

# Likelihood ratio Haar variance stabilization and normalization for Poisson and other non-Gaussian noise removal – online supplement

Piotr Fryzlewicz\*

July 19, 2017

## 1 Links between likelihood ratio Haar wavelets and the Haar-Fisz methodology (with numerical examples)

This section compares the likelihood ratio Haar coefficients  $g_{j,k}$ , defined in the general, Poisson and chi-squared cases in formulae (2), (4) and (5) of the main paper, respectively, to the Fisz coefficients  $f_{j,k}$  (Fryzlewicz and Nason, 2004), which the above work defines as the Haar coefficients  $d_{j,k}$  divided by the maximum likelihood estimates of their own standard deviation under the null hypothesis  $E(d_{j,k}) = 0$ . We start with the Poisson case and note that by Fryzlewicz and Nason (2004),  $f_{j,k}$  is then expressed as

$$f_{j,k} = 2^{j/2-1} \frac{\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^j-1} - \bar{X}_{(k-1)2^j+2^j-1+1}^{k2^j}}{\sqrt{\bar{X}_{(k-1)2^j+1}^{k2^j}}}.$$

We first note that  $\text{sign}(g_{j,k}) = \text{sign}(f_{j,k})$  and that Lemma 3.2, used with  $f(u) = u \log u$ ;  $f(0) = 0$  in the notation of that lemma, reduces to  $|g_{j,k}| \geq |f_{j,k}|$ . Moreover, since the inequality in Lemma 3.2 arises as a simple application of Jensen's inequality to the convex function  $f(\cdot)$ , it is intuitively apparent that the less convexity in  $f(\cdot)$ , the closer  $g_{j,k}$  will be to  $f_{j,k}$ . Noting that  $f''(u) = u^{-1}$  and therefore the degree of convexity in  $f(u)$  decreases as  $u$  increases, it can heuristically be observed that  $g_{j,k}$  and  $f_{j,k}$  should be closer to each other for larger values of  $\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^j-1}$  and  $\bar{X}_{(k-1)2^j+2^j-1+1}^{k2^j}$  (i.e. for high Poisson intensities), and further apart otherwise.

To illustrate this phenomenon and other interesting similarities and differences between the Fisz and the likelihood ratio Haar coefficients in the Poisson case, consider the following two numerical examples, in which we simulate 1000 realisations of  $\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^j-1}$

---

\*Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: [p.fryzlewicz@lse.ac.uk](mailto:p.fryzlewicz@lse.ac.uk). Work supported by the Engineering and Physical Sciences Research Council grant no. EP/L014246/1.

and  $\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}$  and compute the corresponding 1000 realisations of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$ .

- $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ . As is apparent from Figure 1, the values of  $g_{j,k}^{(i)} - f_{j,k}^{(i)}$  are close to zero. Figure 2 provides further evidence that the empirical distributions of  $f_{j,k}^{(i)}$  and  $g_{j,k}^{(i)}$  are difficult to distinguish by the naked eye. Q-q plots (not shown) exhibit good agreement for both  $g_{j,k}^{(i)}$  and  $f_{j,k}^{(i)}$  with the normal distribution, and we have  $\widehat{\text{Var}}(g_{j,k}^{(i)}) = 1.06$  and  $\widehat{\text{Var}}(f_{j,k}^{(i)}) = 1.05$ , which provides evidence that both the likelihood ratio Haar coefficients and the Fisz coefficients achieve good variance stabilization in this high-intensity case.
- $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ . Figures 3 and 4 demonstrate that in this low-intensity case, the distributions of  $f_{j,k}^{(i)}$  and  $g_{j,k}^{(i)}$  are now further apart. The Fisz coefficients and the likelihood ratio Haar coefficients seem to be similarly close to the normal distribution, with the empirical skewness and kurtosis for  $f_{j,k}^{(i)}$  being 0.39 and 2.52 (respectively) and those for  $g_{j,k}^{(i)}$  being 0.35 and 2.53 (respectively). However, the likelihood ratio Haar coefficients achieve far better variance stabilization in this low-intensity example: we have  $\widehat{\text{Var}}(g_{j,k}^{(i)}) = 0.92$  versus  $\widehat{\text{Var}}(f_{j,k}^{(i)}) = 0.68$ .

