ESTIMATION OF RELATIVE ENTROPY USING NOVEL TAYLOR-LIKE REPRESENTATIONS

P. CERONE

Abstract. Sharp bounds are obtained for perturbed generalised Taylor series. The perturbation involves the arithmetic sum of the upper and lower bounds of the $(n+1)^{th}$ derivative. The sharpest bound is in terms of the one norm of the Appell polynomial which constitutes the coefficients of the derivative of the function to be approximated. The results are demonstrated for the estimation of the Kullback-Leibler distance, Shannon entropy and mutual information.

1. Introduction

A number of authors have recently obtained generalisations of the traditional Taylor series expansion of a function $f(x)$ about a point a assuming sufficient differentiability. A Taylor series representation is a fundamental mechanism for estimation in problems arising in many industrial applications. Estimates of bounds on the remainder have also been procured.

In the current article a novel Taylor-like representation is developed and is used to estimate the Kullback-Leibler distance, Shannon entropy and mututal information. These are of importance in general information theory and, in particular, coding theory.

Before proceeding further with more generalisations, let us introduce some notation. We shall term polynomials of degree k, W_k as Appell type and say $W_k \in \mathcal{A}$ if they satisfy the condition

$$W_k' = \xi_k W_{k-1}(t), \; W_0(t) = 1, \;\; t \in \mathbb{R}. \tag{1.1}$$

These are so named since Appell studied (1.1) with $\xi_k = k$ in 1880 (see [1]). Polynomials satisfying (1.1) with $\xi_k = 1$ have been termed harmonic polynomials in Matić et al. [11], however a simple scaling will demonstrate that these are Appell.

The following results were obtained by Matić et al. [11] where $P_n(t)$ satisfy (1.1) with $\xi_k = 1$.

Theorem 1. *Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of polynomials, that satisfy*

$$P_n'(t) = P_{n-1}(t), \; P_0(t) = 1, \; t \in \mathbb{R}, \; n \in \mathbb{N}, \; n \geq 1. \tag{1.2}$$

Further, let $I \subset \mathbb{R}$ be a closed interval and $a \in I$. If $f : I \to \mathbb{R}$ is any function such that for some $n \in \mathbb{N}$, $f^{(n)}$ is absolutely continuous, then for any $x \in I$

$$f(x) = T_n(f; a, x) + R_n(f; a, x) \tag{1.3}$$

Date: June 17, 2002.

1991 *Mathematics Subject Classification.* Primary 26D15, 41A58.

Key words and phrases. Perturbed generalised Taylor's formula, Appell polynomials, Sharp bounds, Čebyšev-type functional, Kullback-Leibler distance, Entropy, Mutual information.

where

$$T_n(f;a,x) = f(a) + \sum_{k=1}^{n} (-1)^{k+1} \left[P_k(x) f^{(k)}(x) - P_k(a) f^{(k)}(a) \right], \tag{1.4}$$

$$R_n(f;a,x) = (-1)^n \int_a^x P_n(t) f^{(n+1)}(t)\, dt. \tag{1.5}$$

They also pointed out the following bounds for the remainder $R_n(f;a,x)$.

Corollary 1. *With the above assumptions of Theorem 1, the following estimations hold. Namely for $x \geq a$,*

$$\begin{aligned} &|R_n(f;a,x)| \\ &\leq \begin{cases} \|P_n\|_\infty \left\| f^{(n+1)} \right\|_1 & \textit{provided} \quad f^{(n+1)} \in L_1[a,x], \\ \|P_n\|_q \left\| f^{(n+1)} \right\|_p & \textit{provided} \quad f^{(n+1)} \in L_p[a,x],\ p>1,\ \frac{1}{p}+\frac{1}{q}=1, \\ \|P_n\|_1 \left\| f^{(n+1)} \right\|_\infty & \textit{provided} \quad f^{(n+1)} \in L_\infty[a,x], \end{cases} \end{aligned} \tag{1.6}$$

where $x \geq a$ and $\|\cdot\|_s$ $(1 \leq s \leq \infty)$ are the usual $s-$Lebesgue norms. That is,

$$\|g\|_s := \left(\int_a^x |g(t)|^s\, dt \right)^{\frac{1}{s}}, \quad s \in [1,\infty)$$

and

$$\|g\|_\infty := ess \sup_{t\in[a,x]} |g(t)|.$$

Let

$$P_n^{c_\lambda}(t) = \frac{(t-\theta(\lambda))^n}{n!},\ \theta(\lambda) = \lambda a + (1-\lambda)x, \quad \lambda \in [0,1], \tag{1.7}$$

represent polynomials involving; a convex combination of the end points. Although the methodology is applicable for general Appell type polynomials such as Euler and Bernoulli, only the above set will be considered in detail here. We should note that the dependence of the polynomials in (1.7), on x is not shown explicitly.

With the polynomials (1.7) then from (1.4) and (1.5)

$$\begin{aligned} &T_n^{c_\lambda}(f;a,x) \\ &\quad = f(a) + \sum_{k=1}^{n} (-1)^{k+1} \frac{(x-a)^k}{k!} \left[\lambda^k f^{(k)}(x) + (-1)^{k+1} (1-\lambda)^k f^{(k)}(a) \right], \end{aligned} \tag{1.8}$$

and the remainder

$$\begin{aligned} R_n^{c_\lambda}(f;a,x) &= \frac{(-1)^{n+1}}{n!} \int_a^x (t-\theta(\lambda))^n f^{(n+1)}(t)\, dt, \\ \theta(\lambda) &= \lambda a + (1-\lambda)x, \quad \lambda \in [0,1], \end{aligned} \tag{1.9}$$

The above expression (1.8) – (1.9) was obtained by Matić et al. [11] only for the equivalent of $\lambda = 0$ and $\frac{1}{2}$ in (1.7).

