Bayesian Statistics HW-4

Weijia Zhao

March 6, 2022

Caution I keep 6 digits for all simulated results. Since it involves random sampling from certain non-degenerative distribution, the results may be slightly different in different trials (and thus from the official solution)

Exercise 1

Metropolis for Correlation Coefficient

Pairs (X_i, Y_i), $i = 1, ..., n$ with density $f(x, y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}$, the prior of ρ given by $\pi(\Sigma) = \frac{1}{|\Sigma|^{3/2}} = \frac{1}{(1-\rho^2)^{3/2}}\mathbb{I}(-1 \le \rho \le 1)$ where $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

(a) When (X_i, Y_i) are observed, the likelihood is given by

$$\begin{aligned} L(\rho|x_1, x_2, ...x_n) &= \Pi_{i=1}^n f(x_i, y_i|\rho) \\ &= \Pi_{i=1}^n \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x_i^2-2\rho x_i y_i+y_i^2)} \\ &= (\frac{1}{2\pi\sqrt{1-\rho^2}})^n e^{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^n(x_i^2-2\rho x_i y_i+y_i^2)} \end{aligned}$$

The posterior distribution is proportional to

$$\begin{aligned} \pi(\rho|(x_i, y_i)_{i=1}^n) &\propto L(\theta|x_1, x_2, ...x_n) \cdot \pi(\rho) \\ &= (\frac{1}{2\pi\sqrt{1-\rho^2}})^n e^{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^n(x_i^2-2\rho x_i y_i+y_i^2)} \cdot \frac{1}{(1-\rho^2)^{3/2}}\mathbb{I}(-1 \le \rho \le 1) \\ &\propto (1-\rho^2)^{-\frac{n}{2}-\frac{3}{2}} e^{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^n(x_i^2-2\rho x_i y_i+y_i^2)}\mathbb{I}(-1 \le \rho \le 1) \end{aligned}$$

(b) Assume $n = 100$ and the observed pairs (X_i, Y_i) have the summary statistics $\sum_{i=1}^{100} x_i^2 = 114.9707$, $\sum_{i=1}^{100} y_i^2 = 105.9196$ and $\sum_{i=1}^{100} x_i y_i = 82.5247$. The Metropolis algorithm is given by the following steps

- Start with an arbitrary initial value ρ_0
- At stage n, generate new proposal $\hat{\rho}$ from the uniform distribution $\mathbb{U}(\rho_n - 0.1, \rho_n + 0.1)$
- With probability $\theta(\rho_n, \hat{\rho})$[1] we accept the new proposal $\rho_{n+1} = \hat{\rho}$ and with probability $1 - \theta(\rho_n, \hat{\rho})$ we reject the new proposal and $\rho_{n+1} = \rho_n$

[1]Note that we flip the notation ρ and θ compared with the lecture notes since ρ is already taken

- Increase n and go to the second step

Here in this case the new proposal is from $\mathbb{U}(\rho_n - 0.1, \rho_n + 0.1)$, thus

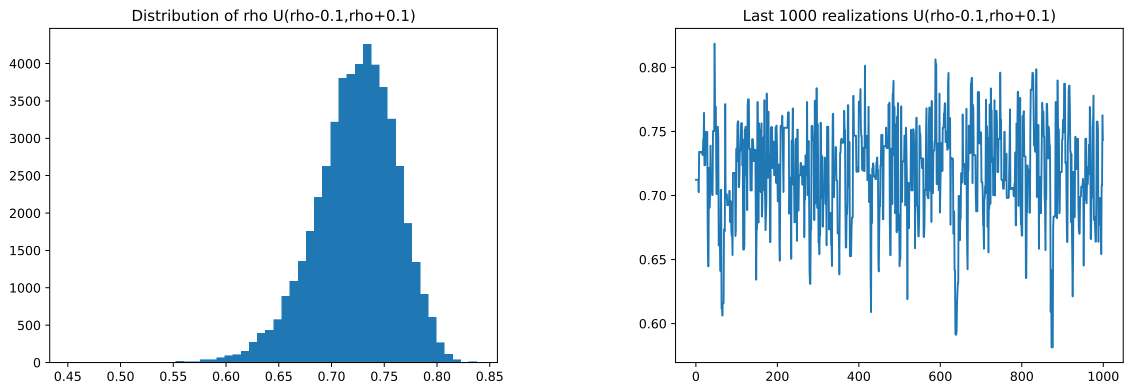
$$\begin{aligned} q(\hat{\rho}|\rho_n) &= \frac{1}{(\rho_n + 0.1) - (\rho_n - 0.1)} \mathbb{I}(\rho_n - 0.1 \le \hat{\rho} \le \rho_n + 0.1) \\ &= 5\mathbb{I}(\rho_n - 0.1 \le \hat{\rho} \le \rho_n + 0.1) \end{aligned}$$