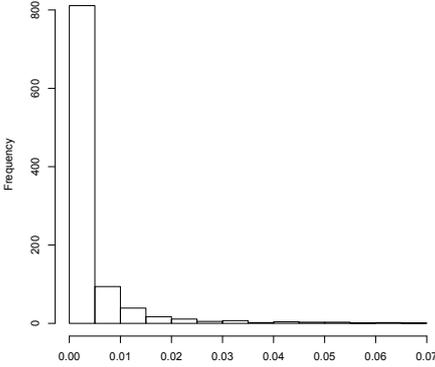


Figure 1: The Poisson case. Histogram of the empirical distribution of  $\{|g_{j,k}^{(i)}| - |f_{j,k}^{(i)}|\}_{i=1}^{1000}$  with  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ .

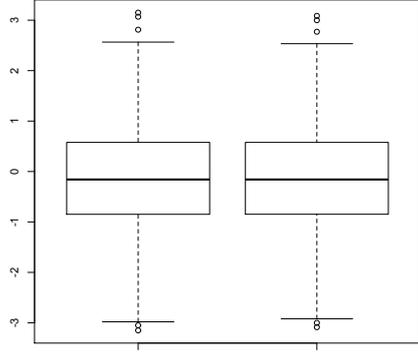


Figure 2: The Poisson case. Boxplots of the empirical distributions of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  (left) and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$  (right) with  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ .

We now turn to the chi-squared distribution. The Fisz coefficients for the  $\sigma^2\chi_1^2$  distribution are derived in Fryzlewicz et al. (2006), those for the exponential distribution ( $\sigma^22^{-1}\chi_2^2$ ) appear in Fryzlewicz et al. (2008) and the general case  $\sigma^2m^{-1}\chi_m^2$  is covered in Fryzlewicz (2008). In the general case of the  $\sigma^2m^{-1}\chi_m^2$  distribution, using the notation from the current

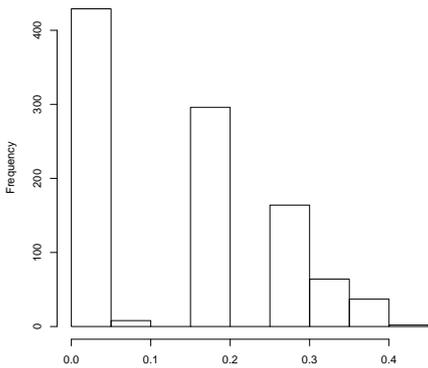


Figure 3: The Poisson case. Histogram of the empirical distribution of  $\{|g_{j,k}^{(i)}| - |f_{j,k}^{(i)}|\}_{i=1}^{1000}$  with  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ .

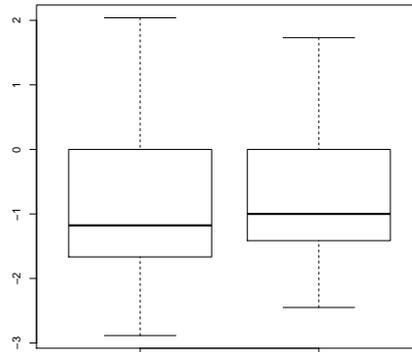


Figure 4: The Poisson case. Boxplots of the empirical distributions of  $\{|g_{j,k}^{(i)}|\}_{i=1}^{1000}$  (left) and  $\{|f_{j,k}^{(i)}|\}_{i=1}^{1000}$  (right) with  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ .

paper, the Fisz coefficients  $f_{j,k}$  are expressed as

$$f_{j,k} = 2^{\frac{j-3}{2}} m^{1/2} \frac{\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} - \bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}}{\bar{X}_{(k-1)2^j+1}^{k2^j}}. \quad (1)$$

As in the Poisson case, we obviously have  $\text{sign}(g_{j,k}) = \text{sign}(f_{j,k})$ . Lemma 3.2, used with  $f(u) = -\log u$  in the notation of that lemma, reduces to  $|g_{j,k}| \geq |f_{j,k}|$ . Moreover, by the same convexity argument as in the Poisson case,  $g_{j,k}$  and  $f_{j,k}$  will be closer to each other for larger values of  $\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}$  and  $\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}$ .

