HEAT-FLOW TRANSFORMATION OF A PROBABILITY DISTRIBUTION

OVIDIU CALIN

Communicated by Lucian Beznea

The first four moments of a distribution are the most significant from the statistical point of view. This paper studies how the mean, variance, skewness and kurtosis change while a probability distribution is deformed according to a heat-flow equation. Closed-form formulas are provided for these four moments of the deformed distributions and some conclusions are drawn. We apply these results to develop a normality test and provide applications to PCA in multivariate case.

AMS 2010 Subject Classification: 92B99, 92B05.

Key words: heat-flow, mean, variance, skewness, kurtosis, normality test, PCA.

INTRODUCTION

In this paper $q(x)$ will denote a probability density function on $\mathbb{R}$, ie., a non-negative function with $\int_{\mathbb{R}} q(x)\,dx = 1$. We shall investigate how the first few moments of $q(x)$, including the mean, variance, skewness and kurtosis are influenced by the the following heat-flow evolution equation

$$\frac{\partial}{\partial t}p_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p_t(x) \tag{1}$$

$$p_0(x) = q(x), \qquad \forall x \in \mathbb{R},\ t > 0. \tag{2}$$

The physical significance of the problem is the following: if $q(x)$ denotes the heat density along an infinite rod, then $p_t(x)$ represents the density in the rod at time $t > 0$, under the assumption that no exterior heat sources are involved and the heat evolves from the initial density $q(x)$. For more details the reader is referred to the book [1].

The geometrical significance consists in the fact that the left side of the equation, $\frac{\partial}{\partial t}p_t(x)$, represents the rate of change of the density at point $x \in \mathbb{R}$, while the right side of the equation, $\frac{1}{2}\frac{\partial^2}{\partial x^2}p_t(x)$, measures the convexity, or curvature of the density function. Along the intervals where the density

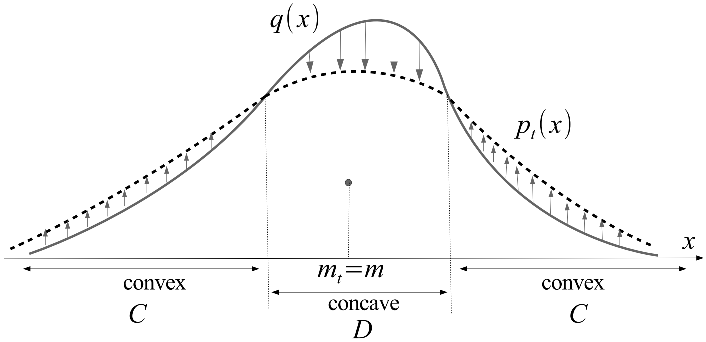

Figure 1: The heat-flow deformation of $q(x)$ into $p_t(x)$. The mean m_t remains invariant, while the spread of the density increases.

function is convex, namely along the set $C = \{x \in \mathbb{R}; \frac{\partial^2}{\partial x^2} p_t(x) > 0\}$ the rate of change is positive, and hence, $p_t(x)$ is increasing in time; while along the set $D = \{x \in \mathbb{R}; \frac{\partial^2}{\partial x^2} p_t(x) < 0\}$ the rate of change is negative and the density decreases in time. This means that $p_t(x)$ becomes thicker on C and more flat on D.

The typical example of continuous distribution density $q(x)$ on $\mathbb{R}$ is fast decreasing to infinity, having the x-axis as a horizontal asymptote to $\pm\infty$, and hence $q(x)$ is convex for $|x|$ large enough. The previous heat equation will provide an evolution towards fatter tails of the distribution and lower peak in the middle, where the distribution is concave. See Fig.1.

One question is what moments of $q(x)$ are preserved during this evolution? Another is how other moments are changing over time? Does the distribution ever become symmetric or close to a normal distribution? This paper answers all these questions. Besides these, it also treats the case of the multivariate case distributions and provides some relation to Principal Component Analysis (PCA). More precisely, we show that the eigenvalues of the covariance matrix are described by the time instances when the heat-flow deformation of the random variable has a degenerate covariance matrix.

It is worthy to note the relation of our problem with the filtering point of view. Since the solution $p_t(x)$ of the heat equation is obtained by a convolution of $q(x)$ with a Gaussian, we may regard $p_t(x)$ as being a blurred version of $q(t)$. In this case the question is to look at how do the moments of $q(x)$ behave under blurring. Looking for the optimal blur can be useful for signal processing. A signal which is too sharp may include too much information, some of it being an effect of the embedded noise, and some filtering is needed to remove the undesired detail. On the other side, a signal which is too blurred may have too little detail left and therefore, due to too much loss of information, it might become useless. So that, somewhere between these two limits might be the optimal situation of the best blur, which can be achieved using the results of this paper.

1. THE PROBABILITY DENSITY

In this section we shall show that the evolution profile $p_t(x)$ is a probability density for any $t > 0$. From the physical point of view, this is a consequence of the conservation of energy law, since there is no exterior heat injected or extracted from the system. More precisely, if $Q = \int_{\mathbb{R}} q(x)\, dx$ represents the total heat in the rod at time $t = 0$, then the heat at time $t > 0$ is given by $Q_t = \int_{\mathbb{R}} p_t(x)\, dx$. Since $q(x)$ is a probability density, then $Q = 1$. If the heat conservation holds, then $Q_t = Q$, and hence $Q_t = 1$, ie., $p_t(x)$ is also a probability density. We shall provide next a mathematical proof of this fact.

