Econ 488, First problem set

2025-01-10

- **Objective:** The purpose of this problem set is to review background knowledge that you (may) have covered in prerequisite classes. If a concept is foreign to you (that is: **after** you have refreshed your memory from a previous class), please indicate this as your response. For example, you could write "No previous class introduced this concept''. You will not lose points if you fail to answer a question for this reason.
- **Due Date:** 11:59pm on January 16. Submission should be via gradescope (accessible via canvas).
- Other instructions:
  - You should work individually on this problem set.
  - You are welcome to practice typewriting the solutions in LaTeX or Rmarkdown to see which option you like best, or whether typewriting is a good choice for you. I have made templates to get you started (See the markdown template file).
  - Unless otherwise stated, please justify your answer. Also, please limit each response to at most two sentences.
  - For questions where I ask for a ''definition'' or ''formulation'', please provide a mathematical definition if you can. Otherwise, try to provide a concise definition in words; or possibly a representative example. I will grade these questions very generously, please just try your best so that I can understand your backgrounds.

Questions:

1. Give a concise definition of a "random sample'' (otherwise known as an *iid* sample).
2. Consider an experiment in which two Amazon.com customers are selected at random from the population of Amazon customers and their total spending during the month of December is measured. Let Y_i denote a random variable that returns the spending of customer i. Is the collection $\{Y_1, Y_2\}$ a random sample? What about $\{Y_1, Y_{10}\}$?
3. Consider the generation of a dataset, drawn randomly from the population of Amazon customers on Amazon.com during themonth of December is measured. Suggest a discrete random variable and a continuous random variable associated with this dataset. No justification is necessary.
4. Let $X \sim f(\cdot)$ with $f(\cdot) = \mathcal{N}(5, 1)$, where $\mathcal{N}(\cdot, \cdot)$ denotes the Normal distribution (also called Gaussian). Let $Z \sim p(\cdot)$ with $p(1) \equiv Pr(Z = 1)$, $p(0) \equiv Pr(Z = 0)$, and $p(1) + p(0) = 1$ (this pmf is called Bernoulli). What is the mean or expected value of X? And that of Z?
5. Consider a random sample $(Y_1, ..., Y_n)$ drawn from the pdf $f(\cdot)$, where n is an integer number. Suppose that this pdf has a *parametric form* i.e. $f(y) = g(y; \theta)$ where θ is called a parameter and $g(.)$ is a function with a known form.[1]

[1] As an example, the RS $\{Y_1, ..., Y_n\}$ may be drawn from the univariate normal distribution function with mean $\mu \in \mathbb{R}$ and variance equal to one i.e. $f(y) = g(y; \mu) = \exp\left(\frac{1}{2\sqrt{\pi}}(y - \mu)^2\right)$ for all $y \in \mathbb{R}$. In this case the parameter θ is a scalar, namely, $\theta = \mu$. As another example, the RS $\{Y_1, ..., Y_n\}$ may be drawn from the univariate normal distribution function with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^{++}$ i.e. $f(y) = g(y; \mu, \sigma) = \exp\left(\frac{1}{2\sqrt{\pi}}\left(\frac{y-\mu}{\sigma}\right)^2\right)$ for all $y \in \mathbb{R}$. In this case the parameter θ is a vector, namely $\theta = (\mu, \sigma)$.

a. Give a formal definition of an estimator of θ. **Hint**: We ask for a definition that applies to the general pdf $f(\cdot)$, not the example pdfs provided in footnote 1.
b. Explain why the estimator is a random variable.
c. What is the difference between an estimator and an estimate?

6. Define an unbiased estimator. Explain why this is considered a 'finite-sample' (also called small-sample) property of an estimator.

7. Consider a RS $\{Y_1, \ldots, Y_n\}$ drawn randomly from $f(\cdot)$ with finite mean and variance. Define 4 estimators: 1) $W_n^a = Y_1$; 2) $W_n^b = Y_n$; 3) $W_n^c = \frac{1}{2}(Y_1 + Y_n)$; and 4) $W_n^d = \frac{1}{n}\sum_{i=1}^n Y_i$. Explain why these are all unbiased estimators of $E[Y]$.[2]

8. Define an consistent estimator. Explain why this is considered a 'large-sample' (also called asymptotic) property of an estimator?

9. Provide a simple formal formulation of the Law of Large Numbers (henceforth LLN).

10. Provide a simple formal formulation of the Central limit theorem.

11. Give a simple formal formulation of the Law of Iterated Expectations (henceforth LIE). **Side Note**: Many derivations in ECON 488 leverage this law so familiarity with this topic is key.

12. You consider senior people. A senior person's health score may be good or bad at the start of a given year. They may or may not have travelled internationally during that year, and they may or may not be deceased by year end. You know that 60% of senior people who travelled abroad had a good health score; instead only 20% of senior people who stayed home had a good health score. You know that 10% of senior people with a good score are deceased by year end; instead 50% of senior people with a bad score are deceased by year end. Finally you know that conditional on the health score, traveling abroad and being alive by year end are independent. What percentage of seniors who travelled abroad is alive by year end? What percentage of seniors who stayed home is alive by year end? **Hint**: Use the LIE.

13. In this question, you will generate a dataset in **R**, and then compute the least squares regression.

    a. First, fix the pseudo random number generator 'seed' by using the function `set.seed(25)`

    b. Then generate X as 100 random draws from the uniform distribution, and **then** generate e as 100 random draws from the standard normal distribution. The functions `runif()` and `rnorm()` may be useful. Finally, generate Y as $Y = 1 + 2 \times X + e$

    c. Compute the least squares regression by using the `lm()` function, and report your estimate for the intercept and the slope.

    d. Recompute the least squares *slope* estimate by using only the `cov()` and `var()` functions.

[2] The mean and variance of a distribution are often called population mean and population variance. We defined the mean in question 4 . The variance of a distribution is often denoted using the short-cut notation $Var[Y_i]$ where Y_i has distribution $f(\cdot)$. We define the variance of $f(\cdot)$ as $Var[Y_i] \equiv E[(Y_i - E[Y_i])^2]$, i.e., as the expected value of the squared deviation from the distribution's mean. As an operator, the variance has properties that you want to remember, e.g., let $Z_i = aY_i + b$ for some pair of scalars a and b. Then, $Var[Z_i] = a^2 Var[Y_i]$ (verify using the definition).