

STAT 542: Homework 1

Sharvi Tomar (stomar2)

Due: Tuesday 11:59 PM CT, Feb 2nd

Contents

Question 1: Basic calculus	1
Question 2: Linear algebra	2
Question 3: Statistics	3
Question 4: Programming	4

Question 1: Basic calculus

1. Calculate the derivative of $f(x)$

(a) $f(x) = e^x$

Solution:

$$f'(x) = e^x$$

(b) $f(x) = \log(1 + x)$

Solution:

$$f'(x) = \frac{1}{\log(1 + x)}$$

(c) $f(x) = \log(1 + e^x)$

Solution:

$$f'(x) = \frac{e^x}{\log(1 + e^x)}$$

2. Taylor expansion. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a twice differentiable function. Please write down the first three terms of its Taylor expansion at point $x = 1$.

Solution:

$$f(x) = f(1) + f'(1)(x - 1) + \frac{f''(1)}{2!}(x - 1)^2$$

3. For the infinite sum $\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}$, where α is a positive real number, give the exact range of α such that the series converges.

Solution: Consider the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ with $p > 0$. Then, the Cauchy condensation test implies that $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if

$$\sum_{n=1}^{\infty} \frac{2^n}{2^{np}} = \sum_{n=1}^{\infty} \left(\frac{1}{2^{p-1}}\right)^n \text{ converges.}$$

We already know that the geometric series $\sum_{n=1}^{\infty} \left(\frac{1}{2^{p-1}}\right)^n$ converges if and only if $\frac{1}{2^{p-1}} < 1$ i.e. $p > 1$

Therefore, $\alpha > 1$ for the given series to converge.

Question 2: Linear algebra

1. What is the eigen-decomposition of a real symmetric matrix $A_{n \times n}$? Write down one form of that decomposition and explain each term in your formula. Based on these terms, suppose all eigenvalues are positive, derive $A^{-1/2}$.

Solution: Every real symmetric $n \times n$ matrix A can be factored as:

$$A = QDQ^T$$

where:

Q is orthogonal with the columns that are an orthonormal set of n eigenvectors

D is $\text{diag}(\lambda_1, \dots, \lambda_n)$ is diagonal with real diagonal elements

A is diagonalizable by an orthogonal similarity transformation: $Q^T A Q = D$

$$A^{-1/2} = QD^{-1/2}Q^T$$

$$D^{-1/2} = \text{diag}\left(\frac{1}{\lambda_1^{1/2}}, \dots, \frac{1}{\lambda_n^{1/2}}\right)$$

2. What is a symmetric positive definite matrix $A_{n \times n}$? Give one of the equivalent definitions and explain your notation.

Solution: A is symmetric and positive definite if $x^T A x > 0$ for all $x \in \mathbb{R}^n, x \neq 0$

Equivalently,

A sufficient condition for a symmetric matrix to be positive definite is that it has positive diagonal elements and is diagonally dominant.

That is,

$$a_{ii} > \sum_{j \neq i} |a_{ij}| \quad \forall i$$

.

3. True/False. If you claim a statement is false, explain why. For two real matrices $A_{m \times n}$ and $B_{n \times m}$

(a) $\text{Rank}(A) = \min\{m, n\}$

Solution: False

The rank of a matrix is defined as (a) the maximum number of linearly independent column vectors in the matrix or (b) the maximum number of linearly independent row vectors in the matrix. Both definitions are equivalent.

(b) If $m = n$, then $\text{trace}(A) = \sum_{i=1}^n A_{ii}$

Solution: True

(c) If A is a symmetric matrix, then all eigenvalues of A are real

Solution: True

(d) If A is a symmetric matrix, λ_1 and λ_2 are two distinct eigen-values and v_1, v_2 are the corresponding eigenvectors, then it is possible that $v_1^T v_2 > 0$.

Solution: False

In case of a real symmetric matrix, its eigenvectors corresponding to different eigenvalues are orthogonal. Hence, $v_1^T v_2 = 0$

(e) If A is a symmetric matrix, v_1, v_2 are two distinct eigen-vectors of λ , then it is possible that $v_1^T v_2 > 0$.

Solution: True

(f) $\text{trace}(ABAB) = \text{trace}(AABB)$.

Solution: False

The multiplication on the right side $AABB$ is not valid due to incorrect dimensions for matrix multiplication.

Question 3: Statistics

1. X_1, X_2, \dots, X_n are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ random variables, where $\mu \in \mathbb{R}$ and $\sigma > 0$ is finite. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

(a) What is an unbiased estimator? Is \bar{X}_n an unbiased estimator of μ ?

Solution: An estimator is said to unbiased if $E(\hat{X}) = X \forall X$

$$E\left(\frac{\sum X_i}{n}\right) = \frac{E(\sum X_i)}{n} = \mu \text{ Since } X_i \text{ are i.i.d.}$$

Therefore, \bar{X}_n is an unbiased estimator.

