Discussion Boards as a Learning, Reflection, and Community Building Tool In Large Statistics Courses

Karen Huynh Wong

Assistant Professor, Teaching Stream

Department of Statistical Sciences

University of Toronto

Content & Context

- Discussion boards and their use to enrich student learning, and support constructivist to teaching
- Motivation and challenges addressed with discussion board activities (DBA)
- Structure and student buy-in
- DBA compared with Piazza for open discussions
- Student exemplars

- Large enrolment introductory courses in probability (12 weeks) and statistics (12 weeks)
- Single section of 400 students, two sections of 240 students
- In-person, online, and hybrid models

Pedagogical Motivation

Tackling the following challenges:

- Lack of student engagement and participation
- Insufficient practice in statistical thinking (per GAISE (2016) recommendations)
- Accessibility
- Flexibility of contact time
- Long queues
- Moderation difficulties

DBA Structure

- Weekly discussion boards per course section
  - General Q&A of weekly topic
  - Separate discussion boards for weekly practice problems
  - Separate board of mixed practice problems during midterm and final exam season
- Discussion boards are ”live” for 1 week where posts are eligible for credit
- Discussion boards remain open and moderated after 1 week for follow-up discussions/questions, but are ineligible for credit
- Student buy-in:
  - Equating to rubber duck debugging for CS students,
  - Transparent conversation on first day on decision and benefits of DBA,
  - Grade incentive: 4-8% of course grade, each post eligible for up to 1%
  - Emphasizing expectations of online etiquette

DBA Rubric

Posted with each discussion board

Points	1 point	0.5 points	0 points
Quality of contribution	Student has made a substantial and unique contribution with detailed explanations and/or clearly outlined process of their approach to a problem. AND/OR Student was involved in follow-up discussions and worked collaboratively with their peers to develop a better understanding of the concepts discussed.	Student has made a contribution that is lacking in detail, not completely unique, or is dismissive. Unable to further the discussion in a way that fosters a collaborative learning environment. e.g., responses such as ‘you just need to integrate this and solve for it’ or ‘I got the same answering doing… (reiterating OP’s process with minimal changes)’	Student has not contributed to the weekly discussion board, or whose posts are off-topic/irrelevant/do not contribute to the discussion OR is not unique to what has already been discussed in the thread. e.g., ‘I got the same answer’, ’How did you get that number’, 'I got the same answering doing… (reiterating OP’s process)’

DBA vs Open Piazza Discussion

Intro Probability - Fall 2021

- 380 students
  - 88% made at least one contribution
  - 59% of students earned 4 points (or more)
  - 11% of students earned 3-4 points
  - 17% earned 1-3 points
  - 45 students did not participate at all (7 of which didn't engage at all in the course)
  - 27% of students far exceeded the required 4 posts
  - More than half of posts were responded by students
- 1738 points earned in total by students (avg: 4.57 points/student)

