© REVISTA DE MATEMÁTICA: TEORÍA Y APLICACIONES 2021 **28**(2) : 261–277
CIMPA – UCR ISSN: 1409-2433 (PRINT), 2215-3373 (ONLINE)
DOI: 10.15517/rmta.v28i2.44507

A MOMENT RECURSIVE FORMULA FOR A CLASS OF DISTRIBUTIONS

UNA FÓRMULA RECURSIVA PARA LOS MOMENTOS DE ALGUNAS DISTRIBUCIONES DE PROBABILIDAD

LUIS RINCÓN *

Received: 16/Nov/2020; Revised: 26/Feb/2021; Accepted: 19/Apr/2021

---

*Universidad Nacional Autónoma de México, Facultad de Ciencias, Departamento de Matemáticas, Ciudad de México, México. E-mail: lars@ciencias.unam.mx

Abstract

We provide a recursive formula for the computation of moments of distributions belonging to a subclass of the exponential family. This subclass includes important cases as the binomial, negative binomial, Poisson, gamma and normal distribution, among others. The recursive formula provides a procedure to sequentially calculate the moments using only elementary operations. The approach makes no use of the moment generating function.

Keywords: moments; exponential family; recursive formula.

Resumen

Se proporciona una fórmula recursiva para calcular los momentos de ciertas distribuciones que pertenecen a una subclase de la familia exponencial. A esta subclase de distribuciones pertenecen las distribuciones binomial, binomial negativa, Poison, gama y normal, entre otras. La fórmula recursiva provee de un procedimiento para calcular los momentos de manera secuencial usando únicamente operaciones elementales. El método no hace uso de la función generadora de momentos.

Palabras clave: momentos; familia exponencial; fórmula recursiva.

Mathematics Subject Classification: 60E05, 97K50, 97K60.

1 Introduction

Finding the moments of a given probability distribution is not an easy problem in most cases. These are defined, in the continuous case and assuming absolute convergence, as the numbers:

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x)\, dx, \quad n = 0, 1, \ldots \tag{1}$$

where X is a random variables with probability density $f(x)$. Frequently, there is no closed form for such quantities. Uncommon cases are the Poisson and the exponential distributions, where closed formulas are well known for their moments of arbitrary order. For the binomial distribution, on the other hand, no formula is known. In 2005, Benyi and Manago [2] provided a simple recursive expression for the k-th moment of the bin(n, θ) distribution in terms of the $(k-1)$-th moment of the bin(n, θ) and bin$(n-1, \theta)$ distributions, but only in the case $\theta = 1/2$. More generally, in 1981, Link [11] provided formulas for the moments of some discrete probability distributions in terms of finite difference operators. In 1982, Chan [3] systematized the method of Link in terms of operator valued probability generating functions and applied it to the hypergeometric and the negative hypergeometric distributions.

In this work we present a general formula to recursively calculate the moments of any probability distribution belonging to a subclass of the so-called exponential family of distributions. We will see that our recursive formula takes a different approach from the classical method to calculate moments using the moment generating function and has some computational advantages.

It is worth mentioning that moments of distributions are important quantities since they are often used in standard statistical procedures to identify a distribution. What underlies in those applications is the fact that, under certain conditions, the set of all moments uniquely determine a distribution [1]. Let us start by defining the probability distributions we will concentrate on.

2 A subclass of the exponential family

Consider the one-parameter exponential family of probability densities:

$$f(x;\theta) = a(\theta)\, b(x)\, \exp\{c(\theta)\, d(x)\}, \tag{2}$$

where $a(\theta)$, $b(x)$, $c(\theta)$ and $d(x)$ are functions depending only on a parameter θ or a variable x, as indicated. This family includes several important distributions, both discrete and continuous, but for simplicity we will call $f(x;\theta)$ a density. The support of these densities is determined by $b(x)$ and $d(x)$, and the specification of these functions may include the indicator function of a certain set of real numbers. General properties are well known for this important collection of distributions, see for example [12]. To define the subclass of distributions we will study, we consider the case $d(x) = x$, and for convenience $c(\theta)$ is replaced by $\ln c(\theta)$. Then, the density (2) reduces to:

$$f(x;\theta) = a(\theta)\, b(x)\, (c(\theta))^x. \tag{3}$$

Table 1 shows some examples of discrete and continuous distributions with densities in the form (3).

