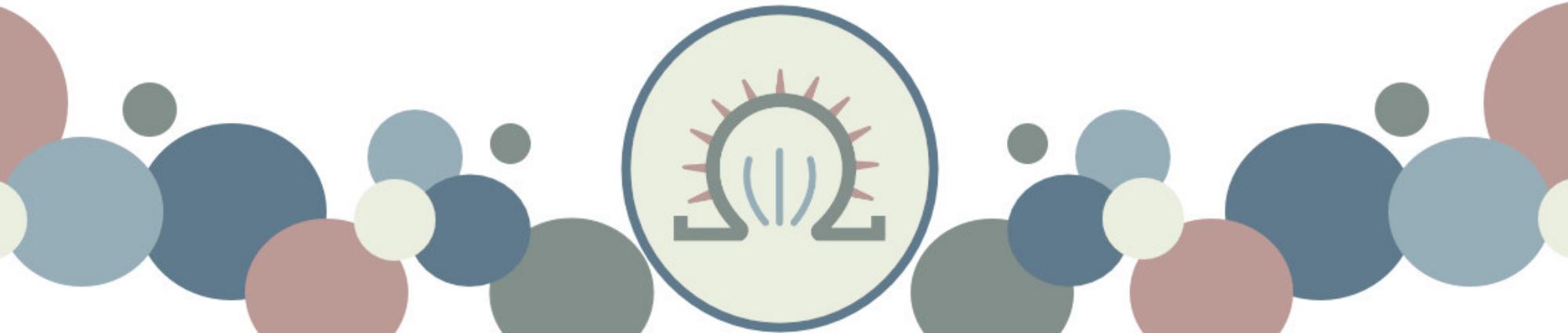
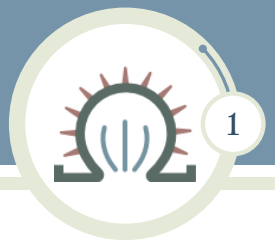
THE MATH CITADEL

Storage Failures Are *Not* Independent.
What Can We Do About It?

Jason Hathcock
www.themathcitadel.com



Contents



Storage Failures Are *Not* Independent

Figure: A simple Poisson process.

Storage Failures Are *Not* Independent

A Faulty Premise

Figure: A simple Poisson process.

Current Storage Failure Analysis

- Typical model says $\mathbb{P}(T_2 = t_2 \mid T_1 = t_1) = \mathbb{P}(T_2 = t_2)$.

Storage Failures Are *Not* Independent

A Faulty Premise

Figure: A simple Poisson process.

Current Storage Failure Analysis

- Typical model says $\mathbb{P}(T_2 = t_2 \mid T_1 = t_1) = \mathbb{P}(T_2 = t_2)$.
  - Just like coin flips.

Storage Failures Are *Not* Independent

A Faulty Premise

Figure: A simple Poisson process.

Current Storage Failure Analysis

- Typical model says $\mathbb{P}(T_2 = t_2 \mid T_1 = t_1) = \mathbb{P}(T_2 = t_2)$.
  - Just like coin flips.
  - Failures have no bearing on each other.

Storage Failures Are *Not* Independent

A Faulty Premise

Figure: A simple Poisson process.

Current Storage Failure Analysis

- This is ***false***.

Storage Failures Are *Not* Independent

A Faulty Premise

Figure: A simple Poisson process.

Current Storage Failure Analysis

- This is ***false***.
  - Bairavasunaram et al (2007) showed latent sector errors are not independent.

Storage Failures Are *Not* Independent

A Faulty Premise

Figure: A simple Poisson process.

Current Storage Failure Analysis

- This is ***false***.
  - Bairavasunaram et al (2007) showed latent sector errors are not independent.
  - Schroeder and Gibson (2007) showed the same goes for disk failures overall.

What Can We Do About It?

The Usual Approach

The Usual Approach

- Typically we use a Poisson process to count random events.

