

Math Camp

Justin Grimmer

Associate Professor
Department of Political Science
University of Chicago

September 15th, 2017

Questions?

Today:

- 1) Properties of Expectations
- 2) Changing Coordinates
- 3) Moment Generating Functions
- 4) Inequalities
- 5) Convergence

Iterated Expectations

Proposition

Suppose X and Y are random variables. Then

$$E[X] = E[E[X|Y]]$$

- Inner Expectation is $E[X|Y] = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx$.
- Outer expectation is over y .

Iterated Expectations

Proof.



Iterated Expectations

Proof.

$$E[E[X|Y]] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy$$



Iterated Expectations

Proof.

$$\begin{aligned} E[E[X|Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dy dx \end{aligned}$$



Iterated Expectations

Proof.

$$\begin{aligned} E[E[X|Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y) dy dx \end{aligned}$$



Iterated Expectations

Proof.

$$\begin{aligned} E[E[X|Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y) dy dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \end{aligned}$$



Iterated Expectations

Proof.

$$\begin{aligned}E[E[X|Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dy dx \\&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y) dy dx \\&= \int_{-\infty}^{\infty} x f_X(x) dx \\&= E[X]\end{aligned}$$



Iterated Expectations

Definition

Suppose Y is a continuous random variable with $Y \in [0, 1]$ and pdf of Y given by

$$f(y) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1-1} (1-y)^{\alpha_2-1}$$

*Then we will say Y is a **Beta** distribution with parameters α_1 and α_2 . Equivalently,*

$$Y \sim \text{Beta}(\alpha_1, \alpha_2)$$

- Beta is a distribution on **proportions**
- Beta is a special case of the **Dirichlet** distribution
- $E[Y] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$

Iterated Expectations

Suppose

$$\begin{aligned}\pi &\sim \text{Beta}(\alpha_1, \alpha_2) \\ Y|\pi, n &\sim \text{Binomial}(n, \pi)\end{aligned}$$

What is $E[Y]$?

Iterated Expectations

Suppose

$$\begin{aligned}\pi &\sim \text{Beta}(\alpha_1, \alpha_2) \\ Y|\pi, n &\sim \text{Binomial}(n, \pi)\end{aligned}$$

What is $E[Y]$?

$$E[Y] = E[E[Y|\pi]]$$

Iterated Expectations

Suppose

$$\begin{aligned}\pi &\sim \text{Beta}(\alpha_1, \alpha_2) \\ Y|\pi, n &\sim \text{Binomial}(n, \pi)\end{aligned}$$

What is $E[Y]$?

$$\begin{aligned}E[Y] &= E[E[Y|\pi]] \\ &= \int_{-\infty}^{\infty} \sum_{j=0}^N \binom{N}{j} j p(j|\pi) f(\pi) d\pi\end{aligned}$$

Iterated Expectations

Suppose

$$\begin{aligned}\pi &\sim \text{Beta}(\alpha_1, \alpha_2) \\ Y|\pi, n &\sim \text{Binomial}(n, \pi)\end{aligned}$$

What is $E[Y]$?

$$\begin{aligned}E[Y] &= E[E[Y|\pi]] \\ &= \int_{-\infty}^{\infty} \sum_{j=0}^N \binom{N}{j} j p(j|\pi) f(\pi) d\pi \\ &= \int_{-\infty}^{\infty} N\pi f(\pi) d\pi\end{aligned}$$

Iterated Expectations

Suppose

$$\begin{aligned}\pi &\sim \text{Beta}(\alpha_1, \alpha_2) \\ Y|\pi, n &\sim \text{Binomial}(n, \pi)\end{aligned}$$

What is $E[Y]$?

$$\begin{aligned}E[Y] &= E[E[Y|\pi]] \\ &= \int_{-\infty}^{\infty} \sum_{j=0}^N \binom{N}{j} j p(j|\pi) f(\pi) d\pi \\ &= \int_{-\infty}^{\infty} N\pi f(\pi) d\pi \\ &= N \frac{\alpha_1}{\alpha_1 + \alpha_2}\end{aligned}$$

Change of Coordinates

Proposition

Suppose X is a random variable and $Y = g(X)$, where $g : \mathbb{R} \rightarrow \mathbb{R}$ that is a monotonic function.

Define $g^{-1} : \mathbb{R} \rightarrow \mathbb{R}$ such that $g^{-1}(g(X)) = X$ and is differentiable. Then,

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \text{ if } y = g(x) \text{ for some } x \\ &= 0 \text{ otherwise} \end{aligned}$$

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$F_Y(y) = P(Y \leq y)$$

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \end{aligned}$$

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(g(X) \leq y) \\&= P(X \leq g^{-1}(y))\end{aligned}$$

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(g(X) \leq y) \\&= P(X \leq g^{-1}(y)) \\&= F_X(g^{-1}(y))\end{aligned}$$

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(g(X) \leq y) \\&= P(X \leq g^{-1}(y)) \\&= F_X(g^{-1}(y))\end{aligned}$$

Now differentiating to get the pdf

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(g(X) \leq y) \\&= P(X \leq g^{-1}(y)) \\&= F_X(g^{-1}(y))\end{aligned}$$

Now differentiating to get the pdf

$$\begin{aligned}\frac{\partial F_Y(y)}{\partial y} &= \frac{\partial F_X(g^{-1}(y))}{\partial y} \\&= f_X(g^{-1}(y)) \frac{\partial g^{-1}(y)}{\partial y}\end{aligned}$$

Change of Coordinates

Proof.

Suppose $g(\cdot)$ is monotonically increasing (WLOG)

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(g(X) \leq y) \\&= P(X \leq g^{-1}(y)) \\&= F_X(g^{-1}(y))\end{aligned}$$

Now differentiating to get the pdf

$$\begin{aligned}\frac{\partial F_Y(y)}{\partial y} &= \frac{\partial F_X(g^{-1}(y))}{\partial y} \\&= f_X(g^{-1}(y)) \frac{\partial g^{-1}(y)}{\partial y}\end{aligned}$$

Then this is a pdf because $\frac{\partial g^{-1}(y)}{\partial y} > 0$.

