

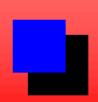


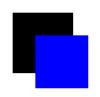
Probabilistic Methods in Combinatorics Lecture 12

Linyuan Lu

University of South Carolina





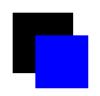


Poisson Paradigm



Let $\{B_i\}_{i\in I}$ be a set of bad events. We will estimate $\Pr(\wedge_{i\in I}\bar{B}_i)$.





Poisson Paradigm

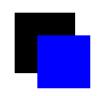


Let $\{B_i\}_{i\in I}$ be a set of bad events. We will estimate $\Pr(\wedge_{i\in I}\bar{B}_i)$.

If B_i 's are "mostly independent", then one may expect

$$\Pr(\wedge_{i\in I}\bar{B}_i)\approx\prod_{i\in I}\Pr(\bar{B}_i).$$





Poisson Paradigm



Let $\{B_i\}_{i\in I}$ be a set of bad events. We will estimate $\Pr(\wedge_{i\in I}\bar{B}_i)$.

If B_i 's are "mostly independent", then one may expect

$$\Pr(\wedge_{i\in I}\bar{B}_i)\approx\prod_{i\in I}\Pr(\bar{B}_i).$$

Let X_i be the random indicator of the event B_i and $X = \sum_{i \in I} X_i$. If $\Pr(B_i)$'s are small and "mostly independent", then one may expect X follows "Poisson-like distribution". In particular,



$$\Pr(X=0) \approx e^{-\mathrm{E}(X)}$$
.





lacksquare U: a finite universal set.







- \blacksquare U: a finite universal set.
- R: a random subset of U given by $\Pr(r \in R) = p_r$.







- lacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.







- \blacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.
- \bullet $i \sim j \text{ if } A_i \cap A_j \neq \emptyset.$







- \blacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.
- \bullet $i \sim j \text{ if } A_i \cap A_j \neq \emptyset.$
- \blacksquare B_i : the event $A_i \subset R$ for $i \in I$.





- \blacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.
- \bullet $i \sim j \text{ if } A_i \cap A_j \neq \emptyset.$
- lacksquare B_i : the event $A_i \subset R$ for $i \in I$.
- \blacksquare X_i : the indicator random variable for B_i .







- lacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.
- \bullet $i \sim j \text{ if } A_i \cap A_j \neq \emptyset.$
- lacksquare B_i : the event $A_i \subset R$ for $i \in I$.
- \blacksquare X_i : the indicator random variable for B_i .
- $X := \sum_{i \in I} X_i; \ \mu = E(X) = \sum_{i \in I} \Pr(B_i).$

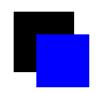






- \blacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.
- \bullet $i \sim j \text{ if } A_i \cap A_j \neq \emptyset.$
- \blacksquare B_i : the event $A_i \subset R$ for $i \in I$.
- \blacksquare X_i : the indicator random variable for B_i .
- $X := \sum_{i \in I} X_i$; $\mu = E(X) = \sum_{i \in I} \Pr(B_i)$.
- $\Delta = \sum_{i \sim j} \Pr(B_i \wedge B_j)$; sum over all ordered pairs $i \sim j$.







- lacksquare U: a finite universal set.
- \blacksquare R: a random subset of U given by $\Pr(r \in R) = p_r$.
- \blacksquare $\{A_i\}_{i\in I}$: a family of subsets of U.
- \bullet $i \sim j \text{ if } A_i \cap A_j \neq \emptyset.$
- \blacksquare B_i : the event $A_i \subset R$ for $i \in I$.
- \blacksquare X_i : the indicator random variable for B_i .
- $X := \sum_{i \in I} X_i; \ \mu = E(X) = \sum_{i \in I} \Pr(B_i).$
- $\Delta = \sum_{i \sim j} \Pr(B_i \wedge B_j)$; sum over all ordered pairs $i \sim j$.
- lacksquare $M = \prod_{i \in I} \Pr(\bar{B}_i).$