It is easy to verify that those definitions of $a(\theta)$, $b(x)$ and $c(\theta)$ yield the corresponding density for each distribution. In particular, for the binomial distribution, it is customary to use the letter n for the number of trials, however, we will reserve that letter to denote the general order of a moment. Also, recall the geometric(θ) distribution is the particular case of neg bin(r,θ) when $r = 1$, and the exp(θ) distribution is included in the gamma(α,θ) case when $\alpha = 1$.

Table 1: Some probability distributions.

Distribution	$a(\theta)$	$b(x)$	$c(\theta)$
$\text{bin}(m,\theta)$	$(1-\theta)^m$	$\binom{m}{x}\cdot 1_{\{0,1,,\ldots,m\}}(x)$	$\dfrac{\theta}{1-\theta}$
neg $\text{bin}(r,\theta)$	θ^r	$\binom{r+x-1}{x}\cdot 1_{\{0,1,\ldots\}}(x)$	$1-\theta$
$\text{Poisson}(\theta)$	$e^{-\theta}$	$\dfrac{1}{x!}\cdot 1_{\{0,1,\ldots\}}(x)$	θ
$\text{gamma}(\alpha,\theta)$	θ^α	$\dfrac{x^{\alpha-1}}{\Gamma(\alpha)}\cdot 1_{(0,\infty)}(x)$	$e^{-\theta}$
$\text{N}(\theta,\sigma^2)$	$e^{-\theta^2/2\sigma^2}$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}, e^{-x^2/2\sigma^2}$	e^{θ/σ^2}

3 General results

Here is our main result and some general immediate consequences.

Theorem 1 *: Let X be a random variable with density (3), where $a(\theta)$ and $c(\theta)$ are differentiable with $c'(\theta)\neq 0$. The n-th moment of X exists and is given by:*

$$E(X^n) = \frac{c(\theta)}{c'(\theta)}\left(-\frac{a'(\theta)}{a(\theta)} + \frac{d}{d\theta}\right) E(X^{n-1}), \quad \textit{for } n\geq 1. \tag{4}$$

There are several ways to prove this formula. For example, considering the continuous case and interchanging the derivative and the integral, we have:

$$\begin{aligned}
E(X^n) &= \int_{-\infty}^{\infty} a(\theta)\, b(x)\, x^{n-1}\left[x\,(c(\theta))^{x-1}\right] c(\theta)\, dx \\
&= \int_{-\infty}^{\infty} b(x)\, x^{n-1}\left[\frac{d}{d\theta}(c(\theta))^x\right] a(\theta)\,\frac{c(\theta)}{c'(\theta)}\, dx \\
&= \int_{-\infty}^{\infty} b(x)\, x^{n-1} \\
&\qquad \left[\frac{d}{d\theta}\left(a(\theta)\,\frac{(c(\theta))^{x+1}}{c'(\theta)}\right) - (c(\theta))^x\,\frac{d}{d\theta}\left(a(\theta)\,\frac{c(\theta)}{c'(\theta)}\right)\right] dx \\
&= \frac{d}{d\theta}\left(\frac{c(\theta)}{c'(\theta)}\, E(X^{n-1})\right) - \left(\frac{1}{a(\theta)}\,\frac{d}{d\theta}\left(a(\theta)\,\frac{c(\theta)}{c'(\theta)}\right)\right) E(X^{n-1}) \\
&= \frac{c(\theta)}{c'(\theta)}\left(-\frac{a'(\theta)}{a(\theta)} + \frac{d}{d\theta}\right) E(X^{n-1}).
\end{aligned} \tag{5}$$

Similar procedure applies in the discrete case. Alternatively, analogous calculations show that the following formula holds:

$$\frac{c(\theta)}{c'(\theta)}\left(-\frac{a'(\theta)}{a(\theta)}+\frac{d}{d\theta}\right) f(x;\theta)=x\cdot f(x;\theta). \tag{6}$$

Multiplying by x^{n-1}, integrating or summing respect to x and interchanging the derivative and the integral, leads to (4). Yet for another proof one can simply expand the right hand side of (4), interchange the derivative and the integral (or sum in the discrete case), and obtain the same result.

