① 

$$P_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sigma} n\left(\frac{x-\mu}{\sigma}\right)$$

Note that

$$P_{x+y}(t) = \int_{\mathbb{R}} P_x(x) \, P_y(t-x) \, dx$$

is the convolution of $P_x$ and $P_y$:

$$P_{x+y}(t) = (P_x * P_y)(t)$$

The convolution theorem states that

$$\mathcal{F}(f*g) = \sqrt{2\pi}\, \mathcal{F}(f) \cdot \mathcal{F}(g)$$

where $\mathcal{F}$ is the Fourier-transform.

Recall that the standard normal distribution is a fixed point of the FT:

For $n(x) = e^{-\frac{x^2}{2}}/\sqrt{2\pi}$,

$$\mathcal{F}(n)(k) = n(k).$$

Using $\mathcal{F}(f(ax))(k) = \frac{1}{|a|} \mathcal{F}(f(x))\left(\frac{k}{a}\right)$

and $\mathcal{F}(f(x-a))(k) = e^{-ika} \mathcal{F}(f(x))(k)$

we get that $\mathcal{F}(P_{\mu,\sigma}(x))(k) = \frac{\sigma}{\sigma} e^{-ik\mu} n(\sigma k)$

$$\Rightarrow \mathcal{F}(P_{x+y})(k) = \frac{\sqrt{2\pi}}{2\pi} e^{-ik(\mu_1+\mu_2)} e^{-\frac{1}{2}k^2(\sigma_1^2 + \sigma_2^2)}$$

$$= e^{-ik(\mu_1+\mu_2)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sigma_1^2+\sigma_2^2)k^2}$$

This is the FT of a normal distribution with $\mu = \mu_1+\mu_2$ and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.

The statement follows from the uniqueness of the FT. $\square$

# Data and Signal Analysis
# Problem Sheet 1

Elise Pilgermann, Max Maschke

November 5, 2024

## 2. Analysing a time series

### 2.a.

The time series is shown in figure 1a. We notice a trend in the data that is clearly linear over the given interval.

### 2.b.

To detrend the time series, we perform a linear regression, i.e. a least squares fit to a function $S(t) = \alpha t + \beta$, using the `linregress` function from the `scipy.stats` module. We obtain the parameters
$$\alpha = 0.00109956 \pm 5.8 \cdot 10^{-7}, \quad \beta = 3.318 \pm 0.034$$

and an $R^2$-value of 0.986. The fit is also shown as the dashed line in figure 1a, and it can clearly be seen that it approximates the underlying trend well.

To detrend the time series, we simply subtract the value of the fitted trend function from the data at every discrete time step. This results in a noisy modified series shown in figure 1b.

### 2.c.

The raw and detrended histograms are shown in figure 2. The histogram of the raw time series in figure 2a is very flat over most of the span of the $x$-range, dropping of quickly at its edges. Assuming a symmetric distribution in the deviation from the linear trend, all values that the trend function assumes should be approximately equally likely to appear in the time series, so the raw histogram should approximately look like a uniform distribution except for the edges where the distribution of the deviation from the trend "shines through".

The suspicion that the deviation follows an approximately symmetric distribution is confirmed by inspecting the detrended histogram, 2b. It has an approximately Gaussian shape (this is not just the flawed judgement of the eye but will be underpinned by later analysis) with a slight skew to the right of the origin.