PROPOSITION 1.1. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$. Then the solution $p_t(x)$ of the system (1)-(2) is a probability density for all $t > 0$.*

Proof. Consider the Gaussian

$$G_t(x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}(x-y)^2}, \qquad t > 0.$$

It is known that the solution of the Cauchy problem (1)-(2) is given by the convolution

$$\begin{aligned} p_t(x) &= (q * G_t)(x) = \int_{\mathbb{R}} q(y) G_t(x,y)\, dy = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} q(y) e^{-\frac{1}{2t}(x-y)^2}\, dy \\ &= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} q(x + z\sqrt{2t}) e^{-z^2}\, dz, \end{aligned} \tag{3}$$

where we employed the substitution $z = (y-x)/\sqrt{2t}$. It is obvious to see now that $p_t(x) \geq 0$, since $q(x) \geq 0$. Using Fubini's theorem and the previous formula for $p_t(x)$, we have

$$\begin{aligned} Q_t(x) &= \int_{\mathbb{R}} p_t(x)\, dx = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \int_{\mathbb{R}} q(x + z\sqrt{2t}) e^{-z^2}\, dz\, dx \\ &= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \Big[\int_{\mathbb{R}} q(x + z\sqrt{2t})\, dx \Big] e^{-z^2}\, dz = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \Big[\underbrace{\int_{\mathbb{R}} q(u)\, du}_{=Q} \Big] e^{-z^2}\, dz \end{aligned}$$

$$= \frac{Q}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-z^2}\,dz = Q.$$

Since $Q = 1$, then $Q_t = 1$ for all $t > 0$, and hence $p_t(x)$ becomes a probability density. $\square$

It is worth noting that $p_t(x)$ tends to zero in the long run. This follows by applying the Dominated Convergence Theorem as

$$\lim_{t\to\infty} p_t(x) = \lim_{t\to\infty} \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} q(x + z\sqrt{2t}) e^{-z^2}\,dz = 0$$

since $q(x)$ tends to zero at infinity.

2. THE MEAN

The means of $q(x)$ and $p_t(x)$ are given respectively by

$$m = \int_{\mathbb{R}} x q(x)\,dx, \qquad m_t = \int_{\mathbb{R}} x p_t(x)\,dx,$$

From the physical point of view, m and m_t represent the x-coordinates of the center of mass of the regions under the density function $q(x)$ and $p_t(x)$, respectively. The next result shows that the center of mass is preserved under the heat-flow; this shows that the first moment is an invariant of the transformation.

PROPOSITION 2.1. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Then $q(x)$ and $p_t(x)$ have the same mean, $m_t = m, \forall t > 0$.*

Proof. We shall employ a direct computation method. Using formula (3) and Fubini's theorem, we have

$$\begin{aligned}
m_t &= \int_{\mathbb{R}} x p_t(x)\,dx = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} x \int_{\mathbb{R}} q(x + z\sqrt{2t}) e^{-z^2}\,dz\;\;dx \\
&= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \int_{\mathbb{R}} x q(x + z\sqrt{2t})\,dx\; e^{-z^2}\;dz \\
&= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \int_{\mathbb{R}} (u - z\sqrt{2t}) q(u)\,du\; e^{-z^2}\;dz \\
&= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \Big[\underbrace{\int_{\mathbb{R}} u q(u)\,du}_{=m} - z\sqrt{2t} \underbrace{\int_{\mathbb{R}} q(u)\,du}_{=1} \Big]\; e^{-z^2}\;dz \\
&= \frac{1}{\sqrt{\pi}} \Big[m \underbrace{\int_{\mathbb{R}} e^{-z^2}\,dz}_{=\sqrt{\pi}} - \sqrt{2t} \underbrace{\int_{\mathbb{R}} z e^{-z^2}\,dz}_{=0} \Big] = m. \quad \square
\end{aligned}$$

Another proof variant would be to notice that m_t is smooth in t with $\lim_{t\to 0} m_t = m$. Then if we show that $\frac{d}{dt}m_t = 0$, it follows that $m_t = m$, for all $t > 0$. Assuming that $p_t(\pm\infty) = 0$ and $x\partial_x p_t(\pm\infty) = 0$, a double integration by parts yields

$$\begin{aligned}\frac{d}{dt}m_t &= \int_{\mathbb{R}} x\frac{d}{dt}p_t(x)\,dx = \frac{1}{2}\int_{\mathbb{R}} x\frac{\partial}{\partial x^2}p_t(x)\,dx\\ &= \frac{1}{2}x\partial_x p_t(x)\Big|_{-\infty}^{\infty} - \frac{1}{2}\int_{\mathbb{R}} x'\partial_x p_t(x)\,dx = 0.\end{aligned}$$

3. THE VARIANCE

Consider the second moments of $q(x)$ and $p_t(x)$ given respectively by

$$m_2 = \int_{\mathbb{R}} x^2 q(x)\,dx, \qquad m_2(t) = \int_{\mathbb{R}} x^2 p_t(x)\,dx.$$

The associated variances are given by

$$V = var(q) = \int_{\mathbb{R}} (x-m)^2 q(x)\,dx, \qquad V_t = var(p_t) = \int_{\mathbb{R}} (x-m_t)^2 p_t(x)\,dx.$$

In this section we shall show that the heat-flow increases the second moment linearly in time, $m_2(t) = m_2 + t$, and we shall show a similar relation for the variance. This means that the blurred density p_t has a dispersion that increases linearly in time.

