and

$$\mathbf{Cov}(\mathbf{I}(p|x), \mathbf{I}(q|x)) = \frac{1}{pq} + O\left(\frac{1}{n}\right) - \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right)\left(\frac{1}{q} + O\left(\frac{1}{n}\right)\right) = O\left(\frac{1}{n}\right).$$

We thus conclude that

$$\mathbf{E}(|B|) = \sum_{p \leq n^{1/10}} \frac{1}{p} + O\left(n^{-9/10}\right)$$

and

$$\mathbf{Var}(|B|) = \sum_{p \leq n^{1/10}} \left(\frac{1}{p} - \frac{1}{p^2}\right) + O\left(n^{-8/10}\right).$$

The expectation and variance estimates now follow from Mertens' theorem (see Proposition 1.51) and the convergence of the sum $\sum_k \frac{1}{k^2}$. □

Exercises

1.2.1 When does equality hold in Chebyshev's inequality?

1.2.2 If X and Y are two random variables, verify the *Cauchy–Schwarz inequality* $|\mathbf{Cov}(X, Y)| \leq \mathbf{Var}(X)^{1/2}\mathbf{Var}(Y)^{1/2}$ and the *triangle inequality* $\mathbf{Var}(X + Y)^{1/2} \leq \mathbf{Var}(X)^{1/2} + \mathbf{Var}(Y)^{1/2}$. When does equality occur?

1.2.3 Prove (1.10).

1.2.4 If $\phi : \mathbf{R} \to \mathbf{R}$ is a convex function and X is a random variable, verify *Jensen's inequality* $\mathbf{E}(\phi(X)) \leq \phi(\mathbf{E}(X))$. If ϕ is strictly convex, when does equality occur?

1.2.5 Generalize Chebyshev's inequality using higher moments $\mathbf{E}(|X - \mathbf{E}(X)|^p)$ instead of the variance.

1.2.6 By obtaining an upper bound on the fourth moment, improve Theorem 1.6 to

$$\frac{1}{N}|\{x \in [1, N] : |\nu(x) - \log\log N| > K\sqrt{\log\log N}\}| = O(K^{-4}).$$

Can you generalize this to obtain a bound of $O_m(K^{-m})$ for any even integer $m \geq 2$, where the constant in the $O()$ notation is allowed to depend on m?

1.3 The exponential moment method

Chebyshev's inequality shows that if one has control of the second moment $\mathbf{Var}(X) = \mathbf{E}(|X - \mathbf{E}(X)|^2)$, then a random variable X takes the value $\mathbf{E}(X) + O(\lambda \mathbf{Var}(X)^{1/2})$ with probability $1 - O(\lambda^{-2})$. If one uses higher moments, one

can obtain better decay of the tail probability than $O(\lambda^{-2})$. In particular, if one can control *exponential* moments[1] such as $\mathbf{E}(e^{tX})$ for some real parameter t, then one can obtain exponential decay in upper and lower tail probabilities, since Markov's inequality yields

$$\mathbf{P}(X \geq \lambda) = \mathbf{P}(e^{tX} \geq e^{t\lambda}) \leq \frac{\mathbf{E}(e^{tX})}{e^{t\lambda}} \tag{1.15}$$

for $t > 0$ and $\lambda \in \mathbf{R}$, and similarly

$$\mathbf{P}(X \leq -\lambda) = \mathbf{P}(e^{-tX} \geq e^{t\lambda}) \leq \frac{\mathbf{E}(e^{-tX})}{e^{t\lambda}} \tag{1.16}$$

for the same range of t, λ. The quantity $\mathbf{E}(e^{tX})$ is known as an *exponential moment* of X, and the function $t \mapsto \mathbf{E}(e^{tX})$ is known as the *moment generating function*, thanks to the Taylor expansion

$$\mathbf{E}(e^{tX}) = 1 + t\mathbf{E}(X) + \frac{t^2}{2!}\mathbf{E}(X^2) + \frac{t^3}{3!}\mathbf{E}(X^3) + \cdots.$$

The application of (1.15) or (1.16) is known as the *exponential moment method*. Of course, to use it effectively one needs to be able to compute the exponential moments $\mathbf{E}(e^{tX})$. A preliminary tool for doing so is

