FYE Problem 1. ID: 6494

1. $P(\theta|Y) = \dfrac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta}$ (By Bayes Theorem)

$P(Y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$ and $P(\theta) = \dfrac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{Beta(\alpha,\beta)}$

Therefore. $P(Y|\theta)P(\theta) = \dfrac{\binom{n}{y}\theta^{\alpha+y-1}(1-\theta)^{n-y+\beta-1}}{Beta(\alpha,\beta)}$

and $\int P(Y|\theta)P(\theta)\,d\theta = \int_0^1 \underbrace{\dfrac{\binom{n}{y}\theta^{\alpha+y-1}(1-\theta)^{n-y+\beta-1}}{Beta(\alpha,\beta)}}_{\text{Kernel of } Beta(\alpha+y,\, n-y+\beta)} d\theta$

$= \dfrac{\binom{n}{y} Beta(\alpha+y, n-y+\beta)}{Beta(\alpha,\beta)}$

Thus. $P(\theta|Y) = \dfrac{\theta^{\alpha+y-1}(1-\theta)^{n-y+\beta-1}}{Beta(\alpha+y, n-y+\beta)}$

which is a Beta distribution with $Beta(\alpha+y, n-y+\beta)$

2. From 1. ~~we~~ we have the marginal distribution of Y as: $m(y) = \dfrac{\binom{n}{y} Beta(\alpha+y, n-y+\beta)}{Beta(\alpha,\beta)}$

Taking $\alpha=\beta=1$, we have:

$m(y) = \dfrac{\binom{n}{y} Beta(y+1, n-y+1)}{Beta(1,1)}$ (1)

Since $Beta(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, we have

$Beta(1,1) = \dfrac{\Gamma(1)\Gamma(1)}{\Gamma(1+1)} = 1$ and

$Beta(y+1, n-y+1) = \dfrac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} = \dfrac{y!(n-y)!}{(n+1)!}$

and finally $\binom{n}{y} = \dfrac{n!}{(n-y)!y!}$. Thus, substitute back to (1) we have:

$m(y) = \dfrac{n!}{(n+1)!} = \dfrac{1}{n+1}$ for $y = 0, 1, \cdots n$.

3. Since $m(y) = \frac{1}{n+1}$ $y = 0, 1, \cdots n$

we have $E(Y) = \sum_{i=0}^{n} \dfrac{i}{n+1} = \dfrac{1}{n+1}\cdot\dfrac{(0+n)\cdot(n+1)}{2} = \dfrac{n}{2}$

4. $E(Y^2) = \sum_{i=0}^{n}\dfrac{i^2}{n+1} = \dfrac{1}{n+1}\cdot\sum_{i=1}^{n} i^2 = \dfrac{1}{n+1}\cdot\dfrac{n(n+1)(2n+1)}{6}$

$= \dfrac{2n^2+n}{6}$

Thus, $Var(Y) = E(Y^2) - (E(Y))^2 = \dfrac{2n^2+n}{6} - \dfrac{n^2}{4}$

$= \dfrac{4n^2+2n-3n^2}{12} = \dfrac{n^2+2n}{12}$

FYE. Problem 2. ID. 6494.

1. Since the p.d.f. of X_i is

$$f(x_i) = \begin{cases} \frac{1}{\theta} & \text{if } x_i \in (0,\theta) \\ 0 & \text{otherwise.} \end{cases}$$

we have the likelihood as

$$L(\theta|x) = \prod_{i=1}^{n} \frac{1}{\theta} 1(x_i \in (0,\theta))$$

$$= \theta^{-n} 1(\min_i X_i > 0)\, 1(\max_i X_i < \theta)$$

Since as a function of θ, $L(\theta|x)$ is a power function with negative power, it is monotonically decreasing function. It achieves the maximum at the minimal value of θ, which is $\max_i x_i$. thus.

$\hat{\theta} = \max_i x_i$

2. Firstly, we show the sufficiency of $\max_i x_i$, from 1 we have: $L(\theta|x) = \theta^{-n} 1(\max_i x_i < \theta) 1(\min_i x_i > 0)$

$$= g(\max_i x_i | \theta)\, h(x)$$

Thus, by factorization theorem we have $\max_i x_i$ is sufficient.

