Likelihood Function: Coin example

Adopting the MLE approach, we aim to find the value of θ which maximizes $\mathcal{L}(\theta) = \theta^7(1-\theta)^3$. We can arrive at this answer using derivatives:

natural

$$\log(\mathcal{L}(\theta)) = \ell(\theta) = 7\log(\theta) + 3\log(1-\theta)$$

$$\ell'(\theta) = \frac{7}{\theta} + \frac{-3}{1-\theta}$$

Set $=0$ solve for θ

$$\frac{7}{\theta} = \frac{3}{1-\theta}$$

$$\Rightarrow \hat{\theta} = \frac{7}{10} \quad \Leftarrow$$ maximum likelihood estimator (M.L.E).

- While the above assumed $n = 10$ independent Bernoulli observations, we could have just as easily modeled this using $X \sim \text{Binomial}(n, \theta)$ having pmf given by:

$$p(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \tag{3}$$

for $x = 0, 1, \ldots, n$ and $\theta \in (0, 1)$

- Since we observed 7 heads in 10 trials the likelihood is:

$$\mathcal{L}(\theta) = \binom{10}{7} \theta^7 (1-\theta)^3 \tag{4}$$

Discrete

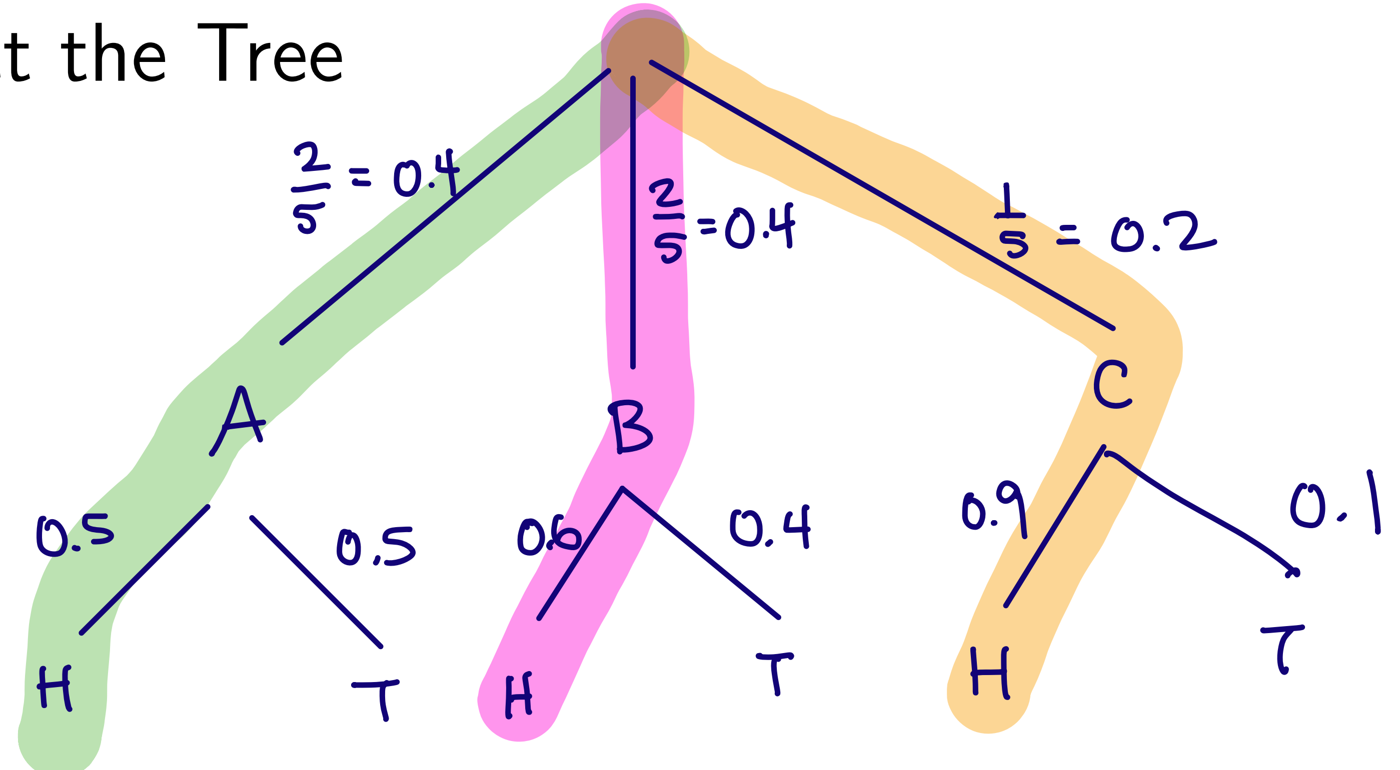
Considers three types of coins: Type A, B, and C. Each has a different probabilities of landing heads when tossed.

