18.177 course project: Invariance Principles

Brendan Juba

1 Introduction

An *invariance principle* is a result permitting us to change our underlying probability space—such as occurs in a central limit theorem. For example, we can suggestively state the Berry-Essen Theorem in the following way:

Theorem 1.1 (Berry-Essen) *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random variables with $\mathbb{E}[\mathbf{X}_i] = 0$, $\mathbb{E}[\mathbf{X}_i^2] = 1$, and $\mathbb{E}[|\mathbf{X}_i|^3] \leq A < \infty$. Let $Q(x_1, \ldots, x_n) = \sum_{i=1}^n c_i x_i$ be a linear form such that $\sum_{i=1}^n c_i^2 = 1$. Then*

$$\sup_t |\mathbb{P}[Q(\mathbf{X}_1, \ldots, \mathbf{X}_n) \leq t] - \mathbb{P}[Q(\mathbf{G}_1, \ldots, \mathbf{G}_n) \leq t]| \leq O\left(A \cdot \max_i |c_i|\right)$$

where $\mathbf{G}_1, \ldots, \mathbf{G}_n$ are independent standard Gaussians.

We think of this as saying that under some mild conditions on our probability space, the distribution of the linear form Q is roughly the same, regardless of the actual underlying distribution. There are, of course, a family of different central limit theorems, with different "mildness conditions" leading to different notions of the closeness of the resulting distributions.

As "central limit theorems," however, these classical results all aimed to show that a linear form on some (arbitrary) distribution was close to the single Gaussian corresponding to that linear form. When these theorems are stated in the form given here, it is natural to wonder if they hold for a richer class of functions, where although the final distribution may not be Gaussian, it may be close to the distribution produced by a function on Gaussian space. Recently, such generalizations of the Berry-Essen Theorem to multilinear polynomials, i.e., Q of the form

$$Q(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i$$

were derived by Mossel et al. [13]. Just as there are numerous "central limit theorems," so exist numerous versions of the multilinear generalization: an earlier generalization that holds under weaker conditions was also proved by Rotar' [15], another generalization under alternative conditions was shown contemporarily by Chatterjee [4], and even in the work of Mossel et al., a host of variants following the same basic argument are stated. The Mossel et al. result is particularly noteworthy because it permitted the authors to state the first solutions to a variety of discrete extremal problems, by reducing them to corresponding optimization problems in Gaussian space where the solutions were known.

Subsequently, these results were generalized to multilinear polynomials in higher dimensions by Mossel [12]. These generalizations permitted Mossel to solve additional discrete extremal problems, by again reducing to the corresponding problems in Gaussian space. In the sequel, we will review the work of Mossel in greater detail. We will state the invariance principle in two dimensions in Section 2; its proof is similar to the one-dimensional invariance principle of Mossel et al. [13], which in turn follows the same outline as Lindeberg's proof of a central limit theorem [11]. We will show an excellent application of the two-dimensional invariance principle to a social choice problem, finding a "fair" voting rule that is optimally predictable from a random sample of votes, which seems to be out of reach of the one-dimensional version. In order to derive this result, in Section 4, we need to apply the invariance principle to the product of two (correlated) polynomials. The derivation of the desired bounds from the invariance principle and existing Gaussian isoperimetric inequalities comprises the bulk of the work done by Mossel, and we give an overview in Section 3.

2 An invariance principle for pairs of correlated functions

In this section, we will state one form of the invariance principle for multilinear polynomials in two dimensions, from the work of Mossel [12]. Actually, since the invariance principles currently require numerous hypotheses, the proof turns out to be more natural than the final statement, so we will give the proof here. Furthermore, as discussed by Mossel et al. [13], essentially the same argument can be used to give a number of different variations on the result, so the argument is in many ways more central than the final results.

2.1 Preliminaries: Influences and hypercontractivity

Mossel [12] uses the same basic notions as Mossel et al. [13] to generalize the central limit theorems from linear forms to multilinear polynomials: *hypercontractivity* to bound the higher moments in terms of a bound on the *influences*. We will review these definitions along with the basic setup and notation here.

2.1.1 Basic notation

We begin with some essential notation:

Definition 2.1 (Multi-index) *A* multi-index *of dimension k is a vector $\mathbf{i} = (i_1, \dots, i_k) \in \mathbb{N}^k$. We write $|\mathbf{i}|$ for $i_1 + \dots + i_k$, $\#\mathbf{i}$ for $|\{j \in [k] : i_j > 0\}|$, and $\mathbf{i}!$ for $i_1! \cdots i_k!$. We will use multi-indices in the following contexts:*

1. *Given a doubly-indexed set of indeterminates $\{x_{i,j}\}_{i \in [k], j \in \mathbb{N}}$, we write $x_\mathbf{i}$ for $\prod_{j=1}^k x_{j,i_j}$.*
2. *Given a vector $x \in \mathbb{R}^k$, we write $x^\mathbf{i}$ for $\prod_{j=1}^k x_j^{i_j}$*
3. *Given a function ψ of k variables, we write $\psi^{(\mathbf{i})}$ for the partial derivative of ψ taken i_j times with with respect to the jth variable.*

Definition 2.2 (Multilinear polynomial) *A k-*dimensional multilinear polynomial *over a set of indeterminates is given by $Q = (Q_1, \dots, Q_k)$ where each Q_i is some expression $Q_i(x) = \sum_{\boldsymbol{\sigma}} c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}$ where each $c_{\boldsymbol{\sigma}} \in \mathbb{R}$, and all but finitely many of the $c_{\boldsymbol{\sigma}}$ are zero. The* degree *of Q is $\max_{Q_i} \max\{|\boldsymbol{\sigma}| : c_{\boldsymbol{\sigma}} \neq 0\}$. We write $\|Q\|_q$ for $\mathbb{E}[\sum_{j=1}^k |Q_i|^q]^{1/q}$ and $\mathsf{Var}[Q]$ for $\|Q - \mathbb{E}[Q]\|_2^2$.*

The use of doubly-indexed sets of indeterminates in our definition of multilinear polynomials may seem quite unusual at this point. The motivation is that we wish to consider functions on product probability spaces, $f : \Omega_1 \times \cdots \times \Omega_k \to \mathbb{R}$; when the probability space Ω_i is finite, we can choose a finite set of random variables $\{\mathbf{X}_{i,j}\}$ that form an orthonormal basis for the space of all functions $\Omega_i \to \mathbb{R}$, in which case every function $f : \prod_{i=1}^k \Omega_i \to \mathbb{R}$ can be written as a multilinear polynomial in $\{\mathbf{X}_{i,j}\}$.

Definition 2.3 (Sequence of ensembles) *We call a collection of finitely many orthonormal real random variables, one of which is the constant 1, an* ensemble. *A* sequence of ensembles *is written* $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_n)$ *where* $\boldsymbol{\mathcal{X}}_i = \{\mathbf{X}_{i,0} = 1, \mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,m_i}\}$. *We call the sequence of ensembles* independent *if the ensembles are independent families of random variables.*

Note that *we only consider orthonormal sequences of ensembles.* We will stress this explicitly in the statements of the main theorems, but this assumption is used throughout.

2.1.2 Properties of multilinear polynomials on orthonormal ensembles

When we consider multilinear polynomials over indeterminates given by a sequence of orthonormal ensembles, the various quantities of interest – moments, influences, etc. – can be expressed simply in terms of the coefficients of the polynomial.

We first recall the crucial definitions for influences of ensembles:

Definition 2.4 (Influence) *For a sequence of ensembles* $\boldsymbol{\mathcal{X}}$ *and a* k*-dimensional multilinear polynomial* Q*, the* influence of the ith ensemble on Q *is given by*

$$\mathrm{Inf}_i(Q(\boldsymbol{\mathcal{X}})) = \sum_{j=1}^{k} \mathbb{E}[\mathsf{Var}[Q_j(\boldsymbol{\mathcal{X}}) | \boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_{i-1}, \boldsymbol{\mathcal{X}}_{i+1}, \ldots, \boldsymbol{\mathcal{X}}_n]]$$

We now have the aforementioned formulas:

Proposition 2.5 *Let* $\boldsymbol{\mathcal{X}}$ *be a sequence of ensembles and* Q *be a* k*-dimensional multilinear polynomial with* $c_{\boldsymbol{\sigma},j}$ *as the coefficients of* Q_j*. Then,*

$$\mathbb{E}[Q(\boldsymbol{\mathcal{X}})] = (c_{\mathbf{0},1}, \ldots, c_{\mathbf{0},k}) \qquad \|Q(\boldsymbol{\mathcal{X}})\|_2^2 = \textstyle\sum_{\boldsymbol{\sigma},j} c_{\boldsymbol{\sigma},j}^2$$

$$\mathsf{Var}[Q(\boldsymbol{\mathcal{X}})] = \textstyle\sum_{j,\#\boldsymbol{\sigma}\neq 0} c_{\boldsymbol{\sigma},j}^2 \qquad \mathrm{Inf}_i(Q(\boldsymbol{\mathcal{X}})) = \textstyle\sum_{j,\boldsymbol{\sigma}:\sigma_i>0} c_{\boldsymbol{\sigma},j}^2$$