Change of Coordinates

Suppose X is a random variable with pdf $f_X(x)$. Suppose $Y = X^n$. Find $f_Y(y)$.

Change of Coordinates

Suppose X is a random variable with pdf $f_X(x)$. Suppose $Y = X^n$. Find $f_Y(y)$.

Then $g^{-1}(x) = x^{1/n}$.

Change of Coordinates

Suppose X is a random variable with pdf $f_X(x)$. Suppose $Y = X^n$. Find $f_Y(y)$.

Then $g^{-1}(x) = x^{1/n}$.

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(Y)}{\partial y} \right|$$

Change of Coordinates

Suppose X is a random variable with pdf $f_X(x)$. Suppose $Y = X^n$. Find $f_Y(y)$.

Then $g^{-1}(x) = x^{1/n}$.

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(Y)}{\partial y} \right| \\ &= f_X(y^{1/n}) \frac{y^{\frac{1}{n}-1}}{n} \end{aligned}$$

Change of Coordinates

Suppose X is a random variable with pdf $f_X(x)$. Suppose $Y = X^n$. Find $f_Y(y)$.

Then $g^{-1}(x) = x^{1/n}$.

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(Y)}{\partial y} \right| \\ &= f_X(y^{1/n}) \frac{y^{\frac{1}{n}-1}}{n} \end{aligned}$$

We've used this to derive many of the pdfs

Change of Coordinates

Suppose X is a random variable with pdf $f_X(x)$. Suppose $Y = X^n$. Find $f_Y(y)$.

Then $g^{-1}(x) = x^{1/n}$.

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(Y)}{\partial y} \right| \\ &= f_X(y^{1/n}) \frac{y^{\frac{1}{n}-1}}{n} \end{aligned}$$

We've used this to derive many of the pdfs

- Normal distribution
- Chi-Squared Distribution

Moment Generating Functions

Definition

Suppose X is a random variable with pdf f . Define,

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx$$

We will call X^n the n^{th} moment of X

- By this definition $\text{var}(X) = \text{Second Moment} - \text{First Moment}^2$
- We are assuming that the integral converges

Moment Generating Functions

Proposition

Suppose X is a random variable with pdf $f(x)$. Call $M(t) = E[e^{tX}]$,

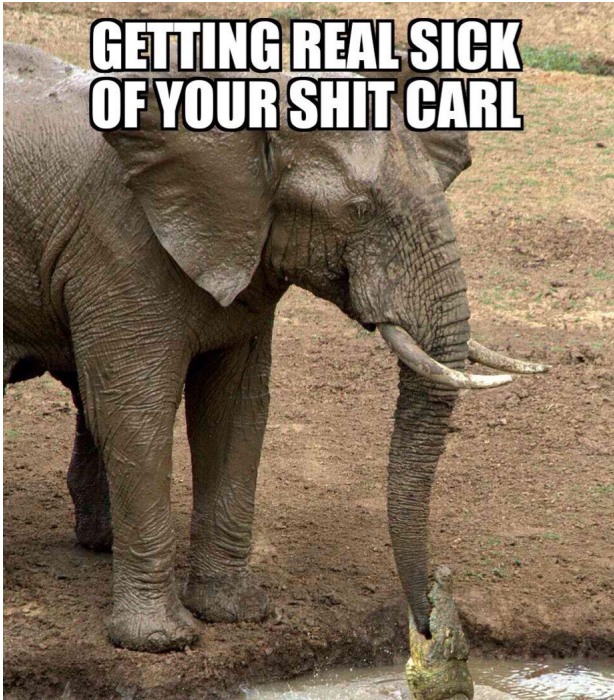
$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \end{aligned}$$

We will call $M(t)$ the moment generating function, because:

$$\left. \frac{\partial^n M(t)}{\partial^n t} \right|_0 = E[X^n]$$

(Assuming that we can interchange derivative and integral)

**GETTING REAL SICK
OF YOUR SHIT CARL**



Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,



Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots$$



Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots$$

Then,



Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots$$

Then,

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \dots$$



Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots$$

Then,

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \dots$$

Differentiate once:



Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots$$

Then,

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \dots$$

Differentiate once:

$$\begin{aligned}\frac{\partial M(t)}{\partial t} &= 0 + E[X] + \frac{2t}{2!}E[X^2] + \dots \\ M'(0) &= 0 + E[X] + 0 + 0 \dots\end{aligned}$$



Proof.

Differentiate n times



Proof.

Differentiate n times

$$\frac{\partial^n M(t)}{\partial^n t} = 0 + 0 + 0 + \dots + \frac{n \times n - 1 \times \dots 2 \times t^0 E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots$$



Proof.

Differentiate n times

$$\begin{aligned}\frac{\partial^n M(t)}{\partial^n t} &= 0 + 0 + 0 + \dots + \frac{n \times n - 1 \times \dots \times 2 \times t^0 E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots \\ &= \frac{n! E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots\end{aligned}$$



Proof.

Differentiate n times

$$\begin{aligned}\frac{\partial^n M(t)}{\partial^n t} &= 0 + 0 + 0 + \dots + \frac{n \times n - 1 \times \dots \times 2 \times t^0 E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots \\ &= \frac{n! E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots\end{aligned}$$

Evaluated at 0, yields $M^n(0) = E[X^n]$



Proof.

Differentiate n times

$$\begin{aligned}\frac{\partial^n M(t)}{\partial^n t} &= 0 + 0 + 0 + \dots + \frac{n \times n - 1 \times \dots \times 2 \times t^0 E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots \\ &= \frac{n! E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \dots\end{aligned}$$

Evaluated at 0, yields $M^n(0) = E[X^n]$



- If two random variables, X and Y have the same moment generating functions, then $F_X(x) = F_Y(x)$ for **almost all** x .

The Moments of the Normal Distribution

Suppose $Z \sim N(0, 1)$.