Now, we show it is also minimal sufficient. Suppose another data y with $\max_i y_i$. The inference of θ is only based on $\theta^{-n} 1(\max_i x_i < \theta)$ so the ratio $\frac{\theta^{-n} 1(\max_i x_i < \theta)}{\theta^{-n} 1(\max_i y_i < \theta)}$ is a constant w.r.t. θ if and only if $\max_i x_i = \max_i y_i$

Thus, $\max_i x_i$ is minimal sufficient.

3. We want the distribution of $X_{(n)} = \max_i X_i = \hat{\theta}$ so ~~[illegible]~~ denote the CDF of $\hat{\theta}$ as G and pdf as g, the cdf of X_i as F and pdf as f, we have.

$$G(\hat{\theta}) = \Pr(X_1 \le \hat{\theta}, \cdots X_n \le \hat{\theta})$$

$$= \Pr(X_1 \le \hat{\theta}) \cdots \Pr(X_n \le \hat{\theta}) = (F(\hat{\theta}))^n$$

Thus, $g(\hat{\theta}) = n[F(\hat{\theta})]^{n-1} f(\hat{\theta})$.

Since $F(x) = \frac{x}{\theta}$ and $f(\theta) = \frac{1}{\theta}$, we have the distribution of $\hat{\theta}$ as

$$g(\hat{\theta}) = \begin{cases} n\frac{\hat{\theta}^{n-1}}{\theta^n} & \text{if } \hat{\theta} > \max_i x_i \\ 0 & \text{otherwise.} \end{cases}$$

written in a more appropriate way.

$$g_{\hat{\theta}}(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{if } t > \max_i x_i \\ 0 & \text{otherwise} \end{cases}$$

4. Since the distribution of X is Uniform$(0,\theta)$, it is a scale family distribution. Thus, we define $Q(x_1, \cdots x_n, \theta) = \frac{\max_i x_i}{\theta}$ and we shall prove Q is a pivot. ~~[illegible]~~ Thus by definition we need to show that the distribution of Q does not depend on θ. Since $\theta > \max_i x_i$. $Q \in (0,1)$. Thus, the CDF of Q is

$$F_Q(q) = P\left(\frac{\max_i x_i}{\theta} \le q\right) = \Pr(x_1 \le q\theta) \cdots \Pr(x_n \le q\theta)$$

$$= \prod_{i=1}^{n} \frac{q\theta}{\theta} = q^n$$

Hence $f_Q(q) = nq^{n-1}$ doesn't depend on θ.

Thus, $\frac{\max_i x_i}{\theta}$ is actually a pivot.

From $\Pr\left(a \le \frac{\max_i x_i}{\theta} \le b\right) = \Pr\left(\frac{\max_i x_i}{\theta} \le b\right) - \Pr\left(\frac{\max_i x_i}{\theta} \le a\right)$

$= F_Q(b) - F_Q(a)$. From previous analysis

we have $\Pr\left(a \le \frac{\max_i x_i}{\theta} \le b\right) = b^n - a^n$.

Thus, for any a, b satisfies $b^n - a^n = 0.95$, $(a\theta, b\theta)$ is a 95% confidence interval of $\hat{\theta} = \max x_i$

5. From $b^n - a^n = 0.95$ we have

$$b = (0.95 + a^n)^{\frac{1}{n}}$$

and the length of the interval is therefore

$$L(a) = b\theta - a\theta = \left[(0.95+a^n)^{\frac{1}{n}} - a\right]\theta.$$

define $l(a) = (0.95+a^n)^{\frac{1}{n}} - a$. Since $\theta > 0$, minimize $L(a)$ is equivalent to minimize $l(a)$.

$$l'(a) = \frac{1}{n}(0.95+a^n)^{\frac{1}{n}-1} \cdot n a^{n-1} - 1$$

set to 0 we have $a =$

Since $l'(a) = \left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{1-n} - 1$

$$l''(a) = (1-n)\left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{-n} \cdot \frac{\left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{1-n} a - \sqrt[n]{0.95+a^n}}{a^2}$$

$$= \frac{(1-n)}{a}\left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{-n} \cdot \left[\left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{1-n} - \frac{\sqrt[n]{0.95+a^n}}{a}\right]$$

whose sign only depends on the term

$$\left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{1-n} - \frac{\sqrt[n]{0.95+a^n}}{a} \quad (*)$$

Since $0.95 > 0$ so $\sqrt[n]{0.95+a^n} > a$

and therefore $\frac{\sqrt[n]{0.95+a^n}}{a} > 1$.

also for $f(x) = x^{1-n} - x$, $n \geq 1$, $f(x) < 0$ if $x > 1$

Thus, we have $(*) < 0$ and $l''(a) < 0$.

