ME 103 — Experimentation and Measurements — **HW1**

Spring 2026 — George Anwar

Due: Monday, February 9th at 11:59 pm

- This homework will cover the basics of probability, statistics, uncertainty, and error propagation.
- In all of the questions, **show your work**, not just the final answer. Unless we explicitly state otherwise, you may expect full credit only if you explain your work succinctly, but clearly and convincingly. For coding questions, attach a screenshot of your code and output.
- Present your answers with **a suitable number of significant figures** for each question. Show your work, including a mathematical formula or the MATLAB or Python code you wrote, before reaching the result. You may need to install the Statistics Toolbox if using MATLAB.
- Throughout this assignment, neglect systematic (bias) errors. Also, assume a normal distribution for the underlying distribution (population) if necessary.
- If you have a confirmed disability that precludes you from complying fully with these instructions or with any other parameter associated with this problem set, please alert us immediately about reasonable accommodations afforded to you by the DSP Office on campus.
- **Start early. Some of the material is prerequisite material not covered in lecture; you are responsible for finding resources to understand it.**

Deliverables

Submit a PDF of your homework to the **Gradescope assignment** entitled "{Your Name} HW1". You **must** typeset your homework in LaTeX (submit PDF format, not .doc/.docx format) . Mac Preview, PDF Expert, and FoxIt PDF Reader, among others, have tools to let you sign a PDF file. We want to make *extra clear* the consequences of cheating.

1 Honor Code

I will adhere to the Berkeley Honor Code: specifically, as a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. Failure to comply with these guidelines can be considered an academic integrity violation. Please email Professor Anwar ganwar@berkeley.edu if you have any questions!

- **List all collaborators. If you worked alone, then you must explicitly state so. Read the following statement and sign below if you agree**:

  "I certify that all solutions in this document are entirely my own and that I have not looked at anyone else's solution. I have given credit to all external sources I consulted."

  Signature : ______________________ *Date* : ______________________

  While discussions are encouraged, *everything* in your solution must be your (and only your) creation. Furthermore, all external material (i.e., *anything* outside lectures and assigned readings, including figures and pictures) should be cited properly. We wish to remind you that consequences of academic misconduct are *particularly severe*!
- Violation of the Code of Conduct will result in a **zero** on this assignment and may also result in disciplinary action.

2 Probabilities and the Normal Distribution [11pts]

A race team is trying to measure the torsional stiffness of their chassis by measuring the deflection given a known identical input force. The sample set they collected from 12 tests is listed below (in mm): Assume the

4.79	4.92	4.58	4.93	4.95	4.31
4.72	4.09	4.86	4.48	4.21	5.12

measurement process follows a normal distribution.

(a) [2 pts] Estimate the population mean of the measurement process.

Solution: The best estimate of the population mean is the sample mean as it is an unbiased estimator of the population mean by MLE. Hence,

$$\hat{\mu} = \bar{x} = 4.6633\text{mm} = 4.66\text{mm} \quad 3 \text{ s.f.}$$

(The rationale for using 3 s.f. is that the data is 3 s.f. , so reporting 3 s.f. retains all the available information in the data.)

Matlab command: if data are `HW1data`, answer is given by `mean(HW1data)`.

Some responses may use the standard error of the mean to report a confidence interval for the population mean, but this is not required. 2 points for working and a final answer shown, whether correct or not. 1 point if some work but no final answer is shown (what is shown above would count as working; there is no need to show the summing up of all the data points and dividing through by 12 as it is accepted that everyone knows how to calculate a mean). 0 points for a final numerical answer with no working.

(b) [2 pts] Estimate the population standard deviation of the measurement process.

Solution: The best estimate of the population standard deviation is the sample standard deviation.

$$\hat{\sigma} = s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 - n\bar{x}^2}{n-1}} = 0.3283\text{mm} = 0.328\text{mm} \quad (3 \text{ s.f.})$$

Matlab command: if data are `HW1data`, answer is given by `std(HW1data)`.

(The rationale for 3 d.p. is that this is the same number of d.p. as in the reported sample mean. Therefore if one adds or subtracts some multiple of the standard deviation from the mean in the process of calculating a z value, the result will not appear to have a level of precision that is unwarranted.)

Note that we are estimating the population standard deviation, not the standard deviation of the sample mean, so we use the sample standard deviation as our estimator, not the standard error of the mean. Dividing by $\sqrt{12}$ would not be correct here.

2 points for working and a final answer shown, whether correct or not. 1 point if some work but no final answer is shown. 0 points for a final numerical answer with no working.

(c) [3 pts] If the team is to collect another measurement, estimate the probability that the next measurement will fall within the range of [4.31 mm, 5.1 mm] based on the sample set.

Solution: Let the next measurement be x_0, and the standard normal CDF evaluated from $-\infty$ to z be $\Phi(z)$. Negative z-score means area to the left of $-z$ is the same as the area to the right of $+z$.

$$\mathbb{P}[4.31\text{mm} < x_0 < 5.1\text{mm}] = \mathbb{P}[x_0 < 5.1\text{mm}] - \mathbb{P}[x_0 < 4.31\text{mm}] = \Phi\left(\frac{5.1 - \hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{4.31 - \hat{\mu}}{\hat{\sigma}}\right)$$

$$= 0.91 - (1 - 0.8599) = 0.7674 = 77\% \quad (2 \text{ s.f.})$$

The rationale for 2 s.f. is that the least precise number given in the range of interest was 5.1 mm (2 s.f), so there is no point reporting a probability to a higher number of s.f. The relevant Matlab function to use for $\Phi(z)$ is `normcdf(z)`.

The z distribution is the correct distribution because we are evaluating the probability of drawing values in a certain range from a normally distributed underlying population. The t distribution is a sampling distribution (used to calculate confidence intervals for a population mean based on a sample mean), and so is not appropriate to use here.

(d) [4 pts] **This question is separate from the above**. Metal balls manufactured by a company have an average diameter of 6.000 `mm` and a tolerance of ±0.200 `mm` with a 95% confidence. Assume the manufactured diameters are normally distributed. What is the probability that the diameter of a metal ball is above 6.050 `mm`?

Solution: First convert the information given to a population standard deviation. A tolerance of ±0.200 mm with a 95% confidence means that

$$0.200\texttt{mm} = 1.96\sigma$$

$$\sigma = \frac{0.200}{1.96}\texttt{mm} = 0.102\texttt{mm}$$

Assuming that $0.200\texttt{mm} = 2\sigma$ is also accepted for full credit.

Then we can use the standard normal CDF to compute the required probability. Let the diameter of one randomly selected metal ball be d_0:

$$\mathbb{P}[d_0 > 6.050\texttt{mm}] = 1 - \mathbb{P}[d_0 \leq 6.050\texttt{mm}] = 1 - \Phi\left(\frac{6.050 - 6.000}{0.102}\right) = 31.2\% \quad \text{(3 s.f.)}$$

If it is assumed that $0.200\texttt{mm} = 2\sigma$, the answer is 30.2%. The rationale for 3 s.f. in the answer is that the tolerance is specified to 3 s.f., which is the smallest number of s.f. in the input variables.