Intro Probability - Fall 2019

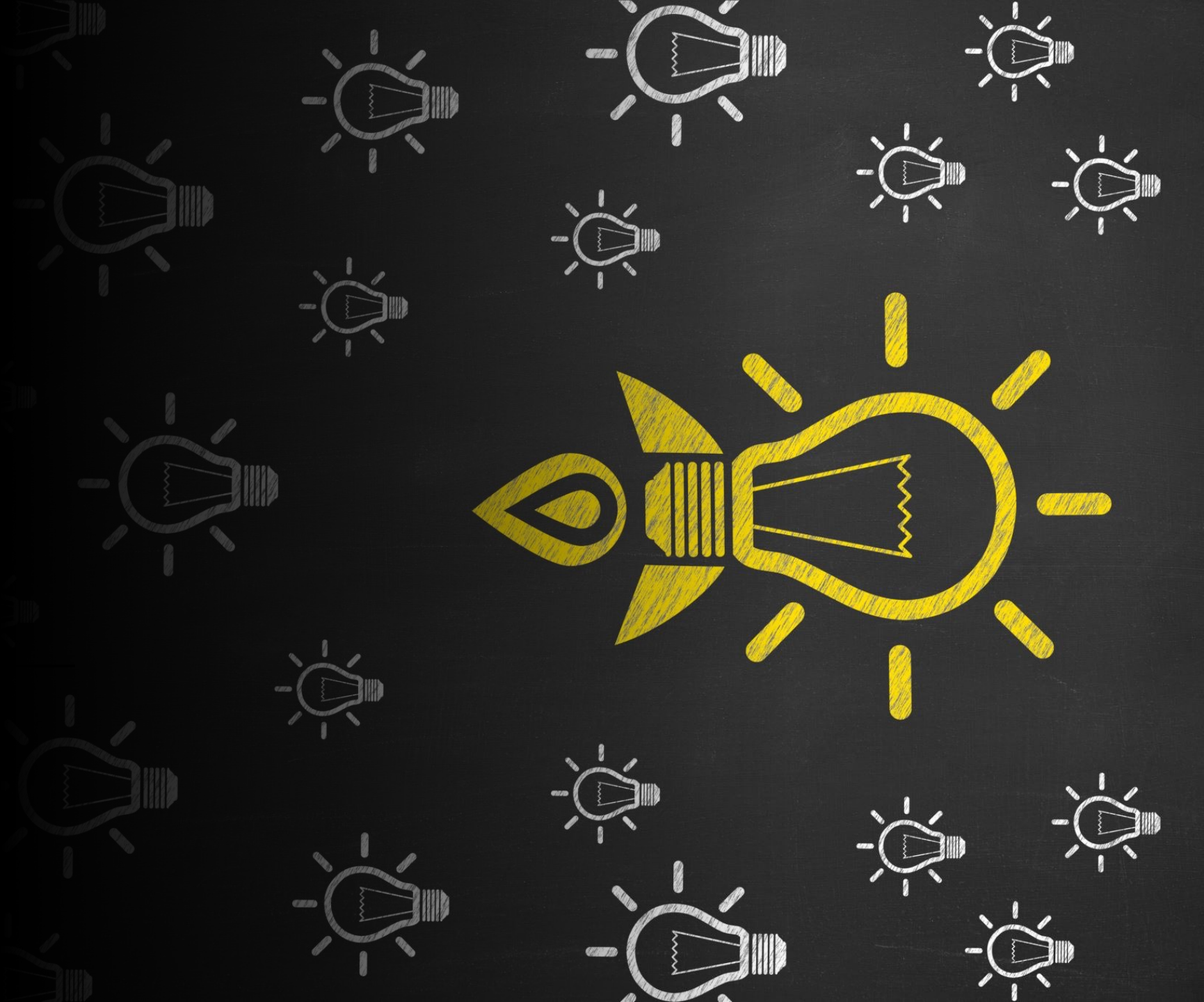
- 288 students
  - 28% made at least one contribution
  - Contributions: posts, responses, edits, follow-ups, comments
  - 781 total contributions
- 239 questions asked by students
  - 86% were responded by instructor/TAs
  - 18% were responded by other students
- 219 instructor responses

Student Feedback

- In institutional course evaluations, a greater number of students specifically name discussion boards and their function as a source of learning support in the class vs when using Piazza, comments were often about response time
- Far fewer students perceiving DBA as additional and unhelpful work than those perceiving DBA to be helpful in their studies from course evaluations and student dialogue
- Students spoke positively on evaluations about:
  - Timely and supportive responses from instructors and TAs
  - Opportunity for open discussion with peers
  - Space to share and discuss answers primarily through peer contributions
  - Being able to assess their own work with others

Thank you!

Creative Contributions



Statistical Sciences
UNIVERSITY OF TORONTO

Sep 23, 2022 9:53pm

Probably one of my favorite questions,

To start off from the basics, we know $n(\Omega) = P(n, k)$. Fairly straightforward.

