Python KM: Kramers–Moyal coefficients at any order for one and two-dimensional stochastic processes

Leonardo Rydin Gorjão, Francisco Meirinhos

Summary

A general problem for measuring and evaluating any stochastic process is the retrieval from data or time-series of the Kramers–Moyal coefficients $\mathcal{M}$, the coefficients of the Kramers–Moyal expansion of the master equation describing a particular stochastic process.

Given a set of continuous and stationary data, i.e., ergodic or quasi-stationary stochastic data, the extensive literature of stochastic processes awards a set of measures, as the Kramers–Moyal coefficients (Risken, H. 1996) or the conditional moments, that link stochastic processes to a probabilistic description of the process or of the family of processes, i.e., to a set of partial differential equations, e.g. the master equation or the Fokker–Planck equations. Of particular relevance is the growing evidence that lower-order stochastic processes do not seem to accurately described phenomena seen in data (Tabar, M. R. R., 2019), thus a necessity to study higher-order processes, an therefore uncover the higher-order Kramers–Moyal coefficients efficiently.

To unravel the Kramers–Moyal coefficients the most straightforward approach is to perform a histogram-based regression to evaluate the conditional moments of the system at hand. Such approach is conventional and requires the most commonly used packages in Python (i.e., `numpy`, `scipy`, or `pandas`), and is directly embodied in MATLAB, Julia or R, usually under the name of `hist` or `histogram`.

The package presented here comprises a manifold of options: A general open-source toolbox for the calculation of Kramers–Moyal coefficients in one and two dimension, for any desired order, and with a variety of different kernels, with the added freedom to implement a specific kernel, if desired.

Mathematics

The probability that an n-dimensional state variable $\boldsymbol{x}'(t) \in \mathbb{R}^n$ is observed at position $\boldsymbol{x}$ is given by the conditional probability of the previous (or future) states of the system (Risken, H. 1996). This probabilistic description takes the form

$$\mathcal{M}^{[\ell]}(\boldsymbol{x}, t) = \int dx'(\boldsymbol{x}(t)' - \boldsymbol{x}(t))^\ell W(\boldsymbol{x}'|\boldsymbol{x}), \tag{1}$$

with $W(\boldsymbol{x}'|\boldsymbol{x})$ the transition probability rate.

The exact evaluation of the Kramers–Moyal coefficients for discrete or discretised datasets $\boldsymbol{y}(t)$—any human measure of a process is discrete, as well as any computer generated data—is bounded by the timewise limit imposed. Taking as an example a two-dimentional case, that one can obtained numerically with this library, with $\boldsymbol{x}(t) = (x_1(t), x_2(t)) \in \mathbb{R}^2$, the Kramers–Moyal coefficients $\mathcal{M}^{[\ell,m]} \in \mathbb{R}^2$ take the form

$$\mathcal{M}^{[\ell,m]}(x_1, x_2) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int \mathrm{d}y_1 \mathrm{d}y_2 (y_1(t+\Delta t) - y_1(t))^\ell (y_2(t+\Delta t) - y_2(t))^m \cdot$$
$$P(y_1, y_2; t+\Delta t|y_1, y_2; t)|_{y_1(t)=x_1, y_2(t)=x_2}, \tag{2}$$

at a certain reference measure point (x_1, x_2). The Kramers–Moyal coefficents $\mathcal{M}^{[\ell,m]}(x_1, x_2)$ are obtained from a two-dimensional time-series $(y_1(t), y_2(t))$.

