Three remarkable properties of the
Normal distribution for sample variance

Eric Benhamou[1], Beatrice Guez[2] and Nicolas Paris[3]

Abstract

In this paper, we present three remarkable properties of the normal distribution: first that if two independent variables 's sum is normally distributed, then each random variable follows a normal distribution (which is referred to as the Levy Cramer theorem), second a variation of the Levy Cramer theorem (new to our knowledge) that states that two independent symmetric random variables with finite variance, independent sum and difference are necessarily normal, and third that normal distribution can be characterized by the fact that it is the only distribution for which sample mean and variance are independent, which is a central property for deriving the Student distribution and referred as the Geary theorem. The novelty of this paper is twofold. First we provide an extension of the Levy Cramer theorem. Second, for the two seminal theorem (the Levy Cramer and Geary theorem), we provide new, quicker or self contained proofs.

[1] A.I. Square Connect, France. E-mail: eric.benhamou@aisquareconnect.com

LAMSADE, Universit Paris Dauphine, France. E-mail: eric.benhamou@dauphine.eu

[2] A.I. Square Connect, France. E-mail: beatrice.guez@aisquareconnect.com

[3] A.I. Square Connect, France.

E-mail: nicolas.paris@aisquareconnect.com

1   Introduction

The normal distribution is central to probability and statistics. It is obviously the asymptotic law for the sum of i.i.d. variables as stated by the Central Limit theorem and its different variants and extensions. It is also well know that the normal distribution is the only Levy α-stable distribution with the largest value of the α coefficient, namely 2. This fundamental properties explains why the Wiener process plays such a central role in stochastic calculus. There are many more properties of the normal distribution. It is for instance the continuous distribution with given mean and variance that has the maximum entropy.

In this paper, we present three remarkable properties of the normal distribution: first that if X_1 and X_2 are two independent variables with $X_1 + X_2$ normally distributed, then both X_1 and X_2 follow a normal distribution (which is referred to the Levy Cramer theorem), second a variation of the Levy Cramer theorem that states that if X_1 and X_2 are independent symmetric random variables with finite variance with $X_1 + X_2$ and $X_1 - X_2$ independent, then both X_1 and X_2 are normal, and third that the normal distribution is characterized by the fact that it is the only distribution for which the sample mean and variance are independent, which is a central property for deriving the Student distribution and referred as the Geary theorem (see [2]).

The novelty of this paper is to provide new, quicker or self contained proofs of theses theorems as well as an extension of the Levy Cramer theorem. In particular, we revisit the proof first provided by [3] to give full details of it as the sketch of the proof provided was rather elusive on technical details in complex analysis. We also provide a quicker proof of the Geary theorem using the log caracteristic function instead of the caracteristic function originally used.

The paper is organized as follows. We first present the Levy Cramer theorem. We provide a thorough proof of the elusive sketch provided by [3]. We then present another remarkable result that combines independence and normal law. We finally present a quick proof of the fact that the independence of the sample mean and variance is a characterization of the normal law.

2 Levy Cramer theorem

The Levy Cramer theorem states that if the sum of two independent non-constant random variables X_1 and X_2 is normally distributed, then each of the summands (X_1 or X_2) is normally distributed. This result stated [4] and proved by [1] admits various equivalent formulations since there is a one to one mapping between variables and their characteristic function:

- if the convolution of two proper distributions is a normal distribution, then each of them is a normal distribution.

- if $\phi_1(t)$ and $\phi_2(t)$ are characteristic functions and if $\phi_1(t)\phi_2(t) = \exp(-\gamma t^2 + i\beta t)$, with $\gamma \geq 0$ and $-\infty < \beta < \infty$, then $\phi_j(t) = \exp(-\gamma_j t^2 + i\beta_j t)$ with $\gamma_j \geq 0$ and $-\infty < \beta_j < \infty$.

