Solutions to the Exercises* on Independent Component Analysis

Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany, EU

4 February 2017

Contents

© 2016 Laurenz Wiskott (homepage https://www.ini.rub.de/PEOPLE/wiskott/). This work (except for all figures from other sources, if present) is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/. Figures from other sources have their own copyright, which is generally indicated. Do not distribute parts of these lecture notes showing figures with non-free copyrights (here usually figures I have the rights to publish but you don't, like my own published figures).

Several of my exercises (not necessarily on this topic) were inspired by papers and textbooks by other authors. Unfortunately, I did not document that well, because initially I did not intend to make the exercises publicly available, and now I cannot trace it back anymore. So I cannot give as much credit as I would like to. The concrete versions of the exercises are certainly my own work, though.

In cases where I reuse an exercise in different variants, references may be wrong for technical reasons.

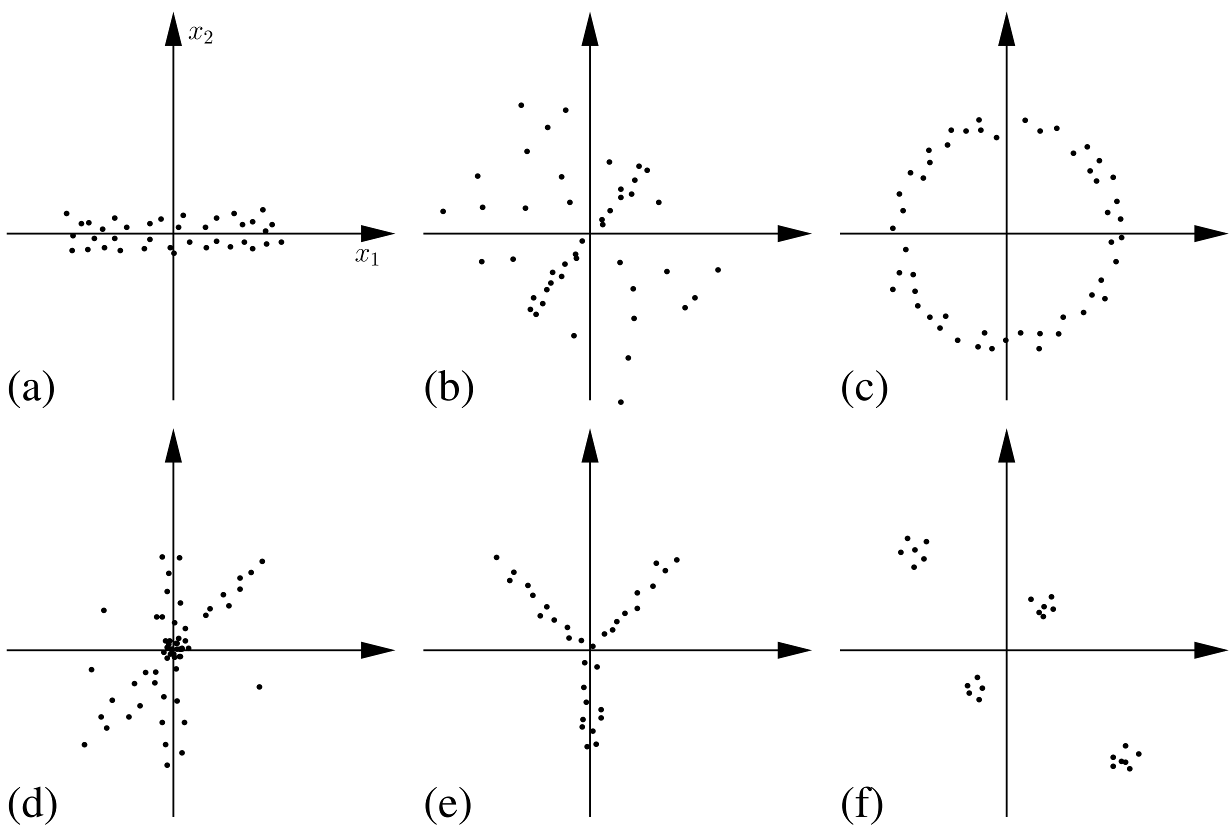
*These exercises complement my corresponding lecture notes available at https://www.ini.rub.de/PEOPLE/wiskott/Teaching/Material/, where you can also find other teaching material such as programming exercises. The table of contents of the lecture notes is reproduced here to give an orientation when the exercises can be reasonably solved. For best learning effect I recommend to first seriously try to solve the exercises yourself before looking into the solutions.

