

Problem 1 (graded by Yiran) - 50 points

(a) 4 points

In a class, among 20 students, 8 are female, and 12 are male. 2 of the female students are taller than 170 cm, and 8 of the male students are taller than 170 cm. Suppose we randomly pick a student, let

x : the student is female;

y : the student is taller than 170 cm.

Then,

$P(x, y)$ is the probability that the student is both female and is taller than 170 cm, which is equal to $2/20 = 0.1$.

$P(x)$ is the probability that the student is female, which is equal to $8/20 = 0.4$.

$P(y|x)$ is the probability that known the student is female, she is taller than 170 cm, which is equal to $2/8 = 0.25$.

$P(y)$ is the probability that the student is taller than 170 cm, which is equal to $(2+8)/20 = 0.5$.

$P(x|y)$ is the probability that known the student is taller than 170 cm, the student is female, which is equal to $2/(2+8) = 0.2$.

We see that:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

(b) 8 points

Independent

Let

x : I get an A for Ge/ESE118.

y : The next president of the U.S. is an Republican.

These two events are independent, because x happens, does not affect the probability of y , vice versa.

Let's assume $P(x) = 3/5$, and $P(y) = 1/2$.

Suppose I get an A with $P(x)$. It doesn't affect the election at all, and there is still 1 in 2 odds that the next president will be an Republican. Therefore, to make both happen, $P(x, y) = P(x)P(y)$. Similary, suppose the Republican wins the election with $P(y)$. It doesn't affect my odd to get an A, and to make both happen, $P(x, y) = P(y)P(x)$.

Intuitively, the rule holds because the two events are independent - one happens does not affect the other; therefore, to make both happen, we need to multiply $P(x)$ and $P(y)$.

Dependent

Let

x : The next president of the U.S. is an Democratic.

y : The next president of the U.S. is an Republican.

These two events are not independent, because either of them happens, will affect the probability of the other.

Let's assume $P(x) = P(y) = \frac{1}{2}$. Because it's impossible that the next president is both an Democratic and an Republican, $P(x, y) = 0 \neq P(x)P(y)$.

(c) 8 points**(c.i) 3 points**

$$E(x) = \int_{-\infty}^{\infty} xP(x)dx$$

Since x is an odd function, and $P(x)$ is an even function, their product is an odd function. Integration of an odd function over symmetric boundaries as $[-\infty, \infty]$ is 0. Therefore,

$$E(x) = 0$$

(c.ii) 5 points

$$\begin{aligned} E(x^2) &= \int_{-\infty}^{\infty} x^2 P(x) dx \\ &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \end{aligned}$$

Since

$$\left[\exp\left(-\frac{x^2}{2\sigma^2}\right) \right]' = -\frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Then

$$\begin{aligned} E(x^2) &= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left[\exp\left(-\frac{x^2}{2\sigma^2}\right) \right]' dx \\ &= -\frac{\sigma}{\sqrt{2\pi}} \left[x \exp\left(-\frac{x^2}{2\sigma^2}\right) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right] \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \end{aligned}$$

Since

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-x^2/2} dx &= \sqrt{2\pi} \\ &= \int_{-\infty}^{\infty} \exp\left[-\frac{\left(\frac{x}{\sigma}\right)^2}{2}\right] d\left(\frac{x}{\sigma}\right) \\ &= \frac{1}{\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \end{aligned}$$

