The St. Petersburg Paradox: The unexpected expectation

Unit 1 Lecture 3

Jonathan Auerbach
STAT 489 Pre-Cap Prof Development
jauerba@gmu.edu

September 16, 2021

Learning Objectives

After this lecture, you will be able to:

1. Describe the St. Petersburg Paradox and understand the solution proposed by Bernoulli.

2. Calculate the expected value or expected utility with a tree diagram.

3. Explain in what sense the expected value fails to fairly represent the outcome of an experiment.

4. Graph a recursive tree diagram using `ggtree`. See Appendix for `R` code.

These slides use the following `R` packages

Setup:

```r
library("tidyverse")
library("treeio")
library("ggtree")
theme_set(theme_bw())
```

The package `ggtree` is not available on the Comprehensive R Archive Network (CRAN). Install it from Bioconductor:

```r
library("BiocManager")
BiocManager::install("ggtree")
```

The St. Petersburg Paradox

- Bernoulli considered the following game:
  - The casino repeatedly flips a "fair" coin until it lands on heads.
  - The casino then pays the player $\$2^n$, where n is the number of times the coin was flipped.
  - What is a fair price for this game? i.e. How much money should a player be willing to pay to play it?

- The paradox is that the expected winnings are infinite—the average amount won has no upper limit—but no reasonable person would pay even $100 to play, let alone their entire wealth.
  - Bernoulli (1738) worked on the paradox in his paper *Exposition of a new theory on the measurement of risk*.
  - His solution introduced the idea of utility: A gambler does not bet based on expected winnings but rather expected utility. As wealth increases, more money does not yield as much utility.

- Expected utility—and equivalent formulations expected loss and expected regret—have become the standard framework for making decisions under uncertainty.

Daniel Bernoulli (1750) and *Exposition* (1738)

SPECIMEN
THEORIAE NOVAE
DE
MENSVRA SORTIS.
AVCTORE
Daniele Bernoulli.

I.

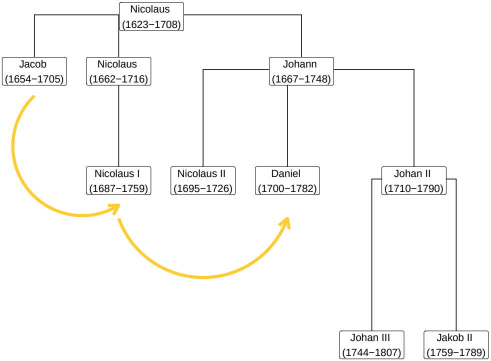
Ex eo tempore, quo Geometrae conſiderare coeperunt menſuras ſortium, affirmarunt omnes, *valorem expectationis obtineri, cum valores ſinguli expectati multiplicentur per numerum caſuum quibus obtingere poſſunt, aggregatumque productorum diuidatur per ſummam omnium caſuum: caſus autem conſiderare iubent, qui ſint inter ſe aeque procliues*: Hacque poſita regula, quodcunque reliquum eſt in iſta doctrina huc redit, vt caſus omnes enumerentur, in aeque procliues reſoluantur atque in debitam claſſem diſponantur.

§. 2. Demonſtrationes huius propoſitionis, quarum quidem in lucem prodierunt multae, ſi recte examines, omnes videbis hac inniti hypotheſi, *quod cum nulla ſit ratio, cur expectanti plus tribui debeat vni quam alteri, vnicuique aequae ſint adiudicandae partes*; rationes autem nullas conſiderari, quae

Source: https://en.wikipedia.org/wiki/Daniel_Bernoulli#/media/File:ETH-BIB-Bernoulli,_Daniel_(1700-1782)-Portrait-Portr_10971.tif_(cropped).jpg

