Noname manuscript No.
(will be inserted by the editor)

Optimization of variance-stabilizing transformations

Alessandro Foi

Received: date / Accepted: date

Abstract Variance-stabilizing transformations are commonly exploited in order to make non-homoskedastic data easily tractable by standard methods. However, for the most common families of distributions (e.g., binomial, Poisson, etc.) exact stabilization is not possible and even achieving some approximate stabilization turns out to be rather challenging. We approach the variance stabilization problem as an explicit optimization problem and propose recursive procedures to minimize a nonlinear stabilization functional that measures the discrepancy between the standard deviation of the transformed variables and a fixed desired constant. Further, we relax the typical requirement of monotonicity of the transformation and introduce optimized nonmonotone stabilizers which are nevertheless invertible in terms of expectations. We demonstrate a number of optimized variance-stabilizing transformations for the most common distribution families. These stabilizers are shown to outperform the existing ones. In particular, optimized variance-stabilizing transformations for low-count Poisson, binomial, and negative-binomial data are presented.

Keywords variance stabilization · heteroskedasticity · transformations

1 Introduction

Let $z \in Z \subseteq \mathbb{R}$ be a random variable distributed according to a one-parameter family of distributions $\boldsymbol{\mathcal{D}} = \{\mathcal{D}_\theta\}$, where $\theta \in \Theta \subseteq \mathbb{R}$ denotes the parameter, and let $\mu(\theta) = E\{z|\theta\}$ and $\sigma(\theta) = \mathrm{std}\{z|\theta\}$ be the conditional expectation and the conditional standard deviation of z given as functions of the parameter θ. For example, $\boldsymbol{\mathcal{D}}$ can be the family of Poisson distributions with mean $\theta \in \Theta = [0, +\infty)$, $\Pr[z = \zeta|\theta] = e^{-\theta}\frac{\theta^\zeta}{\zeta!}$, $\zeta \in \mathbb{N}$; in this case $\mu(\theta) = \theta$ and $\sigma(\theta) = \sqrt{\theta}$. Other important examples are given in Table 1.

We are concerned about the following problem: to find a function $f : Z \to \mathbb{R}$ such that the transformed variable $f(z)$ has constant conditional standard deviation, say, equal to $c > 0$, $\sigma_f(\theta) = \mathrm{std}\{f(z)|\theta\} = c$. In other words, the sought f is a *variance-stabilizing transformation* that transforms the variable z in such a way that the conditional standard deviation does not depend anymore on the distribution parameter, thus turning a heteroskedastic z into a homoskedastic $f(z)$, and, at least for what concerns the variances, f transforms a signal with signal-dependent noise into one where the noise is signal-independent. Using the words of Curtiss (1943), $\sigma_f(\theta)$ becomes *functionally independent* of θ. Of course, f itself should be independent of θ and we are not interested in pathological solutions such as having f identically constant, since often one eventually wants to be able to estimate θ from $f(z)$. Typically, one would require f to be monotone strictly increasing.

The problem of finding variance-stabilizing transformations is widely studied because of their practical usefulness. Starting from the early 1900's, numerous publications have appeared in the mathematical and especially in the applied statistics and engineering literature, where variance-stabilization plays a central role in making non-homoskedastic data easily tractable by standard methods. The fact that for the most common families of distributions (e.g., binomial, Poisson (Curtiss 1943), etc.) exact stabilization is not possible[1] shifted the attention towards finding transformations

A. Foi
Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33720, Tampere, Finland
E-mail: alessandro.foi@tut.fi

[1] For example, the binary samples $z \in \{0,1\} = Z$ of the Bernoulli distribution with parameter $\theta = E\{z|\theta\}$ cannot be stabilized to the same constant variance for different values of θ: indeed, elementary calculations show that $E\{f(z)|\theta\} = \theta f(1) + (1-\theta) f(0)$ and $\mathrm{var}\{f(z)|\theta\} = E\left\{(f(z) - E\{f(z)|\theta\})^2|\theta\right\} =$

that lead at least to some approximate (or asymptotic) stabilization. However, even approximate stabilization turns out to be rather challenging and improved transformations and different approaches are still being proposed nowadays.

1.1 Background

Let us give a brief overview of the existing approaches to the variance-stabilization problem and related techniques. We begin with the classic heuristic stabilizer having the simple indefinite integral form[2]

$$f(z) = \int^{z} \frac{c}{\sigma(\theta)} d\mu(\theta). \tag{1}$$

This expression appears in many early works from the 1930's (e.g., Tippet 1934; Bartlett 1936) but, because of its utmost simplicity, it had presumably been considered and used long before those years. Likewise, it has regularly resurfaced in numerous engineering papers, particularly in the field of signal processing (e.g., Prucnal and Saleh 1981; Arsenault and Denis 1981; Kasturi et al 1983; Hirakawa and Parks 2006; Foi 2009a). It is important to emphasize that although (1) does not follow from a rigorous mathematical derivation, it nevertheless enjoys good asymptotic properties (Curtiss 1943) and was empirically shown to provide reasonable stabilization in various applications, as confirmed by its extensive use. More important, it served as the starting point in suggesting the parametric forms of many classical stabilizers which were later improved through analytical study.

While (1) pretends to be a universal recipe for stabilizing an arbitrary family of distributions, traditionally, most of the efforts and contributions have been targeted at the stabilization of few particular families of distributions, with the Poisson distribution standing out as the most researched one. Let us mention the works of Bartlett (1936), Anscombe (1948), Freeman and Tukey (1950), and Veevers and Tweedie (1971). As matter of fact, the best known variance-stabilizing transformation is actually Anscombe's root transformation (Anscombe 1948) of Poisson data (Anscombe

$(f(0) - f(1))^2\,\theta(1-\theta)$. The last term is constant if and only if $f(z)$ is constant.

[2] The expression (1) follows from considering a local first-order expansion of f at $\mu(\theta)$,

$$f(z) \simeq f(\mu(\theta)) + (z - \mu(\theta))\frac{\partial f}{\partial z}(\mu(\theta)).$$

One can then derive (treating z as distributed according to an impulse centered at $\mu(\theta)$, so-called "delta-method" (Greene 2000))

$$\text{std}\{f(z)|\theta\} \simeq \frac{\partial f}{\partial z}(\mu(\theta))\,\sigma(\theta),$$

and thus, by imposing $\text{std}\{f(z)|\theta\} = c$, obtain the indefinite integral (1).

Table 1 Examples of one-parameter families of distributions.

Family	$\mathcal{D}_\theta$	$\mu(\theta)$	$\sigma(\theta)$
Poisson (unscaled, mean=variance)	$\Pr[z=\zeta\|\theta] = e^{-\theta}\frac{\theta^\zeta}{\zeta!},\ \zeta\in\mathbb{N},\ \theta\in[0,+\infty)$	θ	$\sqrt{\theta}=\sqrt{\mu(\theta)}$
Scaled Poisson (scale $\chi>0$)	$\Pr\left[z=\frac{\zeta}{\chi}\|\theta\right] = e^{-\theta}\frac{\theta^\zeta}{\zeta!},\ \zeta\in\mathbb{N},\ \theta\in[0,+\infty)$	$\frac{\theta}{\chi}$	$\frac{\sqrt{\theta}}{\chi}=\sqrt{\frac{\mu(\theta)}{\chi}}$
Binomial (n trials)	$\Pr[z=\zeta\|\theta] = \binom{n}{\zeta}\theta^\zeta(1-\theta)^{n-\zeta},\ \zeta\in\{0,\dots,n\},\ \theta\in[0,1]$	$n\theta$	$\sqrt{n\theta(1-\theta)}=\sqrt{\frac{\mu(\theta)(n-\mu(\theta))}{n}}$
Scaled binomial (n trials, scale n)	$\Pr\left[z=\frac{\zeta}{n}\|\theta\right] = \binom{n}{\zeta}\theta^\zeta(1-\theta)^{n-\zeta},\ \zeta\in\{0,\dots,n\},\ \theta\in[0,1]$	θ	$\sqrt{\frac{\theta(1-\theta)}{n}}=\sqrt{\frac{\mu(\theta)(1-\mu(\theta))}{n}}$
Negative binomial (exponent λ)	$\Pr[z=\zeta\|\theta] = \frac{\Gamma(\zeta+\lambda)}{\zeta!\Gamma(\lambda)}\left(\frac{\theta}{\theta+\lambda}\right)^\zeta\left(\frac{\lambda+\theta}{\lambda}\right)^{-\lambda},\ \zeta\in\mathbb{N},\ \theta\in[0,+\infty)$	θ	$\sqrt{\frac{\theta(\theta+k)}{k}}=\sqrt{\frac{\mu(\theta)(\mu(\theta)+\lambda)}{k}}$
Scaled negative binomial (exponent λ, scale $\chi>0$)	$\Pr\left[z=\frac{\zeta}{\chi}\|\theta\right] = \frac{\Gamma(\zeta+\lambda)}{\zeta!\Gamma(\lambda)}\left(\frac{\theta}{\theta+\lambda}\right)^\zeta\left(\frac{\lambda+\theta}{\lambda}\right)^{-\lambda},\ \zeta\in\mathbb{N},\ \theta\in[0,+\infty)$	$\frac{\theta}{\chi}$	$\sqrt{\frac{\theta(\theta+\lambda)}{\lambda\chi^2}}=\sqrt{\frac{\mu(\theta)(\mu(\theta)\chi+\lambda)}{\lambda\chi}}$
Multiplicative normal (scale $\chi>0$)	$\text{pdf}[z\|\theta](\zeta) = \frac{\chi}{\theta\sqrt{2\pi}}e^{-\frac{(\zeta-\theta)^2\chi^2}{2\theta^2}}$	θ	$\frac{\theta}{\chi}=\frac{\mu(\theta)}{\chi}$
Doubly censored normal with standard-deviation $\sigma(\theta)$	$\text{pdf}[\tilde{z}\|y](\zeta) = \Phi\left(\frac{-y}{\sigma(y)}\right)\delta_0(\zeta)+\frac{1}{\sigma(y)}\phi\left(\frac{\zeta-y}{\sigma(y)}\right)\chi_{[0,1]}+\Phi\left(\frac{y-1}{\sigma(y)}\right)\delta_0(1-\zeta)$		see text (Section 6.5)

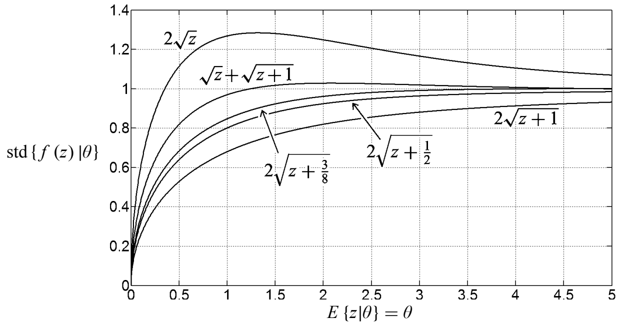


Fig. 1 Conditional standard-deviation $\sigma_f(\theta) = \mathrm{std}\{f(z)|\theta\}$ of the transformed Poisson variables z with parameter $\theta = \mu(\theta)$ after stabilization by five root-type transformations $f(z)$: $2\sqrt{z}$ (1), $2\sqrt{z+1}$, $2\sqrt{z+1/2}$ (Bartlett 1936), $2\sqrt{z+3/8}$ (Anscombe 1948), $\sqrt{z}+\sqrt{z+1}$ (Freeman and Tukey 1950).

attributes the result to A.H.L. Johnson). In their works, Anscombe, Freeman, and Tukey present stabilizers also for the binomial and negative-binomial distributions ("angular" transformations based on the arcsine and hyperbolic arcsine). In a recent paper, Guan (2009) studies the trade-off between variance, skewness and kurtosis of transformed Poisson, binomial and negative binomial distributions, proposing few variations of the transformations of Anscombe (1948) and Freeman and Tukey (1950).

As an illustration of the problem at hand, in Figure 1 we show the conditional standard deviation $\sigma_f(\theta)=\mathrm{std}\{f(z)|\theta\}$ of the transformed Poisson variables z with parameter $\theta = \mu(\theta)$ after stabilization by the following five root-type transformations $f(z)$: $2\sqrt{z}$ (1), $2\sqrt{z+1}$, $2\sqrt{z+1/2}$ (Bartlett 1936), $2\sqrt{z+3/8}$ (Anscombe 1948), $\sqrt{z}+\sqrt{z+1}$ (Freeman and Tukey 1950). Without loss of generality, here and in the sequel we aim at stabilizing to a unitary constant standard deviation $c = 1$.

All the above mentioned works propose parametric transformations which find their origin in the rough stabilizer given by (1). The design and applicability of these transformations is restricted to the respective family of distributions and the stabilization is exact only asymptotically.

There are three seminal works that provide results about the stabilization of a generic family of distributions. First, there is the 1943 paper by Curtiss, where general asymptotic theorems are proved, giving theoretical support to many empirical stabilizers that were already widely used. Second and arguably the most important contribution is Efron's 1981 work (Efron 1981; 1982) on the existence of transformations for exact variance stabilization and/or perfect normalization. Efron not only formalizes sufficient conditions for the existence of the exact transformations (using the framework of so-called "general transformation families"), but also provides their analytical expressions. Third, we mention Tibshirani's 1986 AVAS procedure (Tibshirani 1986b; 1988) for regression, where approximate variance stabilizing transformations are iteratively computed by recursive application of (1). Roughly speaking, this can be interpreted as an iterative refinement of the stabilizer. Although the AVAS is developed for data-driven application, Tibshirani briefly highlights its potential use for random variables. The variance-stabilizing transformations used by Efron and Tibshirani are nonparametric and the stabilization is non-asymptotic.