On the other hand, given $\hat{\rho}$, the conditional distribution of ρ_n is $\mathbb{U}(\hat{\rho} - 0.1, \hat{\rho} + 0.1)$, so $q(\rho_n|\hat{\rho}) = 5\mathbb{I}(\hat{\rho} - 0.1 \le \rho_n \le \hat{\rho} + 0.1)$, thus the acceptance ratio is given by

$$\begin{aligned} \theta(\rho_n, \hat{\rho}) &= \min(1, \frac{q(\rho_n|\hat{\rho})\pi(\hat{\rho})}{q(\hat{\rho}|\rho_n)\pi(\rho_n)}) \\ &= \min(1, \frac{\mathbb{I}(\hat{\rho} - 0.1 \le \rho_n \le \hat{\rho} + 0.1)}{\mathbb{I}(\rho_n - 0.1 \le \hat{\rho} \le \rho_n + 0.1)} \cdot \frac{(1-\hat{\rho}^2)^{-\frac{n}{2}-\frac{3}{2}} e^{-\frac{1}{2(1-\hat{\rho}^2)} \sum_{i=1}^n (x_i^2 - 2\hat{\rho} x_i y_i + y_i^2)} \mathbb{I}(-1 \le \hat{\rho} \le 1)}{(1-\rho_n^2)^{-\frac{n}{2}-\frac{3}{2}} e^{-\frac{1}{2(1-\rho_n^2)} \sum_{i=1}^n (x_i^2 - 2\rho_n x_i y_i + y_i^2)} \mathbb{I}(-1 \le \rho_n \le 1)}) \\ &= \min(1, \frac{(1-\hat{\rho}^2)^{-\frac{n}{2}-\frac{3}{2}} e^{-\frac{1}{2(1-\hat{\rho}^2)} \sum_{i=1}^n (x_i^2 - 2\hat{\rho} x_i y_i + y_i^2)} \mathbb{I}(-1 \le \hat{\rho} \le 1)}{(1-\rho_n^2)^{-\frac{n}{2}-\frac{3}{2}} e^{-\frac{1}{2(1-\rho_n^2)} \sum_{i=1}^n (x_i^2 - 2\rho_n x_i y_i + y_i^2)} \mathbb{I}(-1 \le \rho_n \le 1)}) \end{aligned}$$

Here notice that $\frac{\mathbb{I}(\hat{\rho}-0.1 \le \rho_n \le \hat{\rho}+0.1)}{\mathbb{I}(\rho_n - 0.1 \le \hat{\rho} \le \rho_n + 0.1)} = 1$ and thus cancels, this is because both probability indicates that $|\rho' - \rho_n| \le 0.1$, this is indeed a symmetric Metropolis. We can then substitute $n = 100$ and $\sum_{i=1}^{100} x_i^2 = 114.9707$, $\sum_{i=1}^{100} y_i^2 = 105.9196$ $\sum_{i=1}^{100} x_i y_i = 82.5247$ into the expression above

(c) The For Bayes estimator, we just need to calculate the average of simulated ρ, which is