Bernoulli: most famous family of mathematicians

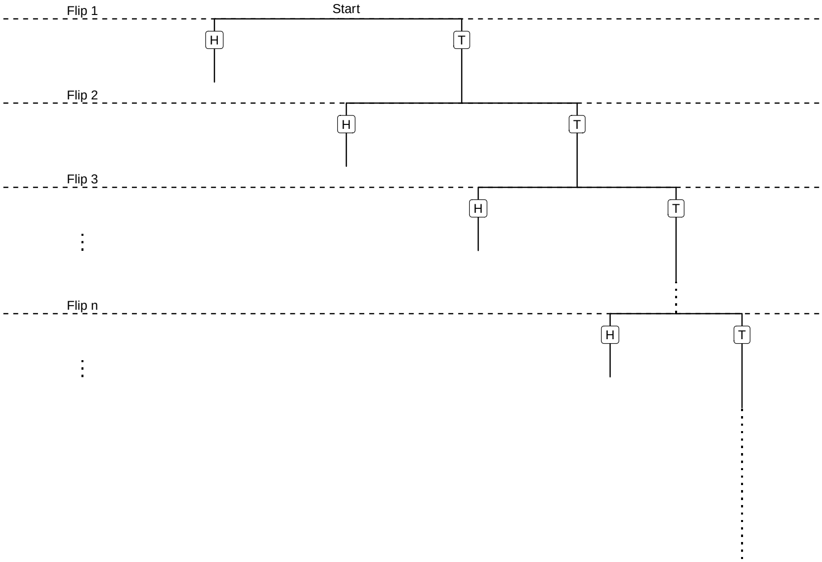
- Three generations of one family generated an incredible amount of knowledge during the 17th and 18th century.
- Relevant for this lecture: Jacob Bernoulli developed and popularized expected value (1713 posthumously, following Pascal and Fermat). Nicolaus I Bernoulli stated the paradox (1713). Daniel Bernoulli invented utility to resolve it (1738, following Gabriel Cramer).



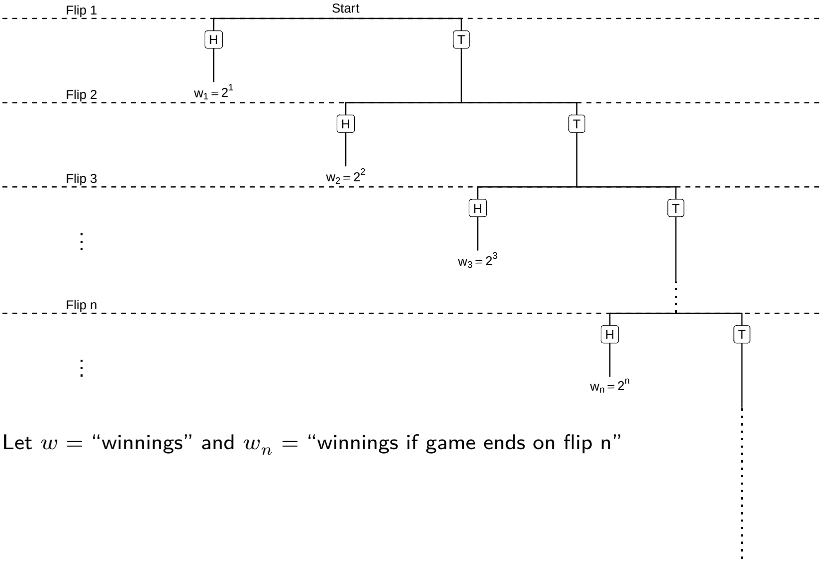
Why was the St. Petersburg Paradox so important?

- Jacob Bernoulli discovered the Law of Large Numbers (posth. 1713)
  - He showed that in large samples, sample frequencies are close to their expectations with high probability. i.e. Given enough experience, one can learn the "degree of certainty" with which future events occur.
  - Many assumed knowing the expected outcome was sufficient for decision making—that it always supported reasonable conclusions.
- Nicholas Bernoulli presented paradox to Pierre de Montmort (1713).
  - It was controversial because it suggested that the expected outcome is not always a reasonable basis for decision making.
- Daniel Bernoulli published his resolution in the annals of the Academy of St. Petersburg (1738, hence "St. Petersburg Paradox").
  - He argued a "fair" price should take into account the diminishing value of money—that money increases utility at a decreasing rate.
  - In this lecture, we use tree diagrams to demonstrate the paradox and D. Bernoulli's resolution.

