Question 1

I. The problem

In this problem, we observe n i.i.d. values from an unknown (μ, σ) normal distribution which we want to find the maximum likelihood estimates. Since the data is rounded into d decimal places, we need to find mle by maximizing the probability of having a point in the interval which could be rounded into our data point.

II. The method

a) In function rnd_norm_mle1, we could find the each probability by integrating the normal density function over the interval [$X_i - 0.5 \times 10^{-d-1}$, $X_i + 0.5 \times 10^{-d-1}$],

In particular,

$$\int_{X_i - 0.5 \times 10^{-d-1}}^{X_i + 0.5 \times 10^{-d-1}} f(x)dx$$

Therefore the log likelihood can be calculated by:

$$L(\mu, \sigma) = \sum_{X_i} log(\int_{X_i - 0.5 \times 10^{-d-1}}^{X_i + 0.5 \times 10^{-d-1}} f(x)dx)$$

We should use the log of σ so that there is no constraint on estimations.

b) In the function rnd_norm_mle2 each probability is generated by

$$\phi(X_i + 0.5 \times 10^{-d-1}) - \phi(X_i - 0.5 \times 10^{-d-1})$$

Where $\phi(X)$ is the normal distribution function.

Similarly, $$L(\mu, \sigma) = \sum_{X_i} log(\phi(X_i + 0.5 \times 10^{-d-1} - \phi(X_i - 0.5 \times 10^{-d-1}))))$$

c) We could use the sample mean and the exponential of sample std as the initial value of nlm function.

III . The output

Table 1 : Data Summary

Dataset	Number of data	Sample mean	Sample standard deviation	True mean	True standard deviation
X1	25	2.408	0.9560858	2.24	1
X2	50	2.238	0.1085902	2.24	0.1
X3	4000	2.2355	0.8515991	2.24	0.8

Table 2: Estimated mean

Dataset	Number of data	Estimate mean using rnd_norm_mle1	Estimate mean using rnd_norm_mle2
X1	25	2.4079988	2.4079988
X2	50	2.237962	2.237962
X3	4000	2.2355587	2.2355587

Table 3: Estimated standard error

Dataset	Number of data	Estimate mean using rnd_norm_mle1	Estimate mean using rnd_norm_mle2
X1	25	0.9363236	0.9363236
X2	50	0.10351	0.10351
X3	4000	0.8007662	0.8007662

Table 4: Operating time

Dataset	Number of data	Estimate using rnd_norm_mle1	Estimate using rnd_norm_mle2	Estimate using sample mean
X1	25	0.03	0	0
X2	50	0.07	0.02	0
X3	4000	4.08	1.01	0

IV. Comments

a) From the table 2 and table 3, we can clearly see that the 2 estimate functions returns exactly the same results. So in terms of precision, the 2 algorithms are indifferent.
b) To compare the mean estimation using mle and sample mean, we can check table 1 and table 2. From the table there is a very little difference between the 2 methods, for all n = 25,50 and 4000. Therefore, using maximum likelihood estimates of μ and simply using the sample mean returns very similar results. In addition, as n gets larger, from 25 to 50, the error of estimates becomes smaller for both sample mean method and mle method. However, as n changes from 50 to 4000, both of the methods return a larger error.
c) To compare the standard deviation estimation, we need to check table 1 and table 3. In the n = 25 case, sample standard deviation makes a better estimate, while in the n = 50, and n =4000 cases, the mle method shows slightly better estimates. What's more, as n getting larger, the error in mle method becomes significantly smaller.
d) From table 4, the operating time using integrate is longer than that using pnorm. The operating time using sample mean method is always 0.

Question 2

I. The problem

In this problem, we observe n pairs of values. The two values within each pair may be dependent, following a bivariate normal distribution. We are also given a prior density distribution with $\sigma \in (0,\infty)$ and $\rho \in (-1, +1)$. We need to compute the normalizing constant for the posterior distribution given data X and the marginal posterior density (with correct normalization) of ρ given data X.

II. The method and functions

$$f(\sigma, \rho | X_1, X_2, \ldots, X_n) = \frac{h(\sigma,\rho)L(X_i|\sigma,\rho)}{\int_{-1}^{1}\int_{0}^{-\infty} h(\sigma,\rho)L(X_i|\sigma,\rho)d\sigma d\rho}$$

Since the prior distribution is singular at ρ = 0, we need to split the integral into 2 parts. Thus the normalizing constant C can be calculated by:

$$C = \int_{-1}^{0}\int_{0}^{-\infty} h(\sigma,\rho)L(X_i|\sigma,\rho)d\sigma d\rho + \int_{0}^{1}\int_{0}^{-\infty} h(\sigma,\rho)L(X_i|\sigma,\rho)d\sigma d\rho$$

