Chapter 22

Linear Combinations; Point Estimation and Bias

Lecture 22 | Devore 9e: 5.5, 6.1

22.1 Linear combinations

Let $X_1, \dots, X_n$ have means $\mathbb{E}[X_i] = \mu_i$ and known covariances $\mathrm{Cov}(X_i, X_j)$, and consider the statistic

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n.$$

Theorem 22.1 (Mean)**.**

$$\mathbb{E}[Y] = \sum_{i=1}^{n} a_i \, \mathbb{E}[X_i] = \sum_{i=1}^{n} a_i \mu_i.$$

This needs **no** independence assumption — expectation is always linear. Variance is different.

Theorem 22.2 (Variance)**.**

$$\mathrm{Var}(Y) = \mathrm{Cov}(Y, Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \; \mathrm{Cov}(X_i, X_j) = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \cdots & \mathrm{Cov}(X_n, X_n) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

If the X_i are independent, all off-diagonal terms vanish and

$$\boxed{\mathrm{Var}\Big(\sum_i a_i X_i\Big) = \sum_{i=1}^{n} a_i^2 \, \sigma_{X_i}^2}$$

Remark 22.3. Note the a_i^2: coefficients enter the mean linearly and the variance *quadratically*. In particular $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ for independent X, Y — subtracting does not subtract variance, it adds it.

Theorem 22.4 (Normal is closed under linear combinations)**.** *If $X_1, \dots, X_n$ are independent and each $X_i \sim \mathrm{Normal}(\mu_i, \sigma_i^2)$, then*

$$\sum_{i=1}^{n} a_i X_i \ \sim\ \mathrm{Normal}\Big(\sum_i a_i\mu_i,\ \sum_i a_i^2\sigma_i^2\Big)$$

exactly — no approximation, no large-n requirement.

This is the fact behind "$\bar{X} \sim \mathrm{Normal}(\mu, \sigma^2/n)$ exactly for a normal sample" from last lecture.

22.2 Point estimation

Definition 22.5 (Point estimator and point estimate)**.** Let θ be an unknown parameter of a distribution. A **point estimator** $\hat{\theta}$ of θ is a statistic — a random variable computed from $X_1, \dots, X_n$. The number it produces from actual data $x_1, \dots, x_n$ is the **point estimate**.

Remark 22.6. Same distinction as always: estimat*or* = random variable (capital letters in); estimat*e* = number (data in). An estimator has a distribution; an estimate does not.

Example 22.7 (The mean)**.** $X_1, \dots, X_n$ i.i.d. with mean μ and variance σ^2. Then $\hat{\mu} = \bar{X}$ is a point estimator of μ, and $\mathbb{E}[\bar{X}] = \mu$.

Example 22.8 (A proportion)**.** Estimate the pass rate p of STAT400. Sample n students and let

$$X_i = \begin{cases} 1, & \text{student } i \text{ passes,} \\ 0, & \text{fails.} \end{cases}$$

Then $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$. Since $\mathbb{E}[X_i] = p$, we get $\mathbb{E}[\hat{p}] = p$.

Example 22.9 (A difference of means)**.** iPhone battery life has mean μ_X; Samsung's has mean μ_Y. To compare them, estimate $\mu_X - \mu_Y$ by

$$\bar{X} - \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

By the linear-combination rules, $\mathbb{E}[\bar{X} - \bar{Y}] = \mu_X - \mu_Y$ and (for independent samples) $\mathrm{Var}(\bar{X} - \bar{Y}) = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$.

Example 22.10 (Two estimators of the variance)**.** For an i.i.d. sample, which of

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad \text{or} \qquad \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

should estimate σ^2? The next definition decides it.

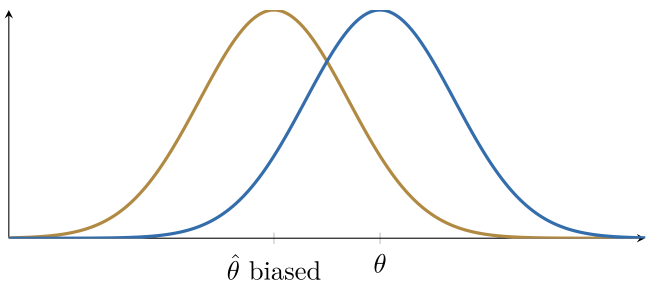
22.3 Bias

Definition 22.11 (Unbiased estimator)**.** $\hat{\theta}$ is an **unbiased** estimator of θ if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

In general $\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$.

Remark 22.12. Unbiased means: *on average over repeated samples*, the estimator hits the target. It says nothing about any single sample — an unbiased estimator can be wildly wrong every time, as long as the errors cancel. That is why the next lecture adds variance to the picture.



Theorem 22.13. *For an i.i.d. sample with variance σ^2,*

$$\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sigma^2,$$

so S^2 is unbiased, while $\frac{1}{n}\sum_i (X_i - \bar{X})^2$ has expectation $\frac{n-1}{n}\sigma^2$ and is biased downward.

Proof sketch (this is Homework). Use the identity $\sum_i (X_i - \bar{X})^2 = \sum_i (X_i - \mu)^2 - n(\bar{X} - \mu)^2$, take expectations, and use $\mathbb{E}[(X_i - \mu)^2] = \sigma^2$ and $\text{Var}(\bar{X}) = \sigma^2/n$:

$$\mathbb{E}\left[\sum_i (X_i - \bar{X})^2\right] = n\sigma^2 - n \cdot \frac{\sigma^2}{n} = (n-1)\sigma^2. \qquad \square$$

22.4 A worked example: estimating the endpoint of a uniform

Example 22.14 (Apple Watch water resistance)**.** The maximum time a watch survives under water is an unknown constant θ. You have n identical watches and test each one, obtaining $X_1, \ldots, X_n$ i.i.d. Uniform$[0, \theta]$. Estimate θ.

The obvious estimator is

$$\hat{\theta} = \max(X_1, \ldots, X_n),$$

and note immediately that $0 \le \hat{\theta} \le \theta$ always — it can never overshoot, which already smells biased. Let us compute.

For $0 \le t \le \theta$, using independence,

$$F_{\hat{\theta}}(t) = \mathbb{P}\{\max(X_1, \ldots, X_n) \le t\} = \mathbb{P}\{X_1 \le t\} \cdots \mathbb{P}\{X_n \le t\} = \left(\frac{t}{\theta}\right)^n,$$

so

$$f_{\hat{\theta}}(t) = \frac{n\,t^{n-1}}{\theta^n}, \qquad 0 \le t \le \theta.$$

Therefore

$$\mathbb{E}[\hat{\theta}] = \int_0^\theta t \cdot \frac{nt^{n-1}}{\theta^n}\,\mathrm{d}t = \frac{n}{\theta^n}\int_0^\theta t^n\,\mathrm{d}t = \frac{n}{\theta^n}\cdot\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\,\theta \;\neq\; \theta.$$

So max is **biased**, with bias $-\theta/(n+1)$ — it systematically underestimates, exactly as the picture suggested. But the bias vanishes as $n \to \infty$, and the fix is immediate: rescale.

Theorem 22.15.

$$\boxed{\hat{\theta}_{unbiased} = \frac{n+1}{n}\max(X_1, \ldots, X_n)}$$

is an unbiased estimator of θ*.*

Remark 22.16. This is the classic "German tank problem": estimating the size of a population of serial numbers from the largest one observed. Notice that the naive estimator was not wrong so much as *systematically* wrong, and a systematic error is exactly the kind you can correct once you have computed it.