Janson inequality



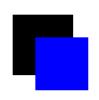
The Janson inequality: Assume all $Pr(B_i) \leq \epsilon$. Then

$$M \leq \Pr(\wedge_{i \in I} \bar{B}_i) \leq M e^{\frac{\Delta}{2(1-\epsilon)}},$$

and, further,

$$\Pr(\wedge_{i \in I} \bar{B}_i) \le e^{-\mu + \frac{\Delta}{2}}.$$





Janson inequality



The Janson inequality: Assume all $Pr(B_i) \leq \epsilon$. Then

$$M \le \Pr(\wedge_{i \in I} \bar{B}_i) \le M e^{\frac{\Delta}{2(1-\epsilon)}},$$

and, further,

$$\Pr(\wedge_{i\in I}\bar{B}_i) \le e^{-\mu + \frac{\Delta}{2}}.$$

The Extended Janson inequality: If further $\Delta \geq \mu$, then

$$\Pr(\wedge_{i\in I}\bar{B}_i) \le e^{\frac{-\mu^2}{2\Delta}}.$$







Proof given by Boppana and Spencer: We will use the following correlation inequality.

lacksquare For all $J \subset I$, $i \notin J$,

$$\Pr(B_i \mid \wedge_{j \in J} \bar{B}_j) \leq \Pr(B_i).$$

 \blacksquare For $J \subset I$, $i, k \notin J$,

$$\Pr(B_i \mid B_k \land \land_{j \in J} \bar{B}_j) \le \Pr(B_i \mid B_k).$$







Proof given by Boppana and Spencer: We will use the following correlation inequality.

lacksquare For all $J \subset I$, $i \not\in J$,

$$\Pr(B_i \mid \wedge_{j \in J} \bar{B}_j) \leq \Pr(B_i).$$

 \blacksquare For $J \subset I$, $i, k \notin J$,

$$\Pr(B_i \mid B_k \land \land_{j \in J} \bar{B}_j) \le \Pr(B_i \mid B_k).$$

Order the index set $I = \{1, 2, \dots, m\}$.

$$\Pr(\wedge_{i\in I}\bar{B}_i) = \prod_{i=1}^m \Pr(\bar{B}_i \mid \wedge_{1\leq j< i}\bar{B}_j) \geq \prod_{i=1}^m \Pr(\bar{B}_i).$$







For a given i renumber, for convenience, so that $i \sim j$ for $1 \leq j \leq d$ and not for $d+1 \leq j < i$. Let $A=B_i$, $B=\bar{B}_1 \wedge \cdots \wedge \bar{B}_d$, and $C=\bar{B}_{d+1} \wedge \cdots \wedge \bar{B}_{i-1}$,

$$\Pr(B_i \mid \wedge_{1 \leq j < i} \overline{B}_j) = \Pr(A \mid B \wedge C)$$

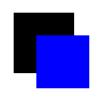
$$\leq \Pr(A \wedge B \mid C)$$

$$= \Pr(A \mid C) \Pr(B \mid A \wedge C).$$

Note $Pr(A \mid C) = Pr(A)$ and

$$\Pr(B \mid A \land C) \ge 1 - \sum_{j=1}^{d} \Pr(B_j \mid B_i \land C) \ge 1 - \sum_{j=1}^{d} \Pr(B_j \mid B_i).$$







For a given i renumber, for convenience, so that $i \sim j$ for $1 \leq j \leq d$ and not for $d+1 \leq j < i$. Let $A=B_i$, $B=\bar{B}_1 \wedge \cdots \wedge \bar{B}_d$, and $C=\bar{B}_{d+1} \wedge \cdots \wedge \bar{B}_{i-1}$,

$$Pr(B_i \mid \wedge_{1 \leq j < i} \bar{B}_j) = Pr(A \mid B \wedge C)$$

$$\leq Pr(A \wedge B \mid C)$$

$$= Pr(A \mid C)Pr(B \mid A \wedge C).$$

Note $Pr(A \mid C) = Pr(A)$ and

$$\Pr(B \mid A \land C) \ge 1 - \sum_{j=1}^{d} \Pr(B_j \mid B_i \land C) \ge 1 - \sum_{j=1}^{d} \Pr(B_j \mid B_i).$$







