CS215 Assignments

Data Analysis and Interpretation

Paarth Jain
190050076

Sahasra Ranjan
190050102

Sibasis Nayak
190050115

September 28, 2020

Assignment 3

1. Solution at the end in handwritten form. In part (c) I have used X_i as the geometric r.v instead of generalised Z. Name of the variable does not matter in context to that (c) part, but for just being rigorous X_i in that part can be considered as Z. The result i.e the expected value and variance of the geometric value remain unchanged though.

2. (a) Distributon function F is reperesented as $F(x) = P(X \leq x)$. Now we know that F is an increasing function for any distribution. In this case however we are given that F is invertible which implies that F(x) is strictly invertible. i.e.

$$F(x_1) > F(x_2) \iff x_1 > x_2$$

Further, by substituting $y1 = F(x_1)$ and $y_2 = F(x_2)$, and using the fact that $F^{-1}(x)$ exists, we can rewrite the above equation as,

$$y_1 > y_2 \iff F^{-1}(y_1) > F^{-1}(y_2)$$

This means that F^{-1} is also an increasing function. Now, let u_i be samples of a random variable U such that $U \sim U(0,1)$ (uniform distribution) (Given in question), and let $v_i = F^{-1}(u_i)$ be the samples of a random variable V whose distribution is to be found.Now,

$$P(U \leq x) = x \qquad \text{(by defn. uniform random variable)}$$

now as F^{-1} is increasing $U \leq x \iff F^{-1}(U) \leq F^{-1}(x)$. Now using this result and substituting $F^{-1}(x) = y$, we have-

$$P(F^{-1}(U) \leq F^{-1}(x)) = x$$
$$P(V \leq F^{-1}(x)) = x$$
$$P(V \leq y) = F(y)$$

Thus V (or equivalently v_i) follow the distribution F.

(b) We know that F is an increasing function (cumulative never decreases) i.e. $x_1 \leq x_2 \iff F(x_1) \leq F(x_2)$

$$P(D \geq d) = P\{max_x|\frac{\sum_i \mathbf{1}(Y_i \leq x)}{n} - F(x)| \geq d\}$$

$$= P\{max_x|\frac{\sum_i \mathbf{1}(F(Y_i) \leq F(x))}{n} - F(x)| \geq d\}$$

$$= P\{max_x|\frac{\sum_i \mathbf{1}(U_i \leq F(x))}{n} - F(x)| \geq d\}$$

substituting $y = F(x)$

$$P(D \geq d) = P\{max_{0 \leq y \leq 1}|\frac{\sum_i \mathbf{1}(U_i \leq y)}{n} - y| \geq d\}$$

$$= P(E \geq d)$$

The random variable $max_x|F_e(x) - F(x)|$, in a way represents the error or deviation from the actual distribution during sampling, or max deviation of samples from the actual distribution.

Now, this error/deviation is same(in distribution) for all distributions as each of them are equal (in distribution; to the one where Y = U(0,1)). Thus by transitivity (of sorts) all are equal(in distribution) to each other). Thus the error in sampling any distribution through uniform is same as error in uniform distribution. That is, no error is added while transforming uniform sample to sample of required distribution.
This result reinforces the correctness of the inverse sampling method using uniform distributions.

3. Solution at the end in handwritten form.

4. (a) Disjoint sets T and V made by taking first 750 elements(T) of Sample and last 250 elements(V) of sample.

   (b) let $\{t_i\}_{i=1}^{750}$ be elements of the set T. Then the pdf estimate can be written as-

$$\hat{p}(x;\sigma) = \frac{\sum_{i=1}^{n} exp(-(x-t_i)^2/2\sigma^2)}{n\sigma\sqrt{2\pi}}$$

