Concentration of Measure

From independent trials to dependent processes

Exponential tails, gradual revelation, and Poissonization

An expectation is only the beginning of an algorithmic guarantee. Exponential moments control independent trials; conditional expectations turn gradual revelation into a martingale. These ideas explain occupancy, amplify randomized algorithms, justify Bloom filters, and make the Poisson heuristic rigorous.

Advanced Algorithms teaching team

Nanjing University

Learning goals and reading routes

This lecture develops concentration tools for randomized algorithms. By the end, you should be able to:

1. derive multiplicative Chernoff and additive Hoeffding bounds, and choose between them;
2. amplify decision and approximation algorithms by a majority or a median;
3. apply McDiarmid's inequality to nonlinear statistics and rigorous Bloom-filter analysis;
4. explain and prove bounded differences through Doob martingales and conditional range widths; and
5. use Poissonization to prove occupancy lower bounds and the coupon-collector limit.

For a first pass, read Sections 1–4 in order: independent sums and amplification; bounded differences; and Bloom filters. McDiarmid's inequality is stated in Section 3.1 and proved in Section 5.5. The applications first show what it accomplishes; Section 5 then explains the mechanism through conditional expectation and martingales. For a proof-first route, read Section 5 after Section 2, consulting Theorem 3.1, and then return to the applications.

Section 6 develops Poissonization, the maximum-load lower bound, and coupon collection. This complementary route uses exponential moments and elementary conditioning, not martingale concentration. Section 7 summarizes method selection. Appendix A supplies elementary estimates; Appendix B gives the optional negative-association extension. All applications and the complete proofs of the main concentration theorems remain in the main text.

Throughout, ln is the natural logarithm and $[n] = \{1, \ldots, n\}$. In occupancy statements, "with high probability" means probability $1 - O(1/n)$ unless another bound is displayed. Asymptotic thresholds apply for sufficiently large n; integer rounding and each application's randomness assumptions are stated where needed.

Contents

1 From expectations to high-probability guarantees

An expectation describes an average over the randomness. An algorithmic guarantee often asks for more: how likely is a large error, an overloaded location, or an unusually long execution? Concentration bounds answer such questions by controlling the probability of a deviation, not merely its average size. A typical goal is a bound of the form

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq \delta,$$

relating a deviation threshold $t > 0$ to an allowed failure probability $0 < \delta < 1$. Other questions require a one-sided tail or a limiting distribution.

Balls into bins

There are $n \geq 2$ bins. Each ball independently chooses a uniform destination in $[n]$. After $m \geq 1$ throws, we ask for the maximum load and the number of empty bins. With a continuing sequence of balls, we ask for the first time that all bins are covered.

Throw m balls independently and uniformly into n bins. Let $B_j \in [n]$ be the destination of ball j, and let $X_i = \sum_{j=1}^{m} \mathbf{1}\{B_j = i\}$ be the load of bin i. The destinations $B_1, \ldots, B_m$ are independent, but the loads $X_1, \ldots, X_n$ are not: their sum is always m.

Symmetry and linearity of expectation give $\mathbb{E}X_i = m/n$. This answers the load of a fixed bin on average, not the maximum load. In particular, $\max_i \mathbb{E}X_i = m/n$ is not a formula for $\mathbb{E} \max_i X_i$. Coupon collection asks when the random map first becomes surjective; maximum load asks how large its largest fiber is.

Three statistics will recur. The maximum load is $M = \max_i X_i$; the number of empty bins counts the indices with $X_i = 0$; and the coupon-collection time is the first throw at which every bin is nonempty. Their most useful representations are different. A fixed load is a sum of independent indicators. The empty-bin count is a function of independent destinations, although its natural summands are dependent. Coverage depends jointly on all the loads.

The union bound provides a simple way to turn a local estimate into a simultaneous guarantee. For events $A_1, \ldots, A_n$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Indeed, the indicator of the union is at most the sum of the individual indicators. No independence is needed. In particular, a failure probability at most n^{-2} for each bin gives a total failure probability at most $1/n$. This explains why exponentially small tails are useful even when only polynomially small failure is required.

Choose the representation before the inequality

For independent sums, factor exponential moments. For a nonlinear function of independent inputs, control what one input can change. Conditional revelation will later explain this principle and extend the method to dependent increments. When a fixed total obstructs independence, consider randomizing that total and then transferring the result back.

2 Independent sums: Chernoff, Hoeffding, and amplification

2.1 Markov's inequality applied to an exponential

For a nonnegative random variable Z and $a > 0$, $\mathbb{P}(Z \geq a) \leq \mathbb{E}Z/a$, since $Z \geq a\mathbf{1}\{Z \geq a\}$. The useful choice is often not $Z = X$, but $Z = e^{\lambda X}$. Define the moment generating function $M_X(\lambda) = \mathbb{E}e^{\lambda X}$ wherever it is finite. For $\lambda > 0$,

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{\lambda X} \geq e^{\lambda a}) \leq e^{-\lambda a} M_X(\lambda). \tag{1}$$

For a lower tail choose $\lambda < 0$: then $X \leq a$ implies $e^{\lambda X} \geq e^{\lambda a}$. The exponential turns an additive statistic into a product, and independence can factor that product. The parameter λ remains free until the final optimization.

For bounded X, the Taylor expansion can be averaged term by term: $M_X(\lambda) = \sum_{r \geq 0} \lambda^r \mathbb{E}[X^r]/r!$. This explains the name "moment generating." For unbounded variables, the corresponding integrability and interchange conditions must be checked rather than assumed.

2.2 The Bernoulli MGF and the multiplicative bounds

Let $X_1, \ldots, X_N$ be independent $\{0, 1\}$-valued random variables, with $p_i = \mathbb{P}(X_i = 1)$. Set $X = \sum_i X_i$ and $\mu = \mathbb{E}X = \sum_i p_i$. The probabilities need not be equal. Such $\{0, 1\}$-valued variables are Bernoulli trials; Poisson-distributed variables, introduced in Section 6, instead take nonnegative integer values.

For every real λ,

$$\begin{aligned} \mathbb{E}e^{\lambda X_i} &= 1 + p_i(e^\lambda - 1) \leq \exp\big(p_i(e^\lambda - 1)\big), \\ \mathbb{E}e^{\lambda X} &= \prod_{i=1}^{N} \mathbb{E}e^{\lambda X_i} \leq \exp\big(\mu(e^\lambda - 1)\big). \end{aligned} \tag{2}$$

The inequality $1 + u \leq e^u$ is valid for negative as well as positive u, so this MGF estimate serves both tails. If $\mu = 0$, all the variables vanish almost surely and the concentration question is trivial.

Theorem 2.1 (Multiplicative Chernoff bounds)**.** *Under the preceding assumptions and with $\mu > 0$,*

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu, \qquad \delta > 0, \tag{3}$$

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu, \qquad 0 < \delta < 1. \tag{4}$$

Proof. For the upper tail, combine exponential Markov with (2):

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp\Big(\mu\big(e^\lambda - 1 - \lambda(1+\delta)\big)\Big), \qquad \lambda > 0.$$

The exponent has derivative $\mu(e^\lambda - (1+\delta))$ and positive second derivative. Its minimum is at $\lambda = \ln(1+\delta)$, which gives (3). For the lower tail take $\lambda < 0$ and obtain

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq \exp\Big(\mu\big(e^{\lambda} - 1 - \lambda(1-\delta)\big)\Big).$$

Now the minimizing parameter is $\lambda = \ln(1-\delta) < 0$, giving (4).

□

The exponential-moment template

Choose an exponential that is large on the bad event. Bound its expectation using the probabilistic structure. Optimize the parameter. A union bound or an algorithmic repetition argument comes afterwards; neither replaces the MGF estimate.

Useful forms. Elementary logarithmic estimates (proved in Appendix A) simplify the exact exponents. For $0 < \delta \leq 1$,

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}, \qquad \mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}. \tag{5}$$

At $\delta = 1$, the lower-tail statement follows by a limit, or directly from $\mathbb{P}(X=0) = \prod_i (1-p_i) \leq e^{-\mu}$. For a threshold $a > \mu$, set $\delta = a/\mu - 1$ to get

$$\mathbb{P}(X \geq a) \leq e^{-\mu}\left(\frac{e\mu}{a}\right)^a \leq 2^{-a} \quad \text{if } a \geq 2e\mu. \tag{6}$$

2.3 Maximum-load upper bounds

Return to m independent uniform balls in $n \geq 2$ bins. The load of a fixed bin satisfies $X_i \sim \mathrm{Bin}(m, 1/n)$ with mean $\mu = m/n$. For any real threshold $L > \mu$, the full Chernoff bound and then the union bound give

$$\mathbb{P}(M \geq L) \leq n\mathbb{P}(X_1 \geq L) \leq ne^{-\mu}\left(\frac{e\mu}{L}\right)^L. \tag{7}$$

Theorem 2.2 (Maximum-load upper bounds)**.** *With probability* $1 - O(1/n)$,

$$M = \begin{cases} O(\ln n/\ln\ln n), & m = n, \\ O(m/n), & m \geq n\ln n. \end{cases}$$

Proof. For $m = n$, take the integer $L = \lceil e\ln n/\ln\ln n\rceil$. As $n \to \infty$,

$$L(\ln L - 1) + 1 = (e + o(1))\ln n \geq 2\ln n$$

for sufficiently large n. Thus $\mathbb{P}(X_i \geq L) \leq e^{L-1}/L^L \leq n^{-2}$, and the union bound gives failure probability at most $1/n$.

For $m \geq n\ln n$, put $L = \lceil 2em/n\rceil$. Equation (6) gives $\mathbb{P}(M \geq L) \leq n2^{-L} \leq 1/n$, since $L \geq 2e\ln n$. If $L > m$, the event is impossible. □

The small-relative-deviation form (5) is not the right tool for a load far above its mean $\mu = 1$. The full exponent retains the $L\ln L$ term that produces $\ln n/\ln\ln n$.