The two reformulations are obvious as the first one just uses the fact that the distribution of the sum is the convolution of the two distributions. The second uses the one to one mapping between distribution and characteristic functions. It is worth mentioning that the LvyCramr theorem can be generalized to the convolution of two signed measures with restrictions on their negative variation. It has also as a consequence the following implication on the stability of distributions. Closeness of the distribution of a sum of independent random variables to the normal distribution implies closeness of the distribution of each of the summands to the normal distribution, qualitative estimates of the stability are known. The LvyCramr theorem can also be reformulated in terms of the Poisson distribution and is named the Raikov's theorem. It can also be extended to the convolution of a Poisson and a normal distribution and to other classes of infinitely-divisible distributions [5]. Mathematically, the Levy Cramer theorem writes as follows:

Theorem 2.1. *If X_1 and X_2 are two independent variables with $X_1 + X_2$ normally distributed, then both X_1 and X_2 follow a normal distribution*

We will provide in the following page a self contained proof that follows the version of [3] but provide full details about the assumptions on complex analysis. For this, we will need 4 different lemmas that are presented now.

2.1 Lemma 1

Lemma 2.2. *Let f be an entire function without any zero (that does not vanish anywhere), then there exists an entire function g such that*

$$f = \exp(g)$$

Proof. This results makes sense as intuitively, we would like to use as a candidate $\log(f)$. However, let us provide a rigorous proof of it and explicitly construct the function g. The exponential function is surjective from $\mathbb{C}$ to $\mathbb{C}^*$. This implies that there exists a complex z_0 such that $f(0) = exp(z_0)$. The ratio of two holomorphic functions with the denominator function without any zero being holomorphic and any non vanishing holomorphic function f admitting an entire function as a primitive, the function $\frac{f'}{f}$ defined as the ratio of two non vanishing holomorphic functions admits an entire function as primitive denoted by

$$g(z) = z_0 + \int_{[0,z]} \frac{f'(u)}{f(u)} du$$

Let us define by $v = \exp(g)$. Trivially, we have $v' = g' \exp(g) = \frac{f'}{f} v$ so that $(v/f)' = \frac{v'f - vf'}{f^2} = 0$, which implies that the ratio $\frac{v}{f}$ is constant and equal to one as by construction, $v(0) = f(0)$. This concludes the proof. $\qquad \square$

Let us also prove another elementary result of complex analysis with the following lemma (which is an extension of the Liouville's theorem). Let us denote by $\mathcal{R}$ the radius of convergence of a function f.

2.2 Lemma 2

Lemma 2.3. *Let f be an entire function and P a polynomial such that:*

$$\forall z \in \mathbb{C}, \mathcal{R}(f(z)) \le |P(z)|,$$

then f is polynomial of degree at most equal to that of P.

Proof. By definition, the entire function writes as a power series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

that converges everywhere in the complex plane (hence uniformly on compact sets). Cauchy formula states that $\forall n \geq 1, \forall r > 0$, we can recover the series coefficients

$$a_n r^n = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) e^{-in\theta} d\theta, \tag{1}$$

as well as

$$0 = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) e^{in\theta} d\theta \tag{2}$$

and

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) d\theta. \tag{3}$$

Summing up equation (1) and the complex conjugate of (2) gives that

$$a_n r^n = \frac{1}{\pi} \int_0^{2\pi} \mathcal{R}(f(re^{i\theta})) e^{-in\theta} d\theta, \tag{4}$$

In particular, combining equations (4) and (3), we have

$$|a_n| r^n + 2|a_0| \leq \frac{1}{\pi} \int_0^{2\pi} |\mathcal{R}(f(re^{i\theta}))| + |f(re^{i\theta})| d\theta \leq 4 \max_{|z|=r}(\mathcal{R}(f(z)), 0). \tag{5}$$

Let us denote the polynomial function $P = \sum_{k=0}^d b_k X^k$. The hypothesis writes $\mathcal{R}(f(z)) \leq |\sum_{k=0}^d b_k X^k| \quad \forall z \in \mathbb{C}$, which combined with inequality (5) leads to:

$$|a_n| r^n + 2|a_0| \leq 4 \max_{|z|=r}(|\sum_{k=0}^d b_k X^k|, 0) \tag{6}$$

The dominating term in the RHS of the inequality (6), for r large, is $|b_d| r^d$. For inequality (6) to hold, all LHS terms of order $n > d$ should therefore be equal to zero: $\forall n > d, a_n = 0$. $\qquad \square$

Let us now prove a central lemma.

