Assignment-1

Kartik Srinivas

August 2022

1 Problem 1

1.1 Discrete - Discrete

Let us write the probability mass function. These stem from the basic definitions of conditional distributions

$$
\begin{aligned}
p(Y=y|X=x) &= \frac{P(Y=y, X=x)}{P(X=x)} \\
p(X=x|Y=y) &= \frac{P(Y=y, X=x)}{P(Y=y)} \\
\therefore p(Y=y|X=x) &= \frac{P(X=x|Y=y)P(Y=y)}{P(X=x)}
\end{aligned}
$$

1.2 Discrete- Continuous

The idea is to take an interval for the continuous random variable so that the denominator does not become zero

$$p(y|x) = \frac{p(Y=y, x<X<x+dx)}{p(x<X<x+dx)} \tag{1}$$

$$p(y|x) = \frac{p(x<X<x+dx, Y=y)}{p(x<X<x+dx)} \tag{2}$$

$$p(y|x) = \frac{p(x<X<x+dx|Y=y)p(Y=y)}{p(x<X<x+dx)} \tag{3}$$

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{4}$$

1.3 Continuous- Continuous

$$
\begin{aligned}
p(y|x) &= p(y<Y<y+dy|x<X<x+dx) \\
p(y|x) &= \frac{p(y<Y<y+dy, x<X<x+dx)}{p(x<X<x+dx)}
\end{aligned}
$$

This has now boiled down to the discrete-discrete case

$$p(y|x) = \frac{p(x < X < x + dx, y < Y < y + dy)}{p(x < X < x + dx)}$$
$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Q.E.D ■

1.4 Total Expectation

Let $f(X, Y)$ be another random variable K (for notational simplicity) then K may or may not be continuous.

1.4.1 K is continuous, X is continuous

The definition of $E[K|X] =$

$$\begin{aligned} &\int_{k=0}^{\infty} y f_{k|x}(k|x)\ dk \\ &= \int_{k=0}^{\infty} k \frac{f(k,x)}{f_X(x)}\ dk \end{aligned}$$

The definition of $E[E[K|X]] =$

$$\begin{aligned} &\int_{x=0}^{x=\infty} \int_{k=0}^{\infty} y f_{k|x}(k|x)\ dk\ f_X(x)\ dx \\ &= \int_{x=0}^{x=\infty} \int_{k=0}^{\infty} k \frac{f(k,x)}{f_X(x)}\ dk\ f_X(x)\ dx \\ &= \int_{x=0}^{x=\infty} \int_{k=0}^{k=\infty} k f(k,x)\ dk\ dx \\ &= E[K] \end{aligned}$$

1.4.2 K is continuous, X is discrete

The only thing that changes here is the definition and evaluation of the probabilities

$$E[K|X=x] = \int_k k f_{K|X}(K=k|X=x) \tag{5}$$

$$E[E[K|X]] = \sum_x \int_k k \times f_{K|X}(K|X=x)P(X=x)\ dk \tag{6}$$

$$= \sum_x \int_k k \times f_{K,X}(K=k, X=x)\ dk \tag{7}$$

$$= \int_k k \sum_x f_{K,X}(k,x)\ dk \tag{8}$$

$$= \int_k f_k(K)\ dk \ldots\ (\because \text{Definition of joint of continuous and discrete}) \tag{9}$$

$$= E[K] \tag{10}$$

1.4.3 K is discrete, X is discrete

$$E[K|X=x] = \sum_k kP_{K|X}(K=k|X=x) \tag{11}$$

$$E[E[K|X]] = \sum_x \sum_k k \times P_{K|X}(K=k|X=x)P(X=x) \tag{12}$$

$$= \sum_x \sum_k k \times P_{K,X}(K=k, X=x) \tag{13}$$

$$= \sum_k k \sum_x P_{K,X}(k,x) \tag{14}$$

$$= \sum_k kP_k(K) \ldots\ (\because \text{Definition of joint of continuous and discrete}) \tag{15}$$

$$= E[K] \tag{16}$$

2 Problem 2

Since we only want the **total** number of injuries per week , let this random variable be X, and let the **total** number of accidents be another random variable Y we know the following

$$\mathbb{E}[Y] = 4 \tag{17}$$

$$\mathbb{E}[(\frac{X}{y})|Y=y] = 2 \tag{18}$$

$$\therefore \mathbb{E}[(X)|Y=y] = 2y \tag{19}$$

Now since y is only a constant for (18) , we get to use this result, since the y can be taken out of the expectation ....(due to linearity of expectation)

$$\mathbb{E}_y[\mathbb{E}[X|Y = y]] = \mathbb{E}_x[X] \tag{20}$$

All that's left is to multiply after substitution of (3)

$$\boxed{\mathbb{E}[X] = \mathbb{E}_y[2y] = 2\,\mathbb{E}_y[y] = 2 \times 4 = 8}$$