Thus, in particular, given that $\boldsymbol{\mathcal{X}}$ is an orthonormal sequence of ensembles, these quantities depend only on Q and not $\boldsymbol{\mathcal{X}}$. For convenience, we may sometimes drop $\boldsymbol{\mathcal{X}}$ in our notation, since it is understood to be an orthonormal sequence of ensembles. In line with this observation, we can define "low-degree influences" as follows:

Definition 2.6 (Low-degree influence) *For a* k*-dimensional multilinear polynomial* Q*, the* d-low-degree influence of the ith ensemble on Q *is given by*

$$\mathrm{Inf}_i^{\leq d}(Q)) = \sum_{j,\boldsymbol{\sigma}:\#\boldsymbol{\sigma}\leq d,\sigma_i>0} c_{\boldsymbol{\sigma},j}^2$$

Since every function on a finite probability space can be written as a multilinear polynomial in some orthonormal basis, this definiton implicitly defines low-degree influences for every function on a finite probability space. This is important in Section 3. Low-degree influences turn out to be useful because there cannot be too many coordinates with high low-degree influences:

Proposition 2.7 *Suppose Q is a k-dimensional multilinear polynomial. Then*

$$\sum_i \mathrm{Inf}_i^{\leq d}(Q) \leq d \cdot \mathsf{Var}[Q].$$

2.1.3 Hypercontractive sequences of ensembles

We will require the sequences of ensembles we consider to be "hypercontractive," a notion which (in this sense) was introduced in the work of Mossel et al. [13], although a similar definition of hypercontractivity for sets of random variables existed in the literature previously, and many of the proofs can be adapted from that setting—see, e.g., Janson [9]. As a starting point, recall that a random variable $\mathbf{X}$ is said to be "(p, q, η)-hypercontractive" for $1 \leq p \leq q < \infty$ and $0 < \eta < 1$ if

$$\|a + \eta \mathbf{X}\|_q \leq \|a + \mathbf{X}\|_p$$

for every $a \in \mathbb{R}$. In our setting, the contraction will be given by the following noise operator on formal multilinear polynomials:

Definition 2.8 (Noise operator) *For $\rho \in [0,1]$ we define the* noise operator T_ρ *on multilinear polynomials as taking $Q(x) = \sum_{\boldsymbol{\sigma}} c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}$ to $(T_\rho Q)(x) = \sum_{\boldsymbol{\sigma}} \rho^{\#\boldsymbol{\sigma}} c_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}$.*

We can now define the key notion of a hypercontractive sequence of ensembles:

Definition 2.9 (Hypercontractivity) *Let $\boldsymbol{\mathcal{X}}$ be a sequence of ensembles. For $1 \leq p \leq q < \infty$ and $0 < \eta < 1$, we say that $\boldsymbol{\mathcal{X}}$ is* (p, q, η)-hypercontractive *if $\|(T_\eta Q)(\boldsymbol{\mathcal{X}})\|_q \leq \|Q(\boldsymbol{\mathcal{X}})\|_p$ for every (one-dimensional) multilinear polynomial Q over $\boldsymbol{\mathcal{X}}$.*

Technically, hypercontractivity will permit us to bound the higher moments of our random variables in terms of the lower moments (which are bounded by orthonormality), as a consequence of the following proposition:

Proposition 2.10 *Let $\boldsymbol{\mathcal{X}}$ be a $(2, q, \eta)$-hypercontractive sequence of ensembles and Q be a (one-dimensional) multilinear polynomial over $\boldsymbol{\mathcal{X}}$ of degree d. Then $\|Q(\boldsymbol{\mathcal{X}})\|_q \leq \eta^{-d} \|Q(\boldsymbol{\mathcal{X}})\|_2$.*

The next proposition is essential in establishing that a sequence of ensembles is hypercontractive:

Proposition 2.11 *Suppose $\boldsymbol{\mathcal{X}}$ is a sequence of n_1 ensembles and $\boldsymbol{\mathcal{Y}}$ is an independent sequence of n_2 ensembles, both of which are (p, q, η)-hypercontractive. Then the sequence of ensembles $(\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_{n_1}, \boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_{n_2})$ is also (p, q, η)-hypercontractive.*

So, for example, to establish that a sequence of ensembles is hypercontractive, it suffices to check that in each ensemble $\boldsymbol{\mathcal{X}}_i = \{\mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,m_i}\}$, all linear combinations of the $\{\mathbf{X}_{i,j}\}$ are hypercontractive in the usual sense.

Finally, we require an analogue of Proposition 2.10 for two-dimensional polynomials in the sequel, as given by the following proposition:

Proposition 2.12 *Let $\boldsymbol{\mathcal{X}}$ be a $(2, q, \eta)$-hypercontractive sequence of ensembles for some integer q and let Q be a two-dimensional multilinear polynomial over $\boldsymbol{\mathcal{X}}$ of degree d. Let $\mathbf{i} = (i_1, i_2)$ be a multi-index with $|\mathbf{i}| = q$. Then*

$$\mathbb{E}[|Q^{\mathbf{i}}(\boldsymbol{\mathcal{X}})|] \leq \eta^{dq}\mathbb{E}[Q_1^2]^{i_1/2}\mathbb{E}[Q_2^2]^{i_2/2}$$

2.2 Main theorem: Invariance for smooth functions

We now prove the invariance principle for two-dimensional multilinear polynomials. We will consider two (orthonormal) sequences of ensembles of random variables, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, on different probability spaces. We would like to say that, for a two-dimensional multilinear polynomial Q, that $Q(\boldsymbol{\mathcal{X}})$ and $Q(\boldsymbol{\mathcal{Y}})$ are close in distribution. The sense in which we will show this is that if $\Psi : \mathbb{R}^2 \to \mathbb{R}$ is a "nice" test function, then

$$|\mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{X}}))] - \mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Y}}))]| \leq \epsilon$$

for some appropriate bound ϵ. Along the way we will indicate additional hypotheses on $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}$, Q, and Ψ, which we will eventually accumulate in the statement of Theorem 2.13 below.

The argument follows the same basic outline as Lindeberg's proof of the Central Limit Theorem [11], with a few modifications. To start, we will define a sequence of ensembles intermediate to $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$: for each $i = 0, 1, \ldots, n$, put $\boldsymbol{\mathcal{Z}}^{(i)} = (\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{X}}_{i+1}, \ldots \boldsymbol{\mathcal{X}}_n)$. We will obtain our desired bound by bounding each difference

$$\left|\mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Z}}^{(i-1)}))] - \mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Z}}^{(i)}))]\right| \leq \varepsilon$$

Notice that the two sequences differ only in the ith ensemble. Suppose we write Q in the following way:

$$\begin{aligned} Q(\{x_{1,j}\}, \ldots, \{x_{n,j}\}) = Q_{-i}(&\{x_{1,j}\}, \ldots, \{x_{i-1,j}\}, \{x_{i+1,j}\}, \ldots, \{x_{n,j}\}) \\ &+ \sum_k x_{i,k} \cdot Q_{+i,k}(\{x_{1,j}\}, \ldots, \{x_{i-1,j}\}, \{x_{i+1,j}\}, \ldots, \{x_{n,j}\}) \end{aligned}$$

where Q_{-i} and $Q_{+i,j}$ are also multilinear polynomials with degree no greater than that of Q. We now have $Q_{-i}(\boldsymbol{\mathcal{Z}}^{(i-1)}) = Q_{-i}(\boldsymbol{\mathcal{Z}}^{(i)})$ and $Q_{+i,j}(\boldsymbol{\mathcal{Z}}^{(i-1)}) = Q_{+i,j}(\boldsymbol{\mathcal{Z}}^{(i)})$; henceforth we will simplify our notation and write the former as Q_{-i}, $\sum_k \boldsymbol{\mathcal{X}}_{i,j} Q_{+i,j}(\boldsymbol{\mathcal{Z}}^{(i-1)})$ as R and $\sum_k \boldsymbol{\mathcal{Y}}_{i,j} Q_{+i,j}(\boldsymbol{\mathcal{Z}}^{(i)})$ as S.

We now wish to apply Taylor's theorem to Ψ at Q_{-i}. In order to do so, however, we need to assume that Ψ is smooth—in particular, we will assume $\Psi \in \mathcal{C}^3$ with $\|\Psi^{(\mathbf{i})}\|_u \leq B$ for some fixed B for every third derivative of Ψ, i.e., $\Psi^{(\mathbf{i})}$ with $|\mathbf{i}| = 3$ (Hypothesis 1). We then find that at each $x, y \in \mathbb{R}^2$,

$$\left|\Psi(x+y) - \sum_{|\mathbf{k}|<3} \frac{\Psi^{(\mathbf{k})}(x) y^{\mathbf{k}}}{\mathbf{k}!}\right| \leq \sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!}|y|^{\mathbf{k}}$$

Now, whenever $\sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!}\mathbb{E}[|\sum_j \boldsymbol{\mathcal{X}}_{i,j} Q_{+i,j}|^{\mathbf{k}}]$ and $\sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!}\mathbb{E}[|\sum_j \boldsymbol{\mathcal{Y}}_{i,j} Q_{+i,j}|^{\mathbf{k}}]$ are finite, we find

$$\left|\mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Z}}^{(i-1)}))] - \sum_{|\mathbf{k}|<3} \mathbb{E}\left[\frac{\Psi^{(\mathbf{k})}(Q_{-i})(R)^{\mathbf{k}}}{\mathbf{k}!}\right]\right| \leq \sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!}\mathbb{E}[|R|^{\mathbf{k}}] \tag{1}$$

and

$$\left|\mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Z}}^{(i)}))] - \sum_{|\mathbf{k}|<3} \mathbb{E}\left[\frac{\Psi^{(\mathbf{k})}(Q_{-i})(S)^{\mathbf{k}}}{\mathbf{k}!}\right]\right| \leq \sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!}\mathbb{E}[|S|^{\mathbf{k}}]. \tag{2}$$

Before we continue, let us consider the bound on the third moments. At this point, we would like the ensembles $\boldsymbol{\mathcal{Z}}^{(i-1)}$ and $\boldsymbol{\mathcal{Z}}^{(i)}$ to be $(2,3,\eta)$-hypercontractive for some η so that we can use Proposition 2.12 to bound the third moments using the second moments – which are finite by the orthonormality condition – provided further that we know a bound d on the degree of Q (this bound is Hypothesis 2).

