Machine Learning (III): Bayes' Rule: Definition, Application, and Bayesian Estimation

Le Wang

Conditional Distribution and Law of Total Probability

We consider constrution of **marginal** distribution from a **joint** distribution.

Alternative way: we can also consider **marginal** probability as a weighted sum of **conditional probabilities**.

Law of Total Probability

Law of Total Probability:

$$\begin{aligned}\Pr[Y] = &\ \Pr[Y, X = x_1] + \Pr[Y, X = x_2 \\ &+ \cdots + \Pr[Y, X = x_n]\end{aligned}$$

Law of Total Probability:

$$\Pr[Y] = \Pr[Y|X = x_1] \cdot \Pr[X = x_1] + \Pr[Y, X = x_2$$
$$+ \cdots + \Pr[Y, X = x_n]$$

Law of Total Probability:

$$\Pr[Y] = \Pr[Y|X = x_1] \cdot \Pr[X = x_1] + \Pr[Y|X = x_2] \cdot \Pr[X = x_2] \\ + \cdots + \Pr[Y, X = x_n]$$

Law of Total Probability:

$$\begin{aligned}\Pr[Y] = {} & \Pr[Y|X = x_1] \cdot \Pr[X = x_1] + \Pr[Y|X = x_2] \cdot \Pr[X = x_2] \\ & + \cdots + \Pr[Y|X = x_n] \cdot \Pr[X = x_n]\end{aligned}$$

Law of Total Probability:

$$
\begin{aligned}
\Pr[Y] &= \Pr[Y|X = x_1] \cdot \Pr[X = x_1] + \Pr[Y|X = x_2] \cdot \Pr[X = x_2] \\
&\quad + \cdots + \Pr[Y|X = x_n] \cdot \Pr[X = x_n] \\
&= \sum \Pr[Y|X = x_i] \cdot \Pr[X = x_i]
\end{aligned}
$$

Law of Total Probability:

$$\begin{aligned}\Pr[Y] &= \Pr[Y|X=x_1]\cdot\Pr[X=x_1] + \Pr[Y|X=x_2]\cdot\Pr[X=x_2] \\ &+ \cdots + \Pr[Y|X=x_n]\cdot\Pr[X=x_n] \\ &= \sum \Pr[Y|X=x_i]\cdot\Pr[X=x_i] \\ &= \sum p(y|x_i)\cdot p(x_i)\end{aligned}$$

Law of Total Probability:

$$\begin{aligned}
\Pr[Y] &= \Pr[Y|X = x_1] \cdot \Pr[X = x_1] \\
&\quad + \Pr[Y|X = x_2] \cdot \Pr[X = x_2] \\
&\quad + \cdots + \Pr[Y|X = x_n] \cdot \Pr[X = x_n] \\
&= \sum \Pr[Y|X = x_i] \cdot \Pr[X = x_i] \\
&= \sum p(y|x_i) \cdot p(x_i)
\end{aligned}$$

Example

Suppose we have a stock of gadgets from two sources: 30% of them are manufactured by factory A, in which one out of 5000 is defective, whereas 70% are manufactured by factory B, in which one out of $10,000$ is defective.

Question: What is the probability that a randomly chosen gadget will be defective?

Note that here we do not have information on the toal number of gadgets.

$$
\begin{aligned}
\Pr[Y] &= \Pr[Y|X = x_1] \cdot \Pr[X = x_1] \\
&+ \Pr[Y|X = x_2] \cdot \Pr[X = x_2] + \cdots + \Pr[Y|X = x_n] \cdot \Pr[X = x_n] \\
&= \sum \Pr[Y|X = x_i] \cdot \Pr[X = x_i] \\
&= \sum p(y|x_i) \cdot p(x_i)
\end{aligned}
$$

$$
\begin{aligned}
\Pr[Y] &= \Pr[Y|X = x_1] \cdot \Pr[X = x_1] \\
&+ \Pr[Y|X = x_2] \cdot \Pr[X = x_2] + \cdots + \Pr[Y|X = x_n] \cdot \Pr[X = x_n] \\
&= \sum \Pr[Y|X = x_i] \cdot \Pr[X = x_i] \\
&= \sum p(y|x_i) \cdot p(x_i)
\end{aligned}
$$