Thus, the value a that provides the smallest C.I. length is the a that satisfies

since $l'(a) = \frac{1}{n}(0.95+a^n)^{\frac{1}{n}-1} \cdot n a^{n-1} - 1$

$$= \left(\frac{\sqrt[n]{0.95+a^n}}{a}\right)^{1-n} - 1$$

since $0.95 > 0$, we have $\sqrt[n]{0.95+a^n} > a$ thus $l'(a) < 0$. so $l(a)$ is a decreasing function of a, which achieve the minimal at the maximum value that a can take.

Since $b = (0.95+a^n)^{\frac{1}{n}} \in [0,1]$.

b as a function of a is increasing, which achieves maximum value 1 at the maximum of a.

denote $\max_a a$ as $\hat{a}$, we have

$$(0.95+\hat{a}^n)^{\frac{1}{n}} = 1 \Rightarrow 0.95 + \hat{a}^n = 1$$

so $\hat{a} = \sqrt[n]{0.05} = (0.05)^{\frac{1}{n}}$

Thus, the value a that achieves the smallest confidence interval length is

$$\hat{a} = (0.05)^{\frac{1}{n}}.$$

FYE Problem 3. ID: 6494.

1. The full posterior distribution for model parameters is:

$$f(\beta,\lambda,\tau \mid data) \propto f(y \mid \beta,\lambda,\tau)\overset{f(\beta\mid\lambda,\tau)}{f(\lambda)}f(\tau)$$

$$\propto \prod_{i=1}^{n} N(y_i \mid x_i^T\beta, 1)\prod_{j=1}^{p} N(\beta_j \mid 0, \lambda_j^2\tau^2)\prod_{j=1}^{p} C^+(\lambda \mid 0,1)\, C^+(\tau \mid 0,1)$$

$$\propto \prod_{i=1}^{n} \exp\left(-\frac{(y_i - x_i^T\beta)^2}{2}\right)\prod_{j=1}^{p}(\lambda_j^2\tau^2)^{-\frac{1}{2}}\exp\left(-\frac{\beta_j^2}{2\lambda_j^2\tau^2}\right)$$

$$\times \prod_{j=1}^{p}\frac{1}{1+\lambda_j^2} \times \frac{1}{1+\tau^2}$$

2. Introduce latent variables η_j, $j=1,\dots p$ and w.

such that: $\lambda_j^2 \mid \eta_j \sim \text{Inv-Gamma}(\frac{1}{2}, \frac{1}{\eta_j})$ for $j=1,\dots,p$

and $\eta_j \sim \text{Inv-Gamma}(\frac{1}{2}, 1)$.

and $\tau^2 \mid w \sim \text{Inv-Gamma}(\frac{1}{2}, \frac{1}{w})$, $w \sim \text{Inv-Gamma}(\frac{1}{2}, 1)$

Then we have the model as:

$$f(\beta,\lambda,\tau,\eta,w \mid data)$$

$$\propto f(y \mid \beta,\lambda,\tau,\eta,w)\, f(\beta \mid \lambda,\tau)\, f(\lambda \mid \eta)\, f(\tau \mid w)\, f(\eta)\, f(w)$$

$$\propto \prod_{i=1}^{n}\exp\left(-\frac{(y_i - x_i^T\beta)^2}{2}\right)\prod_{j=1}^{p}(\lambda_j^2\tau^2)^{-\frac{1}{2}}\exp\left(-\frac{\beta_j^2}{2\lambda_j^2\tau^2}\right)$$

$$\prod_{j=1}^{p}\left(\frac{1}{\eta_j}\right)^{\frac{1}{2}}(\lambda_j^2)^{-\frac{1}{2}-1}\exp\left(-\frac{\frac{1}{\eta_j}}{\lambda_j^2}\right)\cdot(\eta_j)^{-\frac{1}{2}-1}\exp\left(-\frac{1}{\eta_j}\right)$$

$$\times\left(\frac{1}{w}\right)^{\frac{1}{2}}(\tau^2)^{-\frac{1}{2}-1}\exp\left(-\frac{\frac{1}{w}}{\tau^2}\right)\cdot w^{-\frac{1}{2}-1}\exp\left(-\frac{1}{w}\right)$$