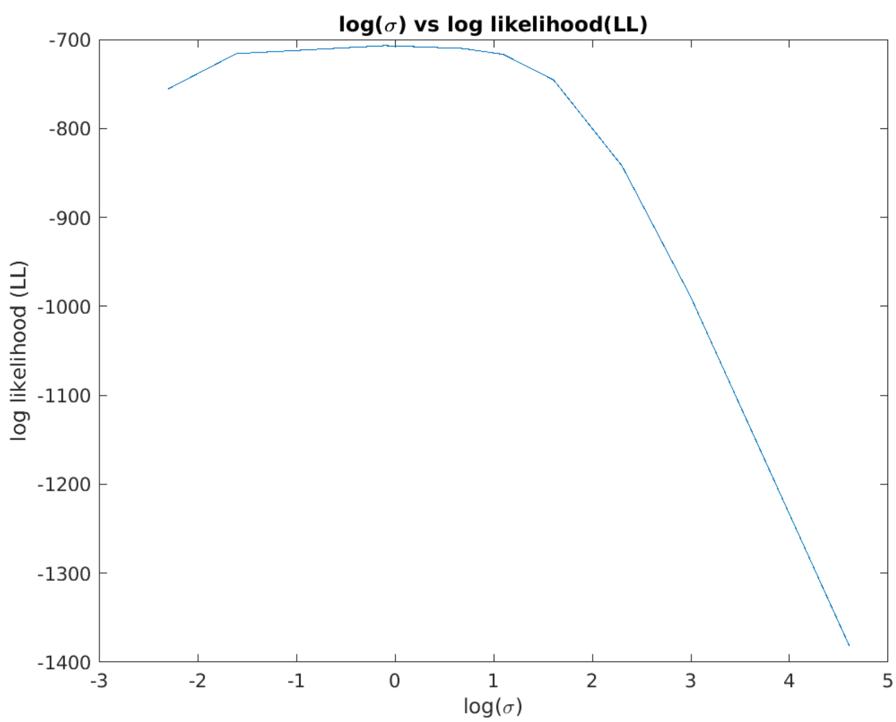
   Thus, for each $v_i \in V$

$$\hat{p}(v_i;\sigma) = \frac{\sum_{i=1}^{n} exp(-(v_i-t_i)^2/2\sigma^2)}{n\sigma\sqrt{2\pi}}$$

   Now, as v_i are independent of each other, the joint pdf

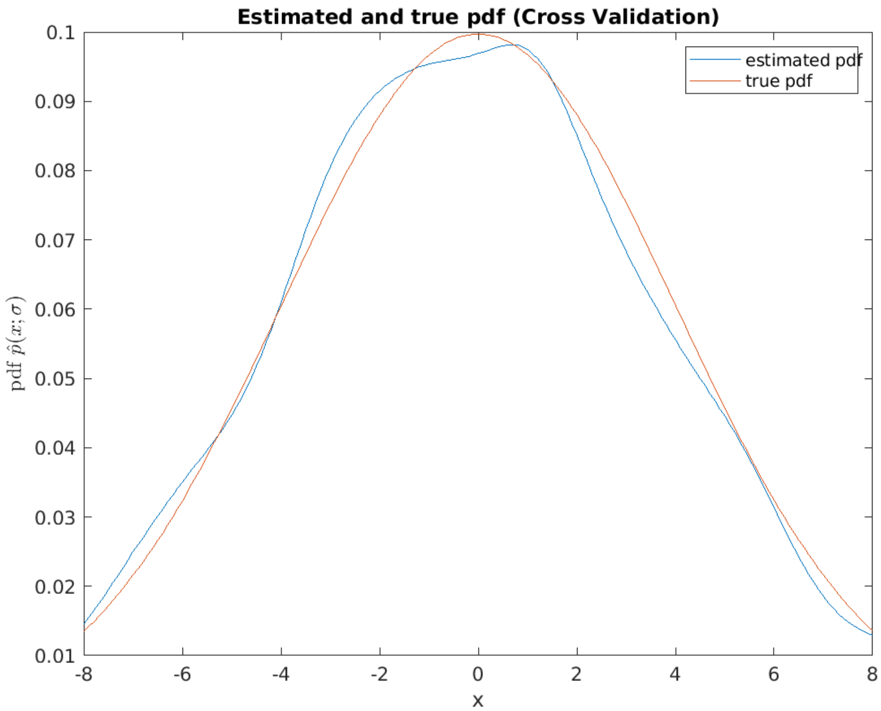
$$\hat{p}(v_1, v_2, ..., v_n;\sigma) = \prod_{i=1}^{n} \hat{p}(v_i;\sigma)$$