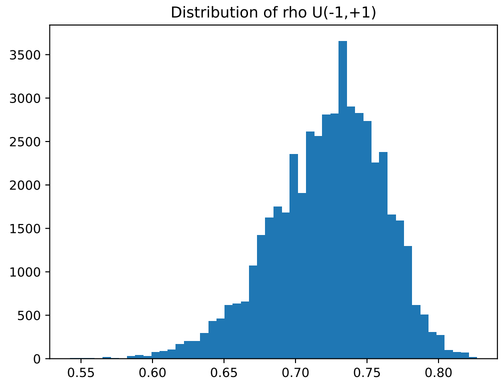
(a) Histogram plot

(b) Last 1000 realizations

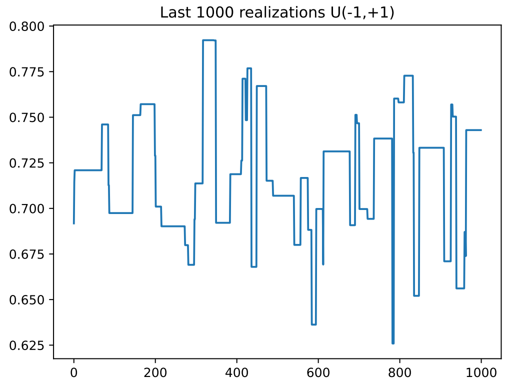
Figure 1: Sampling ρ using $\mathbb{U}(\rho_{i-1} - 0.1, \rho_{i-1} + 0.1)$

given by 0.722986 here (This is simulated result, thus may vary a bit in different experiments)

(d)The graphs are shown below and the Bayes estimator can again be obtained by taking the average for simulated ρ and the value is 0.722697 in this case. The histogram graph looks very similar to the previous case (it is a bit more concentrated around the center). The plot for last 1000 realizations looks more different, we see a lot of flat areas, which means the next simulated ρ takes the same value as the last simulated ρ, i.e. the new proposal is rejected. Here we can see that under the new regime, the probability of accepting a proposal is lower compared to the previous case



(a) Histogram plot



(b) Last 1000 realizations

Figure 2: Sampling ρ using $\mathbb{U}(-1, +1)$

Exercise 2

Gibbs Sampler and Lifetimes with Multiplicative Frailty
Exponentially distributed lifetime with constant hazard rate λ. Introduce heterogeneity of hazard with a multiplicative frailty parameter μ, distribution of lifetimes T_i given by

$$T_i \sim f(t_i|\lambda, \mu) = \lambda\mu e^{-\lambda\mu t_i}, t_i > 0, \lambda, \mu > 0$$

Prior of (λ, μ) $\pi(\lambda, \mu) \propto \lambda^{c-1}\mu^{d-1}e^{-\alpha\lambda-\beta\mu}$, i.e. λ, μ apriori independent with distribution $\Gamma(c, \alpha)$, $\Gamma(d, \beta)$, with c, d, α, β known and positive, observe $t_1, t_2, ...t_n$

(a) The posterior distribution of λ is given by

$$\begin{aligned}
\pi(\lambda|\mu, t_{i\,i=1}^{\,n}) &\propto \pi(\lambda) \cdot \Pi_{i=1}^n f(t_i|\lambda\mu) \\
&= \lambda^{c-1}e^{-\alpha\lambda}\lambda^n\mu^n e^{-\lambda\mu\sum_{i=1}^n t_i} \\
&= \lambda^{n+c-1}\mu^n e^{-\lambda(\alpha+\mu\sum_{i=1}^n t_i)} \\
&\propto \lambda^{n+c-1}e^{-\lambda(\alpha+\mu\sum_{i=1}^n t_i)}
\end{aligned}$$

This is the core of a Gamma distribution and we can see $[\lambda|\mu, t_{i\,i=1}^{\,n}] \sim Gamma(n+c, \alpha+\mu\sum_{i=1}^n t_i)$

Similarly, the posterior distribution of μ is given by

$$\begin{aligned}
\pi(\mu|\lambda, t_{i\,i=1}^{\,n}) &\propto \pi(\mu) \cdot \Pi_{i=1}^n f(t_i|\lambda\mu) \\
&= \mu^{d-1}e^{-\beta\mu}\lambda^n\mu^n e^{-\lambda\mu\sum_{i=1}^n t_i} \\
&= \mu^{n+d-1}\lambda^n e^{-\mu(\beta+\lambda\sum_{i=1}^n t_i)} \\
&\propto \mu^{n+d-1}e^{-\mu(\beta+\lambda\sum_{i=1}^n t_i)}
\end{aligned}$$