2.3 Lemma 3

Lemma 2.4. *If X is a real random variable such that there exists $\eta > 0$ such that $f(\eta) = \int_{\mathbb{R}} e^{\eta^2 x^2} d\mathbb{P}_X(x)$ is finite and if the characteristic function has no zero in the complex domain, then X follows a normal distribution*

Remark 2.1. *This is a remarkable property of the normal distribution. It states that the normal is the only distribution whose characteristic function has no zero such that its transform with the exploding kernel $e^{\eta^2 x^2}$ is finite. The condition of no zero is necessary as for instance the uniform distribution whose characteristic function is cosinus validates the fact that the function $f(\eta)$ is finite but not the no zero condition on the characteristic function. The characterization of the normal distribution is related to the fact that the normal distribution is the only distribution that is Lvy α-stable stable with $\alpha = 2$. Another interesting remark is that this property of the normal distribution is closely related to the amazing features of the holomorphic functions as well as the connection between characteristic function and moments.*

Proof. For $\eta \in \mathbb{R}$ not null, for $\zeta \in \mathbb{C}$ and for $x \in \mathbb{R}$, we have $|\zeta x| = |\eta x\ \eta^{-1}\zeta| \leq \frac{1}{2}(\eta^2 x^2 + \eta^{-2}|\zeta|^2) \leq \eta^2 x^2 + \eta^{-2}|\zeta|^2$ (which is a reformulation of the remarkable identity $2ab \leq a^2 + b^2$). We can upper bound the integral of $\int_{\mathbb{R}} |e^{i\zeta x}| d\mathbb{P}_X(x)$ as follows:

$$\int_{\mathbb{R}} |e^{i\zeta x}| d\mathbb{P}_X(x) \leq \int_{\mathbb{R}} e^{|\zeta x|} d\mathbb{P}_X(x) \leq \int_{\mathbb{R}} e^{\eta^2 x^2 + \eta^{-2}|\zeta|^2} d\mathbb{P}_X(x) \tag{7}$$

The assumption says that for a given $\eta > 0$, $\int_{\mathbb{R}} e^{\eta^2 x^2} d\mathbb{P}_X(x) < \infty$. Hence, for $\zeta \in \mathbb{C}$, the characteristic function of X whose definition in the complex domain is $\phi_X(\zeta) = \int_{\mathbb{R}} e^{i\zeta x} d\mathbb{P}_X(x)$ is bounded, hence defined since we have

$$|\phi_X(\zeta)| \ \leq \ \int_{\mathbb{R}} |e^{i\zeta x}| d\mathbb{P}_X(x) \leq e^{\eta^{-2}|\zeta|^2} \int_{\mathbb{R}} e^{\eta^2 x^2} d\mathbb{P}_X(x) < \infty \tag{8}$$

In particular, inequality (8) provides the dominated convergence necessary to apply the holomorphic theorem under the integral sign to prove that the characteristic function, $\phi_X(\zeta)$ is an entire function. By assumption, it does not have any zero in the complex domain. We can safely apply lemma 2.2. So there exists an entire function g such that $\phi_X(\zeta) = \exp(g(\zeta))$. Inequality (8) states that

$$\exp(g(\zeta)) \leq e^{\eta^{-2}|\zeta|^2} f(\eta) \tag{9}$$

or equivalently

$$\mathcal{R}(g(\zeta)) \leq \eta^{-2}|\zeta|^2 + \ln(f(\eta)) \tag{10}$$

Lemma 2.3 then says that g is polynomial and at most quadratic. Hence it writes as

$$g(\zeta) = a_0 + ia_1\zeta + \frac{a_2}{2}\zeta^2 \tag{11}$$

In addition, since $\phi_X(0) = 1$, we have $a_0 = 0$. Using the fact that from the characteristic function, we can recover moments with $\mathbb{E}[X^k] = (-i)^k \phi_X^{(k)}(0)$, we get immediately that $a_1 = \mathbb{E}[X] = \mu$ and $a_2 = -Var(X) = -\sigma^2$, which leads to the fact that the characteristic function of X is of the form

$$\phi_X(\zeta) = \exp(i\mu\zeta - \frac{1}{2}\sigma^2\zeta^2)$$

which is the characteristic function of the normal. This concludes the proof.

$\square$

Let us prove a final lemma that reformulates the function $f(\eta)$.