1 point for correctly calculating the standard deviation of the population; 1 point for setting up and expression for the probability using the normal CDF; 1 point for correct evaluation; 1 point for reasonable number of s.f.

3 Fatigue Probabilities [13 pts]

THIS ENTIRE QUESTION IS FOR COMPLETION

(a) In a certain lab, 5% of aluminum specimens actually fail under fatigue loading. It is known that a strain-gauge based fatigue monitoring system has a *detection sensitivity* of 65% (i.e. the probability that a specimen is detected as a failure given that it truly failed is 65%). The system's *specificity* is 98% (i.e. the probability that a specimen is correctly classified as non-failure given it did not fail).

(i) [3 pts] What is the probability that a randomly selected specimen is flagged as "failed" by the system?

Solution: Define the events: $F = \{\text{specimen truly fails}\}$, $D = \{\text{system flags specimen as failed}\}$. We are given

$$\mathbb{P}(F) = 0.05, \quad \mathbb{P}(D \mid F) = 0.65, \quad \mathbb{P}(\overline{D} \mid \overline{F}) = 0.98.$$

Hence, we can infer that $\mathbb{P}(\overline{F}) = 0.95$, $\mathbb{P}(D \mid \overline{F}) = 1 - 0.98 = 0.02$. Then, by the law of total probability, we have

$$\mathbb{P}(D) = \mathbb{P}(D \mid F)\mathbb{P}(F) + \mathbb{P}(D \mid \overline{F})\mathbb{P}(\overline{F}) = (0.65)(0.05) + (0.02)(0.95) = 0.0515.$$

(ii) [2 pts] If a specimen is flagged as a "failure", what is the probability that it truly failed?

Solution: Using Bayes' theorem and the event definitions from above, we have

$$\mathbb{P}(F \mid D) = \frac{\mathbb{P}(D \mid F)\mathbb{P}(F)}{\mathbb{P}(D)} = \frac{(0.65)(0.05)}{0.0515} \approx 0.63.$$

(b) Suppose the error in the strain-gauge output is modeled as

- $X \sim \mathcal{N}(0, 4)$, representing sensor noise, and
- $Y \sim \mathcal{N}(3, 9)$, representing bias in the calibration of the gauge.

Assume X and Y are independent.

(i) [2 pts] What is the distribution of the total measurement error $X + 2Y$?

Solution: Since $\mathbb{E}[X + 2Y] = \mathbb{E}[X] + 2\mathbb{E}[Y] = 6$ and $\text{Var}[X + 2Y] + \text{Var}[X] + 4\text{Var}[Y] = 40$, therefore $X + 2Y \sim \mathcal{N}(6, 40)$

(ii) [2 pts] Find the probability that the noise error is larger than the bias, i.e. $\mathbb{P}[X > Y]$.

Solution: Let $Z = X - Y \sim \mathcal{N}(-3, 13)$, we have

$$\mathbb{P}(X > Y) = \mathbb{P}(Z > 0) = \mathbb{P}\left(\frac{Z+3}{\sqrt{13}} > \frac{3}{\sqrt{13}}\right) \approx 1 - \Phi(0.83) = 1 - 0.8 = 0.2$$

(c) [4 pts] Suppose we run $n = 10$ independent fatigue tests, each with the strain gauge described in (a). Let X be a Binomial random variable be the number of specimens correctly classified by the system (either true failures or true non-failures), where the unknown probability of correct classification is p. We use $\hat{p} = \frac{X}{10}$ as an estimator for p. Find the bias, variance, and MSE of $\hat{p}$.

Solution: We have $\mathbb{E}[\hat{p}] = \frac{\mathbb{E}[X]}{10} = \frac{1}{10} \cdot 10p = p$, so $\text{Bias}(\hat{p}) = 0$. Moreover, $\text{Var}[X] = \frac{1}{100}\text{Var}[X] = \frac{p(1-p)}{10}$. Finally, $\text{MSE}(\hat{p}) = (\text{Bias}(\hat{p}))^2 + \text{Var}[\hat{p}] = \frac{p(1-p)}{10}$

4 Probability Potpourri [33 pts]

(a) The probability that an archer hits her target when it is windy is 0.4; when it is not windy, her probability of hitting the target is 0.7. On any shot, the probability of a gust of wind is 0.3. Find the probability that

(i) [2 pts] on a given shot there is a gust of wind and she hits her target.

Solution: For the following solutions let us define event H as hitting the target; let W represent wind being present. $\overline{H}$ and $\overline{W}$ represent the complements respectively. Then,

$$\mathbb{P}[H \cap W] = \mathbb{P}[H|W]\mathbb{P}[W] = (0.4)(0.3) = 0.12$$

Therefore, on a given shot where there is a gust of wind, she hits her target 12% of the time.

(ii) [2 pts] she hits the target with her first shot.

Solution: For 2 disjoint events W and $\overline{W}$ and event H, we can use the theorem of total probabilities, which says

$$\mathbb{P}[H] = \mathbb{P}[H|W]\mathbb{P}[W] + \mathbb{P}[H|\overline{W}]\mathbb{P}[\overline{W}] = (0.4)(0.3) + (0.7)(1 - 0.3) = 0.61$$

(iii) [2 pts] she hits the target exactly once in two shots.

Solution: The probability that she hits the target exactly once in two shots is given by the binomial formula

$$\mathbb{P}[\text{hit exactly once in two}] = \binom{n}{x}\mathbb{P}[H]^x(1-\mathbb{P}[H])^{n-x} = \binom{2}{1}\mathbb{P}[H]^1(1-\mathbb{P}[H])^{2-1} = 2(0.61)(0.39) = 0.48$$

Therefore, she has a 48% to hit the target exactly once in two shots.

(iv) [3 pts] on an occasion when she missed, there was no gust of wind.

Solution: Since we need to compute $\mathbb{P}[\overline{W}|\overline{H}]$, we can use Bayes' Theorem.

$$\mathbb{P}[\overline{W}|\overline{H}] = \frac{\mathbb{P}[\overline{H}|\overline{W}]\mathbb{P}[\overline{W}]}{\mathbb{P}[\overline{H}]}$$

Here, $\mathbb{P}[\overline{W}]$ is calculated using the theorem of total probabilities

$$\mathbb{P}[\overline{H}] = \mathbb{P}[\overline{H}|\overline{W}]\mathbb{P}[\overline{W}] + \mathbb{P}[\overline{H}|W]\mathbb{P}[W] = (1 - 0.7)(1 - 0.3) + (1 - 0.4)(0.3) = 0.39$$

We know that $\mathbb{P}[\overline{H}|\overline{W}]\mathbb{P}[\overline{W}] = (1 - 0.7)(1 - 0.3) = 0.21$ So substituting into Bayes' Theorem

$$\mathbb{P}[\overline{W}|\overline{H}] = \frac{\mathbb{P}[\overline{H}|\overline{W}]\mathbb{P}[\overline{W}]}{\mathbb{P}[\overline{H}]} = \frac{0.21}{0.39} = 0.53846$$

Therefore, there is a 53.8% that on an occasion where she missed, there was no gust of wind.