Lemma 1.7 *Let X be a random variable with $|X| \leq 1$ and $\mathbf{E}(X) = 0$. Then for any $-1 \leq t \leq 1$ we have $\mathbf{E}(e^{tX}) \leq \exp(t^2\mathbf{Var}(X))$.*

Proof Since $|tX| \leq 1$, a simple comparison of Taylor series gives the inequality

$$e^{tX} \leq 1 + tX + t^2X^2.$$

Taking expectations of both sides and using linearity of expectation and the hypothesis $\mathbf{E}(X) = 0$ we obtain

$$\mathbf{E}(e^{tX}) \leq 1 + t^2\mathbf{Var}(X) \leq \exp(t^2\mathbf{Var}(X))$$

as desired. □

This lemma by itself is not terribly effective as it requires both X and t to be bounded. However the power of this lemma can be amplified considerably when applied to random variables X which are *sums* of bounded random variables, $X = X_1 + \cdots + X_n$, provided that we have the very strong assumption of *joint* independence between the $X_1, \ldots, X_n$. More precisely, we have

[1] To avoid questions of integrability or measurability, let us assume for sake of discussion that the random variable X here only takes finitely many values; this is the case of importance in combinatorial applications.

Theorem 1.8 (Chernoff's inequality) *Assume that $X_1, \ldots, X_n$ are jointly independent random variables where $|X_i - \mathbf{E}(X_i)| \leq 1$ for all i. Set $X := X_1 + \cdots + X_n$ and let $\sigma := \sqrt{\mathbf{Var}(X)}$ be the standard deviation of X. Then for any $\lambda > 0$*

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq \lambda\sigma) \leq 2\max\left(e^{-\lambda^2/4}, e^{-\lambda\sigma/2}\right). \tag{1.17}$$

Informally speaking, (1.17) asserts that $X = \mathbf{E}(X) + O(\mathbf{Var}(X)^{1/2})$ with high probability, and $X = \mathbf{E}(X) + O(\ln^{1/2} n\mathbf{Var}(X)^{1/2})$ with extremely high probability ($1 - O(n^{-C})$ for some large C). The bound in Chernoff's theorem provides a huge improvement over Chebyshev's inequality when λ is large. However the joint independence of the X_i is essential (Exercise 1.3.8). Later on we shall develop several variants of Chernoff's inequality in which there is some limited interaction between the X_i.

Proof By subtracting a constant from each of the X_i we may normalize $\mathbf{E}(X_i) = 0$ for each i. Observe that $\mathbf{P}(|X| \geq \lambda\sigma) = \mathbf{P}(X \geq \lambda\sigma) + \mathbf{P}(X \leq -\lambda\sigma)$. By symmetry, it thus suffices to prove that

$$\mathbf{P}(X \geq \lambda\sigma) \leq e^{-t\lambda\sigma/2} \tag{1.18}$$

where $t := \min(\lambda/2\sigma, 1)$.

Applying (1.15) we have

$$\mathbf{P}(X \geq \lambda\sigma) \leq e^{-t\lambda\sigma}\mathbf{E}\left(e^{tX_1} \cdots e^{tX_n}\right).$$

Since the X_i are jointly independent, so are the e^{tX_i}. Using this and Lemma 1.7 we obtain

$$\mathbf{E}\left(e^{tX_1} \cdots e^{tX_n}\right) = \mathbf{E}\left(e^{tX_1}\right) \cdots \mathbf{E}\left(e^{tX_n}\right) \leq \exp(t^2\mathbf{Var}(X_1)) \cdots \exp(t^2\mathbf{Var}(X_n)).$$

On the other hand, from (1.9) we have

$$\mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n) = \sigma^2.$$

Putting all this together, we obtain

$$\mathbf{P}(X \geq \lambda\sigma) \leq e^{-t\lambda\sigma} e^{t^2\sigma^2}.$$

Since $t \leq \lambda/2\sigma$, the claim follows. □

Now let us consider a special, but important case when X_is are independent *boolean* (or *Bernoulli*) variables.

Corollary 1.9 *Let $X = t_1 + \cdots + t_n$ where the t_i are independent boolean random variables. Then for any $\epsilon > 0$*

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq \epsilon\mathbf{E}(X)) \leq 2e^{-\min(\epsilon^2/4, \epsilon/2)\mathbf{E}(X)}. \tag{1.19}$$