Thus, formula (4) expresses the n-th moment in terms of a differential operator applied to the $(n-1)$-th moment, starting from $E(X^0)=1$. In particular, for $n=1$ and $n=2$, general formulas for the expectation and variance of X can be obtained, namely:

$$E(X) = -\frac{c(\theta)}{c'(\theta)}\cdot\frac{a'(\theta)}{a(\theta)}. \tag{7}$$

$$\text{Var}(X) = \frac{c(\theta)}{c'(\theta)}\cdot E'(X). \tag{8}$$

Here the dash in the expectation also means derivative respect to θ. More generally, we can write the n-th moment of X as the n-th power of the differential operator applied to the constant 1, i.e.

$$E(X^n)=\left[\frac{c(\theta)}{c'(\theta)}\left(-\frac{a'(\theta)}{a(\theta)}+\frac{d}{d\theta}\right)\right]^n (1). \tag{9}$$

This is a rather compact formula but calculating the n-th power of the differential operator is not simple since the two terms inside the square bracket do not commute. From (7) and (8), we can alternatively write:

$$E(X^n)=\left[E(X)+\frac{\text{Var}(X)}{E'(X)}\frac{d}{d\theta}\right]^n(1). \tag{10}$$

In [10] there is a nice formula for the m.g.f. of distributions in the exponential family. However, formula (4) takes a different approach in calculating moments and we will demonstrate that this recursive relation has some advantages.

Formula (4) is not completely new. It can be derived from Stein's identity [18] in the case of the normal distribution and its extension given by Hudson [5] for the exponential family. However, its derivation is not straightforward. See also [8]. We have here derived (4) without relying on any previous work. In the rest of this paper we will specialize (4) to the distributions shown in Table 1. These results are summarized in Table 2.

Table 2: Moment recursive formulas.

Distribution	Moment recursive formula
$\text{bin}(m,\theta)$	$E(X^n) = \left[m\theta + \theta(1-\theta)\,\frac{d}{d\theta} \right] E(X^{n-1})$
neg $\text{bin}(r,\theta)$	$E(X^n) = \left[\frac{r(1-\theta)}{\theta} - (1-\theta)\frac{d}{d\theta} \right] E(X^{n-1})$
$\text{Poisson}(\theta)$	$E(X^n) = \left[\theta + \theta\,\frac{d}{d\theta} \right] E(X^{n-1})$
$\text{gamma}(\alpha,\theta)$	$E(X^n) = \left[\frac{\alpha}{\theta} - \frac{d}{d\theta} \right] E(X^{n-1})$
$\text{N}(\theta,\sigma^2)$	$E(X^n) = \left[\theta + \sigma^2\frac{d}{d\theta} \right] E(X^{n-1})$

4 Some particular cases

4.1 Binomial distribution

The n-th moment of $X \sim \text{bin}(m,\theta)$ satisfies:

$$E(X^n) = \left(m\theta + \theta(1-\theta)\,\frac{d}{d\theta} \right) E(X^{n-1}), \quad n \geq 1. \tag{11}$$

Starting from $E(X^0) = 1$ and applying the differential operator successively we have the next expressions:

$$\begin{aligned}
E(X) &= m\theta \qquad (12)\\
E(X^2) &= m\theta + m(m-1)\theta^2\\
E(X^3) &= m\theta + 3m(m-1)\theta^2 + m(m-1)(m-2)\theta^3\\
E(X^4) &= m\theta + 7m(m-1)\theta^2 + 6m(m-1)(m-2)\theta^3\\
&\quad +m(m-1)(m-2)(m-3)\theta^4\\
E(X^5) &= m\theta + 15m(m-1)\theta^2 + 25m(m-1)(m-2)\theta^3\\
&\quad +10m(m-1)(m-2)(m-3)\theta^4\\
&\quad +m(m-1)(m-2)(m-3)(m-4)\theta^5\\
&\vdots
\end{aligned}$$

Although these expressions grow quickly in length, the calculations are straightforward, particularly if a computer is used. See comments in the last section.

These results extend those in [2] and clearly suggest the following general formula which can be proved by induction using the differential operator.

Corollary 1 *: Let $X \sim bin(m,\theta)$. For any integer $n \geq 1$,*

$$E(X^n) = \sum_{j=0}^{n-1} \xi(n,j) \cdot m(m-1)\cdots(m-j) \cdot \theta^{j+1}. \tag{13}$$

where coefficients $\xi(n,j)$ satisfy the difference equation:

$$\xi(n,j) = \xi(n-1,j-1) + (j+1)\cdot\xi(n-1,j), \quad j = 1,\ldots,n-1. \tag{14}$$

with boundary conditions $\xi(n,0) = 1$ and $\xi(n,n-1) = 1$.