   (c) Plotting $log(\sigma)$ vs log likelihood gives-

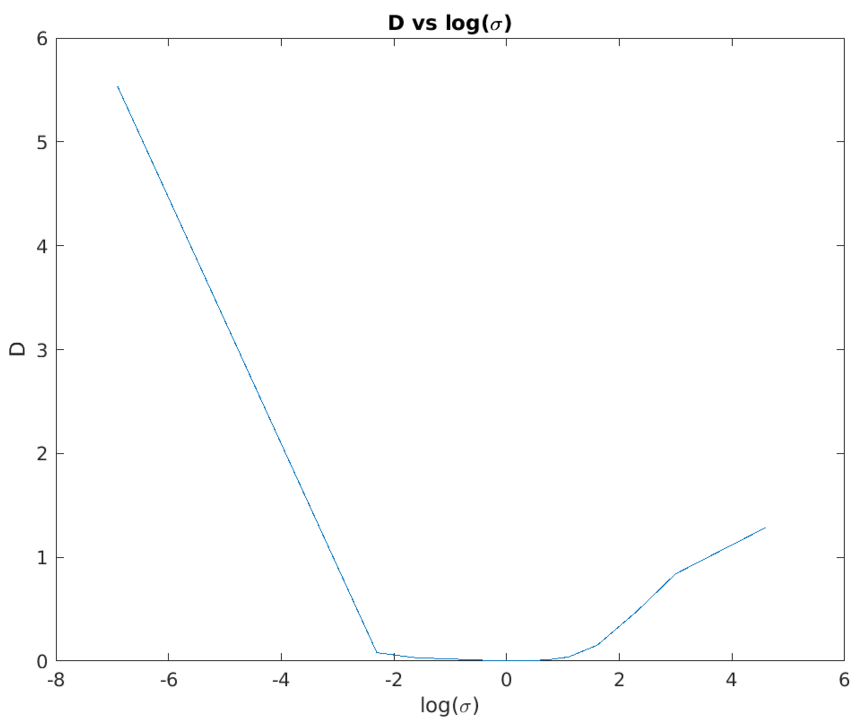


   In this case the σ for which log likelihood is maximum comes out to be 0.9, but in general the best sigma assumes 0.9 or 1 value, occasionally deviations are seen (in different runs of the program).(Seeding the program can remove that discrepancy but the question does not ask us to do so.

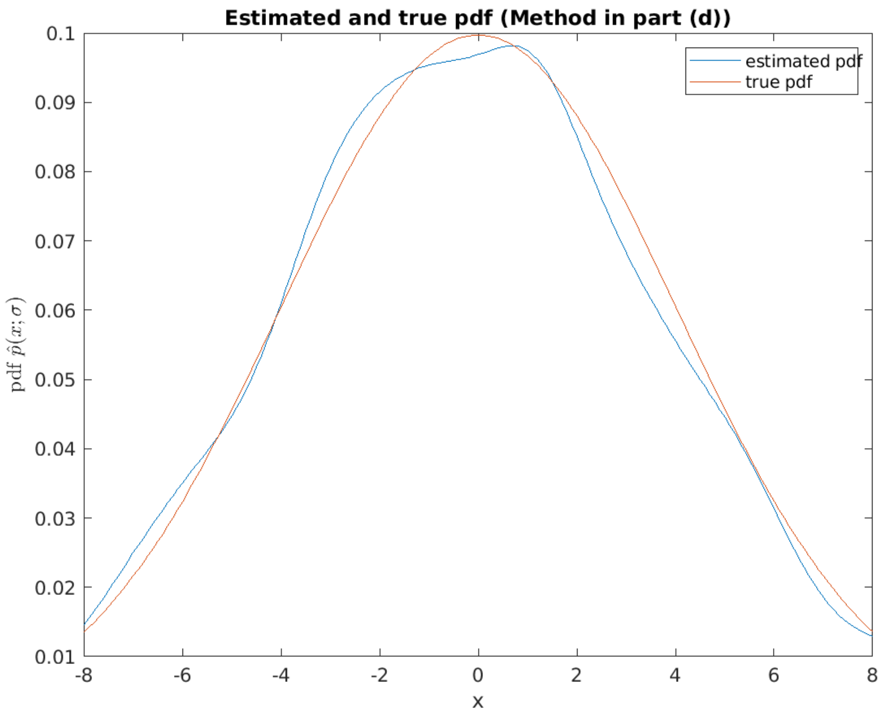
Now plotting the estimated pdf and overlapping it with true pdf.



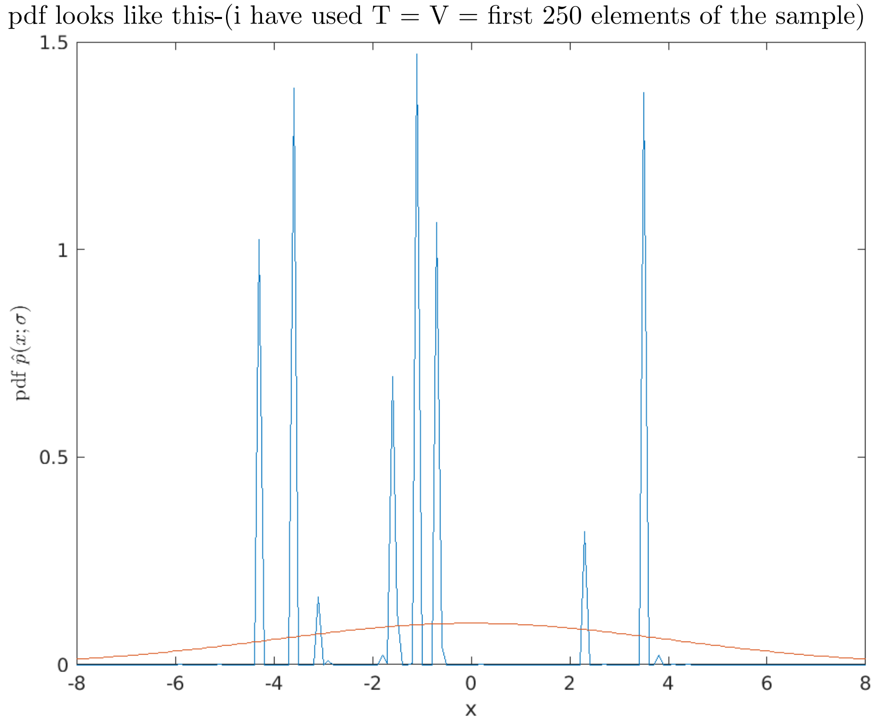
(d) Plotting $log(\sigma)$ vs D gives-



In this case the best sigma (for which D is minimum) comes out to be 1, slightly deviating from the best sigma found in part (c).
D corresponding to this best sigma = 0.0021
D corresponding to best sigma of part c = 0.0024

