Chapter 6

Convergence of Random Processes

In this chapter we study the convergence of discrete random processes. This allows to characterize two phenomena that are fundamental in statistical estimation and probabilistic modeling: the law of large numbers and the central limit theorem.

6.1 Types of convergence

Let us quickly recall the concept of convergence for a deterministic sequence of real numbers x_1, x_2, ... We have

$$\lim_{i \to \infty} x_i = x \tag{6.1}$$

if x_i is arbitrarily close to x as the index i grows. More formally, the sequence converges to x if for any $\epsilon > 0$ there is an index i_0 such that for all indices i greater than i_0 we have $|x_i - x| < \epsilon$. Recall that any realization of a discrete-time random process $\widetilde{X}(\omega, i)$ where we fix the outcome ω is a deterministic sequence. Establishing convergence of such realizations to a fixed number can therefore be achieved by computing the corresponding limit. However, if we consider the random process itself instead of a realization and we want to determine whether it eventually converges to a random variable X, then deterministic convergence no longer makes sense. In this section we describe several alternative definitions of convergence, which allow to extend this concept to random quantities.

6.1.1 Convergence with probability one

Consider a discrete random process $\widetilde{X}$ and a random variable X defined on the same probability space. If we fix an element ω of the sample space Ω, then $\widetilde{X}(i, \omega)$ is a deterministic sequence and $X(\omega)$ is a constant. It is consequently possible to verify whether $\widetilde{X}(i, \omega)$ converges deterministically to $X(\omega)$ as $i \to \infty$ *for that particular value of* ω. In fact, we can ask: what is the probability that this happens? To be precise, this would be the probability that if we draw ω we have

$$\lim_{i \to \infty} \widetilde{X}(i, \omega) = X(\omega). \tag{6.2}$$

If this probability equals one then we say that $\widetilde{X}(i)$ converges to X with probability one.

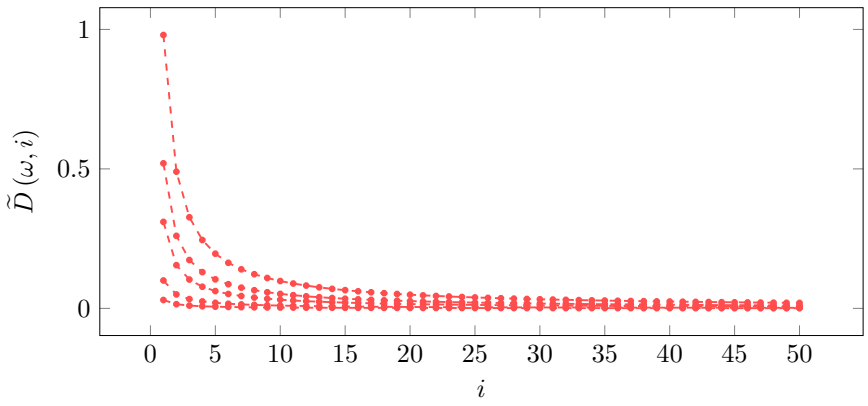


Figure 6.1: Convergence to zero of the discrete random process $\widetilde{D}$ defined in Example 5.1.2.

Definition 6.1.1 (Convergence with probability one)**.** *A discrete random vector $\widetilde{X}$ converges with probability one to a random variable X belonging to the same probability space $(\Omega, \mathcal{F}, \mathrm{P})$ if*

$$\mathrm{P}\left(\left\{\omega \mid \omega \in \Omega, \quad \lim_{i\to\infty} \widetilde{X}(\omega, i) = X(\omega)\right\}\right) = 1. \tag{6.3}$$

Recall that in general the sample space Ω is very difficult to define and manipulate explicitly, except for very simple cases.

Example 6.1.2 (Puddle (continued from Example 5.1.2))**.** Let us consider the discrete random process $\widetilde{D}$ defined in Example 5.1.2. If we fix $\omega \in (0, 1)$

$$\lim_{i\to\infty} \widetilde{D}(\omega, i) = \lim_{i\to\infty} \frac{\omega}{i} \tag{6.4}$$

$$= 0. \tag{6.5}$$

It turns out the realizations tend to zero for all possible values of ω in the sample space. This implies that $\widetilde{D}$ converges to zero with probability one.

△

6.1.2 Convergence in mean square and in probability

To verify convergence with probability one we fix the outcome ω and check whether the corresponding realizations of the random process converge deterministically. An alternative viewpoint is to fix the indexing variable i and consider how close the random variable $\widetilde{X}(i)$ is to another random variable X as we increase i.

A possible measure of the distance between two random variables is the mean square of their difference. If $\mathrm{E}((X - Y)^2) = 0$ then $X = Y$ with probability one by Chebyshev's inequality. The mean square deviation between $\widetilde{X}(i)$ and X is a deterministic quantity (a number), so we can evaluate its convergence as $i \to \infty$. If it converges to zero then we say that the random sequence converges in mean square.

