1.

Known that $X|\{\mu,\sigma^2\}$ ~Normal(μ,σ^2), μ ~Normal$(10,1)$ and $\ln\sigma$ ~Uniform$(-100,100)$, set $t = 2\ln\sigma \Rightarrow \sigma^2 = e^t$, then we can work out the prior for σ^2

$$f_t(t) = f_{\ln\sigma}(\frac{1}{2}t)\cdot\frac{d\frac{1}{2}t}{dt} = \frac{1}{2}\cdot\begin{cases}\dfrac{1}{200} & ,\frac{1}{2}t\in[-100,100]\\ 0 & ,\text{otherwise}\end{cases} = \begin{cases}\dfrac{1}{400} & ,t\in[-200,200]\\ 0 & ,\text{otherwise}\end{cases}$$

$$\begin{aligned}f_{\sigma^2}(\sigma^2) &= f_t(\ln\sigma^2)\cdot\frac{d\ln\sigma^2}{d\sigma^2} = \frac{1}{\sigma^2}\cdot\begin{cases}\dfrac{1}{400} & ,\ln\sigma^2\in[-200,200]\\ 0 & ,\text{otherwise}\end{cases}\\ &= \begin{cases}\dfrac{1}{400\sigma^2} & ,\sigma^2\in[\exp(-200),\exp(200)]\\ 0 & ,\text{otherwise}\end{cases}\end{aligned}$$

and therefore the joint distribution is given by

$$\begin{aligned}f_{\{\mu,\sigma^2\}|X=0.5} =& \frac{f_{X=0.5|\{\mu,\sigma^2\}}\cdot f_\mu(\mu)\cdot f_{\sigma^2}(\sigma^2)}{f_{X=0.5}}\\ \propto&\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(0.5-\mu)^2}{2\sigma^2}\Big)\cdot\exp\Big(-\frac{(\mu-10)^2}{2}\Big)\cdot\frac{1}{\sigma^2}\\ \propto&(\sigma^2)^{(-3/2)}\exp\Big(-\frac{(0.5-\mu)^2}{2\sigma^2}-\frac{(\mu-10)^2}{2}\Big)\end{aligned}$$

(a)

To use Metropolis-Hasting algorithm, we first apply random variable transformation by setting $t = \ln\sigma^2 \in [-200,200]$, then

$$\begin{aligned}f_{\{\mu,t\}|X=0.5} &\propto \exp(-\frac{3}{2}t)\exp\Big(-\frac{(0.5-\mu)^2}{2\exp(t)}-\frac{(\mu-10)^2}{2}\Big)\cdot\exp(t)\\ &\propto\exp(-\frac{1}{2}t)\exp\Big(-\frac{(0.5-\mu)^2}{2\exp(t)}-\frac{(\mu-10)^2}{2}\Big)\end{aligned}$$

and use the proposal

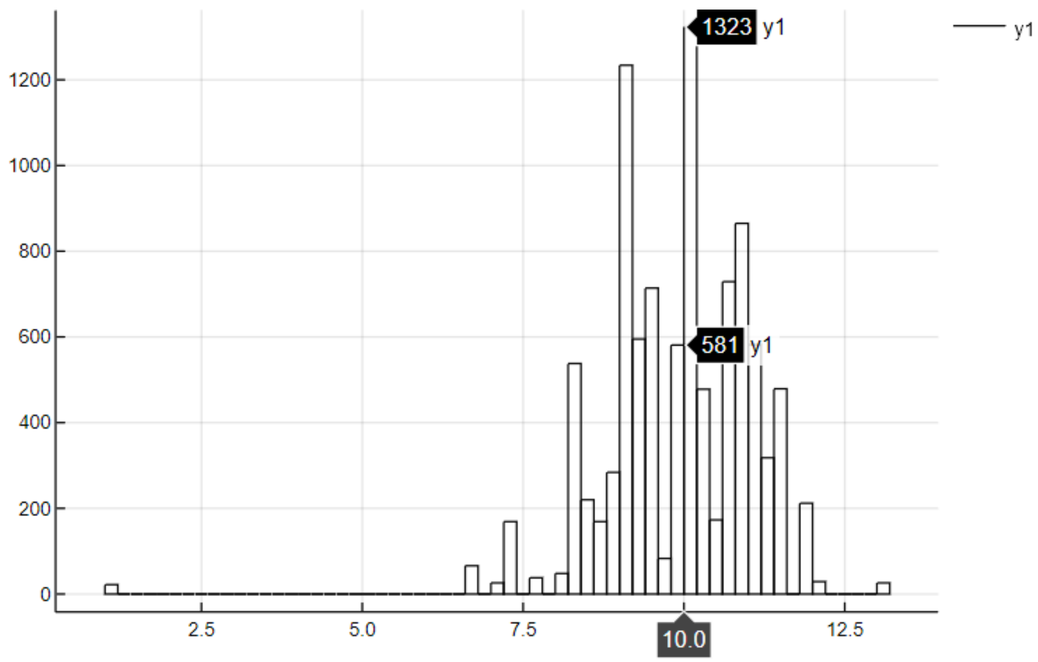
$$g_{\{\mu,t\}|\{\mu^*,t^*\}} = \frac{1}{\sqrt{2\pi\cdot 4}}\exp\Big(-\frac{(\mu-\mu^*)^2}{2\cdot 4}\Big)\cdot\frac{1}{400}$$