Step 1: Enumerate all possible outcomes

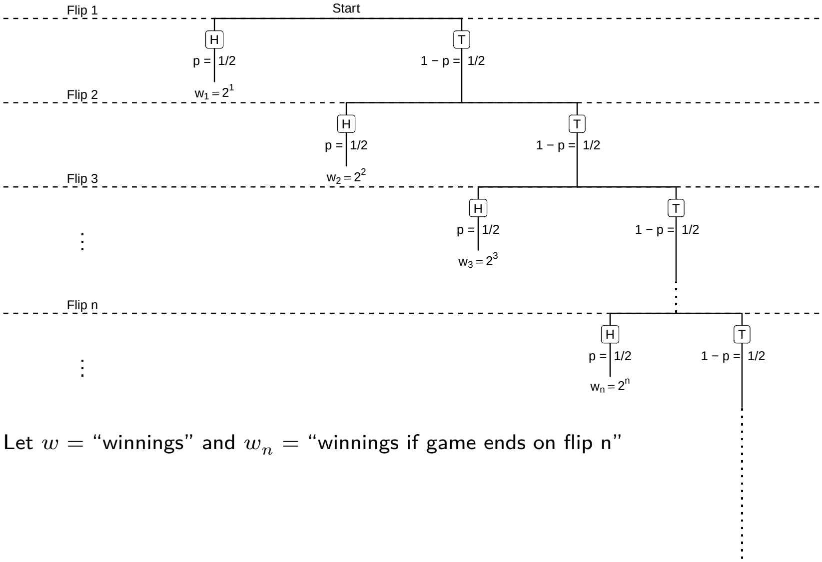


Step 1: Enumerate all possible outcomes



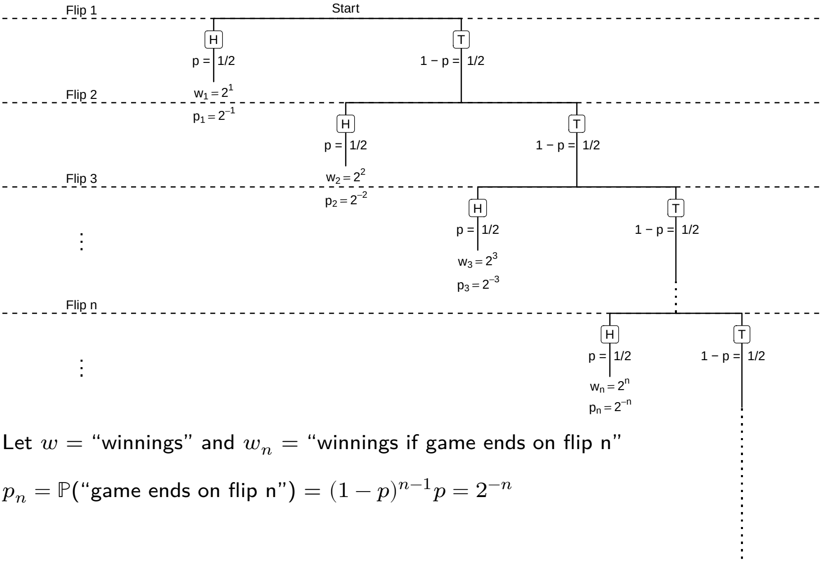
Let w = "winnings" and w_n = "winnings if game ends on flip n"

Step 2: Label the probability of outcomes by stage



Let $w =$ "winnings" and $w_n =$ "winnings if game ends on flip n"

