Skewed superstatistical distributions from a Langevin and Fokker-Planck approach

Erik Van der Straeten and Christian Beck
Queen Mary University of London, School of Mathematical Sciences, Mile End Road, London E1 4NS, UK

The superstatistics concept is a useful statistical method to describe inhomogeneous complex systems for which a system parameter $\beta$ fluctuates on a large spatio-temporal scale. In this paper, we analyze measured time series of wind speed fluctuations and extract the superstatistical distribution function $f(\beta)$ directly from the data. We construct suitable Langevin and Fokker-Planck models with a position dependent $\beta$-field and show that they reduce to standard type of superstatistics in the overdamped limit.

I. INTRODUCTION

The superstatistics concept introduced in 2003 [1] provides a useful tool to describe a large variety of complex systems [2–15]. The basic idea is to characterize the complex system under consideration by a superposition of two statistics, one corresponding to ordinary statistical mechanics (on a mesoscopic level modeled e.g. by a Langevin equation) and the other one corresponding to a slowly varying parameter $\beta(x,t)$ of the system which can be, but need not to be, an inverse temperature. There are many interesting applications of superstatistical techniques. Recent work in this direction includes train delay statistics [16], the distribution of accelerations of test particles in turbulent flows [17], cancer survival statistics [18]. Further applications are described in [19–41].

In this paper, as a working example, we study measured statistics of wind speed fluctuations. For previous work in this direction, see [25, 26]. We will extract the superstatistical distribution function $f(\beta)$ out of our data set and show that in good approximation it is an inverse gamma distribution. As a theoretical model for these and other data sets, we will study Langevin and Fokker-Planck equations with a varying, position-dependent $\beta$. We will show that in the overdamped limit these models reduce to standard type of superstatistics. Skewed stationary distributions arise naturally in these types of models if the overdamped limit is not completely performed.

This paper is organized as follows. In section II we recall the basic concepts of superstatistics, illustrated with the example of a linear Langevin equation. In section III we apply our methods to measured wind speed data, extracting various relevant distributions and parameters out of the data set. In section IV we extend our theoretical approach by studying a Kramers-like equation with spatially varying $\beta$. We show that this model reduces to standard type of superstatistics. Skewed stationary distributions arise naturally in these types of models if the overdamped limit is not completely performed. The last section gives a short discussion of the results of this paper.

II. LINEAR LANGEVIN EQUATION WITH TIME-DEPENDENT $\beta$

Let us study as a simple example the following Langevin equation

$$\frac{d}{dt} v(t) = -\gamma v(t) + \sqrt{2\gamma \frac{1}{m \beta(t)}} g(t),$$

(1)

where $\gamma$ and $m$ are constants and $g(t)$ corresponds to Gaussian white noise with unit variance. $v(t)$ is a stochastic process that can be associated with the velocity of a Brownian particle. If $\beta(t) = \beta$ is a constant, then one ends up with the standard Langevin equation which describes the dynamical process of a Brownian particle with mass $m$ in an environment with constant temperature $1/\beta$. It is well known that the stationary velocity distribution of a Brownian particle is a Gaussian distribution with zero mean and variance $1/\beta$. The relaxation time is $\tau = 1/\gamma$. In superstatistics, one considers a Brownian particle moving through an environment with a slowly fluctuating temperature field whose changes take place on a typical time scale $T$ such that $\tau << T$. When this inequality holds, the local velocity distribution of the system can relax to a Gaussian distribution before the next change of $\beta(t)$ takes place. As such, after a long time, the stationary velocity distribution $P(v)$ of the particle is just a superposition of Gaussian distributions weighted with a function $f(\beta)$

$$P(v) \approx \int_{\beta_{\text{min}}}^{\beta_{\text{max}}} d\beta f(\beta) \sqrt{\frac{m\beta}{2\pi}} \exp \left( -\frac{1}{2} m\beta v^2 \right).$$

(2)

This $f(\beta)$ is the probability density to observe some value of $\beta$. Depending on the properties of $f(\beta)$, different results for the stationary velocity distribution $P(v)$ will occur [19], e.g., power-laws or stretched exponentials. In
a method was introduced to determine $f(\beta)$, given an experimental time series. The method extracts the main superstatistical parameters out of a given data set and examines the validity of the superstatistical model assumptions. Depending on the system under study, one can obtain different results for $f(\beta)$. We will first briefly outline this method to extract $f(\beta)$ from the data.