Solution

$$
\begin{aligned}
\Pr[\mathit{defective}] &= \Pr[\mathit{defective}|X = A] \cdot \Pr[A] \\
&+ \Pr[\mathit{defective}|X = B] \cdot \Pr[B] \\
&= \frac{1}{5000} \cdot 0.30 + \frac{1}{10000} \cdot 0.70
\end{aligned}
$$

$$
\begin{aligned}
\Pr[Y] &= \Pr[Y|X = x_1] \cdot \Pr[X = x_1] \\
&\quad + \Pr[Y|X = x_2] \cdot \Pr[X = x_2] + \cdots + \Pr[Y|X = x_n] \cdot \Pr[X = x_n] \\
&= \sum \Pr[Y|X = x_i] \cdot \Pr[X = x_i] \\
&= \sum p(y|x_i) \cdot p(x_i)
\end{aligned}
$$

Solution

$$
\begin{aligned}
\Pr[\mathit{defective}] &= \Pr[\mathit{defective}|X = A] \cdot \Pr[A] \\
&\quad + \Pr[\mathit{defective}|X = B] \cdot \Pr[B] \\
&= \frac{1}{5000} \cdot 0.30 + \frac{1}{10000} \cdot 0.70 \\
&= 0.00013
\end{aligned}
$$

Further Example: Multivalued Variables

For example, if we are interested in the percentage of female voters in the whole population, we can infer this information from the following

$$
\begin{aligned}
\Pr[\mathit{female}] &= \Pr[\mathit{female}|\mathit{whites}] \cdot \Pr[\mathit{whites}] \\
&+ \Pr[\mathit{female}|\mathit{blacks}] \cdot \Pr[\mathit{blacks}] \\
&+ \Pr[\mathit{female}|\mathit{asians}] \cdot \Pr[\mathit{asians}] \\
&+ \Pr[\mathit{female}|\mathit{hispanics}] \cdot \Pr[\mathit{hispanics}] + \ldots
\end{aligned}
$$

Conditional Distribution and Bayes' Rule

Bayes' Rule or **Theorem** is a surprisingly simple mathematical guide to accurate inference. The theorem, now 250 years old, marked the beginning of statistical inference as a serious scientific subject.

Conditional Distribution and Bayes' Rule

Bayes' Rule or **Theorem** is a surprisingly simple mathematical guide to accurate inference. The theorem, now 250 years old, marked the beginning of statistical inference as a serious scientific subject.

Thomas Bayes, who developed the theorem, was a mid-eighteenth-century nonconfomist English **minister** with substantial mathematical interests.

Conditional Distribution and Bayes' Rule

Bayes' Rule or **Theorem** is a surprisingly simple mathematical guide to accurate inference. The theorem, now 250 years old, marked the beginning of statistical inference as a serious scientific subject.

Thomas Bayes, who developed the theorem, was a mid-eighteenth-century nonconfomist English **minister** with substantial mathematical interests.

Richard Price, a leading figure of letters, science, and politics, had Bayes' theorem published in the 1763 Transcactions of the Royal Society (two years after Bayes' death), his interest being partly theological, with the rule somehow proving the existence of God (Bradley Efron, Trevor Hastie, 2018, p.36).

(General) Bayes' Rule is given by the following equation

$$p(y|x) = \frac{p(x, y)}{p(x)}$$
$$p(x|y) = \frac{p(x, y)}{p(y)} \implies p(x, y) = p(x|y) \cdot p(y)$$

(General) Bayes' Rule is given by the following equation

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$p(x|y) = \frac{p(x,y)}{p(y)} \implies p(x,y) = p(x|y) \cdot p(y)$$

$$\mathbf{p}(\mathbf{y}|\mathbf{x}) = \frac{p(x,y)}{p(x)}$$

.

