3. Bayes' Theorem

3.1 Bayes' Theorem

A straightforward application of conditioning: using $f(x,y) = f(x|y)\,f(y) = f(y|x)\,f(x)$, we obtain Bayes' theorem (also called Bayes' rule)

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

- It applies to discrete and continuous random variables, and a mix of the two.
- Deceptively simple, and the predictions are counterintuitive. *This is a very powerful combination.*

3.2 Examples of Bayes' Theorem

Let's do some "simple" examples, adapted from the book **The Drunkard's Walk – How Randomness Rules Our Lives** by Leonard Mlodinow. *(The solutions to these problems are given at the end of the chapter (Sec. 3.4); they will be made available after the class. One of the purposes of this lecture is to demonstrate how counterintuitive probability can be.)*

3.2.1 Girls named Lulu

Case 1: A family has two children. What is the probability that both are girls?

Case 2: A family has two children. Given that one of them is a girl, what is the probability that both are girls?

Case 3: A family has two children. Given that one of them is a girl named Lulu, what is the probability that both are girls?

3.2.2 Disease diagnosis

Doctors were asked to estimate the probability that a woman with no symptoms, between 40 and 50 years old, who has a positive mammogram (i.e. the test reports cancer), actually has breast cancer. The doctors were given the following facts:

The facts

- 7% of mammograms give false positives.
- 10% of mammograms give false negatives.
- Actual incidence of breast cancer in her age group is 0.8%.

Results

- Germany: A third of doctors said 90%, the median estimate was 70%.
- USA: 95% said the probability was approximately 75%.

What do you think? Aren't you tempted to say 93%?

Date compiled: March 12, 2013

3.2.3 Prosecutor's fallacy

The facts

- Couple's first child died of SIDS (Sudden Infant Death Syndrome) at 11 weeks.
- They had another child, who also died of SIDS at 8 weeks.
- They were arrested and accused of smothering their children, even though there was no physical evidence.

Results

- The prosecutor brought in an expert, who made a simple argument: the odds of SIDS are 1 in 8,543. Therefore, for two children: 1 in 73 million.
- The couple was convicted.

Do you agree with the conviction?

3.3 Combining Priors with Observations

Bayes' theorem:

$$f(x|z) = \frac{f(z|x)\,f(x)}{f(z)}$$

x: unknown quantity of interest (in this class, usually the system *state*).

$f(x)$: prior belief of state.

z: observation related to state (usually a sensor measurement).

$f(z|x)$: observation model: for a given state, what is the probability of observing z?

$f(x|z)$: posterior belief of state, taking observation into account.

$f(z) = \sum_x f(z|x)\,f(x)$, by the total probability theorem (here, for discrete random variables): probability of observation, essentially a normalization constant (does not depend on x).

- This is a systematic way of combining prior beliefs with observations.
- Question: If we observe the state x through z, doesn't that tell us the state directly? No!
  1. Dimension: $\dim(z)$ usually smaller than $\dim(x)$.
  2. Noise: $f(z|x)$ is not necessarily "sharp."

Generalization to Multiple Observations

- We have N observations, each of which can be vector valued: $z_1, \ldots, z_N$.
- Often, we can assume conditional independence:

  $$f(z_1, \ldots, z_N|x) = f(z_1|x) \cdots f(z_N|x)$$

  One interpretation: each measurement is a function of the state x corrupted by noise, and the noise is independent. Example: $z_i = g_i(x, w_i)$; $f(w_1, \ldots, w_N) = f(w_1) \cdots f(w_N)$.
- Then:

$$\underbrace{f(x|z_1, \ldots, z_N)}_{\text{posterior}} = \frac{\overbrace{f(x)}^{\text{prior}} \quad \overbrace{\prod_i f(z_i|x)}^{\text{observation likelihood}}}{\underbrace{f(z_1, \ldots, z_N)}_{\text{normalization}}}$$

- Normalization:

$$f(z_1,\dots,z_N)=\sum_{x\in\mathcal{X}} f(x)\prod_i f(z_i|x)$$

by the total probability theorem (here, for discrete random variables).

Example

- Let $x\in\{0,1\}$ represent the truthful answer to a question, with 0 corresponding to a "No", and 1 to a "Yes". We ask two people the same question, and then we estimate what the truth is. Let z_i be the answer given by person i (an observation), modeled as follows:

$$z_i = x + w_i,\ i=1,2;\ \text{where } w_1 \text{ and } w_2 \text{ are independent; and } w_i=\begin{cases}0: & \text{truth}\\ 1: & \text{lie}\end{cases}$$

and the '+'-operator is defined by

$$\begin{aligned} +:\quad & 0+0=0\\ & 0+1=1\\ & 1+0=1\\ & 1+1=0.\end{aligned}$$

- For the prior, let $f(x)=\frac{1}{2}$ for $x=0,1$. This represents the state of maximum ignorance, where we take all states to be a priori equally likely. We model the truthfulness of person i's answer as:

$$w_i:\ f_{w_i}(0)=p_i,\ f_{w_i}(1)=1-p_i.$$

That is, person i tells the truth with probability p_i.

- By Bayes' rule:

$$f(x|z_1,z_2)=\frac{f(x)\,f(z_1|x)\,f(z_2|x)}{f(z_1,z_2)}.$$

We build tables for $f(x)\,f(z_1|x)\,f(z_2|x)$ and $f(z_1,z_2)$:

x	z_1	z_2	$f(x)\,f(z_1\|x)\,f(z_2\|x)$
0	0	0	$0.5p_1p_2$
0	0	1	$0.5p_1(1-p_2)$
0	1	0	$0.5(1-p_1)p_2$
0	1	1	$0.5(1-p_1)(1-p_2)$
1	0	0	$0.5(1-p_1)(1-p_2)$
1	0	1	$0.5(1-p_1)p_2$
1	1	0	$0.5p_1(1-p_2)$
1	1	1	$0.5p_1p_2$

z_1	z_2	$f(z_1,z_2)$
0	0	$0.5\,(p_1p_2+(1-p_1)\,(1-p_2))$
0	1	$0.5\,(p_1\,(1-p_2)+(1-p_1)\,p_2)$
1	0	$0.5\,((1-p_1)\,p_2+p_1\,(1-p_2))$
1	1	$0.5\,((1-p_1)\,(1-p_2)+p_1p_2)$

- Then

$$f_{x|z_1z_2}(0|0,0)=\frac{p_1p_2}{p_1p_2+(1-p_1)(1-p_2)},\ f_{x|z_1z_2}(1|0,0)=\frac{(1-p_1)(1-p_2)}{p_1p_2+(1-p_1)(1-p_2)},\ \text{etc.}$$

are the probabilities of the true answer being "No" when both persons say "No," the true answer being "Yes" when both persons say "No," etc.

- Special case: $p_1=\frac{1}{2}$. Our intuition is that the answer z_1 is not useful for determining the truth, and this turns out to be correct:
  - $f_{x|z_1z_2}(0|0,0)=p_2$
  - $f_{x|z_1z_2}(0|1,0)=p_2$, etc.