An elementary alternative: two union bounds. Write $M = \max_{1 \le i \le n} X_i$. If bin i has load at least an integer $L \ge 1$, some L-element subset of the balls all landed in it. Thus

$$\begin{aligned} \mathbb{P}(M \ge L) &\le n\mathbb{P}(X_1 \ge L) \\ &\le n\binom{m}{L} n^{-L} \le n\left(\frac{em}{nL}\right)^{L}, \qquad 1 \le L \le m. \end{aligned} \tag{8}$$

The first union is over bins; the second is over subsets of balls. Independence is used only to assign probability n^{-L} to a fixed subset landing in a fixed bin. If $L > m$, the event is impossible. The factorial estimate $L! \ge (L/e)^L$ proves the final inequality.

This counting bound gives the same orders without an exponential-moment calculation. For $m = n$, $L = \lceil 3 \ln n / \ln \ln n \rceil$ makes $(e/L)^L \le n^{-2}$ for sufficiently large n. For $m \ge n \ln n$, $L = \lceil e^2 m/n \rceil$ makes $em/(nL) \le e^{-1}$ and $ne^{-L} \le 1/n$. The two proofs expose the same rare event: many specified balls choosing one bin.

2.4 Hoeffding's lemma and additive concentration

Lemma 2.3 (Hoeffding's lemma)**.** *If $Y \in [a, b]$ almost surely and $\mathbb{E}Y = 0$, then for every $\lambda \in \mathbb{R}$,*

$$\mathbb{E}e^{\lambda Y} \le \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right). \tag{9}$$

Proof. The case $a = b$ is immediate. Otherwise put $d = b - a$ and $p = -a/d \in [0, 1]$. Convexity places $e^{\lambda y}$ below the chord joining its values at a and b. Averaging that chord gives

$$\mathbb{E}e^{\lambda Y} \le (1-p)e^{\lambda a} + pe^{\lambda b} = e^{-p\lambda d}(1 - p + pe^{\lambda d}).$$

For $0 < p < 1$, define $h(s) = \ln(1 - p + pe^s) - ps$. Then $h(0) = h'(0) = 0$ and

$$h''(s) = \frac{p(1-p)e^s}{(1-p+pe^s)^2} = q_s(1-q_s) \le \frac{1}{4}, \qquad q_s = \frac{pe^s}{1-p+pe^s}.$$

Taylor's formula with an integral remainder yields $h(s) \le s^2/8$ for positive and negative s. Substitute $s = \lambda d$ and exponentiate. The endpoint cases $p = 0, 1$ are degenerate and also satisfy the bound.

□

The symmetric interval. For a centered $Y \in [-c, c]$ with $c > 0$, the same chord argument has an especially simple form:

$$\mathbb{E}e^{\lambda Y} \le \frac{e^{\lambda c} + e^{-\lambda c}}{2} = \sum_{r \ge 0} \frac{(\lambda c)^{2r}}{(2r)!} \le \sum_{r \ge 0} \frac{(\lambda c)^{2r}}{2^r r!} = e^{\lambda^2 c^2/2}.$$

The comparison uses $(2r)! \ge 2^r r!$. This is the bounded-increment exponential estimate underlying Azuma's inequality. It agrees with Hoeffding's lemma because the interval width is $2c$.

Theorem 2.4 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_N$ be independent, with $X_i \in [a_i, b_i]$ almost surely. Write $X = \sum_i X_i$ and $V = \sum_i (b_i - a_i)^2$. If $V > 0$, then for every $t > 0$,*

$$\mathbb{P}(X - \mathbb{E}X \geq t) \leq e^{-2t^2/V}, \qquad \mathbb{P}(X - \mathbb{E}X \leq -t) \leq e^{-2t^2/V}, \tag{10}$$

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2e^{-2t^2/V}. \tag{11}$$

If $V = 0$, X is constant.

Proof. Center each variable: $Y_i = X_i - \mathbb{E}X_i$. Centering does not change the interval width. By Lemma 2.3 and independence,

$$\mathbb{E}e^{\lambda \sum_i Y_i} = \prod_i \mathbb{E}e^{\lambda Y_i} \leq e^{\lambda^2 V/8}.$$

Exponential Markov bounds the upper tail by $\exp(-\lambda t + \lambda^2 V/8)$ for $\lambda > 0$. Taking $\lambda = 4t/V$ gives the first bound. Applying the same argument to $-\sum_i Y_i$, or choosing $\lambda = -4t/V$, gives the lower tail. A union bound gives the two-sided statement.

□

For Bernoulli variables every width is 1, so $V = N$. Unlike the multiplicative Chernoff bound, the additive exponent depends on the total range budget, not the mean. Both are useful; neither uniformly dominates the other in all parameter regimes.

2.5 Majority and median amplification

Decision algorithms. Let $f : \{0,1\}^* \to \{0,1\}$ be a decision problem. Suppose a randomized algorithm $\mathcal{A}$ satisfies, for every fixed input x,

$$\mathbb{P}(\mathcal{A}(x) = f(x)) \geq \frac{1}{2} + p, \qquad 0 < p \leq \frac{1}{2}.$$

Fix a desired error probability $0 < \delta < 1$.

Majority amplification

Choose an odd integer $r \geq \ln(1/\delta)/(2p^2)$. Run $\mathcal{A}(x)$ with r independent random seeds and return the majority of the r answers.

Let I_j indicate correctness on run j, and set $S = \sum_j I_j$. Then $\mathbb{E}S \geq r(1/2 + p)$; it need not equal this lower bound. Majority failure implies $S \leq r/2 \leq \mathbb{E}S - rp$. Thus Hoeffding gives

$$\mathbb{P}(\text{majority is wrong}) \leq e^{-2rp^2} \leq \delta \quad \text{if } r \geq \frac{\ln(1/\delta)}{2p^2}. \tag{12}$$

Choose the next odd integer if necessary. Fresh independent randomness across runs is essential to this derivation. Reusing the same random seed need not reduce the error at all.

Numerical approximation. Now suppose each independent run returns a real number lying in a fixed target interval $I_x = [a_x, b_x]$ with probability at least $1/2 + p$.

Median amplification

Choose an odd integer $r \geq \ln(1/\delta)/(2p^2)$. Obtain r independent numerical outputs and return their median.

If more than half the outputs belong to I_x, their median belongs to I_x, regardless of how large the erroneous outputs are. The same correctness indicators prove

$$\mathbb{P}(\text{median} \notin I_x) \leq e^{-2rp^2}. \tag{13}$$

For a nonnegative quantity $f(x)$, take $I_x = [(1-\varepsilon)f(x), (1+\varepsilon)f(x)]$. For a signed quantity, use an explicitly ordered interval, for example $[f(x) - \varepsilon|f(x)|, f(x) + \varepsilon|f(x)|]$. The proof is about interval membership, not about averaging approximate values.

Voting gives another application of the same deviation scale. If N independent fair votes are augmented by b votes all on one side, the probability that this side fails to obtain a strict majority is at most $e^{-b^2/(2N)}$. Indeed, the fair-vote count must deviate below $N/2$ by at least $b/2$. Hence b of order $\sqrt{N \ln N}$ gives polynomially small failure probability in this idealized model.

2.6 Sub-Gaussian variables: the common formulation

Definition 2.5. A centered random variable Y is sub-Gaussian with variance factor $\nu \geq 0$, written $Y \in \mathcal{G}(\nu)$, if

$$\mathbb{E}e^{\lambda Y} \leq e^{\lambda^2 \nu/2} \qquad \text{for every } \lambda \in \mathbb{R}. \tag{14}$$

A variance factor is an MGF upper-bound parameter, not an assertion that $\mathrm{Var}(Y) = \nu$. Hoeffding's lemma says that any centered $Y \in [a,b]$ belongs to $\mathcal{G}((b-a)^2/4)$. The quantifier "every real λ" is part of the definition.

Corollary 2.6. *If centered $Y_i \in \mathcal{G}(\nu_i)$ are independent, then $\sum_i Y_i \in \mathcal{G}(\sum_i \nu_i)$. For $v = \sum_i \nu_i > 0$ and $t > 0$,*

$$\mathbb{P}\left(\sum_i Y_i \geq t\right) \leq e^{-t^2/(2v)}, \qquad \mathbb{P}\left(\sum_i Y_i \leq -t\right) \leq e^{-t^2/(2v)}.$$

Proof. Multiply the marginal MGF bounds using independence, and then optimize $-\lambda t + \lambda^2 v/2$ at $\lambda = t/v$. Negating the variables gives the other tail. □

The definition packages the Gaussian-type exponential estimate used above. We next ask whether the statistic must be a sum: can a general function of independent inputs have the same kind of concentration? Section 3 gives a condition that is easy to check. After its applications, Section 5 explains the mechanism using conditional exponential moments. Appendix B gives a different extension, based on negative association, for which a product inequality still holds.

3 Bounded differences and nonlinear statistics

3.1 From sums to functions of independent inputs

Hoeffding's inequality controls a sum of independent bounded inputs. Its range budget has a simple interpretation: changing input i can change the sum by at most $b_i - a_i$. The next theorem preserves this principle for a general function, even when it is nonlinear or its natural summands are dependent.

