Conditional Expectation

1. Review: joint distribution, LLSE
2. Definition of Conditional expectation
3. Properties of CE
4. Applications: Diluting, Mixing, Rumors
5. CE = MMSE
Conditional Expectation

1. Review: joint distribution, LLSE
2. Definition of Conditional expectation
3. Properties of CE
4. Applications: Diluting, Mixing, Rumors
5. CE = MMSE
**Definitions** Let $X$ and $Y$ be RVs on $\Omega$. 
Review

**Definitions** Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution:** $Pr[X = x, Y = y]$
**Definitions** Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution:** $Pr[X = x, Y = y]$
- **Marginal Distribution:** $Pr[X = x] = \sum_y Pr[X = x, Y = y]$
Review

**Definitions** Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution:** $Pr[X = x, Y = y]
- **Marginal Distribution:** $Pr[X = x] = \sum_y Pr[X = x, Y = y]
- **Conditional Distribution:** $Pr[Y = y|X = x] = \frac{Pr[X=x, Y=y]}{Pr[X=x]}$
Review

**Definitions** Let $X$ and $Y$ be RVs on $\Omega$. 

- **Joint Distribution:** $Pr[X = x, Y = y]$
- **Marginal Distribution:** $Pr[X = x] = \sum_y Pr[X = x, Y = y]$
- **Conditional Distribution:** $Pr[Y = y | X = x] = \frac{Pr[X=x, Y=y]}{Pr[X=x]}$
- **LLSE:**
  \[ L[Y|X] = a + bX \text{ where } a, b \text{ minimize } E[(Y - a - bX)^2]. \]
Review

Definitions Let $X$ and $Y$ be RVs on $\Omega$.

- Joint Distribution: $Pr[X = x, Y = y]$
-Marginal Distribution: $Pr[X = x] = \sum_y Pr[X = x, Y = y]$
-Conditional Distribution: $Pr[Y = y | X = x] = \frac{Pr[X=x, Y=y]}{Pr[X=x]}$

LLSE:
$L[Y|X] = a + bX$ where $a, b$ minimize $E[(Y - a - bX)^2]$.

We saw that
$L[Y|X] = E[Y] + \frac{cov(X, Y)}{var[X]}(X - E[X]).$
Definitions Let \( X \) and \( Y \) be RVs on \( \Omega \).

- **Joint Distribution:** \( Pr[X = x, Y = y] \)
- **Marginal Distribution:** \( Pr[X = x] = \sum_y Pr[X = x, Y = y] \)
- **Conditional Distribution:** \( Pr[Y = y|X = x] = \frac{Pr[X=x,Y=y]}{Pr[X=x]} \)
- **LLSE:**
  \[
  L[Y|X] = a + bX \text{ where } a, b \text{ minimize } E[(Y - a - bX)^2].
  \]

We saw that

\[
L[Y|X] = E[Y] + \frac{cov(X, Y)}{var[X]}(X - E[X]).
\]

Recall the non-Bayesian and Bayesian viewpoints.
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight),
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income),
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear. E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).

Our goal:
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).

Our goal: Derive the best estimate of $Y$ given $X$!
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).

Our goal: Derive the best estimate of $Y$ given $X$!
That is, find the function $g(\cdot)$ so that $g(X)$ is the best guess about $Y$ given $X$. 
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).

Our goal: Derive the best estimate of $Y$ given $X$!

That is, find the function $g(\cdot)$ so that $g(X)$ is the best guess about $Y$ given $X$.

Ambitious!
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).

Our goal: Derive the best estimate of $Y$ given $X$!
That is, find the function $g(\cdot)$ so that $g(X)$ is the best guess about $Y$ given $X$.

Ambitious! Can it be done?
Conditional Expectation: Motivation

There are many situations where a good guess about $Y$ given $X$ is not linear.

E.g., (diameter of object, weight), (school years, income), (PSA level, cancer risk).

Our goal: Derive the best estimate of $Y$ given $X$!
That is, find the function $g(\cdot)$ so that $g(X)$ is the best guess about $Y$ given $X$.

Ambitious! Can it be done? Amazingly, yes!
Definition Let $X$ and $Y$ be RVs on $\Omega$. 

Fact $E[Y | X = x] = \sum_{\omega} Y(\omega) \Pr[\omega | X = x]$. 


Definition Let $X$ and $Y$ be RVs on $\Omega$. The conditional expectation of $Y$ given $X$ is defined as

$$E[Y|X] = g(X)$$
Definition Let $X$ and $Y$ be RVs on $\Omega$. The conditional expectation of $Y$ given $X$ is defined as

$$E[Y|X] = g(X)$$

where

$$g(x) := E[Y|X = x] := \sum_y yPr[Y = y|X = x].$$
**Definition** Let $X$ and $Y$ be RVs on $\Omega$. The conditional expectation of $Y$ given $X$ is defined as

$$E[Y|X] = g(X)$$

where

$$g(x) := E[Y|X = x] := \sum_y yPr[Y = y|X = x].$$

**Fact**

$$E[Y|X = x] = \sum_\omega Y(\omega)Pr[\omega|X = x].$$
**Conditional Expectation**

**Definition** Let $X$ and $Y$ be RVs on $\Omega$. The **conditional expectation** of $Y$ given $X$ is defined as

$$E[Y|X] = g(X)$$

where

$$g(x) := E[Y|X = x] := \sum_y y \Pr[Y = y|X = x].$$

**Fact**

$$E[Y|X = x] = \sum_\omega Y(\omega) \Pr[\omega|X = x].$$

**Proof:** $E[Y|X = x] = E[Y|A]$ with $A = \{\omega : X(\omega) = x\}$. □
Deja vu, all over again?

Have we seen this before?
Deja vu, all over again?

Have we seen this before? Yes.
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new?
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining $g(x) = E[Y|X = x]$ and then $E[Y|X] = g(X)$. Simple but most convenient.
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.

