X_{n-1} \mid X_{n-2}, ...X_1)p(X_{n-2}, ...X_1) && (2)\\ &\dots && (3)\\ &= p(X_n \mid X_{n-1}, ...X_1) \dots p(X_2 \mid X_1)p(X_1) && (4)\\ & && (5) \end{aligned}$$

The ***Markov assumption*** is that the probability of X_k depends on X_{k-1} alone. This simplifies the computation, but at the cost of long term connections. Anecdotally the Markov assumption says:

the future is independent of the past given the present.

So that means we can rewrite the statement above as:

$$p(X_n, X_{n-1}, ...X_1) = p(X_n \mid X_{n-1})p(X_{n-1} \mid X_{n-2}) \dots p(X_2 \mid X_1)p(X_1).$$

This is the basis for ***Hidden Markov Models*** which are used a lot in genetics.

Example. Consider a sequence of DNA nucleotides modeled as a set of random variables $N_1N_2 \dots N_n$, each of which can take on the value $\{\texttt{A}, \texttt{T}, \texttt{C}, \texttt{G}\}$. We want to know if the sequence is a CpG island or not.

Looking at the equation above, if we know the initial distribution (that is the probability distribution for $p(N_1)$) for both cases, and the transition distributions ($p(N_k \mid N_{k-1})$ we can determine which is more probable.

$$P_b = \begin{bmatrix} & \texttt{A} & \texttt{C} & \texttt{G} & \texttt{T} \\ \texttt{A} & 0.30 & 0.20 & 0.29 & 0.21 \\ \texttt{C} & 0.32 & 0.30 & 0.08 & 0.30 \\ \texttt{G} & 0.25 & 0.25 & 0.29 & 0.21 \\ \texttt{T} & 0.18 & 0.24 & 0.29 & 0.29 \end{bmatrix} \quad P_c = \begin{bmatrix} & \texttt{A} & \texttt{C} & \texttt{G} & \texttt{T} \\ \texttt{A} & 0.18 & 0.27 & 0.43 & 0.12 \\ \texttt{C} & 0.17 & 0.37 & 0.27 & 0.19 \\ \texttt{G} & 0.16 & 0.34 & 0.37 & 0.13 \\ \texttt{T} & 0.08 & 0.36 & 0.38 & 0.18 \end{bmatrix}$$

Assuming for both $p(N_1 = \texttt{A}) = p(N_1 = \texttt{C}) = p(N_1 = \texttt{T}) = p(N_1 = \texttt{G}) = 0.25$, what is the sequence `ACTTC` more likely to be a CpG island or not?

Useful References

Degroot and Schervish. "Probability and Statistics" §§3.10