Therefore, denote the full conditionals for parameter α as $\alpha \mid \ell$ to simplify the notation, we have:

① $$\beta_j \mid \ell \propto \exp\left(-\frac{\sum_{i=1}^{n}(x_{ij}\beta_j + x_{i,(-j)}^T\beta_{-j} - y_i)^2}{2}\right)\cdot\exp\left(-\frac{\beta_j^2}{2\lambda_j^2\tau^2}\right)$$

$$\propto \exp\left(-\frac{\lambda_j^2\tau^2\left(\sum_{i=1}^{n}x_{ij}^2\beta_j^2 + 2\sum_{i=1}^{n}x_{ij}(x_{i,(-j)}^T\beta_{-j} - y_i)\beta_j\right) + \beta_j^2}{2\lambda_j^2\tau^2}\right)$$

$$\propto \exp\left(-\frac{\left(\beta_j - \frac{\sum_{i=1}^{n}x_{ij}(x_{i(-j)}^T\beta_{-j} - y_i)}{\lambda_j^2\tau^2\sum_{i=1}^{n}x_{ij}^2+1}\right)^2}{2\lambda_j^2\tau^2/(\lambda_j^2\tau^2\sum_{i=1}^{n}x_{ij}^2+1)}\right)$$

Thus, the full conditionals for β_j is:

$$N\left(\frac{\sum_{i=1}^{n}x_{ij}(x_{i,(-j)}^T\beta_{-j} - y_i)}{\lambda_j^2\tau^2\sum_{i=1}^{n}x_{ij}^2+1},\ \frac{\lambda_j^2\tau^2}{\lambda_j^2\tau^2\sum_{i=1}^{n}x_{ij}^2+1}\right) \quad j=1\dots p$$

where $x_{i,(-j)}$ is the x_i missing jth element and $\beta_{(-j)}$ is β vector missing jth element.

② $$\lambda_j^2 \mid \ell \propto (\lambda_j^2)^{-\frac{1}{2}}\exp\left(-\frac{\beta_j^2}{2\lambda_j^2\tau^2}\right)\cdot(\lambda_j^2)^{-\frac{1}{2}-1}\exp\left(-\left(\frac{1}{\eta_j}\right)\frac{1}{\lambda_j^2}\right)$$

$$\propto (\lambda_j^2)^{-1-1}\exp\left(-\left(\frac{\beta_j^2}{2\tau^2}+\frac{1}{\eta_j}\right)\frac{1}{\lambda_j^2}\right)$$

Thus, $\lambda_j^2 \mid \ell \sim \text{Inverse-Gamma}\left(1, \frac{\beta_j^2}{2\tau^2}+\frac{1}{\eta_j}\right)$ $j=1,\dots p$.

③ $$\eta_j \mid \ell \propto \eta_j^{-\frac{1}{2}}\exp\left(-\left(\frac{1}{\lambda_j}\right)\frac{1}{\eta_j}\right)\cdot\eta_j^{-\frac{1}{2}-1}\exp\left(-\frac{1}{\eta_j}\right)$$

$$\propto \eta_j^{-1-1}\exp\left(-\left(\frac{1}{\lambda_j}+1\right)\frac{1}{\eta_j}\right)$$

Thus, $\eta_j \mid \ell \sim \text{Inverse-Gamma}\left(1, 1+\frac{1}{\lambda_j}\right)$

