University of Washington
Department of Mathematics

Homework 5 Solutions

MATH 394: Probability I

Instructor: Arman Jahangiri

Submission: Single PDF on Gradescope
Deadline: 10:00 PM Pacific Time on the listed due date

This homework is worth **50 points**.
Across the quarter, there are **eight homework assignments**, worth **400 points total**. Homework assignments together account for **40%** of the final course grade.
Further course policies can be seen in the **MATH 394 syllabus.**

Homework Policy

Submission and Deadline

Please:

- Submit homework as a single PDF on Gradescope by **10:00 PM (Pacific Time)** on the listed due date.
- Begin each problem on a new page, and ensure that pages are correctly tagged in Gradescope.
- Write solutions clearly and legibly.
- Typesetting solutions in LaTeX is encouraged, but not required.

Late Days

Each student is allotted **six late days** for the quarter. A late day extends a homework deadline by up to 24 hours without penalty. For example, submitting an assignment anytime between 10:01 PM on the due date and 10:00 PM the following day counts as one late day.

The following rules apply:

- No explanation is required to use late days.
- At most **two late days** may be used on any single homework assignment.
- Late days may not be used on the final homework assignment or on any assignment for which solutions have already been released.
- It is the student's responsibility to keep track of how many late days have been used during the quarter.

Once all late days have been exhausted, additional late submissions will incur a penalty of **10% per day**, up to a maximum deduction of **50%**. Assignments submitted more than five days late, or after solutions have been released, will not be accepted.

Submission Issues and Technical Difficulties

If a serious technical issue prevents a timely Gradescope submission, students may temporarily submit their assignment by email to the instructor at armanjg@uw.edu. In such cases:

- The submission must consist of a **single PDF file**. The email subject line must follow the format:

  `HW# Submission - FULL NAME - STUDENT ID.`

- Once the Gradescope issue has been resolved, the student must upload the *exact same file* to Gradescope as soon as possible.
- The student should clearly indicate in the Gradescope submission that the assignment was previously submitted by email before the deadline.

Email submissions are intended only for genuine technical emergencies and should not be used as a substitute for timely Gradescope submission.

Problem 1.

Source: Blitzstein and Hwang, Introduction to Probability, Chapter 4, Problem 18.

A fair coin is tossed repeatedly, until it has landed Heads at least once and has landed Tails at least once.

Find the expected number of tosses.

Hint: Define T to be the total number of tosses required. After the first toss, only one outcome remains to be observed. Write

$$T = 1 + X,$$

where X is the number of additional tosses needed to obtain the outcome opposite to the first toss. X follows a well-known distribution! So if you find that, the rest is to use the linearity of expectation.

Solution

Let T denote the total number of tosses required until both Heads and Tails have appeared at least once.

After the first toss, exactly one of the two outcomes has appeared. Define

$$X = T - 1,$$

so that X is the number of additional tosses needed to observe the outcome opposite to the first toss.

Conditional on the outcome of the first toss, each subsequent toss produces the opposite outcome with probability $1/2$, independently of all previous tosses. Therefore,

$$X \sim \text{Geom}\left(\frac{1}{2}\right),$$

where X counts the number of trials up to and including the first success. Hence

$$E(X) = \frac{1}{1/2} = 2.$$

Since

$$T = 1 + X,$$

linearity of expectation gives

$$E(T) = 1 + E(X) = 1 + 2 = 3.$$

Problem 2.

Source: Blitzstein and Hwang, Introduction to Probability, Chapter 4, Problem 30.

(a) Use LOTUS to show that for $X \sim \text{Pois}(\lambda)$ and any function g,

$$E\big(Xg(X)\big) = \lambda E\big(g(X+1)\big),$$

assuming that both sides exist. This is called the *Stein–Chen identity* for the Poisson.

(b) Find the third moment $E(X^3)$ for $X \sim \text{Pois}(\lambda)$ by using the identity from part (a) and a bit of algebra to reduce the calculation to the fact that X has mean λ and variance λ.

