2.6. EXAMPLES

In this section we illustrate by examples the applicability of probabilistic notions and tools to other fields. Such applications abound, so we have restricted ourselves to just a handful of representative examples from Analysis, Arithmetic, Graph Theory and Statistics. In the next chapters we shall illustrate this interplay by additional examples, most notably through the interplay between the Brownian Motion process and Partial Differential Equations.

6.1 Exercise : S. Bernštein's proof of the Weierstrass Approximation Theorem. For any continuous function $f : [0,1] \to \mathbf{R}$, there exists a sequence of *polynomials* $\{B_n(\cdot)\}_{n\in\mathbf{N}}$ that converge to f uniformly over $[0,1]$:

$$\sup_{0\le x\le 1} |B_n(x) - f(x)| \longrightarrow 0\,, \qquad \text{as } n \to \infty\,.$$

In fact, one can take as such the *"Bernštein polynomials"*

$$B_n(x) = \sum_{k=0}^{n} f(k/n) \cdot \frac{n!}{k!(n-k)!}\, x^k (1-x)^{n-k}\,, \quad 0 \le x \le 1\,, \quad n \in \mathbf{N}\,. \tag{6.1}$$

If, in addition, the function $f(\cdot)$ satisfies a Lipschitz condition of the type $|f(x) - f(y)| \le K\,|x-y|\,,\ \forall\, x, y \in [0,1]$ for some $K \in (0,\infty)$, then in fact

$$\sup_{0\le x\le 1} |B_n(x) - f(x)| \le \frac{K}{2\sqrt{n}}\,, \quad \text{for all } n \in \mathbf{N}.$$

(*Hint:* Observe that $B_n(p) = \mathbf{E}[f(\overline{X}_n)]$ in the notation of Example 2.1, in particular (2.4), (2.5), and use Exercise 2.3(ii) as well as the Čebyšev Inequality.)

6.1 EXAMPLE : HYPOTHESIS-TESTING, DETECTION. On a measurable space $(\Omega, \mathcal{F})$, suppose that we are given two probability measures $\mathbf{P}_0$ (*hypothesis*/"enemy aircraft") and $\mathbf{P}_1$ (*alternative*/"harmless object"), and that we want to discriminate between them. We can try to do this in terms of a (pure) *test*, that is, a random variable $X : \Omega \to \{0,1\}$, *which rejects* $\mathbf{P}_0$ *on the event* $\{X = 1\}$. With this interpretation, $\mathbf{P}_0(X = 1)$ is the probability of rejecting $\mathbf{P}_0$ when it is true (probability of *type-I-error*, or "failure-to-detect"), whereas $\mathbf{P}_1(X = 0) = 1 - \mathbf{P}_1(X = 1)$ is the probability of accepting $\mathbf{P}_0$ when it is false (probability of *type-II-error*, or "false alarm"). Ideally, one would like to minimize these error probabilities simultaneously, but typically this will not be possible: a more sensitive radar decreases the chance of letting enemy aircraft go undetected, but

also makes false alarms more likely. The next best thing is then to fix a certain number $0 < \alpha < 1$ (say $\alpha = 1\%$ or $\alpha = 5\%$), and try to

$$\text{maximize } \mathbf{P}_1(X = 1), \text{ subject to } \mathbf{P}_0(X = 1) \leq \alpha. \tag{6.2}$$

In other words, one tries to find a test that minimizes the probability of "false alarms", among all tests that keep the probability of "failure to detect" below a given acceptable *significance level* $\alpha \in (0, 1)$. This is the tack taken by the classical Neyman-Pearson theory of Hypothesis Testing; see, for instance, Lehmann (1986), or Ferguson (1967).

The basic results of this theory are as follows. Take a third probability measure μ with $\mathbf{P}_0 < \mu$, $Q < \mu$ (for instance, $\mu = (\mathbf{P}_0 + \mathbf{P}_1)/2$), and set

$$G := \frac{d\mathbf{P}_1}{d\mu} , \quad H := \frac{d\mathbf{P}_0}{d\mu} .$$

6.2 Exercise: Neyman-Pearson Lemma. The problem of (6.2) has a solution, namely $\widehat{X} = \chi_{\{kH<G\}}$, provided that $\mathbf{P}_0(kH < G) = \alpha$ for some $0 < k < \infty$,.