Then

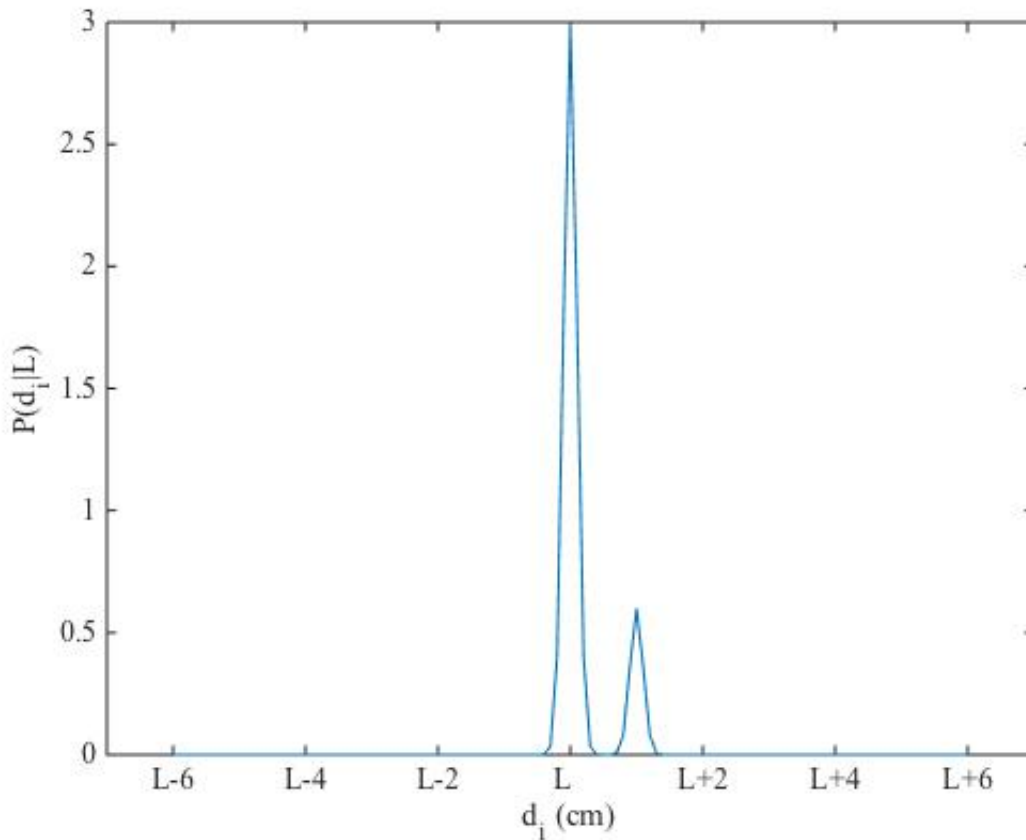
$$E(x^2) = \frac{\sigma}{\sqrt{2\pi}} \sqrt{2\pi} \sigma = \sigma^2$$

(d) 30 points

(d.i) 10 points

$$\begin{aligned}
P(d_i|L) &= \frac{3}{4}\mathcal{N}(L, \sigma) + \frac{1}{4}\mathcal{N}(L+1, \sigma) \\
&= \frac{3}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(d_i - L)^2}{2\sigma^2}\right) + \frac{1}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(d_i - (L+1))^2}{2\sigma^2}\right) \\
&= \frac{15}{2} \frac{1}{\sqrt{2\pi}} \exp(-50(d_i - L)^2) + \frac{3}{2} \frac{1}{\sqrt{2\pi}} \exp(-50(d_i - (L+1))^2)
\end{aligned}$$

where $\sigma = 0.1$ cm, and L, d_i are in cm.



(d.ii) 10 points

From Bayes' theorem,

$$\begin{aligned}
P(L|d_i) &= \frac{P(d_i|L)P(L)}{P(d_i)} \\
&\propto P(d_i|L) = \frac{3}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(d_i - L)^2}{2\sigma^2}\right) + \frac{1}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(d_i - (L+1))^2}{2\sigma^2}\right)
\end{aligned}$$

where we assume the prior distribution $P(L)$ is uniform, and $P(d_i)$ is a constant (but different for different d_i).

The constant ahead of $P(d_i|L)$ (denoted as “ c ”) is determined by

$$\int_{-\infty}^{\infty} P(L|d_i)dL = \int_{-\infty}^{\infty} cP(d_i|L)dL = 1$$

We can re-write $P(d_i|L)$ as

$$P(d_i|L) = \frac{3}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(L-d_i)^2}{2\sigma^2}\right) + \frac{1}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(L-(d_i-1))^2}{2\sigma^2}\right)$$