Theoretically Δt should take the limiting case of $\Delta t \to 0$, but the restriction of any measuring or storing device—or the nature of the observables themselves—permits only time-sampled or discrete recordings. The relevance and importance of adequate time-sampling was extensively studied and discussed in (Lehnertz, K., Zabawa, L. and Tabar, M. R. R. 2018). In the limiting case where Δt is equivalent to the sampling rate of the data, the Kramers–Moyal coefficients take the form

$$\mathcal{M}^{[\ell,m]}(x_1, x_2) = \frac{1}{\Delta t} \langle \Delta y_1^\ell \Delta y_2^m |_{y_1(t)=x_1, y_2(t)=x_2} \rangle, \text{ with } \Delta y_i = y_i(t + \Delta t) - y_i(t).$$

Furthermore, the order of the Kramers–Moyal coefficients is given here by the superscript ℓ and m. For such measure of the Kramers–Moyal coefficients, i.e., a probabilistic measure, a probabilistic space exists, assigned to the process, stemming from the master equation describing the family of such processes. The conventional procedure is to utilise a histogram regression of the observed process and retrieve, via approximation or fitting, the Kramers–Moyal coefficient. The choice of a histogram measure for the Kramers–Moyal coefficient results in

an acceptable measure of the probability density functions of the process but requires a new mathematical space (a distribution space). The usage of a kernel approach, followed here, permits an identical overview without the necessity of a new (descritised) distribution space, given that the equivalent space of the observable can be taken.

Alike the histogram approach for the measure of the Kramers–Moyal coefficients, each single measure of the observable $\boldsymbol{y}(t)$ is averaged, with a designed weight, into the distribution space. The standing difference, in comparison to the histogram approach, is the riddance of a (discrete) binning system. All points are averaged, in a weighted fashion, into the distribution space—aiding specially in cases where the data sets is scarce—and awarding a continuous measurable space (easier for fitting, for example).
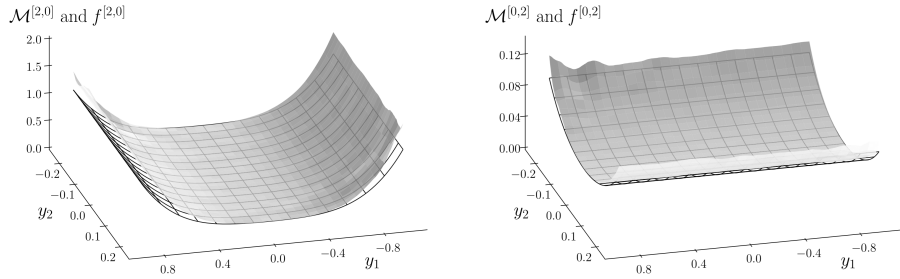

Figure 1: Two exemplary two-dimensional Kramers–Moyal coefficients for a stochastic process were calculated with a package using a $n \times n = 440 \times 440$ array. Plotted as well are the theoretical surfaces according to (Anvari, M., Tabar, M. R. R., Peinke, J. and Lehnertz, K. 2016). An asymmetric two-dimensional Epanechnikov kernel was employed (Epanechnikov 1969).

Library

The presented library is comprised of two separate blocks, `kernels` and `km`, and is a standalone package for a non-parametric retrieval of Kramers–Moyal coefficients, solely dependent on `numpy`. The sub-module `kernels` comprises the kernels for the kernel-based estimation, similarly available in `sklearn`, and `km` performs the desired Kramers–Moyal calculations to any desired power. (Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. 2011).

Acknowledgements

References

Anvari, M., Tabar, M. R. R., Peinke, J. and Lehnertz, K. 2016. "Disentangling the Stochastic Behavior of Complex Time Series." *Scientific Reports* 6: 35435. doi:10.1038/srep35435.

Epanechnikov, V. 1969. "Non-Parametric Estimation of a Multivariate Probability Density." *Theory of Probability & Its Applications* 14 (1): 153–58. doi:10.1137/1114019.

Lehnertz, K., Zabawa, L. and Tabar, M. R. R. 2018. "Characterizing Abrupt Transitions in Stochastic Dynamics." *New Journal of Physics* 20 (11): 113043. doi:10.1088/1367-2630/aaf0d7.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30. http://www.jmlr.org/papers/v12/pedregosa11a.html.

Risken, H. 1996. *The Fokker–Planck Equation.* Springer, Berlin. doi:10.1007/978-3-642-61544-3.

Tabar, M. R. R., 2019. *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems.* Springer International Publishing. doi:10.1007/978-3-030-18472-8.