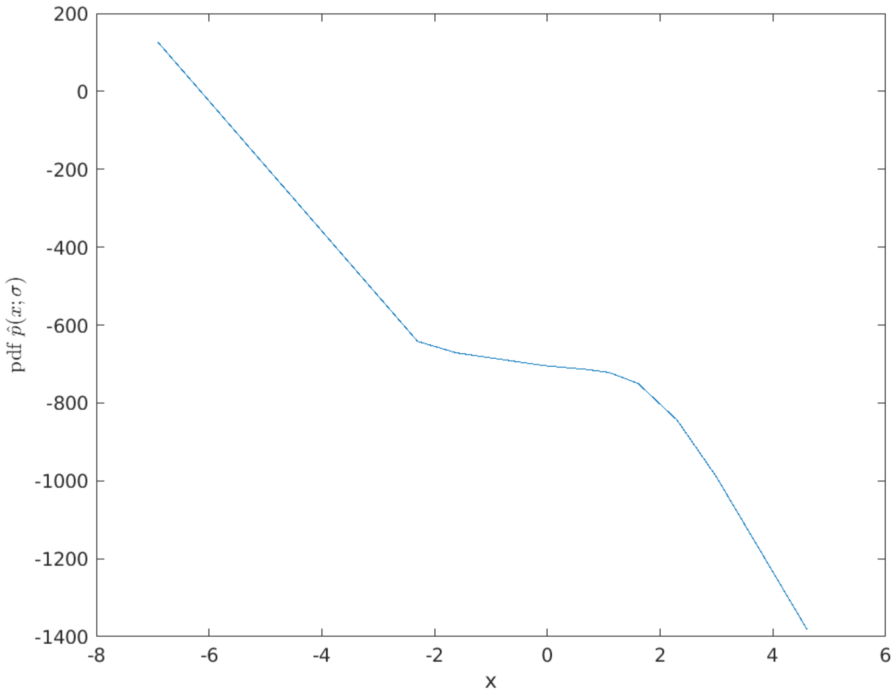


(e) When we take $T = V$, (and use the same joint likelihood function), the CV method, gives $\sigma = 0.001$. And the squared error (D value) for it is very high. The estimated pdf when overlapped with true pdf looks like this-(i have used T = V = first 250 elements of the sample)



Thus the pdf estimation is highly incorrect here.

If we look at the $log(\sigma)$ vs log likelihood graph. Smaller values of sigma, seem to be giving higher log likelihood values.



Here is an explanation to the above observations.
When T equals V. the joint likelihood function looks like-

$$\hat{p}(t_1, t_2, ..., t_n; \sigma) = \prod_{i=1}^{n} \hat{t}(t_i; \sigma)$$

$$\hat{p}(t_1, t_2, ..., t_n; \sigma) = \prod_{i=1}^{n} (\frac{\sum_{j=1}^{n} exp(-(t_i - t_j)^2/2\sigma^2)}{n\sigma\sqrt{2\pi}})$$

Now, for each i (each $\hat{p}(t_i; \sigma)$) there is a term where $t_i = t_j$, thus that term becomes-

$$\frac{1}{n\sigma\sqrt{2\pi}}$$

For large sigma this term is not very big, but for sigma as small as 0.001, this term becomes huge compared to others. Now this comparison of big and small is relative and has a lot of factors in it.

First lets consider the denominator of each term, it is large for large σ and small for small σ.

Second, the numerator is an exponential term. Lets consider the power. (or rather its negative). it is $(t_i - t_j)^2/2\sigma^2$. Now the "behaviour" of difference in values of $t_i and t_j$ remains same in all σ. But we need to keep in mind that the difference varies in accordance with the original pdf. As its std. deviation is 4, the difference can reach order 10 (1-10) mostly. So σ values which are too less, (0.1, 0.001) make the value of the power term higher. Adding the negative sign, results in the exponent term being very small overall.Overshadowing the effect of small denominator.

Thus compared to different V and T, this term where $t_i = t_j$, adds large value to each estimate and the log of their product (the log likelihood) becomes large. and it is chosen as the best sigma.
The correct method to do it when T = V, should be to omit the $t_i = t_j$ term, i.e. the estimate for each t_i is calculated from n-1 variables that are **independent** of t_i. This was verified through a MATLAB program and the results where similar to when $T \cap V = \phi$.