The Moments of the Normal Distribution

Suppose $Z \sim N(0, 1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

The Moments of the Normal Distribution

Suppose $Z \sim N(0, 1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}((x - t)^2 - t^2)$$

The Moments of the Normal Distribution

Suppose $Z \sim N(0, 1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}((x-t)^2 - t^2)$$

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx$$

The Moments of the Normal Distribution

Suppose $Z \sim N(0, 1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}((x-t)^2 - t^2)$$

$$\begin{aligned} E[e^{tX}] &= \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Extracting Moments of the Normal Distribution

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

$$M''(0) = E[X^2] = e^{t^2/2}(t^2 + 1)|_0 = 1$$

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

$$M''(0) = E[X^2] = e^{t^2/2}(t^2 + 1)|_0 = 1$$

$$M'''(0) = E[X^3] = e^{t^2/2}t(t^2 + 3)|_0 = 0$$

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

$$M''(0) = E[X^2] = e^{t^2/2}(t^2 + 1)|_0 = 1$$

$$M'''(0) = E[X^3] = e^{t^2/2}t(t^2 + 3)|_0 = 0$$

$$M''''(0) = E[X^4] = e^{t^2/2}(t^4 + 6t^2 + 3)|_0 = 3$$

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

$$M''(0) = E[X^2] = e^{t^2/2}(t^2 + 1)|_0 = 1$$

$$M'''(0) = E[X^3] = e^{t^2/2}t(t^2 + 3)|_0 = 0$$

$$M''''(0) = E[X^4] = e^{t^2/2}(t^4 + 6t^2 + 3)|_0 = 3$$

$$M^5(0) = E[X^5] = e^{t^2/2}t(t^4 + 10t^2 + 15)|_0 = 0$$

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

$$M''(0) = E[X^2] = e^{t^2/2}(t^2 + 1)|_0 = 1$$

$$M'''(0) = E[X^3] = e^{t^2/2}t(t^2 + 3)|_0 = 0$$

$$M''''(0) = E[X^4] = e^{t^2/2}(t^4 + 6t^2 + 3)|_0 = 3$$

$$M^5(0) = E[X^5] = e^{t^2/2}t(t^4 + 10t^2 + 15)|_0 = 0$$

$$M^6(0) = E[X^6] = e^{t^2/2}(t^6 + 15t^4 + 45t^2 + 15)|_0 = 15$$

Extracting Moments of the Normal Distribution

$$M'(0) = E[X] = e^{t^2/2}t|_0 = 0$$

$$M''(0) = E[X^2] = e^{t^2/2}(t^2 + 1)|_0 = 1$$

$$M'''(0) = E[X^3] = e^{t^2/2}t(t^2 + 3)|_0 = 0$$

$$M''''(0) = E[X^4] = e^{t^2/2}(t^4 + 6t^2 + 3)|_0 = 3$$

$$M^5(0) = E[X^5] = e^{t^2/2}t(t^4 + 10t^2 + 15)|_0 = 0$$

$$M^6(0) = E[X^6] = e^{t^2/2}(t^6 + 15t^4 + 45t^2 + 15)|_0 = 15$$

Proposition

Suppose X_i are a sequence of independent random variables. Define

$$Y = \sum_{i=1}^N X_i$$

Then

$$M_Y(t) = \prod_{i=1}^N M_{X_i}(t)$$

Proof.

$$M_Y(t) = E[e^{tY}]$$



Proof.

$$\begin{aligned} M_Y(t) &= E[e^{tY}] \\ &= E[e^{t \sum_{i=1}^N X_i}] \end{aligned}$$



Proof.

$$\begin{aligned}M_Y(t) &= E[e^{tY}] \\&= E[e^{t\sum_{i=1}^N X_i}] \\&= E[e^{tX_1+tX_2+\dots+tX_N}]\end{aligned}$$



Proof.

$$\begin{aligned}M_Y(t) &= E[e^{tY}] \\&= E[e^{t\sum_{i=1}^N X_i}] \\&= E[e^{tX_1+tX_2+\dots+tX_N}] \\&= E[e^{tX_1}]E[e^{tX_2}]\dots E[e^{tX_N}] \text{ (by independence)}\end{aligned}$$



Proof.

$$\begin{aligned}M_Y(t) &= E[e^{tY}] \\&= E[e^{t\sum_{i=1}^N X_i}] \\&= E[e^{tX_1+tX_2+\dots+tX_N}] \\&= E[e^{tX_1}]E[e^{tX_2}]\dots E[e^{tX_N}] \text{ (by independence)} \\&= \prod_{i=1}^N E[e^{tX_i}]\end{aligned}$$



Inequalities and Limit Theorems

Limit Theorems

- What happens when we consider a long sequence of random variables?
- What can we reasonably infer from data?
 - Laws of large numbers: averages of random variables converge on expected value?
 - Central Limit Theorems: sum of random variables have normal distribution?
- We'll focus on intuition for both, but we'll prove some stuff too.

Weak Law of Large Numbers

Proof plan:

- Markov's Inequality
- Chebyshev's Inequality
- Weak Law of Large Numbers

Markov's Inequality

Proposition

Suppose X is a random variable that takes on non-negative values. Then, for all $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Markov's Inequality

Proof.

For $a > 0$,



Markov's Inequality

Proof.

For $a > 0$,

$$E[X] = \int_0^{\infty} xf(x)dx$$



Markov's Inequality

Proof.

For $a > 0$,

$$\begin{aligned} E[X] &= \int_0^{\infty} xf(x)dx \\ &= \int_0^a xf(x)dx + \int_a^{\infty} xf(x)dx \end{aligned}$$



Markov's Inequality

Proof.

For $a > 0$,

$$\begin{aligned} E[X] &= \int_0^{\infty} xf(x)dx \\ &= \int_0^a xf(x)dx + \int_a^{\infty} xf(x)dx \end{aligned}$$

Because $X \geq 0$,



Markov's Inequality

Proof.

