CMS-EMS
Center for Mathematical Studies in Economics And Management Science

Discussion Paper #1587

"A universal bound on probabilistic influence"

Bruno Strulovici

Northwestern University

October 24, 2016

A universal bound on probabilistic influence

Bruno Strulovici*
Northwestern University

October 24, 2016

Abstract

This paper establishes a new probability result: Fix $\eta \in (0, 1/2]$ and let $\{X_i\}_{i\in\mathbb{N}}$ denote a family of independent random variables such that $Pr(X_i = 1) = 1 - Pr(X_i = 0) = p_i \in [\eta, 1-\eta]$. For any $K > 0$, $\varepsilon > 0$, and function $f : \{0,1\}^n \to [0, K]$,

$$|\{i : E[f(X_1, \ldots, X_n)|X_i = 1] - E[f(X_1, \ldots, X_n)|X_i = 0] > \varepsilon\}| \le N(\varepsilon, K, \eta) + 1 \tag{1}$$

where $N(\varepsilon, K, \eta) = \frac{K^2}{2\pi\varepsilon^2\eta^2}$.

In many environments, a dependent variable potentially depends on a large number of independent binary variables. In economics, an agent's utility may depend on a large number of stochastic consumption events, a market price can depend on a large number of individual decisions, a production function may depend on many stochastic factors, etc. In econometrics, the dependent variable in a nonlinear regression can potentially depend on many independent variables. In genomics, the occurrence of disease can potentially depend on a large number of genes. In quantum mechanics, the realization of a phenomenon can depend on the spin positions of large number of particles, etc. This paper shows, however, that if the variable of interest is bounded, it can only depend *significantly* on at most a fixed number of independent variables. This number depends only on the significance level of interest and on the probability range taken by the independent variables.

Lemma 1 *Let $\{X_i\}_{i\in\mathbb{N}}$ denote a family of independent random variables such that $Pr(X_i = 1) = Pr(X_i = 0) = 1/2$. For any $K > 0$, $\varepsilon > 0$, and function $f : \{0,1\}^n \to [0, K]$,*

$$|\{i \in \{1, \ldots, n\} : E[f(X_1, \ldots, X_n)|X_i = 1] - E[f(X_1, \ldots, X_n)|X_i = 0] > \varepsilon\}| \le N(\varepsilon, K) + 1$$

where $N(\varepsilon, K) = \frac{2}{\pi}\frac{K^2}{\varepsilon^2}$.

This implies in particular that the fraction of independent random variables whose positive influence exceeds ε goes to zero as n gets large, at rate $1/n$. Obviously, a symmetric statement with the same bound holds for negative influence.

*I gratefully acknowledge financial support from an NSF CAREER Award (Grant No. 1151410) and a fellowship from the Alfred P. Sloan Foundation

Proof. Fix some K,ε,n, and f, and let $\mathcal{N}_f = \{i : E[f(X_1,\ldots,X_n)|X_i = 1] - E[f(X_1,\ldots,X_n)|X_i = 0] > \varepsilon\}$ and $n_f = |\mathcal{N}_f|$. We will derive an upper bound N on n_f. Any bound N obtained for the parameters (ε, K) is also a bound for the parameters $(\alpha\varepsilon, \alpha K)$, for any $\alpha > 0$. Therefore, we focus on $K = 1$.

By relabeling coordinates, we can assume without loss of generality that $N_f = \{1,\ldots,n_f\}$. By letting $g(X_1,\ldots,X_{n_f}) = E[f(X_1,\ldots,X_n)|(X_1,\ldots X_{n_f})]$, the law of iterated expectations implies that for $i \in \{1,\ldots,n_f\}$, $E[g(X_1,\ldots,X_{n_f})|X_i = 1] - E[g(X_1,\ldots,X_{n_f})|X_i = 0] = E[f(X_1,\ldots,X_n)|X_i = 1] - E[f(X_1,\ldots,X_n)|X_i = 0] > \varepsilon$ and hence $n_g = n_f$. Therefore we may focus without loss on functions like g for which the influence of each variable exceeds ε.

