Topic 3: Estimation

Ethan P. Marzban University of California, Santa Barbara PSTAT 120B

Outline

1. Consistency

2. Convergence in Distribution

3. Method of Moments

Consistency

Consistency

Definition (Consistent Estimator)

An estimator $\widehat{\theta}_n$ is said to be a **consistent** estimator for θ if

$$\widehat{\theta}_n \xrightarrow{p} \theta$$

That is, if either of the two equivalent conditions hold for any $\varepsilon > 0$:

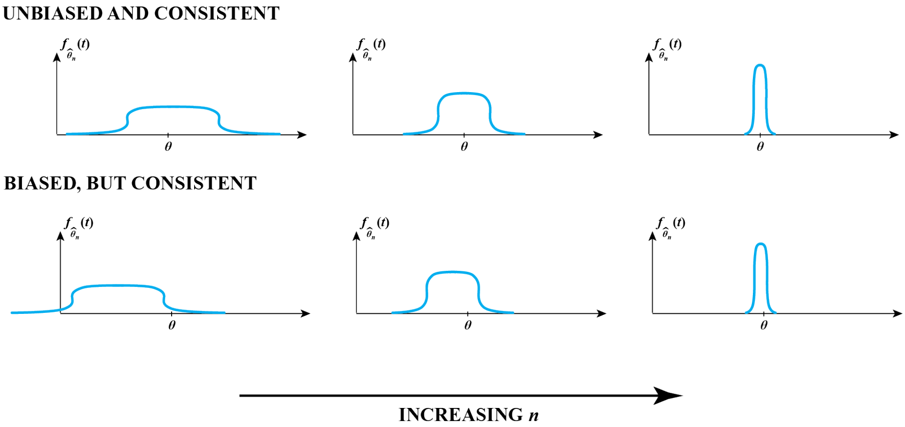
$$\lim_{n \to \infty} \mathbb{P}(|\widehat{\theta}_n - \theta| \geq \varepsilon) = 0$$

$$\lim_{n \to \infty} \mathbb{P}(|\widehat{\theta}_n - \theta| < \varepsilon) = 1$$

Interpretation

- Recall the interpretation of the definition of consistency: as our sample size gets larger, we want to be more and more *certain* that our estimator $\widehat{\theta}_n$ remains *close* to θ.
- Also, recall that it's possible for an estimator to be biased (for a finite sample size) but also consistent.

Bias and Consistency



Convergence in Probability

- By the way, let me quickly explain the notation $\xrightarrow{p}$.
- We use this notation to indicate consistency, because consistency is actually related to something known as **convergence in probability** (which applies to any arbitrary sequence of random variables - not just estimators!)

Convergence in Probability

Definition (Convergence in Probability)

A sequence $\{X_n\}_{n \geq 0}$ of random variables is said to **converge in probability** to a constant x if for every $\varepsilon > 0$ either of the equivalent conditions hold:

$$\lim_{n \to \infty} \mathbb{P}(|X_n - x| \geq \varepsilon) = 0$$

$$\lim_{n \to \infty} \mathbb{P}(|X_n - x| < \varepsilon) = 1$$

Convergence in probability is notated as

$$X_n \xrightarrow{p} x$$

Weak Law of Large Numbers

Theorem (Weak Law of Large Numbers)

Given an i.i.d. sample $\{Y_i\}_{i=1}^{n}$ from a population with (unknown) mean μ and finite variance $\sigma^2 < \infty$,

$$\overline{Y}_n \xrightarrow{p} \mu$$

- In the language of convergence in probability, the WLLN states that the sample mean converges in probability to the population mean.
- In the language of consistency, the WLLN asserts that the sample mean is a consistent estimator for the population mean.