Together with the AVAS, other widely used tools for data-driven variance stabilization are the Box-Cox parametric transformations (for a review, see Sakia 1992) and their modifications (Blaylock and Smallwood 1985; Sakia 1992; DiCiccio et al 2006).

More recently, a number of techniques for dealing with wavelet coefficients of heteroskedastic data have been also proposed. In particular, noting that the Anscombe transformation is not adequate for low-count (photon limited) signals, Starck et al. (Starck et al 1998; Zhang et al 2008) developed a generalization of this transformation applicable to linear combinations of Poissonian variates. It is shown that, provided some smoothness of the underlying data, the variance stabilization of the wavelet coefficients is more effective and, in a sense, easier to achieve than the variance stabilization of the original data. With a similar goal, Fryzlewicz, Nason, et al. introduce so-called wavelet-Fisz transforms (Fryzlewicz and Nason 2004; Fryzlewicz and Delouille 2005; Fryzlewicz 2008; Nunes and Nason 2008; Nason 2008) where stabilization is achieved by dividing each wavelet coefficient by an estimate of its standard deviation. In connection with these latter methods, it is worth citing also the various threshold-correcting schemes by Kolaczyk (1997; 1999). Further references can be found in the review articles (N. 2006; Taylor 2006).

Summarizing, we can say that most of the works have traditionally been about variance stabilizing transformations having an explicit parametric expression (e.g., Bartlett 1936; Anscombe 1948; Freeman and Tukey 1950; Veevers and Tweedie 1971; Blaylock and Smallwood 1985; Sakia 1992; DiCiccio et al 2006; Starck et al 1998; Zhang et al 2008; Guan 2009), while relatively few authors have researched nonparametric stabilizers (e.g., Efron 1981, 1982; Tibshirani 1986b, 1988; Fryzlewicz 2008) and that, quite clearly, the leading interest is for non-asymptotic results (e.g., Bartlett 1936; Freeman and Tukey 1950; Veevers and Tweedie 1971; Efron 1981, 1982; Tibshirani 1986b, 1988; Guan 2009), although also asymptotic ones (e.g., Curtiss 1943; Anscombe 1948; Starck et al 1998; Zhang et al 2008) can be satisfactory for applications.

With such abundance of different transformations, the natural question arises about which transformations provide the best stabilization. Unfortunately, this question remains largely unanswered. Firstly, because, as noted explicitly by Freeman and Tukey (1950) and later by Laubscher (1961), it is typically impossible to achieve simultaneously good stabilization for all parameter values: thus, when a stabilizer appears to be better than another one for some values of the parameter θ, it is likely that for other values it is actually worse. In this sense, there might be no "universally best stabilizer". Secondly, because no objective criterion for assessing the goodness of a stabilizer has ever been formulated. In the light of the first remark, we see that simply demanding $\mathrm{std}\{f(z)|\theta\}$ to be as close as possible to c might be too vague and ambiguous.

1.2 Contributions and outline

The contribution of this paper is five-fold and the remainder of the paper is structured accordingly.

First and foremost (Section 2), we approach the variance stabilization problem as an explicit optimization problem, where we aim at minimizing a nonlinear *weighted stabilization functional* that measures the discrepancy between the conditional standard deviation $\mathrm{std}\{f(z)|\theta\}$ of the transformed variable $f(z)$ and the fixed desired constant c. In particular, we consider functionals having different weights for different values of θ and that can accommodate small stabilization errors, thus enabling a more uniform stabilization overall. The stabilization functional can be used as an objective measure to test the goodness of f as a variance stabilizer.

Second (Section 3), we propose a recursive procedure that at every iteration refines a nonparametric continuous monotone function in such a way to decrease the corresponding value of the stabilization functional.

Third (Section 4), we demonstrate optimization by direct search. In particular, in our experiments we rely on the downhill simplex algorithm (Nelder and Mead 1965). This method is particularly effective for families of discrete distributions.

Fourth (Section 5), we relax the typical assumption of monotonicity of f, and replace it with a somewhat milder requirement on the invertibility of $E\{f(z)|\theta\}$. This is a sufficient condition for successful use of stabilization in many applications such as, for example, regression, where the main interest is about estimating the parameter θ. Optimized nonmonotone stabilizers can provide sensibly better stabilization than monotone ones.

Fifth (Section 6), as a result of the above techniques, we present a number of optimized variance stabilizers for the most common distribution families. These stabilizers are shown to outperform the existing ones. In particular, variance-stabilizing transformations optimized for low-count Poisson data are presented as leading examples throughout the various parts of the paper.

We conclude with discussions and a few open problems (Section 7).

2 Variance stabilization as a minimization problem

Let $c > 0$ be a fixed positive constant (throughout the paper we assume $c = 1$) and

$$e_f(\theta) = \sigma_f(\theta) - c$$

be the *local* error because of inexact stabilization (where locality is understood in terms of the conditioning on θ). A *global* cost functional is hence defined as

$$C_f = \int_\Theta \left|e_f(\theta)\right| d\theta. \tag{2}$$

In principle, one may formulate the variance stabilization problem as the solution of

$$\mathrm{argmin}_f \, C_f, \tag{3}$$

under some constraints on f (e.g., f continuous and strictly increasing). Variance stabilization is exact only when $C_f = 0$ for some f. In most cases of practical interest, this cannot be achieved and the infimum value of C_f is strictly positive and/or a minimum cannot be attained. For example, exact variance stabilization cannot be achieved for the Poisson or binomial families of distributions (Curtiss 1943). Consequently, most works have been devoted to ensuring *asymptotic* stabilization for limiting values of the parameter θ. For these cases, the goodness of stabilization expressed by the rate of decay of the limit

$$e_f(\theta) \underset{\theta\to\infty}{\longrightarrow} 0. \tag{4}$$

The contributions by Curtiss, Barlett, and Anscombe have been of this sort. While these work ignore an integral formulation such as (2), we note that the quadratic decay for (4) achieved by the Anscombe stabilizer at least ensures the convergence of the improper integral (2). However, an unweighted cost functional like (2) is not sufficiently flexible to be useful in the practice. In particular, one has often to deal with parameter values that are far from limiting values and asymptotic stabilization is no longer appealing. Typically, when exact stabilization is not possible, an optimal asymptotic decay prevents $e_f(\theta)$ to be small for ordinary values of θ. Thus, compromises are necessary. For the case of the Poisson distribution, we note that the main goal of Freeman and Tukey (1950) was actually to obtain a stabilizer, $f(z) = \sqrt{z} + \sqrt{z+1}$ (see Figure 1), that yields an $e_f(\theta)$ which has slower decay for $\theta \to \infty$ but which remains "small" also for "small" values of θ. Thus, it appears that the simple formulation given by (2)-(3) is inadequate to properly specify the problem and deal with it.

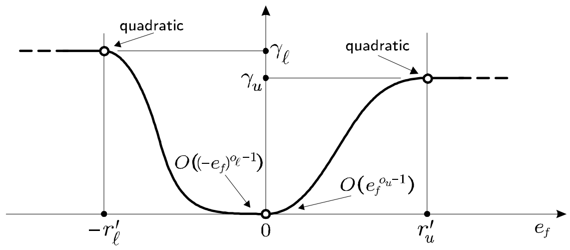


Fig. 2 The weight function φ (8) used for the construction of the stabilization functional (9).

2.1 Weighted stabilization functional

As clear from the previous discussion, it makes little sense to aim at exact variance stabilization simultaneously for all parameter values. Therefore, we consider a *weighted stabilization functional* of the form

$$C_f = \int_\Theta w_\theta(\theta)\, w_e\left(e_f(\theta)\right) d\theta, \tag{5}$$

where the weight functions w_θ and w_e provide different weighting for the different values of θ and the different stabilization errors $e_f(\theta)$, respectively.

In particular, we consider the following special form of the weights w_e, which favor approximate stabilization while ignoring very large stabilization errors, in a fashion roughly similar to Tukey's biweight. Let $\gamma_u, \gamma_\ell \leq 1$, $r'_u, r'_\ell \geq 0$, $r''_u \geq r'_u$, $r''_\ell \geq r'_\ell$, $o_u, o_\ell \geq 1$ be some real constants and χ_Ω be the characteristic (indicator) function of a set $\boldsymbol{\Omega}$. We define the weights w_e as

$$w_e\left(e_f(\theta)\right) = \left|\varphi\left(\overline{e_f}(\theta)\right)\overline{e_f}(\theta)\right|,$$

where $\overline{e_f}$ is the clipped stabilization error

$$\overline{e_f}(\theta) = \overline{\sigma_f}(\theta) - c = \max\left\{-r''_\ell, \min\left\{r''_u, e_f(\theta)\right\}\right\}, \tag{6}$$

$$\overline{\sigma_f}(\theta) = \max\left\{c - r''_\ell, \min\left\{c + r''_u, \sigma_f(\theta)\right\}\right\}, \tag{7}$$

and with the function φ given by (see Figure 2)

$$\begin{aligned}\varphi\left(e_f\right) &= \gamma_u \cdot \chi_{[0,+\infty)}\left(e_f\right) \times \\ &\times \left\{\left[1-\left(\frac{e_f - r'_u}{r'_u}\right)^2\right]^{(o_u-1)} \chi_{(-\infty, r'_u)}\left(e_f\right) + \chi_{[r'_u,+\infty)}\left(e_f\right)\right\} + \\ &+ \gamma_\ell \cdot \chi_{(-\infty,0)}\left(e_f\right) \times \\ &\times \left\{\left[1-\left(\frac{e_f + r'_\ell}{r'_\ell}\right)^2\right]^{(o_\ell-1)} \chi_{(-r'_\ell,+\infty)}\left(e_f\right) + \chi_{(-\infty,-r'_\ell]}\left(e_f\right)\right\}.\end{aligned} \tag{8}$$

The clipped argument $\overline{e_f}(\theta)$ cannot distinguish stabilization errors whose magnitude is larger than r''_ℓ, r''_u, while the multiplication against the function φ increases the polynomial order of the small stabilization errors from 1 (i.e., linear) to o_ℓ, o_u. Note that for a positive (resp. negative) argument, the function φ has a zero of order $o_u - 1$ (resp. $o_\ell - 1$) at zero and becomes constant (with quadratic-smooth joint) equal to γ_u (resp. γ_ℓ) starting from r'_u (resp. r'_ℓ), as shown in Figure 2.

Thus, the cost functional (5) takes the form

$$C_f = \int_\Theta w_\theta(\theta)\left|\varphi\left(\overline{e_f}(\theta)\right)\overline{e_f}(\theta)\right| d\theta, \tag{9}$$

where the factor $w_\theta(\theta)$ is bounded as $0 \leq w_\theta(\theta) \leq 1$ and provides localization to the functional by assigning small weights to the values of θ for which precise stabilization is not required. While the functional (5) can, of course, be given also in other forms, we found the simple (9) to be particularly convenient for demonstrating the significance of a weighted stabilization, as illustrated by the examples given in the later sections and especially by those in Sections 3.1 and 6.3.

Table 2 Recursive integral algorithm for optimizing f.

0. Initialize

$f_0(z) = z$ (i.e. identity), or $f_0(z)$ an arbitrary monotone increasing function.

Iterate the following three stages:

1. Compute statistics

$\vartheta_{f_k}(\theta) = \text{med}\{f_k(z)\,|\theta\}$, (conditional median)

$\sigma_{f_k}(\theta) = \text{std}\{f_k(z)\,|\theta\}$. (conditional st. deviation)

2. Compute stabilization refinement

$r_k(\xi) = \int_{f_k(z_a)}^{\xi} I_{f_k}(\theta)\, d\left[\vartheta_{f_k}(\theta)\right] + a,$ (integration w.r.t. cond. median)

where

$I_{f_k}(\theta) = 1 - \dfrac{w_\theta(\theta)\,\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)}{\overline{\sigma_{f_k}}(\theta)},$ (weighted integrand)

$\overline{e_{f_k}}(\theta)$ and $\overline{\sigma_{f_k}}(\theta)$ are defined by (6) and (7), respectively, and z_a and a are fixed anchoring constants.

3. Compose

$f_{k+1}(z) = r_k\left(f_k(z)\right)$. (refinement of the stabilizer)

3 Recursive integral refinement

We propose an heuristic iterative procedure which aims at minimizing the cost functional (9) through recursive stabilization. The procedure is described by the algorithm in Table 2.

The general structure of this algorithm is analogous to the iterations for stabilization of the variance used in AVAS (Tibshirani 1988). The crucial difference stays however in the definition of the *refinement function* r_k

$$r_k(\xi) = \int_{f_k(z_a)}^{\xi} I_{f_k}(\theta)\, d\left[\vartheta_{f_k}(\theta)\right] + a, \tag{10}$$

which in AVAS is defined by the simple integral (1) (where the integration is computed with respect to the conditional mean $E\{f_k(z)|\theta\}$), whereas we use the *weighted integrand* $I_f(\theta)$,

$$I_f(\theta) = 1 - \frac{w_\theta(\theta)\,\varphi\left(\overline{e_f}(\theta)\right)\overline{e_f}(\theta)}{\overline{\sigma_f}(\theta)}, \quad \theta \in \Theta \subseteq \mathbb{R}, \tag{11}$$

integrated with respect to the conditional medians $\vartheta_f(\theta) = \operatorname{med}\{f(z)|\theta\} = f(\operatorname{med}\{z|\theta\})$. Here and throughout the paper, we consider Riemann-Stieltjes integration (see, e.g., Kolmogorov and Fomin 1975), i.e. given two functions g (integrand) and h (integrator) of θ, the integral of g with respect to h, $\int g(\theta)\, d[h(\theta)]$, can be written as $\int g(\theta) \frac{\partial}{\partial\theta} h(\theta)\, d\theta$, where both the integral and the derivative operator are taken in their distributional sense, thus allowing h to have discontinuities. All numerical computations are enabled by employing trapezoidal integration and linear interpolation and extrapolation.

