Electron. Commun. Probab. **0** (2020), article no. 0, 1–12.
https://doi.org/10.1214/YY-TN
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

The least favorable noise

Philip A. Ernst * Abram M. Kagan † L.C.G. Rogers ‡

We dedicate this work to our colleague, mentor, and friend, Professor Larry Shepp (1936–2013)

Abstract

Suppose that a random variable X of interest is observed perturbed by independent additive noise Y. This paper concerns the "the least favorable perturbation" $\hat{Y}_\varepsilon$, which maximizes the prediction error $E(X - E(X|X+Y))^2$ in the class of Y with $\text{var}(Y) \le \varepsilon$. We find a characterization of the answer to this question, and show by example that it can be surprisingly complicated. However, in the special case where X is infinitely divisible, the solution is complete and simple. We also explore the conjecture that noisier Y makes prediction worse.

Keywords: Least favorable perturbation; self-decomposable random variable; infinitely divisible distributions.
MSC2020 subject classifications: Primary 60E07; 60E10, Secondary 60E05.
Submitted to ECP on March 17, 2021, final version accepted on N/A.

1 Introduction.

Suppose that on a probability space we observe $X + Y$, where X and Y are independent random variables, X being a square-integrable random variable of interest, and Y being an additive noise perturbation. The prediction error

$$E\{X - E(X|X+Y)\}^2 = \text{var}X - \text{var}E(X|X+Y) \tag{1.1}$$

depends on Y of course, and thus a natural question is 'What would be the *worst* noise Y we could add to X?' In other words, given the law of X, how would we choose the law of Y to maximize the prediction error in equation (1.1), or equivalently, how would we find

$$\inf_Y \text{var}E(X|X+Y) \; ? \tag{1.2}$$

Since the mean of $E[X|X+Y]$ is fixed and equal to EX, an equivalent question is to choose the law of Y so as to achieve

$$\inf_Y E\{\, E[X|X+Y]^2 \,\}. \tag{1.3}$$

*Department of Statistics, Rice University, United States of America. E-mail: philip.ernst@rice.edu
†Department of Mathematics, University of Maryland, United States of America. E-mail: amk@math.umd.edu
‡Statistical Laboratory, University of Cambridge, United Kingdom. E-mail: chris@statslab.cam.ac.uk

If we think of what happens when $Y = \sigma Z$, where $Z \sim N(0,1)$, we quickly realize that as $\sigma \to \infty$ we have[1] for any ξ

$$E[X|X+\sigma Z=\xi] = \frac{\int x f(x) \exp\{\,-(\xi-x)^2/2\sigma^2\,\}\,dx}{\int f(x) \exp\{\,-(\xi-x)^2/2\sigma^2\,\}\,dx} \to EX, \tag{1.4}$$

so that the minimization in (1.3) has a trivial solution unless we bound the variance of Y. So we will focus on the problem

$$\inf\{\,E\big(\,E[X|X+Y]^2\,\big):\ \text{var}(Y)\le\varepsilon\,\}, \tag{1.5}$$

where $\varepsilon > 0$ is given. We then have a number of questions:

Question 1: Can we find an explicit solution to (1.5)?

Question 2: Can we characterize the solution to (1.5)?

Question 3: Are there situations with explicit solutions?

Question 4: Does more noise mean worse prediction?

The fact that we asked Question 2 means that the answer to Question 1 has to be ‘No’; however, the answer to Question 2 is ‘Yes’, and we deal with this in Section 2. The answer to Question 3 is also ‘Yes’, as we show in Section 3; if the law of X is infinitely divisible, then we can find the minimizing Y. Question 4 is not very precisely posed, but in Section 4 we give a simple example which shows that the answer is ‘No’. However, if Y is self-decomposable we have a partial result in this direction. In Section 5, we present an analysis of the case where X is binomial and Y is integer-valued, and we give a number of numerical examples which point to the diversity and complexity of the solutions in general.

We conclude this section with some brief remarks about the broader literature. The spirit of the present work is most closely aligned with the lines of inquiry in [2, 3, 4]. We also note that the focus of this paper largely moves in the opposite direction of stochastic filtering, in which one (usually) seeks to get as close as possible to X (see, e.g., [1], and references therein). This being said, the answer to Question 4 should be of particular interest to the stochastic filtering community.