1 Intuition

1.1 Mixing and unmixing

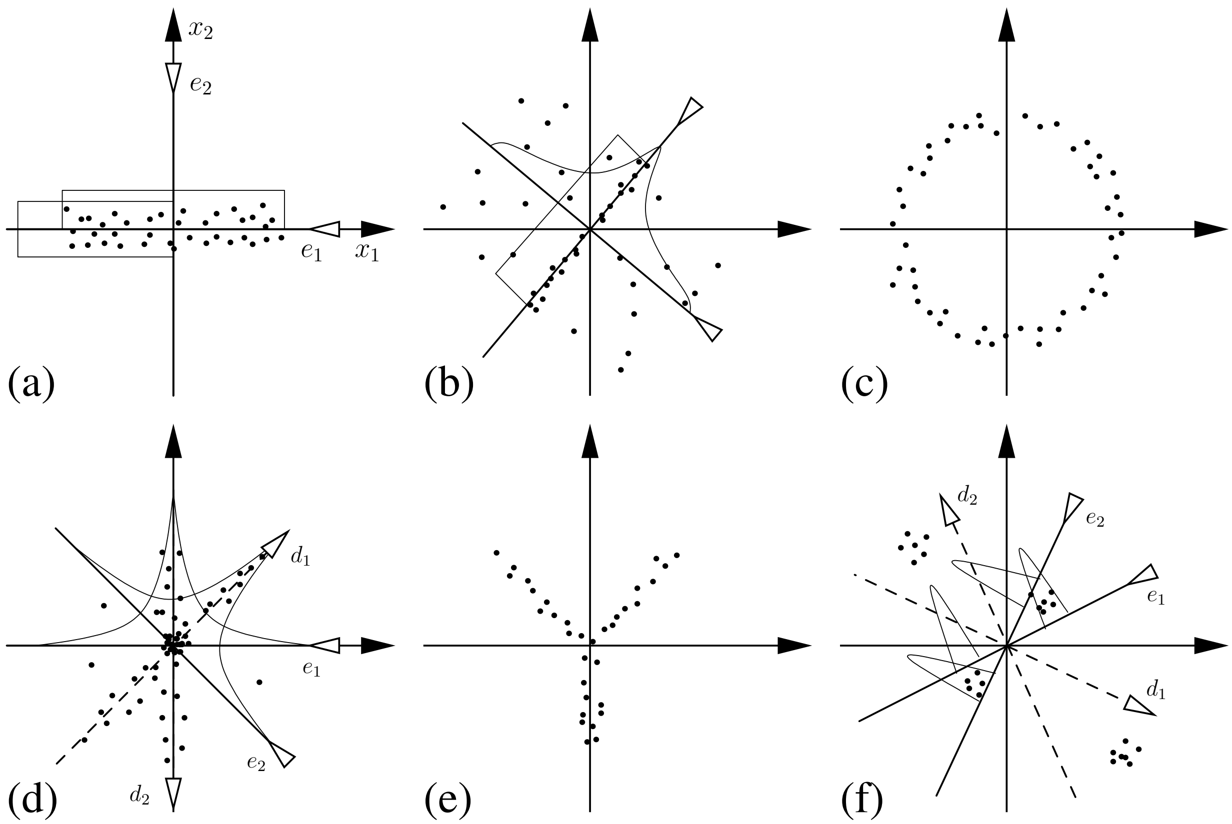
1.1.1 Exercise: Guess independent components and distributions from data

Decide whether the following distributions can be linearly separated into independent components. If yes, sketch the (not necessarily orthogonal) axes onto which the data must be projected to extract the independent components. Draw on these axes also the marginal distributions of the corresponding components.



© CC BY-SA 4.0

Solution:



© CC BY-SA 4.0

Vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ used for extracting the independent components are drawn with inverted arrow heads. Vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ used for mixing the sources are drawn with dashed lines and proper arrow heads, but only if different from $\mathbf{e}_1$ and $\mathbf{e}_2$. All vectors are drawn with same length, which is not correct in some cases, because the variances of the extracted components might be wrong. If unmixing is not possible, no additional vectors are drawn. The distributions of the extracted sources are drawn on the $\mathbf{e}_1$- and $\mathbf{e}_2$-axes. Notice that they should be normalized to 1, i.e. they should have the same area.

Extra question: Elaborate on (f) (do not simplify this to a parallelogram). How exactly do the distribution and extraction vectors come about and relate to each other? How do you write the mixing and unmixing process in matrix notation?

1.2 How to find the unmixing matrix?

1.3 Sources can only be recovered up to permutation and rescaling

1.4 Whiten the data first

1.5 A generic ICA algorithm

2 Formalism based on cumulants

2.1 Moments and cumulants

2.1.1 Exercise: Moments and cumulants

Mixed cumulants κ can be written as sums of products of mixed moments and *vice versa*. More intuitive is the latter. Since cumulants represent a distribution only in exactly one order, since the lower order moments have been subtracted off, it is easy to imagine that a moment can be written by summing over all possible combinations of cumulants that add up exactly to the order of the moment, for instance

$$\begin{aligned} \langle X_1 X_2 X_3 \rangle &= \kappa(X_1, X_2, X_3) \\ &\quad + \kappa(X_1, X_2)\kappa(X_3) + \kappa(X_1, X_3)\kappa(X_2) + \kappa(X_2, X_3)\kappa(X_1) \\ &\quad + \kappa(X_1)\kappa(X_2)\kappa(X_3) \,. \end{aligned} \tag{1}$$

The general rule is

$$\langle X_1 \cdots X_N \rangle = \sum_{\pi} \prod_{B \in \pi} \kappa(X_i : i \in B) \,, \tag{2}$$

with π going through the list of all possible partitions of the N variables into disjoint blocks and B indicating the blocks within one partition.