We now distinguish between $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$—henceforth, $\boldsymbol{\mathcal{X}}$ will be a sequence of ensembles in which each ensemble $\boldsymbol{\mathcal{X}}_i$ is a basis for some independent finite probability space Ω_i in which the minimum nonzero probability of any is at least some constant $\alpha \leq 1/2$ (Hypothesis 3). It then follows from the work of Wolff [17] that $\boldsymbol{\mathcal{X}}$ is $(2,3,\eta)$-hypercontractive for $\eta = \frac{1}{2}\alpha^{1/6}$. On the other hand, we will assume directly that $\boldsymbol{\mathcal{Y}}$ is $(2,3,\eta)$-hypercontractive (Hypothesis 4).

We now obtain from Proposition 2.11 that $\boldsymbol{\mathcal{Z}}^{(i-1)}$ and $\boldsymbol{\mathcal{Z}}^{(i)}$ are $(2,3,\eta)$-hypercontractive; by Propostion 2.12, it then follows that

$$\mathbb{E}[|R|^{\mathbf{k}}] \leq \eta^{-3d}\mathbb{E}[R_1^2]^{k_1/2}\mathbb{E}[R_2^2]^{k_2/2}$$

and similarly

$$\mathbb{E}[|S|^{\mathbf{k}}] \leq \eta^{-3d}\mathbb{E}[S_1^2]^{k_1/2}\mathbb{E}[S_2^2]^{k_2/2}$$

where, since $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are orthonormal sequences of ensembles, notice that by construction and Proposition 2.5,

$$\mathbb{E}[S_j^2] = \mathbb{E}[R_j^2] = \sum_{\boldsymbol{\sigma}:\sigma_i>0} c_{\boldsymbol{\sigma},j}^2 = \mathrm{Inf}_i(Q_j) \leq \mathrm{Inf}_i(Q) \tag{3}$$

So in particular, we see that the third moments are finite, and Inequalities 1 and 2 always hold.

We already assumed that $\boldsymbol{\mathcal{X}}$ was a sequence of independent ensembles (Hypothesis 3); supposing we now additionally assume that $\boldsymbol{\mathcal{Y}}$ is an independent sequence of ensembles (Hypothesis 5), we can see that

$$\sum_{|\mathbf{k}|<3} \mathbb{E}\left[\frac{\Psi^{(\mathbf{k})}(Q_{-i})(S)^{\mathbf{k}}}{\mathbf{k}!}\right] - \sum_{|\mathbf{k}|<3} \mathbb{E}\left[\frac{\Psi^{(\mathbf{k})}(Q_{-i})(R)^{\mathbf{k}}}{\mathbf{k}!}\right] = 0$$

since R and S only differ in substituting $\boldsymbol{\mathcal{X}}_i$ for $\boldsymbol{\mathcal{Y}}_i$, where $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{Y}}_i$ are orthonormal and independent of the other random variables in the expectation, and appear with degree at most two in each term. So, we can combinine Inequalities 1, 2, and 3, to find that

$$\left|\mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Z}}^{(i-1)}))] - \mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Z}}^{(i)}))]\right| \leq 2\sum_{|\mathbf{k}|=3} \frac{B}{\mathbf{k}!}\eta^{-3d}\mathrm{Inf}_i(Q)^{3/2} = \frac{8}{3}B\eta^{-3d}\mathrm{Inf}_i(Q)^{3/2}$$

If we now assume a bound on the influences and variance of Q (Hypothesis 6), since $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{Z}}^{(0)}$ and $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{Z}}^{(n)}$, summing over $i = 0, 1, \ldots, n$ will give a bound of the desired form. If $\mathrm{Inf}_i(Q) \leq \tau$ and $\mathsf{Var}[Q] \leq 1$, we then find that since Q has degree d (by Hypothesis 2),

$$\sum_{i=1}^{n} \mathrm{Inf}_i(Q)^{3/2} \leq \tau^{1/2} \cdot \sum_{i=1}^{n} \mathrm{Inf}_i(Q) = \tau^{1/2} \cdot \sum_{i=1}^{n} \mathrm{Inf}_i^{\leq d}(Q) \leq d\tau^{1/2}$$

by Proposition 2.7. In summary,

Theorem 2.13 (Invariance principle) *Suppose that (orthonormal) sequences of ensembles $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, a two-dimensional multilinear polynomial Q, and a function $\Psi : \mathbb{R}^2 \to \mathbb{R}$ are given that satisfy the following conditions:*

1. *$\Psi \in \mathcal{C}^3$ with $\|\Psi^{(\mathbf{i})}\|_u \leq B$ for every multi-index $\mathbf{i}$ with $|\mathbf{i}| = 3$.*
2. *Q has total degree at most d*
3. *$\boldsymbol{\mathcal{X}}$ is a sequence of ensembles in which each ensemble $\boldsymbol{\mathcal{X}}_i$ is a basis for some independent finite probability space Ω_i in which the minimum nonzero probability of any atom is at least $\alpha \leq 1/2$*
4. *$\boldsymbol{\mathcal{Y}}$ is $(2, 3, \eta)$-hypercontractive for $\eta = \frac{1}{2}\alpha^{1/6}$*
5. *$\boldsymbol{\mathcal{Y}}$ is an independent sequence of ensembles.*
6. *$\mathsf{Var}[Q] \leq 1$ and $\mathrm{Inf}_i(Q) \leq \tau$ for all i.*

Then

$$|\mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{X}}))] - \mathbb{E}[\Psi(Q(\boldsymbol{\mathcal{Y}}))]| \leq \epsilon = \frac{8}{3}B\tau^{1/2}d(8\alpha^{-1/2})^d$$

2.3 Strengthening the conclusions

As stated, Theorem 2.13 only applies to $\mathcal{C}^3$ test functions with bounded third derivatives. In applications, one often wishes to apply invariance to functions that do not satisfy these conditions, notably the following two in our case:

$$f_{[0,1]}(x) = \begin{cases} 0 & x < 0 \\ x & x \in [0, 1] \\ 1 & x > 1 \end{cases}$$
$$\zeta(x_1, x_2) = (x_1 - f_{[0,1]}(x_1))^2 + (x_2 - f_{[0,1]}(x_2))^2$$

We can hand-tailor suitable approximations to these functions using standard techniques. Moreover, we can weaken the bound on the degree of the multilinear polynomial in these applications.

Lemma 2.14 *Suppose that (orthonormal) sequences of ensembles $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, and a two-dimensional multilinear polynomial Q are given that satisfy the following conditions:*

1. *$\boldsymbol{\mathcal{X}}$ is a sequence of ensembles in which each ensemble $\boldsymbol{\mathcal{X}}_i$ is a basis for some independent finite probability space Ω_i in which the minimum nonzero probability of any atom is at least $\alpha \leq 1/2$*
2. *$\boldsymbol{\mathcal{Y}}$ is a $(2, 3, \eta)$-hypercontractive independent sequence of of ensembles for $\eta = \frac{1}{2}\alpha^{1/6}$*
3. *$\mathsf{Var}[Q] \leq 1$ and $\mathrm{Inf}_i^{\leq \log(1/\tau)/\log(1/\alpha)}(Q) \leq \tau$ for all i.*
4. *$\mathsf{Var}[Q^{>d}] < (1-\gamma)^{2d}$ where $0 < \gamma < 1$ and $Q^{>d}(x) = \sum_{\#\boldsymbol{\sigma}>d}(c_{\boldsymbol{\sigma},1}x_{\boldsymbol{\sigma}}, c_{\boldsymbol{\sigma},2}x_{\boldsymbol{\sigma}})$.*

Then

$$|\mathbb{E}[\zeta(Q(\boldsymbol{\mathcal{X}}))] - \mathbb{E}[\zeta(Q(\boldsymbol{\mathcal{Y}}))]| \leq \tau^{\Omega(\gamma/\log(1/\alpha))}$$

The proof proceeds by first constructing f_λ, a $\mathcal{C}^\infty$ approximation to $f(x) = x1_{\{x\geq 0\}}$ in the sup norm with bounded third derivatives, by taking the convolution of f with a $\mathcal{C}^\infty$ bump function. The approximation to ζ is then constructed from f_λ, and bounds for degree d multilinear polynomials follow immediately from Theorem 2.13—e.g., for an approporiate approximation, we get a bound of

$$|\mathbb{E}[\zeta(Q(\boldsymbol{\mathcal{X}}))] - \mathbb{E}[\zeta(Q(\boldsymbol{\mathcal{Y}}))]| \leq O(\alpha^{-d/3}\tau^{-1/3})$$

for degree d polynomials. The refinement for higher-degree polynomials comes from truncating Q at degree $d = \frac{\log(1/\tau)}{\log(1/\alpha)}$, and noticing that due to the orthonormality of the ensembles, the error introduced by the truncation under ζ is bounded by the bound on $\mathsf{Var}[Q^{>d}]$, which is also at most $O(\exp(-\gamma d)) \leq \tau^{\Omega(\gamma/\log(1/\alpha))}$ as needed.