2 Characterizing the solution.

Firstly, we observe that the objective to be minimized,

$$\text{var}\big(\,E[X|X+Y]\,\big) = E\big(\,E[X-\mu|X+Y]^2\,\big) = E\big(\,E[X|X+Y]^2\,\big) - \mu^2 \tag{2.1}$$

is unaltered if we shift X or Y by a constant, so we may assume that the means of X and Y are set to zero. *We shall insist throughout that the mean of Y is 0,* though we make no such restriction on the mean of X, largely for aesthetic reasons. We denote by σ^2 the variance of X, and by M_2 the second moment $E[X^2]$ of X.

Let F denote the (known) distribution of X, and G denote the distribution of Y, which is to be found, subject to the constraints

$$\int G(dy) = 1, \qquad \int yG(dy) = 0, \qquad \int y^2 G(dy) \le \varepsilon. \tag{2.2}$$

[1] Here, f denotes the density of X.

We denote by $\mathcal{P}$ the set of G satisfying (2.2). Once $G \in \mathcal{P}$ has been chosen, we know that there is some measurable function φ_G such that

$$E[X|X+Y] = \varphi_G(X+Y), \tag{2.3}$$

where φ_G has the defining property that

$$E[\,\{X - \varphi_G(X+Y)\}\,\lambda(X+Y)\,] = 0 \tag{2.4}$$

for any measurable λ which is bounded, or indeed for which the expectation is defined. Thus with $G \in \mathcal{P}$ fixed, for any λ for which the integrals in (2.4) are defined,

$$\begin{aligned} \Phi(G) &\equiv E\big[\,E[X|X+Y]^2\,\big] & (2.5)\\ &= E[\,\varphi_G(X+Y)^2\,] \\ &= \iint \varphi_G(x+y)^2\;F(dx)G(dy) \\ &= \iint \big\{\,\varphi_G(x+y)^2 + 2(x-\varphi_G(x+y))\lambda(x+y)\,\big\}\;F(dx)G(dy) \\ &\geq \inf_{\varphi} \iint \big\{\,\varphi(x+y)^2 + 2(x-\varphi(x+y))\lambda(x+y)\,\big\}\;F(dx)G(dy) \\ &= \iint \big\{\,x^2 - (x-\lambda(x+y))^2\,\big\}\;F(dx)G(dy) \\ &\equiv \Phi_{\text{lo}}(G;\lambda). & (2.6) \end{aligned}$$

Thus for any $G \in \mathcal{P}$, any λ and any constants $\gamma \geq 0$, α, β, there is some $z \geq 0$ such that

$$\begin{aligned} \Phi(G) &\geq \Phi_{\text{lo}}(G;\lambda) + \alpha\Big(\int G(dy) - 1\Big) + \beta\int y\,G(dy) + \gamma\Big(\int y^2\,G(dy) + z - \varepsilon\Big) \\ &= M_2 + \iint \big\{\,\alpha + \beta y + \gamma y^2 - (x-\lambda(x+y))^2\,\big\}F(dx)G(dy) - \alpha + \gamma(z-\varepsilon) \\ &\geq M_2 - \alpha - \gamma\varepsilon + \int \Big\{\alpha + \beta y + \gamma y^2 - \int (x-\lambda(x+y))^2 F(dx)\Big\}\;G(dy) & (2.7) \end{aligned}$$

If we define $\mathcal{D}$ to be the space of dual-feasible variables $(\alpha,\beta,\gamma,\lambda)$ satisfying $\gamma \geq 0$ and the condition

$$0 \leq \alpha + \beta y + \gamma y^2 - \int (x-\lambda(x+y))^2 F(dx) \qquad \forall y, \tag{2.8}$$

then it is clear from (2.7) that for $(\alpha,\beta,\gamma,\lambda) \in \mathcal{D}$ we have for any $G \in \mathcal{P}$ the lower bound

$$\Phi(G) \geq M_2 - \alpha - \gamma\varepsilon. \tag{2.9}$$

It follows therefore that

$$\inf_{G\in\mathcal{P}} \Phi(G) \geq \sup_{(\alpha,\beta,\gamma,\lambda)\in\mathcal{D}} \big\{\,M_2 - \alpha - \gamma\varepsilon\,\big\}. \tag{2.10}$$

The inequality (2.10) is a primal-dual inequality familiar from constrained optimization problems. We expect that under technical conditions it is possible to prove that equality in (2.10) is *necessary* for optimality, but we avoid attempting to prove this. We do so because establishing this (if true) does not help us to *identify* an optimal solution; and because *sufficiency* is all we need to prove optimality in examples.