Cerone and Dragomir [4] obtained the following theorem which follows directly from Corollary 1 with $P_n^{c_\lambda}(t)$ as given by (1.7) and, using (1.8) and (1.9).

Theorem 2. *Assume that f is as in Theorem 1 with $x \geq a$, then we have*

$$
\begin{aligned}
(1.10)\qquad & \left|f(x) - T_n^{c_\lambda}(f;a,x)\right| \\
= \; & \left|R_n^{c_\lambda}(f;a,x)\right| \\
\leq \; & \begin{cases}
\frac{1}{n!}(x-a)^n\left[\frac{1}{2}+\left|\lambda-\frac{1}{2}\right|\right]^n \left\|f^{(n+1)}\right\|_1 \quad \text{if } \; f^{(n+1)} \in L_1[a,x]; \\[2ex]
\frac{1}{n!(nq+1)^{\frac{1}{q}}}(x-a)^{n+\frac{1}{q}}\left[(1-\lambda)^{nq+1}+\lambda^{nq+1}\right]^{\frac{1}{q}}\left\|f^{(n+1)}\right\|_p \\[1ex]
\qquad\qquad \text{if } \; f^{(n+1)} \in L_p[a,x], \;\; p>1, \; \frac{1}{p}+\frac{1}{q}=1; \\[2ex]
\frac{1}{(n+1)!}(x-a)^{n+1}\left[(1-\lambda)^{n+1}+\lambda^{n+1}\right]\left\|f^{(n+1)}\right\|_\infty, \\[1ex]
\qquad\qquad \text{if } \; f^{(n+1)} \in L_\infty[a,x].
\end{cases}
\end{aligned}
$$

It was also noted that since $h_1(\lambda) = \left[\frac{1}{2}+\left|\lambda-\frac{1}{2}\right|\right]^n$, $h_2(\lambda) = \left[(1-\lambda)^{nq+1}+\lambda^{nq+1}\right]^{\frac{1}{q}}$ and $h_3(\lambda) = (1-\lambda)^{n+1}+\lambda^{n+1}$ are convex and symmetric about $\frac{1}{2}$, then

$$\inf_{\lambda\in[0,1]} h_i(\lambda) = h_i\left(\frac{1}{2}\right), \quad i = 1,2,3.$$

Hence the best inequality possible in the class, in the sense of providing the tightest bound is obtained from taking $\lambda = \frac{1}{2}$.

Taking $\lambda = 0$ in (1.10) produces the classical Taylor series expansion in terms of the $L_p[a,x]$, $p \geq 1$, Lebesgue norms for the bounds (see for example, Dragomir [8]).

Recently in [7] Dragomir introduced a perturbed Taylor's formula using the Grüss inequality for the Čebyšev functional. Matić et al. [11] obtained generalised Taylor's formulae involving expansions in terms of general polynomials satisfying (1.2) producing in particular Theorem 1 and Corollary 1 above. They also examined perturbed versions of (1.3), namely

$$
\begin{aligned}
(1.11)\qquad f(x) = {} & T_n(f;a,x) \\
& + (-1)^n\left[P_{n+1}(x) - P_{n+1}(a)\right]\left[f^{(n)};a,x\right] + \rho_n(f;a,x),
\end{aligned}
$$

where

$$(1.12)\qquad \left[f^{(n)};a,x\right] := \frac{f^{(n)}(x)-f^{(n)}(a)}{x-a}, \quad \text{the divided difference,}$$

$$(1.13)\qquad \rho_n(f;a,x) \quad \text{is the remainder.}$$

Dragomir [7] developed an estimate of the remainder using the Grüss inequality for $P_n(t) = \frac{(t-x)^n}{n!}$, Matić et al. [11] used a premature or pre-Grüss argument to procure bounds on $\rho_n^{c_{\frac{1}{2}}}(f;a,x)$, $\rho_n^{c_0}(f;a,x)$. Dragomir [8] obtained tighter bounds for the same polynomial generators of the perturbed Taylor series for $f^{(n+1)} \in L_2(I)$ with $x,a \in I \subseteq \mathbb{R}$. In the paper [4], Cerone and Dragomir procured bounds on $\rho_n(f;a,x)$ in terms of $\Delta-$seminorms resulting from the Čebyšev functional and Korkine's identity.

In [7], S.S. Dragomir seems to be the first author to introduce the perturbed Taylor formula

$$f(x) = T_n(f;a,x) + \frac{(x-a)^{n+1}}{(n+1)!}\left[f^{(n)};a,x\right] + \rho_n(f;a,x), \tag{1.14}$$

where

$$T_n(f;a,x) = \sum_{k=0}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) \tag{1.15}$$

and $\left[f^{(n)};a,x\right]$ is as given in (1.12). Dragomir [7] estimated the remainder $\rho_n(f;a,x)$ by using Grüss and Čebyšev type inequalities.

In [11], the authors generalised and improved the results from [7] via a pre-Grüss inequality (see [11, Theorem 3]).