④ $$\tau^2 \mid \ell \propto (\tau^2)^{-\frac{p}{2}}\exp\left(-\left(\sum_{j=1}^{p}\frac{\beta_j^2}{2\lambda_j^2}\right)\cdot\frac{1}{\tau^2}\right)\cdot(\tau^2)^{-\frac{1}{2}-1}\exp\left(-\frac{\frac{1}{w}}{\tau^2}\right)$$

$$\propto (\tau^2)^{-\frac{p+1}{2}-1}\exp\left(-\left(\sum_{j=1}^{p}\frac{\beta_j^2}{2\lambda_j^2}+\frac{1}{w}\right)\frac{1}{\tau^2}\right)$$

Thus, $\tau^2 \mid \ell \sim \text{Inverse-Gamma}\left(\frac{p+1}{2}, \sum_{j=1}^{p}\frac{\beta_j^2}{2\lambda_j^2}+\frac{1}{w}\right)$

Finally, ⑤ $$w \mid \ell \propto w^{-\frac{1}{2}}\exp\left(-\frac{1}{\tau^2}\frac{1}{w}\right)w^{-\frac{1}{2}-1}\exp\left(-\frac{1}{w}\right)$$

$$\propto w^{-1-1}\exp\left(-\left(\frac{1}{\tau^2}+1\right)\frac{1}{w}\right)$$

So $w \mid \ell \sim \text{Inverse-Gamma}\left(1, \frac{1}{\tau^2}+1\right)$

Thus, we have all full conditionals for Gibbs.

3. In the Gibbs sampler, giving initial values of β_j^0, $j=1,\dots p$. λ_j^0, $j=1,\dots p$. τ^0, η_j^0, $j=1,\dots p$, and w^0, in iteration $k=1,\dots S$, we draw parameters (k-th) $(\beta,\lambda,\tau,\eta,w)$ from their full conditionals as follows:

① draw $\beta_j^{(k)}$ from $N\left(\frac{\sum_{i=1}^{n}x_{ij}(x_{i,(-j)}^T\beta_{-j}^{(k-1)} - y_i)}{(\lambda_j^{(k-1)})^2(\tau^{(k-1)})^2\sum_{i=1}^{n}x_{ij}^2+1}, \dots\right)$

② draw $(\lambda_j^{(k)})^2$ from Inverse-Gamma $(1,$

① draw $\beta_j^{(k)}$ from $j=1,\dots p$.

$$N\left(\frac{\sum_{i=1}^{n} x_{ij}\,(X_{i,(-j)}^T \beta_{-j}^{(k-1)} - y_i)}{(\lambda_j^{(k-1)})^2(\tau^{(k-1)})^2 \sum_{i=1}^{n} x_{ij}+1},\ \frac{(\lambda_j^{(k-1)})^2(\tau^{(k-1)})^2}{(\lambda_j^{(k-1)})^2(\tau^{(k-1)})^2\sum_{i=1}^{n} x_{ij}+1}\right)$$

② draw $\lambda_j^{(k)}$ from $j=1,\dots p$

$$\text{Inverse-Gamma}\left(1,\ \frac{(\beta_j^{(k)})^2}{2(\tau^{(k-1)})^2} + \frac{1}{\eta_j^{(k-1)}}\right)$$

③ draw $\eta_j^{(k)}$ from $j=1,\dots p$.

$$\text{Inverse-Gamma}\left(1,\ 1+\frac{1}{\lambda_j^{(k)}}\right)$$

④ draw $(\tau^{(k)})^2$ from Inverse-Gamma$\left(\frac{p+1}{2},\ \sum_{j=1}^{p}\frac{(\beta_j^{(k)})^2}{2(\lambda_j^{(k)})^2}+\frac{1}{w^{(k-1)}}\right)$

⑤ draw $w^{(k)}$ from Invers-Gamma$\left(1,\ 1+\frac{1}{(\tau^{(k)})^2}\right)$.

We can set initial values of β_j $j=1,\dots p$ to be the $\hat{\beta}$ in the linear model $y = X^T\beta+\varepsilon$ with $\varepsilon\sim N(0,I)$.
$=(X^TX)^{-1}X^Ty$.

The initial value for other parameters can be any one in their support. It doesn't matter too much because we will have a burn-in period and the Gibbs sampler is proved to be converged to the true posterior distribution if we run a long enough chain. Thus, we can just set other parameters to be 1.