Theorem 2.1. *Suppose that* $G_* \in \mathcal{P}$, $z_* \geq 0$ *and* $(\alpha_*,\beta_*,\gamma_*,\lambda_*) \in \mathcal{D}$ *satisfy the complementary slackness conditions*

$$\begin{aligned} 0 &\equiv \Big[\alpha_* + \beta_* y + \gamma_* y^2 - \int (x-\lambda_*(x+y))^2 F(dx)\Big]\;G_*(dy) & (2.11)\\ 0 &= \gamma_* z_*, & (2.12) \end{aligned}$$

and

$$z_* = \varepsilon - \int y^2\, G_*(dy), \qquad \lambda_* = \varphi_{G_*}. \tag{2.13}$$

Then G_* is optimal.

Proof. Consider $q \equiv M_2 - \alpha_* - \gamma_* \varepsilon$, which is a lower bound for the right-hand side of (2.10), since $(\alpha_*, \beta_*, \gamma_*, \lambda_*) \in \mathcal{D}$. Now we return to (2.7) and work back through the steps, putting G_* for G and λ_* for λ, noticing that the $\sup$ and $\inf$ are attained everywhere. Because of the conditions in (2.11) and (2.12), the value we start from at (2.5) is q. At every step, we have equality, so we end up with $\Phi(G_*) = q$. Since $G_* \in \mathcal{P}$, G_* is optimal. This concludes the proof.

□

It might appear that the conditions of Theorem 2.1 are too complicated to verify in practice, but upon inspection of (2.11) we realize that *if for some $a, b \in \mathbb{R}$*

$$\lambda_*(s) = a + bs, \tag{2.14}$$

there may be a chance. Indeed, if we assume that $EX = EY = 0$, then the condition

$$\alpha_* + \beta_* y + \gamma_* y^2 = \int (x - \lambda_*(x+y))^2 F(dx) \qquad \forall y, \tag{2.15}$$

combined with (2.14) implies that $a = 0$, $\beta_* = 0$, $b^2 = \gamma_*$, and $\alpha_* = (1-b)^2\sigma^2$.

3 Explicitly soluble situations.

If we took X, Y to be independent zero-mean with the *same* distribution, then it is obvious that

$$E[X|X+Y] = (X+Y)/2. \tag{3.1}$$

Let us now apply Theorem 2.1 to this situation, taking $\lambda_*(s) = s/2$, $\alpha_* = \sigma^2/4$, $\beta_* = 0$, $\gamma_* = 1/4$, and $G = F$. If the bound on the variance of Y is $\varepsilon = \sigma^2 \equiv \text{var}(X)$, then $G = F$ is primal-feasible, $(\alpha_*, \beta_*, \gamma_*, \lambda_*)$ is dual-feasible, and the complementary slackness conditions (2.15) and (2.12) hold. Hence by Theorem 2.1 the law which minimizes var$E[X|X+Y]$ subject to the bound $\text{var}(Y) \le \text{var}(X)$ is $G = F$. The lower bound from (2.9) is seen to be $\sigma^2/2$, which is indeed the variance of $(X+Y)/2$.

By similar reasoning, it is straightforward to see that if $X = \xi_1 + \ldots + \xi_n$, where the ξ_j are IID with zero mean and common variance σ^2, and where we bound $\text{var}(Y) \le m\sigma^2$, then the optimal law of Y is given by $Y = \xi_1 + \ldots + \xi_m$. But this result now points towards a wider result for infinitely divisible distributions, which we state as Proposition 3.1 below.

Proposition 3.1. *Suppose that $(Z_t)_{t\ge 0}$ is a zero-mean square-integrable Lévy process, with $EZ_t^2 = t$. Suppose that $X \sim Z_t$ for some fixed $t > 0$. Then the minimum in* (1.5) *is achieved when $Y \sim Z_\varepsilon$.*

Proof. If we let $Y = Z_\varepsilon$, then $E[X|X+Y] = t(X+Y)/(t+\varepsilon)$, so by setting $\lambda_*(s) = bs$ with $b = t/(t+\varepsilon)$ we ensure that $\lambda_* = \varphi_{G_*}$. The complementary slackness condition (2.11) holds for all y if we take $\gamma_* = b^2$, $\beta_* = 0$, and $\alpha_* = (1-b)^2 t$, as before. With $z_* = 0$, the complementary slackness condition (2.12) holds. The law of Y is primal feasible, and so by Theorem 2.1 the result follows.

