Summary of Intelligent Robotics class of 18/09/12

성기호, 안정도, Peter Muschick, 김유진

1 Beta distribution

In probability theory and statistics, the beta distribution is a family of continuous probability distributions defined on the interval [0, 1] parametrized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution. [1]

The probability density function of beta distribution $\theta \sim Beta\ (\alpha_h, \alpha_t)$:

$$P(\theta) = \frac{\Gamma\ (\alpha_h + \alpha_t)}{\Gamma\ (\alpha_h) * \Gamma\ (\alpha_t)} \theta^{\alpha_{h-1}} (1-\theta)^{\alpha_{t-1}}$$

$$P(x[1] = x^h) = \int_\theta P(x[1] = x^h \mid \theta)\ P(\theta)\ d\theta$$

$$= \frac{\alpha_h}{\alpha_h + \alpha_t}$$

$$P(x[2] = x^h \mid x[1] = x^h) = \frac{\alpha_h + 1}{\alpha_h + 1 + \alpha_t}$$

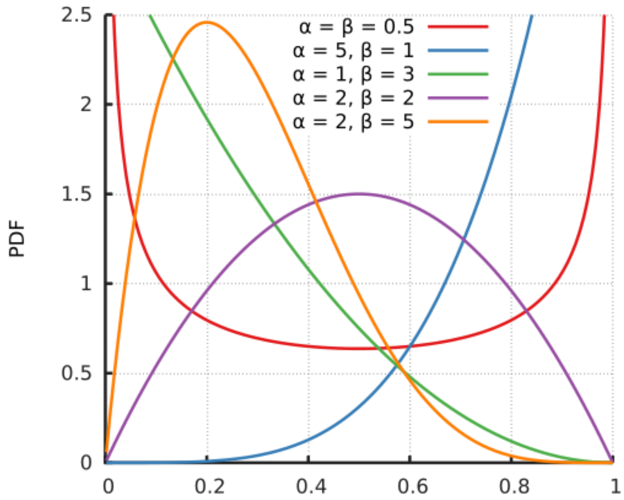


Figure 1: Showing different Probability density function [2]

In general:

With data M^h, M^t

Prior is $\beta\ (\alpha_h, \alpha_t)$

Posterior is $\beta(\alpha_h + M_h, \alpha_t + M_t)$ conjugate family for binomial

$$P(x[M+1] = x^h \mid D) = \frac{\alpha_h + M_h}{\alpha_h + \alpha_t + M_h + M_t}$$

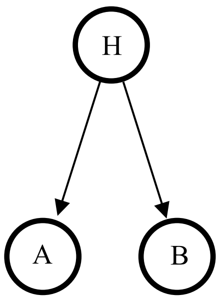
$$= \frac{M'}{M + M'}\,\theta'_h + \frac{M}{M + M'} * \frac{M^h}{M}$$

Where

$$M = M_h + M_t$$

$$M' = \alpha_h + \alpha_t$$

$$\theta'_h = \frac{\alpha_h}{\alpha_h + \alpha_t}$$



$P(H = 1)$ D: A = 1, B = 1, 6 times

$P(A \mid H)$ A = 1, B = 0, 4 times

$P(B \mid H)$

Arg max $P(D \mid \theta) = \sum_n P(D_1 H = h \mid \theta)$

Maximize expected log-likelihood relative to probabilistic distribution over hidden variables.

Arg max $Arg\ max_\theta\ E\ l\left(\theta_j D_i H\right)$

$P(H \mid D_i \theta)$

Expectation Maximization Algorithm:

Pick initial θ_0;

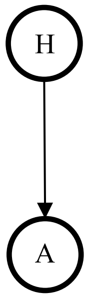
Loop until θ doesn't change much;

$E_{i+1}(\theta) = \sum_n P(H = h \mid D_i \theta_i)\ l\ (\theta_i D_i H)\ Expectation;$

$\theta_{i+j} = argmax_\theta E_{i+1}(\theta)\ maximization;$

Maximization expected log-likelihood is equivalent to maximize MLE in a complete data set where counts are replaced by expected counts.

Example:



$$\left.\begin{array}{l} P(H=1) \\ P(A=1 \mid H=0) \\ P(A=1 \mid H=1) \end{array}\right\} \theta$$

$\theta: P(H=1) = 0.4$

H	$P\,(A=1 \mid H)$
0	0.25
1	0.54

Data D:

A = 0; M (A = 0)

A = 1; M (A = 1)

$P(H \mid A=0) = \frac{P(A \mid H)\,P(H)}{P(A)} = a$

$P(H=1 \mid A=1) = b$

	A	H	Expected counts	
M (A = 0)	0	0	(1 – a) M (A = 0)	v
	0	1	a M (A = 0)	w
M (A = 1)	1	0	(1 – b) M (A = 1)	x
	1	1	b M (A = 1)	y

$\theta: P(H=1) = \frac{w+y}{v+w+x+y}$

$P(A=1 \mid H=0) = \frac{x}{v+x}$

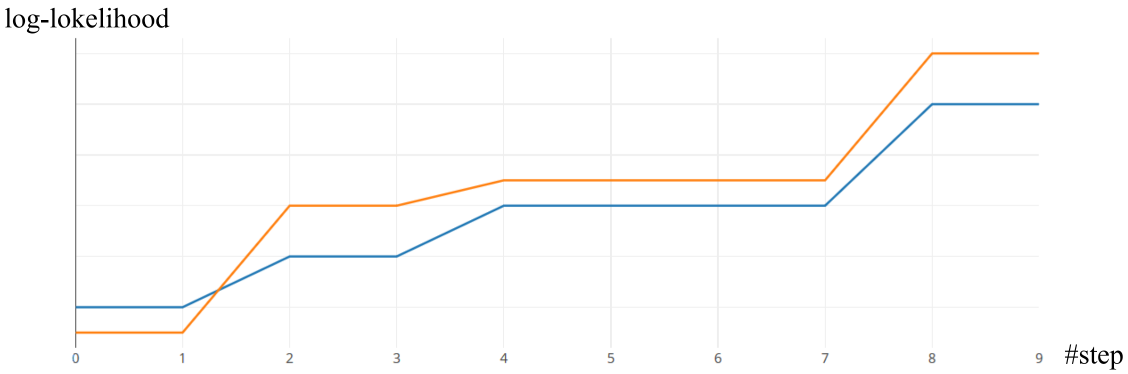


Figure 2: Showing general profiles of expectation maximization

The start of the orange or blue line of Fig. 2 is determined by the initial guess. Since the figure is showing general profiles, every try can look different. What all tries have in common is a monotonic improvement meaning plateaus alternately with more or less steeper parts. If a graph doesn’t fulfil these conditions and/or is even decreasing at one point its certain that an algorithm is calculating no useful results.

$D = \{y_1, \ldots, y_m\}$ want to divide them into 2 groups
Group 1 $\longrightarrow N\ (M_1 \theta_1^2)$
Group 2 $\longrightarrow N\ (M_2 \theta_2^2)$
P (group 1 is taken on) = π
Problem: we have to find parameter values which we don’t know
Hidden variables:

$$H = \{H_1, \ldots, H_m\}$$

2 References

[1] https://en.wikipedia.org/wiki/Beta_distribution
[2] https://en.wikipedia.org/wiki/File:Beta_distribution_pdf.svg