Instructions to run code:

- For q4, run the file q4.m in matlab. The program will print out the required values(these vary with each run as seed is not provided (the question does not ask us to do so)). Also four plots each in different figures are printed out.
- For q1 and q3 run the code in matlab to generate the graph and generate the solutions. q3 prints a column vector corresponding to a,b and c and the noise varince term.

(1)

(a) As we don't have any book already, any book we pick will be a ~~different~~ unique colour, thus only 1 book needs to be picked.

So $X_1 = 1$

When $i-1$ books picked, probability of picking a book with different colour $=$ ~~Probability~~ $\dfrac{\text{No. of unpicked books}}{\text{Total no. of books}}$

$$= \frac{n-(i-1)}{n}$$

$$= \frac{n-i+1}{n}$$

(b) X_i is a geometric random variable, by definition p (parameter) should be the probability of picking a book with a color that has not been picked.

$$\therefore \ p = \frac{n-i+1}{n} \quad \{\text{parameter for } X_i\}$$

(c) $P(X_i = k) = (1-p)^{k-1} p$ {Let X_i be the geometric random variable}

$$E(X_i) = \sum_{k=0}^{\infty} k \, P(X_i = k)$$

$$= \sum_{k=1}^{\infty} k p (1-p)^{k-1} \quad \text{—— (I)}$$

Now multiplying both sides by $1-p$ we get,

$$(1-p) \, E(X_i) = \sum_{k=1}^{\infty} k p (1-p)^{k} \quad \text{—— (II)}$$

Eq^n ① is

$$E(X_i) = p(1-p)^0 + 2p(1-p)^1 + 3p(1-p)^2 + \cdots \infty$$

Eq^n ⑪ is $(1-p)E(X_i) = \qquad p(1-p)^1 + 2p(1-p)^2 + \cdots \infty$

Subtracting the above 2 eq^ns,

$$(1-(1-p))E(X_i) = p(1-p)^0 + p(1-p) + p(1-p)^2 + \cdots \infty$$

$$\Rightarrow \quad pE(X_i) = p\{1 + (1-p) + (1-p)^2 + \cdots \infty\}$$

RHS has an infinite G.P sum with ratio $(1-p)$

$$\therefore \cancel{p}E(X_i) = \cancel{p} \times \frac{1}{1-(1-p)}$$

$$\Rightarrow \quad E(X_i) = 1/p$$

Now,

$$E(X_i^2) = \sum_{K=1}^{\infty} K^2 P(X_i = K)$$

$$= \sum_{K=1}^{\infty} K^2 p(1-p)^{K-1}$$

Now,

$$E(X_i^2) = p(1-p)^0 + 2^2 p(1-p) + \cdots + K^2 p(1-p)^{K-1} + \cdots \infty$$

$$(1-p)E(X_i^2) = \qquad + 1^2 p(1-p) + \cdots + (K-1)^2 p(1-p)^{K-1} + \cdots \infty$$

Subtracting above 2 eq^ns,

$$pE(X_i^2) = p + (2^2-1^2)p(1-p) + (3^2-2^2)(1-p)^2 p + \cdots + (K^2-(K-1)^2)p(1-p)^{K-1} + \cdots \infty$$

$$\Rightarrow pE(X_i^2) = p + (2\times 2-1)p(1-p) + (2\times 3-1)p(1-p)^2 + \cdots + (2K-1)p(1-p)^{K-1} + \cdots \infty$$

$\Rightarrow p\,E(x_i^2) = (2\times1-1)p + (2\times2-1)p(1-p) + \dots + (2k-1)p(1-p)^{k-1} + \dots\infty$

$\Rightarrow (1-p)\,p\,E(x_i^2) = \qquad (2\times1-1)p(1-p) + \dots + (2k-3)p(1-p)^{k-1} + \dots\infty$

$\Rightarrow$ Subtracting above 2 eqns,

$$p^2 E(x_i^2) = p + 2p(1-p) + 2p(1-p)^2 + \dots + 2p(1-p)^{k-1} + \dots + \infty$$

$$\Rightarrow p^2 E(x_i^2) = \{ p + 2p((1-p) + (1-p)^2 + \dots + \infty) \}$$

↳ A infinite G.P sum with ratio 1-p and first term (1-p).

$$\Rightarrow p^2 E(x_i^2) = p + \frac{2p \times (1-p)}{(1-(1-p))}$$

$$\Rightarrow p^2 E(x_i^2) = p + \frac{2\cancel{p}(1-p)}{\cancel{p}}$$

$$= p + 2 - 2p$$

$$\Rightarrow E(x_i^2) = \frac{2-p}{p^2}$$

$$Var(x_i) = E(x_i^2) - (E(x_i))^2$$

$$= \frac{2-p}{p^2} - 1/p^2$$

$$= \frac{1-p}{p^2}$$

d) $E(X^{(n)}) = E(X_1 + X_2 + \dots + X_n)$

$= \sum_{i=1}^{n} E(X_i)$ { since X_is are independent R.V }

$= \sum_{i=1}^{n} 1/p_i$

From (b) we know $p_i = \frac{n-i+1}{n}$

$\Rightarrow E(X^{(n)}) = \sum_{i=1}^{n} \frac{n}{n-i+1}$

$= n\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}\right)$

The above cannot be simplified to a closed form, but for large enough n or $1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \log n$