PROPOSITION 3.1. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Then the variance is increasing with respect to time*

$$V(t) = V + t, \quad \forall t > 0.$$

Proof. We shall show first that $m_2(t) = m_2 + t, \forall t > 0$. A computation involving Fubini's theorem provides

$$\begin{aligned}m_2(t) &= \int_{\mathbb{R}} x^2 p_t(x)\,dx = \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} x^2 \int_{\mathbb{R}} q(x+z\sqrt{2t})e^{-z^2}\,dz\;\,dx\\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\Big[\int_{\mathbb{R}} x^2 q(x+z\sqrt{2t})\,dx\Big]e^{-z^2}\,dz\\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\Big[\int_{\mathbb{R}} (u-z\sqrt{2t})^2 q(u)\,du\Big]e^{-z^2}\,dz\\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\Big[\int_{\mathbb{R}} u^2 q(u)\,du + 2z\sqrt{2t}\int_{\mathbb{R}} uq(u)\,du + 2tz^2\int_{\mathbb{R}} q(u)\,du\Big]e^{-z^2}\,dz\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \Big[m_2 + 2z\sqrt{2t}\, m + 2tz^2 \Big] e^{-z^2}\, dz \\
&= m_2 + 2t \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} z^2 e^{-z^2}\, dz = m_2 + t,
\end{aligned}
$$

where we used $\int_{\mathbb{R}} e^{-z^2}\, dz = \sqrt{\pi}$, $\int_{\mathbb{R}} z e^{-z^2}\, dz = 0$ and $\int_{\mathbb{R}} z^2 e^{-z^2}\, dz = \sqrt{\pi}/2$. Then the relation between variances is given by

$$
V_t = m_2(t) - m_t^2 = m_2 + t - m^2 = V + t,
$$

where we used that $m_t = m$, see Proposition 2.1. $\square$

4. THE SKEWNESS

If X is a random variable of mean m and variance σ^2, then the third moment of the random variable $Z = \frac{X-m}{\sigma}$ is called the skewness of X and represents a measure of non-symmetry for the probability density of X. We have the following equivalent formulas

$$
S(X) = \mathbb{E}\Big[\Big(\frac{X-m}{\sigma}\Big)^3\Big] = \frac{\mathbb{E}[X^3] - 3m\sigma^2 - m^3}{\sigma^3}.
$$

This shows that the skewness of X can be written in terms of the mean, variance and the third moment of X. We shall use a similar formula to define the skewness of a probability density function. If the third centered moments of $q(x)$ and $p_t(x)$ are denoted by

$$
m_3 = \int_{\mathbb{R}} x^3 q(x)\, dx, \qquad m_3(t) = \int_{\mathbb{R}} x^3 p_t(x)\, dx,
$$

the associated skewness functions for $q(x)$ nd $p_t(x)$ are respectively given by

$$
S = S(q) = \frac{m_3 - 3mV - m^3}{V^{3/2}}, \quad S_t = S(p_t) = \frac{m_3(t) - 3m_t V_t - m_t^3}{V_t^{3/2}}.
$$

This section deals with the formula that relates S_t to S. This result is supposed to track the (non)symmetry property of a distribution under the heat-flow.

First we need to establish a relation between the third centered moments.

Lemma 4.1. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Then the relation between their third moments is*

$$
m_3(t) = m_3 + 3mt, \quad \forall t > 0.
$$

Hence, the moment $m_3(t)$ changes linearly in time with slope $3m$.

Proof. First, we note that a change of variables provides

$$\begin{aligned}\int_{\mathbb{R}} x^3 q(x+z\sqrt{2t})\,dx &= \int_{\mathbb{R}} (u-z\sqrt{2t})^3 q(u)\,du \\ &= \int_{\mathbb{R}} \Big(u^3 - 3u^2 z\sqrt{2t} + 3uz^2\cdot 2t - z^3(2t)^{3/2}\Big) q(u)\,du \\ &= m_3 - 3z\sqrt{2t}m_2 + 6tz^2 m - z^3(2t)^{3/2}.\end{aligned}$$

Then the third moment can be evaluated as

$$\begin{aligned}m_3(t) &= \int_{\mathbb{R}} x^3 p_t(x)\,dx = \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} x^3 \int_{\mathbb{R}} q(x+z\sqrt{2t})e^{-z^2}\,dz\,dx \\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} [m_3 - 3z\sqrt{2t}m_2 + 6tz^2 m - z^3(2t)^{3/2}]e^{-z^2}\,dz \\ &= m_3 + 6tm\frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} z^2 e^{-z^2}\,dz = m_3 + 3tm,\end{aligned}$$

where we used that the integral of an odd function is zero. □

It is worth noting that if $q(x)$ has zero mean, $m = 0$, then the third moment is invariant, $m_3(t) = m_3$. If the moment $m > 0$, then the moment $m_3(t)$ is increasing in time, and if $m < 0$, the moment $m_3(t)$ is decreasing in time.