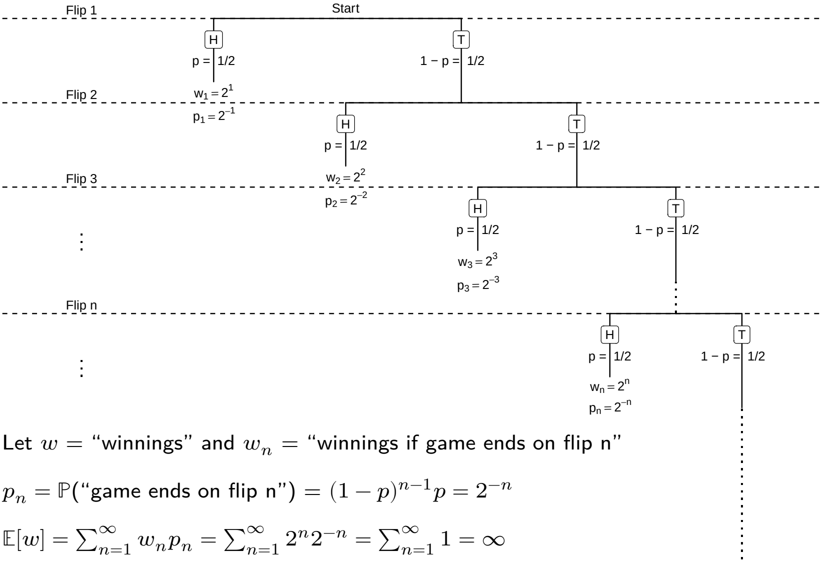
Step 3: Multiply vertical probabilities



Let w = "winnings" and w_n = "winnings if game ends on flip n"

$p_n = \mathbb{P}(\text{"game ends on flip n"}) = (1-p)^{n-1}p = 2^{-n}$

Step 4: Add probability-weighted outcomes



Let $w =$ "winnings" and $w_n =$ "winnings if game ends on flip n"

$p_n = \mathbb{P}(\text{"game ends on flip n"}) = (1-p)^{n-1}p = 2^{-n}$

$\mathbb{E}[w] = \sum_{n=1}^{\infty} w_n p_n = \sum_{n=1}^{\infty} 2^n 2^{-n} = \sum_{n=1}^{\infty} 1 = \infty$

Bernoulli's resolution relies on expected "utility"

- He thought it unrealistic that players value all winnings equally.
  - ▷ In practice, the first million is more valuable than the second million. (Even though two million dollars is more valuable than one million.)
  - ▷ But all winnings have same weight in an expected value calculation.

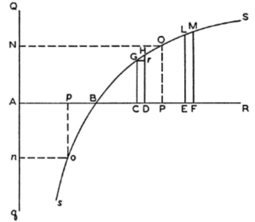
- Let utility, $u(w)$, denote the value derived from winning w.
  - ▷ Bernoulli believed the increase in utility of the winnings should be inversely proportional to total wealth.
  - ▷ This implies utility should be a logarithmic function of winnings:
    - Suppose $\frac{du}{dw} = \frac{c_1}{w + c_2}$, where w is the amount won, c_1 is the relative (marginal) value of wealth, and c_2 is the player's wealth before playing.
    - Integrating yields $u(w) = c_0 + c_1 \log_e(w + c_2)$.
    - To make calculations in this lecture easier, we assume that $c_0 = c_2 = 0$ and $c_1 = \log_2(e)$. In this case, $u(w) = \log_2(w)$.