Definition 6.1.3 (Convergence in mean square)**.** *A discrete random process $\widetilde{X}$ converges in mean square to a random variable X belonging to the same probability space if*

$$\lim_{i\to\infty} \mathrm{E}\left(\left(X - \widetilde{X}\left(i\right)\right)^2\right) = 0. \tag{6.6}$$

Alternatively, we can consider the probability that $\widetilde{X}\left(i\right)$ is separated from X by a certain fixed $\epsilon > 0$. If for any ϵ, no matter how small, this probability converges to zero as $i \to \infty$ then we say that the random sequence converges in probability.

Definition 6.1.4 (Convergence in probability)**.** *A discrete random process $\widetilde{X}$ converges in probability to another random variable X belonging to the same probability space if for any $\epsilon > 0$*

$$\lim_{i\to\infty} \mathrm{P}\left(\left|X - \widetilde{X}\left(i\right)\right| > \epsilon\right) = 0. \tag{6.7}$$

Note that as in the case of convergence in mean square, the limit in this definition is deterministic, as it is a limit of probabilities, which are just real numbers.

As a direct consequence of Markov's inequality, convergence in mean square implies convergence in probability.

Theorem 6.1.5. *Convergence in mean square implies convergence in probability.*

Proof. We have

$$\lim_{i\to\infty} \mathrm{P}\left(\left|X - \widetilde{X}\left(i\right)\right| > \epsilon\right) = \lim_{i\to\infty} \mathrm{P}\left(\left(X - \widetilde{X}\left(i\right)\right)^2 > \epsilon^2\right) \tag{6.8}$$

$$\leq \lim_{i\to\infty} \frac{\mathrm{E}\left(\left(X - \widetilde{X}\left(i\right)\right)^2\right)}{\epsilon^2} \quad \text{by Markov's inequality} \tag{6.9}$$

$$= 0, \tag{6.10}$$

if the sequence converges in mean square. □

It turns out that convergence with probability one also implies convergence in probability. Convergence in probability one does not imply convergence in mean square or vice versa. The difference between these three types of convergence is not very important for the purposes of this course.

6.1.3 Convergence in distribution

In some cases, a random process $\tilde{X}$ does not converge to the value of any random variable, but the cdf of $\tilde{X}\left(i\right)$ converges pointwise to the cdf of another random variable X. In that case, the actual values of $\tilde{X}\left(i\right)$ and X are *not* necessarily close, but in the limit they have the same *distribution*. In this case, we say that $\widetilde{X}$ converges in distribution to X.

Definition 6.1.6 (Convergence in distribution)**.** *A random process $\widetilde{X}$ converges in distribution to a random variable X belonging to the same probability space if*

$$\lim_{i\to\infty} F_{\widetilde{X}(i)}\left(x\right) = F_X\left(x\right) \tag{6.11}$$

for all $x \in \mathbb{R}$ where F_X is continuous.

Note that convergence in distribution is a much weaker notion than convergence with probability one, in mean square or in probability. If a discrete random process $\widetilde{X}$ converges to a random variable X in distribution, this only means that as i becomes large the distribution of $\widetilde{X}(i)$ tends to the distribution of X, not that *the values* of the two random variables are close. However, convergence in probability (and hence convergence with probability one or in mean square) does imply convergence in distribution.

Example 6.1.7 (Binomial converges to Poisson)**.** Let us define a discrete random process $\widetilde{X}(i)$ such that the distribution of $\widetilde{X}(i)$ is binomial with parameters i and $p := \lambda/i$. $\widetilde{X}(i)$ and $\widetilde{X}(j)$ are independent for $i \neq j$, which completely characterizes the n-order distributions of the process for all $n > 1$. Consider a Poisson random variable X with parameter λ that is independent of $\tilde{X}(i)$ for all i. Do you expect the values of X and $\tilde{X}(i)$ to be close as $i \to \infty$?

No! In fact even $\tilde{X}(i)$ and $\tilde{X}(i+1)$ will not be close in general. However, $\tilde{X}$ converges in distribution to X, as established in Example 2.2.8:

$$\lim_{i\to\infty} p_{\widetilde{X}(i)}(x) = \lim_{i\to\infty} \binom{i}{x} p^x (1-p)^{(i-x)} \tag{6.12}$$

$$= \frac{\lambda^x e^{-\lambda}}{x!} \tag{6.13}$$

$$= p_X(x). \tag{6.14}$$

△

6.2 Law of large numbers

Let us define the average of a discrete random process.

Definition 6.2.1 (Moving average)**.** *The moving or running average $\widetilde{A}$ of a discrete random process $\widetilde{X}$, defined for $i = 1, 2, \ldots$ (i.e. 1 is the starting point), is equal to*

$$\widetilde{A}(i) := \frac{1}{i} \sum_{j=1}^{i} \widetilde{X}(j). \tag{6.15}$$

Consider an iid sequence. A very natural interpretation for the moving average is that it is a real-time estimate of the mean. In fact, in statistical terms the moving average is the sample mean of the process up to time i (the sample mean is defined in Chapter 8). The law of large numbers establishes that the average does indeed converge to the mean of the iid sequence.

