Some Information-Theoretic Computations Related to the Distribution of Prime Numbers

Submitted to Jorma Rissanen's Festschrift volume

I. Kontoyiannis*

February 2, 2008

Abstract

We illustrate how elementary information-theoretic ideas may be employed to provide proofs for well-known, nontrivial results in number theory. Specifically, we give an elementary and fairly short proof of the following asymptotic result,

$$\sum_{p\leq n}\frac{\log p}{p}\sim\log n,\quad\text{as } n\to\infty,$$

where the sum is over all primes p not exceeding n. We also give finite-n bounds refining the above limit. This result, originally proved by Chebyshev in 1852, is closely related to the celebrated prime number theorem.

*Department of Informatics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Email: yiannis@aueb.gr. Web: http://pages.cs.aueb.gr/users/yiannisk/.

1 Introduction

The significant depth of the connection between information theory and statistics appears to have been recognized very soon after the birth of information theory [17] in 1948; a book-length exposition was provided by Kullback [12] already in 1959. In subsequent decades much was accomplished, and in the 1980s the development of this connection culminated in Rissanen's celebrated work [14][15][16], laying the foundations for the notion of stochastic complexity and the Minimum Description Length principle, or MDL.

Here we offer a first glimpse of a different connection, this time between information theory and number theory. In particular, we will show that basic information-theoretic arguments combined with elementary computations can be used to give a new proof for a classical result concerning the distribution of prime numbers. The problem of understanding this "distribution" (including the issue of exactly what is meant by that statement) has, of course, been at the heart of mathematics since antiquity, and it has led, among other things, to the development of the field of analytic number theory; e.g., Apostol's text [1] offers an accessible introduction and [2] gives a more historical perspective.

A major subfield is *probabilistic* number theory, where probabilistic tools are used to derive results in number theory. This approach, pioneered by, among others, Mark Kac and Paul Erdös from the 1930s on, is described, e.g., in Kac's beautiful book [11], Billingsley's review [3], and Tenenbaum's more recent text [18]. The starting point in much of the relevant literature is the following setup: For a fixed, large integer n, choose a random integer N from $\{1, 2, \ldots, n\}$, and write it in its unique prime factorization,

$$N = \prod_{p \le n} p^{X_p}, \tag{1}$$

where the product runs over all primes p not exceeding n, and X_p is the largest power $k \ge 0$ such that p^k divides N. Through this representation, the uniform distribution on N induces a joint distribution on the $\{X_p \; ; \; p \le n\}$, and the key observation is that, for large n, the random variables $\{X_p\}$ are distributed approximately like independent geometrics. Indeed, since there are exactly $\lfloor n/p^k \rfloor$ multiples of p^k between 1 and n,

$$\Pr\{X_p \ge k\} = \Pr\{N \text{ is a multiple of } p^k\} = \frac{1}{n}\Big\lfloor \frac{n}{p^k} \Big\rfloor \approx \Big(\frac{1}{p}\Big)^k, \quad \text{for large } n, \tag{2}$$

so the distribution of X_p is approximately geometric. Similarly, for the joint distribution of the $\{X_p\}$ we find,

$$\Pr\{X_{p_i} \ge k_{p_i} \text{ for primes } p_1, p_2, \ldots, p_m \le n\} = \frac{1}{n}\Big\lfloor \frac{n}{p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}} \Big\rfloor \approx \Big(\frac{1}{p_1}\Big)^{k_1} \Big(\frac{1}{p_2}\Big)^{k_2} \cdots \Big(\frac{1}{p_m}\Big)^{k_m},$$

showing that the $\{X_p\}$ are approximately independent.

This elegant approximation is also mathematically powerful, as it makes it possible to translate standard results about collections of independent random variables into important properties that hold for every "typical" integer N. Billingsley in his 1973 Wald Memorial Lectures [3] gives an account of the state-of-the-art of related results up to that point, but he also goes on to make a further, fascinating connection with the *entropy* of the random variables $\{X_p\}$.