Exposition (1738, reprinted Econometrica 1954)

that a rich prisoner who possesses two thousand ducats but needs two thousand ducats more to repurchase his freedom, will place a higher value on a gain of two thousand ducats than does another man who has less money than he. Though innumerable examples of this kind may be constructed, they represent exceedingly rare exceptions. We shall, therefore, do better to consider what usually happens, and in order to perceive the problem more correctly we shall assume that there is an imperceptibly small growth in the individual's wealth which proceeds continuously by infinitesimal increments. Now it is highly probable that *any increase in wealth, no matter how insignificant, will always result in an increase in utility which is inversely proportionate to the quantity of goods already possessed.* To explain this hypothesis it is necessary to define what is meant by the *quantity of goods.* By this expression I mean to connote food, clothing, all things which add to the conveniences of life, and even to luxury—anything that can contribute to the adequate satisfaction of any sort of want. There is then nobody who can be said to possess nothing at all in this sense unless he starves to death. For the great majority the most valuable portion of their possessions so defined will consist in their productive capacity, this term being taken to include even the beggar's talent: a man who is able to acquire ten ducats yearly by begging will scarcely be willing to accept a sum of fifty ducats on condition that he henceforth refrain from begging or otherwise trying to earn money. For he would have to live on this amount, and after he had spent it his existence must also come to an end. I doubt whether even those who do not possess a farthing and are burdened with financial obligations would be willing to free themselves of their debts or even to accept a still greater gift on such a condition. But if the beggar were to refuse such a contract unless immediately paid no less than one hundred ducats and the man pressed by creditors similarly demanded one thousand ducats, we might say that the former is possessed of wealth worth one hundred, and the latter of one thousand ducats, though in common parlance the former owns nothing and the latter less than nothing.

§6. Having stated this definition, I return to the statement made in the previous paragraph which maintained that, in the absence of the unusual, the *utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods previously possessed.* Considering the nature of man, it seems to me that the foregoing hypothesis is apt to be valid for many people to whom this sort of comparison can be applied. Only a few do not spend their entire yearly incomes. But, if among these, one has a fortune worth a hundred thousand ducats and another a fortune worth the same number of semi-ducats and if the former receives from it a yearly income of five thousand ducats while the latter obtains the same number of semi-ducats it is quite clear that to the former a ducat has exactly the same significance as a semi-ducat to the latter, and that, therefore, the gain of one ducat will have to the former no higher value than the gain of a semi-ducat to the latter. Accordingly, if each makes a gain of one ducat the latter receives twice as much utility from it, having been enriched by two semi-ducats. This argument applies to many other cases which, therefore, need not

be discussed separately. The proposition is all the more valid for the majority of men who possess no fortune apart from their working capacity which is their only source of livelihood. True, there are men to whom one ducat means more than many ducats do to others who are less rich but more generous than they. But since we shall now concern ourselves only with one individual (in different states of affluence) distinctions of this sort do not concern us. The man who is emotionally less affected by a gain will support a loss with greater patience. Since, however, in special cases things can conceivably occur otherwise, I shall first deal with the most general case and then develop our special hypothesis in order thereby to satisfy everyone.