The Usual Approach

- Typically we use a Poisson process to count random events.
  - Why Poisson? The model makes sense and the math is easy to carry out.

The Usual Approach

- Typically we use a Poisson process to count random events.
  - Why Poisson? The model makes sense and the math is easy to carry out.
- But the events in a Poisson process are assumed to be independent, which we know to be false.

What Can We Do About It?

The Classic Strategy Fails Us

Is The Poisson Model Bad?

No, it just isn't general enough.

Definition

We say events A and B are **independent** to mean

$$\mathbb{P}(B) = \mathbb{P}(B \mid A).$$

Definition

We say events A and B are **independent** to mean

$$\mathbb{P}(B) = \mathbb{P}(B \mid A).$$

We say A and B are **dependent** if they are *not independent*.

Definition

We say events A and B are **independent** to mean

$$\mathbb{P}(B) = \mathbb{P}(B \mid A).$$

We say A and B are **dependent** if they are *not independent*.

1) This is a definition by negation.

Definition

We say events A and B are **independent** to mean

$$\mathbb{P}(B) = \mathbb{P}(B \mid A).$$

We say A and B are **dependent** if they are *not independent*.

1) This is a definition by negation.
2) It offers no concrete equations.

Definition

We say events A and B are **independent** to mean

$$\mathbb{P}(B) = \mathbb{P}(B \mid A).$$

We say A and B are **dependent** if they are *not independent*.

1) This is a definition by negation.
2) It offers no concrete equations.

We have a new way to quantify dependence.

Dependency Theory

Example: Coin Flips

Example: Coin Flips

- Suppose $\mathbb{P}(\text{heads}) = p$ and $\mathbb{P}(\text{tails}) = 1 - p = q$.

Example: Coin Flips

- Suppose $\mathbb{P}(\text{heads}) = p$ and $\mathbb{P}(\text{tails}) = 1 - p = q$.
- We introduce dependency by taking a number δ between 0 and 1 (inclusive).

Example: Coin Flips

- Suppose $\mathbb{P}(\text{heads}) = p$ and $\mathbb{P}(\text{tails}) = 1 - p = q$.
- We introduce dependency by taking a number δ between 0 and 1 (inclusive).
- Define new values:

$$\begin{aligned} p^+ &= p + \delta q & p^- &= p - \delta p \\ q^+ &= q + \delta p & q^- &= q - \delta q \end{aligned}$$