Thus suppose that $n_f = n$ and let $\mathcal{X} = \{0,1\}^n$. Note that any bound N obtained when n is even implies a bound $N + 1$ for n odd. To economize on notation, we focus on n even. By assumption, we have for each $i \in \{1,\ldots,n\}$

$$\sum_{x\in\mathcal{X}:x_i=1} f(x) - \sum_{x\in\mathcal{X}:x_i=0} f(x) > \varepsilon 2^{n-1}$$

Notice that if x has j zeros and $n - j$ ones, the term $f(x)$ is counted positively in $n - j$ of the previous inequalities and negatively in the j remaining ones. Summing up the n inequalities and rearranging the terms therefore yields

$$\sum_{j=0}^{n}(n-2j) \sum_{x\in\mathcal{X}^j} f(x) > \varepsilon n 2^{n-1},$$

where $\mathcal{X}^j$ denotes the set of all elements of $\{0,1\}^n$ with exactly j zeros. Since $f(x)$ is nonnegative and bounded above by 1, this implies that

$$\sum_{i=0}^{n/2}(n-2i)\binom{n}{i} > \varepsilon n 2^{n-1}. \tag{2}$$

Because $\binom{n}{i} = \binom{n}{n-i}$ and $\sum_{i=0}^{n}\binom{n}{i} = 2^n$, the left-hand side of (2) is equal

$$n\sum_{i=0}^{n/2}\binom{n}{i} - 2\sum_{i=1}^{n/2} i\binom{n}{i} = \frac{n}{2}\left((2^n + \binom{n}{n/2}\right) - 2\sum_{i=1}^{n/2} i\binom{n}{i}.$$

Since $i\binom{n}{i} = n\binom{n-1}{i-1}$ and $\binom{n-1}{i} = \binom{n-1}{n-1-i}$, the last term of the previous expression equals

$$2n\sum_{j=0}^{n/2-1}\binom{n-1}{j} = n\sum_{j=0}^{n-1}\binom{n-1}{j} = n2^{n-1}$$

Therefore, (2) implies that

$$\frac{n}{2}\binom{n}{n/2} > \varepsilon n 2^{n-1},$$

or

$$\frac{n!}{(n/2)!^2} > \varepsilon 2^n.$$

Using Stirling's formula,[1] this implies that

$$\sqrt{\frac{2}{\pi n}} > \varepsilon$$

or

$$n < \frac{2}{\pi \varepsilon^2}.$$

This shows the lemma for

$$N(\varepsilon, K) = \frac{2}{\pi}\frac{K^2}{\varepsilon^2}.$$

LEMMA 2 *Fix* $q \in \mathbb{N}_{++}$ *and* $\eta \in (0, 1/2]$*, and let* $\{X_i\}_{i\in\mathbb{N}}$ *denote a family of independent random variables such that* $Pr(X_i = 1) = 1 - Pr(X_i = 0) = a_i 2^{-q}$ *for some integer* a_i *such that* $a_i 2^{-q} \in [\eta, 1-\eta]$*. For any* $K > 0$*,* $\varepsilon > 0$*, and function* $f : \{0,1\}^n \to [0, K]$*,*

$$|\{i : E[f(X_1, \ldots, X_n)|X_i = 1] - E[f(X_1, \ldots, X_n)|X_i = 0] > \varepsilon\}| \leq N(\varepsilon, K, \eta) + 1$$