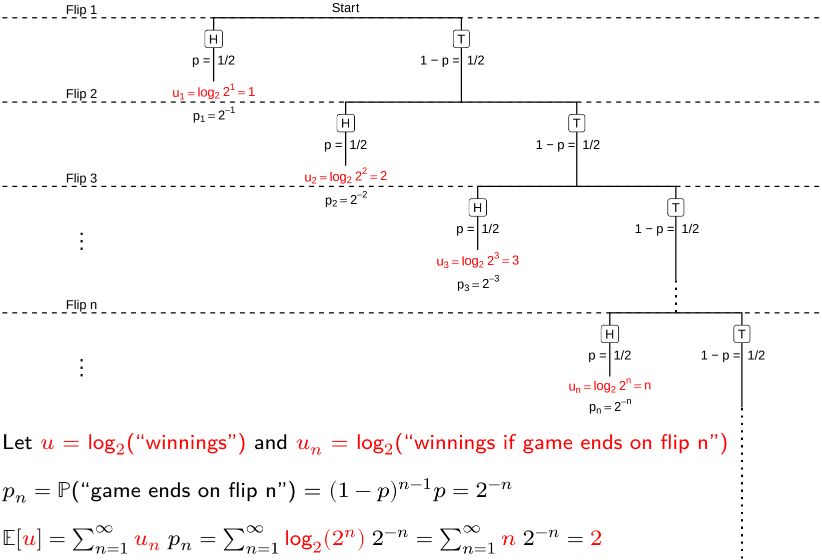


§7. Therefore, let AB represent the quantity of goods initially possessed. Then after extending AB, a curve $BGLS$ must be constructed, whose ordinates CG, DH, EL, FM, etc., designate *utilities* corresponding to the abscissas BC, BD, BE, BF, etc., designating gains in wealth. Further, let m, n, p, q, etc., be the numbers which indicate the number of ways in which gains in wealth BC, BD, BE, BF [misprinted in the original as CF], etc., can occur. Then (in accord with §4) the *moral* expectation of the risky proposition referred to is given by:

$$PO = \frac{m.CG + n.DH + p.EL + q.FM + \cdots}{m + n + p + q + \cdots}$$

Now, if we erect AQ perpendicular to AR, and on it measure off $AN = PO$, the straight line $NO - AB$ represents the gain which may properly be expected, or the value of the risky proposition in question. If we wish, further, to know how

The game is worth only \$4 if $u = \log_2(w)$



Let $u = \log_2(\text{“winnings”})$ and $u_n = \log_2(\text{“winnings if game ends on flip n”})$

$p_n = \mathbb{P}(\text{“game ends on flip n”}) = (1-p)^{n-1}p = 2^{-n}$

$\mathbb{E}[u] = \sum_{n=1}^{\infty} u_n\, p_n = \sum_{n=1}^{\infty} \log_2(2^n)\, 2^{-n} = \sum_{n=1}^{\infty} n\, 2^{-n} = 2$

n.b. $\sum_{n=1}^{\infty} n\, p^n = \frac{p}{(1-p)^2}$ for $0 \le p < 1$ n.b. 2 “utils” = \$4

Bernoulli's utility is the first of many resolutions

- The St. Petersburg Paradox works by placing very large weight on very rare outcomes—outcomes that cannot happen in practice.
  - Bernoulli's resolution reduces the relative weight of those outcomes.
  - However, the log transformation does not solve the problem of infinite expectations in general.
    - In fact, the game can be adjusted to produce infinite expectations by simply increasing the payout. (e.g. set $w_n = 2^{2^n}$)
- Scientists continue to write about the paradox. Other resolutions:
  - Poisson: Arbitrarily large payouts are unrealistic; the world has finite wealth. If payouts are capped, the expected winnings are small.
    - e.g. Suppose a casino only has \$100 million. Then the game must stop before round 27 since $2^{27} > 100{,}000{,}000$
    - If game must stop before round 27, the expected winnings are \$27.
  - Condorcet: Expected winnings do not tell you the value of any one bet, only the value of repeating a bet many times.
    - If the game is repeated a large enough number of times, the average winnings across all plays will exceed any predetermined price.

How much would a real player pay to play?

- We cannot know for sure. The game cannot be played in practice.
  - But when Cox et al. (2011) offered a finite version of the St. Petersburg game—i.e. when players were offered the opportunity to pay \$8.75 to play for a maximum of 9 rounds—83% declined.
    - n.b. There is less than a 0.5 percent chance of getting to round 9.
  - Since the expected winnings of this game are \$9, this suggests most players do not base their decisions on the expected value.
  - Most players are risk averse. i.e. Even though they would make money on average, they do not want to risk the money they have.

- More important than resolving the St. Petersburg Paradox, Bernoulli's insight helped changed our interpretation of data.
  - Decision theory, regression models, and many other statistical tools work by maximizing utility (or an approach similar to maximizing utility like minimizing loss or regret).

References

1. Bernoulli, Daniel. “Exposition of a new theory on the measurement.” Econometrica 22.1 (1954): 23-36.

2. Cox, James C., Vjollca Sadiraj, and Bodo Vogt. “On the empirical relevance of St. Petersburg lotteries.” Petersburg Lotteries (January 1, 2011). Andrew Young School of Policy Studies Research Paper Series 11-04 (2011).