Hint: You cannot assume zero-mean data here.

Hint: In the following it is convenient to write M_{123} and C_{123} etc. instead of $\langle X_1 X_2 X_3 \rangle$ and $\kappa(X_1, X_2, X_3)$ etc.

1. Write with the help of equation (2) all mixed moments up to order four as a sum of products of cumulants, like in equation (1).

   Solution: With equation (2) we get

$$M_1 \stackrel{(2)}{=} C_1 \,, \tag{3}$$

$$\begin{aligned} M_{12} &\stackrel{(2)}{=} C_{12} \\ &\quad + C_1 C_2 \,, \end{aligned} \tag{4}$$

$$\begin{aligned} M_{123} &\stackrel{(2)}{=} C_{123} \\ &\quad + C_{12}C_3 + C_{13}C_2 + C_{23}C_1 \\ &\quad + C_1 C_2 C_3 \,, \end{aligned} \tag{5}$$

$$\begin{aligned} M_{1234} &\stackrel{(2)}{=} C_{1234} \\ &\quad + C_{123}C_4 + C_{124}C_3 + C_{134}C_2 + C_{234}C_1 \\ &\quad + C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23} \\ &\quad + C_{12}C_3C_4 + C_{13}C_2C_4 + C_{14}C_2C_3 + C_{23}C_1C_4 + C_{24}C_1C_3 + C_{34}C_1C_2 \\ &\quad + C_1C_2C_3C_4 \,. \end{aligned} \tag{6}$$

By the way, it is interesting to reflect why it is reasonable that the terms in the sums above contain each random variable exactly once. Why aren't there terms where a variable is missing or a variable appears twice?

2. From the equations found in part 1 derive expressions for the cumulants up to order three written as sums of products of moments.

   Solution: Solving and inserting the equations above yields

$$
\begin{aligned}
C_1 &\overset{(3)}{=} M_1\,, && (7)\\
C_{12} &\overset{(4)}{=} M_{12} \\
&\quad - C_1C_2 && (8)\\
&\overset{(7)}{=} M_{12} \\
&\quad - M_1M_2\,, && (9)\\
C_{123} &\overset{(5)}{=} M_{123} \\
&\quad - C_{12}C_3 - C_{13}C_2 - C_{23}C_1 \\
&\quad - C_1C_2C_3 && (10)\\
&\overset{(7,9)}{=} M_{123} \\
&\quad - (M_{12} - M_1M_2)M_3 - (M_{13} - M_1M_3)M_2 - (M_{23} - M_2M_3)M_1 \\
&\quad - M_1M_2M_3 && (11)\\
&= M_{123} \\
&\quad - M_{12}M_3 - M_{13}M_2 - M_{23}M_1 \\
&\quad + 2M_1M_2M_3\,, && (12)\\
& && (13)
\end{aligned}
$$

and for for the sake of completeness (without any guarantee for correctness)

$$
\begin{aligned}
& C_{1234} \\
\stackrel{(6)}{=} \; & M_{1234} \\
& - C_{123}C_4 - C_{124}C_3 - C_{134}C_2 - C_{234}C_1 \\
& - C_{12}C_{34} - C_{13}C_{24} - C_{14}C_{23} \\
& - C_{12}C_3C_4 - C_{13}C_2C_4 - C_{14}C_2C_3 - C_{23}C_1C_4 - C_{24}C_1C_3 - C_{34}C_1C_2 \\
& - C_1C_2C_3C_4 \qquad (14) \\
\stackrel{(7,9,12)}{=} \; & M_{1234} \\
& - (M_{123} - M_{12}M_3 - M_{13}M_2 - M_{23}M_1 + 2M_1M_2M_3)M_4 \\
& - (M_{124} - M_{12}M_4 - M_{14}M_2 - M_{24}M_1 + 2M_1M_2M_4)M_3 \\
& - (M_{134} - M_{13}M_4 - M_{14}M_3 - M_{34}M_1 + 2M_1M_3M_4)M_2 \\
& - (M_{234} - M_{23}M_4 - M_{24}M_3 - M_{34}M_2 + 2M_2M_3M_4)M_1 \\
& - (M_{12} - M_1M_2)(M_{34} - M_3M_4) \\
& - (M_{13} - M_1M_3)(M_{24} - M_2M_4) \\
& - (M_{14} - M_1M_4)(M_{23} - M_2M_3) \\
& - (M_{12} - M_1M_2)M_3M_4 - (M_{13} - M_1M_3)M_2M_4 - (M_{14} - M_1M_4)M_2M_3 \\
& - (M_{23} - M_2M_3)M_1M_4 - (M_{24} - M_2M_4)M_1M_3 - (M_{34} - M_3M_4)M_1M_2 \\
& - M_1M_2M_3M_4 \qquad (15) \\
= \; & M_{1234} \\
& - M_{123}M_4 + M_{12}M_3M_4 + M_{13}M_2M_4 + M_{23}M_1M_4 - 2M_1M_2M_3M_4 \\
& - M_{124}M_3 + M_{12}M_4M_3 + M_{14}M_2M_3 + M_{24}M_1M_3 - 2M_1M_2M_4M_3 \\
& - M_{134}M_2 + M_{13}M_4M_2 + M_{14}M_3M_2 + M_{34}M_1M_2 - 2M_1M_3M_4M_2 \\
& - M_{234}M_1 + M_{23}M_4M_1 + M_{24}M_3M_1 + M_{34}M_2M_1 - 2M_2M_3M_4M_1 \\
& - M_{12}M_{34} + M_{12}M_3M_4 + M_1M_2M_{34} - M_1M_2M_3M_4 \\
& - M_{13}M_{24} + M_{13}M_2M_4 + M_1M_3M_{24} - M_1M_3M_2M_4 \\
& - M_{14}M_{23} + M_{14}M_2M_3 + M_1M_4M_{23} - M_1M_4M_2M_3 \\
& - M_{12}M_3M_4 + M_1M_2M_3M_4 - M_{13}M_2M_4 + M_1M_3M_2M_4 - M_{14}M_2M_3 + M_1M_4M_2M_3 \\
& - M_{23}M_1M_4 + M_2M_3M_1M_4 - M_{24}M_1M_3 + M_2M_4M_1M_3 - M_{34}M_1M_2 + M_3M_4M_1M_2 \\
& - M_1M_2M_3M_4 \qquad (16) \\
= \; & M_{1234} \\
& - M_{123}M_4 - M_{124}M_3 - M_{134}M_2 - M_{234}M_1 \\
& - M_{12}M_{34} - M_{13}M_{24} - M_{14}M_{23} \\
& + 2M_{12}M_3M_4 + 2M_{13}M_2M_4 + 2M_{14}M_2M_3 + 2M_{23}M_1M_4 + 2M_{24}M_1M_3 + 2M_{34}M_1M_2 \\
& - 6M_1M_2M_3M_4 \, . \qquad (17)
\end{aligned}
$$