Proof

- What we want to show is that, for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) = 0$$

- First note that, by virtue of being a probability,

$$0 \leq \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon)$$

- Additionally, by Chebyshev's Inequality,

$$\mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\overline{Y}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Proof

- So, combining these two statements, we have

$$0 \leq \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

- Note that $[\sigma^2/(n\varepsilon^2)] \to 0$ as $n \to \infty$; additionally, $0 \to 0$ as $n \to \infty$. Hence, by the Squeeze Theorem (from Calculus),

$$\lim_{n \to \infty} \mathbb{P}(|\overline{Y}_n - \mu| \geq \varepsilon) = 0$$

which, by definition, implies

$$\overline{Y}_n \xrightarrow{p} \mu$$

Chalkboard Example

Example

Given an i.i.d. sample $\{Y_i\}_{i=1}^{n}$ from a distribution with mean μ and finite variance $\sigma^2 < \infty$, we define the **second sample moment** to be

$$M_{2,n} := \frac{1}{n} \sum_{i=1}^{n} Y_i^2$$

Does $M_{2,n}$ converge in probability to a constant? If so, find that constant.

Result

Theorem (Consistency and Unbiasedness, I)

Consider an unbiased estimator $\widehat{\theta}_n$ for θ. Then, $\widehat{\theta}_n$ is a consistent estimator for θ if $\lim_{n \to \infty} \mathrm{Var}(\widehat{\theta}_n) = 0$.

- The proof of this theorem utilizes tools similar to those used in the proof of the WLLN; I encourage you to walk through it on your own.
- Also, if we want to use this theorem to prove consistency, we must be careful to check that our estimator is unbiased! That is, if we have a *biased* estimator whose variance goes to 0 as $n \to \infty$, we CANNOT definitively conclude that it is a consistent estimator.

Chalkboard Example

Example

Consider an i.i.d. sample $\{Y_i\}_{i=1}^n$ from a population with mean μ and finite variance $\sigma^2 < \infty$. Show that S_n^2, the sample variance, is a consistent estimator for σ^2.

Properties

Theorem (Properties of Convergence in Probability)

Suppose that $X_n \xrightarrow{p} x$ and $Y_n \xrightarrow{p} y$. Then:

(I) $(X_n + Y_n) \xrightarrow{p} (x + y)$

(II) $(X_n \cdot Y_n) \xrightarrow{p} (x \cdot y)$

(III) $(X_n / Y_n) \xrightarrow{p} (x/y)$ provided that $y \neq 0$

(IV) **Continuous Mapping Theorem**: $g(X_n) \xrightarrow{p} g(x)$ for any real-valued function.

Example

Example

Consider an i.i.d. sample $\{Y_i\}_{i=1}^{n}$ from a population with mean μ and finite variance $\sigma^2 < 0$. Propose a consistent estimator for μ^2, and show explicitly that your estimator *is* consistent.

Leadup

- That's pretty much all I want to say on consistency and convergence in probability.
- If you are planning on pursuing a career in the Actuarial Sciences or Financial Mathematics, you will most *certainly* be using the notion of convergence in probability a lot. (Even Data Scientists find themselves faced with questions of consistency from time to time!)
- I would like to take some time and mention that there is another type of convergence...

Convergence in Distribution

Convergence in Distribution

Definition (Convergence in Distribution)

Consider a sequence $\{X_n\}_{n \geq 1}$ of random variables, where $F_{X_n}(\cdot)$ denotes the CDF of X_n for any $n \in \mathbb{N}$. Consider another random variable X with CDF $F_X(\cdot)$: we say that X_n **converges in distribution** to X, notated $X_n \rightsquigarrow X$ if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

for every x such that both $F_{X_{(n)}}(x)$ and $F_X(x)$ are continuous.