Billingsley's argument essentially begins with the observation that, since the representation (1) is unique, the value of N and the values of the exponents $\{X_p\}$ are in a one-to-one correspondence; therefore, the entropy of N is the same as the entropy of the collection $\{X_p\}$,[1]

$$\log n = H(N) = H(X_p \,;\, p \leq n).$$

And since the random variables $\{X_p\}$ are approximately independent geometrics, we should expect that,

$$\log n = H(X_p \,;\, p \leq n) \approx \sum_{p \leq n} H(X_p) \approx \sum_{p \leq n} \Big[\frac{\log p}{p-1} - \log \Big(1 - \frac{1}{p} \Big) \Big], \tag{3}$$

where in the last equality we simply substituted the well-known expression for the entropy of a geometric random variable (see Section 2 for details on the definition of the entropy and its computation). For large p, the above summands behave like $\frac{\log p}{p}$ to first order, leading to the asymptotic estimate,

$$\sum_{p \leq n} \frac{\log p}{p} \approx \log n, \quad \text{for large } n.$$

Our main goal in this paper is to show that this approximation can indeed be made rigorous, mostly through elementary information-theoretic arguments; we will establish:

Theorem 1. As $n \to \infty$,

$$C(n) := \sum_{p \leq n} \frac{\log p}{p} \sim \log n, \tag{4}$$

where the sum is over all primes p not exceeding n.[2]

As described in more detail in the following section, the fact that the joint distribution of the $\{X_p\}$ is asymptotically close to the distribution of independent geometrics is not sufficient to turn Billingsley's heuristic into an actual proof – at least, we were not able to make the two "$\approx$" steps in (3) rigorous directly. Instead, we provide a proof in two steps. We modify Billingsley's heuristic to derive a *lower bound* on $C(n)$ in Theorem 2, and in Theorem 3 we use a different argument, again going via the entropy of N, to compute a corresponding *upper bound*. These two combined prove Theorem 1, and they also give finer, finite-n bounds on $C(n)$.

In Section 2 we state our main results and describe the intuition behind their proofs. We also briefly review some other elegant information-theoretic arguments connected with bounds on the number of primes up to n. The appendix contains the remaining proofs.

Before moving on to the results themselves, a few words about the history of Theorem 1 are in order. The relationship (4) was first proved by Chebyshev [7][6] in 1852, where he also produced finite-n bounds on $C(n)$, with explicit constants. Chebyshev's motivation was to prove the celebrated prime number theorem (PNT), stating that $\pi(n)$, the number of primes not exceeding n , grows like,

$$\pi(n) \sim \frac{n}{\log n}, \quad \text{as } n \to \infty.$$

[1] For definiteness, we take log to denote the natural logarithm to base e throughout, although the choice of the base of the logarithm is largely irrelevant for our considerations.

[2] As usual, the notation "$a_n \sim b_n$ as $n \to \infty$" means that $\lim_{n \to \infty} a_n / b_n = 1$.

This was conjectured by Gauss around 1792, and it was only proved in 1896; Chebyshev was not able to produce a complete proof, but he used (4) and his finer bounds on $C(n)$ to show that $\pi(n)$ is of order $\frac{n}{\log n}$. Although we will not pursue this direction here, it is actually not hard to see that the asymptotic behavior of $C(n)$ is intimately connected with that of $\pi(n)$. For example, a simple exercise in summation by parts shows that $\pi(n)$ can be expressed directly in terms of $C(n)$:

$$\pi(n) = \frac{n+1}{\log(n+1)}\, C(n) - \sum_{k=2}^{n} \Big(\frac{k+1}{\log(k+1)} - \frac{k}{\log k} \Big) C(k), \quad \text{for all } n \geq 3. \tag{5}$$

For the sake of completeness, this is proved in the appendix.

The PNT was finally proved in 1896 by Hadamard and by de la Vallée-Pousin. Their proofs were not elementary – both relied on the use of Hadamard's theory of integral functions applied to the Riemann zeta function $\zeta(s)$; see [2] for some details. In fact, for quite some time it was believed that no elementary proof would ever be found, and G.H. Hardy in a famous lecture to the Mathematical Society of Copenhagen in 1921 [4] went as far as to suggest that "*if anyone produces an elementary proof of the PNT ... he will show that ... it is time for the books to be cast aside and for the theory to be rewritten.*" It is, therefore, not surprising that Selberg and Erdös' announcement in 1948 that they had produced such an elementary proof caused a great sensation in the mathematical world; see [9] for a survey. In our context, it is interesting to note that Chebyshev's result is again used explicitly in one of the steps of this elementary proof.