Theorem 6.2.2 (Weak law of large numbers)**.** *Let $\widetilde{X}$ be an iid discrete random process with mean $\mu_{\widetilde{X}} := \mu$ such that the variance of $\widetilde{X}(i)$ σ^2 is bounded. Then the average $\widetilde{A}$ of $\widetilde{X}$ converges in mean square to μ.*

Proof. First, we establish that the mean of $\widetilde{A}(i)$ is constant and equal to μ,

$$\mathrm{E}\left(\widetilde{A}(i)\right) = \mathrm{E}\left(\frac{1}{i}\sum_{j=1}^{i}\widetilde{X}(j)\right) \tag{6.16}$$

$$= \frac{1}{i}\sum_{j=1}^{i}\mathrm{E}\left(\widetilde{X}(j)\right) \tag{6.17}$$

$$= \mu. \tag{6.18}$$

Due to the independence assumption, the variance scales linearly in i. Recall that for independent random variables the variance of the sum equals the sum of the variances,

$$\mathrm{Var}\left(\widetilde{A}(i)\right) = \mathrm{Var}\left(\frac{1}{i}\sum_{j=1}^{i}\widetilde{X}(j)\right) \tag{6.19}$$

$$= \frac{1}{i^2}\sum_{j=1}^{i}\mathrm{Var}\left(\widetilde{X}(j)\right) \tag{6.20}$$

$$= \frac{\sigma^2}{i}. \tag{6.21}$$

We conclude that

$$\lim_{i\to\infty}\mathrm{E}\left(\left(\widetilde{A}(i)-\mu\right)^2\right) = \lim_{i\to\infty}\mathrm{E}\left(\left(\widetilde{A}(i)-\mathrm{E}\left(\widetilde{A}(i)\right)\right)^2\right) \quad \text{by (6.18)} \tag{6.22}$$

$$= \lim_{i\to\infty}\mathrm{Var}\left(\widetilde{A}(i)\right) \tag{6.23}$$

$$= \lim_{i\to\infty}\frac{\sigma^2}{i} \quad \text{by (6.21)} \tag{6.24}$$

$$= 0. \tag{6.25}$$

□

By Theorem 6.1.5 the average also converges to the mean of the iid sequence in probability. In fact, one can also prove convergence with probability one under the same assumptions. This result is known as the strong law of large numbers, but the proof is beyond the scope of these notes. We refer the interested reader to more advanced texts in probability theory.

Figure 6.2 shows averages of realizations of several iid sequences. When the iid sequence is Gaussian or geometric we observe convergence to the mean of the distribution, however when the sequence is Cauchy the moving average diverges. The reason is that, as shown in Example 4.2.2, the Cauchy distribution does not have a well defined mean! Intuitively, extreme values have non-negligeable probability under the Cauchy distribution so from time to time the iid sequence takes values with very large magnitudes and this prevents the moving average from converging.

6.3 Central limit theorem

In the previous section we established that the moving average of a sequence of iid random variables converges to the mean of their distribution (as long as the mean is well defined and