where $N(\varepsilon, K, \eta) = \frac{K^2}{2\pi\varepsilon^2\eta^2}$*.*

Proof. As with Lemma 1, we can focus on $K = 1$ and n even and obtain the result for general $K > 0$ by jointly rescaling K and ε and adding 1 to the upper bound to allow for n odd. Fix some integer n and function $f : \mathcal{X} = \{0,1\}^n \to [0,1]$. For each $i \in \{1, \ldots, n\}$ we will decompose X_i into q binary variables $\{\tilde{X}_i^k\}_{k=1,\ldots,q}$ which are i.i.d. with $Pr(\tilde{X}_i^k = 1) = Pr(\tilde{X}_i^k = 1) = 1/2$. For each $i \leq n$, let $\mathcal{A}_i$ denote a subset of $\{0,1\}^q$ with a_i elements, containing all elements such that $X_i^1 = 1$ if $a_i \geq 2^{q-1}$, and covered by these elements otherwise. Letting $\tilde{X}_i = (\tilde{X}_i^1, \ldots, \tilde{X}_i^q)$, notice that $Pr(\tilde{X}_i \in \mathcal{A}_i) = a_i/2^q$. We now show that

$$Pr(\tilde{X}_i \in \mathcal{A}_i|\tilde{X}_i^1 = 1) - Pr(\tilde{X}_i \in \mathcal{A}_i|\tilde{X}_i^1 = 0) \geq 2\eta. \tag{3}$$

Suppose first that $a_i \geq 2^{q-1}$. In this case, by construction $\tilde{X}_i$ surely belongs to $\mathcal{A}_i$ if $\tilde{X}_i^1 = 1$. Let $\delta = Pr(\tilde{X}_i \in \mathcal{A}_i|\tilde{X}_i^1 = 0)$. We have $1/2 + \delta/2 = a_i/2^q$. By assumption, $a_i/2^q \leq 1 - \eta$. Combining this yields $1 - \delta \geq 2\eta$, as desired. If instead $a_i < 2^{q-1}$, $\tilde{X}_i$ cannot belong to $\mathcal{A}_i$ if $\tilde{X}_i^1 = 0$, and belongs to it with probability $a_i/2^{q-1} \geq 2\eta$ if $\tilde{X}_i^1 = 1$, which yields again (3).

Let $\tilde{f} : \{0,1\}^{np} \to [0,1]$ be defined by

$$\tilde{f}(\tilde{x}_1, \ldots, \tilde{x}_n) = f(1_{\tilde{x}_1\in\mathcal{A}_1}, \ldots, 1_{\tilde{x}_n\in\mathcal{A}_n}).$$

[1]While Stirling's formula is an approximation, the bound obtained here is exact: from Robbins (1955), we have $\sqrt{2\pi}n^{n+1/2}e^{-n}e^{1/(12n+1)} < n! < \sqrt{2\pi}n^{n+1/2}e^{-n}e^{1/12n}$. This implies that $n!/((n/2)!)^2$ is bounded above by $(2\pi n)^{-1/2}2^{n+1}e^{1/(12n)-2/(6n+1)}$. The last factor is less than one for all $n \geq 1$ and can thus be dropped from the upper bound.

By construction, the random variable $\tilde{f}(\tilde{X})$ has the same distribution as $f(X)$. Moreover,

$$E[f(X)|X_i = 1] = E[f(\tilde{X})|\tilde{X}_i \in \mathcal{A}_i]$$

and

$$E[f(X)|X_i = 0] = E[f(\tilde{X})|\tilde{X}_i \notin \mathcal{A}_i].$$

The function $\tilde{f}$ depends on nq iid symmetric Bernoulli variables. Lemma 1 can thus be applied to $\tilde{f}$, and shows that

$$E[\tilde{f}(\tilde{X})|\tilde{X}_i^k = 1] - E[\tilde{f}(\tilde{X})|\tilde{X}_i^k = 0] \leq \varepsilon$$

except for at most $N = \lceil 2/\pi\varepsilon^2 \rceil$ of these np variables. If $n > N$, this implies that there is a subset $\mathcal{N}$ of $\{1, \ldots, n\}$ with at least $n - N$ elements such that for all $i \in \mathcal{N}$

$$\Delta_i^k = E[\tilde{f}(\tilde{X})|\tilde{X}_i^k = 1] - E[\tilde{f}(\tilde{X})|\tilde{X}_i^k = 0] \leq \varepsilon$$

for all $k \in \{1, \ldots, q\}$, and in particular for $k = 1$. Notice that for $r \in \{0, 1\}$,

$$E[\tilde{f}(\tilde{X})|\tilde{X}_i^k = r] = Pr(\tilde{X}_i \in \mathcal{A}_i|\tilde{X}^1 = r)E[f(X)|X_i = 1] + (1 - Pr(\tilde{X}_i \in \mathcal{A}_i|\tilde{X}^1 = r))E[f(X)|X_i = 0].$$