(General) Bayes' Rule is given by the following equation

$$p(y|x) = \frac{p(x,y)}{p(x)}$$
$$p(x|y) = \frac{p(x,y)}{p(y)} \implies p(x,y) = p(x|y) \cdot p(y)$$

$$\mathbf{p(y|x)} = \frac{p(x|y)p(y)}{p(x)}$$

.

(General) Bayes' Rule is given by the following equation

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$p(x|y) = \frac{p(x,y)}{p(y)} \implies p(x,y) = p(x|y) \cdot p(y)$$

$$\begin{aligned}\mathbf{p(y|x)} &= \frac{p(x|y)p(y)}{p(x)} \\ &= \frac{\mathbf{p(x|y)p(y)}}{\sum_{\mathbf{j}} \mathbf{p(x|y_j) \cdot p(y_j)}}\end{aligned}$$

The last equality comes from the law of total probability.

Example (We will use this later!)

$$\Pr[\text{surname}|\text{race}] = \frac{\Pr[\text{race}|\text{surname}]\,\Pr[\text{surname}]}{\Pr[\text{race}]}$$

Example (We will use this later!)

$$\Pr[\text{surname}|\text{race}] = \frac{\Pr[\text{race}|\text{surname}]\Pr[\text{surname}]}{\Pr[\text{race}]}$$

We can obtain these terms:

1. $\Pr[\text{race}|\text{surname}]$ and $\Pr[\text{surname}]$ from Census data
2. $\Pr[\text{race}] = \sum_{\text{residence}} \Pr[\text{race}|\text{residence}]\Pr[\text{residence}]$ from Florida Census data

Example (We will use this later!)

$$\Pr[\text{surname}|\text{race}] = \frac{\Pr[\text{race}|\text{surname}]\Pr[\text{surname}]}{\Pr[\text{race}]}$$

We can obtain these terms:

1. $\Pr[\text{race}|\text{surname}]$ and $\Pr[\text{surname}]$ from Census data
2. $\Pr[\text{race}] = \sum_{\text{residence}} \Pr[\text{race}|\text{residence}]\Pr[\text{residence}]$ from Florida Census data

For example,

$$\Pr[\text{Smith}|\text{race=white}] = \frac{\Pr[\text{white}|\text{Smith}]\Pr[\text{Smith}]}{\Pr[\text{race=white}]}$$

A More Complicated Version of Bayes' Rule (Skip)

$$\mathbf{p(y|x)} = \frac{\mathbf{p(x|y)p(y)}}{\mathbf{p(x)}} = \frac{\mathbf{p(x|y)p(y)}}{\sum_{\mathbf{j}} \mathbf{p(x|y_j) \cdot p(y_j)}}$$

A More Complicated Version of Bayes' Rule (Skip)

$$\mathbf{p(y|x)} = \frac{\mathbf{p(x|y)p(y)}}{\mathbf{p(x)}} = \frac{\mathbf{p(x|y)p(y)}}{\sum_{\mathbf{j}} \mathbf{p(x|y_j) \cdot p(y_j)}}$$

$$\mathbf{p(y|x,z)} = \frac{\mathbf{p(x|y,z)p(y|z)}}{\mathbf{p(x|z)}} = \frac{\mathbf{p(x|y,z)p(y|z)}}{\sum_{\mathbf{j}} \mathbf{p(x|y_j,z) \cdot p(y_j|z)}}$$

Bayes' Rule (continuous variable) (Skip)

$$
\begin{aligned}
f(y|x) &= \frac{f(x|y) \cdot f(y)}{f(x)} \\
&= \frac{f(x|y) \cdot f(y)}{\int f(x|y) f(y) dy}
\end{aligned}
$$