Let us give an intuition why (11) can be useful for minimizing (9). If $w_\theta(\theta)\left|\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)\right|$ is small then $I_{f_k}(\theta)$ is close to 1, the refinement function r_k has a derivative which is close to 1 and, thus, composition of f_k with r_k (Step 3) does not alter essentially the conditional standard deviation $\sigma_{f_k}(\theta)$. Conversely, if $w_\theta(\theta)\left|\varphi\left(\overline{e_{f_k}}(\theta)\right)\right|$ is close to 1 (which is the upper bound for these weights), then $w_\theta(\theta)\left|\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)\right|$ is close to $\left|\overline{e_{f_k}}(\theta)\right|$ and $I_f(\theta)$ is close to $\frac{c}{\sigma_f(\theta)}$, which by relaxation of (7) coincides with the integrand of (1), thus composition of f_k with r_k should make $\sigma_{f_k}(\theta)$ close to c. The same reasoning can be installed even when γ_{u} or γ_{l} are equal to some $\gamma < 1$, because when $w_\theta(\theta)\left|\varphi\left(\overline{e_{f_k}}(\theta)\right)\right|$ is close to γ, then $I_f(\theta) \approx 1 - \frac{\gamma(\sigma_{f_k}(\theta)-c)}{\sigma_{f_k}(\theta)}$ and $\sigma_{f_{k+1}}(\theta) \approx \sigma_{f_k}(\theta) - \gamma\left(\sigma_{f_k}(\theta) - c\right)$ leading to a geometric convergence (with rate $(1-\gamma)^k$) of σ_{f_k} towards c. Further discussion about the particular form of (10) is provided in Section 7.1.

Observe that, since $w_\theta \leq 1$ and $\varphi \leq \max\left\{\gamma_{\mathrm{u}}, \gamma_{\mathrm{l}}\right\} \leq 1$, we have that the weighted integrand (11) is positive, $I_f \geq 0$, which means that composition of a monotone function with the refinement r_k returns another monotone function. Thus, the refined transformations f_k are all monotone increasing, from which we have $\operatorname{med}\{f_k(z)|\theta\} = f_k(\operatorname{med}\{z|\theta\})$ for all $k \in \mathbb{N}$. Note also that the anchoring constants in (10) are used to maintain $f_{k+1}(z_a) = a$ and thus avoid drift of the refined stabilizer while the algorithm progresses. This is a mere technicality, because an arbitrary additive constant does not obviously influence the stabilization properties of a transformation.

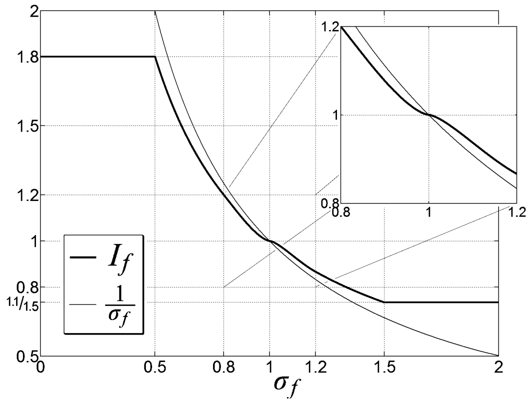


Fig. 3 Illustration of the weighted integrand I_f (11), shown as function of σ_f for $w_\theta = 1$ and w_e defined by the parameters $o_{\mathrm{u}}, o_{\mathrm{l}} = 1.5$, $r'_{\mathrm{u}}, r'_{\mathrm{l}} = 0.2$, $r''_{\mathrm{u}}, r''_{\mathrm{l}} = 0.5$, $\gamma_{\mathrm{u}}, \gamma_{\mathrm{l}} = 0.8$. The weighted integrand (thick line) is compared with the basic integrand $1/\sigma_f$ used in (1) (thin line).

Overall, this particular algorithm is obtained mostly out of the above heuristic speculations. While we cannot claim that it can exactly minimize the cost (9) (a few counterexamples are given in the later sections), our extensive experiments suggest that the cost is at least progressively decreased through the recursions, i.e. the sequence

$$C_{f_k} = \int_\Theta w_\theta(\theta)\left|\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)\right| d\theta. \tag{12}$$

is decreasing. Thus, the algorithm may be treated as a pseudo-minimizer for (9).

On this point, we wish to note that performing the integration with respect to the conditional median (as opposed to the conditional mean) appears to be crucial for enabling the decrease of the sequence of costs (12). In the following sections, we give few examples demonstrating that if the integration with respect to the conditional means were used (i.e. defining $\vartheta_{f_k}(\theta) = E\{f_k(z)|\theta\}$ instead of $\vartheta_{f_k}(\theta) = \operatorname{med}\{f_k(z)|\theta\}$ in Table 2), the sequence (12) would not be decreasing and instead it would converge to a limit sensibly larger than its minimum. The examples are discussed in Section 7.2.

3.1 Example: Poisson

Let us give an example of the application of this algorithm to the stabilization of the Poisson family. We consider a stabilization functional where $w_\theta = \chi_{[0,15]}$ (characteristic function of the closed interval $[0, 15]$) and where w_e is defined by the following parameters: $o_{\mathrm{u}}, o_{\mathrm{l}} = 1.5$, $r'_{\mathrm{u}}, r'_{\mathrm{l}} = 0.2$, $r''_{\mathrm{u}}, r''_{\mathrm{l}} = 0.5$, $\gamma_{\mathrm{u}}, \gamma_{\mathrm{l}} = 0.8$. Figure 3 illustrates the weighted integrand I_f (11) corresponding to these parameters as a function of σ_f when $w_\theta = 1$, i.e. for $\theta \in [0, 15]$. For the computations it suffices to regard only $0 \leq z \leq 54$, because the probability $\Pr[z > 54|\theta]$ is numerically negligible for any $\theta \in [0, 15]$,

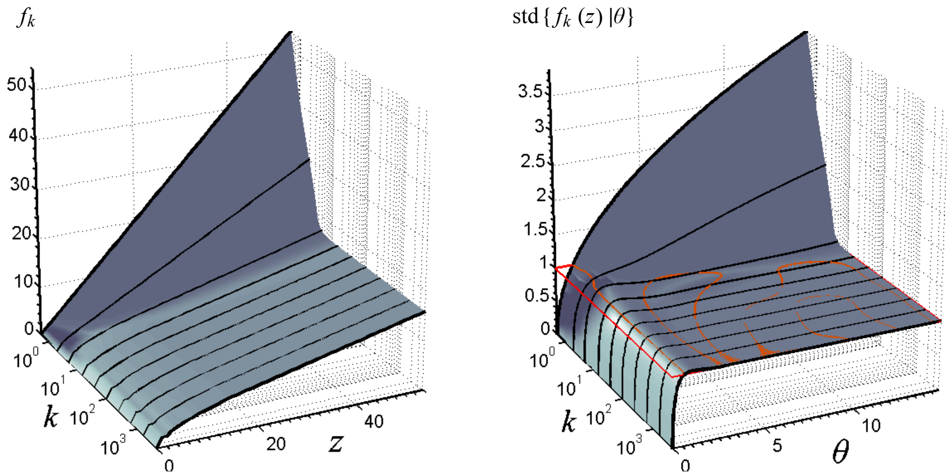


Fig. 4 Optimization of Poisson variance stabilizer using the recursive integral algorithm (Table 2). Left: sequence of transformations $f_k(z)$; Right: sequence of the corresponding conditional standard deviations std $\{f_k(z)|\theta\}$ (the red contour lines indicate where stabilization is exact).

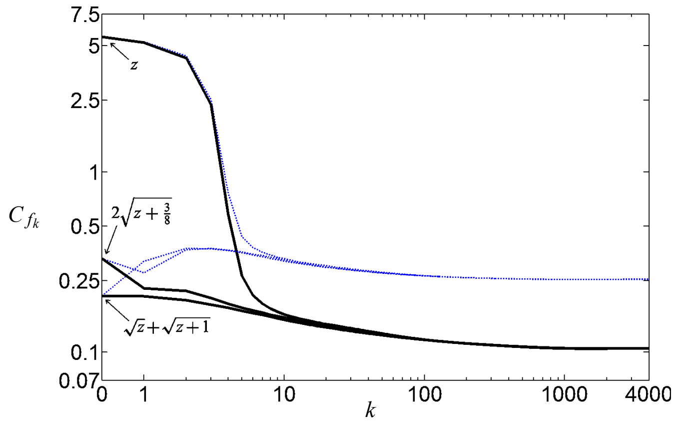


Fig. 5 Optimization of the Poisson stabilizer by the recursive integral algorithm. Solid lines: value of the stabilization functional vs. iterations (logarithmic scale). For all three initializations the final value is $C_{f_{4000}} = 0.1051$. Dotted lines: sequence obtained by integration with respect to conditional mean (see Section 7.2).

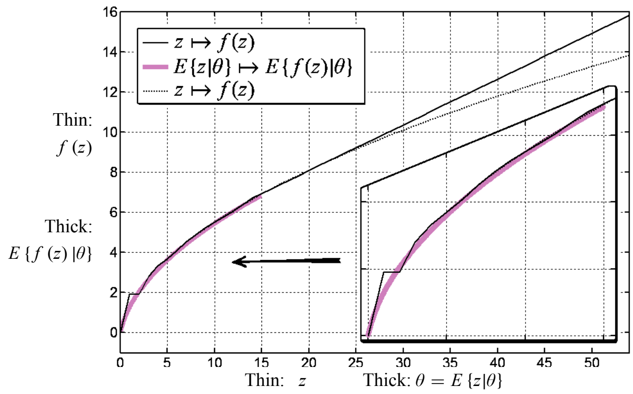


Fig. 6 Optimized Poisson stabilizer f obtained after 4000 iterations of the algorithm in Table 2 initialized by $f_0 = z$ (thin solid line) and the corresponding expectation mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ (thick line). The thin dotted line shows the optimized stabilizers obtained by initialization by either $f_0 = \sqrt{z} + \sqrt{z+1}$ (Freeman and Tukey 1950) or $f_0 = 2\sqrt{z+3/8}$ (Anscombe 1948) (the plots of these two optimized stabilizers coincide and, for $z \leq 25$, overlap almost perfectly with the one obtained from $f_0 = z$).

with $\Pr[z > 54|\theta] < 2\Pr[z = 55|15] = 2.3 \cdot 10^{-15}$. In Figure 4 we show the sequence f_k of transformations produced by 4000 iterations of the algorithm and the corresponding conditional standard deviations std $\{f_k(z)|\theta\}$. The algorithm had been initialized by $f_0(z) = z$ with the anchoring constants $z_a = a = 0$. We can see that after the first 10 iterations the algorithm reaches a near steady-state, with a slow but progressive improvement. The decreasing sequence of costs C_{f_k} (12) is plotted in Figure 5. In this figure we also show the sequence C_{f_k} obtained from different initializations, using $f_0 = 2\sqrt{z+3/8}$ (Anscombe 1948) and $f_0 = \sqrt{z} + \sqrt{z+1}$ (Freeman and Tukey 1950): as one can see from the plots, in practice, it does not really matter which had been the particular initialization, since the same stabilization accuracy is eventually achieved with approximately the same rate (for all three cases, $C_{f_{4000}}$=1.051). The optimized stabilizer $f = f_{4000}$ obtained from $f_0 = z$ is shown in Figure 6. It is interesting to observe few peculiarities of this transformation. We notice that the first 16 samples $0 \leq z \leq 15$ are roughly following a root-type progression, pretty much like the stabilizers cited in Section 1.1. For $z > 15$ the transformations f_k remain affine like $f_0 = z$, because (due to the linear extrapolation) the refinement itself is always affine for $z > 15$, since med $\{z|\theta\} \leq 15$ if $\theta \leq 15$ (Adell and Jodrá 2005). Close inspection shows the curious fact that $f(1) \approx f(2)$. In other words, the optimization leads to a non-invertible transformation. As demonstrated in the next sections, this peculiarity turns out to be quite common for optimized non-decreasing transformations. It is important to emphasize that the lack of invertibility of f does not necessarily compromise the invertibility and smoothness of the mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ (see the thick dashed

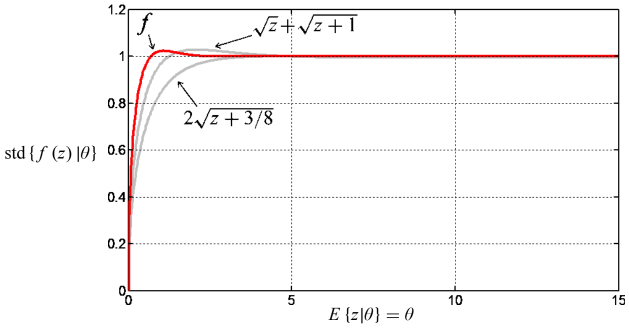


Fig. 7 Conditional standard deviation of the Poisson variance stabilizer f optimized by the recursive integral algorithm (Table 2) vs. those corresponding to the classical stabilizers $2\sqrt{z+3/8}$ (Anscombe 1948) and $\sqrt{z}+\sqrt{z+1}$ (Freeman and Tukey 1950).

line in Figure 6), which is the actual inverse mapping that one uses after processing the stabilized data (e.g., filtering in a regression application). Note that monotonicity of f is anyway guaranteed by the nonnegativity of I_{f_k} in the integral (10).

