THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics
Lecture 21: Continuous random variables- max of two independent r.v.'s, iterated expectation

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Roadmap

- Function of two independent random variables
- Conditional expectation as a random variable, law of iterated expectations
- Useful inequalities and the normal approximation to Binomial

Functions of two random variables

- If X and Y are both random variables, then $Z = g(X, Y)$ is also a random variable.
- In the discrete case, we could easily find the PMF of the new random variable:

$$p_Z(z) = \sum_{x,y|g(x,y)=z} p_{X,Y}(x, y)$$

- For example, if I roll two fair dice, what is the probability that the sum is 6?

Functions of two random variables

- If X and Y are both random variables, then $Z = g(X, Y)$ is also a random variable.
- In the discrete case, we could easily find the PMF of the new random variable:

$$p_Z(z) = \sum_{x,y|g(x,y)=z} p_{X,Y}(x, y)$$

- For example, if I roll two fair dice, what is the probability that the sum is 6?
- Each possible ordered pair has probability $1/36$.
- The options that sum to 6 are $(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)$... or in other words $(k, 6 - k)$ for $k = 1, \dots, 5$.
- So, $p_Z(5) = 5/36$

Functions of two independent continuous random variables

- I know two guests will both arrive in the next hour. The arrival times are independent random variables which follow a uniform distribution. What is the PDF of the arrival time of the last guest to arrive?
- $X \sim \textit{Uniform}([0,1])$ and $Y \sim \textit{Uniform}([0,1])$.
- In other words, what is the PDF of $Z = \max(X, Y)$?

Functions of two independent continuous random variables

- I know two guests will both arrive in the next hour. The arrival times are independent random variables which follow a uniform distribution. What is the PDF of the arrival time of the last guest to arrive?
- $X \sim \mathit{Uniform}([0,1])$ and $Y \sim \mathit{Uniform}([0,1])$.
- In other words, what is the PDF of $Z = \max(X, Y)$?
- Let's first think about the CDF...

Functions of two independent continuous random variables

- I know two guests will both arrive in the next hour. The arrival times are independent random variables which follow a uniform distribution. What is the PDF of the arrival time of the last guest to arrive?
- $X \sim \mathit{Uniform}([0,1])$ and $Y \sim \mathit{Uniform}([0,1])$.
- In other words, what is the PDF of $Z = \max(X, Y)$?
- Let's first think about the CDF...

Functions of two independent continuous random variables

- I know two guests will both arrive in the next hour. The arrival times are independent random variables which follow a uniform distribution. What is the PDF of the arrival time of the last guest to arrive?
- $X \sim \mathit{Uniform}([0,1])$ and $Y \sim \mathit{Uniform}([0,1])$.
- In other words, what is the PDF of $Z = \max(X, Y)$?
- Let's first think about the CDF...

$$P(Z \leq z) = P(\{X \leq z\} \cap \{Y \leq z\})$$

Functions of two independent continuous random variables

- I know two guests will both arrive in the next hour. The arrival times are independent random variables which follow a uniform distribution. What is the PDF of the arrival time of the last guest to arrive?
- $X \sim \mathit{Uniform}([0,1])$ and $Y \sim \mathit{Uniform}([0,1])$.
- In other words, what is the PDF of $Z = \max(X, Y)$?
- Let's first think about the CDF...

$$
\begin{aligned}
P(Z \leq z) =& P(\{X \leq z\} \cap \{Y \leq z\}) \\
=& P(X \leq z) P(Y \leq z) = \begin{cases} 0 & z < 0 \\ z^2 & 0 \leq z \leq 1 \\ 1 & z > 1 \end{cases}
\end{aligned}
$$

Functions of two independent continuous random variables

- I know two guests will both arrive in the next hour. The arrival times are independent random variables which follow a uniform distribution. What is the PDF of the arrival time of the last guest to arrive?
- $X \sim \mathit{Uniform}([0,1])$ and $Y \sim \mathit{Uniform}([0,1])$.
- In other words, what is the PDF of $Z = \max(X, Y)$?
- Let's first think about the CDF...