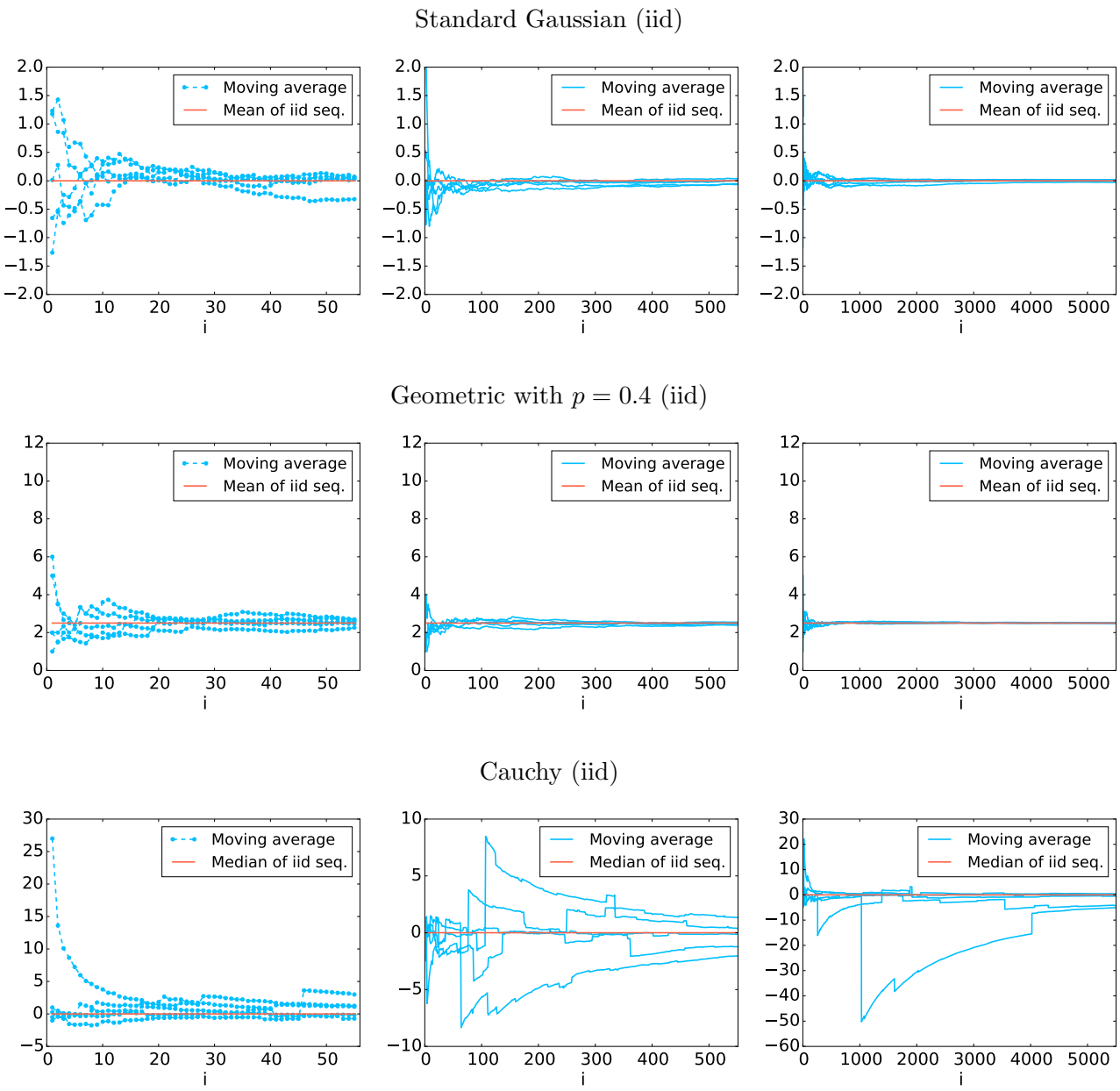


Figure 6.2: Realization of the moving average of an iid standard Gaussian sequence (top), an iid geometric sequence with parameter $p = 0.4$ (center) and an iid Cauchy sequence (bottom).

the variance is finite). In this section, we characterize the *distribution* of the average $\widetilde{A}(i)$ as i increases. It turns out that $\widetilde{A}$ converges to a Gaussian random variable in distribution, which is very useful in statistics as we will see later on.

This result, known as the central limit theorem, justifies the use of Gaussian distributions to model data that are the result of many different independent factors. For example, the distribution of height or weight of people in a certain population often has a Gaussian shape– as illustrated by Figure 2.13– because the height and weight of a person depends on many different factors that are roughly independent. In many signal-processing applications noise is well modeled as having a Gaussian distribution for the same reason.

Theorem 6.3.1 (Central limit theorem)**.** *Let $\widetilde{X}$ be an iid discrete random process with mean $\mu_{\widetilde{X}} := \mu$ such that the variance of $\widetilde{X}(i)$ σ^2 is bounded. The random process $\sqrt{n}(\widetilde{A} - \mu)$, which corresponds to the centered and scaled moving average of $\widetilde{X}$, converges in distribution to a Gaussian random variable with mean 0 and variance σ^2.*

Proof. The proof of this remarkable result is beyond the scope of these notes. It can be found in any advanced text on probability theory. However, we would still like to provide some intuition as to why the theorem holds. Theorem 3.5.2 establishes that the pdf of the sum of two independent random variables is equal to the convolutions of their individual pdfs. The same holds for discrete random variables: the pmf of the sum is equal to the convolution of the pmfs, as long as the random variables are independent.

If each of the entries of the iid sequence has pdf f, then the pdf of the sum of the first i elements can be obtained by convolving f with itself i times

$$f_{\sum_{j=1}^{1} \widetilde{X}(j)}(x) = (f * f * \cdots * f)(x). \tag{6.26}$$

If the sequence has a discrete state and each of the entries has pmf p, the pmf of the sum of the first i elements can be obtained by convolving p with itself i times

$$p_{\sum_{j=1}^{1} \widetilde{X}(j)}(x) = (p * p * \cdots * p)(x). \tag{6.27}$$

Normalizing by i just results in scaling the result of the convolution, so the pmf or pdf of the moving mean $\widetilde{A}$ is the result of repeated convolutions of a fixed function. These convolutions have a smoothing effect, which eventually transforms the pmf/pdf into a Gaussian! We show this numerically in Figure 6.3 for two very different distributions: a uniform distribution and a very irregular one. Both converge to Gaussian-like shapes after just 3 or 4 convolutions. The central limit theorem makes this precise, establishing that the shape of the pmf or pdf becomes Gaussian asymptotically. □

In statistics the central limit theorem is often invoked to justify treating averages as if they have a Gaussian distribution. The idea is that for large enough n $\sqrt{n}(\widetilde{A} - \mu)$ is approximately Gaussian with mean 0 and variance σ^2, which implies that $\widetilde{A}$ is approximately Gaussian with mean μ and variance σ^2/n. It's important to remember that we have *not* established this rigorously. The rate of convergence will depend on the particular distribution of the entries of the iid sequence.

