HOMEWORK-3 | MATH 541A

Problem 1: MLE for Poisson Regression

Suppose $y_i \in \mathbb{N}$ and $x_i \in \mathbb{R}$. Assume y_i are independent Poisson random variables such that $P(y_i | x_i) = \frac{e^{-\lambda_i} \lambda^{y_i}}{y_i!}$ where $\lambda_i > 0$.

Let $\eta_i = \ln \lambda_i$ and assume $\eta_i = \langle x_i, \theta \rangle$

1. Write the pmf $\mathbb{P}(y_i | \lambda_i)$ in an exponential form.

Sol^n: $\mathbb{P}(y | \lambda) = \frac{1}{y!} \exp\{-\lambda\} \lambda^{y}$. Since $\eta = \ln \lambda$, $\lambda = e^{\eta}$.

$$\therefore \mathbb{P}(y|\lambda) = \frac{1}{y!} \exp\{-e^{\eta}\} e^{\eta y} = \frac{1}{y!} \exp\{\eta y - e^{\eta}\}$$

$$\therefore \mathbb{P}(y|\lambda) = h(y) \exp\{\eta \cdot T - A(\eta)\}$$ where $h(y) = 1/y!$, $\eta(\lambda) = \ln \lambda$

$T(y) = y$ and $A(\eta) = e^{\eta}$.

2. Given n independent data $(x_1, y_1) \ldots (x_n, y_n)$, write down the likelihood function $L(\theta | x_1, y_1, \ldots x_n, y_n)$.

Sol^n: Recall that $\eta = \ln \lambda = \langle x, \theta \rangle$ which implies $\lambda = e^{\langle x, \theta \rangle}$

$$L(\theta) := L(\theta | x_1, y_1, x_2, y_2 \ldots x_n, y_n) = p(x_1, y_1, x_2 y_2 \ldots x_n, y_n | \theta)$$

$$L(\theta) = p(x_1 y_1 | \theta) \cdot p(x_2, y_2 | \theta) \cdots p(x_n, y_n | \theta).$$

$$p(x, y | \theta) = p(y | x, \theta) \cdot p(x | \theta) = p(y | \lambda) \cdot p(x | \theta).$$

As data is independent of η (hence θ), $p(x|\theta) = p(x)$.

$$\therefore p(x, y | \theta) = p(y|\lambda) \cdot p(x|\theta) = p(y|\lambda) p(x).$$

$$\Rightarrow p(x_1, y_1 | \theta) \cdots p(x_n, y_n | \theta) = p(y_1 | \lambda_1) \cdots p(y_n | \lambda_n) \cdot \prod_{i=1}^{n} p(x_i)$$

$$\Rightarrow \prod_{i=1}^{n} p(x_i, y_i | \theta) = \prod_{i=1}^{n} p(y_i | \lambda_i) \cdot \prod_{i=1}^{n} p(x_i)$$

$$\Rightarrow L(\theta) = \underbrace{\prod_{i=1}^{n} h(y_i)}_{H(y)} \cdot \exp\left\{\sum_{i=1}^{n} \eta_i y_i - \sum_{i=1}^{n} e^{\eta_i}\right\} \underbrace{\prod_{i=1}^{n} p(x_i)}_{G(x)}$$

$$\Rightarrow L(\theta) = H(y) G(x) \exp\left\{\sum_{i=1}^{n} [\eta_i y_i - e^{\eta_i}]\right\}$$

$$\Rightarrow L(\theta) = H(y) G(x) \exp\left\{\sum_{i=1}^{n} [\langle x_i, \theta \rangle y_i - e^{\langle x_i, \theta \rangle}]\right\}$$

3. Write the MLE for θ as a solution to an optimization problem.

Soln: $\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\ L(\theta)$

$$= \underset{\theta}{\operatorname{argmax}}\ H(y)\, G(x) \exp\left\{ \sum_{i=1}^{n} \left[\langle x_i, \theta \rangle y_i - e^{\langle x_i, \theta \rangle} \right] \right\}$$

$$= \underset{\theta}{\operatorname{argmax}}\ \exp\left\{ \sum_{i=1}^{n} \left[\langle x_i, \theta \rangle y_i - e^{\langle x_i, \theta \rangle} \right] \right\}$$

Problem 2: Consider the following questions:

Bayesian Estimator

1. Assume $X_1, X_2, \ldots, X_n$ are sampled i.i.d. from Poisson(λ).

a) Find the conjugacy family for a Poisson(λ) distribution.