In the same figure, the thin dotted line shows the two stabilizers f_{4000} obtained from initialization with the root transformations $f_0 = 2\sqrt{z+3/8}$ (Anscombe 1948) and $f_0 = \sqrt{z}+\sqrt{z+1}$ (Freeman and Tukey 1950) (with the maximum absolute difference of 0.0073, their respective plots are so close to each other, that they cannot be distinguished in the figure). These two stabilizers practically coincide with that obtained from $f_0 = z$ for all $z \leq 25$ (maximum absolute difference of 0.0506).

Figure 7 presents the conditional standard deviation $\text{std}\{f(z)|\theta\}$ for $f = f_{4000}$ in comparison with that for $f = 2\sqrt{z+3/8}$ (Anscombe 1948), $f = \sqrt{z}+\sqrt{z+1}$ (Freeman and Tukey 1950), and $f = z$. The improvement of the stabilization is here clearly visualized.

Stable convergence of the iterative integral algorithm was verified experimentally, up to the numerical precision of the algorithm, in extensive tests. However, its limit does not generally coincide with the infimum of the stabilization functional. In particular, in the above example we have seen that the refinement outside of the conditional medians, i.e. for $z \geq 15 = \max_{\theta \leq 15} \text{med}\{z|\theta\}$, is bound to be affine and thus it restricts the space of functions which can be spanned by the iterations of the algorithm.

4 Optimization by direct search

A practical way to circumvent the above issues as well as the computational aspects involved in the evaluation of the integrals and in the composition with the refinement functions, is to approach the minimization by direct search, i.e. by explicit evaluation of C_f. This method is particularly feasible for discrete distributions, for which the search space can be identified directly with the values of $f(z)$, but it can be used also with continuous distributions provided a suitable representation of f (e.g., optimizing over the wavelet coefficients or other multiscale decompositions of f) or by using interpolation over an adaptive sampling grid.

In principle, any iterative algorithm for direct search or derivative-free optimization may be used (Kolda et al 2003). In our simulations, we utilize the downhill simplex algorithm (Nelder and Mead 1965; Conn et al 2009). We choose this algorithm mainly because of its ease of applicability. Other more sophisticated methods (e.g., Kelley 1999; Conn et al 2009) can be considered as well. The stabilizer found upon convergence of the recursive integral algorithm may be taken as initialization for the search.

First, we minimize the cost functional (9) within the class of monotone nondecreasing[3] functions, thus constraining the search to these functions. A more general case is then considered in Section 5.

4.1 Example: Poisson

Again, we give here an example of the application of this algorithm to the stabilization of the Poisson family, for the same weighted stabilization functional as in the previous section.

The optimized stabilizer found by direct search is shown in Figure 8. Observe that this transformation is constant on many subintervals of [0, 54], while the mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ is smooth and invertible for all $\theta \in [0, 15]$. The value of the stabilization functional achieved by this transformation is 0.0944, well below the cost of 0.1051 obtained from the iterative integral solutions (see also Figure 5). Part of this improvement stems from the direct search succeeding in optimizing also the samples $z > 15$. As observed in Efron (1982), ideal stabilization and ideal normalization typically cannot coexist and the distribution of stabilized variables needs to depart from a normal one. In Figure 10, we can see that indeed the conditional cumulative distributions functions (c.d.f.) after direct search optimization allow a poorer fit by the standard normal c.d.f. Φ compared to those after recursive integral optimization.

As in the previous section, we ignore the optimization for $z > 54$. Firstly, because we aim at restricting the dimensionality of the search space, optimizing only over the most influential samples for the considered stabilization functional, which is supported on the closed interval [0, 15]. Secondly, because, as noted above, with a growing $z > 54$, $\Pr[z|\theta]$ for $\theta \in [0, 15]$ becomes so small that the optimization of $f(z)$ would eventually be badly conditioned and

[3] We note that it is practically equivalent to restrict to either nondecreasing or strictly increasing functions: provided φ is continuous and $\|w_\theta\|_1 < \infty$, one can always introduce a small enough perturbation to a nondecreasing f, making it strictly increasing and without increasing C_f more than an arbitrary positive constant.

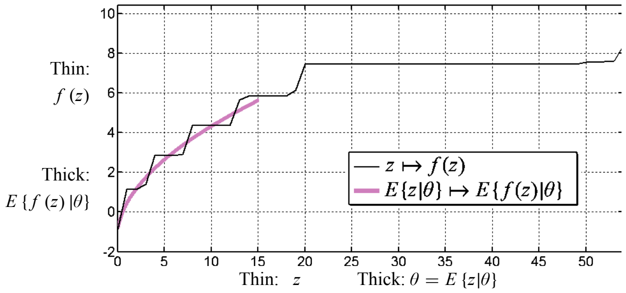


Fig. 8 Optimized monotone stabilizer for the Poisson family obtained by direct search (thin line) and the corresponding expectation mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ (thick line).

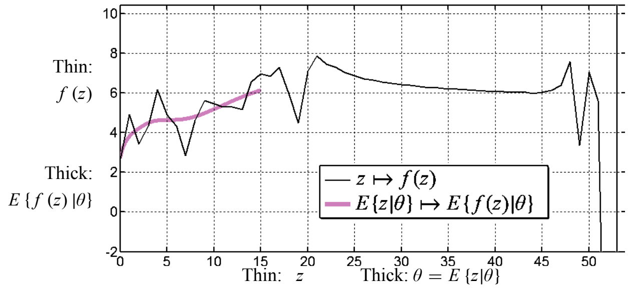


Fig. 9 Optimized nonmonotone stabilizer for the Poisson family obtained by direct search (thin line) and the corresponding monotone expectation mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ (thick line). The values outside of the plot are $f(52) = -21.697$, $f(53) = -35.938$, and $f(54) = 134462.365$, while $\max_{\theta\in[0,15]} \Pr[z = 54|\theta] = 4.28 \cdot 10^{-15}$ (see text).

hence numerically unstable. Thus, restricting the optimization to $z \leq 54$ can be formally interpreted as a form numerical regularization. We shall discuss in Section 6.6 how stabilizers optimized on bounded subsets can be easily extended for stabilization over unbounded subsets.

5 Relaxing the monotonicity of *f*

The example shown in Figure 8 highlights the maybe surprising fact that optimization can lead to transformations which have concentrated "jumps" between a number of stationary segments. It is quite natural to suspect that improved stabilization may be achieved by allowing this pattern to be enhanced (e.g., larger jumps between segments where the transformation is decreasing).

Therefore, here we relax the requirement on the monotonicity of f (undertaken in Sections 3 and 4) and replace it with that of the invertibility of the mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$. The choice of this latter requirement follows from the idealized scenarios where variance stabilization is utilized. For instance, the principal aim of regression is estimating the conditional expectation $E\{z|\theta\}$ from the given noisy observations. Thus, when the regression problem is translated to the stabilized variables, one estimates their conditional expectation $E\{f(z)|\theta\}$; the inverse mapping $E\{f(z)|\theta\} \mapsto E\{z|\theta\}$ is then exploited to express the solution for the untransformed variables.

We note that the monotonicity of $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ is not necessarily a weaker requirement than the monotonicity of f, in the sense that the latter does not imply the monotonicity of $E\{f(z)|\theta\}$ as a function of θ or $E\{z|\theta\}$. For instance, in Foi (2009a) we have shown that monotonicity of f is alone not sufficient to ensure monotonicity of the mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$ for particular families of normal distributions and have formulated sufficient conditions which require hypotheses on both f and $\sigma(\theta)$.

5.1 Example: Poisson

Figure 9 shows the optimized nonmonotone stabilizer found by direct search. Observe that this stabilizer has indeed a number of oscillations and that there is a visible correspondence between the main stationary segments of the monotone stabilizer in Figure 8 and the segments where this nonmonotone stabilizer is decreasing.

The reduction of the stabilization functional C_f after relaxation of the monotonicity is substantial, going further down from 0.0944 to 0.0771. In connection with the remark made at the end of Section 4.1, let us note that the large-magnitude values found here for $z = 52, 53, 54$ have no practical impact to the actual stabilization, as they are anyway several orders of magnitude smaller than the corresponding probabilities $\Pr[z|\theta]$ for $\theta \in [0, 15]$.

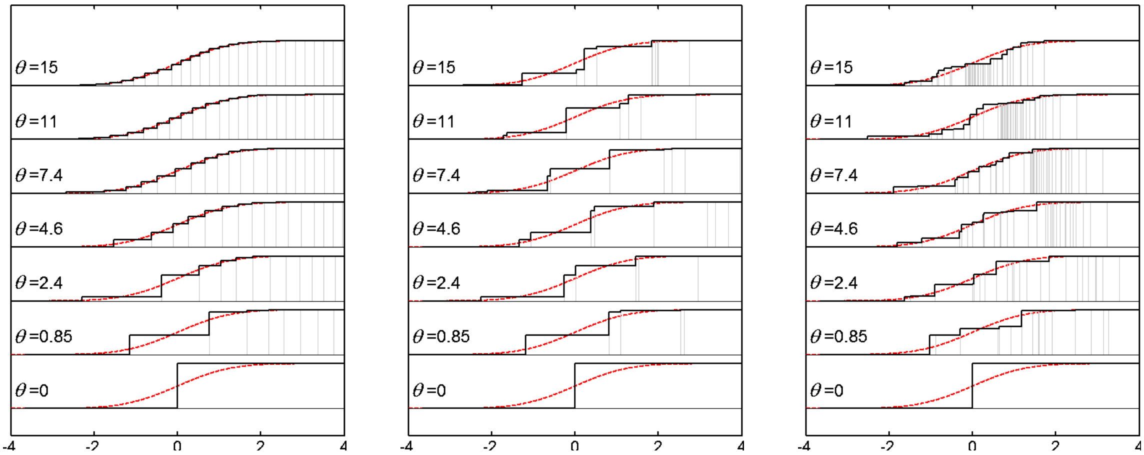
We wish to observe also that, for $\theta = E\{z|\theta\} \approx 5$, the plot of $E\{f(z)|\theta\}$ has a stationary point (i.e., there is a value of θ such that $\partial E\{f(z)|\theta\} / \partial E\{z|\theta\} = 0$). This suggests that better stabilization could be reached by further relaxing the constraint on the invertibility of the mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$.

Figures 11 and 12 show the conditional standard deviation $\operatorname{std}\{f(z)|\theta\}$ for our three optimized stabilizers in comparison with that for the classical stabilizers shown in Figure 1. As a result of the optimization, all our stabilizers, and particularly the nonmonotone one, have a negligible maximum overshooting[4] of the desired stabilized deviation $c = 1$ and yet a very slow decay at zero[5].

It is remarkable that the additional freedom of a nonmonotone f also contributes to improving the approximate normality of the conditional distributions of $f(z)|\theta$, as shown in Figure 10.

[4] About 3%, comparable to that achieved by $f = \sqrt{z} + \sqrt{z+1}$ (Freeman and Tukey 1950).

[5] Comparable to that produced by $f = 2\sqrt{z}$, which instead overshoots dramatically.



Recursive integral (Section 3.1) | Direct search - Monotone (Section 4.1) | Direct search - Nonmonotone (Section 5.1)

Fig. 10 Stabilization vs. normalization: the conditional cumulative distribution functions (c.d.f.) of the stabilized Poisson variables after our three proposed optimization procedures. The distributions are drawn centered about their mean $E\{f(z)|\theta\}$. The dashed line is the c.d.f. Φ of the standard normal $\mathcal{N}(0,1)$.

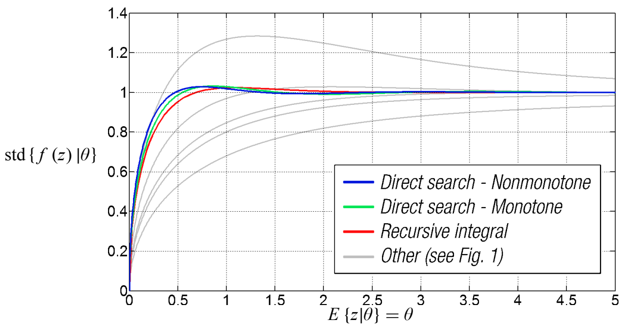


Fig. 11 Conditional standard-deviation $\text{std}\{f(z)|\theta\}$ of the transformed Poisson variates after stabilization by our optimized stabilizers compared against the other stabilizers from Figure 1. Although the optimization is made for $\theta \in [0, 15]$, here we detail only $0 \le \theta \le 5$. See Figure 12 for the logarithmically scaled plot with $0.001 \le \theta \le 15$.

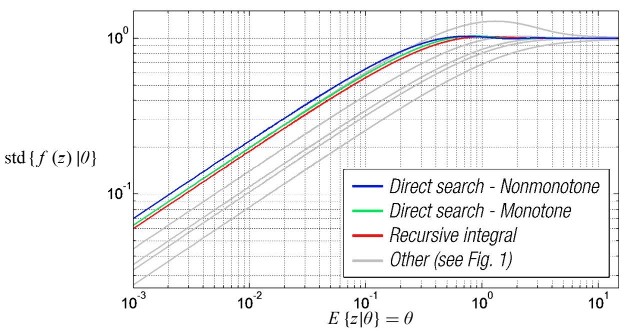


Fig. 12 Same as in Figure 11, but using logarithmic scale on both axes.

6 Examples of optimized transformations

In this section, to further illustrate to potential of the proposed techniques, we present a number of optimized stabilizers for the most common families of distributions. When relevant, comparisons are made against some of the best stabilizers found in the literature.

Matlab and ASCII files implementing all optimized transformations presented in this paper can be downloaded from the author's website at `http://www.cs.tut.fi/~foi/optvst` .

