

## SAMPLING MODEL CLARIFICATIONS

NELSON RAY

Stein’s “Theorem 1” (page 35 of “Approximate Computation of Expectations”) holds very generally, for any random variable  $W$  that satisfies the conditions:

**Theorem 1.** *Let  $W, W'$  be an exchangeable pair with  $\mathbb{E}[W'|W] = (1 - \lambda)W$  and  $0 < \lambda < 1$ . Then*

$$|P(W \leq w) - \Phi(w)| \leq 2\sqrt{\mathbb{E}[1 - \frac{1}{2\lambda}E[(W' - W)^2|W]]} + (2\pi)^{-1/4}\sqrt{\frac{1}{\lambda}\mathbb{E}|W' - W|^3}$$

We have the more general (I’ll just show the result with the extra condition to get the better rate) result with the approximate regression condition:

**Theorem 2.** *If  $W, W'$  are mean 0 exchangeable random variables with variance  $\mathbb{E}W^2$  satisfying*

$$\mathbb{E}[W' - W|W] = -\lambda(W - R)$$

*for some  $\lambda \in (0, 1)$  and some random variable  $R$  and  $|W' - W| \leq \delta$ , then*

$$\begin{aligned} \sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| &\leq \frac{.41\delta^3}{\lambda} + 3\delta(\sqrt{\mathbb{E}W^2} + \mathbb{E}|R|) + \frac{1}{2\lambda}\sqrt{\text{Var}(\mathbb{E}[(W' - W)^2|W])} \\ &\quad + |\mathbb{E}W^2 - 1| + \mathbb{E}|WR| + \mathbb{E}|R| \end{aligned}$$

In the specific condition of letting the random variable  $W$  be the sum of a random diagonal of an appropriately-scaled matrix, Stein’s theorem is specialized to

**Theorem 3.** *Let  $A = \{a_{ij}\}_{i,j \in \{1, \dots, n\}}$  be a square array of numbers such that  $\sum_j a_{ij} = 0$  for all  $i$ ,  $\sum_i a_{ij} = 0$  for all  $j$ , and  $\sum_i \sum_j a_{ij}^2 = n - 1$ . Then with  $F_n(x) = P(\sum_i a_{i\Pi(i)} \leq x)$ ,*

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sqrt{n}} \left( \sqrt{\sum_{i,j} a_{ij}^4} + \sqrt{\sum_{i,j} |a_{ij}|^3} \right)$$

Similarly, we apply our theorem to the situation of observing fixed data  $u_i$ , rescaling it, and taking  $W$  to be the  $t$ -statistic under the randomization distribution.

Then we get the bound

$$N^{-1/2}c_1''^* + N^{-1}f_1'(\mathbf{u})^* + N^{-1}f_2(\mathbf{u}) + N^{-1}f_3(\mathbf{u}) + N^{-1/2}f_4(\mathbf{u}) + N^{-1/2}f_5(\mathbf{u}).$$

This is similar to the result in the HCCLT because there’s an  $N^{-1/2}$  rate, universal constants, and functions that depend on the observed data.

Up to here, all the results we have are conditional on the data. I think I might have confused myself in trying to relate the “conditional on the data” results with the “asymptotic, Berry–Esseen-like” results.

I took another look through “Approximate Computation of Expectations” and “Normal Approximation by Stein’s Method” and couldn’t find anything beyond this.

However, we can attempt to bridge the gap between the two types of results if we incorporate a sampling model.

Let’s say instead of having fixed data  $u_i$ , we draw  $u_i$  from some distribution. Now, we can calculate our upper bounds because they are deterministic functions of the data. But if we say  $U_i \sim p$ , then our upper bounds will now be random variables with some (probably very hard to calculate) distributions.

The simplest situation to analyze is just to take  $u_i = i$  and see how the various  $f_i(\mathbf{u})$  change as  $N$  gets large (this is what I tried to do with  $\delta = \max |T' - T|$ ).

I'm happy to try to develop this further, but it would be great to have a reference or a guide of some sort. The HCCLT result and proof were fantastic in this respect, since the situation is so similar and a bit more basic.

If I understood your last e-mail about using Slutsky's theorem and postulating moments for the distribution correctly, I'd start by doing simulations with a sampling model (say,  $u_i$  are drawn from the Normal). The results look good, so I can assume something like with high probability, the various functions of the data  $f_i$  will be well-behaved (say, bounded). It's always necessary to include the "with high probability" qualifier when sampling in this case because I can always construct a sequence  $u_i$  with positive probability that is pathological.

I haven't checked yet, but probably the Cauchy wouldn't work out so well because its moments aren't finite. So a moment condition would certainly get rid of that bad case. But I'm not really sure where to go next. Let's say for simplicity that we only allow sampling from the Normal (a very strict moment condition). Also for simplicity, let's look at the HCCLT. If  $A_{i,j} \sim \mathcal{N}(0, 1)$ , then we still need to scale them so they fit the conditions of the theorem. Let the scaled random variables be  $\tilde{A}_{i,j}$ . Then the bound has

$$\sqrt{\sum_{i,j} \tilde{A}_{i,j}^4} + \sqrt{\sum_{i,j} |\tilde{A}_{i,j}|^3}.$$

If the  $\tilde{A}_{i,j}$  were independent (they aren't because of the scaling), then it wouldn't be so bad. After all, the sum of square of i.i.d. standard Normals is  $\chi^2$ . But I'm not sure how to treat the dependence and where to go from here.

Also, I don't understand how Slutsky's theorem fits in. Can you clarify? I think one of my problems is that Slutsky's theorem concerns convergence in distribution, but I'm not sure how I can use those results in trying to prove rates of convergence bounds.