Appendix A20

Averaging

HERE WE REVIEW some elementary notions of probability theory that will be useful to you no matter what you do with the rest of your life.

A20.1 Moments, cumulants

The exact *mean* μ (or expectation or expected value $\mathbb{E}[a]$) is the integral of the random variable a with respect to its probability measure ρ, commonly denoted

$$\mu = \mathbb{E}[a] = \langle a \rangle = \int_{\mathcal{M}} dx\, \rho(x)\, a(x)\,. \tag{A20.1}$$

In ChaosBook we use $\langle \cdots \rangle_\rho$ or simply $\langle \cdots \rangle$ to denote an integral over state space weighted by ρ, while $\overline{\cdots}$ denotes a time average. If the average is over a (finite or infinite) set of states labeled by discrete labels π, each state contributing with a weight t_π, the expectation is given by

$$\langle a \rangle = \sum_\pi a_\pi t_\pi\,, \tag{A20.2}$$

with probabilities in either case normalized so that $\langle 1 \rangle = 1$.

The expectation $\langle a^k \rangle$ is called the kth *moment*. The first moment is the mean μ defined in (A20.1). For $k > 1$, it is more natural to consider the moments about the mean, $\langle (a - \langle a \rangle)^k \rangle$, called *central moments*. The second, and all-important central moment is known as the *variance*,

$$\sigma^2 = \langle (a - \langle a \rangle)^2 \rangle = \langle a^2 \rangle - \langle a \rangle^2\,, \tag{A20.3}$$

or, in probabilist notation,

$$\mathbb{E}[a^2] = \mu^2 + \sigma^2\,. \tag{A20.4}$$

Its positive square root σ is called the *standard deviation* σ. As a mnemonic, think of the width of a Gaussian being $\approx 2\sigma$.

Standardized moment

$$\langle (a - \langle a \rangle)^k \rangle / \sigma^k \tag{A20.5}$$

is the kth central moment divided by σ^k, a dimensionless representation of the distribution of variance 1, independent of translations and linear changes of scale.

Moments can be collected into the (exponential) moment-generating function

question A20.1

$$\langle e^{\beta a} \rangle = 1 + \sum_{k=1}^{\infty} \frac{\beta^k}{k!} \langle a^k \rangle \,. \tag{A20.6}$$

Why the prefactor $1/k!$ (a Taylor series), and not $1/k$ (a logarithmic series), or 1 (discrete Laplace transform or Z-transform) generating function? In statistical, stochastic and quantum mechanics / quantum field theory applications one is solving linear ODEs or PDEs, and their solutions are always exponential in form.

Hardly any experiment measures a^k for $k > 2$ -that might require a lot of data- and raising approximate numbers to high powers is not smart: if $|a| < 1$, a^k gets very small very fast, and conversely If $|a| > 1$, a^k gets very big. Still, with a bit of hindsight, one finds that moments do play a natural, fundamental role if folded into the *cumulant-generating function*

$$\ln \langle e^{\beta a} \rangle = \sum_{k=1}^{\infty} \frac{\beta^k}{k!} \langle a^k \rangle_c \,, \tag{A20.7}$$

where the subscript c indicates a *cumulant*, or, in statistical mechanics and quantum field theory contexts, the 'connected Green's function'. Were $\langle a^k \rangle = \langle a \rangle^k$, we would have only one term in the series (A20.7), $\ln \langle e^{\beta a} \rangle = \ln e^{\beta \langle a \rangle} = \beta \langle a \rangle$, and that would be that. So cumulants $\langle a^k \rangle_c$ measure fluctuations about the mean $\langle a \rangle$. Indeed, expanding the logarithm of the series (A20.6), it is easy to check that the first cumulant is the mean, the second is the variance,

$$\langle a^2 \rangle_c = \langle (a - \langle a \rangle)^2 \rangle = \langle a^2 \rangle - \langle a \rangle^2 = \sigma^2 \,, \tag{A20.8}$$

and $\langle a^3 \rangle_c$ is the third central moment, or the *skewness*,

$$\langle a^3 \rangle_c = \langle (a - \langle a \rangle)^3 \rangle = \langle a^3 \rangle - 3 \langle a^2 \rangle \langle a \rangle + 2 \langle a \rangle^3 \,. \tag{A20.9}$$

The higher cumulants, however, are *not* central moments. The fourth cumulant,

$$\begin{aligned} \langle a^4 \rangle_c &= \langle (a - \langle a \rangle)^4 \rangle - 3 \langle (a - \langle a \rangle)^2 \rangle^2 \\ &= \langle a^4 \rangle - 4 \langle a^3 \rangle \langle a \rangle - 3 \langle a^2 \rangle^2 + 12 \langle a^2 \rangle \langle a \rangle^2 - 6 \langle a \rangle^4 \,, \end{aligned} \tag{A20.10}$$

rewritten in terms of standardized moments, is known as the *kurtosis*:

$$\frac{1}{\sigma^4} \langle a^4 \rangle_c = \frac{1}{\sigma^4} \langle (a - \langle a \rangle)^4 \rangle - 3 \,. \tag{A20.11}$$

The deep reason why cumulants are preferable to moments is that for a normalized Gaussian distribution all cumulants beyond the second one vanish, so they are a measure of deviation of statistics from Gaussian (see example 24.3). For a 'free' or 'Gaussian' field theory the only non-vanishing cumulant is the second one; for field theories with interactions the derivatives of $\ln\langle\exp(\beta a)\rangle$ with respect to β then yield cumulants, or the Burnett coefficients (24.14), or 'effective' n-point Green functions or n-point correlations.

exercise A20.1
question A20.2

So, what's so special about Gaussians?

example A20.1
p. 950

A20.1.1 Covariance matrix

For a multi-component observable the second central moment is called the *covariance matrix*

$$Q_{ij} = \langle (a_i - \langle a_i \rangle)(a_j - \langle a_j \rangle) \rangle \,. \tag{A20.12}$$

As Q is a symmetric, diagonalizable matrix, with eigenvalues σ_k^2 and orthogonal eigenvectors $\mathbf{e}^{(k)}$, you can visualize such *multivariate normal distribution* as a cigar-shaped cloud of points, with orthonormal principal axes of standard deviation (singular value) lengths σ_k. A cigar fat in a few directions, negligibly thin in the remaining directions motivates reduced-dimensional, linear modeling of the data by a retaining only a hyperplane spanned by the dominant directions; depending on the community, this is called the principal component analysis (PCA), the proper orthogonal decomposition (POD), the singular value decomposition (SVD), or the Karhunen–Loéve transform.

section 6.1

A20.1.2 Empirical means

Given a set of N iid (independently identically distributed) data samples $\{a_i\}$, where "iid" means that probability measures ρ factorize,

$$\rho(a_i, a_j) = \rho(a_i)\,\rho(a_j)\,, \qquad i \neq j\,, \tag{A20.13}$$

the *empirical mean* of observable a is the average

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} a_i \,. \tag{A20.14}$$

example A20.2
p. 951

$\hat{\mu}$ is *unbiased* if $\mathbb{E}[\hat{\mu}] = \mu$; we verify that in example A20.3. However, the *unbiased sample variance* $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ of observable a is defined differently, as

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \hat{a})^2 \,. \tag{A20.15}$$

What's up with the $N - 1$ divisor? See

example A20.3 p. 953 example A20.4 p. 953

Commentary

Question A20.1. Henriette Roux asks
Q Isn't expectation value (A20.6) the *characteristic function*?
A With imaginary exponent, $\beta \to it$ and the observable defined in the momentum space, $a = a(p)$, expectation value (A20.6) does have the form of a characteristic function, i.e., the Fourier transform of the probability density function

$$\mathbb{E}[e^{ip\,x}] = \int_{\mathcal{M}} dx\,\rho(x)\,e^{ipx}\,. \tag{A20.16}$$

Remark A20.1. Gaussian integrals. Kadanoff [2] has a nice discussion of Gaussian integrals, the central limit theorem and large deviations in *Chap. 3 Gaussian Distributions*, available online here.

section A20.1

Question A20.2. Henriette Roux muses
Q Somehow cumulants seem to fill my head with ideas — only I don't exactly know what they are!
A A scholarly aside, safely ignored, on where the characteristic state function $s(\beta)$ (20.10) fits into the grander scheme of things: in statistical mechanics and field theory, the partition function and the Helmholtz free energy have form

$$Z(\beta) = \exp(-\beta F)\,, \quad F(\beta) = -\frac{1}{\beta}\ln Z(E)\,, \tag{A20.17}$$

so in that sense $\langle e^{\beta a}\rangle$ is a 'partition function', and $s(\beta)$ in (20.10) is the associated 'free energy'. Expanding the logarithm of the series (A20.6) is easy for the first few terms, but it quickly gets old. The smart way to do this, explained in ref. [1], is to write down the the Dyson-Schwinger equations that generate recursively the terms in the Helmholtz free energy expansion (connected Green's functions), and Gibbs free energy (1-particle irreducible Green's functions).

References

[1] P. Cvitanović, *Field Theory*, Notes prepared by E. Gyldenkerne (Nordita, Copenhagen, 1983).

[2] L. P. Kadanoff, *Statistical Physics: Statics, Dynamics and Renormalization* (World Scientific, Singapore, 2000).

A20.2 Examples

Example A20.1. Gaussian minimizes information. Shannon information entropy is given by

$$S[\rho] = -\langle \ln \rho \rangle = -\int_{\mathcal{M}} dx\, \rho(x)\, \ln \rho(x)\,, \tag{A20.18}$$

where ρ is a probability density. Shannon thought of $-\ln \rho$ as 'information', very roughly in the sense that if -for example- $\rho(x) = 2^{-6}$, it takes $-\ln \rho = 6$ binary bits of 'information' to specify the probability density ρ at the point x. Information entropy (A20.18) is the expectation value of (or average) information.

A function $\rho \geq 0$ is an arbitrary function, of which we only require that it is normalized as a probability,

$$\int_{\mathcal{M}} dx\, \rho(x) = 1\,, \tag{A20.19}$$

has a mean value,

$$\int_{\mathcal{M}} dx\, x\rho(x) = \mu\,, \tag{A20.20}$$

and has a variance

$$\int_{\mathcal{M}} dx\, x^2\rho(x) = \mu^2 + \sigma^2\,. \tag{A20.21}$$

As ρ can be arbitrarily wild, it might take much "information" to describe it. Is there a function $\rho(x)$ that contains the *least* information, i.e., that minimizes the information entropy (A20.18)?

To find it, we minimize (A20.18) subject to constraints (A20.19)-(A20.21), implemented by adding Lagrange multipliers λ_j

$$\begin{aligned} C[\rho] \;=\; & \int_{\mathcal{M}} dx\, \rho(x)\, \ln \rho(x) \\ & +\lambda_0 \left(\int_{\mathcal{M}} dx\, \rho(x) - 1 \right) + \lambda_1 \left(\int_{\mathcal{M}} dx\, x\rho(x) - \mu \right) \\ & +\lambda_2 \left(\int_{\mathcal{M}} dx\, x^2\rho(x) - \mu^2 - \sigma^2 \right), \end{aligned} \tag{A20.22}$$

and looking for the extremum $\delta C = 0$,

$$\frac{\delta C[\rho]}{\delta\rho(x)} = (\ln \rho(x) + 1) + \lambda_0 + \lambda_1 x + \lambda_2 x^2 = 0\,, \tag{A20.23}$$

so

$$\rho(x) = e^{-(1+\lambda_0+\lambda_1 x+\lambda_2 x^2)}\,. \tag{A20.24}$$

The Lagrange multipliers λ_j can be expressed in terms of distribution parameters μ and σ by substituting this $\rho(x)$ into the constraint equations (A20.19)-(A20.21). We find that the probability density that minimizes information entropy is the Gaussian

$$\rho(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x-\mu)^2}{2\sigma^2}}\,. \tag{A20.25}$$

Participant	Stress (X)	Satisfaction (Y)
1	11	7
2	25	1
3	19	4
4	7	9
5	23	2
6	6	8
7	11	8
8	22	3
9	25	3
10	10	6

Table A20.1: Stress (1 to 30 scale) vs. happiness (1 to 10 scale) for a sample of 10 participants.

In what sense is that the distribution with the 'least information'? As we saw in the derivation of the cumulant expansion eq. (20.17), for a Gaussian distribution all cumulants but the mean μ and the variance σ^2 vanish, it is a distribution specified by only two 'informations', the location of its peak and its width.

click to return: p. 948

Sara A. Solla

Example A20.2. I get stress, but I can't get no satisfaction. A group of participants in a study of the correlation between stress and life satisfaction completed a questionnaire on how stressed they felt, and how satisfied they felt with their lives. Participants' scores are given in table A20.1.

We start our statistical analysis in the usual way, by evaluating the empirical means (A20.14) of the stress and satisfaction,

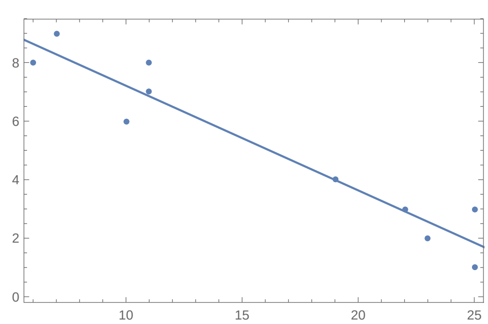
$$\hat{\mu}_X = \frac{1}{10}\sum_{i=1}^{10} X_i = 15.9\,, \qquad \hat{\mu}_Y = \frac{1}{10}\sum_{i=1}^{10} Y_i = 5.1\,,$$

and the *unbiased* variances (A20.15) and standard deviations,

$$\hat{\sigma}_X^2 = \frac{1}{10-1}\sum_{i=1}^{10}(X_i - \hat{\mu}_X)^2 = 58.1\,, \quad \hat{\sigma}_X = 7.6\,.$$

$$\hat{\sigma}_Y^2 = \frac{1}{10-1}\sum_{i=1}^{10}(Y_i - \hat{\mu}_Y)^2 = 8.1\,, \quad \hat{\sigma}_Y = 2.8\,.$$

The means are halfway their respective ranges, but the standard deviations are huge, they span across the available ranges. To figure out what is going on, one should always start with *visualizing* the data:



So the empirical means are meaningless - the subjects are either unhappy or happy, there is nobody in between. Standard deviations of such bimodal distributions are not helpful either, as they are measuring deviations from the non-existent average participant. However, the linear fit

$$Y = 11 - .36\,X \tag{A20.26}$$

is pretty good. Clearly this is 2-dimensional data set, so we compute the stressed/happy covariance

$$Q_{XY} = \frac{1}{10-1}\sum_{i=1}^{10}(X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y) = -20.8\,.$$

The ellipsoid given by the covariance matrix

$$\mathrm{Q} = \begin{pmatrix} \hat{\sigma}_X^2 & Q_{XY} \\ Q_{XY} & \hat{\sigma}_Y^2 \end{pmatrix},$$

with singular values (square roots of eigenvalues) and eigenvectors

$$\{\sigma_1, \sigma_2\} = \{8.10, 0.76\} \; : \quad \mathbf{e}_{(1)} = (0.94, -0.34)\,, \quad \mathbf{e}_{(2)} = (0.34, 0.94)\,,$$

gives a good description of the data, aligned along $\mathbf{e}_{(1)}$ (of slope close to the linear fit (A20.26)), with small transverse fluctuations along $\mathbf{e}_{(2)}$. The only problem is that we are plotting lemons vs. roses.

For this reason, statisticians like to study pairwise Pearson *correlation coefficients*, such as

$$\rho_{XY} = \frac{Q_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y} = -0.9573\,,$$

for which are the deviation $1 - |\rho_{XY}|$ is a measure for how well the data is fit by a linear fit.

One might be tempted to study the full *correlation coefficients matrix*, a somewhat contrived "standardized" or "whitened" rescaling (A20.5) of the covariance matrix (A20.12),

$$\mathrm{Corr}(X,Y) = \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_X & 0 \\ 0 & \hat{\sigma}_Y \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_X^2 & Q_{XY} \\ Q_{XY} & \hat{\sigma}_Y^2 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_X & 0 \\ 0 & \hat{\sigma}_Y \end{pmatrix}^{-1}.$$

Its eigenvalues $\{1 + \rho_{XY}, 1 - \rho_{XY}\}$ and eigenvectors are a *dimensionless* least-squares fit to the data, with the ellipsoid's principal axes along the diagonals

$$\{\sigma_1, \sigma_2\} = \{1.40, 0.207\} \; : \quad \frac{1}{\sqrt{2}}\mathbf{e}_{(1)} = (1, -1)\,, \quad \frac{1}{\sqrt{2}}\mathbf{e}_{(2)} = (1, 1)\,.$$

The pairwise *correlation coefficients* have some utility in singling out directions and signs of slopes in which data is nearly linear ($\rho_{X_iX_j}$ close to ± 1), The transformation from the covariance matrix to the correlations matrix is not a similarity transformation, so while the covariance matrix is a fundamental object in the multivariate cumulant expansions, the correlation matrices are not used in physics, only the 2-dimensional planes spanned by $\rho_{X_iX_j}$ are informative.

Bonus reading: the article on "The Economist" (if you can get past the paywall), or, more seriously, D. Kahneman and A. Deaton -the 2002 Nobel Memorial Prize in Economic Sciences- about the correlation between income and happiness. Penny for your thoughts.

click to return: p. 948

I. Taboada, P. Cvitanović & S.A. Solla

Example A20.3. Unbiased sample variance.

Why is the empirical estimate for the *unbiased sample variance*

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_i-\hat{\mu})^2 \tag{A20.27}$$

defined with the $N-1$ divisor?

At this point your instructor mumbled something about "degrees of freedom" and moved on, but why mumble if you can compute? By the definition (A20.1), expectations of unbiased estimates are exact,

$$\mathbb{E}[\hat{\mu}] = \mu\,, \qquad \mathbb{E}[\hat{\sigma}^2] = \sigma^2\,. \tag{A20.28}$$

That is true of the empirical mean (A20.14),

$$\mathbb{E}[\hat{\mu}] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[a_i] = \frac{1}{N}\sum_{i=1}^{N}\mu = \mu\,,$$

but the empirical estimate for the sample variance written as average over the sum of deviations square does not quite work out. Assume first that the empirical variance is given by the usual average

$$\begin{aligned}\bar{\sigma}^2 &= \frac{1}{N}\sum_{i=1}^{N}(a_i-\hat{\mu})^2 = \frac{1}{N}\sum_{i=1}^{N}\left(a_i^2-2\hat{\mu}a_i+\hat{\mu}^2\right) = \frac{1}{N}\sum_{i=1}^{N}a_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{N}a_i\right)^2\\ &= \frac{1}{N}\sum_{i=1}^{N}a_i^2 - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_ia_j = \frac{N-1}{N^2}\sum_{i=1}^{N}a_i^2 - \frac{1}{N^2}\sum_{i\neq j}^{N}a_ia_j\,.\end{aligned} \tag{A20.29}$$

By the iid independence of individual measurements (A20.13), and $\sigma^2 = \mathbb{E}[a^2]-\mu^2$ relation (A20.4), the expectation of $\bar{\sigma}^2$ is

exercise A20.2

$$\begin{aligned}\mathbb{E}\left[\bar{\sigma}^2\right] &= \frac{N-1}{N^2}\sum_{i=1}^{N}\mathbb{E}\left[a_i^2\right] - \frac{1}{N^2}\sum_{i\neq j}^{N}\mathbb{E}\left[a_ia_j\right] = \frac{N-1}{N}\left(\mathbb{E}\left[a_i^2\right]-\mathbb{E}\left[a_i\right]\mathbb{E}\left[a_j\right]\right)\\ &= \frac{N-1}{N}\left(\mathbb{E}\left[a^2\right]-\mu^2\right) = \frac{N-1}{N}\,\sigma^2\,.\end{aligned} \tag{A20.30}$$

This attempt at a definition of empirical variance $\bar{\sigma}^2$ thus violates the 'unbiased' condition (A20.28). The *unbiased empirical variance* (A20.15), $\hat{\sigma}^2 = N\bar{\sigma}^2/(N-1)$, is correct for any sample size, not only in the $n\to\infty$ limit. What happened? a_i, a_j are idd only for the N^2-N off-diagonal covariance elements, the squares a_j^2 along the diagonal do not contribute to "covariance."

click to return: p. 949

(continued in example A20.4) Sara A. Solla

Example A20.4. Standard error of the mean. (Continued from example A20.3)

Think now of estimating the empirical mean (A20.14) of observable a as $j = 1, 2, \cdots, N$ attempts to estimate the mean $\hat{\mu}_j$, each based on M data samples

$$\hat{\mu}_j = \frac{1}{M}\sum_{i=1}^{M}a_i\,. \tag{A20.31}$$

Every attempt yields a different sample mean, so $\hat{\mu}_j$ itself is an idd random variable, with unbiased expectation $\mathbb{E}[\hat{\mu}] = \mu$. What is its variance

$$\mathrm{Var}[\hat{\mu}] = \mathbb{E}[(\hat{\mu} - \mu)^2] = \mathbb{E}[\hat{\mu}^2] - \mu^2 \,?$$

This calculation is very much the same as the one carried out in example A20.3, resulting in

exercise A20.3

$$\mathrm{Var}[\hat{\mu}] = \frac{1}{N}\sigma^2$$

The quantity $\sqrt{\mathrm{Var}[\hat{\mu}]} = \sigma/\sqrt{N}$ is called the *standard error of the mean* (SEM); it tells us that the accuracy of the determination of the mean μ increases as the $1/\sqrt{N}$, where N is the number of estimate attempts, each based on the same number of data points.

click to return: p. 949

Sara A. Solla

Exercises

A20.1. **Cumulants.** Show that for a Gaussian probability distribution (a) all odd moments vanish, and (b) all cumulants in (A20.7) vanish for $n \geq 3$, $\langle a^n \rangle_c = 0$.

(P. Cvitanović)

A20.2. **Unbiased sample variance.** Empirical estimates of the mean $\hat{\mu}$ and the variance $\hat{\sigma}^2$ are said to be "unbiased" if their expectations equal the exact values,

$$\mathbb{E}[\hat{\mu}] = \mu\,, \qquad \mathbb{E}[\hat{\sigma}^2] = \sigma^2\,. \tag{A20.32}$$

(a) Verify that the empirical mean

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} a_i \tag{A20.33}$$

is unbiased.

(b) Show that the naive empirical estimate for the *sample variance*

$$\bar{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \hat{\mu})^2 = \frac{1}{N}\sum_{i=1}^{N} a_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{N} a_i\right)^2$$

is biased. Hint: note that in evaluating $\mathbb{E}[\cdots]$ you have to separate out the diagonal terms in

$$\left(\sum_{i=1}^{N} a_i\right)^2 = \sum_{i=1}^{N} a_i^2 + \sum_{i\neq j}^{N} a_i a_j\,. \tag{A20.34}$$

(c) Show that the empirical estimate of form

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \hat{\mu})^2\,, \tag{A20.35}$$

is unbiased.

(d) Is this empirical sample variance unbiased for any finite sample size, or is it unbiased only in the $n \to \infty$ limit?

Sara A. Solla

A20.3. **Standard error of the mean.** Now, estimate the empirical mean (A20.33) of observable a by $j = 1, 2, \cdots, N$ attempts to estimate the mean $\hat{\mu}_j$, each based on M data samples

$$\hat{\mu}_j = \frac{1}{M}\sum_{i=1}^{M} a_i\,. \tag{A20.36}$$

Every attempt yields a different sample mean.

(a) Argue that $\hat{\mu}_j$ itself is an idd random variable, with unbiased expectation $\mathbb{E}[\hat{\mu}] = \mu$.

(b) What is its variance

$$\mathrm{Var}[\hat{\mu}] = \mathbb{E}[(\hat{\mu} - \mu)^2] = \mathbb{E}[\hat{\mu}^2] - \mu^2$$

as a function of variance expectation (A20.32) and N, the number of $\hat{\mu}_j$ estimates? Hint; one way to do this is to repeat the calculations of exercise A20.2, this time for $\hat{\mu}_j$ rather than a_i.

(c) The quantity $\sqrt{\mathrm{Var}[\hat{\mu}]} = \sigma/\sqrt{N}$ is called the *standard error of the mean* (SEM); it tells us that the accuracy of the determination of the mean μ. How does SEM decrease as the N, the number of estimate attempts, increases?

Sara A. Solla