3 Problem 3

let $z \in \mathbb{R}^{m+n}$, represent the Random variable that has a joint probability density function $p(\vec{z}, \mu_z, \Sigma_z)$ of random vectors $\vec{x}$ and $\vec{y}$

$$\vec{z} = \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} \tag{21}$$

$$p(\vec{z}, \mu_z, \Sigma_z) = \frac{1}{\sqrt{(2\pi)^{m+n} |\,\Sigma_z\,|}} \exp\left(-\frac{1}{2}(z - \mu_z)^T\, \Sigma_z^{-1}(z - \mu_z)\right) \tag{22}$$

We must find Σ_z^{-1} in terms of Σ_x and Σ_y 1 in order to calculate the exact marginals

Firstly note that the covariances of each entry of x lies inside the Σ_z. It is trivial to see that (since the corss multiplication remains the same)

$$\Sigma_z = \mathbb{E}[\begin{bmatrix} \vec{x} - \mu_x \\ \vec{y} - \mu_y \end{bmatrix} \times \begin{bmatrix} \vec{x} - \mu_x & \vec{y} - \mu_y \end{bmatrix}] \tag{23}$$

$$\Sigma_z = \mathbb{E}[\begin{bmatrix} (\vec{x} - \mu_x)(\vec{x} - \mu_x) & (\vec{x} - \mu_x)(\vec{y} - \mu_y) \\ (\vec{y} - \mu_y)(\vec{x} - \mu_x) & (\vec{y} - \mu_y)(\vec{x} - \mu_x) \end{bmatrix}] \tag{24}$$

$$\Sigma_z = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \tag{25}$$

Now using these equations we can write down the problem as follows.

$$p(\vec{y}) = \int \cdots \int_{\vec{\mathbf{x}} \in \mathbb{R}^m} \frac{1}{\sqrt{C}} \exp\left(-\frac{1}{2}(z - \mu_z)^T\, \Sigma_z^{-1}(z - \mu_z)\right) d\vec{\mathbf{x}} \tag{26}$$

3.1 Separating $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ in the integral

The trick to separate x and y in integral is to first assume the inverse of the the covariance matrix of Z to be

$$\Sigma_z^{-1} = \begin{bmatrix} \mathrm{A}_{xx} & \mathrm{A}_{xy} \\ \mathrm{A}_{yx} & \mathrm{A}_{yy} \end{bmatrix} \tag{27}$$

We use an important fact form linear algebra . i.e

$$\begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix}^T \begin{bmatrix} \mathrm{A}_{xx} & \mathrm{A}_{xy} \\ \mathrm{A}_{yx} & \mathrm{A}_{yy} \end{bmatrix} \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \vec{x}^T\, \mathrm{A}_{xx}\, \vec{x} + \vec{y}^T\, \mathrm{A}_{yx}\, \vec{x} + \vec{x}^T\, \mathrm{A}_{xy}\, \vec{y} + \vec{y}^T\, \mathrm{A}_{yy}\, \vec{y} \tag{28}$$

We use this expansion along with the following "Completing the square trick"

$$\frac{1}{2}z^T A z + b^T Z + c = \frac{1}{2}(z + A^{-1}b)^T A(z + A^{-1}b) + c - \frac{1}{2}b^T A^{-1} b \tag{29}$$