A major difference between the Poisson and the chi-square cases is that in the chi-square case,  $f_{j,k}$  is a compactly supported random variable (see formula (1)), whereas  $g_{j,k}$  is not. This difference does not apply in the Poisson case, in which neither  $f_{j,k}$  nor  $g_{j,k}$  are compactly supported. This has implications for how quickly  $f_{j,k}$  and  $g_{j,k}$  approach the normal distribution (with increasing  $j$  or  $m$ ) in the chi-square case, and we illustrate this numerically below.

As before, we simulate 1000 realisations of  $\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}$  and  $\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}$  and compute the corresponding 1000 realisations of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$ . We consider the following four cases.

- $m = 1$ ,  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ . In this case, the likelihood ratio Haar coefficients provide far better variance stabilization and normalization than the Fisz coefficients. For  $f_{j,k}^{(i)}$ , we have the following empirical values: variance 0.67, skewness 0.03, kurtosis 1.81. For  $g_{j,k}^{(i)}$ , we have variance 1.29, skewness 0.03, kurtosis 3.06. Figure 5 confirms the superiority of the likelihood ratio

Haar coefficients over the Fisz coefficients as regards their closeness to the normal distribution.

- $m = 1, j = 2, E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2, E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ . This low-sigma case differs from the previous one mainly in that both the likelihood ratio Haar coefficients and the Fisz coefficients are skewed to the right, although the Fisz coefficients (much) more so. For  $f_{j,k}^{(i)}$ , we have the following empirical values: variance 0.59, skewness 0.89, kurtosis 2.70. For  $g_{j,k}^{(i)}$ , we have variance 1.23, skewness 0.46, kurtosis 3.1. Figure 6 provides further visual evidence of the higher degree of symmetry in the likelihood ratio Haar coefficients and its closeness to the normal distribution.
- $m = 2, j = 2, E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10, E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ . As  $m$  increases, both the likelihood ratio Haar coefficients and the Fisz coefficients move closer towards variance-one normality, although again the likelihood ratio Haar coefficients beat Fisz. For  $f_{j,k}^{(i)}$ , we have the following empirical values: variance 0.81, skewness 0.05, kurtosis 2.19. For  $g_{j,k}^{(i)}$ , we have variance 1.16, skewness 0.03, kurtosis 2.97. Figure 7 shows both empirical distributions.
- $m = 2, j = 2, E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2, E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ . In this low-sigma case also, the likelihood ratio Haar coefficients appear to be far closer to variance-one normality than the Fisz coefficients. For  $f_{j,k}^{(i)}$ , we have the following empirical values: variance 0.57, skewness 1.15, kurtosis 4.08. For  $g_{j,k}^{(i)}$ , we have variance 1.04, skewness 0.45, kurtosis 3.64. Figure 8 shows both empirical distributions.

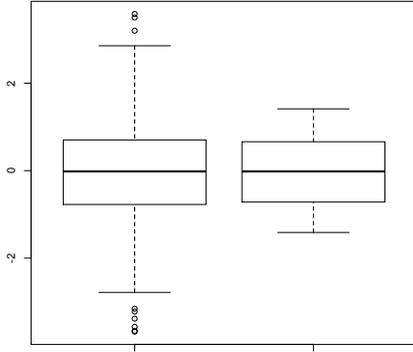


Figure 5: The chi-squared case. Boxplots of the empirical distributions of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  (left) and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$  (right) with  $m = 1, j = 2, E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10, E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ .

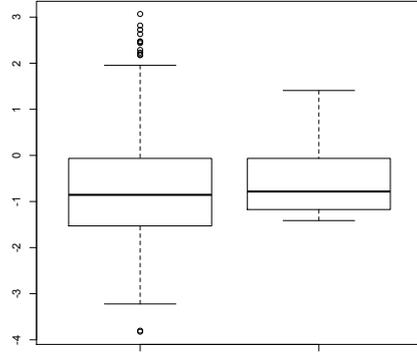


Figure 6: The chi-squared case. Boxplots of the empirical distributions of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  (left) and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$  (right) with  $m = 1, j = 2, E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2, E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ .

Overall, our empirical observations from the above (and other unreported) numerical exercises are as follows. For fine scales (i.e. those for which  $j$  is small) and for low degrees

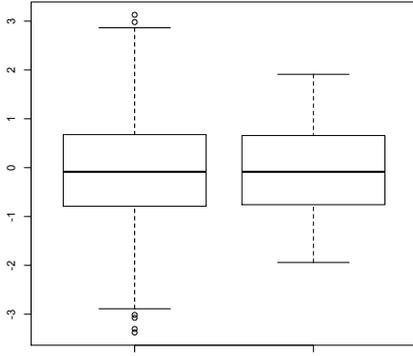


Figure 7: The chi-squared case. Boxplots of the empirical distributions of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  (left) and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$  (right) with  $m = 2$ ,  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 10$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 10.5$ .

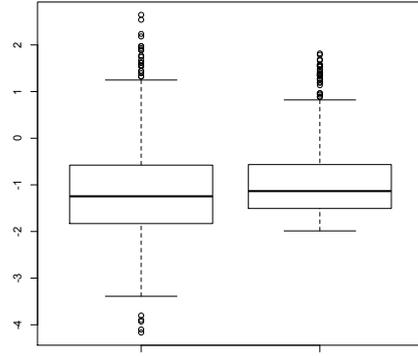


Figure 8: The chi-squared case. Boxplots of the empirical distributions of  $\{g_{j,k}^{(i)}\}_{i=1}^{1000}$  (left) and  $\{f_{j,k}^{(i)}\}_{i=1}^{1000}$  (right) with  $m = 2$ ,  $j = 2$ ,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}) = 0.2$ ,  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}) = 0.7$ .

of freedom  $m$ , the likelihood ratio Haar coefficients are much closer to a normal variable with variance one than the corresponding Fisz coefficients. From the properties of the chi-squared distribution, the effect of increasing  $j$  while keeping  $m$  constant is similar to the effect of increasing  $m$  while keeping  $j$  constant. As  $m$  or  $j$  increases, the likelihood ratio Haar coefficients appear to move closer to the normal distribution with variance one. However, for the same to happen with Fisz coefficients, the two means,  $E(\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}})$  and  $E(\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j})$ , need to be relatively close to each other. The latter phenomenon can also be observed in the Poisson case for increasing  $j$ . This is not unexpected as the results from Fisz (1955) suggest that the asymptotic normality with variance one arises when the two means approach each other asymptotically; no results are provided in Fisz (1955) on the case in which the two means diverge.

We end this section with an interesting interpretation of Lemmas 3.2 and 3.4 in the case of the Poisson distribution. Note that together, they imply

$$2^{j/2-1} \frac{|\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} - \bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}|}{\sqrt{\frac{2}{\frac{1}{\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}} + \frac{1}{\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}}}}} \geq |g_{j,k}| \geq 2^{j/2-1} \frac{|\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} - \bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}|}{\sqrt{\frac{1}{2} \left( \bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} + \bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j} \right)}}$$

on in other words, the magnitude of the likelihood ratio Haar coefficient is bounded from below by the magnitude of the corresponding Fisz coefficient and from above by the magnitude of a ‘‘Fisz-like’’ coefficient in which the arithmetic mean in the denominator has been replaced by the harmonic mean.