Theorem 3.1 (McDiarmid's inequality). *Let $X = (X_1, \ldots, X_N)$ have independent coordinates, with X_i taking values in a set Ω_i. Let $f : \prod_{i=1}^N \Omega_i \to \mathbb{R}$ be measurable. Suppose deterministic constants $c_i \geq 0$ satisfy*

$$|f(x_1, \ldots, x_i, \ldots, x_N) - f(x_1, \ldots, x_i', \ldots, x_N)| \leq c_i \tag{15}$$

for every i, every $x \in \prod_j \Omega_j$, and every $x_i' \in \Omega_i$. Put $C = \sum_{i=1}^N c_i^2$. If $C > 0$, then for every $t > 0$,

$$\begin{aligned} \mathbb{P}(f(X) - \mathbb{E}f(X) \geq t) &\leq \exp(-2t^2/C), \\ \mathbb{P}(f(X) - \mathbb{E}f(X) \leq -t) &\leq \exp(-2t^2/C). \end{aligned} \tag{16}$$

Consequently,

$$\mathbb{P}\big(|f(X) - \mathbb{E}f(X)| \geq t\big) \leq 2\exp(-2t^2/C). \tag{17}$$

If $C = 0$, then f is constant on the product domain, so each deviation probability is zero.

The hypothesis bounds the change between *any two* values of one coordinate while all other coordinates are fixed. This is a coordinatewise Lipschitz condition; the constants c_i need not be equal. The theorem is also called the bounded-differences inequality [2].

We first explore its consequences. The complete proof is in Section 5.5, where revealing the independent inputs one at a time explains why their squared coordinate bounds add.

How it extends Chernoff–Hoeffding. For a sum of independent $X_i \in [a_i, b_i]$, set

$$f(x_1, \ldots, x_N) = \sum_{i=1}^N x_i, \qquad c_i = b_i - a_i.$$

McDiarmid then gives exactly Hoeffding's inequality, with the same constants. For independent Bernoulli inputs, each $c_i = 1$, so the one-sided additive bound is $e^{-2t^2/N}$. Thus the generalization is from *sums* to *functions* of independent inputs, not from independent inputs to arbitrary dependent ones.

The full multiplicative Chernoff bound remains a separate tool: it retains the mean μ, whereas the bounded-differences budget only records the coordinate changes. McDiarmid therefore extends the additive Chernoff–Hoeffding bound, but does not subsume every mean-sensitive Chernoff estimate.

The quantity that controls the deviation

Identify the independent random choices, write the statistic as their function, and bound the change caused by replacing each choice. Input i contributes c_i^2 to $C = \sum_i c_i^2$. For $C > 0$ and $0 < \delta < 1$, a deviation of size $\sqrt{(C/2)\ln(2/\delta)}$ has probability at most δ. The calculation requires neither independent summands nor an additive representation of the statistic.

3.2 Dependent indicators, independent choices

Empty bins. In the m-ball, n-bin model, let Y be the number of empty bins. For a fixed bin, emptiness has probability $(1-1/n)^m$, so

$$\mathbb{E}Y = n(1-1/n)^m. \tag{18}$$

The emptiness indicators are dependent, but the ball destinations are independent. Write

$$Y = f(B_1,\ldots,B_m) = n - \big|\{B_1,\ldots,B_m\}\big|.$$

Move one ball from bin u to bin v. If $u \neq v$, its departure may create one empty bin, and its arrival may eliminate one empty bin. When both changes occur, they cancel in the total. Therefore the net change in Y has absolute value at most 1, not 2. Applying (17) with m coordinates and $c_j = 1$ yields

$$\mathbb{P}(|Y - n(1-1/n)^m| \geq t) \leq 2e^{-2t^2/m}. \tag{19}$$

For example, substituting $t = \sqrt{cm \ln n/2}$ gives failure probability at most $2n^{-c}$. This is an additive estimate. When the expected number of empty bins is very small, the resulting additive window may be much larger than the mean; concentration around a mean is not automatically relative concentration.

Occurrences of an overlapping pattern. Let $X_1 \cdots X_n$ be a uniformly random string over an alphabet Σ of size $q \geq 1$, and fix a pattern $\pi \in \Sigma^k$, with $1 \leq k \leq n$. Let Y count all occurrences, including overlapping ones:

$$Y = \sum_{j=1}^{n-k+1} \mathbf{1}\{X_j X_{j+1} \cdots X_{j+k-1} = \pi\}, \qquad \mathbb{E}Y = \frac{n-k+1}{q^k}.$$

A single character participates in at most k candidate windows. Changing it can alter at most k indicator values, so Y is a function of n independent letters with coordinate bounds $c_i \leq k$. Therefore

$$\mathbb{P}\left(\left|Y - \frac{n-k+1}{q^k}\right| \geq t\right) \leq 2e^{-2t^2/(nk^2)}. \tag{20}$$

The overlap prevents treating the window indicators as independent. The proof avoids analyzing their correlation altogether.

3.3 Nonadditive statistics

Distance to a set in the hypercube. Fix a nonempty set $S \subseteq \{0,1\}^n$ and choose a uniform random $X \in \{0,1\}^n$. Define the Hamming distance $H(x,y) = \sum_i |x_i - y_i|$ and the distance to the set

$$D = f_S(X) = \min_{y \in S} H(X,y).$$

Changing one bit changes the distance to each fixed y by at most 1, and therefore changes their minimum by at most 1. The independent bits give

$$\mathbb{P}(|D - \mathbb{E}D| \geq t) \leq 2e^{-2t^2/n}, \qquad \mathbb{P}\Big(|D - \mathbb{E}D| \geq \sqrt{cn \ln n/2}\Big) \leq 2n^{-c}. \tag{21}$$

Thus most points have nearly the same distance to S on this additive scale. The argument does not locate $\mathbb{E}D$ and does not assert that most points are close to S. Requiring $S \neq \varnothing$ makes the minimum well-defined.

Random signs in a normed space. Fix unit vectors $u_1, \ldots, u_n$ in any normed space, and choose independent uniform signs $\varepsilon_i \in \{-1, 1\}$. Set $Z = \|\sum_i \varepsilon_i u_i\|$. Flipping sign i changes the vector sum by $2u_i$, so the reverse triangle inequality gives a change in the norm of at most 2. Hence

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2e^{-t^2/(2n)}, \qquad \mathbb{P}(|Z - \mathbb{E}Z| \geq s\sqrt{n}) \leq 2e^{-s^2/2}. \tag{22}$$

No assumption about an inner product or Euclidean geometry is needed. Once again, the inequality controls fluctuations, not the location of the mean.

3.4 Choosing the independent coordinates: graph coloring

A random graph $G \sim G(n, p)$ has vertex set $[n]$, and each of the $\binom{n}{2}$ possible edges is independently present with probability p. Its chromatic number $\chi(G)$ is the minimum number of colors in a proper vertex coloring, where adjacent vertices receive different colors. We can prove concentration without computing its mean. The useful choice is to group the random edges by vertex, rather than treat each edge as a separate coordinate.

Theorem 3.2 (Chromatic-number concentration)**.** *For $n \geq 2$, $0 \leq p \leq 1$, $G \sim G(n, p)$, and every $s > 0$,*

$$\mathbb{P}\big(|\chi(G) - \mathbb{E}\chi(G)| \geq s\sqrt{n}\big) \leq 2e^{-2s^2}. \tag{23}$$

Proof. Partition the edge indicators into independent blocks B_i, where B_i contains the edges from vertex i to vertices $1, \ldots, i-1$. Each edge belongs to exactly one block, so the blocks are independent and together determine G.

Fix the other blocks and compare two choices for B_i. The resulting graphs G, G' differ only in edges incident to vertex i. After deleting i, both become the same graph H. Since

$$\chi(H) \leq \chi(G) \leq \chi(H) + 1, \qquad \chi(H) \leq \chi(G') \leq \chi(H) + 1,$$

their chromatic numbers differ by at most 1: color H optimally and assign a new color to i if necessary. Thus the coordinate bound for each independent block is 1. Applying McDiarmid's inequality to these independent blocks with $C \leq n$ and $t = s\sqrt{n}$ proves (23).

□

Using each edge as a coordinate also gives a unit change bound, but its budget is $\binom{n}{2}$ instead of at most n. The grouped representation is better because changing all the relevant edges at one vertex still changes the chromatic number by at most one. Neither calculation determines the mean. Section 5.6 will interpret these two representations as edge- and vertex-exposure martingales; this perspective is used in work of Shamir and Spencer [6].

4 Bloom filters: a rigorous error analysis

4.1 The probability model and the exact conditional formula

Approximate membership under ideal hashing

Let U be a finite key universe. Fix $n \geq 1$ distinct keys $S \subseteq U$, an array length $m \geq 2$, and an integer $k \geq 1$. The set and the nonmember to be queried are fixed before the randomness is chosen. The aim is to bound the false-positive probability of this one query.

Choose independent uniformly random functions $h_1, \ldots, h_k : U \to [m]$.

Shared-array Bloom filter

Initialize $A[1], \ldots, A[m]$ to zero. For every inserted key $y \in S$, set $A[h_j(y)] = 1$ for all $j \in [k]$. On query x, answer yes if and only if all k tested bits $A[h_j(x)]$ equal one.

Inserted keys have no false negatives. Fix a nonmember $x \notin S$ before the hash functions are sampled. Under this ideal random-function model, all kn insertion locations are independent uniform samples, and the k query locations are independent uniform samples independent of the inserted array. These assertions use more than pairwise independence.

Let $W = \sum_{a=1}^{m} A[a]$ be the number of occupied bits, and let

$$q_m = \frac{\mathbb{E}W}{m} = 1 - \left(1 - \frac{1}{m}\right)^{kn}. \tag{24}$$

Fix a realized array A. Each of the independent query locations lands on a 1 with probability W/m, so the conditional false-positive probability is $(W/m)^k$. To remove the conditioning, average over the possible arrays. Arrays with the same occupied-bit count w have the same conditional error, so grouping them by w gives

$$\begin{aligned} \mathbb{P}(\text{false positive} \mid A) &= \left(\frac{W}{m}\right)^k, \\ p_{\mathrm{FP}} &= \sum_{w=0}^{m} \left(\frac{w}{m}\right)^k \mathbb{P}(W = w) = \mathbb{E}\left[\left(\frac{W}{m}\right)^k\right]. \end{aligned} \tag{25}$$

This is the law of total probability, written as a finite weighted average. No independence among the occupied bits is being assumed.

