Course Notes

Part IV

Probabilistic Combinatorics

and

Algorithms

J. A. Verstraete
Department of Mathematics
University of California San Diego
9500 Gilman Drive
La Jolla California 92037-0112

jacques@ucsd.edu
1 Correlation Inequalities

Events $A_1, A_2, \ldots, A_n$ are positively correlated if for any set $I \subset [n]$,

$$P(\bigcap_{i \in I} A_i) \geq \prod_{i \in I} P(A_i).$$

The events are negatively correlated if the inequality is reversed for every set $I \subset [n]$. There are many problems in which positive and negative correlation arise in a very simple way. Our main concern is with events in product probability spaces called upsets and downsets. The basic framework is as follows. Let $Q_N$ denote the $N$-dimensional cube – equivalently the graph whose vertices are the binary strings of length $N$ and where two binary strings are adjacent if they differ in one position. It is sometimes convenient to consider the vertices of $Q_N$ as subsets of $[N]$ by considering the binary strings as characteristic function of these subsets. A set $A \subset Q_N$ is a downset if $x \in A$ and $y \subset x$ implies $y \in A$, and an upset if $x \in A$ and $y \supset x$ implies $y \in A$. An example of a downset is the set of graphs on $n$ vertices which do not contain a complete graph of order $k$. For we may consider this set as a subset of $Q_N$ where $N = \binom{n}{2}$. In fact, any property of graphs on $n$ vertices closed under taking subgraphs (called a monotone property) is a downset in $Q_N$. On the other hand, the set of graphs which contain neither a complete graph on $k$ vertices nor an independent set of $k$ vertices is neither a downset nor an upset. A natural probability measure on $Q_N$ is, for $x \in Q_N$,

$$P(x) = \prod_{i \in x} p_i \cdot \prod_{j \notin x} (1 - p_j)$$

where $p_1, p_2, \ldots, p_N \in [0, 1]$ are prescribed real numbers. In other words, we regard $Q_N$ as the product space $\prod_{i=1}^N \Omega_i$, where $\Omega_i = \{0, 1\}$ and the marginal distribution on $\Omega_i$ is a Bernoulli random variable with probability $p_i$. Henceforth, with slight abuse of notation, we just refer to the probability space above as the probability space $Q_N$. As mentioned above, if $N = \binom{n}{2}$ and $p_i = p$ for $i \in [n]$ we obtain the Erdős-Rényi space of random graphs $G_{n,p}$.

1.1 Harris Kleitman inequality

Here is a fundamental inequality concerning downsets and upsets in $Q_N$: this inequality states that if $A_1, A_2, \ldots, A_n$ are all upsets, or all downsets, then they are positively correlated. We give the proof for $n = 2$, and leave general $n$ as an exercise. Precisely, Harris (1960) and Kleitman (1966) proved the following inequality:

**Theorem 1** Let $A, B$ be upsets and $C, D$ be downsets in the probability space $Q_N$. Then $A$ and $B$ are positively correlated, $C$ and $D$ are positively correlated, whereas $A$ and $C$, and $B$ and $D$, are negatively correlated.

**Proof** It is sufficient to prove that $A$ and $B$ are positively correlated – we proceed by induction on $N$. The statement is clear if $N = 1$. Suppose $N > 1$, and define

$$A_0 = \{ x \cap \{1, 2, \ldots, N - 1\} : x \in A, N \notin x \}$$
$$A_1 = \{ x \cap \{1, 2, \ldots, N - 1\} : x \in A : N \in x \}.$$
Similar definitions are made for $B_0$ and $B_1$. It follows that $A_0 \subset A_1$ and $B_0 \subset B_1$, since $A$ and $B$ are upsets. Therefore

$$ (\mathbb{P}(A_1) - \mathbb{P}(A_0))(\mathbb{P}(B_1) - \mathbb{P}(B_0)) \geq 0. $$

Multiplying this out gives

$$ \mathbb{P}(A_0)\mathbb{P}(B_0) + \mathbb{P}(A_1)\mathbb{P}(B_1) \geq \mathbb{P}(A_0)\mathbb{P}(B_1) + \mathbb{P}(A_1)\mathbb{P}(B_0). \quad (1) $$