If all variables are identical, i.e. $X_1 = X_2 = X_3 = X_4 = X$, then the first four mixed cumulants become the mean, variance, skewness, and kurtosis.

Extra question: Can one get an intuition for what higher order mixed moments and cumulants mean?

2.1.2 Exercise: Moments and cumulants of concrete distributions

1. What can you say in general about moments and cumulants of symmetric distributions (even functions) of a single variable?

   Solution: For symmetric distributions of a single variable all moments of odd order vanish, since positive and negative contributions cancel out each other for symmetry reasons. In particular such distributions have zero mean. The same is true for cumulants of odd order, which, as we know, can be written as sums of products of moments. Each of these products must have the same order as the cumulant, i.e. must be of odd overall order. This implies that at least one moment in each product must be odd, so that each product vanishes, because moments of odd order vanish. (One could not argue like that for even cumulants, since it is easy to generate an even order product with odd moments only. For instance, the product of two moments of order 3 (odd) is of order 6 (even).) Moments of even order are positive, because of the even exponent. That does not necessarily hold for even cumulants.

2. Calculate all moments up to order ten and all cumulants up to order four for the following distributions.

   Hint: First derive a closed form or a recursive formula for $\langle x^n \rangle$ and then insert the different values for n.

   Solution: As a reminder here are the definitions of the first four cumulants again.

$$
\begin{aligned}
\text{Mean:} \qquad \operatorname{mean}(x) &:= \langle x \rangle && (1)\\
&= 0 \quad \text{(if the data have zero mean)}\,, && (2)\\
\text{Variance:} \qquad \operatorname{var}(x) &:= \langle x^2 \rangle - \langle x \rangle^2 && (3)\\
&:= \langle x^2 \rangle \quad \text{(if the data have zero mean)}\,, && (4)\\
\text{Skewness:} \qquad \operatorname{skew}(x) &:= \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3\,, && (5)\\
&:= \langle x^3 \rangle \quad \text{(if the data have zero mean)}\,, && (6)\\
\text{Kurtosis:} \qquad \operatorname{kurt}(x) &:= \langle x^4 \rangle - 3\langle x^2 \rangle^2 \quad \text{(if the data have zero mean)}\,. && (7)
\end{aligned}
$$

   Now to the exercises.

(a) Uniform distribution (D: Gleichverteilung):

$$
p(x) := \begin{cases} 1/2 & \text{if } x \in [-1, +1] \\ 0 & \text{otherwise} \end{cases} \quad . \tag{8}
$$

Solution: First we calculate the moments of the distribution and then the cumulants. Since the distribution is symmetric, all odd moments and cumulants vanish. In particular the distribution has zero mean. Thus, we confine our considerations to even n.