Theorem 3. *Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of polynomials satisfying (1.2). Let $I \subset \mathbb{R}$ be a closed interval and $a \in I$. Suppose $f : I \to \mathbb{R}$ is as in Theorem 1. Then for all $x \in I$ we have the perturbed generalised Taylor formula, (1.11), where $x \geq a$, the remainder $\rho(f;a,x)$ satisfies the estimate*

$$|\rho_n(f;a,x)| \leq \frac{x-a}{2}\sqrt{\mathcal{T}(P_n,P_n)}\left[\Gamma(x) - \gamma(x)\right], \tag{1.16}$$

provided that $f^{(n+1)}$ is bounded and

$$\gamma(x) := \inf_{t\in[a,x]} f^{(n+1)}(t) > -\infty, \quad \Gamma(x) := \sup_{t\in[a,x]} f^{(n+1)}(t) < \infty. \tag{1.17}$$

In (1.16), $\mathcal{T}(\cdot,\cdot)$ is the Čebyšev functional on the interval $[a,x]$. That is,

$$\mathcal{T}(g,h) := \frac{1}{x-a}\int_a^x g(t)\,h(t)\,dt - \frac{1}{x-a}\int_a^x g(t)\,dt \cdot \frac{1}{x-a}\int_a^x h(t)\,dt. \tag{1.18}$$

It is the intention in the current article to produce perturbed generalised Taylor series like (1.11), however the perturbation involves the arithmetic average of the upper and lower bounds of the $f^{(n+1)}(t)$, $t \in I$. The bounds for the expansion involve the norms of the Appell polynomials with the one norm, which is shown to provide the tightest bound.

A novel Čebyšev functional and its bounds are presented in Section 2, the results of which, are applied to perturbed generalised Taylor series. The approximation of the logarithmic function using the results of Section 2 is presented in Section 3. The results are demonstrated by estimating some key entities in information and coding theory. In particular, the Kullback-Leibler distance, Shannon's entropy and mutual information are approximated by rational functions providing sharp bounds. This application in Information Theory will be discussed in detail in Section 4.

2. The Čebyšev-like Functional and Perturbed Taylor Results with Bounds

Let $f, g : [a,b] \to \mathbb{R}$ be two integrable functions and define the functional

$$T(f,g;a,b) := \mathcal{M}(fg;a,b) - \mathcal{M}(f;a,b)\,\mathcal{M}(g;a,b), \tag{2.1}$$

where the integral mean is given by

$$\mathcal{M}(f;a,b) := \frac{1}{b-a}\int_a^b f(x)\,dx. \tag{2.2}$$

The functional $T(f,g;a,b)$ as defined in (2.1) – (2.2) is widely known in the literature as Čebyšev's functional. The reader is referred to [12], Chapters IX and X and, to the work by Dragomir [6] and Fink [10] for extensive treatments of the functional.

We now define a Čebyšev-like functional

$$C(f,g;a,b) := \mathcal{M}(fg;a,b) - \frac{M+m}{2}\mathcal{M}(f;a,b), \tag{2.3}$$

where $-\infty < m \le g(t) \le M < \infty$, for $t \in [a,b]$ and $\mathcal{M}(f;a,b)$ is as given by (2.2).

The following theorem holds providing bounds for the functional $C(f,g;a,b)$ (see Cerone [3]). The proof involves using the identity

$$\mathcal{M}(fg;a,b) - \frac{M+m}{2}\mathcal{M}(f;a,b) = \mathcal{M}\left(f\left(g(\cdot) - \frac{M+m}{2}\right)\right). \tag{2.4}$$

Theorem 4. *Let $f,g : [a,b] \to \mathbb{R}$ be integrable functions and $-\infty < m \le g(t) \le M < \infty$, then*

$$\begin{aligned} |C(f,g;a,b)| &= \left|\mathcal{M}(fg;a,b) - \frac{M+m}{2}\mathcal{M}(f;a,b)\right| \\ &\le \frac{M-m}{2}\cdot\frac{1}{b-a}\|f\|_1, \quad f \in L_1[a,b] \\ &\le \frac{M-m}{2}\cdot\frac{1}{(b-a)^{\frac{1}{p}}}\|f\|_p, \quad f \in L_p[a,b],\ 1<p<\infty \\ &\le \frac{M-m}{2}\|f\|_\infty = \frac{M-m}{2}\max\{|N|,|n|\}, \quad f \in L_\infty[a,b], \\ &-\infty < n \le f(t) \le N < \infty,\ t \in [a,b], \end{aligned} \tag{2.5}$$

where $\|f\|_p$ are the usual Lebesgue norms for $f \in L_p[a,b]$ defined by

$$\|f\|_p := \left(\int_a^b |f(t)|^p\,dt\right)^{\frac{1}{p}}, \quad 1 \le p < \infty$$

and

$$\|f\|_\infty := ess \sup_{t\in[a,b]} |f(t)|.$$

The $\frac{1}{2}$ in the three inequalities in (2.5) are sharp.

Remark 1. *The inequalities in (2.5) are in the order of increasing coarseness although each of them are sharp for $f \in L_p[a,b]$, $p \ge 1$.*

We now apply the above results to obtain sharp bounds for perturbed Taylor-like formulae.

Theorem 5. *Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of polynomials that satisfy (1.2). Further, let $I \subset \mathbb{R}$ be a closed interval and $a \in I$. If $f : I \to \mathbb{R}$ is a function which for some $n \in \mathbb{N}$, $f^{(n)}$ is absolutely continuous and $-\infty < \gamma_{n+1}(x) \le f^{(n+1)}(t) \le \Gamma_{n+1}(x) < \infty$, then for any $a \le x \in I$*

$$\begin{aligned} f(x) = T_n(f;a,x) + \left(\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right)(-1)^n \frac{[P_{n+1}(x) - P_{n+1}(a)]}{n+1} \\ + G_n(f;a,x) \end{aligned} \tag{2.6}$$

and

$$|G_n(f;a,x)| \tag{2.7}$$
$$\begin{aligned}
&\le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\left\|P_n\right\|_{1,[a,x]}, \qquad P_n \in L_1[I] \\
&\le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{1}{(x-a)^{\frac{1}{p}-1}}\left\|P_n\right\|_{p,[a,x]},\ P_n \in L_p[I],\ 1<p<\infty \\
&\le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\left\|P_n\right\|_{\infty,[a,x]}, \qquad P_n \in L_\infty[I],
\end{aligned}$$

where

$$G_n(f;a,x) = (-1)^n \int_a^x P_n(t)\left[f^{(n+1)}(t) - \frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right]dt \tag{2.8}$$

and $T_n(f;a,x)$ is as given by (1.4).