Sorry

e) $Var(X^{(n)}) = \sum_{i=1}^{n} Var(X_i)$ { $\because X_i$s are independent }

$= \sum_{i=1}^{n} \frac{1 - \frac{n-i+1}{n}}{\left(\frac{n-i+1}{n}\right)^2}$

$= \sum_{i=1}^{n} \frac{\frac{i-1}{n}}{\frac{(n-i+1)^2}{n^2}}$

$= \sum_{i=1}^{n} \frac{n(i-1)}{(n-i+1)^2}$

$Var(X^{(n)}) = \sum_{i=1}^{n} \frac{n(i-1)}{(n-i+1)^2} < \sum_{i=1}^{n} \frac{n^2}{(n-i+1)^2}$

$$\operatorname{Var}(X^{(n)}) < \sum_{i=1}^{n} \frac{n^2}{(n-i+1)^2}$$

$$< n^2 \left\{ \frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{n^2} \right\}$$

$$< n^2 \left\{ \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots \infty \right\}$$

$$< \frac{n^2 \pi^2}{6}$$

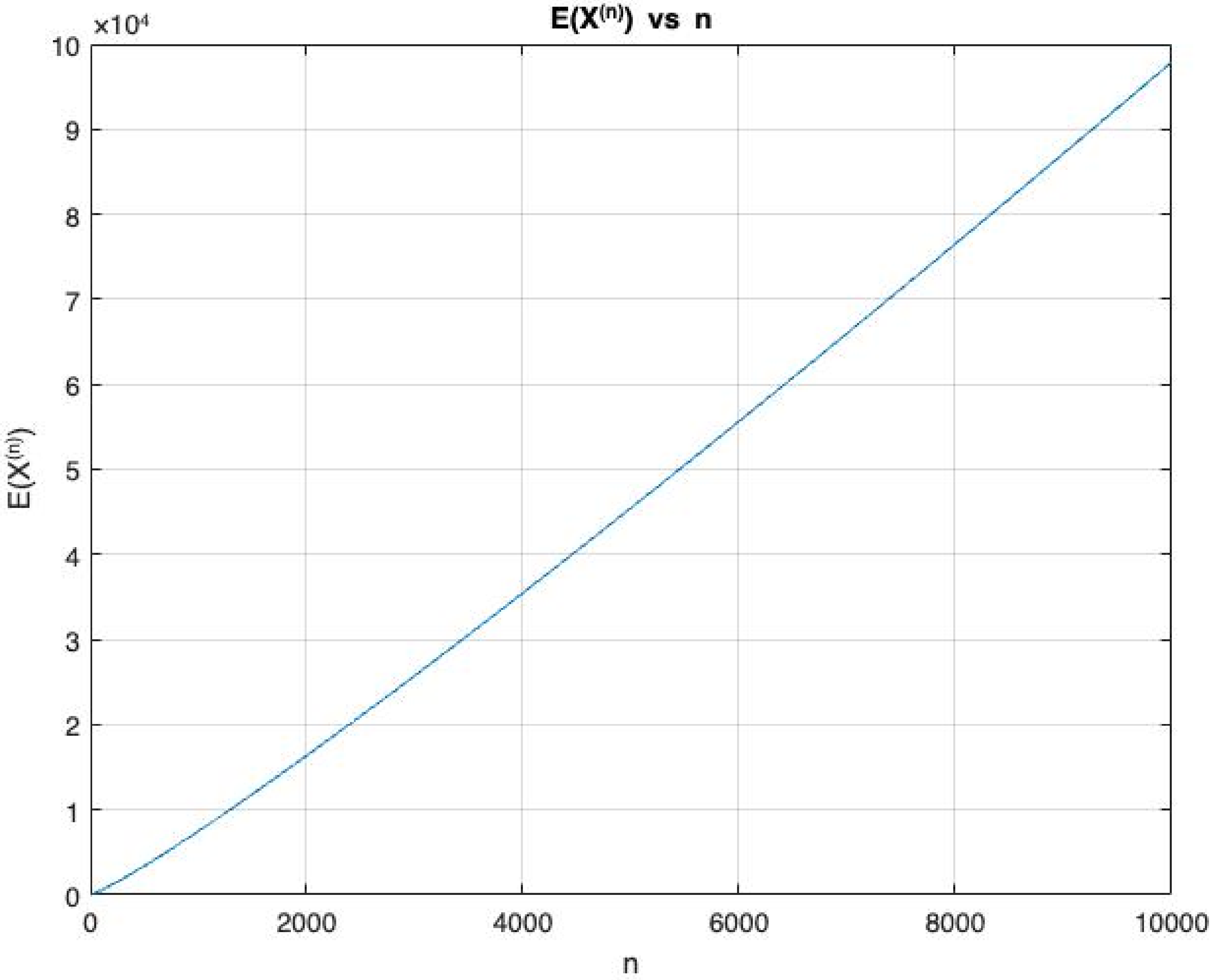
(f) $E(X^{(n)}) = n(1 + 1/2 + 1/3 + \cdots + 1/n)$