□

4 Does more noise mean worse prediction?

We begin by recording some simple facts:

1. For any random variables $U,\ V,\ W$ with $E|U| < \infty$ and $(U,\ V)$ independent of W,
$$E(U|V,\ W) = E(U|V). \tag{4.1}$$
2. For any $U,\ V,\ W$ with $E(U^2) < \infty$,
$$\mathrm{var}E(U|V,\ W) \geq \mathrm{var}E(U|V+W). \tag{4.2}$$
3. If Z is independent of (X, Y) then
$$\mathrm{var}E(X|X+Y) \geq \mathrm{var}E(X|X+Y+Z), \tag{4.3}$$

The third fact is a consequence of the first two.

As we saw at (1.4), if $Y \sim N(0, \sigma^2)$ is independent of X, then as $\sigma \to \infty$, for any ξ
$$E[X|X+Y=\xi] \to EX, \tag{4.4}$$
so in this situation, *adding a larger-variance noise to X decreases*[2] *the variance of* $E[X|X+Y]$. One might conjecture that this holds more generally, but a little thought shows that this is not so. Indeed, if $X, Y \sim B(1, \frac{1}{2})$, then we have[3] $E[X|X+2Y] = X$, which has larger variance than $E[X|X+Y]$.

This being said, a result in the direction of (4.4) is valid if Y is *self-decomposable*, as defined in Definition 1 below.

Definition 1. A random variable Y is self-decomposable (belongs to class $\mathcal{L}$), if for any $c,\ 0 < c < 1$ there exists a random variable U_c independent of Y such that Y is equal in law to $cY + U_c$.

All $Y \in \mathcal{L}$ are infinitely divisible. Not all infinitely divisible random variables are in $\mathcal{L}$, but the random variables having stable distributions are in $\mathcal{L}$. See Chapter 5 of [5] for properties of the class $\mathcal{L}$.

Theorem 4.1. *Let X be a random variable with $\mathrm{var}X < \infty$ and $Y \in \mathcal{L}$. Let $V(\lambda) := \mathrm{var}E(X|X+\lambda Y)$. Then $V(\lambda)$ is monotone decreasing on $(0,\ \infty)$ and monotone increasing on $(-\infty,\ 0)$.*

Proof. Let $0 < \lambda_1 < \lambda_2$ and set $\lambda_1 = c\lambda_2$ with $0 < c < 1$. Suppose that $X,\ Y,\ U_c$ are independent random variables with the self-decomposable property
$$Y \sim cY + U_c.$$

Then
$$\begin{aligned} \mathrm{var}E[X|X+\lambda_2 Y] &= \mathrm{var}E[X|X+\lambda_2 cY + \lambda_2 U_c] \\ &= \mathrm{var}E[X|X+\lambda_1 Y + \lambda_2 U_c] \\ &\leq \mathrm{var}E[X|X+\lambda_1 Y], \end{aligned}$$

[2] To see that $\mathrm{var}(E[X|X+Y])$ is decreasing with σ in (4.4), we use Fact 3, (4.3), noticing that a $N(0, v_1+v_2)$ is the independent sum of a $N(0, v_1)$ and a $N(0, v_2)$.

[3] Notice that $X = X + 2Y \mod 2$

where the last step follows by (4.3). Monotonicity in $(-\infty, 0)$ follows because $-Y \in \mathcal{L}$.
□

An interesting question raised by the referee of the paper is the following. Consider the original problem (1.5) with the constraint ε on the variance of Y; *will an optimal Y always satisfy the variance bound with equality?* We do not have a definitive answer to this question, but the following little result shows that we may restrict our search to random variables Y which do satisfy the bound with equality.