Proof. If we identify $(-1)^n P_n(t)$ with $f(t)$ and $f^{(n+1)}(t)$ with $g(t)$ in Theorem 4, then from identity (2.4) we have

$$\begin{aligned}
G_n(f;a,x) &= (-1)^n \int_a^x P_n(t)\left[f^{(n+1)}(t) - \frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right]dt \\
&= (x-a)\,C\left((-1)^n P_n(t), f^{(n+1)}(t); a, x\right).
\end{aligned} \tag{2.9}$$

That is,

$$\begin{aligned}
&G_n(f;a,x) \\
&= (-1)^n \int_a^x P_n(t) f^{(n+1)}(t)\,dt \\
&\qquad - (-1)^n \frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\int_a^x P_n(t)\,dt \\
&= R_n(f;a,x) - (-1)^n\left(\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right)\frac{[P_{n+1}(x)-P_{n+1}(a)]}{n+1},
\end{aligned} \tag{2.10}$$

where $R_n(f;a,x)$ satisfies (1.3) – (1.5). Using identity (1.3) in (2.10) produces the stated result (2.6).

For the bound on the remainder $|G_n(f;a,x)|$ we have from the first inequality in (2.5) and (2.9)

$$|G_n(f;a,x)| \le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\int_a^x |P_n(t)|\,dt,$$

and hence (2.7). Utilising the second and third inequality in (2.5) produces the respective bounds in (2.7). ∎

Corollary 2. *Let the conditions of Theorem 5 persist, and assume $a \le x$, then we have from (2.6) and (2.7), for $P_n^{c_\lambda}(t)$ given by (1.7)*

$$\begin{aligned}(2.11)\quad & \left| f(x) - T_n^{c_\lambda}(f;a,x) - \frac{1}{(n+1)!}\left[(\theta(\lambda)-a)^{n+1} \right.\right. \\ & \left.\left. + (-1)^n (x-\theta(\lambda))^{n+1}\right]\left(\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right)\right| \\ & \le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{1}{n!}\Psi_1(\lambda;a,x) \\ & \le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{1}{(x-a)^{\frac{1}{p}-1}}\cdot\frac{1}{n!}\left[\Psi_p(\lambda;a,x)\right]^{\frac{1}{p}} \\ & \le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{(x-a)}{n!}\cdot\left[\frac{x-a}{2}+\left|\theta(\lambda)-\frac{a+x}{2}\right|\right]^n,\end{aligned}$$

where

$$(2.12)\quad \begin{cases} \Psi_p(\lambda;a,x) = \dfrac{(x-\theta(\lambda))^{np+1}+(\theta(\lambda)-a)^{np+1}}{np+1} \\ \\ \theta(\lambda) = \lambda a + (1-\lambda)x, \quad \lambda\in[0,1]. \end{cases}$$

Proof. We need to evaluate for $P_n^{c_\lambda}(t)$ given by (1.7)

$$\int_a^x P_n^{c_\lambda}(t)\,dt \quad\text{and}\quad \int_a^x \left|P_n^{c_\lambda}(t)\right|^p dt,\quad p\ge 1.$$

Thus, from (1.7), we have,

$$\int_a^x P_n^{c_\lambda}(t)\,dt = \frac{1}{(n+1)!}\left[(x-\theta(\lambda))^{n+1}-(\theta(\lambda)-a)^{n+1}\right]$$

and

$$\begin{aligned}\int_a^x \left|P_n^{c_\lambda}(t)\right|^p dt &= \frac{1}{n!}\int_a^x |t-\theta(\lambda)|^n\,dt \\ &= \frac{1}{n!}\left[\int_a^{\theta(\lambda)}(\theta(\lambda)-t)^n\,dt + \int_{\theta(\lambda)}^b (t-\theta(\lambda))^n\,dt\right] \\ &= \frac{1}{(n+1)!}\left[(\theta(\lambda)-a)^{n+1}+(x-\theta(\lambda))^{n+1}\right],\end{aligned}$$

producing (2.11).

Further,

$$\begin{aligned}\left\|P_n^{c_\lambda}\right\|_{p,[a,x]}^p = \int_a^x \left|P_n^{c_\lambda}(t)\right|^p dt &= \frac{1}{n!}\int_a^x |t-\theta(\lambda)|^{np}\,dt \\ &= \frac{1}{n!}\left[\int_a^{\theta(\lambda)}(\theta(\lambda)-t)^{np}\,dt+\int_{\theta(\lambda)}^x (t-\theta(\lambda))^{np}\,dt\right] \\ &= \frac{1}{n!}\left[\frac{(\theta(\lambda)-a)^{np+1}+(x-\theta(\lambda))^{np+1}}{np+1}\right]\end{aligned}$$

and so from (2.7) the second inequality in (2.11) is procured.