2.4   Lemma 4

Lemma 2.5. *The function $f(\eta)$ can be computed in terms of the probability of the absolute value of X, $\mathbb{P}(|X| \geq x)$, as follows:*

$$f(\eta) = 1 + \int_0^{+\infty} 2x\eta^2 e^{\eta^2 x^2} \mathbb{P}(|X| \geq x)dx \tag{12}$$

Proof. Let us split the integral between negative and positive reals as follows

$$f(\eta) \;=\; \int_{-\infty}^0 e^{\eta^2 x^2} d\mathbb{P}_X(x) + \int_0^{+\infty} e^{\eta^2 x^2} d\mathbb{P}_X(x) \tag{13}$$

An integration by parts for the first integral, with the two primitives functions chosen judiciously as $e^{\eta^2 x^2}$ and $\mathbb{P}(X \leq x)$, leads to

$$\int_{-\infty}^0 e^{\eta^2 x^2} d\mathbb{P}_X(x) \;=\; \mathbb{P}(X \leq 0) - \int_{-\infty}^0 2x\eta^2 e^{\eta^2 x^2} \mathbb{P}(X \leq x)dx \tag{14}$$

$$\;=\; \mathbb{P}(X \leq 0) + \int_0^{\infty} 2x\eta^2 e^{\eta^2 x^2} \mathbb{P}(X \leq -x)dx \tag{15}$$

Similarly, for the second integral in equation (13), an integration by parts, with the two primitives functions chosen as still $e^{\eta^2 x^2}$ but $-\mathbb{P}(X \geq x)$, leads to

$$\int_0^{+\infty} e^{\eta^2 x^2} d\mathbb{P}_X(x) \;=\; \mathbb{P}(X \geq 0) + \int_0^{+\infty} 2x\eta^2 e^{\eta^2 x^2} \mathbb{P}(X \geq x)dx \tag{16}$$

Combining equations (15) and (16) gives the result. $\square$

We are now able to prove the Levy Cramer theorem that states that if X_1 and X_2 are two independent variables with $X_1 + X_2$ normally distributed, then both X_1 and X_2 follow a normal distribution.

Proof. A first trivial case is when the variance of $X_1 + X_2$ is equal to zero. This means that each of two variables X_1 and X_2 has a null variance (since they are independent) and the result is trivial as they are all constant variables. We therefore assume that the variance of $X_1 + X_2$ is not zero. It is worth noticing that since $X_1 + X_2$ is a continuous distribution without any atom, we cannot have that both X_1 and X_2 have an atoms since if it was the case with these two atoms given by a_1 and $a2$, we would have

$$0 < \mathbb{P}(X_1 = a_1)\mathbb{P}(X_2 = a_2) = \mathbb{P}(X_1 = a_1, X_2 = a_2) \leq \mathbb{P}(X_1 + X_2 = a_1 + a_2) = 0$$

which would result in a contradiction! Hence, we can safely assume that X_2 has no atom[4] and find μ that splits the distribution of X_2 equally: $\mathbb{P}(X_2 \leq \mu) = \mathbb{P}(X_2 \geq \mu) = \frac{1}{2}$. Hence, for $x > 0$, we have

$$
\begin{align}
\mathbb{P}(|X_1| \geq x) &= 2\mathbb{P}(X_1 \geq x, X_2 \geq \mu) + 2\mathbb{P}(X_1 \leq -x, X_2 \leq \mu) \tag{17} \\
&\leq 2\mathbb{P}(X_1 + X_2 \geq x + \mu) + 2\mathbb{P}(X_1 + X_2 \leq -x + \mu) \tag{18} \\
&\leq 2\mathbb{P}(|X_1 + X_2 - \mu| \geq x) \tag{19}
\end{align}
$$

Lemma 2.5 states that

$$f_{X_1}(\eta) = 1 + \int_0^{+\infty} 2x\eta^2 e^{\eta^2 x^2} \mathbb{P}(|X_1| \geq x)dx \tag{20}$$

Combining these two results gives us a way to upper bound $f_{X_1}(\eta)$:

$$
\begin{align}
|f_{X_1}(\eta)| &\leq 2 + 2\int_0^{+\infty} 2x\eta^2 e^{\eta^2 x^2} \mathbb{P}(|X_1 + X_2 - \mu| \geq x)dx \tag{21} \\
&\leq 2f_{X_1 + X_2 - \mu}(\eta) \tag{22}
\end{align}
$$