$$\Pr(\bar{B}_i \mid \wedge_{1 \leq j < i} \bar{B}_j) \leq \Pr(\bar{B}_i) + \sum_{j=1}^{a} \Pr(B_j \wedge B_i)$$

$$\leq \Pr(\bar{B}_i) \left(1 + \frac{1}{2} \sum_{j=1}^{d} \Pr(B_j \wedge B_j) \right)$$

$$\leq \Pr(\bar{B}_i) \left(1 + \frac{1}{1 - \epsilon} \sum_{j=1}^d \Pr(B_j \wedge B_i) \right)$$

$$\leq \Pr(\bar{B}_i)e^{\frac{1}{1-\epsilon}\sum_{j=1}^d\Pr(B_j\wedge B_i)}.$$







$$\Pr(\bar{B}_i \mid \wedge_{1 \leq j < i} \bar{B}_j) \leq \Pr(\bar{B}_i) + \sum_{j=1}^d \Pr(B_j \wedge B_i)$$

$$\leq \Pr(\bar{B}_i) \left(1 + \frac{1}{1 - \epsilon} \sum_{j=1}^d \Pr(B_j \wedge B_i) \right)$$

$$\leq \Pr(\bar{B}_i)e^{\frac{1}{1-\epsilon}\sum_{j=1}^d\Pr(B_j\wedge B_i)}.$$

Plug it into $\Pr(\wedge_{i\in I}\bar{B}_i)=\prod_{i=1}^m\Pr\left(\bar{B}_i\mid \wedge_{1\leq j< i}\bar{B}_j\right)$; we get the first inequality. The second inequality use the following estimation.

$$\Pr(\bar{B}_i \mid \wedge_{1 \leq j < i} \bar{B}_j) \leq \Pr(\bar{B}_i) + \sum_{j=1}^{d} \Pr(B_j \wedge B_i)$$





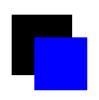
Proof of second Theorem



From the Jansen inequality, we have

$$-\ln(\Pr(\wedge_{i\in I}\bar{B}_i)) \ge \sum_{i\in I}\Pr(B_i) - \frac{1}{2}\sum_{i\sim j}\Pr(B_i \wedge B_j).$$





Proof of second Theorem



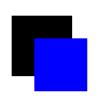
From the Jansen inequality, we have

$$-\ln(\Pr(\wedge_{i\in I}\bar{B}_i)) \ge \sum_{i\in I}\Pr(B_i) - \frac{1}{2}\sum_{i\sim j}\Pr(B_i \wedge B_j).$$

For any set $S \subset I$, the same inequality applied to $\{B_i\}_{i \in S}$:

$$-\ln(\Pr(\wedge_{i\in S}\bar{B}_i)) \ge \sum_{i\in S} \Pr(B_i) - \frac{1}{2} \sum_{i,j\in S, i\sim j} \Pr(B_i \wedge B_j).$$





Proof of second Theorem



From the Jansen inequality, we have

$$-\ln(\Pr(\wedge_{i\in I}\bar{B}_i)) \ge \sum_{i\in I}\Pr(B_i) - \frac{1}{2}\sum_{i\sim j}\Pr(B_i \wedge B_j).$$

For any set $S \subset I$, the same inequality applied to $\{B_i\}_{i \in S}$:

$$-\ln(\Pr(\wedge_{i\in S}\bar{B}_i)) \ge \sum_{i\in S} \Pr(B_i) - \frac{1}{2} \sum_{i,j\in S, i\sim j} \Pr(B_i \wedge B_j).$$

Now take S be a random subset of I given by $Pr(i \in S) = p$, and take the expectation.

$$E\left[-\ln(\Pr(\wedge_{i\in S}\bar{B}_i))\right] \ge p\mu - p^2\frac{\delta}{2}.$$







Now choose $p = \mu/\Delta$.

$$E\left[-\ln(\Pr(\wedge_{i\in S}\bar{B}_i))\right] \ge \frac{\mu^2}{2\Delta}.$$







Now choose $p = \mu/\Delta$.