6.1 Binomial

Let us consider the stabilization of the binomial family of distributions with n trials and parameter $\theta \in \Theta = [0, 1]$. Similar to the Poisson family, we compare against the angular transformations by Anscombe (1948) and by Freeman and Tukey (1950):

$$f(z) = 2\sqrt{n+\tfrac{1}{2}}\arcsin\sqrt{\tfrac{z+3/8}{n+3/4}}, \tag{13}$$

$$f(z) = \sqrt{n+\tfrac{1}{2}}\left(\arcsin\sqrt{\tfrac{z}{n+1}} + \arcsin\sqrt{\tfrac{z+1}{n+1}}\right). \tag{14}$$

Here $z \in Z = [0, n]$ and, for any $\theta \in (0, 1)$, both these transformation are asymptotically exact as $n \to \infty$. Good stabilization for θ close to 0 or 1 is difficult to obtain when n is small, thus in the following examples we consider $n = 4, 7, 15, 30$. In Figure 13 we show the stabilizers (13)-(14), the optimized stabilizers found by the proposed algorithms, and the corresponding conditional standard deviations (in both linear and logarithmic scale). The optimized transformations are found by minimization of a stabilization functional where the weight w_e is the same as that used in Section 3.1 while $w_\theta = \chi_\Theta = \chi_{[0,1]}$. As can be seen from the figure, optimization leads to more accurate stabilization than it is achieved using classical stabilizers of the form

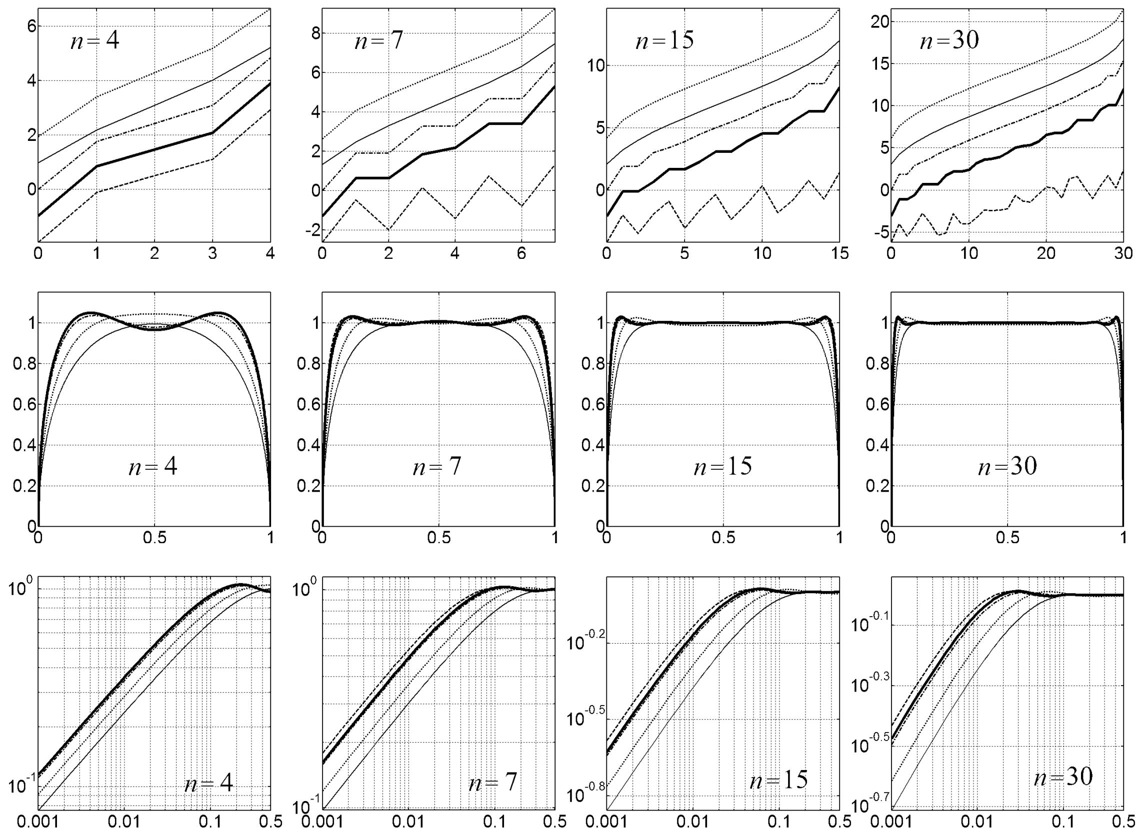


Fig. 13 Stabilization of the binomial families with number of trials $n = 4, 7, 15, 30$. Variance-stabilizing transformations f (top row) and the corresponding conditional standard deviations of the stabilized variables std $\{f(z)|\theta\}$ (middle and bottom row, in logarithmic scale). Solid thick line: optimized monotone stabilizer found by direct search. Dashed line: optimized nonmonotone stabilizer found by direct search. Dash-dotted line: optimized stabilizer found by recursive integral algorithm. Solid thin line: angular stabilizer (13) (Anscombe 1948). Dotted line: angular stabilizer (14) (Freeman and Tukey 1950). The transformations are plotted vertically shifted one with respect to the other, to improve discrimination.

(13)-(14). Figure 14 shows the values C_{f_k} of the stabilization functional for $n = 7$ during the recursive integral optimization. Similar to Figure 5, we show plots for three different initializations f_0: identity, (13), and (14). The sequences (solid lines) converge to 0.0293, which can be further decreased by direct search minimization to 0.0284 (monotone) and 0.0235 (nonmonotone). In Figure 13, we can observe for the binomial family the same qualitative behavior found for the Poisson stabilization, i.e. the optimized monotone stabilizers present a number of constant segments and relaxing monotonicity allows to improve stabilization by introducing oscillations in correspondence with those segments. However, this is improvement is here less pronounced than with the Poisson family (indeed, the differences can be clearly seen only in the plots with logarithmic scale). Curiously, while all monotone stabilizers are found to be symmetric about to their middle point, optimization of nonmonotone stabilizers produced a few asymmetric transformations for particular values of n: e.g., $n = 30$ (as can be seen in Figure 13, top-right) or $n = 12$ (not shown).

Stabilization for the scaled binomial families is equivalent and involves only an obvious rescaling of the stabilizer.

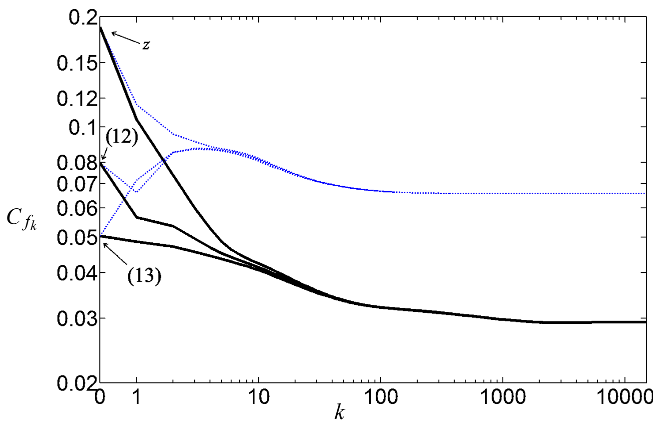


Fig. 14 Optimization of the binomial stabilizer by the recursive integral algorithm. Solid lines: value of the stabilization functional vs. iterations (logarithmic scale). For all three initializations the final value is $C_{f_{15000}} = 0.0293$. Dotted lines: sequence obtained by integration with respect to conditional mean (see Section 7.2).

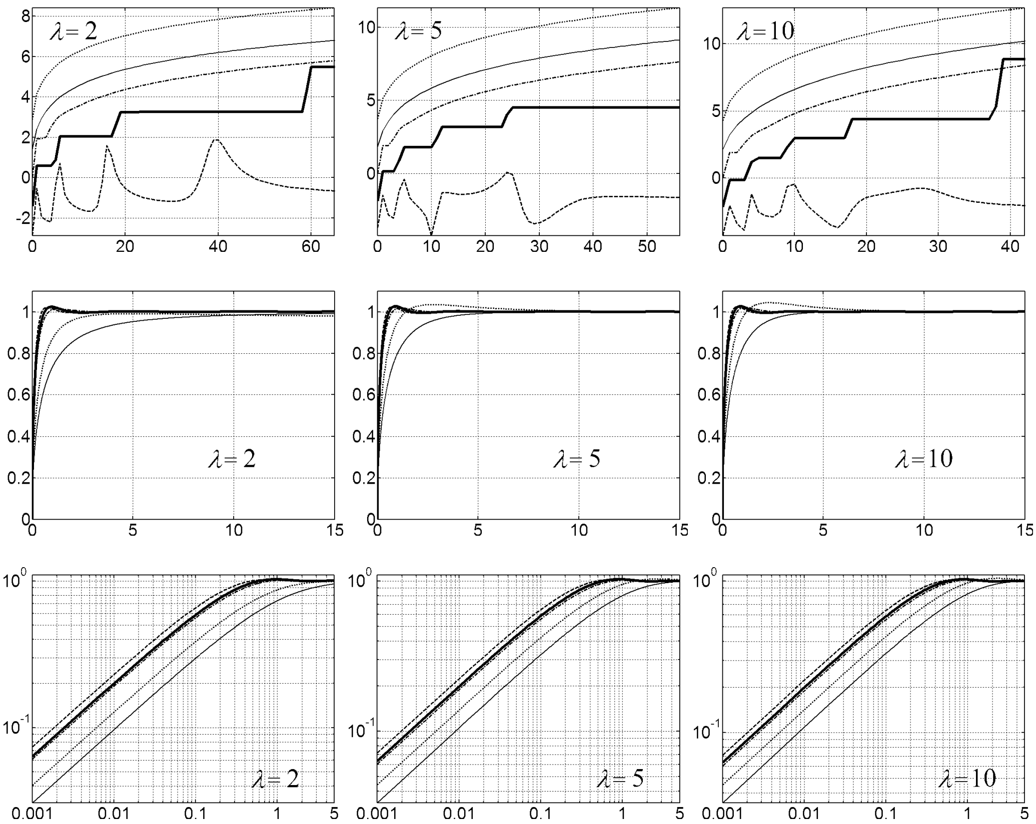


Fig. 15 Stabilization of the negative binomial family with exponent $\lambda = 2, 5, 10$. Variance-stabilizing transformations f (top row) and the corresponding conditional standard deviations of the stabilized variables std $\{f(z)\,|\theta\}$ (middle and bottom row, in logarithmic scale). Solid thick line: optimized monotone stabilizer found by direct search. Dashed line: optimized nonmonotone stabilizer found by direct search. Dash-dotted line: optimized stabilizer found by recursive integral algorithm. Solid thin line: angular stabilizer (15) (Anscombe 1948). Dotted line: angular stabilizer (16) (Laubscher 1961). The transformations are plotted vertically shifted one with respect to the other, to improve discrimination.

6.2 Negative binomial

For the negative-binomial families, we consider the stabilizers by Anscombe (1948) and Laubscher (1961):

$$f(z) = \sqrt{\lambda - \tfrac{1}{2}} \arcsin \sqrt{\tfrac{z+3/8}{\lambda-3/4}}, \qquad (15)$$

$$f(z) = \sqrt{\lambda} \arcsin \sqrt{\tfrac{z}{\lambda}} + \sqrt{\lambda - 1} \arcsin \sqrt{\tfrac{z+3/4}{\lambda-3/2}}. \qquad (16)$$

Here $z \in \mathbb{N}$ and, for any $\lambda > 0$, both these transformation are asymptotically exact as $\theta \to \infty$. Negative-binomial variates share many features with the Poisson ones. In particular, the smaller gets θ, the harder becomes achieving good stabilization. For the following examples, we take $\lambda = 2, 5, 10$, which are the values investigated in the two cited papers. Figure 15 shows the stabilizers (15)-(16), the optimized stabilizers found by the proposed algorithms, and the corresponding conditional standard deviations (in both linear and logarithmic scale). Here, for the optimization we use the same stabilization functional used in Section 3.1. The benefit of optimization can be clearly seen in the figure, particularly for smaller λ values. Again, we can observe that there is a rough correspondence between the locations of the flat segments in the optimized monotone stabilizers and those of the oscillations in the nonmonotone ones.

Stabilizers for the scaled negative binomial families are obtained by an obvious rescaling.

6.3 Poisson

To illustrate the flexibility of the stabilization functional, we now show some additional examples for the Poisson family, using different weights in the definition of the functional (5). In particular, we compare the examples given in Sections 3.1,4.1,5.1 (cost functional "A", Section 3.1) with those obtained by either modifying the localization on θ, replacing $w_\theta = \chi_{[0,15]}$ with $w_\theta = \chi_{[0,3]}$ (cost functional "B"), or by allowing larger stabilization errors, replacing $o_u, o_l = 1.5$, $r'_u, r'_l = 0.2$, with $o_u, o_l = 3$, $r'_u, r'_l = 0.4$ (cost functional "C"). These two modifications can be meaningful, respectively, either when one wants stabilization only for small values of the parameter (in this case $\theta \leq 3$), or when better stabilization for smaller values of the parameter (which is harder to achieve) is preferred to an accurate stabilization overall.

In Figure 16 we show the best monotone and nonmonotone stabilizers found by minimizing these three functionals. These plots allow to distinguish various trade-offs enabled by the weighting. In particular, it is remarkable how improving the stabilization for small θ corresponds to an increase in the amplitude of the oscillations of the conditional standard deviation (functional "C") while relaxing the stabilization for $\theta > 3$ does not provide a conspicuous benefit to the stabilization achieved for $\theta \in [0, 3]$. We discuss on this particular issue in Section 6.6.