(b) [6 pts] An archery target is made of 3 concentric circles of radii $1/\sqrt{3}$, 1 and $\sqrt{3}$ feet. Arrows striking within the inner circle are awarded 4 points, arrows within the middle ring are awarded 3 points, and arrows within the outer ring are awarded 2 points. Shots outside the target are awarded 0 points.

Consider a random variable X, the distance of the strike from the center in feet, and let the probability density function of X be

$$f(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected value of the score of a single strike?

Solution: Given that X is a continuous random variable, the expected value is given as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

the function is defined as 0 for values of $x \leq 0$ so we can change our lower bound to 0 and the upper bounds will vary from $1/\sqrt{3}, 1$, and $\sqrt{3}$ feet. We write the expected value now as three separate integrals

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{1/\sqrt{3}} 4 \cdot \frac{2}{\pi(1+x^2)} dx + \int_{1/\sqrt{3}}^{1} 3 \cdot \frac{2}{\pi(1+x^2)} dx + \int_1^{\sqrt{3}} 2 \cdot \frac{2}{\pi(1+x^2)} dx \\
&= \int_0^{1/\sqrt{3}} \frac{8}{\pi(1+x^2)} dx + \int_{1/\sqrt{3}}^{1} \frac{6}{\pi(1+x^2)} dx + \int_1^{\sqrt{3}} \frac{4}{\pi(1+x^2)} dx \\
&= \frac{8}{\pi} \int_0^{1/\sqrt{3}} \frac{1}{1+x^2} dx + \frac{6}{\pi} \int_{1/\sqrt{3}}^{1} \frac{1}{1+x^2} dx + \frac{4}{\pi} \int_1^{\sqrt{3}} \frac{1}{1+x^2} dx \\
&= \frac{8}{\pi} \arctan(x)\Big|_0^{1/\sqrt{3}} + \frac{6}{\pi} \arctan(x)\Big|_{1/\sqrt{3}}^{1} + \frac{4}{\pi} \arctan(x)\Big|_1^{\sqrt{3}} \\
&= \frac{4}{3} + \frac{1}{2} + \frac{1}{3} \Rightarrow \mathbb{E}[X] = \frac{13}{6}
\end{aligned}$$

Therefore, the expected value of the score of a single strike is 13/6.

(c) [6 pts] You are testing a suspension damper on a shock dynamometer. During a test run you collect two independent summaries. In Dyno Run A, for each of 100 compression strokes, the dyno's DAQ bins the peak force into one of 6 discrete integer-coded levels $\{1, \ldots, 6\}$ with equal likelihood per stroke, due to coarse quantization. You **sum** the 100 coded levels. Call this random variable X. For Dyno Run B, over the same session, the dyno's trigger channel flags whether a shock event occurred in each of 600 time windows. Each window independently records 1 (event) or 0 (no event) with probability 0.5 each. You count the total number of flagged events. Call this random variable Y. Use the Central Limit Theorem to approximate $\mathbb{P}(X < Y)$.

Solution: First compute the mean and variance of X and Y.

Dyno Run A. Let X_i denote the integer-coded force level recorded on the ith compression stroke, where $X_i \in \{1, 2, 3, 4, 5, 6\}$ with equal probability. Then

$$\mathbb{E}[X_i] = \frac{1+2+3+4+5+6}{6} = 3.5, \qquad \mathrm{Var}(X_i) = \frac{6^2-1}{12} = \frac{35}{12}.$$

Since $X = \sum_{i=1}^{100} X_i$, note that $X \sim \mathrm{Uniform}(n)$

$$\mathbb{E}[X] = 100(3.5) = 350, \qquad \mathrm{Var}(X) = 100 \cdot \frac{35}{12} = \frac{3500}{12}.$$

Dyno Run B. Let Y_i denote the indicator of a shock event in the ith time window, where $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = 0) = 0.5$. Since $Y = \sum_{i=1}^{600} Y_i$, note that $Y \sim \mathrm{Binomial}(n = 600, p = 0.5)$

$$\mathbb{E}[Y] = 600(0.5) = 300, \qquad \mathrm{Var}(Y) = 600(0.5)(1-0.5) = 150.$$

CLT approximation. By the Central Limit Theorem,

$$X \approx \mathcal{N}\left(350, \frac{3500}{12}\right), \qquad Y \approx \mathcal{N}(300, 150).$$

Define new Random Variable Z, where:

$$Z = X - Y$$

Since X and Y are independent, their variances and means act as linear operators:

$$E[Z] = E[X - Y] = E[X] - E[Y] = 50$$

$$Var[Z] = Var[X] + Var[Y] \approx 442$$

Covariance is also zero becuase of independence. Therefore,

$$\mathbb{P}(X < Y) = \mathbb{P}(X - Y < 0) = \mathbb{P}\left(Z < \frac{0 - 50}{\sqrt{5300/12}}\right), \quad Z \sim \mathcal{N}(0, 1).$$

Numerically,

$$\mathbb{P}(X < Y) = \Phi(-2.38) \approx 0.0087.$$

(d) [6 pts] You meet two students in Hesse Hall. Assume that each student is a Senior or a Sophomore with equal probability, and each student takes ME103 with probability $1/10$, independent of each other and independent of their class standing. What is the probability that both students are Seniors, given at least one of them is a Senior currently taking ME103?

Solution: Let A_i be the event that student i is a Senior, and let B_i be the event that student i is currently taking ME103, for $i \in \{1, 2\}$. We are asked to compute

$$\mathbb{P}\big(A_1 \cap A_2 \,\big|\, (A_1 \cap B_1) \cup (A_2 \cap B_2)\big) = \frac{\mathbb{P}\Big(A_1 \cap A_2 \cap \big((A_1 \cap B_1) \cup (A_2 \cap B_2)\big)\Big)}{\mathbb{P}\big((A_1 \cap B_1) \cup (A_2 \cap B_2)\big)}.$$

Since $A_1 \cap A_2$ implies A_1 and A_2, the numerator simplifies to the following using $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$

$$\mathbb{P}\big(A_1 \cap A_2 \cap (B_1 \cup B_2)\big).$$

Using independence (between students and between A_i and B_i),

$$\mathbb{P}\big(A_1 \cap A_2 \cap (B_1 \cup B_2)\big) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(B_1 \cup B_2).$$

Also, since $(A_1 \cap B_1)$ and $(A_2 \cap B_2)$ are independent,

$$\mathbb{P}\big((A_1 \cap B_1) \cup (A_2 \cap B_2)\big) = 1 - \Big(1 - \mathbb{P}(A_1 \cap B_1)\Big)\Big(1 - \mathbb{P}(A_2 \cap B_2)\Big).$$

Now compute each term:

