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.

$var(x) = \mathbb{E}[(x - E[x])^2]$

$= (1-\mu)(0-\mu)^2 + \mu(1-\mu)^2$

$= (1-\mu)(\mu)^2 + \mu(1-\mu)^2$

$= \mu - \mu^2 = \mu(1-\mu)$

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.
$\sum\limits_{x=-1,1} p(x|\mu) = \frac{(1-\mu)^1}{2}\frac{(1+\mu)^0}{2} + \frac{(1-\mu)^0}{2}\frac{(1+\mu)^1}{2} = \frac{(1-\mu)}{2} + \frac{(1+\mu)}{2} = \frac{2}{2} = 1$
Evaluate its mean:
$= \sum\limits_x f(x)\;\; p(x) = (-1)(\frac{(1-\mu)}{2}) + (1)(\frac{(1+\mu)}{2}) = \frac{2\mu}{2} = \mu$
Evaluate its Variance:
$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = (-1)^2(\frac{1-\mu}{2}) + (1)^2(\frac{1+\mu}{2}) + \mu^2 = 1 - \mu^2$
Evaluate its Entropy:
$H[x] = -\sum\limits_x p(x|\mu)\, ln[p(x|\mu)] = -[((\frac{1-\mu}{2})\, ln((\frac{1-\mu}{2})) + ((\frac{1+\mu}{2})\, ln((\frac{1+\mu}{2}))] =$
$-((\frac{1-\mu}{2})\, ln((\frac{1-\mu}{2}))\; - \;((\frac{1+\mu}{2})\, ln((\frac{1+\mu}{2}))$

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 $\lambda(\frac{a}{a+b}) + (1-\lambda)(\frac{m}{m+l}) = \frac{a+m}{a+b+l+m}$ where $0 \leq \lambda \leq 1$.
So solve for λ and argue its range is $0 \leq \lambda \leq 1$:

$\lambda(\frac{a}{a+b}) + (1-\lambda)(\frac{m}{m+l}) = \frac{a+m}{a+b+l+m}$
$(\frac{a\lambda}{a+b}) + (\frac{m-m\lambda}{m+l}) = \frac{a+m}{a+b+l+m}$
$\frac{(m+l)(a\lambda)+(m-m\lambda)(a+b)}{(a+b)(m+l)} = \frac{a+m}{a+b+l+m}$
$\frac{a\lambda l+am+bm-b\lambda m}{(a+b)(m+l)} = \frac{a+m}{a+b+l+m}$

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

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, lambda will tend to 0. 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 $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$.

$E[x] = \int_x p(x)\ x\ dx$
$= \int_x \int_y p(x|y)\ \ p(y)\ \ x\ dy\ dx$ (using product rule)
$= \int_y \int_x p(x|y)\ \ p(y)\ \ x\ dx\ dy$
$= \int_y p(y) \int_x p(x|y)\ x\ dx\ dy$
$= \int_y p(y)\ E_x[x|y]\ dy$
$= E_y[E[x|y]]$

Prove: $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$
$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$
$= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]]^2$
$= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]^2] + \mathbb{E}_y[\mathbb{E}_x[x|y]^2] - \mathbb{E}_y[\mathbb{E}_x[x|y]]^2$
$= \mathbb{E}_y[\mathbb{E}_x[x^2|y]] - \mathbb{E}_y[\mathbb{E}_x[x|y]^2] + var_y[\mathbb{E}_x[x|y]]$
Since Expectation is a linear operator:
$= \mathbb{E}_y[\mathbb{E}_x[x^2|y] - \mathbb{E}_x[x|y]^2] + var_y[\mathbb{E}_x[x|y]]$
$= \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

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

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

Show that the covariance atrix of y is given by the sum of the covariance matrices x and z.
Proof: Using formula 2.63:
$cov[y] = \mathbb{E}[(y - E[y])(y - E[y])^T]$
$= \mathbb{E}[((x + z) - E[(x + z)])((x + z)) - E[(x + z)])^T]$
$= \mathbb{E}[(x - E[x] + z - E[x])(x - E[x] + z - E[x])^T] = \mathbb{E}[(\ (x - E[x]) + (z - E[x])\ )(\ (x - E[x])^T + (z - E[x])^T)] = \mathbb{E}[(x - E[x])(x - E[x])^T + (x - E[x])(z - E[x])^T + (z - E[x])(x - E[x])^T + (z - E[x])(z - E[x])^T]$
$= \mathbb{E}[(x - E[x])(x - E[x])^T] + \mathbb{E}[(x - E[x])(z - E[x]^T)] + \mathbb{E}[(z - E[x])(x - E[x])^T] + \mathbb{E}[(z - E[x])(z - E[x])^T)]$
$= cov[x] + cov[z] + cov[x, z] + cov[z, x]$
$= cov[x] + cov[y]$

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 excercise 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 varince so 1.29 holds.

6 2.28

Consider a joint distribution over the variable $z = \binom{x}{y}$ whoose mean and covairiance are given by (2.108) and (2.105) respectively.

I. By making use of the results (2.92) and (2.93) show that the marginal distribution p(x) is given by $p(x) = N(x|\mu, \Lambda^{-1})$
So find the mean and covariance of the marginal distribution:
By 2.92 we know that marginal ditribution p(x) has the mean given by $E[x] = \mu_x$ but we also know that the mean vector for z is givien by $E[z] = \binom{\mu}{A\mu+b}$. Therefore $\mu_x = \mu$ and so the mean of the marginal distribution p(x) is μ. From 2.93, we know that the covariance of the marginal distribution p(x) is given by $cov[x] = \Sigma_{xx}$. But from 2.105 we know that covariance matrix for z is $cov[z] = \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L{-}1{+}A\Lambda^{-1}A^T \end{pmatrix}$
So: $\Lambda^{-1} = \Sigma_{xx} = cov[x]$ then Λ^{-1} is the covariance of the marginal distribution. Therefore $p(x) = N(x|\mu, \Lambda^{-1})$

II. Similarly, by making useof the results (2.81) and (2.82) show that the conditional distribution $p(y|x) = N(y|Ax + b, L^{-1})$
So find the mean and covariance of the conditonal distribution:
By 2.81:
$\mu_{y|x} = \mu_y + \Sigma_{xy}\Sigma_{xx}^{-1}(x - \mu_x)$
$= A\mu + b + (A\Lambda^{-1})(\Lambda^{-1})^{-1}(x - \mu)$ (using 2.108 and 2.105)
$= A\mu + b + (A\Lambda^{-1})(\Lambda)(x - \mu)$
$= A\mu + b + (A)(x - \mu)$
$= A\mu + b + (Ax - A\mu)$
$= Ax + b$
And so the mean for the conditonal distribution is $Ax + b$

By 2.82:
$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{yx}^{-1}\Sigma_{xy}$
$= L^{-1} + A\Lambda^{-1}A^T - (A\Lambda^{-1})(\Lambda^{-1})^{-1}(\Lambda^{-1}A^T)$ (using 2.105)
$= L^{-1} + A\Lambda^{-1}A^T - (A\Lambda^{-1})(\Lambda)(\Lambda^{-1}A^T)$
$= L^{-1} + A\Lambda^{-1}A^T - (A\Lambda^{-1}A^T)$
$= L^{-1}$
And so the variance for the conditonal distribution is L^{-1} Therefore the conditonal distribution is $N(y|Ax + b, L^{-1})$

7 2.31