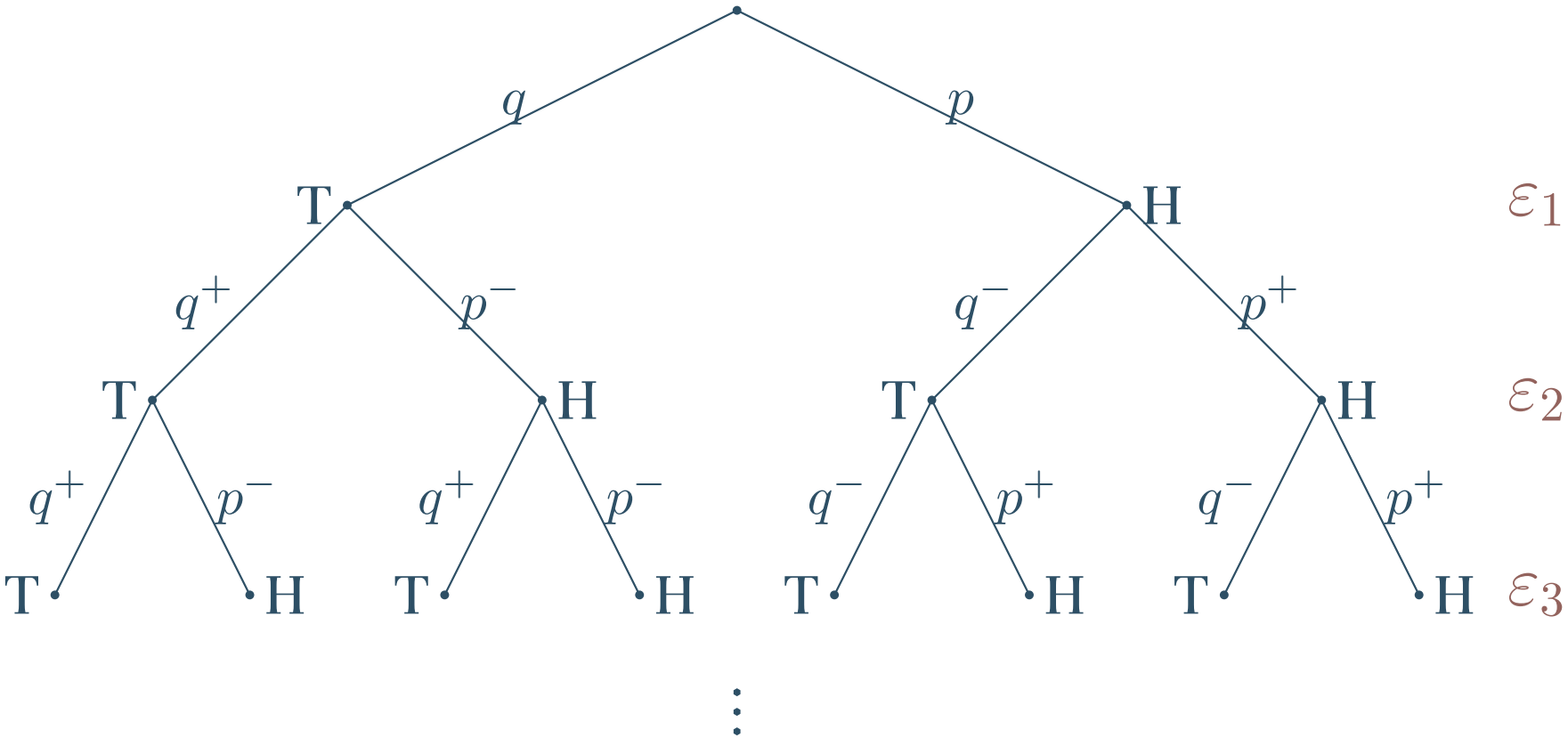


Figure: First-Kind Dependence. All flips depend on the first one.