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \frac{1}{2}, \qquad \mathbb{P}(A_i \cap B_i) = \mathbb{P}(A_i)\mathbb{P}(B_i) = \frac{1}{2} \cdot \frac{1}{10} = \frac{1}{20},$$

and

$$\mathbb{P}(B_1 \cup B_2) = 1 - \mathbb{P}(\overline{B_1} \cap \overline{B_2}) = 1 - \left(\frac{9}{10}\right)^2 = \frac{19}{100}.$$

Substituting,

$$\mathbb{P}\big(A_1 \cap A_2 \,\big|\, (A_1 \cap B_1) \cup (A_2 \cap B_2)\big) = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{19}{100}\right)}{1 - \left(1 - \frac{1}{20}\right)^2} = \frac{\frac{19}{400}}{1 - \left(\frac{19}{20}\right)^2} = \frac{19}{39}.$$

(e) [6 pts] There are n identical looking sensors in Hesse 122 and n boxes to store them, where sensor i is supposed to go into box i. After a hectic cleanup, the sensors get thrown into the boxes uniformly at random (every permutation is equally likely). Assume n is a positive integer. What is the probability that **no sensor** ends up in its correct bin? Also find the limit of this probability as $n \to \infty$.

Hint: Use the principle of inclusion-exclusion and the power series representation of e^x. Assume that each box gets exactly one sensor.

Solution: It may be easier to first find the probability of the complementary event, namely that *at least one sensor* ends up in its correct box. Let A_i be the event that sensor i is placed into box i. Then the event that at least one sensor is correctly placed is $\bigcup_{i=1}^{n} A_i$.

By the principle of inclusion–exclusion, we may write

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n-1}\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

We now compute the terms appearing in this sum. If exactly one specified sensor i is fixed in its correct box, then the remaining $n-1$ sensors may be permuted arbitrarily among the remaining $n-1$ boxes. Thus,

$$\mathbb{P}(A_i) = \frac{(n-1)!}{n!}.$$

Consider next the second summation. For any particular pair $i < j$, the event $A_i \cap A_j$ occurs when sensors i and j are both placed in their correct boxes. The remaining $n-2$ sensors may then be arranged in $(n-2)!$ ways, giving

$$\mathbb{P}(A_i \cap A_j) = \frac{(n-2)!}{n!}.$$

There are $\binom{n}{2}$ choices of such pairs $i < j$.

Generalizing this reasoning, if k specified sensors are fixed in their correct boxes, then the remaining $n-k$ sensors may be arranged in $(n-k)!$ ways. Since there are $\binom{n}{k}$ choices of which k sensors are fixed, the inclusion–exclusion sum becomes

$$\sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!}.$$

Therefore, the probability that *no sensor* is placed into its correct box is

$$1 - \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Using the power series expansion $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, we conclude that

$$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} \approx 0.368.$$

Alternate solution (optional). Instead of using inclusion–exclusion, we can derive a recursion. Let B_n be the event that no sensor is placed into its correct box when there are n sensors, and define $p_n = \mathbb{P}(B_n)$. Clearly, $p_0 = 1$ since the statement is vacuously true, and $p_1 = 0$ since a single sensor must go into its own box.

For $n \geq 2$, consider sensor 1. With probability $\frac{n-1}{n}$, it is placed into a box $i \neq 1$. If sensor i does not go into box 1, then none of the remaining $n-1$ sensors must be correctly placed, which occurs with probability p_{n-1}. If sensor i does go into box 1 (which happens with probability $\frac{1}{n-1}$), then we only need the remaining $n-2$ sensors to be incorrectly placed, which occurs with probability p_{n-2}. Combining these cases yields

$$p_n = \frac{n-1}{n} p_{n-1} + \frac{1}{n} p_{n-2}.$$

To solve this recursion, define the differences $d_n = p_n - p_{n-1}$. Then

$$d_n = \left(\left(1 - \frac{1}{n}\right) p_{n-1} + \frac{1}{n} p_{n-2}\right) - p_{n-1} = -\frac{1}{n}(p_{n-1} - p_{n-2}) = -\frac{1}{n} d_{n-1}.$$

The base case is $d_1 = p_1 - p_0 = -1$. Iterating the recursion gives

$$d_n = -\frac{1}{n} d_{n-1} = \frac{1}{n(n-1)} d_{n-2} = \cdots = \frac{(-1)^{n-1}}{n!} d_1 = \frac{(-1)^n}{n!}.$$

Finally, writing p_n as a telescoping sum and taking $d_0 := p_0$,

$$p_n = (p_n - p_{n-1}) + \cdots + (p_1 - p_0) + p_0 = \sum_{i=0}^{n} d_i = \sum_{i=0}^{n} \frac{(-1)^i}{i!}.$$

5 Confidence Intervals [17 pts]

Use the provided data (`HW1data.mat`) for this question. Each column in the matrix called data represents one data set collected by running nominally identical tests on one system. Note that the sizes of the data sets distinguished in each column representing `test1`, `test2` and `test3` are 20, 14, and 17, respectively. Do not count the 0's in any of the datasets. Use MATLAB for calculations and look-up tables (z or t) if appropriate.

(a) [3 pts] For each of the three data sets, complete the table by calculating the mean and standard deviation, and stating the number of degrees of freedom. The data provided have 6 significant figures.

Data set, i	**Sample size n_i**	**Sample mean $\bar{x}_i$**	**Sample standard deviation $S_{x,i}$**	**Degrees of Freedom ν_i**
1	20	12081.3	1105.08	19
2	14	11971.2	1281.83	13
3	17	12141.9	1247.84	16

Table 1: Summary of sample statistics for three data sets

Solution: Sample means are reported to 6 s.f. because that is the same number of s.f. as the sum of the data points. Std dev are reported to 6 s.f. although it would also be OK to report to 1 d.p. to match the sample means.

1 point for sample mean numerical calculations; 1 point for sample std dev numerical calculations; 0.5 point for correct use of s.f.; 0.5 point correct number of degrees of freedom.

Partial credit within each grading element above is possible if not all data sets are correct. By now we assume that students can operate the `mean()` and `std()`, or equivalent, functions, so no working is required here beyond completion of the table.

(b) Calculate the following quantities

(i) [1 pt] The mean of three mean values computed from above.

Solution: Mean of means without weighting is 12064.8 (6 s.f.) (reached using a command like `mean([mean1 mean2 mean3])`, where `mean1` etc. are the mean values calculated in the table above).

(ii) [1 pt] The "pooled" mean of three means (= the weighted mean with respect to sample size).

Solution: The pooled mean is found as: `pooled_mean = (20*mean1 + 14*mean2 + 17*mean3) / (20+14+17)` which evaluates to 12071.3 (6 s.f.)

(iii) [1 pt] The mean of all the combined data sets.

Solution: The mean of all combined data is effectively the same as computing the pooled mean. It can be computed by: `mean_all = mean([data(:,1); data(1:14,2); data(1:17,3)])` which also evaluates to 12071.3 (6 s.f.)