3 Gaussian noise stability bounds

The main contribution of Mossel [12] is bounds on the products of bounded functions on correlated probability spaces; indeed, in some sense that we will consider further briefly in Section 4, these bounds are a strict generalization of the noise stability bounds derived by Mossel et al. [13]. In order to establish these powerful bounds, Mossel employs quite a bit of groundwork which we will examine next.

Actually, Mossel goes further in demonstrating the wide applicability of his approaches. In order to give a better sense of what ingredients are actually necessary for the simple application of Section 4, we will restrict our attention to only the pieces that are actually used in that result. We will follow the same presentation in general, but where possible we will simplify the statements and proofs of some theorems.

In the next section, we will cover the basic tools that Mossel uses extensively in the remaining sections, where the invariance principle is actually applied. Section 3.2 describes the solution to an optimization problem in correlated Gaussian spaces, which is then combined with the invariance principle to obtain bounds for discrete probability spaces in Section 3.3. It is these bounds that will immediately yield our desired results in Section 4.

3.1 Preliminaries: Correlated spaces and Efron-Stein decompositions

We begin this section by giving the definitions of correlated probability spaces, Markov operators, and Efron-Stein decompositions, and stating a few of their important properties. We closely follow Mossel's presentation [12] in introducing these central notions from that work.

We begin by defining the notion of correlation between two spaces. The notion used here is the same as the "maximum correlation corefficient" introduced by Hirschfeld and Gebelein [8, 7].

Definition 3.1 (Correlated spaces) *For a probability space $(\mathbb{P}, \prod_{i=1}^k \Omega^{(i)}, \mathcal{F})$, we say that $\Omega^{(1)}, \ldots, \Omega^{(k)}$ are* correlated spaces. *Given $A \subseteq \Omega^{(i)}$ such that $\prod_{j<i} \Omega^{(j)} \times A \times \prod_{j>i} \Omega^{(j)} \in \mathcal{F}$, $\mathbb{P}[A]$ is just the marginal distribution of A.*

Definition 3.2 (Correlation) *For any two linear subspaces A and B of $L^2(\mathbb{P})$, we define the* correlation *of A and B to be*

$$\rho(A, B; \mathbb{P}) = \rho(A, B) = \sup\{\mathsf{Cov}[f, g] : f \in A, g \in B, \mathsf{Var}[f] = \mathsf{Var}[g] = 1\}$$

For correlated spaces $\Omega^{(1)} \times \Omega^{(2)}$*, we define the* correlation $\rho(\Omega^{(1)} \times \Omega^{(2)})$ *to be*

$$\rho(\Omega^{(1)} \times \Omega^{(2)}) = \rho(L^2(\Omega^{(1)}, \mathbb{P}), L^2(\Omega^{(2)}, \mathbb{P}); \mathbb{P})$$

3.1.1 Markov operators

Markov operators play an important role in the relationship between correlated probability spaces. We will give one example of a Markov operator here, and state a couple basic properties.

Definition 3.3 (Markov operator) *Given correlated spaces* $\Omega^{(1)}$ *and* $\Omega^{(2)}$*, if* $(X, Y) \in \Omega^{(1)} \times \Omega^{(2)}$ *are distributed according to* $\mathbb{P}$*, then the* Markov operator *associated with* $(\Omega^{(1)}, \Omega^{(2)})$ *is the operator mapping* $f \in L^p(\Omega^{(2)}, \mathbb{P})$ *to* $Tf \in L^p(\Omega^{(1)}, \mathbb{P})$ *given by*

$$(Tf)(x) = \mathbb{E}[f(Y)|X = x]$$

for $x \in \Omega^{(1)}$.

Proposition 3.4 *Suppose that for* $i = 1, \ldots, n$, $\Omega_i^{(1)}$ *and* $\Omega_i^{(2)}$ *are correlated spaces under distribution* μ_i*, with associated Markov operator* T_i*. Then* $\prod_{i=1}^n \Omega_i^{(1)}, \prod_{i=1}^n \Omega_i^{(2)}$ *are correlated spaces under distribution* $\prod_{i=1}^n \mu_i$*, with associated Markov operator* $T = \bigotimes_{i=1}^n T_i$.

In particular, the Bonami-Beckner operator, which we define next, is our primary example of a Markov operator. It will play an important role in the developments that follow.

Definition 3.5 (Bonami-Beckner operator) *Given a probability space* $(\Omega, \mu, \mathcal{F})$*, the* Bonami-Beckner (noise) operator T_ρ *is the Markov operator associated with the correlated space* (Ω, Ω) *with probabilty measure*

$$\nu(x, y) = (1 - \rho)\mu(x)\mu(y) + \rho\delta(x, y)\mu(x)$$

where $\delta(x, y) = 1$ *if* $x = y$ *and* $\delta(x, y) = 0$ *otherwise, i.e., such that the conditional distribution of* Y *given* $X = x$ *is* $\rho\delta_x + (1 - \rho)\mu$ *where* δ_x *is the delta measure on* x.

On product spaces $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i, \mathcal{F})$*, we take the Bonami-Beckner operator* T_ρ *to be given by the Markov operator associated with* $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \Omega_i)$ *and* $\prod_{i=1}^n \nu_i$*. By Proposition 3.4, explicitly this is* $T_\rho = \bigotimes_{i=1}^n T_\rho^i$.

Of course, we have overloaded the notation T_ρ; it now denotes both the noise operator on formal polynomials (cf. Definition 2.8) and an operator on functions on a probability space. It is easiliy verified, though, that when we interpret the formal polynomials as functions on a probability space, that the two definitions coincide.

A more concrete example of correlated spaces and their associated Markov operator, corresponding to random samples of votes, will be given in Section 4. The Markov operator also appears in the solution of the following key maximization problem:

Lemma 3.6 *Let* $\Omega^{(1)} \times \Omega^{(2)}$ *be correlated spaces under* $\mathbb{P}$*, and let* f *be a* $\Omega^{(2)}$ *measurable function satisfying* $\mathbb{E}[f] = 0$ *and* $\mathbb{E}[f^2] = 1$*. Put* $\mathcal{G} = \{g : \Omega^{(1)} \text{ measurable}, \ \mathbb{E}[g^2] = 1\}$*. Then*

$$\arg\max_{g \in \mathcal{G}} |\mathbb{E}[fg]| = \frac{Tf}{\sqrt{\mathbb{E}[(Tf)^2]}}$$

where T is the Markov operator associated with $\Omega^{(1)} \times \Omega^{(2)}$ and $\mathbb{P}$. Moreover,

$$|\mathbb{E}[fg]| = \frac{|\mathbb{E}[f(Tf)]|}{\sqrt{\mathbb{E}[(Tf)^2]}} = \sqrt{\mathbb{E}[(Tf)^2]}$$

This lemma is used in the next section to derive bounds on Markov operators on correlated spaces. It is also used directly in Section 4.

3.1.2 Efron-Stein decompositions

We now define the Efron-Stein decomposition of a function, introduced by Efron and Stein [5]. This representation will be useful to us when we work with Markov operators.

Given $x \in \prod_{i=1}^n \Omega_i$ and $S \subseteq [n]$, we let x_S denote $(x_i : i \in S)$ and $\Omega_S = \prod_{i \in S} \Omega_i$. Similarly, suppose that X_i is a random variable on Ω_i for $i \in [n]$; then we write X_S for $(X_i : i \in S)$, a random variable on Ω_S.

Proposition 3.7 (Efron-Stein decomposition) *Let $\Omega = \prod_{i=1}^n \Omega_i$, $\mu = \prod_{i=1}^n \mu_i$, $(\Omega, \mu, \mathcal{F})$ be a probability space, and let f be a real-valued function on Ω. Then there exists a unique collection of functions $f_S : \Omega_S \to \mathbb{R}$ for each $S \subseteq [n]$ such that*

1. $f(x) = \sum_{S \subseteq [n]} f_S(x_S)$
2. $\forall S \not\subseteq S', x_{S'} \in \Omega_{S'}$ $\mathbb{E}[f_S | X_{S'} = x_{S'}] = 0$

We refer to the representation of f in item 1 as the *Efron-Stein decomposition of f*.

