Algorithms and computations in physics (Oxford Lectures 2025)

Werner Krauth*
Laboratoire de Physique, Ecole normale supérieure, Paris (France)
Rudolf Peierls Centre for Theoretical Physics
University of Oxford (UK)

Special topics A, 28 Jan 2025

We supplement the first Lecture with two all-important special topics that explore the meaning of convergence in statistics, and the fundamental usefulness of statistical reasoning. We can discuss them here, in the direct-sampling framework, but they are more generally relevant. The strong law of large numbers, for example, will turn into the famous ergodic theorem for Markov chains.

Contents

1 Direct sampling (contd.)

Here, we continue our discussion of the first lecture, and treat a number of special subjects.

1.4 Direct sampling: Sums of random variables, limit theorems

Many fundamental aspects of sampling already manifest themselves in the direct-sampling framework, and then translate, *mutatis mutandis*, to the much more complicated Markov chains of subsequent lectures. For example, the strong law of large numbers that we need to understand for direct samples will turn into the famous ergodic theorem for Markov chains. We also discuss importance sampling that permeates all of Monte Carlo and discuss the frequentist interpretation of probabilities at the core of the method.

*werner.krauth@ens.fr, werner.krauth@physics.ox.ac.uk

1.4.1 Sums of random variables—Chebychev inequality: Weak law of large numbers

On the Monte Carlo beach, as we discussed in detail, the pebbles are "samples", and the hits (inside the circle) and the nonhits (in the square but outside the circle) represent random variables. However, what we computed was the average over many hits and nonhits, that is, the sum of random variables. Let's write the numer of hits "Nhits" as ξ and the individual stone-throws as ξ_i, to arrive at

$$\xi = \xi_1 + \cdots + \xi_N.$$

This, of course, opens up a Pandora's box of subjects in probability theory. To make it short, the expectation (mean value) of ξ is equal to the sum of the mean values of the individual terms (something that does not require the independence of the random variables),

$$\langle \xi \rangle = \langle \xi_1 \rangle + \cdots + \langle \xi_N \rangle \,.$$

Furthermore, for identical, independent random variables ξ_i (as we have them with the independent pebble-throws) with the same finite variance $\mathrm{Var}\,(\xi_i)$, we have

$$\begin{aligned} \mathrm{Var}\,(\xi_1 + \cdots + \xi_N) &= N\,\mathrm{Var}\,(\xi_i)\,, \\ \mathrm{Var}\left(\frac{\xi_1 + \cdots + \xi_N}{N}\right) &= \frac{1}{N}\,\mathrm{Var}\,(\xi_i)\,, \end{aligned}$$

where the quintessential variance is given by:

$$\mathrm{Var}\,(\xi) = \left\{ \begin{matrix} \text{average squared distance} \\ \text{from the mean value} \end{matrix} \right\} = \big\langle (\xi - \langle \xi \rangle)^2 \big\rangle \tag{17}$$

For concreteness, we shall now apply these two formulas to the children's game, with $N_{\mathrm{hits}} = \xi_1 + \cdots + \xi_N$:

$$\mathrm{Var}\,(N_{\mathrm{hits}}) = \left\langle \left(N_{\mathrm{hits}} - \frac{\pi}{4} N \right)^2 \right\rangle = N\,\mathrm{Var}\,\xi_i = N \overbrace{\theta(1-\theta)}^{0.169},$$

$$\mathrm{Var}\left(\frac{N_{\mathrm{hits}}}{N}\right) = \left\langle \left(\frac{N_{\mathrm{hits}}}{N} - \frac{\pi}{4} \right)^2 \right\rangle = \frac{1}{N}\,\mathrm{Var}\,(\xi_i) = \frac{\overbrace{\theta(1-\theta)}^{0.169}}{N}. \tag{18}$$

Here, we transform the definition of the variance, for the case of a distribution with zero mean, into a statement of probabilities:

$$\mathrm{Var}\,(\xi) = \int_{-\infty}^{\infty} \mathrm{d}x\; x^2 \pi(x) \geq \int_{|x|>\varepsilon} \mathrm{d}x\; x^2 \pi(x) \geq \varepsilon^2 \underbrace{\int_{|x|>\varepsilon} \mathrm{d}x\; \pi(x)}_{\substack{\text{prob. that} \\ |x-\langle x \rangle|>\epsilon}}.$$

This gives the famous Chebychev inequality

$$\left\{ \begin{matrix} \text{Chebyshev} \\ \text{inequality} \end{matrix} \right\}: \quad \left\{ \begin{matrix} \text{probability that} \\ |x - \langle x \rangle\,| > \varepsilon \end{matrix} \right\} < \frac{\mathrm{Var}\,\xi}{\varepsilon^2}. \tag{19}$$