Proposition 4.2. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Then the relation between their skewness functions is*

$$S_t = S\Big(\frac{V}{V+t}\Big)^{3/2}, \forall t > 0.$$

Proof. Using Lemma 5.1, Proposition 3.1 and proposition 2.1 we have

$$\begin{aligned}S_t &= \frac{m_3(t) - 3m_t V_t - m_t^3}{V_t^{3/2}} = \frac{m_3 + 3mt - 3m(V+t) - m^3}{(V+t)^{3/2}} \\ &= \frac{m_3 - 3mV - m^3}{(V+t)^{3/2}} = \frac{m_3 - 3mV - m^3}{V^{3/2}}\Big(\frac{V}{V+t}\Big)^{3/2} \\ &= S\Big(\frac{V}{V+t}\Big)^{3/2}. \quad \square\end{aligned}$$

The previous result has a few important consequences regarding the skewness invariance sign under the heat-flow:

(i) If the probability density $q(x)$ is symmetric, then $p_t(x)$ is symmetric, for all $t > 0$. This follows from the fact that if $S = 0$ then $S_t = 0$.

(ii) If there is $t_0 > 0$ such that $p_{t_0}(x)$ is symmetric, then $q(x)$ is symmetric. The proof is similar with the one done for part (i).

(iii) If the probability density $q(x)$ is skewed to the left (right), then $p_t(x)$ is also skewed to the left (right). This follows from the fact that if $S > 0$ then $S_t > 0$.
(iv) The density $p_t(x)$ becomes more symmetric in the long run. This follows from the fact that S_t is decreasing to zero for t large.

5. THE KURTOSIS

If X is a random variable of mean m and variance σ^2, then the fourth moment of the standardized random variable $Z = \frac{X-m}{\sigma}$ is called the kurtosis of X and represents a measure of the tail thickness for the probability density of X. Also, kurtosis tells the height and sharpness of the central peak, relative to that of a standard bell curve. We have the following kurtosis formula

$$K(X) = \mathbb{E}\Big[\Big(\frac{X-m}{\sigma}\Big)^4\Big] = \frac{1}{\sigma^4}\mathbb{E}[(X-m)^4].$$

Expanding, yields

$$\begin{aligned}\mathbb{E}[(X-m)^4] &= \mathbb{E}[X^4 - 4X^3m + 6X^2m^2 - 4Xm^3 + m^4] \\ &= m_4 - 4mm_3 + 6m^2m_2 - 3m^4.\end{aligned}$$

This shows that the kurtosis of X can be written in terms of the first four moments of X. We shall use a similar formula to define the kurtosis for a probability density. Thus, the kurtosis of $q(x)$ becomes

$$K = K(q) = \frac{m_4 - 4mm_3 + 6m^2m_2 - 3m^4}{V^2},$$

while the kurtosis of $p_t(x)$ is given by a similar formula

$$K_t = K(p_t) = \frac{m_4(t) - 4m_tm_3(t) + 6m_t^2m_2(t) - 3m_t^4}{V_t^2},$$

where the fourth moments of q and p_t are defined respectively by

$$m_4 = \int_{\mathbb{R}} x^4 q(x)\,dx, \qquad m_4(t) = \int_{\mathbb{R}} x^4 p_t(x)\,dx.$$

We shall establish a relation between the fourth moments of $q(x)$ and $p_t(x)$.

Lemma 5.1. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Then the relation between their third moments is*

$$m_4(t) = m_4 + 6tm_2 + 3t^2, \forall t > 0.$$

Proof. We evaluate the following integral using a change of variables

$$\begin{aligned}\int_{\mathbb{R}} x^4 q(x+z\sqrt{2t})\,dx &= \int_{\mathbb{R}} (u-z\sqrt{2t})^4 q(u)\,du\\ &= \int_{\mathbb{R}} [u^4 - 4u^3 z\sqrt{2t} + 6u^2 z^2\cdot 2t - 4uz^3(2t)^{3/2} + z^4(2t)^2]q(u)\,du\\ &= m_4 - 4z\sqrt{2t}m_3 + 12z^2 t m_2 - 4z^3(2t)^{3/2}m + 4z^4t^2.\end{aligned}$$

Then we further compute the fourth moment of p_t as

$$\begin{aligned} m_4(t) &= \int_{\mathbb{R}} x^4 p_t(x)\,dx = \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} x^4 \int_{\mathbb{R}} q(x+z\sqrt{2t})e^{-z^2}\,dz\,dx\\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\Big(\int_{\mathbb{R}} x^4 q(x+z\sqrt{2t})\,dx\Big)e^{-z^2}\,dz\\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}}\Big(m_4 - 4z\sqrt{2t}m_3 + 12z^2tm_2 - 4z^3(2t)^{3/2}m + 4z^4t^2\Big)e^{-z^2}\,dz\\ &= m_4 + 12tm_2\frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} z^2e^{-z^2}\,dz + \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} z^4 e^{-z^2}\,dz\cdot 4t^2\\ &= m_4 + 6tm_2 + 3t^2,\end{aligned}$$

since $\int_{\mathbb{R}} z^2e^{-z^2}\,dz = \frac{\sqrt{\pi}}{2}$ and $\int_{\mathbb{R}} z^4e^{-z^2}\,dz = \frac{3\sqrt{\pi}}{4}$. □