The idea of defining \( g(x) = E[Y|X = x] \) and then \( E[Y|X] = g(X) \).
Big deal?
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining \( g(x) = E[Y|X = x] \) and then \( E[Y|X] = g(X) \).
Big deal? Quite!
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining \( g(x) = E[Y|X = x] \) and then \( E[Y|X] = g(X) \).
Big deal? Quite! Simple but most convenient.
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining \( g(x) = E[Y|X = x] \) and then 
\( E[Y|X] = g(X) \).
Big deal? Quite! Simple but most convenient.
Recall that \( L[Y|X] = a + bX \) is a function of \( X \).
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining \( g(x) = E[Y|X = x] \) and then \( E[Y|X] = g(X) \).
Big deal? Quite! Simple but most convenient.
Recall that \( L[Y|X] = a + bX \) is a function of \( X \).
This is similar: \( E[Y|X] = g(X) \) for some function \( g(\cdot) \).
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining $g(x) = E[Y|X = x]$ and then $E[Y|X] = g(X)$.
Big deal? Quite! Simple but most convenient.
Recall that $L[Y|X] = a + bX$ is a function of $X$.
This is similar: $E[Y|X] = g(X)$ for some function $g(\cdot)$.
In general, $g(X)$ is not linear, i.e., not $a + bX$. 
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining $g(x) = E[Y|X = x]$ and then $E[Y|X] = g(X)$.
Big deal? Quite! Simple but most convenient.
Recall that $L[Y|X] = a + bX$ is a function of $X$.
This is similar: $E[Y|X] = g(X)$ for some function $g(\cdot)$.
In general, $g(X)$ is not linear, i.e., not $a + bX$. It could be that $g(X) = a + bX + cX^2$. 
Deja vu, all over again?

Have we seen this before? Yes.

Is anything new? Yes.

The idea of defining $g(x) = E[Y|X = x]$ and then $E[Y|X] = g(X)$.

Big deal? Quite! Simple but most convenient.

Recall that $L[Y|X] = a + bX$ is a function of $X$.

This is similar: $E[Y|X] = g(X)$ for some function $g(\cdot)$.

In general, $g(X)$ is not linear, i.e., not $a + bX$. It could be that $g(X) = a + bX + cX^2$. Or that $g(X) = 2\sin(4X) + \exp\{-3X\}$. 
Deja vu, all over again?

Have we seen this before? Yes.
Is anything new? Yes.
The idea of defining \( g(x) = E[Y|X = x] \) and then \( E[Y|X] = g(X) \).

Big deal? Quite! Simple but most convenient.
Recall that \( L[Y|X] = a + bX \) is a function of \( X \).
This is similar: \( E[Y|X] = g(X) \) for some function \( g(\cdot) \).
In general, \( g(X) \) is not linear, i.e., not \( a + bX \). It could be that \( g(X) = a + bX + cX^2 \). Or that \( g(X) = 2\sin(4X) + \exp\{-3X\} \). Or something else.
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]
Properties of CE

\[ E[Y|X = x] = \sum_y y \Pr[Y = y|X = x] \]

Theorem
Properties of CE

\[ E[Y|X = x] = \sum_y y \cdot Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
Properties of CE

\[ E[Y|X = x] = \sum_y y Pr[Y = y|X = x] \]

**Theorem**

(a) \(X, Y\) independent \(\Rightarrow E[Y|X] = E[Y];\)

(b) \(E[aY + bZ|X] = aE[Y|X] + bE[Z|X];\)
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]

**Theorem**
(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
Properties of CE

\[ E[Y|X = x] = \sum_y y \Pr[Y = y|X = x] \]

**Theorem**
(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
Properties of CE

\[ E[Y|X = x] = \sum_{y} y \Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);

(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);

(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);

(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);

(e) \( E[E[Y|X]] = E[Y] \).
Properties of CE

\[ E[Y|X = x] = \sum_y y \Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent ⇒ \( E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X] \), \( \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y] \), \( \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:**
Properties of CE

\[ E[Y|X = x] = \sum_{y} y Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow \) \( E[Y|X] = E[Y] \);

(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);

(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);

(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);

(e) \( E[E[Y|X]] = E[Y] \).

**Proof:**

(a),(b) Obvious
Properties of CE

\[ E[Y|X = x] = \sum_{y} yPr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:**

(a), (b) Obvious

(c) \( E[Yh(X)|X = x] = \sum_{\omega} Y(\omega)h(X(\omega)Pr[\omega|X = x] \)
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);

(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);

(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);

(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);

(e) \( E[E[Y|X]] = E[Y] \).

**Proof:**

(a), (b) Obvious

(c) \[ E[Yh(X)|X = x] = \sum_\omega Y(\omega)h(X(\omega))Pr[\omega|X = x] \]

\[ = \sum_\omega Y(\omega)h(x)Pr[\omega|X = x] \]
Properties of CE

\[ E[Y|X = x] = \sum_{y} y Pr[Y = y|X = x] \]

**Theorem**
(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:**
(a), (b) Obvious
(c) \( E[Yh(X)|X = x] = \sum_{\omega} Y(\omega) h(X(\omega)) Pr[\omega|X = x] \)
\[ = \sum_{\omega} Y(\omega) h(x) Pr[\omega|X = x] = h(x)E[Y|X = x] \]
Properties of CE

\[ E[Y | X = x] = \sum_y y \Pr[Y = y | X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y | X] = E[Y] \);
(b) \( E[aY + bZ | X] = aE[Y | X] + bE[Z | X] \);
(c) \( E[Yh(X) | X] = h(X)E[Y | X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y | X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y | X]] = E[Y] \).