The trouble is that $n(\text{increasing order})$ cannot be easily found as, the most straight forward method, would be to choose a value and then figure out the number of cases from there. A method that, given the question, is impossible as n and k are arbitrary.

So, taking a step back, permutations would not work since, as mentioned earlier, there would be impossible number of considerations to account for.

then how would $n(\text{increasing order})$ be determined for an arbitrary n and k?

Consider, combinations, more specifically what they represent. Since we're following a very strict order of increasing numbers, then we can ignore the conditions of which numbers are chosen to focus on the numbers themselves. Or, in another way to look at it, imagine that all the numbers are always shifted into order. If that's the given case, then we'll only need to focus on the combinations as we can, effectively, throw ordering out the window.

(Source: Tenor.com)

and after chucking order (Dr. Fate, calm down) out of the equation, we are now left with figuring out the number of combinations that can be shifted to increasing order, meaning that $n(\text{increasing order}) = \binom{n}{k}$

thus we can reach the conclusion that: $P(\text{increasing order}) = \frac{\binom{n}{k}}{P(n,k)}$

Feel free to correct me on anything or, if your confused, ask questions or clarifications.

Dialogue & Self-Assessment

Statistical Sciences
UNIVERSITY OF TORONTO

Jasmine Liu

Feb 22, 2022 5:58pm Last reply Feb 27, 2022 4:01pm

Estimator analysis - Unbiasedness, Consistency, Efficiency, Variability

I still don't quite understand the differences (and perhaps similarities) between these concepts, so I'm hoping that someone could ***explain their differences*** to me. More specifically:

1) What is the difference between ***consistency*** and ***variability*** and ***efficiency***?

2) Is efficiency basically variability, but we use the term ***"efficiency"*** when we ***compare the variability*** of two estimators that we know are ***unbiased***?

3) Does ***"consistency" incorporate both variability and unbiasedness***? i.e. Does an estimator have to be ***unbiased and non-variable(variability=0)*** to be consistent? I am confused because in the example problems we always prove unbiasedness before consistency, so I thought maybe unbiasedness is one of the premises of consistency.

Here are some understandings that I already have, and I hope someone could tell me ***if I am correct about them***.

4) There is a ***"yes or no"*** answer to "unbiasedness" and "consistency". E.g., if $E(T) = \theta$, the estimator is unbiased, otherwise it is biased. There is not a "scale" of unbiasedness and consistency.

5) Whereas, "variability" and "efficiency" can ***range*** from 0 to infinity.

Reply | | 3 Replies

YX

Yushan Xie

Feb 22, 2022 10:33pm

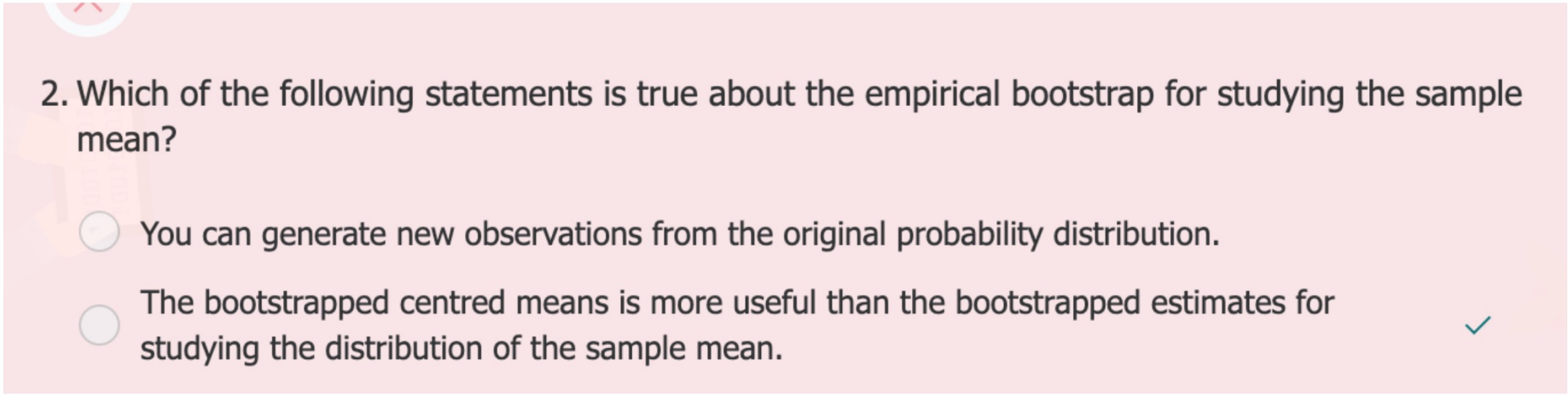
Hello Jasmine, I have some thoughts for your first question:

Firstly, variability refers to the divergence of data from its mean value. For example, a simple measure of variability is the range, the difference between the highest and lowest scores in a set. Secondly, consistency means a statistic tends toward the parameter it was supposed to be estimating as power increase. And if a statistic does not estimate the population parameter in the long run, it cannot be used much at all.
For efficiency, a measure of efficiency is the ratio of the theoretically minimal variance to the actual variance of the estimator. The most efficient estimator among a group of unbiased estimators is the one with the smallest variance.

Hope this helps

Reply |

I am a little confused about option 2 of this question in video 4A. Does it mean those bootstrapped centred means are useful because we could use them to infer the parameter of the population? And bootstrapped estimates just refer to each test statistic of each bootstrap sample?



Reply | | 1 Reply

ZM Zilan Mo

Feb 7, 2022 10:42am

Hi [redacted]

You are correct in that bootstrapped estimates correspond to each bootstrap sample, however, these bootstrapped centered means are used to make inference about the distribution of the sample mean (if we had many different samples, then we can use their distribution to infer about the population). Based on my understanding, since bootstrap samples contain different sets of observations from the original sample, their estimates vary greatly (they can be closer to or further away from the sample mean). The bootstrapped centered means gives the distribution of these estimates and shows how they differ from the sample mean. If we plot these estimates, we would expect for the values to be centered close to the sample mean. We can compare bootstrap estimate and sample mean to the sample mean ad population mean, that is, the deviation or $\delta^* = \overline{x_n}^* - \mu^*$ should be approximately the same as $\delta = \overline{x_n} - \mu$ (deviation of sample mean from population). I'm not sure if I explained the concept clearly, please correct me if I'm wrong.

Reply |

DK

Dohyun Kim

Mar 18, 2022 5:12pm Last reply Mar 21, 2022 11:53pm

From my understanding, α represents the rate of incorrectly rejected true (type 1 error), β represents the rate of incorrectly accepted false (type 2 error), $power$ represents the rate of correctly rejected false, and n represents the sample size.

a) From this week's lecture (L0201 Week 9), the late slide states:

$\alpha \downarrow \Rightarrow \ \beta \uparrow \Rightarrow \ power \downarrow$.

Since less false hypothesis are rejected, more hypothesis will be accepted. Then, more false hypothesis will be accepted, which will result less false hypothesis will be rejected.

Is this explanation correct? Feel free to add more details to this brief explanation.

b) Also from the last slide, it states:

$n \uparrow \Rightarrow \ \beta \downarrow$, when α stays the same.

However, I cannot see why the increase in sample size will result less incorrectly accepted false hypothesis. Can someone explain this? Thank you.

Reply | | 3 Replies

ZM

Zilan Mo

Mar 18, 2022 5:55pm

Hi, I was a bit confused as to whether there is always an inverse relationship between α and β. I'm guessing that if we are trying to understand the errors that could occur in hypothesis testing, there would only be two options (otherwise it is that the test provides correct outcomes). Hence, if the probability of one error increases then the other should decrease. With regards to your last question, I think the reason could be that, according to the LLN, as the sample size n increases, sample values grow closer to the expected value. Hence we are less likely to make an error in our assessment since our claims contain known values for H_0 and H_A. I'm not sure if this is true, correct me if I'm wrong.

