

SKELLAMSHRINK: POISSON INTENSITY ESTIMATION FOR VECTOR-VALUED DATA

Keigo Hirakawa and Patrick J. Wolfe

Statistics and Information Sciences Laboratory
Harvard University, School of Engineering and Applied Sciences
Oxford Street, Cambridge, MA 02138 USA
{hirakawa, wolfe}@stat.harvard.edu

ABSTRACT

Owing to the stochastic nature of discrete processes such as photon counts in imaging, a variety of real-world data are well modeled as Poisson random variables whose means are in turn proportional to an underlying vector-valued signal of interest. Certain wavelet and filterbank transform coefficients corresponding to measurements of this type are distributed as sums and differences of Poisson counts, taking in the simplest case the so-called *Skellam distribution*. We show that a Skellam mean estimator provides a Poisson intensity estimation method based on shrinkage of filterbank coefficients, and a means of estimating the risk of any Skellam mean estimator is derived in closed form under a frequentist model.

Index Terms—Filterbank transforms, Poisson distribution, Skellam distribution, SkellamShrink, wavelets.

1. INTRODUCTION

Real-world sensing devices are subject to various types of measurement noise; for example, losses in resolution (e.g., quantization effects), randomness inherent in the signal of interest (e.g., photon or packet arrival), and variabilities in physical devices (e.g., thermal noise, electron leakage) can all contribute significantly to signal degradation. Estimation of a vector-valued signal $f \in \mathbb{R}^N$ given noisy observations $g \in \mathbb{R}^N$ therefore plays a prominent role in a variety of engineering applications such as signal processing, digital communications, and astronomical and biomedical imaging.

Motivated by both prior knowledge and empirical studies, statistical modeling of latent variables in linear transform domains has enjoyed tremendous popularity across these diverse applications—in particular, wavelets and other filterbank transforms provide convenient platforms; as is by now universally acknowledged, such classes of transform coefficients tend to exhibit temporal and spectral locality, sparsity, and energy compaction properties for a variety of natural data sets. In this setting, the special case of additive white Gaussian noise (AWGN) is by far the most studied scenario, as the distribution of transform coefficients is readily accessible when the likelihood function has a closed form in the transform domain. The twin assumptions of additivity and Gaussianity, however, are clearly inadequate for many genuine engineering applications; for instance, measurement noise is often dependent on the range space of the signal f , effects of which permeate across multiple transform coefficients and subbands [1]. As an important example, the number of electrons or photons g_i measured over the i th spatio-temporal integration region in a sensor array is well modeled as a Poisson random

variable $g_i \sim \mathcal{P}(f_i)$, where f_i is proportional to the electric current or the light intensity in that region. Recall that for $g_i \sim \mathcal{P}(f_i)$ we have that $\mathbb{E} g_i = \text{Var}(g_i) = f_i$, and so in the case at hand f_i is proportional to the expected electron or photon count per i th integration region, with the resultant “noise” in the form of variability being signal-dependent and heteroscedastic rather than homoscedastic (i.e., of constant variance).

Classical strategies overcome the problems associated with Poisson intensity estimation by exploiting the asymptotic behavior of the Poisson distribution. In an approach termed *variance stabilization*, one seeks an invertible nonlinear operator that (approximately) maps the heteroscedastic process realizations to the familiar additive Gaussian setting [2–8]. Based on techniques such as cross-validation, level-dependent shrinkage thresholds adjusted for Poisson distribution have also been proposed [9–11]. An alternative to the asymptotic approximation approach is to work with the heteroscedastic noise directly, either by leveraging the independence of Poisson variables [12–14] or explicitly encoding the dependencies between Poisson variables in the context of Haar frames [15, 16].