3. Diaconis, Persi, and Brian Skyrms. “Ten great ideas about chance.” Ten Great Ideas about Chance. Princeton University Press, 2017.

4. Gigerenzer, Gerd, Zerno Swijtnik, Theodore Porter, Lorraine Daston, John Beatty, and Lorenz Kruger. “The empire of chance: How probability changed science and everyday life.” Cambridge University Press, 1990.

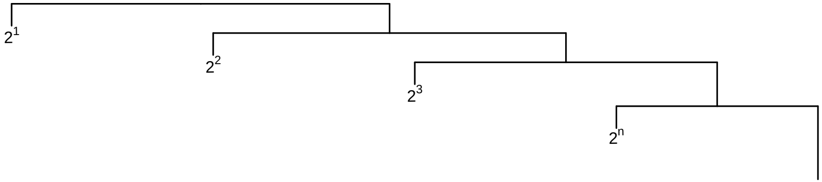
5. Hacking, Ian. The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference. Cambridge University Press, 2006.

Appendix: Newick Representation of Decision Tree

```r
# Tree coded using Newick format
## parens. denote grouping of terminal nodes
## c.f. https://en.wikipedia.org/wiki/Newick_format

tree_text <- "(b:1.5,(c:1.5,(d:1.5,(e:1.5, f:5))))a;"
tree_data <- treeio::read.newick(text = tree_text)
tree_data$edge.length[c(2, 4, 6)] <- c(2, 2, 3)
tree_labels <- tibble(label = letters[1:6],
                      label_text = c(paste0("2^",0:3),"2^n",""))

ggtree(tree_data) %<+% tree_labels + layout_dendrogram() +
  geom_tiplab(geom = "text", aes(label = label_text),
              parse = T, vjust = 1, hjust = .5, angle = 1)
```



Appendix: Graph a lightly annotated tree

```
(decision_tree_unlabeled <-
  ggtree(tree_data) %<+% tree_labels +
  layout_dendrogram() +
  annotate("text", x = -12.25, y = 2, label = "Start") +
  annotate("label", label = rep(c("H", "T"), 4),
      x = -c(11.5, 11.5, 9.5, 9.5, 7.5, 7.5, 4.5, 4.5),
      y = c(1, 2.875, 2, 3.75, 3, 4.5, 4, 5)) +
  xlim(0, -12.25) +
  geom_tiplab(geom = "text", aes(label = label_text),
              parse = T, vjust = 1, hjust = .5,
              angle = 1) +
  geom_linerange(y = 5, xmin = -1, xmax = 2.75,
                 color = "white", size = 2) +
  geom_linerange(y = 4.5, xmin = 5.02, xmax = 5.75,
                 color = "white", size = 2) +
  geom_linerange(y = 5, xmin = -1, xmax = 2.75,
                 linetype = "dotted",
                 size = .75) +
  geom_linerange(y = 4.5, xmin = 5.02, xmax = 5.75,
                 linetype = "dotted", size = .75))
```

Appendix: Graph an annotated tree

```
(decision_tree_labeled <-
  decision_tree_unlabeled +
  annotate("text", y = 0,
           x = -(c(5, 8, 10, 12)) -.2,
           label = paste("Flip", c("n", 3:1))) +
  geom_vline(xintercept = -(c(5, 8, 10, 12)),
             linetype = 2) +
  annotate("text",
           x = -c(11.5, 11.5, 9.5, 9.5,
                  7.5, 7.5, 4.5, 4.5) + .5,
           y = c(1, 2.875, 2, 3.75, 3, 4.5, 4, 5),
           label = rep(c("p =  1/2",
                         "1 - p =  1/2     "), 4))) +
  geom_linerange(y = 0, xmin = 6.5, xmax = 7,
                 linetype = "dotted", size = .75) +
  geom_linerange(y = 0, xmin = 3.5, xmax = 4,
                 linetype = "dotted", size = .75)
```