This implies that

$$\Delta_i = (E[f(X)|X_i = 1] - E[f(X)|X_i = 0])(Pr(\tilde{X}_i \in \mathcal{A}_i|X_i^1 = 1) - Pr(\tilde{X}_i \in \mathcal{A}_i|X_i^1 = 0))$$

Since the left-hand side is less than ε for $i \in \mathcal{N}$ and the probability difference on the right-hand side exceeds η, we conclude that

$$E[f(X)|X_i = 1] - E[f(X)|X_i = 0] \leq \varepsilon/2\eta.$$

Setting $\varepsilon' = \varepsilon/2\eta$, this proves the lemma with $N(\varepsilon', 1, \eta) = 1/2\pi\varepsilon'^2\eta^2$ and, by rescaling, $N(\varepsilon, K, \eta) = K^2/2\pi\varepsilon^2\eta^2$.

THEOREM 1 *Fix* $\eta \in (0, 1/2]$ *and let* $\{X_i\}_{i \in \mathbb{N}}$ *denote a family of independent random variables such that* $Pr(X_i = 1) = 1 - Pr(X_i = 0) = p_i \in [\eta, 1 - \eta]$. *For any* $K > 0$, $\varepsilon > 0$, *and function* $f : \{0, 1\}^n \to [0, K]$,

$$|\{i : E[f(X_1, \ldots, X_n)|X_i = 1] - E[f(X_1, \ldots, X_n)|X_i = 0] > \varepsilon\}| \leq N(\varepsilon, K, \eta) + 1 \qquad (4)$$

where $N(\varepsilon, K, \eta) = \frac{K^2}{2\pi\varepsilon^2\eta^2}$.

Proof. Fix some n even, function $f : \{0, 1\} \to [0, K]$ and n-dimensional vector p such that $p_i \in [\eta, 1 - \eta]$ for all i. For each $q \in \mathbb{N}_{++}$, let p^q denote an dimensional vector taking values in the set $\{1/2^p, 2/2^p, \ldots, 1 - 1/2^p\}$ such that $p_i^q \in [\eta, 1 - \eta]$ and $|p_i^q - p_i| \leq 1/2^q$ for all i. Let E_q denote the expectation taken with respect to the probability vector p^q. From Lemma 2 applied to p^q, we have

$$E_q[f(X)|X_i = 1] - E_q[f(X)|X_i = 0] \leq \varepsilon \tag{5}$$

for all i, except for a subset $\mathcal{N}_q$ of $\{1, \ldots, n\}$ which contains at most $N(\varepsilon, K, \eta) = K^2/(2\pi\varepsilon^2\eta^2)$ elements. Since the set $\{1, \ldots, n\}$ is finite, there exists an infinite subsequence of $\{\mathcal{N}_q\}_{q\geq 1}$ which is constant, equal to some set $\mathcal{N}$ of cardinality at most $N(\varepsilon, K, \eta)$.

The conditional expectations entering (5) are polynomial and, hence, continuous in q. Therefore, $E_q[f(X)|X_i = 1] - E_q[f(X)|X_i = 0]$ converges to $E[f(X)|X_i = 1] - E[f(X)|X_i = 0]$ for all i as $q \to +\infty$. Applying this observation to $i \in \mathcal{N}$ for the converging subsequence proves the result.

References

ROBBINS, H. (1955) "A Remark on Stirling's Formula," *The American Mathematical Monthly*, Vol. 62, pp. 26-29.