(c) [1 pt] Based on the results, discuss why it is favorable to maintain the same size of sample during experiments.

Solution: In the first case computed above (averaging the three sample means), individual data points in the smaller samples had a disproportionately large effect on the final value. This would make it very hard to carry out fair statistical tests that depended on the use of sampling distributions. Therefore, data analysis may be slightly simpler if one can compare samples with equal sizes.

However, it should also be noted that in real experiments, data are often 'messy' and it can often be the case that one has more data points for a given set of conditions than for another set of conditions. There are completely valid ways of comparing samples that have different sizes (e.g. the two-sample t-test that was shown in lecture). It would certainly not be advisable, for example, to discard data from larger samples to bring all samples down to a consistent size – that would reduce the power of any statistical test that may later be carried out to compare them. For example, imagine preparing two sets of 10 dogbones of two different materials for testing to compare their strength. If one of the dogbones in one of the sets was accidentally dropped on the floor and broke before it could be tested, you wouldn't want to throw away a random dogbone from the other set to make the two sets have 9 data points each – that would be foregoing potentially perfectly useful data.

1 point for any reasonable comment that is well articulated.

(d) [4 pt] Considering only the data in `test1`, compute a 95% confidence interval for the true mean of the underlying distribution of measurements.(Note that because there are 20 data points in `test1`, the t-distribution is needed to determine an accurate confidence interval.) Give your answer to 4 significant figures.

Solution: Central estimate of population mean is $\hat{\mu} = 12081.3$ (6 s.f.). Standard error of the mean, is $\hat{\sigma}_x = 1105.08/\sqrt{20} = 247.104$ (6 s.f.)

We require a 95% confidence interval that will be symmetrical about our central estimate of the mean. Since there are 20 data points, there are 19 degrees of freedom and the t distribution must be used. The t value corresponding to the 97th percentile of the t-distribution is given by `t_CI=tinv(0.975, 19)` in MATLAB, which is 2.093024. Note we want 2.5% in each tail of the sampling distribution outside, therefore the confidence interval is:

$$\hat{\mu} \pm t_{CI}\hat{\sigma}_{\bar{x}} = 12081.3 \pm 2.093024 \times 247.104 = 12080 \pm 517.2 \text{ (4 s.f. as requested)}$$

It would also be acceptable to report the bounds of the confidence interval: $[1156E1 \quad 1260E1]$,

- 2 points for correct approach to calculating the size of the confidence interval
- 1 point for correct numerical evaluation
- 1 point for using 4 s.f. as requested.

(e) [4 pt] If a normal distribution of the sample mean was assumed instead of the t-distribution in calculating the confidence interval in the previous part, by what factor would the size of the confidence interval differ? (First calculate the confidence interval would you have computed if you had (wrongly) assumed that the sample was 'large' instead of having a finite size of 20 samples.)

Solution: In this case, the z values of the extremities of the CI would be $\pm$`norminv(0.975)` $= 1.96$ The confidence interval would then be smaller by a factor of $1.96/2.09 = 0.9364$ if a normal distribution of the sample mean was wrongly assumed, so it would underestimate the uncertainty by about 6.4% (2 s.f.)

- 1 point for correctly identifying that a z value is needed for this computation, and using either tables or a function like `norminv()` to find it.
- 1 point for using correct numerical value of z (1.96; 1.96 or 2 also acceptable for full credit)
- 1 point for looking at the ratio of the relevant z and t values to calculate the factor by which the confidence intervals would differ in size.
- 1 point for correct numerical evaluation of the ratio (or percentage change also acceptable). Accept answers with any number of s.f. in this case, and also give full credit to those that used z values of 1.96 or 2 as well as less-rounded values.

(f) [2 pt] Briefly comment on possible practical implications of assuming the normal distribution instead of the t-distribution for the sample mean when computing the confidence interval. To frame this question more concretely, imagine that the data represent the diameters of separate instances of a manufactured component that are designed to be integrated with a running clearance fit into a machine (recall E29 material).

Solution: Using the z distribution when not warranted will result in specifying a confidence interval that is narrower than it should be for the stated level of confidence. Consider the situation of using measurements of the diameters of a sample of manufactured components to decide whether the average diameter being manufactured falls within acceptable limits (often referred to as the 'tolerance zone', as written, e.g., on the working drawing). Consider two specific situations:

(a) The true population average diameter output by the production process lies just within the tolerance zone, but the sample mean for a given sample happens to lie 2.03 standard errors of the mean outside the tolerance zone due to production variability and/or measurement errors. If the CI is too narrow because the z distribution has been used, a 95% confidence interval would not overlap the specified tolerance zone, whereas the slightly wider CI given by the t-distribution would just overlap with the tolerance zone. This example shows how using an erroneously tight CI could lead to one concluding at a particular confidence level that a process was operating out of tolerance, when there actually wasn't evidence at that confidence level that it was out of tolerance.

(b) Conversely, if the true process (population) mean was actually out of the tolerance zone, but a sample mean happened to lie inside the tolerance zone because of part-to-part variability and/or measurement error, a CI that was too tight would give an excessive level of confidence that the process was operating, on average, within spec.

2 points for a reasonably clear articulation of the fact that using the z distribution when not justified could lead to erroneous conclusions, one way or another, depending on how the random variation of the data occurs. An explanation does not have to be as detailed as above to get 2 points; the general idea that one would have slightly too much confidence in one's conclusions is enough.

6 Uncertainty Propagation [12 pts]

For this problem, use the error propagation formula from lecture

$$u_y = \sqrt{\left(\frac{\partial y}{\partial x_1} u_1\right)^2 + \left(\frac{\partial y}{\partial x_2} u_2\right)^2 + \cdots + \left(\frac{\partial y}{\partial x_n} u_n\right)^2}$$

where y is a function of $x_1, \ldots, x_n$, and u_k represents the uncertainty of variable k. This formula assumes that the input variable uncertainties are small relative to the magnitudes of the variables, and that they are independent of each other.

(a) [4 pts] displacement sensor outputs a voltage, v, that is analogous to the displacement, x, it is measuring. To allow the user to determine the displacement based on the recorded voltage, a parabolic calibration relationship has been established:

$$x = 2\left[\frac{\mathtt{mm}}{\mathtt{V}^2}\right] \cdot v^2 + 3\left[\frac{\mathtt{mm}}{\mathtt{V}}\right] \cdot v + 3[\mathtt{mm}]$$

Suppose, for this part of the question, that the coefficients are exactly known and have no uncertainty. Calculate the uncertainty in displacement at $v = 2.00[\mathtt{V}]$, given that the uncertainty in the voltage measurement v is ± 0.25 [V] at 95% confidence.

Solution:

$$u_x = \frac{\partial x}{\partial v} u_v = \left(4[\frac{mm}{V^2}]v + 3[\frac{mm}{V}]\right) u_v = \left(4[\frac{mm}{V^2}] \times 2[V] + 3[\frac{mm}{V}]\right) \times (\pm 0.25[V]) = \pm\frac{11}{4}[mm] = \pm 2.75[mm]$$

- 6 points for a reasonably thorough attempt leading to a final answer, whether correct or not (noting that the numerical answer on its own does not get any credit).