The starting point is a discrete time series $v$ containing $n$ data points. Using the definitions proposed in Refs. [2, 21], one can estimate values for the two different time scales $\tau$ and $T$. In case that the inequality $\tau/T << 1$ holds, one proceeds by dividing the time series $v$ in $N$ different time slices of length $t$ with $N = \lfloor n/t \rfloor$, where $\lfloor x \rfloor$ means rounding the value of $x$ to the nearest lower integer. Then, one calculates the variance of $v$ in each of these sub-intervals. The inverse of the variance is an estimator for the value of $\beta$. As such, one obtains a new series containing $N$ points which are denoted as $\beta_i$ with $i = 1, 2, \ldots, N$. Within this assumption, the distribution $P(v)$ is approximated by

$$P(v) \approx \frac{1}{N} \sum_{i=1}^{N} \sqrt{\beta_i} \frac{1}{2\pi} e^{-\frac{1}{2}\beta_i v^2}.$$  

(3)

When $N$ is large enough, one can replace expression (3) by (2) with $f(\beta)$ being the probability density that the value of the inverse variance in a randomly chosen time slice of length $T$ equals $\beta$. Notice that the superstatistical approach includes two approximations. In the first step, one assumes the existence of two time scales $\tau$ and $T$ such that in every time slice 'local' equilibrium is reached. Then, $P(v)$ can be approximated by (3). In the second step one assumes the existence of a distribution $f(\beta)$ replacing the summation in expression (3) by an integral. Then, $P(v)$ can be approximated by (2).

III. APPLICATION TO WIND SPEED FLUCTUATIONS

In the following, as a working example, we apply our method to an experimentally measured time series of the horizontal component $v(t)$ of wind speeds recorded at the Lammefjord site at a height of 10m (cup 1) during the year 1987 [33]. The incremental distribution of $u(t) := v(t + \delta) - v(t)$ exhibits non-Gaussian behavior that can be modelled using superstatistics [41]. As an illustration, we calculated the superstatistical approximation of the incremental distribution ($\delta = 8$) of the wind velocities at day 191. The measuring frequency is 8Hz. This means that the number of data points is $n = 24 \times 3600 \times 8 \approx 7 \times 10^5$. The two different time scales extracted from the data are $\tau \approx 4.1$ and $T \approx 112$, obtained by using similar techniques as in Refs. [2, 23]. This illustrates that the data set shows clear time scale separation, in agreement with the results of Ref. [26]. Knowing the value of $T$, one can construct the distribution $f(\beta)$ and try to approximate this histogram with some well-known distributions such as the gamma distribution, the lognormal distribution or the inverse gamma distribution. The latter one is given by

$$f(\beta) = \frac{\theta^\alpha}{\Gamma(\alpha)} \beta^{\alpha-1} e^{-\theta/\beta}.$$  

(4)

Here $\alpha$ and $\theta$ are parameters. The relevance of the above three distributions was motivated in Ref. [2]. The results of our calculations are shown in Figure 1. There is an excellent agreement between the histogram extracted from the data and the inverse gamma distribution. The right part of the figure also shows the empirical distribution $P(u)$ together with the first and second approximation of superstatistics given by eqs. (3) and (2), respectively. The first approximation of superstatistics well models the fat tails of the empirical distribution $P(u)$, see Figure 1. The excellent agreement between $P(u)$ and expression (2), with $f(\beta)$ being the inverse gamma distribution, gives further evidence that this latter distribution is well able to represent the fluctuations of $\beta$. Summarizing, our simple superstatistical model discussed so far is a good first-order approximation of the process of wind speeds increments.