We shall make a few remarks.
(i) Completing the square, we obtain

$$m_4(t) = 3(t+m_2)^2 + (m_4 - 3m_2^2),$$

so the minimum of $m_4(t)$ is realized for the value $t = -m_2$, which is a negative value, since $m_2 > 0$. Consequently, we have the inequality $m_4(t) \ge m_4$, because the function $f(t) = m_4(t)$ is increasing for $t > 0$ and $f(0) = m_4 > 0$.
(ii) The behavior of the 4th moment is different from the second and third moment, since its rate of change is not constant. This follows from $m_2'(t) = 1$, $m_3'(t) = 3m$, and $m_4'(t) = 6(m_2 + t)$. This means the 4th moment changes a lot faster than the lower order moments do.

PROPOSITION 5.2. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Then the relation between their kurtosis functions is*

$$K_t = 3 + (K-3)\Big(\frac{V}{V+t}\Big)^2, \quad \forall t > 0.$$

Proof. Using the kurtosis formula as well as the formulas for $m_2(t)$, $m_3(t)$ and $m_4(t)$, we have

$$K_t = \frac{m_4(t) - 4m_tm_3(t) + 6m_t^2m_2(t) - 3m_t^4}{V_t^2}$$

$$
\begin{aligned}
&= \frac{m_4 + 6tm_2 + 3t^2 - 4m(m_3 + 3mt) + 6m^2(m_2 + t) - 3m^4}{(V+t)^2} \\
&= \frac{(m_4 - 4mm_3 + 6m^2 m_2 - 3m^4) + 3t^2 + 6t(m_2 - m^2)}{(V+t)^2} \\
&= \frac{V^2K + 3t^2 + 6tV}{(V+t)^2}.
\end{aligned}
$$

Completing the square

$$V^2K + 3t^2 + 6tV = 3(V+t)^2 + V^2(K-3),$$

the previous relation becomes

$$K_t = \frac{3(V+t)^2 + V^2(K-3)}{(V+t)^2} = 3 + (K-3)\Big(\frac{V}{V+t}\Big)^2. \quad \square$$

Proposition 5.2 has a few interesting consequences.
(i) If the probability density $q(x)$ is a standard normal density, then $p_t(x)$ has $K_t = 3$. This follows from the fact that a standard normal density has kurtosis $K = 3$.
(ii) If the probability density $q(x)$ has the tails thicker (thinner) than a standard normal distribution, then $p_t(x)$ has the same property. This follows from the following analysis:

(a) If $K < 3$ then $K < K_t$, for all $t > 0$; the kurtosis increases asymptotically to 3;

(b) If $K = 3$ then $K = K_t$, for all $t > 0$; the kurtosis is constant equal to 3;

(c) If $K > 3$ then $K > K_t$, for all $t > 0$; the kurtosis decreases asymptotically to 3.

In all cases, the kurtosis of $p_t(x)$ tends to 3 in the long run, $\lim_{t\to\infty} K_t = 3$.
(iii) We have the following relations between variance, skewness and kurtosis

$$\frac{K_t - 3}{K - 3} = \Big(\frac{V}{V+t}\Big)^2 = \Big(\frac{S_t}{S}\Big)^{4/3},$$

provided $K \neq 3$ and $S \neq 0$.
(iv) The difference between the kurtosis and 3 is called excess kurtosis, $K - 3$. For normal distributions this is equal to zero. The proposition shows that the excess kurtosis of p_t is equal to the the excess kurtosis of q multiplied by the factor $\left(\frac{V}{V+t}\right)^2$. In the long run the excess kurtosis of p_t, i.e. $K_t - 3$, tends to zero.

6. NORMALITY TEST

A normality test is needed in statistics to determine whether a data set is modeled by a normal distribution. The most well-known normality tests are Kolmogorov-Smirnov [2], [4], and Shapiro-Wilk [3].

Two numerical measures of shape, skewness and kurtosis, can be also used to test for normality. If skewness is not close to zero, then the data set is not normally distributed. Also, if kurtosis is not close to 3 the data is also not normally distributed.

We shall assume, for the beginning, that the initial probability density $q(x)$ is normal, with $q \sim \mathcal{N}(m, \sigma^2)$. This implies $V = \sigma^2$, $S = 0$ and $K = 3$. Consider the heat-flow evolution of $q(x)$ at time t, which is $p_t(x)$. Then, according to the results proved in the previous sections, the new distribution will have $m_t = m$, $V_t = V + t$, $S_t = 0$ and $K_t = 3$. Therefore, the test for normality is satisfied. Actually, in this case (when $q(x)$ is normal) it can be shown that $p_t(x)$ is actually a normal density function, with $p_t \sim \mathcal{N}(m, \sigma^2 + t)$. The proof is based on the fact that the convolution of two Gaussian densities is also a Gaussian density. More precisely, the result states that $G_{\sigma_1} * G_{\sigma_2} = G_\sigma$, with $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. In our case $q = G_\sigma = G_{\sigma_1}$ and $p_t = G_{\sqrt{t}} = G_{\sigma_2}$.

