Conditional Probability

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$$

$$P(X \mid Y)P(Y) = P(X,Y)$$

$$P(X \mid Y)P(Y) = P(Y \mid X)P(X)$$

Bayes Rule

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

$P(cause)$: prior probability of cause
$P(cause \mid effect)$: posterior probability of cause.
$P(effect \mid cause)$: likelihood of effect

Prior ∝ posterior × likelihood

$$P(cause \mid effect) \propto P(effect \mid cause)P(cause)$$

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

$P(h)$: prior probability of hypothesis
$P(h \mid D)$: posterior probability of hypothesis.
$P(D \mid h)$: likelihood of data, given hypothesis

Prior ∝ posterior × likelihood

$$P(h \mid D) \propto P(D \mid h)P(h)$$

$$\begin{aligned} &\operatorname{argmax}_h P(h \mid D) \\ &= \operatorname{argmax}_h \frac{P(D \mid h)P(h)}{P(D)} \\ &= \arg\max_h P(D \mid h)P(h) \end{aligned}$$

$P(h)$: prior probability of hypothesis
$P(h \mid D)$: posterior probability of hypothesis.
$P(D \mid h)$: likelihood of data, given hypothesis

Maximum Likelihood

We assume a uniform prior $P(h)$.
We then choose the hypothesis that assigns the highest likelihood to the data

$$h_{ML} = argmax_h P(D|h)$$

$$P(X \mid \mathbf{D}) = P(X \mid h_{ML})$$

This is commonly used in machine learning.

Bayes Nets

A Bayes Net defines a joint distribution $P(X_1 \ldots X_n)$ over a set of random variables $X_1 \ldots X_n$

Using the chain rule, we can factor $P(X_1 \ldots X_n)$ into a product of n conditional distributions:

$$P(X_1 \ldots X_n) = \prod_j P(X_i \mid X_1 \ldots X_{i-1}).$$

A Bayes Net makes a number of (conditional) independence assumptions:

$$P(X_1 \ldots X_n) =_{def} \prod_j P(X_i \mid \text{Parents}(X_i) \subseteq \{X_{1\ldots} X_{i-1}\})$$

Learning Bayes Nets

Parameter estimation: Given some data D over a set of random variables X and a Bayes Net (with empty conditional probability tables or CPTs) estimate the parameters (= fill in the CPTs) of the Bayes Net.

Structure learning: Given some data D over a set of random variables X, find a Bayes Net (define its CPTs) and estimate its parameters.

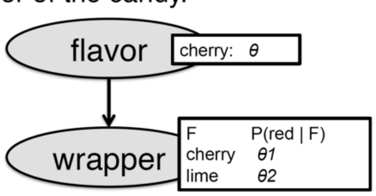
Let's look at an example

Given data ***D***, we want to find the parameters that maximize $P(\boldsymbol{D} \mid \theta)$.

We have a data set with N candies.
c are cherry. $l = (N-c)$, are lime.
Parameter θ = probability of cherry

Maximum likelihood estimate: $\theta = c/N$

Now the candy has two kinds of wrappers (red or green).
The wrapper is chosen probabilistically, depending on the flavor of the candy.



Out of N candies, c are cherry. r_c are cherry with a red wrapper, r_l are lime with a red wrapper

The likelihood of this data set:
$$P(d \mid \theta, \theta_1, \theta_2) = \theta^{c}(1-\theta)^{N-c}\,\theta_1^{\,rc}\,(1-\theta_1)^{c-rc}\,\theta_2^{\,rl}\,(1-\theta_1)^{(N-c)-rl}$$

The log likelihood of this data set:
$$\begin{aligned} L(d \mid \theta, \theta_1, \theta_2) = & [c \log\theta + (N-c)\log(1-\theta)] \\ & +[r_c \log\theta_1 + (c-r_c)\log(1-\theta_1)] \\ & +[l_c \log\theta_2 + (N-c-l_c)\log(1-\theta_2)] \end{aligned}$$

The ML parameter estimates:
$\theta = c/N \quad \theta_1 = r_c/c \quad \theta_2 = r_l/(N-c)$