The latter inequality is finite as $X_1 + X_2 - \mu$ is normal so that $|f_{X_1}(\eta)| \leq \infty$. In addition, since we have $\phi_{X_1}\phi_{X_2} = \phi_{X_1 + X_2}$ by independence, the characteristic function of X_1 is never equal to zero in the complex domain since $\phi_{X_1 + X_2}$ is never null in $\mathbb{C}$. Using lemma 2.4, we conclude that X_1 is normal. If X_1 is of null variance, then $X_2 = X_2 + X_1 - X_1$ is a normal as the sum of a normal and a constant. Otherwise, X_1's variance is strictly positive and X_1 has no atom. Interchanging X_1 and X_2 and using previous reasoning, we get immediately that X_2 follows also a normal distribution, which concludes the proof.   □

---

[4]if it was not the case, we can interchange X_1 and X_2 to make this happen

3 An extension of the Levy Cramer theorem

It is a well known and easy to prove result (using for instance characteristic function) that if X_1 and X_2 are independent normal standard variables, then $X_1 + X_2$ and $X_1 - X_2$ are independent. But in fact, we can prove somehow the opposite that requires less assumption than the Levy Cramer theorem. This is the following result

Theorem 3.1. *If X_1 and X_2 are two independent symmetric variables with mean 0 and variance 1, such that $X_1 + X_2$ and $X_1 - X_2$ are independent, then X_1 and X_2 follow a normal distribution*

Remark 3.1. *Compared to the Levy Cramer theorem, we do not assume in the hypothesis that any of the variable is normal which makes the result quite remarkable.*

Proof. Noticing that $2X_1 = (X_1 + X_2) + (X_1 - X_2)$, we have using characteristic function and independence that

$$\phi_{2X_1}(t) = \phi_{X_1+X_2}(t)\phi_{X_1-X_2}(t)$$

Since X_1 and X_2 are two symmetric independent variables, we have that $\phi_{X_1+X_2}(t) = \phi_{X_1}(t)^2$ and likewise $\phi_{X_1-X_2}(t) = \phi_{X_1}(t)\phi_{\bar{X}_1}(t)$

We also have that $\phi_{2X_1}(t) = \phi_{X_1}(2t)$ and $\phi_{\bar{X}_1}(t) = \overline{\phi_{X_1}}(t)$.

Combining all results, we get

$$\phi_{X_1}(2t) = \phi_{X_1}(t)^3 \overline{\phi_{X_1}}(t) \tag{23}$$

The convex conjugate of previous equation writes $\overline{\phi_{X_1}}(2t) = \overline{\phi_{X_1}}(t)^3 \phi_{X_1}(t)$, which leads to

$$\phi_{X_1}(2t)\overline{\phi_{X_1}}(2t) = \phi_{X_1}(t)^4 \overline{\phi_{X_1}}(t)^4 \tag{24}$$

Denoting by $\psi(t) = \phi_{X_1}(t)\overline{\phi_{X_1}}(t)$, equation (24) states that $\psi(2t) = \psi(t)^4$ or iteratively

$$\psi(t) = \psi(\frac{t}{2^n})^{2^n} \tag{25}$$

Clearly, $\phi_{X_1}(t) \neq 0$ for all $t \in \mathbb{R}$, since if it was the case for a given $t_0 \in \mathbb{R}$, then $\psi(t_0) = 0$ (as $\psi(t_0) = \phi_{X_1}(t_0)\overline{\phi_{X_1}}(t_0)$), which would imply that for all $n \in \mathbb{N}$, we would have $\psi(\frac{t}{2^n}) = 0$, which would imply in particular the following limit $\lim_{n \to \infty} \psi(\frac{t}{2^n}) = 0$ which will be in contradiction with the continuity of the characteristic function and the fact that $\phi_{X_1}(0) = 1$. Hence, for all $t \in \mathbb{R}$, we have $\phi_{X_1}(t) \neq 0$ and $\overline{\phi_{X_1}}(t) \neq 0$. We can define safely

$$\gamma(t) = \frac{\phi_{X_1}(t)}{\overline{\phi_{X_1}}(t)}.$$

Doing basic computations, we have

$$\gamma(2t) \quad = \quad \frac{\phi_{X_1}(2t)}{\overline{\phi_{X_1}}(2t)} = \frac{\phi_{X_1}(t)^3 \overline{\phi_{X_1}}(t)}{\overline{\phi_{X_1}}(t)^3 \phi_{X_1}(t)} = \frac{\phi_{X_1}(t)^2}{\overline{\phi_{X_1}}(t)^2} = \gamma(t)^2 \qquad (26)$$