Soln: We know from the lecture notes that the conjugacy family of Possion distribution is $\lambda \sim \text{Gamma}(\alpha, \beta)$, i.e. a Gamma distribution

$$f(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta}, \quad \lambda > 0, \quad \alpha, \beta >$$

b) Assuming the prior of λ is from the Conjugacy family, compute the posterior distribution.

Let $x \sim$ Poisson(λ) and the prior be $\lambda \sim$ Gamma(α, β)

$$f(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad f(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}$$

$$\text{posterior} = f(\lambda \mid \vec{x}) \propto f(\vec{x} \mid \lambda) f(\lambda)$$

$$\Rightarrow f(\lambda \mid \vec{x}) \propto \lambda^{\sum x_i} e^{-\lambda n} \cdot \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\propto \underbrace{\lambda^{\sum x_i + \alpha - 1} \cdot e^{-\lambda(\beta+n)}}_{\text{Gamma}\left(\sum_{i=1}^{n} x_i + \alpha - 1,\ \beta + n\right)}$$

c) Compute the posterior mean and the MAP estimators for λ.

Soln: Since the mean of Gamma(α, β) is α/β, the posterior mean estimate is

$$\bar{\lambda} = \frac{\sum_{i=1} x_i + \alpha}{\beta + n}$$

The MAP estimate is $\hat{\lambda}_{MAP} = \underset{\lambda}{\operatorname{argmax}}\ f(\lambda \mid \vec{x}) = \underset{\lambda}{\operatorname{argmax}}\ \lambda^{\sum x_i + \alpha - 1} \cdot e^{-\lambda(\beta+n)}$

$$\therefore \hat{\lambda}_{MAP} = \underset{\lambda}{\operatorname{argmax}}\ \lambda^{\sum x_i + \alpha - 1} \cdot e^{-\lambda(\beta+n)} = \underset{\lambda}{\operatorname{argmax}}\ \text{Gamma}\left(\sum x_i + \alpha,\ \beta + n\right)$$

Since the mode of a Gamma distibution is $\frac{\alpha - 1}{\beta}$ and MAP gives us this mode, $\hat{\lambda}_{MAP} = \frac{(\sum x_i + \alpha - 1) - 1}{\beta + n} = \frac{\sum x_i + \alpha - 2}{\beta + n}$

2. Let $X_1 \ldots\ldots X_n$ be iid samples from $N(\theta, \sigma^2)$ where σ^2 is a known constant. Let the prior distribution of θ be a Laplace distribution, that is, $\pi(\theta) = \frac{1}{2a} e^{-|\theta|/a}$,

where a is a known constant. Note that in this case, the prior distribution is not from the conjugacy family.

a) Find the posterior distribution $\pi(\theta \mid X_1 \ldots X_n)$.

Soln: $x \sim N(\theta, \sigma^2)$ and $\theta \sim \pi(\theta)$

$$\text{posterior} = \pi(\theta \mid x_1 \ldots x_n) = f(x_1 \ldots x_n \mid \theta) \cdot \pi(\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x_i-\theta)^2}{2\sigma^2}\right\} \cdot \frac{1}{2a} \exp\{-|\theta|/a\}$$

$$= \left[\frac{1}{2\pi\sigma^2}\right]^n \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\theta)^2\right\} \cdot \frac{1}{2a} \exp\left\{\frac{-|\theta|}{a}\right\}$$

$$= \frac{1}{2a} \cdot \left[\frac{1}{2\pi\sigma^2}\right]^n \exp\left\{\frac{-1}{2}\left(\frac{+1}{\sigma^2} \sum_{i=1}^{n} (x_i-\theta)^2 + \frac{2|\theta|}{a}\right)\right\}$$

