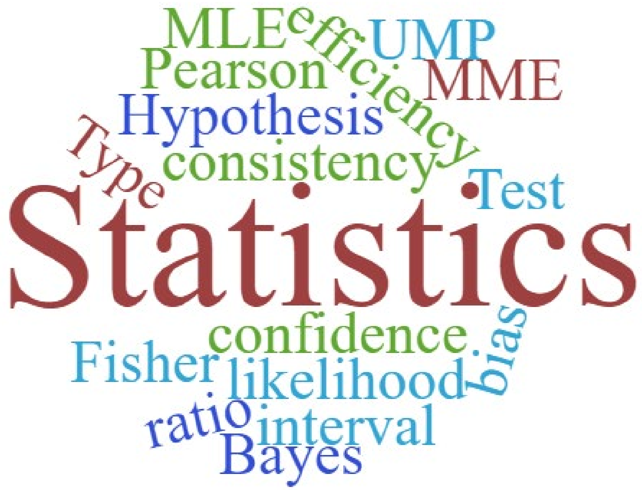
Math 181A: Mathematical Statistics



David Quarfoot
Associate Teaching Professor
Email: dquarfoot@ucsd.edu
Office: APM 5220

Learning Goals:

- See the need for the estimation of the parameters of a model
- Define iid random variables (AKA random sample) and visualize what this means
- Use the Law of Large Numbers to notice that sample moments are approximately equal to theoretical moments, and from this, devise a technique known as method of moments estimators (MME)
- Find MMEs for both discrete and continuous RVs

Method of Moments Estimators/Estimates (MME)

Some questions to get us thinking:

- Suppose you believe the IQ scores of UCSD students are normally distributed via $N(\mu, 15^2)$. How would you estimate μ assuming you can't ask everyone to take an IQ test?

- Suppose a hat contains identical tokens labeled $1, 2, \ldots, N$. You pull one out at random and get the number 3. What's your best estimate for N?

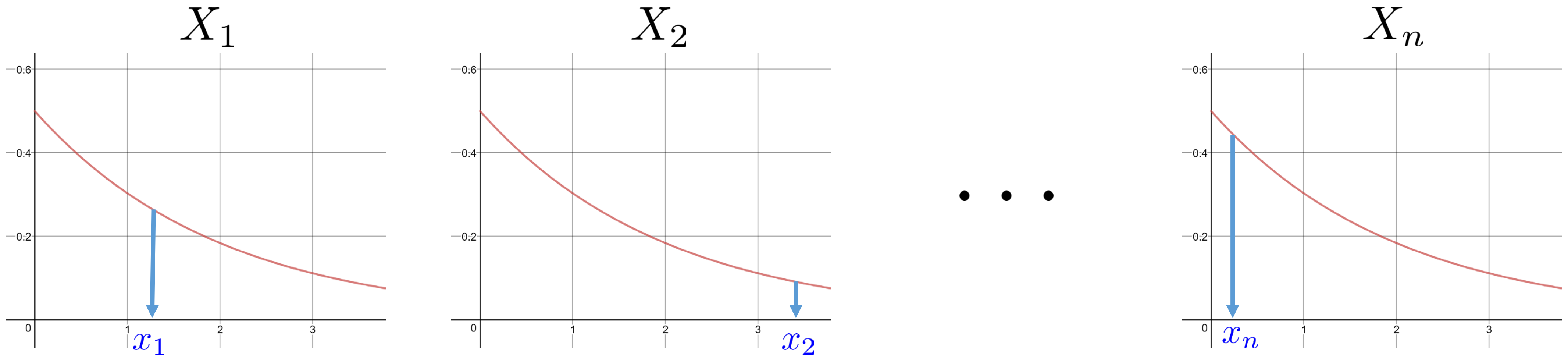
Estimation: The act of using data to guess at the parameters (e.g., $\mu, \sigma, \lambda, p, N$, etc.) of models ($N(\mu, \sigma^2), Pois(\lambda)$, etc.) believed to have generated those data.

There are many ways to estimate based on what you want to prioritize.