In practice convergence is usually very fast. Figure 6.4 shows the empirical distribution of the moving average of an exponential and a geometric iid sequence. In both cases the approximation obtained by the central limit theory is very accurate even for an average of 100 samples. The

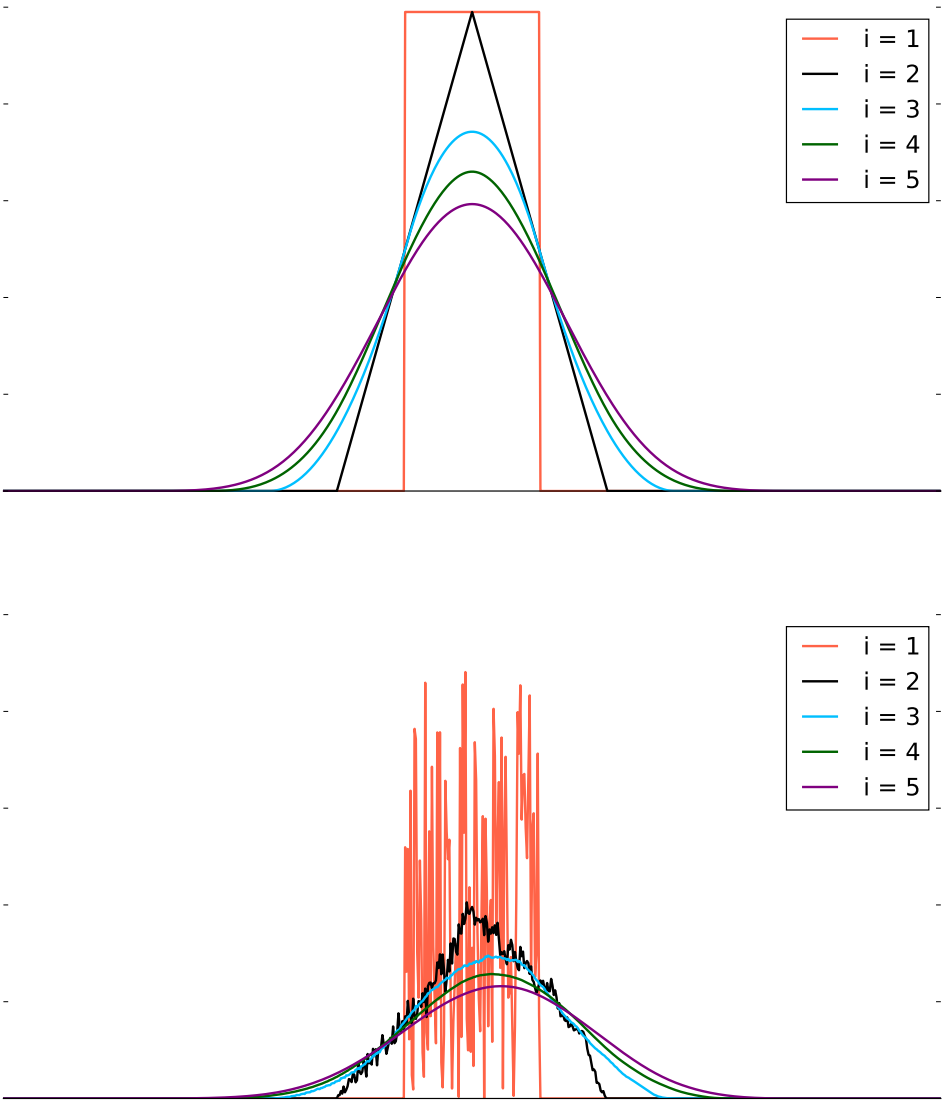


Figure 6.3: Result of convolving two different distributions with themselves several times. The shapes quickly become Gaussian-like.

