ME 103 Discussion 3

Week of 2/2

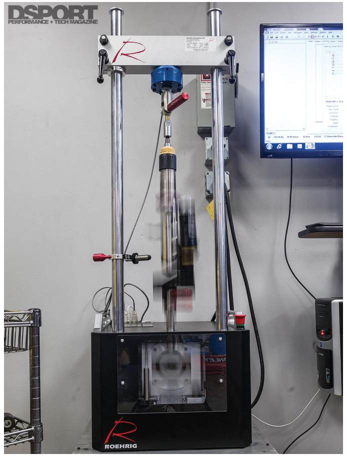
Announcements

- Fill out the Lab Safety Form before lab next week.
- Continuing our work on HW 1, problem 4, 5, 6, 7
    - Due Feb 9, 11:59 PM
    - Larry has posted a video covering problem 4d and 4e.

Shock Dyno

BIG PICTURE: This problem is asking you to compare two measurements from a shock test and find the chance that one total is smaller than the other, using the Central Limit Theorem.

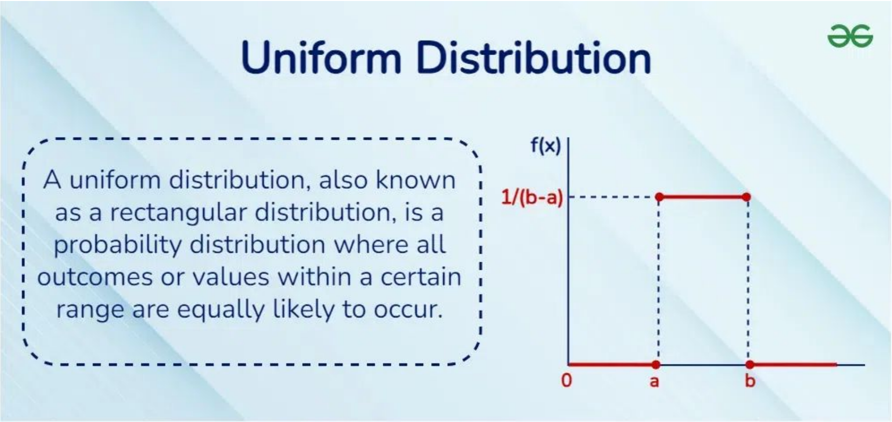
1. In Run A, each compression stroke gets an integer level from 1 to 6, equally likely, and you sum these 100 levels to get a total force code X.
2. In Run B, over 600 time windows, each window either detects an event (1) or not (0) with probability 0.5, and you sum the 600 flags to get a total event count Y.



Approximate the distributions of X and Y as normal (via CLT) and use that to estimate the probability that the total coded force from Run A is less than the total event count from Run B, i.e. P(X<Y). For CDF purposes, this can be rewritten as Z = X-Y and P(Z) < 0.

Shock Dyno - Run A

- Let's call X the random variable for Run A.
- Just like any other probability distribution problem, need to convert to **z scores and find CDF.**
- X totals 100 independent force readings from compression strokes. Each stroke gets binned into one level from 1 to 6, equally likely.
    - **6 numerical bins, equally likely outcomes → a discrete uniform distribution.**
    - Find E[x], or the mean of one stroke.
    - Find var of x.
- Then find the Expected value of the TOTAL force. For 100 trials,
    - E[Total] = 100 * E[x]
    - Var[Total] = 100 * Var[x]



<u>Discrete Uniform</u>: $X \sim \text{Uniform}(n)$

$$p_x = \frac{1}{n} \, (x = 1, 2, \ldots, n) \quad \mu = (n+1)/2, \; \sigma^2 = (n^2 - 1)/12$$

Linearity of expectation: $\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$

Shock Dyno - Run B

- Y totals the number of "event detected" flags across 600 independent time windows. Each window has two outcomes: 1 (event, p=0.5) or 0 (no event),
    - This indicates **Binomial Distribution.**
- Find E[Total] and Var[Total]:
    - Find mean and variance, remember var * 600 trials = total variance.
- **Define random variable Z = X - Y.**
    - Find E[Z] = E[X] -E[Y] (linearity)
    - Var[Z] = Var[X] + Var[Y]. Cov = 0 for independent variables.