The following two propositions are used crucially in the proof of an important Lemma described in Section 3.3. The first says that the noise operator commutes with the Efron-Stein decomposition;

Proposition 3.8 *For $i \in [n]$, let $\Omega_i^{(1)} \times \Omega_i^{(2)}$ be correlated spaces under μ_i, and let T_i be the associated Markov operator. Put $\Omega^{(1)} = \prod_{i=1}^n \Omega_i^{(1)}$, $\Omega^{(2)} = \prod_{i=1}^n \Omega_i^{(2)}$, $\mu = \prod_{i=1}^n \mu_i$, and $T = \bigotimes_{i=1}^n T_i$. Suppose $f \in L^2(\Omega^{(2)}, \mu)$ has Efron-Stein decomposition $\sum_{S \subseteq [n]} f_S(x_S)$. Then the Efron-Stein decomposition of (Tf) satisfies $(Tf)_S = T(f_S)$.*

Proposition 3.9 *For $i \in [n]$, let $\Omega_i^{(1)} \times \Omega_i^{(2)}$ be correlated spaces under μ_i such that $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mu_i) \leq \rho_i$, with associated Markov operator T_i. Put $\Omega^{(1)} = \prod_{i=1}^n \Omega_i^{(1)}$, $\Omega^{(2)} = \prod_{i=1}^n \Omega_i^{(2)}$ and $T = \bigotimes_{i=1}^n T_i$ as before. Then for all f, $\|T(f)_S\|_2 \leq \left(\prod_{i \in S} \rho_i\right) \|f_S\|_2$.*

This proposition is crucial later, and its proof is slightly more involved, so we briefly sketch it here. If we consider the operator $T^{(r)}$ for $r = 1, \ldots, n$ that switches the rth index from $\Omega^{(1)}$ to $\Omega^{(2)}$, i.e., given by

$$(T^{(r)}g)(w_1, \ldots, w_n) = \mathbb{E}[g(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1}, \mathbf{Y}_r, \ldots, \mathbf{Y}_n) | (\mathbf{X}_1, \ldots, \mathbf{X}_r, \mathbf{Y}_{r+1}, \ldots, \mathbf{Y}_n) = (w_1, \ldots, w_n)]$$

then $T = T^{(1)} \cdots T^{(n)}$. Now, using the fact that $\mathbb{E}[f_S | X_{S'} = x_{S'}] = 0$ for any $S' \subsetneq S$ and $x_{S'}$, and that $T^{(r)}$ is a Markov operator, we can repeatedly apply Lemma 3.6 to find that $\|T^{(r)}g\|_2 \leq \rho_r \|g\|_2$ (and the property that $\mathbb{E}[T^{(r)}g | X_{S'} = x_{S'}] = 0$ is preserved so we can continue).

3.2 A Gaussian extremal problem

The utility of an invariance principle in solving a discrete problem such as the one in Section 4 is that it permits us to switch the underlying probability space to one where the problem at hand can be solved. In this case, the problem happened to have been solved in Gaussian space, which is not to say that the Gaussian version of the problem is necessarily easier, though the known proofs seem to rely essentially on symmetrization. We will review the solution to the relevant problem in Gaussian space at a high level next.

Let γ be the standard Gaussian measure, γ_n be the n-dimensional Gaussian measure, and let Φ be the Gaussian distribution function.

Definition 3.10 (Ornstein-Uhlenbeck operator) *Given $f \in L^2(\mathbb{R}^n, \gamma_n)$, the* Ornstein-Uhlenbeck operator *U_ρ is given by $(U_\rho f)(x) = \mathbb{E}_{y\sim\gamma_n}[f(\rho x + \sqrt{1-\rho^2}y)]$.*

Notice that when f is a multilinear polynomial over a Gaussian sequence of ensembles, $(U_\rho f)(x) = (T_\rho f)(x)$. On the other hand, for any $f, g \in L^2(\mathbb{R}^n, \gamma_n)$, we also have

$$\mathbb{E}_{\gamma_n}[fU_\rho g] = \mathbb{E}_{\gamma_n}[gU_\rho f] = \mathbb{E}[f(\mathbf{X}_1, \dots, \mathbf{X}_n)g(\mathbf{Y}_1, \dots, \mathbf{Y}_n)]$$

where each $(\mathbf{X}_i, \mathbf{Y}_i)$ is an independent two-dimensional Gaussian with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Thus, we describe the following quantities as "Gaussian noise stability" bounds:

Definition 3.11 (Gaussian noise stability) *Given $\mu \in [0, 1]$, let $\chi_\mu : \mathbb{R} \to \{0, 1\}$ be the indicator function of the interval $(-\infty, \Phi^{-1}(\mu)]$ so that $\mathbb{E}_\gamma[\chi_\mu] = \mu$. Then, if (X, Y) is a two-dimensional Gaussian with covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, we define*

$$\overline{\Gamma}_\rho(\mu, \nu) = \mathbb{P}[X \le \Phi^{-1}(\mu), Y \le \Phi^{-1}(\nu)]$$
$$\underline{\Gamma}_\rho(\mu, \nu) = \mathbb{P}[X \le \Phi^{-1}(\mu), Y \ge \Phi^{-1}(1-\nu)]$$

Notice that, e.g., if $\rho > 0$, then $\underline{\Gamma}_\rho(\mu, \nu) < \overline{\Gamma}_\rho(\mu, \nu)$. These two quantities will give us upper and lower bounds later.

Mossel et al. [13] argue that bounds on the noise stability of $L^2(\mathbb{R}^n, \gamma_n)$ functions (with range bounded by $[0, 1]$) are given by bounds on the noise stability of indicator functions having the same mean (these are the extremal points). Thus, the relevant problem in Gaussian space turns out to be a generalization of an isoperimetric problem, first solved (in greater generality still) by Borell [2]. Borell's paper is not the best place for the uninitiated to start—we would instead recommend the interested reader consult Ledoux's lecture notes [10], starting with a review of isoperimetric problems in Chapter 1, and continuing with the summary of the literature in this area (and this result in particular) in Chapter 8. In particular, the proofs given by Beckner [1] or Carlen and Loss [3] (referenced there) may be more directly accessible.

The solution to the isoperimetric problem in Gaussian space – the set of given Gaussian measure with greatest Gaussian noise stability – turns out to be a halfspace, analogous to how the solution to the isoperimetric problem on a sphere turns out to be a cap (since these are Euclidean balls on the surface). Indeed, if one is only concerned with the case where the number of dimensions $n \to \infty$, then we know (e.g., by a law of large numbers or simply Gaussian tail bounds) that an

n-dimensional Gaussian is close to being uniformly distributed on the surface of a sphere of radius $\sqrt{n}$, and Feige and Schectman [6] give a rough, hands-on argument for a cap (half-space) being the set with the best noise stability. In any case, since the indicator function of a half-space under the measure γ_n behaves like a half-line under the one-dimensional Gaussian measure, we can express our bounds in terms of Φ^{-1} (cf. Definition 3.11):

Theorem 3.12 *Let $f, g : \mathbb{R}^n \to [0,1]$ be two γ_n measurable functions with $\mathbb{E}[f] = \mu$ and $\mathbb{E}[g] = \nu$. Then, for all $\rho \in [0,1]$ we have $\underline{\Gamma}_\rho(\mu,\nu) \le \mathbb{E}[fU_\rho g] \le \overline{\Gamma}_\rho(\mu,\nu)$.*

The theorem is connected to arbitrary correlated spaces via the following corollary:

Corollary 3.13 *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ be jointly Gaussian such that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent, $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ are independent, and*

$$\sup_{\|\alpha\|_2 = \|\beta\|_2 = 1} \left| \mathsf{Cov}\left[\sum_{i=1}^n \alpha_i \mathbf{X}_i, \sum_{i=1}^m \beta_i \mathbf{Y}_i\right]\right| \le \rho.$$

Let $f : \mathbb{R}^n \to [0,1]$ with $\mathbb{E}[f] = \mu$ and $g : \mathbb{R}^m \to [0,1]$ with $\mathbb{E}[g] = \nu$. Then,

$$\underline{\Gamma}_\rho(\mu,\nu) \le \mathbb{E}[f(\mathbf{X}_1, \ldots, \mathbf{X}_n) g(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)] \le \overline{\Gamma}_\rho(\mu,\nu).$$

Roughly, one first constructs a new basis $\{\tilde{\mathbf{X}}_i\}$, $\{\tilde{\mathbf{Y}}_i\}$ for each of the probability spaces by repeatedly taking linear combinations that maximize the covariance on the portion of the space orthogonal to the previously constructed elements of the basis; the constructed basis then has covariances $\mathsf{Cov}[\tilde{\mathbf{X}}_i, \tilde{\mathbf{Y}}_j] = 0$ for $i \neq j$ and $|\mathsf{Cov}[\tilde{\mathbf{X}}_i, \tilde{\mathbf{Y}}_i]| \le \rho$. In this new basis, we can rewrite g as $(U_\rho g')$ where $\mathbb{E}[g'] = \mathbb{E}[g]$, and apply Theorem 3.12.

3.3 Noise stability bounds via invariance

We now state and prove the main results of Mossel [12], using the solution of the Gaussian isoperimetric problem of the previous section to bound the product of two discrete functions with bounded ranges, influences, and correlation.

We begin with a couple of lemmas that are used in the proof of the main result, Theorem 3.16. Roughly following Mossel, we have broken the main result into two stages, where the first lemma below is used in the first stage, and the second lemma is used in the second stage.

