Homework 1
86-595 Neural Data Analysis
Name: Shashank Singh
Email: sss1@andrew.cmu.edu
Due: Tuesday, September 18, 2012

Problem 1

By given definition of mutual information,

$$
\begin{aligned}
MI(X,Y;S) &= \sum_{(x,y)\in X\times Y}\sum_{s\in S} p(x,y,s)\log_2\left(\frac{p(x,y,s)}{p(x,y)p(s)}\right) \\
&= \sum_{x\in X}\sum_{y\in Y}\sum_{s\in S} p(x,y,s)\log_2\left(\frac{p(y,s|x)p(x)p(s|x)}{p(y|x)p(x)p(s)p(s|x)}\right) \\
&= \sum_{x\in X}\sum_{y\in Y}\sum_{s\in S} p(x,y,s)\left(\log_2\left(\frac{p(x,s)}{p(x)p(s)}\right)+\log_2\left(\frac{p(y,s|x)}{p(y|s)p(s|x)}\right)\right) \\
&= \sum_{x\in X}\sum_{y\in Y}\sum_{s\in S} p(x,y,s)\log_2\left(\frac{p(x,s)}{p(x)p(s)}\right)+\sum_{x\in X}\sum_{y\in Y}\sum_{s\in S} p(y,s|x)p(x)\log_2\left(\frac{p(y,s|x)}{p(y|s)p(s|x)}\right) \\
&= \sum_{x\in X}\sum_{s\in S} p(x,s)\log_2\left(\frac{p(x,s)}{p(x)p(s)}\right)+\sum_{x\in X} p(x)\sum_{y\in Y}\sum_{s\in S} p(y,s|x)p(x)\log_2\left(\frac{p(y,s|x)}{p(y|s)p(s|x)}\right) \\
&= MI(X;S)+MI(Y;S|X). \quad \blacksquare
\end{aligned}
$$

Problem 2

By definition of entropy,

$$
\begin{aligned}
H(Y) &= -\sum_{n=0}^{\infty} p(y=n)\log(p(y=n)) = -\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\log\left(\frac{\lambda^n}{n!}e^{-\lambda}\right) \qquad \left(p(Y=n)=\frac{\lambda^n}{n!}e^{-\lambda}\right) \\
&= \sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\left(-\log(\lambda^n)-\log(e^{-\lambda})+\log(n!)\right) \\
&= -\lambda\log(\lambda)\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda}+\lambda\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}+\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\log(n!) \\
&= \lambda(1-\log(\lambda))+e^{-\lambda}\sum_{n=0}^{\infty}\frac{\log(n!)\lambda^n}{n!}. \quad \blacksquare \qquad \left(\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}=1\right)
\end{aligned}
$$

Problem 3

a. Note that $P(Y = 0) = \cdots = P(Y = 4) = 0.2$. Thus, by Problem 4a., $H(Y) = \log_2(5)$.

b.

$$\begin{aligned} MI(T;Y) &= \sum_{t=1,2}\sum_{y=0,1,2,3,4} p(T = t, Y = y)\log_2\left(\frac{p(T = t, Y = y)}{p(T = y)p(Y = y)}\right) \\ &= \sum_{t=1,2}\sum_{y=0,1,2,3,4} p(T = t, Y = y)\log_2\left(\frac{p(T = t, Y = y)}{(0.5)(0.2)}\right) \end{aligned}$$

Problem 4

a. By definition of entropy,

$$H(X) = -\sum_{x\in X}\frac{1}{N}\log_2\left(\frac{1}{N}\right) = -N\frac{1}{N}\log_2\left(\frac{1}{N}\right) = \log_2(N). \quad \blacksquare$$

b. Suppose $p : X \to \mathbb{R}$ is a probability distribution on some finite set X, and suppose q is the uniform distribution on X. Then, by definition of entropy,

$$\begin{aligned} 0 \le D(p||q) &= \sum_{x\in X} p(x)\log_2\left(\frac{p(x)}{q(x)}\right) = \sum_{x\in X} p(x)\log_2\left(\frac{p(x)}{1/|X|}\right) \\ &= \sum_{x\in X} p(x)\left(\log_2(p(x)) + \log(|X|)\right) \\ &= \log_2(|X|)\sum_{x\in X} p(x) + \sum_{x\in X} p(x)\log_2(p(x)) = \log_2(|X|) - H(p). \end{aligned}$$

Thus, $D(p||q)$ is minimized when $H(p)$ is maximized. Since $D(p||q) \ge 0$, with equality holding if and only if $p = q$ on X, $D(p||q)$ is minimized (so that $H(p)$ is maximized) when $p = q$. $\blacksquare$

Problem 5

a. By the result of part a. of problem 4, $H(S) = \log_2(4) = \boxed{2.}$

b.