so we can sample from

$$g_{\mu|\mu^*,t^*} = \frac{1}{\sqrt{2\pi\cdot 4}}\exp\Big(-\frac{(\mu-\mu^*)^2}{2\cdot 4}\Big)$$

$$g_{t|\mu^*,t^*} = \begin{cases}\dfrac{1}{400} & ,t\in[-200+t^*,200+t^*]\\ 0 & ,\text{otherwise}\end{cases}$$

```julia
using Distributions
using Random
using StatsBase
gui()
n = 10000;
mu = Array{Float64, 1}(undef, n + 1);
sigma2 = Array{Float64, 1}(undef, n + 1);
mu[1] = 1;
sigma2[1] = 1;
for t in range(2, n + 1)
    sigma2z = rand() * 400 - 200 + sigma2[t-1];
    muz = mu[t-1] + rand(Normal(0, 4));
    alpha = 1;
    a = exp(-0.5*sigma2z) * exp(-(0.5-muz)^2/2/exp(sigma2z)-(muz-10)^2/2);
    c = exp(-0.5*sigma2[t-1]) * exp(-(0.5-mu[t-1])^2/2/exp(sigma2[t-1])-(mu[t-1]-10)^2/2);
    alpha = min(1, a/c);
    v = rand();
    if v <= alpha
        mu[t] = muz;
        sigma2[t] = sigma2z;
    else
        mu[t] = mu[t-1];
        sigma2[t] = sigma2[t-1];
    end
end
histogram(mu, fillcolor=[:black], fillalpha=0);
println("Mean of mu found by Metropolis-Hasting Method is: E[mu|sigma2,X=0.5]=",  mean(mu));
```




```
Mean of mu found by Metropolis-Hasting Method is: E[mu|sigma2,X=0.5]=9.957921789699299
```

So we can see that the mode of μ is in $[10, 10.2]$, and the mean is 9.96.

(b)

To use Gibbs sampling, the conditional distribution is given by and the conditional distributions are

$$\begin{aligned}
f_{\mu|\{\sigma^2,X=0.5\}} =& \frac{f_{\{\mu,\sigma^2\}|X=0.5}}{f_{\sigma^2|X=0.5}} \\
\propto& f_{\{\mu,\sigma^2\}|X=0.5} \\
\propto& \exp\Big(- \frac{(0.5-\mu)^2}{2\sigma^2} - \frac{(\mu-10)^2}{2}\Big) \\
\propto& \exp\Big(- \Big(\frac{1}{2\sigma^2} + \frac{1}{2}\Big)\mu^2 + \Big(\frac{1}{2\sigma^2} + 10\Big)\mu\Big) \\
\propto& \exp\Big(- \frac{1}{2}\Big(\frac{\sigma^2}{\sigma^2+1}\Big)^{-1}\Big(\mu - \frac{1+20\sigma^2}{2+2\sigma^2}\Big)^2\Big) \\
\sim& \text{Normal}(\frac{1+20\sigma^2}{2+2\sigma^2}, \frac{\sigma^2}{\sigma^2+1})
\end{aligned}$$

$$\begin{aligned}
f_{\sigma^2|\{\mu,X=0.5\}} =& \frac{f_{\{\mu,\sigma^2\}|X=0.5}}{f_{\mu|X=0.5}} \\
\propto& f_{\{\mu,\sigma^2\}|X=0.5} \\
\propto& \begin{cases} (\sigma^2)^{-3/2}\exp\Big(- \frac{(0.5-\mu)^2}{2\sigma^2}\Big) & ,\sigma^2 \in [\exp(-200), \exp(200)] \\ 0 & ,\text{otherwise} \end{cases}
\end{aligned}$$

so for μ we can sample from a Normal distribution and for σ^2 we try to use rejection sampling with uniform proposal. However, the interval $[\exp(-200), \exp(200)]$ is too large which is not suitable for sampling, we use random variable transformation by setting $t = \ln\sigma^2 \in [-200, 200]$, then