This distribution, the posterior, seems to be from exponential family. Simplifying further,

$$\pi(\theta \mid x_1 x_2 x_3 \ldots x_n) \propto \exp\left\{\frac{-1}{2}\left(\frac{1}{\sigma^2} \sum (x_i-\theta)^2 + \frac{2|\theta|}{a}\right)\right\}$$

$$\Rightarrow \pi(\theta \mid x \ldots x_n) \propto \exp\left\{\frac{-1}{2}\left(\frac{1}{\sigma^2}\left[\sum x_i^2 + n\theta^2 - 2\theta \sum x_i\right] + 2|\theta|/a\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2} \cdot \sum x_i^2\right\} \exp\left\{\frac{-1}{2}\left(\frac{1}{\sigma^2}\left[n\theta^2 - 2\theta \sum x_i\right] - 2|\theta|/a\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{n\theta^2 - 2\theta \sum x_i}{\sigma^2} - \frac{2|\theta|}{a}\right)\right\}$$

$$\propto \exp\left\{\frac{-n\theta^2 + 2\theta \sum x_i}{2\sigma^2}\right\} \cdot \exp\left\{\frac{-|\theta|}{a}\right\}$$

The posterior's mean can be computed as $\bar{\theta} = \int_{-\infty}^{\infty} \theta \, \pi(\theta \mid x_1 \ldots x_n) \, d\theta$.

This would be the posterior mean estimate. I am not sure if there is a closed form value of this integral but one can always use numerical methods.

For the MAP estimate, $\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ \pi(\theta \mid x_1, \dots x_n)$

$$\Rightarrow \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ \exp\left\{ +\frac{1}{2}\left(\frac{-n\theta^2 + 2\theta \sum x_i}{\sigma^2} + \frac{2|\theta|}{a} \right) \right\}$$

$$= \underset{\theta}{\operatorname{argmax}} \left(\frac{2\theta \sum x_i - n\theta^2}{\sigma^2} + \frac{2|\theta|}{a} \right)$$

$$= \underset{\theta}{\operatorname{argmax}} \left(\frac{n\theta[2\bar{x} - \theta]}{\sigma^2} + \frac{2|\theta|}{a} \right), \quad \bar{x} := \frac{\sum x_i}{n}$$

Let $g(\theta) = \left[\frac{n\theta[2\bar{x} - \theta]}{\sigma^2} + \frac{2|\theta|}{a} \right]$

$$g'(\theta) = \left[\frac{n}{\sigma^2}\left(-\theta + 2\bar{x} - \theta \right) + \frac{2}{a} \operatorname{sgn}(\theta) \right]$$

where $\operatorname{sgn}(\theta) = \begin{cases} 1, & \theta \geq 0 \\ -1, & \theta \leq 0 \end{cases}$

$$\Rightarrow g'(\theta) = \left[\frac{n}{\sigma^2}\left(-2\theta + 2\bar{x} \right) + \frac{2}{a} \operatorname{sgn}(\theta) \right]$$

Setting this derivative to zero,

$$g'(\theta) = 0 = \frac{2n}{\sigma^2}(\bar{x} - \theta) + \frac{2}{a}\operatorname{sgn}(\theta)$$

$$\Rightarrow \frac{n}{\sigma^2}(\bar{x} - \theta) = -\frac{1}{a}\operatorname{sgn}(\theta)$$

Since $\operatorname{sgn}(\theta) \neq 0 \ \forall \theta$, we can multiply both sides by $\operatorname{sgn}(\theta)$.