Lemma 4.2. *If Y_0 achieves the minimum in* (1.5) *and* $\mathrm{var}(Y_0) < \varepsilon$*, then there exists Y with variance ε which achieves the same value in* (1.5)*.*

Proof. Choose some random variable Z independent of Y_0 with mean zero and variance $\varepsilon - \mathrm{var}(Y_0)$. Then $Y = Y_0 + Z$ satisfies the variance constraint with equality, and using (4.3) we have

$$\mathrm{var}E(X|X+Y_0) \geq \mathrm{var}E(X|X+Y_0+Z) = \mathrm{var}E(X|X+Y). \tag{4.5}$$

But Y is feasible for the problem, so the right-hand side of (4.5) must be at least the value of the problem, which by assumption is the left-hand side of (4.5). Hence Y is also optimal, and satisfies the variance constraint with equality.
□

The conjecture that the optimal Y must *always* satisfy the variance bound with equality would follow from the argument just given if it were true that adding a independent random variable Z of strictly positive variance to Y_0 necessarily strictly reduces the variance of $E(X|X+Y_0)$. However, this is not true, as we see if we take $X \sim B(1, \frac{1}{2})$ and $Y_0, Z \sim U(-a, a)$ for any $a < \frac{1}{4}$. In that example,

$$E(X|X+Y_0) = X = E(X|X+Y_0+Z).$$

5 Examples.

Our first example is $X \sim B(1,p)$, which is simple enough to allow fairly complete analysis for small ε. Thereafter we take a few examples where X has a symmetric discrete distribution and present numerical solutions.

To begin with, suppose that X is an integer-valued random variable, and Y is an independent random variable with integer part $[Y]$, fractional part $\{Y\}$. Then it is clear that

$$\mathcal{F}_1 \equiv \sigma(X+Y) = \sigma(X+[Y], \{Y\}) \supseteq \mathcal{F}_2 \equiv \sigma(X+[Y]). \tag{5.1}$$

Accordingly, the variance of $E[X|\mathcal{F}_1]$ will be larger than the variance of $E[X|\mathcal{F}_2]$, so if we are seeking to minimize the variance of $E[X|X+Y]$ we may restrict attention to random variables Y which take integer values. But a word of caution is in order: the variance of $[Y]$ may be greater or smaller than the variance of Y, so there is no guarantee that $[Y]$ will satisfy the variance bound even if Y does.

For the remainder of this section, we shall explore numerically some examples where both X and Y are integer-valued. As just explained, we should not expect that the best Y that we find here (which are limited by the constraint that Y be integer-valued) will be overall optimal, and indeed in the case of $X \sim B(1,p)$ which we discuss first, it can be shown that the optimal Y will not in general be integer-valued. Nevertheless, the numerical examples presented serve to illustrate the surprising complexity of the solutions obtained.

5.1 Binomial distribution.

Suppose that $P(X = 1) = p = 1 - q = 1 - P(X = 0)$, and that the variance of Y is bounded by $\varepsilon > 0$ as before. We shall assume without loss of generality that $p \geq q$.

To begin with, assuming that $Y \sim B(1, t)$, and setting $s = 1 - t$, notice that

$$E[X|X + Y = 1] = \frac{ps}{ps + qt}$$

so that

$$\begin{aligned} h(t) &\equiv E\big[\, E[\, X \,|\, X + Y\,]^2 \,\big], \\ &= \frac{p^2 s^2}{ps + qt} + pt \\ &= \frac{p(ps + qt^2)}{ps + qt}. \end{aligned} \tag{5.2}$$

As a function of $t \in [0, 1]$, this is convex and has a unique minimum at

$$t_* = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{q}} \geq \tfrac{1}{2}. \tag{5.3}$$

Since

$$\frac{t_*}{1 - t_*} = \frac{\sqrt{p}}{\sqrt{q}} \leq \frac{p}{q} = \frac{p}{1 - p} \tag{5.4}$$

it follows that $p \geq t_*$, which will be needed later.

If Y is variance-constrained, then in order to minimize this objective we will take t as close as possible to t_*, subject to the variance constraint. But the variance of Y is $t(1 - t)$, so for a given variance, do we take $t > \frac{1}{2}$ or $t < \frac{1}{2}$? If we set

$$w = 2p - 1 \in [0, 1], \qquad u = 2t - 1 \in [-1, 1],$$

then

$$h(t) - h(1 - t) = -\frac{wu(1 - u^2)(1 - w^2)}{4(1 - u^2 w^2)}, \tag{5.5}$$

from which it is clear that for $t > \frac{1}{2}$ we have $h(t) \leq h(1 - t)$, so to minimize the objective we will be looking at $t \in [t_*, 1]$. We also expect that the value

$$\varepsilon_* \equiv t_*(1 - t_*) = \frac{\sqrt{pq}}{(\sqrt{p} + \sqrt{q})^2} \tag{5.6}$$

will be a critical value for the variance bound on Y.

