EL2320 - Baysian Inference

John Folkesson

Baysian Inference
(Chap 3.1-3.2 in Thrun)

Maximum Likelihood Estimate=MLE.

- We make many measurements, $\{z_i\}$, from some unknown distribution.
- We make an assumption on the form of that distribution (e.g. its Gaussian).
- We write an expression for the probability of our measurements as a function of the unknown distribution parameters λ, (e.g for Gaussians λ are μ and σ^2). $P(\{z_i\}|\lambda)$
- We find the λ that maximizes this so called likelihood. $\lambda_{MLE} = \arg\max_\lambda P(\{z_i\}|\lambda)$

Paramenter Estimation - Maximum Likelihood Estimate

Your thermometer measurement τ is modeled:

$\tau \sim N(T, 1)$ where T is the 'true' temperature.
(That is it has a pdf that goes like $e^{-\frac{(\tau - T)^2}{2}}$)

You make a measurement of $\tau = 20$.

Use Bayes rule to deduce (in writing) the MLE for T.

You make 2 independent measurements of 20 and 19.

What is the MLE for T?

MLE and MAP

MLE = Maximum Likelihood Estimate
MAP = Maximum a Posteriori Estimate

Very similar with a minor difference.

Both estimate parameters using a distribution.

Both use the point where a distribution is highest as the estimate.

$$\lambda_{MLE} = \arg\max_{\lambda} p(\{z_i\}|\lambda).$$
$$\lambda_{MAP} = \arg\max_{\lambda} p(\lambda|\{z_i\}).$$

MLE and MAP

MAP uses some prior knowledge of the distribution, $p(\lambda) =$some distribution over *lambda*.

$\lambda_{MAP} = \arg\max_{\lambda} p(\lambda|\{z_i\})$.

$\lambda_{MAP} = \arg\max_{\lambda} \frac{p(\{z_i\}|\lambda)p(\lambda)}{p(\{z_i\})}$.

$\lambda_{MAP} = \arg\max_{\lambda} p(\{z_i\}|\lambda)p(\lambda)$.

MLE and MAP are the same if there is no a priori knowledge of the distribution. It starts with an uniform distribution, $p(x) =$constant.

MLE and MAP Are not invariant to coordinate changes

Both estimates can change if we make non-linear transformations $y = f(x)$:

$p(x)dx = \tilde{p}(y)dy =>$

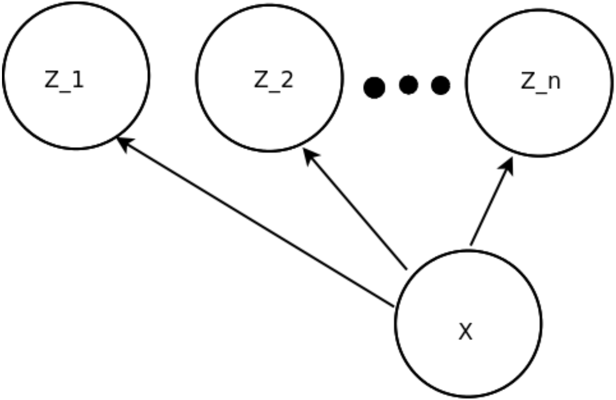
$p(x) = \tilde{p}(f(x))\frac{dy}{dx} = \tilde{p}(f(x))f'(x) =>$

$\frac{p(x)}{f'(x)} = \tilde{p}(y) =>$

$\arg\max_y \tilde{p}(y)$ does not need to be $f(m)$.

On the other hand, we often find it easier to maximize $\ln p(x)$, the so called log likelihood. The x that maximizes the log-likelihood is the same as the x that maximizes the likelihood (ie $p(\lambda)$).

Bayes Inference



$$p(x|z_1, \ldots, z_n) \propto \frac{p(z_n|x)p(x|z_1,\ldots,z_{n-1})}{p(z_n|z_1,\ldots z_{n-1})}$$

Generalize to iterative filtering of independent measurements z_i of quantity x:

$$p(z_1, \ldots, z_n|x) = p(z_1|x) \ldots p(z_n|x)$$

Apply Bayes Thereom:

$$p(x|z_1, \ldots, z_n) = \frac{p(z_n|x,z_1,\ldots,z_{n-1})p(x,z_1,\ldots,z_{n-1})}{p(z_n,z_1,\ldots z_{n-1})}$$

$$= \frac{p(z_n|x,z_1,\ldots,z_{n-1})p(x|z_1,\ldots,z_{n-1})p(z_1,\ldots,z_{n-1})}{p(z_n|z_1,\ldots z_{n-1})p(z_1,\ldots,z_{n-1})}$$

$$\propto p(z_n|x)p(x|z_1, \ldots, z_{n-1})$$

Baysian Inference

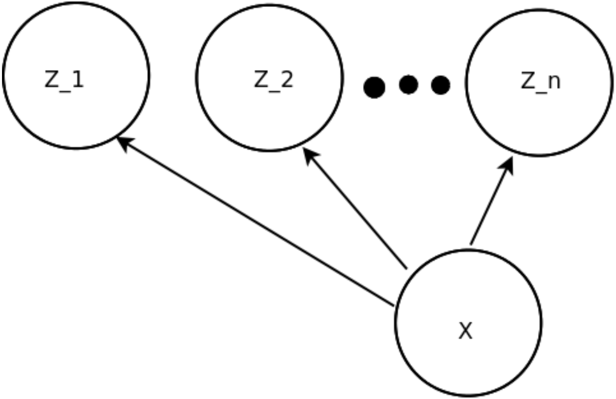
$p(x|z_1,\ldots,z_n) \propto p(z_n|x)p(x|z_1,\ldots,z_{n-1})$

Which is the basic idea of Bayesian inference from a series of measurements. (Step 4 of Table 2.1 Bayes filter).

We call $p(z_n|x)$ a measurement model.
We call $p(x|z_1,\ldots,z_n)$ the a posteriori distribution.

This is also notated as: $bel(x_t) = p(x|z_1,\ldots,z_n)$ the 'belief'.

Bayes Inference



$$p(x|z_1,\ldots,z_n) \propto p(z_n|x)p(x|z_1,\ldots,z_{n-1})$$

Baysian Inference

Psychologist use a test to ID seriously ill people:

- Person sits in front of desk. Doctor sits behind desk.
- Two jars on the desk, jar X with 80% red balls and 20% blue. And jar Y with 80% blue and 20% red.
- The doctor then places the jars on the floor and tells the person that one of the jars will be drawn from at random.
- The person is to stop the process when he or she feels confident in which jar is being drawn from.
- The sequence is in fact fixed with RBRBRBBRBB or something similar so a resonable person would not be able to judge for at least 7-8 balls.
- The deranged will "know" after one ball.

What is (as far as the test subject knows) probability of it being jar X given a read ball was drawn, $P(X|R) = ?$ $P(Y|R) = ?$

What about $P(X|B,R) = ?$ (Watch out for the denominator (use normalization))