Figure 2 shows the extracted time scale ratio $T/\tau$ as a function of $\delta$. Apparently, for small $\delta$ the time scale separation between $T$ and $\tau$ is less pronounced. The right part of the figure also shows a parameter $\epsilon$ as a function of $\delta$. This $\epsilon$ was defined in Ref. [20] as a measure of quality of the superstatistical approximation. The smaller $\epsilon$, the better the superstatistical model assumptions are satisfied for the given time series.

IV. SUPERSTATISTICS AND OVERDAMPED MOTION

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations

$$\frac{d}{dt}v(t) = -\gamma v(t) + \frac{1}{m} F(x) + \sqrt{\frac{2\gamma}{m}} \frac{1}{\beta(x)} g(t).$$

We will now generalize the Langevin approach to superstatistics. We start from the following set of equations
\[ \frac{d}{dt} x(t) = v(t), \]  
(5)

with \( F(x) = -\frac{\partial V(x)}{\partial x} \). Here \( V(x) \) is a confining potential like, e.g., a harmonic potential, and \( x \) is a position variable. Crucial for our approach is that \( \beta \) depends on \( x \) but not on the velocity \( v \). We are interested in the stationary velocity distribution \( P(v) \). It is well known that in case \( \beta(x) = \beta \) is a constant, this distribution becomes a Gaussian distribution

\[ P(v) = \sqrt{\frac{m\beta}{2\pi}} \exp \left( -\frac{1}{2} m\beta v^2 \right). \]  
(6)

However, we are interested in the general problem where the temperature is position dependent. Then, in order to obtain an exact expression for \( P(v) \), one has to solve the corresponding stationary Fokker-Planck equation for the joint distribution \( P(v, x) \),

\[ 0 = \left[ \frac{\partial}{\partial v} \left( \gamma v - \frac{1}{m} F(x) \right) - v \frac{\partial}{\partial x} + \frac{\gamma}{m} \frac{1}{\beta(x)} \frac{\partial^2}{\partial^2 v} \right] P(v, x) \]  
(7)
and integrate out the dependence of the variable $x$

\[ P(v) = \int_{x_{\text{min}}}^{x_{\text{max}}} dx P(v, x). \]  

(8)

The problem with this procedure is that usually it is not possible to obtain an analytical expression for $P(v, x)$. However, in order to make the connection with the superstatistical approach outlined in the previous section, it is not necessary to have the exact form of $P(v, x)$. It is sufficient to study (5) in the high friction limit, i.e., $\gamma \to \infty$. Taking this overdamped limit of (5) results in a stochastic differential equation in the variable $x(t)$ only:

\[ \frac{d}{dt}x(t) = \frac{1}{m\gamma} F(x) + \sqrt{\frac{2}{m\gamma}} \frac{1}{\beta(x)} g(t) \]  

(9)

The associated Fokker-Planck equation for the stationary distribution $P_0(x)$ becomes

\[ 0 = -\frac{\partial}{\partial x} F(x) P_0(x) + \frac{\partial^2}{\partial^2 x} \frac{1}{\beta(x)} P_0(x) \]  

(10)

(we used Ito’s interpretation). The lower index $_o$ is used to denote the overdamped limit. Equation (10) can be solved analytically

\[ P_0(x) = Z^{-1} \beta(x) \exp \left( \int dx F(x) \beta(x) \right), \]  

(11)

where $Z$ is a normalization constant. Taking the overdamped limit physically means that the velocity will thermalise very quickly. As a consequence, for large values of $\gamma$, the solution $P(v, x)$ of the complete Fokker-Planck equation (7) is approximated by (12)

\[ P(v, x) \approx P_0(v, x) = P_0(x) P_0(v|x). \]  

Here $P_0(v|x)$ denotes the conditional distribution of $v$ given the local value of the slow variable $x$. This is just a Gaussian distribution (6) with $\beta$ replaced by $\beta(x)$. As a consequence, one obtains the following approximation for the distribution of interest:

\[ P(v) \approx \int_{x_{\text{min}}}^{x_{\text{max}}} dx P_0(v, x) = \int_{x_{\text{min}}}^{x_{\text{max}}} dx P_0(x) \sqrt{\frac{m\beta(x)}{2\pi}} \exp \left( -\frac{1}{2} m\beta(x)v^2 \right). \]  

(13)

After a change of variables, one ends up with (14) in which the superstatistical distribution $f(\beta)$ is associated with

\[ \left( \frac{d\beta(x)}{dx} \right)^{-1} P_0(\beta(x)). \]  

(14)

This shows that the superstatistical approximation is formally equivalent to taking the high friction limit of a stochastic differential equations of type (5). The latter model describes the dynamics of a Brownian particle in an inhomogeneous heat bath in the presence of a confining potential.