Binomial distribution equations

$$\text{Mean} = \mu = E(x) = np$$

$$\text{Variance} = \sigma^2 = np(1-p)$$

$$\text{Standard Deviation} = \sigma = \sqrt{np(1-p)}$$

where

n = number of trials

p = probability of success

$1\text{-}p$ = probability of other outcome (failure)

$$\mathbb{E}[\cdot] \text{ is a linear operator: } E[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X,Y)$$

- CLT: Both X and Y sums are large → Z ≈ Normal(μ=E[Z], σ=√Var(Z)).
- P(X < Y) = P(Z < 0)
    - Convert 0 to z score and use CDF/z table to find (0 - E[Z])/√Var[Z]

Confidence Intervals

- Helpful matlab commands:
  - mean([mean1 mean2 mean3])
  - If you had an array named "data"
    - data(:,1);
      - all rows, first col.
    - data(1:14,2);
      - rows 1-14 of the second col

Question: What are the drawbacks of wrongly assuming normal distribution with a sample size less than 30?

T- distribution confidence interval: for random variable X with mean $\boldsymbol{\mu}$ and variance σ^2, a (1-α)*100% confidence interval is:

$$\bar{x} \pm t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

- $\bar{x}$ is the estimate of the value you want the confidence interval to be around
- t* is the t-score for α/2 significance and n-1 degrees of freedom
- s= sample std. dev

Error/ Uncertainty Propagation

Big Picture: When measured values multiply and/or divide, **fractional uncertainties** don't add linearly. Instead, add their squares, then take the square root.

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

$$V = lwh$$

$$u_V = \pm \sqrt{\left(\frac{\partial V}{\partial l} u_l\right)^2 + \left(\frac{\partial V}{\partial w} u_w\right)^2 + \left(\frac{\partial V}{\partial h} u_h\right)^2}$$

$$\frac{\partial V}{\partial l} = wh \; ; \quad \frac{\partial V}{\partial w} = lh \; ; \quad \frac{\partial V}{\partial h} = lw$$

$$u_V = \pm \sqrt{(wh\,u_l)^2 + (lh\,u_w)^2 + (lw\,u_h)^2}$$

← Here is a familiar example, with Volume = lwh. Uncertainties from length, width, and height all "propagate" through the volume uncertainty with respect to their fraction of the total volume.

Maximum Likelihood Estimators

Big Picture: I'm running a failure test on an IC Engine. I am counting how many firing cycles it lasts until failure. **Counting how many trials until the** first **success is modeled with a geometric distribution.**

In each cycle, the engine either fails (bad event) with probability **ρ**, or survives with probability **1−ρ.**

What is a MLE?

A statistical model says: "If the true parameter were ρ, the data would follow this distribution."

The likelihood reverse engineers this: for the fixed data you already observed, how likely is this data if the parameter were ρ?"

The MLE ρ^ is the value of ρ that maximizes that likelihood function.

Maximum Likelihood Estimators

Big Picture: I'm running a failure test on an IC Engine. I am counting how many firing cycles it lasts until failure. **Counting how many trials until the** first **success is modeled with a geometric distribution.**

In each cycle, the engine either fails (bad event) with probability **ρ**, or survives with probability **1−ρ.**

What is a MLE?

A statistical model says: "If the true parameter were ρ, the data would follow this distribution."

The likelihood reverse engineers this: for the fixed data you already observed, how likely is this data if the parameter were ρ?"

The MLE ρ^ is the value of ρ that maximizes that likelihood function.

Part c: Choose p^ that maximizes the log of L(p).

1. Start with L(p). Take the log of this function to make the math easier.
2. $\prod$ is like sum, but for multiplication.
3. Differentiate with respect to p and set derivative = 0
4. Rearrange and solve for p
5. This will give you one critical point (which we know to be the maximum)
6. You will most if not all the properties to the right.
7. Keep in mind the definition of the sample mean (equation 4)

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

Some Math rules you may have forgotten:

① $\prod_{i=1}^{n} a^{x_i} = a^{\sum_{i=1}^{n} x_i}$

② $\prod_{i=1}^{n} b = b \cdot b \cdot b \cdots = b^n$

③ $\log(a^b) = b \log a$

④ $\bar{x}$ (sample mean) $= \dfrac{\sum_{i=1}^{n} x_i}{n}$

⑤ if c is a const: $\sum_{i=1}^{n} c = nc$

⑥ $\sum_{i=1}^{n} (x - y) = \sum_{i=1}^{n} x - \sum_{i=1}^{n} y$

Part d: using the estimator on example data.

1. Use the p^ from MLE part c)
2. With that p^, compute probability an engine fails within first 3 cycles.

$P(X \le 3) = P(X=1) + P(X=2) + P(X=3)$.

- $P(X=1) = (1-p)^0 p = p$
- And so on for X=2 and X=3.

Use solved p^ parameter and plug into equation

Part e): Fisher Information

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

1. Take the second derivative of the log of f(x;p)
2. Then find the expected value.
   a. Wolfram, matlab, desmos, etc

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

Larry Time

Larry’s Supplemental Video :p