In this paper, we treat Poisson intensity estimation for vector-valued data in the Haar wavelet and filterbank transform domains directly by way of the *Skellam distribution* [17], whose use to date appears largely limited [18–20]. Our approach to combining Skellam distribution with the Haar transform domain differs from the prior art in that we derive level-dependent optimal shrinkage based on an exact rather than asymptotic analysis. After briefly reviewing wavelet and filterbank coefficient models in Section 2, we consider in Section 3 a means of estimating the risk of shrinkage operators under the Skellam distribution. Simulation studies presented in Section 4 verify the effectiveness of our approach, and a brief discussion of its relationship to SUREShrink [21] and the Haar-Fisz algorithm [8] is provided in Section 5. Though optimal soft-threshold shrinkage is given as an illustrative example in this paper (“SkellamShrink”), the reader is reminded that risk estimate results apply to *any* Haar-based Poisson intensity estimation methods.

2. WAVELET AND FILTERBANK COEFFICIENT MODELS

2.1. Haar wavelet and filterbank transforms

Below, we review the requisite notions of discrete wavelet and filterbank transforms—complete details can be found in [22]. In the Haar transform, the sequence $\{f_j\} \in \ell^2(\mathbb{Z})$ is represented in terms of a recursive relationship as follows:

$$\begin{cases} s_{k,i} = s_{k-1,2i} + s_{k-1,2i+1} \\ x_{k,i} = s_{k-1,2i} - s_{k-1,2i+1}, \end{cases} \quad (1)$$

Based upon work supported in part by the National Science Foundation under Grant No. DMS-0652743.

where $s_{0,i} := f_i$. In the formulation of (1), each sequence $\{s_{k-1,i}\}_i$ is decomposed into low-pass and high-pass components $\{s_{k,i}, x_{k,i}\}_i$ in turn. A recursive application of the map $\{s_{k-1,i}\} \mapsto \{s_{k,i}, x_{k,i}\}$ yields the Haar wavelet transform (HWT), whereas the same transform applied to high-pass component $x_{k-1,i}$ further decomposes it into narrower bands. Recursive decomposition of *both* low-pass and high-pass sequences in this way yields the Hadamard transform, otherwise known as the Haar filterbank transform (HFT). The computational requirements of these transforms makes them attractive alternatives to other joint time-frequency analysis techniques possessing better frequency localization. The Haar transforms enjoy orthonormality, compact spatial support, and computational simplicity, with the HWT satisfying the axioms of a multiresolution analysis. We later demonstrate how their simplicity serves to admit analytical tractability that in turn enables efficient inference and estimation procedures.

We omit subband index k in the sequel, as wavelet coefficients $x_{k,i}$ are always aggregated within a given scale 2^k . Notation such as $\sum_i x_i$ refers to summation over i within a single subband.

2.2. The Skellam distribution

Let $W \in \{0, \pm 1\}^{N \times N}$ denote the HWT or HFT matrix as specified in Section 2, where components of $x := Wf$ are sums and differences of elements of f . To this end, define $x_i = x_i^+ - x_i^-$:

$$x_i^+ = \sum_{j: [W]_{i,j}=1} f_j, \quad x_i^- = \sum_{j: [W]_{i,j}=-1} f_j. \quad (2)$$

Defined similarly with respect to the observed data $g_i \sim \mathcal{P}(f_i)$ and its transform $y := Wg$, each observed transform coefficient $y_i = y_i^+ - y_i^-$ is a difference of two independent Poisson counts

$$y_i^+ \sim \mathcal{P}(x_i^+), \quad y_i^- \sim \mathcal{P}(x_i^-),$$

taking the so-called *Skellam distribution* [17] with parameters x_i^+, x_i^- :

$$\begin{aligned} \Pr(Y_i = y_i; x_i^+, x_i^-) &= e^{-(x_i^+ + x_i^-)} \sum_{n \geq \max(0, -y_i)} \frac{(x_i^+)^{y_i+n} (x_i^-)^n}{(y_i+n)! n!} \\ &= e^{-(x_i^+ + x_i^-)} \left(\frac{x_i^+}{x_i^-} \right)^{\frac{y_i}{2}} I_{y_i} \left(2\sqrt{x_i^+ x_i^-} \right), \end{aligned}$$