To illustrate the kind of solutions we arrive at, we show in Figures 1, 2 and 3 below the probability mass functions for X and the optimal Y in the case where $X \sim B(1, 0.6)$ and Y has to satisfy a low variance bound $\varepsilon = 0.5\varepsilon_*$, the critical variance bound ε_*, and a higher variance bound $2\varepsilon_*$ respectively. The probability mass function (PMF) of Y is shown shifted to the left for clarity - as we have already remarked, such a shift makes no difference to the objective. Notice how the objective decreases as the bound on the variance of Y becomes more relaxed, as it should do. Notice also that the PMF of Y in the final plot gives non-zero weight to *more than two* values, again as we should expect from the preceding analysis.

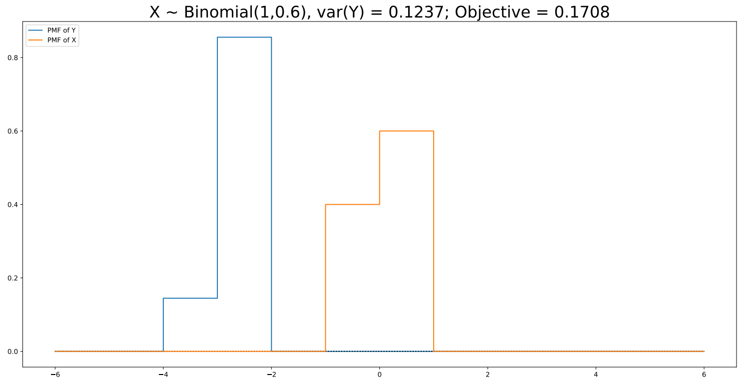

Figure 1: $X \sim B(1, 0.6)$ with low bound on var(Y).

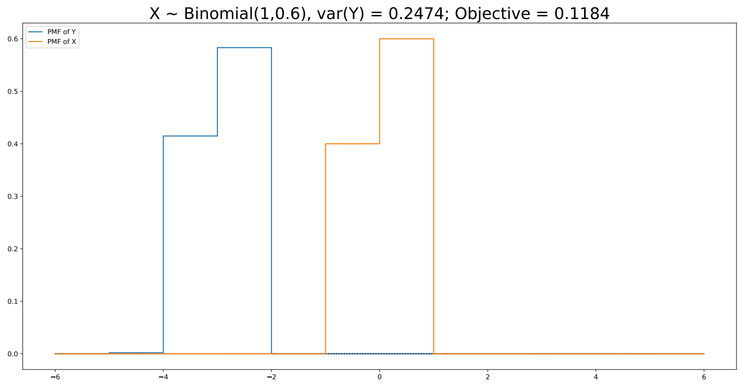

Figure 2: $X \sim B(1, 0.6)$ with critical bound on var(Y).

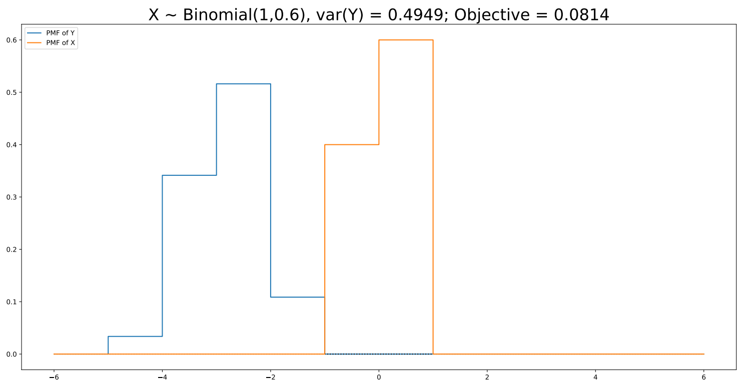

Figure 3: $X \sim B(1, 0.6)$ with high bound on var(Y).

