Notes from Week 3

STATS 118

Andrea Yu

January 24, 2026

Here are some extra notes on topics that we discussed in tutorial this week.

- On **Problem 32.11**:
    - Make sure you understand the symmetry argument for the i.i.d. case. In particular, why is it true that all X_i have the same variance, even though the problem only explicitly states that they share a mean μ?
    - The general case is an extension from what the original problem asks. The Lagrange multiplier method should be familiar to you. (Please review Lagrange multipliers if you need a refresher.)
    - The Cauchy-Schwarz method is an alternative way to solve the general case. This inequality is useful to know but not strictly necessary. Don't worry too much about this; I just wanted to include it since we mentioned it in class.
- On **uniform convolutions**: like I mentioned in class, these questions show up a lot.
    - They can be a little confusing at first, but the good news is that once you've figured out how to work through them, you'll be able to solve all problems of the same type.
    - I worked through one example fully and included a few additional problems.

#32.11 Linear unbiased estimators of the mean

Exercise

Let $X_1, \dots, X_n$ be independent random variables with common mean $E[X_i] = \mu$ and different variances $V[X_i] = \sigma_i^2$.

(a) What condition on constants $w_1, \dots, w_n$ is necessary for $\hat{\mu} = w_1 X_1 + \cdots + w_n X_n$ to be an unbiased estimator for μ?

(b) What values of $w_1, \dots, w_n$ ensure that $\hat{\mu}$ is both unbiased and minimizes $V[\hat{\mu}]$?

Note. The original problem in the textbook assumes that X_i are i.i.d.—that is, they all have the same mean and the same variance. We'll solve the general case (featuring **heteroskedasticity** or differing variances) here.

The logic for the general case easily applies to the i.i.d. case. We will get the same answer of $w_1 = \cdots = w_n = 1/n$ in this special case, but hopefully it will be motivated by a more satisfying mathematical argument than simply "by symmetry".

Part (a)

By linearity of expectation, the expected value of $\hat{\mu}$ is

$$E[\hat{\mu}] = E\left[\sum_{i=1}^n w_i X_i\right] = \sum_{i=1}^n w_i E[X_i] = \sum_{i=1}^n w_i \mu = \mu\left(\sum_{i=1}^n w_i\right).$$

In order for $\hat{\mu}$ to be unbiased, we need $E[\hat{\mu}] = \mu$. Thus, our constants w_i must satisfy

$$\sum_{i=1}^n w_i = w_1 + \cdots + w_n = 1. \qquad \square$$

Part (b)

We know from Part (a) that $\sum_{i=1}^n w_i$ must equal 1 in order for $\hat{\mu}$ to be unbiased.

Since X_i are independent, the variance of $\hat{\mu}$ is given by

$$V[\hat{\mu}] = V\left[\sum_{i=1}^n w_i X_i\right] = \sum_{i=1}^n w_i^2 V[X_i] = \sum_{i=1}^n w_i^2 \sigma_i^2.$$

Our goal is to minimize this expression, where σ_i^2 are constants and w_i are subject to the constraint that guarantees unbiasedness. This is a natural application for **Lagrange multipliers**. We can also use a clever substitution to apply the **Cauchy-Schwarz inequality**.

Lagrange multipliers

Recall that the method of Lagrange multipliers allows us to optimize (i.e., minimize or maximize) a **target function** $f(w_1, \dots, w_n)$ subject to the **constraint** $g(w_1, \dots, w_n) = c$. We do this by computing the gradients and finding where they align. We solve the system $\nabla f = \lambda \nabla g$ (which actually gives us n equations in the n components) and $g = c$.

Here, we are trying to minimize our target function $f(w_1, \dots, w_n) = \sum_{i=1}^n w_i^2 \sigma_i^2$ subject to the constraint $g(w_1, \dots, w_n) = \sum_{i=1}^n w_i = 1$. Taking gradients yields

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{pmatrix} = \begin{pmatrix} 2w_1\sigma_1^2 \\ 2w_2\sigma_2^2 \\ \vdots \\ 2w_n\sigma_n^2 \end{pmatrix}, \qquad \lambda \nabla g = \lambda \begin{pmatrix} \frac{\partial g}{\partial w_1} \\ \frac{\partial g}{\partial w_2} \\ \vdots \\ \frac{\partial g}{\partial w_n} \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda \\ \lambda \\ \vdots \\ \lambda \end{pmatrix}.$$