with $I_{y_i}(\cdot)$ the y_i th-order modified Bessel function of the first kind. Defining the scaling coefficient $s_i = x_i^+ + x_i^-$, we may write the Skellam likelihood in terms of transform coefficients as follows:

$$\begin{aligned} \phi(y_i; x_i, s_i) &:= \Pr(Y_i = y_i; x_i, s_i) \\ &= e^{-s_i} \left(\frac{s_i + x_i}{s_i - x_i} \right)^{\frac{y_i}{2}} I_{y_i} \left(\sqrt{s_i^2 - x_i^2} \right), \quad (3) \end{aligned}$$

where the above is undefined if $|x_i| \geq s_i$. Figure 1 confirms that $\phi(y_i; x_i, s_i)$ tends toward a Normal as s_i increases.

Defining the scaling coefficient $t_i = y_i^+ + y_i^-$, where $t_i \sim \mathcal{P}(s_i)$, we observe the following properties of Skellam random variables.

Lemma 2.1. *Let $y_i \sim \text{Skellam}(x_i, s_i)$, with x_i and s_i unknown, and $t_i = y_i^+ + y_i^-$ as defined above. Then we have that*

$$\begin{aligned} \mathbb{E} Y_i &= x_i \\ \mathbb{E}(Y_i - x_i)^2 &= s_i \\ \mathbb{E}[T_i \pm Y_i | Y_i = y_i] &= (s_i \pm x_i) \frac{\phi(y_i \mp 1; x_i, s_i)}{\phi(y_i; x_i, s_i)}. \quad (4) \end{aligned}$$

Proof. See Appendix A. \square

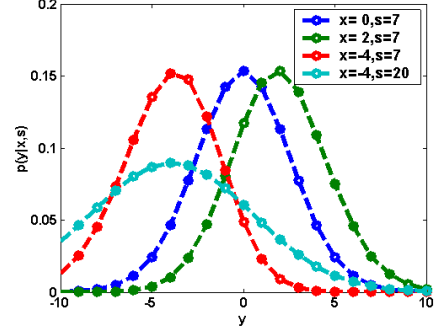


Fig. 1. Illustrations of Skellam probability mass functions

3. RISK ESTIMATES FOR SHRINKAGE OPERATORS

The main conclusion of Section 2 is that the observed wavelet or filterbank coefficient y_i corresponding to the ‘‘Poisson-corrupted’’ vector-valued data is the ‘‘Skellam-corrupted’’ version of an ideal coefficient x_i . Thus the Poisson intensity estimation problem is equivalent to estimating the Skellam mean in the transform domain. The main result of this section is the following theorem, which yields a procedure for unbiased ℓ^2 risk estimation in the context of shrinkage operators. An example using soft-threshold operator follows.

Theorem 3.1. *Let $y_i \sim \text{Skellam}(x_i, s_i)$, with x_i, s_i unknown, and $t_i = y_i^+ + y_i^-$. Fix an estimator $\hat{x}(Y) = Y + \theta(Y)$ where $\theta : \mathbb{Z}^N \rightarrow \mathbb{R}$. Then the resultant risk $\mathbb{E} \|\hat{x}(Y) - x\|_2^2$ may be formulated as*

$$\mathbb{E} [\|T\|_1 + \|\theta(Y)\|_2^2 + 2Y^T \theta(Y) - (T+Y)^T \theta(Y - \vec{1}) + (T-Y)^T \theta(Y + \vec{1})],$$

where $\vec{1} = [1, \dots, 1]^T$, with

$$R(y) = \|t\|_1 + \|\theta(y)\|_2^2 + 2y^T \theta(y) - (t+y)^T \theta(y - \vec{1}) + (t-y)^T \theta(y + \vec{1})$$

an unbiased estimate thereof.