The final inequality is obtained from (1.7) and (2.7) giving

$$\begin{aligned}\left\|P_n^{c_\lambda}\right\|_{\infty,[a,x]} &= ess \sup_{t\in[a,x]} \left|P_n^{c_\lambda}(t)\right| \\ &= ess \sup_{t\in[a,x]} \frac{\left|t-\theta(\lambda)\right|^n}{n!} \\ &= \frac{1}{n!}\left[\max\left\{x-\theta(\lambda),\theta(\lambda)-a\right\}\right]^n.\end{aligned}$$

∎

Remark 2. *The bounds in (2.11) are in order of increasing coarseness. This was commented upon in Remark 1 referring to the results (2.5), of which (2.11) is a specialisation. For* $\lambda = \frac{1}{2}$, $\theta\left(\frac{1}{2}\right) = \frac{a+x}{2}$ *then from (2.11) and (2.12)*

$$\begin{aligned}(2.13)\quad & \left|f(x) - T_n^{c_{\frac{1}{2}}}(f;a,x) - \frac{1+(-1)^n}{(n+1)!}\cdot\left(\frac{x-a}{2}\right)^{n+1}\left(\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right)\right| \\ &\le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{(x-a)^{n+1}}{2^n(n+1)!} \\ &\le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{(x-a)^{n+1}}{2^n(np+1)^{\frac{1}{p}}n!} \\ &\le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{(x-a)^{n+1}}{2^n n!}\end{aligned}$$

where, from (1.8),

$$(2.14)\qquad T_n^{c_{\frac{1}{2}}}(f;a,x) = f(a) + \sum_{k=1}^{n}\frac{(x-a)^k}{2^k k!}\left[f^{(k)}(a) + (-1)^{k+1}f^{(k)}(x)\right]$$

from which we may confirm, for this case at least, the fact that the bounds are in order of increasing coarseness since $\frac{1}{n+1} < \frac{1}{(np+1)^{\frac{1}{p}}} < 1$, $1<p<\infty$.

We further note that for n odd, the perturbation in (2.13) vanishes, giving the tightest bound

$$(2.15)\qquad \left|f(x) - T_n^{c_{\frac{1}{2}}}(f;a,x)\right| \le \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}\cdot\frac{(x-a)^{n+1}}{2^n(n+1)!},\ \ n \text{ odd}.$$

For n odd, the result (2.15) may be compared with the last result in (1.10) with $\lambda = \frac{1}{2}$, demonstrating that it is tighter since

$$\left\|f^{(n+1)}\right\|_\infty \ge \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2}$$

where $\gamma_{n+1}(x) \le f(t) \le \Gamma_{n+1}(x)$, $t\in[a,x]$.

For n even, the bound is still tighter in (2.13), however, the perturbation is now present.

If $\lambda = 0$ in (2.11) then $\theta(0) = x$ and we obtain a perturbed version of the traditional Taylor series expansion about a point a. That is, from (1.8),

$$\left| f(x) - T_n^{c_0}(f;a,x) - \frac{(x-a)^{n+1}}{(n+1)!}\left(\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right)\right| \tag{2.16}$$
$$= \left| f(x) - \sum_{k=0}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) - \frac{(x-a)^{n+1}}{(n+1)!}\left(\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}\right)\right|$$
$$\leq \frac{\Gamma_{n+1}(x)-\gamma_{n+1}(x)}{2} \cdot \frac{(x-a)^{n+1}}{(n+1)!},$$

where

$$T_n^{c_0}(f;a,x) = f(x) + \sum_{k=0}^{n} (-1)^{k+1} \frac{(x-a)^k}{k!} f^{(k)}(a).$$

We notice that the bound in (2.16) is inferior to that in the first inequality in (2.13) where $T_n^{c_{\frac{1}{2}}}(f;a,x)$ is as given by (2.14). However, (2.14) requires information involving $f^{(k)}(x)$ being available in order to approximate $f(x)$.

3. Applications to the Logarithm

We shall apply the results of the previous sections to the logarithm function to illustrate the results.

Let $f : (0,\infty) \to \mathbb{R}$, $f(t) = \ln t$ then

$$f^{(k)}(t) = (-1)^{k-1} \frac{(k-1)!}{t^k},\ t > 0,\ k \in \mathbb{N}. \tag{3.1}$$

We note that $f^{(n+1)}$ is strictly monotonic on $(0,\infty)$ so that

$$\frac{\gamma_{n+1}(x) \mp \Gamma_{n+1}(x)}{2} \tag{3.2}$$
$$= \frac{1}{2}\left[\max\left\{f^{(n+1)}(a), f^{(n+1)}(x)\right\} \mp \min\left\{f^{(n+1)}(a), f^{(n+1)}(x)\right\}\right]$$
$$= \frac{1}{2}\left|f^{(n+1)}(a) \mp f^{(n+1)}(x)\right|$$
$$= \frac{n!}{2}\left(\frac{1}{a^{n+1}} \mp \frac{1}{x^{n+1}}\right), \quad \text{for} \quad x \geq a.$$

We shall utilise the first inequality in (2.11), which is the least coarse of the three, to illustrate the results to give

$$\Bigg| f(x) - T_n^{c_\lambda}(\ln;a,x) - \frac{1}{2(n+1)}\left[(\theta(\lambda)-a)^{n+1} + (-1)^n (x-\theta(\lambda))^{n+1}\right]\Bigg| \tag{3.3}$$
$$\leq \frac{1}{2(n+1)}\left(\frac{1}{a^{n+1}} - \frac{1}{x^{n+1}}\right)\left[(\theta(\lambda)-a)^{n+1} + (x-\theta(\lambda))^{n+1}\right],$$

where $\theta(\lambda) = \lambda a + (1-\lambda)x$, $\lambda \in [0,1]$, and from (1.8),

$$T_n^{c_\lambda}(\ln;a,x) = \ln a + \sum_{k=1}^{n} \frac{(x-a)^k}{k}\left[\left(\frac{\lambda}{x}\right)^k + (-1)^{k+1}\left(\frac{1-\lambda}{a}\right)^k\right].$$