Lemma 3.14 *Let $\Omega_1, \ldots, \Omega_n, \Lambda_1, \ldots, \Lambda_n$ be a collection of finite probability spaces such that $\rho(\Omega_i, \Lambda_i) \le \rho$, and let $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}$ be sequences of ensembles of random variables such that for each i, the ensembles $(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{Y}}_i)$ are independent, $\boldsymbol{\mathcal{X}}_i$ is a basis for the functions on Ω_i, and $\boldsymbol{\mathcal{Y}}_i$ is a basis for the functions on Λ_i. Let P and Q be multi-linear polynomials, and let $\epsilon > 0$ be given. Then there is some constant C such that for $\eta = C\frac{(1-\rho)\epsilon}{\log(1/\epsilon)} \le 1 - (1-\epsilon)^{\log\rho/(\log\epsilon + \log\rho)}$ we have*

$$|\mathbb{E}[P(\boldsymbol{\mathcal{X}})Q(\boldsymbol{\mathcal{Y}})] - \mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{X}})T_{1-\eta}Q(\boldsymbol{\mathcal{Y}})]| \le 2\epsilon \mathsf{Var}[P]\mathsf{Var}[Q]$$

Proof We first obseve that when we treat the noise operator as a Markov operator, it follows from the properties of conditional expectations that we can write $\mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{X}})T_{1-\eta}Q(\boldsymbol{\mathcal{Y}})]$ as $\mathbb{E}[P(\boldsymbol{\mathcal{X}})T_{1-\eta}Q(\boldsymbol{\mathcal{Y}})]$. We then need only to bound $\mathbb{E}[P(\boldsymbol{\mathcal{X}})(I - T_{1-\eta})Q(\boldsymbol{\mathcal{Y}})]$, where by using the

Markov operator T for our correlated spaces spanned by $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, we see that we can write this as $\mathbb{E}[P(\boldsymbol{\mathcal{X}})(T(I - T_{1-\eta})Q)(\boldsymbol{\mathcal{X}})]$. Expanding Q into its Efron-Stein decomposition and applying Propositions 3.8 and 3.9, we find that

$$\|T(I - T_{1-\eta})Q_S\|_2 \leq \rho^{|S|}(1 - (1 - \eta)^{|S|})\|Q_S\|_2 \leq \epsilon\|Q_S\|_2$$

If we now also expand P into its Efron-Stein decomposition and note that P_S and $(T(I-T_{1-\eta})Q)_{S'}$ are orthogonal for $S \neq S'$, we can use Cauchy-Schwarz to obtain the desired bound. ■

Lemma 3.15 *Let $f, g : \Omega^n \to [0,1]$ and let $S \subset [n]$ be a set of coordinates such that for each $i \in S$,* $\min(\mathrm{Inf}_i(f), \mathrm{Inf}_i(g)) \leq \epsilon$. *Define*

$$f'(x) = \mathbb{E}[f(\mathbf{Y})|\mathbf{Y}_{[n]\setminus S} = x_{[n]\setminus S}],\ \ g'(x) = \mathbb{E}[g(\mathbf{Y})|\mathbf{Y}_{[n]\setminus S} = x_{[n]\setminus S}]$$

Then f' and g' do not depend on the coordinates in S and

$$|\mathbb{E}[fg] - \mathbb{E}[f'g']| \leq 2|S|\sqrt{\epsilon}$$

Proof As f' and g' are defined by averages of the values of f and g, it is clear that f' and g' defined in this way are functions with range in $[0,1]$ and satisfying the influence conditions. It therefore suffices to treat the case $|S| = 1$ and apply the result inductively to extend it to larger sets. Suppose without loss of generality that $S = \{1\}$ and $\mathrm{Inf}_1(g) \leq \epsilon$; then $\mathbb{E}[(g - g')^2] \leq \epsilon$, and by Cauchy-Schwarz, $\mathbb{E}[|g - g'|] \leq \sqrt{\epsilon}$. Now, since $|f| \leq 1$,

$$|\mathbb{E}[fg] - \mathbb{E}[fg']| \leq \mathbb{E}[|g - g'|] \leq \sqrt{\epsilon}$$

On the other hand, if $\mathbb{E}_1$ denotes the expected value with respect to the first coordinate, we then find that since g' is not a function of the first coordinate, $\mathbb{E}_1[fg'] = g'\mathbb{E}_1[f] = f'g'$. Thus, $\mathbb{E}[fg'] = \mathbb{E}[f'g']$, and the claim for $S = \{1\}$ follows. ■

We now turn to the main result, applying the invariance principle (using Lemma 2.14) with the Gaussian noise stability bounds (via Corollary 3.13) to obtain a bound on the product of two functions on correlated spaces:

Theorem 3.16 *For $i = 1, \ldots, n$, let $\Omega_i^{(1)} \times \Omega_i^{(2)}$ be a sequence of independent finite correlated spaces under $\mathbb{P}_i$ such that the minimum $\mathbb{P}_i$ probability of any atom is at least $\alpha \leq 1/2$ and such that $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mathbb{P}_i) \leq \rho$. Then for every $\epsilon > 0$ there is an absolute constant C such that for $\tau = \tau(\epsilon) = \epsilon^{C^{\frac{\log(1/\alpha)\log(1/\epsilon)}{\epsilon(1-\rho)}}} < 1/2$, if $f : \prod_{i=1}^n \Omega_i^{(1)} \to [0,1]$ and $g : \prod_{i=1}^n \Omega_i^{(2)} \to [0,1]$ satisfy*

$$\forall i\ \ \min(\mathrm{Inf}_i^{\leq \log^2(1/\tau)/\log(1/\alpha)}(f), \mathrm{Inf}_i^{\leq \log^2(1/\tau)/\log(1/\alpha)}(g)) \leq \tau$$

then

$$\underline{\Gamma}_\rho(\mathbb{E}[f], \mathbb{E}[g]) - \epsilon \leq \mathbb{E}[fg] \leq \overline{\Gamma}_\rho(\mathbb{E}[f], \mathbb{E}[g]) + \epsilon$$

The proof of this theorem proceeds in two stages. In the first stage, the conclusion is derived under the weaker hypothesis that

$$\max_i \{\mathrm{Inf}_i^{\leq \log(1/\tau)/\log(1/\alpha)}(f), \mathrm{Inf}_i^{\leq \log(1/\tau)/\log(1/\alpha)}(g)\} \leq \tau$$

which is then removed in the second stage using Lemma 3.15.

Proof Stage 1. Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be sequences of ensembles such that $\boldsymbol{\mathcal{X}}_i$ spans $\Omega_i^{(1)}$ and $\boldsymbol{\mathcal{Y}}$ spans $\Omega_i^{(2)}$, so that f and g can be written as (formal) multilinear polynomials P and Q over $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$. By Lemma 3.14, for $\eta = O(\epsilon(1-\rho)/\log(1/\epsilon))$, we have

$$|\mathbb{E}[P(\boldsymbol{\mathcal{X}})Q(\boldsymbol{\mathcal{Y}})] - \mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{X}})T_{1-\eta}Q(\boldsymbol{\mathcal{Y}})]| \leq \epsilon/4 \tag{4}$$

For Stage 2, it will also be important to note that, by choosing our constant C in the definition of τ sufficiently large, we can guarantee

$$(1-\eta)^{\log(1/\tau)/\log(1/\alpha)} \leq \epsilon/4 \tag{5}$$

Let us now choose Gaussian ensembles $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{H}}$ such that for all i, the covariance matrix of $\boldsymbol{\mathcal{G}}_i$ and $\boldsymbol{\mathcal{H}}_i$ matches that of $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{Y}}_i$, and $(\boldsymbol{\mathcal{G}}_i, \boldsymbol{\mathcal{H}}_i)$ is independent of $(\boldsymbol{\mathcal{G}}_j, \boldsymbol{\mathcal{H}}_j)$ for $i \neq j$. Recalling that $T_{1-\eta}P$ and $T_{1-\eta}Q$ are multilinear polynomials, it is immediate from the choice of $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{H}}$ that $\mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{G}})] = \mu$, $\mathbb{E}[T_{1-\eta}Q(\boldsymbol{\mathcal{H}})] = \nu$, and

$$\mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{X}})T_{1-\eta}Q(\boldsymbol{\mathcal{Y}})] = \mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{G}})T_{1-\eta}P(\boldsymbol{\mathcal{H}})] \tag{6}$$

We now recall the functions

$$f_{[0,1]}(x) = \begin{cases} 0 & x < 0 \\ x & x \in [0,1] \\ 1 & x > 1 \end{cases}$$

$$\zeta(x_1, x_2) = (x_1 - f_{[0,1]}(x_1))^2 + (x_2 - f_{[0,1]}(x_2))^2$$

Since f and g were functions taking values in $[0,1]$, recalling the definition of $T_{1-\eta}$ as a Markov operator, we see that $(T_{1-\eta}P(\boldsymbol{\mathcal{X}}), T_{1-\eta}Q(\boldsymbol{\mathcal{Y}}))$ is a $[0,1]^2$-valued random variable, and hence

$$\zeta(T_{1-\eta}P(\boldsymbol{\mathcal{X}}), T_{1-\eta}Q(\boldsymbol{\mathcal{Y}})) = 0.$$

We now apply the invariance principle. Put $P' = f_{[0,1]}T_{1-\eta}P$ and $Q' = f_{[0,1]}T_{1-\eta}Q$ and likewise put $\mu' = \mathbb{E}[P'(\boldsymbol{\mathcal{G}})]$ and $\nu' = \mathbb{E}[Q'(\boldsymbol{\mathcal{H}})]$. Note that we have assumed that the low-degree influences are low (this is Stage 1) and applied a noise operator so that the high-degree influences are low, bounding the tails of the polynomials. It now follows from Lemma 2.14 that

$$\mathbb{E}[\zeta(T_{1-\eta}P(\boldsymbol{\mathcal{G}}), T_{1-\eta}Q(\boldsymbol{\mathcal{H}}))] = \|(T_{1-\eta}P, T_{1-\eta}Q) - (P', Q')\|_2^2 \leq \tau^{\Omega(\eta/\log(1/\alpha))} \tag{7}$$

where Cauchy-Schwarz then gives

$$|\mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{G}})T_{1-\eta}Q(\boldsymbol{\mathcal{H}}) - P'(\boldsymbol{\mathcal{G}})Q'(\boldsymbol{\mathcal{H}})]| \leq \tau^{\Omega(\eta/\log(1/\alpha))}. \tag{8}$$

Recalling that $\mathbb{E}[T_{1-\eta}P(\boldsymbol{\mathcal{G}})] = \mu$ and $\mathbb{E}[T_{1-\eta}Q(\boldsymbol{\mathcal{H}})] = \nu$, we see that Cauchy-Schwarz on the individual components of Inequality 7 also gives $|\mu - \mu'| \leq \tau^{\Omega(\eta/\log(1/\alpha))}$ and similarly for ν and ν'.