Proof. See Appendix A. \square

Estimators of *any* parametric form may be optimized by evaluating the above expectation over the observed vectors y and t . For example, let $\hat{x}(Y; \tau)$ denote a soft-thresholding operator [23]:

$$\hat{x}_i(Y_i; \tau) := \text{sgn}(Y_i)(|Y_i| - \tau)_+. \quad (5)$$

Writing $\hat{x}(Y) = Y + \theta(Y; \tau)$, we have:

$$\theta(Y_i; \tau) = \begin{cases} -\text{sgn}(Y_i) \tau & \text{if } |Y_i| \geq \tau \\ -Y_i & \text{if } |Y_i| < \tau. \end{cases} \quad (6)$$

The optimal threshold for the shrinkage estimator of this form is the minimizer of the ‘‘SkellamShrink’’ risk estimate

$$\sum_i \text{sgn}(|y_i| - \tau) t_i + \sum_i \min(y_i^2, \tau^2) - \tau \#\{i : |y_i| = \tau\}. \quad (7)$$

This estimate is derived from Theorem 3.1 by evaluating the risk with respect to (6) under three separate scenarios:

- Suppose $|Y_i| < \tau$. Then

$$\begin{aligned} R(y_i; \tau) &= t_i + y_i^2 - 2y_i^2 - (t_i + y_i)(-y_i + 1) + (t_i - y_i)(-y_i - 1) \\ &= -t_i + y_i^2; \quad (8) \end{aligned}$$

- Suppose $|Y_i| > \tau$. Then

$$\begin{aligned} R(y_i; \tau) &= t_i + \tau^2 - [2y_i + (t_i + y_i) - (t_i - y_i)](\text{sgn}(y_i)\tau) \\ &= t_i + \tau^2; \end{aligned} \quad (9)$$

- Suppose $|Y_i| = \tau$. Then (e.g., for $Y_i = \tau$)

$$\begin{aligned} R(y_i; \tau) &= t_i + \tau^2 - 2\tau^2 + (t_i + \tau)(\tau - 1) - (t_i - \tau)\tau \\ &= \tau^2 - \tau. \end{aligned} \quad (10)$$

The risk estimate of (7) then follows from combining (8)–(10).

4. EXPERIMENTAL RESULTS

A variety of previous Poisson intensity estimation methods have been evaluated using a standard set one-dimensional wavelet test functions—intended to cover a wide range of signal classes and applications, including image signals, network traffic data, and Gamma-ray burst. One hundred realizations of Poisson counts $g_i \sim \mathcal{P}(f_i)$ were generated synthetically, and reconstruction errors are reported in Figure 2 for SkellamShrink, the Poisson-adjusted thresholding methods in [9, 11] ([10] behaves similarly to [11]), and the combination of variance stabilization and SUREShrink [4, 23]. For a full comparison of these approaches in the Poisson setting to other methods, the reader is referred to [24]. All methods were implemented using a five-level translation-invariant wavelet decomposition.

As can be seen from Figure 2, SkellamShrink measures well against all alternatives despite the diversity in the underlying test functions considered. Overall, performance is comparable to the soft-thresholding method in [9]—the best among these alternatives. Indeed, SkellamShrink consistently outperformed the combination of variance stabilization and SUREShrink, despite the similarities in estimator structure and formulation described below.

5. RELATION OF SKELLAMSHRINK TO SURESHRINK

The risk estimate of Theorem 3.1 is remarkably similar to the that of the SUREShrink approach [23] derived in the additive white Gaussian noise setting:

$$\sum_i \text{sgn}(|y_i| - \tau)\sigma^2 + \sum_i \min(y_i^2, \tau^2). \quad (11)$$

Interpreting T_i in Theorem 3.1 as a proxy for the variance of y_i —analogous to σ^2 —connections to the additive white Gaussian noise regime are obvious. However, it follows that the risk estimate equivalent to Theorem 3.1 after scaling y_i by $t_i^{-1/2}$ —a proxy for the standard deviation of y_i [8]—is

$$\sum_i \text{sgn}(|y_i| - \tau) + \sum_i \min\left(\frac{y_i^2}{t_i}, \frac{\tau^2}{t_i}\right) - \frac{\tau}{\sqrt{t_i}} \#\{i : |y_i| = \tau\},$$

where y_i^2/t_i represents the stabilized variance. The most significant difference between this risk and (11) is the appearance of τ^2/t_i in the second term, pointing to the fact that “optimality” in Haar-Fisz variance stabilization domain can be a poor approximation.