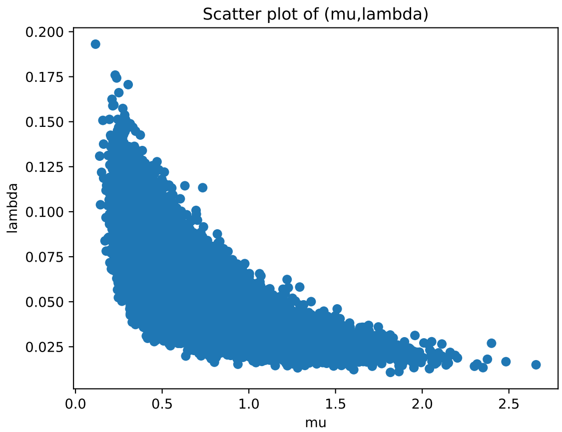
This is also the core of a Gamma distribution and we can see that $[\mu|\lambda, t_{i\,i=1}^{\,n}] \sim Gamma(n + d, \beta + \lambda\sum_{i=1}^n t_i)$

(b) The Gibbs sampler follows the following several steps

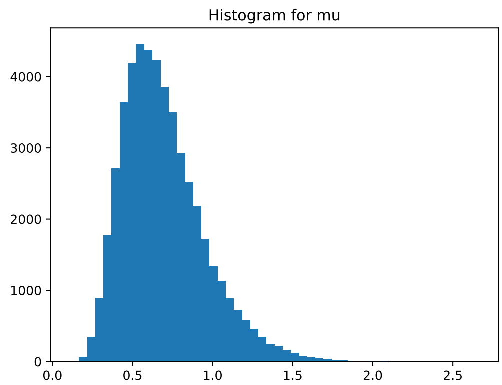
- Start with $\mu = 0.1$
- Sample $\lambda' \sim Gamma(n + c, \alpha + \mu\sum_{i=1}^n t_i)$, set $\lambda = \lambda'$

- Sample $\mu' \sim Gamma(n + d, \beta + \lambda \sum_{i=1}^{n} t_i)$, set $\mu = \mu'$
- With the newly obtained μ and λ, got back to step (2) and repeat until we get enough observations (ignore the first 1000 observations)

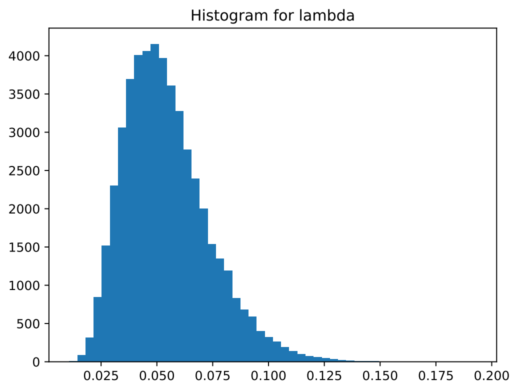
(c) The graphs are shown below. Note that we remove one more observation from the series of μ since it has one more initial value compared with the series of λ. The posterior mean of μ is given by 0.695968 and posterior mean of λ is given by 0.054395. The posterior variance of μ is given by 0.066731 and posterior variance of λ is given by 0.000361. The 95% equitailed credible set (calculated by percentile of the simulated values) for μ is (0.316929, 1.311192) and the 95% equitailed credible set for λ is (0.025294, 0.098936) (These are simulated results, thus may vary a bit in different experiments)



(a) Histogram plot



(b) Histogram for μ



(c) Histogram for λ

Figure 3: Gibbs Sampling

(d) Bayes estimator of the product is given by $\frac{1}{50000} \sum_{i=1}^{50000} \lambda_i \cdot \mu_i = 0.034253$ (This is simulated result, thus may vary a bit in different experiments)