For $a > 0$,

$$\begin{aligned} E[X] &= \int_0^{\infty} xf(x)dx \\ &= \int_0^a xf(x)dx + \int_a^{\infty} xf(x)dx \end{aligned}$$

Because $X \geq 0$,

$$E[X] \geq \int_a^{\infty} xf(x)dx \geq \int_a^{\infty} af(x)dx = aP(X \geq a)$$



Markov's Inequality

Proof.

For $a > 0$,

$$\begin{aligned} E[X] &= \int_0^{\infty} xf(x)dx \\ &= \int_0^a xf(x)dx + \int_a^{\infty} xf(x)dx \end{aligned}$$

Because $X \geq 0$,

$$\begin{aligned} E[X] &\geq \int_a^{\infty} xf(x)dx \geq \int_a^{\infty} af(x)dx = aP(X \geq a) \\ \frac{E[X]}{a} &\geq P(X \geq a) \end{aligned}$$



Chebyshev's Inequality

Proposition

If X is a random variable with mean μ and variance σ^2 , then, for any value $k > 0$,

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Chebyshev's Inequality

Proof.

Define the random variable

$$Y = (X - \mu)^2$$

Where $\mu = E[X]$.



Chebyshev's Inequality

Proof.

Define the random variable

$$Y = (X - \mu)^2$$

Where $\mu = E[X]$.

Then we know Y is a non-negative random variable. Set $a = k^2$.



Chebyshev's Inequality

Proof.

Define the random variable

$$Y = (X - \mu)^2$$

Where $\mu = E[X]$.

Then we know Y is a non-negative random variable. Set $a = k^2$.

Applying the inequality:



Chebyshev's Inequality

Proof.

Define the random variable

$$Y = (X - \mu)^2$$

Where $\mu = E[X]$.

Then we know Y is a non-negative random variable. Set $a = k^2$.

Applying the inequality:

$$P(Y \geq k^2) \leq \frac{E[Y]}{k^2}$$



Chebyshev's Inequality

Proof.

Define the random variable

$$Y = (X - \mu)^2$$

Where $\mu = E[X]$.

Then we know Y is a non-negative random variable. Set $a = k^2$.

Applying the inequality:

$$P(Y \geq k^2) \leq \frac{E[Y]}{k^2}$$
$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2}$$



Chebyshev's Inequality

Proof.

Define the random variable

$$Y = (X - \mu)^2$$

Where $\mu = E[X]$.

Then we know Y is a non-negative random variable. Set $a = k^2$.

Applying the inequality:

$$P(Y \geq k^2) \leq \frac{E[Y]}{k^2}$$

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2}$$

$$P((X - \mu)^2 \geq k^2) \leq \frac{\sigma^2}{k^2}$$



Chebyshev's Inequality

Further we know that,

$$(X - \mu)^2 \geq k^2$$

Chebyshev's Inequality

Further we know that,

$$(X - \mu)^2 \geq k^2$$

Implies that

Chebyshev's Inequality

Further we know that,

$$(X - \mu)^2 \geq k^2$$

Implies that

$$|X - \mu| \geq k$$

Chebyshev's Inequality

Further we know that,

$$(X - \mu)^2 \geq k^2$$

Implies that

$$|X - \mu| \geq k$$

Thus, we have shown

Chebyshev's Inequality

Further we know that,

$$(X - \mu)^2 \geq k^2$$

Implies that

$$|X - \mu| \geq k$$

Thus, we have shown

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Sequence of Random Variables

Sequence of Independent and Identically, Distributed Random variables.

- Sequence: $X_1, X_2, \dots, X_n, \dots$
- Think of a sequence as sampled **data**:
 - Suppose we are drawing a sample of N observations
 - Each observation will be a random variable, say X_i
 - With realization x_i

Mean/Variance of Sample Mean

Proposition

Let X_1, X_2, \dots, X_n be a random sample from a distribution with mean μ and variance σ^2 . Let \bar{X}_n be the sample mean. Then $E[\bar{X}_n] = \mu$ and $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$

Proof.



Mean/Variance of Sample Mean

Proposition

Let X_1, X_2, \dots, X_n be a random sample from a distribution with mean μ and variance σ^2 . Let \bar{X}_n be the sample mean. Then $E[\bar{X}_n] = \mu$ and $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$

Proof.

$$E[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n E[X_i]$$



Mean/Variance of Sample Mean

Proposition

Let X_1, X_2, \dots, X_n be a random sample from a distribution with mean μ and variance σ^2 . Let \bar{X}_n be the sample mean. Then $E[\bar{X}_n] = \mu$ and $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$

Proof.

$$\begin{aligned} E[\bar{X}_n] &= \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \frac{1}{n} n\mu = \mu \end{aligned}$$



Mean/Variance of Sample Mean

Mean/Variance of Sample Mean

$$\text{var}(\bar{X}_n) = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right)$$

Mean/Variance of Sample Mean

$$\begin{aligned}\text{var}(\bar{X}_n) &= \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i)\end{aligned}$$

Mean/Variance of Sample Mean

$$\begin{aligned}\text{var}(\bar{X}_n) &= \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) \\ &= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

Weak Law of Large Numbers

Proposition

Suppose X_1, X_2, \dots, X_n is a random sample from a distribution with mean μ and $\text{Var}(X_i) = \sigma^2$. Then, for all $\epsilon > 0$,

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq \epsilon \right\} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Weak Law of Large Numbers

Proof.

From our previous proposition



Weak Law of Large Numbers

Proof.

From our previous proposition

$$\frac{E[X_1 + X_2 + \cdots + X_n]}{n} = \frac{\sum_{i=1}^n E[X_i]}{n} = \mu$$



Weak Law of Large Numbers

Proof.

From our previous proposition

$$\frac{E[X_1 + X_2 + \cdots + X_n]}{n} = \frac{\sum_{i=1}^n E[X_i]}{n} = \mu$$

Further,



Weak Law of Large Numbers

Proof.