5.2 X is uniform.

Here we compute the optimal distribution for Y when X is uniform. We consider two cases: the first low-variance case has $\text{var}(Y) = 2\text{var}(X)/\pi$ and the second high-variance case has $\text{var}(Y) = \pi\text{var}(X)/2$. The two corresponding figures, Figures 4 and 5 below, display the PMFs of X and Y, along with a diagnostic plot[4] in red and green markers of the computed function

$$y \mapsto \int \lambda_*(y+x)\{\, \lambda_*(y+x) - 2x \,\}\, F(dx) \tag{5.7}$$

which according to (2.8) must be dominated by a quadratic[5], and equal to that quadratic wherever the PMF of Y is positive. From our discussion in Section 3, if we set $\varepsilon = 2\text{var}(X)$ then the optimal choice would be to take Y to be the sum of two independent copies of X, which in this case would be the sum of two independent uniforms; the resulting PMF would be a symmetric piecewise-linear 'tent', and looking at Figure 5 we see something that looks approximately like that.

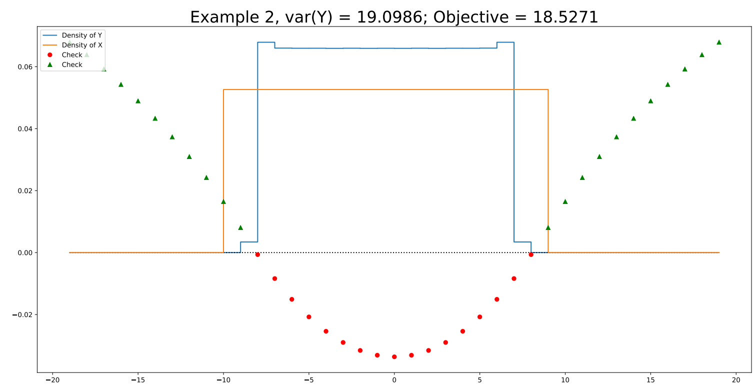

Figure 4: X uniform, with $\text{var}(Y) = 2\text{var}(X)/\pi$.

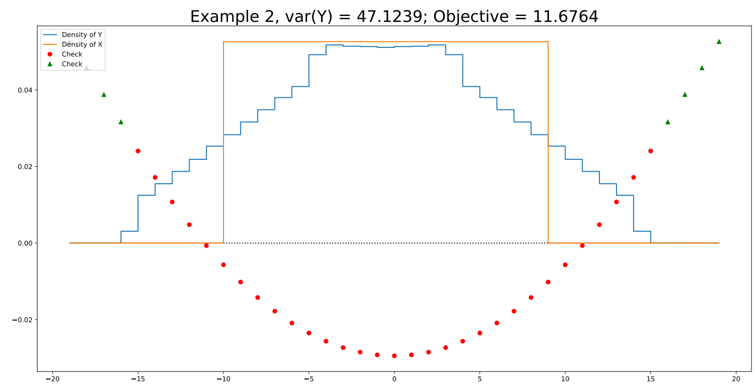

Figure 5: X uniform, with $\text{var}(Y) = \pi\text{var}(X)/2$.

[4]... scaled to fit the plot of the PMFs...
[5]Recall that f is symmetric.

5.3 X is the sum of two uniforms.

Again we compute the optimal Y for two values of ε. Notice how strange the solution is in both cases, particularly for the high variance case, where we see that the distribution of the optimal Y has a hole at the center!

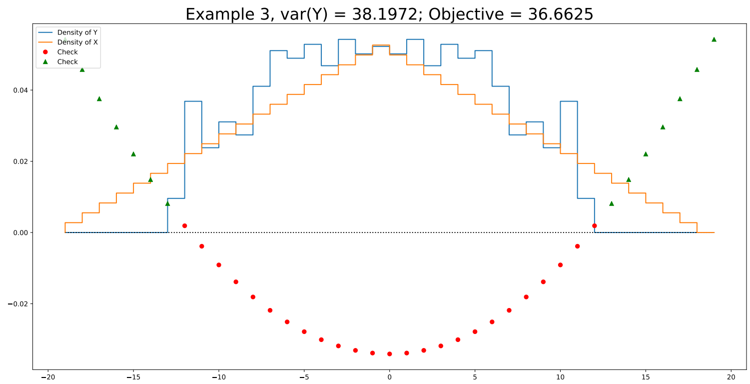

Figure 6: X is the sum of two uniforms, with $\text{var}(Y) = 2\text{var}(X)/\pi$.