Finally, with $p_N = p$ and $q = 1 - p$,

$$ \mathbb{P}(A \cap B) = q\mathbb{P}(A_0 \cap B_0) + p\mathbb{P}(A_1 \cap B_1) $$

$$ \geq q\mathbb{P}(A_0)\mathbb{P}(B_0) + p\mathbb{P}(A_1)\mathbb{P}(B_1) \quad \text{by induction} $$

$$ = \{q\mathbb{P}(A_0)\mathbb{P}(B_0) + p\mathbb{P}(A_1)\mathbb{P}(B_1)\} \cdot (q + p) $$

$$ = q^2\mathbb{P}(A_0)\mathbb{P}(B_0) + p^2\mathbb{P}(A_1)\mathbb{P}(B_1) + pq \cdot \{\mathbb{P}(A_0)\mathbb{P}(B_0) + \mathbb{P}(A_1)\mathbb{P}(B_1)\} $$

$$ \geq \{q\mathbb{P}(A_0) + p\mathbb{P}(A_1)\} \cdot \{q\mathbb{P}(B_0) + p\mathbb{P}(B_1)\} \quad \text{by (1)} $$

$$ = \mathbb{P}(A)\mathbb{P}(B). $$

Some interesting combinatorial consequences of the Harris-Kleitman inequality concerning families of finite sets may be found in Bollobás’ Combinatorics text.

### 1.2 The Four Functions Theorem and FKG inequality

The Four Functions Theorem of Ahlswede and Daykin (1978) is one of the most general results in combinatorics. It generalizes the Harris-Kleitman inequality, and its proof is similar to the proof of that inequality. For a lattice $L$ and sets $A, B \subset L$, we write $A \vee B = \{a \vee b : a \in A, b \in B\}$ and define $A \wedge B$ similarly.

**Theorem 2** Let $\alpha, \beta, \delta, \gamma$ be four non-negative real valued functions on a distributive lattice $L$, and suppose that

$$ \alpha(a)\alpha(b) \leq \gamma(a \vee b)\delta(a \wedge b) $$

for any elements $a, b \in L$. Then

$$ \alpha(A)\beta(B) \leq \gamma(A \vee B)\delta(A \wedge B) $$

for any sets $A, B \subset L$.

One of the famous inequalities from probability and statistical physics is the FKG inequality. The inequality was proved by Fortuin, Kastelyn and Ginibre (1971), and has had a major impact on random and statistical physical modelling. The general form of the FKG inequality is stated for distributive lattices—a lattice $L$ is distributive if $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ for all $x, y, z \in L$. For functions $\mu, f : L \to \mathbb{R}$, we define $\mu(f) = \sum_{x \in L} \mu(x)f(x)$ (to be though of as an expectation). A function $f$ is monotone increasing if $x < y$ in the lattice $L$ implies $f(x) \leq f(y)$. Denote by 1 the function $f(x) = 1$ for all $x \in L$. We are now ready to state the FKG inequality, which is a straightforward consequence of the Four Functions Theorem.
**Theorem 3** Let $L$ be a distributive lattice, and let $\mu$ be a function $L \rightarrow \mathbb{R}^+$ such that
\[
\mu(x \lor y)\mu(x \land y) \geq \mu(x)\mu(y) \text{ for all } x, y \in L.
\]
Then $\mu(fg)\mu(1) \geq \mu(f)\mu(g)$ for all increasing functions $f, g : L \rightarrow \mathbb{R}^+$. 

Note that in the case that $\mu$ is a probability measure and $f = \chi_A$ and $g = \chi_B$ for upsets $A$ and $B$, we recover the Harris-Kleitman inequality $\mu(A \cap B) \geq \mu(A)\mu(B)$ but for general distributive lattices. It is clear that the FKG inequality is reversed when one of $f$ and $g$ is increasing and the other is decreasing.