From our previous proposition

$$\frac{E[X_1 + X_2 + \cdots + X_n]}{n} = \frac{\sum_{i=1}^n E[X_i]}{n} = \mu$$

Further,

$$E\left[\left(\frac{\sum_{i=1}^n X_i - n\mu}{n}\right)^2\right] = \frac{\text{Var}(X_1 + X_2 + \cdots + X_n)}{n^2} = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2} = \frac{\sigma^2}{n}$$



Weak Law of Large Numbers

Proof.

From our previous proposition

$$\frac{E[X_1 + X_2 + \cdots + X_n]}{n} = \frac{\sum_{i=1}^n E[X_i]}{n} = \mu$$

Further,

$$E\left[\left(\frac{\sum_{i=1}^n X_i - n\mu}{n}\right)^2\right] = \frac{\text{Var}(X_1 + X_2 + \cdots + X_n)}{n^2} = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2} = \frac{\sigma^2}{n}$$

Apply Chebyshev's Inequality:



Weak Law of Large Numbers

Proof.

From our previous proposition

$$\frac{E[X_1 + X_2 + \cdots + X_n]}{n} = \frac{\sum_{i=1}^n E[X_i]}{n} = \mu$$

Further,

$$E\left[\left(\frac{\sum_{i=1}^n X_i - n\mu}{n}\right)^2\right] = \frac{\text{Var}(X_1 + X_2 + \cdots + X_n)}{n^2} = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2} = \frac{\sigma^2}{n}$$

Apply Chebyshev's Inequality:

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$



Suppose X_1, X_2, \dots are iid normal distributions,

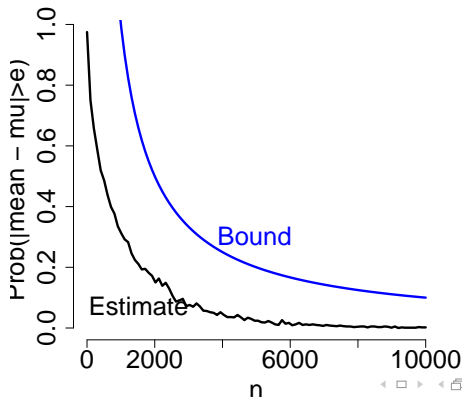
$$X_i \sim \text{Normal}(0, 10)$$

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq 0.1 \right\} \text{ as } n \rightarrow \infty$$

Suppose X_1, X_2, \dots are iid normal distributions,

$$X_i \sim \text{Normal}(0, 10)$$

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq 0.1 \right\} \text{ as } n \rightarrow \infty$$



Suppose X_1, X_2, \dots are iid normal distributions,

$$X_i \sim \text{Normal}(0, 10)$$

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq 0.1 \right\} \text{ as } n \rightarrow \infty$$

Suppose we want to guarantee that we have at most a 0.01 probability of being more than 0.1 away from the true μ . How big do we need n ?

Suppose X_1, X_2, \dots are iid normal distributions,

$$X_i \sim \text{Normal}(0, 10)$$

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq 0.1 \right\} \text{ as } n \rightarrow \infty$$

Suppose we want to guarantee that we have at most a 0.01 probability of being more than 0.1 away from the true μ . How big do we need n ?

$$0.01 = \frac{10}{n(0.1^2)}$$

Suppose X_1, X_2, \dots are iid normal distributions,

$$X_i \sim \text{Normal}(0, 10)$$

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq 0.1 \right\} \text{ as } n \rightarrow \infty$$

Suppose we want to guarantee that we have at most a 0.01 probability of being more than 0.1 away from the true μ . How big do we need n ?

$$\begin{aligned} 0.01 &= \frac{10}{n(0.1^2)} \\ n &= \frac{1000}{0.01} \end{aligned}$$

Suppose X_1, X_2, \dots are iid normal distributions,

$$X_i \sim \text{Normal}(0, 10)$$

$$P \left\{ \left| \frac{X_1 + X_2 + \dots + X_n}{n} - \mu \right| \geq 0.1 \right\} \text{ as } n \rightarrow \infty$$

Suppose we want to guarantee that we have at most a 0.01 probability of being more than 0.1 away from the true μ . How big do we need n ?

$$\begin{aligned} 0.01 &= \frac{10}{n(0.1^2)} \\ n &= \frac{1000}{0.01} \\ n &= 100,000 \end{aligned}$$

Sequences and Convergence

Sequence (refresher):

$$\{a_i\}_{i=1}^{\infty} = \{a_1, a_2, a_3, \dots, a_n, \dots, \}$$

Definition

We say that the sequence $\{a_i\}_{i=1}^{\infty}$ converges to real number A if for each $\epsilon > 0$ there is a positive integer N such that for $n \geq N$, $|a_n - A| < \epsilon$

Sequences and Convergence

Sequence of functions:

$$\{f_i\}_{i=1}^{\infty} = \{f_1, f_2, f_3, \dots, f_n, \dots, \}$$

Definition

Suppose $f_i : X \rightarrow \mathfrak{R}$ for all i . Then $\{f_i\}_{i=1}^{\infty}$ converges *pointwise* to f if, for all $x \in X$ and $\epsilon > 0$, there is an N such that for all $n \geq N$,

$$|f_n(x) - f(x)| < \epsilon$$

This is as strong of a statement as we're likely to make in statistics

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Sequence of estimators: increasing sample size

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Sequence of estimators: increasing sample size

$$\left\{ \hat{\theta}_i \right\}_{i=1}^n = \left\{ \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_n \right\}$$

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Sequence of estimators: increasing sample size

$$\left\{ \hat{\theta}_i \right\}_{i=1}^n = \left\{ \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_n \right\}$$

Question: What can we say about $\left\{ \hat{\theta}_i \right\}_{i=1}^n$ as $n \rightarrow \infty$?

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Sequence of estimators: increasing sample size

$$\left\{ \hat{\theta}_i \right\}_{i=1}^n = \left\{ \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_n \right\}$$

Question: What can we say about $\left\{ \hat{\theta}_i \right\}_{i=1}^n$ as $n \rightarrow \infty$?