V. EXAMPLE

To illustrate the theoretical considerations of the previous section, we study an explicit example here. Assume for simplicity that the temperature is linear in the position $x$. In order to ensure that the temperature remains positive we introduce an arbitrary small positive constant $a$ and define

\[ \beta(x) = \frac{1}{|x| + a}, \quad V(x) = \frac{\theta}{2} \left[ |x| + \frac{\alpha}{\theta} \right]^2. \]  

(15)

With this choice one obtains for the velocity distribution in the overdamped limit (13)

\[ P(v) \approx Z^{-1} \int_{-\infty}^{+\infty} dx \beta(x) \exp \left( -\int dx \frac{\partial V(x)}{\partial x} \beta(x) \right) \sqrt{\frac{m\beta(x)}{2\pi}} \exp \left( -\frac{1}{2} m\beta(x)v^2 \right). \]
\[ P(v) \approx \frac{1}{\Gamma(\alpha)} \left( \frac{2\theta}{\pi} \right)^{\alpha-1/2} K_{\alpha-1/2} \left( \sqrt{2\theta |v|} \right) \]

where \( K_{\nu}(x) \) is the modified Bessel function of the second kind. Notice that \( f(\beta) \) becomes an inverse gamma distribution in this limit. This particular example connects the results of sections III and IV because it shows that the theoretical model of section IV together with the choice (15) for the temperature and the potential reproduces the observed empirical distributions discussed in section III. However, this is only one possibility. In principle, for every choice of \( \beta(x) \), a possible candidate for the force that results in an inverse gamma distribution for \( f(\beta) \) can be calculated by evaluating the following expression

\[ F(x) = -\alpha + 2 \frac{\beta'(x)}{\beta(x)} - \theta \frac{\beta'(x)}{\beta^2(x)} + \frac{\beta''(x)}{\beta'(x)}. \]

Where the accent means differentiating with respect to \( x \). The reason to chose the particular form of \( T(x) \) and \( V(x) \) as in (15) is that this is probably the easiest choice that both fulfills equation (18) and is physically meaningful. The latter is also the reason that we introduced the constant \( a \) in the expression for the temperature (15). This is to ensure that the particle can pass the origin in the overdamped limit. This problem disappears if one takes finite friction into account. Therefore, it is appropriate to restrict our analysis to the special case \( a \to 0 \).

VI. SKEWNESS

We showed in section III that the superstatistical approach is well suited to describe the fat tails that are exhibited in the wind speed distribution under study. However, it is well known, and clearly visible in figure 1, that the distribution \( P(u) \) is slightly skewed (with vanishing first moment). Clearly, the superstatistical distribution (2) is symmetric and can not represent the skewness that occurs in the experimental data. Figure 3 (right part) shows the skewness and kurtosis as a function of \( \delta \) for the data set of measured velocity increments \( u_\delta(t) \). It is possible to model the kurtosis with the superstatistical approach described up till now, however it is impossible to model the skewness.