## 2 Percolation in the plane

Percolation is the study of the properties of random subgraphs of lattices. In this section, we consider only the integer lattice $\mathbb{Z}^d$, where edges are independently selected with probability $p$ – in percolation these edges are called open edges. Formally, we have a probability triple $(\Omega, \mathcal{F}, P)$ where $\Omega = 2^E(L)$, $\mathcal{F}$ the $\sigma$-field generated by finite-dimensional cylinders in $\Omega$, and the product measure $P$, with expectation operator $E_p$. We write $\omega$ for a generic point in $\Omega$, and call these points configurations. A cluster is a collection of open edges which span a connected graph. The main quantity of interest is $\theta(p)$, the probability that the origin lies in an infinite cluster of open edges. For example, if instead of $\mathbb{Z}^d$ we consider percolation on an infinite tree $T$, then $1 - \theta(p)$ is just the probability of extinction in the standard Galton-Watson branching process on $T$, and in this case $\theta(p)$ can often be determined. Since the lattice $\mathbb{Z}^d$ is locally finite, the existence of an infinite cluster necessarily means that there is an infinite open path in the lattice. It is not hard to see that the function $\theta(p)$ is non-negative, increasing in $p$, and continuous on its support, and $\theta$ is identically zero on some interval of the form $[0, t)$. In view of this, the critical probability of $L$ is defined to be
\[
p_c(L) = \sup\{p : \theta(p) = 0\}.
\]

For example, it is clear that $p_c(\mathbb{Z}) = 1$. We shall determine the value of $p_c(\mathbb{Z}^2)$ – this is Kesten’s Theorem – while the value of $p_c(\mathbb{Z}^d)$ is unknown even for $d = 3$.

To determine $p_c(\mathbb{Z}^2)$, we make use of the following result, showing that the number of infinite components (or clusters) in $\mathbb{Z}^2$ is zero or one almost surely.

**Theorem 4** For any $p \in [0, 1]$, $P(I = 0) = 1$ or $P(I = 1) = 1$, where $I$ denotes the number of infinite open clusters in $\mathbb{Z}^2$.

**Proof** It is not hard to prove that since $I$ is translation invariant, $I$ is almost surely a constant\(^1\)

In other words, for some positive integer $k$ or $k = \infty$,
\[
P[I = k] = 1.
\]

Suppose first, for a contradiction, that $k$ is finite and at least two. Let $A_n$ be the event that the box $B(n)$ intersects at least two infinite clusters. Since $P[A_n] \rightarrow 1$ as $n \rightarrow \infty$, we can find an $N$ such that $P[A_N] > 1 - P[B_N]$, where $B_N$ is the event of an open path in $B(N)$ joining the two

\(^1\)This is a simple consequence of the ergodic theorem, but can be proved directly. We leave it as an exercise.
infinite clusters. Note that this event has constant positive probability – depending only on \( p \) and the distance between two of the infinite clusters (which is constant). This means that with positive probability, \( A_N \) and \( B_N \) occur. But

\[
P[I < k] \geq P[A_N \cap B_N] > 0.
\]

This contradicts \( P[I = k] = 1 \). So the number of infinite clusters is almost surely zero, one, or infinite. We refer the reader to Grimmett for the interesting proof that \( I \) is almost surely not infinite.

We now determine \( p_c(\mathbb{Z}^2) \).

**Theorem 5** \( p_c(\mathbb{Z}^2) = 1/2 \).

**Proof** By Theorem 4, if there exists an infinite open cluster then it is unique with probability one. First we show \( p_c(\mathbb{Z}^2) \geq \frac{1}{2} \). In fact, we will show more: that \( \theta(\frac{1}{2}) = 0 \). Set \( p = \frac{1}{2} \) and let \( B(n)^{+} \) denote the non-negative box \([0, n] \times [0, n] \). Let \( A \) be the event that there is an infinite open path crossing the boundary of \( B(n)^{+} \), and let \( A^{*} \) be the event that there is a closed infinite path crossing the boundary of \( B(n)^{+} \) in the dual of \( \mathbb{Z}^2 \). Label the sides of \( B(n)^{+} \) with elements of \( \{1, 2, 3, 4\} \) in increasing order clockwise from the top side. Let \( A_i \) be the event that the \( i \)th side of \( B(n)^{+} \) is crossed by an infinite open path in \( \mathbb{Z}^2 \setminus \text{int}(B(n)^{+}) \) and let \( A_i^{*} \) be the event that the \( i \)th side of \( B(n)^{+} \) in the dual of \( \mathbb{Z}^2 \setminus \text{int}(B(n)^{+}) \) is crossed by an infinite closed path. Then

\[
P[A_1 \cap A_3 \cap A_2^{*} \cap A_4^{*}] = 0
\]