Using (29) and (28) the integral (26) using the following shorthand notation

$$\vec{\mathbf{x}} = \vec{\mathbf{x}} = \mu_x$$

and

$$\vec{\mathbf{y}} = \vec{\mathbf{y}} - \mu_y$$

$$p(\vec{y}) = \int \cdots \int_{\vec{\mathbf{x}} \in \mathbb{R}^m} \frac{1}{\sqrt{C}} \exp\left(\frac{-1}{2} (\vec{x}^T \mathrm{A}_{xx}\, \vec{x} + \vec{y}^T \mathrm{A}_{yx}\, \vec{x} + \vec{x}^T \mathrm{A}_{xy}\, \vec{y} + \vec{y}^T \mathrm{A}_{yy}\, \vec{y}) \right) d\vec{\mathbf{x}}$$

removing the parts that pertain to y outside the integral we get

$$p(\vec{y}) = \frac{1}{\sqrt{C}} \exp\big(-1/2 * \vec{y}^T \mathrm{A}_{yy}\, \vec{y}\big) \int \cdots \int_{\vec{\mathbf{x}} \in \mathbb{R}^m} \exp\left(\frac{-1}{2} (\vec{y}^T \mathrm{A}_{yx}\, \vec{x} + \vec{x}^T \mathrm{A}_{xy}\, \vec{y} + \vec{x}^T \mathrm{A}_{xx}\, \vec{x}) \right)) \, d\vec{\mathbf{x}}$$

Now we use (29) to simplify the quadratic form inside the integral

$$p(\vec{y}) = \frac{1}{\sqrt{C}} \exp\left(\frac{-1}{2} \vec{\mathbf{y}}^T A_{yy} \vec{\mathbf{y}} + 1/2 * \vec{\mathbf{y}}^T A_{yx} A_{yy}^{-1} A_{xx} \vec{\mathbf{y}} \right) \times \tag{30}$$

$$\boxed{\int \exp\big((\vec{\mathbf{x}} + \mathrm{A}_{xx}^{-1}\, A_{yx} y)^T A_{xx} (\vec{\mathbf{x}} + \mathrm{A}_{xx}^{-1}\, A_{yx} y) \big) \, d\vec{\mathbf{x}}} \tag{31}$$

The boxed integral on the right looks very very similar to the integral of the pdf of a gaussian distribution (leaving only the normalization constant) This gaussian integral can be evaluated and will just be equal to the constant on the LHS

$$\sqrt{(2\pi)^{m+n} |\, \Sigma_z \,|} = \int_{\vec{\mathbf{z}} \in \mathbb{R}^d} \exp\left(-\frac{1}{2} (z - \mu_z)^T \, \Sigma_z^{-1} (z - \mu_z) \right)$$

So we finally get p(y), by using this trick to solve the integral on the right!(we also use the fact that quadratic forms in one variable add to give quadratic forms in the same variable) and resubstituting values of $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ from the shorthand notation below (29)

$$p(\vec{y}) = \boxed{\frac{(2\pi)^{m/2} |A_{xx}|^{1/2}}{\sqrt{C}} \exp\left(\frac{-1}{2} ((\vec{\mathbf{y}} - \mu_y)^T (A_{yy} - A_{yx} A_{yy}^{-1} A_{xx}) (\vec{\mathbf{y}} - \mu_y)) \right)} \tag{32}$$

This is clearly a gaussian if the position of the covariance matrix matches Σ_{yy}. The term in the middle $(A_{yy} - A_{yx} A_{yy}^{-1} A_{xx})$ can be later on showed to be equal to Σ_y using properties of inverses of block matrices. The properties that we will use are as follows

$$\begin{aligned}\begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} &= \begin{bmatrix} A_{yy} & A_{yx} \\ A_{xy} & A_{xx} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \left(A_{yy} - A_{yx}A_{xx}^{-1}A_{xy}\right)^{-1} & -\left(A_{yy} - A_{yx}A_{xx}^{-1}A_{xy}\right)^{-1} A_{yx}A_{xx}^{-1} \\ -A_{xx}^{-1}A_{xy}\left(A_{yy} - A_{yx}A_{xx}^{-1}A_{xy}\right)^{-1} & \left(A_{xx} - A_{xy}A_{yy}^{-1}A_{yx}\right)^{-1} \end{bmatrix}\end{aligned}$$

We immediately see that $\left(A_{yy} - A_{yx}A_{xx}^{-1}A_{xy}\right)^{-1} = \Sigma_{yy}$, this proves that the probability distributions is a gaussian, moreover it contains exactly that sub-block of the mean and covariance -matrices as that of the joint.

$$\boxed{p(\vec{y}, \mu_y, \Sigma_y) = \frac{1}{\sqrt{(2\pi)^n |\,\Sigma_y\,|}} \exp\left(-\frac{1}{2}(y-\mu_y)^T\,\Sigma_y^{-1}(y-\mu_y)\right)} \tag{33}$$