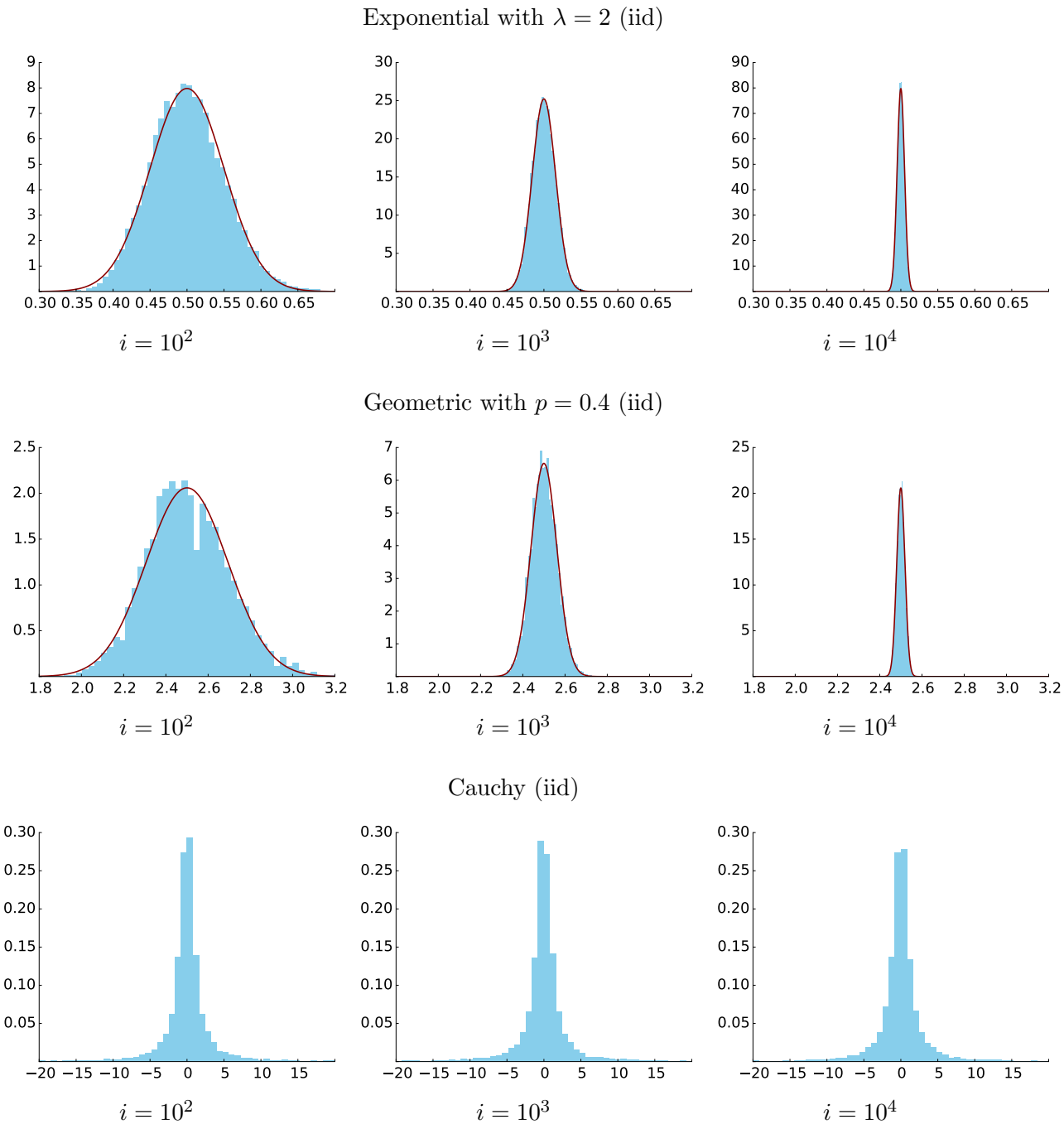


Figure 6.4: Empirical distribution of the moving average of an iid standard Gaussian sequence (top), an iid geometric sequence with parameter $p = 0.4$ (center) and an iid Cauchy sequence (bottom). The empirical distribution is computed from 10^4 samples in all cases. For the two first rows the estimate provided by the central limit theorem is plotted in red.

figure also shows that for a Cauchy iid sequence, the distribution of the moving average does not become Gaussian, which does not contradict the central limit theorem as the distribution does not have a well defined mean. To close this section we derive a useful approximation to the binomial distribution using the central limit theorem.

Example 6.3.2 (Gaussian approximation to the binomial distribution)**.** Let X have a binomial distribution with parameters n and p, such that n is large. Computing the probability that X is in a certain interval requires summing its pmf over all the values in that interval. Alternatively, we can obtain a quick approximation using the fact that for large n the distribution of a binomial random variable is approximately Gaussian. Indeed, we can write X as the sum of n independent Bernoulli random variables with parameter p,

$$X = \sum_{i=1}^{n} B_i. \tag{6.28}$$

The mean of B_i is p and its variance is $p\left(1-p\right)$. By the central limit theorem $\frac{1}{n}X$ is approximately Gaussian with mean p and variance $p\left(1-p\right)/n$. Equivalently, by Lemma 2.5.1, X is approximately Gaussian with mean np and variance $np\left(1-p\right)$.

Assume that a basketball player makes each shot she takes with probability $p = 0.4$. If we assume that each shot is independent, what is the probability that she makes more than 420 shots out of 1000? We can model the shots made as a binomial X with parameters 1000 and 0.4. The exact answer is

$$\begin{aligned} \mathrm{P}\left(X \geq 420\right) &= \sum_{x=420}^{1000} p_X\left(x\right) && (6.29)\\ &= \sum_{x=420}^{1000} \binom{1000}{x} 0.4^x 0.6^{(n-x)} && (6.30)\\ &= 10.4\,10^{-2}. && (6.31) \end{aligned}$$

If we apply the Gaussian approximation, by Lemma 2.5.1 X being larger than 420 is the same as a standard Gaussian U being larger than $\frac{420-\mu}{\sigma}$ where μ and σ are the mean and standard deviation of X, equal to $np = 400$ and $\sqrt{np(1-p)} = 15.5$ respectively.

$$\begin{aligned} \mathrm{P}\left(X \geq 420\right) &\approx \mathrm{P}\left(\sqrt{np\left(1-p\right)}U + np \geq 420\right) && (6.32)\\ &= \mathrm{P}\left(U \geq 1.29\right) && (6.33)\\ &= 1 - \Phi\left(1.29\right) && (6.34)\\ &= 9.85\,10^{-2}. && (6.35) \end{aligned}$$

△

6.4 Monte Carlo simulation

Simulation is a powerful tool in probability and statistics. Probabilistic models are often too complex for us to derive closed-form solutions of the distribution or expectation of quantities of interest, as we do in homework problems.