$$\begin{aligned} P(Z \leq z) =& P(\{X \leq z\} \cap \{Y \leq z\}) \\ =& P(X \leq z)P(Y \leq z) = \begin{cases} 0 & z < 0 \\ z^2 & 0 \leq z \leq 1 \\ 1 & z > 1 \end{cases} \end{aligned}$$

- Differentiating, $f_Z(z) = \begin{cases} 2z & 0 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases}$

Functions of two random variables: Summary

- If $Y = g(X)$, in order to get the PDF of Y we first looked at the CDF, $P(Y \leq y) = P(g(X) \leq y)$ and then differentiated with respect to y.
- For functions $Z = g(X, Y)$ of two random variables, the same general idea applies:
- First we look at the CDF, $P(Z \leq z) = P(g(X, Y) \leq x)$
- Then we differentiate with respect to z.
- We looked at a special case: The maximum of two independent r.v.'s.

Conditional expectation

- Recall that the conditional expectation of a discrete random variable is given by
$$E[X|Y=y] = \sum_x xP(X=x|Y=y)$$
- Recall that the conditional expectation of a continuous random variable is given by
$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$
- For a given value of y $E[X|Y=y]$ gives a numerical value.

Example 1

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

- Let X denote the amount of time (in hours) until the miner reaches safety, and let Y denote the door he initially chooses

Example 1

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

- Let X denote the amount of time (in hours) until the miner reaches safety, and let Y denote the door he initially chooses
- $$\begin{aligned} E[X] &= E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) + E[X|Y=3]P(Y=3) \\ &= \frac{1}{3}\left(3 + (5 + E[X]) + (7 + E[X])\right) \\ &= \frac{1}{3}(15 + 2E[X]) \\ E[X] &= 15 \end{aligned}$$

Example 1

- John's tank holds 15 gallons of gas, and he always refills his tank when he gets down to 5 gallons. For y gallons of gas in John's tank, total number of miles he will run is $X \sim N(30Y, 1)$.
- $Y \sim Uniform([5, 15])$
- Whats $E[X]$?
- $E[X|Y = y] = 30y$.

Conditional expectation as a random variable

- In the last example, we saw that $E[X|Y = y] = 30y$.
- For any number y $E[X|Y = y]$ is also a number.
- As y varies so does $E[X|Y = y]$. So, $E[X|Y = y]$, so we can view $E[X|Y = y]$ as a function of Y.
- Since a function of a random variable, $E[X|Y = y]$ is also a random variable.
- Concretely we will define $E[X|Y]$ as a function of Y, such that:

$$E[X|Y] = E[X|Y = y] \quad \text{When } Y = y$$

- In the last example, $E[X|Y] = 30Y$.
- Fun fact. $E[Xh(Y)|Y] = E[X|Y]h(Y)$. Why?

Law of iterated expectations

Since $E[X|Y]$ is a random variable, its expectation can be calculated as $E[E[X|Y]]$.

- $$E[E[X|Y]] = \begin{cases} \sum_y E[X|Y=y]P(Y=y) & Y \text{ discrete} \\ \int_y E[X|Y=y]f_Y(y) & Y \text{ continuous} \end{cases}$$
- But we know the RHS of the above, don't we?

Law of iterated expectations

Since $E[X|Y]$ is a random variable, its expectation can be calculated as $E[E[X|Y]]$.

- $$E[E[X|Y]] = \begin{cases} \sum_y E[X|Y=y]P(Y=y) & Y \text{ discrete} \\ \int_y E[X|Y=y]f_Y(y) & Y \text{ continuous} \end{cases}$$
- But we know the RHS of the above, don't we?
- Total expectation theorem! So $E[E[X|Y]] = E[X]$

Law of iterated expectations

Since $E[X|Y]$ is a random variable, its expectation can be calculated as $E[E[X|Y]]$.