Solution

(a) By LOTUS and the Poisson PMF,

$$E\big(Xg(X)\big) = \sum_{x=0}^{\infty} xg(x)e^{-\lambda}\frac{\lambda^x}{x!}.$$

The term for $x = 0$ is zero, so

$$E\big(Xg(X)\big) = \sum_{x=1}^{\infty} g(x)e^{-\lambda}\frac{\lambda^x}{(x-1)!}.$$

Letting $y = x - 1$, we get

$$E\big(Xg(X)\big) = \lambda\sum_{y=0}^{\infty} g(y+1)e^{-\lambda}\frac{\lambda^y}{y!} = \lambda E\big(g(X+1)\big).$$

(b) To find $E(X^3)$, take $g(x) = x^2$. Then the identity gives

$$E(X^3) = \lambda E\big((X+1)^2\big).$$

Expanding the square,

$$E(X^3) = \lambda\{E(X^2) + 2E(X) + 1\}.$$

For a Poisson random variable, $E(X) = \lambda$ and $\text{Var}(X) = \lambda$, so

$$E(X^2) = \text{Var}(X) + \{E(X)\}^2 = \lambda + \lambda^2.$$

Therefore

$$E(X^3) = \lambda(\lambda^2 + \lambda + 2\lambda + 1) = \lambda^3 + 3\lambda^2 + \lambda.$$

Problem 3.

Source: Blitzstein and Hwang, Introduction to Probability, Chapter 4, Problem 54.

Note. This problem introduces several fundamental ideas in data analysis, including sampling, estimation, uncertainty, bias, and statistical precision. Pay particular attention to this problem, specially if you'd like to smile, remembering this course, years later in a job that contains data collection/analysis. :)

A population consists of N people, with ID numbers $1, \dots, N$. For each person j, let y_j be the value of some numerical variable, and define the population mean by

$$\bar{y} = \frac{1}{N} \sum_{j=1}^{N} y_j.$$

A quantity such as $\bar{y}$ is called a *population parameter*. Once the population is fixed, its value is fixed, although it may be unknown to us. Population parameters are often denoted by symbols such as θ, μ, or $\bar{y}$.

In contrast, a quantity computed from a random sample is called a *sample statistic*. For example, if $W_1, \dots, W_n$ are the values observed in a random sample, then

$$\bar{W} = \frac{1}{n} \sum_{r=1}^{n} W_r$$

is a sample statistic. Since the selected sample is random, $\bar{W}$ is a random variable whose value may vary from one sample to another. More generally, a statistic may be written as

$$T = T(W_1, \dots, W_n),$$

emphasizing that it is a function of the observed sample.

One of the primary goals of statistics is to use sample statistics to estimate unknown population parameters.

Suppose a researcher wants to estimate the population parameter

$$\theta := \bar{y},$$

but it is not feasible to measure every person in the population (which is required for computing $\bar{y}$ based on its formula). The researcher therefore selects a simple random sample of size n, where $n \ll N$, without replacement. Thus, every subset of n people is equally likely to be selected.

Let $W_1, \dots, W_n$ denote the values observed for the selected people, where the sampled people are listed in a uniformly random order, and let

$$\bar{W} = \frac{1}{n} \sum_{r=1}^{n} W_r$$

be the sample mean.

For each population member $1 \leq j \leq N$, define

$$I_j = \begin{cases} 1, & \text{if person } j \text{ is included in the sample,} \\ 0, & \text{otherwise.} \end{cases}$$

Define the finite-population variance by

$$S_y^2 = \frac{1}{N-1}\sum_{j=1}^{N}(y_j - \bar{y})^2.$$

Answer the following questions.

(a) Explain carefully why $y_1, \ldots, y_N$ are not random variables, while $W_1, \ldots, W_n$, $I_1, \ldots, I_N$, and $\bar{W}$ are random variables.

(b) Show by symmetry that, for every sample position r,

$$P(W_r = y_j) = \frac{1}{N}, \qquad j = 1, \ldots, N.$$

Use this fact to prove that

$$E(\bar{W}) = \bar{y}.$$

(c) Show that

$$\bar{W} = \frac{1}{n}\sum_{j=1}^{N} I_j y_j.$$

Then compute $E(I_j)$ and use linearity of expectation to give a second proof that

$$E(\bar{W}) = \bar{y}.$$

(d) Since the sample contains exactly n people, show that

$$\sum_{j=1}^{N} I_j = n$$

with probability 1. Explain why this identity does not imply that the individual indicators $I_1, \ldots, I_N$ are constants.