In summary, we have presented here a technique for Poisson intensity estimation using the Skellam distribution in conjunction with Haar wavelet/filterbank transforms. We derived a means to estimating the risk of any Skellam mean estimator under shrinkage rules and frequentist model. Application of the proposed method to a soft-thresholding operator confirms that this approach offers robust alternative to existing methods with a substantial improvement in some cases.

A. PROOFS OF THE LEMMA AND THE THEOREM

Proof of Lemma 2.1. The first two properties are obvious; to obtain the third property of (4) we apply Bayes’ rule as follows:

$$\begin{aligned} \mathbb{E}[T_i \pm Y_i | Y_i = y_i] &= \sum_{t_i \in 2\mathbb{Z}_+ + |y_i|} (t_i \pm y_i) p(t_i | y_i, x_i, s_i) \\ &= \sum_{t_i \in 2\mathbb{Z}_+ + |y_i|} (t_i \pm y_i) \frac{p(t_i, y_i; x_i, s_i)}{\phi(y_i; x_i, s_i)} \\ &= 2\phi(y_i; x_i, s_i)^{-1} \sum_{t_i > 0} \frac{e^{-s_i} \binom{s_i \pm x_i}{2} t_i + \frac{|y_i| \pm y_i}{2} \binom{s_i \mp x_i}{2} t_i + \frac{|y_i| \mp y_i}{2}}{(t_i + \frac{|y_i| \pm y_i}{2} - 1)! (t_i + \frac{|y_i| \mp y_i}{2})!}. \end{aligned}$$

If $|y_i| \mp y_i = 0$, then by definition of the Skellam distribution,

$$\begin{aligned} 2\phi(y_i; x_i, s_i)^{-1} \sum_{t_i > 0} \frac{e^{-s_i} \binom{s_i \pm x_i}{2} t_i \pm y_i \binom{s_i \mp x_i}{2} t_i}{(t_i \pm y_i - 1)! t_i!} \\ = (s_i \pm x_i) \frac{\phi(\pm y_i - 1 | \pm x_i, s_i)}{\phi(y_i; x_i, s_i)} = (s_i \pm x_i) \frac{\phi(y_i \mp 1; x_i, s_i)}{\phi(y_i; x_i, s_i)}. \end{aligned}$$

The same conclusion holds for the case $|y_i| \pm y_i = 0$. \square

Proof of Theorem 3.1. Rewriting the ℓ^2 risk, we obtain

$$\mathbb{E} \|Y - x\|_2^2 + 2 \mathbb{E}(Y - x)^T \theta(Y) + \mathbb{E} \|\theta(Y)\|_2^2.$$

We substitute $\mathbb{E}(Y_i - x_i)^2 = s_i = \mathbb{E} T_i$ from Lemma 2.1 and expand the second term as follows:

$$\begin{aligned} \mathbb{E} \left[\mathbb{E}[(Y_i - x_i)^T \theta(Y_i) | Y_i] \right] &= \mathbb{E} \left[\mathbb{E} \left[Y_i - \frac{s_i + x_i}{2} + \frac{s_i - x_i}{2} | Y_i \right]^T \theta(Y_i) \right] \\ &= \mathbb{E} \left[Y_i^T \theta(Y_i) - \frac{(Y_i + T_i)^T}{2} \theta(Y_i - \vec{1}) + \frac{(Y_i - T_i)^T}{2} \theta(Y_i + \vec{1}) \right], \end{aligned}$$

by way of the following equality, based on (4):