The stabilization of the scaled Poisson family can be achieved by a simple change of variables in the argument of the stabilizer.

6.4 Filtered Poisson

It is interesting to look also at the stabilization of filtered Poisson data. Here, we consider the filtered Poisson distributions through the 2-D B_3-spline low-pass filter $h = h_{B_3} \otimes h_{B_3}$, given as tensor product of the 1-D kernels $h_{B_3} = [1\ 4\ 6\ 4\ 1]/16$. This particular case have been studied in Zhang et al (2008), where Anscombe-type stabilizers are utilized. In the same paper, the authors evaluate also the stabilization produced for these distributions by the conditional variance stabilization (CVS) technique by Jansen (2006). Observe that $h \geq 0$, $\|h\|_1 = 1$, and $\|h\|_2 = 5 \times 7 \times 2^{-7} = 0.2734375$. Thus, the filter is mean preserving and $\mu(\theta) = \theta$ just like for the Poisson family, while $\sigma(\theta) = \sqrt{\theta}\,\|h\|_2$. The $5 \times 5 = 25$ Poisson samples which are combined by the filter yield distributions which are much closer to a normal one than the original Poisson distribution is (because of the central-limit theorem). Though the filtered distributions are still obviously discrete, they have much finer granularity and, thus, give way to increased freedom in the optimization of the stabilizer.

In Figure 17, we compare the stabilization achieved by our optimized monotone transformation found by direct search (using the same cost functional "A" as in Section 3.1) against that achieved by the stabilizers of Zhang et al (2008) and Jansen (2006). It can be seen that the optimized stabilizer provides significantly better stabilization, especially for low values of θ.

6.5 Clipped heteroskedastic normal

In Foi et al (2008); Foi (2009a), we proposed a noise model for the digital raw data output of imaging sensors, which can be expressed by a clipped heteroskedastic normal family of distributions. Let $z \sim \mathcal{N}\left(y, \sigma^2(y)\right)$, where $\sigma(y) = \sqrt{ay+b}$, with $a, b \in \mathbb{R}^+$ and $y \geq -\frac{b}{a}$, and define the clipped observations $\tilde{z}$ as

$$\tilde{z} = \max\{0, \min\{z, 1\}\}. \tag{17}$$

Therefore, $\tilde{z} \in \tilde{Z} = [0, 1]$ are distributed according to a doubly censored Gaussian distribution having a generalized probability density function (p.d.f.) of the form

$$\begin{aligned}&\text{pdf}\left[\tilde{z}|y\right](\zeta) = \\ &= \Phi\left(\tfrac{-y}{\sigma(y)}\right)\delta_0(\zeta) + \tfrac{1}{\sigma(y)}\phi\left(\tfrac{\zeta-y}{\sigma(y)}\right)\chi_{[0,1]} + \Phi\left(\tfrac{y-1}{\sigma(y)}\right)\delta_0(1-\zeta),\end{aligned} \tag{18}$$

where δ_0 is the Dirac delta impulse at 0. The conditional expectation and variance of $\tilde{z}$ are

$$\begin{aligned}E\{\tilde{z}|y\} = \tilde{y} = \Phi\left(\tfrac{y}{\sigma(y)}\right)y - \Phi\left(\tfrac{y-1}{\sigma(y)}\right)(y-1) + \\ + \sigma(y)\,\phi\left(\tfrac{y}{\sigma(y)}\right) - \sigma(y)\,\phi\left(\tfrac{y-1}{\sigma(y)}\right),\end{aligned} \tag{19}$$

$$\begin{aligned}&\text{var}\{\tilde{z}|y\} = \tilde{\sigma}^2(\tilde{y}) = \Phi\left(\tfrac{y}{\sigma(y)}\right)\left(y^2 - 2\tilde{y}y + \sigma^2(y)\right) + \\ &+ \tilde{y}^2 - \Phi\left(\tfrac{y-1}{\sigma(y)}\right)\left(y^2 - 2\tilde{y}y + 2\tilde{y} + \sigma^2(y) - 1\right) + \\ &+ \sigma(y)\,\phi\left(\tfrac{y-1}{\sigma(y)}\right)(2\tilde{y} - y - 1) - \sigma(y)\,\phi\left(\tfrac{y}{\sigma(y)}\right)(2\tilde{y} - y),\end{aligned} \tag{20}$$

where ϕ and Φ are the p.d.f. and c.d.f. of the standard normal $\mathcal{N}(0, 1)$, respectively.

We are interested in the stabilization of $\tilde{z}$, i.e. for the family of distributions given by (18). Thus, in this section, the clipped variable $\tilde{z}$ plays the role that the variable z has in the rest of the paper. Because there exists a smooth invertible mapping which links y and $\tilde{y}$ (Foi 2009a), we can use $\theta = \tilde{y}$ as the reference parameter which identifies the distribution within the family.

As an illustrative example, we consider the case given by $a = 1/30$, $b = 0.01$. The conditional standard deviation $\text{std}\{\tilde{z}|\theta\}$ is plotted in Figure 18(middle) (dotted line).

First, let us compare the stabilizing transformation produced by the simple integral (analogous to that in (1))

$$f_0(\tilde{z}) = \int_0^{\tilde{z}} \frac{1}{\tilde{\sigma}(\theta)} d\theta \tag{21}$$

against the optimized transformation obtained by the recursive integral algorithm of Section 3[6]. For this, we use the same stabilization functional used for the binomial family in Section 6.1. Figure 19 shows the evolution of the variance stabilizing transform f_k through 15000 iterations of the recursive algorithm initialized by f_0 (21). The figure shows also the sequence of the conditional standard deviations $\text{std}\{f_k(\tilde{z})|\theta\}$. The initial and optimized stabilizers f_0 and f_{15000} and their respective conditional standard deviations after stabilization are plotted also in Figure 18 (thin

[6] We obviously replace z by $\tilde{z}$ in the formulas in Table 2. The boundedness of the integral (21) is proved in Foi (2009a). Similar arguments can be used to prove the boundedness of the integral (10) applied to the clipped $\tilde{z}$.

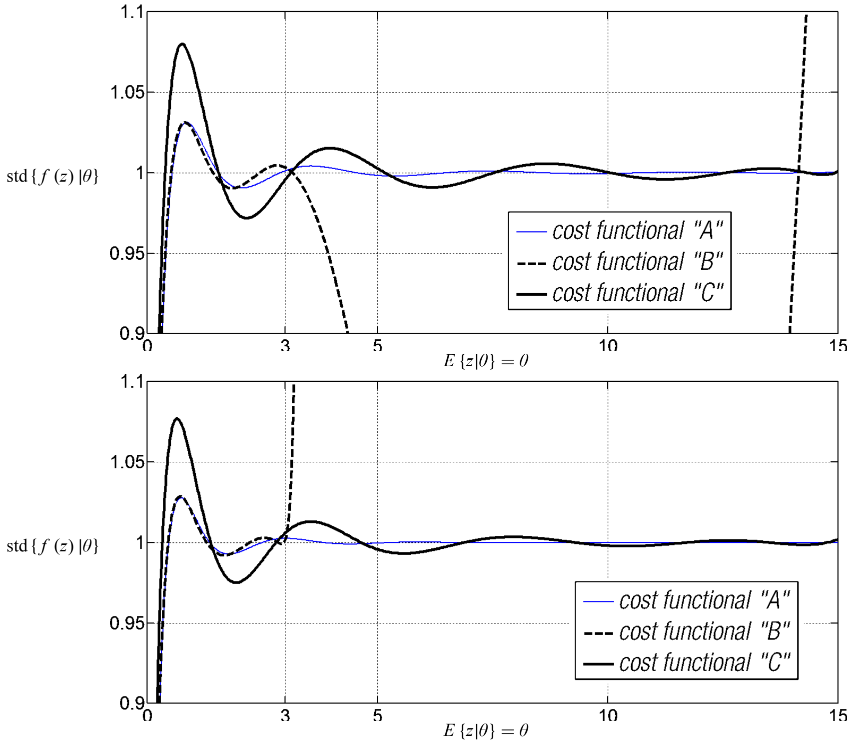


Fig. 16 Conditional standard-deviation std $\{f(z)\,|\theta\}$ of the transformed Poisson variates after stabilization by monotone (top) and nonmonotone (bottom) stabilizers minimizing (direct search) three different cost functionals (see text).

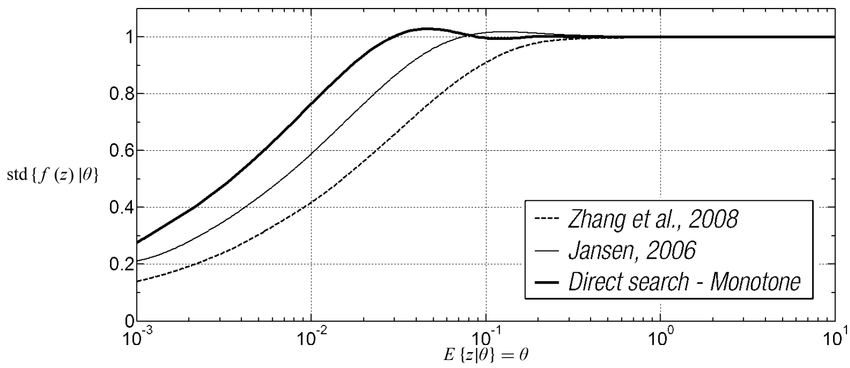


Fig. 17 Stabilization of Poisson data filtered through a 2-D B_3-spline low-pass filter. Comparison between the conditional standard-deviation std $\{f(z)\,|\theta\}$ after stabilization using the Anscombe-type stabilizer (Zhang et al 2008), the CVS technique (Jansen 2006), and the proposed optimized monotone stabilizer found by direct search.

solid and dash-dotted lines). As can be seen in the figures, the stabilizer f_0 produced by (21) is much overshooting the desired unit deviation. This should not surprise: we have already seen in Figure 1, for the Poisson case, that this simple kind of stabilization can overshoot higher than the other more sophisticated approaches.

We now turn to the direct-search optimization. To deal with continuous domain $\tilde{Z} = [0, 1]$ of the distribution family, we use linear resampling of f on an adaptive sampling grid which is progressively refined during the optimization. The optimized stabilizer and the conditional standard deviation of the stabilized variables are plotted in Figure 18 (thick solid lines). A comparison between the conditional standard deviations associated to the two optimized stabilizers shows that the latter has wider oscillations around the desired value $c = 1$, but at the same time its decay towards zero is slower, thus achieving smaller value of the stabilization functional: 0.6646 (direct search) versus 0.6784 (recursive integral).

It is interesting to observe that, while the conditional standard deviations which appear at the denominator of the integrands in (21) and (10) are continuous functions (and thus (21) is continuous also), both optimized stabilizers present discontinuities at both 0 and 1, presumably reflecting the discontinuities at 0 and 1 in the c.d.f. associated to (17),(18). In particular, the value at $\tilde{z} = 1$ of the stabilizer

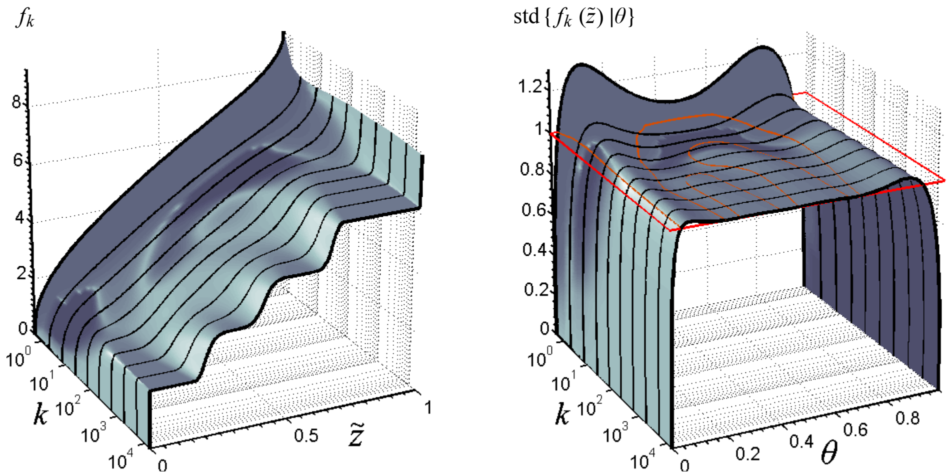


Fig. 19 Recursive integral optimization (algorithm of Table 2) of the stabilizer for the clipped heteroskedastic normal $\tilde{z} = \max\{0, \min\{z, 1\}\}$, $z \sim \mathcal{N}(y, \sigma^2(y))$, in the case $\sigma(y) = \sqrt{y/30 + 0.01}$. The algorithm is initialized by f_0 of Equation (21). Left: sequence of transformations $f_k(z)$; Right: sequence of the corresponding conditional standard deviations std $\{f_k(z)|\theta\}$ (the red contour lines indicate where stabilization is exact).

optimized by the recursive integral algorithm is 7.898, while that of the stabilizer optimized by direct search is 7.921.

Successful application of the optimized stabilizers to the denoising of raw data from digital camera has been demonstrated in Foi (2008, 2009b) and is the subject of a forthcoming publication.

6.6 Splicing transformations

The examples considered so far are optimized over compact parameter supports, using weights w_θ defined as characteristic functions of closed intervals. Outside such a support, the stabilization is not guaranteed and, as seen in Section 6.3 (particularly in Figure 16) stabilization errors can be very large.

While the optimization can be always recomputed for wider supports (i.e. with a different w_θ), in some practical situations, it can be instead useful to extend an optimized transformation so to be applicable for any parameter value, without seriously compromising the stabilization accuracy achieved on the support of w_θ.