for large n, $1 + 1/2 + 1/3 + \cdots + 1/n$ approximates to ~~1/n~~ log
{by approximating it to $\int_1^n \frac{1}{x}\,dx$ }

So $E(X^{(n)})$ is bounded by $n \log n$.

Therefore if $E(X^{(n)}) = \Theta(f(n))$,

$f(n) = n \log n$



E(X^(n)) vs n
×10^4
10
9
8
7
6
5
4
3
2
1
0
0
2000
4000
6000
8000
10000
n
E(X^(n))

3) a) We have n such (x_i, y_i, z_i) such that

$\forall i \quad z_i = ax_i + by_i + c + \varepsilon_i$ {where a, b, c are params}

and $\varepsilon_i \sim N(0, \sigma^2)$

ε_is are drawn independent random variables. {Noise}

we know $\{x_i\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$ and noisy $\{z_i\}_{i=1}^{n}$

We have to determine a, b, c.

Now z_i can be represents as a ~~normal~~ gaussian distribution s.t.,

$$z_i \sim N(ax_i + by_i + c, \sigma^2)$$

The joint distribution is the product of distributions, since z_is are independent.

Let p be the joint distribution,

~~$p(z_i; x_i, a, b, c, y_i) =$~~

$$p(\{z_i\}; \{x_i\}, \{y_i\}, a, b, c) = \prod_{i=1}^{n} N(ax_i + by_i + c, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{e^{-\frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$\log P = \sum_{i=1}^{n} -\frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2} - n\log\sigma - n\log\sqrt{2\pi}$$

~~Diffe~~

Differentiating the above w.r.t. a and setting to 0 we get,

$$\frac{\partial \log P}{\partial a} = \sum_{i=1}^{n} \frac{2x_i(z_i - ax_i - by_i - c)}{2\sigma^2} = 0$$

$$\Rightarrow a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i z_i \quad \text{— (I)}$$

Differentiating P w.r.t. b and setting to 0 we get,

$$\frac{\partial \log P}{\partial b} = \sum_{i=1}^{n} \frac{2y_i(z_i - ax_i - by_i - c)}{2\sigma^2} = 0$$

$$\Rightarrow a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i^2 + c\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i z_i \quad \text{— (II)}$$

Differentiating P w.r.t. c and setting it to 0 we get,

$$\frac{\partial \log P}{\partial c} = \sum_{i=1}^{n} \frac{2(z_i - ax_i - by_i - c)}{2\sigma^2} = 0$$

$$\Rightarrow a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i + cn = \sum_{i=1}^{n} z_i \quad \text{— (III)}$$

(I), (II), (III) are the 3 required linear equations. These equations can be represented by following:-

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} y_i^2 & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i z_i \\ \sum_{i=1}^{n} y_i z_i \\ \sum_{i=1}^{n} z_i \end{bmatrix}$$

(b) Now we have for ~~ino~~ n values,

$$\forall i \quad z_i = a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6 + \epsilon_i$$

{where $a_1, a_2, a_3, a_4, a_5, a_6$ are params and ϵ_is are normally distributed noise}

Now $\quad z_i \sim N\left(a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6, \sigma^2\right)$

The joint distribution will be the product of distributions since z_is are independent.

$$P\left\{ \{z_i\}; \{x_i\}, \{y_i\}, a_1, a_2, a_3, a_4, a_5, a_6 \right\} = \prod_{i=1}^{n} N\left(a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6, \sigma^2\right)$$

~~lo~~ $\quad \cancel{= \prod_{i=1}^{n} e^{-(a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6}}$

$$= \prod_{i=1}^{n} \frac{e^{-(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)^2 / 2\sigma^2}}{\sqrt{2\pi}\,\sigma}$$

$$\log P = \sum_{i=1}^{n} \frac{-(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)^2}{2\sigma^2} - n \log(\sqrt{2\pi}\,\sigma)$$

Differentiating P w.r.t. a_1 and setting it to 0,

$$\frac{\partial \log P}{\partial a_1} = + \sum_{i=1}^{n} \frac{2 x_i^2 (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{2\sigma^2} = 0$$

$$\Rightarrow a_1 \sum_{i=1}^{n} x_i^4 + a_2 \sum_{i=1}^{n} x_i^2 y_i^2 + a_3 \sum_{i=1}^{n} x_i^3 y_i + a_4 \sum_{i=1}^{n} x_i^3 + a_5 \sum_{i=1}^{n} x_i^2 y_i + a_6 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 z_i \quad \text{—(1)}$$