Setting Notation

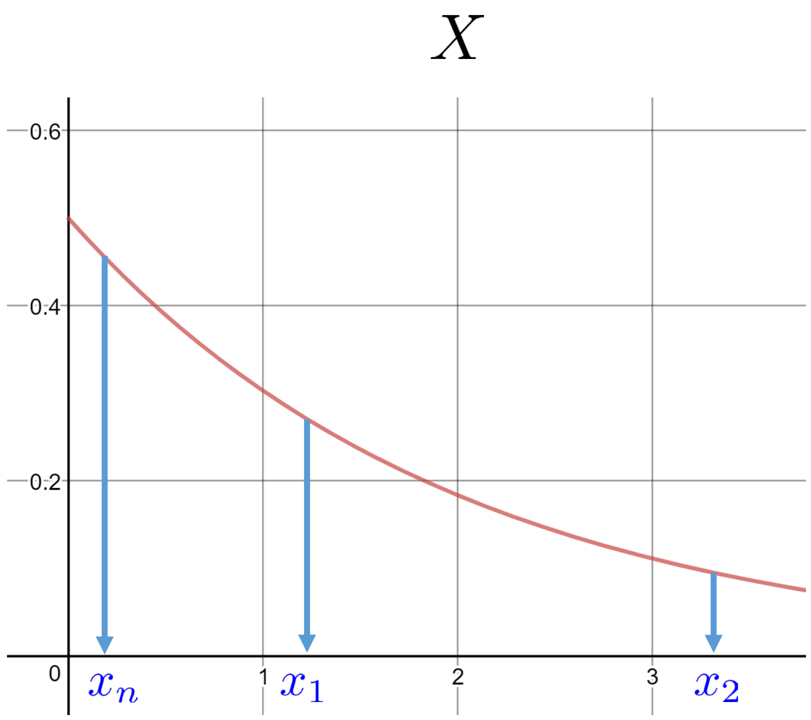
We will often write: "Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d., or iid)", or equivalently "Let $X_1, \ldots, X_n$ be a *random sample*".

One way to think about iid data:



Use lowercase $(x_1, x_2, \ldots)$ to think about the actual values in your sample.

... Or Think This Way



"Identically distributed" is nice because it forces:

$$E(X_1) = E(X_2) = \cdots = E(X_n) = E(X)$$
$$Var(X_1) = Var(X_2) = \cdots = Var(X_n) = Var(X)$$

"Independence" is nice because it forces ($i \neq j$):

$$E(X_i \cdot X_j) = E(X_i) \cdot E(X_j)$$
$$Var(X_i + X_j) = Var(X_i) + Var(X_j)$$

The Idea of Estimation

This notation stands for a pmf or pdf that depends on some parameters $\theta_1, \ldots, \theta_k$. Some write $f(x|\theta_1, \ldots, \theta_k)$ (OK to use in this class).

For example, the normal distribution has density $f(x|\mu, \sigma^2)$ (here $\theta_1 = \mu, \theta_2 = \sigma^2$, so $k = 2$).

$$X \sim \boxed{f(x; \theta_1, \theta_2, \ldots, \theta_k)} \xrightarrow[\text{Draw a random sample}]{} x_1, x_2, \ldots, x_n$$

Use the data to estimate the θ_is

$$\text{IQ Example: } X \sim N(\mu, 15^2) \xrightarrow[\text{Draw a random sample}]{} x_1, x_2, \ldots, x_n$$

Earlier, your intuition suggested $\mu \approx \dfrac{\sum x_i}{n} = \overline{x} = \overline{X}$

We need a back-deriving approach that moves beyond intuition and works when multiple parameters must be estimated (e.g., both μ and σ^2 in $N(\mu, \sigma^2)$).

Method of Moments Estimators/Estimates (MME)

History: Chebyshev (1887)
Pearson (1930s)

	"first moments"	"second moments"		"i^{th} moments"	
Theoretical Moments: Involve distribution parameters	$E(X)$	$E(X^2)$	$\cdots$	$E(X^i)$	$\cdots$
		$E(X^2) = Var(X) + E(X)^2$			
Example: $X \sim N(\mu, \sigma^2)$	$E(X) = \mu$	$= \sigma^2 + \mu^2$			
Sample Moments: Involve sample data	$\frac{1}{n}\sum_{j=1}^{n} X_j$	$\frac{1}{n}\sum_{j=1}^{n} X_j^2$	$\cdots$	$\frac{1}{n}\sum_{j=1}^{n} X_j^i$	$\cdots$
	roughly equal!	roughly equal!		roughly equal!	

Key Idea: LLN says $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{p} E(X)$. Similarly, $\frac{1}{n}\sum_{j=1}^{n} X_j^i \xrightarrow{p} E(X^i)$.

MME: Set $E(X^i) = \frac{1}{n}\sum_{j=1}^{n} X_j^i$ (for $i = 1, 2, \ldots$) to get formulas linking the data and the parameters. These show how to estimate the parameters using the data!