As mentioned before, the complete Fokker-Planck equation (14) is usually not analytically solvable. However, relations between the moments of the distribution \( P(v, x) \) can be obtained, using the Fokker-Planck operator \( L \). One
\[ \langle L v^i x^j \rangle = 0 \text{ for } i, j = 0, 1, 2, \ldots \text{ with } L := \left( \gamma v - \frac{1}{m} F(x) \right) \frac{\partial}{\partial v} - v \frac{\partial}{\partial x} - \frac{\gamma}{m} \beta(x) \frac{\partial^2}{\partial v^2} \]  

Applying this with \( i = 0, j = 1 \) results in \( \langle v \rangle = 0 \), while depending on the particular choices of \( F(x) \) and \( \beta(x) \) one can obtain \( \langle v^3 \rangle \neq 0 \). This means that the stationary velocity distribution \( P(v) \) can have non-vanishing skewness, while at the same time it has a vanishing first moment. However notice that the overdamped approximation \([12]\) of \( P(v, x) \) is symmetric in \( v \), regardless of the choice of \( \beta(x) \). This means that in order to introduce skewness, one has to go beyond the overdamped limit. Perturbation expansions of \( P(v, x) \) around the overdamped limit can be performed \([12]\). The first order correction is

\[
P(v, x) = \sqrt{\frac{m\beta(x)}{2\pi}} \exp\left(-\frac{1}{2} m\beta(x)v^2 \right) P_o(x) \left( 1 + \frac{1}{6} v [m\beta(x)v^2 - 3] \frac{1}{\beta(x)} \frac{d\beta(x)}{dx} \right). \tag{20}
\]

Notice that the contribution of the first order correction vanishes for \( \langle v^i \rangle \) with \( i = 0, 1, 2 \). The third order moment becomes

\[
\langle v^3 \rangle = \int_{-\infty}^{+\infty} dv \int_{-\infty}^{+\infty} dx v^3 P(v, x) = \frac{1}{\gamma m^2} \int_{-\infty}^{+\infty} dx P_o(x) \frac{1}{\beta^3(x)} \frac{d\beta(x)}{dx}. \tag{21}
\]

This shows that the theoretical model introduced in section IV can be used to systematically describe the experimental data. The superstatistical approximation (the overdamped limit) is used to model the fat tails (the kurtosis) of the empirical distribution. By going beyond the overdamped limit, skewness is naturally introduced in the theoretical distribution.

VII. DISCUSSION

In previous papers dealing with superstatistics, the superstatistical concept was completely probabilistic in nature. One was facing the question what the relevant temperature distributions \( f(\beta) \) are. In this paper, we have shown that the superstatistical model is equivalent to a dynamical process with a position dependent temperature field in the overdamped limit. This opens up the possibility to relate the choice of \( f(\beta) \) to a concrete question with a clear physical interpretation, namely what are appropriate choices for the potential and the temperature fields.

One such example was studied in this paper. If one assumes a temperature field that is linear in the position, in combination with basically a harmonic potential, one ends up with an inverse gamma distribution for \( f(\beta) \). This distribution is indeed observed in empirical data of wind speed fluctuations. We emphasize that other choices for temperature field and potential are of course possible. It is an interesting topic for further research to examine other physically meaningful combinations of \( \beta(x) \) and \( V(x) \). One can also try to extract the position dependence of \( \beta(x) \) and \( V(x) \) immediately out of empirical data, see e.g. \([14]\) where such analysis is performed for financial time series.

Finally, we showed that the theoretical approach to superstatistics presented in this paper naturally leads to skewed probability distributions if one goes beyond the overdamped limit. Slightly skewed distributions are observed in several scientific fields for various variables, e.g., wind speed fluctuations (see section VI), log returns of prices in the stock market \([12]\), and velocity increments of turbulent flows \([10]\). Apart from the observed skewness in the data, these three examples have in common that the symmetric part of the empirical distribution can be well approximated by \([2]\) with different expressions for \( f(\beta) \), see, e.g., the present paper and \([24]\). In the theoretical model developed in this paper, expression \([2]\) is the zeroth-order approximation of the true superstatistical distribution which is obtained by fully taking into account the position dependence of \( \beta(x) \) and \( V(x) \). Therefore, the current approach is very promising, since one can now go beyond \([2]\) and try to calculate successive higher-order corrections in order to describe all the essential features of an empirical distribution at hand. Apart from the aforementioned examples, also dense granular flows are interesting experimental systems from a superstatistical point of view, because position dependent temperature profiles can be estimated for these kinds of systems. Also, it has been shown that the particle velocity distribution shows skewness \([17]\).