Conditionally independent does not mean unconditionally independent

The k tests share the same random array. Replacing $\mathbb{E}[(W/m)^k]$ by $(\mathbb{E}W/m)^k$ discards that shared randomness. In fact, convexity gives $\mathbb{E}[(W/m)^k] \geq (\mathbb{E}W/m)^k$ for $k \geq 1$. The familiar expression $(1 - e^{-kn/m})^k$ is a heuristic until the occupancy fluctuation is controlled.

4.2 A finite-size concentration bound

The occupied-bit count W is 1-Lipschitz in each of the kn insertion locations, by the empty-bin argument in Section 3.2. For $t > 0$, the upper-tail part of McDiarmid's inequality gives

$$\mathbb{P}(W > \mathbb{E}W + t) \le \exp\left(-\frac{2t^2}{kn}\right).$$

On the complementary event, $(W/m)^k \le \min\{1, q_m + t/m\}^k$. On the exceptional event it is at most 1. Splitting the expectation in (25) proves the following finite statement.

Theorem 4.1 (Bloom-filter false positives)**.** *For a fixed set and a fixed nonmember under the ideal model above,*

$$q_m^k \le p_{\mathrm{FP}} \le \min\left\{1, q_m + \frac{t}{m}\right\}^k + \exp\left(-\frac{2t^2}{kn}\right) \qquad (t > 0). \tag{26}$$

For any $0 < \eta < 1$, choosing $t = \sqrt{(kn/2)\ln(1/\eta)}$ makes the exceptional term at most η. This gives a finite error budget without requiring an asymptotic regime.

For example, take $t = \sqrt{50kn\ln(2n)}$. The exceptional probability is at most $(2n)^{-100} \le n^{-100}$, giving

$$p_{\mathrm{FP}} \le \min\left\{1, q_m + \frac{\sqrt{50kn\ln(2n)}}{m}\right\}^k + n^{-100}. \tag{27}$$

This is an explicit guarantee, not an independence approximation. The lower bound in (26) shows why replacing a random occupied fraction by its mean tends to underestimate the error.

4.3 The asymptotic formula and the choice of k

Fix $c > 0$ and an integer $k \ge 1$, and take $m = \lceil cn \rceil$. The occupied fraction satisfies $q_m \to 1 - e^{-k/c}$, while the fluctuation term $\sqrt{50kn\ln(2n)}/m$ tends to zero. The upper and lower bounds squeeze the false-positive probability to

$$p_{\mathrm{FP}} = (1 - e^{-k/c})^k + o(1). \tag{28}$$

For an explicit error accounting, $q_m = 1 - e^{-k/c} + O(1/n)$ and the map $u \mapsto u^k$ is k-Lipschitz on $[0, 1]$. Thus the upper error from this argument is $O_{c,k}(\sqrt{\ln n/n}) + n^{-100}$. The small fluctuation term cannot simply be deleted from a finite-n bound.

To select the number of hashes, consider the continuous relaxation $F_c(k) = (1 - e^{-k/c})^k$. Put $u = e^{-k/c} \in (0, 1)$. Then $\ln F_c(k) = -c \ln u \ln(1 - u)$. The product $\ln u \ln(1 - u)$ is maximized at $u = 1/2$ (a calculus verification is in Appendix A). Hence the relaxed optimum is

$$k = c\ln 2, \qquad F_c(k) = e^{-c(\ln 2)^2} = (0.6185\ldots)^c. \tag{29}$$

The data structure must use an integer k. Compare the positive integers adjacent to $c\ln 2$ and choose the one with smaller $F_c(k)$; when $c\ln 2 < 1$, use $k = 1$. Formula (29) is the relaxed value, not an exact guarantee for an arbitrary integer rounding.

The scope of the guarantee

The set and nonmember query are fixed independently of the random hash functions. This proof does not establish the same result for adaptively chosen queries or for a merely pairwise-independent hash family. The array uses m bits and a query uses k hash evaluations; the ideal fully random functions are a modeling assumption, not a short-seed implementation.

The analysis has used only the statement of McDiarmid's inequality, a one-coordinate occupancy bound, and elementary conditioning. We now turn to the common mechanism behind this and the preceding nonlinear concentration bounds.

5 Why bounded differences work: martingales

The preceding applications reduced concentration to a bound on what one independent input can change. Why should that condition suffice even when the statistic is nonlinear? Reveal the inputs successively and track the conditional estimate of the final value. Its total change is a sum of small, conditionally centered increments. This will prove Theorem 3.1 and also explain how concentration can survive dependence.

For independent sums, exponential moments factor. The new increments need not be independent, so we will replace factorization by successive conditioning. We first establish the averaging rules needed for this step.

5.1 Conditional expectation and revealed information

For an integrable discrete random variable Z and an event E with $\mathbb{P}(E) > 0$, define

$$\mathbb{E}[Z \mid E] = \sum_z z\mathbb{P}(Z = z \mid E) = \frac{\mathbb{E}[Z\mathbf{1}_E]}{\mathbb{P}(E)}. \tag{30}$$

For a discrete X, set $g(x) = \mathbb{E}[Z \mid X = x]$ at every value of positive probability. Then $\mathbb{E}[Z \mid X] = g(X)$. Values assigned to g outside the support of X do not affect this random variable almost surely. Conditioning on a tuple $(X_1, \ldots, X_i)$ is the same construction with a vector-valued label.

A population example distinguishes a function from its random evaluation. Choose a person uniformly from a finite population, let Z be height and X the country label. Then $g(x)$ is the average height within country x, and $g(X)$ is the average height of the country of the sampled person. Equivalently, choose a country with probability proportional to its population and return its mean height. Choosing a country uniformly would generally give a different distribution.

For a finite information partition $\mathcal{P}$, $\mathbb{E}[Z \mid \mathcal{P}]$ assigns to each outcome the average of Z over its partition cell. In general notation this is $\mathbb{E}[Z \mid \mathcal{F}]$, where $\mathcal{F}$ is the corresponding σ-algebra. All identities below hold almost surely. Integrability of the quantities and products being averaged is understood.

Linearity and taking out what is known. Conditional expectation is linear. For constants a, b and integrable U, V,

$$\mathbb{E}[aU + bV \mid X] = a\mathbb{E}[U \mid X] + b\mathbb{E}[V \mid X].$$

If $h(X)$ is known once X is known, then

$$\mathbb{E}[h(X)Z \mid X] = h(X)\mathbb{E}[Z \mid X]. \tag{31}$$

To verify this, condition on $X = x$. The factor $h(X)$ becomes the constant $h(x)$ and can be taken outside the finite or absolutely convergent sum defining the conditional mean. No independence between X and Z is required. In particular,

$$\mathbb{E}[h(X) \mid X] = h(X), \qquad \mathbb{E}[X \mid X] = X.$$

If Z is independent of X, then its conditional law equals its original law, and $\mathbb{E}[Z \mid X] = \mathbb{E}Z$. This is a separate use of independence, not a consequence of linearity alone.

Total expectation and the tower property.

Proposition 5.1. *For integrable Z and discrete labels X, U,*

$$\mathbb{E}\big[\mathbb{E}[Z \mid X]\big] = \mathbb{E}Z, \tag{32}$$
$$\mathbb{E}\big[\mathbb{E}[Z \mid X, U] \mid U\big] = \mathbb{E}[Z \mid U]. \tag{33}$$

More generally, if $\mathcal{G} \subseteq \mathcal{F}$, then

$$\mathbb{E}\big[\mathbb{E}[Z \mid \mathcal{F}] \mid \mathcal{G}\big] = \mathbb{E}[Z \mid \mathcal{G}]. \tag{34}$$

Proof. For total expectation, sum over the conditioning label:

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[Z \mid X]\big] &= \sum_x \sum_z z\mathbb{P}(Z = z \mid X = x)\mathbb{P}(X = x) \\ &= \sum_z z \sum_x \mathbb{P}(Z = z, X = x) = \mathbb{E}Z.\end{aligned}$$

For the tower property, perform the same calculation within each positive-probability event $U = u$. Averaging the finer conditional means over $X \mid U = u$ gives the mean of $Z \mid U = u$. In the partition formulation, first average within fine cells and then across the fine cells making up a coarse cell. Each cell is weighted by its conditional probability.

□

In the population example, one may first fix a gender and then average the country-specific mean heights within that group, weighting countries by their populations within the group. The result is the mean height conditional on that gender. This is the same averaging rule applied inside one conditional population.

The direction matters. The outer conditioning retains the coarser information. If $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i$, averaging the refined estimate $\mathbb{E}[Z \mid \mathcal{F}_i]$ using the old information returns the old estimate. This observation motivates the process below.

Combining the tower property with (31) gives a useful iterated-expectation identity:

$$\mathbb{E}\big[\mathbb{E}[f(X)g(X,Y) \mid X]\big] = \mathbb{E}\big[f(X)\mathbb{E}[g(X,Y) \mid X]\big]. \tag{35}$$

In an exponential-moment argument, the known factor will be the exponential of the past sum; only the new increment remains to be averaged.

5.2 Doob martingales: refine an estimate

Think of revealing information over time. A filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ records what is known at each step. In the finite settings here, it can be regarded as a sequence of successively finer partitions of the sample space. A variable is $\mathcal{F}_i$-measurable when its value is determined by the information available at step i.

Let $Z = f(X_1, \ldots, X_N)$ be integrable, without assuming independence of the inputs. Set $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$, with trivial $\mathcal{F}_0$, and define

$$Y_i = \mathbb{E}[Z \mid \mathcal{F}_i], \qquad Y_0 = \mathbb{E}Z, \qquad Y_N = Z. \tag{36}$$

The ith value averages over the still-hidden coordinates, conditional on the revealed prefix. It is not obtained by merely substituting zero for the hidden coordinates. At time zero no randomness has been observed; at the last step the value of f is fully known.