Finally we remark that, although the simple arguments in this work fall short of giving estimates precise enough for an elementary information-theoretic proof of the PNT, it may not be entirely unreasonable to hope that such a proof may exist.

2 Primes and Bits: Heuristics and Results

2.1 Preliminaries

For a fixed (typically large) $n \geq 2$, our starting point is the setting described in the introduction. Take N to be a uniformly distributed integer in $\{1, 2, \ldots, n\}$ and write it in its unique prime factorization as in (1),

$$N \;=\; \prod_{p \leq n} p^{X_p} \;=\; p_1^{X_1} \cdot p_2^{X_2} \cdot \cdots \cdot p_{\pi(n)}^{X_{\pi(n)}},$$

where $\pi(n)$ denotes the number of primes $p_1, p_2, \ldots, p_{\pi(n)}$ up to n, and X_p is the largest integer power $k \geq 0$ such that p^k divides N. As noted in (2) above, the distribution of X_p can be described by,

$$\Pr\{X_p \geq k\} = \frac{1}{n} \Big\lfloor \frac{n}{p^k} \Big\rfloor. \quad \text{for all } k \geq 1, \tag{6}$$

This representation also gives simple upper and lower bounds on its mean $E(X_p)$,

$$\mu_p \;:=\; E(X_p) \;=\; \sum_{k \geq 1} \Pr\{X_p \geq k\} \leq \sum_{k \geq 1} \Big(\frac{1}{p}\Big)^k = \frac{1/p}{1 - 1/p} = \frac{1}{p-1}, \tag{7}$$

$$\text{and} \quad \mu_p \;\geq\; \Pr\{X_p \geq 1\} \geq \frac{1}{p} - \frac{1}{n}. \tag{8}$$

Recall the important observation that the distribution of each X_p is close to a geometric. To be precise, a random variable Y with values in $\{0, 1, 2, \ldots\}$ is said to have a geometric distribution with mean $\mu > 0$, denoted $Y \sim \text{Geom}(\mu)$, if $\Pr\{Y = k\} = \mu^k/(1+\mu)^{k+1}$, for all $k \geq 0$. Then Y of course has mean $E(Y) = \mu$ and its entropy is,

$$h(\mu) := H(\text{Geom}(\mu)) = -\sum_{k\geq 0} \Pr\{Y = k\} \log \Pr\{Y = k\} = (\mu+1)\log(\mu+1) - \mu \log \mu. \tag{9}$$

See, e.g., [8] for the standard properties of the entropy.

2.2 Billingsley's Heuristic and Lower Bounds on $C(n)$

First we show how Billingsley's heuristic can be modified to yield a lower bound on $C(n)$. Arguing as in the introduction,

$$\log n \overset{(a)}{=} H(N) \overset{(b)}{=} H(X_p \,;\, p \leq n) \overset{(c)}{\leq} \sum_{p\leq n} H(X_p) \overset{(d)}{\leq} \sum_{p\leq n} H(\text{Geom}(\mu_p)) \overset{(e)}{=} \sum_{p\leq n} h(\mu_p), \tag{10}$$

where (a) is simply the entropy of the uniform distribution, (b) comes from the fact that N and the $\{X_p\}$ are in a one-to-one correspondence, (c) is the well-known subadditivity of the entropy, (d) is because the geometric has maximal entropy among all distributions on the non-negative integers with a fixed mean, and (e) is the definition of $h(\mu)$ in (9). Noting that $h(\mu)$ is nondecreasing in μ and recalling the upper bound on μ_p in (7) gives,

$$\log n \leq \sum_{p\leq n} h(\mu_p) \leq \sum_{p\leq n} h(1/(p-1)) = \sum_{p\leq n} \Big[\frac{p}{p-1}\log\Big(\frac{p}{p-1}\Big) - \frac{1}{p-1}\log\Big(\frac{1}{p-1}\Big)\Big]. \tag{11}$$

Rearranging the terms in the sum proves:

Theorem 2. For all $n \geq 2$,

$$T(n) := \sum_{p\leq n} \Big[\frac{\log p}{p-1} - \log\Big(1 - \frac{1}{p}\Big)\Big] \geq \log n.$$