Convergence in Distribution

- Basically, if the distributions of X_n as $n \to \infty$ get closer and closer to the distribution of X, we say that X_n converges in distribution to X.
- If the distribution of X has a familiar name, it's common to replace X in the notation $X_n \rightsquigarrow X$ with the name of the distribution. For example, you might see things like

$$X_n \rightsquigarrow \mathcal{N}(0, 1)$$

to mean

$$\lim_{n \to \infty} F_{X_n}(x) = \Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, \mathrm{d}z$$

Convergence in Distribution

- As a concrete example, consider a collection of independent random variables $\{X_n\}$ where

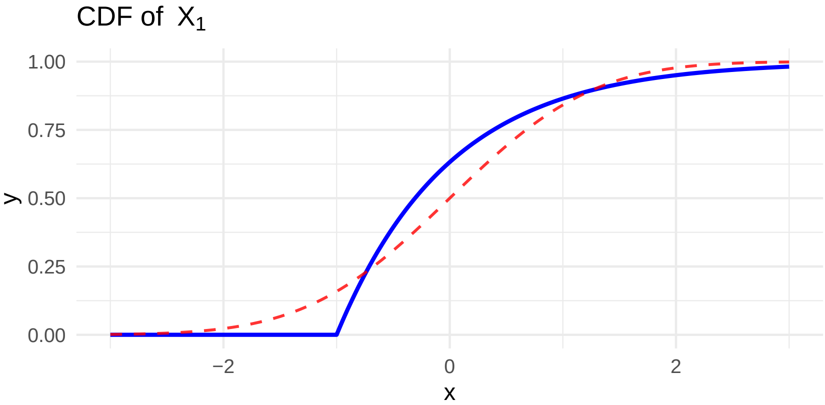
$$f_{X_n}(x) = \frac{n^{n/2}}{(n-1)!}(x + \sqrt{n})^{n-1} e^{-(x\sqrt{n}+u)} \cdot \mathbb{1}_{\{x \geq -\sqrt{n}\}}$$

  - Don’t worry about where I got this from - it’ll actually make sense in a minute!

- Though this distribution’s CDF doesn’t have a simple closed-form expression, we can use a computer software to plot it.

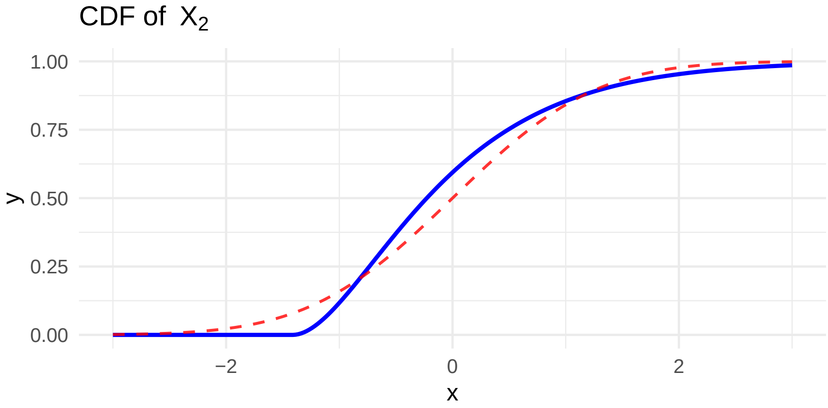
Convergence in Distribution

- For instance, the CDF of X_1 looks like this:



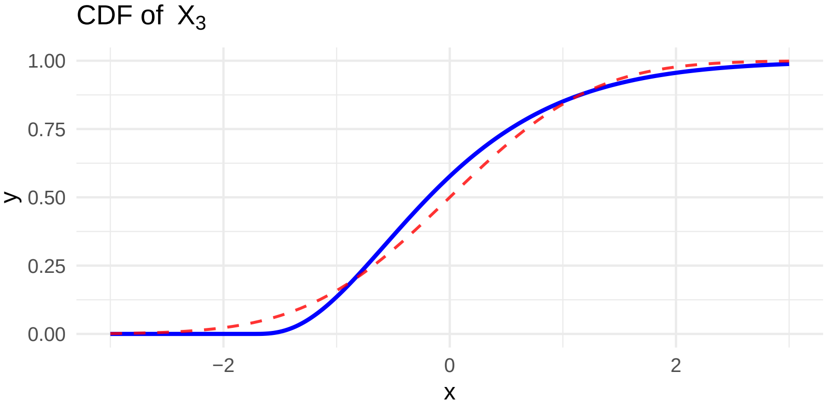
Convergence in Distribution

- The CDF of X_2 looks like this:



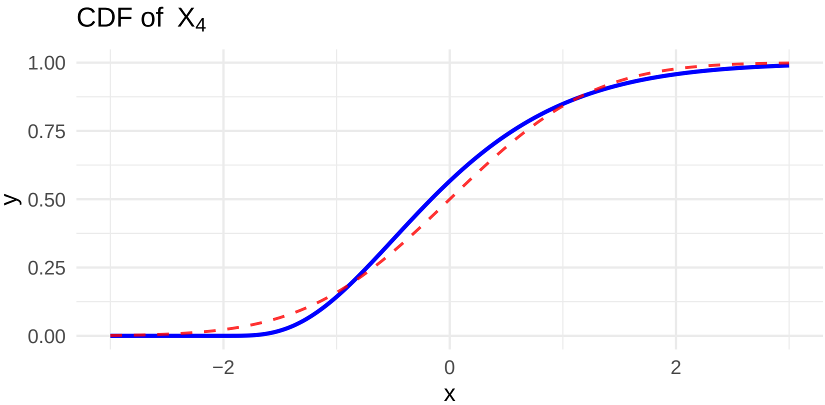
Convergence in Distribution

- The CDF of X_3 looks like this:



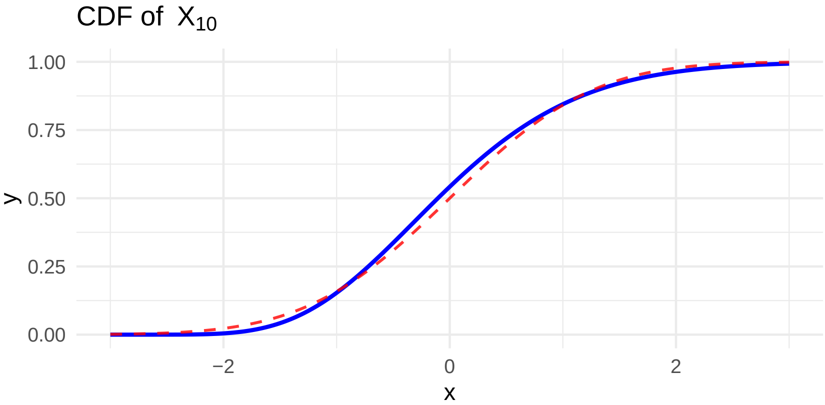
Convergence in Distribution

- The CDF of X_4 looks like this:



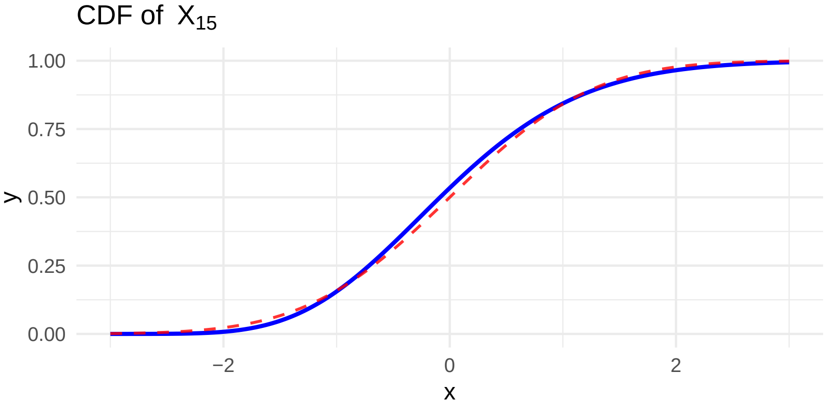
Convergence in Distribution

- The CDF of X_{10} looks like this:



Convergence in Distribution

- The CDF of X_{15} looks like this:



Convergence in Distribution

- As $n \to \infty$, doesn't it look like the blue curve gets closer and closer to the red dashed curve?
- This is exactly what we mean by convergence in distribution: the sequence of random variables $\{X_n\}$ is converging in distribution to the distribution whose CDF is given by the red dashed curve.
    - If you're curious, the red dashed CDF is actually just $\Phi(\cdot)$.