Applying previous equation iteratively, we get

$$\gamma(t) = \gamma(\frac{t}{2^n})^{2^n} \qquad (27)$$

$\gamma(t)$'s value in 0 is 1 as $\phi_{X_1}(0) = 1$. As X_1 is a variable with zero mean and unit variance, we have that $\phi_{X_1}(t) = 1 - \frac{t^2}{2} + o(t^2)$. Hence, $\gamma(t) = 1 + o(t^2)$, which implies in particular that $\forall t \in \mathbb{R}$:

$$\gamma(t) = \gamma(\frac{t}{2^n})^{2^n} = \lim_{n \to \infty} \gamma(\frac{t}{2^n})^{2^n} = \lim_{n \to \infty} \left(1 + o(\frac{t}{2^n})\right)^{2^n} = 1 \qquad (28)$$

By definition of $\gamma(t)$, ϕ_{X_1} is a real function. Equation (23) becomes $\phi_{X_1}(2t) = \phi_{X_1}(t)^4$ or iteratively $\phi_{X_1}(t) = \phi_{X_1}(\frac{t}{2^n})^{2^n}$. Remember that $\phi_{X_1}(t) = 1 - \frac{t^2}{2} + o(t^2)$ as X_1 is a variable with zero mean and unit variance. Hence, we get

$$\phi_{X_1}(t) = \lim_{n \to \infty} \phi_{X_1}(\frac{t}{2^n})^{2^n} = \lim_{n \to \infty} \left(1 - \frac{1}{2}\frac{t^2}{2^n} + o(t^2)\right)^{2^n} = \exp(-\frac{t^2}{2}) \qquad (29)$$

This shows that X_1 's characteristic function is the one of a normal distribution with zero mean and unit variance, which concludes the proof. $\qquad \square$

4   Independence between sample mean and variance and normal distribution

We finally tackle the question of the condition for the sample mean and variance to be independent. This is a strong result that for instance enables

us to derive the Student distribution as in the normal case of i.i.d. variables, the sample mean and variance are clearly independent. We are interested in the opposite. What is the condition to impose on our distribution for iid variable to make our sample mean and variance independent? We shall prove that it is only in the case of normal distribution that these two estimators are independent as the following proposal states

Proposition 1. *The sample mean and variance are independent if and only if the underlying (parent) distribution is normal.*

Remark 4.1. *This result was first proved by [2] and later by [7]. We provide a new proof which is simpler as we work with the log characteristic function and the unbiased sample variance. This makes the resulting differential equation trivial to solve as this is just a constant second order derivative constraint.*

This result implies consequently that it will not be easy to derive the underlying distribution of the t-statistic for a non normal distribution. Indeed the t-statistic is defined as the ratio of the sample mean over the sample variance. If the sample mean and sample variance are not independent, the computation of the underlying distribution does not decouple. This makes the problem of the computation of the underlying distribution an integration problem that has no closed form. This kills in particular any hope to derive trivially other distribution that generalizes the case of the Student distribution to non normal underlying assumptions.

Proof. The assumption of i.i.d. sample for $(x_1, \ldots, c_n)$ implies that the joint distribution of $(x_1, \ldots, x_n)$ denoted by $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ is equal to $\prod_{i=1}^{n} f_X(x_i)$, which we will write $\prod_{i=1}^{n} f(x_i)$ dropping the $._X$ to make notation lighter.