A coins are fair, with probability 0.5 of heads $\theta = 0.5$
B coins are bent and have probability 0.6 of heads $\theta = 0.6$
C coins are bent and have probability 0.9 of heads $\theta = 0.9$

Suppose I have a drawer containing 5 coins: 2 of type A, 2 of type B, and 1 of type C. I reach into the drawer and pick a coin at random. Without showing you the coin, I flip it once and get heads. What is the probability it is type A? Type B? Type C?

Source: Jeremy Orloff, and Jonathan Bloom. 18.05 Introduction to Probability and Statistics. Spring 2014. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu.

Construct the Tree



$P(A \cap H) = P(H|A) \cdot P(A)$
$= (0.5)(0.4)$
$= 0.20$

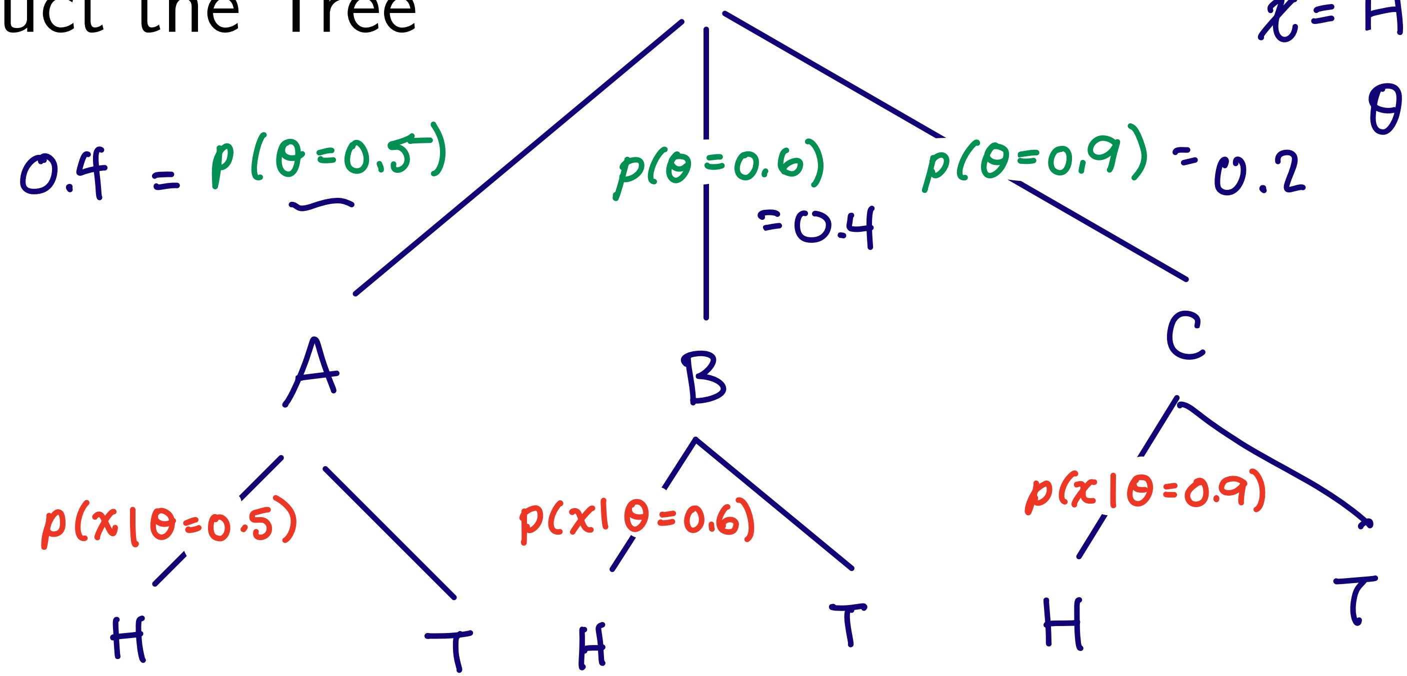
$P(B \cap H) = P(H|B) \cdot P(B)$
$= (0.6)(0.4)$
$= 0.24$