Reply |

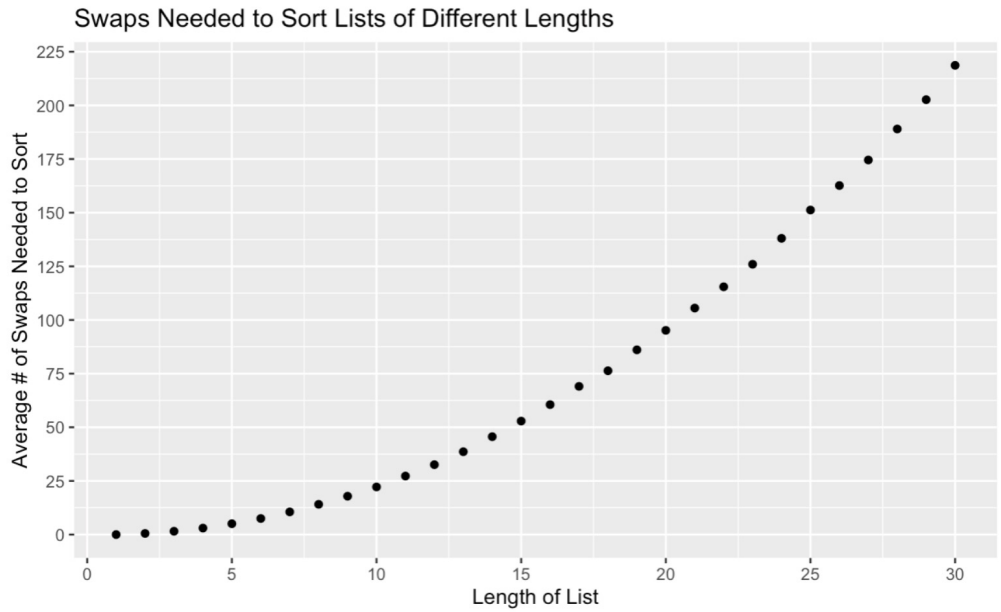
Opportunity for Experimenting

JC

Jamie Chew

Oct 27, 2022 10:09am Last reply Oct 31, 2022 10:13am

Hi, so after class, I was curious about how many swaps it takes to sort a list of length n with an insertion sort, so I made a graph below modelling it. I was expecting the relationship between list length and number of swaps to be quadratic since the efficiency of an insertion sort is $O(n^2)$, and I think that's shown pretty well in my graph. Basically what I did was I took 1000 samples for each list length (2 to 30 in this case) and averaged the number of swaps it took to sort them. I also have the code below the graph that's a bit messy, but I left it here just in case anyone wanted to see. I was also wondering, does anyone know how to do a quadratic regression for something like this?




```
#upper bound of x-axis
maxListLength = 30
listsPerSample = 1000
#The number of different values a list can contain (i.e., each list samples numbers from 1 to 1000 without replacement)
sampleSize = 1000
sampleList = c()
avgSwaps = c(0)
listLengths = c(1)

for(i in 2:maxListLength){
  for(j in 1:listsPerSample){
    currList = sample(1:sampleSize, i)
    sampleList[j] = insertionSort(currList)[1]
  }
  avgSwaps[i] = mean(sampleList)
  listLengths[i] = i
}

#create a table with the data
avgSwapsTable = tibble(listLength = listLengths, meanSwaps = avgSwaps)

#plot the table
ggplot(avgSwapsTable)+
  geom_point(aes(x=listLength, y=meanSwaps))+
  scale_x_continuous(breaks = seq(0, max(avgSwapsTable$listLength), by = 5))+
  scale_y_continuous(breaks = seq(0, (ceiling(max(avgSwapsTable$meanSwaps)/25))*25, by = 25))+
  labs(x="Length of List",
       y="Average # of Swaps Needed to Sort",
       title="Swaps Needed to Sort Lists of Different Lengths")
```

