ME 103 Experimentation and Measurements
Spring 2026 George Anwar
HW1

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:

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

Solution:

(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:

(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:

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: TODO

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

Solution: TODO

(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: TODO

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

Solution: TODO

(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: TODO

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: TODO

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

Solution: TODO

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

Solution: TODO

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

Solution: TODO

(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: TODO

(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: TODO

(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: TODO

(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: TODO

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				
2				
3				

Table 1: Summary of sample statistics for three data sets

(b) Calculate the following quantities

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

Solution: TODO

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

Solution: TODO

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

Solution: TODO

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

Solution: TODO

(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: TODO

(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: TODO

(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: TODO

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:

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

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

Solution: TODO

(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: TODO

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

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

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: TODO

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 $X_1, \ldots, X_n$ 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

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

(c) [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, $X_1, \ldots, X_n$, 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

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

Solution: TODO

(d) [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: TODO

(e) [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: TODO

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

Solution: TODO

(g) [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: TODO

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: TODO

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: TODO

(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: TODO