More Examples o Applications

1. Managerial Decisions
2. Medical Tests
3. Political Science
4. Natural Language Processing
5. etc..

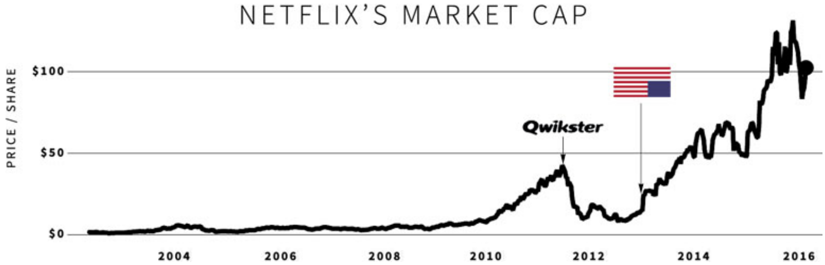
Netflix

FU '16
FRANK UNDERWOOD

the HUSTLE

HBO showed that addicting, original content for a network can boost ratings. Not a huge surprise there. But when Netflix did it with ‘House of Cards’, the total number of paid subscribers dramatically increased. Something clicked and, for a technology company, that’s the type of realization where you start seeing dollar signs.

Netflix made a big bet on the unproven show to the tune of \$ 100m for 26 episodes. That's roughly \$4m an episode or \$77,000 per minute (assuming 50-minute episodes). Everyone was hooked, everyone is watching, and the show is, by all accounts, a smashing success.



Question: Yes, how do we know whether we are going to have a hit?!

Question: Yes, how do we know whether we are going to have a hit?!

Basic Problem:

1. Your company is developing a new product and will be test marketing to better gauge the sales of the new product.
2. Based on positive, neutral or negative reactions, what are the probability of high and low sales?

Example: Suppose you are given the following information

New products introduced in the marketplace have high sales 8% of the time and low sales 92% of the time.

Example: Suppose you are given the following information

New products introduced in the marketplace have high sales 8% of the time and low sales 92% of the time.

A marketing test has the following accuracies:

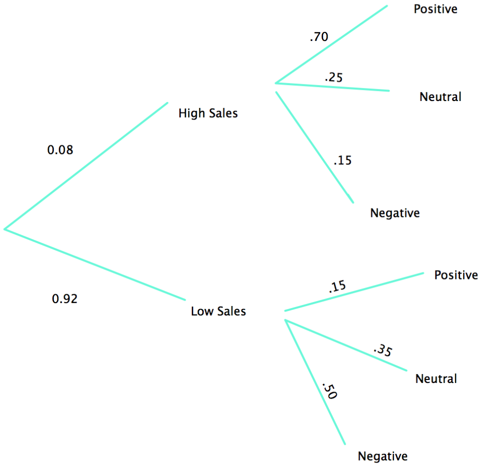
– If sales are high, then consumer test reaction is positive 70%, neutral 25% and negative 5%.

– If sales are low, then consumer test reaction is positive 15%, neutral 35% and negative 50%.

Question: What information do we have?

Question: What information do we have?

$$\Pr[\text{Sales}] \quad \Pr[\text{Reaction} \mid \text{Sales}]$$



Positive
.70
.25
Neutral
High Sales
0.08
.15
Negative
Positive
.15
0.92
Low Sales
.35
Neutral
.50
Negative

Solution:

What should we get?

$$\Pr[\text{high sales}|\text{positve}] = \frac{p(\text{positve}|\text{high}) \cdot p(\text{high})}{p(\text{positive})}$$

Solution:

What should we get?

$$\begin{aligned}\Pr[\text{high sales}|\text{positve}] &= \frac{p(\text{positve}|\text{high}) \cdot p(\text{high})}{p(\text{positive})} \\ &= \frac{0.70 \cdot 0.08}{p(\text{positive})}\end{aligned}$$

What is $p(\text{positive})$?

$$\begin{aligned} p(\text{positive}) &= p(\text{positive}|\text{high}) \cdot p(\text{high}) + p(\text{positive}|\text{low}) \cdot p(\text{low}) \\ &= 0.70 \cdot 0.08 + 0.15 \cdot 0.92 \\ &= 0.194 \end{aligned}$$

We can calculate

$$\text{Pr[high sales|positve]} = \frac{0.70 \cdot 0.08}{0.194} = 0.288$$

Conclusions

Hence, 28.8% you will have high sales in the market given that you receive positive reactions from marketing tests.

Not perfect, but better forecasts relative to our initial probabiity of only 8%!