$$E\left[-\ln(\Pr(\wedge_{i\in S}\bar{B}_i))\right] \ge \frac{\mu^2}{2\Delta}.$$

Then there is a specific $S \subset I$ for which

$$-\ln(\Pr(\wedge_{i\in S}\bar{B}_i)) \ge \frac{\mu^2}{2\Delta}.$$







Now choose $p = \mu/\Delta$.

$$E\left[-\ln(\Pr(\wedge_{i\in S}\bar{B}_i))\right] \ge \frac{\mu^2}{2\Delta}.$$

Then there is a specific $S \subset I$ for which

$$-\ln(\Pr(\wedge_{i\in S}\bar{B}_i)) \ge \frac{\mu^2}{2\Delta}.$$

$$\Pr(\wedge_{i \in S} \bar{B}_i) \le e^{-\frac{\mu^2}{2\Delta}}.$$







Now choose $p = \mu/\Delta$.

$$E\left[-\ln(\Pr(\wedge_{i\in S}\bar{B}_i))\right] \ge \frac{\mu^2}{2\Delta}.$$

Then there is a specific $S \subset I$ for which

$$-\ln(\Pr(\wedge_{i\in S}\bar{B}_i)) \ge \frac{\mu^2}{2\Delta}.$$

$$\Pr(\wedge_{i \in S} \bar{B}_i) \le e^{-\frac{\mu^2}{2\Delta}}.$$

$$\Pr(\wedge_{i \in I} \bar{B}_i) \le \Pr(\wedge_{i \in S} \bar{B}_i) \le e^{-\frac{\mu^2}{2\Delta}}.$$





Brun's sieve



- lacksquare X_i : the indicator random variable for B_i , for $i \in I$.
- $lacksquare X := \sum_{i=1}^m X_i$.
- \blacksquare m=m(n), $B_i=B_i(n)$, and X=X(n).
- Let

$$S^{(r)} = \sum \Pr(B_{i_1} \wedge \cdots \wedge B_{i_r}),$$

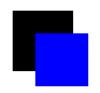
where the sum is over all sets

$$\{i_1,\ldots,i_r\}\subset\{1,2\ldots,m\}.$$

Let

$$X^{(r)} = X(X-1)\cdots(X-r+1).$$





Brun's sieve



- $lacksquare X_i$: the indicator random variable for B_i , for $i \in I$.
- $lacksquare X := \sum_{i=1}^m X_i$.
- \blacksquare m=m(n), $B_i=B_i(n)$, and X=X(n).
- Let

$$S^{(r)} = \sum \Pr(B_{i_1} \wedge \cdots \wedge B_{i_r}),$$

where the sum is over all sets

$$\{i_1,\ldots,i_r\}\subset\{1,2\ldots,m\}.$$

Let

$$X^{(r)} = X(X-1)\cdots(X-r+1).$$

By inclusion-exclusion principle,

$$\Pr(X = 0) = \Pr(\bar{B}_1 \wedge \dots \wedge \bar{B}_m) = \sum_{r>0} (-1)^r S^{(r)}.$$





Brun's sieve



Theorem: Suppose there is a constant μ so that for every fixed r,

$$E\binom{X}{r} = S^{(r)} \to \frac{\mu^r}{r!}.$$

Then

$$\Pr(X=0) \to e^{-\mu},$$

and for every t

$$\Pr(X=t) \to \frac{\mu^t}{t!} e^{-\mu}.$$





Proof: We only prove the case t = 0. Fix $\epsilon > 0$. Choose s so that

$$\left| \sum_{r=0}^{2s} (-1)^r \frac{\mu^r}{r!} - e^{-\mu} \right| \le \frac{\epsilon}{2}.$$

Select n_0 so that for $n \geq n_0$,

$$|S^{(r)} - \frac{\mu^r}{r!}| \le \frac{\epsilon}{2s(2s+1)}$$

for $0 \le r \le 2s$.