Setting these gradients equal tells us that $2w_i\sigma_i^2 = \lambda$ for all $1 \le i \le n$. Thus, each $w_i = \lambda/(2\sigma_i^2)$, and substituting back into our constraint $g(w_1, \dots, w_n) = 1$ gives us

$$1 = \frac{\lambda}{2\sigma_1^2} + \cdots + \frac{\lambda}{2\sigma_n^2} = \frac{\lambda}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \qquad \Longrightarrow \qquad \lambda = \frac{2}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}.$$

This means the minimizing values of w_i are

$$w_i = \frac{\lambda}{2\sigma_i^2} = \frac{\frac{2}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}}{2\sigma_i^2} = \frac{1}{\sigma_i^2} \cdot \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} = \boxed{\frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}}.$$

Alternatively, if we write $S = \sum_{i=1}^n \frac{1}{\sigma_i^2}$, then we have

$$w_i = \frac{\frac{1}{\sigma_i^2}}{S} = \boxed{\frac{1}{\sigma_i^2 \cdot S}}. \qquad \square$$

These are the **minimizers** of $V[\hat{\mu}]$. We compute the **minimum value** of the variance as

$$\sum_{i=1}^n w_i^2 \sigma_i^2 = \sum_{i=1}^n \left(\frac{1}{\sigma_i^2 \cdot S} \right)^2 \sigma_i^2 = \sum_{i=1}^n \frac{1}{\left(\sigma_i^2\right)^2 \cdot S^2} \cdot \sigma_i^2 = \frac{1}{S^2} \underbrace{\sum_{i=1}^n \frac{1}{\sigma_i^2}}_{\text{exactly } S} = \frac{1}{S^2} \cdot S = \boxed{\frac{1}{S}}.$$

Special case. If all X_i are i.i.d., then all the variances σ_i^2 are equal to some common variance σ^2. The minimizing coefficients w_i are given by

$$S = \sum_{i=1}^n \frac{1}{\sigma^2} = n \cdot \frac{1}{\sigma^2} = \frac{n}{\sigma^2} \qquad \Longrightarrow \qquad w_i = \frac{1}{\sigma_i^2 \cdot S} = \frac{1}{\sigma_i^2 \cdot \frac{n}{\sigma_i^2}} = \boxed{\frac{1}{n}}.$$

Cauchy-Schwarz inequality

Theorem (General Cauchy-Schwarz)
Let u, v be vectors in an inner product space endowed with $\langle \cdot, \cdot \rangle$. Then

$$|\langle u, v\rangle|^2 \leq \langle u, u\rangle \cdot \langle v, v\rangle.$$

Alternatively, since $\langle \cdot, \cdot \rangle$ induces the norm $\|u\| := \sqrt{\langle u, u\rangle}$, we can write

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|.$$

Don't worry too much about this version of Cauchy-Schwarz, which is stated in full generality. The version that we care about—and which you will almost always use, unless you're a math major or *super* interested in analysis—is stated as follows:

Theorem (Cauchy-Schwarz in $\mathbb{R}^n$)
Let $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$ be vectors in $\mathbb{R}^n$ (i.e., vectors with n components which are all real numbers). Then

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right).$$

We return to our variance-minimizing problem and make a clever choice of a_i and b_i in order to leverage this inequality. In particular, let

$$a_i = \frac{1}{\sigma_i} \qquad \text{and} \qquad b_i = \sigma_i w_i \qquad\qquad \text{for all } 1 \leq i \leq n.$$

Recalling the constraint $\sum_{i=1}^n w_i = 1$, the left side of the inequality becomes

$$\left(\sum_{i=1}^n a_i b_i\right)^2 = \left(\sum_{i=1}^n \frac{1}{\sigma_i} \cdot \sigma_i w_i\right)^2 = \left(\sum_{i=1}^n w_i\right)^2 = 1^2 = 1.$$

Defining $S = \sum_{i=1}^n \frac{1}{\sigma_i^2}$ as before, the right side of the inequality becomes

$$\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right) = \underbrace{\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)}_{S} \cdot \underbrace{\left(\sum_{i=1}^n w_i^2 \sigma_i^2\right)}_{\text{variance}} = S \cdot V[\hat{\mu}].$$