Definition 5.2 (Martingale)**.** An integrable sequence $Y_0, Y_1, \ldots$ is a martingale with respect to $(\mathcal{F}_i)$ if Y_i is $\mathcal{F}_i$-measurable and

$$\mathbb{E}[Y_i \mid \mathcal{F}_{i-1}] = Y_{i-1} \qquad (i \geq 1), \tag{37}$$

Equivalently, its conditional expected increment is zero. Replacing equality by $\leq$ defines a supermartingale, and replacing it by $\geq$ defines a submartingale.

Here $\sigma(X_0, \ldots, X_i)$ denotes all information determined by the listed variables. Taking $\mathcal{F}_i = \sigma(X_0, \ldots, X_i)$ defines a martingale with respect to the information revealed by an underlying sequence. Taking $\mathcal{F}_i = \sigma(Y_0, \ldots, Y_i)$ gives the definition in terms of the process's own past. Integrability is included explicitly so these expectations are well-defined.

Proposition 5.3 (Doob's construction)**.** *The sequence in* (36) *is a martingale with respect to* $(\mathcal{F}_i)$*.*

Proof. Each Y_i is determined by the revealed prefix. The tower property (34) gives

$$\mathbb{E}[Y_i \mid \mathcal{F}_{i-1}] = \mathbb{E}\big[\mathbb{E}[Z \mid \mathcal{F}_i] \mid \mathcal{F}_{i-1}\big] = \mathbb{E}[Z \mid \mathcal{F}_{i-1}] = Y_{i-1}.$$

□

No independence entered this proof. To prove the bounded-differences theorem from Section 3, independence will instead be used to control the conditional range of each refined estimate.

A nonlinear statistic becomes a sum of increments. Let $D_i = Y_i - Y_{i-1}$. Then

$$Z - \mathbb{E}Z = Y_N - Y_0 = \sum_{i=1}^{N} D_i, \qquad \mathbb{E}[D_i \mid \mathcal{F}_{i-1}] = 0.$$

The identity is telescoping; conditional centering follows from the martingale property. For the empty-bin count, Y_i is the expected final number of empty bins after seeing the first i destinations, not the current number of empty bins. We have recovered a sum, but its increments can depend on one another. The next concentration argument uses their conditional means and ranges in place of independence.

Fair increments and adaptive choices. As a basic example, let $\varepsilon_i \in \{-1, 1\}$ be independent fair signs and let $Y_i = \sum_{j \le i} \varepsilon_j$. Since ε_i is independent of the past and has mean zero, $\mathbb{E}[Y_i \mid \varepsilon_1, \ldots, \varepsilon_{i-1}] = Y_{i-1}$. More generally, a bounded stake b_i chosen from the past gives the martingale $\sum_{j \le i} b_j \varepsilon_j$. The fair-game intuition is that the choices may adapt to the past, but the conditional expected gain remains zero.

Doubling does not remove the risk

Consider a bettor who doubles the stake after each loss. A first win after r losses gives net gain $2^r - \sum_{j=0}^{r-1} 2^j = 1$, but requires stakes that grow without bound as r grows. An arbitrarily long losing streak has positive probability. With a finite bankroll, the strategy can fail before the next win; the formal gain calculation is not a guarantee of risk-free profit.

The fair random walk is also Markov, but the martingale definition does not require the Markov property. It constrains a conditional mean, not the entire conditional law of the next state.

5.3 The conditional-range Azuma–Hoeffding inequality

The conditional range need not be centered at zero, even when the conditional mean is zero. Its width is the quantity retained by Hoeffding's lemma.

Theorem 5.4 (Azuma–Hoeffding: conditional-range form)**.** *Let $(Y_i)_{i=0}^N$ be a martingale with respect to $(\mathcal{F}_i)$, and put $D_i = Y_i - Y_{i-1}$. Suppose there are $\mathcal{F}_{i-1}$-measurable endpoints a_i, b_i and deterministic constants $w_i \ge 0$ such that, almost surely,*

$$a_i \le D_i \le b_i, \qquad b_i - a_i \le w_i.$$

Put $W = \sum_{i=1}^N w_i^2$. If $W > 0$, then for every $t > 0$,

$$\begin{aligned} \mathbb{P}(Y_N - Y_0 \ge t) &\le e^{-2t^2/W}, \\ \mathbb{P}(Y_N - Y_0 \le -t) &\le e^{-2t^2/W}, \\ \mathbb{P}(|Y_N - Y_0| \ge t) &\le 2e^{-2t^2/W}. \end{aligned} \tag{38}$$

When $W = 0$, every increment is zero almost surely.

Proof. By the martingale property, $\mathbb{E}[D_i \mid \mathcal{F}_{i-1}] = 0$. Once the past is known, the interval endpoints are fixed. Applying Hoeffding's lemma to the conditional distribution gives, for every real λ,

$$\mathbb{E}[e^{\lambda D_i} \mid \mathcal{F}_{i-1}] \le \exp(\lambda^2 w_i^2/8). \tag{39}$$

The endpoints may depend on the past, but the width bound w_i is deterministic.

Let $S_i = \sum_{j=1}^i D_j = Y_i - Y_0$, with $S_0 = 0$. The increments need not be independent, so we do not multiply their unconditional moment generating functions. Instead, S_{i-1} is already known at time $i - 1$, and the tower property and the known-factor rule yield

$$\begin{aligned} \mathbb{E}e^{\lambda S_i} &= \mathbb{E}\left[\mathbb{E}[e^{\lambda S_{i-1}} e^{\lambda D_i} \mid \mathcal{F}_{i-1}]\right] \\ &= \mathbb{E}\left[e^{\lambda S_{i-1}} \mathbb{E}[e^{\lambda D_i} \mid \mathcal{F}_{i-1}]\right] \\ &\le e^{\lambda^2 w_i^2/8} \mathbb{E}e^{\lambda S_{i-1}}. \end{aligned} \tag{40}$$

Iterating from $i = N$ down to 1 gives

$$\mathbb{E}e^{\lambda(Y_N - Y_0)} \leq \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^{N} w_i^2\right) = e^{\lambda^2 W/8}.$$

Exponential Markov bounds the upper tail by $\exp(-\lambda t + \lambda^2 W/8)$ for $\lambda > 0$. The minimizing value is $\lambda = 4t/W$, giving $e^{-2t^2/W}$. Apply the same argument to the negated process for the lower tail, then add the two tail probabilities. If $W = 0$, each conditional interval has width zero; conditional centering then forces each increment to be zero. □

One conditional exponential-moment argument

More generally, if $\mathbb{E}[e^{\lambda D_i} \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2 v_i/2}$ for deterministic $v_i \geq 0$ and every real λ, the same iteration makes $\sum_i D_i$ sub-Gaussian with variance factor $\sum_i v_i$. For positive total factor, its one-sided tails are at most $e^{-t^2/(2\sum_i v_i)}$. The range-width theorem uses $v_i = w_i^2/4$.

5.4 Azuma's inequality for bounded increments

Corollary 5.5 (Azuma's inequality). *Let $(Y_i)_{i=0}^N$ be a martingale and suppose $|Y_i - Y_{i-1}| \leq c_i$ almost surely for deterministic $c_i \geq 0$ and $i \geq 1$. Put $C = \sum_i c_i^2 > 0$. Then for every $t > 0$,*

$$\mathbb{P}(Y_N - Y_0 \geq t) \leq e^{-t^2/(2C)}, \qquad \mathbb{P}(Y_N - Y_0 \leq -t) \leq e^{-t^2/(2C)}, \tag{41}$$

$$\mathbb{P}(|Y_N - Y_0| \geq t) \leq 2e^{-t^2/(2C)}. \tag{42}$$

Proof. Conditional on the past, $D_i = Y_i - Y_{i-1}$ lies in $[-c_i, c_i]$, an interval of width $2c_i$. Use $w_i = 2c_i$ in Theorem 5.4. In exponential-moment terms, the conditional factor is

$$\mathbb{E}[e^{\lambda D_i} \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2 c_i^2/2}. \tag{43}$$

Thus the variance factor is $\sum_i c_i^2$, rather than $\sum_i c_i^2/4$. □

The absolute-increment bound and the range-width bound are different hypotheses. Keeping the actual conditional width can improve the exponent by a factor of four; the distinction is essential in McDiarmid's inequality.

5.5 Proof of McDiarmid's inequality

We now complete the proof deferred in Section 3.1. The only remaining task is to translate a bound on changing one input into a bound on the conditional range of one Doob increment. Ordinary Azuma alone, applied with $|D_i| \leq c_i$, would lose a factor of four in the exponent; the range-width form retains the needed information.

Proof of Theorem 3.1. Use the Doob martingale $M_i = \mathbb{E}[f(X) \mid X_1, \ldots, X_i]$, with $M_0 = \mathbb{E}f(X)$ and $M_N = f(X)$. It remains to bound the conditional range of $D_i = M_i - M_{i-1}$.