(b) [5 pts] Now, assume that instead of using the parabolic calibration curve from previous part, a simplified calibration relationship is used, as follows:

$$x = A \cdot v + B$$

where $A = 2.00 \pm 0.12$ [mm/V] and $B = 7.00 \pm 0.15$ [mm] at 95% confidence. Given the same uncertainty in the voltage measurement as in the previous part, what is the uncertainty in the displacement at $v = 2.00$ [V] now? Note that in this part of the question, there is uncertainty in the coefficients A and B which must be taken into account.

Solution:

$$u_x = \sqrt{\left(\frac{\partial x}{\partial A} u_A\right)^2 + \left(\frac{\partial x}{\partial v} u_v\right)^2 + \left(\frac{\partial x}{\partial B} u_B\right)^2}$$

$$= \sqrt{\left[(2V) \times 0.12\,\frac{\mathrm{mm}}{V}\right]^2 + \left[2\,\frac{\mathrm{mm}}{V} \times 0.25V\right]^2 + [1 \times 0.15\,\mathrm{mm}]^2}$$

$$= \pm 0.57454\ldots\,\mathrm{mm} \;=\; \pm 0.57\,\mathrm{mm}\;\;(2\,\mathrm{s.f.})$$

Note that 2 s.f. are appropriate because the minimum number of s.f. in the input variables is 2.

- 6 points for a reasonably complete attempt (numerical answer only does not get credit).

(c) [3 pts] The voltage output of the differential pressure transducer depends on both air density ρ (kg·m^{-3}) and dynamic pressure q (N·m^{-2})

$$V = k\rho^{1/2}q^{1/2}$$

where k is a constant. If the relative uncertainty $u_\rho/\rho = 2\%$ and $u_q/q = 1\%$. Estimate the **relative uncertainty** in V i.e. u_V/V.

Solution: Starting with the error propagation formula, we have

$$u_V^2 = \left(\frac{\partial V}{\partial \rho}u_\rho\right)^2 + \left(\frac{\partial V}{\partial q}u_q\right)^2$$

Then, computing the partial derivatives leads us to

$$\frac{\partial V}{\partial \rho} = k \cdot \frac{1}{2}\rho^{-1/2}q^{1/2} = \frac{1}{2}k\rho^{-1/2}q^{-1/2} = \frac{1}{2}\frac{V}{\rho}$$
$$\frac{\partial V}{\partial q} = k\rho^{1/2} \cdot \frac{1}{2}q^{-1/2} = \frac{1}{2}k\rho^{1/2}q^{-1/2} = \frac{1}{2}\frac{V}{q}$$

Substituting into the propagation formula and divind both sides by V^2 yields

$$\begin{aligned} u_V^2 &= \left(\frac{1}{2}\frac{V}{\rho}\right)^2 u_\rho^2 + \left(\frac{1}{2}\frac{V}{q}\right)^2 u_q^2 \\ &= \frac{V^2}{4}\left(\frac{u_\rho^2}{\rho^2} + \frac{u_q^2}{q^2}\right) \\ \left(\frac{u_V}{V}\right)^2 &= \frac{1}{4}\left(0.02^2 + 0.01^2\right) \implies \frac{u_V}{V} = 1.1 = 1\% \quad 1 \text{ s.f.} \end{aligned}$$

7 Maximum Likelihood Estimators [14 pts]

The Maximum Likelihood Estimator (MLE) finds the model, or set of parameters, that maximizes the probability of the data. In other words, it maximizes the likelihood of some model θ given the data we obtain and seek to fit a model to. Recall for independent and identically distributed (i.i.d) random variables $X_1, \dots, X_n$ with probability mass function $f(x;p)$, the likelihood function is

$$\mathcal{L}(\theta;p) = \prod_{i=1}^{n} f(X_i;p) \implies \text{MLE is } \hat{\theta}_{\text{MLE}} = \arg\max \mathcal{L}(\theta;p)$$

The log-likelihood function is then

$$\ell(p) = \log \mathcal{L}(\theta;p) = \sum_{i=1}^{n} \log f(X_i;p)$$

The maximum likelihood estimator $\hat{p}$ is obtained by solving and verifying via the 2nd derivative that the solution below corresponds to a maximum

$$\frac{d}{dp}\ell(p) = 0$$

(a) [4 pts] Steph is running a durability test on a small internal combustion (IC) engine to quantify reliability under repeated firing cycles. For each engine, she operates it until the first failure event occurs (i.e. misfire that persists, loss of compression). She models the measured number of successful engine cycles until failure using a geometric distribution. She denotes by $1-p$ the probability that an engine completes a given cycle successfully and by p the probability that it fails on that cycle.

Given a random sample of n engines, $X_1, \dots, X_n$, find the MLE estimator for the parameter p in the geometric distribution. Remember to verify that the critical point is a maximum. Recall that the probability mass function of the geometric distribution is

$$f(x;p) = (1-p)^{x-1}p, \quad x = 1, 2, \dots$$

Solution: Likelihood is given by:

$$L(p) = \prod_{i=1}^{n}(1-p)^{x_i-1}p = (1-p)^{\sum_{i=1}^{n}(x_i-1)}p^n = (1-p)^{\sum_{i=1}^{n}x_i - n}p^n.$$

Log-likelihood is given by:

$$\log L(p) = \left(\sum_{i=1}^{n} x_i - n\right)\log(1-p) + n\log(p).$$

Differentiate with respect to p and set the derivative equal to zero:

$$\frac{\partial \log L(p)}{\partial p} = \left(\sum_{i=1}^{n} x_i - n\right)\frac{-1}{1-p} + n\frac{1}{p} = 0.$$

Multiplying through by $p(1-p)$ and rearranging

$$-\left(\sum_{i=1}^{n} x_i - n\right)p + n(1-p) = 0 \implies p\left(-\sum_{i=1}^{n} x_i\right) + n = 0 \implies p = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}.$$

The second derivative of the log-likelihood with respect to p is

$$\frac{\partial^2 \log L(p)}{\partial p^2} = \left(\sum_{i=1}^{n} x_i - n\right)\frac{-1}{(1-p)^2} - n\frac{1}{p^2},$$

which is negative because

$$\sum_{i=1}^{n} x_i - n \geq \sum_{i=1}^{n} 1 - n = 0.$$

Hence, the critical point is indeed a maximum. The maximum likelihood estimator for p is

$$\hat{p} = \frac{1}{\bar{x}}.$$

(b) [4 pts] Suppose that we now use 7 IC engines and test them on an engine dynamometer. For each engine, Steph recorded the number of successful firing cycles completed until the first failure event. The measured cycle-to-failure data were: $14, 17, 2, 3, 29, 25, 3$. Use the estimator in part (a), estimate p. Using the estimated value of p, estimate the probability that an engine breaks down during one of the first three cycles.

