2.4. Some probability estimates

Lemma 2.10 (Borel Cantelli lemmas). *Let A_n be events on a common probability space.*

(1) *If $\sum_n \mathbf{P}(A_n) < \infty$, then $\mathbf{P}(A_n \text{ i.o}) = 0$.*
(2) *If A_n are independent and $\sum_n \mathbf{P}(A_n) = \infty$, then $\mathbf{P}(A_n \text{ i.o}) = 1$.*

PROOF. (1) For any N, $\mathbf{P}\left(\cup_{n=N}^{\infty} A_n\right) \leq \sum_{n=N}^{\infty} \mathbf{P}(A_n)$ which goes to zero as $N \to \infty$. Hence $\mathbf{P}(\limsup A_n) = 0$.

(2) For any $N < M$, $\mathbf{P}(\cup_{n=N}^{M} A_n) = 1 - \prod_{n=N}^{M} \mathbf{P}(A_n^c)$. Since $\sum_n \mathbf{P}(A_n) = \infty$, it follows that $\prod_{n=N}^{M}(1 - \mathbf{P}(A_n)) \leq \prod_{n=N}^{M} e^{-\mathbf{P}(A_n)} \to 0$, for any fixed N as $M \to \infty$. Hence $\mathbf{P}\left(\cup_{n=N}^{\infty} A_n\right) = 1$ for all N, implying that $\mathbf{P}(A_n \text{ i.o}) = 1$. ■

Lemma 2.11 (First and second moment methods). *Let $X \geq 0$ be a r.v.*

(1) *(**Markov's inequality a.k.a first moment method**) For any $t > 0$, we have $\mathbf{P}(X \geq t) \leq t^{-1}\mathbf{E}[X]$.*
(2) *(**Paley-Zygmund inequality a.k.a second moment method**) For any non-negative r.v. X,*

$$(i)\, \mathbf{P}(X > 0) \geq \frac{\mathbf{E}[X]^2}{\mathbf{E}[X^2]}. \qquad (ii)\, \mathbf{P}(X > \alpha \mathbf{E}[X]) \geq (1-\alpha)^2 \frac{\mathbf{E}[X]^2}{\mathbf{E}[X^2]}.$$

PROOF. (1) $t\mathbf{1}_{X \geq t} \leq X$. Positivity of expectations gives the inequality.

(2) $\mathbf{E}[X]^2 = \mathbf{E}[X\mathbf{1}_{X>0}]^2 \leq \mathbf{E}[X^2]\mathbf{E}[\mathbf{1}_{X>0}] = \mathbf{E}[X^2]\mathbf{P}(X > 0)$. Hence the first inequality follows. The second inequality is similar. Let $\mu = \mathbf{E}[X]$. By Cauchy-Schwarz, we have $\mathbf{E}[X\mathbf{1}_{X>\alpha\mu}]^2 \leq \mathbf{E}[X^2]\mathbf{P}(X > \alpha\mu)$. Further, $\mu = \mathbf{E}[X\mathbf{1}_{X<\alpha\mu}] + \mathbf{E}[X\mathbf{1}_{X>\alpha\mu}] \leq \alpha\mu + \mathbf{E}[X\mathbf{1}_{X>\alpha\mu}]$, whence, $\mathbf{E}[X\mathbf{1}_{X>\alpha\mu}] \geq (1-\alpha)\mu$. Thus,

$$\mathbf{P}(X > \alpha\mu) \geq \frac{\mathbf{E}[X\mathbf{1}_{X>\alpha\mu}]^2}{\mathbf{E}[X^2]} \geq (1-\alpha)^2 \frac{\mathbf{E}[X]^2}{\mathbf{E}[X^2]}.$$ ■

Remark 2.12. Applying these inequalities to other functions of X can give more information. For example, if X has finite variance, $\mathbf{P}(|X - \mathbf{E}[X]| \geq t) = \mathbf{P}(|X - \mathbf{E}[X]|^2 \geq t^2) \leq t^{-2}\text{Var}(X)$, which is called *Chebyshev's inequality*. Higher the moments that exist, better the asymptotic tail bounds that we get. For example, if $\mathbf{E}[e^{\lambda X}] < \infty$ for some $\lambda > 0$, we get exponential tail bounds by $\mathbf{P}(X > t) = \mathbf{P}(e^{\lambda X} < e^{\lambda t}) \leq e^{-\lambda t}\mathbf{E}[e^{\lambda X}]$.

2.5. Applications of first and second moment methods

The first and second moment methods are immensely useful. This is somewhat surprising, given the very elementary nature of these inequalities, but the following applications illustrate the ease with which they give interesting results.