Our First MME

Let $X_1, \ldots, X_n$ be iid based on $X \sim Exp(\lambda)$. Find the MME for λ.

With only one unknown (λ), we try MME using first moments.

Recall that $E(X) = \dfrac{1}{\lambda}$. Equating moments gives $\dfrac{1}{\widehat{\lambda}} = \overline{X}$.

Put hats on the unknown parameter(s) when doing this step. This notation signals an *estimator*.

Thus, $\widehat{\lambda} = \dfrac{1}{\overline{X}}$, assuming $\overline{X} \neq 0$.

A statistic, or estimator, is any function of a sample $X_1, \ldots, X_n$ (here: $1/\overline{X}$) whose objective is to approximate a parameter. If λ is the parameter, $\widehat{\lambda}$ is the estimator. An *estimate* is the number you get from an estimator when actual data are plugged in (compare: $g(x)$ vs. $g(12)$).

Your First MME: Notational Variety

As always, outside of the specified interval, we assume the density function has height 0.

Let $y_1, y_2, \ldots, y_n$ be a random sample from the density $f_Y(y;\theta) = \dfrac{2y}{\theta^2}, 0 \leq y \leq \theta$. Find the MME for θ.

We need $E(Y) = \displaystyle\int_0^\theta y \cdot \frac{2y}{\theta^2}\, dy = \frac{2}{\theta^2}\left(\frac{1}{3}y^3\right)\Big|_0^\theta = \frac{2}{3}\theta.$

Equating moments gives $\overline{y} = \dfrac{2}{3}\widehat{\theta}$. Thus, $\boxed{\widehat{\theta} = \dfrac{3}{2}\overline{y}}$.

Alternative notation – Left: $\widehat{\theta}_{\mathrm{MME}}, \theta_e$; Right: $\dfrac{3}{2n}\displaystyle\sum_{i=1}^n y_i, \frac{3}{2}\overline{Y}, \frac{3}{2}\overline{Y}_n$

With actual data (say, $y_1 = 3, y_2 = 4, y_3 = 1.5, y_4 = 3.5$), find an estimate:

$$\widehat{\theta} = \frac{3}{2} \cdot \frac{3 + 4 + 1.5 + 3.5}{4} = 4.5.$$

Step Up The Difficulty!

With 2 parameters, we need 2 moments.
With k parameters, use k moments.

Find the MME for μ and σ^2 for $X \sim N(\mu, \sigma^2)$ assuming both are unknown.

From before: $E(X) = \mu$ and $E(X^2) = Var(X) + E(X)^2 = \sigma^2 + \mu^2$.

Equating first moments: $\boxed{\widehat{\mu} = \overline{X}}$.

This confirms your guess from the first slide!

Equating second moments gives: $\widehat{\sigma^2} + \widehat{\mu}^2 = \frac{1}{n}\sum X_i^2 \overset{\text{def.}}{=} \overline{X^2}$

Careful: $\overline{X}^2 \neq \overline{X^2}$

Since $\widehat{\mu} = \overline{X}$, we get $\widehat{\sigma^2} = \overline{X^2} - \overline{X}^2 \overset{\text{alg.}}{=} \boxed{\frac{1}{n}\sum \left(X_i - \overline{X}\right)^2}$.

At home, try to simplify this expression to the one before.

Helpful definition: $\overline{f(X)} = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$

Example: $\overline{\ln X} = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$

Further Practice

I suggest working the examples that follow before attempting the homework.

Back to the Beginning

Suppose a hat contains identical tokens labeled $1, 2, \ldots, N$. You pull one out at random and get the number 3. What's the MME for N?

Set up the probability function: $P(X = k) = \dfrac{1}{N}$ for $k = 1, \ldots, N$.

Note that $E(X) = \sum_k k \cdot P(X = k) = \sum_{k=1}^{N} k \cdot \dfrac{1}{N} = \dfrac{1}{N} \sum_{k=1}^{N} k = \dfrac{1}{N} \dfrac{N(N+1)}{2}$.

Equating first moments gives $\overline{X} = \dfrac{\widehat{N} + 1}{2}$. Thus, $\widehat{N} = 2\overline{X} - 1$.

With one piece of data, $\overline{X} = 3$. So $\boxed{\widehat{N} = 2 \cdot 3 - 1 = 5}$.

Intuitively, the number you pull will be, on average, in the middle of the numbers possible. The middle of $\{1, 2, 3, 4, 5\}$ is 3, so this MME just reverses the logic to get from 3 to 5. MMEs generally have an intuitive feel.