- $$E[E[X|Y]] = \begin{cases} \sum_y E[X|Y=y]P(Y=y) & Y \text{ discrete} \\ \int_y E[X|Y=y]f_Y(y) & Y \text{ continuous} \end{cases}$$
- But we know the RHS of the above, don't we?
- Total expectation theorem! So $E[E[X|Y]] = E[X]$

For the last problem where $X \sim N(30y, 1)$, find $E[X|Y=y]$

Law of iterated expectations

Since $E[X|Y]$ is a random variable, its expectation can be calculated as $E[E[X|Y]]$.

- $$E[E[X|Y]] = \begin{cases} \sum_y E[X|Y=y]P(Y=y) & Y \text{ discrete} \\ \int_y E[X|Y=y]f_Y(y) & Y \text{ continuous} \end{cases}$$
- But we know the RHS of the above, don't we?
- Total expectation theorem! So $E[E[X|Y]] = E[X]$

For the last problem where $X \sim N(30y, 1)$, find $E[X|Y=y]$

- We can do it by calculating $f_X(x)$, or we can just use the iterated expectation theorem.

Law of iterated expectations

Since $E[X|Y]$ is a random variable, its expectation can be calculated as $E[E[X|Y]]$.

- $E[E[X|Y]] = \begin{cases} \sum_y E[X|Y=y]P(Y=y) & Y \text{ discrete} \\ \int_y E[X|Y=y]f_Y(y) & Y \text{ continuous} \end{cases}$
- But we know the RHS of the above, don't we?
- Total expectation theorem! So $E[E[X|Y]] = E[X]$

For the last problem where $X \sim N(30y, 1)$, find $E[X|Y=y]$

- We can do it by calculating $f_X(x)$, or we can just use the iterated expectation theorem.
- We already saw that $Y \sim Uniform([5, 15])$.
- $E[X] = E[E[X|Y]] = 30E[Y] = 30 \times 10 = 300$.

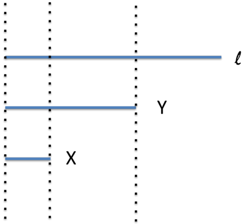
Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

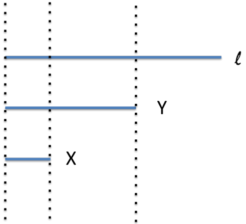
- Let Y is the length of the stick after we break it for the first time and X be the length after the second break.



Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

- Let Y is the length of the stick after we break it for the first time and X be the length after the second break.

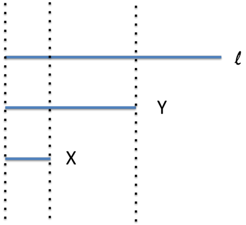


- What is $E[X|Y = y]$? For that we need $f_{X|Y}(x|y)$. Uniform in $[0, y]$.

Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

- Let Y is the length of the stick after we break it for the first time and X be the length after the second break.

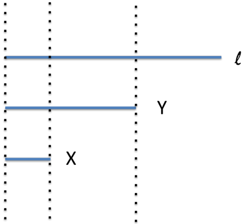


- What is $E[X|Y=y]$? For that we need $f_{X|Y}(x|y)$. Uniform in $[0,y]$.
- So $E[X|Y=y]=y/2$, and $E[X|Y]=Y/2$.

Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

- Let Y is the length of the stick after we break it for the first time and X be the length after the second break.

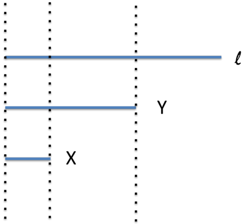


- What is $E[X|Y = y]$? For that we need $f_{X|Y}(x|y)$. Uniform in $[0, y]$.
- So $E[X|Y = y] = y/2$, and $E[X|Y] = Y/2$.
- So $E[X]$ is $E[E[X|Y]] = E[Y/2]$.

Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

- Let Y is the length of the stick after we break it for the first time and X be the length after the second break.

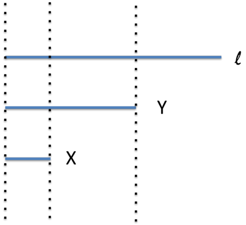


- What is $E[X|Y=y]$? For that we need $f_{X|Y}(x|y)$. Uniform in $[0,y]$.
- So $E[X|Y=y]=y/2$, and $E[X|Y]=Y/2$.
- So $E[X]$ is $E[E[X|Y]]=E[Y/2]$.
- But Y is also Uniform in the interval $[0,\ell]$

Stick breaking

We are breaking a stick of length ℓ at a point which is chosen uniformly over its length and keep the piece that contains the left end and then we repeat the process with the piece we have. What is the expected length of the remaining stick?