The weak law of large numbers follows (with an upper bound of $1/4$ for the variance):

$$\left\{ \begin{matrix} \text{weak law of} \\ \text{large numbers} \end{matrix} \right\}: \quad \left\{ \begin{matrix} \text{probability that} \\ |N_{\mathrm{hits}}/N - \pi/4| < \varepsilon \end{matrix} \right\} > 1 - \frac{1}{4\varepsilon^2 N}.$$

In this equation, we can keep the interval parameter ε fixed. The probability inside the interval approaches 1 as $N \to \infty$. We can also bound the interval containing, say, 99% of the probability, as a function of N. Setting $0.01 = 1/(4\epsilon^2 N$, we arrive at

$$\left\{\begin{matrix}\text{size of interval containing} \\ 99\% \text{ of probability}\end{matrix}\right\}: \quad \epsilon < \frac{5}{\sqrt{N}}.$$

Chebyshev's inequality shows that a (finite) variance plays the role of a scale delimiting an interval of probable values of x: whatever the distribution, it is improbable that a sample will be more than a few standard deviations away from the mean value. This basic regularity property of distributions with a finite variance must be kept in mind in practical calculations. In particular, we must keep this property separate from the

1.4.2 Strong laws of large numbers

The weak law of large numbers, that we just discussed, is quite misleading in the context of Markov-chain calculations (we will discuss this in later lectures). What we need is the strong law of large numbers, that we discuss in the context of the Gamma integral that we discussed in Lecture 1. It keeps its meaning, in the form of the ergodic theorem, for the Markov chains of the second part of lesson 2, and the rest of the lecture series. Here it is, the γ integral:

$$I(\gamma) = \int_0^1 \mathrm{d}x \; x^\gamma = \frac{1}{\gamma+1} \quad \text{for } \gamma > -1 \tag{20}$$

(see [1, Sect. 1.4.2] for the full context). We attempt to compute the integral in a sample space $\Omega^{[0,1]}$, the unit interval between 0 and 1.

$$I(\gamma) = \int_0^1 \mathrm{d}x \; x^\gamma = \int_0^1 \underbrace{(1\mathrm{d}x\,)}_{x=\texttt{ran}(0,1)} \overbrace{x^\gamma}^{\mathcal{O}} \tag{21}$$

As we discussed before, the random variable $\mathcal{O}$ has its own probability distribution:

$$\pi(\mathcal{O}) = (\alpha-1)\mathcal{O}^{-\alpha}, \tag{22}$$

with $\alpha = 1 - 1/\gamma$. Its mean value of the random variable $\mathcal{O}$ can be equivalently written with $\pi(\mathcal{O})$ or in the original sample space:

$$\langle \mathcal{O} \rangle = (\alpha-1)\int_1^\infty \mathrm{d}\mathcal{O} \; \mathcal{O}\mathcal{O}^{-\alpha} = \int_0^1 \mathrm{d}x x^\gamma. \tag{23}$$

The same holds for any higher moments.

After these preliminaries, let us now actually compute the γ integral with a running average of a sum of uniform random numbers to the power of γ (see Algorithm 8). We compute the integral $I(\gamma)$ as the mean value of x^γ, with $x = \texttt{ran}(0,1)$, and likewise "try" to compute the error through the Gaussian error formula

$$\text{Error} \stackrel{?}{=} \frac{\sqrt{\langle \mathcal{O}^2 \rangle - \langle \mathcal{O} \rangle^2}}{\sqrt{N}} \tag{24}$$

This is implemented in Alg. 8 (`direct-gamma`). The calculation works well, most of the time, but is clearly in trouble for $\gamma = -0.8$. Nevertheless, as the integral $I(\gamma)$ exists for $\gamma > -1$, we

can rely on the strong law of large numbers which states that:

$$\mathbb{P}\left[\lim_{i\to\infty}\frac{1}{i}\Sigma_i = I(\gamma)\right] = 1 \tag{25}$$

This radical theorem tells us that, with probability 1, we can build little boxes, as in Fig. 1, and the running average will never leave this box until $i = \infty$. The strong law of large numbers (best described in a blog post of Terence Tao) shows us that that an individual trajectory of running averages converges to the ensemble mean. Any fluctuations mean that the data are not yet as good as they eventually will become.

procedure `direct-gamma`
$\Sigma \leftarrow 0;\ \Sigma_{\rm sq} \leftarrow 0$
for $i = 1, \ldots, N$**:**
$\left\{\begin{array}{l} x_i \leftarrow \texttt{ran}(0,1) \\ \Sigma \leftarrow \Sigma + x_i^\gamma \text{ (running average: } \Sigma/i) \\ \Sigma_{\rm sq} \leftarrow \Sigma_{sq} + x_i^{2\gamma} \end{array}\right.$
Error $\leftarrow \left[\Sigma_{sq}/N - (\Sigma/N)^2\right]/\sqrt{N}$ (⚠)
output $\Sigma/N \pm$ Error

Algorithm 8: `direct-gamma`. Computing the γ-integral in eq. (20) by direct sampling.

γ	$\Sigma/N \pm$ Error	$1/(\gamma+1)$
2.0	0.334 ± 0.003	$0.333\ldots$
1.0	0.501 ± 0.003	0.5
0.0	1.000 ± 0.000	1
−0.2	1.249 ± 0.003	1.25
−0.4	1.682 ± 0.014	$1.666\ldots$
−0.8	3.959 ± 0.110	5.0

Table 1: Output of Alg. 8 (`direct-gamma`) for various values of γ ($N = 10\,000$, standard empirical error shown). The computation for $\gamma = -0.8$ is in trouble.

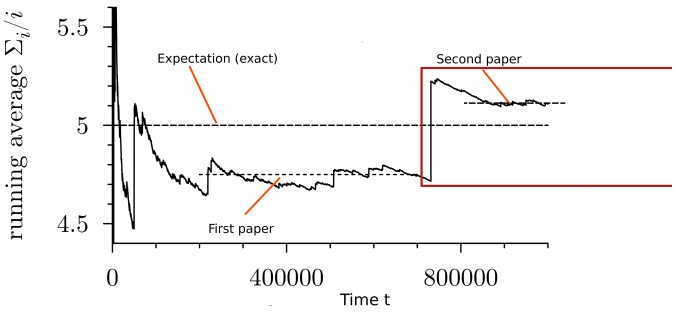

Figure 1: Running average of Alg. 8 (`direct-gamma`) for $\gamma = -0.8$. The strong law of large numbers guarantees that any individual trajectory of running averages Σ_i/i converges *almost surely* towards 5, that is, that, for any $\epsilon > 0$ we can draw red boxes, starting at i_ϵ, from which the running average will never exit.

γ	ζ	Σ/N	$\frac{\zeta+1}{\gamma+1}$
-0.4	0.0	1.685 ± 0.017	1.66
-0.6	-0.4	1.495 ± 0.008	1.5
-0.7	-0.6	1.331 ± 0.004	1.33
-0.8	-0.7	1.508 ± 0.008	1.5

Table 2: Output of Alg. 9 (`direct-gamma-zeta`) with $N = 10\,000$. All pairs $\{\gamma, \zeta\}$ satisfy $2\gamma - \zeta > -1$ so that $\langle \mathcal{O}^2 \rangle < \infty$.

1.4.3 Importance sampling

As we have seen, the Gamma integral is in trouble for $\gamma = -0.8$ although, in the limit of very long times, it will converge towards a mean value of 5. More generally, sampling problems where exceptional values of the observable $\mathcal{O}$ play too much of a role, are problematic. This situation does not have to be accepted, because of the concept of importance sampling. There,

$$I(\gamma) = \int_0^1 \mathrm{d}x \; x^\gamma = \int_0^1 \underbrace{(1\mathrm{d}x\;)}_{x=\texttt{ran}(0,1)} \overbrace{x^\gamma}^{\mathcal{O}} = \cdots = \int_0^1 x^\zeta \mathrm{d}x \; x^{\gamma-\zeta} \tag{26}$$

The problem has to do that for $\gamma < -\frac{1}{2}$, the variance of $\mathcal{O}$ is infinite, so that the Gaussian error analysis no longer applies. A modified program, Alg. 9 (`direct-gamma-zeta`), uses importance sampling to reduce the non-Gaussian fluctuations. It computes, not $I(\gamma)$, but $I(\gamma)/I(\zeta)$, simply because the integral $\int_0^1 x^\zeta \mathrm{d}x$ is not correctly normalized:

$$\Sigma/N = \frac{1}{N}\sum_{i=1}^N \mathcal{O}_i \simeq \langle \mathcal{O} \rangle = \frac{\int_0^1 \mathrm{d}x \; \pi(x)\mathcal{O}(x)}{\int_0^1 \mathrm{d}x \; \pi(x)}$$
$$= \frac{\int_0^1 \mathrm{d}x \; x^\zeta x^{\gamma-\zeta}}{\int_0^1 \mathrm{d}x \; x^\zeta} = \frac{\int_0^1 \mathrm{d}x \; x^\gamma}{\int_0^1 \mathrm{d}x \; x^\zeta} = \frac{I(\gamma)}{I(\zeta)} = \frac{\zeta+1}{\gamma+1}. \tag{27}$$

The calculation of the variance now involves the integral

$$\int_0^1 x^\zeta \mathrm{d}x \; \left(x^{\gamma-\zeta}\right)^2, \tag{28}$$

which is finite if $\zeta + 2(\gamma - \zeta) > -1$. (see [1, eq. 1.74] for details).

procedure `direct-gamma-zeta`
$\Sigma \leftarrow 0$
for $i = 1, \ldots, N$**:**
$\quad x_i \leftarrow \texttt{ran}(0,1)^{1/(\zeta+1)}$ $(\pi(x_i) \propto x_i^\zeta$, see eq. (??))
$\quad \Sigma \leftarrow \Sigma + x_i^{\gamma-\zeta}$
output Σ/N

Algorithm 9: `direct-gamma-zeta`. Using importance sampling to compute the γ-integral (see eq. (27)).

1.5 Statistics and the finite number of samples

What can we learn (about π) by throwing pebbles? This question about the role and the value of statistics, in other contexts, and there is a lot of heat in the battle. Let us add a little light. Our first Monte Carlo simulation, on the Monte Carlo beach, generated 3156 hits for 4000 trials (see Lecture 1). We shall now see what this result tells us about π, without adding any assumptions about N being large, the distribution being almost Gaussian, etc.. We discuss here the frequentist approach to statistics, and point to where the Bayesian approach, which is weaker, but more malleable.

1.5.1 Frequentist statistics on the Monte Carlo beach

In the children's Monte Carlo game, hits and nonhits were generated by the Bernoulli distribution:

$$\xi_i = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{with probability } (1-\theta) \end{cases}, \tag{29}$$

but of the value $\pi/4 = \theta = \langle \xi_i \rangle$, we only know that it is between 0 and 1, and that its variance $\theta(1-\theta)$ cannot exceed $1/4$. Instead of the original variables ξ_i, we consider random variables η_i shifted by this unknown mean value:

$$\eta_i = \xi_i - \theta.$$

The shifted random variables η_i now have zero mean and the same variance as the original variables ξ_i:

$$\langle \eta_i \rangle = 0, \quad \text{Var}\, \eta_i = \text{Var}\, \xi_i = \theta(1-\theta) \leq \frac{1}{4}.$$

Without invoking the limit $N \to \infty$, we can use the Chebyshev inequality to obtain an interval around zero containing at least 68% of the probability. This goes as follows: With 68% probability and remembering that $0.68 + 0.32 = 1$, we have

$$\frac{\text{Var}\,(\xi)}{\epsilon^2} = 0.32 \implies \epsilon = \sqrt{\frac{\theta(1-\theta)}{0.32N}} = \frac{1}{2\sqrt{N}}\text{`}\frac{1.77}{2\sqrt{4000}} = 0.014. \tag{30}$$

Therefore, with more than 68% probability, we have

$$-0.014 < \frac{1}{4000}\sum_i \eta_i < 0.014 \tag{31}$$

$$-0.014 < -0.789 + \frac{\pi}{4} < 0.014 \tag{32}$$

$$-0.014 + 0.789 < \frac{\pi}{4} < 0.014 + 0.789 \tag{33}$$

This has implications for the difference between our experimental result, 0.789 and the mathematical constant π. The difference between the two, with more than 68% chance, is smaller than 0.014:

$$\frac{\pi}{4} = 0.789 \pm 0.014 \Leftrightarrow \pi = 3.156 \pm 0.056, \tag{34}$$

where the value 0.056 is an upper bound for the 68% confidence interval that in physics is called an error bar (there are better estimates using the Hoeffding inequality). The quite abstract reasoning leading from eq. (29) to eq. (34)—in other words from the experimental result 3156 to the estimate of π with an error bar—is extremely powerful, and rarely understood. To derive the error bar, we did not use the central limit theorem. No limit $N \to \infty$ was implied, no Gaussian approximation was made. We only used an upper bound for the variance, and we supposed that our random numbers were perfect. With this, we obtain the marvellous result that, among an infinite number of beach parties, at which participants would play the same game of 4000 as we

described in Lecture 1 and which would yield Monte Carlo results analogous to ours, more than 68% would hold the mathematical value π inside their error bars. In arriving at this result, we did not treat the number π as a random variable—that would be inappropriate, because π is a mathematical constant.

1.5.2 Bayesian statistics on the Monte Carlo beach

The Bayesian approach to statistics is treated in Ref. [1, Section 1.3.3].

References

[1] W. Krauth, *Statistical Mechanics: Algorithms and Computations*. Oxford University Press, 2006.