$$\begin{aligned} \mathbb{E} \left[\mathbb{E}[(Y_i \pm T_i)^T \theta(Y_i \mp \vec{1}) | Y_i] \right] &= \mathbb{E} \left[(s_i \pm x_i) \frac{\phi(y_i \mp \vec{1}; x_i, s_i)}{\phi(y_i; x_i, s_i)} \theta(Y_i \mp \vec{1}) \right] \\ &= \sum_{y_i} \left[(s_i \pm x_i) \frac{\phi(y_i \mp \vec{1}; x_i, s_i)}{\phi(y_i; x_i, s_i)} \theta(Y_i \mp \vec{1}) \right] \phi(y_i; x_i, s_i) \\ &= \sum_{y_i} \left[(s_i \pm x_i) \phi(y_i \mp \vec{1}; x_i, s_i) \theta(Y_i \mp \vec{1}) \right] = \mathbb{E}[(s_i \pm x_i) \theta(Y_i)]. \end{aligned}$$

\square

B. REFERENCES

- [1] K. Hirakawa, “Fourier and filterbank analysis of signal-dependent noise,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 3517–3520.
- [2] M. S. Bartlett, “The square root transformation in analysis of variance,” *J. R. Statist. Soc. (Suppl.)*, vol. 3, pp. 68–78, 1936.
- [3] M. Fisz, “The limiting distribution of a function of two independent random variables and its statistical application,” *Colloq. Mathemat.*, vol. 3, pp. 138–146, 1955.
- [4] F. J. Anscombe, “The transformation of Poisson, Binomial and Negative Binomial data,” *Biometrika*, vol. 35, pp. 246–254, 1948.
- [5] M. F. Freeman and J. W. Tukey, “Transformations related to the angular and the square root,” *Ann. Math. Statist.*, vol. 21, pp. 607–611, 1950.