The log of the characteristic function of the joint variable $(\bar{X}_n, s_n^2)$ writes

$$\ln(\phi_{(\bar{X}_n, s_n^2)}(t_1, t_2)) = \ln\left(\iiint e^{it_1\bar{x}_n + it_2 s_n^2} \prod_{i=1}^{n} f(x_i)dx_i\right). \qquad (30)$$

Similarly, the log of the characteristic function for the sample mean $\bar{X}_n$ writes

$$\ln(\phi_{\bar{X}_n}(t_1)) = \ln\left(\iiint e^{it_1\bar{x}_n} \prod_{i=1}^{n} f(x_i)dx_i\right), \qquad (31)$$

and similarly for the sample variance

$$\ln(\phi_{s_n^2}(t_2)) = \ln\left(\iiint e^{it_2 s_n^2} \prod_{i=1}^{n} f(x_i)dx_i\right). \qquad (32)$$

The assumption of independence between sample mean $\bar{X}_n$ and variance s_n^2 is equivalent to the fact that the characteristic function of the couple decouples, or that the log characteristic functions sum up.

$$\ln(\phi_{(\bar{X}_n,s_n^2)}(t_1,t_2)) = \ln(\phi_{\bar{X}_n}(t_1)) + \ln(\phi_{s_n^2}(t_2)). \tag{33}$$

Differentiating condition 33 with respect to t_2 in $t_2 = 0$ leads to

$$\frac{1}{\phi_{(\bar{X}_n,s_n^2)}(t_1,t2)} \left.\frac{\partial\phi_{(\bar{X}_n,s_n^2)}(t_1,t_2)}{\partial t_2}\right|_{t_2=0} = \frac{1}{\phi_{s_n^2}(t2)} \left.\frac{\partial\phi_{s_n^2}(t_2)}{\partial t_2}\right|_{t_2=0}. \tag{34}$$

Noticing that $\phi_{s_n^2}(0) = 1$ and $\phi_{(\bar{X}_n,s_n^2)}(t_1,0) = \phi_{\bar{X}_n}(t_1)$, the condition 33 writes

$$\frac{1}{\phi_{\bar{X}_n}(t_1)} \left.\frac{\phi_{(\bar{X}_n,s_n^2)}(t_1,t_2)}{\partial t_2}\right|_{t_2=0} = \left.\frac{\partial\phi_{s_n^2}(t_2)}{\partial t_2}\right|_{t_2=0}. \tag{35}$$

Using the fact that $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$, it is easy to see that

$$\phi_{\bar{X}_n}(t_1) = \prod_{i=1}^n \int e^{it_1 x_i/n} f(x_i) dx_i = [\phi_X(t_1/n)]^n \tag{36}$$

For the sample variance, we can use the "U-statistic" (or symmetric) form to see that

$$s_n^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \frac{\sum_{i\neq j} X_i X_j}{n(n-1)} \tag{37}$$

Hence, the derivative of the characteristic function of the couple $(\bar{X}_n, s_n^2)$ writes

$$\left.\frac{\partial\phi_{(\bar{X}_n,s_n^2)}(t_1,t_2)}{\partial t_2}\right|_{t_2=0} = \iiint is_n^2 \prod_{i=1}^n e^{it_1 x_i/n} f(x_i) dx_i \tag{38}$$

$$= i \iiint \left(\frac{\sum_{i=1}^n x_i^2}{n} - \frac{\sum_{i\neq j} x_i x_j}{n(n-1)}\right) \prod_{i=1}^n e^{it_1 x_i/n} f(x_i) dx_i \tag{39}$$

$$= i[\phi_X(\frac{t_1}{n})]^{n-2} \left(\phi_X(\frac{t_1}{n}) \int x^2 e^{\frac{it_1 x}{n}} f(x) dx - (\int x^2 e^{\frac{it_1 x}{n}} f(x) dx)^2\right) \tag{40}$$

In the latter equation, if we set $t_1 = 0$, we get in particular that

$$\left.\frac{\partial\phi_{s_n^2}(t_1,t_2)}{\partial t_2}\right|_{t_2=0} = \left.\frac{\partial\phi_{(\bar{X}_n,s_n^2)}(0,t_2)}{\partial t_2}\right|_{t_2=0} = i\sigma^2 \tag{41}$$

Hence, condition (35) writes

$$\frac{\phi_X(\frac{t_1}{n}) \int x^2 e^{\frac{it_1 x}{n}} f(x) dx - (\int x^2 e^{\frac{it_1 x}{n}} f(x) dx)^2}{[\phi_X(\frac{t_1}{n})]^2} = \sigma^2 \tag{42}$$