## 2 Invertibility of the likelihood Haar transform

Inverting the standard Haar transform proceeds by transforming each pair of coefficients  $(s_{j,k}, d_{j,k})$  into  $(s_{j-1,2k-1}, s_{j-1,2k})$ , hierarchically for  $j = J, \dots, 1$  (note that  $s_{0,k} = X_k$ ). Similarly, to demonstrate that the likelihood Haar transform is invertible, we need to show that it is possible to transform  $(s_{j,k}, g_{j,k})$  into  $(s_{j-1,2k-1}, s_{j-1,2k})$ .

We first show the invertibility of the Poisson likelihood ratio Haar transform. Denoting for brevity  $u = \bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}$ ,  $v = \bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}$  and ignoring some multiplicative constants and the square-root operation in  $g_{j,k}$ , which are irrelevant for invertibility, this amounts to showing that  $(u, v)$  can be uniquely determined from  $(u+v)/2$  and  $\text{sign}(u-v)\{u \log u + v \log v - (u+v) \log((u+v)/2)\}$ . The term  $\text{sign}(u-v)$  determines whether  $u \leq v$  or vice versa, so assume that  $u \leq v$  w.l.o.g. Denoting by  $a$  the known value of  $u+v$ , observe that the function  $(a-v) \log(a-v) + v \log v$  is strictly increasing for  $v \in [a/2, a]$ , which means that  $v$  can be determined uniquely and therefore that the Poisson likelihood ratio Haar transform is invertible.

We now show the invertibility of the chi-squared likelihood ratio Haar transform. We denote  $u = \bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}$ ,  $v = \bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}$  and ignore some multiplicative constants and the square-root operation in  $g_{j,k}$  which are irrelevant for invertibility. Assume that  $u \leq v$  w.l.o.g. Denoting by  $a$  the known value of  $u+v$ , observe that the function  $-\log(a-v) - \log v$  is strictly increasing for  $v \in [a/2, a)$ , which means that  $v$  can be determined uniquely and therefore that the chi-squared likelihood ratio Haar transform is invertible.

## 3 Technical results including proof of Theorem 4.1 from the main paper

**Lemma 3.1** *Let function  $f : [u, v] \rightarrow \mathbb{R}$  be such that  $f'$  is continuous on  $[u, v]$  and  $f''$  is continuous on  $(u, v)$ . There exists a point  $\xi \in (u, v)$  such that*

$$f(u) - 2f\left(\frac{u+v}{2}\right) + f(v) = \frac{(u-v)^2}{4} f''(\xi).$$

**Proof.** Let  $z = (u+v)/2$  and  $\delta = (v-u)/2$ , then

$$f(u) - 2f\left(\frac{u+v}{2}\right) + f(v) = f(z-\delta) - 2f(z) + f(z+\delta).$$

Defining  $g(x) = f(z-x) - 2f(z) + f(z+x)$ , Taylor's theorem yields

$$g(\delta) = g(0) + \delta g'(0) + \frac{\delta^2}{2} g''(\xi') = \frac{\delta^2}{2} g''(\xi') = \frac{\delta^2}{2} \{f''(z+\xi') + f''(z-\xi')\}, \quad (2)$$

where  $\xi' \in (0, \delta)$ . By the intermediate value theorem, there exists a  $\xi \in (z-\xi', z+\xi') \subset [u, v]$  such that  $\{f''(z+\xi') + f''(z-\xi')\}/2 = f''(\xi)$ , which by (2) completes the result.