(b) What is $E[(\bar{X}_n)^2]$ in terms of n, μ, σ ?

Solution: $E(X^2) = \text{Var}(X) + [E(X)]^2$

$$\text{Var}\left(\frac{\sum X_i}{n}\right) = \text{Var}(X_1 + X_2 + \dots + X_n)$$

$$\implies \text{Var}\left(\frac{\sum X_i}{n}\right) = \text{Var}\left(\frac{X_1}{n}\right) + \text{Var}\left(\frac{X_2}{n}\right) + \dots + \text{Var}\left(\frac{X_n}{n}\right)$$

$$\text{Var}\left(\frac{\sum X_i}{n}\right) = \frac{\sigma^2}{n}$$

$$E(X^2) = \frac{\sigma^2}{n} + \mu$$

(c) Give an unbiased estimator of σ^2 .

Solution: Unbiased estimator of $\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2$

(d) What is a consistent estimator? Is \bar{X}_n a consistent estimator of μ ?

Solution: An estimator is consistent if, as the sample size increases, the estimates “converge” to the true value of the parameter being estimated. To be slightly more precise - consistency means that, as the sample size increases, the sampling distribution of the estimator becomes increasingly concentrated at the true parameter value.

By using Weak Law of Large Numbers,

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \bar{X}_i \rightarrow \mu$$

Therefore, \bar{X}_n a consistent estimator of μ

2. Suppose $X_{p \times 1}$ is a vector of covariates, $\beta_{p \times 1}$ is a vector of unknown parameters, ϵ is the unobserved random noise and we assume the linear model relationship $y = X^T \beta + \epsilon$. Suppose we have n i.i.d. samples from this linear model, and the observed data can be written using the matrix form: $\mathbf{y}_{n \times 1} = \mathbf{X}_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1}$.

(a) If we want to estimate the unknown β using a least square method, what is the objective/loss function $L(\beta)$ to obtain $\hat{\beta}$?

Solution:

$$L(\beta) = \frac{1}{2} (X\beta - y)^T (X\beta - y)$$

(b) What is the solution of $\hat{\beta}$? Represent the solution using the observed data \mathbf{y} and $\mathbf{X}_{n \times p}$. Note that you may assume that $\mathbf{X}^T \mathbf{X}$ is invertible.

Solution:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Question 4: Programming

1. Use the following code to generate a set of n observations $\mathbf{y}_{n \times 1}$ and $\mathbf{X}_{n \times p}$. Follow the previously established formula to solve for the least square estimator $\hat{\beta}$. Note that you must write your own code, instead of using existing functions such as `lm()`. In addition, what should you do if you are asked to add an intercept term β_0 into your estimation (even the true $\beta_0 = 0$ in our data generator)?

```
set.seed(1)
n = 100
p = 5
X = matrix(rnorm(n * p), n, p)
y = X %*% c(1, 0, 0, 1, -1) + rnorm(n)
```

Calculating $\hat{\beta}$

```
library('pracma')
beta_hat = inv(t(X) %*% X) %*% t(X) %*% y
beta_hat
```

```
##           [,1]
## [1,]  0.88369785
## [2,] -0.01725878
## [3,] -0.02658416
## [4,]  0.90233834
## [5,] -1.05639331
```

If we are asked to add an intercept term β_0 into your estimation, we can change the design matrix by adding a column of 1s and thus changing its dimensions to $n * (p + 1)$. We can go ahead with calculating the $\hat{\beta}$ in a similar way and the first value will be indicative of β_0 .

2. Perform a simulation study to check the consistency of a sample mean estimator \bar{X}_n . Please save your random seed so that the results can be replicated by others.

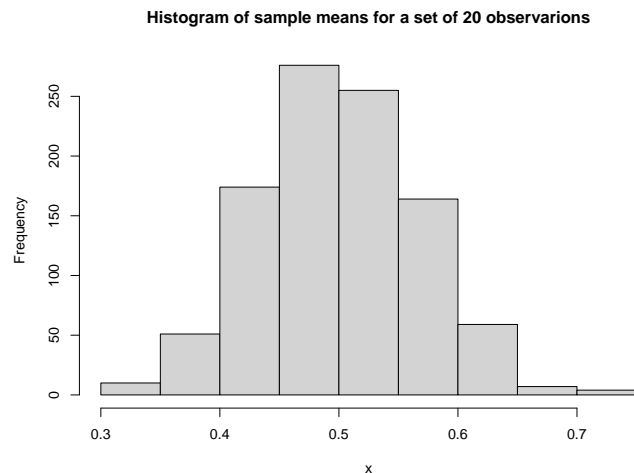
- (a) Generate a set of $n = 20$ i.i.d. observations from $\text{uniform}(0, 1)$ distribution and calculate the sample mean \bar{X}_n

```
set.seed(1)
a = runif(20, min = 0, max = 1)
sample_mean = mean(a)
print(paste(sample_mean, "is the sample mean for a set of 20 observations"))
```

```
## [1] "0.5551671132911 is the sample mean for a set of 20 observations"
```