$$
\begin{aligned}
\langle x^n \rangle &= \int_{-\infty}^{+\infty} x^n p(x)\,\mathrm{d}x && (9)\\
&\overset{(8)}{=} \int_{-1}^{+1} x^n \cdot 1/2\,\mathrm{d}x && (10)\\
&= \left[x^{n+1}/(n+1)\right]_{-1}^{+1} \Big/ 2 && (11)\\
&= \left((+1)^{n+1} - (-1)^{n+1}\right) \Big/ (2\,(n+1)) && (12)\\
&= 2 \Big/ (2\,(n+1)) \quad \text{(since } n \text{ is even)} && (13)\\
&= 1/(n+1) && (14)\\
\Longrightarrow \quad \langle x^2 \rangle &\overset{(14)}{=} 1/3 \approx 0.33 && (15)\\
\wedge \quad \langle x^4 \rangle &\overset{(14)}{=} 1/5 = 0.5 && (16)\\
\wedge \quad \langle x^6 \rangle &\overset{(14)}{=} 1/7 \approx 0.14 && (17)\\
\wedge \quad \langle x^8 \rangle &\overset{(14)}{=} 1/9 \approx 0.11 && (18)\\
\wedge \quad \langle x^{10} \rangle &\overset{(14)}{=} 1/11 \approx 0.09\,, && (19)\\
\operatorname{var}(x) &\overset{(3)}{=} \langle x^2 \rangle - \langle x \rangle^2 \overset{(15)}{=} 1/3 \approx 0.33 \quad (\text{since } \langle x \rangle = 0)\,, && (20)\\
\operatorname{kurt}(x) &\overset{(7)}{=} \langle x^4 \rangle - 3\langle x^2 \rangle^2 \overset{(16,15)}{=} 1/5 - 3 \cdot (1/3)^2 && (21)\\
&= 3/15 - 5/15 = -2/15 \approx -0.13\,. && (22)
\end{aligned}
$$

As expected the kurtosis is negative, since the distribution is less peaky (D: spitz(?)) than a Gaussian distribution.

(b) Laplace- or double exponential distribution (D: doppelt exponentielle Verteilung):

$$p(x) := \frac{\exp(-|x|)}{2} \,. \tag{23}$$

Solution: First we calculate the moments of the distribution and then the cumulants. Since the distribution is symmetric, all odd moments and cumulants vanish. In particular the distribution has zero mean. Thus, we confine our considerations to even n.

$$\begin{aligned}
\langle x^n \rangle &= \int_{-\infty}^{+\infty} x^n p(x)\,\mathrm{d}x && (24)\\
&\overset{(23)}{=} \int_{-\infty}^{+\infty} x^n \exp(-|x|)/2\,\mathrm{d}x && (25)\\
&= \int_{-\infty}^{0} x^n \exp(x)\,\mathrm{d}x \quad \text{(since } n \text{ is even)} && (26)\\
&= \underbrace{[x^n \exp(x)]_{-\infty}^{0}}_{=0} - n\int_{-\infty}^{0} x^{n-1}\exp(x)\,\mathrm{d}x && (27)\\
&\qquad \text{(integration by parts)} &&\\
&= -n\left(\underbrace{[x^{n-1}\exp(x)]_{-\infty}^{0}}_{=0} - (n-1)\int_{-\infty}^{0} x^{n-2}\exp(x)\,\mathrm{d}x\right) &&\\
&\qquad \text{(integration by parts)} && (28)\\
&= n(n-1)\int_{-\infty}^{0} x^{n-2}\exp(x)\,\mathrm{d}x && (29)\\
&\overset{(26)}{=} n(n-1)\cdot\langle x^{n-2}\rangle \quad \text{(since } (n-2) \text{ is even)} && (30)\\
&\overset{(26)}{=} n! \quad \text{(as one can show with mathematical induction)} && (31)\\
\Longrightarrow \quad \langle x^0 \rangle &= 1 \quad \text{(since the distribution must be normalized)} && (32)\\
\wedge \quad \langle x^2 \rangle &\overset{(31)}{=} 2! = 2 && (33)\\
\wedge \quad \langle x^4 \rangle &\overset{(31)}{=} 4! = 24 && (34)\\
\wedge \quad \langle x^6 \rangle &\overset{(31)}{=} 6! = 720 && (35)\\
\wedge \quad \langle x^8 \rangle &\overset{(31)}{=} 8! = 40{,}320 && (36)\\
\wedge \quad \langle x^{10} \rangle &\overset{(31)}{=} 10! = 3{,}628{,}800\,, && (37)\\
\mathrm{var}(x) &\overset{(3)}{=} \langle x^2\rangle - \langle x\rangle^2 = 2 \quad \text{(since } \langle x\rangle = 0)\,, && (38)\\
\mathrm{kurt}(x) &\overset{(7)}{=} \langle x^4\rangle - 3\langle x^2\rangle^2 \overset{(34,33)}{=} 24 - 3\cdot 2^2 = 12\,. && (39)
\end{aligned}$$

As expected the kurtosis is positive, since the distribution is more peaky (D: spitz(?)) than a Gaussian distribution.

2.1.3 Exercise: Moments and cumulants of scaled distributions

Assume all moments and cumulants of a scalar random variable X are given. How do moments and cumulants change if the variable is scaled by a factor s around the origin, i.e. simply multiplied by s ? Prove your result.

Hint: Cumulants can always be written as a sum of products of moments of identical overall order.