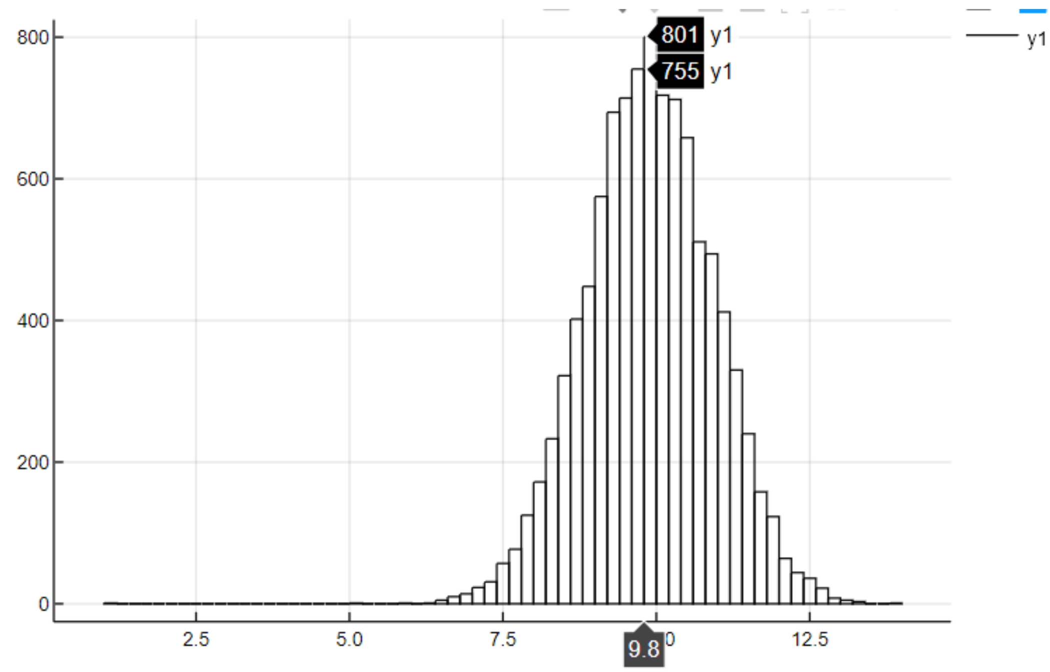
$$\begin{aligned}
f_{t|\{\mu,X=0.5\}} \propto& \begin{cases} \exp(-\frac{3}{2}t)\exp\Big(- \frac{(0.5-\mu)^2}{2\exp(t)}\Big)\cdot\exp(t) & ,t \in [-200, 200] \\ 0 & ,\text{otherwise} \end{cases} \\
\propto& \begin{cases} \exp(-\frac{1}{2}t)\exp\Big(- \frac{(0.5-\mu)^2}{2\exp(t)}\Big) & ,t \in [-200, 200] \\ 0 & ,\text{otherwise} \end{cases}
\end{aligned}$$

```julia
using Distributions
using Random
using Plots
using StatsBase
gui()
n = 10000;
mu = Array{Float64, 1}(undef, n + 1);
sigma2 = Array{Float64, 1}(undef, n + 1);
mu[1] = 1;
sigma2[1] = 1;
for t in range(2, n + 1)
    a = (1+20*sigma2[t-1])/(2+2*sigma2[t-1]);
    b = sigma2[t - 1]/(1+sigma2[t-1]);
    muza = rand(Normal(a, b), 1);
```

```
    mu[t] = muza[1];
    sigma2z = rand() * 400 - 200;
    v = rand();
    while v > exp(-0.5*sigma2z)*exp(-(0.5-mu[t])^2/2/exp(sigma2z))*400
        sigma2z = rand() * 400 - 200;
        v = rand();
    end
    sigma2[t] = exp(sigma2z);

end

histogram(mu, fillcolor=[:black], fillalpha=0);
println("Mean of mu found by Gibbs Sampling is: E[mu|sigma2,X=0.5]=",  mean(mu));
```




```
Mean of mu found by Gibbs Sampling is: E[mu|sigma2,X=0.5]=9.894546225145948
```

So we can see that the mode of μ is in $[9.8, 10]$, and the mean is 9.89.