In general, if $q(x)$ is not a normal density, then p_t is not normal either, but for t large it gets very close to a normal density, in the sense that the skewness and excess kurtosis can be made arbitrary small.

PROPOSITION 6.1. *Let $q(x)$ be a continuous probability density on $\mathbb{R}$ and $p_t(x)$ be the solution of the system (1)-(2). Denote by S_t and K_t the skewness and kurtosis of p_t.*
Then for any $\epsilon_1 > 0$, $\epsilon_2 > 0$, there is $T > 0$ such that

$$|S_t| < \epsilon_1, \qquad |K_t - 3| < \epsilon_2, \qquad \forall t > T.$$

Proof. Denote $\rho_t = \dfrac{V}{V+t}$. We note that $0 < \rho_t < 1$. Using that

$$\begin{aligned} S_t &= S\rho_t^{3/2} \\ K_t - 3 &= (K-3)\rho_t^2, \end{aligned}$$

it suffices to find a $T > 0$ such that

$$\begin{aligned} |S|\rho_t^{3/2} &< \epsilon_1 \\ |K-3|\rho_t^2 &< \epsilon_2, \qquad \forall t > T. \end{aligned}$$

This is equivalent to

$$\rho_t < \left(\frac{\epsilon_1}{|S|}\right)^{2/3}$$

$$\rho_t \quad < \quad \Big(\frac{\epsilon_2}{|K-3|}\Big)^{1/2}, \qquad \forall t > T.$$

Let $\epsilon = \min\Big\{\Big(\frac{\epsilon_1}{|S|}\Big)^{2/3}, \Big(\frac{\epsilon_2}{|K-3|}\Big)^{1/2}\Big\}$. It suffices to find a $T > 0$ such that

$$\rho_t < \epsilon, \qquad \forall t > T.$$

Using the definition of ρ_t, the previous inequality is satisfied for any $t > V(\frac{1}{\epsilon} - 1)$. Hence, we may take

$$T = \max\Big\{0, V(\frac{1}{\epsilon} - 1)\Big\}. \qquad \square$$

The previous result has a downside. The heat-flow deformation of $q(x)$ can make the density to have skewness and excess kurtosis as close to zero as possible. However, this is done at the expense of the variance, which increases linearly in t. For $t > T$ we obtain the following lower bound of the variance

$$V_t = V + t > V + T > V + V(\frac{1}{\epsilon} - 1) = \frac{V}{\epsilon}.$$

Hence, the variance explodes as ϵ tends to zero. An upper bound constraint on the variance V_t will forbid ϵ to get too small. Therefore, there is a trade-off between the variance size and order of smallness for the skewness and excess kurtosis.

7. MULTIVARIATE CASE

Some of the previous results do extend to the case of several variables. In this case $q(\mathbf{x})$ denotes a probability density function on $\mathbb{R}^n$, with $\mathbf{x} = (x_1, \cdots, x_n)$. The heat-flow evolution equation becomes

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) \quad = \quad \Delta_{\mathbf{x}} p_t(\mathbf{x}) \tag{4}$$

$$p_0(\mathbf{x}) \quad = \quad q(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathbb{R}^n,\ t > 0, \tag{5}$$

where

$$\Delta_{\mathbf{x}} = \frac{1}{2}\Big(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\Big)$$

is the n-dimensional Laplacian on $\mathbb{R}^n$. The solution of the Cauchy problem (4)-(5) is given by the convolution

$$p_t(\mathbf{x}) = (q * G)(\mathbf{x}),$$

where the multivariate Gaussian is given by

$$G(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi t)^{n/2}} e^{-\frac{1}{2t}\|\mathbf{x}-\mathbf{y}\|^2} = \frac{1}{(2\pi t)^{n/2}} e^{-\frac{1}{2t}\sum_{i=1}^{n}(x_i-y_i)^2},$$

with $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$. A computation similar with the one done for the one-dimensional case shows

$$\begin{aligned} p_t(\mathbf{x}) &= \int_{\mathbb{R}^n} q(\mathbf{y}) G(\mathbf{x},\mathbf{y})\, d\mathbf{y} = \int_{\mathbb{R}^n} q(\mathbf{y}) \frac{1}{(2\pi t)^{n/2}} e^{-\frac{1}{2t}\|\mathbf{x}-\mathbf{y}\|^2}\, d\mathbf{y} \\ &= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} q(\mathbf{x} + \mathbf{z}\sqrt{2t}) e^{-\|\mathbf{z}\|^2}\, d\mathbf{z}. \end{aligned}$$

We assume that $q(\mathbf{x})$ is the probability density of the random variable $X = (X_1, \cdots, X_n)$. The mean of q is the vector $\mu = (\mu_1, \cdots, \mu_n)$, with $\mu_j = \mathbb{E}[X_j] = \int_{\mathbb{R}^n} x_j q(\mathbf{x})\, d\mathbf{x}$. Similarly, the mean of the density p_t is the vector $\mu(t) = (\mu_1(t), \cdots, \mu_n(t))$, with $\mu_j(t) = \int_{\mathbb{R}^n} x_j p_t(\mathbf{x})\, d\mathbf{x}$.