4. The Bayes factor is defined as

$$D^{(\pi)} = \underset{d}{\arg\min} \int_{\Theta} L(\theta,d)\,\pi(\theta|X)\,d\theta.$$

Thus, we first need to calculate the posterior of θ, which is β in our case. When fixing other parameters, we have.

$f(\beta|data) \propto f(y|\beta,X)f(\beta) = \exp\left(-\frac{(y-\beta X)^T(y-\beta X)}{2}\right)$
$\times \exp\left(-\frac{\beta^T\Sigma\beta}{2}\right)$ where Σ is diagonal matrix $\Sigma = (\lambda_1^2\tau^2, \dots \lambda_p^2\tau^2)$. Thus, by complete the square we have.

$$f(\beta|data) \propto \exp\left(\frac{(\beta-(X^TX+\Sigma)^{-1}X^Ty)(X^TX+\Sigma)^{-1}(\beta-(X^TX+\Sigma)^{-1}X^Ty)}{2}\right)$$

Thus. $\beta\sim N\left((X^TX+\Sigma)^{-1}X^Ty,\ (X^TX+\Sigma)^{-1}\right)$

For quadratic loss function, we can prove the best decision is posterior mean. Because we can do $E[(\beta-D+D-d)^T(\beta-D+D-d)]$

$$= E[(\beta-d)^T(\beta-d)]$$
$$= E[(\beta-D)^T(\beta-D)] + E[(D-d)^T(D-d)]$$
$$\geq E[(\beta-D)^T(\beta-D)]$$

if D is posterior mean.

Thus, the Bayes estimator is

$$\boxed{D^{(\pi)} = (X^TX+\Sigma)^{-1}X^Ty}.$$

FYE Problem 4 ID: 6494

1. The original model has likelihood:

$f(y_i \mid \beta, X) = \Phi(x_i^T\beta)^{y_i}(1-\Phi(x_i^T\beta))^{1-y_i}$

If we introduce latent variable $z_1 \cdots z_n$, $z_i \sim N(x_i^T\beta, 1)$

Such that $y_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{if } z_i \le 0 \end{cases}$ then we have:

$f(y_i=1, z_i \mid \beta, X) = f(y_i=1 \mid z_i) f(z_i \mid \beta, X) = 1_{(z_i>0)} N(z_i \mid x_i^T\beta, 1)$

Integrate out z_i we have: $f(y_i=1 \mid \beta, X) =$

$\int f(y_i=1, z_i \mid \beta, X)\, dz_i = \int_{z_i>0} N(z_i \mid x_i^T\beta, 1)\, dz_i$

$= \Phi(x_i^T\beta)$ if Φ is standard normal c.d.f.

Similarly, we have

$f(y_i=0 \mid \beta, X) = \int f(y_i=0, z_i \mid \beta, X) dz_i = \int_{z_i \le 0} N(z_i \mid x_i^T\beta, 1) dz$

$= \Phi(-x_i^T\beta) = 1 - \Phi(x_i^T\beta)$

Since under two models, we always have

$f^{(1)}(y_i=1 \mid \beta, X) = f^{(2)}(y_i=1 \mid \beta, X)$ and $f^{(1)}(y_i=0 \mid \beta, X) = f^{(2)}(y_i=0 \mid \beta, X)$

for $i=1, \cdots n$ and the data are i.i.d.

we have the two models are equivalent

~~2. We have the posterior distribution as~~

~~$f(\beta, z \mid data) \propto f(y \mid z) f(z \mid X, \beta) f(\beta)$~~

~~$\propto \prod_{i=1} 1(z_i > 0) \exp(-\frac{(z_i - x_i^T\beta)^2}{2}) \cdot 1$~~

~~Therefore, for each z_i, we have the full conditionals as~~

~~$f(z_i \mid \rho) \propto 1_{(z_i>0)} \exp(-\frac{(z_i - x_i^T\beta)^2}{2})$~~

~~which is a truncated normal distribution.~~

~~Thus, the full conditionals can be written as:~~

~~$f(z_i \mid \rho) \propto \frac{1}{\sqrt{2\pi}\Phi(x_i^T\beta)} \exp(-\frac{(z_i - x_i^T\beta)^2}{2})$~~