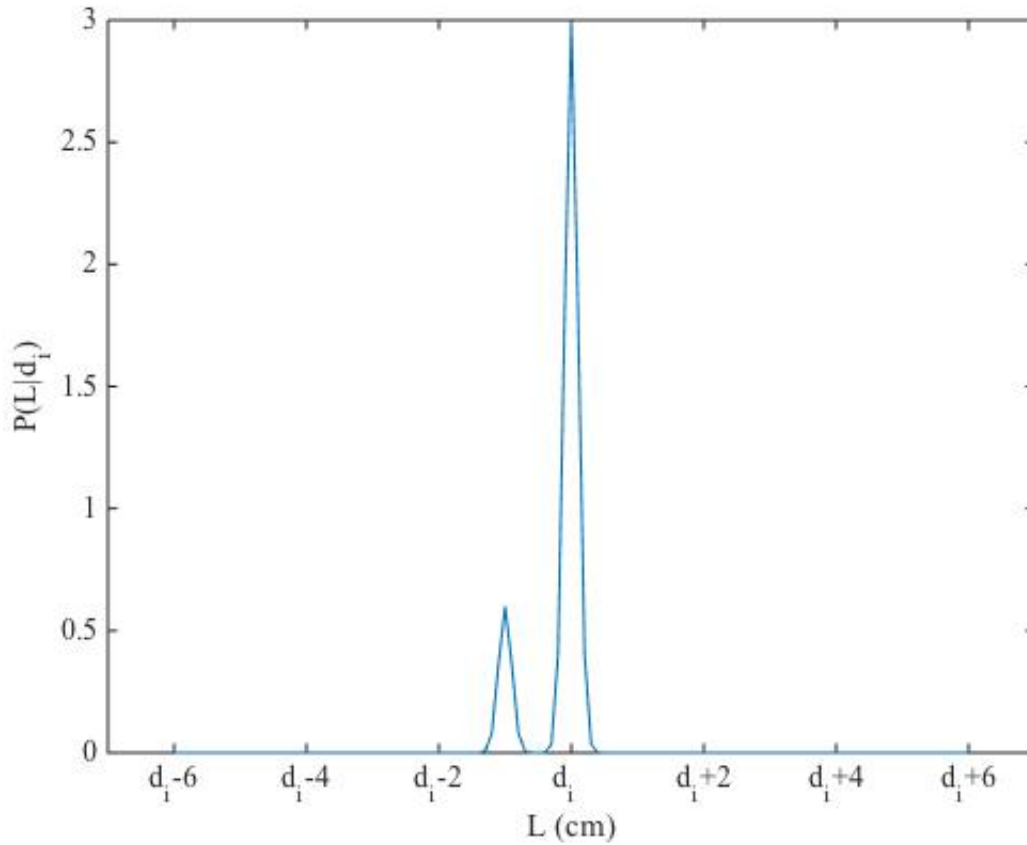
which is a summation of two normal distributions (centered at d_i , and $d_i - 1$), weighted by $\frac{3}{4}$ and $\frac{1}{4}$.

Then

$$\int_{-\infty}^{\infty} P(d_i|L)dL = 1$$

Therefore, the constant $c = 1$

$$\begin{aligned} P(L|d_i) &= P(d_i|L) \\ &= \frac{3}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(L-d_i)^2}{2\sigma^2}\right) + \frac{1}{4} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(L-(d_i-1))^2}{2\sigma^2}\right) \\ &= \frac{15}{2} \frac{1}{\sqrt{2\pi}} \exp(-50(L-d_i)^2) + \frac{3}{2} \frac{1}{\sqrt{2\pi}} \exp(-50(L-(d_i-1))^2) \end{aligned}$$