We next apply our noise stability bounds: Corollary 3.13 gives

$$\underline{\Gamma}_\rho(\mu', \nu') \le \mathbb{E}[P'(\boldsymbol{\mathcal{G}})Q'(\boldsymbol{\mathcal{H}})] \le \overline{\Gamma}_\rho(\mu', \nu')$$

where, recalling the definition of $\overline{\Gamma}_\rho$, we see

$$\begin{aligned}
|\overline{\Gamma}_\rho(\mu, \nu) - \overline{\Gamma}_\rho(\mu', \nu')| &= |\mathbb{P}[\mathbf{X} \le \Phi^{-1}(\mu), \mathbf{Y} \le \Phi^{-1}(\nu)] - \mathbb{P}[\mathbf{X} \le \Phi^{-1}(\mu'), \mathbf{Y} \le \Phi^{-1}(\nu')]| \\
&= |\mathbb{P}[\mathbf{X} \le \Phi^{-1}(\mu), \mathbf{Y} \le \Phi^{-1}(\nu)] - \mathbb{P}[\mathbf{X} \le \Phi^{-1}(\mu'), \mathbf{Y} \le \Phi^{-1}(\nu)] \\
&\qquad + \mathbb{P}[\mathbf{X} \le \Phi^{-1}(\mu'), \mathbf{Y} \le \Phi^{-1}(\nu)] - \mathbb{P}[\mathbf{X} \le \Phi^{-1}(\mu'), \mathbf{Y} \le \Phi^{-1}(\nu')]| \\
&\le |\mu - \mu'| + |\nu - \nu'| \\
&\le \tau^{\Omega(\eta/\log(1/\alpha))}.
\end{aligned}$$

and similarly for $|\underline{\Gamma}_\rho(\mu, \nu) - \underline{\Gamma}_\rho(\mu', \nu')|$.

Recalling Equation 6 and Inequalities 4 and 8, we now find

$$\underline{\Gamma}_\rho(\mu, \nu) - \tau^{\Omega(\eta/\log(1/\alpha))} - \epsilon/4 \le \mathbb{E}[P(\boldsymbol{\mathcal{X}})Q(\boldsymbol{\mathcal{Y}})] \le \overline{\Gamma}_\rho(\mu, \nu) + \tau^{\Omega(\eta/\log(1/\alpha))} + \epsilon/4$$

where a sufficiently large choice of C for τ gives a bound of

$$\underline{\Gamma}_\rho(\mu, \nu) - \epsilon/2 \le \mathbb{E}[P(\boldsymbol{\mathcal{X}})Q(\boldsymbol{\mathcal{Y}})] \le \overline{\Gamma}_\rho(\mu, \nu) + \epsilon/2.$$

Stage 2. Put $R = \frac{\log(1/\tau)}{\log(1/\alpha)}$, $R' = R\log(1/\tau)$, and $\tau' = \tau^{C'}$ for some large C' to be determined. Note that $R' > R$ and $\tau' < \tau$. Now suppose that f and g satisfy

$$\forall i \ \min(\mathrm{Inf}_i^{\le R'}(f), \mathrm{Inf}_i^{\le R'}(g)) \le \tau'$$

as in the statement of the theorem.

Put $S_f = \{i : \mathrm{Inf}_i^{\le R}(f) > \tau\}$, $S_g = \{i : \mathrm{Inf}_i^{\le R}(g) > \tau\}$, and $S = S_f \cup S_g$. Note that S_f and S_g are disjoint.

Consider the smoothed versions of f and g, $T_{1-\eta}P$ and $T_{1-\eta}Q$ we constructed in Stage 1 satisfying Inequality 4. Notice that whenever $i \in S_f$, our hypothesis guarantees that g must satisfy $\mathrm{Inf}_i^{\le R'}(g) \le \tau'$ (and similarly when $i \in S_g$). Since the noise operator only decreases influences, this then gives

$$\mathrm{Inf}_i(T_{1-\eta}Q) \le \tau' + (1-\eta)^{2R'}$$

and similarly for $\mathrm{Inf}_i(T_{1-\eta}P)$.

Therefore, by averaging over the influential coordinates in S as in Lemma 3.15, we can construct $\tilde{P}$ and $\tilde{Q}$ such that

$$\max_i \{\mathrm{Inf}_i^{\le R}(\tilde{P}), \mathrm{Inf}_i^{\le R}(\tilde{Q})\} \le \tau$$

so we can apply the result from Stage 1.

By Proposition 2.7 and the fact that f and g had bounded range, we see that $\sum_i \mathrm{Inf}_i^{\le R}(f) \le R$, so S_f contains at most R/τ coordinates (and similarly for S_g). Thus, recalling our choice of τ in Inequality 5, Lemma 3.15 tells us that

$$|\mathbb{E}[(T_{1-\eta}P)(T_{1-\eta}Q)] - \mathbb{E}[\tilde{P}\tilde{Q}]| \le \frac{2R}{\tau}\sqrt{\tau' + (1-\eta)^{2R'}} \le \frac{\epsilon}{2}$$

for C' large enough so this additional error is still within the permitted bound. ■

4 An application: Majority is most predictable

The net result of Section 3 (and indeed, the bulk of the work by Mossel [12]) was Theorem 3.16, giving bounds on the product of a pair of functions having low coordinate-wise influences on probability spaces with bounded correlation. We now show, following Mossel, an application of this theorem to a rather natural social choice problem, which we motivate with the following story:

Suppose that in the aftermath of the 2000 US Presidential election, a commision was formed to find an electoral system that featured optimal resillience to uncounted votes—that is, an electoral system that could be best predicted from a random sample of the votes. We will suppose that there are n voters, and that our electoral system is modeled by a function $f : \{-1, 1\}^n \to \{-1, 1\}$. We would like to require that the election does not discriminate between the candidates, so $f(-x) = -f(x)$ for all x, and we would like to ensure that no single voter has excessive influence over the outcome—supposing that the voters choose uniformly at random between the two candidates, this condition corresponds to a bound on the influences: $\mathrm{Inf}_i(f) \le \tau$ for every $i \in [n]$ and some τ. We will refer to such a function f as a *social choice function.*

We will show absolute bounds on how well any social choice function f can be predicted from a sample of votes; we will then show that the simple majority function approaches this bound as the number of voters $n \to \infty$. It will be helpful to recall the definitions of correlated spaces and Markov operators from Section 3.1. We will consider the following correlated space: let $\Omega^{(1)} = \{-1, +1\}$, corresponding to a vote; let $\Omega^{(2)} = \{-1, 0, +1\}$ corresponding to the sampled vote, with 0 indicating "no sample;" and let the joint probability measure be given by

$$\mu(x, y) = \frac{1}{2}(\rho\delta(x, y) + (1 - \rho)\delta_0(y))$$

Intuitively, independent of y, x is uniformly distributed over 1 and -1; furthermore, with probability ρ we sample the vote, and then $y = x$, and otherwise (with probability $1-\rho$) $y = 0$. We will briefly confirm these intuitions: notice now that the conditional distribution of our sample from $\Omega^{(1)}$ given a sample from $\Omega^{(2)}$ is

$$\mu(x| + 1) = \delta_{+1}(x)$$
$$\mu(x| - 1) = \delta_{-1}(x)$$
$$\mu(x|0) = \frac{1}{2}$$

and likewise, the marginal distribution $\mu^{(1)}(x) = 1/2$. On the other hand, the marginal distribution of $\mu^{(2)}$ is

$$\mu^{(2)}(y) = (1 - \rho)\delta_0 + \rho\frac{1}{2}(\delta_{-1} + \delta_1)$$

so μ describes the desired correlated space for a single vote. We extend the distribution to an election involving n independent voters using the probability measure $\mu = \prod_{i=1}^n \mu_i$ on the correlated spaces $(\prod_{i=1}^n \Omega_i^{(1)}, \prod_{i=1}^n \Omega_i^{(2)})$ where each μ_i is given by the μ above on $\Omega_i^{(1)} \times \Omega_i^{(2)}$.