Application 1: Borel-Cantelli lemmas: The first B-C lemma follows from Markov's inequality. In fact, applied to $X = \sum_{k=N}^{\infty} \mathbf{1}_{A_k}$, Markov's inequality is the same as the union bound $\mathbf{P}(\cup_{k=N}^{\infty} A_k) \leq \sum_{k=N}^{\infty} \mathbf{P}(A_k)$ which is what gave us the first Borel-Cantelli.

The second is more interesting. Suppose Fix $n < m$, define $X = \sum_{k=n}^{m} \mathbf{1}_{A_k}$. Then $\mathbf{E}[X] = \sum_{k=n}^{m} \mathbf{P}(A_k)$. Also,

$$\begin{aligned} \mathbf{E}[X^2] &= \mathbf{E}\left[\sum_{k=n}^{m}\sum_{\ell=n}^{m} \mathbf{1}_{A_k}\mathbf{1}_{A_\ell}\right] = \sum_{k=n}^{m} \mathbf{P}(A_k) + \sum_{k\neq\ell} \mathbf{P}(A_k)\mathbf{P}(A_\ell) \\ &\leq \left(\sum_{k=n}^{m} \mathbf{P}(A_k)\right)^2 + \sum_{k=n}^{m} \mathbf{P}(A_k). \end{aligned}$$

Apply the second moment method to se that for any fixed n, as $m \to \infty$,

$$\mathbf{P}(X \geq 1) \geq \frac{\left(\sum_{k=n}^{m} \mathbf{P}(A_k)\right)^2}{\left(\sum_{k=n}^{m} \mathbf{P}(A_k)\right)^2 + \sum_{k=n}^{m} \mathbf{P}(A_k)} = \frac{1}{1 + \left(\sum_{k=n}^{m} \mathbf{P}(A_k)\right)^{-1}} \to 1,$$

by assumption that $\sum \mathbf{P}(A_k) = \infty$. This shows that $\mathbf{P}(\cup_{k\geq n} A_k) = 1$ for any n and hence $\mathbf{P}(\limsup A_n) = 1$.

Application 2: Coupon collector problem: A bookshelf has (large number) n books numbered $1, 2, \ldots, n$. Every night, before going to bed, you pick one of the books at random to read. The book is replaced in the shelf in the morning. How many days pass before you have picked up each of the books at least once?

Theorem 2.13. *Let T_n denote the number of days till each book is picked at least once. Then T_n is "concentrated around $n \log n$ in a window of size n" by which we mean that for any sequence $\theta_n \to \infty$, we have*

$$\mathbf{P}(|T_n - n\log n| < n\theta_n) \to 1.$$

Remark 2.14. In the following proof and many other places, we shall have occasion to make use of the elementary estimate

$$1 - x \leq e^{-x} \quad \forall x, \qquad 1 - x \geq e^{-x-x^2} \quad \forall |x| < \frac{1}{2}. \tag{2.1}$$

The first inequality follows by expanding e^{-x} while the second follows by expanding $\log(1-x) = -x - x^2/2 - x^3/3 - \ldots$ (valid for $|x| < 1$).

PROOF. Fix an integer $t \geq 1$ and let $X_{t,k}$ be the indicator that the k^{th} book is not picked up on the first t days. Then, $\mathbf{P}(T_n > t) = \mathbf{P}(S_{t,n} \geq 1)$ where $S_{t,n} = X_{t,1} + \ldots + X_{t,n}$. As $\mathbf{E}[X_{t,k}] = (1 - 1/n)^k$ and $\mathbf{E}[X_{t,k}X_{t,\ell}] = (1-2/n)^k$ for $k \neq \ell$, we also compute that thefirst two moments of $S_{t,n}$ and use (2.1) to get

$$n e^{-\frac{t}{n} - \frac{t}{n^2}} \leq \mathbf{E}[S_{t,n}] = n\left(1 - \frac{1}{n}\right)^t \leq n e^{-\frac{t}{n}}. \tag{2.2}$$

$$\mathbf{E}[S_{t,n}^2] = n\left(1 - \frac{1}{n}\right)^t + n(n-1)\left(1 - \frac{2}{n}\right)^t \leq n e^{-\frac{t}{n}} + n(n-1)e^{-\frac{2t}{n}}. \tag{2.3}$$

The left inequality on the first line is valid only for $n \geq 2$ which we assume.

