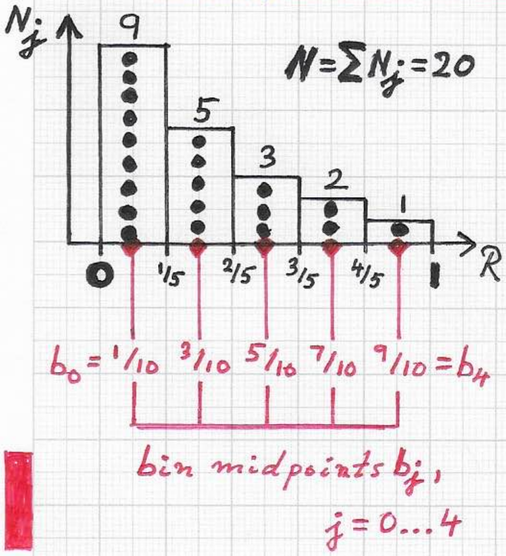
■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks:... Our goal is to use statistical characteristics to a discrete representation of the numbers N_i and associated intervals $[R_i, R_{i+1})$. For simplicity, we assume in the following that these intervals have the same width, i.e., $(R_{i+1} - R_i) = \Delta_i = \Delta = \text{constant}$. Using 10 R-intervals for the partitioning of $[0,1]$, the R_i-values are $R_0 = 0$, $R_1 = 1/10$, $R_2 = 2/10, \ldots, R_{10} = 1$. The computed statistical characteristics – mean, variance, skewness and kurtosis – can be used to determine whether to subdivide/split a point set or not. We briefly review these characteristics for the case of discrete distribution data ("histogram data for constant-width bins"). We consider a simple example to calculate the needed moments and desired derived statistical characteristics (left figure). The raw moments are:



$$M_1 = 1/20 \sum_{j=0}^{4} r_j = 1/20 \sum_{j=0}^{4} N_j \cdot b_j$$
$$= 1/20 \, (9b_0 + 5b_1 + 3b_2 + 2b_3 + 1b_4)$$
$$= 1/20 \left(9 \cdot \tfrac{1}{10} + 5 \cdot \tfrac{3}{10} + 3 \cdot \tfrac{5}{10} + 2 \cdot \tfrac{7}{10} + 1 \cdot \tfrac{9}{10}\right)$$
$$= 1/200 \, (9 + 15 + 15 + 14 + 9)$$
$$= 62/200 = 31/100 = \mathbf{0.31}$$

$$M_2 = 1/20 \sum_{j=0}^{4} N_j \cdot b_j^2 = 1/20 \, (9b_0^2 + 5b_1^2 + 3b_2^2 + 2b_3^2 + 1b_4^2)$$
$$= 1/20 \cdot 1/100 \cdot (9 + 5 \cdot 3^2 + 3 \cdot 5^2 + 2 \cdot 7^2 + 1 \cdot 9^2) = 1/2000 \, (9 + 45 + 75 + 98 + 81)$$
$$= 308/2000 = 154/1000 = \mathbf{0.154}$$

• • •

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• **Laplacian eigenfunctions and neural networks:...**

$M_3 = 1/20\ \Sigma_{j=0}^{4}\ N_j \cdot b_j^{\,3} = 1/20\,(9 b_0^{\,3} + \ldots + 1 b_4^{\,3})$
$= 1/20 \cdot 1/1000\,(9 + 5\cdot 27 + 3\cdot 125 + 2\cdot 343 + 729)$
$= 1/20000 \cdot 1934 = 967/10000 = 0.0967$

$M_4 = 1/20\ \Sigma_{j=0}^{4}\ N_j \cdot b_j^{\,4} = 1/20\,(9 b_0^{\,4} + \ldots + 1 b_4^{\,4}) = 1/200000\,(9 + 5\cdot 81 + \ldots + 1\cdot 9^4)$
$= 1/200000\,(9 + 405 + 1875 + 4802 + 6561) = 13652/200000 = 6826/100000 = 0.06826$

The raw moment M_1 is the mean. By performing mean subtraction and performing the same computations again — "about the mean" — one obtains the moments after mean subtraction; they are:

$m_1 = 1/20\ \Sigma_{j=0}^{4}\ N_j \cdot (b_j - M_1) = 1/20\ \Sigma_{j=0}^{4}\ (N_j \cdot b_j - N_j \cdot M_1) = M_1 - M_1 = 0$
$m_2 = \ldots = M_2 - M_1^{\,2} = 0.154 - 0.0961 = 0.0579$

The moment m_2 is the variance. The square root of it is the standard deviation s, i.e., $s = \sqrt{m_2}$. In our example, we obtain $s = 0.240624$. Further,

$m_3 = \ldots = M_3 - 3 M_1 M_2 + 2 M_1^{\,3} = 0.013062$
$m_4 = \ldots = M_4 - 4 M_1 M_3 + 6 M_1^{\,2} M_2 - 3 M_1^{\,4} = 0.00944277$

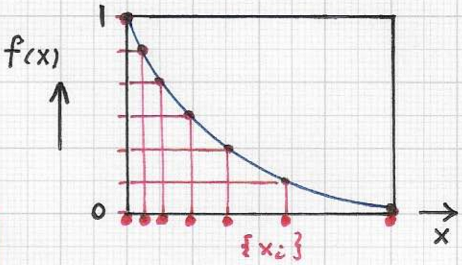
By performing division by the standard deviation s and performing the same computations again one obtains the standardized moments; they are:

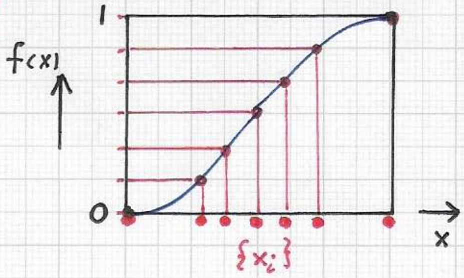
$a_1 = 1/20\ \Sigma_{j=0}^{4}\ N_j \cdot (b_j - M_1)/s = 1/s\ m_1 = 0$
$a_2 = \ldots = m_2/s^2 = m_2/m_2 = 1$
$a_3 = \ldots = m_3/s^3 = 0.013062 / 0.0139321405 = 0.93754438$
$a_4 = \ldots = m_4/m_2^{\,2} = 0.00944277 / 0.00335241 = 2.816711$

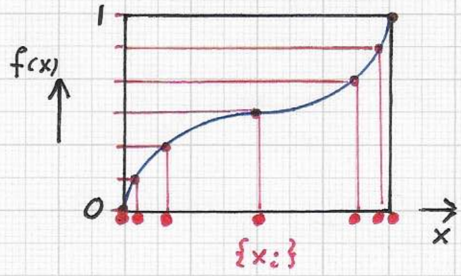
■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks:... The standardized moment a_3 is the skewness. One can define three cases: (i) $a_3 = 0$: symmetrical case; (ii) $a_3 < 0$: negatively skewed case; (iii) $a_3 > 0$: positively skewed case. The value $a_4 - 3$ is the value of kurtosis. One can interpret the kurtosis value as follows: $a_4 - 3 < (>) \, 0$: the "peakness"/"peakedness" is less (more) than that of a "normal curve."

⇒ The mean, variance, skewness and kurtosis values are the main shape characteristics.







One can also view such "shape characteristics" of a given discrete sample set $\{x_i\}$ in the context of sampling and reconstructing a function $f(x)$, see left figures. All functions are rather "simple"; they are bijective, and the last two functions could potentially be cumulative distribution functions. Here, the function $f(x)$ has been sampled for $f = j/6$, $j = 0 \ldots 6$, yielding the associated x_j-values. The sets $\{x_i\}$ can be understood – together with their mean, variance, skewness and kurtosis values – as input for a reconstruction method to generate approximations of f.

...

■ OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

• Laplacian eigenfunctions and neural networks:... • Note. Distributions – Moments – Reconstruction. It is a fact that the univariate distributions / probability densities (in the univariate case) are "relatively simple functions" — in the sense that they can often be sufficiently characterized by a small number of parameters, i.e., moments. This is true for both discrete and continuous distributions, including binomial, exponential, normal and Poisson distributions. Nevertheless, it appears that the reconstruction of an "appropriate continuous probability model function" from a finite, discrete set of observations, generated experimentally or computationally, is a difficult task that requires more research. Research publications concerned with this reconstruction problem, including reconstruction via moments, include:

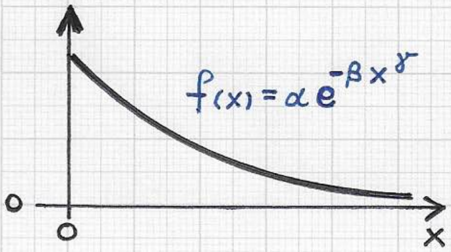
- Lebaz, Cockx, Spérandio and Morchain, "Reconstruction of a distribution from a finite number of its moments:...," 2015.
- John, Angelov, Öncül and Thévenin, "Techniques for the reconstruction of a distribution from a finite number of its moments," 2007.
- Dang and Xu, "Novel algorithm for reconstruction of a distribution by fitting its first-four statistical moments," 2019.

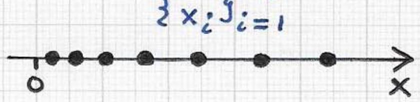
Typically, one assumes that an observed distribution has an underlying continuous model of a specific kind of model, rather than allowing the model type to be constructed as well.

•••

OBJECT AND MATERIAL EIGENFUNCTIONS – Cont'd.

- Laplacian eigenfunctions and neural networks: ...





For completeness, we briefly describe the general principle underlying the so-called Method of Moments (MOM) that is often used for the numerical approximation of (usually) a small number of values needed to "complete" the definition of a distribution function of a specific kind.

The figure sketches a simple scenario. For example, one "knows a priori" or "postulates for theoretical reasons" that it is an exponentially decreasing distribution function of the kind $f(x) = \alpha e^{-\beta x^{\gamma}}$ that is the underlying continuous function determining an experimentally observed discrete, finite sample set $\{x_i\}_{i=1}^{n}$. The values of the three parameters α, β and γ are to be determined, i.e., approximated, via a sample set $\{x_i\}_{i=1}^{n}$, with n ideally being "large." The MOM works as follows: For example, let us assume that the first three moments of $f(x)$, called μ_1, μ_2 and μ_3, can be written as functions of α, β and γ; i.e., $\mu_j = f_j(\alpha, \beta, \gamma)$, $j = 1, 2, 3$. The finite sample set yields the moment values (approximations) $\bar{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} (x_i)^j$, $j = 1, 2, 3$. One assumes that $\bar{\mu}_j = f_j(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$, i.e., $\bar{\mu}_j$ is also defined via f_j, having the approximations $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ as arguments. Solving the equation system $\bar{\mu}_j = f_j(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ for $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ yields α, β and γ estimates.

...