- What is the probability $\hat{\theta}_n$ differs from θ ?

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Sequence of estimators: increasing sample size

$$\left\{ \hat{\theta}_i \right\}_{i=1}^n = \left\{ \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_n \right\}$$

Question: What can we say about $\left\{ \hat{\theta}_i \right\}_{i=1}^n$ as $n \rightarrow \infty$?

- What is the probability $\hat{\theta}_n$ differs from θ ?
- What is the probability $\left\{ \hat{\theta}_i \right\}_{i=1}^n$ converges to θ ?

Convergence Definitions

Define $\hat{\theta}_n$ to be estimator for θ based on n observations.

Sequence of estimators: increasing sample size

$$\left\{\hat{\theta}_i\right\}_{i=1}^n = \left\{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_n\right\}$$

Question: What can we say about $\left\{\hat{\theta}_i\right\}_{i=1}^n$ as $n \rightarrow \infty$?

- What is the probability $\hat{\theta}_n$ differs from θ ?
- What is the probability $\left\{\hat{\theta}_i\right\}_{i=1}^n$ converges to θ ?
- What is sampling distribution of $\hat{\theta}_n$ as $n \rightarrow \infty$?

Convergence in Probability

Convergence in Probability

Definition

We will say the sequence $\hat{\theta}_n$ converges in probability to θ (perhaps a non-degenerate RV) if,

$$\lim_{n \rightarrow \infty} \text{Prob}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

For any $\epsilon > 0$

Convergence in Probability

Definition

We will say the sequence $\hat{\theta}_n$ converges in probability to θ (perhaps a non-degenerate RV) if,

$$\lim_{n \rightarrow \infty} \text{Prob}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

For any $\epsilon > 0$

- ϵ is a tolerance parameter: how much error around θ ?

Convergence in Probability

Definition

We will say the sequence $\hat{\theta}_n$ converges in probability to θ (perhaps a non-degenerate RV) if,

$$\lim_{n \rightarrow \infty} \text{Prob}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

For any $\epsilon > 0$

- ϵ is a tolerance parameter: how much error around θ ?
- In the limit, convergence in probability implies sampling distribution collapses on a spike at θ

Convergence in Probability

Definition

We will say the sequence $\hat{\theta}_n$ converges in probability to θ (perhaps a non-degenerate RV) if,

$$\lim_{n \rightarrow \infty} \text{Prob}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

For any $\epsilon > 0$

- ϵ is a tolerance parameter: how much error around θ ?
- In the limit, convergence in probability implies sampling distribution collapses on a spike at θ
- $\{\hat{\theta}_i\}$ need not actually converge to θ , only $P(|\theta_n - \theta| > \epsilon) = 0$

Example (Cassella and Burger)

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$\begin{aligned} X_1(s) &= s + I(s \in [0, 1]) & , & & X_2(s) &= s + I(s \in [0, 1/2]) \\ X_3(s) &= s + I(s \in [1/2, 1]) & , & & X_4(s) &= s + I(s \in [0, 1/3]) \end{aligned}$$

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

$$X_3(s) = s + I(s \in [1/2, 1]) \quad , \quad X_4(s) = s + I(s \in [0, 1/3])$$

$$X_5(s) = s + I(s \in [1/3, 2/3]) \quad , \quad X_6(s) = s + I(s \in [2/3, 1])$$

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

$$X_3(s) = s + I(s \in [1/2, 1]) \quad , \quad X_4(s) = s + I(s \in [0, 1/3])$$

$$X_5(s) = s + I(s \in [1/3, 2/3]) \quad , \quad X_6(s) = s + I(s \in [2/3, 1])$$

Does $X_n(s)$ pointwise converge to $X(s)$?

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

$$X_3(s) = s + I(s \in [1/2, 1]) \quad , \quad X_4(s) = s + I(s \in [0, 1/3])$$

$$X_5(s) = s + I(s \in [1/3, 2/3]) \quad , \quad X_6(s) = s + I(s \in [2/3, 1])$$

Does $X_n(s)$ pointwise converge to $X(s)$?

Does $X_n(s)$ converge in probability to $X(s)$?

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

$$X_3(s) = s + I(s \in [1/2, 1]) \quad , \quad X_4(s) = s + I(s \in [0, 1/3])$$

$$X_5(s) = s + I(s \in [1/3, 2/3]) \quad , \quad X_6(s) = s + I(s \in [2/3, 1])$$

Does $X_n(s)$ pointwise converge to $X(s)$?

Does $X_n(s)$ converge in probability to $X(s)$?

$$P(|X_n - X| > \epsilon) = P(s \in [l_n, u_n])$$

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$. Define $X(s) = s$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

$$X_3(s) = s + I(s \in [1/2, 1]) \quad , \quad X_4(s) = s + I(s \in [0, 1/3])$$

$$X_5(s) = s + I(s \in [1/3, 2/3]) \quad , \quad X_6(s) = s + I(s \in [2/3, 1])$$

Does $X_n(s)$ pointwise converge to $X(s)$?

Does $X_n(s)$ converge in probability to $X(s)$?

$$P(|X_n - X| > \epsilon) = P(s \in [l_n, u_n])$$

Length of $[l_n, u_n] \rightarrow 0 \Rightarrow P(s \in [L_n, U_n]) = 0$

Almost Sure Convergence

Definition

Almost Sure Convergence

Definition

We will say the sequence $\hat{\theta}_n$ converges almost surely to θ if,

$$Prob\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| > \epsilon\right) = 0$$

Almost Sure Convergence

Definition

We will say the sequence $\hat{\theta}_n$ converges almost surely to θ if,

$$Prob\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| > \epsilon\right) = 0$$

- Stronger: says that sequence converges to θ (almost everywhere))

Almost Sure Convergence

Definition

We will say the sequence $\hat{\theta}_n$ converges almost surely to θ if,

$$Prob\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| > \epsilon\right) = 0$$

- Stronger: says that sequence converges to θ (almost everywhere))
- Think about definition of random variable: $\hat{\theta}_n$ is a function from sample space to real line.

