

Muhammet Yağcıoğlu - 290204042

May 25, 2024

## Contents

## Question 1

### *Problem Statement:*

Q1) (30p) The sources of electrical power for a region nuclear, fossil, and hydroelectric. The respective generating capacities of these sources can be described as independent Gaussian random variables as follows (in megawatt power):

Nuclear:  $N(100, 15)$

Fossil:  $N(200, 40)$

Hydro:  $N(400, 100)$

(a) Determine the total power supply for the region; i.e., define the probability distribution of the power supply with the corresponding mean and standard deviation. (b) The power demand of the region during normal weather is 400 megawatts, whereas during extreme weather the demand would be 600 megawatts. In any given year, normal weather is twice as likely to occur as extreme weather. With the generating capacity of (a), what is the probability that power shortage for the region will occur during the year?

### Part a

Let  $X_n$ ,  $X_f$ , and  $X_h$  denote the random variables representing the generating capacities of nuclear, fossil, and hydroelectric sources, respectively. We are given that these variables are independent and normally distributed:

$$X_n \sim \mathcal{N}(100, 15^2), \quad X_f \sim \mathcal{N}(200, 40^2), \quad X_h \sim \mathcal{N}(400, 100^2).$$

Let the total power supply, denoted by  $S$ , is the sum of these individual capacities:  $S = X_n + X_f + X_h$ . We invoke the Theorem of Sum of Independent Normal Random Variables, stipulating that if  $X_1, X_2, \dots, X_n$  are independent normal random variables with means  $\mu_1, \mu_2, \dots, \mu_n$  and variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ , then their sum  $Y = X_1 + X_2 + \dots + X_n$  is also normally distributed with mean  $\mu_Y = \mu_1 + \mu_2 + \dots + \mu_n$  and variance  $\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$ . Applying this theorem to our scenario, we deduce:

$$\begin{aligned} \mu_S &= \mu_{X_n} + \mu_{X_f} + \mu_{X_h} = 100 + 200 + 400 = 700 \text{ MW}, \\ \sigma_S &= \sqrt{\sigma_{X_n}^2 + \sigma_{X_f}^2 + \sigma_{X_h}^2} = \sqrt{15^2 + 40^2 + 100^2} \approx 109 \text{ MW}. \end{aligned}$$



### Part b

Let  $D$  denote the power demand of the region. Given that normal weather ( $N$ ) is twice as likely as extreme weather ( $\bar{N}$ ), we establish  $P(N) = \frac{2}{3}$  and  $P(\bar{N}) = \frac{1}{3}$ . The probability of a power shortage

conditioned on normal weather is

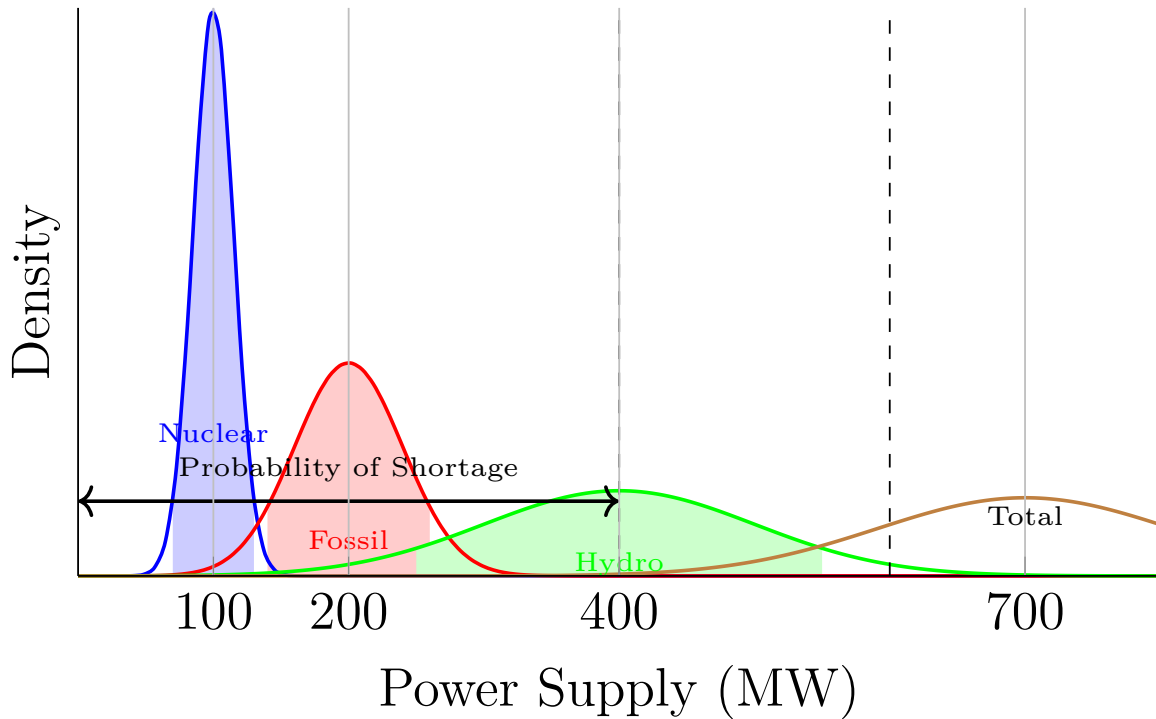
$$P(S < D \mid N) = P(S < d_N) = \Phi\left(\frac{d_N - \mu_S}{\sigma_S}\right),$$