$$\Rightarrow \frac{n}{\sigma^2}(\bar{x} - \theta)\operatorname{sgn}(\theta) = -\frac{1}{a}\underbrace{\operatorname{sgn}^2(\theta)}_{+1} = \frac{-1}{a}$$

$$\Rightarrow \frac{an}{\sigma^2}(\bar{x} - \theta)\operatorname{sgn}(\theta) = -1$$

$$\Rightarrow \frac{an}{\sigma^2}(\bar{x} - \theta)\operatorname{sgn}(\theta) + 1 = 0.$$

$$\Rightarrow (\bar{x} - \theta)\operatorname{sgn}(\theta) = \frac{-\sigma^2}{na}$$

For $\theta > 0$, $\theta_1 = \bar{x} + \frac{\sigma^2}{na}$ & for $\theta < 0$, $\theta_2 = \bar{x} - \frac{\sigma^2}{na}$.

We will need to check at which of these points $g(\theta)$ is maximized.

i.e $\hat{\theta}_{MLE} = \underset{\theta \in \{\theta_1, \theta_2\}}{\operatorname{argmax}}\ g(\theta)$

Problem 3: The following questions are from textbook.

1. Let $X_1, \dots X_n$ be i.i.d random variables with the probability density function $f(x|\theta)$, where if $\theta = 0$, then

$$f(x|\theta) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

while if $\theta = 1$, then

$$f(x|\theta) = \begin{cases} \frac{1}{2\sqrt{x}}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the MLE of θ.

Sol^n: We only know $f(x|\theta)$ when $\theta = 0$ or $\theta = 1$. So we can only say if $\theta = 0$ or $\theta = 1$ maximizes likelihood of some given Sample $X_1, \dots X_n$.

$$L(\theta|\vec{x}) = f(\vec{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Let us try to represent $f(x|\theta)$ in a single expression instead of the piecewise definition. $f(x|\theta) = \mathbb{I}_{\{x \in (0,1)\}} \left[\mathbb{I}_{\{\theta=0\}} + \frac{1}{2\sqrt{x}} \mathbb{I}_{\{\theta=1\}} \right]$

indicator fn.

When any one of x_i lies outside $(0,1)$, we get $L(\theta|\vec{x}) = 0$ for both $\theta = 0$ & $\theta = 1$. Hence both are equally likely. We could break tie and pick any one in this situation.

When all x_i are in the interval $(0,1)$, then

$$f(x|\theta) = \mathbb{I}_{\{\theta=0\}} + \frac{1}{2\sqrt{x}} \mathbb{I}_{\{\theta=1\}}.$$

Notice that $\mathbb{I}_{\{\theta=0\}} \mathbb{I}_{\{\theta=1\}} = 0$ as θ cannot be 0 & 1 at the same time.

$$\therefore L(\theta|\vec{x}) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \left[\mathbb{I}_{\{\theta=0\}} + \frac{1}{2\sqrt{x_i}} \mathbb{I}_{\{\theta=1\}} \right]$$

$$= \prod_{i=1}^{n} \mathbb{I}_{\{\theta=0\}} + \prod_{i=1}^{n} \left(\frac{1}{2\sqrt{x_i}} \mathbb{I}_{\{\theta=1\}} \right)$$

(this happens because the crossterms contain $\mathbb{I}_{\{\theta=0\}} \cdot \mathbb{I}_{\{\theta=1\}}$ and hence go to zero.

$$\therefore \quad L(\theta|x) = \prod_{i=1}^{n} \mathbb{I}_{\{\theta=0\}} + \prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}} \mathbb{I}_{\{\theta=1\}}$$

$$= \mathbb{I}^{n}_{\{\theta=0\}} + \left(\prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}}\right)\left(\prod_{i=1}^{n} \mathbb{I}_{\{\theta=1\}}\right)$$

$$= \mathbb{I}^{n}_{\{\theta=0\}} + \left(\prod_{i=1}^{n} \mathbb{I}_{\{\theta=1\}}\right)\left(\prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}}\right)$$

