

In the paper, *Bayesian Approach to the Choice of Smoothing Parameter in Kernel Density Estimation*, authors Gangopadhyay and Cheung investigate the bandwidth selection problem for variable bandwidth kernels. The paper first discusses work done in global bandwidth selection procedures such as Least Squares Cross Validation (LSCV) and improvements with Biased Cross Validation (BCV) techniques using the leave one kernel density estimate. They note, however, that choosing a bandwidth globally fundamentally contradicts the essence of nonparametric smoothing. If the data is collected at a mode of the underlying distribution, the amount of smoothing needed will be less than at points far from the modes. This motivates variable bandwidth procedures which was introduced by Fan et al. 1996. The authors note that this procedure does not perform well in moderate sample sizes. In their paper, Gangopadhyay and Cheung attempt to address this issue by introducing a prior distribution of the bandwidth h that will replace any missing information that the sample fails to capture.

Following the motivation above, the authors try to derive the posterior distribution of h , but seeing that h is not a model parameter, they first devise a way to introduce h as a parameter to the model in question. They achieve this by introducing *truncated version of* $f(x)$ defined as

$$f_h(x) = \int f(x)K_h(x - u)du$$

h therefore acts as a scale parameter to f . That is as h decreases, $f_h(x) \rightarrow f(x)$. Having introduced h as a model parameter, it is now possible to discuss the Bayesian approach to the bandwidth selection procedure. Letting $h \sim \pi(h)$ be the prior distribution of h then the posterior is given by

$$\pi(h|x) = \frac{f_h(x)\pi(h)}{\int f_h(x)\pi(h)dh}$$

We see that $\pi(h|x)$ relies on $f_h(x)$. Therefore, a plug in estimate of f is needed. The authors

cleverly notice that $f_h(x) = \mathbb{E}(K_h(x - X))$ so it is reasonable to estimate this function by

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)$$

which is just the kernel estimator of f . So the approximated posterior distribution is given by

$$\hat{\pi}(h|X_1, \dots, X_n, x) = \frac{\hat{f}_h(x)\pi(h)}{\int \hat{f}_h(x)\pi(h)dh}$$

This then gives rise to the posterior mean, and hence the estimate of h ,

$$h^* = \int h\hat{\pi}(h|X_1, \dots, X_n, x)dh$$

The authors note that this procedure is quite computationally intensive in that none of the integrals need have a closed form solution. Instead of addressing these computational issues, the authors focus on a case where a closed form solution *does* exist. In the case where K is a normal kernel, then h serves as a standard deviation of $K_h(X_i - x)$. Under this assumption, h has conjugate prior given by the Inverse Gamma distribution.

$$\tau(h) = \frac{2}{\Gamma(\alpha)\beta^\alpha} \frac{1}{h^{\alpha+1}} \exp\left\{-\frac{1}{\beta h^2}\right\} \quad h > 0$$

Using this nice case, they authors successfully show that

$$h^*(x) = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha + 1/2)} \frac{\sum_{i=1}^n [1/(\beta(X_i - x)^2 + 2)]^\alpha}{\sum_{i=1}^n [1/(\beta(X_i - x)^2 + 2)]^{\alpha+1/2}}$$

which gives a close form solution to the estimated mean of the posterior distribution of h . They go on to point out that the posterior does in fact depend on parameters of the prior, so a sensitivity analysis of the posterior's behavior for different (α, β) combinations is necessary. Moreover, they note that by deriving $\hat{\pi}(h|X_1, \dots, X_n, x)$ a more robust sense of appropriate values of h can be achieved.

To assess the effectiveness of their method, Gangopadhyay and Cheung devise a simulation experiment in which they measure the goodness of fit of global bandwidth kernels selected by *BCV* and *LSCV* as well as the smoothing method based on *K-NN* against their newly proposed local bandwidth selection method. They measure the effectiveness of each method by the average squared error

$$ASE = \frac{1}{m} \sum_{i=1}^m (\hat{f}(v_i) - f(v_i))^2$$

For the synthetic data generated from $0.6N(4, 1) + 0.4N(0, 1)$ they show that their method outperforms all others considered for a large set of (α, β) . The authors claim that this simulation shows their method is relatively stable with respect to prior parameter choices. In their first application, for a dataset with $n = 63$, they show their method successfully identifies all three modes in comparison to the noisy estimate of *K-NN* and the over smoothed *BCV* and *LSCV*. Lastly, they show that in the Claw data example, their method is the only one to come close to identifying the underlying structure of the distribution. It should be noted, however, that the prior model parameters are not near the parameters found in the authors' simulations. Therefore while it does fit the data quite well, the authors cannot claim model stability for such large values of (α, β) . Moreover, in applications we do not see a comparison to other local bandwidth estimators such as those introduced by Fan et al. 1996. I would be interested to see a comparison of the performance of variable bandwidth estimators in moderate sample sizes.

This paper improves on the variable kernel density estimation in moderate sample sizes by introducing h as a scale parameter to the kernel estimate. By developing a Bayesian framework to estimate h , the authors show that by supplying additional information in the form of the Bayesian prior $\pi(\cdot)$ may stabilize the estimation of h . Through simulation study and applications, they show that their variable bandwidth estimate outperforms global bandwidth selection procedures and is stable with respect to the prior model parameters.