**Lemma 3.2** *Let function  $f : [u, v] \rightarrow \mathbb{R}$  be such that  $f'$  is continuous on  $[u, v]$  and  $f''$  is convex on  $(u, v)$ . Then*

$$f(u) - 2f\left(\frac{u+v}{2}\right) + f(v) \geq \frac{(u-v)^2}{4} f''\left(\frac{u+v}{2}\right).$$

**Proof.** Straightforward from the convexity of  $f''$  and (2).

**Lemma 3.3** *Let function  $f : [u, v] \rightarrow \mathbb{R}$  be such that  $f'$  is continuous on  $[u, v]$  and  $f''$  is nonincreasing on  $[u, v]$ . Then*

$$f(u) - 2f\left(\frac{u+v}{2}\right) + f(v) \leq \frac{(u-v)^2}{8} \left\{ f''\left(\frac{u+v}{2}\right) + f''(u) \right\}.$$

**Proof.** Straightforward from (2) and the fact that  $f''$  is nonincreasing on  $[u, v]$ .

**Lemma 3.4** *Let function  $f : [u, v] \rightarrow \mathbb{R}$  be such that  $f'$  is continuous on  $[u, v]$  and  $f''$  is convex on  $[u, v]$ . Then*

$$f(u) - 2f\left(\frac{u+v}{2}\right) + f(v) \leq \frac{(u-v)^2}{8} \{f''(v) + f''(u)\}.$$

**Proof.** Straightforward from the convexity of  $f''$  and (2).

**Lemma 3.5** *The Poisson distribution satisfies Cramer's conditions.*

**Proof.** The Poisson distribution is log-concave, and Schudy and Sviridenko (2011), Lemma 7.4, show that all log-concave random variables  $Z$  are central moment bounded with real parameter  $L > 0$ , that is, satisfy for any integer  $i \geq 1$ ,

$$E|Z - E(Z)|^i \leq i L E|Z - E(Z)|^{i-1}.$$

Moreover, again by Schudy and Sviridenko (2011), Lemma 7.5, we have

$$L = 1 + \max(E(|Z - E(Z)| \mid Z \geq E(Z)), E(|Z - E(Z)| \mid Z < E(Z))),$$

which for the  $\text{Pois}(\lambda)$  distribution gives  $L = O(\lambda^{1/2})$ . But

$$\begin{aligned} E|Z - E(Z)|^i &\leq i L E|Z - E(Z)|^{i-1} \\ &\leq i! L^{i-2} E(Z - E(Z))^2, \end{aligned}$$

which completes the proof of the lemma.

**Proof of Theorem 4.1 from the main paper.**

We first show that  $P(\mathcal{A} \cap \mathcal{B}) \rightarrow 1$ . We have

$$P(\mathcal{A}^c) \leq \sum_{j=J_0+1}^J \sum_{k=1}^{2^{J-j}} P((\bar{\lambda}_{(k-1)2^j+1}^{k2^j})^{-1/2} |d_{j,k} - \mu_{j,k}| \geq t_1). \quad (3)$$

Since by Lemma 3.5, the Poisson distribution satisfies Cramer's conditions,  $\Lambda$  is bounded from above and away from zero, and  $2^{J_0} = O(n^\beta)$  for  $\beta \in (0, 1)$ , the strong asymptotic normality from the Corollary underneath the proof of Theorem 1 in Rudzki et al. (1978) can be used, which in our context implies that if  $t_1 = O(\log^{1/2} n)$ , then

$$P((\bar{\lambda}_{(k-1)2^j+1}^{k2^j})^{-1/2} |d_{j,k} - \mu_{j,k}| \geq t_1) \leq C\Phi(t_1), \quad (4)$$

where  $\Phi(\cdot)$  is the cdf of the standard normal distribution and  $C$  is a universal constant. Using (4), Mills' ratio inequality and the fact that  $t_1 = C_1 \log^{1/2} n$ , we bound (3) from above by  $\tilde{C} \log^{-1/2} n n^{1-\beta-C_1^2/2}$ , where  $\tilde{C}$  is a constant, which proves that  $P(\mathcal{A}) \rightarrow 1$ . The proof that  $P(\mathcal{B}) \rightarrow 1$  is identical.