$$= \underbrace{\mathbb{I}^{n}_{\{\theta=0\}}}_{\mathbb{I}_{\{\theta=0\}}} + \underbrace{\mathbb{I}^{n}_{\{\theta=1\}}}_{\mathbb{I}_{\{\theta=1\}}}\left(\prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}}\right)$$

$$= \underbrace{1}_{A} \cdot \mathbb{I}_{\{\theta=0\}} + \mathbb{I}_{\{\theta=1\}} \underbrace{\left(\prod \frac{1}{2\sqrt{x_i}}\right)}_{B} = \underbrace{A}_{1} \mathbb{I}_{\{\theta=0\}} + B\,\mathbb{I}_{\{\theta=1\}}$$

Note: If $g(z) = a\,\mathbb{I}_{\{z=z_1\}} + b\,\mathbb{I}_{\{z=z_2\}}$,

then $\underset{z}{\operatorname{argmax}}\, g(z) = \begin{cases} z_1, & a>b \\ z_2, & a<b \end{cases}$

i.e. $\operatorname{argmax}\, g(z) = z_1 \mathbb{I}_{\{a>b\}} + z_2 \mathbb{I}_{\{b>a\}}$

$$= z_1 \mathbb{I}_{\{a>b\}} + \left(1 - \mathbb{I}_{\{a>b\}}\right) z_2$$

$$= z_2 + (z_1 - z_2)\mathbb{I}_{\{a>b\}}$$

Also note that in this case, we are dealing with $x_i \in (0,1)$ as for the case where $\exists\, x_i \notin (0,1)$, we mentioned that there is a tie betn $\theta=0$ & $\theta=1$ & any one of them is a good MLE estimate.

So $0 < x_i < 1 \Rightarrow 0 < \sqrt{x_i} < 1 \Rightarrow 1 < \frac{1}{\sqrt{x_i}} < \infty$

or $1 < \frac{1}{\sqrt{x_i}}$

$$\therefore \quad \underbrace{\prod_{i=1}^{n} \frac{1}{\sqrt{x_i}}}_{B} > \underbrace{1}_{A} \Rightarrow B > A = 1 > 0$$

$$\therefore \quad \hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, L(\theta|x) = \underset{\theta}{\operatorname{argmax}}\, A\,\mathbb{I}_{\{\theta=0\}} + B\,\mathbb{I}_{\{\theta=1\}} = 1$$

$$\therefore \quad \hat{\theta}_{MLE} = \begin{cases} 0 \text{ or } 1 & ; \quad \exists\, x_i \notin (0,1) \\ 1 & ; \quad \forall\, x_i \in (0,1) \end{cases}$$

2. Let $X_1, \dots X_n$ be iid random variables with distribution $N(\theta, \sigma^2)$ and suppose the parameter θ is random with a prior distribution $N(\mu, \tau^2)$. Assume that σ^2, μ, τ^2 are all known.

a) Find the joint probability density function of $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and θ.

b) Show that the marginal distribution $m(\bar{x} \mid \sigma^2, \mu, \tau^2)$ of $\bar{X}$ is $N(\mu, (\sigma^2/n) + \tau^2))$.

c) Derive the posterior distribution $\pi(\theta \mid x_1, \dots x_n)$.

Sol^n a) Since $X_i \sim N(\theta, \sigma^2)$, the sample mean $\bar{X}$ is also normally distributed with the same mean but smaller variance σ^2/n. $\therefore \bar{X} \sim N(\theta, \sigma^2/n)$

$$\therefore p(\bar{x} \mid \theta) = \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left\{-\frac{(\bar{x}-\theta)^2}{2(\sigma^2/n)}\right\}$$

Need to find the joint density function $p(\bar{x}, \theta)$.

$$\theta \sim N(\mu, \tau^2) \therefore p(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