(e) Compute $\mathrm{Var}(I_j)$. Then use the indicator representation of $\bar{W}$ and the definition of variance to show that

$$\mathrm{Var}(\bar{W}) = \frac{N-n}{Nn} S_y^2 = \left(1 - \frac{n}{N}\right)\frac{S_y^2}{n}.$$

You may use the fact that, for $i \neq j$,

$$P(I_i = 1 \text{ and } I_j = 1) = \frac{n(n-1)}{N(N-1)}.$$

The factor

$$1 - \frac{n}{N}$$

is called the *finite-population correction.*

(f) Explain what the variance formula predicts in each of the following cases:

(i) $n = 1$;

(ii) $n = N$;

(iii) all population values are equal;

(iv) n increases while the population remains fixed.

(g) Suppose instead that the researcher samples n people independently *with replacement*. Let $\bar{W}_{\text{rep}}$ be the resulting sample mean. Show that

$$\operatorname{Var}(\bar{W}_{\text{rep}}) = \frac{N-1}{N}\frac{S_y^2}{n}.$$

Compare this with the variance under sampling without replacement. Why is sampling without replacement more precise?

(h) The bias and mean squared error of an estimator T for a population parameter θ are defined by

$$\operatorname{Bias}(T) = E(T) - \theta$$

and

$$\operatorname{MSE}(T) = E\big[(T-\theta)^2\big].$$

Show that

$$\operatorname{MSE}(T) = \operatorname{Var}(T) + \operatorname{Bias}(T)^2.$$

Use this identity to find $\operatorname{Bias}(\bar{W})$ and $\operatorname{MSE}(\bar{W})$.

(i) A researcher says:

> "Because $\bar{W}$ is unbiased, the sample mean must equal the population mean for every possible sample."

Explain why this statement is false. What does unbiasedness actually mean?

Solution

(a) The population values $y_1, \dots, y_N$ are fixed characteristics of the population. Once the population is specified, these values do not depend on which sample is selected. Therefore, under the sampling procedure, they are constants rather than random variables.
By contrast, W_r depends on which population member appears in sample position r, and I_j depends on whether person j is selected. Therefore, their values can vary across possible random samples. Since $\bar{W}$ is computed from the random observations $W_1, \dots, W_n$, it is also a random variable.

(b) By symmetry, every population member is equally likely to appear in sample position r. Therefore,

$$P(W_r = y_j) = \frac{1}{N}, \qquad j = 1, \dots, N.$$

By LOTUS,

$$E(W_r) = \sum_{j=1}^{N} y_j P(W_r = y_j) = \frac{1}{N}\sum_{j=1}^{N} y_j = \bar{y}.$$

Using linearity of expectation,

$$\begin{aligned} E(\bar{W}) &= E\left(\frac{1}{n}\sum_{r=1}^{n} W_r\right) \\ &= \frac{1}{n}\sum_{r=1}^{n} E(W_r) \\ &= \frac{1}{n}\sum_{r=1}^{n} \bar{y} \\ &= \bar{y}. \end{aligned}$$

Thus, the sample mean is an unbiased estimator of the population mean.

(c) For each j, the product $I_j y_j$ equals y_j if person j is selected and equals 0 otherwise. Therefore,

$$\sum_{j=1}^{N} I_j y_j$$

is exactly the sum of the values belonging to the sampled people. Hence

$$\sum_{r=1}^{n} W_r = \sum_{j=1}^{N} I_j y_j,$$

and therefore

$$\bar{W} = \frac{1}{n}\sum_{j=1}^{N} I_j y_j.$$

Every population member has the same probability of being included in a simple random sample of size n. Thus,

$$P(I_j = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}.$$

Since I_j is an indicator,

$$E(I_j) = P(I_j = 1) = \frac{n}{N}.$$

Consequently,

$$\begin{aligned} E(\bar{W}) &= E\left(\frac{1}{n}\sum_{j=1}^{N} I_j y_j\right) \\ &= \frac{1}{n}\sum_{j=1}^{N} y_j E(I_j) \\ &= \frac{1}{n}\sum_{j=1}^{N} y_j \frac{n}{N} \\ &= \frac{1}{N}\sum_{j=1}^{N} y_j \\ &= \bar{y}. \end{aligned}$$

(d) Every possible sample contains exactly n people. Since $I_j = 1$ precisely when person j is selected, the number of selected people is

$$\sum_{j=1}^{N} I_j.$$

Therefore,

$$\sum_{j=1}^{N} I_j = n$$

for every possible sample, and hence with probability 1.
This does not imply that the individual indicators are constants. For example, one sample may have $I_1 = 1$, while another may have $I_1 = 0$. The indicators vary across samples, but they vary in a coordinated way so that their sum is always n.

