Asymptotic enumeration of correlation-immune functions

E. Rodney Canfield

Jason Gao

Catherine Greenhill

Brendan D. McKay

Robert W. Robinson

Correlation-immune functions

Suppose we have a secret Boolean function of n Boolean variables.

Suppose a malicious eavesdropper is able to observe function values while monitoring any k of the variables.

We would like these observations to give the eavesdropper as little information as possible.

Correlation-immune functions

Suppose we have a secret Boolean function of n Boolean variables.

Suppose a malicious eavesdropper is able to observe function values while monitoring any k of the variables.

We would like these observations to give the eavesdropper as little information as possible.

The function is correlation-immune of order k if the function value is uncorrelated with any k of the arguments.

Suppose the fraction λ of all 2^n argument values give a function value 1. Then correlation-immune means that if any arbitrary k of the arguments are fixed to arbitrary values, the same fraction λ of the remaining 2^{n-k} argument values give a function value 1.

The weight of the function is $\lambda 2^n$ — the number of argument lists that give function value 1.

Example (Sloane): $n = 12$, $k = 3$, $\lambda = 24/2^{12}$, weight $= 24 = 2^k 3$.

The rows of the table give the argument lists for which the function value is 1.

0	1	1	1	1	1	1	1	1	1	1	1
0	0	1	0	1	1	1	0	0	0	1	0
0	0	0	1	0	1	1	1	0	0	0	1
0	1	0	0	1	0	1	1	1	0	0	0
0	0	1	0	0	1	0	1	1	1	0	0
0	0	0	1	0	0	1	0	1	1	1	0
0	0	0	0	1	0	0	1	0	1	1	1
0	1	0	0	0	1	0	0	1	0	1	1
0	1	1	0	0	0	1	0	0	1	0	1
0	1	1	1	0	0	0	1	0	0	1	0
0	0	1	1	1	0	0	0	1	0	0	1
0	1	0	1	1	1	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	0	1	1	1	0	1
1	1	1	0	1	0	0	0	1	1	1	0
1	0	1	1	0	1	0	0	0	1	1	1
1	1	0	1	1	0	1	0	0	0	1	1
1	1	1	0	1	1	0	1	0	0	0	1
1	1	1	1	0	1	1	0	1	0	0	0
1	0	1	1	1	0	1	1	0	1	0	0
1	0	0	1	1	1	0	1	1	0	1	0
1	0	0	0	1	1	1	0	1	1	0	1
1	1	0	0	0	1	1	1	0	1	1	0
1	0	1	0	0	0	1	1	1	0	1	1

This is an orthogonal array of 2 levels, 12 variables, 24 runs and strength 3.

Our task

Since the weight is a multiple of 2^k, let's define it to be $2^k q$, where $0 \leq q \leq 2^{n-k}$.

Our task

Since the weight is a multiple of 2^k, let's define it to be $2^k q$, where $0 \le q \le 2^{n-k}$.

Define $N(n, k, q)$ to be the number of n-variable correlation-immune functions of order k and weight $2^k q$.

Also define $N(n, k) = \sum_q N(n, k, q)$.

We seek the asymptotic values of $N(n, k, q)$ and $N(n, k)$ as $n \to \infty$, with k and q being some functions of n.

Our task

Since the weight is a multiple of 2^k, let's define it to be $2^k q$, where $0 \le q \le 2^{n-k}$.

Define $N(n, k, q)$ to be the number of n-variable correlation-immune functions of order k and weight $2^k q$.

Also define $N(n, k) = \sum_q N(n, k, q)$.

We seek the asymptotic values of $N(n, k, q)$ and $N(n, k)$ as $n \to \infty$, with k and q being some functions of n.

Define

$$M = \sum_{i=0}^{k} \binom{n}{i} \quad \text{and} \quad Q = \sum_{i=1}^{k} i \binom{n}{i}.$$

Theorem (Denisov, 1992)
If $k \ge 1$ is a constant integer, then as $n \to \infty$,

$$N(n, k) \sim 2^{2^n + Q - k} (2^{n-1} \pi)^{-(M-1)/2}. \quad \square$$

Denisov's method (translated)

For $S \subseteq \{1, 2, \ldots, n\}$, let β_S be the number of rows $(\beta_1, \beta_2, \ldots, \beta_n)$ of the matrix such that $\beta_i = 1$ for $i \in S$. Also $\beta_\emptyset = 2^k q$ (i.e., all the rows).