By definition of joint density;

$$p(\bar{x}, \theta) = p(\bar{x} \mid \theta) p(\theta)$$

$$= \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left\{-\frac{(\bar{x}-\theta)^2}{2(\sigma^2/n)}\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$\Rightarrow p(\bar{x}, \theta) = \frac{1}{2\pi\sqrt{\frac{(\sigma\tau)^2}{n}}} \exp\left\{-\frac{1}{2}\left(\frac{(\bar{x}-\theta)^2}{\sigma^2/n} + \frac{(\theta-\mu)^2}{\tau^2}\right)\right\}$$

b) By definition of marginal distribution, $m(\bar{x} \mid \sigma^2, \mu, \tau^2) = \int p(\bar{x}, \theta)\, d\theta$

$$m(\bar{x} \mid \sigma^2, \mu, \tau^2) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{\frac{(\sigma\tau)^2}{n}}} \exp\left\{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2/n} - \frac{(\theta-\mu)^2}{\tau^2}\right\} d\theta$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{\frac{(\sigma\tau)^2}{n}}} \exp\left\{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2/n}\right\} \exp\left\{\frac{-1}{2}\frac{(\theta-\mu)^2}{\tau^2}\right\} d\theta$$

This expression is the convolution of pdf of two normal distributions.

We know that the convolution of two pdf is the distribution of the sum of two random variables whose pdf are involved in the convolution. i.e. if $X_1 \sim f_1$, $X_2 \sim f_2$, then $X_1 + X_2 \sim f_1 * f_2$

where $*$ is the convolution operator.

We also know that the convolution of two normal pdfs is another normal pdf. i.e. if $Z_1 \sim N(a_1, b_1^2)$

$$Z_2 \sim N(a_2, b_2^2)$$

then $Z_1 + Z_2 \sim N(a_1 + a_2, b_1^2 + b_2^2)$

Recall that $p(\bar{x} \mid \theta) = N(\theta, \frac{\sigma^2}{n})$ or $\bar{x} \sim N(\theta, \frac{\sigma^2}{n})$

and $p(\theta) = N(\mu, \tau^2)$ or $\theta \sim N(\mu, \tau^2)$

$$p(\bar{x}, \theta) = p(\bar{x} \mid \theta)\, p(\theta)$$

$$= N(\theta, \tfrac{\sigma^2}{n})\, N(\mu, \tau^2)$$

And $\int p(\bar{x}, \theta)\, d\theta = N(\theta, \frac{\sigma^2}{n}) * N(\mu, \tau^2)$

Ignore this

By inspecting the previous integral, we can tell that the integration will yeild anothere normal ditribution. Hence, if we can find the mean & variance of this new normal distribution, we have all the information.

In general, we want to show that if $X \mid Y \sim N(Y, \sigma^2)$ and $Y \sim N(\mu, \tau^2)$, then $X \sim N(\mu, \sigma^2 + \tau^2)$.

proof: We assume that we know the distribution of X is normal.

Then $\text{mean}(X) = E[X] = E[E[X \mid Y]] = E[Y] = \mu$

$\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E[X \mid Y]) = E[\sigma^2] + \text{Var}(Y) = \sigma^2 + \tau^2$

$E(X) = \mu$

Since $\text{mean}(X) = \mu$ & $\text{Var}(X) = \sigma^2 + \tau^2$ and we assumed that X is normally distributed, we have

$$X \sim N(\mu, \sigma^2 + \tau^2).$$

Now, in our case, we have $\bar{X}|\theta \sim N(\theta, \sigma^2/n)$ & $\theta \sim N(\mu, \tau^2)$

$$\because f(\bar{x}) = \underbrace{\bar{m}(\bar{x}|\sigma^2, \mu, \tau^2)}_{\int f(\bar{x},\theta)d\theta = f(\bar{x})} = N\left(\mu, \frac{\sigma^2}{n} + \tau^2\right)$$