Using the recursive relation (11) applied to (13), we obtain an equality of two polynomials in θ. Equating the corresponding coefficients, one arrives at the difference equation (14). The boundary coefficients being the first and the last equality in the power series. Observe the coefficients $\xi(n,j)$ do not depend on the parameters of the distribution and the first few of them can be written in a triangle array:

$$\begin{array}{lllll} 1 & & & & \\ 1 & 1 & & & \\ 1 & 3 & 1 & & \\ 1 & 7 & 6 & 1 & \\ 1 & 15 & 25 & 10 & 1 \\ \ldots\ldots & & & & \end{array} \tag{15}$$

As an example, let us take the number 25 in the array (15). This corresponds to the fifth moment $(n = 5)$ and index $j = 2$. Then $25 = 7 + (j+1)\cdot 6$. Solving the difference equation (14) seems to be a difficult task and we will not attempt to do that here. However, for $j = 1$, it is not difficult to check that $\xi(n,1) = 2^{n-1} - 1$, for $n \geq 2$. This yields the values $1, 3, 7, 15$ in the second column of (15).

It is also interesting to note that the power series of the n-th moment of X has n different terms as shown in (12), or (13), as long as $n \leq m$.

For moments of order larger than the parameter m, the power series (13) has at most m terms since there is a vanishing factor when the index j reaches the value m and beyond. For example, suppose $m = 2$. Then the list of moments (12) reduces to:

$$\begin{aligned} E(X) &= m\theta \\ E(X^2) &= m\theta + m(m-1)\theta^2 \\ E(X^3) &= m\theta + 3m(m-1)\theta^2 \\ E(X^4) &= m\theta + 7m(m-1)\theta^2 \\ E(X^5) &= m\theta + 15m(m-1)\theta^2 \\ &\vdots \end{aligned} \tag{16}$$

The coefficients are as in (15) but restricted to only the first two columns. The rest of the array is discarded. These coefficients still satisfy the difference equation (14), namely for $j = 1$, we have $3 = 1 + (j+1)\cdot 1$, $7 = 1 + (j+1)\cdot 3$ and $15 = 1 + (j+1)\cdot 7$. Thus, formula (13) and the difference equation (14) hold for all possible values of the parameter m, although several summands of (13) are null. The extreme case $m = 1$ reduces all moments to the same value θ as expected.

4.2 Negative binomial distribution

The n-th moment of $X \sim$ neg bin(r, θ) satisfies:

$$E(X^n) = \left(\frac{r(1-\theta)}{\theta} - (1-\theta)\frac{d}{d\theta} \right) E(X^{n-1}), \quad n \geq 1. \tag{17}$$

Starting from $E(X^0) = 1$, one obtains:

$$\begin{aligned} E(X) &= (r - r\theta)/\theta \\ E(X^2) &= ((r + r^2) - (r + 2r^2)\theta + r^2\theta^2)/\theta^2 \\ E(X^3) &= ((2r + 3r^2 + r^3) - 3(r + 2r^2 + r^3)\theta + (r + 3r^2 + 3r^3)\theta^2 - r^3\theta^3)/\theta^3 \\ &\vdots \end{aligned} \tag{18}$$

The case $r = 1$ reduces to the geometric(θ) distribution and we can write a few more moments:

$$\begin{aligned} E(X) &= (1-\theta)/\theta \\ E(X^2) &= (2-3\theta+\theta^2)/\theta^2 \\ E(X^3) &= (6-12\theta+7\theta^2-\theta^3)/\theta^3 \\ E(X^4) &= (24-60\theta+50\theta^2-15\theta^3+\theta^4)/\theta^4 \\ E(X^5) &= (120-360\theta+390\theta^2-180\theta^3+31\theta^4-\theta^5)/\theta^5 \\ &\vdots \end{aligned} \tag{19}$$

These expressions suggest the following result.

Corollary 2 *: Let $X \sim$ geometric(θ). For any integer $n \geq 1$,*

$$E(X^n) = \sum_{j=0}^{n} \xi(n,j) \cdot \frac{\theta^j}{\theta^n}. \tag{20}$$

where coefficients $\xi(n,j)$ satisfy the difference equation:

$$\xi(n,j) = (n-j) \cdot \xi(n-1,j) + (j-n-1) \cdot \xi(n-1,j-1). \tag{21}$$

for $j = 1, 2, \ldots, n-1$, with boundary conditions $\xi(n,0) = n!$ and $\xi(n,n) = (-1)^n$.