since the event \( A_1 \cap A_3 \cap A_2^{*} \cap A_4^{*} \) implies that there are at least two infinite clusters. Now we derive a contradiction by showing that, when \( p = \frac{1}{2} \), the above probability is positive. Since the \( A_i \) are upsets and the \( A_i^{*} \) are upsets, the \( A_i \) are positively correlated and the \( A_i^{*} \) are positively correlated, by the FKG inequality. Moreover, they all have the same probability. Similar statements are true for the events \( A_i^{*} \). It follows that for all \( i \in \{1, 2, 3, 4\} \),

\[
P[A] = P[A_1 \cap A_3 \cap A_2^{*} \cap A_4^{*}] \geq P[A_i]^{4}
\]

Suppose, for the contradiction, that there is an infinite cluster almost surely. Then \( P[A] \to 1 \) as \( n \) tends to infinity. Considering percolation on the closed edges in the dual of \( \mathbb{Z}^2 \), we also have \( p = \frac{1}{2} \) and so \( P[A^*] \to 1 \). Using the inequalities above, we deduce that \( P[A_i] \to 1 \) and \( P[A_i^{*}] \to 1 \) for all \( i \in \{1, 2, 3, 4\} \). In particular,

\[
P[A_1 \cap A_3 \cap A_2^{*} \cap A_4^{*}] \to 1.
\]

This contradiction shows \( \theta(\frac{1}{2}) = 0 \) and \( p_c(\mathbb{Z}^2) \geq \frac{1}{2} \). Now we show that \( \theta(p) > 0 \) for all \( p > \frac{1}{2} \). Suppose, for a contradiction, that \( p_c(\mathbb{Z}^2) > \frac{1}{2} \). Let \( R(n) \) denote the rectangle \([0, n + 1] \times [0, n]\), minus the edges which are on the left and right sides of \( R(n) \). Then at \( p = \frac{1}{2} < p_c \), the probability of the event \( L \) that the origin is joined to the right side of \( R(n) \) is exponentially small as a function of \( n \). This means that the chance that the left side of \( R(n) \) is joined to the right tends to zero. On the other hand, \( P[L] = \frac{1}{2} \), since \( \overline{L} \) is the event that the top side of \( R(n) \) is joined to the bottom by a closed path in the dual of \( \mathbb{Z}^2 \), and therefore \( P[\overline{A}] = P[A] \). This completes the proof. □
3 Janson’s inequality

The main result we present is Janson’s inequality. Let $X \subset Q_N$, where $Q_N$ has the product measure given in the last section. For $x \in X$, define the event $A_x$ that $x$ is chosen, so the probability of $A_x$ is just $\mathbb{P}(x)$. Define the following measure of dependence for the events $A_x : x \in X$:

$$\triangle = \sum_{x \cap y \neq \emptyset} \mathbb{P}(A_x \cap A_y),$$

where the sum is over ordered pairs of distinct elements $x, y \in X$ which intersect. Janson’s inequality gives a good estimate on the probability that none of the events $A_x : x \in X$ occur.

**Theorem 6 (Janson’s inequality)** Let $X$ be a subset of $Q_N$, and suppose that each $A_x$ has probability at most $\alpha < 1$ and $\mu$ is the expected number of $A_x$ which occur. Then

$$\prod_{x \in X} \mathbb{P}(\overline{A}_x) \leq \mathbb{P}(\bigcap_{x \in X} \overline{A}_x) \leq e^{\alpha \triangle} \cdot \prod_{x \in X} \mathbb{P}(\overline{A}_x).$$

Furthermore,

$$\mathbb{P}(\bigcap_{x \in X} \overline{A}_x) \leq e^{-\mu + \frac{\triangle}{2}}.$$

**Proof** The Harris-Kleitman inequality gives the lower bound. For the upper bound, we’re going to find a lower estimate $\mathbb{P}(A_x \cap \bigcap_{y < x} \overline{A}_y)$ for each $x \in X$, and then use the multiplication principle to obtain the upper bound given by Janson’s inequality. Let $x$ be a fixed element of $X$ and let $<$ be any ordering of the elements of $X$. Define

$$Y = \{ y < x : x \cap y \neq \emptyset \} \cap X \quad \text{and} \quad Z = \{ y < x : x \cap y = \emptyset \} \cap X$$

$$B = \bigcap_{y \in Y} \overline{A}_y \quad \text{and} \quad C = \bigcap_{z \in Z} \overline{A}_z.$$