This can be operated by *splicing* together two stabilizing transformations: one optimized over a limited range of parameter values (e.g., like those introduced in the previous sections), another one providing stabilization for all parameter values (e.g., a transformation with asymptotical properties, such as those by Anscombe 1948). Let us denote these two stabilizers as f_{opt} and f_{asy}, respectively. Splicing can be realized as

$$f_{\text{spliced}}(z) = \\ = \int_{z_a}^{z} w_{\text{mix}}(v)\, f'_{\text{opt}}(v) + (1 - w_{\text{mix}}(v))\, f'_{\text{asy}}(v)\, dv + a, \quad (22)$$

where $w_{\text{mix}} : Z \to [0, 1]$ is a mixing window function that regulates the transition from f_{opt} to f_{asy} and $z_a \in Z$ and $a \in \mathbb{R}$ are arbitrary constants. In case of discrete distributions, the integral and derivatives in (22) are assumed in generalized sense.

We illustrate this by considering the particular situation where the monotone Poisson stabilizer optimized over $\chi_{[0,15]}$ by direct search (Section 4.1) f_{opt} is spliced together with Anscombe (1948) root transformation $f_{\text{asy}}(z) = 2\sqrt{z + 3/8}$. The derivatives in (22) are here computed as the finite differences $f'_{\text{opt}}(z) = f_{\text{opt}}(z+1) - f_{\text{opt}}(z)$ and $f'_{\text{asy}}(z) = f_{\text{asy}}(z+1) - f_{\text{asy}}(z)$, while the integral becomes a cumulative sum.

In choosing the mixing window function, we first observe that the Anscombe transformation f_{asy} provides excellent stabilization for $\theta \geq 4$ (see Figures 1 and 7), while for smaller values of θ f_{opt} is a much better stabilizer (see Figure 11). However, f_{opt} ensures optimal stabilization for all $\theta \leq 15$. We then analyze what is the influence of the samples at z to the conditional distributions. More precisely, for each $\zeta \in Z$, we look at $\max_\theta \Pr[z = \zeta|\theta]$ over specific ranges of parameter θ. If this quantity is very small at a certain ζ, then modifications of the value of $f(\zeta)$ will not result in significant changes to the conditional standard deviation std$\{f_{\text{spliced}}(z)|\theta\}$. In Figure 20 we show plots of $\max_{\theta\in\Theta_j} \Pr[z = \zeta|\theta]$ as function of ζ for $\Theta_1 = [0, 4)$, $\Theta_2 = [4, 50)$, $\Theta_3 = [50, +\infty)$. We can see that the stabilization on Θ_1 is practically not affected by the transformation for $\zeta \geq 20$, with the most influential values of ζ being the first few smallest. Likewise, the stabilization on Θ_3 is practically not affected by $\zeta \leq 20$.

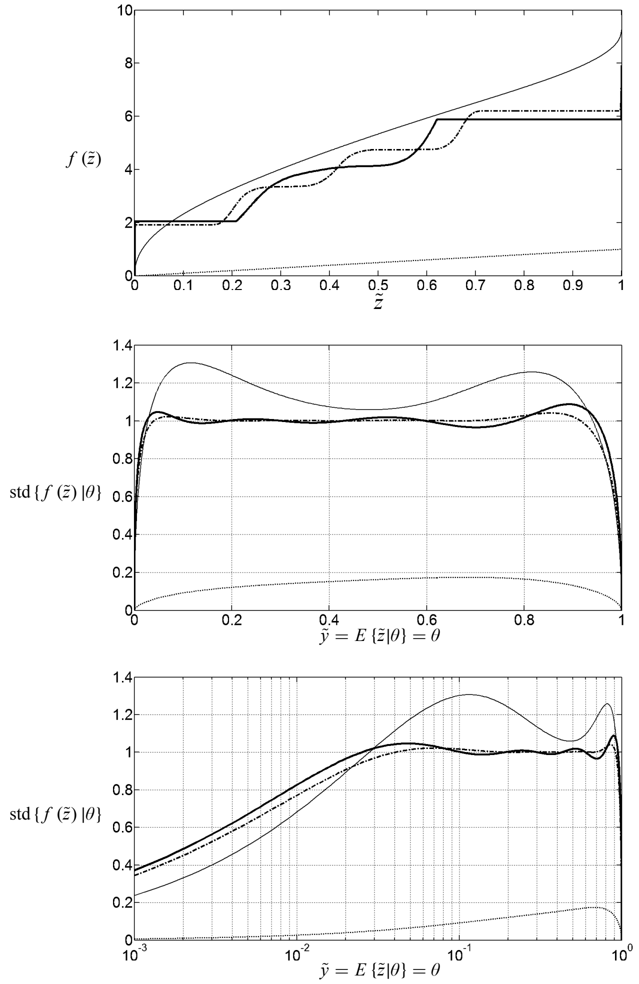


Fig. 18 Stabilization of the clipped heteroskedastic normal $\tilde{z} = \max\{0, \min\{z, 1\}\}$, $z \sim \mathcal{N}(y, \sigma^2(y))$, in the case $\sigma(y) = \sqrt{y/30 + 0.01}$: transformations and conditional standard-deviations. Solid thin line: stabilizer (21). Dash-dotted line: optimized stabilizer found by recursive integral algorithm. Solid thick line: optimized monotone stabilizer found by direct search. Dotted line: identity (i.e. no stabilization).

Thus, as mixing window we take the smooth function (see Figure 21)

$$w_{\mathrm{mix}}(\zeta) = \begin{cases} 1 & 0 \leq \zeta \leq 5, \\ \frac{1}{2}\left(1 + \cos\left(\frac{(\zeta-5)\pi}{14}\right)\right) & 5 \leq \zeta \leq 19, \\ 0 & 19 \leq \zeta. \end{cases} \quad (23)$$

This choice ensures that f_{spliced} inherits the good stabilization of f_{opt} for small $\theta \in \Theta_1$ and at the same time have practically the same stabilizing properties of f_{asy} for large $\theta \in \Theta_3$. The in-between values $\theta \in \Theta_2$ are those where splicing might ruin the stabilization offered by f_{asy} or f_{opt} and therefore are those where the stabilization produced by f_{spliced} (i.e. $\mathrm{std}\{f_{\mathrm{spliced}}(z)|\theta\}$) deserves most to be inspected. The transformations f_{asy}, f_{opt}, f_{spliced}, and the conditional standard deviations $\mathrm{std}\{f_{\mathrm{asy}}(z)|\theta\}$, $\mathrm{std}\{f_{\mathrm{opt}}(z)|\theta\}$, $\mathrm{std}\{f_{\mathrm{spliced}}(z)|\theta\}$ are shown in Figure 22 and Figure 23,

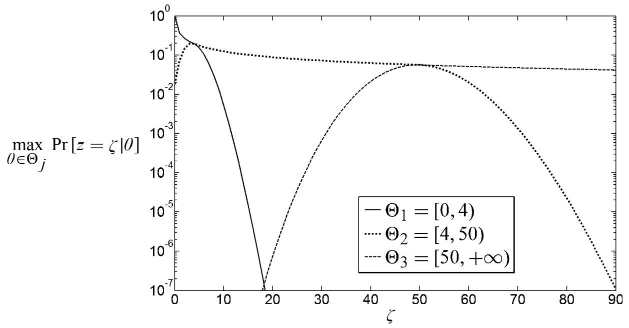


Fig. 20 $\max_{\theta\in\Theta_j} \Pr[z = \zeta|\theta]$ over the parameter ranges $\Theta_1 = [0, 4)$, $\Theta_2 = [4, 50)$, $\Theta_3 = [50, +\infty)$.

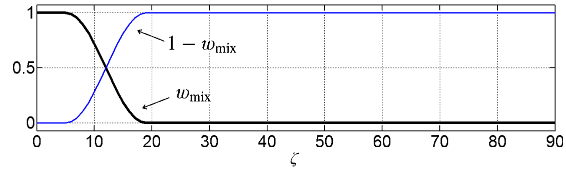


Fig. 21 Mixing window function w_{mix} (23).

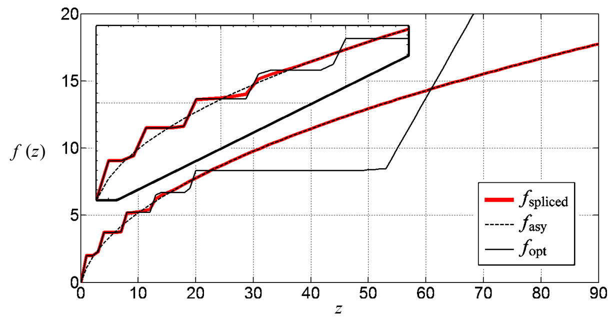


Fig. 22 Splicing together the Anscombe root transformation f_{asy} with the optimized monotone stabilizer f_{opt} obtained by direct search. The enlarged fragment shows the portion $0 \leq z \leq 25$.

respectively. The loss of stabilization outside of the interval $[0, 15]$ is clearly visible in the plot of $\mathrm{std}\{f_{\mathrm{opt}}(z)|\theta\}$. As can be seen in the enlarged fragment shown in Figure 23, the spliced transformation f_{spliced} yields stabilization which, for all $\theta \geq 0$, is within a 1.5% tolerance of the best among $\mathrm{std}\{f_{\mathrm{asy}}(z)|\theta\}$, $\mathrm{std}\{f_{\mathrm{opt}}(z)|\theta\}$. In particular, the behavior of $\mathrm{std}\{f_{\mathrm{spliced}}(z)|\theta\}$ and $\mathrm{std}\{f_{\mathrm{opt}}(z)|\theta\}$ for very small θ values (i.e. $\theta \leq 4$) is essentially identical, while

$$e_{f_{\mathrm{spliced}}}(\theta) = \mathrm{std}\{f_{\mathrm{spliced}}(z)|\theta\} - 1 \underset{\theta\to+\infty}{\longrightarrow} 0$$

and

$$e_{f_{\mathrm{asy}}}(\theta) = \mathrm{std}\{f_{\mathrm{asy}}(z)|\theta\} - 1 \underset{\theta\to+\infty}{\longrightarrow} 0$$

converge to zero at the same rate.

7 Discussion and conclusions

In this paper we aimed at fulfilling a number of goals. We have introduced a clear formulation of variance stabilizati-

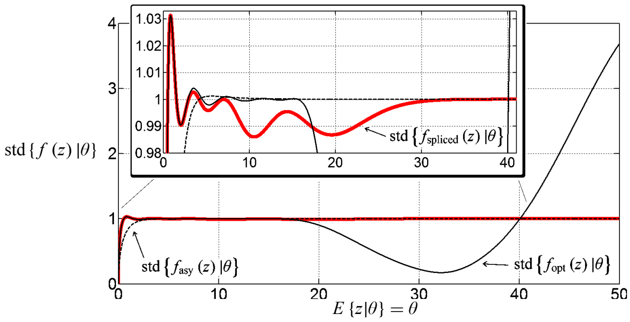


Fig. 23 Conditional standard-deviations of the transformed Poisson variates after stabilization by the Anscombe root transformation f_{asy} (dashed line), by the optimized monotone stabilizer f_{opt} (solid thin line), and by the spliced transformation f_{spliced} (thick red line).

on as a cost-minimization problem, under which different variance-stabilizing transformations can be evaluated and compared. While such an approach is widespread, with successful application to many optimization problems in applied mathematics and engineering, it had surprisingly never been formulated in the context of variance stabilization. This formalism, allows one to look for approximate and localized stabilization, using weights which allow a certain tolerance for small stabilization errors as well as to restrict around only some specified range of the parameter values. We have then presented two ways to perform the minimization of the cost functional: first by an algorithm based on recursive integration, and then by direct search. In both cases, the variance-stabilizing transformations are nonparametric. Further, we have demonstrated that improved stabilization can be achieved by relaxing the commonly assumed requirement of invertibility of the variance-stabilizing transformation, supplanting it with the invertibility of the transformation of the expectations, which is the actual requirement in regression. We have also shown the application of the proposed approach to various distribution families, resulting in optimized stabilizers that outperform the conventional stabilizers found in the literature.

This paper can be partly seen as a continuation of the works by Efron (1982) and especially Tibshirani (1988, 1986a), in the sense that we provide an alternative and practical method for designing optimal stabilizers based on recursive refinement of the variance-stabilizing transform using an approximate technique. The main formal difference between our work and all previous ones, is that we treat variance stabilization as a cost-minimization problem. We note that Tibshirani does conjecture a minimization problem solved by AVAS (Tibshirani 1988, Section 5.3). However, his formulation is not relevant to the stabilization problem itself, since in Tibshirani (1988) the minimization is of a quadratic fit for the regression constrained to the class of exact stabilizers (which in most situations happens to be empty) and not a minimization of the stabilization error.

The stepwise behavior of all the presented optimized monotone stabilizers underlines the inadequacy of the classical approach to stabilization based on asymptotic calculus with smooth functions in explicit parametric form.

7.1 Specific features of the proposed recursive integral stabilizer

Let us discuss the particular refinement function used in our recursive integral algorithm of Section 3, in relation to other stabilizers in integral form previously proposed in the litterature. For convenience, these stabilizers are summarized in Table 3. They can be mainly categorized according to four features: iterative or non-iterative, integrand based on $\text{pdf}[z|\theta](\theta)$ or on $1/\sigma(\theta)$, use of weights for the integrand or not, integration with respect to conditional mean or with respect to conditional median.