Your Turn: A Break In Continuity

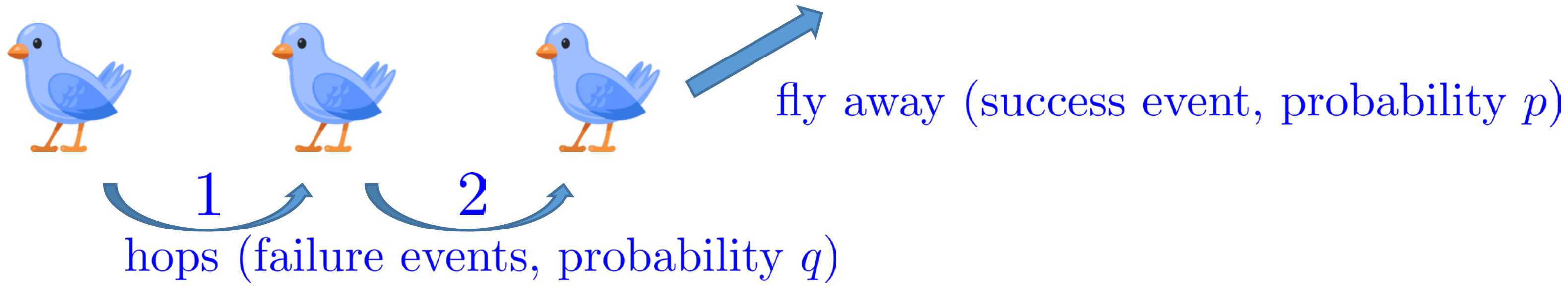
Let X be a discrete RV modeled by $P(X=0)=\frac{2}{3}\theta, P(X=1)=\frac{1}{3}\theta$, $P(X=2)=\frac{2}{3}(1-\theta)$, and $P(X=3)=\frac{1}{3}(1-\theta)$ where $0\le\theta\le 1$. Given the eight independent observations $(3,0,0,2,1,2,3,1)$, find the MME for θ.

Note that $E(X)=\sum_x x\cdot P(X=x)=0\cdot\frac{2}{3}\theta+1\cdot\frac{1}{3}\theta+2\cdot\frac{2}{3}(1-\theta)+3\cdot\frac{1}{3}(1-\theta)$

Simplifying gives $E(X)=\frac{7}{3}-2\theta$. Equate first moments: $\overline{X}=\frac{7}{3}-2\widehat{\theta}$.

So, $\widehat{\theta}=\frac{7}{6}-\frac{\overline{X}}{2}$. Our data give $\overline{X}=\frac{3}{2}$, so $\boxed{\widehat{\theta}=\frac{5}{12}}$.

Your Turn: Hop To It!



Researchers counted the numbers of hops 114 random birds made between landing and taking off again. If these data fit a Geometric model, find the MME for p, the probability a bird will fly away. Then make a table of expected counts based on your value of p.

Numbers of hops	Number of birds
0	48
1	31
2	20
3	9
4	6

If $X \sim Geom(p)$, then $E(X) = \dfrac{1}{p}$.

Equate first moments: $\overline{X} = \dfrac{1}{\widehat{p}}$.

Thus, $\widehat{p} = \dfrac{1}{\overline{X}} = \dfrac{n}{\sum X_i}$. Here, $\widehat{p} = \dfrac{114}{48 \cdot 1 + 31 \cdot 2 + 20 \cdot 3 + 9 \cdot 4 + 6 \cdot 5} \approx \boxed{0.483}$.

We add 1 to each hop count to include the success (flying way) in addition to the failures.

To find the number of birds we expect to make exactly 0 hops before take-off, multiply 114 by $P(X = 1)$ (immediately choose flight!)

Since $P(X = k) = (1 - p)^{k-1}p$, we get $114 \cdot p \approx 114 \cdot 0.483 \approx 55.062$ birds.

Similarly, we expect $114qp \approx 28.467$ birds to make exactly 1 hop.

Continuing in this way, we get the below table. Note that we get the 8.146 not by doing $114 \cdot P(X = 4)$, but by making the column sum to 114. Since the Geometric distribution has infinite support, the last column won't sum to 114 unless we force it to. The data roughly seem to follow $Geom(p)$.

Numbers of hops	Number of birds	MME predictions
0	48	55.062
1	31	28.467
2	20	14.717
3	9	7.608
4	6	8.146