Combining yields $1 \leq S \cdot V[\hat{\mu}]$ and thus $V[\hat{\mu}] \geq 1/S$. This gives us the minimum value of the variance immediately—precisely when the inequality becomes an *equality*—and it is the same as the minimum value that we computed using Lagrange multipliers. Furthermore, equality holds if and only if $a_i = cb_i$ for every $1 \leq i \leq n$, where c is a fixed constant. This means $\frac{1}{\sigma_i} = c\sigma_i w_i$ or $w_i = \frac{c}{\sigma_i^2}$. Substituting into our constraint, we have

$$1 = \sum_{i=1}^n w_i = \sum_{i=1}^n \frac{c}{\sigma_i^2} = c\sum_{i=1}^n \frac{1}{\sigma_i^2} = c \cdot S \qquad \Longrightarrow \qquad w_i = \frac{c}{\sigma_i^2} = \frac{1/S}{\sigma_i^2} = \boxed{\frac{1}{\sigma_i^2 \cdot S}}. \qquad \square$$

#33.5 (and more) Uniform convolution problems

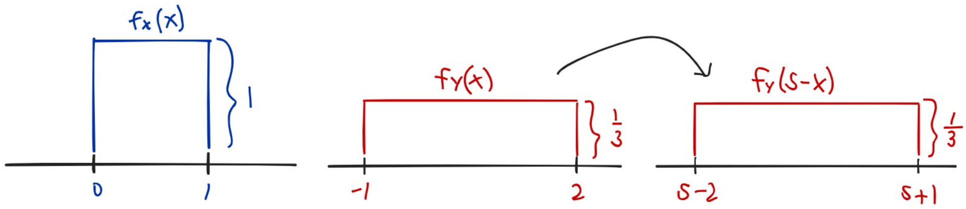
Exercise

Let $X \sim \text{Uni}(0,1)$ and $Y \sim \text{Uni}(-1,2)$ be independent uniform random variables. What is the distribution of $S = X + Y$?

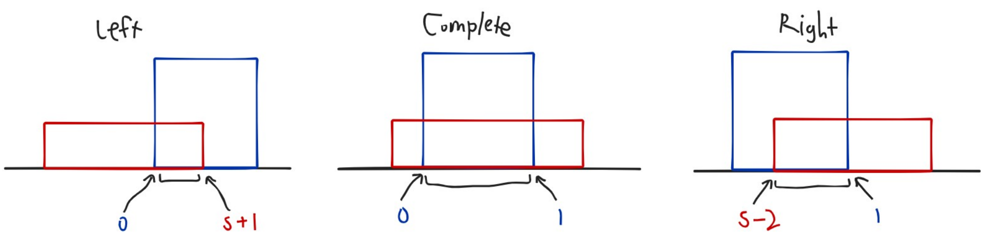
Recall the formula for the convolution of two independent continuous random variables X and Y, denoting their sum as $S = X + Y$:

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x)\, dx.$$

(1) Sketch the densities of the two uniform random variables. Note that one of these densities will be in terms of the argument x and one of them will be in terms of the argument $s - x$ (or $s - y$ and y respectively if you convolve the other way). Notice how the bounds of $f_Y(s-x)$ change in terms of s.



(2) Consider where the densities overlap. Be sure to watch your bounds carefully and identify which bounds contribute in each case. You can think about this by "fixing" one density (e.g., the one that does not depend on s) in place and "sliding" the other density (e.g., the one that depends on s) from left to right, which is equivalent to varying s.



(3) Solve for the conditions that determine each case. In other words, figure out how "overlap on the left", "complete overlap", and "overlap on the right" occur.

- **Left.** The bounds of the overlapping piece are 0 (blue) and $s+1$ (red).

  The higher/right bound of the red density must be greater than the lower/left bound of the blue density; otherwise, they will not overlap at all. This means $s + 1 > 0$ and thus $s > -1$.

  The higher/right bound of the red density must be less than the higher/right bound of the blue density; otherwise, we will be in the complete overlap case. This means $s + 1 < 1$ and thus $s < 0$.

  Combining tells us that left overlap occurs when $\boxed{-1 < s < 0}$.

- **Complete.** The bounds of the overlapping piece are 0 (blue) and 1 (blue).

  The lower/left bound of the red density must be less than the lower/left bound of the blue density; otherwise, we will be in the right overlap case. This means $s - 2 < 0$ and thus $s < 2$.

  The higher/right bound of the red density must be greater than the higher/right bound of the blue density; otherwise, we will be in the left overlap case. This means $s + 1 > 1$ and thus $s > 0$.

  Combining tells us that complete overlap occurs when $\boxed{0 < s < 2}$.

- **Right.** The bounds of the overlapping piece are $s - 2$ (red) and 1 (blue).

  The lower/left bound of the red density must be greater than the lower/left bound of the blue density; otherwise, we will be in the complete overlap case. This means $s - 2 > 0$ and thus $s > 2$.