Now set $t = n\log n + n\theta_n$ and apply Markov's inequality to get

$$\mathbf{P}(T_n > n\log n + n\theta_n) = \mathbf{P}(S_{t,n} \geq 1) \leq \mathbf{E}[S_{t,n}] \leq n e^{-\frac{n\log n + n\theta_n}{n}} \leq e^{-\theta_n} = o(1). \tag{2.4}$$

On the other hand, taking $t < n\log n - n\theta_n$ (where we take $\theta_n < \log n$, of course!), we now apply the second moment method. For any $n \geq 2$, by using (2.3) we get

$\mathbf{E}[S_{t,n}^2] \le e^{\theta_n} + e^{2\theta_n}$. The first inequality in (2.2) gives $\mathbf{E}[S_{t,n}] \ge e^{\theta_n - \frac{\log n - \theta_n}{n}}$. Thus,

$$\mathbf{P}(T_n > n\log n - n\theta_n) = \mathbf{P}(S_{t,n} \ge 1) \ge \frac{\mathbf{E}[S_{t,n}]^2}{\mathbf{E}[S_{t,n}^2]} \ge \frac{e^{2\theta_n - 2\frac{\log n - \theta_n}{n}}}{e^{\theta_n} + e^{2\theta_n}} = 1 - o(1) \tag{2.5}$$

as $n \to \infty$. From (2.4) and (2.5), we get the sharp bounds

$$\mathbf{P}(|T_n - n\log(n)| > n\theta_n) \to 0 \text{ for any } \theta_n \to \infty. \qquad \blacksquare$$

Application 3: Branching processes: Consider a Galton-Watson branching process with offsprings that are i.i.d ξ. Let Z_n be the number of offsprings in the n^{th} generation. Take $Z_0 = 1$.

Theorem 2.15. (1) *If $m < 1$, then w.p.1, the branching process dies out. That is $\mathbf{P}(Z_n = 0$ for all large $n) = 1$.*

(2) *If $m > 1$, then with positive probability, the branching process survives. That is $\mathbf{P}(Z_n \ge 1$ for all $n) > 0$.*

PROOF. The proof uses elementary conditioning concepts. By conditioning on what happens in the $(n-1)^{\text{st}}$ generation, we write Z_n as a sum of Z_{n-1} independent copies of ξ. From this, one can compute that $\mathbf{E}[Z_n|Z_{n-1}] = mZ_{n-1}$ and if we assume that ξ has variance σ^2 we also get $\mathrm{Var}(Z_n|Z_{n-1}) = Z_{n-1}\sigma^2$. Therefore, $\mathbf{E}[Z_n] = \mathbf{E}[\mathbf{E}[Z_n|Z_{n-1}]] = m\mathbf{E}[Z_{n-1}]$ from which we get $\mathbf{E}[Z_n] = m^n$. Similarly, from the formula $\mathrm{Var}(Z_n) = \mathbf{E}[\mathrm{Var}(Z_n|Z_{n-1})] + \mathrm{Var}(\mathbf{E}[Z_n|Z_{n-1}])$ we can compute that

$$\begin{aligned}
\mathrm{Var}(Z_n) &= m^{n-1}\sigma^2 + m^2\mathrm{Var}(Z_{n-1}) \\
&= \left(m^{n-1} + m^n + \ldots + m^{2n-1}\right)\sigma^2 \qquad \text{(by repeating the argument)} \\
&= \sigma^2 m^{n-1}\frac{m^{n+1}-1}{m-1}.
\end{aligned}$$

(1) By Markov's inequality, $\mathbf{P}(Z_n > 0) \le \mathbf{E}[Z_n] = m^n \to 0$. Since the events $\{Z_n > 0\}$ are decreasing, it follows that $\mathbf{P}(\text{extinction}) = 1$.

(2) If $m = \mathbf{E}[\xi] > 1$, then as before $\mathbf{E}[Z_n] = m^n$ which increases exponentially. But that is not enough to guarantee survival. Assuming that ξ has finite variance σ^2, apply the second moment method to write

$$\mathbf{P}(Z_n > 0) \ge \frac{\mathbf{E}[Z_n]^2}{\mathrm{Var}(Z_n) + \mathbf{E}[Z_n]^2} \ge \frac{1}{1 + \frac{\sigma^2}{m-1}}$$

which is a positive number (independent of n). Again, since $\{Z_n > 0\}$ are decreasing events, we get $\mathbf{P}(\text{non-extinction}) > 0$.

The assumption of finite variance of ξ can be removed as follows. Since $\mathbf{E}[\xi] = m > 1$, we can find A large so that setting $\eta = \min\{\xi, A\}$, we still have $\mathbf{E}[\eta] > 1$. Clearly, η has finite variance. Therefore, the branching process with η offspring distribution survives with positive probability. Then, the original branching process must also survive with positive probability! (A coupling argument is the best way to deduce the last statement: Run the original branching process and kill every child after the first A. If inspite of the violence the population survives, then ...) $\blacksquare$

Remark 2.16. The fundamental result of branching processes also asserts the a.s extinction for the critical case $m = 1$. We omit this for now.

