Notes

1 Testing

(a) Geweke

Given the model

$$p(\theta, \mathbf{y}) = p(\theta)P(\mathbf{y}|\theta) \tag{1}$$

Define the test function $g : \mathbf{\Theta} \times \mathbf{Y} \to \mathbb{R}$ such that $\mathrm{Var}(g(\theta, \mathbf{y})) < \infty$. The Geweke joint distribution test compares two estimates of $\bar{g} = \mathbb{E}[g(\theta, \mathbf{y})]$ using samples from the joint simulators in Algorithms 1 and 2.

Algorithm 1 marginal-conditional

Initialize $\mathbf{g}_1 \in \mathbb{R}_{M\times 1}$
for $m = 1, \dots, M$ **do**
    $\theta_m \sim p(\theta)$
    $\mathbf{y}_m \sim p(\mathbf{y}|\theta_m)$
    $\mathbf{g}_1[m] = g(\theta_m, \mathbf{y}_m)$
end for
return $\mathbf{g}_1$

Algorithm 2 successive-conditional

Initialize $\mathbf{g}_2 \in \mathbb{R}_{M\times 1}$
$\theta_0 \sim p(\theta)$
for $m = 1, \dots, M$ **do**
    $\mathbf{y}_m \sim p(\mathbf{y}|\theta_{m-1})$
    $\theta_m \sim q(\theta|\theta_{m-1}, \mathbf{y}_m)$
    $\mathbf{g}_2[m] = g(\theta_m, \mathbf{y}_m)$
end for
return $\mathbf{g}_2$

In particular,

$$\frac{\hat{\bar{g}}_1 - \hat{\bar{g}}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{M_1} + \frac{\hat{\sigma}_2^2}{M_2}}} \xrightarrow{d} \mathcal{N}(0, 1) \tag{2}$$

with the mean estimates given by

$$\hat{\bar{g}}_1 = \frac{1}{M}\sum_{m=1}^{M} g_1^{(m)}, \qquad \hat{\bar{g}}_2 = \frac{1}{M}\sum_{m=1}^{M} g_2^{(m)}$$

The variance estimate for the marginal-conditional samples is straightforward

$$\hat{\sigma}_1^2 = \frac{1}{M}\sum_{m=1}^{M} (g_1^{(m)} - \hat{\bar{g}}_1)^2$$

However, the successive-conditional variance estimate is not so simple, since the samples are dependent. One choice is the window estimator

$$\hat{\sigma}_2^2 = \frac{1}{M} \sum_{t=-\infty}^{\infty} w(t)\hat{\gamma}(t)$$

$$\hat{\gamma}(t) = \hat{\gamma}(-t) = \frac{1}{M} \sum_{i=1}^{M-t} (g_2^i - \hat{\bar{g}}_2)(g_2^{i+t} - \hat{\bar{g}}_2)$$

where w is a weight function (lag window). Geweke (1999) chooses

$$w(t) = \max\left(\frac{L-t}{L}, 0\right), \quad L > 0$$

$$L \in \{0.04, 0.08, 0.15\} \times M$$

Alternatively one might use a batch mean estimator. We divide the samples into B non-overlapping batches of size m. Then, for large m, the batch means $\{\bar{g}_j\}_{j=1}^{B}$ are approximately independent and

$$\bar{g}_j \sim \mathcal{N}(\bar{g}, \frac{\sigma^2}{m})$$

$$\hat{\sigma}_{2,BM}^2 = \frac{m}{B-1} \sum_{j=1}^{B} (\bar{g}_j - \bar{g})^2$$

Performance may be improved using multiple chains. However, the choice of batch size B may be challenging — we need to choose B large enough for the batch means to be approximately independent, but not so large that the confidence interval of the estimator explodes.

Still another alternative is initial sequence estimators (Geyer 1992).

For a significance level α, the testing procedure is

- Draw $\mathbf{g}_1, \mathbf{g}_2$
- Calculate $z = \frac{\hat{\bar{g}}_1 - \hat{\bar{g}}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{M_1} + \frac{\hat{\sigma}_2^2}{M_2}}}$
- If $|z| \geq \Phi^{-1}(1 - \alpha/2)$, reject the null hypothesis that the distributions are the same

When the number of test functions grows large, we expect some of the joint distribution tests to fail by chance. To compensate for this, we might introduce a Bonferroni correction and scale down the significance levels. However, this may be too conservative (reduce test power too much), especially if the test statistics are positively correlated.

A less principled but more intuitive approach is to examine the PP plot of the empirical marginal-conditional and successive-conditional distributions. If the points are close to the unit line, then we fail to reject the null.

(b) MMD

(i) Wild bootstrap

This approach is most similar to the Geweke test.

Given n_x τ-dependent samples from $p(X)$ and n_y τ-dependent samples from $p(Y)$, the biased empirical MMD is

$$\begin{aligned}\widehat{\mathrm{MMD}}_k^2 = \tfrac{1}{n_x^2} \textstyle\sum_{i=1}^{n_x} \sum_{j=1}^{n_z} k\left(x_i, x_j\right) + \tfrac{1}{n_y^2} \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} k\left(y_i, y_j\right) \\ -\tfrac{2}{n_x n_y} \textstyle\sum_{i=1}^{n_s} \sum_{j=1}^{n_y} k\left(x_i, y_j\right)\end{aligned} \tag{3}$$

Define the wild bootstrap process $\{W_{t,n}\}_{1\leq t\leq n}$ as

$$W_{t,\mathrm{n}} = e^{-1/t_n} W_{t-1,n} + \sqrt{1-e^{-2/l_n}}\epsilon_t \tag{4}$$

with $W_{0,m}, \epsilon_t \sim \mathcal{N}(0,1)$, satisfying the bootstrap assumption from Chwialkowski et al. (2016).