Code for Q1

```
import numpy as np
import matplotlib.pyplot as plt

#define statistics
np.random.seed(123456)
sum_x2=114.9707
sum_y2=105.9196
sum_xy=82.5247
n=100
#define density function
def f(rho):
    term1=(1-rho**2)**(-n/2-3/2)
    term2=1/(2*(1-rho**2))*(sum_x2-2*rho*sum_xy+sum_y2)
    return term1*np.exp(-term2) if (rho<=1 and rho>=-1) else 0

#sampling
rho0=0
results=[rho0]
for i in range(51000):
    new_rho=np.random.uniform(results[-1]-0.1,results[-1]+0.1)
    cut=min(f(new_rho)/f(results[-1]),1)
    if np.random.uniform(0,1)<cut:
        results.append(new_rho)
    else:
        results.append(results[-1])
results=results[1000:]
print('bayes estimator is '+str(np.mean(results)))
#histogram for all rhos after removing the first 1000 obs
plt.hist(results,bins=50)
plt.title('Distribution of rho U(rho-0.1,rho+0.1)')
plt.savefig('q1_partc1.pdf')
plt.show()

#realization of the last 1000 obs
plt.plot(results[-1000:])
plt.title('Last 1000 realizations U(rho-0.1,rho+0.1)')
plt.savefig('q1_partc2.pdf')
plt.show()


#sampling
rho0=0
results=[rho0]
for i in range(51000):
    new_rho=np.random.uniform(-1,1)
    cut=min(f(new_rho)/f(results[-1]),1)
    if np.random.uniform(0,1)<cut:
```

```
        results.append(new_rho)
    else:
        results.append(results[-1])
results=results[1000:]
print('bayes estimator is '+str(np.mean(results)))


#histogram for all rhos after removing the first 1000 obs
plt.hist(results,bins=50)
plt.title('Distribution of rho U(-1,+1)')
plt.savefig('q1_partd1.pdf')
plt.show()

#realization of the last 1000 obs
plt.plot(results[-1000:])
plt.title('Last 1000 realizations U(-1,+1)')
plt.savefig('q1_partd2.pdf')
plt.show()
```

Code for Q2

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(1234)
#define statistics
sum_t=512
n=20
c=3
d=1
alpha=100
beta=5

mu0=0.1

mu_collector=[mu0]
lambda_collector=[]
for i in range(51000):
    #sample lambda
    cur_lambda=np.random.gamma(n+c,1/(mu_collector[-1]*sum_t+alpha))
    lambda_collector.append(cur_lambda)
    #sample mu
    cur_mu=np.random.gamma(n+d,1/(lambda_collector[-1]*sum_t+beta))
    mu_collector.append(cur_mu)

mu_collector=mu_collector[1001:]
lambda_collector=lambda_collector[1000:]

print('Posterior mean of mu is '+str(np.mean(mu_collector)))
print('Posterior variance of mu is '+str(np.var(mu_collector)))

print('Posterior mean of lambda is '+str(np.mean(lambda_collector)))
print('Posterior variance of lambda is '+str(np.var(lambda_collector)))


print('The 95% equitailed credible set for mu is ('+
      str(np.percentile(mu_collector,2.5))+', '+
      str(np.percentile(mu_collector,97.5))+')')
print('The 95% equitailed credible set for lambda is ('+
      str(np.percentile(lambda_collector,2.5))+', '+
      str(np.percentile(lambda_collector,97.5))+')')
print('Bayes estimator of the product is given by '+
      str(np.mean([mu_collector[i]*lambda_collector[i]
                   for i in range(len(mu_collector))]))+')')
#scatter plot
plt.scatter(mu_collector,lambda_collector)
plt.title('Scatter plot of (mu,lambda)')
plt.xlabel('mu')
```

```
plt.ylabel('lambda')
plt.savefig('q2_partc_scatter.pdf')
plt.show()

plt.hist(mu_collector,bins=50)
plt.title('Histogram for mu')
plt.savefig('q2_partc_hist_mu.pdf')
plt.show()

plt.hist(lambda_collector,bins=50)
plt.title('Histogram for lambda')
plt.savefig('q2_partc_hist_lambda.pdf')
plt.show()
```