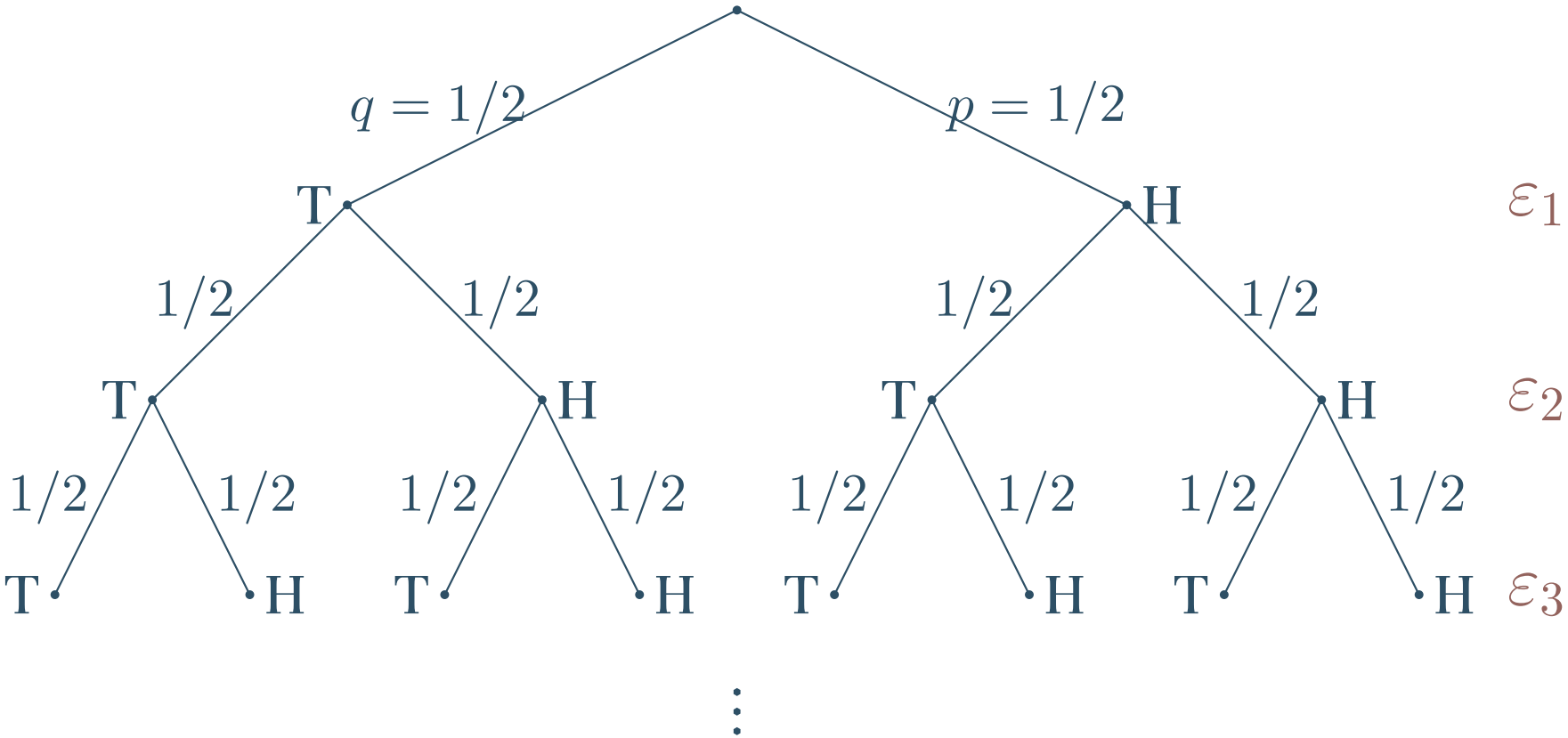


Figure: Example 1. (Complete Independence) $p = 1/2 = q$, $\delta = 0$.