Central Limit Theorem

Theorem (Central Limit Theorem)

Given an i.i.d. collection $\{Y_i\}_{i=1}^n$ of random variables with $\mathbb{E}[Y_i] = \mu$ and $\mathrm{Var}(Y_i) = \sigma^2 < \infty$, define

$$U_n := \frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma} = \frac{(\sum_{i=1}^n Y_i) - n\mu}{\sigma\sqrt{n}}$$

Then $U_n \rightsquigarrow \mathcal{N}(0, 1)$. That is,

$$\lim_{n \to \infty} \mathbb{P}(U_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, \mathrm{d}z$$

Central Limit Theorem

- By the way, this is how I got the peculiar-looking density from before!
- Specifically, I set $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(2)$ so that $\overline{Y}_n \sim \text{Gamma}(n, 2/n)$. The density of X_n in the above example is just the exact density (derived using the CDF method) of

$$X_n := \frac{\sqrt{n}(\overline{Y}_n - 2)}{2} = \left(\frac{\sqrt{n}}{2} \right) \overline{Y}_n - \sqrt{n}$$

  which, by the CLT, converges in distribution to the $\mathcal{N}(0, 1)$ distribution as $n \to \infty$.

Example

Example

The weight of a randomly-selected *GauchoPlay*-brand toy (in oz) follows some distribution with mean 8.2 oz and standard deviation 0.8 oz. If a random sample of 64 *GauchoPlay* toys is taken, what is the approximate probability that the average weight of toys in the sample lies within 0.1 oz of the true average weight of 8.2 oz?

Solution

- Let Y_i denote the weight of a randomly-selected *GauchoPlay*-brand toy. We don't know exactly what the distribution of Y_i is, but we do know that $\mathbb{E}[Y_i] = 8.2$ and $\mathrm{Var}(Y_i) = 0.8^2 = 1.6$.
- The quantity we seek is

$$\mathbb{P}(|\overline{Y}_{64} - 8.2| < 0.1) = \mathbb{P}(8.1 < \overline{Y}_{64} < 8.3)$$

- Now, the issue is that we don't know what distribution $\overline{Y}_{64}$ follows. However, by the CLT, we do know that

$$\frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma} = \frac{\sqrt{n}(\overline{Y}_n - 8.2)}{0.8} \rightsquigarrow \mathcal{N}(0, 1)$$

Solution

- So, what this means is that if our sample size n is large enough, the quantity
$$\frac{\sqrt{n}(\overline{Y}_n - 8.2)}{0.8}$$
will be well-approximated by something following the standard normal distribution.
- A sample size of $n = 64$ is relatively large, so we can safely conclude that
$$\frac{\sqrt{64}(\overline{Y}_{64} - 8.2)}{0.8} = \frac{\overline{Y}_{64} - 8.2}{0.1} \overset{\mathrm{d}}{\approx} \mathcal{N}(0, 1)$$

Solution

- Therefore, we write

$$\begin{aligned}
\mathbb{P}(|\overline{Y}_{64} - 8.2| < 0.1) &= \mathbb{P}(-0.1 < \overline{Y}_{64} - 8.2 < 0.1) \\
&= \mathbb{P}\left(-\frac{0.1}{0.1} < \frac{\overline{Y}_{64} - 8.2}{0.1} < \frac{0.1}{0.1} \right) \\
&= \mathbb{P}\left(\frac{\overline{Y}_{64} - 8.2}{0.1} < 1 \right) - \mathbb{P}\left(\frac{\overline{Y}_{64} - 8.2}{0.1} < -1 \right) \\
&\approx \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 68.26\%
\end{aligned}$$

Notes

- A couple of notes:
- For this class we won't concern ourselves too much with how large a sample size must be in order for the CLT to apply.
  - As a rule of thumb, $n > 20$ is usually sufficient to invoke the CLT however other cutoffs are sometimes used ($n > 10$, or even $n > 35$).
- Also, the CLT can be used to construct intervals that cover certain events with a given probability - take a look at MT02 Practice Problem 7, which we will go over together tomorrow (Tuesday).