Computational Statistics: Bayesian Estimation

1. Computer Age Statistical Inference by Efron and Hastie
2. Machine Learning: A Probablistic Perspective

Computational Statistics: Bayesian Estimation

The fundamental unit of statistical inference both for frequentists and for Bayesians is a family of probability densities

$$\mathcal{F} = \{f_\theta(x); x \in \chi, \theta \in \Omega\}$$

where x the observed data is a point in the sample space χ, while the unobserved parameter θ is a point in the parameter space, Ω. The econometrician observes x from $f_\theta(x)$, and infers the value of θ.

Computational Statistics: Bayesian Estimation

The fundamental unit of statistical inference both for frequentists and for Bayesians is a family of probability densities

$$\mathcal{F} = \{f_\theta(x); x \in \chi, \theta \in \Omega\}$$

where x the observed data is a point in the sample space χ, while the unobserved parameter θ is a point in the parameter space, Ω. The econometrician observes x from $f_\theta(x)$, and infers the value of θ.

For example,

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$$

Bayesian Estimation/Inference

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

Bayesian inference requires one crucial assumption in addition to the probability family $\mathcal{F}$, the knowledge of a **prior** density

$$f(\theta), \quad \theta \in \Omega$$

Prior information concerning the parameter θ. For example,

1. It could be known that θ is positive
2. uniform, $f(\theta) = \frac{1}{10} \quad \theta \in [0, 10]$

Notations:

Posterior Distribution

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

In practice, x is observed and hence fixed, while θ varies. We can rewrite this as follows

$$f(\theta|x) = c_x \cdot L_x(\theta) \cdot f(\theta)$$

1. **Likelihood function**: $L_x(\theta)$

Notations:

Posterior Distribution

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

In practice, x is observed and hence fixed, while θ varies. We can rewrite this as follows

$$f(\theta|x) = c_x \cdot L_x(\theta) \cdot f(\theta)$$

1. **Likelihood function**: $L_x(\theta)$
2. **Prior distribution**: $f(\theta)$

Notations:

Posterior Distribution

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

In practice, x is observed and hence fixed, while θ varies. We can rewrite this as follows

$$f(\theta|x) = c_x \cdot L_x(\theta) \cdot f(\theta)$$

1. **Likelihood function**: $L_x(\theta)$
2. **Prior distribution**: $f(\theta)$

(a) **Laplace's "principle of insufficient reason"**: $f(\theta)$ is a uniform distribution

(b) **Uninformative priors**: e.g., Jeffreys' prior $f(\theta) = \frac{1}{1-\theta^2}$ for correlation

(c) **Shrinkage Prior**, one designed to favor smaller values of θ. e.g. $f(\theta) = 1 - |\theta|$

As opposed to frequentist estimation, our **goal** here is to estimate a **entire distribution** for the unknow parameter.

Lets look at one concrete example:

Example: Let's suppose that we would like to know whether or not a coin is biased, $\theta = \Pr[H]$. We observed the following data

$$D = (3H, 9T) = (H, H, H, T, T, T, T, T, T, T, T, T)$$

What is θ?

Example: Let's suppose that we would like to know whether or not a coin is biased, $\theta = \Pr[H]$. We observed the following data

$$D = (3H, 9T) = (H, H, H, T, T, T, T, T, T, T, T, T)$$

What is θ?

Frequentist Estimate: $\theta = \frac{3}{9} = \frac{1}{3}$

Bayesian Estimation: Every number in (0,1) is possible! just with different probabilities.

Bayesian Estimation: Every number in (0,1) is possible! just with different probabilities.

$$\Pr[\theta = 0.0000000001]$$
$$\Pr[\theta = 0.00000000011111]$$
$$\Pr[\theta = 0.00000000011112]$$
$$\vdots$$
$$\Pr[\theta = 0.99999999999999]$$
$$\vdots$$

Bayesian Estimation:

1. I have some prior information about the distribution ($f(\theta)$)
2. Then I observe the data (evidence) $f(x)$ and liklihood that we observe the evidence **given a particular** θ**:** $f(x \mid \theta)$
3. I will update my prior with the new evidence presented to me to form a new view about the distribution of θ

How?

