Ming Hsieh Department of Electrical Engineering

USC Viterbi School of Engineering

Bayesian Inference with Adaptive Fuzzy Priors and Likelihoods

Osonde Osoba, Sanya Mitaim, Bart Kosko

Motivation

❑ Bayesian Inference uses Bayes theorem to update prior knowledge about a parameter θ of observed data x.

$$f(\theta|x) = \frac{g(x|\theta)h(\theta)}{\int g(x|u)h(u)\,du} \propto g(x|\theta)h(\theta)$$

❑ The posterior pdf f(θ|x) gives all the probabilistic information about the parameter given the available evidence x.

❑ The prior pdf h(θ) describes the unknown parameter. It can inject "subjective" information into the inference process.

❑ The key problem is most applications use unsuitable closed-form priors either for ease of computation or for convenience.

Questions:
1. Is there a way to approximate arbitrary priors and likelihood functions?
2. Do such approximations produce good posterior pdf approximations?

Theory

❑ The Fuzzy Approximation Theorem (FAT) states that fuzzy systems can uniformly approximate arbitrary priors and likelihood functions under the standard additive model:

$$H(\theta) \;=\; \frac{\sum_{j=1}^{m} a_j(\theta)V_j c_j}{\sum_{j=1}^{m} a_j(\theta)V_j} \;\approx\; h(\theta)$$

❑ The new Bayesian Approximation Theorem[1] (BAT) guarantees that uniform approximators for the prior (H) and likelihood function (G) lead to a uniform posterior pdf approximation (F) :

$$F \xrightarrow{u} f$$

$$\text{when } H \xrightarrow{u} h, \quad G \xrightarrow{u} g, \text{and} \quad F = \frac{G(x|\theta)H(\theta)}{\int_\theta G(x|\theta)H(\theta)}$$