Solution: It is easy to see that the moments of order n simply scale with s^n if one goes from the original data X to the scaled data sX,

$$\langle (sX)^n \rangle = s^n \langle X^n \rangle \quad \text{(since } s \text{ is a constant)}\,. \tag{1}$$

The same is true for the cumulants, because all terms in a cumulant are of same overall order, i.e. contain X and therefore s the same number of times, for instance

$$\begin{aligned}
\mathrm{skew}(sX) &= \langle (sX)^3 \rangle - 3\langle (sX)^2 \rangle \langle sX \rangle + 2\langle (sX) \rangle^3 && (2)\\
&= s^3 \langle X^3 \rangle - 3s^2 \langle X^2 \rangle s \langle X \rangle + 2(s\langle X \rangle)^3 && (3)\\
&= s^3 (\langle X^3 \rangle - 3\langle X^2 \rangle \langle X \rangle + 2\langle X \rangle^3) && (4)\\
&= s^3 \, \mathrm{skew}(X)\,. && (5)
\end{aligned}$$

2.1.4 Exercise: Moments and cumulants of shifted distributions

Assume all moments and cumulants of a (non zero-mean) scalar random variable X are known. How do the first four moments and the first three cumulants change if the variable is shifted by m?

Hint: See exercise 2.1.1 for the cumulants of non-zero-mean data.

Solution: For the moments we find

$$\begin{aligned}
\langle m + X \rangle &= m + \langle X \rangle\,, && (1)\\
\langle (m+X)^2 \rangle &= \langle m^2 + 2mX + X^2 \rangle && (2)\\
&= m^2 + 2m\langle X \rangle + \langle X^2 \rangle\,, && (3)\\
\langle (m+X)^3 \rangle &= \langle m^3 + 3m^2 X + 3mX^2 + X^3 \rangle && (4)\\
&= m^3 + 3m^2 \langle X \rangle + 3m\langle X^2 \rangle + \langle X^3 \rangle\,, && (5)\\
\langle (m+X)^4 \rangle &= \langle m^4 + 4m^3 X + 6m^2 X^2 + 4mX^3 + X^4 \rangle && (6)\\
&= m^4 + 4m^3 \langle X \rangle + 6m^2 \langle X^2 \rangle + 4m\langle X^3 \rangle + \langle X^4 \rangle\,. && (7)
\end{aligned}$$

Here is a pattern emerging. With a bit of thought one can see that the weighting factors for the monomials $m^\mu \langle X^{4-\mu} \rangle$ grow like the numbers in Pascal's triangle.

For the cumulants the situation is conceptionally simpler but computationally more complex. Intuitively one would expect that the mean is the only cumulant that changes, because higher-order cumulants are blind to what is represented by lower-order cumulants already. Thus they capture the shape of the distribution and should be blind to the mean. The formal verification is a bit tedious, but for order two we find

$$\begin{aligned}
\mathrm{var}(m+X) &= \langle (m+X)^2 \rangle - \langle (m+X) \rangle^2 && (8)\\
&= \langle m^2 + 2mX + X^2 \rangle - \langle m + X \rangle^2 && (9)\\
&= \langle m^2 \rangle + \langle 2mX \rangle + \langle X^2 \rangle - (\langle m \rangle + \langle X \rangle)^2 && (10)\\
&= m^2 + 2m\langle X \rangle + \langle X^2 \rangle - m^2 - 2m\langle X \rangle - \langle X \rangle^2 && (11)\\
&= \langle X^2 \rangle - \langle X \rangle^2 && (12)\\
&= \mathrm{var}(X)\,, && (13)
\end{aligned}$$

and for three

$$
\begin{aligned}
\text{skew}(m+X) &= \langle (m+X)^3 \rangle - 3\langle (m+X)^2 \rangle \langle (m+X) \rangle + 2\langle (m+X) \rangle^3 && (14)\\
&= \langle m^3 + 3m^2X + 3mX^2 + X^3 \rangle - 3\langle m^2 + 2mX + X^2 \rangle \langle m+X \rangle + 2\langle m+X \rangle^3 && (15)\\
&= m^3 + 3m^2\langle X \rangle + 3m\langle X^2 \rangle + \langle X^3 \rangle \\
&\quad - 3(m^2 + 2m\langle X \rangle + \langle X^2 \rangle)(m + \langle X \rangle) + 2(m + \langle X \rangle)^3 && (16)\\
&= m^3 + 3m^2\langle X \rangle + 3m\langle X^2 \rangle + \langle X^3 \rangle \\
&\quad - 3m^2m - 3\cdot 2m\langle X \rangle m - 3\langle X^2 \rangle m - 3m^2\langle X \rangle - 3\cdot 2m\langle X \rangle\langle X \rangle - 3\langle X^2 \rangle\langle X \rangle \\
&\quad + 2m^3 + 2\cdot 3m^2\langle X \rangle + 2\cdot 3m\langle X \rangle^2 + 2\langle X \rangle^3 && (17)\\
&= (m^3 - 3m^2m + 2m^3) \\
&\quad + (3m^2 - 3\cdot 2mm - 3m^2 + 2\cdot 3m^2)\langle X \rangle \\
&\quad + (3m - 3m)\langle X^2 \rangle + (-3\cdot 2m + 2\cdot 3m)\langle X \rangle^2 \\
&\quad + \langle X^3 \rangle + (-3)\langle X^2 \rangle\langle X \rangle + 2\langle X \rangle^3 && (18)\\
&= \langle X^3 \rangle - 3\langle X^2 \rangle\langle X \rangle + 2\langle X \rangle^3 && (19)\\
&= \text{skew}(X)\,. && (20)
\end{aligned}
$$

Scaling, however, changes all cumulants (unless they have been normalized) as well as the moments.