In other words, the test $\widehat{X}$ of Exercise 6.2 rejects the hypothesis, if and only if the "likelihood ratio" $G/H = (d\mathbf{P}_1/d\mu)/(d\mathbf{P}_0/d\mu)$ is sufficiently large. When a number k with these properties cannot be found, one has to consider *randomized tests*, that is, random variables $X : \Omega \to [0, 1]$. The new interpretation is that, if the outcome $\omega \in \Omega$ is observed, then the hypothesis $\mathbf{P}_0$ is rejected (respectively, accepted) with probability $X(\omega)$ (resp., $1 - X(\omega)$), independently of everything else. This way, $\mathbf{E}_1(X) = \int X(\omega)\, d\mathbf{P}_1(\omega)$ is then the *power of the randomized test* X, that is, the probability of rejecting the hypothesis $\mathbf{P}_0$ when it is false; and $\mathbf{E}_0(X) = \int X(\omega)\, d\mathbf{P}_0(d\omega)$ is the probability of type-I-error for the randomized test X (i.e., of rejecting $\mathbf{P}_0$ when it is true). By analogy with (6.2), one seeks a randomized test $\widehat{X}$ which

$$\left\{ \begin{array}{c} \text{maximizes } \mathbf{E}_1(X), \text{ over all randomized tests} \\ X : \Omega \to [0, 1] \text{ with } \mathbf{E}_0(X) \leq \alpha \end{array} \right\}. \tag{6.3}$$

The advantage of this "randomized" formulation is that the problem of (6.3) has a solution for *any* given significance level $\alpha \in (0, 1)$.

6.3 Exercise: Generalized Neyman-Pearson Lemma. With $\mathcal{X}_\alpha := \{X : \Omega \to [0, 1] \mid \mathbf{E}_0(X) \leq \alpha\}$, the supremum $\sup_{X \in \mathcal{X}_\alpha} \mathbf{E}_1(X)$ is attained by the randomized test

$$\widehat{X} \;=\; \chi_{\{kH<G\}} + b \cdot \chi_{\{kH=G\}} \;\in\; \mathcal{X}_\alpha ,$$

where we have set (with the convention $0/0 = 0$): $k := \inf\{u \geq 0 \mid \mathbf{P}_0(uH < G) \leq \alpha\}$ and

$$b := \frac{\alpha - \mathbf{P}_0(kH < G)}{\mathbf{P}_0(kH = G)} = \frac{\alpha - \mathbf{P}_0(kH < G)}{\mathbf{P}_0(kH \leq G) - \mathbf{P}_0(kH < G)} \quad \in \quad [0, 1].$$

6.2 EXAMPLE: PRIME DIVISORS. For any integer m, let us denote by $\xi(m)$ the number of prime divisors of m (without multiplicities). A result of Hardy and Ramanujan states, roughly, that "*almost every* integer m has *approximately* $\log\log m$ prime divisors, and that the variance in the uncertainty of this statement is of the same order" $\log\log m$:

$$\lim_{n\to\infty} \frac{1}{n} \cdot \# \left\{ 1 \le m \le n \;\middle|\; \frac{|\xi(m) - \log\log m|}{\sqrt{\log\log m}} > z_m \right\} = 0 \qquad \text{and} \tag{6.4}$$

$$\lim_{n\to\infty} \frac{1}{n} \cdot \# \left\{ 1 \le m \le n \;\middle|\; \frac{|\xi(m) - \log\log n|}{\sqrt{\log\log n}} > z_n \right\} = 0\,, \tag{6.5}$$

for any sequence $\{z_n\}_{n\in\mathbf{N}}$ of positive numbers with $\lim_{n\to\infty} z_n = \infty$. This statement was later refined by Erdös and Kac, who showed that in the notation of (2.7) we have

$$\lim_{n\to\infty} \frac{1}{n} \cdot \# \left\{ 1 \le m \le n \;\middle|\; \frac{\xi(m) - \log\log n}{\sqrt{\log\log n}} \le z \right\} = \Phi(z)\,, \qquad \forall\, z \in \mathbf{R}\,, \tag{6.6}$$

$$\lim_{n\to\infty} \frac{1}{n} \cdot \# \left\{ 1 \le m \le n \;\middle|\; \frac{\xi(m) - \log\log m}{\sqrt{\log\log m}} \le z \right\} = \Phi(z)\,, \qquad \forall\, z \in \mathbf{R}\,. \tag{6.7}$$