Uniform Fuzzy Approximation

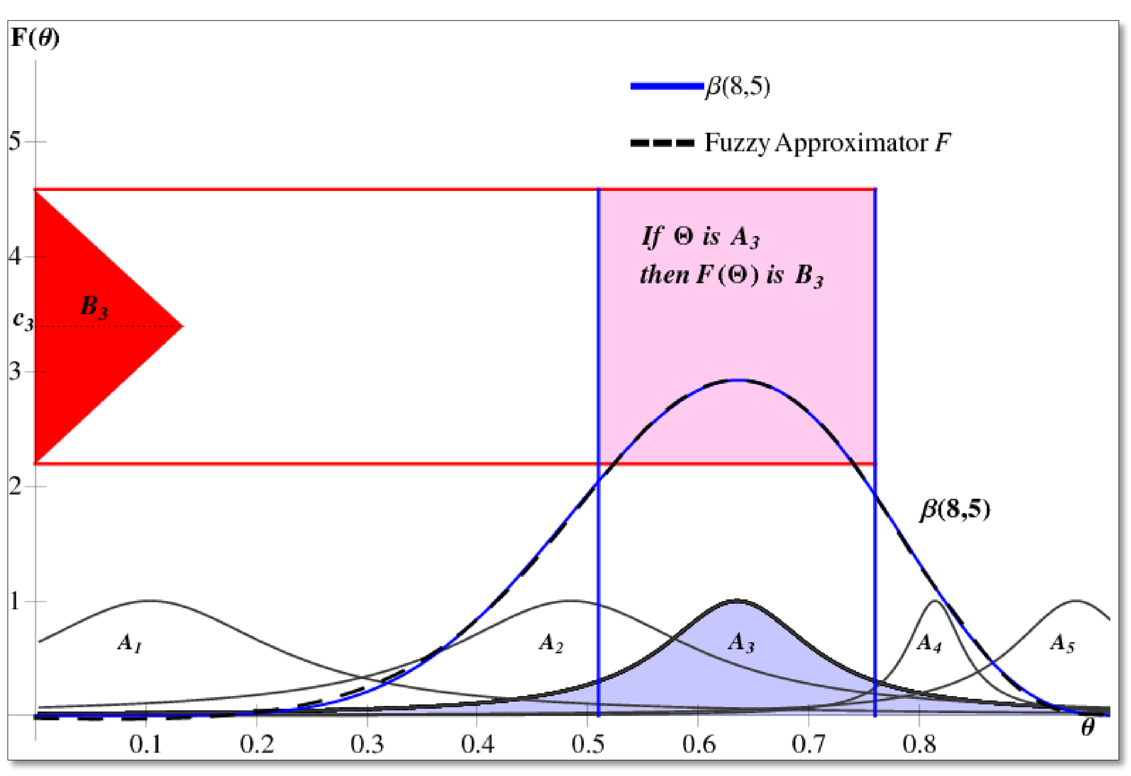

Figure 1: Five fuzzy if-then rules approximate the beta prior h(θ)=β(8, 5). The five if-part fuzzy sets are truncated Cauchy bell curves. An adaptive Cauchy SAM fuzzy system tuned the sets' location and dispersion parameters to give a nearly exact approximation of the beta prior. Each fuzzy rule defines a patch or 3-D surface in the input-output state space. The third rule has the form "If Θ is in A_3 then B_3" where then-part set B_3 is a fuzzy number centered at centroid c_3. This rule might have the linguistic form "If Θ is larger than 0.5 then F(θ) is large." The adaptive system tuned the centroids and areas of all five then-part sets (not pictured).

Posterior Approximation

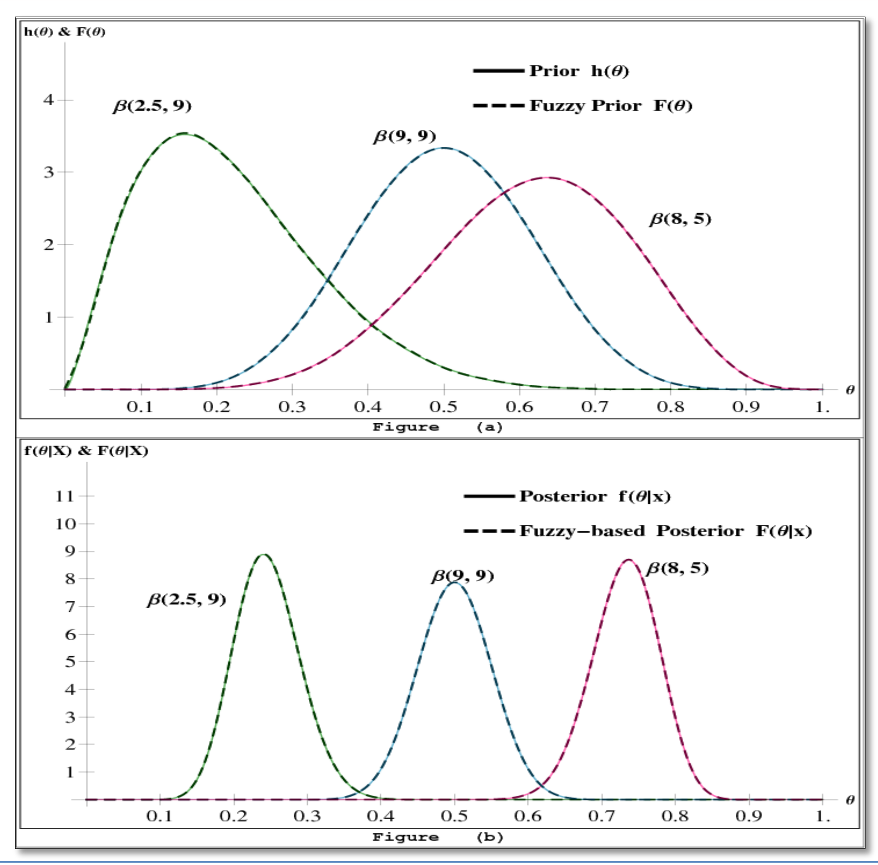

Figure 2: Comparison of conjugate beta priors and posteriors with their fuzzy approximators. (a) an adapted sinc-SAM fuzzy system F(θ) with 15 rules approximates the three conjugate beta priors h(θ): β(2.5,9), β(9, 9), and β(8, 5). (b) the sinc-SAM fuzzy priors F(θ) in (a) produce the SAM-based approximators F(θ|x) of the three corresponding beta posteriors f(θ|x) for the three corresponding binomial likelihood pdfs g(x|θ) with n = 80: b(20, 80), b(40, 80).