~~which is a truncated normal distribution with mean $x_i^T\beta$, variance 1 and truncated to be $z_i > 0$.~~

~~The full conditionals for β is then,~~

~~$f(\beta, z \mid data)$~~

~~$f(\beta \mid \rho) \propto$~~

2. We have the posterior distribution as

$f(\beta, Z \mid data) \propto f(y \mid z) f(z \mid X, \beta) f(\beta)$

$\propto \prod_{i=1}^{n} \left(\exp(-\frac{(z_i - x_i^T\beta)^2}{2}) 1_{(z_i>0)}\right)^{y_i} \left(\exp(-\frac{(z_i - x_i^T\beta)^2}{2}) 1_{(z_i\le 0)}\right)^{1-y_i}$

Thus, we have the full conditionals for z_i

as $f(z_i \mid \beta, X, y_i=1) \propto \exp(-\frac{(z_i - x_i^T\beta)^2}{2}) 1_{(z_i>0)}$

which is a truncated normal distribution with mean $x_i^T\beta$ and variance 1, truncated at $z_i > 0$.

Similarly, $f(z_i \mid \beta X, y_i=0) \propto \exp(-\frac{(z_i - x_i^T\beta)^2}{2}) 1_{(z_i \le 0)}$

which is a truncated normal distribution with mean $x_i^T\beta$ and variance 1, truncated at $z_i \le 0$

As for β, we have

$f(\beta \mid \rho) \propto \exp(-\frac{\sum_i^n (z_i - x_i^T\beta)^2}{2})$

Thus, $\beta \mid \rho \sim N((X^TX)^{-1}X^T Z, (X^TX)^{-1})$.

3. The model can be written as:

$y_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{if } z_i \le 0. \end{cases}$

$z_i \overset{ind}{\sim} N(x_i^T\beta, \frac{1}{\lambda})$

$\lambda \sim \text{Gamma}(\frac{\alpha}{2}, \frac{\alpha}{2})$

By the fact from the problem, this representation is true.

Thus, the posterior distribution is

$f(Z, \beta, \lambda \mid data) \propto \prod_{i=1}^{n} \left(\exp(-\frac{(z_i - x_i^T\beta)^2}{2\frac{1}{\lambda}}) \cdot 1_{(z_i>0)} \cdot \left(\frac{1}{\lambda}\right)^{-\frac{1}{2}}\right)^{y_i}$

$\left(\exp(-\frac{(z_i - x_i^T\beta)^2}{2\frac{1}{\lambda}}) \cdot 1_{(z_i\le 0)} \cdot \left(\frac{1}{\lambda}\right)^{-\frac{1}{2}}\right)^{1-y_i}$

Therefore, we have the full conditionals as follows.

$$f(z_i|\ell) \propto \exp\left(-\frac{(z_i-x_i^T\beta)^2}{2\frac{1}{\lambda}}\right) \cdot 1_{(z_i>0)} \text{ if } y_i=1$$

and $f(z_i|\ell) \propto \exp\left(\frac{-(z_i-x_i^T\beta)^2}{2\frac{1}{\lambda}}\right) 1_{(z_i \le 0)}$

if $y_i=0$. Thus, the full conditionals for z_i is

$$f(z_i|\ell) = \begin{cases} \text{truncated normal with mean } x_i^T\beta \text{, variance } \frac{1}{\lambda} \text{ and truncated at } (0,\infty), \text{ if } y_i=1 \\ \text{truncated normal with mean } x_i^T\beta \text{, variance } \frac{1}{\lambda} \text{ and truncated at } (-\infty,0], \text{ if } y_i=0. \end{cases}$$

Now, for β, we have:

$f(\beta|\ell) \propto \prod_{i=1}^{n} \exp\left(-\frac{(z_i-x_i^T\beta)^2}{2\frac{1}{\lambda}}\right)$ Thus, β has a normal distribution with mean $\frac{1}{\lambda}(X^TX)^{-1}X^TZ$ and variance $\frac{1}{\lambda}(X^TX)^{-1}$

Thus, $\beta|\ell \sim N\left(\frac{(X^TX)^{-1}X^TZ}{\lambda}, \frac{(X^TX)^{-1}}{\lambda}\right)$