Then the matrix is that of a correlation-immune function of weight $2^k q$ iff

$$\beta_S = 2^{k-|S|} q \quad \text{for} \quad |S| \leq k.$$

Denisov's method (translated)

For $S \subseteq \{1, 2, \ldots, n\}$, let β_S be the number of rows $(\beta_1, \beta_2, \ldots, \beta_n)$ of the matrix such that $\beta_i = 1$ for $i \in S$. Also $\beta_\emptyset = 2^k q$ (i.e., all the rows).

Then the matrix is that of a correlation-immune function of weight $2^k q$ iff

$$\beta_S = 2^{k-|S|} q \quad \text{for} \quad |S| \leq k.$$

Consider

$$\prod_{\beta \in \{0,1\}^n} \left(1 + \prod_{|S| \leq k} x_S^{\prod_{i \in S} \beta_i}\right),$$

where $\{x_S \mid S \subseteq \{1, 2, \ldots, n\}\}$ are indeterminates.

Then $N(n, k, q)$ is the coefficient of the monomial

$$\prod_{|S| \leq k} x_S^{2^{k-|S|} q}.$$

Denisov extracts $N(n, k)$ by Fourier inversion.

The inversion integral is concentrated at two equivalent places, where it is approximately gaussian.

Expansion near the critical points together with bounds away from the critical points establishes the asymptotics.

The inversion integral is concentrated at two equivalent places, where it is approximately gaussian.

Expansion near the critical points together with bounds away from the critical points establishes the asymptotics.

Denisov's retraction

In 2000, Denisov published a retraction of his 1992 result.

He wrote that he had "made a mistake", and gave a new asymptotic value of $N(n, k)$.

The inversion integral is concentrated at two equivalent places, where it is approximately gaussian.

Expansion near the critical points together with bounds away from the critical points establishes the asymptotics.

Denisov's retraction

In 2000, Denisov published a retraction of his 1992 result.

He wrote that he had "made a mistake", and gave a new asymptotic value of $N(n, k)$.

This is unfortunate, since the 1992 result is correct and the 2000 result is incorrect!

Alternative approach

For a boolean function $g(x_1, \ldots, x_n)$, the Walsh transform of g is the real-valued function $\hat{g}$ over $\{0,1\}^n$ defined by

$$\hat{g}(w_1, \ldots, w_n) = \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n} g(x_1, \ldots, x_n)(-1)^{w_1x_1+\cdots+w_nx_n}.$$

It is known that g is correlation-immune of order k iff $\hat{g}(w_1, \ldots, w_n) = 0$ whenever the number of 1s in $w_1, \ldots, w_n$ is between 1 and k.

Alternative approach

For a boolean function $g(x_1, \ldots, x_n)$, the Walsh transform of g is the real-valued function $\hat{g}$ over $\{0,1\}^n$ defined by

$$\hat{g}(w_1, \ldots, w_n) = \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n} g(x_1, \ldots, x_n)(-1)^{w_1x_1+\cdots+w_nx_n}.$$

It is known that g is correlation-immune of order k iff $\hat{g}(w_1, \ldots, w_n) = 0$ whenever the number of 1s in $w_1, \ldots, w_n$ is between 1 and k.

Put $R = \lambda/(1-\lambda)$. Define

$$F(\boldsymbol{x}) = \prod_{\alpha\in\{\pm1\}^n} \Big(1 + R \prod_{|S|\le k} x_S^{\alpha_S}\Big),$$

where

$$\alpha_S = \prod_{i\in S} \alpha_i.$$

Theorem: $N(n,k,q)$ is the constant term of $(Rx_\emptyset)^{-2^kq}F(\boldsymbol{x})$.

Apply the Cauchy coefficient formula, using unit circles as contours, and change variables as $x_S = e^{i\theta_S}$ for each S.