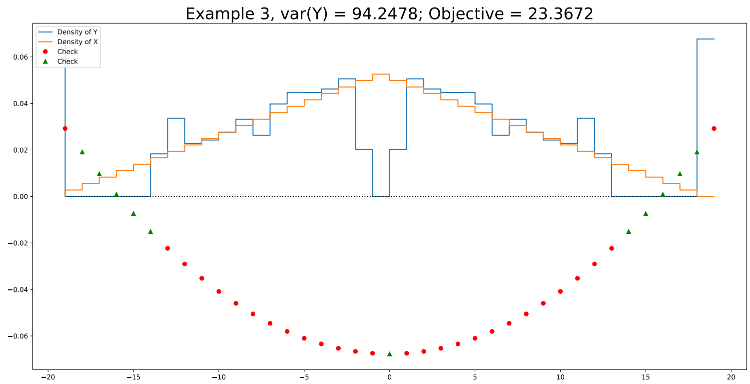

Figure 7: X is the sum of two uniforms, with $\text{var}(Y) = \pi\text{var}(X)/2$.

5.4 The density of X is the square of that in section 5.3

This time we take the density of X from Section 5.3 and square it (of course, renormalizing to sum to 1). Once again, the distribution of the optimal Y has a form which would be difficult to guess - the PMF is not monotone in $\mathbb{Z}^+$, for example.

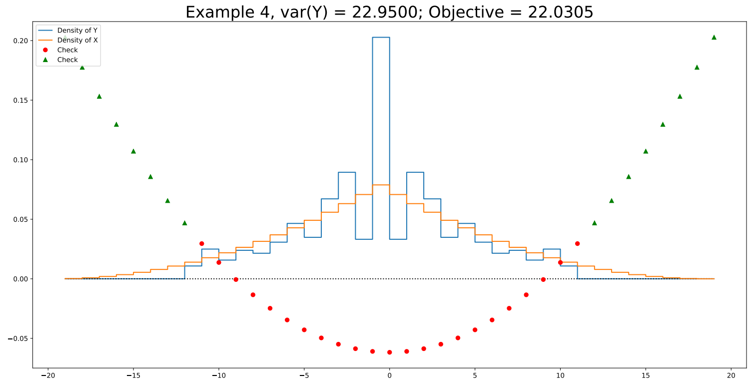

Figure 8: The density of X is the square of the example in Figure 6, with var$(Y) = 2$var$(X)/\pi$.

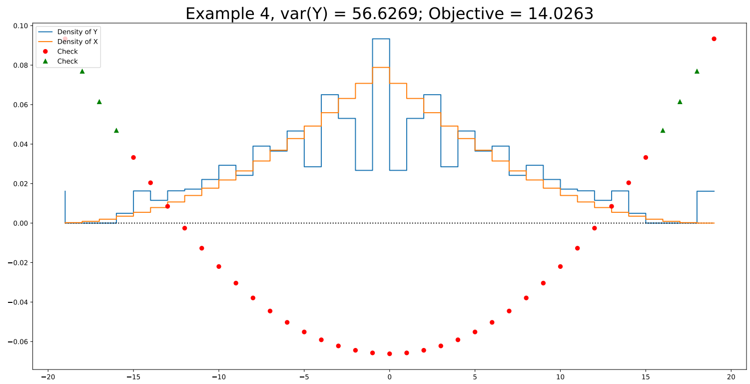

Figure 9: The density of X is the square of the example in Figure 7, with var$(Y) = \pi$var$(X)/2$.

References

[1] Bain, A. and Crisan, D.: Fundamentals of Stochastic Filtering (Vol. 60). *Springer Science & Business Media*, New York, 2008.

[2] Bryc, W. and Smolenski, W.: On the stability problem for conditional expectation. *Statist. Probab. Lett.* **15**, (1992), 41–46.

[3] Bryc, W., Dembo, A., and Kagan, A.M.: On the maximum correlation coefficient. *Theory Probab. Appl.* **49**, (2005), 132–138.

[4] Dembo, A., Kagan, A.M. and Shepp, L.A.: Remarks on the maximum correlation coefficient. *Bernoulli* **7**, (2001), 343–350.

[5] Lukacs, E.: Characteristic Functions. *Charles Griffin & Co., Ltd.*, Glasgow, 1970.

Acknowledgments. We thank Professor Dan Crisan (Imperial College London) and Dongzhou Huang (Rice University) for helpful discussions.