Simplification of (3.1) gives

$$\left|\ln x-\ln a-\sum_{k=1}^{n}\frac{(x-a)^{k}}{k}\left[\left(\frac{\lambda}{x}\right)^{k}+(-1)^{k+1}\left(\frac{1-\lambda}{a}\right)^{k}\right]\right. \\ \left.-\frac{(x-a)^{n+1}}{2(n+1)}\left[(1-\lambda)^{n+1}+\lambda^{n+1}\right]\left(\frac{1}{a^{n+1}}+\frac{1}{x^{n+1}}\right)\right| \\ \leq\frac{(x-a)^{n+1}}{2(n+1)}\left[(1-\lambda)^{n+1}+\lambda^{n+1}\right]\left(\frac{1}{a^{n+1}}-\frac{1}{x^{n+1}}\right). \tag{3.4}$$

where the sharpest bound results from taking $\lambda=\frac{1}{2}$.

In particular, taking $\lambda=0$ gives

$$\left|\ln x-\ln a-\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k}\left(\frac{x-a}{a}\right)^{k}-\frac{(x-a)^{n+1}}{2(n+1)}\left(\frac{1}{a^{n+1}}+\frac{1}{x^{n+1}}\right)\right| \\ \leq\frac{(x-a)^{n+1}}{2(n+1)}\left(\frac{1}{a^{n+1}}-\frac{1}{x^{n+1}}\right):=R, \tag{3.5}$$

which may be compared with the result (see case in Section 4 of Matić et al. [11])

$$\ln x=\ln a+\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k}\left(\frac{x-a}{a}\right)^{k}+(-1)^{n}\frac{(x-a)^{n}}{n(n+1)}\left(\frac{1}{a^{n}}-\frac{1}{x^{n}}\right)+B, \tag{3.6}$$

where $|B|\leq\frac{n}{\sqrt{2n+1}}R$.

It may be seen that the bound, R, obtained here, is much tighter, especially for large n since $\frac{n}{\sqrt{2n+1}}>1$. It must be remembered, however, that a different perturbation is present in (3.5) than in (3.6). There is very little difference in complexity between the two perturbations. The perturbation used in [11] from the traditional Čebyšev functional (2.1) giving rise to (1.11) rather than the novel Čebyšev functional (2.3) producing (2.6). That is, the perturbation in (1.11) involves $\left[f^{(n+1)};a,x\right]=\frac{f^{(n)}(x)-f^{(n)}(a)}{x-a}$ whereas that in (2.6) contains $\frac{\gamma_{n+1}(x)+\Gamma_{n+1}(x)}{2}$, where $\gamma_{n+1}(x)\leq f^{(n+1)}(t)\leq\Gamma_{n+1}(x)$.

4. Approximation of the Kullback-Leibler Distance and Related Measures

The results of Section 3 will now be applied to approximate the *Kullback-Leibler distance* or *relative entropy* $D(p||q)$ where for probability mass functions $p(x)$, $q(x)$

$$D(p||q):=\sum_{x\in\chi}p(x)\ln\left(\frac{p(x)}{q(x)}\right), \tag{4.1}$$

with $0\ln\left(\frac{0}{q}\right)=0$ and $p\ln\left(\frac{p}{0}\right)=\infty$,

The relative entropy $D(p||q)$ is a measure of inefficiency in assuming a distribution $q(\cdot)$ when the true distribution is $p(\cdot)$. In coding theory, for example, if the true distribution of a random variable X with mass function $p(x)$, $x\in\chi$, were known, then a code could be constructed with average description length given by the *Shannon entropy*

$$H(p) := -\sum_{x\in\chi} p(x)\ln p(x). \tag{4.2}$$

If the code for a mass function $q(x)$ were used instead, then on average $H(p) + D(p\|q)$ bits would be required to describe the random variable [5, p. 18]. The *mutual information,* $I(X,Y)$ is a measure of the amount of information contained by one random variable about the other. It represents the effect in reducing uncertainty of one random variable from knowledge of the other. *Mutual information* is defined by

$$I(X,Y) := D(\rho(x,y)\,\|p(x)\,q(y)), \tag{4.3}$$

where $\rho(x,y)$ is the joint probability mass function and $p(x), q(y)$ are its marginals.

Theorem 6. *Let X and Y be random variables with probability mass functions $p(x)$ and $q(x)$ respectively with $x\in\chi$. Then we have the following approximation for $D(p\|q)$. Namely,*

$$\begin{aligned}(4.4)\quad &\Bigg|D(p\|q) - \sum_{x\in\chi} p(x)\sum_{k=1}^{n}\frac{(p(x)-q(x))^k}{k}\left[\left(\frac{\lambda}{p(x)}\right)^k + (-1)^{k+1}\left(\frac{1-\lambda}{q(x)}\right)^k\right] \\ &-\frac{(1-\lambda)^{n+1}+\lambda^{n+1}}{2(n+1)}\sum_{x\in\chi} p(x)(p(x)-q(x))^{n+1}\left(\frac{1}{q^{n+1}(x)}+\frac{1}{p^{n+1}(x)}\right)\Bigg| \\ &\le \frac{(1-\lambda)^{n+1}+\lambda^{n+1}}{2(n+1)}\sum_{x\in\chi} p(x)\,|p(x)-q(x)|^{n+1}\left|\frac{1}{q^{n+1}(x)}-\frac{1}{p^{n+1}(x)}\right|.\end{aligned}$$