Extension

❑ The BAT also extends[2] to hierarchical Bayesian applications. So when the model includes a hyperprior, π:

$$f(\theta,\tau|x) = \frac{g(x|\theta)h(\theta|\tau)\pi(\tau)}{\int_{\theta,\tau} g(x|\theta)h(\theta|\tau)\pi(\tau)}$$

then:

$$F(\theta,\tau|x) \xrightarrow{u} f(\theta,\tau|x)$$

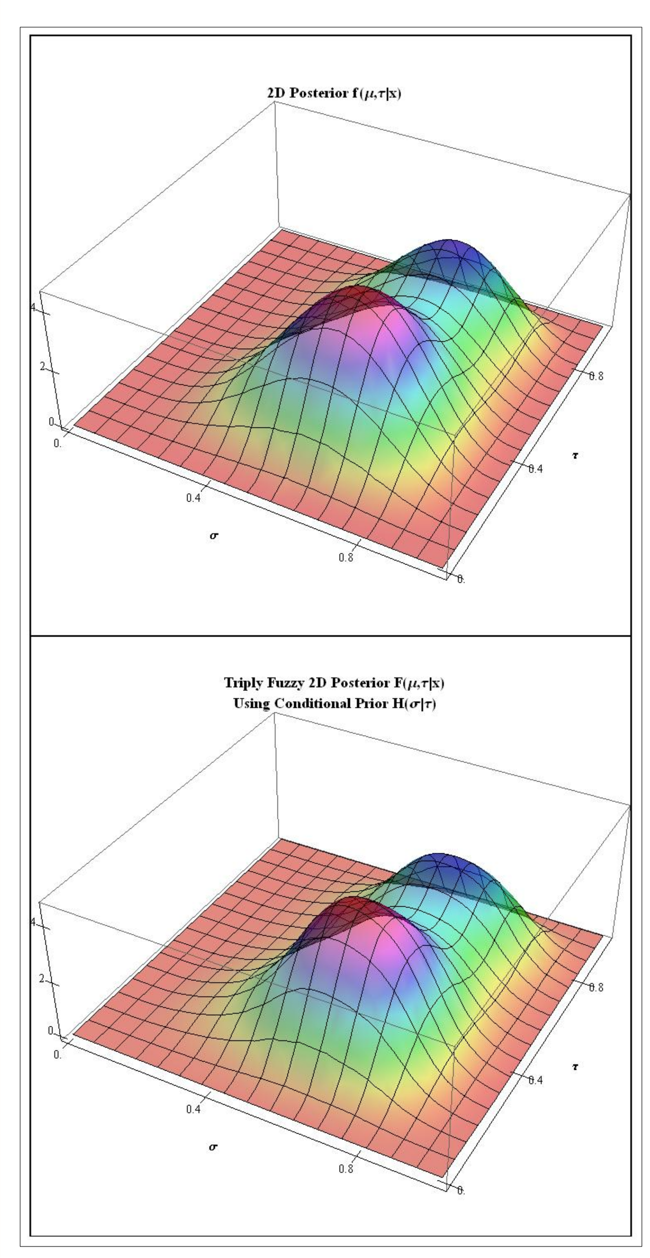

Figure 3: Triply fuzzy Bayesian inference: comparison of a 2-D non-conjugate posterior and its triply fuzzy approximator F. The first panel shows the approximand f. The second panel shows a triply fuzzy approximator F that used a fuzzy approximators for the likelihood, prior, and hyperprior.

Discussion and Conclusion

❑ Fuzzy systems allow users to encode prior information through fuzzy rules or training data rather than through the choice of a handful of closed-form probability densities.

❑ Watkins Representation Theorem[3] guarantees that such fuzzy systems can exactly represent any closed-form bounded function. Thus this approximation scheme subsumes all previous closed-form Bayesian inference schemes.

❑ Figure 4 shows how fuzzy systems can also perform robust probability density estimation[1].

❑ The Bayesian Approximation Theorem (BAT) guarantees that posterior approximations under this scheme are uniform. This guarantee also applies if we replace fuzzy approximators with any other uniform approximator (like neural networks, Bernstein polynomials etc.).

❑ The BAT extends to hierarchical Bayesian models.