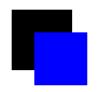
For such n,

$$\Pr[X = 0] \leq \sum_{r=0}^{2s} (-1)^r S^{(r)}$$

$$\leq \sum_{r=0}^{2s} (-1)^r \frac{\mu^r}{r!} + \frac{\epsilon}{2}$$

$$\leq e^{-\mu} + \epsilon.$$







For such n,

$$\Pr[X = 0] \leq \sum_{r=0}^{2s} (-1)^r S^{(r)}$$

$$\leq \sum_{r=0}^{2s} (-1)^r \frac{\mu^r}{r!} + \frac{\epsilon}{2}$$

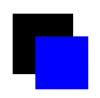
$$\leq e^{-\mu} + \epsilon.$$

Similarly, taking the sum to 2s+1, we can find n_0 so that for $n \ge n_0$,

$$\Pr[X=0] \ge e^{-\mu} - \epsilon.$$



As ϵ was arbitrary $\Pr(X=0) \to e^{-\mu}$.



An application



Let G = G(n, p), and EPIT represent the statement that every vertex lies in a triangle.

Theorem (a special case of Spencer's Theorem): Let c>0 be fixed and let p=p(n), $\mu=\mu(n)$ satisfy

$$\binom{n-1}{2}p^3 = \mu,$$

$$e^{-\mu} = \frac{c}{n}.$$

Then

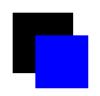
$$\lim_{n\to\infty} Pr(G(n,p) \text{ satisfies } EPIT) = e^{-c}.$$





First fix $x \in V(G)$. For each unordered $y, z \neq x$ let B_{xyz} be the event that $\{x, y, z\}$ is a triangle of G. Let C_x be the event $\wedge_{y,z} \bar{B}_{xyz}$ and X_x the corresponding indicator random variable. Apply Janson's Inequality to bound $E(X_x) = \Pr(\wedge_{y,z} \bar{B}_{xyz})$.







First fix $x \in V(G)$. For each unordered $y, z \neq x$ let B_{xyz} be the event that $\{x, y, z\}$ is a triangle of G. Let C_x be the event $\wedge_{y,z} \bar{B}_{xyz}$ and X_x the corresponding indicator random variable. Apply Janson's Inequality to bound $E(X_x) = \Pr(\wedge_{y,z} \bar{B}_{xyz})$.

$$\Delta = \sum_{y,z,z'} \Pr(B_{xyz} \wedge B_{xyz'}) = O(n^3 p^5) = o(1)$$

since $p = n^{-2/3 + o(1)}$.





First fix $x \in V(G)$. For each unordered $y, z \neq x$ let B_{xyz} be the event that $\{x, y, z\}$ is a triangle of G. Let C_x be the event $\wedge_{y,z} \bar{B}_{xyz}$ and X_x the corresponding indicator random variable. Apply Janson's Inequality to bound $E(X_x) = \Pr(\wedge_{y,z} \bar{B}_{xyz})$.

$$\Delta = \sum_{y,z,z'} \Pr(B_{xyz} \wedge B_{xyz'}) = O(n^3 p^5) = o(1)$$

since $p = n^{-2/3 + o(1)}$. Thus

$$E(X_x) \approx e^{-\mu} = \frac{c}{n}.$$





continue



Let $X = \sum_{x} X_x$, which is the number of vertices x no lying a triangle.

$$E(X) = \sum_{x} E(X_x) \to c.$$

We need to show that the Poisson Paradigm applies to X.





continue



Let $X = \sum_{x} X_x$, which is the number of vertices x no lying a triangle.

$$E(X) = \sum_{x} E(X_x) \to c.$$

We need to show that the Poisson Paradigm applies to X. Fix r and consider

$$\mathrm{E}\binom{X}{r} = S^{(r)} = \sum \mathrm{Pr}(C_{x_1} \wedge \cdots \wedge C_{x_r}),$$

where the sum is over all sets $\{x_1, \ldots, x_r\}$.





continue



Let $X = \sum_{x} X_x$, which is the number of vertices x no lying a triangle.

$$E(X) = \sum_{x} E(X_x) \to c.$$

We need to show that the Poisson Paradigm applies to X. Fix r and consider

$$\mathrm{E}\binom{X}{r} = S^{(r)} = \sum \mathrm{Pr}(C_{x_1} \wedge \cdots \wedge C_{x_r}),$$

where the sum is over all sets $\{x_1, \ldots, x_r\}$. Note

$$C_{x_1} \wedge \cdots \wedge C_{x_r} = \wedge_{1 \leq i \leq r, y, z} \overline{B_{x_i y z}}.$$







We apply Janson's Inequality again.

$$\sum \Pr(B_{x_i yz}) = p^3 \left(r \binom{n-1}{2} - O(n) \right) = r\mu + O(n^{-1+o(1)}).$$

As before Δ is p^5 times the number of pairs $x_iyz \sim x_jyz$; $\Delta = O(n^3p^5) = o(1)$.