**Proof:** (continued)
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:** (continued)

(d) \( E[h(X)E[Y|X]] = \sum_x h(x)E[Y|X = x]Pr[X = x] \)
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y]; \)
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X]; \)
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot); \)
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot); \)
(e) \( E[E[Y|X]] = E[Y]. \)

**Proof:** (continued)

(d) \[ E[h(X)E[Y|X]] = \sum_x h(x)E[Y|X = x]Pr[X = x] \]
\[ = \sum_x h(x)\sum_y yPr[Y = y|X = x]Pr[X = x] \]
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:** (continued)

(d) \[
E[h(X)E[Y|X]] = \sum_x h(x)E[Y|X = x]Pr[X = x]
= \sum_x h(x)\sum_y yPr[Y = y|X = x]Pr[X = x]
= \sum_x h(x)\sum_y yPr[X = x, y = y]
\]
Properties of CE

\[ E[Y|X = x] = \sum_y y Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:** (continued)

(d) \( E[h(X)E[Y|X]] = \sum_x h(x)E[Y|X = x]Pr[X = x] \)

\[ = \sum_x h(x)\sum_y y Pr[Y = y|X = x]Pr[X = x] \]

\[ = \sum_x h(x)\sum_y y Pr[X = x, y = y] \]

\[ = \sum_{x,y} h(x)y Pr[X = x, y = y] \]
Properties of CE

\[ E[Y|X = x] = \sum_y yPr[Y = y|X = x] \]

Theorem
(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

Proof: (continued)
(d) \[ E[h(X)E[Y|X]] = \sum_x h(x)E[Y|X = x]Pr[X = x] \]
\[ = \sum_x h(x)\sum_y yPr[Y = y|X = x]Pr[X = x] \]
\[ = \sum_x h(x)\sum_y yPr[X = x, y = y] \]
\[ = \sum_{x,y} h(x)yPr[X = x, y = y] = E[h(X)Y]. \]
Properties of CE

\[ E[Y|X = x] = \sum_y y \Pr[Y = y|X = x] \]

**Theorem**

(a) \( X, Y \) independent \( \Rightarrow E[Y|X] = E[Y] \);
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X] \);
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot) \);
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot) \);
(e) \( E[E[Y|X]] = E[Y] \).

**Proof:** (continued)
Properties of CE

\[ E[Y|X = x] = \sum_{y} y \Pr[Y = y|X = x] \]

Theorem
(a) \( X, Y \) independent \( \implies E[Y|X] = E[Y]; \)
(b) \( E[aY + bZ|X] = aE[Y|X] + bE[Z|X]; \)
(c) \( E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot); \)
(d) \( E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot); \)
(e) \( E[E[Y|X]] = E[Y]. \)

Proof: (continued)
(e) Let \( h(X) = 1 \) in (d).
Properties of CE

**Theorem**

(a) $X, Y$ independent $\Rightarrow E[Y|X] = E[Y]$;
(b) $E[aY + bZ|X] = aE[Y|X] + bE[Z|X]$;
(c) $E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot)$;
(d) $E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot)$;
(e) $E[E[Y|X]] = E[Y]$.

Note that (d) says that $E[(Y - E[Y|X])h(X)] = 0$. We say that the estimation error $Y - E[Y|X]$ is orthogonal to every function $h(\cdot)$ of $X$. We call this the projection property. More about this later.
Properties of CE

Theorem
(a) $X, Y$ independent $\Rightarrow E[Y|X] = E[Y]$;
(b) $E[ay + bZ|X] = aE[Y|X] + bE[Z|X]$;
(c) $E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot)$;
(d) $E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot)$;
(e) $E[EE[Y|X]] = E[Y]$.

Note that (d) says that

$$E[(Y - E[Y|X])h(X)] = 0.$$
Properties of CE

**Theorem**

(a) $X, Y$ independent $\Rightarrow E[Y|X] = E[Y]$;
(b) $E[aY + bZ|X] = aE[Y|X] + bE[Z|X]$;
(c) $E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot)$;
(d) $E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot)$;
(e) $E[E[Y|X]] = E[Y]$.

Note that (d) says that

$$E[(Y - E[Y|X])h(X)] = 0.$$ 

We say that the estimation error $Y - E[Y|X]$ is orthogonal to every function $h(X)$ of $X$. 
Properties of CE

**Theorem**
(a) $X, Y$ independent $\Rightarrow E[Y|X] = E[Y]$;
(b) $E[aY + bZ|X] = aE[Y|X] + bE[Z|X]$;
(c) $E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot)$;
(d) $E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot)$;
(e) $E[E[Y|X]] = E[Y]$.

Note that (d) says that

$$E[(Y - E[Y|X])h(X)] = 0.$$ 

We say that the estimation error $Y - E[Y|X]$ is orthogonal to every function $h(X)$ of $X$.

We call this the projection property.
Properties of CE

Theorem
(a) $X$, $Y$ independent $\Rightarrow E[Y|X] = E[Y]$;
(b) $E[aY + bZ|X] = aE[Y|X] + bE[Z|X]$;
(c) $E[Yh(X)|X] = h(X)E[Y|X], \forall h(\cdot)$;
(d) $E[h(X)E[Y|X]] = E[h(X)Y], \forall h(\cdot)$;
(e) $E[E[Y|X]] = E[Y]$.

Note that (d) says that

$$E[(Y - E[Y|X])h(X)] = 0.$$ 

We say that the estimation error $Y - E[Y|X]$ is orthogonal to every function $h(X)$ of $X$.

We call this the projection property. More about this later.
Application: Calculating $E[Y|X]$

Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1.
Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1. We want to calculate

$$E[2 + 5X + 7XY + 11X^2 + 13X^3 Z^2 | X].$$
Application: Calculating $E[Y|X]$

Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1. We want to calculate

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X].$$

We find

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X]$$
Application: Calculating $E[Y|X]$

Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1. We want to calculate

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X].$$

We find

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X] \quad = \quad 2 + 5X + 7XE[Y|X] + 11X^2 + 13X^3E[Z^2|X]$$
Application: Calculating $E[Y|X]$

Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1. We want to calculate

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X].$$

We find

Application: Calculating $E[Y|X]$

Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1. We want to calculate

$$E[2 + 5X + 7XY + 11X^2 + 13X^3 Z^2 | X].$$

We find

$$E[2 + 5X + 7XY + 11X^2 + 13X^3 Z^2 | X] = 2 + 5X + 7XE[Y|X] + 11X^2 + 13X^3 E[Z^2 | X]$$

$$= 2 + 5X + 7XE[Y] + 11X^2 + 13X^3 E[Z^2]$$

$$= 2 + 5X + 11X^2 + 13X^3(\text{var}[Z] + E[Z]^2)$$
Application: Calculating $E[Y|X]$

Let $X, Y, Z$ be i.i.d. with mean 0 and variance 1. We want to calculate

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X].$$

We find

$$E[2 + 5X + 7XY + 11X^2 + 13X^3Z^2|X]$$
$$= 2 + 5X + 7XE[Y|X] + 11X^2 + 13X^3E[Z^2|X]$$
$$= 2 + 5X + 7XE[Y] + 11X^2 + 13X^3E[Z^2]$$
$$= 2 + 5X + 11X^2 + 13X^3(var[Z] + E[Z]^2)$$
$$= 2 + 5X + 11X^2 + 13X^3.$$
**Application: Diluting**

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$.