Then bootstrapped MMD is

$$\begin{aligned}\widehat{\mathrm{MMD}}_{k,b}^2 = \tfrac{1}{n_x^2} \textstyle\sum_{i=1}^{n_x} \sum_{j=1}^{n_z} \tilde{W}_i^{(x)} \tilde{W}_j^{(x)} k\left(x_i, x_j\right) + \tfrac{1}{n_y^2} \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} \tilde{W}_i^{(y)} \tilde{W}_j^{(y)} k\left(y_i, y_j\right) \\ -\tfrac{2}{n_x n_y} \textstyle\sum_{i=1}^{n_s} \sum_{j=1}^{n_y} \tilde{W}_i^{(x)} \tilde{W}_j^{(y)} k\left(x_i, y_j\right)\end{aligned} \tag{5}$$

with $\tilde{W}_t^{(x)} = W_t^{(x)} - \frac{1}{n_z}\sum_{i=1}^{n_x} W_i^{(x)}, \tilde{W}_t^{(y)} = W_t^{(y)} - \frac{1}{n_y}\sum_{j=1}^{n_y} W_j^{(y)}$, though we don't have to center the wild bootstrap process.

Under the null hypothesis $p(X) = p(Y)$

$$\varphi\left(\rho_x\rho_y n \widehat{MMD}_k^2, \rho_x\rho_y n \widehat{MMD}_{k,b}^2\right) \xrightarrow{p} 0, \quad n\to\infty$$

where $\rho_x = \frac{n_x}{n_x+n_y}$, $\rho_y = \frac{n_y}{n_x+n_y}$.

For a significance level α and B bootstrap samples, the testing procedure is

- Draw $\{\mathbf{y}_1^{(n)}, \theta_1^{(n)}\}_{n=1}^{n_1}, \{\mathbf{y}_2^{(n)}, \theta_2^{(n)}\}_{n=1}^{n_2}$
- Simulate $\{\rho_1\rho_2 n \widehat{\mathrm{MMD}}_{k,b}^2\}_{b=1}^B$
- Calculate c_α, the $1-\alpha$ empirical quantile of $\{\rho_1\rho_2 n\widehat{\mathrm{MMD}}_{k,b}^2\}_{b=1}^B$
- If $\rho_1\rho_2 n \widehat{MMD}_k^2 \geq c_\alpha$, reject the null hypothesis that the distributions are the same

(ii) Backward burn-in

A major disadvantage of the successive-conditional sampler is that it cannot be parallelized, i.e., we must draw one sample at a time. Instead, we could draw from the marginal distribution of $\mathbf{y}$ and burn in the posterior simulator to get θ.

Algorithm 3 backward-conditional

Initialize $\mathbf{g}_3 \in \mathbb{R}_{M\times 1}$
for $m = 1, \ldots, M$ **do**
    $\theta_0 \sim p(\theta)$
    $\mathbf{y}_m \sim p(\mathbf{y}|\theta_0)$
    for $n = 1, \ldots, N$ **do**
        $\theta_n \sim q(\theta|\theta_{n-1}, \mathbf{y}_m)$
    end for
    $\mathbf{g}_3[m] = g(\theta_n, \mathbf{y}_m)$
end for
return $\mathbf{g}_3$

Since the samples are independent, we can then apply the Geweke test without the spectral variance estimator, or apply a test based on the unbiased MMD

$$\widehat{\mathrm{MMD}}_{\mathrm{U}}^{2}(X, Y) = \frac{1}{\binom{m}{2}} \sum_{i \neq i'} k\left(X_i, X_{i'}\right) + \frac{1}{\binom{m}{2}} \sum_{j \neq j'} k\left(Y_j, Y_{j'}\right) - \frac{2}{\binom{m}{2}} \sum_{i \neq j} k\left(X_i, Y_j\right) \tag{6}$$

For a significance level α and B bootstrap samples, the testing procedure is

- Draw $\{\mathbf{y}_1^{(m)}, \theta_1^{(m)}\}_{m=1}^{n}, \{\mathbf{y}_3^{(m)}, \theta_3^{(m)}\}_{m=1}^{n}$
- Simulate the null distribution of $n\widehat{\mathrm{MMD}}_{\mathrm{U}}^{2}$ via permutation and calculate the $1-\alpha$ empirical quantile c_α
- If $n\widehat{\mathrm{MMD}}_{\mathrm{U}}^{2}(\{\mathbf{y}_1^{(m)}, \theta_1^{(m)}\}_{m=1}^{n}, \{\mathbf{y}_3^{(m)}, \theta_3^{(m)}\}_{m=1}^{n}) \geq c_\alpha$, reject the null hypothesis that the distributions are the same

2 Experiments

See BayesianLassoDemo.ipynb

3 Notes TODO

- Unify notation
    - MMD tests on g rather than $\{y, \theta\}$?
- Experiments