P. Turán gave an elementary probabilistic argument for (6.4)-(6.5), that runs as follows. One starts by introducing, for each $n \in \mathbf{N}$, a random variable M_n with *uniform distribution on* $\{1, \cdots, n\}$, namely $\mathbf{P}[M_n = m] = 1/n$, $m = 1, \cdots, n$. Writing $\xi(m) = \sum_{p\le n} \eta_p(m)$, where the sum extends over primes and $\eta_p(m) = 1$ if p/m, $\eta_p(m) = 0$ otherwise, we obtain $\frac{1}{p} - \frac{1}{n} < \mathbf{E}\,[\eta_p(M_n)] = \frac{1}{n} \sum_{m=1}^{n} \eta_p(m) \le \frac{1}{p}$, as well as

$$\begin{aligned} \mathrm{Cov}\,(\eta_p(M_n), \eta_q(M_n)) &= \mathbf{E}\,[\eta_p(M_n)\,\eta_q(M_n)] - \mathbf{E}\,[\eta_p(M_n)] \cdot \mathbf{E}\,[\eta_q(M_n)] \\ &\le \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right) \le \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right) \end{aligned}$$

for any prime numbers $p \ne q$ (since then $\eta_p(m)\,\eta_q(m) = 1 \;\Leftrightarrow\; \eta_{pq}(m) = 1$). Using the basic information from the Prime-Number Theorem

$$\pi(n) := \sum_{p\le n} 1 \;\sim\; \frac{n}{\log n}\,, \qquad \sum_{p\le n} \frac{1}{p} \;\sim\; \log\log n + A + O(1/\log n)$$

as $n \to \infty$ (e.g. Apostol (1976), Chapter 4), we obtain that the sum of covariances

$$\sum\sum\nolimits_{1\le p\ne q\le n} \mathrm{Cov}\,(\eta_p(M_n), \eta_q(M_n)) \;\le\; \frac{1}{n} \cdot \sum\sum\nolimits_{p\ne q} \left(\frac{1}{p} + \frac{1}{q}\right) \;=\; \frac{\pi(n) - 1}{n} \cdot \sum_{p\le n} (2/p)$$

$$\sim\; 2\left(\frac{1}{\log n} - \frac{1}{n}\right)\,[\,\log\log n + A + O(1/\log n)\,] \;\longrightarrow\; 0$$

in the inequality of Exercise 2.2(c) is negligible as $n \to \infty$, yielding $\mathbf{E}(\xi(M_n)) = \sum_{p \le n} \mathbf{E}\,[\eta_p(M_n)] = \log\log n + O(1)$ and $\mathrm{Var}(\xi(M_n)) \le \log\log n + O(1)$. It follows then from the Čebyšev inequality that the left-hand side of (6.5) is asymptotically equivalent to

$$\mathbf{P}\left[\, |\xi(M_n) - \log\log n| > z_n \sqrt{\log\log n}\,\right] \le \frac{\mathrm{Var}(\xi(M_n))}{z_n^2 \log\log n} = \frac{1}{z_n^2} + O\,(\,1\,/\,\log\log n\,)$$

as $n \to \infty$, and the result (6.5) follows.

6.4 Exercise: Deduce (6.4) from (6.5), using the very slow increase of $\log\log n$. (*Hint:* Let $0 < \alpha < 1$, and consider only integers in the range $n^\alpha \le m \le n$; show that every integer m in this range that satisfies the condition of (6.5), also satisfies the condition of (6.4) for an appropriate increasing sequence $\{z_n\}_{n \in \mathbf{N}} \to \infty$.)