Proof. From (3.4) we have for all $a,b>0$

$$\begin{aligned}(4.5)\quad &\Bigg|\ln b - \ln a - \sum_{k=1}^{n}\frac{(b-a)^k}{k}\left[\left(\frac{\lambda}{b}\right)^k + (-1)^{k+1}\left(\frac{1-\lambda}{a}\right)^k\right] \\ &- \frac{(1-\lambda)^{n+1}+\lambda^{n+1}}{2(n+1)}(b-a)^{n+1}\left(\frac{1}{a^{n+1}}+\frac{1}{b^{n+1}}\right)\Bigg| \\ &\le \frac{(1-\lambda)^{n+1}+\lambda^{n+1}}{2(n+1)}|b-a|^{n+1}\left|\frac{1}{a^{n+1}}-\frac{1}{b^{n+1}}\right|,\end{aligned}$$

where the modulus in the bound has been included to accommodate the lack of knowledge about the relative position of a and b.

Choose $a = q(x)$ and $b = p(x)$ to give, after multiplication by $p(x)$ and summation over $x\in\chi$, the stated expression (4.4). ∎

Corollary 3. *Let the conditions of Theorem 6 hold, then*

$$\begin{aligned}(4.6)\quad &\Bigg| D\left(p\|q\right)-\sum_{x\in\chi}p\left(x\right)\sum_{k=1}^{n}\frac{\left(p\left(x\right)-q\left(x\right)\right)^{k}}{k\cdot 2^{k}}\left[\frac{1}{p^{k}\left(x\right)}+\frac{\left(-1\right)^{k+1}}{q^{k}\left(x\right)}\right]\\ &-\frac{1}{\left(n+1\right)2^{n+1}}\sum_{x\in\chi}p\left(x\right)\left(p\left(x\right)-q\left(x\right)\right)^{n+1}\left(\frac{1}{q^{n+1}\left(x\right)}+\frac{1}{p^{n+1}\left(x\right)}\right)\Bigg|\\ &\leq\frac{1}{\left(n+1\right)2^{n+1}}\sum_{x\in\chi}p\left(x\right)\left|p\left(x\right)-q\left(x\right)\right|^{n+1}\left|\frac{1}{q^{n+1}\left(x\right)}-\frac{1}{p^{n+1}\left(x\right)}\right|.\end{aligned}$$

Proof. Taking $\lambda=\frac{1}{2}$ provides the tightest bound in (4.4) giving the result (4.6). ∎

Corollary 4. *For the conditions of Theorem 6 continuing to hold then*

$$0\leq D\left(p\|q\right)\leq\sum_{x\in\chi}\frac{p^{2}\left(x\right)}{q\left(x\right)}-1, \tag{4.7}$$

$$\begin{aligned}(4.8)\quad &\Bigg| D\left(p\|q\right)-\left(1-\lambda\right)\left[\sum_{x\in\chi}p\left(x\right)\left(\frac{p\left(x\right)}{q\left(x\right)}\right)-1\right]\\ &-\frac{1}{2}\left[\frac{1}{4}+\left(\lambda-\frac{1}{2}\right)^{2}\right]\sum_{x\in\chi}p\left(x\right)\left(p\left(x\right)-q\left(x\right)\right)^{2}\left[\frac{1}{q^{2}\left(x\right)}+\frac{1}{p^{2}\left(x\right)}\right]\Bigg|\\ &\leq\frac{1}{2}\left[\frac{1}{4}+\left(\lambda-\frac{1}{2}\right)^{2}\right]\sum_{x\in\chi}p\left(x\right)\left(p\left(x\right)-q\left(x\right)\right)^{2}\left|\frac{1}{q^{2}\left(x\right)}-\frac{1}{p^{2}\left(x\right)}\right|.\end{aligned}$$

Proof. Taking $n=0$ in (4.4) recaptures the result (4.7) of [9] and [5]. The result (4.8) follows after some simplification for (4.4) on taking $n=1$. ∎

Remark 3. *Equation (4.8) may be rearranged to give upper and lower bounds for $D\left(p\|q\right)$ in terms of rationals functions of $p\left(x\right)$ and $q\left(x\right)$.*

The following theorem provides approximations for the Shannon entropy $H\left(p\right)$, as defined by (4.2) together with bounds.

Theorem 7. *Let X be a random variable with probability mass function $p\left(x\right)$ for $x\in\chi$*

$$\begin{aligned}(4.9)\quad &\Bigg| H\left(p\right)-\ln\left|\chi\right|+\sum_{x\in\chi}\sum_{k=1}^{n}\frac{\left(\left|\chi\right|p\left(x\right)-1\right)^{k}}{k\left|\chi\right|^{k}p^{k-1}\left(x\right)}\left[\lambda^{k}+\left(-1\right)^{k+1}\left|\chi\right|^{k}\left(1-\lambda\right)^{k}\right]\\ &+\frac{\left(1-\lambda\right)^{n+1}+\lambda^{n+1}}{2\left(n+1\right)}\sum_{x\in\chi}\frac{\left(\left|\chi\right|p\left(x\right)-1\right)^{n+1}}{\left|\chi\right|^{2\left(n+1\right)}p^{n}\left(x\right)}\left(p^{n+1}\left(x\right)+\left|\chi\right|^{n+1}\right)\Bigg|\\ &\leq\frac{\left(1-\lambda\right)^{n+1}+\lambda^{n+1}}{2\left(n+1\right)}\sum_{x\in\chi}\frac{\left|\left|\chi\right|p\left(x\right)-1\right|^{n+1}}{\left|\chi\right|^{2\left(n+1\right)}p^{n}\left(x\right)}\left|p^{n+1}\left(x\right)+\left|\chi\right|^{n+1}\right|.\end{aligned}$$

Proof. From (4.1) and (4.2) we note that

$$H\left(p\right)=\ln\left|\chi\right|-D\left(p\|u\right),$$

where $u(x) = \frac{1}{|\chi|}$, the uniform probability mass function on χ. Thus, from (4.4) taking $q(x) = u(x) = \frac{1}{|\chi|}$ gives (4.9) upon some simplification. ∎

The following theorem gives an approximation to the *mutual information* as defined by (4.3) with (4.1).