(c)

Through the Expectation-Maximization method,

$$
\begin{aligned}
&\arg\max_{\mu} \int_{\mathcal{D}} \ln(f_{\{\mu,\sigma^2\}|X=0.5}) \cdot f_{\sigma^2|\{\mu,X=0.5\}} d\sigma^2 \\
=&\arg\max_{\mu} \int_{\mathcal{D}} \left(-\frac{3}{2}\ln(\sigma^2) - \frac{(0.5-\mu)^2}{2\sigma^2} - \frac{(\mu-10)^2}{2} \right) \cdot (\sigma^2)^{-3/2} \exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2\sigma^2} \Big) d\sigma^2 \\
=&\arg\max_{\mu} -\int_{\mathcal{D}} \frac{(0.5-\mu)^2}{2(\sigma^2)^{5/2}} \exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2\sigma^2} \Big) d\sigma^2 - \int_{\mathcal{D}} \frac{(\mu-10)^2}{2(\sigma^2)^{3/2}} \exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2\sigma^2} \Big) d\sigma^2 \\
=&\arg\max_{\mu} -A(0.5-\mu)^2 - B(\mu-10)^2 \\
=&\frac{A+20B}{2(A+B)}
\end{aligned}
$$

where

$$A=\int_{\exp(-200)}^{\exp(200)}\frac{1}{2(\sigma^2)^{5/2}}\exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2\sigma^2}\Big)d\sigma^2$$

$$B=\int_{\exp(-200)}^{\exp(200)}\frac{1}{2(\sigma^2)^{3/2}}\exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2\sigma^2}\Big)d\sigma^2$$

Applying variable transformation by setting $t=\dfrac{1}{\sigma^2}$,

$$A=\int_{\exp(-200)}^{\exp(200)}\frac{1}{2}t^{1/2}\exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2}t\Big)dt\approx\frac{\Gamma(\frac{3}{2})}{2((0.5-\mu^{(t-1)})^2/2)^{3/2}}$$

$$B=\int_{\exp(-200)}^{\exp(200)}\frac{1}{2}t^{-1/2}\exp\Big(-\frac{(0.5-\mu^{(t-1)})^2}{2}t\Big)dt\approx\frac{\Gamma(\frac{1}{2})}{2((0.5-\mu^{(t-1)})^2/2)^{1/2}}$$

so $\dfrac{B}{A}=\dfrac{(0.5-\mu^{(t-1)})^2/2}{1/2}=(0.5-\mu^{(t-1)})^2$ and therefore

$$\mu^{(t)}=\frac{1+20(0.5-\mu^{(t-1)})^2}{2+2(0.5-\mu^{(t-1)})^2}$$

```julia
using Distributions
using Random

n = 10000;
mu = Array{Float64, 1}(undef, n + 1);
mu[1] = 5;
muz = Array{Float64, 1}(undef, 1);
muz[1] = 0;
for t in range(2, n + 1)
    mu[t] = (20*(0.5-mu[t-1])^2+1)/2/((0.5-mu[t-1])^2+1);
    if abs(mu[t]-mu[t-1])<0.0001
        muz[1]=mu[t];
        break;
    end
end
println("Mode of mu found by Expectation Maximization method is: ",  muz[1]);
```

```
Mode of mu found by Expectation Maximization method is: 9.893543814369174
```

So the mode of $f_{\mu|X}$ is 9.89.

2.

