Chapter 10

Randomization via quantile coupling

SECTION 1 explains why the method from the previous Chapter cannot quite capture Nussbaum's asymptotic equivalence over his full range of smoothness parameters.

SECTION 2 describes Andrew Carter's second method of randomization, based on quantile coupling, which solves the problem described in Section 1.

SECTION 3 derives Carter's inequality for randomization via the quantile coupling.

1. Randomization of Binomials

Carter's method, as described in Chapter 9, relied on repeated convolution smoothing as the randomization to make the (conditional) Binomials close to normals in total variation. The key inequality was:

<1>
$$H^2\left(\text{Bin}(n,\,p)\star\mathfrak{U},\,N(np,npq)\right)\le\frac{C}{(1+n)pq},$$

where $\mathfrak{U}$ denotes the Uniform distribution on $(-1/2,1/2)$ and C is a universal constant. After some further randomizations involving the normal models, the method led to a bound

<2>
$$\Delta(\mathcal{M},\mathcal{N}_{\text{stabil}})\le C'_\Theta\frac{m\log m}{\sqrt{n}}\qquad\text{provided }\sup_{\theta\in\Theta}\frac{\max_i\theta_i}{\min_i\theta_i}<\infty,$$

where $\mathcal{M}:=\{\mathcal{M}(n,\theta):\theta\in\Theta\}$, a collection of multinomials with m cells, and $\mathcal{N}_{\text{stabil}}:=\{\widetilde{\mathbb{N}}_\theta:\theta\in\Theta\}$ with $\widetilde{\mathbb{N}}_\theta=\otimes_{i\le m}N(\sqrt{n\theta_i},1/4)$.

In order to recover the result of Nussbaum (1996), we work with a value of m that increases with n. For example, if $\alpha\le 3/2$ then the methods like those described in the last Section of Chapter 9 give approximations to the multinomial white noise processes of order $m^{-\alpha}n^{1/2}$. For the approximation in <2> to be of value, we need $m\log m=o\left(n^{1/2}\right)$. Thus the convolution method can reproduce Nussbaum's asymptotic equivalence when $\alpha>1$.

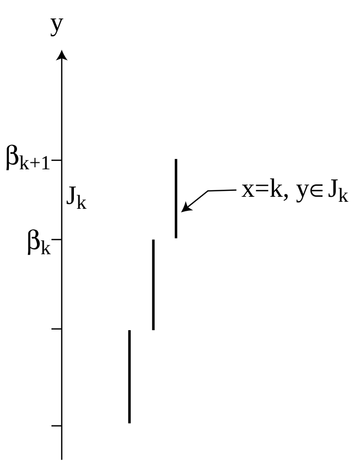
Check rate for m

REMARK. Of course the method also works for $\alpha>3/2$, but then edge effects prevent faster rates of convergence than are achieved when $\alpha=3/2$.

For the range $1/2 < \alpha \leq 1$, we need an improvement over randomization based on convolution smoothing of conditional Binomials, at least when m is large. In fact, Carter (2001) showed that the construction underlying <2> could be invoked to derive a preliminary approximation based on an m of order $n^{1/(1+2\alpha)}$. At that stage all the cells correspond to very short subintervals of $[0, 1]$. The smoothness condition on the density ensures that the θ_i for neighboring cells are close, which means that the parameters p_i for the conditional $\text{Bin}(s_i, p_i)$ distributions are all close to $1/2$. For those cases a randomization based on the quantile coupling of $\text{Bin}(s_i, 1/2)$ and $N(s_i/2, s_i/4)$ distributions gives a superior approximation. The result is based on slight modification by Carter & Pollard (2000) of a bound due to Tusnády (1977).

REMARK. See Csörgő & Révész (1981, Section 4.4) and Bretagnolle & Massart (1989) for a discussion of the role of Tusnády's lemma in the so-called Hungarian construction. The Appendix to the latter actually contains a full proof of the lemma. Chapter 10 of Pollard (2001) also contains an exposition of the Hungarian construction, with an Appendix that explains in a more leisurely fashion the Carter-Pollard results.