We now turn to the estimator. Due to the orthonormality of the Haar transform, we have

$$n^{-1} \|\hat{\Lambda} - \Lambda\|^2 = n^{-1} \sum_{j=1}^J \sum_{k=1}^{2^{J-j}} (\hat{\mu}_{j,k} - \mu_{j,k})^2 + n^{-1} (s_{J,1} - \tilde{\lambda})^2, \quad (5)$$

where  $\tilde{\lambda} = n^{-1/2} \sum_{k=1}^n \lambda_k$ .

We first consider scales  $j = 1, \dots, J_0$ , for which  $\hat{\mu}_{j,k} = 0$ . At each scale  $j$ , there are at most  $N$  indices  $k$  for which  $\mu_{j,k} \neq 0$ . From the definition of  $d_{j,k}$ , for those  $\mu_{j,k}$ , we have  $\mu_{j,k} \leq 2^{j/2-1} \Lambda'$ , which gives

$$\sum_{j=1}^{J_0} \sum_{k=1}^{2^{J-j}} (\hat{\mu}_{j,k} - \mu_{j,k})^2 \leq N(\Lambda')^2 \sum_{j=1}^{J_0} 2^{j-2} = N(\Lambda')^2 (2^{J_0-1} - \frac{1}{2}). \quad (6)$$

We now consider the remaining scales  $j = J_0 + 1, \dots, J$  and first take an arbitrary index  $(j, k)$  for which  $\lambda_i$  is not constant for  $i = (k-1)2^j + 1, \dots, k2^j$ . For such a  $(j, k)$ , we have (using Lemma 3.2 in the second inequality)

$$\begin{aligned} (\hat{\mu}_{j,k} - \mu_{j,k})^2 &= (d_{j,k} \mathbb{I}(|g_{j,k}| > t) - \mu_{j,k})^2 \\ &\leq 2d_{j,k}^2 \mathbb{I}(|g_{j,k}| \leq t) + 2(d_{j,k} - \mu_{j,k})^2 \\ &\leq 2d_{j,k}^2 \mathbb{I}(|d_{j,k}| \leq t(\bar{X}_{(k-1)2^j+1}^{k2^j})^{1/2}) + 2(d_{j,k} - \mu_{j,k})^2 \\ &\leq 2t^2 \bar{X}_{(k-1)2^j+1}^{k2^j} + 2(d_{j,k} - \mu_{j,k})^2 \\ &\leq 2t^2 (\bar{\lambda}_{(k-1)2^j+1}^{k2^j} + t_2 2^{-j/2} (\bar{\lambda}_{(k-1)2^j+1}^{k2^j})^{1/2}) + 2t_1^2 \bar{\lambda}_{(k-1)2^j+1}^{k2^j}. \end{aligned}$$

Summing the bound over the at most  $N$  indices  $k$  within each scale for which  $\lambda_i$  is not constant for  $i = (k-1)2^j + 1, \dots, k2^j$ , as well as over scales  $j = J_0 + 1, \dots, J$ , and noting that  $\bar{\lambda}_{(k-1)2^j+1}^{k2^j} \leq \bar{\Lambda}$ , gives the upper bound of

$$2N\bar{\Lambda}^{1/2} \left\{ (J - J_0)(t^2 + t_1^2) \bar{\Lambda}^{1/2} + t^2 t_2 (1 + 2^{-1/2}) 2^{-\frac{J_0+1}{2}} \right\}. \quad (7)$$

We finally consider again the scales  $j = J_0 + 1, \dots, J$  and those indices  $(j, k)$  for which  $\lambda_i$  is constant for  $i = (k-1)2^j + 1, \dots, k2^j$ , which implies  $\mu_{j,k} = 0$ . For each such  $(j, k)$ , we have

$$(\hat{\mu}_{j,k})^2 = d_{j,k}^2 \mathbb{I}(|g_{j,k}| > t).$$

Consider the following sequence of inequalities, with the first one being implied by Lemma

3.4, and the second using the fact that  $\bar{\lambda}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} = \bar{\lambda}_{(k-1)2^j+2^{j-1}+1}^{k2^j} = \bar{\lambda}_{(k-1)2^j+1}^{k2^j}$ .