Fix the revealed prefix $x_{<i} = (x_1, \ldots, x_{i-1})$. For a possible value a of coordinate i, set

$$g_i(a) = \mathbb{E}[f(x_{<i}, a, X_{>i})], \qquad X_{>i} = (X_{i+1}, \ldots, X_N).$$

Independence allows the *same distribution of the future* in this average for every value of a. For any $a, b \in \Omega_i$,

$$|g_i(a) - g_i(b)| \leq \mathbb{E}\big[|f(x_{<i}, a, X_{>i}) - f(x_{<i}, b, X_{>i})|\big] \leq c_i.$$

Thus the possible values of g_i lie in an interval of width at most c_i. Given the prefix, $M_i = g_i(X_i)$ and M_{i-1} is its conditional mean. Subtracting that mean translates the interval without changing its width. The Doob increment therefore has conditional mean zero and conditional range width at most c_i; equivalently, conditional Hoeffding gives

$$\mathbb{E}[e^{\lambda D_i} \mid X_1, \ldots, X_{i-1}] \leq e^{\lambda^2 c_i^2/8}. \tag{44}$$

Apply Theorem 5.4 with $w_i = c_i$ to obtain the two one-sided bounds and their union. If $C = 0$, changing coordinates one at a time shows that f is constant. □

What the proof explains

A coordinate change bound survives averaging over the same independent future. Subtracting a conditional mean does not enlarge the resulting interval. Conditional Hoeffding and repeated conditioning then combine the squared widths. This proves all the bounded-differences applications above without assuming that their apparent summands are independent.

5.6 Beyond independent inputs: averaged bounded differences

Let $Z = f(X_1, \ldots, X_N)$ be integrable; the input coordinates may be dependent. Doob's construction and Azuma's inequality apply if, for deterministic $c_i \geq 0$, almost surely

$$|\mathbb{E}[f(X) \mid X_1, \ldots, X_i] - \mathbb{E}[f(X) \mid X_1, \ldots, X_{i-1}]| \leq c_i. \tag{45}$$

Writing $C = \sum_i c_i^2 > 0$, the conclusion is

$$\mathbb{P}\big(|f(X) - \mathbb{E}f(X)| \geq t\big) \leq 2\exp\left(-\frac{t^2}{2C}\right). \tag{46}$$

This is the method of averaged bounded differences. "Averaged" refers to the conditional averages inside each difference; the bound on their difference must hold almost surely, not just after taking an expectation. A single tail is bounded by $e^{-t^2/(2C)}$.

For dependent inputs, a coordinatewise Lipschitz condition alone does not imply (45). Conditioning on one coordinate can change the distribution of all the unrevealed coordinates. For example, if $X_1 = \cdots = X_N = B$ for a fair bit B and $f(x) = \sum_i x_i$, changing one coordinate of f changes its value by at most 1, but revealing X_1 moves the conditional estimate from $N/2$ to either 0 or N. The first Doob increment has absolute value $N/2$.

Edge and vertex exposure revisited. For a random graph G on $[n]$ and an integrable statistic $f(G)$, number the possible edges and write their indicators as $I_1, \ldots, I_m$, where $m = \binom{n}{2}$. The edge-exposure martingale is

$$Y_j = \mathbb{E}[f(G) \mid I_1, \ldots, I_j], \qquad 0 \leq j \leq m.$$

Alternatively, write $G[i]$ for the subgraph induced by $[i]$. Since the induced subgraphs are nested, they reveal successively more information, and $\mathbb{E}[f(G) \mid G[i]]$, $0 \leq i \leq n$, is the vertex-exposure martingale. Take $G[0]$ to be the empty graph, giving the initial value $\mathbb{E}f(G)$. These constructions do not require the edges to be independent.

For $G(n,p)$, the independent blocks B_i in Section 3.4 reveal exactly $G[i]$ after the first i blocks. Thus the earlier graph-coloring argument controls the conditional widths of the vertex-exposure increments. With a different graph distribution, the Doob sequence is still a martingale, but one must establish its increment or range bounds under that distribution; the independent-block proof cannot simply be reused.

6 Poissonization and occupancy thresholds

6.1 Randomize the total to recover independence

The upper bound on maximum load required only a one-bin tail and a union bound. A lower bound asks something different: is it unlikely that *all* bins have small load? Multiplying their probabilities would be convenient, but the fixed identity $\sum_i X_i = m$ prevents independence.

Poissonization changes the experiment. Instead of fixing the number of throws, choose a suitable random total and then throw that many independent uniform balls. The individual loads become independent Poisson variables. Conditioning their total back to m recovers the original joint distribution exactly. The main work is to determine how much the probability of the event of interest changes when this conditioning is removed.

6.2 Poisson variables and the exact conditioning identity

A Poisson random variable $Z \sim \mathrm{Pois}(\mu)$, $\mu > 0$, has mass function

$$\mathbb{P}(Z = j) = e^{-\mu}\frac{\mu^j}{j!}, \qquad j = 0, 1, 2, \ldots.$$

Direct summation gives the exact MGF, for every real λ,

$$\mathbb{E}e^{\lambda Z} = e^{-\mu}\sum_{j\geq 0}\frac{(\mu e^{\lambda})^j}{j!} = \exp\big(\mu(e^{\lambda}-1)\big). \tag{47}$$

This is exactly the upper bound used for a sum of Bernoulli trials. Differentiating at zero gives $\mathbb{E}Z = \mu$ and $\mathrm{Var}\, Z = \mu$. Products of these MGFs also show that sums of independent Poisson variables are Poisson, with the means added.

Proposition 6.1 (Poisson tails)**.** *For a real threshold $a > 0$,*

$$\mathbb{P}(Z \geq a) \leq e^{-\mu}\left(\frac{e\mu}{a}\right)^a, \qquad a > \mu, \tag{48}$$

$$\mathbb{P}(Z \leq a) \leq e^{-\mu}\left(\frac{e\mu}{a}\right)^a, \qquad 0 < a < \mu. \tag{49}$$

Proof. Use (47) in exponential Markov and minimize $\mu(e^{\lambda}-1) - \lambda a$. The stationary point is $\lambda = \ln(a/\mu)$, positive for the upper tail and negative for the lower tail. At $a = 0$, the lower-tail probability is exactly $e^{-\mu}$. □

The same simplifications as in (5) imply, for $0 < t \le \mu$,

$$\mathbb{P}(Z \ge \mu + t) \le e^{-t^2/(3\mu)}, \qquad \mathbb{P}(Z \le \mu - t) \le e^{-t^2/(2\mu)}. \tag{50}$$

The joint law after conditioning. Let $X = (X_1, \ldots, X_n)$ be the load vector from exactly m uniform throws. The following identity gives its exact relation to independent Poisson loads.

Theorem 6.2 (Exact Poisson conditioning)**.** *For any $\lambda > 0$, let $Y_1, \ldots, Y_n$ be independent* $\mathrm{Pois}(\lambda)$ *variables and $S = \sum_i Y_i$. Then*

$$(X_1, \ldots, X_n) \overset{d}{=} (Y_1, \ldots, Y_n \mid S = m). \tag{51}$$

Proof. For nonnegative integers $x_1, \ldots, x_n$ summing to m, the load vector has multinomial mass

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \frac{m!}{x_1! \cdots x_n!}\, n^{-m}.$$

Since $S \sim \mathrm{Pois}(n\lambda)$, the corresponding conditional Poisson mass is

$$\frac{\prod_{i=1}^n (e^{-\lambda} \lambda^{x_i}/x_i!)}{e^{-n\lambda}(n\lambda)^m/m!} = \frac{m!}{x_1! \cdots x_n! n^m}.$$

Both distributions vanish outside the vectors with total m. This proves equality of the entire joint laws, not just their marginals.

□

An equivalent construction first draws $S \sim \mathrm{Pois}(m)$ and then throws S independent uniform balls. Unconditionally the loads are independent $\mathrm{Pois}(m/n)$ variables; conditionally on $S = m$ they recover the original experiment.

6.3 A general de-Poissonization comparison

Theorem 6.3 (The $e\sqrt{m}$ comparison)**.** *Let $m \ge 1$, let X be the load vector for m balls in n bins, and let Y_i be independent* $\mathrm{Pois}(m/n)$ *variables. For every nonnegative function f,*

$$\mathbb{E} f(X) \le e\sqrt{m}\, \mathbb{E} f(Y). \tag{52}$$

In particular, an event with Poissonized probability q has fixed-total probability at most $e\sqrt{m}\, q$.

Proof. Use nonnegativity to keep only the summand with total m:

$$\mathbb{E} f(Y) = \sum_{s \ge 0} \mathbb{E}[f(Y) \mid S = s]\mathbb{P}(S = s) \ge \mathbb{E} f(X) \mathbb{P}(S = m).$$

The factorial estimate $m! \le e\sqrt{m}(m/e)^m$ from Appendix A gives

$$\mathbb{P}(S = m) = e^{-m} \frac{m^m}{m!} \ge \frac{1}{e\sqrt{m}}.$$

Rearranging proves the claim. Taking f to be an event indicator gives the probability version.

□

This theorem is a one-sided comparison, not a claim that all event probabilities are asymptotically equal. It is especially useful when a bad event already has extremely small Poissonized probability. A nontrivial constant-probability limit requires a finer argument, which we give for coupon collection below.

6.4 The maximum-load lower bound

Take $m = n$, so the Poissonized loads are independent $\mathrm{Pois}(1)$ variables. If $L \geq 1$ is an integer, then

$$\mathbb{P}\left(\max_i Y_i < L\right) = \mathbb{P}(Y_1 < L)^n \leq (1 - \mathbb{P}(Y_1 = L))^n = \left(1 - \frac{1}{eL!}\right)^n \leq \exp\left(-\frac{n}{eL!}\right). \tag{53}$$

Independence now makes the event that all bins have small load easy to analyze. Put $L = \lfloor \ln n / \ln\ln n \rfloor$. To evaluate the factorial term, write $A = \ln n$, $B = \ln\ln n$, and $D = \ln\ln\ln n$. For sufficiently large n, $L \leq A/B$ and $\ln L \leq B - D$. The elementary bound $L! \leq L^L$ already yields

$$\ln \frac{n}{eL!} \geq A - 1 - L \ln L \geq \frac{AD}{B} - 1.$$

Since $(AD/B)/B \to \infty$, this implies $n/(eL!) \geq 3 \ln n$ for sufficiently large n. Thus (53) is at most n^{-3}. Applying Theorem 6.3,

$$\mathbb{P}\left(\max_i X_i < L\right) \leq e\sqrt{n}\, n^{-3} \leq \frac{1}{n}.$$

Theorem 6.4 (The typical maximum load)**.** *When n balls are thrown independently and uniformly into n bins,*

$$\max_i X_i = \Theta\left(\frac{\ln n}{\ln\ln n}\right) \quad \textit{with probability } 1 - O(1/n). \tag{54}$$

The lower bound is the Poissonization argument just proved; the upper bound is Theorem 2.2. As a separate immediate consequence, when $m \geq n \ln n$, the deterministic inequality $\max_i X_i \geq m/n$ combines with the upper bound to give $\Theta(m/n)$ with high probability.

