

CS 512, Spring 2014

Assignment 1

Shan Sikdar

Due Monday January 27th

1 (*) 2.1

Verify $\sum_{x=0}^1 p(x|\mu) = 1$

Proof. $\sum_{x=0}^1 p(x|\mu) = p(x=0|\mu) + p(x=1|\mu) = \mu + 1 - \mu = 1$

Verify: $\mathbb{E}[x] = \mu$

Proof. $\mathbb{E}[x] = \sum_x x p(x|\mu) = 1 * \mu + 0 * (1 - \mu) = \mu$

Verify $var[x] = \mu(1 - \mu)$

Proof.

$$\begin{aligned} var(x) &= \mathbb{E}[(x - E[x])^2] \\ &= (1 - \mu)(0 - \mu)^2 + \mu(1 - \mu)^2 \\ &= (1 - \mu)(\mu)^2 + \mu(1 - \mu)^2 \\ &= \mu - \mu^2 = \mu(1 - \mu) \end{aligned}$$

Show entropy of $H[x]$ of a Bernoulli distributed random variable x is given

by : $H[x] = -\mu \ln \mu - (1 - \mu) \ln(1 - \mu)$

By definition of entropy: $H[x] = -\sum_x p(x|\mu) \ln[p(x|\mu)]$

$$= -[(1 - \mu) \ln(1 - \mu) + (\mu \ln \mu)] = -\mu \ln \mu - (1 - \mu) \ln(1 - \mu)$$

2 2.2

Show that the distribution (2.261) is normalized and evaluate its mean, variance, and entropy.

Show normalized: $\sum_{x=-1,1} p(x|\mu) = 1$

proof.

$$\sum_{x=-1,1} p(x|\mu) = \frac{(1-\mu)^1}{2} \frac{(1+\mu)^0}{2} + \frac{(1-\mu)^0}{2} \frac{(1+\mu)^1}{2} = \frac{(1-\mu)}{2} + \frac{(1+\mu)}{2} = \frac{2}{2} = 1$$

Evaluate its mean:

$$= \sum_x f(x) p(x) = (-1) \left(\frac{1-\mu}{2} \right) + (1) \left(\frac{1+\mu}{2} \right) = \frac{2\mu}{2} = \mu$$

Evaluate its Variance:

$$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = (-1)^2 \left(\frac{1-\mu}{2} \right) + (1)^2 \left(\frac{1+\mu}{2} \right) + \mu^2 = 1 - \mu^2$$

Evaluate its Entropy:

$$H[x] = - \sum_x p(x|\mu) \ln[p(x|\mu)] = - \left[\left(\frac{1-\mu}{2} \right) \ln \left(\frac{1-\mu}{2} \right) + \left(\frac{1+\mu}{2} \right) \ln \left(\frac{1+\mu}{2} \right) \right] = - \left(\left(\frac{1-\mu}{2} \right) \ln \left(\frac{1-\mu}{2} \right) \right) - \left(\left(\frac{1+\mu}{2} \right) \ln \left(\frac{1+\mu}{2} \right) \right)$$

3 2.7

Show that the posterior mean value of x lies between the prior mean and the maximum likelihood estimate for μ . To do this show the posterior mean can be written as λ times the prior mean plus $(1 - \lambda)$ times the maximum likelihood estimate where $0 \leq \lambda \leq 1$.

From 2.15 the mean of the prior is $\frac{a}{a+b}$

From 2.20 the mean of the posterior is $\frac{m+a}{m+a+l+b}$

From class the maximum likelihood estimate is $\frac{m}{m+l}$

So want to show $\lambda \left(\frac{a}{a+b} \right) + (1 - \lambda) \left(\frac{m}{m+l} \right) = \frac{a+m}{a+b+l+m}$ where $0 \leq \lambda \leq 1$.

So solve for λ and argue its range is $0 \leq \lambda \leq 1$:

$$\begin{aligned} \lambda \left(\frac{a}{a+b} \right) + (1 - \lambda) \left(\frac{m}{m+l} \right) &= \frac{a+m}{a+b+l+m} \\ \left(\frac{a\lambda}{a+b} \right) + \left(\frac{m-m\lambda}{m+l} \right) &= \frac{a+m}{a+b+l+m} \\ \frac{(m+l)(a\lambda) + (a+b)(m-m\lambda)}{(a+b)(m+l)} &= \frac{a+m}{a+b+l+m} \\ \frac{a\lambda l + am + bm - b\lambda m}{(a+b)(m+l)} &= \frac{a+m}{a+b+l+m} \end{aligned}$$