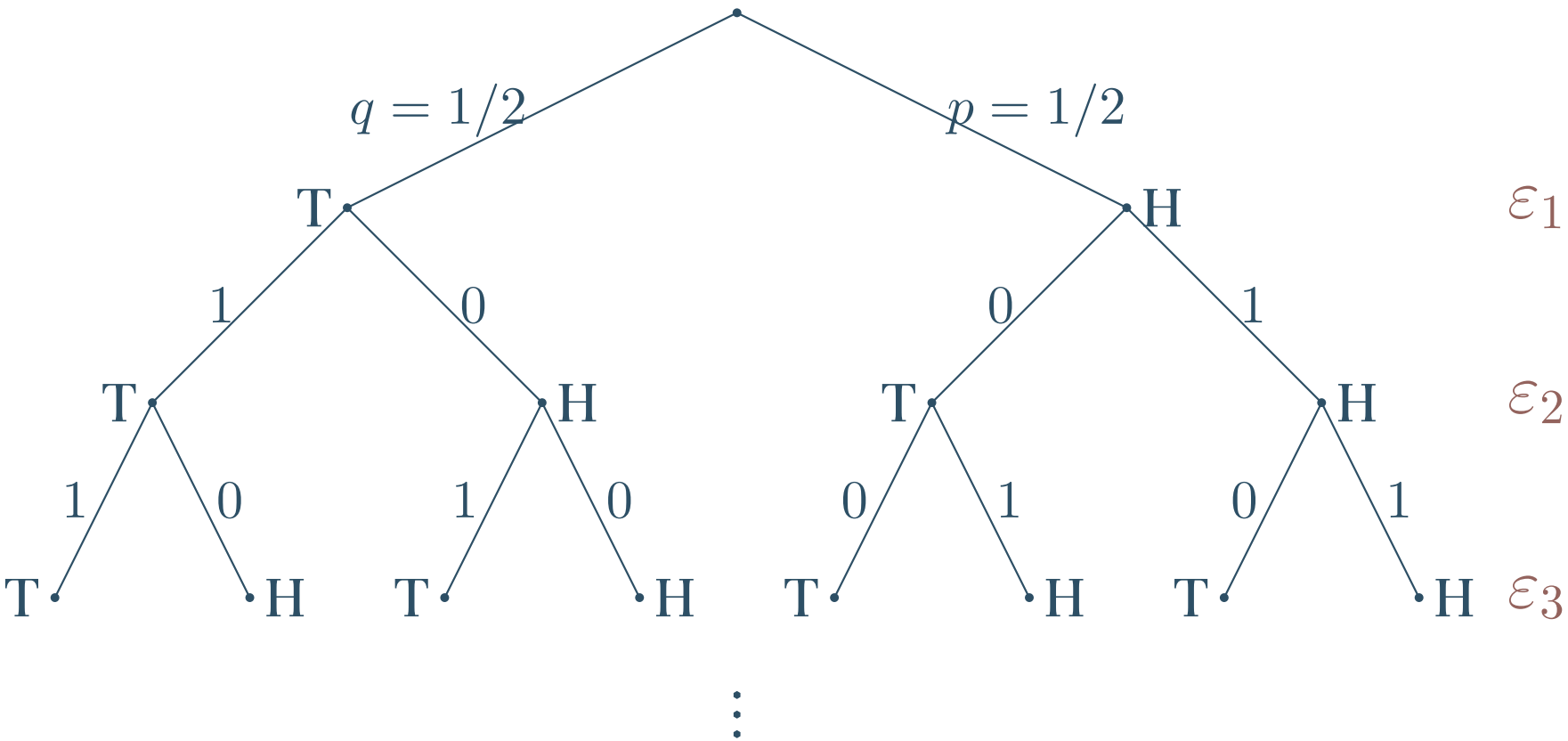


Figure: Example 2. (Total Dependence) $p = 1/2 = q$, $\delta = 1$.