$P(C \cap H) = P(H|C) \cdot P(C)$
$= (0.9)(0.2)$
$= 0.18$

Construct the Tree

$x = H$

$\theta =$ bias towards H



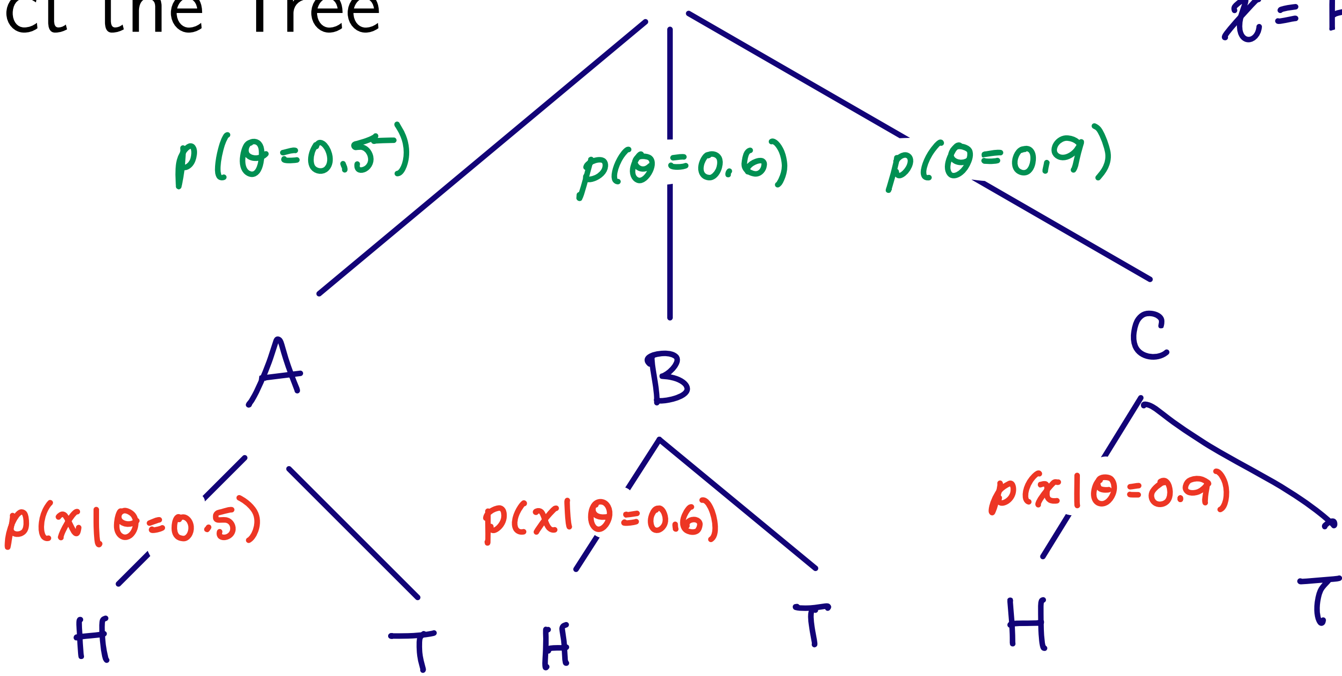
$$P(A \cap H) = P(H|A) \cdot P(A) = (0.5)(0.4) = 0.20$$

$$P(B \cap H) = P(H|B) \cdot P(B) = (0.6)(0.4) = 0.24$$

$$P(C \cap H) = P(H|C) \cdot P(C) = (0.9)(0.2) = 0.18$$

Construct the Tree

$x = H$



$= P(H|A)\cdot P(A)$
$= (0.5)(0.4)$
$= 0.20$

$= P(H|B)\cdot P(B)$
$= (0.6)(0.4)$
$= 0.24$

$= P(H|C)\cdot P(C)$
$= (0.9)(0.2)$
$= 0.18$

Bayes numerator
(unnormalized posterior)

Step 2: Likelihood

If we define a random variable $X \sim \text{Bernoulli}(\theta)$ then we can summarizing our data $x = 1$ in the following likelihood.

$$\begin{aligned}
\mathcal{L}(\theta) &= \prod_{i=1}^{1} \theta^{x_i}(1-\theta)^{1-x_i} \\
&= \begin{cases} p(x \mid \theta = 0.5) = 0.5^1(1-0.5)^{1-0} = 0.5 \\ p(x \mid \theta = 0.6) = 0.6^1(1-0.6)^{1-0} = 0.6 \\ p(x \mid \theta = 0.9) = 0.9^1(1-0.9)^{1-0} = 0.9 \end{cases}
\end{aligned}$$

Notice how these probabilities do NOT sum to 1.