Since the summands above behave like $\frac{\log p}{p}$ for large p, it is not difficult to deduce the following lower bounds on $C(n) = \sum_{p\leq n} \frac{\log p}{p}$:

Corollary 1. [Lower Bounds on $C(n)$]

$$(i) \qquad \liminf_{n\to\infty} \frac{C(n)}{\log n} \geq 1;$$

$$(ii) \qquad C(n) \geq \frac{86}{125}\log n - 2.35, \qquad \text{for all } n \geq 16.$$

Corollary 1 is proved in the appendix. Part (i) proves half of Theorem 1, and (ii) is a simple evaluation of the more general bound derived in equation (15) in the proof: For any $N_0 \geq 2$, we have,

$$C(n) \geq \Big(1 - \frac{1}{N_0}\Big)\Big(1 - \frac{1}{1+\log N_0}\Big)\log n + C(N_0) - T(N_0), \quad \text{for all } n \geq N_0.$$

2.3 A Simple Upper Bound on $C(n)$

Unfortunately, it is not clear how to reverse the inequalities in equations (10) and (11) to get a corresponding upper bound on $C(n)$ – especially inequality (c) in (10). Instead we use a different argument, one which is less satisfying from an information-theoretic point of view, for two reasons. First, although again we do go via the entropy of N, it is not necessary to do so; see equation (13) below. And second, we need to use an auxiliary result, namely, the following rough estimate on the sum, $\vartheta(n) := \sum_{p \le n} \log p$:

$$\vartheta(n) := \sum_{p \le n} \log p \le (2 \log 2) n, \quad \text{for all } n \ge 2. \tag{12}$$

For completeness, it is proved at the end of this section.

To obtain an upper bound on $C(n)$, we note that the entropy of N, $H(N) = \log n$, can be expressed in an alternative form: Let Q denote the probability mass function of N, so that $Q(k) = 1/n$ for all $1 \le k \le n$. Since $N \le n = 1/Q(N)$ always, we have,

$$H(N) = E[-\log Q(N)] \ge E[\log N] = E\Big[\log \prod_{p \le n} p^{X_p}\Big] = \sum_{p \le n} E(X_p) \log p. \tag{13}$$

Therefore, recalling (8) and using the bound (12),

$$\log n \ge \sum_{p \le n} \left(\frac{1}{p} - \frac{1}{n}\right) \log p = \sum_{p \le n} \frac{\log p}{p} - \frac{\vartheta(n)}{n} \ge \sum_{p \le n} \frac{\log p}{p} - 2 \log 2,$$

thus proving:

Theorem 3. [Upper Bound] For all $n \ge 2$,

$$\sum_{p \le n} \frac{\log p}{p} \le \log n + 2 \log 2.$$

Theorem 3 together with Corollary 1 prove Theorem 1. Of course the use of the entropy could have been avoided entirely: Instead of using that $H(N) = \log n$ in (13), we could simply use that $n \ge N$ by definition, so $\log n \ge E[\log N]$, and proceed as before.

Finally (paraphrasing from [10, p. 341]) we give an elegant argument of Erdös that employs a cute, elementary trick to prove the inequality on $\vartheta(n)$ in (12). First observe that we can restrict attention to odd n, since $\vartheta(2n) = \vartheta(2n-1)$, for all $n \ge 2$ (as there are no even primes other than 2). Let $n \ge 2$ arbitrary; then every prime $n+1 < p \le 2n+1$ divides the binomial coefficient,

$$B := \binom{2n+1}{n} = \frac{(2n+1)!}{n!(n+1)!},$$

since it divides the numerator but not the denominator, and hence the product of all these primes also divides B. In particular, their product must be no greater than B, i.e.,

$$\prod_{n+1 < p \le 2n+1} p \le B = \frac{1}{2}\binom{2n+1}{n} + \frac{1}{2}\binom{2n+1}{n+1} \le \frac{1}{2}(1+1)^{2n+1} = 2^{2n},$$

or, taking logarithms,

$$\vartheta(2n+1) - \vartheta(n+1) = \sum_{n+1<p\le 2n+1} \log p = \log\Big[\prod_{n+1<p\le 2n+1} p\Big] \le (2\log 2)n.$$

Iterating this bound inductively gives the required result.