6.5 Coupon collection: expectation, tails, and the limit law

The expectation and an elementary upper tail. Continue throwing balls indefinitely, and let T be the first time all n bins are nonempty. Equivalently, each independent purchase gives one uniformly random coupon among n types, and T is the number of purchases needed to collect every type. Let T_i count the throws from the moment $i - 1$ bins are occupied up to and including the next new bin.

Proposition 6.5. *The coupon-collection time satisfies*

$$T = \sum_{i=1}^{n} T_i, \qquad T_i \sim \mathrm{Geom}\left(\frac{n-i+1}{n}\right), \qquad \mathbb{E}T = nH_n = n \ln n + O(n), \tag{55}$$

where $H_n = \sum_{i=1}^{n} 1/i$ and the geometric distribution counts trials including the first success.

Proof. While exactly $i-1$ bins are occupied, each fresh throw hits a new bin with probability $p_i = (n-i+1)/n$. The mean waiting time is $1/p_i$ (Appendix A). Summing these expectations gives $\mathbb{E}T = \sum_i n/(n-i+1) = nH_n$. The bounds $\ln n \le H_n \le 1 + \ln n$ follow by comparison with the integral of $1/x$. Linearity itself does not require independence of the waiting times.

□

An expectation does not say that a particular run finishes near that time. Here the upper tail has an elementary proof that does not analyze the sum of waiting times.

Proposition 6.6 (Coupon-collector upper tail)**.** *For every integer $m \ge 0$ and $n \ge 2$,*

$$\mathbb{P}(T > m) \le n\left(1 - \frac{1}{n}\right)^m \le ne^{-m/n}. \tag{56}$$

Consequently, $\mathbb{P}(T > \lceil n \ln n + cn \rceil) \le e^{-c}$ for $c > 0$.

Proof. A fixed bin misses all m balls with probability $(1-1/n)^m$. The event $T > m$ means that at least one bin does so. Apply the union bound over the n bins and use $1 - x \le e^{-x}$.

□

The event $T > m$ is exactly "not covered after m throws." The expectation gives the average completion time, and the upper tail gives a high-confidence stopping bound. Neither by itself determines the fluctuations within an $O(n)$ window around $n \ln n$.

The sharp window. Fix $c \in \mathbb{R}$ and let $m = \lfloor n \ln n + cn \rfloor$, taking n sufficiently large that $m > 0$. For independent $Y_i \sim \mathrm{Pois}(m/n)$, the chance that every bin is nonempty is exactly

$$\mathbb{P}(Y_i > 0\ \forall i) = \left(1 - e^{-m/n}\right)^n = \left(1 - \frac{e^{-c+o(1)}}{n}\right)^n \longrightarrow e^{-e^{-c}}. \tag{57}$$

This is only the Poissonized calculation. To transfer the limit to exactly m balls, define $q_n(s)$ to be the probability that s uniform throws cover all n bins. Coupling all experiments by one infinite sequence of throws shows that $q_n(s)$ is nondecreasing in the integer s. Exact conditioning gives

$$\mathbb{P}(Y_i > 0\ \forall i) = \mathbb{E}[q_n(S)], \qquad S = \sum_i Y_i \sim \mathrm{Pois}(m). \tag{58}$$

Choose the integer window radius $a = \lceil \sqrt{2m \ln m} \rceil$. Because $m \sim n \ln n$, we have $a = o(n)$ and $a < m$ for sufficiently large n. By (50),

$$\eta_n := \mathbb{P}(|S - m| > a) \le e^{-a^2/(3m)} + e^{-a^2/(2m)} \le m^{-2/3} + m^{-1} = o(1). \tag{59}$$

Lemma 6.7 (Completion changes slowly with the number of throws)**.** *For integers $s, r \ge 0$,*

$$0 \le q_n(s+r) - q_n(s) \le r/n. \tag{60}$$

Proof. Couple the two experiments by using the same first s throws. If all bins are already covered, their completion indicators agree. Otherwise choose one distinguished empty bin, for example the smallest-index empty bin. Conditional on the first s throws, this is a fixed bin. For coverage to become complete during the next r throws, this particular bin must be hit. By a union bound, that conditional probability is at most r/n. Averaging proves the bound.

□

The distinguished bin is important. The probability that the new balls hit *some* empty bin need not be at most r/n when many bins are empty. Completing the cover, however, requires hitting the one we singled out.

On $|S-m| \le a$, monotonicity places both $q_n(S)$ and $q_n(m)$ between $q_n(m-a)$ and $q_n(m+a)$. Lemma 6.7 bounds the width of this interval by $2a/n$. Outside the window, both probabilities are in $[0,1]$. Therefore

$$\left|\mathbb{E}q_n(S) - q_n(m)\right| \le \frac{2a}{n} + \eta_n = o(1). \tag{61}$$

Combining (57)–(61) proves the claimed limit.

Theorem 6.8 (Coupon-collector limit law)**.** *For each fixed real c,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{T - n\ln n}{n} \le c\right) = e^{-e^{-c}}. \tag{62}$$

The fluctuation window is of order n around $n \ln n$. Its upper-tail form is $1 - e^{-e^{-c}}$. Switching a threshold between strict and non-strict inequalities changes the completion probability by at most $1/n$, by Lemma 6.7, so it does not change the limit.

Two different uses of Poissonization

For maximum load, independence makes a bad event tiny, and the general $e\sqrt{m}$ comparison preserves that conclusion. For the coupon-collector limit, a constant probability must be preserved accurately: concentration of the total, monotonicity, and a slow-change coupling remove the conditioning with an $o(1)$ error.

7 Choosing a concentration argument

The choice of a concentration inequality begins with the random representation, not the theorem's name. The following guide summarizes the arguments established in this lecture.

Structure	What to control	Method
Independent indicators	The sum's mean and the full multiplicative exponent; retain it for large relative deviations.	Chernoff (Section 2)
Independent bounded summands	The sum of squared range widths.	Hoeffding (Section 2.4)
A function of independent choices	The effect of changing each coordinate; choose useful blocks when possible.	McDiarmid (Section 3)
Conditionally centered increments	Their conditional range widths, or their absolute bounds.	Martingale concentration (Section 5)
Loads with a fixed total	Independent Poisson loads, followed by a justified comparison to the fixed-total model.	Poissonization (Section 6)

Match the scale to the question. An additive window need not be small relative to the mean. For a very rare pattern or a small expected empty-bin count, a valid bounded-differences estimate may still be too coarse. Conversely, a bound around $\mathbb{E}\chi(G)$ or $\mathbb{E}\|\sum_i \varepsilon_i u_i\|$ need not locate that expectation. Concentration and estimation of the center are separate tasks.

Account for every source of error. A one-sided bound has no extra factor 2; a two-sided statement follows by adding the two tails. A simultaneous guarantee requires a union bound over the objects being controlled. In a Bloom filter, the occupancy fluctuation and the exceptional probability are separate terms. In de-Poissonization, an exact conditional identity does not itself justify replacing a conditional probability by an unconditional one.

The common proof pattern

Identify a bad event, represent the randomness in a useful way, and control an exponential moment or an independent product. Preserve the hypotheses and constants when translating back to the original statistic. The most useful refinement is often a better representation, not a more general inequality.

8 Exercises

These exercises follow the order of the exposition: first use the concentration bounds, then examine their martingale proof and its hypotheses.

Independent sums and amplification

1. **Tune the failure probability of the maximum load.** For $m = n$ and fixed $\alpha > 0$, show that any fixed $K > \alpha + 1$ gives $\mathbb{P}(M \geq \lceil K \ln n / \ln \ln n \rceil) \leq n^{-\alpha}$ for sufficiently large n. Identify where the additional 1 in $\alpha + 1$ comes from.

   Hint. Take logarithms in $n(e/L)^L$.

2. **Majority and median are not the same estimator.** An algorithm is correct with probability at least $2/3$. How many independent runs suffice to reduce the error to 10^{-6} by the bounds of this lecture? Give an example of numerical outputs showing why their arithmetic mean need not stay in the target interval even when a strict majority does.

   Hint. Use $p = 1/6$, choose an odd repetition count, and allow one arbitrarily large outlier.

Bounded differences and Bloom filters

3. **The exact coordinate budget for pattern matching.** For each position i, count the number r_i of length-k windows containing it. Replace the denominator nk^2 in (20) by $\sum_i r_i^2$. Discuss why even this improved additive estimate may be weak when the pattern is very rare.

   Hint. The start position lies in $[1, n-k+1] \cap [i-k+1, i]$.

4. **Edges versus vertex blocks.** Apply McDiarmid with one coordinate per edge, then compare its chromatic-number fluctuation scale with the vertex-block bound (23). Verify that two graphs differing only in edges incident to a single vertex have chromatic numbers differing by at most one.

   Hint. Delete that vertex in both graphs before comparing their colorings.

5. **A finite Bloom-filter calculation.** Set $m = 10n$ and $k = 7$. Write down the finite upper bound (27) without replacing $(1 - 1/m)^{kn}$ by an exponential. Identify separately the occupancy fluctuation, the Poisson-type approximation error, and the exceptional probability.

   Hint. The first two errors are different; only the last is n^{-100}.