Since $A_x$ and $C$ are independent, we have $\mathbb{P}(A_x \mid C) = \mathbb{P}(A_x)$. Therefore

$$\mathbb{P}(A_x \mid B \cap C) = \frac{\mathbb{P}(A_x \cap B \cap C)}{\mathbb{P}(B \cap C)} \geq \frac{\mathbb{P}(A_x \cap B \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(A_x \cap B \cap C)}{\mathbb{P}(A_x \cap C)} \mathbb{P}(A_x \cap C) = \mathbb{P}(A_x) \cdot \mathbb{P}(B \mid A_x \cap C)$$

by independence of $A$ and $C$. Now we use the important fact that for $y \in Y$, $A_x \cap A_y$ and $A_x \cap C$ are negatively correlated in the cube of subsets of $[n] \setminus x$, by the Harris-Kleitman inequality, because one is an upset and the other is a downset. Therefore

$$\mathbb{P}(A_x \mid \bigcap_{y < x} \overline{A}_y) = \mathbb{P}(A_x \mid B \cap C) \geq \mathbb{P}(A_x) \cdot \mathbb{P}(B \mid A_x \cap C) = \mathbb{P}(A_x) \{ 1 - \mathbb{P}(\overline{B} \mid A_x \cap C) \} \geq \mathbb{P}(A_x) \{ 1 - \sum_{y \in Y} \mathbb{P}(A_y \mid A_x \cap C) \} = \mathbb{P}(A_x) \{ 1 - \sum_{y \in Y} \mathbb{P}(A_x \cap A_y \mid A_x \cap C) \} \geq \mathbb{P}(A_x) - \sum_{y \in Y} \mathbb{P}(A_x \cap A_y).$$

5
So we have determined, as we wanted, a lower estimate for \( P(A_x | \bigcap_{y < x} A_y) \) for each \( x \in X \). It follows that

\[
P(\bigcap_{x \in X} \overline{A}_x) = \prod_{x \in X} P(\overline{A}_x | \bigcap_{y < x} A_y)
\]
\[
\leq \prod_{x \in X} P(\overline{A}_x) \cdot \prod_{x \in X} \left\{ 1 + \frac{1}{1 - P(A_x)} \sum_{y \in Y} P(A_x \cap A_y) \right\}
\]
\[
\leq \prod_{x \in X} P(\overline{A}_x) \cdot \prod_{x \in X} \left\{ 1 + \frac{1}{1 - \alpha} \sum_{y \in Y} P(A_x \cap A_y) \right\}
\]
\[
\leq \prod_{x \in X} P(\overline{A}_x) \cdot \exp \left( \frac{\sum_{x \in X} \frac{1}{1 - \alpha} \sum_{y \in Y} P(A_x \cap A_y)}{2(1 - \alpha)} \right).
\]

This completes the proof of the first part. The second inequality comes from

\[
P(\bigcap_{x \in X} \overline{A}_x) \leq \prod_{x \in X} (1 - P(A_x) + \sum_{y \in Y} P(A_x \cap A_y)) \leq \exp(-\mu + \frac{\Delta}{2})
\]

using the lower estimate for \( P(A_x | \bigcap_{y < x} A_y) \).

We can express Janson’s inequality in terms of \( \mu \), the expected number of \( A_x : x \in X \) which occur: the upper bound in Janson’s inequality is at most \( \exp(-\mu + \frac{\Delta}{2}) \) and the lower bound is generally close to \exp(-\mu). Thus Janson’s inequality can be expected to give a tight bound for the probability that none of the \( A_x : x \in X \) occur provided \( \Delta \) is much smaller than \( \mu \): in this case the probability is close to \exp(-\mu). The following corollary goes a little further:

**Corollary 7** Let \( X \subseteq Q_N \), and suppose that each \( A_x \) has probability at most \( \alpha < 1 \). Then

\[
P(\bigcap_{x \in X} \overline{A}_x) \leq e^{-\frac{\mu^2}{2\alpha}}.
\]

where \( \mu \) is the expected number of \( A_x : x \in X \) which occur.