Given a social choice function f, an optimal prediction of the outcome of the election from a sample of votes y is given by $\mathrm{sgn}((Tf)(y))$ where Tf is the Markov operator associated with $\prod_{i=1}^n \Omega_i^{(1)} \times \prod_{i=1}^n \Omega_i^{(2)}$ under μ, $(Tf)(y) = \mathbb{E}_\mu[f(X)|Y = y]$. It is easy to see that this is optimal, since in our setting

$$\mathbb{E}_\mu[f(X)|Y = y] = \mu[f(X) = 1|Y = y] - \mu[f(X) = -1|Y = y].$$

4.1 Passing correlation bounds back to the discrete space

To begin, we need to calculate the correlation between $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$ under μ so that we can apply invariance on the product of our voting rule f and its optimal prediction $\mathrm{sgn}(Tf)$ by changing the underlying probability space to a two-dimensional Gaussian having the same covariance matrix and obtaining the bounds there, encapsulated in Theorem 3.16.

Lemma 4.1 $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mu) = \sqrt{\rho}$

Proof Recalling the definition of corelation of probability spaces, we first observe that it suffices to consider the functions $f \in L^2(\Omega_i^{(1)}, \mu)$ with $\mathbb{E}[f] = 0$ (since given any $h \in L^2(\Omega_i^{(1)}, \mu)$, $f = h - \mathbb{E}[h]$ satisfies $\mathsf{Cov}[f, g] = \mathsf{Cov}[h, g]$). Next, we observe that the only $f \in L^2(\Omega_i^{(1)}, \mu)$ with $\mathbb{E}_\mu[f] = 0$ and $\mathsf{Var}_\mu[f] = 1$ is given by $f(x, y) = x$, where by Lemma 3.6 we see that it suffices to calculate the value of $\sqrt{\mathbb{E}_\mu[(Tf)^2]}$.

This is easily accomplished: with probability $1 - \rho$, $y = 0$, and then $(Tf)(0) = \mathbb{E}_\mu[f(X, Y)|Y = 0] = 0$; on the other hand, with probability ρ, $y \neq 0$, in which case $X = Y$, and then $(Tf)(y)^2 = \mathbb{E}_\mu[f(X, Y)|Y = y] = 1$. Thus, $\mathbb{E}_\mu[(Tf)^2] = \rho$, and $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mu) = \sqrt{\rho}$. ■

Notice that when the voters choose their votes uniformly, our condition on fairness to the two candidates guarantees that $\mathbb{E}[f] = 0$ and $\mathbb{E}[\mathrm{sgn}(Tf)] = 0$. We can now obtain the following main result:

Theorem 4.2 *Let $0 \leq \rho < 1$ and $\epsilon > 0$ be given. There exists $\tau > 0$ such that if $f : \{-1, 1\}^n \to [-1, 1]$ satisfies $\mathbb{E}[f] = 0$ and $\mathrm{Inf}_i(f) \leq \tau$ for all i, then*

$$\mu[f = \mathrm{sgn}(Tf)] \leq \frac{1}{2} + \frac{1}{\pi} \arcsin \sqrt{\rho} + \epsilon$$

Proof We begin with some elementary manipulations. We first notice that

$$\mu[f = \mathrm{sgn}(Tf)] = \frac{1}{2}(1 + \mathbb{E}[f\mathrm{sgn}(Tf)]).$$

Now, for $g = \frac{1}{2}(1 + f)$ and $h = \frac{1}{2}(1 + \mathrm{sgn}(f))$, g and h have range in $[0, 1]$, $\mathbb{E}[g] = \mathbb{E}[h] = 1/2$, and

$$\mathbb{E}[gh] = \frac{1}{4}(1 + \mathbb{E}[f] + \mathbb{E}[\mathrm{sgn}(Tf)] + \mathbb{E}[f\mathrm{sgn}(Tf)])$$

so we can write $\mu[f = \mathrm{sgn}(Tf)] = 2\mathbb{E}[gh]$.

We now apply Theorem 3.16 to find that there exists a $\tau > 0$ such that

$$\mathbb{E}[gh] \leq \overline{\Gamma}_{\sqrt{\rho}} \left(\frac{1}{2}, \frac{1}{2} \right) + \epsilon/2$$

Sheppard's formula [16] now gives

$$\overline{\Gamma}_{\sqrt{\rho}} \left(\frac{1}{2}, \frac{1}{2} \right) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \sqrt{\rho}$$

and the claim follows immediately. ■

Since we noted in the previous section that $\mathrm{sgn}(Tf)$ gives the optimal prediction of f, this gives an absolute bound on how well a social choice function may be predicted. The title, "majority is most predictable," is derived from the fact that this bound is asymptotically achieved by the majority function, $\mathrm{Maj}_n(x_1, \ldots, x_n) = \mathrm{sgn}\left(\sum_{i=1}^n x_i\right)$:

Theorem 4.3

$$\lim_{n\to\infty} \mu[\mathrm{Maj}_n = \mathrm{sgn}(T\mathrm{Maj}_n)] = \frac{1}{2} + \frac{1}{\pi}\arcsin\sqrt{\rho}$$

We note that a version of this calculation was done with great care by O'Donnell [14].

Proof We put $\mathbf{X} = \frac{1}{\sqrt{n}}\sum_{i=1}^n x_i$ and $\mathbf{Y} = \frac{1}{\sqrt{\rho n}}\sum_{i=1}^n x_i y_i$ where $y_i \in \{0,1\}$ indicates whether or not the ith voter was included in the sample and $x_i \in \{-1,1\}$ is the ith voter's vote. Observe that $\mathbf{X}$ gives the bias of the actual vote, and $\mathbf{Y}$ gives the bias of the sample.

Now, both $\mathbb{E}[x_i]$ and $\mathbb{E}[x_i y_i]$ are zero, and for each ith voter, the covariance matrix of $(x_i, \rho^{-1/2}x_i y_i)$ is given by

$$\Sigma = \begin{pmatrix} 1 & \sqrt{\rho} \\ \sqrt{\rho} & 1 \end{pmatrix}$$

Therefore, it follows from the multidimensional Central Limit Theorem that $(\mathbf{X}, \mathbf{Y})$ converges to a two-dimensional Gaussian with covariance matrix Σ. Recalling the definition of Gaussian noise stability now,

$$\lim_{n\to\infty} \mu[\mathrm{sgn}(\mathbf{X}) = -1, \mathrm{sgn}(\mathbf{Y}) = -1] = \overline{\Gamma}_{\sqrt{\rho}}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{4} + \frac{1}{2\pi}\arcsin\sqrt{\rho}$$

by Sheppard's formula again. By symmetry, we also have

$$\lim_{n\to\infty} \mu[\mathrm{sgn}(\mathbf{X}) = 1, \mathrm{sgn}(\mathbf{Y}) = 1] = \frac{1}{4} + \frac{1}{2\pi}\arcsin\sqrt{\rho}$$

and hence, since these are the two values taken by the sgn function,

$$\lim_{n\to\infty} \mu[\mathrm{sgn}(\mathrm{Maj}_n) = \mathrm{sgn}(T\mathrm{Maj}_n)] = \lim_{n\to\infty} \mu[\mathrm{sgn}(\mathbf{X}) = \mathrm{sgn}(\mathbf{Y})] = \frac{1}{2} + \frac{1}{\pi}\arcsin\sqrt{\rho}$$

■

We now briefly remark that similar calculations for the spaces given by the Bonami-Beckner operator will also provide proofs of the results of Mossel et al. [13], such as the celebrated "Majority is Stablest" Theorem. Thus, the correlated spaces framework is strictly stronger than the noise stability framework developed there.

References

[1] William Beckner. Sobolev inequalities, the Poisson semigroup, and analysis on the sphere s^n. *Proc. Natl. Acad. Sci.*, 89:4816–4819, 1992.

[2] Christer Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related Fields*, 70(1):1–13, 1985.

[3] Eric A. Carlen and Michael Loss. Extremals of functionals with competing symmetries. *J. Funct. Anal.*, 88:437–456, 1990.

[4] Sourav Chatterjee. A generalization of the Lindeberg principle. *Ann. Probability*, 34(6):2061–2076, 2006.

[5] B. Efron and C. Stein. The jacknofe estimate of variance. *Ann. Statistics*, 9(3):586–596, 1981.

[6] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for MAX CUT. *Random Structures & Algorithms*, 20(3):403–440, 2002.

[7] H. Gebelein. Das statistische Problem der Korrelation als Variations und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Z. Angew. Math. Mech.*, 21:364–379, 1941.

[8] H. O. Hirschfeld. A connection between correlation and contingency. *Mathematical Proc. Cambridge Phil. Soc.*, 31(4):520–524, 1935.

[9] Svante Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1997.

[10] Michel Ledoux. Isoperimetry and gaussian analysis. In *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIV-1994*, volume 1648, pages 165–294. Springer, 1996.

[11] Jarl Waldemar Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Zeitschrift*, 15:211–225, 1922.

[12] Elchanan Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *Proc. 49th Symp. Found. Computer Sci.*, 2008. Full version: arXiv:math/0703683.

[13] Elchanan Mossel, Ryan O'Donnell, and Krzystof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th Symp. Found. Computer Sci.*, 2005. Full version: arXiv:math/0503503, to appear in Ann. Math. 2009.

[14] Ryan O'Donnell. Hardness amplification within NP. *J. Computer and System Sciences*, 69:68–94, 2004.

[15] V. I. Rotar'. Limit theorems for polylinear forms. *J. Multivariate Analysis*, 9(4):511–530, 1979.

[16] W. F. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London*, 192:101–168, 1899.

[17] Pawel Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, 180(3):219–236, 2007.