Solution: We first compute the sample mean:

$$\bar{x} = \frac{14 + 17 + 2 + 3 + 29 + 25 + 3}{7} = 13.29.$$

Using the estimator from part (i),

$$\hat{p} = \frac{1}{\bar{x}} = \frac{1}{13.29} \approx 0.075.$$

For a geometric random variable X, the probability that the device breaks down during one of the first three trials is

$$\mathbb{P}(X \leq 3) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3).$$

Using the geometric distribution with parameter $\hat{p}$,

$$\mathbb{P}(X \leq 3) = p + p(1-p) + p(1-p)^2 \approx 0.209.$$

(c) [2 pts] The Fisher information, $I(p)$ quantifies how much information a measured random variable carries about an unknown parameter. In the context of this experiment, it characterizes how informative each measured cycle-to-failure observation is about the true engine failure probability p. Find the Fisher information for a geometric random variable.

$$I(p) = -\mathbb{E}\left[\frac{\partial^2 \log f(X;p)}{\partial p^2}\right]$$

Solution: We have, since $\mathbb{E}[X] = \frac{1}{p}$,

$$I(p) = -\mathbb{E}\left[(X-1)\left(\frac{-1}{(1-p)^2} - \frac{1}{p^2}\right)\right] = \left(\frac{1}{p} - 1\right)\left(\frac{1}{(1-p)^2} + \frac{1}{p^2}\right) = \frac{1}{p^2(1-p)}$$

(d) [2 pts] Show that the sample mean $\overline{X} = \frac{1}{N}\sum_{i=1}^{n} X_i$ is a sufficient statistic.

Solution: Denoting the statistic $t = \bar{x}$ we may write the likelihood

$$L(p) = h(x_1, \ldots, x_n)\, g(t, p),$$

with $h(x_1, \ldots, x_n) = 1$ and

$$g(t, p) = (1-p)^{nt-n} p^n,$$

so by the Neyman–Fisher Factorisation Theorem it is a sufficient statistic.

(e) [2 pts] Show that a geometric random variable X has the memoryless property, which reflects the assumption that an IC engine's probability of failure in the next cycle is independent of how many cycles it has already completed. Mathematically, show

$$\mathbb{P}(X > m + n | X > n) = \mathbb{P}(X > m), \quad m, n = 0, 1, 2, \ldots$$

Solution:

$$\mathbb{P}(X > m + n \mid X > n) = \mathbb{P}(X > m), \qquad m, n = 0, 1, 2, \ldots$$

Solution: We have

$$\mathbb{P}(X > m + n \mid X > n) = \frac{\mathbb{P}(X > m + n)}{\mathbb{P}(X > n)} = \frac{\sum_{i=m+n+1}^{\infty} p(1-p)^{i-1}}{\sum_{i=n+1}^{\infty} p(1-p)^{i-1}}.$$

Evaluating each geometric series,

$$\sum_{i=m+n+1}^{\infty} p(1-p)^{i-1} = p(1-p)^{m+n} \sum_{k=0}^{\infty} (1-p)^k = p(1-p)^{m+n} \frac{1}{1-(1-p)},$$

and

$$\sum_{i=n+1}^{\infty} p(1-p)^{i-1} = p(1-p)^{n} \sum_{k=0}^{\infty} (1-p)^k = p(1-p)^{n} \frac{1}{1-(1-p)}.$$

Hence

$$\mathbb{P}(X > m + n \mid X > n) = \frac{p(1-p)^{m+n} \frac{1}{1-(1-p)}}{p(1-p)^{n} \frac{1}{1-(1-p)}} = (1-p)^m.$$

The right-hand side equals

$$\mathbb{P}(X > m) = \sum_{i=m+1}^{\infty} p(1-p)^{i-1} = p(1-p)^m \frac{1}{1-(1-p)} = (1-p)^m,$$

which finishes the proof.

8 Moment Generating Functions and Multivariate Gaussians [21 pts]

THIS ENTIRE QUESTION IS FOR EXTRA CREDIT - NOT ON MIDTERMS

Moment-generating functions provide a compact way to characterize the distribution of a random variable and to compute its moments (i.e. 1st central moment always 0, 2nd central moment is variance, 3rd is skewness, 4th is kurtosis etc.). In probalem (a), you will examine the moment-generating function of a zero-mean Gaussian random variable and show that it takes a particularly simple closed form. This result is fundamental in probability and statistics and highlights why the normal distribution plays a central role in modeling random phenomena. Recall that the MFG of a random variable X is defined as

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \int_{-\infty}^{\infty} e^{\lambda x} f_X(x) dx$$

where $f_X(x)$ is the probability density function of X. For a Gaussian random variable $X \sim \mathcal{N}(0, \sigma^2)$, the probability density function is given below. On the right, you may also find the following identity to be helpful:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad \int_{-\infty}^{\infty} \exp(-a(x+b)^2) dx = \sqrt{\frac{\pi}{a}}$$

(a) [7 pts] Prove that $\mathbb{E}[e^{\lambda X}] = e^{\sigma^2\lambda^2/2}$, where $\lambda \in \mathbb{R}$ is a constant, and $X \sim \mathcal{N}(0, \sigma^2)$. As a function of λ, $M_X(\lambda) = \mathbb{E}[e^{\lambda X}]$ is also known as the *moment-generating function*.

Solution: The pdf of a normal distribution with mean μ and variance σ^2 is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

However, since $X \sim \mathcal{N}(0, \sigma^2)$, we can simplify the pdf to

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Using the definition of the moment generating function and expected value

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

By manipulating the exp($\cdot$) argument and completing the square we have

$$\begin{aligned} M_X(\lambda) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x^2 - 2\sigma^2\lambda x)\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[(x-\sigma^2\lambda)^2 - \left(\frac{\cancel{2}\sigma^2\lambda}{\cancel{2}}\right)^2\right]\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[(x-\sigma^2\lambda)^2 - \sigma^4\lambda^2\right]\right) dx \end{aligned}$$

Expanding the exponent and factoring out the constant we get

$$M_X(\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\sigma^2\lambda)^2}{2\sigma^2}\right) \exp\left(\frac{\sigma^2\lambda^2}{2}\right) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{\sigma^2\lambda^2}{2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\sigma^2\lambda)^2}{2\sigma^2}\right) dx$$

An arbitrary Gaussian can be integrated to get the following solution according to Wikipedia

$$\int_{-\infty}^{\infty} \exp(-a(x+b)^2)dx = \sqrt{\frac{\pi}{a}}$$

So, for our integral, $a = \frac{1}{2\sigma^2}, b = -\sigma^2\lambda$ so

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\sigma^2\lambda)^2\right) dx = \sqrt{\pi \cdot 2\sigma^2}$$
$$= \sigma\sqrt{2\pi}$$

Substituting this back into the MGF we get

$$M_X(\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{\sigma^2\lambda^2}{2}\right) \cdot \sigma\sqrt{2\pi}$$
$$= \exp\left(\frac{\sigma^2\lambda^2}{2}\right)$$

Therefore, $M_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \exp\left(\frac{\sigma^2\lambda^2}{2}\right)$

The multivariate normal distribution with mean $\mu \in \mathbb{R}^d$ and positive definite ($\Sigma \succ 0$) covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, denoted $\mathcal{N}(\mu, \Sigma)$, has the probability density function

$$f(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

Here $|\Sigma|$ denotes the determinant of Σ. You may use the following facts without proof.