Reply | 1 | 2 Replies

Karen Huynh Wong

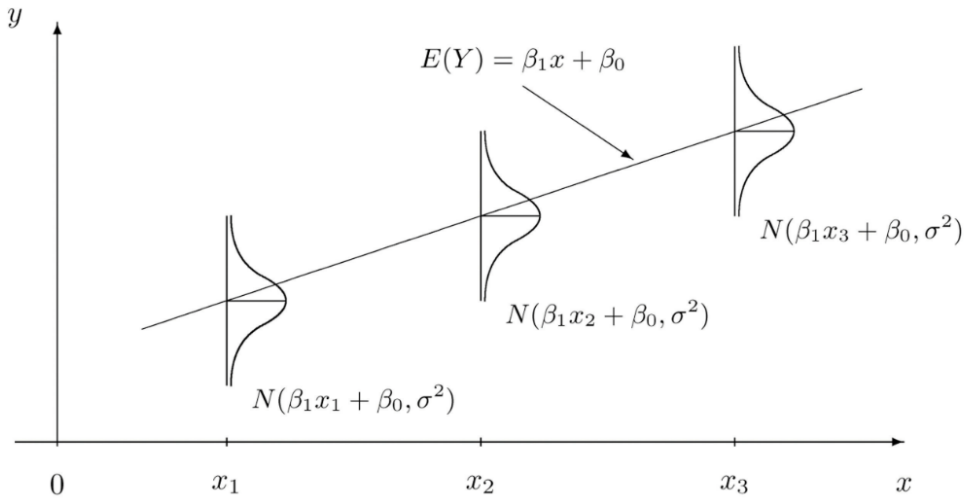
Teacher

Oct 28, 2022 6:23pm

Hey! This is pretty cool way to take the simulation and tie it back to the efficiency of the sorting method. There are quadratic regressions, but that's more STA302 level but you can start to play around with the idea once you learn simple linear regression. It's actually pretty cool to think of regression in this application because you've plotted the mean number of swaps for each list size, because the basis of regression is that we model the **average of one variable** as some linear function of another. In this case, a quadratic regression model would try to find some function in the form:

$E[Swaps] = a*list.size^2 + b*list.size + c$

Rationale being that at any given list size, the number of swaps required to fully sort is random with its own distribution (like all the simulated histograms we generated in class) BUT when you plotted the mean number of swaps, we see that there's a general trend: the average number of swaps increases quadratically to the list size. You can almost imagine at every point along your graph, there is a mini distribution around each point representing all the possible random # of swaps for each given list size. A bit hard to describe over a textual discussion post, so maybe a visual might help, in place of a straight line you have a parabola, and each xi is a list size, and # of swaps is on your y-axis:



Modelling Linear Relationships with Randomness Present: https://saylordotorg.github.io/text_introductory-statistics/s14-03-modelling-linear-relationships.html

Reply | 👍

JC

Jamie Chew

Oct 31, 2022 10:13am

Oh wow that's really cool! I wasn't really thinking about that visualization of distributions when I was making the plot, but yeah it makes a lot of sense looking at it. I might consider taking STA302 :)

Reply | 👍

Course Evaluation – Student Comments

The tutorials and the discussion boards were very effective at reinforcing all the concepts introduced in class. The discussion boards filled all the missing gaps from assigned work, and the tutorials took a more in depth approach of what was introduced in the lectures.

Discussions boards were the most helpful to clear any doubts relating to the course concepts and textbook problems.

Discussion boards were great tools, allowed open discussion amongst students regarding course content and practice questions.

The main assistance came from discussion boards which were monitored by Prof. Wong, but mostly contributed by students. It was a great way to get help and share answers, especially during online learning.

The discussion boards were helpful to check over my answers (especially for the PAR textbook, which had abysmal solutions).

Discussion boards were very helpful

Discussion groups were excellent. I used them a lot, especially for midterm prep.

* Discussion board was great for questions. I didn't ask many qusetions myself, but it was very useful for me to go through the questions of my peers, and it helped clear up a lot of my own questions.