(a)

Known that $\mathbf{Y}|\{\boldsymbol{\beta},\boldsymbol{\sigma}^2\}\sim$Normal$(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$, $\boldsymbol{\beta}\sim$Normal$(\boldsymbol{\beta_0},\boldsymbol{\Sigma_0})$, $\dfrac{1}{\sigma^2}\sim$Gamma$\Big(\dfrac{\nu_0}{2},\dfrac{\nu_0\sigma_0^2}{2}\Big)$, i.e.

$$f_{\mathbf{Y}|\{\mathbf{X},\boldsymbol{\beta},\frac{1}{\sigma^2}\}}=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Big(-\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\Big)$$

$$f_{\boldsymbol{\beta}} = \frac{1}{(2\pi\det(\boldsymbol{\Sigma_0}))^{n/2}}\exp\Big(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta_0})^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\Big)$$

$$f_{\frac{1}{\sigma^2}} = \frac{\Big(\frac{\nu_0\sigma_0^2}{2}\Big)^{\frac{\nu_0}{2}}}{\Gamma\Big(\frac{\nu_0}{2}\Big)}\cdot\Big(\frac{1}{\sigma^2}\Big)^{\frac{\nu_0}{2}}\exp\Big(-\frac{\nu_0\sigma_0^2}{2}\cdot\frac{1}{\sigma^2}\Big)$$

the joint distribution is given by

$$f_{\{\mathbf{Y},\boldsymbol{\beta},\frac{1}{\sigma^2}|\mathbf{X}\}} = f_{\mathbf{Y}|\{\mathbf{X},\boldsymbol{\beta},\frac{1}{\sigma^2}\}}\cdot f_{\boldsymbol{\beta}}\cdot f_{\frac{1}{\sigma^2}}$$

and the conditional distribution for $\frac{1}{\sigma^2}$ given $\mathbf{Y},\boldsymbol{\beta}$ and $\mathbf{X}$ is

$$\begin{aligned}
f_{\frac{1}{\sigma^2}|\{\boldsymbol{\beta},\mathbf{Y},\mathbf{X}\}} =& \frac{f_{\{\mathbf{Y},\boldsymbol{\beta},\frac{1}{\sigma^2}|\mathbf{X}\}}}{f_{\{\mathbf{Y},\boldsymbol{\beta}\}|\mathbf{X}}}\\
\propto& f_{\mathbf{Y}|\{\mathbf{X},\boldsymbol{\beta},\frac{1}{\sigma^2}\}}\cdot f_{\boldsymbol{\beta}}\cdot f_{\frac{1}{\sigma^2}}\\
\propto& \frac{1}{(\sigma^2)^{n/2}}\exp\Big(-\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\Big)\cdot\Big(\frac{1}{\sigma^2}\Big)^{\frac{\nu_0}{2}}\exp\Big(-\frac{\nu_0\sigma_0^2}{2}\cdot\frac{1}{\sigma^2}\Big)\\
\propto& \Big(\frac{1}{\sigma^2}\Big)^{\frac{n+\nu_0}{2}}\exp\Big(-\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})+\nu_0\sigma_0^2)}{2}\cdot\frac{1}{\sigma^2}\Big)\\
\sim& \text{Gamma}\Big(\frac{n+\nu_0}{2},\frac{\nu_0\sigma_0^2+\text{RSS}(\boldsymbol{\beta})}{2}\Big)
\end{aligned}$$

So the conditional posterior of σ^2 is given by

$$\sigma^2|\{\boldsymbol{\beta},\mathbf{Y},\mathbf{X}\}\sim\text{Inverse-Gamma}\Big(\frac{n+\nu_0}{2},\frac{\nu_0\sigma_0^2+\text{RSS}(\boldsymbol{\beta})}{2}\Big)$$

(b)

Known that the conditional distribution for $\boldsymbol{\beta}$ given $\mathbf{Y},\frac{1}{\sigma^2}$ and the conditional distribution for $\frac{1}{\sigma^2}$ given $\mathbf{Y},\boldsymbol{\beta}$ are

$$\begin{aligned}
\boldsymbol{\beta}|\{\sigma^2,\mathbf{Y},\mathbf{X}\}&\sim\text{Normal}\Big(\mathbf{m},\mathbf{V}\Big)\\
\sigma^2|\{\boldsymbol{\beta},\mathbf{Y},\mathbf{X}\}&\sim\text{Inverse-Gamma}\Big(\alpha,\beta\Big)
\end{aligned}$$

where

$$\mathbf{m} = (\boldsymbol{\Sigma_0}^{-1}+\mathbf{X^TX}/\sigma^2)^{-1}(\boldsymbol{\Sigma_0}\boldsymbol{\beta_0}^{-1}+\mathbf{X^TY}/\sigma^2)^{-1}$$

$$\mathbf{V} = (\boldsymbol{\Sigma_0}^{-1}+\mathbf{X^TX}/\sigma^2)^{-1}$$

$$\alpha = \frac{n+\nu_0}{2}, \beta = \frac{\nu_0\sigma_0^2+\text{RSS}(\boldsymbol{\beta})}{2}$$

and positivity is satisfied, so we can use Gibbs sampling.