3.2 Conditional of variables with Joint gaussian

The approach is quite similar, I will basically use the same quadratic trick, the only change is that the marginal in the denominator will get absorbed into the constant

$$p(y|x) = \frac{p(y,x)}{\int_{\vec{\mathbf{y}} \in \mathbb{R}^n} p(x,y)\ d\vec{\mathbf{y}}} \tag{34}$$

$$\text{The denominator will not depend on } \vec{\mathbf{y}} \tag{35}$$

Hence the denominator can be absorbed into the normalization constant of the Numerator On doing this we obtain

$$p\,(y \mid x) = \frac{p\,(x, y; \mu, \Sigma)}{\int_{y \in \mathbf{R}^n} p\,(x, y; \mu, \Sigma)\,dy} \tag{36}$$

$$= \frac{1}{Z'} \exp\left(-\frac{1}{2}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}^T \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}\right) \tag{37}$$

Now we use the **same completion of squares trick** in (29) that we used in the previous proof after assuming the inverse of the covariance to be a matrix **A** Shorthand notation like last time:-

$$\vec{\mathbf{x}} = \vec{\mathbf{x}} = \mu_x$$

and

$$\vec{\mathbf{y}} = \vec{\mathbf{y}} - \mu_y$$

on splitting the matrix A using 29 we obtain this integral

$$p(\vec{y}|\vec{x}) = \frac{1}{Z'} \exp \boxed{\left(\frac{-1}{2}\vec{\mathbf{x}}^T A_{xx}\vec{\mathbf{x}} + 1/2 * \vec{\mathbf{x}}^T A_{xy}A_{xx}^{-1}A_{yy}\vec{\mathbf{x}}\right)} \times \exp\left((\vec{\mathbf{y}} + \mathrm{A}_{yy}^{-1}\,A_{yx}x)^T A_{yy}(\vec{\mathbf{y}} + \mathrm{A}_{yy}^{-1}\,A_{yx}x)\right) \tag{38}$$

The boxed part is a constant and can be absorbed, expanding the shorthand we get

$$\boxed{p\left(y \mid x\right) = \frac{1}{Z''} \exp\left(-\frac{1}{2}\left(y-\mu_y+A_{yy}^{-1}A_{yx}\left(x-\mu_x\right)\right)^T A_{yy}\left(y-\mu_y+A_{yy}^{-1}A_{yx}\left(x-\mu_x\right)\right)\right)} \tag{39}$$

This is a gaussian with the Covariance equal to A_{yy}^{-1} and mean $\mu_y - A_{yy}^{-1}A_{yx}\left(x-\mu_x\right)$

4 Problem 4

Let X, Y be a random vectors that satisfy

$$Y = WX + b \tag{40}$$

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ . \\ . \\ X_n \end{bmatrix}$$

The expectations follow linearity, i.e expectation of a linear in 2 random variables, is the linear of an expectation of 2 random variables

$$Y_i = \sum_{j=1}^{j=n} W_{ij} X_j + b_j \tag{41}$$

$$\mathbb{E}[Y_j] = \sum_j W_{ij}\,\mathbb{E}[X_j] + b_j \ldots \text{(Linearity of expectation of 2 Random Variables)} \tag{42}$$

$$\boxed{\therefore \mathbb{E}[\vec{Y}] = W\,\mathbb{E}[\vec{X}] + \vec{b}} \tag{43}$$

The Covariances are defined as

$$\mathbb{E}[(X-\mu_x)(X-\mu_x)^T] = \Sigma_X \tag{44}$$

$$\begin{aligned} Y-\mu_y &= W(X-\mu_x) \ldots (40)\ \&\ (43) \\ \therefore \Sigma_y &= \mathbb{E}[W(X-\mu_x)(W(X-\mu_x))^T] \\ \therefore \Sigma_y &= W\,\mathbb{E}[(X-\mu_x)(X-\mu_x)^T]W^T \end{aligned}$$

$$\boxed{\Sigma_y = W\,\Sigma_x\,W^T} \tag{45}$$

5 Problem 1

We have already proved in class that for a 0-1 loss the function that is modelled achieves the mode of the posterior distribution, the mode is

$$M_o = \arg\max(p(0|x), p(1|x)) \tag{46}$$

Now the bayes optimal

$$f(x) = \arg\min(p(1|x) \mid_{y=0}, p(0|x) \mid_{y=1})$$

$$f(x) = \arg\min_{y} \frac{(1, e^{(x-2)(x-3)})}{1 + e^{(x-2)(x-3)}}$$

The quadratic assumes negative values between 2 and 3, and is positive elsewhere, that solves the problem

$$f(x) = \begin{cases} 0 & \text{if } x \geq 3 \text{ or } x \leq 2 \\ 1 & \text{if } x \leq 3 \text{ and } x \geq 2 \end{cases} \tag{47}$$