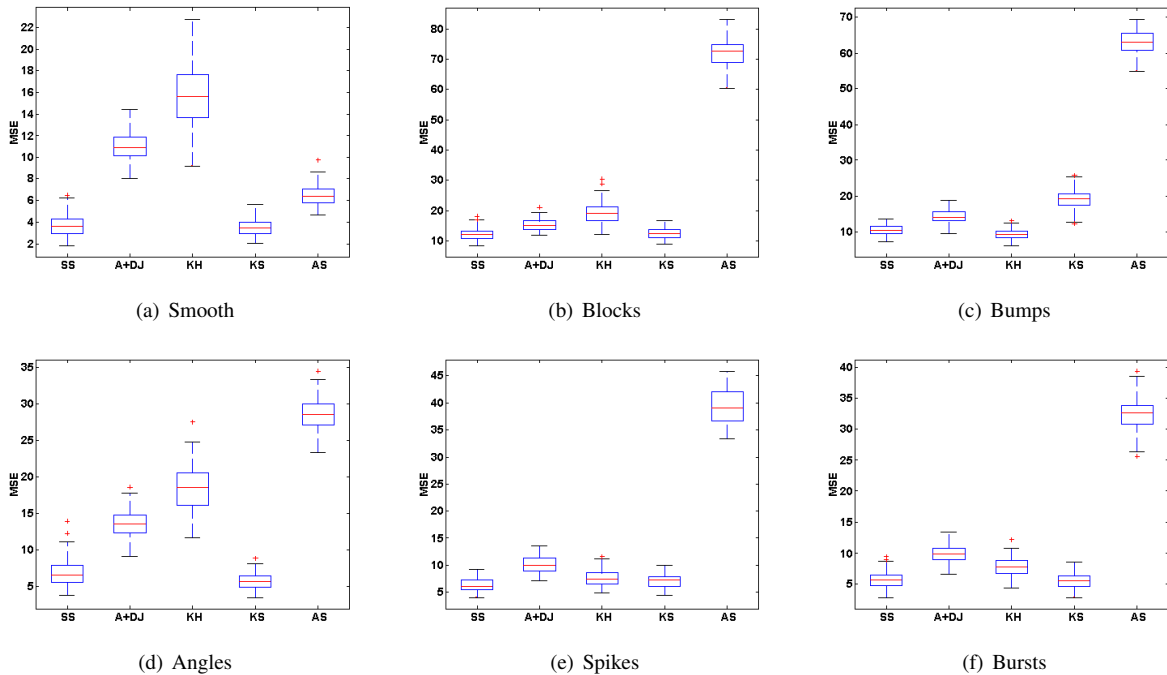


Fig. 2. Mean-squared error (averaged over 100 trials) for the prototype functions in [24]. Here, SS=SkellamShrink; A+DJ=SUREShrink [23] with variance stabilization [4]; KH/KS=corrected threshold (hard/soft) [9]; and AS=asymptotic minimax [11]

- [6] A. Veevers and M. C. K. Tweedie, "Variance-stabilizing transformation of a Poisson variate by a Beta function," *Appl. Statist.*, vol. 20, pp. 304–308, 1971.
- [7] M. C. K. Tweedie and A. Veevers, "The inversion of cumulant operators for power-series distributions, and the approximate stabilization of variance by transformations," *J. Am. Statist. Ass.*, vol. 63, pp. 321–328, 1968.
- [8] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *J. Computat. Graph. Statist.*, vol. 13, pp. 621–638, 2004.
- [9] E. D. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statist. Sinica*, vol. 9, pp. 119–135, 1999.
- [10] R. D. Nowak and R. G. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Trans. Image Process.*, vol. 8, pp. 666–678, 1999.
- [11] A. Antoniadis and T. Sapatinas, "Wavelet shrinkage for natural exponential families with quadratic variance functions," *Biometrika*, vol. 88, pp. 805–820, 2001.
- [12] M. L. Clevenston and J. V. Zidek, "Simultaneous estimation of the means of independent Poisson laws," *J. Am. Statist. Ass.*, vol. 70, pp. 698–705, 1975.
- [13] J. T. Hwang, "Improving upon standard estimators in discrete exponential families with applications to Poisson and Negative Binomial cases," *Ann. Statist.*, vol. 10, pp. 857–867, 1982.
- [14] M. Raphan and E. P. Simoncelli, "Learning to be Bayesian without supervision," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds., pp. 1145–1152. MIT Press, Cambridge, 2007.
- [15] K. E. Timmermann and R. D. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging," *IEEE Trans. Info. Theory*, vol. 45, pp. 846–862, 1999.
- [16] E. D. Kolaczyk, "Bayesian multiscale models for Poisson processes," *J. Am. Statist. Ass.*, vol. 94, pp. 920–921, 1999.
- [17] J. G. Skellam, "The frequency distribution of the difference between two Poisson variates belonging to different populations," *J. R. Statist. Soc.*, vol. 109, pp. 296, 1946.
- [18] D. Karlis and I. Ntzoufras, "Analysis of sports data using bivariate Poisson models," *J. R. Statist. Soc. D*, vol. 52, pp. 381–393, 2003.
- [19] Y. Hwang, I.-S. Kweon, and J.-S. Kim, "Color edge detection using the Skellam distribution as a sensor noise model," *Proc. Soc. Indust. Control Eng. Annual Conf.*, pp. 1972–1979, 2007.
- [20] B. Zhang, M. J. Fadili, J. L. Starck, and S. W. Digei, "Fast Poisson noise removal by biorthogonal Haar domain hypothesis testing," *Statist. Methodol.*, vol. 5, pp. 387–396, 2008.
- [21] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [22] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, second edition, 1999.
- [23] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Am. Statist. Ass.*, vol. 90, pp. 1200–1224, 1995.
- [24] P. Besbeas, I. De Feis, and T. Sapatinas, "A comparative simulation study of wavelet shrinkage estimators for Poisson counts," *Int. Statist. Rev.*, vol. 72, pp. 209–237, 2004.