where  $d_N = 400$  MW. Evaluating this yields  $P(S < D \mid N) = \Phi(-2.75) \approx 0.0029$ . Similarly, the probability of a power shortage conditioned on extreme weather is:

$$P(S < D \mid \bar{N}) = P(S < d_{\bar{N}}) = \Phi\left(\frac{d_{\bar{N}} - \mu_S}{\sigma_S}\right) = \Phi\left(\frac{600[MW] - 700[MW]}{109[MW]}\right) = \Phi(-0.91),$$

where  $d_{\bar{N}} = 600$  MW. The computation results in  $P(S < D \mid \bar{N}) = \Phi(-0.91) \approx 0.18$ . Finally, the probability of a power shortage,  $P(S)$ , is determined by the law of total probability:

$$\begin{aligned} P(S) &= P(N \cap S < d_N \cup \bar{N} \cap S < d_{\bar{N}}) = P(S < D \mid N)P(N) + P(S < D \mid \bar{N})P(\bar{N}), \\ &= P(S \mid N)P(N) + P(S \mid \bar{N})P(\bar{N}) \\ &= 0.0029 \frac{2}{3} + 0.18 \frac{1}{3} = \underline{0.6156} \end{aligned}$$



**ans.**  $X_S \sim \mathcal{N}(700, 109^2)$ ,  $P(S) = 0.6156$ .

## Question 2

### *Problem Statement:*

Q2) (30p) The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of two integrated layers. The layers are normally distributed with means of 0.5 and 1 millimeters, respectively. The standard deviations of layer thickness are 0.1 and 0.2 millimeters, respectively. The correlation between layers is 0.7 (a) Determine the mean and variance of the total thickness of the two layers. (b) What is the probability that the total thickness is less than 1 millimeter? (c) Let  $X_1$  and  $X_2$  denote the thickness of layers 1 and 2, respectively. A measure of performance of the membrane is a function  $2X_1 + 3X_2$  of the thickness. Determine the mean and variance of this performance measure.

We introduce the random variables  $X_1$  and  $X_2$  to denote the thicknesses of the first and second layers of the membrane, respectively. These variables are modeled as Gaussian distributed, with  $X_1 \sim \mathcal{N}(0.5, 0.1^2)$  and  $X_2 \sim \mathcal{N}(1, 0.2^2)$ . The correlation coefficient between  $X_1$  and  $X_2$  is given as  $\rho = 0.7$ .

### Part a

The total thickness of the membrane is represented by the random variable  $Y = X_1 + X_2$ . Since  $X_1$  and  $X_2$  are linear combinations of Gaussian variables,  $Y$  is also Gaussian. The mean of  $Y$ , denoted as  $\mathbb{E}[Y]$ , is the sum of the means of  $X_1$  and  $X_2$ :

$$\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 0.5 + 1 = 1.5 \text{ mm.}$$

The variance of  $Y$ , denoted as  $\text{Var}(Y)$ , incorporates both individual variances and the covariance between  $X_1$  and  $X_2$ , given by  $\text{Cov}(X_1, X_2) = \rho\sigma_{X_1}\sigma_{X_2}$ . Thus,

$$\text{Var}(Y) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2) = (0.1)^2 + (0.2)^2 + 2 \times 0.7 \times 0.1 \times 0.2 = 0.078 \text{ mm}^2.$$



### Part b

To find the probability that the total thickness is less than 1 mm, we consider the standardized variable  $Z = \frac{Y - \mathbb{E}[Y]}{\sqrt{\text{Var}(Y)}}$ . The probability in question is  $P(Y < 1)$ , which translates to  $P\left(Z < \frac{1 - 1.5}{\sqrt{0.078}}\right)$ .