This can be proved by induction using the differential operator. The difference equation follows after equating the coefficients of the two polynomials obtained when the power series (20) is substituted in the recursive formula (17) with $r = 1$.

4.3 Poisson distribution

The n-th moment of $X \sim$ Poisson(θ) satisfies:

$$E(X^n) = \left(\theta + \theta \frac{d}{d\theta}\right) E(X^{n-1}), \quad n \geq 1. \tag{22}$$

Starting from $E(X^0) = 1$ one successively obtains:

$$\begin{aligned} E(X) &= \theta \\ E(X^2) &= \theta + \theta^2 \\ E(X^3) &= \theta + 3\theta^2 + \theta^3 \\ E(X^4) &= \theta + 7\theta^2 + 6\theta^3 + \theta^4 \\ E(X^5) &= \theta + 15\theta^2 + 25\theta^3 + 10\theta^4 + \theta^5 \\ &\vdots \end{aligned} \tag{23}$$

Again, this sequence of polynomials suggests that the n-th moment of this distribution can be written as follows.

Corollary 3 *Let $X \sim$ Poisson(θ). For any integer $n \geq 1$,*

$$E(X^n) = \sum_{j=0}^{n-1} \xi(n,j) \cdot \theta^{j+1}. \tag{24}$$

where coefficients $\xi(n,j)$ satisfy the difference equation:

$$\xi(n,j) = \xi(n-1,j-1) + (j+1) \cdot \xi(n-1,j). \quad j = 1, \ldots, n-1, \tag{25}$$

with boundary conditions $\xi(n,0) = 1$ and $\xi(n,n) = 0$.

As in the previous examples, this result can be proved by induction using the differential operator. The difference equation (25) is a consequence of equating the coefficients of the two polynomials obtained when the power series (24) is substituted in the recursive formula (22).

It is interesting to note that coefficients $\xi(n,j)$ for the Poisson distribution are exactly the same as in the binomial distribution. This is really no surprise since the moments of the binomial distribution converge to the moments of the Poisson distribution. Indeed, we can readily verify this classical result using (13). The factor $m(m-1)\cdots(m-j)\, p^{j+1}$ which appears in the n-th moment of the bin(m,p) distribution can be written as $mp \cdot (m-1)p \cdots (m-j)p$. This converges to θ^{j+1} when $m \to \infty$ and $p \to 0$ in such a way that $mp \to \theta$.

To finish this section, let us recall that for the Poisson(θ) distribution the following formula is known: $E(X^n) = \theta \sum_{k=0}^{n-1} \binom{n-1}{k} E(X^k)$. This can be proved by induction on n using the recursive formula (22). See also [15] for an alternative formula in this case.

4.4 Gamma distribution

The n-th moment of the gamma(α,θ) distribution satisfies:

$$E(X^n) = \left(\frac{\alpha}{\theta} - \frac{d}{d\theta} \right) E(X^{n-1}). \quad n \geq 1, \tag{26}$$

which yields the sequence:

$$\begin{aligned} E(X) &= \alpha/\theta, \\ E(X^2) &= (\alpha+\alpha^2)/\theta^2, \\ E(X^3) &= (2\alpha+3\alpha^2+\alpha^3)/\theta^3 \\ E(X^4) &= (6\alpha+11\alpha^2+6\alpha^3+\alpha^4)/\theta^4 \\ E(X^5) &= (24\alpha+50\alpha^2+35\alpha^3+10\alpha^4+\alpha^5)/\theta^5 \\ &\vdots \end{aligned} \tag{27}$$

It is known, for this distribution, that $E(X^n) = \alpha(\alpha+1)\cdots(\alpha+n-1)/\theta^n$. The case $\alpha = 1$ reduces to the exp(θ) distribution, for which $E(X^n) = n!/\theta^n$. These two formulas can now be proved by induction using (26). The above sequence of polynomials suggest that the n-th moment of the gamma distribution can also be written as follows.

Corollary 4 *: Let $X \sim$ gamma(α,θ). For any integer $n \geq 1$,*

$$E(X^n) = \sum_{j=0}^{n-1} \xi(n,j)\cdot\frac{\alpha^{j+1}}{\theta^n}. \tag{28}$$

where coefficients $\xi(n,j)$ satisfy the difference equation:

$$\xi(n,j) = \xi(n-1,j-1) + (n-1)\cdot\xi(n-1,j). \quad j = 1,\ldots,n-2, \tag{29}$$

with boundary conditions $\xi(n,0) = (n-1)!$ and $\xi(n,n-1) = 1$.