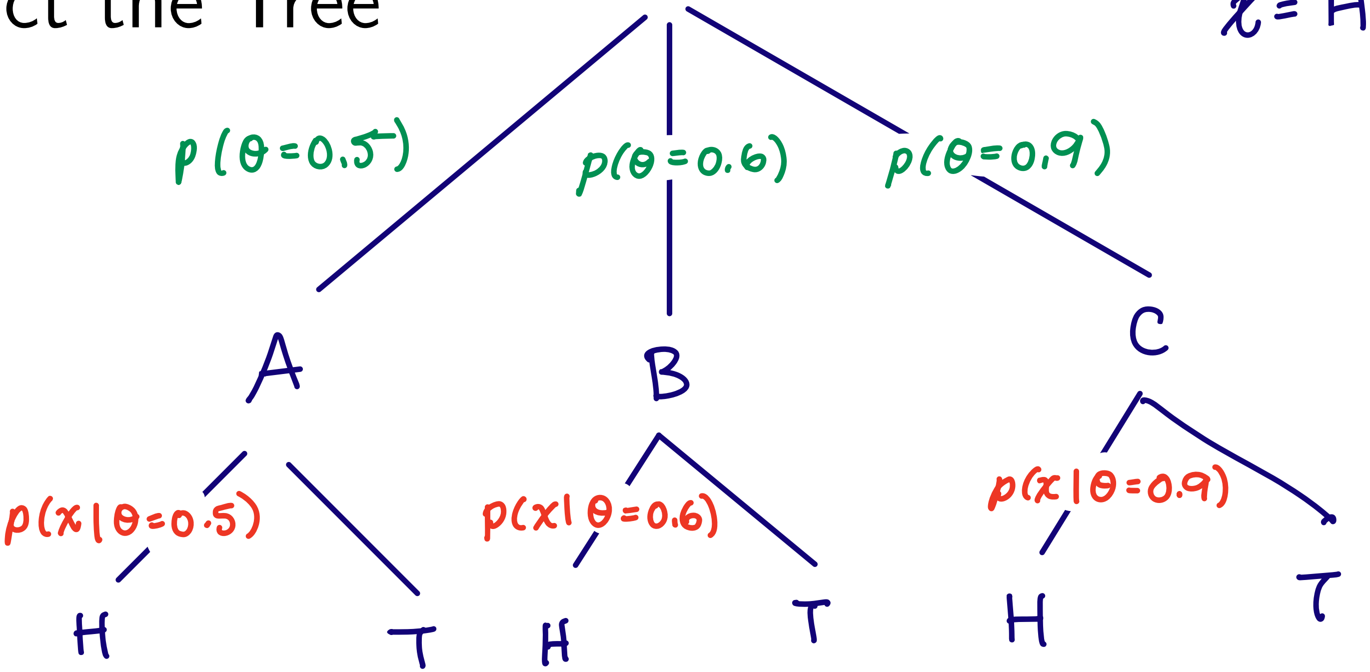
support

H_i	θ	prior	likelihood	Bayes num.	posterior*
A	0.5	$\frac{2}{5}$	0.5	$\frac{2}{5} \times 0.5 = 0.2$	0.2
B	0.6	$\frac{2}{5}$	0.6	$\frac{2}{5} \times 0.6 = 0.24$	0.24
C	0.9	$\frac{1}{5}$	0.9	$\frac{\cancel{2}\,1}{5} \times 0.9 = 0.18$	0.18
Total:		1	2	0.62	0.62

*this is an *unnormalized posterior* because it is not yet a proper pmf that sums to 1.

Construct the Tree

$x = H$



$$P(A \cap H) = P(H|A) \cdot P(A) = (0.5)(0.4) = 0.20$$

$$P(B \cap H) = P(H|B) \cdot P(B) = (0.6)(0.4) = 0.24$$

$$P(C \cap H) = P(H|C) \cdot P(C) = (0.9)(0.2) = 0.18$$

$$p(x) = P(H \cap A) + P(H \cap B) + P(H \cap C)$$
$$= P(H|A) \cdot P(A) + P(H|B) \cdot P(B) + P(H|C) \cdot P(C)$$
$$= 0.20 + 0.24 + 0.18 = 0.62$$

- More generally, the marginal distribution $p(x)$ is given by:

$$p(x) = \sum_{\cancel{y}\,\theta} p(x, \cancel{y}\,\theta) = \sum_{y} p(x \mid \cancel{y}\,\theta) p(\cancel{y}\,\theta) \qquad \text{for discrete RVs}$$

$$p(x) = \int p(x, \cancel{y}\,\theta) dy = \int p(x \mid \cancel{y}\,\theta) p(\cancel{y}\,\theta) dy \qquad \text{for continuous RVs}$$

- Often this marginal is hard if not impossible to compute so we'd like to avoid calculating it if we can.

where RV stands for random variable

Second study:

$\mathcal{D}$ata $x_2 = 1$ and using a prior equal to the the unnormalized posterior obtained from the first experiment (i.e. 3rd column from slide 67)

	prior $p(\theta)$	likelihood $p(x_2 = 1 \mid \theta)$	Bayes numerator $p(x_2 = 1 \mid \theta)p(\theta)$	posterior $p(\theta \mid x_2)$
	0.20	0.5	0.5*0.20 $= 0.100$	$\frac{0.100}{0.406} = 0.2463$
	0.24	0.6	0.6*0.24 $= 0.144$	$\frac{0.144}{0.406} = 0.3547$
	0.18	0.9	0.9*0.18 $= 0.162$	$\frac{0.162}{0.406} = 0.3990$
$\sum$	1	–	0.406	1