This is a basic computation involving the standard normal distribution:

$$P(Y < 1) = P\left(Z < \frac{1 - 1.5}{\sqrt{0.078}}\right) = P(Z < -1.79) \approx 0.0367.$$



**Part c**

The performance measure is a function of the membrane's thickness, given by  $T = 2X_1 + 3X_2$ . To determine the mean  $\mathbb{E}[T]$  and variance  $\text{Var}(T)$  of this measure, we apply linear transformation properties on  $X_1$  and  $X_2$ . The mean of  $T$  is:

$$\mathbb{E}[T] = \mathbb{E}[2X_1 + 3X_2] = 2\mathbb{E}[X_1] + 3\mathbb{E}[X_2] = 2 \times 0.5 + 3 \times 1 = 4 \text{ mm}.$$

The variance of  $T$  incorporates the variances and covariance of  $X_1$  and  $X_2$ :

$$\begin{aligned} \text{Var}(T) &= \text{Var}(2X_1 + 3X_2) = 4\text{Var}(X_1) + 9\text{Var}(X_2) + 12\text{Cov}(X_1, X_2) \\ &= 4 \times (0.1)^2 + 9 \times (0.2)^2 + 12 \times 0.7 \times 0.1 \times 0.2 = 0.568 \text{ mm}^2. \end{aligned}$$



**ans.**  $\mathbb{E}[Y] = 1.5 \text{ mm}$ ,  $\text{Var}(Y) = 0.078 \text{ mm}^2$ .  $P(Y < 1) \approx 0.0367$ .  $\mathbb{E}[T] = 4 \text{ mm}$ ,  $\text{Var}(T) = 0.568 \text{ mm}^2$ .

## Question 3

### **Problem Statement:**

Q3) (40p) Suppose the annual maximum stream flow of a given river has been observed for 10 years yielding the following statistics:

Sample mean:  $10^6 \text{ m}^3/\text{h}$

Sample variance:  $9 \times 10^8 \left( \text{m}^3/\text{h} \right)^2$

(a) Establish the two-sided 90% confidence interval on the mean annual maximum stream flow. Assume a normal population. (b) If the desired to estimate the mean annual maximum stream flow to within  $\pm 30000 \text{ m}^3/\text{h}$  with 90% confidence, how many additional years of observation will be require? Assume the sample variance based on the new set of data will be approximately  $9 \times 10^8 \left( \text{m}^3/\text{h} \right)^2$ .

Let us denote the annual maximum stream flow as a random variable  $X$  in the probability space  $(\Omega, \mathcal{F}, P)$ , where  $\Omega$  is the sample space,  $\mathcal{F}$  a sigma-algebra of events, and  $P$  the probability measure. The random variable  $X : \Omega \rightarrow \mathbb{R}$  is assumed to be normally distributed, which situates its range within the manifold of real numbers,  $\mathbb{R}$ , under the Gaussian measure. Given a sample of size  $n$  from this distribution, represented as  $\{X_1, X_2, \dots, X_n\}$ , the sample mean  $\bar{X}$  and sample variance  $S^2$  are estimators for the population parameters  $\mu$  and  $\sigma^2$ , respectively. In this scenario,  $\bar{X} = 10^6 \text{ m}^3/\text{h}$  and  $S^2 = 9 \times 10^8 \left( \text{m}^3/\text{h} \right)^2$  with  $n = 10$ .

### **Part a**

To construct a 90% confidence interval for  $\mu$ , we utilize the t-distribution, a pivotal quantity stemming from the fact that  $\frac{\bar{X} - \mu}{S/\sqrt{n}}$  follows a Student's t-distribution with  $n - 1$  degrees of freedom under the assumption of normality. The two-sided 90% confidence interval for  $\mu$  is given by:

$$CI = \left[ \bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \right]$$

Where  $\alpha = 0.1$  and  $t_{\alpha/2, n-1}$  is the critical value from the t-distribution. This interval embodies the essence of inferential statistics: it provides a range within which the true parameter  $\mu$  is expected to lie with a specified level of confidence. In applying these principles, we extract  $t_{0.05, 9}$  from the t-distribution, recognizing this critical value as a fundamental threshold in the realm of probability distributions. The substitution of these values into the formula yields the desired confidence interval:

$$CI = \left[ 10^6 - t_{0.05, 9} \frac{\sqrt{9 \times 10^8}}{\sqrt{10}}, 10^6 + t_{0.05, 9} \frac{\sqrt{9 \times 10^8}}{\sqrt{10}} \right] \cong [8261 \text{ m}^3/\text{h}, 11739 \text{ m}^3/\text{h}] .$$