Then

$$N(n,k,q) = \frac{(1+R)^{2^n}}{(2\pi)^M R^{2^k q}}\, I(n,k,q),$$

where

$$I(n,k,q) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} G(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

$$G(\boldsymbol{\theta}) = e^{-i2^k q \theta_\emptyset} \prod_{\alpha \in \{\pm 1\}^n} \frac{1 + R e^{i f_\alpha(\boldsymbol{\theta})}}{1+R},$$

$$f_\alpha(\boldsymbol{\theta}) = \sum_{|S| \le k} \alpha_S \theta_S.$$

Here $\boldsymbol{\theta}$ is a vector of the variables θ_S, $|S| \le k$, in arbitrary order.

Analysis of the domain of integration

The integrand

$$G(\boldsymbol{\theta}) = e^{-i2^k q\theta_\emptyset} \prod_{\alpha\in\{\pm1\}^n} \frac{1 + Re^{if_\alpha}}{1 + R}$$

has greatest absolute value 1 when

$$f_\alpha = f_\alpha(\boldsymbol{\theta}) = \sum_{|S|\le k} \alpha_S \theta_S$$

is a multiple of 2π for each S. When does that happen?

Analysis of the domain of integration

The integrand

$$G(\boldsymbol{\theta}) = e^{-i2^k q\theta_\emptyset} \prod_{\alpha\in\{\pm1\}^n} \frac{1 + Re^{if_\alpha}}{1+R}$$

has greatest absolute value 1 when

$$f_\alpha = f_\alpha(\boldsymbol{\theta}) = \sum_{|S|\le k} \alpha_S \theta_S$$

is a multiple of 2π for each S. When does that happen?

Define the difference operator

$$\delta_j f_{(\alpha_1,\ldots,\alpha_j,\ldots,\alpha_n)} = f_{(\alpha_1,\ldots,\alpha_j,\ldots,\alpha_n)} - f_{(\alpha_1,\ldots,-\alpha_j,\ldots,\alpha_n)}.$$

and in general $\delta_S = \prod_{j\in S} \delta_j$.

If each f_α is a multiple of 2π, then so are all the differences. Now we compute

$$\delta_S f_\alpha = 2^{|S|} \sum_{T\supseteq S} \alpha_T \theta_T.$$

and apply this with decreasing $|S|$.

Conclusion:

$|G(\boldsymbol{\theta})| = 1$ iff there are integers j_S such that

$$\sum_{T \supseteq S} \theta_T = 2^{-|S|+1} j_S \pi$$

for every $S \subseteq \{1, 2, \ldots, n\}$ with $|S| \leq k$.

There are 2^Q such critical points, where $Q = \sum_{i=1}^{k} i \binom{n}{i}$.

Conclusion:

$|G(\boldsymbol{\theta})| = 1$ iff there are integers j_S such that

$$\sum_{T \supseteq S} \theta_T = 2^{-|S|+1} j_S \pi$$

for every $S \subseteq \{1, 2, \ldots, n\}$ with $|S| \leq k$.

There are 2^Q such critical points, where $Q = \sum_{i=1}^{k} i \binom{n}{i}$.

Define the critical region $\mathcal{R}$ to be the set of points $\boldsymbol{\theta}$ such that, for some critical point $\hat{\boldsymbol{\theta}}$

$$|\theta_S - \hat{\theta}_S| \leq \Delta (2n)^{-|S|}$$

for each S, where $\Delta = 2^{-n/2+k+3} \lambda^{-1/2} n^{k+1/2} M^{1/2}$.

These 2^Q cuboids are disjoint and equivalent.

The integrand outside the critical region

If $\boldsymbol{\theta}$ is not in the critical region,

$$|G(\boldsymbol{\theta})| < \exp(-\tfrac{4}{5}nM).$$

The integrand outside the critical region

If $\boldsymbol{\theta}$ is not in the critical region,

$$|G(\boldsymbol{\theta})| < \exp(-\tfrac{4}{5}nM).$$

Proof:

(1) There is some S such that, outside the critical region, $\delta_S f_\alpha$ is at least $(2 - e^{1/2})\Delta n^{-|S|}$ from any multiple of 2π for all α.