What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence, $E[X_{n+1} | X_n = m] = m - (m/N) = X_n \rho$.

Consequently, $E[X_{n+1}] = E[E[X_{n+1} | X_n]] = \rho E[X_n]$.

$E[X_n] = \rho^{n-1} E[X_1] = N(N-1/N)^{n-1}, \quad n \geq 1$. 

$X_1 = N$ red balls
Application: Diluting

At each step, pick a ball from a well-mixed urn.

\[ X_1 = N \]
\[ X_2 = N - 1 \]
\[ X_3 = N - 2 \]
\[ X_4 = N - 2 \]

red balls
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball.
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. 

\[
\begin{align*}
X_1 &= N \\
X_2 &= N - 1 \\
X_3 &= N - 2 \\
X_4 &= N - 2
\end{align*}
\]

red balls
At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?
At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball)
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise.
At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N)$$
**Application: Diluting**

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N) = m(N - 1)/N = X_n \rho,$$

with $\rho := (N - 1)/N$. 

Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N) = m(N - 1)/N = X_n \rho,$$

with $\rho := (N - 1)/N$. Consequently,
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N) = m(N - 1)/N = X_n\rho,$$

with $\rho := (N - 1)/N$. Consequently,

$$E[X_{n+1}] = E[E[X_{n+1}|X_n]]$$
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N) = m(N - 1)/N = X_n\rho,$$

with $\rho := (N - 1)/N$. Consequently,

$$E[X_{n+1}] = E[E[X_{n+1}|X_n]] = \rho E[X_n], \quad n \geq 1.$$
At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N) = m(N - 1)/N = X_n \rho,$$

with $\rho := (N - 1)/N$. Consequently,

$$E[X_{n+1}] = E[E[X_{n+1}|X_n]] = \rho E[X_n], n \geq 1.$$

$$\implies E[X_n] = \rho^{n-1} E[X_1]$$
Application: Diluting

At each step, pick a ball from a well-mixed urn. Replace it with a blue ball. Let $X_n$ be the number of red balls in the urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m - 1$ w.p. $m/N$ (if you pick a red ball) and $X_{n+1} = m$ otherwise. Hence,

$$E[X_{n+1}|X_n = m] = m - (m/N) = m(N - 1)/N = X_n \rho,$$

with $\rho := (N - 1)/N$. Consequently,

$$E[X_{n+1}] = E[E[X_{n+1}|X_n]] = \rho E[X_n], \quad n \geq 1.$$

$$\implies E[X_n] = \rho^{n-1} E[X_1] = N\left(\frac{N-1}{N}\right)^{n-1}, \quad n \geq 1.$$
Diluting

Here is a plot:
Diluting

Here is a plot:

\( E[X_n] \)
By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}, n \geq 1$. 
Diluting

By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}$, $n \geq 1$.

Here is another argument for that result.
By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}, n \geq 1$.

Here is another argument for that result.

Consider one particular red ball, say ball $k$. 

Diluting
By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N(N-1)^{n-1}, n \geq 1$.

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N-1)/N$, when another ball is picked.
By analyzing $E[X_{n+1}|X_n]$, we found that 
$$E[X_n] = N \left( \frac{N-1}{N} \right)^{n-1}, \ n \geq 1.$$ 

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N-1)/N$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $[(N-1)/N]^{n-1}$. 
Diluting

By analyzing $E[X_{n+1}|X_n]$, we found that

$$E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}, \ n \geq 1.$$ 

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $\frac{N-1}{N}$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $\left(\frac{N-1}{N}\right)^{n-1}$. Let

$$Y_n(k) = 1\{\text{ball } k \text{ is red at step } n\}.$$
By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}$, $n \geq 1$.

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N-1)/N$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $\left[(N - 1)/N\right]^{n-1}$. Let

$$Y_n(k) = 1\{\text{ball } k \text{ is red at step } n\}.$$ 

Then, $X_n = Y_n(1) + \cdots + Y_n(N)$. 
By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}, n \geq 1$.

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N-1)/N$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $\left((N-1)/N\right)^{n-1}$. Let

$$Y_n(k) = 1\{\text{ball } k \text{ is red at step } n\}.$$ 

Then, $X_n = Y_n(1) + \cdots + Y_n(N)$. Hence,

$$E[X_n] = E[Y_n(1) + \cdots + Y_n(N)]$$
Diluting

By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N\left(\frac{N-1}{N}\right)^{n-1}, n \geq 1$.

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N-1)/N$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $[(N-1)/N]^{n-1}$. Let

$$Y_n(k) = 1\{\text{ball } k \text{ is red at step } n\}.$$ 

Then, $X_n = Y_n(1) + \cdots + Y_n(N)$. Hence,

$$E[X_n] = E[Y_n(1) + \cdots + Y_n(N)] = NE[Y_n(1)]$$
By analyzing $E[X_{n+1}|X_n]$, we found that $E[X_n] = N(\frac{N-1}{N})^{n-1}$, $n \geq 1$.

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N - 1)/N$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $[(N - 1)/N]^{n-1}$. Let

$$Y_n(k) = 1\{\text{ball } k \text{ is red at step } n\}.$$

Then, $X_n = Y_n(1) + \cdots + Y_n(N)$. Hence,

$$E[X_n] = E[Y_n(1) + \cdots + Y_n(N)] = NE[Y_n(1)] = NPr[Y_n(1) = 1]$$
By analyzing $E[X_{n+1}|X_n]$, we found that

$$E[X_n] = N \left( \frac{N-1}{N} \right)^{n-1}, \quad n \geq 1.$$ 

Here is another argument for that result.