2. Randomization via quantile coupling



Let Y have a $N(n/2, n/4)$ distribution. There is an increasing function ψ_n for which $\psi_n(Y)$ has exactly a $\text{Bin}(n, 1/2)$ distribution. The function is defined by the sequence of cutpoints

$$-\infty = \beta_0 < \beta_1 < \ldots < \beta_n < \beta_{n+1} = \infty$$

for which $\mathbb{P}\{\text{Bin}(n, \tfrac{1}{2}) \geq k\} = \mathbb{P}\{Y > \beta_k\}$ for each k. Define $J_k := (\beta_k, \beta_{k+1}]$. Then

$$\psi_n(y) := \sum\nolimits_{k=0}^{n} \{y \in J_k\} k$$

is the desired function. Symmetry of Q_0 about $n/2$ implies a similar symmetry for the J_k intervals: $y \in J_k$ if and only if $n - y \in J_{n-k}$. We can also think of the joint distribution of $\psi_n(Y)$ and Y as a probability measure Γ on $\Omega := \{0, 1, \ldots, n\} \times \mathbb{R}$ with marginals $P_0 := \text{Bin}(n, 1/2)$ and $Q_0 := N(n/2, n/4)$. That is,

<3> $$\Gamma^{x,y} h(x, y) = Q_0^y h(\psi_n(y), y) = \sum\nolimits_{k=0}^{n} \{y \in J_k\} h(k, y) \qquad \text{for } h \in \mathcal{M}^+(\Omega).$$

The joint distribution can also be specified by a conditional distribution for the Y given $\psi_n(Y)$. That is, $\Gamma = P_0 \otimes K$, where K is the probability kernel from $\{0, 1, \ldots, n\}$ to $\mathbb{R}$ for which $K_x = Q_0(\cdot \mid J_x)$.

Write P_δ for the $\text{Bin}(n, (1+\delta)/2)$ and Q_δ for the $N(n(1+\delta)/2, n/4)$ distribution. The kernel K is defined so that $KP_0 = Q_0$. If δ is close to 0, then we might hope that $KP_\delta \approx Q_\delta$ in a total variation sense. Indeed, Carter showed there is a neat bound for $D(KP_\delta \| Q_\delta)$, which makes it worthwhile to replace convolution smoothing by the randomization K when the probabilities for neighboring multinomial cells are close.

> REMARK. Notice that the variance for Q_δ was chosen as fixed, not depending on δ. When δ is close enough to 0, this simplification will add only small terms in the recursive step when we need to compare the randomized conditional Binomials with the appropriate conditional normals.

I will not give all the details required for modification of the recursive argument from Chapter 9, but instead I refer you to Carter (2001). Once we have a bound for $H^2(KP_\delta, Q_\delta)$, or for the larger quantity $D(KP_\delta \| Q_\delta)$, the argument is similar to the argument based on <2>.

3. Bound for the quantile randomization

The calculation starts from the densities,

$$f_\delta(x) := \frac{dP_\delta}{dP_0} = (1+\delta)^x(1-\delta)^{n-x}$$
$$g_\delta(x) := \frac{dQ_\delta}{dQ_0} = \exp\left(2\delta(y-n/2) - n\delta^2/2\right).$$

We first need to determine the density of KP_δ with respect to Q_0. For $h \in \mathcal{M}^+(\mathbb{R})$,

$$\begin{aligned}(KP_\delta)^y h(y) := P_\delta^x K_x^y h(y) &= P_0^x f_\delta(x) K_x^y h(y) \\ &= \Gamma^{x,y} f_\delta(x) h(y) = Q_0^y \textstyle\sum_{k=0}^n \{y \in J_k\} f_\delta(k) h(y).\end{aligned}$$

Thus

<4> $$\frac{d(KP_\delta)}{dQ_0} = \textstyle\sum_{k=0}^n \{y \in J_k\} f_\delta(k) = f_\delta(\psi_n(y)).$$

Notice that $f_{-\delta}(n-k) = f_\delta(k)$ and $g_{-\delta}(n-y) = g_\delta(y)$. Together with the symmetries for Q_0 and the J_k, these equalities imply that $H(KP_\delta, Q_\delta)$ is a symmetric functions of δ. Thus we need only consider positive δ.