(2) Divide the 2^n vectors α into $2^{n-|S|}$ classes of size $2^{|S|}$, where two vectors are in the same class iff they agree outside S.

(3) For each class,

$$\prod_\alpha \left| \frac{1 + Re^{if_\alpha}}{1 + R} \right| \leq \exp(-stuff).$$

(4) That does it.

The integrand inside the critical region

Since the 2^Q components of the critical region are all equivalent, consider the component containing the origin.

If $\boldsymbol{\theta}$ is in the critical region near the origin,

$$G(\boldsymbol{\theta}) = \exp\Big(-\tfrac{1}{2}\lambda(1-\lambda)2^n \sum_{|S|\le k} \theta_S^2 + O(\lambda 2^n \Delta^3)\Big).$$

The integrand inside the critical region

Since the 2^Q components of the critical region are all equivalent, consider the component containing the origin.

If $\boldsymbol{\theta}$ is in the critical region near the origin,

$$G(\boldsymbol{\theta}) = \exp\Big(-\tfrac{1}{2}\lambda(1-\lambda)2^n \sum_{|S|\le k} \theta_S^2 + O(\lambda 2^n \Delta^3)\Big).$$

Proof:

Use Taylor expansion.

The linear term vanishes thanks to the choice of R.

Conclusion

Theorem: If $\omega\big(2^{5k}n^{6k+3}M^3\big) \le q \le 2^{n-k} - \omega\big(2^{5k}n^{6k+3}M^3\big)$, then

$$N(n,k,q) \sim \frac{2^{Q-(n+1)M/2}}{\pi^{M/2}\big(\lambda^\lambda(1-\lambda)^{(1-\lambda)}\big)^{2^n+M/2}}.$$

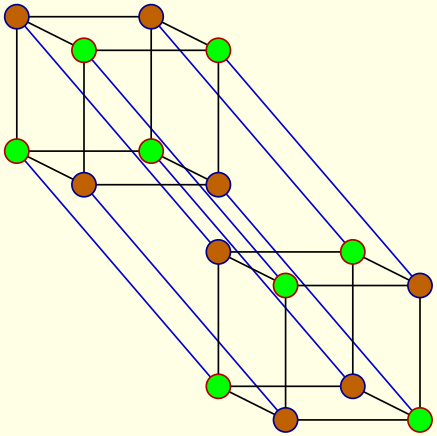
This allows *some* values of q if $k \le cn/\log n$ (compared to constant k for Denisov).

In that case:

$$N(n,k) \sim 2^{2^n+Q-k}(2^{n-1}\pi)^{-(M-1)/2}.$$

More on the case $k = 1$

A balanced colouring of a hypercube is a colouring with two colours such that the center of mass is at the center of the hypercube.



0 0 0 0
1 0 0 0
0 1 1 0
1 1 1 0
1 0 0 1
0 1 0 1
0 0 1 1
1 1 1 1

This corresponds to colouring according to the value of $f(x_1, x_2, \ldots, x_n)$ for a correlation-immune function of order 1.

Palmer, Read and Robinson did an exact enumeration (1992) but it seems unsuitable for asymptotics.

Naive estimate ($2q$ uses of first colour):

Choose, uniformly at random, a set of $2q$ distinct elements of $\{0,1\}^n$. The event of any particular column having exactly q 0s and q 1s has probability

$$\binom{2^{n-1}}{q}^2 \Big/ \binom{2^n}{2q}.$$

Therefore, if these n events are close to being independent,

$$N(n,1,q) \sim \binom{2^n}{2q} \left(\frac{\binom{2^{n-1}}{q}^2}{\binom{2^n}{2q}} \right)^n .$$

Naive estimate ($2q$ uses of first colour):

Choose, uniformly at random, a set of $2q$ distinct elements of $\{0,1\}^n$. The event of any particular column having exactly q 0s and q 1s has probability

$$\binom{2^{n-1}}{q}^2 \Big/ \binom{2^n}{2q}.$$

Therefore, if these n events are close to being independent,

$$N(n,1,q) \sim \binom{2^n}{2q} \left(\frac{\binom{2^{n-1}}{q}^2}{\binom{2^n}{2q}} \right)^n .$$

For small q, actually $q = o(2^{n/2})$, we can estimate $N(n,1,q)$ probabilistically: make each column randomly with q zeros and q ones. Then use Bonferroni to show that the rows are distinct with probability $1 - o(1)$.

