Homework 0

Kai Qu
kq4ff

September 3, 2019

1. (1) The sample space of X is {000, 001, 010, 011, 100, 101, 110, 111}.

   (2) $P(X = 0, 0, 0) = (1 - 0.6) * (1 - 0.6) * (1 - 0.6) = 0.064$

   Let E = the outcome has two heads and one tail

   $P(E) = \binom{3}{2} * 0.6 * 0.6 * (1 - 0.6) = 0.432$

2. (1) Let B = there is a burglar in the house, A = the alarm is triggered

$$
\begin{aligned}
P(B) &= 0.005 \\
P(A \mid B) &= 0.97 \\
P(A, B) = P(B) * P(A \mid B) &= 0.00485 \\
P(A \mid B_C) &= 0.002 \\
P(A, B_C) = P(A \mid B_C) * (1 - P(B)) &= 0.002 * (1 - 0.005) = 0.00199 \\
P(A) = P(A, B_C) + P(A, B) &= 0.00684 \\
P(B \mid A) = P(A, B)/P(A) &= 0.00485/0.00684 = 0.709
\end{aligned}
$$

   Therefore, the probability that there was a burglar in his house, given the alarm triggers, is 70.9%.

(2) Let E = there is an earthquake

$$\begin{aligned}
P(A \mid E) &= 0.01 \\
P(E) &= 0.001 \\
P(A, E) &= P(A \mid E)P(E) = 1 \times 10^{-5} \\
P(B_C) &= 0.995 \\
P(B_C, E) = P(B_C)P(E) &= 9.95 \times 10^{-4} \\
P(A \mid B_C, E) &= 0.002 \\
P(A, B_C, E) &= P(A \mid B_C, E)P(B_C, E) = 1.99 \times 10^{-6} \\
P(A, B, E) &= P(A, E) - P(A, B_C, E) = 8.01 \times 10^{-6} \\
P(B \mid A, E) &= P(A, B, E)/P(A, E) \\
&= 0.801
\end{aligned}$$

Therefore, the probability that there is a burglar given that the earthquake occurs and the alarm triggers, is 80.1%.

3. In this problem, the function that we want to maximize and the constraints are

$$\ln f(p) = \sum_{n=1}^{N} \ln P(x^{(n)})$$

$$\sum_{k=1}^{K} p_k = 1$$

By Lagrange Multiplier, we have

$$\frac{\partial}{\partial p_k}(\sum_{n=1}^{N} \ln P(X^{(n)}) - \lambda(\sum_{k=1}^{K} p_k - 1)) = 0 \tag{*}$$

Note that

$$P(X^{(n)}) = \prod_{k=1}^{K} p_k^{\delta(X^{(n)},k)} \tag{1}$$

$$\ln P(X^{(n)}) = \sum_{k=1}^{K} \ln(p_k)\delta(X^{(n)}, k) \tag{2}$$

Substitute (2) into (*), we get

$$\frac{\partial}{\partial p_k}(\sum_{n=1}^{N} \sum_{k=1}^{K} \ln(p_k)\delta(X^{(n)}, k) - \lambda(\sum_{k=1}^{K} p_k - 1)) = 0$$

Taking the partial derivative with respect to p_k and using the constraints, we get

$$p_k = \sum_{n=1}^{N} \delta(X^{(n)}, k)/\lambda \tag{.1}$$
$$\sum_{k=1}^{K} p_k = 1$$
$$\sum_{n=1}^{N}\sum_{k=1}^{K} \delta(X^{(n)}, k)/\lambda = 1$$
$$N/\lambda = 1$$
$$\lambda = N$$

Finally, substitute $\lambda = N$ into (.1) above, we get

$$p_k = \sum_{n=1}^{N} \delta(X^{(n)}, k)/N$$

4. (1) $E_{P(X)} = 0.7 * 1 + 0.3 * 0 = 0.7$

(2) $P(X) = -p\log_2 p - (1-p)\log_2(1-p) = 0.8812$

(3) $KL(P \mid\mid Q) = \sum_{i=1}^{N} P(X_i)(ln(P(X_i) - ln(Q(X_i)) = 0.1837$

(4) To prove $KL \geq 0$, we try to prove $-KL \leq 0$.

First, note that $\ln x \leq x - 1$. Let $f(x) = \ln x - x + 1$ for $x > 0$

$$f'(x) = 1/x - 1$$
$$f'(1) = 0$$
$$f''(x) = -1/x^2$$
$$f''(1) < 0$$

Therefore, f(1) = 0 is a local maximum.

Also,

$$\lim_{x\to 0^+} = -\infty = \lim_{x\to\infty}$$

Therefore, f(1) = 0 is the global maximum, and thus $\ln x \leq x - 1$.

$$\begin{aligned}
-KL(P \mid\mid Q) &= \sum_{i=1}^{N} P(X_i) log(Q(X_i)/P(X_i)) \\
&\leq \sum_{i=1}^{N} P(X_i)(Q(X_i)/P(X_i) - 1) \\
&= \sum_{i=1}^{N} Q(X_i) - P(X_i) \\
&= 1 - 1 \\
&= 0
\end{aligned}$$

Thus, since $-KL(P \mid\mid Q) \leq 0$, $KL(P \mid\mid Q) \geq 0$.
If P = Q ($\rightarrow$):

$$\begin{aligned}
KL(P \mid\mid P) &= \sum_{i=1}^{N} P(X_i) log(P(X_i)/P(X_i)) \\
&= \sum_{i=1}^{N} P(X_i) log(1)) \\
&= \sum_{i=1}^{N} 0 \\
&= 0
\end{aligned}$$

If $KL(P \mid\mid Q) = 0(\leftarrow)$:
First note that $f(x) = \log(x)$ is a convex function. By Jensen's inequality, we have $f(E(x)) \leq E(f(x))$. Therefore,

$$\begin{aligned}
KL(P||Q) &= -\sum_{x} P(x) \log(Q(x)/P(x)) \\
&\geq -\log(\sum_{x} P(x)Q(x)/P(x)) \\
&= -\log(\sum_{x} Q(x)) \\
&= -\log(1) \\
&= 0
\end{aligned}$$

The equality is equal to 0, which means that f is either linear or, for every x

$(Q(x)/P(x))$ is 1. Since $f(x) = log(x)$ is convex, the $Q(x)/P(x) = 1$ for all x. Therefore, $P(X) = Q(X)$. Thus, $KL(P||Q) = 0$ iff $P = Q$.