Once more, this result is obtained after equating the coefficients of the two polynomials resulting when (28) is substituted into the recursive formula (26). Considering the exponential case ($\alpha = 1$), note that $n!$ should be equal to the sum of the numeric coefficients of $E(X^n)$, for example, for the fifth moment, $5! = 1 + 10 + 35 + 50 + 24$. This is indeed the case in general and we prove it next.

Corollary 5 *: The coefficients $\xi(n,j)$ given by (29) satisfy* $\displaystyle\sum_{j=0}^{n-1}\xi(n,j) = n!$

Let S_n be the sum of all numeric coefficients of the n-th moment. Using the difference equation (29) and the boundary conditions it is easy to obtain that $S_n = n \cdot S_{n-1}$, from which the result follows.

4.5 Normal distribution

The n-th moment of the $\mathrm{N}(\theta, \sigma^2)$ distribution satisfies:

$$E(X^n) = \left(\theta + \sigma^2 \frac{d}{d\theta}\right) E(X^{n-1}). \quad n \geq 1. \tag{30}$$

Starting from $E(X^0) = 1$, one obtains successively:

$$\begin{aligned}
E(X) &= \theta, \\
E(X^2) &= \theta^2 + \sigma^2, \\
E(X^3) &= \theta^3 + 3\theta\sigma^2, \\
E(X^4) &= \theta^4 + 6\theta^2\sigma^2 + 3\sigma^4, \\
E(X^5) &= \theta^5 + 10\theta^3\sigma^2 + 15\theta\sigma^4, \\
E(X^6) &= \theta^6 + 15\theta^4\sigma^2 + 45\theta^2\sigma^4 + 15\sigma^6, \\
E(X^7) &= \theta^7 + 21\theta^5\sigma^2 + 105\theta^3\sigma^4 + 105\theta\sigma^6, \\
E(X^8) &= \theta^8 + 28\theta^6\sigma^2 + 210\theta^4\sigma^4 + 420\theta^2\sigma^6 + 105\sigma^8, \\
E(X^9) &= \theta^9 + 36\theta^7\sigma^2 + 378\theta^5\sigma^4 + 1260\theta^3\sigma^6 + 945\theta\sigma^8, \\
E(X^{10}) &= \theta^{10} + 45\theta^8\sigma^2 + 630\theta^6\sigma^4 + 3150\theta^4\sigma^6 + 4725\theta^2\sigma^8 + 945\sigma^{10} \\
&\vdots
\end{aligned} \tag{31}$$

We have written a longer list of moments here as a careful examination of these expressions was crucial to derive the following general formula for even moments.

Corollary 6 *: Let $X \sim N(\theta, \sigma^2)$. For any integer $n \geq 0$,*

$$E(X^{2n}) \;=\; \sum_{k=0}^{n} \binom{2n}{2k}\binom{2k}{k} \frac{k!}{2^k}\, \theta^{2n-2k}\, \sigma^{2k}. \tag{32}$$

A proof by induction follows after some lengthy algebraic operations using the differential operator twice. Assuming the formula holds for some integer $n \geq 0$, the squared differential operator applied to the above sum yields four sums not all of them with index from 0 to n due to the derivatives. Making a change of index when appropriate so that the powers in each sum appear as $\theta^{2n+2-2k}\,\sigma^{2k}$ and joining the sums produce the result for $2n+2$. In particular, when $\theta = 0$, only the last term ($k = n$) of the sum in (32) remains and the well known formula $E(X^{2n}) = [(2n)!/(2^n\, n!)]\,(\sigma^2)^n$ is recovered. Also, from (32) we can now derive a general formula for the odd moments.

Corollary 7 *: Let $X \sim N(\theta, \sigma^2)$. For any integer $n \geq 0$,*

$$E(X^{2n+1}) \;=\; \sum_{k=0}^{n} \binom{2n+1}{2k} \binom{2k}{k} \frac{k!}{2^k}\, \theta^{2n-2k+1}\, \sigma^{2k}. \tag{33}$$

This formula follows after applying the differential operator to the expression of $E(X^{2n})$. For centralized normal distributions ($\theta = 0$), odd moments are zero as expected.