Diff. ~~wrt~~ P wrt a_2 and setting it to 0,

$$\frac{\partial \log P}{\partial a_2} = + \sum_{i=1}^{n} \frac{2 y_i^2 (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{2\sigma^2} = 0$$

$$\Rightarrow a_1 \Sigma x_i^2 y_i^2 + a_2 \Sigma y_i^4 + a_3 \Sigma x_i y_i^3 + a_4 \Sigma x_i y_i^2 + a_5 \Sigma y_i^3 + a_6 \Sigma y_i^2 = \Sigma z_i y_i^2 \quad \text{— (II)}$$

Diff P wrt a_3 and setting it to 0,

$$\frac{\partial \log P}{\partial a_3} = \sum_{i=1}^{n} \frac{2 x_i y_i (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{2\sigma^2} = 0$$

$$\Rightarrow a_1 \Sigma x_i^3 y_i + a_2 \Sigma x_i y_i^3 + a_3 \Sigma x_i^2 y_i^2 + a_4 \Sigma x_i^2 y_i + a_5 \Sigma x_i y_i^2 + a_6 \Sigma x_i y_i = \Sigma x_i y_i z_i \quad \text{— (III)}$$

Diff P wrt a_4 and setting it to 0,

$$\frac{\partial \log P}{\partial a_4} = \sum_{i=1}^{n} \frac{2 x_i (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{2\sigma^2} = 0$$

$$\Rightarrow a_1 \Sigma x_i^3 + a_2 \Sigma x_i y_i^2 + a_3 \Sigma x_i^2 y_i + a_4 \Sigma x_i^2 + a_5 \Sigma x_i y_i + a_6 \Sigma x_i = \Sigma x_i z_i \quad \text{— (IV)}$$

Diff wrt a_5 and setting it to 0,

$$\frac{\partial \log P}{\partial a_5} = \sum_{i=1}^{n} \frac{2 y_i (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{2\sigma^2} = 0$$

$$\Rightarrow a_1 \Sigma x_i^2 y_i + a_2 \Sigma y_i^3 + a_3 \Sigma x_i y_i^2 + a_4 \Sigma x_i y_i + a_5 \Sigma y_i^2 + a_6 \Sigma y_i = \Sigma y_i z_i \quad \text{— (V)}$$

Diff wrt. a_6 and setting it to 0,

$$\frac{\partial \log P}{\partial a_6} = \sum_{i=1}^{n} \frac{2(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{2\sigma^2} = 0$$

$\Rightarrow \cancel{a_1 x_i^2 + a_2}$

$$\Rightarrow a_1 \Sigma x_i^2 + a_2 \Sigma y_i^2 + a_3 \Sigma x_i y_i + a_4 \Sigma x_i + a_5 \Sigma y_i + n a_6 = \Sigma z_i \quad \text{--- 6 (VI)}$$

Eqns (I), (II), (III), (IV), (V) and (VI) are the required linear equations. Represented as follows in M.V form -

$$\begin{bmatrix} \Sigma x_i^4 & \Sigma x_i^2 y_i^2 & \Sigma x_i^3 y_i & \Sigma x_i^3 & \Sigma x_i^2 y_i & \Sigma x_i^2 \\ \Sigma x_i^2 y_i^2 & \Sigma y_i^4 & \Sigma x_i y_i^3 & \cancel{\Sigma y_i^3}\ \Sigma x_i y_i^2 & \Sigma y_i^3 & \Sigma y_i^2 \\ \Sigma x_i^3 y_i & \Sigma x_i y_i^3 & \Sigma x_i^2 y_i^2 & \Sigma x_i^2 y_i & \Sigma x_i y_i^2 & \Sigma x_i y_i \\ \Sigma x_i^3 & \Sigma x_i y_i^2 & \Sigma x_i^2 y_i & \Sigma x_i^2 & \Sigma x_i y_i & \Sigma x_i \\ \Sigma x_i^2 y_i & \Sigma y_i^3 & \Sigma x_i y_i^2 & \Sigma x_i y_i & \Sigma y_i^2 & \Sigma y_i \\ \Sigma x_i^2 & \Sigma y_i^2 & \Sigma x_i y_i & \Sigma x_i & \Sigma y_i & n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma x_i^2 z_i \\ \Sigma z_i y_i^2 \\ \Sigma x_i y_i z_i \\ \Sigma x_i z_i \\ \Sigma y_i z_i \\ \Sigma z_i \end{bmatrix}$$

c) The equation of plane predicted from MATLAB code

$$z = 10.0022x + 19.9980y + 29.9516$$

The noise variance calculated is 23.0685