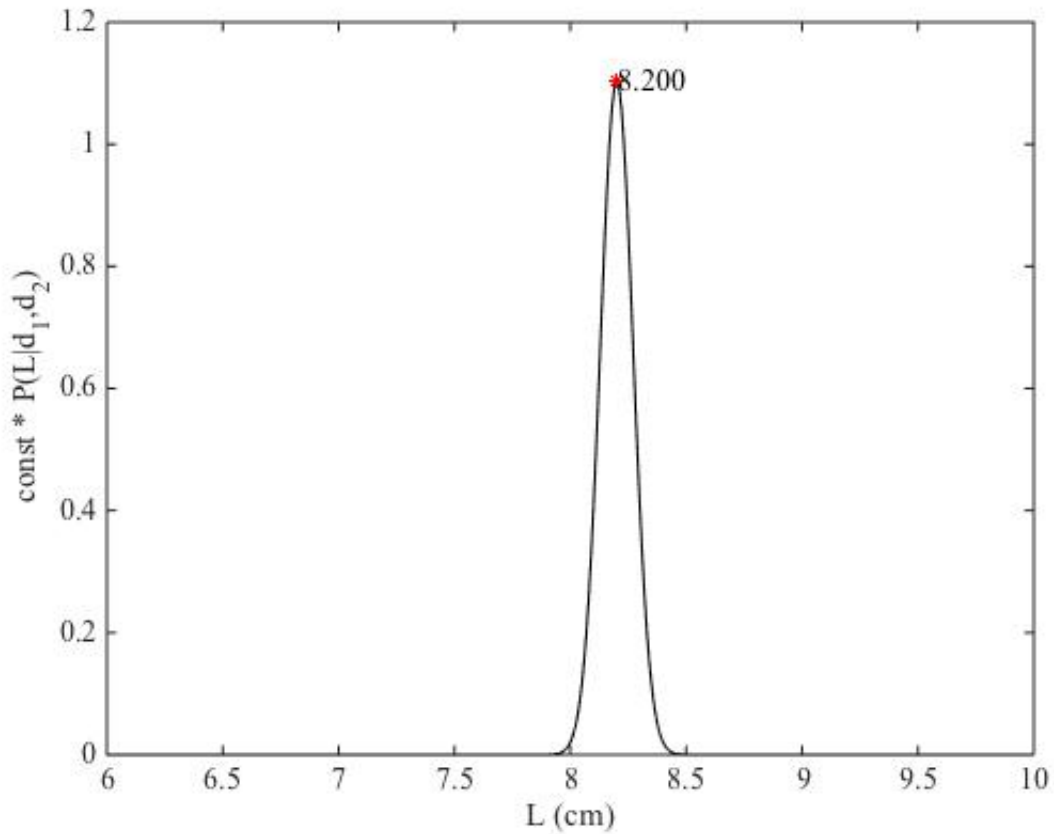
(d.iii) 10 points

We can use the distribution of L from the first measurement as prior distribution to compute a posterior distribution after the second measurement

$$P(L|d_1 = 8.3, d_2 = 9.1) \propto P(d_2 = 9.1|L)P(L|d_1 = 8.3)$$

Since we only care about the maximum of the LHS, instead of its value; for the RHS, absorbing the $\frac{1}{4\sigma\sqrt{2\pi}}$ terms into the constant

$$P(L|d_1 = 8.3, d_2 = 9.1) \propto [3 \exp(-50(9.1 - L)^2) + \exp(-50(8.1 - L)^2)] \cdot [3 \exp(-50(L - 8.3)^2) + \exp(-50(L - 7.3)^2)]$$



From the plot, we see that the best estimate of L is 8.2.

Because the two measurement differs ≈ 1 , it's likely that in the second measurement, the one quarter chance of additional 1 cm happens. After the 1 cm correction, the second measurement should be 8.1. The mean of 8.3 and 8.1 is 8.2, which is our estimation through the analysis above.

Problem 2 (graded by Kangchen) - 50 points

(a)12 points

$$P(\mathbf{m}) \propto e^{-F(\mathbf{m})} \text{ where } F(\mathbf{m}) = \sum \frac{(d_k - g_k(\mathbf{m}))^2}{2\sigma_k^2}$$

(b)12 points

Since $d'_k = d_k/\sigma_k$, the relation between \mathbf{d}' and \mathbf{d} can be written in matrix form $\mathbf{d}' = \mathbf{W}\mathbf{d}$ where $W_{ij} = \delta_{ij} \frac{1}{\sigma_j}$ ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$).

(c)12 points

$$F = \frac{1}{2}(\mathbf{d}' - \mathbf{g}'(\mathbf{m}))^T(\mathbf{d}' - \mathbf{g}'(\mathbf{m}))$$

the gradient: $\nabla F = \hat{\mathbf{G}}'^T(\mathbf{d}' - \mathbf{g}'(\mathbf{m}))$

the approximated hessian: $\mathbf{H} = (\hat{\mathbf{G}}'^T \hat{\mathbf{G}}')$

So the least squares solution: $\Delta \mathbf{m} = (\hat{\mathbf{G}}'^T \hat{\mathbf{G}}')^{-1} \hat{\mathbf{G}}'^T(\mathbf{d}' - \mathbf{g}'(\mathbf{m}))$

Substitute $\hat{\mathbf{G}}' = \mathbf{W}\hat{\mathbf{G}}$, $\mathbf{d}' = \mathbf{W}\mathbf{d}$, $\mathbf{g}' = \mathbf{W}\mathbf{g}$, $\mathbf{C} = \mathbf{W}^T \mathbf{W}$

$$\Delta \mathbf{m} = (\hat{\mathbf{G}}^T \mathbf{W}^T \mathbf{W} \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T \mathbf{W}^T \mathbf{W}(\mathbf{d} - \mathbf{g}(\mathbf{m})) = (\hat{\mathbf{G}}^T \mathbf{C} \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T \mathbf{C}(\mathbf{d} - \mathbf{g}(\mathbf{m}))$$