As an example, imagine that you set up a probabilistic model to determine the probability of winning a game of solitaire. If the cards are well shuffled, this probability equals

$$\mathrm{P}\left(\text{Win}\right) = \frac{\text{Number of permutations that lead to a win}}{\text{Total number}}. \tag{6.36}$$

The problem is that characterizing what permutations lead to a win is very difficult without actually playing out the game to see the outcome. Doing this for every possible permutation is computationally intractable, since there are $52! \approx 8\,10^{67}$ of them. However, there is a simple way to approximate the probability of interest: simulating a large number of games and recording what fraction result in wins. The game of solitaire was precisely what inspired Stanislaw Ulam to propose simulation-based methods, known as the Monte Carlo method (a code name, inspired by the Monte Carlo Casino in Monaco), in the context of nuclear-weapon research in the 1940s:

The first thoughts and attempts I made to practice (the Monte Carlo Method) were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays.

This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later, I described the idea to John von Neumann, and we began to plan actual calculations.[1]

Monte Carlo methods use simulation to estimate quantities that are challenging to compute exactly. In this section, we consider the problem of approximating the probability of an event $\mathcal{E}$, as in the *game of solitaire* example.

Algorithm 6.4.1 (Monte Carlo approximation)**.** *To approximate the probability of an event $\mathcal{E}$, we:*

1. *Generate n independent samples from the indicator function $1_{\mathcal{E}}$ associated to the event: I_1, I_2, $\ldots$, I_n.*
2. *Compute the average of the n samples*

$$\widetilde{A}\left(n\right) := \frac{1}{n}\sum_{i=1}^{n} I_i \tag{6.37}$$

which is the estimate for the probability of $\mathcal{E}$

The probability of interest can be interpreted as the expectation of the indicator function $1_{\mathcal{E}}$ associated to the event,

$$\mathrm{E}\left(1_{\mathcal{E}}\right) = \mathrm{P}\left(\mathcal{E}\right). \tag{6.38}$$

By the law of large numbers, the estimate $\widetilde{A}$ converges to the true probability as $n \to \infty$. The following example illustrates the power of this simple technique.

[1] http://en.wikipedia.org/wiki/Monte_Carlo_method#History

Game outcomes			Rank			Probability
1-2	1-3	2-3	R_1	R_2	R_3	
1	1	2	1	2	3	1/6
1	1	3	1	3	2	1/6
1	3	2	1	1	1	1/12
1	3	3	2	3	1	1/12
2	1	2	2	1	3	1/6
2	1	3	1	1	1	1/6
2	3	2	3	1	2	1/12
2	3	3	3	2	1	1/12

Probability mass function

	R_1	R_2	R_3
1	7/12	1/2	5/12
2	1/4	1/4	1/4
3	1/6	1/4	1/3

Table 6.1: The table on the left shows all possible outcomes in a league of three teams ($m = 3$), the resulting ranks for each team and the corresponding probability. The table on the right shows the pmf of the ranks of each of the teams.

Example 6.4.2 (Basketball league)**.** In an intramural basketball league m teams play each other once every season. The teams are ordered according to their past results: team 1 being the best and team m the worst. We model the probability that team i beats team j, for $1 \leq i < j \leq m$ as

$$\mathrm{P}\left(\text{team } j \text{ beats team } i\right) := \frac{1}{j-i+1}. \tag{6.39}$$

The best team beats the second with probability 1/2 and the third with probability 2/3, the second beats the third with probability 1/2, the fourth with probability 2/3 and the fifth with probability 3/4, and so on. We assume that the outcomes of the different games are independent.

At the end of the season, after every team has played with every other team, the teams are ranked according to their number of wins. If several teams have the same number of wins, then they share the same rank. For example, if two teams have the most wins, they both have rank 1, and the next team has rank 3. The goal is to compute the distribution of the final rank of each team in the league, which we model as the random variables $R_1, R_2, \ldots, R_m$. We have all the information to compute the joint pmf of these random variables by applying the law of total probability. As shown in Table 6.1 for $m = 3$, all we need to do is enumerate all the possible outcomes of the games and sum the probabilities of the outcomes that result in a particular rank.

Unfortunately, the number of possible outcomes grows dramatically with m. The number of games equals $m\left(m-1\right)/2$, so the possible outcomes are $2^{m(m-1)/2}$. When there are just 10 teams, this is larger than 10^{13}. Computing the exact distribution of the final ranks for leagues that are not very small is therefore very computationally demanding. Fortunately, Algorithm 6.4.1 offers a more tractable alternative: We can sample a large number of seasons n by simulating each game as a Bernoulli random variable with a parameter given by equation (6.39) and approximate the pmf using the fraction of times that each team ends up in each position. Simulating a whole season only requires sampling $m\left(m-1\right)/2$ games, which can be done very fast.