**Proof** Take a random subset \( Y \) of \( X \) in which each element is present independently and uniformly with probability \( p \), and let \( \Gamma[Y] \) denote the dependency graph for the events \( A_y : y \in Y \). Then the expected number of \( A_y : y \in Y \) which occur is the random variable \( Y_\mu = \sum_{y \in Y} P(A_y) \), and the analog of \( \Delta \) in \( Y \) is the random variable

\[
Y_\Delta = \sum_{xy \in \Gamma[Y]} P(A_x \cap A_y).
\]

Let \( Z = -Y_\mu + \frac{1}{2}Y_\Delta \). Then

\[
\mathbb{E}(Z) = -\mu p + \frac{p^2 \Delta}{2}.
\]
Therefore we can find a set $Y$ for which $Z \leq -p\mu + \frac{1}{2}p^2\Delta$. Applying Janson’s inequality to this $Y$, we obtain
\[
P(\bigcap_x \overline{A}_x) \leq e^{-p\mu + \frac{1}{2}p^2\Delta}.
\]
The stated inequality now obtains from taking $p = \frac{\mu}{\Delta}$.

### 3.1 Two applications

Let $G_{n,p}$ denote the random graph with edge-probability $p$, and let $Z$ be the number of triangles in $G_{n,p}$. Then $\mu := \mathbb{E}(Z) = p\binom{n}{3}$. Let $A_z$ be the event that a triple $z$ of vertices form a triangle in $G_{n,p}$. Then $P(Z = 0) = P(\bigcap_z \overline{A}_z)$. The event $A_w \cap A_z$ is precisely the appearance of two triangles sharing an edge in $G_{n,p}$ if $w \cap z \neq \emptyset$. Therefore
\[
\Delta = \sum_{w \cap z \neq \emptyset} P(A_w \cap A_z) = 6\binom{n}{4}p^5.
\]
By Janson’s inequality,
\[
\prod_z (1 - p^3) \leq P(Z = 0) \leq e^{-\mu + \frac{\Delta}{2}}.
\]
If $pn^{1/2} \to 0$, then we deduce
\[
\log P(Z = 0) \sim -p^3\binom{n}{3}.
\]
In particular, if $pn = c$, then the probability that $G_{n,p}$ contains no triangles is asymptotic to $e^{-c^3/6}$. The problem of accurately estimating $P(Z = 0)$ for large values of $p$ has received much attention – see Janson, Luczak and Rucinski.

For our second example, a set $S \subset \mathbb{Z}$ is called a Sidon set if all the differences of distinct pairs of elements in $S$ are distinct. The aim of this example is to give an estimate for the probability that a subset of $[n]$ is a Sidon set. Suppose we choose elements of $[n]$ independently with probability $q$. For reasons we will see later, we insist that $qn^{2/3} \to 0$. The probability that the set $S$ we obtain is a Sidon set may be computed via Janson’s inequality. For a given quadruple $x = (a, b, c, d)$ of elements of $[n]$ with $a - b = c - d$, define $A_x$ to be the event $X \subset S$. Let $N \sim \alpha n^3$ be the number of quadruples $x = (a, b, c, d)$ with $a - b = c - d$, where $\alpha$ is an absolute constant. By Janson’s inequality, with $\mu = q^4 N$,
\[
-\mu \lesssim \log P(\bigcap_x \overline{A}_x) \lesssim -\mu + \frac{\Delta}{2}.
\]
Clearly $\Delta$ is a polynomial of the form $\sum_{i=2}^7 a_i q^i n^{i-2}$ where $a_i$ are absolute constants. It is easily checked that $\Delta \ll qn^{2/3}\mu$ and this is much smaller than $\mu$, since $qn^{2/3} \to 0$. Therefore from Janson’s inequality,
\[
\log P(S \text{ is Sidon}) \sim q^4 N.
\]
It follows that the probability that $S$ is Sidon tends to zero if $qn^{3/4} \to \infty$ and tends to one if $qn^{3/4} \to 0$. So a “threshold” for a random set in $[n]$ to be Sidon is $q = n^{-3/4}$.
4 Generalizations

The following result is a generalization of Janson’s inequality. A proof can be found in the text on random graphs by Janson, Łuczak and Ruciński.