1. For $\boldsymbol{\beta}$, since $\mathbf{V} = (\boldsymbol{\Sigma_0}^{-1} + \mathbf{X^T X}/\sigma^2)^{-1}$ is symmetric positive definite, we can find Cholesky Decomposition of it

$$\mathbf{V} = \mathbf{CC^T}$$

and therefore

$$\mathbf{V} = \mathbf{CC^T} \Rightarrow \mathbf{C^{-1}V(C^T)^{-1}} = \mathbb{I} \Rightarrow \mathbf{CVC^T} = \mathbb{I}$$

Assume $\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\beta'}$, then $\boldsymbol{\beta'}$ $\sim$Normal$(\mathbf{C^{-1}m}, \mathbf{C^{-1}V(C^{-1})^T})$ $\sim$Normal$(\mathbf{C^{-1}m}, \mathbb{I})$ and therefore we can sample β'_i $\sim$Normal$((\mathbf{C^{-1}M})_i, 1)$ where $(\mathbf{C^{-1}M})_i$ denotes the i^{th} element of $\mathbf{C^{-1}M}$. And the sample of $\boldsymbol{\beta}$ can be obtained through

$$\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\beta'}$$

2. For σ^2, we know that

$$\frac{1}{\sigma^2}|\{\boldsymbol{\beta}, \mathbf{Y}, \mathbf{X}\} \sim \text{Gamma}\Big(\frac{n+\nu_0}{2}, \frac{\nu_0\sigma_0^2 + \text{RSS}(\boldsymbol{\beta})}{2}\Big) \sim \text{Gamma}\Big(\alpha, \beta\Big)$$

then apply random variable transformation by setting $t = \arctan\left(\frac{1}{\sigma^2}\right) \in [0, \frac{\pi}{2})$,

$$f_{t|\{\boldsymbol{\beta},\mathbf{Y},\mathbf{X}\}} = \frac{\beta^\alpha}{\Gamma(\alpha)} \tan^{\alpha-1}(t) \exp(-\beta \tan(t)) \cdot (1 + \tan^2(t)) < \infty$$

So we can apply rejection sampling to get t, which leads to a sample of $\sigma^2 = \dfrac{1}{\tan(t)}$.

With the sample obtained through Gibbs sampling, we can calculate obtain a point estimate of $\boldsymbol{\beta}|\{\mathbf{X}, \mathbf{Y}\}$.

(c)

For the predictive distribution of $\mathbf{Y}^*$,

$$\begin{aligned}
f_{\mathbf{Y}^*|\{\mathbf{X},\mathbf{X}^*\}} &= \int_{\mathcal{D}_1}\int_{\mathcal{D}_2} f_{\{\mathbf{Y}^*,\boldsymbol{\beta},\sigma^2\}|\{\mathbf{X}^*,\mathbf{X},\mathbf{Y}\}} d\boldsymbol{\beta} d\sigma^2 \\
&= \int_{\mathcal{D}_1}\int_{\mathcal{D}_2} f_{\mathbf{Y}^*|\{\mathbf{X}^*,\mathbf{X},\mathbf{Y},\boldsymbol{\beta},\boldsymbol{\sigma^2}\}} \cdot f_{\{\boldsymbol{\beta},\sigma^2\}|\{\mathbf{X}^*,\mathbf{X},\mathbf{Y}\}} d\boldsymbol{\beta} d\sigma^2 \\
&= \int_{\mathcal{D}_1}\int_{\mathcal{D}_2} f_{\mathbf{Y}^*|\{\mathbf{X}^*,\boldsymbol{\beta},\boldsymbol{\sigma^2}\}} \cdot f_{\{\boldsymbol{\beta},\sigma^2\}|\{\mathbf{X},\mathbf{Y}\}} d\boldsymbol{\beta} d\sigma^2
\end{aligned}$$

with the predictive distribution we can find the prediction interval for $X_k = x_k^*$ is

$$[\ell_y, u_y]$$

with probability

$$\int_{\ell_y}^{u_y} f_{\mathbf{Y}^*|\{\mathbf{X},\mathbf{X}^*\}} dy = 1 - \alpha$$