From <4>,

$$\begin{aligned}D(KP_\delta \| Q_\delta) &= Q_0^y\left(f_\delta(\psi_n(y)) \log \frac{f_\delta(\psi_n(y))}{g_\delta(y)}\right) \\ &= \Gamma^{x,y}\left(f_\delta(x) \log \frac{f_\delta(x)}{g_\delta(y)}\right) \\ &= P_0^x f_\delta(x)\left(n\delta + \tfrac{1}{2}n\delta^2 + \log f_\delta(x) - 2\delta x\right) + \Gamma^{x,y}\left(2\delta f_\delta(x)\,(x-y)\right).\end{aligned}$$

The first integral equals

$$\begin{aligned}&P_\delta\left(\tfrac{1}{2}n\delta^2 + \log f_\delta(x) - \delta(2x-n)\right) \\ &\quad = -\tfrac{1}{2}n\delta^2 + \tfrac{1}{2}n\left((1+\delta)\log(1+\delta) + (1-\delta)\log(1-\delta)\right) \\ &\quad = \frac{n}{12}\left(\delta^4 + O(\delta^6)\right).\end{aligned}$$

The second integral equals

$$2\delta \sum\nolimits_{k=0}^{n} f_\delta(k) Q_0\left((k-y)\{y \in J_k\}\right).$$

There is a partial cancellation between the kth and $(n-k)$th terms, because y and $n-y$ have the same distribution under Q_0:

$$\begin{aligned} f_\delta(n-k)Q_0\big(\{y\in J_{n-k}\}(n-k-y)\big) &= f_{-\delta}(k)Q_0\big(\{n-y\in J_k\}(n-y-k)\big) \\ &= -f_{-\delta}(k)Q_0\big(\{y\in J_k\}(k-y)\big). \end{aligned}$$

If n is even, the central term, for $k=n/2$, is identically zero. Thus

<5>
$$2\delta\Gamma f_\delta(x)(k-y) = 2\delta\sum\nolimits_{k>n/2}\left(f_\delta(k) - f_{-\delta}(k)\right)\left(Q_0 J_k\right) Q_0\left(k-y \mid y\in J_k\right).$$

Notice that

<6>
$$0 \le 1 - \frac{f_{-\delta}(k)}{f_\delta(k)} = 1 - \left(\frac{1-\delta}{1+\delta}\right)^{2k-n} \le \frac{2\delta(2k-n)}{1+\delta}.$$

Carter & Pollard (2000) showed that there is a universal constant c_0 for which

<7>
$$\beta_k \ge k - \tfrac{1}{2} - c_0 n^{-1/2} \qquad \text{when } k > n/2.$$

I think I have made a silly mistake somewhere in the calculations that follow. Somehow I have gotten and extra factor of $n^{-1/2}$ into one of the terms derived by Carter. Problem [2] seems the likely suspect. I will track down the mistake and correct it in the revised version of these Paris notes.

REMARK. Check to see whether the original Tusnády bound, $\beta_k \ge k-1$, would be enough. Probably not, unless my extra $n^{-1/2}$ factor survives.

By Problem [1],

$$\begin{aligned} Q_0\left(k-y \mid y\in J_k\right) &= Q_0\big(k-y \mid k-\beta_{k+1} < k-y \le k-\beta_k\big) \\ &\le Q_0\left(k-y \mid k-y\in I\right), \end{aligned}$$

where $I := [c_0 n^{-1/2} - \frac{1}{2}, c_0 n^{-1/2} + \frac{1}{2})$. By Problem [2], the final conditional expectation is less than

$$2\left(\tfrac{1}{2}\right)^2 \left| \frac{c_0}{\sqrt{n}} - \left(k - \frac{n}{2}\right)\right| / (n/4).$$

From this upper bound and <6> we get an upper bound for the right-hand side of <5>,

$$\begin{aligned} &\frac{(2\delta)^2}{n} Q_0 \sum\nolimits_{k>n/2} f_\delta(k)\{y\in J_k\}\left(\frac{2c_0|2k-n|}{\sqrt{n}} + |2k-n|^2\right) \\ &\qquad \le \frac{(2\delta)^2}{n} P_\delta\left(\frac{2c_0|2k-n|}{\sqrt{n}} + |2k-n|^2\right). \end{aligned}$$