The fact that one should use integration with respect to conditional median is explained by Efron (1982): it essentially follows from the median of a distribution being invariant through transformation by a monotone function f (i.e., $\text{med}\{f(z)|\theta\} = f(\text{med}\{z|\theta\})$), whereas the mean is typically not invariant through a nonlinear transformation. In particular, Efron (1982) proves that for the so-called "general transformation families" of distributions (i.e. one-parameter families of distributions which are related to the standard normal through some smooth transformations) the transformation that achieves exact stabilization exists and has the form

$$f(\xi) = \int_{z_a}^{\xi} \frac{\text{pdf}[z|\theta](\theta)}{\phi(0)}\, d\left[\,\text{med}\{z|\theta\}\,\right] + a, \tag{24}$$

where ϕ denotes the p.d.f. of the standard normal distribution $\mathcal{N}(0,1)$. It is easy to realize that for normal distributions $\text{pdf}[z|\theta](\theta) = \phi(0)/\sigma(\theta)$, from which follows the commonly employed variance stabilizing transformation (1). In this sense, by relying on a normal approximation of $\text{pdf}[z|\theta]$ for the integrand, (24) becomes

$$f(\xi) = \int_{z_a}^{\xi} \frac{1}{\sigma(\theta)}\, d\left[\,\text{med}\{z|\theta\}\,\right] + a, \tag{25}$$

where the integration is still with respect to conditional median. A non-iterative algorithm based on (25) is described in a technical report by Tibshirani (1986a, Section 3). However, only a very brief note on this is eventually given in Section 5.6 of the paper Tibshirani (1988), where single iteration algorithms based on (24),(25) are declared as "not useful". Ultimately, the AVAS algorithm presented in Tibshirani (1988) uses recursive stabilization with a refinement function,

$$r_k(\xi) = \int_{f_k(z_a)}^{\xi} \frac{1}{\sigma_{f_k}(\theta)}\, d\left[\,E\{f_k(z)|\theta\}\,\right] + a, \tag{26}$$

Table 3 Different variance-stabilizing transforms in integral form. See Sections 7.1 and 7.2.

NON ITERATIVE	
eq. (1) (e.g., Bartlett 1936)	$f(\xi) = \int_{z_a}^{\xi} \frac{1}{\sigma(\theta)} d\left[E\{f_k(z)\mid\theta\}\right] + a$
exact stabilizer for GTFs (Efron 1982)	$f(\xi) = \int_{z_a}^{\xi} \frac{\mathrm{pdf}[z\mid\theta](\theta)}{\phi(0)} d\left[\ \mathrm{med}\{z\mid\theta\}\ \right] + a$
normal approximation in the integrand (Tibshirani 1986a)	$f(\xi) = \int_{z_a}^{\xi} \frac{1}{\sigma(\theta)} d\left[\ \mathrm{med}\{z\mid\theta\}\ \right] + a$
ITERATIVE $f_{k+1}(z) = r_k(f_k(z))$	
AVAS (Tibshirani 1988)	$r_k(\xi) = \int_{f_k(z_a)}^{\xi} \frac{1}{\sigma_{f_k}(\theta)} d\left[E\{f_k(z)\mid\theta\}\right] + a$
proposed (Table 2 of Section 3)	$r_k(\xi) = \int_{f_k(z_a)}^{\xi} 1 - \frac{w_\theta(\theta)\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)}{\sigma_{f_k}(\theta)} d\left[\mathrm{med}\{f_k(z)\mid\theta\}\right] + a$
modified, integration w.r.t. to mean (Section 7.2)	$r_k(\xi) = \int_{f_k(z_a)}^{\xi} 1 - \frac{w_\theta(\theta)\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)}{\sigma_{f_k}(\theta)} d\left[E\{f_k(z)\mid\theta\}\right] + a$

thus following a normal approximation of the p.d.f. for both the integrand and the integrator.

Our refinement function (Table 2, Section 3)

$$r_k(\xi) = \int_{f_k(z_a)}^{\xi} 1 - \frac{w_\theta(\theta)\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)}{\sigma_{f_k}(\theta)} d\left[\mathrm{med}\{f_k(z)\mid\theta\}\right] + a \tag{27}$$

can be interpreted as a recursive form of (25), where the weighted integrand (11) is used instead of the basic $\sigma_{f_k}^{-1}(\theta)$, which appears as the integrand in (26).

We have already shown that indeed multiple iterations are necessary in order to achieve good stabilization (see Figures 4, 5, 14, 19). Likewise, the effect of weighting has been shown in Section 6.3. In the next section we emphasize the importance of the integration with respect to the conditional medians.

7.2 Conditional median vs. conditional mean

If, in our recursive integral algorithm, instead of integrating with respect to the conditional median, we integrate with respect to the conditional mean, i.e. replacing (27) by

$$r_k(\xi) = \int_{f_k(z_a)}^{\xi} 1 - \frac{w_\theta(\theta)\varphi\left(\overline{e_{f_k}}(\theta)\right)\overline{e_{f_k}}(\theta)}{\sigma_{f_k}(\theta)} d\left[E\{f_k(z)\mid\theta\}\right] + a,$$

the recursion fails to effectively decrease the stabilization functional. For example, we can see in Figure 5, that for the Poisson family and using exactly the same settings as in Section 3.1, the sequence C_{f_k} converges, regardless of the particular initialization, to 0.2543. Although it is obviously much better than leaving the data untransformed, this value is larger than that corresponding to the root stabilizer by Freeman and Tukey (1950) while the improvement with respect to the Anscombe (1948) transformation is marginal. The same situation can be found for the other families and, in particular, for the binomial (Figure 14), with C_{f_k} converging to 0.0659.

We wish also to emphasize that our recursive integral algorithm with integration with respect to the conditional median, though not reaching the optimum stabilizers in the class of monotone transformations, takes us often extremely close to this optimum, as much that it is hard to find significant differences between the plots of the conditional standard deviations corresponding to the optimum found by direct search and the solution produced by the recursive integral algorithm.

7.3 Optimization of higher-order moments and additional penalties

Here we have exclusively considered optimization of transformations for the conditional variance. However, it is evident that, provided inclusion of higher-order moments in the stabilization functional, its minimization can be used for the joint optimization of variance, skewness, kurtosis, etc. While, in its current form, the recursive integral algorithm is addressing only the stabilization of the variance, the direct-search procedures are directly applicable to the joint optimization of multiple moments, as they explicitly work on the minimization of the functional. By proper choice of weights in the functional, one can easily look for solutions which favor normality over stability.

One would typically demand also some good conditioning for the inversion of the mapping $E\{z|\theta\} \mapsto E\{f(z)|\theta\}$, in order to avoid amplification of estimation errors. For instance, the mapping could be constrained to have derivative included within a closed strictly positive range.

Since the direct search becomes practically unfeasible when the number of dimensions is large, additional constraints might be used as a pragmatic way for dimensionality reduction of the search space. In particular, without imposing additional constraints or penalties on f (such as bounded variation or some explicit or implicit parametric representation), it makes little sense to consider a nonmonotone f for continuous distributions, as an unconstrained optimization will likely introduce oscillations of large amplitude and high frequency.

7.4 Open problems

We have focused on computational procedures to decrease the stabilization functional, mainly ignoring the fundamental problem of ensuring the existence of the minimizer in particular function classes.

The transformations obtained by direct search give evidence of strong structures underlying the optimal stabilizers, which suggests a potential use of ad-hoc optimization algorithms for the minimization of the stabilization functional. Exploiting these structures might be critical for dealing with transformations with large supports, for which the Nelder-Mead algorithm is inefficient. Moreover, the proposed optimization techniques necessarily yield transformations in nonparametric form, whereas it is apparent that in many cases an underlying simple parametric model might exist (see, e.g., Figure 18), which could be more practical both for further analysis and for use in applications.

8 Acknowledgments

This work was supported by the Academy of Finland (Finnish Programme for Centres of Excellence in Research 2006-2011, Finland Distinguished Professor Programme 2007-2010, and Postdoctoral Researcher's Project 2009-2011).

References

Adell JA, Jodrá P (2005) The median of the Poisson distribution. Metrika 61(3):337–346

Anscombe FJ (1948) The transformation of Poisson, binomial and negative-binomial data. Biometrika 35(3/4):246–254

Arsenault HH, Denis M (1981) Integral expression for transforming signal-dependent noise into signal-independent noise. Opt Lett 6(5):210–212

Bartlett MS (1936) The square root transformation in analysis of variance. J R Statist Soc Suppl 3(1):68–78

Blaylock JR, Smallwood DM (1985) Box-Cox transformations and a heteroscedastic error variance: Import demand equations revisited. International Statistical Review 53(1):91–97

Conn AR, Scheinberg K, Vicente LN (2009) Introduction to derivative-free optimization. MPS-SIAM series on optimization

Curtiss JH (1943) On transformations used in the analysis of variance. The Annals of Mathematical Statistics 14(2):107–122

DiCiccio TJ, Monti AC, Young GA (2006) Variance stabilization for a scalar parameter. J R Statist Soc B, 68(2)

Efron B (1981) Transformation theory: How normal is a family of distributions? Technical Report 69, Department of Statistics, Stanford University

Efron B (1982) Transformation theory: How normal is a family of distributions? The Annals of Statistics 10(2):323–339

Foi A (2008) Direct optimization of nonparametric variance-stabilizing transformations. presented at 8èmes Rencontres de Statistiques Mathématiques, CIRM, Luminy, France

Foi A (2009a) Clipped noisy images: Heteroskedastic modeling and practical denoising. Signal Processing 89(12):2609–2629, http://dx.doi.org/10.1016/j.sigpro.2009.04.035

Foi A (2009b) Removal of signal-dependent noise: the BM3D filter and optimized variance-stabilizing transformations. presented at "Patch-based Image Representation, Manifolds and Sparsity" Minisymposium, INRIA Centre de Rennes Bretagne Atlantique, IRISA, Rennes, France, material online, http://www.irisa.fr/videos/seminaire_Irisa/Vista

Foi A, Trimeche M, Katkovnik V, Egiazarian K (2008) Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. IEEE Trans Image Process 17(10):1737–1754

Freeman M, Tukey J (1950) Transformations related to the angular and the square root. The Annals of Mathematical Statistics 21(4):607–611

Fryzlewicz P (2008) Data-driven wavelet-Fisz methodology for nonparametric function estimation. Electronic Journal of Statistics 2:863–896

Fryzlewicz P, Delouille V (2005) A data-driven Haar-Fisz transform for multiscale variance stabilization. Proc 13th IEEE/SP Workshop on Statistical Signal Processing pp 539–544

Fryzlewicz P, Nason GP (2004) A Haar-Fisz algorithm for Poisson intensity estimation. J Comp Graph Stat 13:621–638

Greene W (2000) Econometric Analysis, 4th ed. Prentice Hall

Guan Y (2009) Variance stabilizing transformations of Poisson, binomial and negative binomial distributions. Statistics and Probability Letters doi:10.1016/j.spl.2009.04.010

Hirakawa K, Parks TW (2006) Image denoising using total least squares. IEEE Trans Image Process 15(9):2730–2742

Jansen M (2006) Multiscale Poisson data smoothing. J R Statist Soc B, 68(2):27–48

Kasturi R, Walkup JF, Krile TF (1983) Image restoration by transformation of signal-dependent noise to signal-independent noise. Applied Optics 22(22):3537–3542

Kelley CT (1999) Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition. SIAM J Optim 10(1):43–55

Kolaczyk ED (1997) Nonparametric estimation of gamma-ray burst intensities using Haar wavelets. The Astrophysical Journal 483:340–349

Kolaczyk ED (1999) Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. Statistica Sinica 9:119–135

Kolda TG, Lewis RM, Torczon V (2003) Optimization by direct search: New perspectives on some classical and modern methods. SIAM Review 45(3):385–482

Kolmogorov AN, Fomin SV (1975) Introductory Real Analysis. Dover Publications, Inc., New York

Laubscher NF (1961) On stabilizing the binomial and negative binomial variances. Journal of the American Statistical Association 56(293):143–150

N N (2006) Variance stabilization. In: Encyclopedia of Statistical Sciences, John Wiley & Sons, http://doi.wiley.com/10.1002/0471667196.ess6048

Nason GP (2008) Wavelet Methods in Statistics with R. Springer

Nelder JA, Mead R (1965) A simplex method for function minimization. The Computer Journal 7:308–313

Nunes MA, Nason GP (2008) A multiscale variance stabilization for binomial sequence proportion estimation. Research Report – Statistics 3, University of Bristol

Prucnal PR, Saleh BEA (1981) Transformation of image-signal-dependent noise into image signal-independent noise. Optics Letters 6(7)

Sakia RM (1992) The Box-Cox transformation technique: A review. The Statistician 41(2):169–178

Starck JL, Murtagh F, Bijaoui A (1998) Image Processing and Data Analysis. Cambridge University Press

Taylor JMG (2006) Transformations—II. In: Encyclopedia of Statistical Sciences, John Wiley & Sons, http://doi.wiley.com/0471667196.ess0746.pub2

Tibshirani R (1986a) Estimating transformations for regression. Technical Report 12, Department of Statistics, University of Toronto

Tibshirani R (1986b) Estimating transformations for regression: a variation on ACE. Technical Report 4, Department of Statistics, University of Toronto

Tibshirani R (1988) Estimating transformations for regression via additivity and variance stabilization. Journal of the American Statistical Association 83(402):394–405

Tippet LHC (1934) Statistical methods in textile research. Part 2, Uses of the binomial and Poisson distributions. Shirley Inst Mem 13:35–72

Veevers A, Tweedie MCK (1971) Variance-stabilizing transformation of a Poisson variate by a beta function. Applied Statistics 20(3):304–308

Zhang B, Fadili MJ, Starck JL (2008) Wavelets, ridgelets and curvelets for Poisson noise removal. IEEE Trans Image Process 17(7):1093–1108