- Let Y is the length of the stick after we break it for the first time and X be the length after the second break.



- What is $E[X|Y=y]$? For that we need $f_{X|Y}(x|y)$. Uniform in $[0,y]$.
- So $E[X|Y=y] = y/2$, and $E[X|Y] = Y/2$.
- So $E[X]$ is $E[E[X|Y]] = E[Y/2]$.
- But Y is also Uniform in the interval $[0,\ell]$
- So, $E[Y] = \ell/2$. and $E[X] = E[Y]/2 = \ell/4$

Some useful inequalities-Markov

So far we have looked at expectations and variances of sums of independent random variables. Today we will also look at there behavior when the number of random variables is increasing.

- For a positive random variable X and $t > 0$,

$$P(X > t) \leq E[X]/t.$$

- How do we show this?

$$\begin{aligned} E[X] &= E[X|X \geq t]P(X \geq t) + E[X|X < t]P(X < t) \\ &\geq E[X|X \geq t]P(X \geq t) \geq tP(X \geq t) \\ P(X \geq t) &\leq E[X]/t \end{aligned}$$

- All this comes in handy to show that a random variable cannot be too far from its expectation if the variance is small.

Some useful inequalities

So far we have looked at expectations and variances of sums of independent random variables. Today we will also look at there behavior when the number of random variables is increasing.

- Remember markov's inequality? For a positive random variable X and some $t > 0$, we said that $P(X \geq t) \leq \frac{E[X]}{t}$
- We can use this to bound $P(|X - E[X]| \geq c)$.

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

- This is the famous Chebyshev inequality.
- All this comes in handy to show that a random variable cannot be too far from its expectation if the variance is small.

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$
- $\mathrm{var}(M_n) = \dfrac{\mathrm{var}[X_1] + \cdots + \mathrm{var}[X_n]}{n^2} = \dfrac{\sigma^2}{n}$

Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$
- $\mathrm{var}(M_n) = \dfrac{\mathrm{var}[X_1] + \cdots + \mathrm{var}[X_n]}{n^2} = \dfrac{\sigma^2}{n}$
- So $P(|M_n - \mu| \geq \epsilon) \leq \dfrac{\sigma^2}{n\epsilon^2}$

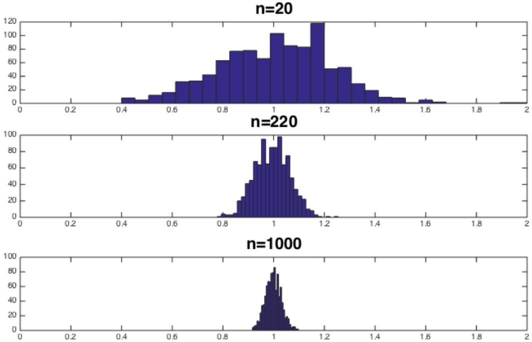
Weak law of large numbers

The WLLN basically states that the sample mean of a large number of random variables is very close to the true mean with high probability.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $M_n = \dfrac{X_1 + \cdots + X_n}{n}$.
- $E[M_n] = \dfrac{E[X_1] + \cdots + E[X_n]}{n} = \mu$
- $\text{var}(M_n) = \dfrac{\text{var}[X_1] + \cdots + \text{var}[X_n]}{n^2} = \dfrac{\sigma^2}{n}$
- So $P(|M_n - \mu| \geq \epsilon) \leq \dfrac{\sigma^2}{n\epsilon^2}$
- For large n this probability is small.

Illustration

Consider the mean of n independent Poisson(1) random variables. For each n, we plot the distribution of the average.



Can we say more? Central Limit Theorem

Turns out that not only can you say that the sample mean is close to the true mean, you can actually predict its distribution using the famous Central Limit Theorem.