(e) Since I_j is Bernoulli with success probability n/N,

$$\mathrm{Var}(I_j) = \frac{n}{N}\left(1 - \frac{n}{N}\right) = \frac{n(N-n)}{N^2}.$$

To find the variance of the sample mean, begin with

$$\bar{W} - \bar{y} = \frac{1}{n}\sum_{j=1}^{N} I_j y_j - \bar{y}.$$

Since

$$\sum_{j=1}^{N} I_j = n$$

for every possible sample, we may write

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{N} I_j \bar{y}.$$

Therefore,

$$\bar{W} - \bar{y} = \frac{1}{n}\sum_{j=1}^{N} I_j (y_j - \bar{y}).$$

Because $E(\bar{W}) = \bar{y}$,

$$\mathrm{Var}(\bar{W}) = E\big[(\bar{W} - \bar{y})^2\big].$$

Thus,

$$\begin{aligned}
\mathrm{Var}(\bar{W}) &= \frac{1}{n^2} E\left[\left(\sum_{j=1}^{N} I_j (y_j - \bar{y})\right)^2\right] \\
&= \frac{1}{n^2} E\left[\sum_{j=1}^{N} I_j^2 (y_j - \bar{y})^2 + 2 \sum_{1 \le i < j \le N} I_i I_j (y_i - \bar{y})(y_j - \bar{y})\right].
\end{aligned}$$

Since $I_j^2 = I_j$,

$$E(I_j^2) = E(I_j) = \frac{n}{N}.$$

Also, for $i \neq j$,

$$E(I_i I_j) = P(I_i = 1 \text{ and } I_j = 1) = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}.$$

Hence

$$\operatorname{Var}(\bar{W}) = \frac{1}{n^2}\left[\frac{n}{N}\sum_{j=1}^{N}(y_j - \bar{y})^2 + \frac{2n(n-1)}{N(N-1)}\sum_{i<j}(y_i - \bar{y})(y_j - \bar{y})\right].$$

Now,

$$\sum_{j=1}^{N}(y_j - \bar{y}) = 0.$$

Squaring this identity gives

$$0 = \sum_{j=1}^{N}(y_j - \bar{y})^2 + 2\sum_{i<j}(y_i - \bar{y})(y_j - \bar{y}).$$

Therefore,

$$2\sum_{i<j}(y_i - \bar{y})(y_j - \bar{y}) = -\sum_{j=1}^{N}(y_j - \bar{y})^2.$$

Substituting,

$$\begin{aligned}\operatorname{Var}(\bar{W}) &= \frac{1}{n^2}\left[\frac{n}{N} - \frac{n(n-1)}{N(N-1)}\right]\sum_{j=1}^{N}(y_j - \bar{y})^2 \\ &= \frac{1}{n^2}\frac{n(N-n)}{N(N-1)}\sum_{j=1}^{N}(y_j - \bar{y})^2.\end{aligned}$$

Since

$$S_y^2 = \frac{1}{N-1}\sum_{j=1}^{N}(y_j - \bar{y})^2,$$

we conclude that

$$\operatorname{Var}(\bar{W}) = \frac{N-n}{Nn}S_y^2.$$

Equivalently,

$$\operatorname{Var}(\bar{W}) = \left(1 - \frac{n}{N}\right)\frac{S_y^2}{n}.$$

(f)

(i) If $n = 1$, then

$$\operatorname{Var}(\bar{W}) = \frac{N-1}{N}S_y^2.$$

In this case, the sample mean consists of only one observation and generally has substantial variability.

(ii) If $n = N$, then

$$\mathrm{Var}(\bar{W}) = 0.$$

The entire population is observed, so

$$\bar{W} = \bar{y}$$

for every possible sample.