Consider one particular red ball, say ball $k$. At each step, it remains red w.p. $(N-1)/N$, when another ball is picked. Thus, the probability that it is still red at step $n$ is $\left( (N-1)/N \right)^{n-1}$. Let

$$Y_n(k) = 1 \{ \text{ball } k \text{ is red at step } n \}.$$

Then, $X_n = Y_n(1) + \cdots + Y_n(N)$. Hence,

$$E[X_n] = E[Y_n(1) + \cdots + Y_n(N)] = NE[Y_n(1)] = NPr[Y_n(1) = 1] = N\left[ (N-1)/N \right]^{n-1}.$$
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$ where $p = \left(1 - \frac{m}{N}\right)^2$ (B goes up, R down) and $q = \left(\frac{m}{N}\right)^2$ (R goes up, B down).

Thus, $E[X_{n+1} | X_n] = X_n + p - q = X_n + 1 - 2X_n/N = 1 + \rho X_n$, $\rho = \left(1 - \frac{2}{N}\right)$. 
Application: Mixing

At each step, pick a ball from each well-mixed urn.
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn.
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. 

$$X_1 = N$$

$$X_2 = N - 1$$

$$X_3 = N - 1$$
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?
At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$

where $p = (1 - m/N)^2$ (B goes up, R down)
At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$

where $p = (1 - m/N)^2$ (B goes up, R down) and $q = (m/N)^2$ (R goes up, B down).
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$

where $p = (1 - m/N)^2$ (B goes up, R down) and $q = (m/N)^2$ (R goes up, B down).

Thus,

$$E[X_{n+1} | X_n] = X_n + p - q$$
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$

where $p = (1 - m/N)^2$ (B goes up, R down) and $q = (m/N)^2$ (R goes up, B down).

Thus,

$$E[X_{n+1}|X_n] = X_n + p - q = X_n + 1 - 2X_n/N$$
Application: Mixing

At each step, pick a ball from each well-mixed urn. We transfer them to the other urn. Let $X_n$ be the number of red balls in the bottom urn at step $n$. What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m + 1$ w.p. $p$ and $X_{n+1} = m - 1$ w.p. $q$

where $p = (1 - m/N)^2$ (B goes up, R down) and $q = (m/N)^2$ (R goes up, B down).

Thus,

$$E[X_{n+1}|X_n] = X_n + p - q = X_n + 1 - 2X_n/N = 1 + \rho X_n, \quad \rho := (1 - 2/N).$$
Mixing

We saw that $E[X_{n+1} | X_n] = 1 + \rho X_n$, $\rho := (1 - 2/N)$. 
We saw that $E[X_{n+1}|X_n] = 1 + \rho X_n$, $\rho := (1 - 2/N)$. Hence,

$$E[X_{n+1}] = 1 + \rho E[X_n]$$
We saw that $E[X_{n+1}|X_n] = 1 + \rho X_n$, $\rho := (1 - 2/N)$. Hence,

\[
E[X_{n+1}] = 1 + \rho E[X_n]
\]

$E[X_2] = 1 + \rho N$; $E[X_3] = 1 + \rho (1 + \rho N) = 1 + \rho + \rho^2 N$
Mixing

We saw that $E[X_{n+1}|X_n] = 1 + \rho X_n$, $\rho := (1 - 2/N)$. Hence,

$$E[X_{n+1}] = 1 + \rho E[X_n]$$
$$E[X_2] = 1 + \rho N; E[X_3] = 1 + \rho (1 + \rho N) = 1 + \rho + \rho^2 N$$
$$E[X_4] = 1 + \rho (1 + \rho + \rho^2 N) = 1 + \rho + \rho^2 + \rho^3 N$$
We saw that $E[X_{n+1} | X_n] = 1 + \rho X_n$, $\rho := (1 - 2/N)$. Hence,

\[
E[X_{n+1}] = 1 + \rho E[X_n]
\]

\[
E[X_2] = 1 + \rho N; \quad E[X_3] = 1 + \rho (1 + \rho N) = 1 + \rho + \rho^2 N
\]

\[
E[X_4] = 1 + \rho (1 + \rho + \rho^2 N) = 1 + \rho + \rho^2 + \rho^3 N
\]

\[
E[X_n] = 1 + \rho + \cdots + \rho^{n-2} + \rho^{n-1} N.
\]
We saw that $E[X_{n+1}|X_n] = 1 + \rho X_n$, $\rho := (1 - 2/N)$. Hence,

\[
E[X_{n+1}] = 1 + \rho E[X_n]
\]

\[
E[X_2] = 1 + \rho N; \quad E[X_3] = 1 + \rho (1 + \rho N) = 1 + \rho + \rho^2 N
\]

\[
E[X_4] = 1 + \rho (1 + \rho + \rho^2 N) = 1 + \rho + \rho^2 + \rho^3 N
\]

\[
E[X_n] = 1 + \rho + \cdots + \rho^{n-2} + \rho^{n-1} N.
\]

Hence,

\[
E[X_n] = \frac{1 - \rho^{n-1}}{1 - \rho} + \rho^{n-1} N, \quad n \geq 1.
\]
Application: Mixing

Here is the plot.
Application: Mixing

Here is the plot.
Application: Going Viral

Consider a social network (e.g., Twitter).
Application: Going Viral

Consider a social network (e.g., Twitter).

You start a rumor
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends.
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$. 
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$.
Each of your friends has $d$ friends, etc.
Application: Going Viral

Consider a social network (e.g., Twitter).

You start a rumor (e.g., Walrand is really weird).

You have $d$ friends. Each of your friend retweets w.p. $p$.

Each of your friends has $d$ friends, etc.

Does the rumor spread?
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$.
Each of your friends has $d$ friends, etc.
Does the rumor spread? Does it die out
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$.
Each of your friends has $d$ friends, etc.
Does the rumor spread? Does it die out (mercifully)?
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$.
Each of your friends has $d$ friends, etc.
Does the rumor spread? Does it die out (mercifully)?
Application: Going Viral

Consider a social network (e.g., Twitter).
You start a rumor (e.g., Walrand is really weird).
You have $d$ friends. Each of your friend retweets w.p. $p$.
Each of your friends has $d$ friends, etc.
Does the rumor spread? Does it die out (mercifully)?