Namely, for random variables X and Y with joint probability mass function $\rho(x,y)$ and marginals $p(x)$, $q(y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

$$I(X,Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \rho(x,y) \ln \frac{\rho(x,y)}{p(x)\,q(y)}. \tag{4.10}$$

Theorem 8. *With the above assumptions regarding X and Y we have*

$$\begin{aligned} &\Bigg| I(X,Y) - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \rho(x,y) \sum_{k=1}^{n} \rho(x,y) \frac{(\rho(x,y) - p(x)\,q(y))^k}{k} \\ &\quad \times \left[\left(\frac{\lambda}{\rho(x,y)} \right)^k + (-1)^{k+1} \left(\frac{1-\lambda}{p(x)\,q(y)} \right)^k \right] - \frac{(1-\lambda)^{n+1} + \lambda^{n+1}}{2(n+1)} \\ &\quad \times \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \rho(x,y) (\rho(x,y) - p(x)\,q(y))^{n+1} \left(\frac{1}{p^n(x)\,q^n(x)} + \frac{1}{\rho^n(x,y)} \right) \Bigg| \\ &\leq \frac{(1-\lambda)^{n+1} + \lambda^{n+1}}{2(n+1)} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \rho(x,y) \left| \rho(x,y) - p(x)\,q(y) \right|^{n+1} \\ &\qquad \times \left| \frac{1}{p^{n+1}(x)\,q^{n+1}(x)} - \frac{1}{\rho^{n+1}(x,y)} \right|. \end{aligned} \tag{4.11}$$

Proof. Choose $a = p(x)\,q(y)$ and $h = \rho(x,y)$ and $b = \rho(x,y)$ to give, from (4.5), after multiplication by $\rho(x,y)$ and summation over $x \in \mathcal{X}$, and $y \in \mathcal{Y}$, the result as stated, (4.11). ∎

5. Concluding Remarks

A Čebyšev-type functional introduced in [3] has been applied to obtain perturbed generalised Taylor series together with sharp bounds of the approximations. The approximation of the logarithmic function is given as an example which is further used to estimate the Kullback-Leibler distance, Shannon entropy and mutual information. These are of importance in information theory and, in particular, coding theory.

Acknowledgement 1. *The author would like to thank Professor S.S. Dragomir for bringing the paper* [9] *to his attention and for the useful discussions.*

References

[1] P. APPELL, Sur une classe de polnômes, *Ann. Sci. Ecole Norm. Sup.,* **9**(2) (1880), 119-144.

[2] P. CERONE, Generalised Taylor's formula with estimates of the remainder, submitted to *Inequality Theory and Applications,* (2002). [ONLINE: `http://rgmia.vu.edu.au/v5n2.html`]

[3] P. CERONE, Perturbed generalised Taylor's formulae with sharp bounds, *RGMIA Res. Rep. Coll.,* **5**(4) (2002), Article . [ONLINE `http://rgmia.vu.edu.au/v5n4.html`]

[4] P. CERONE and S.S. DRAGOMIR, New bounds for a perturbed generalised Taylor's formula, *East Asian Mathematical Journal,* **17**(2) (2001), 197-215.

[5] T.M. COVER and J.A. THOMAS, Elements of Information Theory, John Wiley & Son Inc., 1991.

[6] S.S. DRAGOMIR, Some integral inequalities of Grüss type, *Indian J. of Pure and Appl. Math.,* **31**(4) (2000), 397-415.

[7] S.S. DRAGOMIR, New estimation of the remainder in Taylor's formula using Grüss type inequalities and applications, *Math. Ineq. & Appl.,* **2**(2) (1999), 183-193.

[8] S.S. DRAGOMIR, An improvement of the remainder estimate in the generalised Taylor's formula, *RGMIA Res. Rep. Coll.,* **3**(1) (2000), Article 1.

[9] S.S. DRAGOMIR and V. GLUŠČEVIĆ, New estimates of the Kullback-Leibler distance and applications, *Inequality Theory and Applications - Volume 1*, Y.J. Cho, J.K. Kim and S.S. Dragomir (Eds.), Nova Science Publishers, Inc., Huntington, New York, 2001.

[10] A.M. FINK, A treatise on Grüss inequality, *Analytic and Geometric Inequalities and Applications, Math. Appl.,* **478** (1999), Kluwer Academic Publishers, Dordrecht, 93-114.

[11] M. MATIĆ, J.E. PEČARIĆ and N. UJEVIĆ, On new estimation of the remainder in generalised Taylor's formula, *Math. Ineq. & Appl.,* **2**(3) (1999), 343-361.

[12] D.S. MITRINOVIĆ, J.E. PEČARIĆ and A.M. FINK, *Classical and New Inequalities in Analysis,* Kluwer Academic Publishers, 1993.

[13] J.E. PEČARIĆ, F. PROCHAN and Y.L. TONG, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press, 1992.

School of Communications and Informatics, Victoria University of Technology, PO Box 14428, MCMC 8001, VIC, Australia.

E-mail address: `pc@matilda.vu.edu.au`

URL: `http://rgmia.vu.edu.au/sofo`