Almost Sure Convergence

Definition

We will say the sequence $\hat{\theta}_n$ converges almost surely to θ if,

$$Prob\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| > \epsilon\right) = 0$$

- Stronger: says that sequence converges to θ (almost everywhere))
- Think about definition of random variable: $\hat{\theta}_n$ is a function from sample space to real line.
- Almost sure says that, for all outcomes (s) in sample space (S)
 $s \in S$,

Almost Sure Convergence

Definition

We will say the sequence $\hat{\theta}_n$ converges almost surely to θ if,

$$Prob\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| > \epsilon\right) = 0$$

- Stronger: says that sequence converges to θ (almost everywhere))
- Think about definition of random variable: $\hat{\theta}_n$ is a function from sample space to real line.
- Almost sure says that, for all outcomes (s) in sample space (S)
 $s \in S$,

$$\hat{\theta}_n(s) \rightarrow \theta(s)$$

Almost Sure Convergence

Definition

We will say the sequence $\hat{\theta}_n$ converges almost surely to θ if,

$$Prob\left(\lim_{n \rightarrow \infty} |\hat{\theta}_n - \theta| > \epsilon\right) = 0$$

- Stronger: says that sequence converges to θ (almost everywhere))
- Think about definition of random variable: $\hat{\theta}_n$ is a function from sample space to real line.
- Almost sure says that, for all outcomes (s) in sample space (S)
 $s \in S$,

$$\hat{\theta}_n(s) \rightarrow \theta(s)$$

Except for a subset $\mathcal{N} \subset S$ such that $P(\mathcal{N}) = 0$.

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$.

Suppose X_n is define as follows:

$$X_1(s) = s + I(s \in [0, 1]) \quad , \quad X_2(s) = s + I(s \in [0, 1/2])$$

$$X_3(s) = s + I(s \in [1/2, 1]) \quad , \quad X_4(s) = s + I(s \in [0, 1/3])$$

$$X_5(s) = s + I(s \in [1/3, 2/3]) \quad , \quad X_6(s) = s + I(s \in [2/3, 1])$$

Does $X_n(s)$ converge almost surely to $X(s) = s$?

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$.

Suppose X_n is define as follows:

$$\begin{aligned} X_1(s) &= s + I(s \in [0, 1]) & , & \quad X_2(s) = s + I(s \in [0, 1/2]) \\ X_3(s) &= s + I(s \in [1/2, 1]) & , & \quad X_4(s) = s + I(s \in [0, 1/3]) \\ X_5(s) &= s + I(s \in [1/3, 2/3]) & , & \quad X_6(s) = s + I(s \in [2/3, 1]) \end{aligned}$$

Does $X_n(s)$ converge almost surely to $X(s) = s$?

No!: the sequence doesn't converge for each s

Example (Cassella and Burger)

Suppose $S \sim \text{Uniform}(0,1)$.

Suppose X_n is define as follows:

$$\begin{aligned} X_1(s) &= s + I(s \in [0, 1]) & , & \quad X_2(s) = s + I(s \in [0, 1/2]) \\ X_3(s) &= s + I(s \in [1/2, 1]) & , & \quad X_4(s) = s + I(s \in [0, 1/3]) \\ X_5(s) &= s + I(s \in [1/3, 2/3]) & , & \quad X_6(s) = s + I(s \in [2/3, 1]) \end{aligned}$$

Does $X_n(s)$ converge almost surely to $X(s) = s$?

No!: the sequence doesn't converge for each s

For each value of s the sequence varies between s and $s + 1$ infinitely often

Convergence in Distribution

We've talked about $\hat{\theta}_n$'s sampling distribution converging to a normal distribution.

Definition

Convergence in Distribution

We've talked about $\hat{\theta}_n$'s sampling distribution converging to a normal distribution.

This is **convergence in distribution**

Definition

Convergence in Distribution

We've talked about $\hat{\theta}_n$'s sampling distribution converging to a normal distribution.

This is **convergence in distribution**

Definition

$\hat{\theta}_n$, with cdf $F_n(x)$, converges in distribution to random variable Y with cdf $F(x)$ if

$$\lim_{n \rightarrow \infty} |F_n(x) - F(x)| = 0$$

For all $x \in \mathbb{R}$ where $F(x)$ is continuous.

Convergence in Distribution

We've talked about $\hat{\theta}_n$'s sampling distribution converging to a normal distribution.

This is **convergence in distribution**

Definition

$\hat{\theta}_n$, with cdf $F_n(x)$, converges in distribution to random variable Y with cdf $F(x)$ if

$$\lim_{n \rightarrow \infty} |F_n(x) - F(x)| = 0$$

For all $x \in \mathbb{R}$ where $F(x)$ is continuous.

- Weakest form of convergence almost sure \rightarrow probability \rightarrow distribution

Convergence in Distribution

We've talked about $\hat{\theta}_n$'s sampling distribution converging to a normal distribution.

This is **convergence in distribution**

Definition

$\hat{\theta}_n$, with cdf $F_n(x)$, converges in distribution to random variable Y with cdf $F(x)$ if

$$\lim_{n \rightarrow \infty} |F_n(x) - F(x)| = 0$$

For all $x \in \mathbb{R}$ where $F(x)$ is continuous.

- Weakest form of convergence almost sure \rightarrow probability \rightarrow distribution
- Says that cdfs are equal, says nothing about convergence of underlying RV

Convergence in Distribution

We've talked about $\hat{\theta}_n$'s sampling distribution converging to a normal distribution.

This is **convergence in distribution**

Definition

$\hat{\theta}_n$, with cdf $F_n(x)$, converges in distribution to random variable Y with cdf $F(x)$ if

$$\lim_{n \rightarrow \infty} |F_n(x) - F(x)| = 0$$

For all $x \in \mathbb{R}$ where $F(x)$ is continuous.