6.5 Exercise: Prove the central-limit-theorem-type results (6.6)-(6.7) of Erdös and Kac.

6.3 EXAMPLE : TOURNAMENTS. Consider a set $\mathcal{V}$ of n vertices, and on it a *Tournament* $\mathcal{T}_n$, that is, a complete directed graph. In other words, suppose we have n players, each of whom faces every other player, in a competition where no draws are allowed; we direct an edge from vertex (player) i to vertex (player) j, if i beats j. The schedule of the tournament does not matter, only the results. For a given integer $k < n$, we say that the tournament $\mathcal{T}_n$ *has property* $\mathcal{S}_k$, if for every set of k players $\{x_1, \cdots, x_k\}$, there is some other player y, who beats every player in the set.

Schütte was the first to pose the following problem: *Is it true that for every integer* k, *there exists a set* $\mathcal{V}$ *of* $n > k$ *vertices, and on it a Tournament* $\mathcal{T}_n$ *with the property* $\mathcal{S}_k$? And if so, what is the smallest necessary number $f(k)$ of players?

The (affirmative) answer of (6.8) below to Schütte's question, was given by P. Erdös (1963), and illustrates the "Probabilistic Method" that he introduced in Combinatorics and in Graph Theory. The rough idea is that, for $n \ge f(k)$ sufficiently large, an appropriately defined "random tournament" on the set $\mathcal{V} = \{1, \cdots, n\}$ of players is "very likely" to have the property $\mathcal{S}_k$.

For every $k \in \mathbf{N}$, *there exists a finite tournament* $\mathcal{T}_n$, $n > k$ *with the property* $\mathcal{S}_k$. (6.8)

Proof of (6.8): Consider a set $\mathcal{V} = \{1, \cdots, n\}$ of $n > k$ players, and a "random graph" $\mathcal{T}_n$ on it that corresponds to the idea of "deciding each game by tossing a coin independently from game to game". More formally, one takes as the sample space

$$\Omega = \{\, (\omega_{ij})_{1 \le i,j \le n} \;\mid\; \omega_{ij} = 1 \ (i \text{ beats } j) \quad \text{or} \quad \omega_{ij} = -1 \ (j \text{ beats } i)\,\}\,,$$

and assigns the probability 2^{-N}, $N = n(n-1)/2$ to each of its 2^N elements.

For any given subset $\mathcal{X} \subset \mathcal{V}$ with k elements, denote by $A_{\mathcal{X}}$ the property "no player $y \in \mathcal{V} \setminus \mathcal{X}$ beats all players in $\mathcal{X}$". Clearly, $\mathbf{P}[\, v \text{ beats all players in } \mathcal{X}\,] = 1 - 2^{-k}$, for every $v \in \mathcal{V} \setminus \mathcal{X}$, and thus by independence: $\mathbf{P}(A_{\mathcal{X}}) = \left(1 - 2^{-k}\right)^{n-k}$. Therefore,

$$\mathbf{P}\left[\bigcup_{\substack{\mathcal{X}\subset\mathcal{V}\\ \#\mathcal{X}=k}} A_{\mathcal{X}}\right] \le \sum_{\substack{\mathcal{X}\subset\mathcal{V}\\ \#\mathcal{X}=k}} \mathbf{P}(A_{\mathcal{X}}) = \frac{n!}{k!(n-k)!}\left(1-2^{-k}\right)^{n-k} < 1\,, \qquad \text{for } n \ge f(k)\,,$$

where $n \ge f(k) := \min\{m \ge k \,|\, (m!/k!(m-k)!) \cdot \left(1-2^{-k}\right)^{m-k} < 1\}$. In other words,

$$\mathbf{P}\left[\bigcap_{\substack{\mathcal{X}\subset\mathcal{V}\\ \#\mathcal{X}=k}} (A_{\mathcal{X}})^c\right] = \mathbf{P}[\text{the property } A_{\mathcal{X}} \text{ holds}] > 0\,, \tag{6.9}$$

and thus *there is a point in the probability space* Ω (i.e., a tournament $\mathcal{T}_n$) with the property $\mathcal{S}_k$. $\diamond$

It can be checked that $f(1) = 3$, $f(2) = 7$, and careful asymptotics give $c \cdot k2^k \le f(k) \le k^2 2^k\,(1 + o(1))$, as $k \to \infty$. The argument leading to (6.9) is a typical illustration of P. Erdös's "probabilistic method" in Combinatorial Analysis: one shows the existence of an object possessing a certain property by constructing an appropriate probability space and showing that the event corresponding to the property under consideration has positive probability.