2.4 Other Information-Theoretic Bounds on the Primes

Billingsley in his 1973 Wald Memorial Lectures [3] appears to have been the first to connect the entropy with properties of the asymptotic distribution of the primes. Although there are no results in that work based on information-theoretic arguments, he does suggest the heuristic upon which part of our proof of Theorem 2 was based, and he also goes in the opposite direction: He uses probabilistic techniques and results about the primes to compute the entropy of several relevant collections of random variables.

Chaitin in 1979 [5] gave a proof of the fact that there are infinitely many primes, using algorithmic information theory. Essentially the same argument proves a slightly stronger result, namely that, $\pi(n) \ge \frac{\log n}{\log\log n+1}$, for all $n \ge 3$. Chaitin's proof can easily be translated into our setting as follows. Recall the representation (1) of a uniformly distributed integer N in $\{1, 2, \ldots, n\}$. Since p^{X_p} divides N, we must have $p^{X_p} \le n$, so that each X_p lies in the range,

$$0 \le X_p \le \Big\lfloor \frac{\log n}{\log p} \Big\rfloor \le \frac{\log n}{\log p},$$

and hence, $H(X_p) \le \log\big(\frac{\log n}{\log p} + 1\big)$. Therefore, arguing as before,

$$\log n = H(N) = H(X_p\,;\, p \le n) \le \sum_{p\le n} H(X_p) \le \sum_{p\le n} \log\Big(\frac{\log n}{\log 2} + 1\Big) \le \pi(n)(\log\log n + 1),$$

where the last inequality holds for all $n \ge 3$.

It is interesting that the same argument applied to a different representation for N yields a marginally better bound: Suppose we write,

$$N = M^2 \prod_{p\le n} p^{Y_p},$$

where $M \ge 1$ is the largest integer such that M^2 divides N, and each of the Y_p are either zero or one. Then $H(Y_p) \le \log 2$ for all p, and the fact that $M^2 \le n$ implies that $H(M) \le \log\lfloor\sqrt{n}\rfloor$. Therefore,

$$\log n = H(N) = H(M, Y_{p_1}, Y_{p_2}, \ldots, Y_{p_{\pi(n)}}) \le H(M) + \sum_{p\le n} H(Y_p) \le \frac{1}{2}\log n + \pi(n)\log 2,$$

which implies that $\pi(n) \ge \frac{\log n}{2\log 2}$, for all $n \ge 2$.

Finally we mention that in Li and Vitányi's text [13], an elegant argument is given for a more accurate lower bound on $\pi(n)$. Using ideas and results from algorithmic information theory, they show that, $\pi(n) = \Omega\big(\frac{n}{(\log n)^2}\big)$. But the proof (which they attribute to unpublished work by P. Berman (1987) and J. Tromp (1990)) is somewhat involved, and uses tools very different to those developed here.

Appendix

PROOF OF THE SUMMATION-BY-PARTS FORMULA (5). Note that, since $\pi(k)-\pi(k-1)$ is zero unless k is prime, $C(n)$ can be expressed as a sum over all integers $k \le n$,

$$C(n) = \sum_{2\le k\le n} [\pi(k)-\pi(k-1)]\frac{\log k}{k}.$$

Each of the following steps is obvious, giving,

$$\begin{aligned}
\pi(n) &= \sum_{k=2}^{n}[\pi(k)-\pi(k-1)] \\
&= \sum_{k=2}^{n}[\pi(k)-\pi(k-1)]\frac{\log k}{k}\frac{k}{\log k} \\
&= \sum_{k=2}^{n}\Big[C(k)-C(k-1)\Big]\frac{k}{\log k} \\
&= \sum_{k=2}^{n}C(k)\frac{k}{\log k} - \sum_{k=2}^{n}C(k-1)\frac{k}{\log k} \\
&= \sum_{k=2}^{n}C(k)\frac{k}{\log k} - \sum_{k=1}^{n-1}C(k)\frac{k+1}{\log(k+1)} \\
&= \frac{n+1}{\log(n+1)}\,C(n) - \sum_{k=2}^{n}\Big(\frac{k+1}{\log(k+1)} - \frac{k}{\log k}\Big)C(k) \;-\; \frac{2}{\log 2}\,C(1),
\end{aligned}$$