5.1 Problem 2

The only difference in this function is the loss, the loss will appear only when the assumed value of $\mathbf{y}$ differs from that of $\mathbf{Y}$

$$f(x) = \underset{y \in \{0,1\}}{\operatorname{argmin}} \mathbb{E}[l(\mathbf{y}, \mathbf{Y})] \tag{48}$$

$$\tag{49}$$

Now we basically only take cases where $\vec{\mathbf{y}}$ differs from $\vec{\mathbf{Y}}$,rest will be 0 loss

$$\mathbb{E}[l(\mathbf{y}, \mathbf{Y})] = \begin{cases} 0.5p(1|x) & \text{if } \vec{\mathbf{y}} = 0 \\ 2p(0|x) & \text{if } \vec{\mathbf{y}} = 1 \end{cases}$$

$$f(x) = \begin{cases} 1 & \text{if } 0.5p(1|x) > 2p(0|x) \\ 0 & \text{if } 2p(0|x) > 0.5p(1|x) \end{cases} \tag{50}$$

$$f(x) = \begin{cases} 1 & \text{if } 0.5p(1|x) > 2p(0|x) \\ 0 & \text{if } 2p(0|x) > 0.5p(1|x) \end{cases}$$

On solving the inequality and plotting we get

$$(x-2)(x-3) \geq -ln(4) \quad \forall x \in \mathbb{R}$$

$$\boxed{\therefore f^*(x) = 0 \quad \forall x \in \mathbb{R}}$$

5.2 Problem 3

The distribution is **exponential**, the $\Lambda(x)$ is always greater than 0. Because it is the maximum eigen value of a sum of positive semidefinite matrices(multiplied by a nonzero x_i) and added to a positive definite matrix. therefore $\Lambda(x) \geq 0$

$$p(y|x) = \Lambda(x)e^{-\Lambda(x)y}$$

The baye's optimal will assume the mean of the geometric distribution which is the reciprocal of the decay factor

$$\boxed{f^*(x) = \frac{1}{\Lambda(x)}}$$

5.3 Problem 4

This means we have to minimize the "sum" of distances , in a way this is the median of the distribution(For more see the part below this proof) the median can be calculated at the value of y for which $F_{y|x}(y) = 0.5$

$$1 - e^{-\Lambda(x)y} = 0.5 \tag{51}$$

$$\therefore e^{-\Lambda(x)y} = 0.5 \tag{52}$$

$$y = \frac{\ln(2)}{\Lambda(x)} \tag{53}$$

$$\boxed{f^*(x) = \frac{\ln(2)}{\Lambda(x)}} \tag{54}$$

5.3.1 Why is it the median ?

$$f^*(x) = \arg\min_y \int_0^\infty |(y - y')| p(y'|x) dy'$$

We will take the derivative of this function w.r.t y and set it to 0

$$f^*(x) = y| \quad \frac{\partial}{\partial y} \int_0^\infty |(y - y')| p(y'|x) dy' = 0$$

We will evaluate the integral using Leibnitz rule after splitting the integral into two pieces

$$I'(x) = \int_{u(x)}^{v(x)} \frac{\partial f}{\partial x}(t, x) \mathrm{d}t + f(v(x), x)v'(x) - f(u(x), x)u'(x)$$

$$f^*(x) = y| \quad \frac{\partial}{\partial y} \int_0^y (y - y')p(y'|x)dy' + \frac{\partial}{\partial y} \int_y^\infty (y' - y)p(y'|x)dy' = 0$$

Applying Leibniz rule and simplifying , we obtain an interesting condition

$$2 \int_0^y p(y'|x)dx = 1 \implies \int_0^y p(y'|x)dx = 0.5$$

$$\boxed{F_{y|x}(y|x) = 0.5}$$

This is the median!