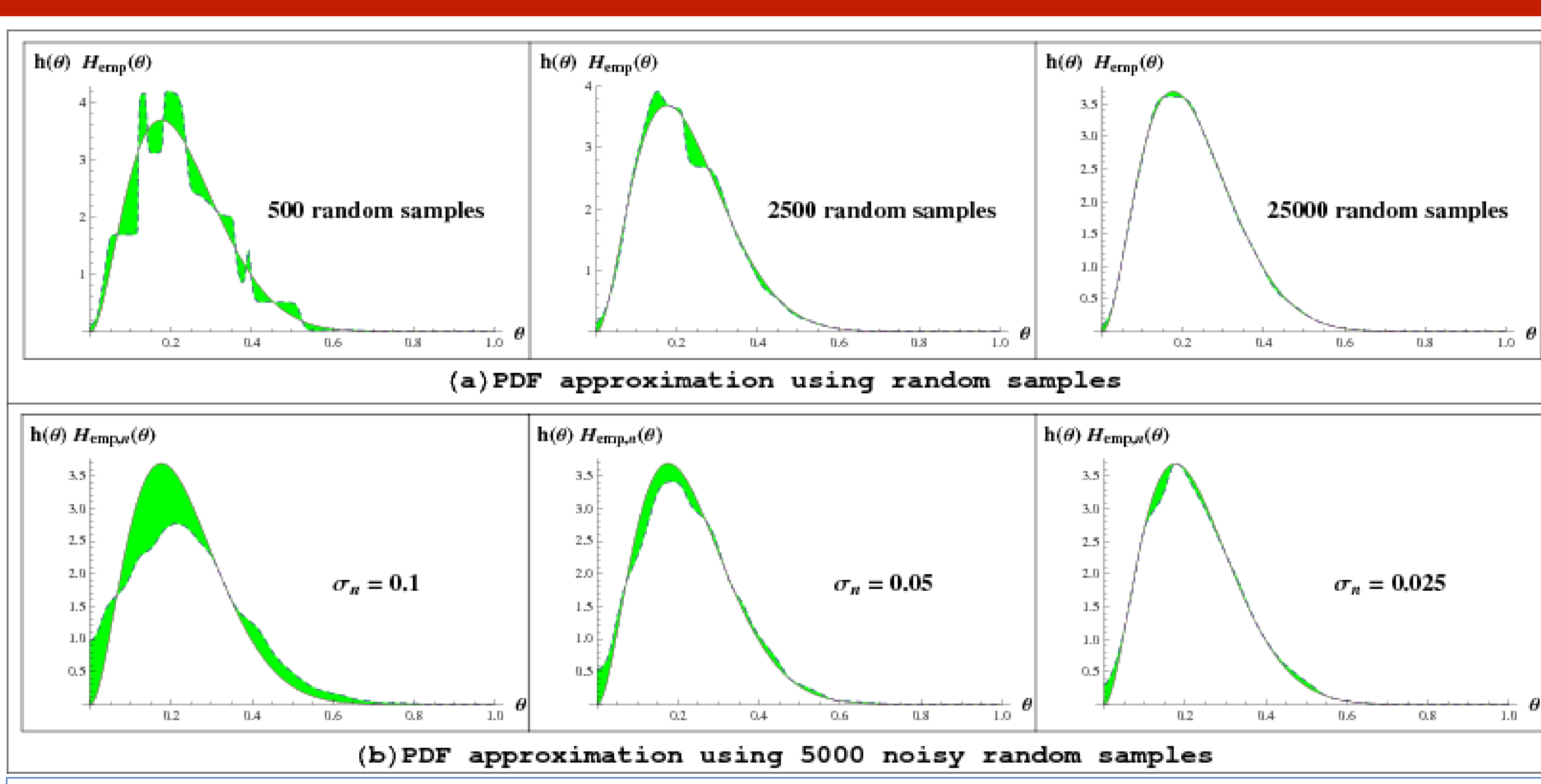

Figure 4: ASAMs can use a limited number of random samples or noisy random samples to estimate the sampling pdf. The ASAMs approximate empirical pdfs from random samples. The shaded regions represent the approximation error between the ASAM estimate and the sampling pdf. Part (a) compares the β(3, 10.4) pdf with ASAM approximations for three β(3, 10.4) empirical pdfs. The figure shows comparisons for different sample sizes N. Part (b) compares the β(3, 10.4) pdf with ASAM approximations of three β(3, 10.4) random samples corrupted by independent noise. The plots show that the ASAM estimate gets better as the number of samples increases and as noise power reduces.

References

[1] O. Osoba, S. Mitaim, and B. Kosko, "Bayesian Inference with Adaptive Fuzzy Priors and Likelihoods", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 41, no. 5, pp. 1183-1197, 2011

[2] O. Osoba, S. Mitaim, B. Kosko, "Triply Fuzzy Function Approximation for Bayesian Inference," International Joint Conference on Neural Networks, pp.3105-3111, August 2011

[3] F. A. Watkins, "The representation problem for additive fuzzy systems," Proc. IEEE Int. Conf. Fuzzy Syst. (IEEE FUZZ), Mar. 1995, vol. 1, pp. 117–122.

[4] B. Kosko, "Fuzzy Engineering", Prentice Hall, 1996.

Contact: kosko@usc.edu

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering