Consistent Bayes with Stochastic Maps: Write-Up Summaries

Tian Yu Yen

Due: *March 7, 2018*

Stability Analysis and Proofs:

Joint Stability $\Rightarrow$ Marginal Stability

Let $Q(\lambda)$ be a stochastic map with random parameters of interest λ and stochastic error ξ. Then let $\widetilde{Q}(\lambda, \xi)$ is a deterministic map with random parameters λ and ξ.

We prove that if the joint posterior $P^{post}(\lambda, \xi)$ is stable as in Def. 4.1 or 4.2 (i.e, with respect to either perturbations in data or in the prior), then the distribution of the marginal $P^{post}(\lambda)$ will also be stable.

Proof: Let $\epsilon > 0$ and $P^{post}(\lambda, \xi)$ be stable in the sense of either 4.1 or 4.2. Then there exists a δ such that:

$$d_{TV}(P^{post}(\lambda, \xi), \widehat{P^{post}(\lambda, \xi)}) < \epsilon$$

Now consider the marginals $P^{post}(\lambda)$ and $\widehat{P^{post}(\lambda)}$:

$$\begin{aligned}
d_{TV}(P^{post}(\lambda), \widehat{P^{post}(\lambda)}) &= \int_\Lambda |\pi^{post}(\lambda) - \widehat{\pi^{post}}(\lambda)| d\mu_\Lambda \\
&= \int_\Lambda \left| \int_\Xi \pi^{post}(\lambda, \xi) - \int_\Xi \widehat{\pi^{post}}(\lambda, \xi) \right| d\mu_\Lambda \\
&= \int_\Lambda \left| \int_\Xi \left(\pi^{post}(\lambda, \xi) - \widehat{\pi^{post}}(\lambda, \xi) \right) \right| d\mu_\Lambda \\
&\leq \int_\Lambda \int_\Xi \left| \left(\pi^{post}(\lambda, \xi) - \widehat{\pi^{post}}(\lambda, \xi) \right) \right| d\mu_\Xi \, d\mu_\Lambda \\
&= \int_{\Lambda \times \Xi} \left| \left(\pi^{post}(\lambda, \xi) - \widehat{\pi^{post}}(\lambda, \xi) \right) \right| d\mu_{\Lambda,\Xi} \quad \text{(Fubini's Thm.)} \\
&= d_{TV}(P^{post}(\lambda, \xi), \widehat{P^{post}(\lambda, \xi)}) < \epsilon
\end{aligned}$$

Therefore, the marginal posterior $P^{post}(\lambda)$ is stable.

Other proof concerns:

Do we need to show that this result holds when the posterior is constructed WITHOUT using a prior distribution assumption on Q? In other words, the previous proof showed that if we calculate the posterior from the joint, then we are good for certain, but what if we don't calculate the prior (as we have discussed in previous meetings).

Questions:

1. Is it still guaranteed to be a measure? (Yes, but maybe some technical trickiness)

   Can we show that $\pi^{prior}(\lambda) \cdot \dfrac{\pi^{obs}(\widetilde{Q}(\lambda, \xi))}{\pi^{pf}(\widetilde{Q}(\lambda, \xi))}$ is still a probability measure? I think so. This is like using Baye's theorem to condition on only one parameter...

Sampling Explanation

Random variables (X_i, Y_i) for $i = 1, \ldots, n$ are a random sample if they are mutually independent with respect to i and all come from the same distribution $f(x, y)$ [adapted from Casella & Berger].

Given the random sample (X_i, Y_i), each Y_i has the marginal distribution $f_y(y) = \int_X f(x, y)$ by the definition of marginals. So, since every Y_i comes from the same distribution $f_y(y)$, Y_i for $i = 1, \ldots, n$ is a random sample of the marginal distribution $f_y(y)$.

Accept-Reject Explanation

It is possible to use accept-reject to sample from the posterior distribution, without explicitly constructing the posterior distribution. We walk through this procedure in general and then apply it to this context.

Suppose we want to draw a sample from a multivariate distribution function $f(x_1, \ldots, x_n)$. Suppose we also have a multivariate distribution function $g(x_1, \ldots, x_n)$ that we can already draw samples from.

Let M be a constant such that $f(x_1, \ldots, x_n) \leq M \cdot g(x_1, \ldots, x_n)$, for all $x_1, \ldots, x_n$. Then the accept-reject algorithm does the following:

1. Generate a test vector $(t_1, \ldots, t_n)$ from the distribution we are able to sample from $g(x_1, \ldots, x_n)$.

2. Statistically, a large sample size of f would contain the test vector $(t_1, \ldots, t_n)$ in the proportion $\frac{f}{Mg}$. We simulate this by drawing a new random number y from $U[0, 1]$ and comparing it to the ratio $\frac{f}{Mg}$. Specifically, if:

$$y < \frac{f(t_1, \ldots, t_n)}{M \cdot g(t_1, \ldots, t_n)} \Rightarrow \text{ we accept } (t_1, \ldots, t_n)$$
$$y \geq \frac{f(t_1, \ldots, t_n)}{M \cdot g(t_1, \ldots, t_n)} \Rightarrow \text{ we reject } (t_1, \ldots, t_n)$$

3. Repeat steps 1 and 2 until we've accepted the desired number of samples.

In the context of consistent bayes, we can similarly construct a sample from the joint posterior $\pi^p(\lambda, \xi)$ using the accept reject algorithm. Consider:

$$\pi^p(\lambda, \xi) = \pi^{prior}(\lambda, \xi) \cdot \frac{\pi^{obs}(Q(\lambda, \xi))}{\pi^{pf}(Q(\lambda, \xi))}$$

Then let $M > \sup \dfrac{\pi^{obs}(Q(\lambda, \xi))}{\pi^{pf}(Q(\lambda, \xi))}$.

This implies that $\pi^p(\lambda, \xi) \leq \pi^{prior}(\lambda, \xi) \cdot M$.

Then to follow the accept-reject given above, we first get a test vector (λ^*, ξ^*) and calculate its corresponding QoI value q^*. We draw a random number y from $U[0, 1]$ and compare it the following ratio:

$$\text{ratio} = \frac{\pi^p(\lambda^*, \xi^*)}{M \cdot \pi^{prior}(\lambda^*, \xi^*)} = \frac{\pi^{prior}(\lambda^*, \xi^*) \cdot \pi^{obs}(q^*)}{M \cdot \pi^{prior}(\lambda^*, \xi^*) \cdot \pi^{pf}(q^*)} = \frac{\pi^{obs}(q^*)}{M\pi^{pf}(q^*)}$$

$$\begin{aligned} y < \text{ratio} &\Rightarrow \text{ we accept } (\lambda^*, \xi^*) \\ y > \text{ratio} &\Rightarrow \text{ we reject } (\lambda^*, \xi^*) \end{aligned}$$

We can repeat this process multiple times until the desired sample size is achieved. Then because samples from the joint distribution also act as samples from the marginal distributions, we can take just the $\lambda_1, \ldots, \lambda_n$ values as our sample from the posterior of the parameter of interest, ignoring the ξ sample values.

One beneficial consequence of using this accept-reject method for generating samples from the posterior distribution is that it works even when the stochastic error is embedded in the map Q and it is not possible to explicitly generate or save these values.

For example, suppose that the model map Q generates some sort of random variable ξ during its execution. In this case, we can first generate a λ value from our prior distribution on the parameter of interest, λ^*, and then calculate the QoI q^* using $Q(\lambda^*)$. The map implicitly uses a random value ξ^*, but we have no access to it.

In this case, our ratio can be rewritten in the following manner:

$$\begin{aligned} \text{ratio} &= \frac{\pi^p(\lambda^*, \xi^*)}{M \cdot \pi^{prior}(\lambda^*, \xi^*)} \\ &= \frac{\pi^\xi(\xi \mid \lambda^*)\pi^{prior}(\lambda^*) \cdot \pi^{obs}(q^*)}{M\pi^\xi(\xi \mid \lambda^*) \cdot \pi^{prior}(\lambda^*, \xi^*) \cdot \pi^{pf}(q^*)} \\ &= \frac{\pi^{obs}(q^*)}{M\pi^{pf}(q^*)} \end{aligned}$$

which is our desired result. Thus we have calculated the accept-reject ratio, even though we did not have access to the generation of ξ. Since we are only interested in λ values, the loss of the ξ variable is of no concern to us.

VERRRYY HAND-WAAVY!!!