2.1.5 Exercise: Kurtosis is additive

Kurtosis for a zero-mean random variable x is defined as

$$\text{kurt}(x) := \langle x^4 \rangle - 3\langle x^2 \rangle^2\,. \tag{1}$$

Show that for two statistically independent zero-mean random variables x und y

$$\text{kurt}(x+y) = \text{kurt}(x) + \text{kurt}(y) \tag{2}$$

holds. Make clear where you use which argument for simplifications.

Solution: We can show directly that

$$
\begin{aligned}
\text{kurt}(x+y) &\overset{(1)}{=} \langle (x+y)^4 \rangle - 3\langle (x+y)^2 \rangle^2 && (3)\\
&= \langle x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \rangle - 3\langle x^2 + 2xy + y^2 \rangle^2 && (4)\\
&= \langle x^4 \rangle + 4\langle x^3 \rangle\langle y \rangle + 6\langle x^2 \rangle\langle y^2 \rangle + 4\langle x \rangle\langle y^3 \rangle + \langle y^4 \rangle \\
&\quad - 3\left(\langle x^2 \rangle + 2\langle x \rangle\langle y \rangle + \langle y^2 \rangle\right)^2 && (5)\\
&\quad\ \text{(since } x \text{ and } y \text{ are statistically independent)} \\
&= \langle x^4 \rangle + 6\langle x^2 \rangle\langle y^2 \rangle + \langle y^4 \rangle - 3\left(\langle x^2 \rangle + \langle y^2 \rangle\right)^2 && (6)\\
&\quad\ \text{(since } \langle x \rangle = \langle y \rangle = 0\text{)} \\
&= \langle x^4 \rangle + 6\langle x^2 \rangle\langle y^2 \rangle + \langle y^4 \rangle - 3\langle x^2 \rangle^2 - 3\cdot 2\langle x^2 \rangle\langle y^2 \rangle - 3\langle y^2 \rangle^2 && (7)\\
&= \langle x^4 \rangle - 3\langle x^2 \rangle^2 + \langle y^4 \rangle - 3\langle y^2 \rangle^2 && (8)\\
&\overset{(1)}{=} \text{kurt}(x) + \text{kurt}(y)\,. && (9)
\end{aligned}
$$

This additivity for statistically independent random variables holds true for any cumulant.

2.1.6 Exercise: Moments and cumulants are multilinear

1. Show that cross-moments, such as $\langle x_1 x_2 x_3 \rangle$, are multilinear, i.e. linear in each of their arguments, e.g.

$$\langle (ax_1 + bx_1')x_2x_3 \ldots \rangle \quad = \quad a\langle x_1x_2x_3 \ldots \rangle + b\langle x_1'x_2x_3 \ldots \rangle \tag{1}$$

with a and b being constants and x_1' being another random variable.

Solution: This follows directly from the linearity of the averaging process. Without loss of generality we consider the first out of an arbitrary number of variables and find

$$\begin{aligned} \langle (ax_1 + bx_1')x_2x_3...\rangle &= \langle (ax_1)x_2x_3...\rangle + \langle (bx_1')x_2x_3...\rangle && (2)\\ &= a\langle x_1x_2x_3...\rangle + b\langle x_1'x_2x_3...\rangle\,, && (3)\end{aligned}$$

which proves the assertion.

2. Show that cross-cumulants, such as $\kappa(x_1, x_2, x_3)$, are multilinear, i.e. linear in each of their arguments.

   Solution: A cross-cumulant can always be written as a sum of products of moments, each term containing the same variables as the cumulant, for instance

$$\kappa(x_1, x_2, x_3) = \langle x_1x_2x_3\rangle - \langle x_1x_2\rangle\langle x_3\rangle - \langle x_1x_3\rangle\langle x_2\rangle - \langle x_2x_3\rangle\langle x_1\rangle + 2\langle x_1\rangle\langle x_2\rangle\langle x_3\rangle\,. \quad (4)$$

   Each term in this sum is linear in each of its arguments, because the moment it appears in is linear and the other moments are simple constant factors. Since a sum of linear functions is again a linear function, cumulants are linear in each of its elements.

3. As we have seen in exercise 2.1.5, the kurtosis is additive for two statistically independent zero-mean random variables x und y, i.e.

$$\mathrm{kurt}(x + y) = \mathrm{kurt}(x) + \mathrm{kurt}(y)\,. \quad (5)$$

   Why is statistical independence required for the additivity of kurtosis of the signals while it is not for the multilinearity of cross-cumulants.