In this example, $d = 4$. 
Application: Going Viral

Fact:

$$X = \sum_{n=1}^{\infty} X_n.$$  Then, $$E[X] < \infty$$ iff $$p_d < 1.$$  

Proof:

Given $$X_n = k,$$ then $$X_{n+1} = B(kd, p).$$  Hence, $$E[X_{n+1} | X_n = k] = kp_d.$$  Thus, $$E[X_n] = p_d X_n.$$  Consequently, $$E[X_n] = (pd)^{n-1}, n \geq 1.$$  If $$pd < 1,$$ then $$E[X_1 + \cdots + X_n] \leq \left(1 - pd\right)^{n-1} \Rightarrow E[X] \leq \left(1 - pd\right)^{-1}.$$  If $$pd \geq 1,$$ then for all $$C$$ one can find $$n$$ s.t. $$E[X] \geq E[X_1 + \cdots + X_n] \geq C.$$  In fact, one can show that $$pd = 1 \Rightarrow \Pr[X = \infty] > 0.$$
Application: Going Viral

Fact:

\[ X = \sum_{n=1}^{\infty} X_n. \]

Then,

\[ \mathbb{E}[X] < \infty \iff p < 1. \]

Proof:

Given \( X_n = k \),

\[ X_{n+1} = B(kp, p). \]

Hence,

\[ \mathbb{E}[X_{n+1} | X_n = k] = kp. \]

Thus,

\[ \mathbb{E}[X_{n+1} | X_n] = p \mathbb{E}[X_n]. \]

Consequently,

\[ \mathbb{E}[X_n] = (pd)^{n-1}, \quad n \geq 1. \]

If \( pd < 1 \), then

\[ \mathbb{E}[X_1 + \cdots + X_n] \leq (1 - pd)^{-1} = \Rightarrow \mathbb{E}[X] \leq (1 - pd)^{-1}. \]

If \( pd \geq 1 \), then for all \( C \) one can find \( n \) s.t.

\[ \mathbb{E}[X] \geq \mathbb{E}[X_1 + \cdots + X_n] \geq C. \]

In fact, one can show that

\[ pd \geq 1 \Rightarrow \Pr[X = \infty] > 0. \]
Application: Going Viral

Fact: Let $X = \sum_{n=1}^{\infty} X_n$. 
Application: Going Viral

Fact: Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$. 
**Fact:** Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

**Proof:**
Given $X_n = k$, $X_{n+1} = B(kd, p)$. 
Application: Going Viral

Fact: Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

Proof:
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1}|X_n = k] = kpd$. 
**Application: Going Viral**

**Fact:** Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

**Proof:**
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1}|X_n = k] = kpd$.
Thus, $E[X_{n+1}|X_n] = pdX_n$. 

**Fact:** Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

**Proof:**
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1}|X_n = k] = kpd$.

Thus, $E[X_{n+1}|X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$. 
Fact: Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

Proof:
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1} | X_n = k] = kpd$.
Thus, $E[X_{n+1} | X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$.
If $pd < 1$, then $E[X_1 + \cdots + X_n] \leq (1 - pd)^{-1}$.
**Application: Going Viral**

**Fact:** Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

**Proof:**
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1}|X_n = k] = kpd$.
Thus, $E[X_{n+1}|X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$.
If $pd < 1$, then $E[X_1 + \cdots + X_n] \leq (1 - pd)^{-1} \implies E[X] \leq (1 - pd)^{-1}$. 
Fact: Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

Proof:
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1}|X_n = k] = kpd$.
Thus, $E[X_{n+1}|X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$.
If $pd < 1$, then $E[X_1 + \cdots + X_n] \leq (1 - pd)^{-1} \implies E[X] \leq (1 - pd)^{-1}$.
If $pd \geq 1$, then for all $C$ one can find $n$ s.t. $E[X] \geq E[X_1 + \cdots + X_n] \geq C$. 

□
**Fact:** Let $X = \sum_{n=1}^{\infty} X_n$. Then, $E[X] < \infty$ iff $pd < 1$.

**Proof:**
Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1} | X_n = k] = kpd$.

Thus, $E[X_{n+1} | X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$.

If $pd < 1$, then $E[X_1 + \cdots + X_n] \leq (1 - pd)^{-1} \implies E[X] \leq (1 - pd)^{-1}$.

If $pd \geq 1$, then for all $C$ one can find $n$ s.t.

$E[X] \geq E[X_1 + \cdots + X_n] \geq C$.

In fact, one can show that $pd \geq 1 \implies Pr[X = \infty] > 0$.  

\[ \square \]
Application: Going Viral

An easy extension:

Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d_i$.

Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_n + 1 = B(d_1 + \cdots + d_k, p)$.

Hence, $E[X_n + 1 | X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k)$.

Consequently, $E[X_n + 1 | X_n = k] = p E[d_1 + \cdots + d_k]$.

Finally, $E[X_n + 1 | X_n] = pE[X_n]$, and $E[X_n + 1] = pE[X_n]$. We conclude as before.
Application: Going Viral

An easy extension:
An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. 

Application: Going Viral
Application: Going Viral

An easy extension: Assume that everyone has an independent number \( D_i \) of friends with \( E[D_i] = d \). Then, the same fact holds.
Application: Going Viral

An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people,
An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. 
An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1} | X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$
An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1} | X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p (d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1} | X_n = k, D_1, \ldots, D_k] = p (D_1 + \cdots + D_k)$. 
An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1}|X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1}|X_n = k, D_1, \ldots, D_k] = p(D_1 + \cdots + D_k)$.

Consequently, $E[X_{n+1}|X_n = k] = E[p(D_1 + \cdots + D_k)] = pdk$. 


An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1}|X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1}|X_n = k, D_1, \ldots, D_k] = p(D_1 + \cdots + D_k)$.