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

How?

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

Note: if my *prior* is that $f(\theta) = 0$ (no chance a value would ever occur), then

$$f(\theta|x) = \frac{f(x|\theta) \cdot 0}{f(x)} = 0$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and
$p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and $p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot \Pr(\theta = .25)}{f(x)}$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and
$p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot \Pr(\theta = .25)}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot \Pr(\theta = .5)}{f(x)}$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and
$p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot \Pr(\theta = .25)}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot \Pr(\theta = .5)}{f(x)}$$

$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot \Pr(\theta = .75)}{f(x)}$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and
$p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$
$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot .25}{f(x)}$$
$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot \Pr(\theta = .5)}{f(x)}$$
$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot \Pr(\theta = .75)}{f(x)}$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and
$p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$
$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot .25}{f(x)}$$
$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot .5}{f(x)}$$
$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot \Pr(\theta = .75)}{f(x)}$$

Step 1: **Prior** $(f(\theta))$\$: we believe that only three parameters are possible $\theta = .25, .5, , .75$ and $p(\theta = .25) = .25, p(\theta = .5) = .5, p(\theta = .75) = .25.$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot .25}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot .5}{f(x)}$$

$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot .25}{f(x)}$$

Our data:

$$D = (3H, 9T) = (H, H, H, T, T, T, T, T, T, T, T, T)$$

Step 2: Likelihood of Evidence for each θ $(f(x \mid \theta))$\$

the probability of coming up with a H is θ, then that of T is $(1 - \theta)$

Our data:

$$D = (3H, 9T) = (H, H, H, T, T, T, T, T, T, T, T, T)$$

Step 2: Likelihood of Evidence for each θ ($f(x \mid \theta)$)$

the probability of coming up with a H is θ, then that of T is $(1 - \theta)$

$$f(x|\theta) = \theta^3(1 - \theta)^9$$

$$\begin{aligned} f(x|\theta = 0.25) &= 0.25^3(1 - .25)^9 = .0011732 \\ f(x|\theta = 0.5) &= 0.5^3(1 - .5)^9 = 0.0002441406 \\ f(x|\theta = 0.75) &= 0.75^3(1 - .75)^9 = 1.609325e - 06 \end{aligned}$$

What is the total probability of oberserving our evidence (data)?

$$D = (3H, 9T) = (H, H, H, T, T, T, T, T, T, T, T, T)$$

Evidence:

$$\begin{aligned} f(x) &= f(x|\theta_1) \cdot p(\theta_1) + f(x|\theta_2) \cdot p(\theta_2) + f(x|\theta_3) \cdot p(\theta_3) \\ &= .00041577 \end{aligned}$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot \Pr(\theta = .25)}{f(x)}$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot \Pr(\theta = .25)}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot \Pr(\theta = .5)}{f(x)}$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot \Pr(\theta = .25)}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot \Pr(\theta = .5)}{f(x)}$$

$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot \Pr(\theta = .75)}{f(x)}$$

$$
\begin{aligned}
f(\theta|x) &= \frac{f(x|\theta) \cdot f(\theta)}{f(x)} \\
\Pr(\theta = .25 \mid x) &= \frac{f(x|\theta = .25) \cdot .25}{f(x)} \\
\Pr(\theta = .5 \mid x) &= \frac{f(x|\theta = .5) \cdot \Pr(\theta = .5)}{f(x)} \\
\Pr(\theta = .75 \mid x) &= \frac{f(x|\theta = .75) \cdot \Pr(\theta = .75)}{f(x)}
\end{aligned}
$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot .25}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot .5}{f(x)}$$