- The volume under the normal PDF is 1.

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right\} \mathrm{d}x = 1.$$

- The change-of-variables formula for integrals: let f be a smooth function from $\mathbb{R}^d \to \mathbb{R}$, let $A \in \mathbb{R}^{d\times d}$ be an invertible matrix, and let $b \in \mathbb{R}^d$ be a vector. Then, performing the change of variables $x \mapsto z = Ax + b$,

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(A^{-1}z - A^{-1}b)\,|A^{-1}|\,\mathrm{d}z.$$

(a) [7 pts] Let $X \sim \mathcal{N}(\mu, \Sigma)$. Use a suitable change of variables to show that $\mathbb{E}[X] = \mu$.

Solution:

The expected value for a multivariate normal distribution is defined as

$$\mathbb{E}[X] = \int_{\mathbb{R}^d} x f(x; \mu, \Sigma) dx$$

let $z = x - \mu \Rightarrow x = z + \mu$, then the PDF of the MVN is

$$f(z+\mu; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(z+\mu-\mu)^\top\Sigma^{-1}(z+\mu-\mu)\right)$$
$$= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}z^\top\Sigma^{-1}z\right)$$

Given our transformation $x \mapsto z = Ax+b$ is affine, where $A = I_d$ and $b = -\mu$, then $A^{-1} = I^{-1} = I$ so

$$\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} f(I_d^{-1}z + I_d^{-1}\mu)|I_d^{-1}|dz = \int_{\mathbb{R}^d} f(z+\mu)dz$$

Then the expected value becomes

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\mathbb{R}^d} (z+\mu) f(z+\mu;\mu,\Sigma)dz \\
&= \int_{\mathbb{R}^d} z f(z+\mu;\mu,\Sigma)dz + \mu \int_{\mathbb{R}^d} f(z+\mu;\mu,\Sigma)dz \\
&= \int_{\mathbb{R}^d} z \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2} z^\top \Sigma^{-1} z\right) dz + \mu \int_{\mathbb{R}^d} \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2} z^\top \Sigma^{-1} z\right) dz
\end{aligned}$$

Notice that the second integral is just the volume under the normal PDF which is 1. The first integral is z multiplied by the PDF which is even in z. So multiplying an odd function by an even function which is odd. Therefore, the integral should just evaluate to 0, hence

$$\mathbb{E}[X] = 0 + \mu(1) = \mu$$

(b) [7 pts] Use a suitable change of variables to show that $\mathrm{Var}(X) = \Sigma$, where the variance of a vector-valued random variable X is

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X) = \mathbb{E}[(X-\mu)\,(X-\mu)^\top] = \mathbb{E}[XX^\top] - \mu\,\mu^\top.$$

Hints: Every symmetric, positive semidefinite matrix Σ has a symmetric, positive definite square root $\Sigma^{1/2}$ such that $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$. Note that Σ and $\Sigma^{1/2}$ are invertible. After the change of variables, you will have to find another variance $\mathrm{Var}(Z)$; if you've chosen the right change of variables, you can solve that by solving the integral for each diagonal component of $\mathrm{Var}(Z)$ and a second integral for each off-diagonal component. The diagonal components will require integration by parts.

Solution:

Given the pdf of X as

$$f(x;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

Let's use the following change of variables $z = \Sigma^{-1/2}(x-\mu)$. Given our transformation $x \mapsto z = Ax+b$ is affine, where $A = \Sigma^{-1/2}$ and $b = -\Sigma^{-1/2}\mu$, then using the formula

$$\begin{aligned}
\int_{\mathbb{R}^d} f(x)dx &= \int_{\mathbb{R}^d} f(A^{-1}z - A^{-1}b)|A^{-1}|dz \\
&= \int_{\mathbb{R}^d} f(\Sigma^{1/2}z + \Sigma^{1/2}\Sigma^{-1/2}\mu)|\Sigma^{1/2}|dz \\
&= \int_{\mathbb{R}^d} f(\Sigma^{1/2}z + \mu)\sqrt{|\Sigma|}dz
\end{aligned}$$

This shows that $x = \Sigma^{1/2}z + \mu$. Now let's rewrite the pdf with the change of variables

$$\begin{aligned}
f(x;\mu,\Sigma) &= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) \\
f(\Sigma^{1/2}z+\mu;\mu,\Sigma) &= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(\Sigma^{1/2}z+\mu-\mu)^\top \Sigma^{-1}(\Sigma^{1/2}z+\mu-\mu)\right)\sqrt{|\Sigma|} \\
&= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}(z^\top \Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}z)\right)
\end{aligned}$$

$$f_Z(z) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2} z^\top z\right) \Rightarrow Z \sim \mathcal{N}(0, I)$$

Then since $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)(X-\mu)^\top]$, notice how $X - \mu = \Sigma^{1/2} z$ we can write

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(\Sigma^{1/2} z)(\Sigma^{1/2} z)^\top] \\ &= \Sigma^{1/2} \mathbb{E}[z z^\top] \Sigma^{1/2} \end{aligned}$$

Here $\mathbb{E}[zz^\top] = \mathrm{Var}(zz^\top)$, and we calculate it using 2 integrals for the diagonal and off-diagonal terms. Starting with the diagonal terms

$$\mathbb{E}[z_i^2] = \int_{\mathbb{R}^d} z_i^2 f_Z(z) dz_i = \int_{\mathbb{R}^d} z_i^2 \frac{1}{\sqrt{2\pi}} \exp\left(-z_i^2/2\right) dz_i = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^d} z_i^2 \exp\left(-z_i^2/2\right) dz_i$$

Integrating by parts with $u = z_i \Rightarrow du = dz_i$ and $dv = z_i \exp(-z_i^2/2) \Rightarrow v = -\exp(-z_i^2/2)$

$$\int_{\mathbb{R}^d} z_i^2 \exp(-z_i^2/2) dz_i = -z_i \exp(-z_i^2/2) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp(-z_i^2/2) dz_i = \sqrt{2\pi}$$

Therefore, $\mathbb{E}[z_i^2] = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1$ For the off diagonal components, the integrand is odd so integrating yields 0. The Gaussian integral was evaluated using the same formula as above

$$\mathbb{E}[z_i z_j] = \int_{\mathbb{R}^d} z_i z_j f_Z(z) dz = 0$$

Combining both, $\mathbb{E}[zz^\top] = I$ so $\mathrm{Var}(X) = \Sigma^{1/2} \mathbb{E}[zz^\top] \Sigma^{1/2} = \Sigma^{1/2} I \Sigma^{1/2} \Rightarrow \mathrm{Var}(X) = \Sigma$