Finally, for λ we have:

$$f(\lambda|\ell) \propto \prod_{i=1}^{n} \exp\left(-\frac{(z_i-x_i^T\beta)^2}{2\frac{1}{\lambda}}\right) \cdot \left(\frac{1}{\lambda}\right)^{-\frac{1}{2}} \cdot \lambda^{\frac{d}{2}-1} e^{-\frac{d}{2}\lambda}$$

$$\propto \lambda^{\frac{d+1}{2}-1} e^{-\left(\frac{d}{2}+\frac{\sum_{i=1}^{n}(z_i-x_i^T\beta)^2}{2}\right)\lambda}.$$

So $\lambda|\ell \sim Gamma\left(\frac{d+1}{2}, \frac{d}{2}+\frac{\sum_{i=1}^{n}(z_i-x_i^T\beta)^2}{2}\right)$

FYE Problem 5. ID. 6494.

1. Since $\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0,\Gamma) \end{cases}$ we have $Y \sim N(X\beta, \Gamma)$.

Thus, $\Gamma^{-\frac{1}{2}}Y$ also has a multivariate normal distribution

we let with ~~showing that~~ finding mean and Variance of $\Gamma^{-\frac{1}{2}}Y$.

$E(\Gamma^{-\frac{1}{2}}Y) = \Gamma^{-\frac{1}{2}}E(Y) = \Gamma^{-\frac{1}{2}}X\beta$.

$Var(\Gamma^{-\frac{1}{2}}Y) = \Gamma^{-\frac{1}{2}} Var(Y)(\Gamma^{-\frac{1}{2}})^t = (\Gamma^{-\frac{1}{2}})\Gamma(\Gamma^{-\frac{1}{2}})^t = I$.

Thus. $\Gamma^{-\frac{1}{2}}Y \sim N(\Gamma^{-\frac{1}{2}}X\beta, I)$

2. From the OLS model, that is, if

$Y^* = X^*\beta + \varepsilon^*$, $\varepsilon^* \sim N(0, I)$

then $\beta^{OLS} = ((X^*)^T X^*)^{-1}(X^*)^T Y^*$.

Define $Y^* = \Gamma^{-\frac{1}{2}}Y$, $X^* = \Gamma^{-\frac{1}{2}}X$ and $\varepsilon^* = \Gamma^{-\frac{1}{2}}\varepsilon$.

~~then~~ we can change the original weighted least problem to the OLS model as

$Y^* = X^*\beta + \varepsilon^*$, where $Var(\varepsilon^*) = (\Gamma^{-\frac{1}{2}})\Gamma(\Gamma^{-\frac{1}{2}})^t = I$

satisfies the OLS setting.

Thus, $\hat\beta = ((X^*)^T X^*)^{-1}(X^*)^T Y^*$ Plug in the form of X^*, Y^* we have

$$\hat\beta^{(WLS)} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$$

3. ~~From the OLS~~ ~~Since~~

We have:

$E(\hat\beta) = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}E(Y) = (X^T\Sigma X)^{-1}X^T\Sigma^{-1}X\beta$

$= \beta$

and $Var(\hat\beta) = Var((X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y)$

$= (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Var(Y)\Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}$

$= (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\Sigma\Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}$

$= (X^T\Sigma^{-1}X)^{-1}$

4. Since Y has a normal distribution, we have $\hat\beta \sim MVN(\beta, (X^T\Sigma^{-1}X)^{-1})$

and therefore, the jth component of $\hat\beta$, denote as $\hat\beta_j$, follows a normal distribution,

$\hat\beta_j \sim N(\beta_j, \sigma_j^2)$ where σ_j^2 is the jth diagonal component of matrix $(X^T\Sigma^{-1}X)^{-1}$

Thus, the test statistic is $Z = \dfrac{\hat\beta_j - 0}{\sqrt{\sigma_j^2}} = \dfrac{\hat\beta_j}{\sigma_j}$

with σ_j^2 defined above.

It is compared with $Z_{\frac{\alpha}{2}}$ where Z is standard normal distributed r.v. and α is confindence level.

we can also calculate the p-value of $\frac{\hat\beta_j}{\sigma_j}$ with respect to standard normal distribution and make decisions accordingly.