(iii) If all population values are equal, then

$$S_y^2 = 0,$$

and hence

$$\mathrm{Var}(\bar{W}) = 0.$$

Every possible sample has the same sample mean.

(iv) As n increases, both $1/n$ and the finite-population correction

$$1 - \frac{n}{N}$$

decrease. Therefore, the variance of the sample mean decreases as a larger fraction of the population is observed.

(g) Under sampling with replacement, the observations

$$W_1, \ldots, W_n$$

are independent and identically distributed, with

$$P(W_r = y_j) = \frac{1}{N}, \qquad j = 1, \ldots, N.$$

Therefore,

$$E(W_r) = \bar{y}.$$

Also,

$$\begin{aligned}
\mathrm{Var}(W_r) &= E\big[(W_r - \bar{y})^2\big] \\
&= \frac{1}{N} \sum_{j=1}^{N} (y_j - \bar{y})^2 \\
&= \frac{N-1}{N} S_y^2.
\end{aligned}$$

Since the observations are independent,

$$\begin{aligned}
\mathrm{Var}(\bar{W}_{\mathrm{rep}}) &= \mathrm{Var}\left(\frac{1}{n} \sum_{r=1}^{n} W_r\right) \\
&= \frac{1}{n^2} \sum_{r=1}^{n} \mathrm{Var}(W_r) \\
&= \frac{1}{n} \frac{N-1}{N} S_y^2.
\end{aligned}$$

Under sampling without replacement,

$$\mathrm{Var}(\bar{W}) = \frac{N-n}{Nn} S_y^2.$$

Since

$$N - n \le N - 1,$$

we have

$$\mathrm{Var}(\bar{W}) \le \mathrm{Var}(\bar{W}_{\text{rep}}).$$

In fact, when $n > 1$, the inequality is strict unless $S_y^2 = 0$.
Sampling without replacement is more precise because it avoids duplicate observations. For a fixed sample size, it therefore gathers information from more distinct population members.

Final Answer for Part (g)

$$\mathrm{Var}(\bar{W}_{\text{rep}}) = \frac{N-1}{N} \frac{S_y^2}{n}.$$

(h) Let θ be the parameter being estimated. Write

$$T - \theta = \big(T - E(T)\big) + \big(E(T) - \theta\big).$$

Squaring both sides gives

$$\begin{aligned}(T-\theta)^2 = {} & \big(T - E(T)\big)^2 \\ & + 2\big(T - E(T)\big)\big(E(T) - \theta\big) \\ & + \big(E(T) - \theta\big)^2.\end{aligned}$$

Taking expectations,

$$\begin{aligned}E\big[(T-\theta)^2\big] = {} & E\big[(T - E(T))^2\big] \\ & + 2\big(E(T) - \theta\big)E\big[T - E(T)\big] \\ & + \big(E(T) - \theta\big)^2.\end{aligned}$$

Since

$$E\big[T - E(T)\big] = 0,$$

the middle term vanishes. Hence

$$\mathrm{MSE}(T) = \mathrm{Var}(T) + \mathrm{Bias}(T)^2.$$

For the sample mean,

$$\mathrm{Bias}(\bar{W}) = E(\bar{W}) - \bar{y} = 0.$$

Therefore,

$$\mathrm{MSE}(\bar{W}) = \mathrm{Var}(\bar{W}) = \frac{N-n}{Nn} S_y^2.$$

Final Answer for Part (h)

$$\mathrm{Bias}(\bar{W}) = 0,$$

and

$$\mathrm{MSE}(\bar{W}) = \frac{N-n}{Nn} S_y^2.$$

(i) The statement is false. Unbiasedness is a statement about the average value of an estimator over all possible random samples, not about the value obtained from every individual sample.
A particular sample may contain an unusually large number of people with high values, in which case

$$\bar{W} > \bar{y}.$$

Another sample may contain an unusually large number of people with low values, in which case

$$\bar{W} < \bar{y}.$$

The equality

$$E(\bar{W}) = \bar{y}$$

means that if the sampling procedure were repeated many times, the average of the resulting sample means would equal the population mean. It does not imply that

$$\bar{W} = \bar{y}$$

for every possible sample.