Consequently, $E[X_{n+1}|X_n = k] = E[p(D_1 + \cdots + D_k)] = pdk$.

Finally, $E[X_{n+1}|X_n] = pdX_n$, 
Application: Going Viral

An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1}|X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1}|X_n = k, D_1, \ldots, D_k] = p(D_1 + \cdots + D_k)$.

Consequently, $E[X_{n+1}|X_n = k] = E[p(D_1 + \cdots + D_k)] = pdk$.

Finally, $E[X_{n+1}|X_n] = pdX_n$, and $E[X_{n+1}] = pdE[X_n]$. 
Application: Going Viral

An easy extension: Assume that everyone has an independent number $D_i$ of friends with $E[D_i] = d$. Then, the same fact holds.

To see this, note that given $X_n = k$, and given the numbers of friends $D_1 = d_1, \ldots, D_k = d_k$ of these $X_n$ people, one has $X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1}|X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1}|X_n = k, D_1, \ldots, D_k] = p(D_1 + \cdots + D_k)$.

Consequently, $E[X_{n+1}|X_n = k] = E[p(D_1 + \cdots + D_k)] = pdk$.

Finally, $E[X_{n+1}|X_n] = pdX_n$, and $E[X_{n+1}] = pdE[X_n]$.

We conclude as before.
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in \{0, 1, 2, \ldots\}
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in $\{0, 1, 2, \ldots\}$ and $E[X_n] = \mu$ for all $n \geq 1$. 

Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in $\{0, 1, 2, \ldots\}$ and $E[X_n] = \mu$ for all $n \geq 1$.

Then,

$$E[X_1 + \cdots + X_Z] = \mu E[Z].$$
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in $\{0, 1, 2, \ldots\}$

and $E[X_n] = \mu$ for all $n \geq 1$.

Then,

$$E[X_1 + \cdots + X_Z] = \mu E[Z].$$

**Proof:**

Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in $\{0, 1, 2, \ldots\}$ and $E[X_n] = \mu$ for all $n \geq 1$.

Then,

$$E[X_1 + \cdots + X_Z] = \mu E[Z].$$

**Proof:**

$$E[X_1 + \cdots + X_Z | Z = k] = \mu k.$$
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that \( X_1, X_2, \ldots \) and \( Z \) are independent, where \( Z \) takes values in \( \{0, 1, 2, \ldots\} \) and \( E[X_n] = \mu \) for all \( n \geq 1 \).

Then,

\[
E[X_1 + \cdots + X_Z] = \mu E[Z].
\]

**Proof:**

\[
E[X_1 + \cdots + X_Z|Z = k] = \mu k.
\]

Thus, \( E[X_1 + \cdots + X_Z|Z] = \mu Z \).
Application: Wald’s Identity

Here is an extension of an identity we used in the last slide.

**Theorem** Wald’s Identity

Assume that $X_1, X_2, \ldots$ and $Z$ are independent, where $Z$ takes values in $\{0, 1, 2, \ldots\}$ and $E[X_n] = \mu$ for all $n \geq 1$.

Then,

$$E[X_1 + \cdots + X_Z] = \mu E[Z].$$

**Proof:**

$E[X_1 + \cdots + X_Z|Z = k] = \mu k$.

Thus, $E[X_1 + \cdots + X_Z|Z] = \mu Z$.

Hence, $E[X_1 + \cdots + X_Z] = E[\mu Z] = \mu E[Z]$.  

Theorem
$E[Y|X]$ is the ‘best’ guess about $Y$ based on $X$. 
Theorem

$E[Y|X]$ is the ‘best’ guess about $Y$ based on $X$.

Specifically, it is the function $g(X)$ of $X$ that

minimizes $E[(Y - g(X))^2]$. 

Theorem

$E[Y|X]$ is the ‘best’ guess about $Y$ based on $X$. Specifically, it is the function $g(X)$ of $X$ that minimizes $E[(Y - g(X))^2]$. 
Theorem $CE = MMSE$

$g(X) := \mathbb{E}[Y | X]$ is the function of $X$ that minimizes $\mathbb{E}[(Y - g(X))^2]$.

Proof: Let $h(X)$ be any function of $X$. Then $\mathbb{E}[(Y - h(X))^2] = \mathbb{E}[(Y - g(X) + g(X) - h(X))^2] = \mathbb{E}[(Y - g(X))^2] + \mathbb{E}[(g(X) - h(X))^2] + 2\mathbb{E}[(Y - g(X))(g(X) - h(X))]$. But, $\mathbb{E}[(Y - g(X))(g(X) - h(X))] = 0$ by the projection property. Thus, $\mathbb{E}[(Y - h(X))^2] \geq \mathbb{E}[(Y - g(X))^2]$. 
CE = MMSE

Theorem CE = MMSE

\( g(X) := E[Y|X] \) is the function of \( X \) that minimizes \( E[(Y - g(X))^2] \).
**Theorem** \( CE = \text{MMSE} \)

\( g(X) := E[Y|X] \) is the function of \( X \) that minimizes \( E[(Y - g(X))^2] \).

**Proof:**
**Theorem** \( CE = \text{MMSE} \)

\( g(X) := \mathbb{E}[Y|X] \) is the function of \( X \) that minimizes \( \mathbb{E}[(Y - g(X))^2] \).

**Proof:**
Let \( h(X) \) be any function of \( X \).
**Theorem** CE = MMSE

\[ g(X) := E[Y|X] \] is the function of \( X \) that minimizes \( E[(Y - g(X))^2] \).

**Proof:**

Let \( h(X) \) be any function of \( X \). Then

\[
E[(Y - h(X))^2] =
\]
Theorem CE = MMSE

\[ g(X) := E[Y|X] \] is the function of \( X \) that minimizes \( E[(Y - g(X))^2] \).

Proof:
Let \( h(X) \) be any function of \( X \). Then

\[
E[(Y - h(X))^2] = E[(Y - g(X) + g(X) - h(X))^2]
\]
Theorem CE = MMSE

\( g(X) := E[Y|X] \) is the function of \( X \) that minimizes \( E[(Y - g(X))^2] \).