as claimed, since $C(1)=0$, by definition. □

PROOF OF COROLLARY 1. Choose and fix any $N_0 \ge 2$ and let $n \ge N_0$ arbitrary. Then,

$$\log n \le T(n) = T(N_0) + \sum_{N_0<p\le n}\Big[\frac{\log p}{p-1} - \log\Big(1-\frac{1}{p}\Big)\Big] \le T(N_0) + \sum_{N_0<p\le n}\Big[\frac{1}{1-\frac{1}{p}}\frac{\log p}{p} + \frac{N_0}{N_0-1}\frac{1}{p}\Big],$$

where the last inequality follows from the inequality $-\log(1-x) \le x/(1-\delta)$, for all $0 \le x \le \delta < 1$, with $\delta = 1/N_0$. Therefore,

$$\begin{aligned}
\log n &\le T(N_0) + \sum_{N_0<p\le n}\Big[\frac{1}{1-\frac{1}{N_0}}\frac{\log p}{p} + \frac{N_0}{N_0-1}\frac{\log p}{\log N_0}\frac{1}{p}\Big] \\
&= T(N_0) + \Big(\frac{N_0}{N_0-1}\Big)\Big(1+\frac{1}{\log N_0}\Big)\Big(C(n)-C(N_0)\Big). \qquad (14)
\end{aligned}$$

Dividing by $\log n$ and letting $n \to \infty$ yields,

$$\liminf_{n\to\infty}\frac{C(n)}{\log n} \ge \frac{(N_0-1)\log N_0}{N_0(1+\log N_0)},$$

and since N_0 was arbitrary, letting now $N_0 \to \infty$ implies (i).

For all $n \ge N_0$, (14) implies,

$$C(n) \ge \Big(1-\frac{1}{N_0}\Big)\Big(1-\frac{1}{1+\log N_0}\Big)\log n + C(N_0) - T(N_0), \qquad (15)$$

and evaluating this at $N_0 = 16$ gives (ii). □

Acknowledgments

Thanks to Peter Harremoës for spotting a small error in an earlier version of this paper.

References

[1] T.M. Apostol. *Introduction to Analytic Number Theory*. Springer-Verlag, New York, 1976.

[2] P.T. Bateman and H.G. Diamond. A hundred years of prime numbers. *Amer. Math. Monthly*, 103(9):729–741, 1996.

[3] P. Billingsley. The probability theory of additive arithmetic functions. *Ann. Probab.*, 2:749–791, 1974.

[4] H. Bohr. Address of Professor Harold Bohr. In *Proceedings of the International Congress of Mathematicians (Cambridge, 1950) vol. 1*, pages 127–134. Amer. Math. Soc., Providence, RI, 1952.

[5] G.J. Chaitin. Toward a mathematical definition of "life". In *Maximum entropy formalism (Conf., Mass. Inst. Tech., Cambridge, Mass., 1978)*, pages 477–498. MIT Press, Cambridge, Mass., 1979.

[6] P.L. Chebychev. Mémoire sur les nombres premiers. *J. de Math. Pures Appl.*, 17:366–390, 1852.

[7] P.L. Chebychev. Sur la totalité des nombres premiers inférieurs à une limite donnée. *J. de Math. Pures Appl.*, 17:341–365, 1852.

[8] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. J. Wiley, New York, 1991.

[9] H.G. Diamond. Elementary methods in the study of the distribution of prime numbers. *Bull. Amer. Math. Soc. (N.S.)*, 7(3):553–589, 1982.

[10] G.H. Hardy and E.M. Wright. *An Introduction to the Theory of Numbers*. The Clarendon Press Oxford University Press, New York, fifth edition, 1979.

[11] M. Kac. *Statistical Independence in Probability, Analysis and Number Theory*. Published by the Mathematical Association of America. Distributed by John Wiley and Sons, Inc., New York, 1959.

[12] S. Kullback. *Information Theory and Statistics*. Dover Publications Inc., Mineola, NY, 1997. Reprint of the second (1968) edition.

[13] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, New York, second edition, 1997.

[14] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Ann. Statist.*, 11(2):416–431, 1983.

[15] J. Rissanen. Stochastic complexity. *J. Roy. Statist. Soc. Ser. B*, 49(3):223–239, 253–265, 1987. With discussion.

[16] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore, 1989.

[17] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[18] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.