We apply Janson's Inequality again.

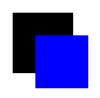
$$\sum \Pr(B_{x_i y z}) = p^3 \left(r \binom{n-1}{2} - O(n) \right) = r \mu + O(n^{-1+o(1)}).$$

As before Δ is p^5 times the number of pairs $x_iyz \sim x_jyz$; $\Delta = O(n^3p^5) = o(1)$.

$$\Pr(C_{x_1} \wedge \cdots \wedge C_{x_r}) \sim e^{-r\mu}$$

$$\operatorname{E}\binom{X}{r} \approx \binom{n}{r} e^{-r\mu} \approx \frac{c^r}{r!}$$







We apply Janson's Inequality again.

$$\sum \Pr(B_{x_i yz}) = p^3 \left(r \binom{n-1}{2} - O(n) \right) = r\mu + O(n^{-1+o(1)}).$$

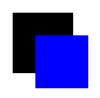
As before Δ is p^5 times the number of pairs $x_iyz \sim x_jyz$; $\Delta = O(n^3p^5) = o(1)$.

$$\Pr(C_{x_1} \wedge \cdots \wedge C_{x_r}) \sim e^{-r\mu}$$

$$\operatorname{E}\binom{X}{r} \approx \binom{n}{r} e^{-r\mu} \approx \frac{c^r}{r!}$$

Applying Brun's Sieve method, we have $\Pr(X=0) \to e^{-c}$.





Generalization



A sufficient condition for Janson's Inequality:

- I: a dependency digraph; if for each $i \in I$ the event B_i is mutually independent of $\{B_j : i \not\sim j\}$.
- $\Delta := \sum_{i \sim j} \Pr(B_i \wedge B_j).$
- \blacksquare For all $J \subset I$, $i \notin J$,

$$\Pr(B_i \mid \wedge_{j \in J} \bar{B}_j) \leq \Pr(B_i).$$

 \blacksquare For $J \subset I$, $i, k \notin J$,

$$\Pr(B_i \mid B_k \land \land_{j \in J} \bar{B}_j) \leq \Pr(B_i \mid B_k).$$

Then Janson's inequality holds.



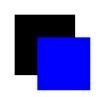
Suen's theorem



An binary relation \sim on I is **superdenpendency digraph** if the following holds:

Suppose that $J_1, J_2 \subset I$ are disjoint subsets so that there is no edge between J_1 and J_2 . Let B^1 be any Boolean combination of the events $\{B_j\}_{j\in J_1}$ and B^2 be any Boolean combination of the events $\{B_j\}_{j\in J_2}$. Then B^1 and B^2 are independent.





Suen's theorem



An binary relation \sim on I is **superdenpendency digraph** if the following holds:

Suppose that $J_1, J_2 \subset I$ are disjoint subsets so that there is no edge between J_1 and J_2 . Let B^1 be any Boolean combination of the events $\{B_j\}_{j\in J_1}$ and B^2 be any Boolean combination of the events $\{B_j\}_{j\in J_2}$. Then B^1 and B^2 are independent.

Theorem [Suen]: Under the above conditions,

$$\left| \Pr(\wedge_{i \in I} \bar{B}_i) - M \right| \le M(e^{\sum_{i \sim j} y(i,j)} - 1),$$

where

$$y_{i,j} = (\operatorname{Pr}(B_i \wedge B_j) + \operatorname{Pr}(B_i)\operatorname{Pr}(B_j)) \prod_{l \sim i \text{ or } l \sim j} (1 - \operatorname{Pr}(B_l))^{-1}.$$