- Consider a sequence of i.i.d random variables $X_1, \ldots X_n$ with mean μ and variance σ^2.
- Let $\bar{X}_n = \dfrac{X_1 + \cdots + X_n}{n}$. Remember $E[\bar{X}_n] = \mu$ and $\text{var}(\bar{X}_n) = \sigma^2/n$
- Standardize $\bar{X}_n$ to get $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$
- As n gets bigger, $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ behaves more and more like a $Normal(0,1)$ random variable.
- $P\left(\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < z\right) \approx \Phi(z)$

Can we say more? Central Limit Theorem

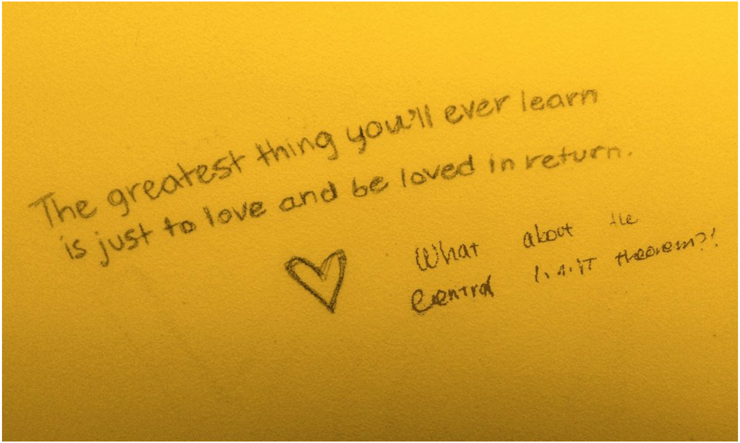


Figure: (Courtesy: Tamara Broderick) You bet!

Example

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to 95% sure that his estimated distance is accurate to within $\pm .5$ lightyears?

Example

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to 95% sure that his estimated distance is accurate to within $\pm .5$ lightyears?

- Let $\bar{X}_n$ be the mean of the measurements.
- How large does n have to be so that $P(|\bar{X}_n - d| \leq .5) = 0.95$

$$P\left(\frac{|\bar{X}_n - d|}{2/\sqrt{n}} \leq 0.25\sqrt{n}\right) \approx P(|Z| \leq 0.25\sqrt{n}) = 1 - 2P(Z \leq -0.25\sqrt{n}) = 0.95$$

- $$\begin{aligned} P(Z \leq -0.25\sqrt{n}) &= 0.025 \\ -0.25\sqrt{n} &= -1.96 \\ \sqrt{n} &= 7.84 \\ n &\approx 62 \end{aligned}$$

Normal Approximation to Binomial

The probability of selling an umbrella is 0.5 on a rainy day. If there are 400 umbrellas in the store, whats the probability that the owner will sell at least 180?

- Let X be the total number of umbrellas sold.
- $X \sim Binomial(400, .5)$
- We want $P(X > 180)$. Crazy calculations.

Normal Approximation to Binomial

The probability of selling an umbrella is 0.5 on a rainy day. If there are 400 umbrellas in the store, whats the probability that the owner will sell at least 180?

- Let X be the total number of umbrellas sold.
- $X \sim Binomial(400, .5)$
- We want $P(X > 180)$. Crazy calculations.

- But can we approximate the distribution of X/n?
- $X/n = (\sum_i Y_i)/n$ where $E[Y_i] = 0.5$ and $\mathrm{var}(Y_i) = 0.25$.
- Sure! CLT tells us that for large n, $\frac{X/400 - 0.5}{\sqrt{0.25/400}} \sim N(0, 1)$
- So $P(X > 180) = P((X - 200)/\sqrt{100} > -2) \approx P(Z \geq -2) = 1 - \Phi(-2) = 0.97$

Summary

- We looked at conditional expectation as a random variable.
- For two r.v.'s X, Y $E[X|Y]$ is a random variable which is a function of Y
- Law of Iterated expectation: $E[E[X|Y]] = E[X]$.
- We also saw Chebyshev's inequality, WLLN and the CLT.