(d)14 points

```

1 function [ Grad, Hess] = compute_gradient_approx_hess( x,y,M,residue,weight)
2
3 W=diag(weight);
4
5 C=W'*W;
6
7 xs = M(1);
8 ys = M(2);
9 zs = M(3);
10 p = M(4);
11
12 R = ((x - xs).^2 + (y - ys).^2 + zs.^2);
13
14 dx = x-xs;
15 dy = y-ys;
16
17 Ghat(:,1) = (3.*p.*zs.*(dx))./((R).^(5/2));
18 Ghat(:,2) = (3.*p.*zs.*(dy))./((R).^(5/2));
19 Ghat(:,3) = p./(R).^(3/2) - (3*p.*zs.^2)./(R).^(5/2);
20 Ghat(:,4) = zs./(R).^(3/2);
21
22 Grad = (residue')*C*Ghat;
23
24
25
26 Hess = (Ghat')*C*Ghat;
27 %this is the approximated Hessian;
28
29
30
31 end

1
2 function [M]=nonlinear_solver(x,y,d,Minit,w)
3
4 M=Minit;
5 r = 0;
6 r_old = 0;
7
8 for ii = 1:1:1000
9 r_old = r;
10 r=compute_residue(x,y,M,d);
11
12
13 [Grad,Hess]=compute_gradient_approx_hess(x,y,M,r,w);

```

```

14 deltaM= (Hess)\Grad';
15
16 M=M+deltaM;
17
18
19 if (norm(r-r_old)<1e-7)
20     break;
21 end
22 end
23 disp(r)
24 disp(ii)
25 end

1 function [ r ] = compute_residue( x,y,M,d )
2
3 xs = M(1);
4 ys = M(2);
5 zs = M(3);
6 p = M(4);
7 r =d- p*zs./((x - xs).^2 + (y - ys).^2 + zs^2).^(3/2);
8 end

1 %%problem 1d
2 x = [0 11 15 6 -7 3]';
3 y = [0 0 6 13 10 -7]';
4 d = [0.103 0.162 0.065 0.036 0.025 0.169]';
5 M0 = [8 -5 10 30]'; %initial guess
6 Ms = nonlinear_solver(x,y,d,M0,[1,1,0.2,1,2.5,2.5]);

```

$$\mathbf{m} = [8.3068, -5.3425, 11.8179, 31.8569]^T$$

$$\mathbf{error} = [-4.99 \times 10^{-5}, 1.20 \times 10^{-5} - 2.95 \times 10^{-3}, 3.59 \times 10^{-4}, -2.87 \times 10^{-5}, 2.74 \times 10^{-6}]$$

We can find that since we put a smaller weight on station 3, its error is the largest. Since we put a larger weight on station 5 6 , their errors are smaller.

(Extra Credit) Problem 3 (graded by Yiran) - 25 points**(a) 5 points**

Since \mathbf{g}_i has M elements, the maximum dimension of \mathbf{G} spanned by $\{\mathbf{g}_i, i = 1 \dots N\}$ is M. Therefore,

$$\dim(R(\mathbf{G})) \leq M < N = \dim(\mathbb{R}^N)$$

(b) 10 points

$$\mathbf{H}_{ij} = \mathbf{g}_i^T \mathbf{g}_j$$

The diagonal elements of \mathbf{H} are the squared lengths of the column vectors of \mathbf{G} , the off-diagonal elements measure how much the column vectors of \mathbf{G} project onto each other. Suppose we project a vector \mathbf{d} into the column space of \mathbf{G} , the coordinates are (m_1, \dots, m_M) . If the length of certain column vector \mathbf{g}_i is small, an error in \mathbf{d} will cause a big error in m_i . The error in \mathbf{d} also tends to affect those m_i similarly if their corresponding column vectors \mathbf{g}_i are near parallel.

(c) 5 points

$$\mathbf{G}^T \mathbf{d} = (\mathbf{g}_1^T \mathbf{d}, \dots, \mathbf{g}_M^T \mathbf{d})^T$$

which is the projection of \mathbf{d} into the model space.

(d) 5 points

This equation implies that $\mathbf{G}\mathbf{m}$ equals the projection of \mathbf{d} in the model space. Thus, the least-squares solution \mathbf{m} is the coordinate of \mathbf{d} in the model space.