$$\begin{aligned}
\frac{a\lambda l - b\lambda m}{(a+b)(m+l)} + \frac{am+bm}{(a+b)(m+l)} &= \frac{a+m}{a+b+l+m} \\
\frac{a\lambda l - b\lambda m}{(a+b)(m+l)} + \frac{m}{(m+l)} &= \frac{a+m}{a+b+l+m} \\
\frac{a\lambda l - b\lambda m}{(a+b)(m+l)} &= \frac{a+m}{a+b+l+m} - \frac{m}{(m+l)} \\
\frac{a\lambda l - b\lambda m}{(a+b)(m+l)} &= \frac{a+m}{a+b+l+m} - \frac{m}{(m+l)} \\
\frac{a\lambda l - b\lambda m}{(a+b)(m+l)} &= \frac{(a+m)(m+l)}{(a+b+l+m)(m+l)} - \frac{m(a+b+l+m)}{(a+b+l+m)(m+l)} \\
\frac{a\lambda l - b\lambda m}{(a+b)(m+l)} &= \frac{(al-bm)}{(a+b+l+m)(m+l)} \\
\lambda &= \frac{(al-bm)(a+b)(m+l)}{(al-bm)(a+b+l+m)(m+l)} \\
\lambda &= \frac{(a+b)}{(a+b+l+m)} \\
\lambda &= \frac{1}{1+\frac{l+m}{a+b}}
\end{aligned}$$

All variables in this equation are greater than 0. If $\frac{l+m}{a+b}$ gets larger, λ will tend to 0. If $\frac{l+m}{a+b}$ gets smaller, λ will tend to 1. So $0 \leq \lambda \leq 1$

4 2.8

Prove $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$ and $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

Prove $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$.

$$\begin{aligned}
E[x] &= \int_x p(x) x dx \\
&= \int_x \int_y p(x|y) p(y) x dy dx \text{ (using product rule)} \\
&= \int_y \int_x p(x|y) p(y) x dx dy \\
&= \int_y p(y) \int_x p(x|y) x dx dy \\
&= \int_y p(y) E_x[x|y] dy \\
&= E_y[E_x[x|y]]
\end{aligned}$$

Prove: $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

$$\begin{aligned}
var[x] &= \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\
&= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]]^2 \\
&= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]^2] + \mathbb{E}_y[\mathbb{E}_x[x|y]^2] - \mathbb{E}_y[\mathbb{E}_x[x|y]]^2 \\
&= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]^2] + var_y[\mathbb{E}_x[x|y]]
\end{aligned}$$

Since Expectation is a linear operator:

$$\begin{aligned}
&= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]^2] + var_y[\mathbb{E}_x[x|y]] \\
&= \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]
\end{aligned}$$

5 2.27

Let x and z be two independent random vectors, so that $p(x,z) = p(x)p(z)$.

Show that the mean of their sum $y = x + z$ is given by the sum of the means of each of the variables separately.

Proof: Let $x = [x_1, \dots, x_n]$ and $z = [z_1, \dots, z_n]$ then:

$$\frac{x_1+x_2+\dots+x_n}{n} + \frac{z_1+z_2+\dots+z_n}{n} = \frac{(x_1+z_1)+(x_2+z_2)+\dots+(x_n+z_n)}{n} = \frac{y_1+y_2+\dots+y_n}{n}$$

Show that the covariance matrix of y is given by the sum of the covariance matrices x and z .

Proof: Using formula 2.63:

$$\begin{aligned} \text{cov}[y] &= \mathbb{E}[(y - E[y])(y - E[y])^T] \\ &= \mathbb{E}[(x + z - E[x + z])(x + z - E[x + z])^T] \\ &= \mathbb{E}[(x - E[x] + z - E[z])(x - E[x] + z - E[z])^T] = \mathbb{E}[(x - E[x]) + (z - E[z])][(x - E[x])^T + (z - E[z])^T] \\ &= \mathbb{E}[(x - E[x])(x - E[x])^T + (x - E[x])(z - E[z])^T + (z - E[z])(x - E[x])^T + (z - E[z])(z - E[z])^T] \\ &= \mathbb{E}[(x - E[x])(x - E[x])^T] + \mathbb{E}[(x - E[x])(z - E[z])^T] + \mathbb{E}[(z - E[z])(x - E[x])^T] + \mathbb{E}[(z - E[z])(z - E[z])^T] \\ &= \text{cov}[x] + \text{cov}[z] + \text{cov}[x, z] + \text{cov}[z, x] \\ &= \text{cov}[x] + \text{cov}[y] \end{aligned}$$

Facts I used:

- (1) x, z independent, E is a linear operator
- (2) $(A + B)^T = A^T + B^T$
- (3) Since x and z are independent, the covariance matrices with respect to each other will be 0.

Confirm that this result agrees with that of exercise 1.10.

You can consider one variable to be a vector with one element so the first proof shows that 1.28 holds. In the single variable case, covariance reduces to just normal variance so 1.29 holds.