**Theorem 8** Let $S$ be a family of sets and let $A$ be a random subset of $[n]$. Let $X$ denote the number of events $A_S = (A \subseteq S)$ which occur with $S \in S$, and let $\mu = E(X)$. Then

$$P(X \leq \mu - t) \leq \exp\left(-\frac{t^2}{\Delta + \mu}\right).$$

As an exercise, one can construct examples where $P(X \geq t E(X))$ is a polynomial of bounded degree in $t$, so the theorem does not hold for “upper tails”. For upper tails, there is a fairly simple inequality based on dependency graphs: let $\Gamma$ be the dependency graph of the events $A_S$ in the last theorem. We denote by $d$ the maximum degree of this graph. The following was proved by Rödl and Ruciński (1994).

**Theorem 9**

$$P(X \geq \mu + t) \leq (d + 1) \exp\left(-\frac{3t^2}{4(d + 1)(3\mu + t)}\right).$$

**Proof** For simplicity we consider only the case that $|A| = r$ for all $A \in S$ and $P(x \in S) = p$ for all $x \in \Omega$. We find many independent events. Let $n = |V(\Gamma)|$. By the Hajnal-Szemerédi Theorem\(^2\), $V(\Gamma)$ admits a partition into $d + 1$ independent sets $I_1, I_2, \ldots, I_{d+1}$, each of size $\lceil n/(d + 1) \rceil$ or $\lfloor n/(d + 1) \rfloor$. For each independent set $I_j$, we have a family of pairwise disjoint sets $S_j \subseteq S$ of size $|I_j|$. Now if $X \geq \mu + t$, then there is a $j$ such that the number of events $I_A$ which occur with $A \in S_j$ is at least $p^r |S_j| + t |S_j|/|S|$. Since the expected number of these which occur is $p^r |S_j|$, we can use the Chernoff Bound to deduce

$$P(X \geq \mu + t) \leq \sum_{i=1}^{d+1} \exp\left(\frac{-t^2|S_j|}{2n(\mu + t/3)}\right),$$

which gives the required bound.

Another variation of Janson’s inequality was given by Suen:

**Theorem 10** Let $X_1, X_2, \ldots, X_n$ be Bernoulli random variables with probabilities $p_1, p_2, \ldots, p_n$, and let $X$ be their sum, with mean $\mu$. Let $\Gamma$ be a dependency graph of $X_1, X_2, \ldots, X_n$, $\Delta = \sum_{ij \in \Gamma} P(X_i = X_j = 1)$, and $\gamma = \max_j \sum_{ij \in \Gamma} p_i$. Then

$$P(X = 0) \leq \exp(-\mu + \Delta e^{2\gamma}).$$

\(^2\)Every graph of maximum degree $d$ has a proper colouring with $d + 1$ colours so that the numbers of times each colour is used differ by at most one. This is a theorem in extremal graph theory, which we shall not prove. A relatively short proof was given by Kierstead and Kostochka.
4.1 Poisson Approximation

In the setting of Janson’s inequality, let $Z$ be the number of events $A_x : x \in X$ which occur. If $E(Z) = \mu$ and $\Delta$ is much less than $\mu$, then Janson’s inequality gives $P(X = 0) \approx \exp(-\mu)$, which is the probability that a Poisson random variable with mean $\mu$ is zero. The topic of Poisson approximation was an original motivation for Janson’s inequality, and we now state conditions on the moments of random variables which guarantee Poisson approximation. More general machinery for convergence to other distributions is given by studying the moment generating function or characteristic function of a random variable, for example, as given by Lévy’s Convergence Theorem from which the central limit theorem follows. The following result is a version of Brun’s Sieve.

We write $X_r$ for the falling factorial $X(X-1) \cdots (X-r+1)$, and notice that if $X \sim \text{POI}(\mu)$, then $E_x[X] = \mu^r$.

**Theorem 11** Suppose that $X_1, X_2, \ldots, X_n$ are non-negative integer-valued random variables, and $E_x(X_i) := E(X_i(X_i-1) \cdots (X_i-r+1))$, and

$$E_x(X_i) \to \mu^r$$

where $\mu \in \mathbb{R}$. Then for every $t$, as $n \to \infty$,

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

The proof uses alternating sums (inclusion-exclusion) and is left as an exercise, or one can use generating functions – recall that the moment generating function of the Poisson distribution with mean $\lambda$ is $\exp(\lambda(e^t - 1))$. 