  The lower/left bound of the red density must be less than the upper/right bound of the blue density; otherwise, they will not overlap at all. This means $s - 2 < 1$ and thus $s < 3$.

  Combining tells us that right overlap occurs when $\boxed{2 < s < 3}$.

There are a few sanity checks you can do after you have solved for these conditions.

First, what are the maximum and minimum values that $S = X + Y$ can take on? Your conditions should cover every possible value in that range.

- The minimum possible value of S is $\min X + \min Y = 0 + (-1) = -1$.
- The maximum possible value of S is $\max X + \max Y = 1 + 2 = 3$.
- Indeed, our conditions cover the ranges $(-1, 0)$, $(0, 2)$, and $(2, 3)$. The union of these ranges forms $(0, 3)$, the entire support of S. *
- We also notice that the conditions are symmetric; this makes sense if you think about the "sliding" visual that we introduced earlier.

Note. We are being a bit handwavy and excluding the boundary values at the edges of each range. If we have solved correctly, we will find later that the piecewise densities of S that we find on each segment should line up at these endpoints.

As a second sanity check, sketch what happens to the densities exactly at the endpoints. (For instance, what happens when $s = -1$ or $s = 0$? You should see the "extreme" versions of the left overlap case. If you slide to the left from $s = -1$, you will end up with no overlap. If you slide to the right from $s = 0$, you will end up with complete overlap.) This will help you build intuition for how to track and solve for all these conditions.

(4) Integrate along each piece using the convolution formula. The great thing about uniform densities is that they are constant or zero. Therefore, whenever there is overlap (whether left, complete, or right), the integrand $f_X(x)f_Y(s - x)$ will be a constant value. In this example, $f_X(x) = 1$ when nonzero and $f_Y(s - x) = 1/3$ when nonzero. When they are both nonzero (i.e., they overlap), the integrand is always $1 \cdot 1/3 = 1/3$.

The bounds on each piece are exactly the bounds that we identified in the previous step. **Be careful not to mix up your bounds (on x) and your conditions (on s)!**

On the left overlap piece, our bounds are $(0, s+1)$. We integrate

$$f_S(s) = \int_0^{s+1} \frac{1}{3}\,dx = \frac{1}{3}x\Big|_0^{s+1} = \frac{1}{3}(s+1) - \frac{1}{3}(0) = \frac{s+1}{3}.$$

On the complete overlap piece, our bounds are $(0, 1)$. We integrate

$$f_S(s) = \int_0^{1} \frac{1}{3}\,dx = \frac{1}{3}x\Big|_0^{1} = \frac{1}{3}(1) - \frac{1}{3}(0) = \frac{1}{3}.$$

On the right overlap piece, our bounds are $(s-2, 1)$. We integrate

$$f_S(s) = \int_{s-2}^{1} \frac{1}{3}\,dx = \frac{1}{3}x\Big|_{s-2}^{1} = \frac{1}{3}(1) - \frac{1}{3}(s-2) = \frac{3-s}{3}.$$

As we noted before, we should check the endpoints between ranges to verify that we get the same value no matter which piecewise density we use. (For example, check $s = 0$ in both $(s+1)/3$ and $1/3$, and check $s = 2$ in both $1/3$ and $(3-s)/3$.)

(5) Write down the complete density. This is where we use the **conditions** from Step 3. Don't forget to specify that everything outside the support of S has density 0. We get

$$f_S(s) = \begin{cases} \dfrac{s+1}{3}, & -1 \le s \le 0, \\ \dfrac{1}{3}, & 0 \le s \le 2, \\ \dfrac{3-s}{s}, & 2 \le s \le 3, \\ 0, & \text{otherwise.} \end{cases}$$ □

Exercise

Let X and Y be independent uniform random variables. Find the distribution of $S = X + Y$ in each of the following cases:

- $X \sim \text{Uni}(0, 4)$ and $Y \sim \text{Uni}(-2, 5)$;
- $X \sim \text{Uni}(-1, 3)$ and $Y \sim \text{Uni}(0, 1)$;
- $X \sim \text{Uni}(-2, 3)$ and $Y \sim \text{Uni}(-3, 2)$.

Exercise (#33.4 – Challenge)

Let $X, Y, Z \sim \text{Uni}(0, 1)$ be independent uniform random variables. What is the distribution of $S = X + Y + Z$?

Hint. First derive the distribution of $X + Y$; this is explained in **Example 33.3**. How many pieces do you have to consider when working with $S = X + Y + Z$?