θ	$0°$	$45°$	$90°$	$135°$
$H(R\|S = \lambda_\theta)$	27.45	7094.62	869114.22	78436652.48

The above values were computed by the following MATLAB code:

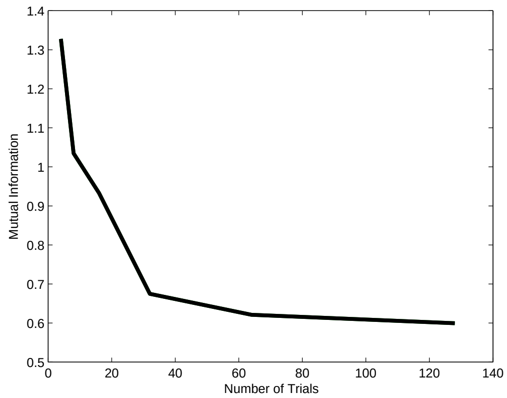
```
>> L = (2:2:8)';
>> N = 1:15;
>> for j=1:15, S(:,j) = gammaln(j).*(L.^(N(j)))./factorial(j); end
>> H = L.*(1 - log(L)) + exp(L).*sum(S,2);
```

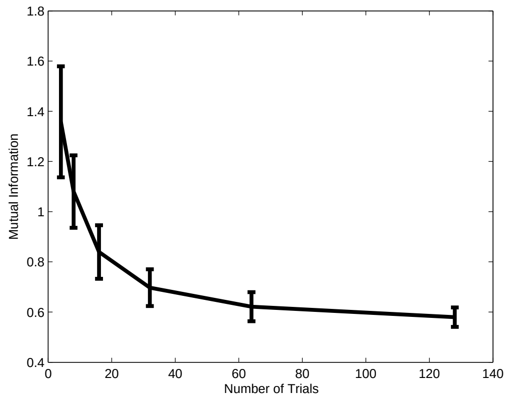
c.

$$H(R|S) = \frac{1}{4}\sum_{s\in S} H(R|S=s) = \boxed{19828222.19.}$$

d.

e. $MI(R;S) = H(R) - H(R|S)$

f. 

g. 

h. Both the mean and standard deviation in estimated mutual information decrease as the number of trials increases, sharply at first, and then more gradually.

i.

j.

Problem 6

The simulation took me quite a while this time, and I'm not sure I understand, theoretically, why data insufficiency causes mutual information to be overestimated. 1,2, and 4 were easy enough (4b. was especially nifty), but 3 involved a lot of calculation. The math for information theory seems really ugly...

The following MATLAB code was used to empirically compute mutual information for problem 5:

```
MI = zeros(100,6);

for trial=1:100
  num = 1;
  for N=[4 8 16 32 64 128];
    data = [poissrnd(2,1,N);
            poissrnd(4,1,N);
            poissrnd(6,1,N);
            poissrnd(8,1,N)];

    for i=1:4
      u = unique(data(i,:));
      prob{i} = zeros(2,length(u));
      prob{i}(1,:) = u;

      for j=1:length(u)
        prob{i}(2,j) = length(find(data(i,:) == prob{i}(1,j)))./(4*N);
      end
    end

    all = [prob{1} prob{2} prob{3} prob{4}];
    u = unique(all(1,:));
    pr = zeros(2,length(u));
    pr(1,:) = u;

    for i=1:length(u)
      pr(2,i) = sum(all(2,find(all(1,:) == pr(1,i))));
    end

    HR = -sum(pr(2,:).*log2(pr(2,:)));
    HRS = -sum(all(2,:).*log2(4.*all(2,:)));
    MI(trial,num) = HR - HRS;

    num = num + 1;
  end
end
```