5 Approximations to probability densities

As an application of our results, in this section we briefly recall a few moment-based schemes to approximate a probability density. For a more detailed discussion see [6], [14] and [16].

Let $f(x)$ be a probability density with bounded support (a, b). In [13], Munkhammar *et al.* discuss the method of moments to approximate $f(x)$ by a polynomial:

$$p(x) = w_0 + w_1 x + w_2 x^2 + \cdots + w_n x^n. \quad a < x < b, \tag{34}$$

for some coefficients $w_0, w_1, \ldots, w_n$ and some natural number n. Equating the first $n+1$ moments of $f(x)$ with those of $p(x)$ (as if $p(x)$ were a genuine density on (a, b)), yields a system of $n+1$ linear equations for the unknown weights $w_0, w_1, \ldots, w_n$. Assuming a unique solution for such a linear system exists, we have the approximation $f(x) \approx p(x)$. Thus, $p(x)$ is a polynomial that best approximates $f(x)$ in the sense defined above. As n increases, more moments of $f(x)$ are needed and the approximation becomes more accurate. It is clear that knowing an arbitrary number of moments of $f(x)$ is crucial in this procedure.

In most practical situations $f(x)$ is unknown, a sample is then required and the sampling moments are used instead of the theoretical moments. In our case, we can assume $f(x)$ is known and belongs to the exponential family in its reduced form (3). Its moments of any order can be calculated using the recursive formula (4).

The advantage of such an approximation lies in the fact that the calculation of some probabilities could be simpler using $p(x)$ instead of $f(x)$. For example, this approach can be used to approximate probabilities for the $\mathrm{N}(0, 1)$ distribution truncated on the bounded interval $(-4, 4)$.

There are other more elaborate polynomial expansions for a probability density $f(x)$. One of such expansions is the Gram-Charlier series of type A. This series is defined in terms of standard normal distribution and the Hermite polynomials. Again, the moments of $f(x)$ are used in such a construction. See [4] and [9].

More particularly, in the classical theory of risk, it is important to determine the distribution of the aggregate loss defined by the compound random variable:

$$X = \sum_{i=1}^{N} U_i. \tag{35}$$

where N counts the random number of losses over a definite period of time and $U_1, U_2, \ldots$ denote the losses. It is customary to assume that the random variables U_i are positive, independent, identically distributed and independent of N. Finding the distribution of X is a hard problem and any approximation is of practical importance.

Jin *et al.* [6] and Nadarajah*et al.* [14] apply a moment-based method introduced by Provost [16] to approximate the density function of (35). In this case, the distribution of the random variables N and U_i are known but the density function $f(x)$ of (35) is completely unknown. However, its moments can be computed in terms of the moments of N and U_i, assuming we know the latter. The approximation reads:

$$f(x) \approx g(x)\,[\,c_0 + c_1\,x + \cdots + c_t\,x^t\,]. \tag{36}$$

where t is a natural number and $g(x)$ is the density of a gamma distribution. This is called the base distribution and is one of several options available. The two parameters of the gamma density are chosen so that the first and second moment of $g(x)$ coincide with those of $f(x)$. Further moments of $f(x)$ are then used to determine the coefficients $c_0, \ldots, c_t$. Now, if N and U_i follow a distribution for which formulas are known for their moments, or belong to the family (3) so that we can recursively calculate their moments, then the moments of X can be found and the approximation (36) can be carried out. This shows a practical application where moments can help approximate an unknown distribution.

6 Concluding remarks

We have provided a general recursive formula for the n-th moment of a probability distribution in a subclass of the exponential family. Classically, if one needs to calculate such quantity, one can try solving the corresponding integral (or sum) or calculate the n-th derivative of the moment generating function (assuming this is given) and evaluate at zero. The alternative procedure here proposed requires only the computation of derivatives of polynomials and some simple algebraic operations. This is perfectly suitable for use in a computer algebra system (CAS), assuming it can calculate derivatives and reduce expressions. An example of such system is `Sage` [17], where the command `diff()` can be used to calculate derivatives. For example, the following sequence of `Sage` commands was used to produce the first 10 moments of the normal distribution:

```
# Moments normal(t,s2)
t,s2 = var('t,s2')
m=1
for n in range(10):
    m=t*m+s2*diff(m,t)
    print 'Moment',n+1,'is',m.simplify_full()
```

Here `t` stands for θ and `s2` for σ^2. The command `diff(m,t)` calculates the derivative of `m` respect to `t`. Short computer programs like these make the process of finding the moments a breeze thanks to the recursive perspective. They also prevent errors in the handling of long expressions and may even yield full expansions, or adequate factorization, of the results if needed.