In the limit of $\alpha \longrightarrow 0$, both histograms will be identical. Conversely, in the limit of $\alpha \longrightarrow \infty$, the distribution of the raw time series will become even broader and tend to zero (while staying normalised). As long as a good fit of the trend is still possible, the histogram of the deviation should not change noticably.
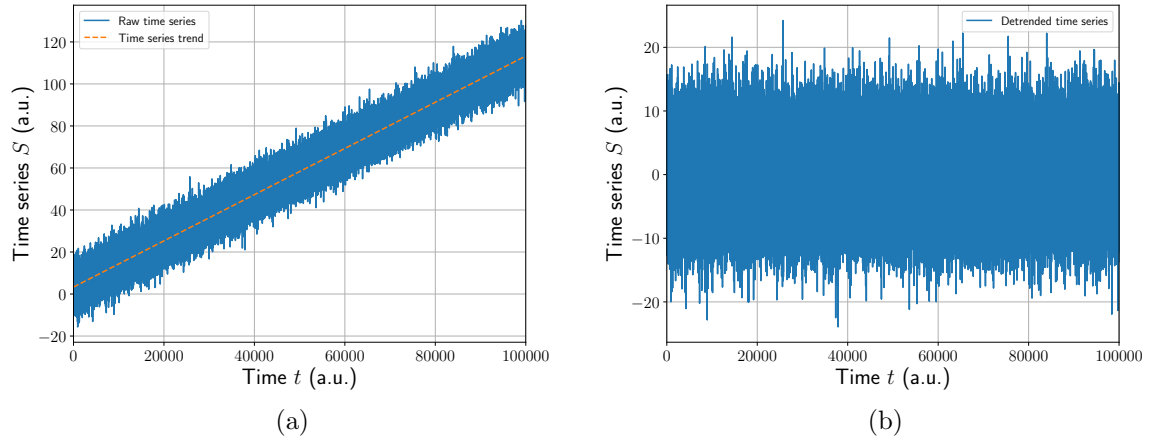
Figure 1: Time series given in `timeseries.txt` including trend obtained through linear regression 1a and detrended using said fit 1b.
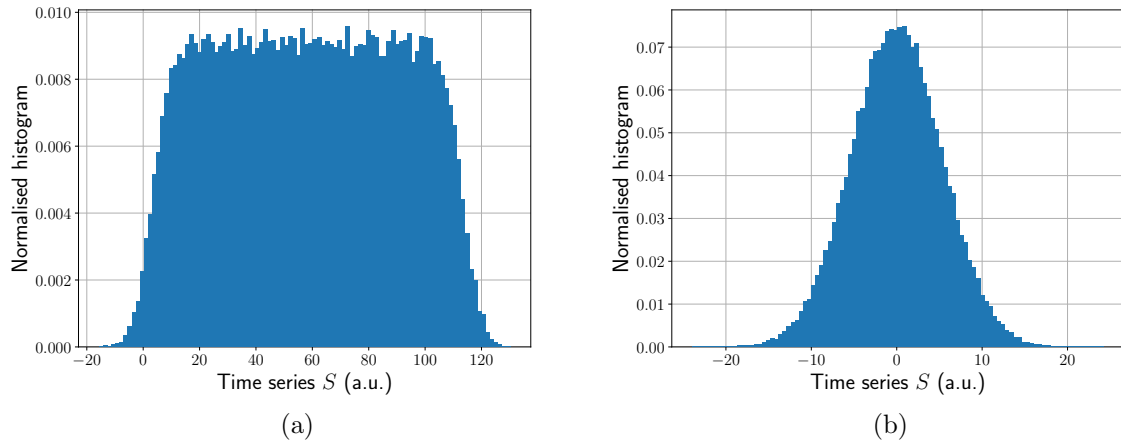


Figure 2: Histogram of the raw time series 2a and of the detrended time series 2b.

Table 1: Empirical centered moments of the histograms of the raw and detrended time series.

| Order | Raw time series | Detrended time series |
|:-----:|:---------------:|:---------------------:|
| 0 | 1 | 1 |
| 1 | 0.00118 | $-8.8 \cdot 10^{-5}$ |
| 2 | 1040 | 28.2 |
| 3 | 22.0 | 0.645 |
| 4 | 2010000 | 2400 |

## 2.d.

Normalised empirical histograms for datasets can be conveniently obtained using the `histogram` function from the `numpy` module (care must be taken to call it with the optional parameter `density=True`).

The histogram is returned as a vector containing the values of the histogram in the bins, so the empirical centered moments can be conveniently calculated as

$$m_n = \mathrm{d}x \cdot (x - \mu)^n \cdot h \tag{1}$$

where $h$ is the histogram vector, $x$ is a linearly spaced vector spanning the range of the dataset, $\mu$ is the mean of the dataset and $\mathrm{d}x$ is the separation of the bins.

The moments obtained in this way are shown in table 1. The 0-th order moments are equal to unity as expected. The first order moment is approximately equal to zero for both histograms because they are centered, otherwise they should yield the mean. The second order moments give the variance. The third order moment is the skewness of the histogram, which is small for both but especially that of the deviation, i.e. the detrended time series. The positive sign indicates that both show a small skew to the right of the mean. The fourth order moment, the kurtosis, divided by the variance squared is exactly equal to three for a Gauussian. In our case, we get a value of 3.011, showing that it is very close to a Gaussian and was likely sampled from one.