Martingales and the proof of bounded differences

6. **Conditional range width.** In Theorem 5.4, explain why the interval endpoints may depend on the past, but must be known before the next increment. For $B \sim \text{Bernoulli}(p)$ with $0 < p < 1$, compute the exact range of $B - p$ and compare its width with twice its largest absolute value. Use the range-width bound to recover Hoeffding's exponent for a sum of independent bounded variables.

7. **Why independence cannot be dropped.** Let B be a fair bit and $X_1 = \cdots = X_n = B$. The function $f(x) = \sum_i x_i$ is 1-Lipschitz in each coordinate. Compute the first nontrivial Doob increment and the probability $\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq n/2)$. Explain exactly which step of the independent-input proof fails.

   Hint. Revealing one coordinate reveals all of them.

Poissonization and occupancy thresholds

8. **A high-confidence collection time.** Given $0 < \delta < 1$, find an explicit integer number of throws that guarantees coverage with probability at least $1 - \delta$. Compare its order with $\mathbb{E}T$. Explain why this one-sided guarantee does not determine the limiting distribution of $(T - n \ln n)/n$.

   Hint. Solve $ne^{-m/n} \leq \delta$ and use the correct event $T > m$.

9. **A simpler lower-load constant.** Use $L = \lfloor \ln n/(2 \ln\ln n) \rfloor$ and $L! \le L^L$ to show that the Poissonized probability of all loads being below L is at most $\exp(-\sqrt{n}/e)$ for sufficiently large n. Transfer this to the fixed-total experiment.

10. **What de-Poissonization actually needs.** Show that any integer window $a_n = o(n)$ with $\mathbb{P}(|S - m| > a_n) = o(1)$ suffices for the coupon-collector limit. Explain why the generic $e\sqrt{m}$ comparison alone does not recover that limit, and why "a distinguished empty bin" cannot be replaced by "any empty bin" in the slow-change proof.

A Elementary probability and analytic estimates

This appendix supplies the elementary estimates used in the main arguments.

A.1 Geometric waiting times and factorial estimates

If G counts independent Bernoulli trials of success probability $p > 0$ up to the first success, then $\mathbb{P}(G \ge r) = (1-p)^{r-1}$ for $r \ge 1$. Since $G = \sum_{r \ge 1} \mathbf{1}\{G \ge r\}$, taking expectations of this nonnegative sum gives

$$\mathbb{E}G = \sum_{r \ge 1} \mathbb{P}(G \ge r) = \sum_{r \ge 1} (1-p)^{r-1} = 1/p.$$

For every integer $r \ge 1$, the factorial estimates needed here are

$$(r/e)^r \le r! \le e\sqrt{r}\,(r/e)^r. \tag{63}$$

For the lower bound, monotonicity of $\ln x$ gives $\ln r! \ge \int_1^r \ln x\,dx = r\ln r - r + 1 \ge r \ln r - r$. For the upper bound, concavity of $\ln x$ places its graph above each chord, so

$$\int_1^r \ln x\,dx \ge \sum_{j=1}^{r-1} \frac{\ln j + \ln(j+1)}{2} = \ln r! - \tfrac{1}{2}\ln r.$$

Rearrange and exponentiate. These elementary inequalities suffice for the $e\sqrt{m}$ comparison; the full Stirling asymptotic formula is unnecessary. Also, $r! \le r^r$ is often a useful simpler upper bound.

A.2 Simplifying the Chernoff exponents

Define $h_+(\delta) = (1+\delta)\ln(1+\delta) - \delta$ and $h_-(\delta) = \delta + (1-\delta)\ln(1-\delta)$. The exact upper and lower bounds are $e^{-\mu h_+(\delta)}$ and $e^{-\mu h_-(\delta)}$.

For $x \ge 0$, $\ln(1+x) \ge 2x/(2+x)$: the difference is zero at 0 and has derivative $x^2/((1+x)(2+x)^2) \ge 0$. In particular, for $0 \le x \le 1$, $\ln(1+x) \ge 2x/3$. Since $h_+(0) = 0$ and $h'_+(x) = \ln(1+x)$,

$$h_+(\delta) \ge \int_0^\delta \frac{2x}{3}\,dx = \frac{\delta^2}{3} \qquad (0 \le \delta \le 1).$$

Similarly, $h'_-(x) = -\ln(1-x) \ge x$ for $0 \le x < 1$, so

$$h_-(\delta) \geq \int_0^\delta x\,dx = \frac{\delta^2}{2} \qquad (0 \leq \delta < 1).$$

These are the inequalities used in both the Bernoulli and Poisson small-deviation bounds. The endpoint $\delta = 1$ is handled by continuity or a direct zero-count calculation.

A.3 The one-variable calculation behind Bloom's relaxed optimum

To complete the minimization in Section 4, let $R(u) = \ln u \ln(1-u)$ for $0 < u < 1$. Then

$$R'(u) = \frac{(1-u)\ln(1-u) - u\ln u}{u(1-u)}.$$

The numerator $H(u)$ has $H(0+) = H(1/2) = 0$ and $H''(u) = (2u-1)/(u(1-u)) < 0$ on $(0, 1/2)$. It is strictly concave there and hence positive between its endpoint zeros. Thus R increases up to $1/2$; symmetry $R(u) = R(1-u)$ makes it decrease afterwards. Its unique maximum is $(\ln 2)^2$. Because $\ln F_c(k) = -cR(e^{-k/c})$, the relaxed minimizer is $k = c\ln 2$.

B Negative association and exponential-moment bounds

The independent-sum proofs use a product rule for exponential moments. Equality is not necessary: an upper bound by the product of the marginal moments suffices. Negative association is a joint condition that supplies this inequality.

Definition B.1 (Negative association)**.** A family $(X_i)_{i\in I}$ is negatively associated if the following inequality holds for all disjoint subsets $A, B \subseteq I$ and coordinatewise nondecreasing functions f, g with finite relevant expectations:

$$\mathbb{E}[f(X_A)g(X_B)] \leq \mathbb{E}f(X_A)\,\mathbb{E}g(X_B), \tag{64}$$

The comparison concerns increasing functions of disjoint coordinate groups.

For a negatively associated family with the required exponential moments finite,

$$\mathbb{E}\prod_{i=1}^N e^{\lambda X_i} \leq \prod_{i=1}^N \mathbb{E}e^{\lambda X_i} \qquad (\lambda \in \mathbb{R}). \tag{65}$$

The same covariance inequality holds when both functions are nonincreasing: apply the definition to $-f$ and $-g$. If f_i are nonnegative functions of single coordinates and are all nondecreasing, then the product of any initial group is itself nondecreasing. Induction on the number of factors yields

$$\mathbb{E}\prod_i f_i(X_i) \leq \prod_i \mathbb{E}f_i(X_i).$$

The same induction works when all factors are nonincreasing. Taking $f_i(x) = e^{\lambda x}$ proves (65): the factors increase together for $\lambda > 0$ and decrease together for $\lambda < 0$. For unbounded factors, apply the bounded argument to nonnegative truncations and then pass to the limit.

A useful additional check separates this condition from pairwise independence. Choose uniformly from the four bit vectors $001, 010, 100, 111$. Every pair of coordinates is independent and each bit is fair, but $\mathbb{P}(X_1 = X_2 = X_3 = 1) = 1/4 > 1/8 = \prod_i \mathbb{P}(X_i = 1)$. Thus even a product of three increasing single-coordinate functions need not obey the product inequality under pairwise independence alone.

Extensions of the independent-sum bounds. Equation (65) replaces the independent product identity in each proof of Section 2. Consequently, the multiplicative Chernoff bounds hold for negatively associated Bernoulli variables; Hoeffding's inequality holds for negatively associated bounded variables with the same range widths; and a sum of centered negatively associated sub-Gaussian variables with variance factors ν_i has variance factor $\sum_i \nu_i$. Centering each coordinate preserves negative association, since translating the arguments preserves monotonicity.

These extensions concern sums and their product inequalities. They do not remove the independent-coordinate assumption from McDiarmid's theorem. Merely nonpositive pairwise covariances, or pairwise independence, do not establish the joint product inequality. Unless negative association has actually been proved for the family at hand, the independence hypotheses used in the algorithmic applications remain in force.

Bibliographic notes

Hoeffding [1] develops exponential bounds for sums of bounded variables; McDiarmid [2] presents the bounded-differences method for functions of independent inputs. Azuma [3] treats concentration for bounded martingale differences. The conditional-range argument explains how these inequalities fit into the same exponential-moment framework.

Bloom [4] introduced the approximate-membership data structure. Bose et al. [5] analyze its finite false-positive probability and the distinction between conditional and unconditional independence. Shamir and Spencer [6] use martingale methods to study concentration of the chromatic number of random graphs.

References

[1] Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi:10.1080/01621459.1963.10500830.

[2] Colin McDiarmid. On the Method of Bounded Differences. In J. Siemons (ed.), *Surveys in Combinatorics, 1989*, pages 148–188. Cambridge University Press, 1989. doi:10.1017/CBO9781107359949.008.

[3] Kazuoki Azuma. Weighted Sums of Certain Dependent Random Variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967. doi:10.2748/tmj/1178243286.

[4] Burton H. Bloom. Space/Time Trade-offs in Hash Coding with Allowable Errors. *Communications of the ACM*, 13(7):422–426, 1970. doi:10.1145/362686.362692.

[5] Prosenjit Bose, Hua Guo, Evangelos Kranakis, Anil Maheshwari, Pat Morin, Jason Morrison, Michiel Smid, and Yihui Tang. On the False-Positive Rate of Bloom Filters. *Information Processing Letters*, 108(4):210–213, 2008. doi:10.1016/j.ipl.2008.05.018.

[6] Eli Shamir and Joel Spencer. Sharp Concentration of the Chromatic Number on Random Graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987. doi:10.1007/BF02579208.