Proof:
Let \( h(X) \) be any function of \( X \). Then

\[
E[(Y - h(X))^2] = E[(Y - g(X) + g(X) - h(X))^2] \\
= E[(Y - g(X))^2] + E[(g(X) - h(X))^2] + 2E[(Y - g(X))(g(X) - h(X))].
\]
Theorem $CE = MMSE$

$g(X) := E[Y|X]$ is the function of $X$ that minimizes $E[(Y - g(X))^2]$.

Proof:
Let $h(X)$ be any function of $X$. Then

$$E[(Y - h(X))^2] = E[(Y - g(X) + g(X) - h(X))^2]$$
$$= E[(Y - g(X))^2] + E[(g(X) - h(X))^2] + 2E[(Y - g(X))(g(X) - h(X))].$$

But,

$$E[(Y - g(X))(g(X) - h(X)))] = 0 \text{ by the projection property}.$$
CE = MMSE

**Theorem** CE = MMSE

\[ g(X) := E[Y|X] \] is the function of \( X \) that minimizes \( E[(Y - g(X))^2] \).

**Proof:**

Let \( h(X) \) be any function of \( X \). Then

\[
E[(Y - h(X))^2] = E[(Y - g(X) + g(X) - h(X))^2]
\]

\[
= E[(Y - g(X))^2] + E[(g(X) - h(X))^2] + 2E[(Y - g(X))(g(X) - h(X))].
\]

But,

\[
E[(Y - g(X))(g(X) - h(X))] = 0 \text{ by the projection property.}
\]

Thus, \( E[(Y - h(X))^2] \geq E[(Y - g(X))^2] \).
$E[Y|X]$ and $L[Y|X]$ as projections

$\hat{Y} = L[Y|X]$  
$\{c + dX, c, d \in \mathbb{R}\}$

$E[Y|X]$ is the projection of $Y$ on $\{g(X), g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}\}$
$E[Y|X]$ and $L[Y|X]$ as projections

$L[Y|X]$ is the projection of $Y$ on $\{a + bX, a, b \in \mathbb{R}\}$:
\( E[Y|X] \) and \( L[Y|X] \) as projections

\( L[Y|X] \) is the projection of \( Y \) on \( \{a + bX, a, b \in \mathbb{R}\} \): LLSE
$E[Y|X]$ and $L[Y|X]$ as projections

$L[Y|X]$ is the projection of $Y$ on $\{a + bX, a, b \in \mathbb{R}\}$: LLSE

$E[Y|X]$ is the projection of $Y$ on $\{g(X), g(\cdot) : \mathbb{R} \to \mathbb{R}\}$:
$E[Y|X]$ and $L[Y|X]$ as projections

$L[Y|X]$ is the projection of $Y$ on $\{a + bX, a, b \in \mathbb{R}\}$: LLSE

$E[Y|X]$ is the projection of $Y$ on $\{g(X), g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}\}$: MMSE.
Summary

Conditional Expectation

Definition:
\[ E[Y | X] := \sum_y y \Pr[Y = y | X = x] \]

Properties: Linearity, \( Y - E[Y | X] \perp h(X) \);
\[ E[E[Y | X]] = E[Y] \]

Some Applications:
Calculating \( E[Y | X] \), Diluting, Mixing, Rumors, Wald, MMSE:
\[ E[Y | X] \text{ minimizes } E[(Y - g(X))^2] \text{ over all } g(·) \]
Summary

Conditional Expectation

- Definition: \( E[Y|X] := \sum_y yPr[Y = y|X = x] \)
Conditional Expectation

Definition: \( E[Y|X] := \sum_y yPr[Y = y|X = x] \)

Properties: Linearity, \( Y - E[Y|X] \perp h(X) \)
Conditional Expectation

Definition: \( E[Y|X] := \sum_y yPr[Y = y|X = x] \)

Properties: Linearity,
\( Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \)

Some Applications:
- Calculating \( E[Y|X] \)
- Diluting
- Mixing
- Rumors
- Wald
- MMSE: \( E[Y|X] \) minimizes \( E[(Y - g(X))^2] \) over all \( g(\cdot) \)
Conditional Expectation

Definition: \( E[Y|X] := \sum_y yPr[Y = y|X = x] \)

Properties: Linearity,
\( Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \)

Some Applications:
Conditional Expectation

Definition: \( E[Y|X] := \sum_y yPr[Y = y|X = x] \)

Properties: Linearity, 
\( Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \)

Some Applications:
- Calculating \( E[Y|X] \)
Summary

Conditional Expectation

Definition: \( E[Y|X] := \sum_y y \Pr[Y = y|X = x] \)

Properties: Linearity, 
\( Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \)

Some Applications:
- Calculating \( E[Y|X] \)
- Diluting
Summary

Conditional Expectation

Definition:  
\[ E[Y|X] := \sum_y y \Pr[Y = y|X = x] \]

Properties: Linearity,  
\[ Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \]

Some Applications:
  
  - Calculating \( E[Y|X] \)
  - Diluting
  - Mixing
Conditional Expectation

Definition: 
$$E[Y|X] := \sum_y y \Pr[Y = y | X = x]$$

Properties: Linearity,
$$Y - E[Y|X] \perp h(X); \quad E[E[Y|X]] = E[Y]$$

Some Applications:
- Calculating $E[Y|X]$
- Diluting
- Mixing
- Rumors
Summary

Conditional Expectation

Definition: \( E[Y|X] := \sum_y y \Pr[Y = y|X = x] \)

Properties: Linearity,
\( Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \)

Some Applications:
- Calculating \( E[Y|X] \)
- Diluting
- Mixing
- Rumors
- Wald
Summary

**Conditional Expectation**

- **Definition:** \( E[Y|X] := \sum_y yPr[Y = y|X = x] \)

- **Properties:** Linearity,
  \[ Y - E[Y|X] \perp h(X); \ E[E[Y|X]] = E[Y] \]

- **Some Applications:**
  - Calculating \( E[Y|X] \)
  - Diluting
  - Mixing
  - Rumors
  - Wald

- **MMSE:** \( E[Y|X] \) minimizes \( E[(Y - g(X))^2] \) over all \( g(\cdot) \)