Leadup

- Alright, so up until now we've concerned ourselves primarily with determining what properties a "good" estimator might possess.
- We started off with the notions of unbiasedness and small MSE (both of which are good!), and then talked about consistency as yet another potential property we'd like an estimator to satisfy.
- But, in everything we've done so far, we've been provided with an estimator and asked to test its efficacy.
- We now turn our attention to answering the question - how do we go about *constructing* estimators?

Method of Moments

Method of Moments

- Allow me to introduce the Method of Moments (MoM) by way of an example.
- Say we have $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim}$ Unif$[0, \theta]$ for some (unknown) $\theta > 0$, and suppose we want to construct an estimator for θ.
- Note that $\mathbb{E}[Y_1] = \theta/2$.
- Now, by the WLLN, it makes sense to assume that the sample mean $(\overline{Y}_n)$ will be relatively close to the population mean $(\mu := \theta/2)$.
- So, what if we set the sample mean to the population mean, solve for θ, and call the resulting quantity an estimator for θ?

Method of Moments

- That is, we assert our estimator $\widehat{\theta}_{\mathrm{MM}}$ must satisfy

$$\overline{Y}_n = \frac{\widehat{\theta}_{\mathrm{MM}}}{2}$$

- This in turn yields the estimator

$$\widehat{\theta}_{\mathrm{MM}} = 2\overline{Y}_n$$

Method of Moments

- First and foremost, is this a “good” estimator?
- Well, note that it is unbiased:

$$\mathbb{E}[\widehat{\theta}_{\mathrm{MM}}] = \mathbb{E}[2\overline{Y}_n] = 2 \cdot \mathbb{E}[\overline{Y}_n] = 2 \cdot \frac{\theta}{2} = \theta \ \checkmark$$

- Additionally, it is a consistent estimator for θ:
  - By the WLLN, $\overline{Y}_n \xrightarrow{p} (\theta/2)$
  - Hence, by the Continuous Mapping Theorem (CMT) with $g(x) = 2x$, we have that

$$\widehat{\theta}_{\mathrm{MM}} = 2\overline{Y}_n \xrightarrow{p} 2 \cdot \frac{\theta}{2} = \theta \ \checkmark$$

Moments

- So, it seems we've constructed a pretty good estimator for θ!
- Before posing the general scheme for uing this strategy to construct estimators, let me quickly remind us of some terminology.
- If Y_i follows some distribution, recall that the ***k*^th^ population moment** of that distribution, notated μ_k, is given by $\mu_k := \mathbb{E}[Y_i^k]$.
- Given a sample $\{Y_i\}_{i=1}^n$, the ***k*^th^ sample moment**, notated $M_{k,n}$, is defined as

$$M_{k,n} := \frac{1}{n} \sum_{i=1}^{n} Y_i^k$$

Method of Moments, One-Parameter Case

- We're finally ready to pose the first version of the Method of Moments!
- Again, remember that this method is designed to yield an estimator $\widehat{\theta}_n$ for θ.

1. Equate the first sample moment (i.e. the sample mean) to the first population moment.
2. Replace all instances of θ with $\widehat{\theta}_{\mathrm{MM}}$, and then solve for $\widehat{\theta}_{\mathrm{MM}}$ in terms of the first sample moment.

Method of Moments, General Case

- What do we do if our distribution has multiple parameters?
- Well, we follow a very similar procedure:

1. Set $\mu_k = M_{k,n}$ for $k = 1, \cdots p$ where p is the number of parameters.
2. Replace all instances of θ_k with $\widehat{\theta}_{\mathrm{MM},k}$ (again, for $k = 1, \cdots, p$), and then solve the **system** of equations for $\widehat{\theta}_{\mathrm{MM},k}$ in terms of the first sample moment.