The differential operator in the method proposed takes the form $D(\theta) = A(\theta) + B(\theta)\, d/d\theta$, where $A(\theta)$ and $B(\theta)$ are functions of θ specified in (10). Finding conditions on these two functions such that the sequence of functions generated by the powers of $D(\theta)$ resulted in a truly sequence of moments is a natural question. Thus, characterization of a distribution via this differential operator might be possible and this is something worth exploring. The result we have found provides yet another link between probability theory and differential difference equations.

Acknowledgments

Thanks to the referees for their constructive and valuable observations that allowed an improvement of the manuscript.

Funding

This work was supported by UNAM, México.

References

[1] N.I. Akhiezer, *The classical moment problem and some related questions in analysis*, Oliver & Boyd, Edinburgh, 1965.

[2] A. Bényi, S. M. Manago, *A recursive formula for moments of a binomial distribution*, The College Mathematics Journal, **36**(2005), no. 1, 68–72. Doi: 10.2307/30044825

[3] B. Chan, *Derivation of moment formulas by operator valued probability generating functions*, The American Statistician **36**(1982), no. 3a, 179–181. Doi: 10.1080/00031305.1982.10482826

[4] H. Cramér, *Mathematical Methods of Statistics.* Princeton University Press, Princeton NJ, 1946. Doi: 10.1515/9781400883868

[5] H.M. Hudson, *A natural identity for exponential families with applications in multiparameter estimation*, The Annals of Statistics, **6**(1978), no. 3, 473–484. Doi: 10.1214/aos/1176344194

[6] T. Jin, S.B. Provost, J. Ren, *Moment-based density approximations for aggregate losses*, Scandinavian Actuarial Journal, (2016), no. 3, 216–245. Doi: 10.1080/03461238.2014.921640

[7] J.S.J. Kang, S.B. Provost, J. Ren, *Moment-based density approximation techniques as applied to heavy-tailed distributions*, International Journal of Statistics and Probability, **8**(2019), no. 3. Doi: 10.5539/ijsp.v8n3p1

[8] S.K. Kattumannil, *On Stein's identity and its applications*, Statistics and Probability Letters, **79** (2009) 1444–1449. Doi: 10.1016/j.spl.2009.03.021

[9] M.G. Kendall, *The Advanced Theory of Statistics*. Charles Griffin & Company, London, 1945.

[10] E.L. Lehmann, G. Casella, *Theory of Point Estimation*, 2nd edition, Springer, New York NY, 1988. Doi: 10.1007/b98854

[11] F.R. Link, *Moments of discrete probability distributions derived using finite difference operators*, The American Statistician, **35**(1981), no.1, 44–46. Doi: 10.2307/2683584

[12] N. Mukhopadhyay, *Introductory Statistical Inference*, Chapman and Hall/CRC, New York NY, 2006. Doi: 10.1201/b16497

[13] J. Munkhammar, L. Mattsson, J. Rydén, *Polynomial probability distribution estimation using the method of moments*, PLOS ONE, **12**(2017), no.4. Doi: 10.1371/journal.pone.0174573

[14] S. Nadarajah, J. Chu, X. Jiang, *On moment based density approximations for aggregate losses*, Journal of Computational and Applied Mathematics, **298**(2016), 152–166. Doi: 10.1016/j.cam.2015.11.048

[15] C. Philipson, *A note on moments of a Poisson probability distribution*, Scandinavian Actuarial Journal, (1963), no.3-4, 243-244. Doi: 10.1080/03461238.1963.10410613

[16] S.B. Provost, *Moment-based density approximants*, The Mathematica Journal, **9**(2005), no.4, 727–756.

[17] SageMath, Open-source Mathematics Software System. In: `https://www.sagemath.org`. Accessed by 10/07/2020.

[18] C.M. Stein, *Estimation of the mean of a multivariate normal distribution*, The Annals of Statistics **9**(1981), no. 6, 1135–1151. Doi: 10.1214/aos/1176345632

[19] B. Sundt, W.S. Jewell, *Further results on recursive evaluation of compound distributions*, ASTIN Bulletin: The Journal of the International Actuarial Association **12**(1981), no.1, 27–39. Doi: 10.1017/S0515036100006802