   Solution: The additivity of kurtosis is not a linearity in the proper sense. For instance, $\mathrm{kurt}(ax) = a\mathrm{kurt}(x)$ with a constant a does not hold but would be required for linearity. Zero mean and statistical independence are required to eliminate all the mixed terms that would otherwise arise.

2.1.7 Exercise: Mixing statistically independent sources

Given some scalar and statistically independent random variables (signals) s_i with zero mean, unit variance, and a value a_i for the kurtosis that lies between $-a$ and $+a$, with arbitrary but fixed value of $0 < a$. The s_i shall be mixed like

$$x := \sum_i w_i s_i \quad (1)$$

with constant weights w_i.

1. Which constraints do you have to impose on the weights w_i to guarantee that the mixture has unit variance as well?

Solution: The variance of the mixture is

$$\begin{aligned}
\operatorname{var}(x) &= \left\langle (x - \langle x \rangle)^2 \right\rangle && (2)\\
&= \left\langle x^2 \right\rangle - \langle x \rangle^2 && (3)\\
&\overset{(1)}{=} \left\langle \left(\sum_i w_i s_i \right)^2 \right\rangle - \left\langle \sum_i w_i s_i \right\rangle^2 && (4)\\
&= \left\langle \left(\sum_i w_i s_i \right)^2 \right\rangle - \left(\sum_i w_i \langle s_i \rangle \right)^2 && (5)\\
&= \left\langle \left(\sum_i w_i s_i \right) \left(\sum_j w_j s_j \right) \right\rangle - \left(\sum_i w_i \langle s_i \rangle \right) \left(\sum_j w_j \langle s_j \rangle \right) && (6)\\
&= \left\langle \sum_{i,j} w_i w_j s_i s_j \right\rangle - \sum_{i,j} w_i w_j \langle s_i \rangle \langle s_j \rangle && (7)\\
&= \sum_{i,j} w_i w_j \langle s_i s_j \rangle - \sum_{i,j} w_i w_j \langle s_i \rangle \langle s_j \rangle && (8)\\
&= \sum_i w_i w_i (\langle s_i s_i \rangle - \langle s_i \rangle \langle s_i \rangle) + \sum_{i,j:i \neq j} w_i w_j (\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle) && (9)\\
&= \sum_i w_i^2 \underbrace{(\langle s_i s_i \rangle - \langle s_i \rangle^2)}_{=\operatorname{var}(s_i)=1} + \sum_{i,j:i \neq j} w_i w_j \underbrace{(\langle s_i \rangle \langle s_j \rangle - \langle s_i \rangle \langle s_j \rangle)}_{=0} && (10)\\
&\quad \text{(since } s_i \text{ and } s_j \text{ are statistically independent for } i \neq j\text{)}\\
&= \sum_i w_i^2 \,. && (11)
\end{aligned}$$

Thus, the unit variance contraint for the mixture translates into the constraint

$$\sum_i w_i^2 = 1 \qquad (12)$$

for the weights.

2. Prove that the kurtosis of an equally weighted mixture ($w_i = w_j \forall i, j$) of N signals converges to zero as N goes to infinity.

   Hints: (i) For the kurtosis and two statistically independent random variables s_1 and s_2

   $$\operatorname{kurt}(s_1 + s_2) = \operatorname{kurt}(s_1) + \operatorname{kurt}(s_2) \qquad (13)$$

   holds. (ii) Use the constraint from part 1.

   Solution: Due to the equal weights and the constraint (12) we get

   $$x := \frac{1}{\sqrt{N}} \sum_i s_i \,. \qquad (14)$$

This implies the following estimate for the kurtosis

$$
\begin{aligned}
|\text{kurt}(x)| &\overset{(14)}{=} \left| \text{kurt}\left(\frac{1}{\sqrt{N}} \sum_i s_i \right) \right| && (15) \\
&= \left| \left(\frac{1}{\sqrt{N}} \right)^4 \text{kurt}\left(\sum_i s_i \right) \right| && (16) \\
&\qquad \text{(since the kurtosis is of fourth order)} \\
&\overset{(13)}{=} \frac{1}{N^2} \left| \sum_i \text{kurt}(s_i) \right| && (17) \\
&\leq \frac{1}{N^2} \sum_i |\text{kurt}(s_i)| && (18) \\
&\leq \frac{a}{N^2} \underbrace{\sum_i 1}_{=N} \quad \text{(since } \text{kurt}(s_i) \in [-a, +a]\text{)} && (19) \\
&= \frac{a}{N}\,, && (20) \\
& && (21)
\end{aligned}
$$

which goes to zero as N goes to infinity, so that $|\text{kurt}(x)|$ and therefore also $\text{kurt}(x)$ goes to zero.

This is a weak version of the well known central limit theorem, which says that the distribution of a mixture of infinitely many random variables converges to a Gauss distribution, which has zero kurtosis.

2.2 Cross-cumulants of statistically independent components are zero

2.3 Components with zero cross-cumulants are statistically independent

2.4 Rotated cumulants

2.5 Contrast function

2.6 Givens-rotations

2.7 Optimizing the contrast function

2.8 The algorithm