We also have that the derivative of the characteristic function $\phi_X(\frac{t_1}{n})$ with respect to $u = t_1/n$ gives

$$\frac{\partial \phi_X(u)}{\partial u} = \int ix e^{ixu} f(x) dx \tag{43}$$

To simplify notation, we drop the index in ϕ_X and writes this function ϕ. Using equation (43), condition (42) writes

$$\frac{-\phi(u)\frac{\partial^2 \phi(u)}{\partial u^2} + \left(\frac{\partial \phi(u)}{\partial u}\right)^2}{\phi(u)^2} = \sigma^2 \tag{44}$$

The log of the characteristic function of $\phi(u) = \mathbb{E}[e^{iuX}]$, denoted by $\Psi(u) = \ln \phi(u)$, first and second derivatives with respect to u are given by:

$$\frac{\partial \Psi(u)}{\partial u} = \frac{\partial \ln \phi(u)}{\partial u} = \frac{1}{\phi(u)} \frac{\partial \phi(u)}{\partial u} \tag{45}$$

$$\frac{\partial^2 \Psi(u)}{\partial^2 u} = \frac{\partial}{\partial u} \frac{\partial \Psi(u)}{\partial u} = \frac{1}{\phi(u)} \frac{\partial^2 \phi(u)}{\partial u^2} - \frac{1}{\phi(u)^2} \left(\frac{\partial \phi(u)}{\partial u}\right)^2 \tag{46}$$

Hence, condition (44) writes

$$\frac{\partial^2 \Psi(u)}{\partial^2 u} = -\sigma^2 \tag{47}$$

Using the boundary conditions $\Psi(0) = 0$ and $\Psi'(0) = i\mathbb{E}[X] = i\mu$, it is easy to integrate condition (47) which is a constant second order derivative to get

$$\Psi(u) = i\mu u - \frac{\sigma^2 u^2}{2} \tag{48}$$

Condition (48) states that a necessary and sufficient condition for the sample mean and variance to be independent is that the log characteristic function of X is a quadratic form. But a quadratic form for the log characteristic function of X is a characterization of a normal distribution, which concludes the proof. $\qquad \square$

5   Conclusion

In this paper, we have presented three remarkable properties of the normal distribution: first that if X_1 and X_2 are two independent variables with X_1+X_2 normally distributed, then both X_1 and X_2 follow a normal distribution, which is referred to as the Cramer theorem, second a variation of the Levy Cramer theorem that states that if X_1 and X_2 are independent symmetric random variables with finite variance with X_1+X_2 and X_1-X_2 independent, then both X_1 and X_2 are normal, and third that the normal distribution is characterized by the fact that it is the only distribution for which the sample mean and variance are independent (which is a central property for deriving the Student distribution and referred as the Geary theorem (see [2])).

The novelty of this paper has been to provide new, quicker or self contained proofs of theses theorems. In particular, we revisited the proof first provided by [3] to give full details of it as the sketch of the proof provided was rather elusive on technical details in complex analysis. Also using the log characteristic function turns out to provide a quicker proof of the result on the characterization of the normal distribution as the only distribution for which sample mean and variance are independent.

References

[1] H. Cramr, Sur une proprit de la loi de Gauss, *C. R. Acad. Sci.*, (1936), 615–616.

[2] R. C. Geary, Distribution of Student's ratio for nonnormal samples, *Royal Statistical Society Journal, Supplement* , Taylor & Francis, **3**, (1936).

[3] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, **2**, 2001.

[4] P. Lvy, Proprits asymptotiques des sommes de variables alatoires indpendantes ou enchanes, *J. Math. Pures Appl.*, (1935), 347-402.

[5] Linnik, Ju. V. and Ostrovskii, I. V., *Decomposition of random variables and vectors*, Providence, American Mathematical Society, 1977.

[6] Linnik, Yu. V. and Spira, Robert, On the decomposition of the convolution of Gaussian and Poissonian laws, *Theory of Probability and its Applications*, **2**, (1957), 31–57.

[7] Eugene Lukacs, A Characterization of the Normal Distribution, *Ann. Math. Statist.*, The Institute of Mathematical Statistics, **13**(1), (1942), 91–93, https://doi.org/10.1214/aoms/1177731647.