6.4 EXAMPLE : BOREL's NORMAL NUMBERS. One of the first strong laws of large numbers was proved by E. Borel, who observed that the Rademacher functions $\{r_k\}_{k\in\mathbf{N}}$ of Example 2.2 satisfy

$$\lim_{n\to\infty} \frac{r_1(\omega) + \cdots + r_n(\omega)}{n} = 0\,, \quad \text{for } \lambda - \text{a.e. } \omega \in [0,1)\,. \tag{6.10}$$

Borel used essentially the method of proof for the Markov and Cantelli stong laws; cf. Exercises 4.3 and 4.5.

To wit: if $\{f_n\}_{n\in\mathbf{N}}$ is a sequence of non-negative and integrable functions on [0,1), then the convergence of $\sum_{n=1}^{\infty}\int_0^1 f_n(\omega)\,d\omega$ implies the convergence of the series $\sum_{n=1}^{\infty} f_n(\omega)$ for $\lambda-$a.e. $\omega \in [0,1)$; recall Exercise 2.4 (iv). Now take $f_n(\omega) = \left(\frac{r_1(\omega)+\cdots+r_n(\omega)}{n}\right)^4$ and observe that for this choice

$$\int_0^1 f_n(\omega)\,d\omega = \frac{n + \frac{4!}{2!2!}\frac{n!}{2!(n-2)!}}{n^4}$$

is the general term of a convergent series. We conclude now that for $\lambda-$a.e. $\omega \in [0,1)$ we have $\sum_{n=1}^{\infty} \int_0^1 f_n(\omega)\, d\omega < \infty$, thus also $\lim_{n\to\infty} f_n(\omega) = 0$, and we are done.

In the notation of Example 2.2 we have $r_k = 1 - 2\varepsilon_k$, so (6.10) can be written equivalently as

$$\lim_{n\to\infty} \frac{\varepsilon_1(\omega) + \cdots + \varepsilon_n(\omega)}{n} = \frac{1}{2}, \quad \text{for } \lambda - \text{a.e. } \omega \in [0,1). \tag{6.11}$$

In other words, *almost every number* $\omega \in [0,1)$ *has asymptotically the same proportion of 0's and 1's in its binary expansion.* We express this property as **normality to base 2**.

From a probabilistic point of view the statement (6.11) is just the strong law of large numbers applied to the sequence of independent Bernoulli variables $\varepsilon_1, \varepsilon_2, \cdots$ of Example 2.2, for which $\lambda(\varepsilon_n = 0) = \lambda(\varepsilon_n = 1) = 1/2$.

Of course, nothing about the particular base 2 is sacrosanct. If $b \geq 2$ is an integer, we also have a unique expansion

$$\omega = \frac{\zeta_1(\omega)}{b} + \frac{\zeta_2(\omega)}{b^2} + \cdots + \frac{\zeta_n(\omega)}{b^n} + \cdots$$

for every $\omega \in [0,1)$, where each digit $\zeta_n(\omega)$ takes values in $\{0, 1, \cdots, b-1\}$. Then one shows that every given digit $k \in \{0, 1, \cdots, b-1\}$ occurs with the same asymptotic frequency in this expansion, namely:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \chi_{\{k\}}(\zeta_j(\omega)) = \frac{1}{b}, \tag{6.12}$$

for $\lambda-$a.e. $\omega \in [0,1)$. This is *normality to base b.*

But of course, the union of countably many null sets is a null set, so we conclude that (Lebesgue) *almost every number in* [0,1) *is* **normal**, meaning that it satisfies (6.12) for all digits $k = 0, 1, \cdots, b-1$ and all bases $b \geq 2$.

It is ironic that it is not actually easy to exhibit even one member of this overwhelming majority! No rational number is normal, though it might be to a particular basis (for example, $\frac{1}{3} = \frac{0}{2} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \frac{0}{2^5} + \cdots$ is normal to base 2). The simplest known example of a normal number is written in usual decimal notation as

0.12345678910111213141516171819202122232425262 7.....

by listing all positive integers in succession after the decimal point; but even for this number normality is no trivial matter to establish!