$\int f(\bar{x},\theta)d\theta = f(\bar{x}) \leftarrow$ the marginal of $f(\bar{x},\theta)$ gives the density of $f(\bar{x})$ and we know that the density of $\bar{X}$ is $N(\mu, \frac{\sigma^2}{n} + \tau^2)$

c) Derive the posterior distribution $\pi(\theta|x_1 \dots x_n)$

Sol^n:
$$\pi(\theta|x_1 \dots x_n) \propto f(x_1 \dots x_n|\theta)\,\pi(\theta)$$

$$\propto \prod_{i=1}^{n} f(x_i|\theta)\,\pi(\theta)$$

$$\propto \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2}(x_i-\theta)^2\right\}\right) \cdot \pi(\theta)$$

$$\propto \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\} \cdot \pi(\theta)$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \theta)^2\right\} \cdot \pi(\theta)$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(\bar{x}-\theta)^2 + \underbrace{2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\theta)}_{(\bar{x}-\theta)\sum_{i=1}^{n}(x_i-\bar{x}) = 0}\right)\right\} \cdot \pi(\theta)$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left(\underbrace{\sum(x_i-\bar{x})^2}_{\text{independent of }\theta} + n(\bar{x}-\theta)^2\right)\right\} \cdot \pi(\theta)$$

$$\propto \exp\left\{\frac{-n}{2\sigma^2} \cdot (\bar{x}-\theta)^2\right\} \cdot \pi(\theta)$$

$$\propto \exp\left\{\frac{-n}{2\sigma^2} \cdot (\bar{x}-\theta)^2\right\} \cdot \underbrace{\frac{1}{\sqrt{2\pi\tau^2}}}_{\text{independent of }\theta} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{\frac{-n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\bar{x}-\theta)^2 + \frac{1}{\tau^2}(\theta-\mu)^2\right]\right.$$

$$\therefore \pi(\theta \mid x_1 \ldots x_n) \propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\bar{x}-\theta)^2 + \frac{1}{\tau^2}(\theta-\mu)^2\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\bar{x}^2+\theta^2-2\bar{x}\theta) + \frac{1}{\tau^2}(\theta^2+\mu^2-2\mu\theta)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2 + \underbrace{\left(\frac{n\bar{x}^2}{\sigma^2}+\frac{\mu^2}{\tau^2}\right)}_{\text{indep. of } \theta} - 2\left[\frac{n}{\sigma^2}\bar{x}+\frac{\mu}{\tau^2}\right]\theta\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\underbrace{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}_{A}\theta^2 - 2\underbrace{\left[\frac{n}{\sigma^2}\bar{x}+\frac{\mu}{\tau^2}\right]}_{B}\theta\right]\right\}$$

This is a normal distribution. We want to know the mean & std.

One could complete the squares but it is easier to find the mode & for normal mean = mode.

$$\text{Mode} = \underset{\theta}{\operatorname{argmax}}\ A\theta^2 - 2B\theta = A\left(\theta^2 - \frac{2B}{A}\theta\right) = A\left(\theta - \frac{2B}{A}\right)\theta$$

This is a parabola and the peak is the $\underset{\text{center}}{\underline{\text{mean}}}$ of the roots.

The roots are $\theta = 0$ & $\theta = \frac{2B}{A}$, Center $= \frac{\theta_1+\theta_2}{2} = B/A$.

$$\therefore \mu_{\text{posterior}} = B/A$$

Similarly, for a Normal distribution, the coefficient of x^2 is $\frac{-1}{2\sigma^2}$

$$-\frac{1}{2\sigma^2_{\text{posterior}}} = -\frac{1}{2}A \Rightarrow A = \frac{1}{\sigma^2_{\text{posterior}}} \Rightarrow \sigma_{\text{posterior}} = \frac{1}{A}.$$

$$\therefore \pi(\theta \mid x_1 \ldots x_n) = N(\mu_{\text{posterior}}, \sigma_{\text{posterior}})$$

where $\mu_{\text{posterior}} = \dfrac{\frac{n}{\sigma^2}\bar{x} + \frac{\mu}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$ & $\sigma_{\text{posterior}} = \dfrac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$