Table 6.2 illustrates the Monte Carlo approach for $m = 3$. The approximation is quite coarse if we only use $n = 10$ simulated seasons, but becomes very accurate when $n = 2,000$. Figure 6.5

Game outcomes			Rank		
1-2	1-3	2-3	R_1	R_2	R_3
1	3	2	1	1	1
1	1	3	1	3	2
2	1	2	2	1	3
2	3	2	3	1	2
2	1	3	1	1	1
1	1	2	1	2	3
2	1	3	1	1	1
2	3	2	3	1	2
1	1	2	1	2	3
2	3	2	3	1	2

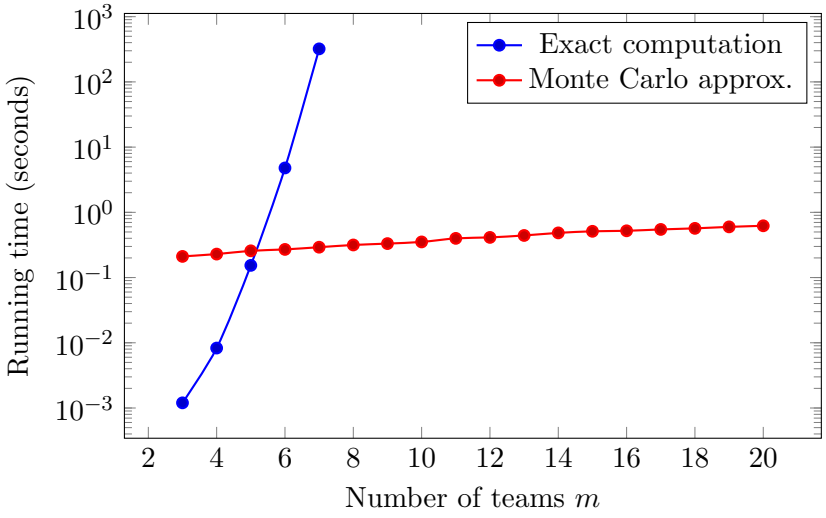
Estimated pmf ($n = 10$)

	R_1	R_2	R_3
1	0.6 (0.583)	0.7 (0.5)	0.3 (0.417)
2	0.1 (0.25)	0.2 (0.25)	0.4 (0.25)
3	0.3 (0.167)	0.1 (0.25)	0.3 (0.333)

Estimated pmf ($n = 2,000$)

	R_1	R_2	R_3
1	0.582 (0.583)	0.496 (0.5)	0.417 (0.417)
2	0.248 (0.25)	0.261 (0.25)	0.244 (0.25)
3	0.171 (0.167)	0.245 (0.25)	0.339 (0.333)

Table 6.2: The table on the left shows 10 simulated outcomes of a league of three teams ($m = 3$) and the resulting ranks. The tables on the right show the estimated pmf obtained by Monte Carlo simulation from the simulated outcomes on the left (top) and from 2,000 simulated outcomes (bottom). The exact values are included in brackets for comparison.



Monte Carlo error

m	Average error
3	$9.28\,10^{-3}$
4	$12.7\,10^{-3}$
5	$7.95\,10^{-3}$
6	$7.12\,10^{-3}$
7	$7.12\,10^{-3}$

Figure 6.5: The graph on the left shows the time needed to obtain the exact pmf of the final ranks in Example 6.4.2 and to approximate them by Monte Carlo approximation using 2,000 simulated league outcomes. The table on the right shows the average error per entry of the Monte Carlo approximation.

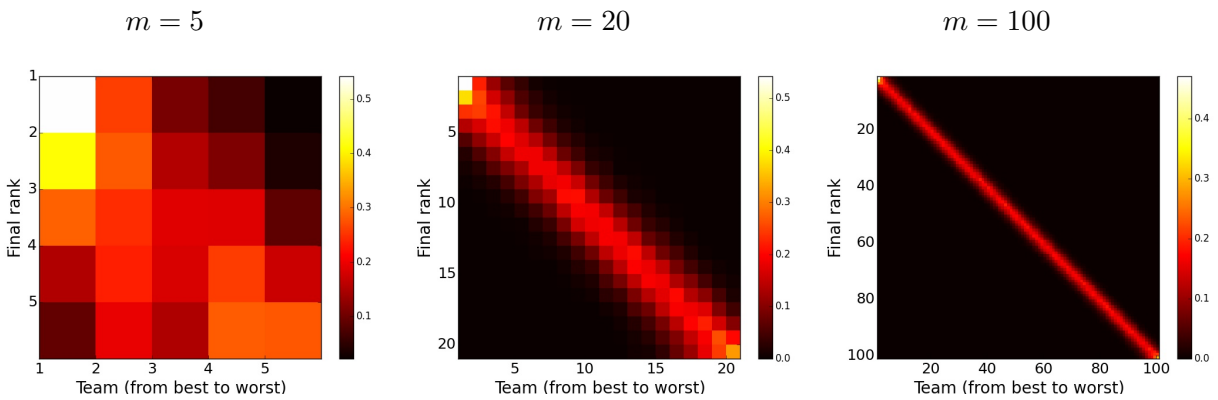


Figure 6.6: Approximate pmf of the final ranks in Example 6.4.2 using 2,000 simulated league outcomes.

shows the running time needed to compute the exact pmf and to approximate it with the Monte Carlo approach for different numbers of teams. When the number of teams is very small the exact computation is very fast, but the running time increases exponentially with m as expected, so that for 7 teams the computation already takes 5 and a half minutes. In contrast, the Monte Carlo approximation is dramatically faster. For $m = 20$ it just takes half a second. Figure 6.6 shows the approximate pmf of the final ranks for 5, 20 and 100 teams. Higher ranks have higher probabilities because when two teams are tied they are awarded the higher rank.

△