- Weakest form of convergence almost sure \rightarrow probability \rightarrow distribution
- Says that cdfs are equal, says nothing about convergence of underlying RV
- Useful for justifying use of some sampling distributions

Convergence in Distribution $\not\Rightarrow$ Convergence in Probability

Define $X \sim N(0, 1)$ and each $X_n = -X$. Then:
 $X_n \sim N(0, 1)$ for all n so X_n trivially converges to X . But,

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|X + X| > \epsilon) \\ &= P(|2X| > \epsilon) \\ &= P(|X| > \epsilon/2) \not\rightarrow 0 \end{aligned}$$

Central Limit Theorem

Proposition

Let X_1, X_2, \dots be a sequence of independent random variables with mean μ and variance σ^2 . Let X_i have a cdf $P(X_i \leq x) = F(x)$ and moment generating function $M(t) = E[e^{tX_i}]$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - \mu n}{\sigma \sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{z^2}{2}\right) dz$$

Central Limit Theorem

Proposition

Let X_1, X_2, \dots be a sequence of independent random variables with mean μ and variance σ^2 . Let X_i have a cdf $P(X_i \leq x) = F(x)$ and moment generating function $M(t) = E[e^{tX_i}]$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - \mu n}{\sigma \sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{z^2}{2}\right) dz$$

Proof plan:

- 1) Rely on Fact that convergence of MGFs \rightsquigarrow convergence in CDFs
- 2) Show that MGFs, in limit, converge on normal MGF

Proposition

Let F_n be a sequence of cumulative distribution functions with the corresponding moment generating functions M_n . F be a cdf with the moment generating functions M . If $\lim_{n \rightarrow \infty} M_n(t) \rightarrow M(t)$ for all t in some interval, then $F_n(x) \rightsquigarrow F(x)$ for all x (when F is continuous).

Proposition

Let F_n be a sequence of cumulative distribution functions with the corresponding moment generating functions M_n . F be a cdf with the moment generating functions M . If $\lim_{n \rightarrow \infty} M_n(t) \rightarrow M(t)$ for all t in some interval, then $F_n(x) \rightsquigarrow F(x)$ for all x (when F is continuous).

Proposition

Suppose $\lim_{n \rightarrow \infty} a_n \rightarrow a$, then

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

Proposition

Let F_n be a sequence of cumulative distribution functions with the corresponding moment generating functions M_n . F be a cdf with the moment generating functions M . If $\lim_{n \rightarrow \infty} M_n(t) \rightarrow M(t)$ for all t in some interval, then $F_n(x) \rightsquigarrow F(x)$ for all x (when F is continuous).

Proposition

Suppose $\lim_{n \rightarrow \infty} a_n \rightarrow a$, then

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

Proposition

Suppose $M(t)$ is a moment generating function some random variable X . Then $M(0) = 1$.

Proof of Central Limit Theorem (Courtsey of Swarthmore Notes)

Proof. Suppose X_1, \dots, X_n are iid variables with $E[X] = 0$, variance σ_x^2 , Moment Generating Function (MGF) $M_x(t)$.

Proof of Central Limit Theorem (Courtsey of Swarthmore Notes)

Proof. Suppose X_1, \dots, X_n are iid variables with $E[X] = 0$, variance σ_x^2 , Moment Generating Function (MGF) $M_x(t)$.

Let $S_n = \sum_{i=1}^n X_i$ and $Z_n = \frac{S_n}{\sigma_x \sqrt{n}}$.

Proof of Central Limit Theorem (Courtsey of Swarthmore Notes)

Proof. Suppose X_1, \dots, X_n are iid variables with $E[X] = 0$, variance σ_x^2 , Moment Generating Function (MGF) $M_x(t)$.

Let $S_n = \sum_{i=1}^n X_i$ and $Z_n = \frac{S_n}{\sigma_x \sqrt{n}}$.

$$M_{S_n} = (M_x(t))^n \text{ and } M_{Z_n}(t) = \left(M_x \left(\frac{t}{\sigma_x \sqrt{n}} \right) \right)^n$$

Proof of Central Limit Theorem (Courtsey of Swarthmore Notes)

Proof. Suppose X_1, \dots, X_n are iid variables with $E[X] = 0$, variance σ_x^2 , Moment Generating Function (MGF) $M_x(t)$.

Let $S_n = \sum_{i=1}^n X_i$ and $Z_n = \frac{S_n}{\sigma_x \sqrt{n}}$.

$$M_{S_n} = (M_x(t))^n \text{ and } M_{Z_n}(t) = \left(M_x \left(\frac{t}{\sigma_x \sqrt{n}} \right) \right)^n$$

Using Taylor's Theorem we can write

$$M_x(s) = M_x(0) + sM'_x(0) + \frac{1}{2}s^2M''_x(0) + e_s$$

$$e_s/s^2 \rightarrow 0 \text{ as } s \rightarrow 0.$$

$$M_x(s) = M_x(0) + sM'_x(0) + \frac{1}{2}s^2M''_x(0) + e_s$$

Filling in the values we have

$$M_x(s) = 1 + 0 + \frac{\sigma_x^2}{2}s^2 + \underbrace{e_s}_{\text{Goes to zero}}$$

Set $s = \frac{t}{\sigma_x\sqrt{n}}$ $\lim_{n \rightarrow \infty} s \rightarrow 0$. Then

$$\begin{aligned} M_{Z_n}(t) &= \left(1 + \frac{\sigma_x^2}{2} \left(\frac{t}{\sigma_x\sqrt{n}} \right)^2 \right)^n \\ &= \left(1 + \frac{t^2/2}{n} \right)^n \\ \lim_{n \rightarrow \infty} M_{Z_n}(t) &= e^{\frac{t^2}{2}} \end{aligned}$$

Congratulations!

Next steps:

- 1) Statistics/Linear Regression
- 2) Causal Inference (Me, in Spring)
- 3) Machine Learning (Me, in Winter, but 2019)
- 4) Take a methods class each quarter