Application 4: How many prime divisors does a number typically have? For a natural number k, let $\nu(k)$ be the number of (distinct) prime divisors of n. What is the typical size of $\nu(n)$ as compared to n? We have to add the word typical, because if p is a prime number then $\nu(p) = 1$ whereas $\nu(2 \times 3 \times \ldots \times p) = p$. Thus there are arbitrarily large numbers with $\nu = 1$ and also numbers for which ν is as large as we wish. To give meaning to "typical", we draw a number at random and look at its ν-value. As there is no natural way to pick one number at random, the usual way of making precise what we mean by a "typical number" is as follows.

Formulation: Fix $n \geq 1$ and let $[n] := \{1, 2, \ldots, n\}$. Let μ_n be the uniform probability measure on $[n]$, i.e., $\mu_n\{k\} = 1/n$ for all $k \in [n]$. Then, the function $\nu : [n] \to \mathbb{R}$ can be considered a random variable, and we can ask about the behaviour of these random variables. Below, we write $\mathbf{E}_n$ to denote expectation w.r.t μ_n.

Theorem 2.17 (Hardy, Ramanujan). *With the above setting, for any $\delta > 0$, as $n \to \infty$ we have*

$$\mu_n \left\{ k \in [n] : \left| \frac{\nu(k)}{\log\log n} - 1 \right| > \delta \right\} \to 0. \tag{2.6}$$

PROOF. (**Turan**). Fix n and for any prime p define $X_p : [n] \to \mathbb{R}$ by $X_p(k) = \mathbf{1}_{p|k}$. Then, $\nu(k) = \sum_{p \leq k} X_p(k)$. We define $\psi(k) := \sum_{p \leq \sqrt[4]{k}} X_p(k)$. Then, $\psi(k) \leq \nu(k) \leq \psi(k) + 4$ since there can be at most four primes larger than $\sqrt[4]{k}$ that divide k. From this, it is clearly enough to show (2.6) for ψ in place of ν (why?).

We shall need the first two moments of ψ under μ_n. For this we first note that $\mathbf{E}_n[X_p] = \frac{\left\lfloor \frac{n}{p} \right\rfloor}{n}$ and $\mathbf{E}_n[X_p X_q] = \frac{\left\lfloor \frac{n}{pq} \right\rfloor}{n}$. Observe that $\frac{1}{p} - \frac{1}{n} \leq \frac{\left\lfloor \frac{n}{p} \right\rfloor}{n} \leq \frac{1}{p}$ and $\frac{1}{pq} - \frac{1}{n} \leq \frac{\left\lfloor \frac{n}{pq} \right\rfloor}{n} \leq \frac{1}{pq}$.

By linearity $\mathbf{E}_n[\psi] = \sum_{p \leq \sqrt[4]{n}} \mathbf{E}[X_p] = \sum_{p \leq \sqrt[4]{n}} \frac{1}{p} + O(n^{-\frac{3}{4}})$. Similarly

$$\begin{aligned} \mathrm{Var}_n[\psi] &= \sum_{p \leq \sqrt[4]{n}} \mathrm{Var}[X_p] + \sum_{p \neq q \leq \sqrt[4]{n}} \mathrm{Cov}(X_p, X_q) \\ &= \sum_{p \leq \sqrt[4]{n}} \left(\frac{1}{p} - \frac{1}{p^2} + O(n^{-1}) \right) + \sum_{p \neq q \leq \sqrt[4]{n}} O(n^{-1}) \\ &= \sum_{p \leq \sqrt[4]{n}} \frac{1}{p} - \sum_{p \leq \sqrt[4]{n}} \frac{1}{p^2} + O(n^{-\frac{1}{2}}). \end{aligned}$$

We make use of the following two facts. $\sum_{p \leq \sqrt[4]{n}} \frac{1}{p} \sim \log\log n$ and $\sum_{p=1}^{\infty} \frac{1}{p^2} < \infty$. The second one is obvious, while the first one is not hard and we leave it as exercise. Thus, we get $\mathbf{E}_n[\psi] = \log\log n + O(n^{-\frac{3}{4}})$ and $\mathrm{Var}_n[\psi] = \log\log n + O(1)$. Thus, by Chebyshev's inequality,

$$\mu_n \left\{ k \in [n] : \left| \frac{\psi(k) - \mathbf{E}_n[\psi]}{\log\log n} \right| > \delta \right\} \leq \frac{\mathrm{Var}_n(\psi)}{\delta^2 (\log\log n)^2} = O\left(\frac{1}{\log\log n} \right).$$

From the asymptotics $\mathbf{E}_n[\psi] = \log\log n + O(n^{-\frac{3}{4}})$ we also get (for n large enough)

$$\mu_n \left\{ k \in [n] : \left| \frac{\psi(k)}{\log\log n} - 1 \right| > \delta \right\} \leq \frac{\mathrm{Var}_n(\psi)}{\delta^2 (\log\log n)^2} = O\left(\frac{1}{\log\log n} \right). \qquad \blacksquare$$