Proposition 7.1. *Probability densities $p_t(\mathbf{x})$ and $q(\mathbf{x})$ have the same mean, $\mu(t) = \mu$.*

Proof. This is an adaptation of the proof of Proposition 2.1 to several variables:

$$\begin{aligned} \mu_j(t) &= \int_{\mathbb{R}^n} x_j p_t(\mathbf{x})\, d\mathbf{x} = \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} x_j \int_{\mathbb{R}^n} q(\mathbf{x} + \mathbf{z}\sqrt{2t}) e^{-\|z\|^2}\, d\mathbf{z}\ \ d\mathbf{x} \\ &= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} x_j q(\mathbf{x} + \mathbf{z}\sqrt{2t})\, d\mathbf{x}\ e^{-\|\mathbf{z}\|^2}\, d\mathbf{z} \\ &= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (u_j - z_j\sqrt{2t}) q(\mathbf{u})\, d\mathbf{u}\ e^{-\|\mathbf{z}\|^2}\, d\mathbf{z} \\ &= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} \Big[\underbrace{\int_{\mathbb{R}^n} u_j q(\mathbf{u})\, d\mathbf{u}}_{=\mu_j} - z_j\sqrt{2t} \underbrace{\int_{\mathbb{R}^n} q(\mathbf{u})\, d\mathbf{u}}_{=1} \Big]\ e^{-\|\mathbf{z}\|^2}\, d\mathbf{z} \\ &= \frac{1}{\pi^{n/2}} \Big[\mu_j \underbrace{\int_{\mathbb{R}^n} e^{-\|\mathbf{z}\|^2}\, d\mathbf{z}}_{=\pi^{n/2}} - \sqrt{2t} \underbrace{\int_{\mathbb{R}^n} z_j e^{-\|\mathbf{z}\|^2}\, d\mathbf{z}}_{=0} \Big] = \mu_j. \quad \square \end{aligned}$$

In the following we shall discuss the covariance matrix. Consider the n-dimensional continuous stochastic process $X_t = (X_1(t), \cdots, X_n(t))$, with $t \geq 0$. Let $X_0 = X = (X_1, \cdots, X_n)$ be the initial state of the process. The probability density function of X is denoted by $q(\mathbf{x})$. We assume the probability density of X_t is denoted by p_t and it satisfies the heat-flow equation (4).

The question of concern now is: what is the relation between the covariance matrices $Cov(X_i, X_j)$ and $Cov(X_i(t), X_j(t))$? For the definition and properties of the covariance the reader can consult Wackerly et. al [5].

PROPOSITION 7.2. *The covariance matrix of X_t modifies only along the main diagonal*

$$Cov(X_i(t), X_j(t)) = Cov(X_i, X_j) + t\delta_{ij}, \forall t > 0.$$

Proof. We start by recalling the formula definitions of covariance

$$\begin{aligned} Cov(X_i, X_j) &= \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \\ &= \int_{\mathbb{R}^n} x_ix_jq(\mathbf{x})\,dx - \int_{\mathbb{R}^n} x_iq(\mathbf{x})\,dx \int_{\mathbb{R}^n} x_jq(\mathbf{x})\,dx \\ &= \int_{\mathbb{R}^n} x_ix_jq(\mathbf{x})\,dx - \mu_i\mu_j, \end{aligned}$$

$$\begin{aligned} Cov(X_i(t), X_j(t)) &= \mathbb{E}[X_i(t)X_j(t)] - \mathbb{E}[X_i(t)]\mathbb{E}[X_j(t)] \\ &= \int_{\mathbb{R}^n} x_ix_jp_t(\mathbf{x})\,d\mathbf{x} - \int_{\mathbb{R}^n} x_ip_t(\mathbf{x})\,d\mathbf{x} \int_{\mathbb{R}^n} x_jp_t(\mathbf{x})\,d\mathbf{x} \\ &= \int_{\mathbb{R}^n} x_ix_jp_t(\mathbf{x})\,d\mathbf{x} - \mu_i(t)\mu_j(t) \\ &= \int_{\mathbb{R}^n} x_ix_jp_t(\mathbf{x})\,d\mathbf{x} - \mu_i\mu_j. \end{aligned}$$

In order to evaluate $\int_{\mathbb{R}^n} x_ix_jp_t(\mathbf{x})\,d\mathbf{x}$, we first compute the integral $I = \displaystyle\int_{\mathbb{R}^n} x_ix_j\, q(\mathbf{x} + \mathbf{z}\sqrt{2t})\,d\mathbf{x}$:

$$\begin{aligned} I &= \int_{\mathbb{R}^n} (u_i - z_i\sqrt{2t})(u_j - z_j\sqrt{2t})q(\mathbf{u})\,d\mathbf{u} \\ &= \int_{\mathbb{R}^n} (u_iu_j - z_iu_j\sqrt{2t} - z_ju_i\sqrt{2t} + 2tz_iz_j)q(\mathbf{u})\,d\mathbf{u} \\ &= \int_{\mathbb{R}^n} u_iu_jq(\mathbf{u})\,d\mathbf{u} - \sqrt{2t}(z_j\mu_i + z_i\mu_j) + 2tz_iz_j. \end{aligned}$$