## Part b

We consider a random variable  $X$ , representing the annual maximum stream flow, which follows a normal distribution as established earlier. The sample mean  $\bar{X}$  and sample variance  $S^2$  are estimators of the population mean  $\mu$  and variance  $\sigma^2$ . Our goal is to determine the required sample size  $n_{\text{new}}$  such that the margin of error  $E$  in estimating  $\mu$  does not exceed  $\pm 30000 \text{ m}^3/\text{h}$  with 90% confidence. The margin of error in the context of a t-distribution is given by:

$$E = t_{\alpha/2, n_{\text{new}}-1} \cdot \frac{S}{\sqrt{n_{\text{new}}}}$$

where  $\alpha = 0.1$  and  $t_{\alpha/2, n_{\text{new}}-1}$  is the critical value from the t-distribution for  $n_{\text{new}} - 1$  degrees of freedom. To solve for  $n_{\text{new}}$ , we recognize that  $t_{\alpha/2, n_{\text{new}}-1}$  varies with  $n_{\text{new}}$ , necessitating an iterative approach. Rearranging the margin of error formula, we obtain:

$$n_{\text{new}} = \left( \frac{t_{\alpha/2, n_{\text{new}}-1} \cdot S}{E} \right)^2$$

Given  $S^2 = 9 \times 10^8 (\text{m}^3/\text{h})^2$  and  $E = 30000 \text{ m}^3/\text{h}$ , we iteratively adjust  $n_{\text{new}}$  and compute the corresponding  $t_{\alpha/2, n_{\text{new}}-1}$  until the margin of error is less than or equal to the desired threshold.

```
% Define the parameters
alpha = 0.1; % Significance level for 90%
    confidence (two-tailed)
n = 15; % Starting sample size

% Initialize arrays to store the results
sample_sizes = [];
t_values = [];
ratios = [];

% Iteratively calculate t-values and ratios
while true
    t_value = tinv(1 - alpha/2, n - 1); %
        t-value for the current sample size
    ratio = t_value / sqrt(n); % Ratio
        t/sqrt(n)

    % Store the results
    sample_sizes = [sample_sizes, n];
    t_values = [t_values, t_value];
    ratios = [ratios, ratio];

    % Check if the ratio is less than 1/3,
        if so, stop the loop
    if ratio < 1/3
        break;
    end
    n = n + 1;
end

% Create a table with the results
results_table = table(sample_sizes',
    t_values', ratios', ...
        'VariableNames',
            {'Sample_Size',
                't_0_05_n_minus_1',
                'Ratio_t_over_sqrt_n'});

disp(results_table);
writetable(results_table,
    'SampleSizeCalculations.csv');
```

**Output**

Sample_size	t_0_05_n_minus_1	Ratio_t_over_sqrt_n
15	1.7613	0.45477
16	1.7531	0.43826
17	1.7459	0.42344
18	1.7396	0.41003
19	1.7341	0.39782
20	1.7291	0.38665
21	1.7247	0.37636
22	1.7207	0.36686
23	1.7171	0.35805
24	1.7139	0.34984
25	1.7109	0.34218
26	1.7081	0.33499
27	1.7056	0.32825

Following the iterative calculation using MATLAB, it has been established that a sample size of 27 is sufficient to ensure that the margin of error in estimating the mean annual maximum stream flow does not exceed  $\pm 30000 \text{ m}^3/\text{h}$  at a 90% confidence level. Given that the initial dataset encompassed observations over 10 years, the requirement for additional observations can be deduced. To reach the requisite sample size of 27, an additional  $27 - 10 = 17$  years of observation will be necessary.