Naive estimate ($2q$ uses of first colour):

Choose, uniformly at random, a set of $2q$ distinct elements of $\{0,1\}^n$. The event of any particular column having exactly q 0s and q 1s has probability

$$\binom{2^{n-1}}{q}^2 \Big/ \binom{2^n}{2q}.$$

Therefore, if these n events are close to being independent,

$$N(n,1,q) \sim \binom{2^n}{2q} \left(\frac{\binom{2^{n-1}}{q}^2}{\binom{2^n}{2q}} \right)^n .$$

For small q, actually $q = o(2^{n/2})$, we can estimate $N(n,1,q)$ probabilistically: make each column randomly with q zeros and q ones. Then use Bonferroni to show that the rows are distinct with probability $1 - o(1)$.

Larger q is covered by the analytic results.

Naive estimate ($2q$ uses of first colour):

Choose, uniformly at random, a set of $2q$ distinct elements of $\{0,1\}^n$. The event of any particular column having exactly q 0s and q 1s has probability

$$\binom{2^{n-1}}{q}^2 \Big/ \binom{2^n}{2q}.$$

Therefore, if these n events are close to being independent,

$$N(n,1,q) \sim \binom{2^n}{2q} \left(\frac{\binom{2^{n-1}}{q}^2}{\binom{2^n}{2q}} \right)^n .$$

For small q, actually $q = o(2^{n/2})$, we can estimate $N(n,1,q)$ probabilistically: make each column randomly with q zeros and q ones. Then use Bonferroni to show that the rows are distinct with probability $1 - o(1)$.

Larger q is covered by the analytic results.

Conclusion: The naive estimate is correct for all q.

Extensions

(1) Correlation-immune functions can be defined over sets other than $\{0, 1\}$. The asymptotic techniques can be generalized (but hasn't been, yet).

Extensions

(1) Correlation-immune functions can be defined over sets other than $\{0, 1\}$.
The asymptotic techniques can be generalized (but hasn't been, yet).

(2) A Hadamard matrix is an $n \times n$ matrix H over $\{-1, +1\}$ such that $H^T H = nI$.

```
-1  1  1  1  1 -1 -1 -1
 1 -1  1  1 -1  1 -1 -1
 1  1 -1  1 -1 -1  1 -1
 1  1  1 -1 -1 -1 -1  1
 1 -1 -1 -1 -1  1  1  1
-1  1 -1 -1  1 -1  1  1
-1 -1  1 -1  1  1 -1  1
-1 -1 -1  1  1  1  1 -1
```

Hadamard conjecture: A Hadamard matrix exists iff $n = 2$ or n is a multiple of 4.

Extensions

(1) Correlation-immune functions can be defined over sets other than $\{0, 1\}$.
The asymptotic techniques can be generalized (but hasn't been, yet).

(2) A Hadamard matrix is an $n \times n$ matrix H over $\{-1, +1\}$ such that $H^T H = nI$.

```
-1  1  1  1  1 -1 -1 -1
 1 -1  1  1 -1  1 -1 -1
 1  1 -1  1 -1 -1  1 -1
 1  1  1 -1 -1 -1 -1  1
 1 -1 -1 -1 -1  1  1  1
-1  1 -1 -1  1 -1  1  1
-1 -1  1 -1  1  1 -1  1
-1 -1 -1  1  1  1  1 -1
```

Hadamard conjecture: A Hadamard matrix exists iff $n = 2$ or n is a multiple of 4.

Multiply rows by -1 as needed so the first column is 1 then delete the first column. Then change -1 into 0. The result is a correlation-immune function of $n-1$ variables, order 2, and weight n.

Are there any??
De Launey and Levin used similar methods to show that at least $n^{1/12-\epsilon}$ rows of a Hadamard rectangle always exist.