- (b) Repeat step (a) 1000 times to collect 1000 such sample means and plot them using a histogram.

```
x = c()
for (i in 1:1000) {
  x[i] = mean(runif(20, min = 0, max = 1))
}
hist(x, main = paste("Histogram of sample means for a set of 20 observations"))
```



- (c) How many of such sample means (out of 1000) are at least 0.1 away from true mean parameter, which is 0.5 for uniform (0, 1)?

```
count = sum(x > 0.6) + sum(x < 0.4)
print(
  paste(
    count,
    "sample means that are at least 0.1 away from true mean for
      a set of 20 observations"
  )
)
```

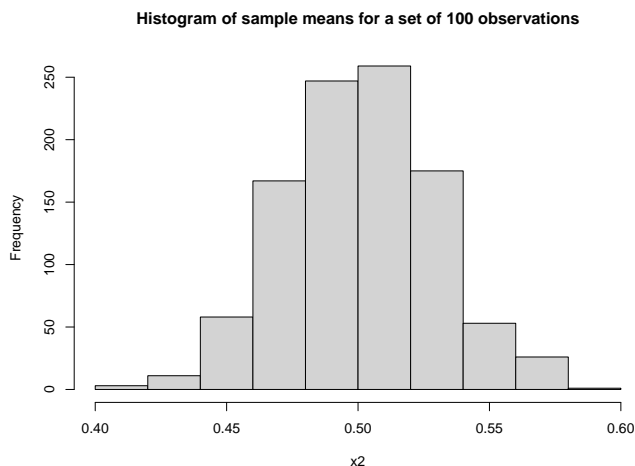
```
## [1] "131 sample means that are at least 0.1 away from true mean for\n      a set of 20 observations"
```

- (d) Repeat steps (a) to (c) with $n = 100$ and $n = 500$. What conclusion can you make?

```
set.seed(12)
b = runif(100, min = 0, max = 1)
sample_mean2 = mean(b)
print(paste(sample_mean2, "is the sample mean for a set of 100 observations"))
```

```
## [1] "0.479715390205383 is the sample mean for a set of 100 observations"
```

```
x2 = c()
for (i in 1:1000) {
  x2[i] = mean(runif(100, min = 0, max = 1))
}
hist(x2, main = paste("Histogram of sample means for a set of 100 observations"))
```



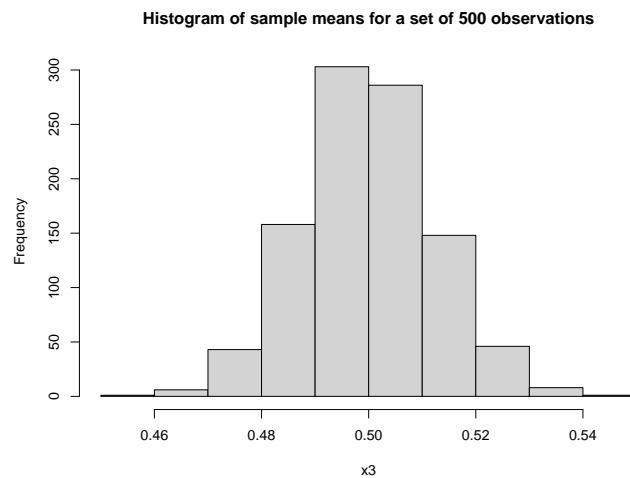
```
count2 = sum(x2 > 0.6) + sum(x2 < 0.4)
print(
  paste(
    count2,
    "sample means that are at least 0.1 away from true mean for
      a set of 100 observations"
  )
)
```

```
## [1] "0 sample means that are at least 0.1 away from true mean for\n          a set of 100 observations"
```

```
set.seed(123)
c = runif(500, min = 0, max = 1)
sample_mean3 = mean(c)
print(paste(sample_mean3, "is the sample mean for a set of 500 observations"))
```

```
## [1] "0.495283688393887 is the sample mean for a set of 500 observations"
```

```
x3 = c()
for (i in 1:1000) {
  x3[i] = mean(runif(500, min = 0, max = 1))
}
hist(x3, main = paste("Histogram of sample means for a set of 500 observations"))
```



```
count3 = sum(x3 > 0.6) + sum(x3 < 0.4)
print(
  paste(
    count3,
    "sample means that are at least 0.1 away from true mean for
      a set of 500 observations"
  )
)
```

```
## [1] "0 sample means that are at least 0.1 away from true mean for\n          a set of 500 observations"
```

Conclusions:

1. On increasing the value of n , we can see that the sample means gets closer to the true mean of 0.5.
2. On increasing the value of n , we can see that fewer samples mean are far away from the true mean of 0.5.