**Algorithm Table for Adaptation to Other Languages****Algorithm 1** Determination of Required Sample Size on a Statistical Manifold

1: $\alpha \leftarrow 0.1$	▷ Significance level
2: $n \leftarrow 15$	▷ Initial sample size
3: $\theta \leftarrow \frac{1}{3}$	▷ Ratio limit
4: <b>while</b> True <b>do</b>	
5: $t_{\text{value}} \leftarrow \text{TINV}(1 - \frac{\alpha}{2}, n - 1)$	
6: $\rho \leftarrow \frac{t_{\text{value}}}{\sqrt{n}}$	
7: <b>if</b> $\rho < \theta$ <b>then</b>	
8: <b>break</b>	
9: <b>end if</b>	
10: $n \leftarrow n + 1$	
11: <b>end while</b>	
12: <b>return</b> $n$	

**ans.**  $CI \cong [8261 \text{ m}^3/\text{h}, 11739 \text{ m}^3/\text{h}] , 17.$



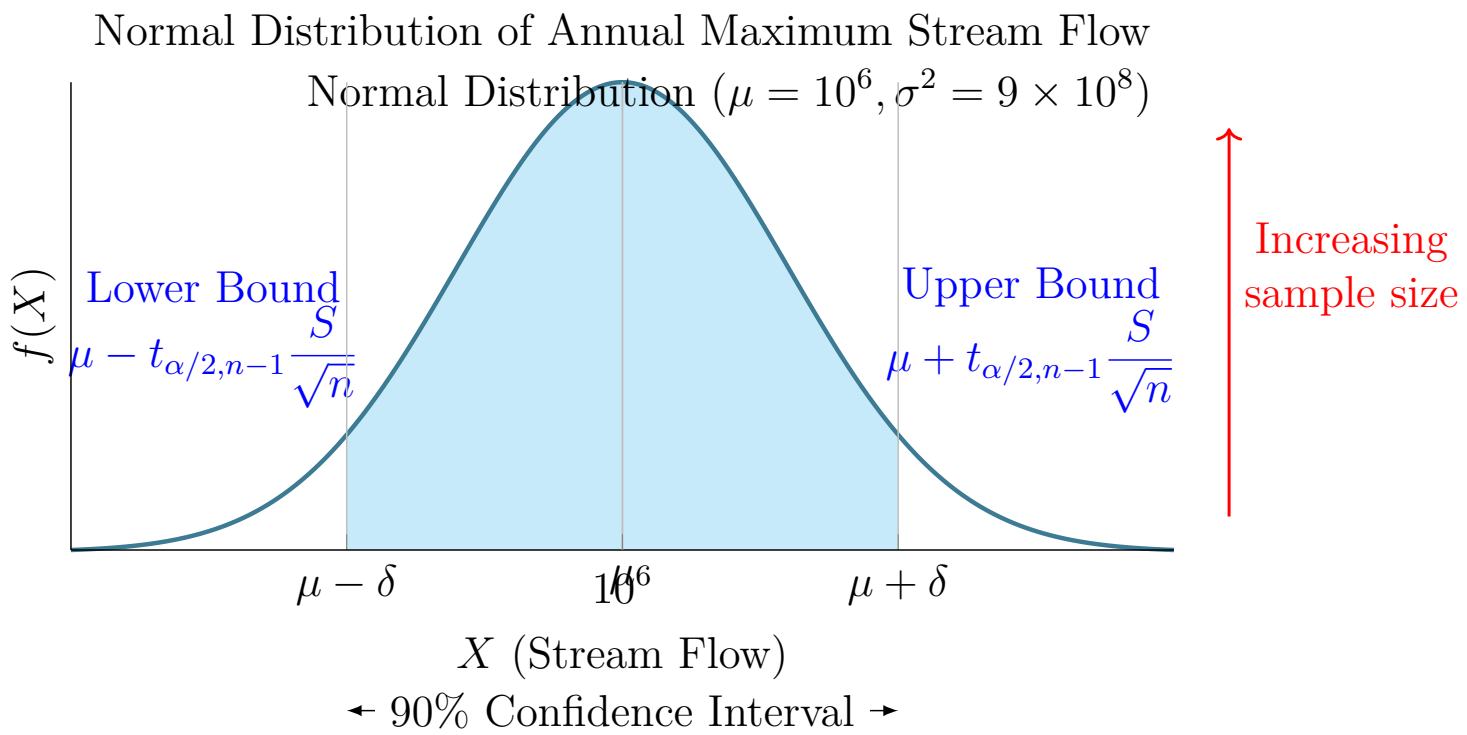


Figure 1

ans.  $CI \cong [8261 \text{ m}^3/\text{h}, 11739 \text{ m}^3/\text{h}] , 17.$