Table of functions

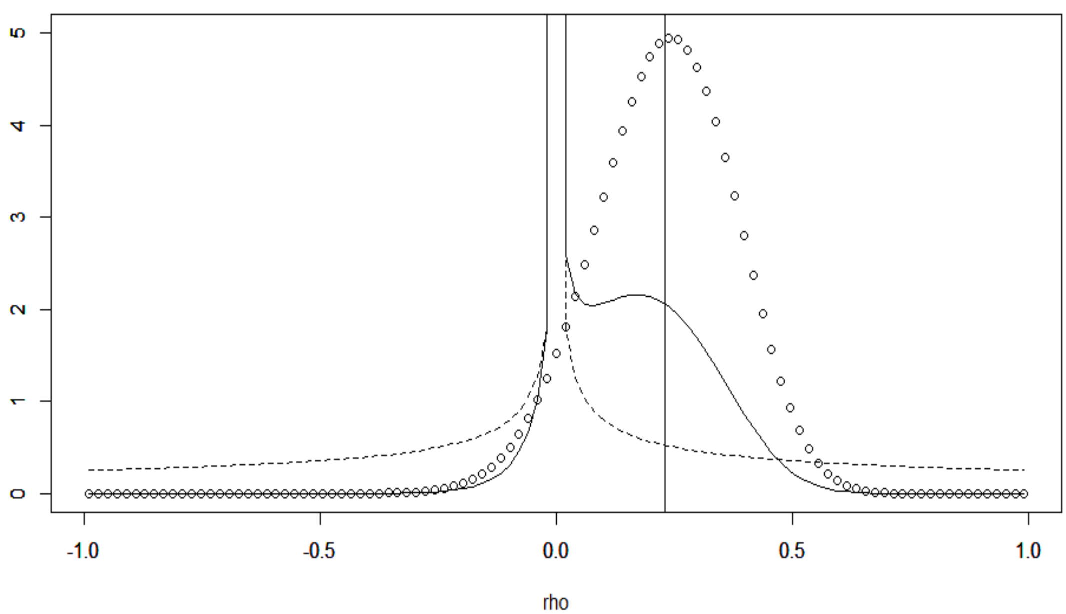
Function Name	Function description
jd	Calculate the joint density of single data pairs given σ and ρ
bvn_prior	Calculate the prior density, return a vector.
bnv_likelihood	Returns the likelihood for given σ and ρ values and the data
lik_v	Given data, returns a vector of likelihood for different σ values and a single ρ value.
m_rho	Marginal posterior density (before normalization) of ρ, given data. Return vectors.
bvn_normalize	Compute the normalizing constant for the posterior distribution given data X
bvn_posterior_rho	Marginal posterior density (with normalization) of ρ, given data.
lik_rho	Marginal likelihood function of ρ with given data.

III. The output

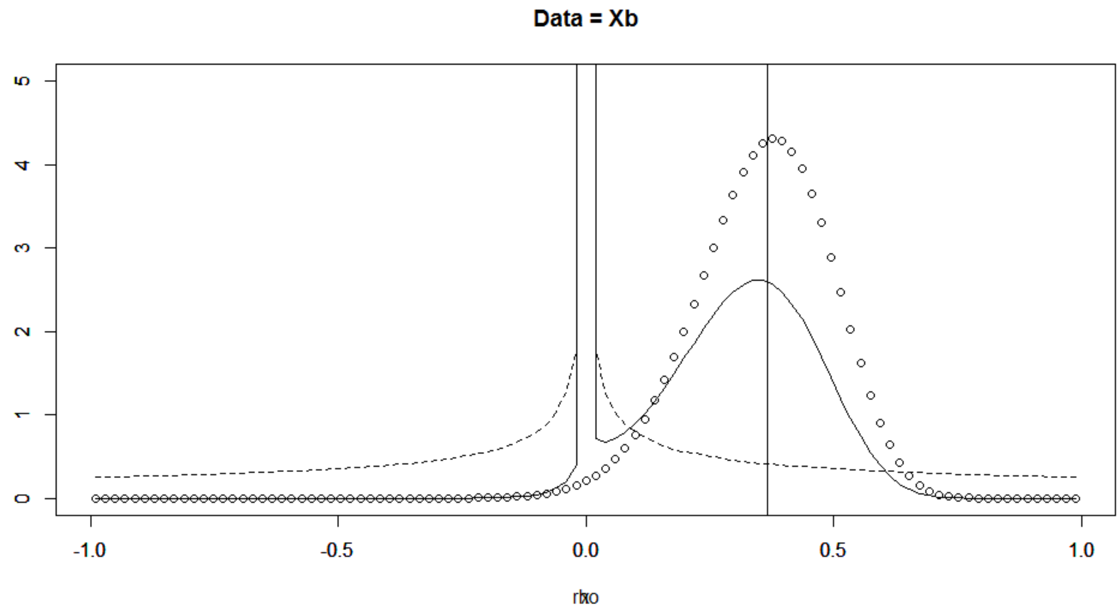
Normalizing constant for Xa	1.204831e-47
Normalizing constant for Xb	1.871503e-50

Plot 1

Data = Xa



Plot 2



IV. Comments

a) From plot 1 and 2, we can find that the marginal likelihood distribution has local maximum at around sample correlation.

b) The marginal posterior function also attains local maximum at around sample correlation. However the local maximum for data Xa is a little bit divergent from x = sample correlation.

The reason might be that the marginal prior decreases around the sample correlation, which draws the local maximum of function $L(Xa, \rho)h(\rho)$ leftward, since $f(\rho|Xa) = \frac{L(Xa,\rho)h(\rho)}{Ca}$

c) As we know, the true values of correlations for Xa and Xb are 0 and 0.25. The ρ estimate using posterior is not effective when the data pairs are uncorrelated. For non-zero ρ value, the posterior gives a reasonable estimation.

Bonus Question

To find out the size of n which the function start to fail due to underflow, I wrote a function to search which n makes the normalizing constant zero. My function shows that the n is between 250 and 258. The likelihood is too closed to zero which had caused the underflow. To alleviate the issue, we can multiply a large constant (i.e 10^50) to the likelihood function. This has no effect on the posterior distribution, since

$$f(\sigma, \rho | X_1, X_2, \ldots, X_n) = \frac{h(\sigma,\rho) \times C \times L(X_i|\sigma,\rho)}{\int_{-1}^{1} \int_{0}^{-\infty} h(\sigma,\rho) C L(X_i|\sigma,\rho) d\sigma d\rho}$$

By doing this, the overflow issue can be alleviated.