$$\begin{aligned}
|g_{j,k}| > t &\Rightarrow |d_{j,k}|2^{-1/2} \left| \frac{1}{\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}} + \frac{1}{\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j}} \right|^{1/2} > t \\
&\Rightarrow \frac{|d_{j,k}|}{(\bar{\lambda}_{(k-1)2^j+1}^{k2^j} - \delta)^{1/2}} > t \quad \vee \quad |\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} - \bar{\lambda}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}| \geq \delta \\
&\quad \vee \quad |\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j} - \bar{\lambda}_{(k-1)2^j+2^{j-1}+1}^{k2^j}| \geq \delta \\
&\Leftrightarrow \frac{|d_{j,k}|}{(\bar{\lambda}_{(k-1)2^j+1}^{k2^j})^{1/2}} > t \left( 1 - \frac{\delta}{\bar{\lambda}_{(k-1)2^j+1}^{k2^j}} \right)^{1/2} \\
&\quad \vee \quad 2^{j/2} (\bar{\lambda}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}})^{-1/2} |\bar{X}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}} - \bar{\lambda}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}}| \geq \delta 2^{j/2} (\bar{\lambda}_{(k-1)2^j+1}^{(k-1)2^j+2^{j-1}})^{-1/2} \\
&\quad \vee \quad 2^{j/2} (\bar{\lambda}_{(k-1)2^j+2^{j-1}+1}^{k2^j})^{-1/2} |\bar{X}_{(k-1)2^j+2^{j-1}+1}^{k2^j} - \bar{\lambda}_{(k-1)2^j+2^{j-1}+1}^{k2^j}| \geq \\
&\quad \delta 2^{j/2} (\bar{\lambda}_{(k-1)2^j+2^{j-1}+1}^{k2^j})^{-1/2}. \tag{8}
\end{aligned}$$

Let us set  $\delta = t_2 2^{-j/2} (\bar{\lambda}_{(k-1)2^j+1}^{k2^j})^{1/2}$ , then if

$$t_1 \leq t(1 - t_2 2^{-j/2} (\bar{\lambda}_{(k-1)2^j+1}^{k2^j})^{-1/2})^{1/2}, \tag{9}$$

then the right-hand side of the implication (8) is negated on  $\mathcal{A} \cap \mathcal{B}$ , which implies that so is the left-hand side, and therefore  $\hat{\mu}_{j,k} = 0$ . Note (9) is satisfied if (6) from the main paper holds.

Putting together (6) and (7) and noting that  $n^{-1}(s_{J,1} - \tilde{\lambda})^2 \leq n^{-1}t_1^2 \bar{\lambda}_1^n$  on  $\mathcal{A}$ , we bound (5) by

$$\frac{1}{2} n^{-1} N(\Lambda')^2 (n^\beta - 1) + 2n^{-1} N \bar{\Lambda}^{1/2} \left\{ (J - J_0)(t^2 + t_1^2) \bar{\Lambda}^{1/2} + t^2 t_2 (2 + 2^{1/2}) n^{-\beta/2} \right\} + n^{-1} t_1^2 \bar{\lambda}_1^n$$

on condition that (6) from the main paper holds, which completes the proof.

## References

- Fisz, M. (1955). The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3:138–146.
- Fryzlewicz, P. (2008). Data-driven wavelet-Fisz methodology for nonparametric function estimation. *Electronic Journal of Statistics*, 2:863–896.
- Fryzlewicz, P. and Nason, G. (2004). A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13:621–638.
- Fryzlewicz, P., Nason, G., and von Sachs, R. (2008). A wavelet-Fisz approach to spectrum estimation. *Journal of Time Series Analysis*, 29:868–880.
- Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93:687–704.

- Rudzkis, R., Saulis, L., and Statulevicius, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Math. J.*, 18:226–238.
- Schudy, W. and Sviridenko, M. (2011). Bernstein-like concentration and moment inequalities for polynomials of independent random variables: Multilinear case. *Preprint*.