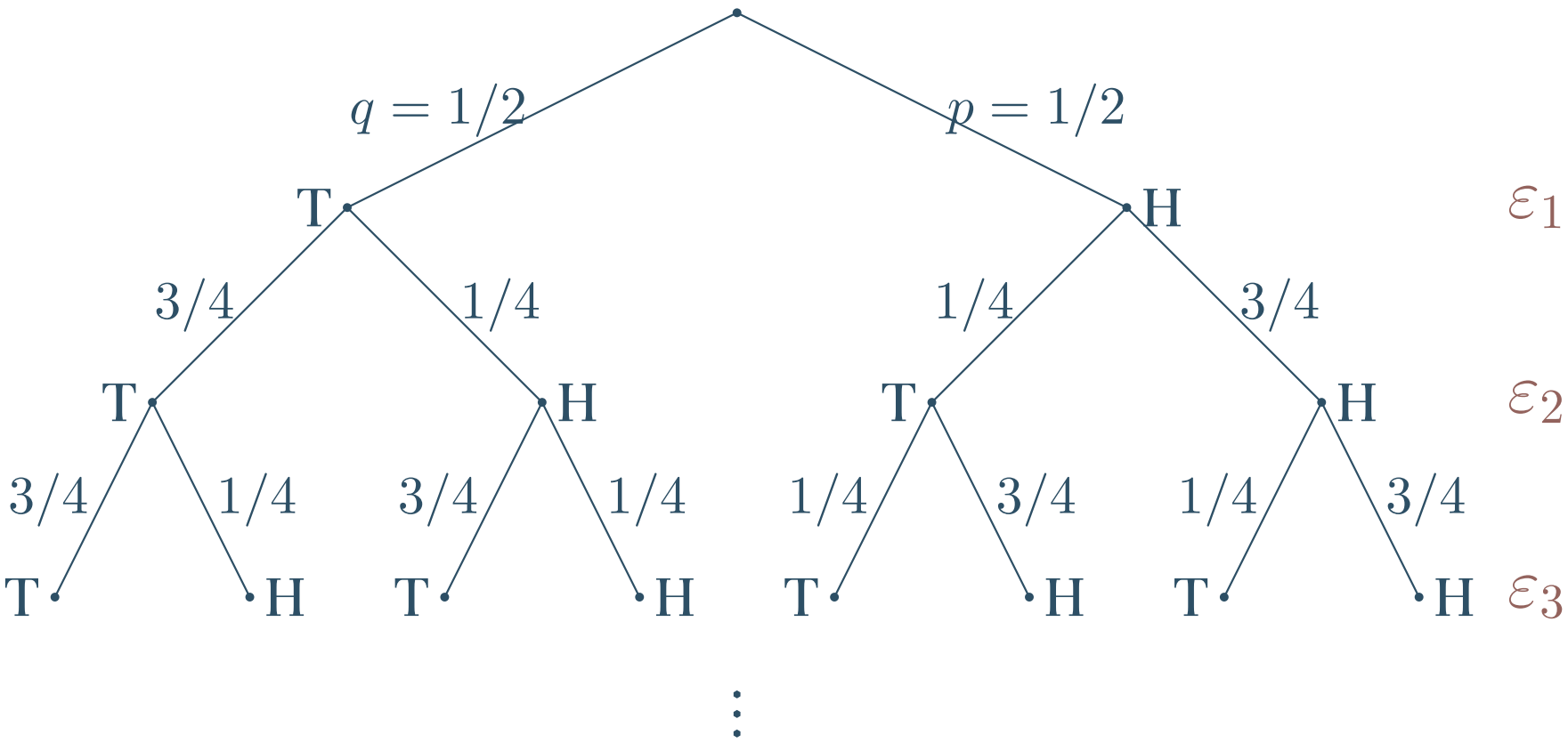


Figure: Example 3. $p = 1/2 = q$, $\delta = 1/2$.

First-Kind Dependence May Not Fit

First-Kind Dependence May Not Fit

- Dependence may follow a different scheme in practice.

First-Kind Dependence May Not Fit

- Dependence may follow a different scheme in practice.
- We have developed a rich collection of alternative dependence patterns.

First-Kind Dependence May Not Fit

- Dependence may follow a different scheme in practice.
- We have developed a rich collection of alternative dependence patterns.

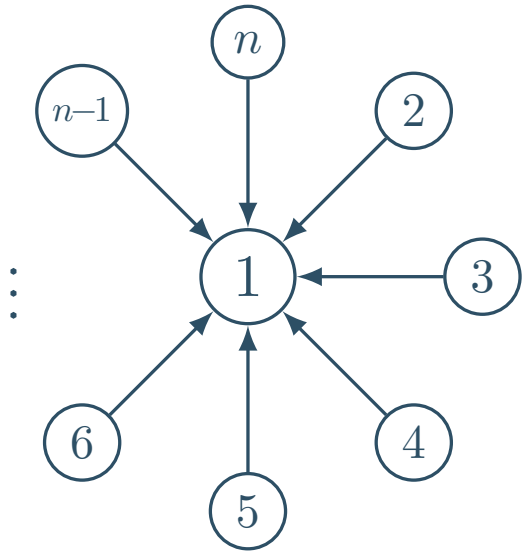


Figure: First-Kind Dependency Diagram, all trials point back to the first.

Figure: Sequential Dependency Diagram, each trial points back to the previous one.