$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot \Pr(\theta = .75)}{f(x)}$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{f(x|\theta = .25) \cdot .25}{f(x)}$$

$$\Pr(\theta = .5 \mid x) = \frac{f(x|\theta = .5) \cdot .5}{f(x)}$$

$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot .25}{f(x)}$$

$$
\begin{aligned}
f(\theta|x) &= \frac{f(x|\theta) \cdot f(\theta)}{f(x)} \\
\Pr(\theta = .25 \mid x) &= \frac{.0011732 \cdot .25}{f(x)} \\
\Pr(\theta = .5 \mid x) &= \frac{f(x|\theta = .5) \cdot .5}{f(x)} \\
\Pr(\theta = .75 \mid x) &= \frac{f(x|\theta = .75) \cdot .25}{f(x)}
\end{aligned}
$$

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{f(x)}$$

$$\Pr(\theta = .25 \mid x) = \frac{.0011732 \cdot .25}{f(x)}$$

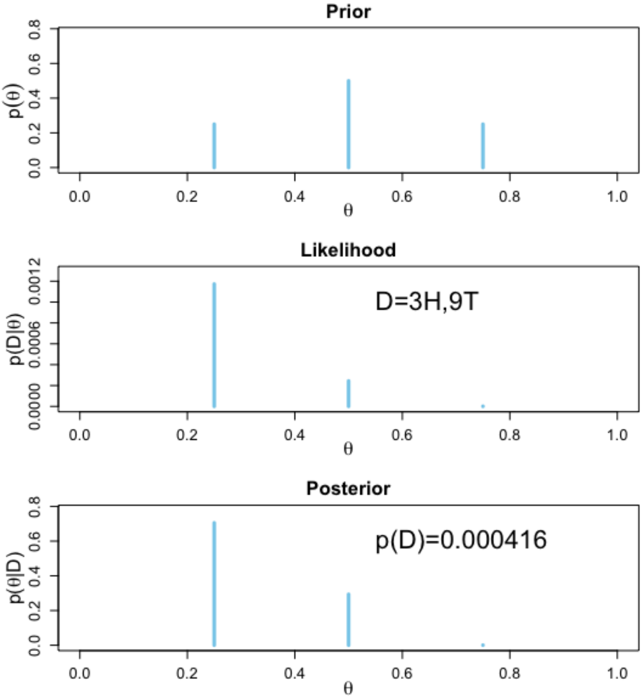
$$\Pr(\theta = .5 \mid x) = \frac{0.0002441406 \cdot .5}{f(x)}$$

$$\Pr(\theta = .75 \mid x) = \frac{f(x|\theta = .75) \cdot .25}{f(x)}$$

$$
\begin{aligned}
f(\theta|x) &= \frac{f(x|\theta)\cdot f(\theta)}{f(x)} \\
\Pr(\theta = .25 \mid x) &= \frac{.0011732 \cdot .25}{f(x)} \\
\Pr(\theta = .5 \mid x) &= \frac{0.0002441406 \cdot .5}{f(x)} \\
\Pr(\theta = .75 \mid x) &= \frac{1.609325e-06 \cdot .25}{f(x)}
\end{aligned}
$$

$$
\begin{aligned}
f(\theta|x) &= \frac{f(x|\theta)\cdot f(\theta)}{f(x)}\\
\Pr(\theta=.25 \mid x) &= \frac{.0011732\cdot .25}{.00041577}\\
\Pr(\theta=.5 \mid x) &= \frac{0.0002441406\cdot .5}{.00041577}\\
\Pr(\theta=.75 \mid x) &= \frac{1.609325e-06\cdot .25}{.00041577}
\end{aligned}
$$

$$
\begin{aligned}
f(\theta|x) &= \frac{f(x|\theta)\cdot f(\theta)}{f(x)} \\
\Pr(\theta = .25 \mid x) &= \frac{.0011732 \cdot .25}{.00041577} = .7054381 \\
\Pr(\theta = .5 \mid x) &= \frac{0.0002441406 \cdot .5}{.00041577} = 0.2936005 \\
\Pr(\theta = .75 \mid x) &= \frac{1.609325e-06 \cdot .25}{.00041577} = 0.0009676777
\end{aligned}
$$



Prior
p(θ)
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood
p(D|θ)
0.0000
0.0006
0.0012
D=3H,9T
θ
Posterior
p(θ|D)
p(D)=0.000416
θ