We further evaluate using Fubini

$$\begin{aligned}
\int_{\mathbb{R}^n} x_i x_j p_t(\mathbf{x})\,d\mathbf{x} &= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} x_i x_j q(\mathbf{x}+\mathbf{z}\sqrt{2t}) e^{-\|\mathbf{z}\|^2}\,d\mathbf{z}\,d\mathbf{x}\\
&= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} x_i x_j q(\mathbf{x}+\mathbf{z}\sqrt{2t})\,d\mathbf{x}\right] e^{-\|\mathbf{z}\|^2}\,d\mathbf{z}\\
&= \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} u_i u_j q(\mathbf{u})\,d\mathbf{u} - \sqrt{2t}(z_j\mu_i + z_i\mu_j) + 2t z_i z_j\right] e^{-\|\mathbf{z}\|^2}\,d\mathbf{z}\\
&= \int_{\mathbb{R}^n} u_i u_j q(\mathbf{u})\,d\mathbf{u} + \frac{2t}{\pi^{n/2}} \int_{\mathbb{R}^n} z_i z_j e^{-\|\mathbf{z}\|^2}\,d\mathbf{z},
\end{aligned}$$

because $\int_{\mathbb{R}^n} e^{-\|\mathbf{z}\|^2}\,d\mathbf{z} = \pi^{n/2}$, $\int_{\mathbb{R}^n} z_j e^{-\|\mathbf{z}\|^2}\,d\mathbf{z} = 0$.

We shall compute next the integral $J_{ij} = \int_{\mathbb{R}^n} z_i z_j e^{-\|\mathbf{z}\|^2}\,d\mathbf{z}$. We distinguish two cases:

(1) Case $i \neq j$: In this case the integral vanishes. We assume for simplicity that $i = 1$ and $j = 2$. Applying Fubini's theorem we obtain

$$J_{12} = \int_{\mathbb{R}} z_1 e^{-z_1^2}\,dz_1 \int_{\mathbb{R}} z_2 e^{-z_2^2}\,dz_2 \int_{\mathbb{R}^{n-2}} e^{-(z_3^2+\cdots+z_n^2)}\,dz_3\cdots dz_n = 0,$$

since the first two integrals vanish.

(2) Case $i = j$: We assume for simplicity that $i = j = 1$.

$$J_{11} = \int_{\mathbb{R}} z_1^2 e^{-z_1^2}\,dz_1 \int_{\mathbb{R}^{n-1}} e^{-(z_2^2+\cdots+z_n^2)}\,dz_3\cdots dz_n = \frac{\sqrt{\pi}}{2}(\sqrt{\pi})^{n-1} = \frac{1}{2}\pi^{n/2}.$$

since the first two integrals vanish. Substituting into (6) we obtain

$$\int_{\mathbb{R}^n} x_i x_j p_t(\mathbf{x})\,d\mathbf{x} = \begin{cases} \int_{\mathbb{R}^n} u_i u_j q(\mathbf{u})\,d\mathbf{u}, & \text{if } i \neq j \\ \int_{\mathbb{R}^n} u_i u_j q(\mathbf{u})\,d\mathbf{u} + t, & \text{if } i = j \end{cases}$$

$\square$

The previous result has applications to Principal Component Analysis (PCA). According to PCA, the information stored in the n-dimensional random variable X can be analysed on principal components, which are the eigenvectors of the covariance matrix $Cov(X_i, X_j)$ (which being a symmetric matrix has real eigenvalues). The corresponding eigenvalues represent the information weight contained within each principal component.

Proposition 7.3. *Let X_t be the stochastic process obtained by applying the heat-flow to the random variable X. Then each value t, for which the covariance matrix $Cov(X_i(t), X_j(t))$ becomes singular, represents the negative of an eigenvalue $\lambda = -t$ of matrix $Cov(X_i, X_j)$.*

Proof. Applying Proposition 7.1, we have

$$\det Cov(X_i(t), X_j(t)) = \det(Cov(X_i, X_j) + t\delta_{ij}) = \det(Cov(X_i, X_j) - \lambda\delta_{ij}).$$

Hence, $\det Cov(X_i(t), X_j(t)) = 0$ if and only if $\lambda = -t$ is an eigenvalue of $Cov(X_i, X_j)$. $\square$

It is worth noting that if the components of X are independent random variables, then $Cov(X_i, X_j) = Var(X_i)\delta_{ij}$ is a diagonal matrix with only positive eigenvalues. Therefore, in this case, $\det Cov(X_i(t), X_j(t)) \neq 0$ for all $t > 0$.

REFERENCES

[1] R. B. Guenther and J. W. Lee, *Partial Differential equations of Mathematical Physics and Integral Equations*. Pretince-Hall, London, 1988.

[2] A. Kolmogorov, *Sulla determinazione empirica di una legge di distribuzione*, G. Ist. Ital. Attuari. **4** (1933), 83–91

[3] S. S. Shapiro and M. B. Wilk, *An analysis of variance test for normality (complete samples)*. Biometrika **52** (1965), 591–611.

[4] N. Smirnov, *Table for estimating the goodness of fit of empirical distributions*. Annals of Mathematical Statistics **19** (1948), 279–281.

[5] D. Wackerly, W. Meddenhall and R. Scheaffer, *Mathematical Statistics with Applications*, Brooks/Cole Cengage Learning, 7th ed., 2008.

Received January 15, 2020

Eastern Michigan University
Department of Mathematics and Statistics
Ypsilanti, MI, 48197, U.S.A.
ocalin@emich.edu