Example

Example

Consider $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown parameters. Find $\widehat{\mu}_{\text{MM}}$ and $\widehat{\sigma^2}_{\text{MM}}$, the Method of Moments estimators for μ and σ^2, respectively.

Example

- Since we have two parameters, we will end up with a system of two equations to solve.
- Furthermore, we'll need both the first and second population moments, so let's compute those.
- We know that $\mu_1 := \mathbb{E}[Y_1] = \mu$; additionally, $\mu_2 = \mathrm{Var}(Y_1) + \mu_1^2 = \sigma^2 + \mu^2$.
- Therefore, $\widehat{\mu}_{\mathrm{MM}}$ and $\widehat{\sigma^2}_{\mathrm{MM}}$ must satisfy the system of equations

$$\begin{cases} \widehat{\mu}_{\mathrm{MM}} & = \overline{Y}_n \\ \widehat{\mu}_{\mathrm{MM}}^2 + \widehat{\sigma^2}_{\mathrm{MM}} & = M_{2,n} \end{cases}$$

Example

- The first equation immediately gives us the Method of Moments estimator for μ:

$$\widehat{\mu}_{\mathrm{MM}} = \overline{Y}_n$$

- Substituting this into the second equation yields the Method of Moments estimator for σ^2:

$$\begin{aligned}\widehat{\sigma}^2{}_{\mathrm{MM}} &= M_{2,n} - \widehat{\mu}^2_{\mathrm{MM}} \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - (\overline{Y}_n)^2\end{aligned}$$

Example

Example

An i.i.d. sample of 8 cats is taken from the population of all DSH cats; their observed weights (in lbs) are given by

$$(8.2, 9.2, 8.7, 10.2, 15.2, 7.2, 16.1, 9.5)$$

Based on this sample, what is an appropriate estimate for the true average weight of all DSH cats? What is an appropriate estimate for the true variance of weights among all DSH cats? Assume the weight of a randomly-selected DSH cat is normally distributed.

Example

- Let's use the Method of Moments estimator.
- Define Y_i to be the weight of a randomly-selected DSH cat; the problem then tells us that

$$Y_1, Y_2, \cdots, Y_8 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

  where μ denotes the true average weight of all DSH cats and σ^2 denotes the true variance among all DSH cat weights.
- All we need to do is to plug into our MoM estimators we derived in the previous example!

Example

$$
\begin{aligned}
\vec{\boldsymbol{y}} &= (8.2, 9.2, 8.7, 10.2, 15.2, 7.2, 16.1, 9.5) \\
\overline{y}_n &= \frac{1}{8}(8.2 + 9.2 + 8.7 + \cdots + 16.1 + 9.5) = 10.5375 \\
m_{2,8} &= \frac{1}{8}(8.2^2 + 9.2^2 + 8.7^2 + \cdots + 16.1^2 + 9.5^2) = 120.4938 \\
\implies \widehat{\mu}_{\text{MM}} &= \text{10.5375 lbs} \\
\widehat{\sigma^2}_{\text{MM}} &= 120.4938 - (10.5375)^2 \approx \text{9.4549 lbs}^2
\end{aligned}
$$

Zero-Moment

- If a particular moment is equal to zero, move to the next moment.
- For example, if $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[-\theta, \theta]$ then $\mathbb{E}[Y_1] = 0$. Naturally, $\widehat{\theta}_{\text{MM}}$ doesn't satisfy the equation $\widehat{\theta}_{\text{MM}} = 0$; rather, you should consider finding the *second* population moment $\mathbb{E}[Y_i^2]$ in terms of θ, replace θ with $\widehat{\theta}_{\text{MM}}$, and solve.

Result

Theorem (Consistency of MoM Estimators)

In general, method of moments estimators are consistent.

- There are certain exceptions, but we won’t concern ourselves with those in this class.
- Please keep in mind: though you are welcome to use this result (unless a problem explicitly asks you not to), you can usually appeal to a combination of the WLLN and CMT to establish consistency for MoM estimators.