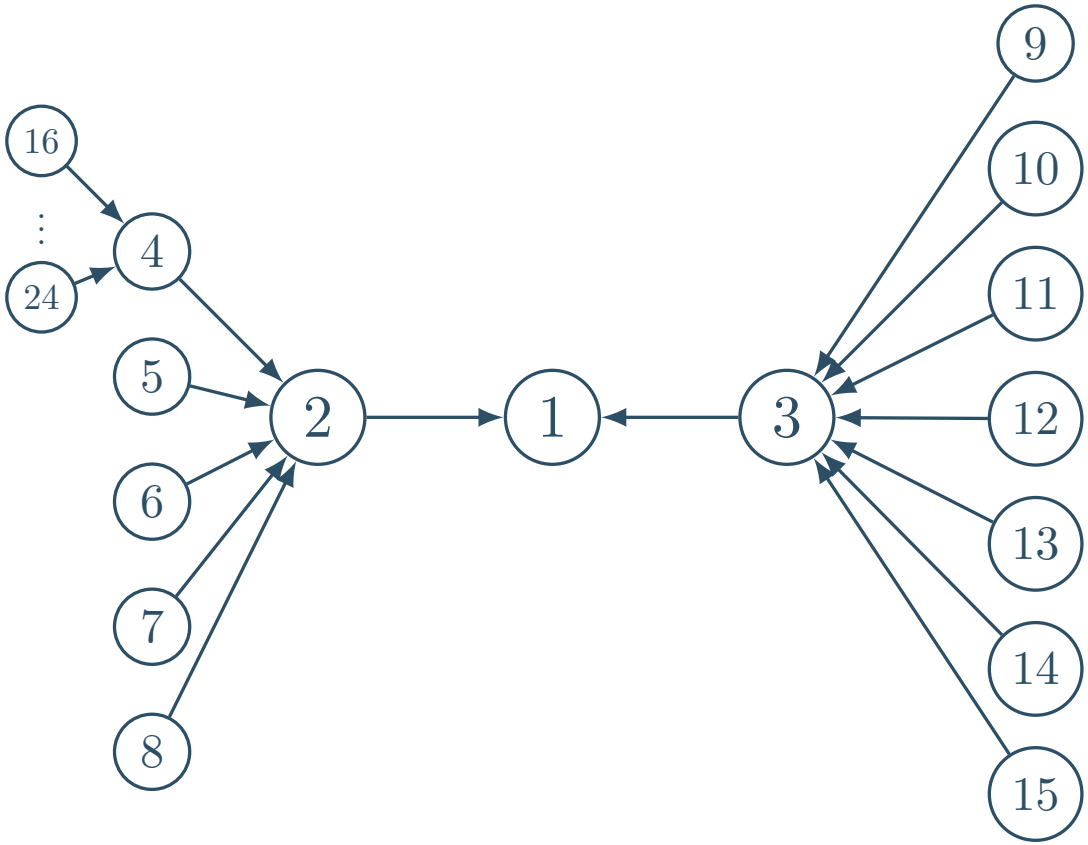


Figure: Square Root Dependency Diagram, governed by $\alpha(n) = \lfloor \sqrt{n} \rfloor$.

Why Do You Care?

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.
  - The features of the Poisson process are desirable.

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.
  - The features of the Poisson process are desirable.
  - We build in the machinery to handle dependency.

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.
  - The features of the Poisson process are desirable.
  - We build in the machinery to handle dependency.
- The results? Better failure models:

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.
  - The features of the Poisson process are desirable.
  - We build in the machinery to handle dependency.
- The results? Better failure models:
  - More understandable.

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.
  - The features of the Poisson process are desirable.
  - We build in the machinery to handle dependency.
- The results? Better failure models:
  - More understandable.
  - Improved reliability predictions.

Why Do You Care?

Our Work

- We are using dependency theory to generalize the Poisson process.
  - The features of the Poisson process are desirable.
  - We build in the machinery to handle dependency.
- The results? Better failure models:
  - More understandable.
  - Improved reliability predictions.
  - Explains *how* the failures are dependent, not just *if*.

The patentable applications of
this fundable research
are unclaimed.

Questions?