Under P_δ, the difference $2k-n$ has mean $n\delta$ and variance $n(1-\delta^2)$, which suggests the simple bounds

$$\begin{aligned} P_\delta|2k-n| &\le n\delta + \sqrt{\text{var}(2k)} \le n\delta + \sqrt{n} \\ P_\delta|2k-n|^2 &\le (n\delta)^2 + \text{var}(2k) \le n + n^2\delta^2, \end{aligned}$$

leading to the inequality

$$2\delta\Gamma f_\delta(x)(k-y) \le C(\delta^2 + n\delta^4) \qquad \text{for some constant } C.$$

If the my calculations are to be believed, the final inequality takes the form

$$D(KP_\delta \| Q_\delta) \le C'(\delta^2 + n\delta^4) \qquad \text{for some constant } C'.$$

For the sake of comparison, Carter (2001) asserts that

$$D(KP_\delta \| Q_\delta) \le C'(\delta^2 + \sqrt{n}|\delta|^3 + n\delta^4) \qquad \text{for some constant } C'.$$

He needs to use his bound for cases where $|\delta| \le c_1 m^{-\alpha}$ with $\sqrt{n}/m^\alpha \to 0$, that is, for values with $\delta = o(n^{-1/2})$. For those cases, the δ^2 dominates the bound. Thus,

$$H^2(KP_\delta, Q_\delta) = O(\delta^2) \qquad \text{for } \delta = o(n^{-1/2}).$$

Compare with <1>, which gives a bound of order n^{-1} for the randomization based on convolution smoothing.

4. Problems

[1] Let Q be a probability measure on the real line, such that $QI > 0$ for every nondegenerate interval. Show that $Q\,(y \mid a \le y < b)$ is an increasing function of both a and b. Hint: For $t \in (a,b)$, show that

$$Q\,(y \mid a \le y < b) = \frac{Q[a,t)}{Q[a,b)} Q\,(y \mid a \le y < t) + \frac{Q[t,b)}{Q[a,b)} Q\,(y \mid t \le y < b)\,.$$

Then argue from the fact that $Q\,(y \mid a \le y < t) \le t \le Q\,(y \mid t \le y < b)$.

This result seems too good to be true

[2] Suppose W has a $N(\mu, \sigma^2)$ distribution. For each $h > 0$ and each $x \in \mathbb{R}$, show that

$$|\mathbb{P}\,(W \mid x-h \le W < x+h) - x| \le \frac{2|x-\mu|h^2}{\sigma^2}.$$

Hint: Reduce to the case where $\mu = 0$ and $\sigma = 1$. For $s > 0$, define $F(s) := \int_{z-s}^{z+s} (t-z)\phi(t)\,dt$ and $G(h) := \int_{z-s}^{z+s} \phi(t)\,dt$. Show that

$$|F'(s)/G'(s)| = s\left|\frac{\phi(z+s) - \phi(z-s)}{\phi(z+s) - \phi(z-s)}\right| \le 2|z|s^2\{|zs| \le 1\} + |s|\{|zs| > 1\}.$$

Invoke the mean value theorem for $F(h/\sigma)/G(h/\sigma)$.

References

Bretagnolle, J. & Massart, P. (1989), 'Hungarian constructions from the nonasymptotic viewpoint', *Annals of Probability* **17**, 239–256.

Carter, A. (2001), Le Cam distance between multinomial and multivariate normal experiments under smoothness constraints on the parameter set, Technical report, University of California, Santa Barbara. http://www.pstat.ucsb.edu/~carter.

Carter, A. & Pollard, D. (2000), Tusnády's inequality revisited, Technical report, Yale University. http://www.stat.yale.edu/~pollard.

Csörgő, M. & Révész, P. (1981), *Strong Approximations in Probability and Statistics*, Academic Press, New York.

Nussbaum, M. (1996), 'Asymptotic equivalence of density estimation and gaussian white noise', *Annals of Statistics* **24**, 2399–2430.

Pollard, D. (2001), *A User's Guide to Measure Theoretic Probability*, Cambridge University Press. Some samples at http://www.stat.yale.edu/˜pollard.

Tusnády, G. (1977), A study of Statistical Hypotheses, PhD thesis, Hungarian Academy of Sciences, Budapest. In Hungarian.