Acta Appl Math (2009) 106: 455–472
DOI 10.1007/s10440-008-9307-2

Gaussian DCT Coefficient Models

Saralees Nadarajah

Received: 5 May 2008 / Accepted: 21 August 2008 / Published online: 11 September 2008

Abstract It has been known that the distribution of the discrete cosine transform (DCT) coefficients of most natural images follow a Laplace distribution. However, recent work has shown that the Laplace distribution may not be a good fit for certain type of images and that the Gaussian distribution will be a realistic model in such cases. Assuming this alternative model, we derive a comprehensive collection of formulas for the distribution of the actual DCT coefficient. The corresponding estimation procedures are derived by the method of moments and the method of maximum likelihood. Finally, the superior performance of the derived distributions over the Gaussian model is illustrated. It is expected that this work could serve as a useful reference and lead to improved modeling with respect to image analysis and image coding.

Keywords Discrete cosine transform (DCT) · Gaussian distribution · Generalized hypergeometric function · Image analysis · Image coding · Incomplete gamma function · Kummer function · Modified Bessel function

1 Introduction

A discrete cosine transform (DCT) expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. The DCT was first introduced by Ahmed et al. [1]. Later Wang and Hunt [2] introduced a complete set of variants of the DCT. The DCT is included in many mathematical packages, such as Matlab, Mathematica and GNU Octave.

DCTs are important to numerous applications in science and engineering, from lossy compression of audio and images (where small high-frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations to Chebyshev approximation of arbitrary functions by series of Chebyshev polynomials. The use of cosine rather than sine functions is critical in these applications: for compression, it

S. Nadarajah (✉)
University of Manchester, Manchester M13 9PL, UK
e-mail: saralees.nadarajah@manchester.ac.uk

turns out that cosine functions are much more efficient, whereas for differential equations the cosines express a particular choice of boundary conditions. We refer the readers to Jain [3] and Rao and Yip [4] for comprehensive accounts of the theory and applications of the DCT. For a tutorial account see Duhamel and Vetterli [5].

The DCT is widely used in image coding and processing systems, especially for lossy data compression, because it has a strong “energy compaction” property (Rao and Yip [4]). DCT coding relies on the premise that pixels in an image exhibit a certain level of correlation with their neighboring pixels. Similarly in a video transmission system, adjacent pixels in consecutive frames show very high correlation. (Frames usually consist of a representation of the original data to be transmitted, together with other bits which may be used for error detection and control. In simplistic terms, frames can be referred to as consecutive images in a video transmission.) Consequently, these correlations can be exploited to predict the value of a pixel from its respective neighbors. The DCT is, therefore, defined to map this spatial correlated data into transformed uncorrelated coefficients. Clearly, the DCT utilizes the fact that the information content of an individual pixel is relatively small, i.e. to a large extent visual contribution of a pixel can be predicted using its neighbors.

A typical image/video transmission system is outlined in Fig. 1. The objective of the source encoder is to exploit the redundancies in image data to provide compression. In other words, the source encoder reduces the entropy, which in our case means decrease in the average number of bits required to represent the image. On the contrary, the channel encoder adds redundancy to the output of the source encoder in order to enhance the reliability of the transmission. The source encoder has three sub-blocks: the transformation sub-block, quantizer sub-block and the entropy sub-block. The transformation sub-block refers to the DCT. The quantizer sub-block utilizes the fact that the human eye is unable to perceive some visual information in an image. Such information is deemed redundant and can be discarded without introducing noticeable visual artifacts. The entropy encoder employs its knowledge

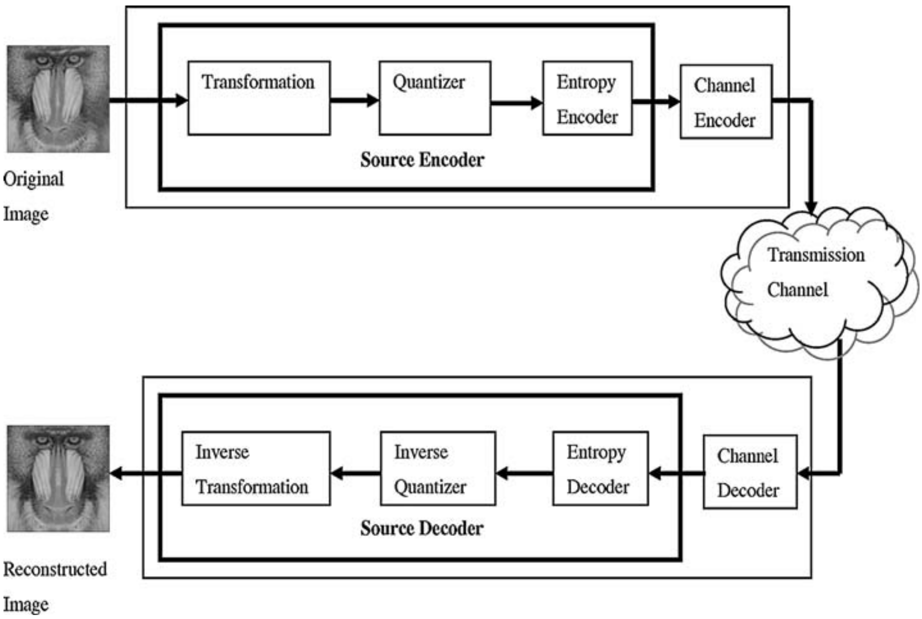

Fig. 1 Components of a typical image/video transmission system

of the DCT and quantization processes to reduce the number of bits required to represent each symbol at the quantizer output. The source and channel decoders reconstruct the image by doing all the above operations in reverse.

Efficient encoders and decoders are based on source models. Reininger and Gibson [6] used Kolmogorov–Smirnov tests to show that most DCT coefficients are reasonably modeled by the Laplace distribution. This knowledge has been used to improve decoder design. However, recent work has shown that the Laplace model does not give a good fit for certain images such as text documents (see Lam [7] and Lam and Goodman [8]). In these cases, it is suggested that the Gaussian distribution is a realistic model. In fact, using a doubly stochastic model, Lam and Goodman [8] argue that within an 8×8 block used for the DCT, assuming that the pixels are identically distributed, the DCT coefficient is approximately Gaussian. The probability density function (pdf) of the Gaussian distribution is given by

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) \tag{1}$$

for $-\infty < y < \infty$, where $\sigma^2 > 0$ represents the variance of the block.

When modeling the DCT densities of a big image, the block variance σ^2 in (1) is likely to vary over different parts of the image and so one should consider σ itself to be a random variable. This means that the actual DCT coefficient distribution will be given by the compound form:

$$f(y) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) g(\sigma^2) d\sigma^2, \tag{2}$$

where $g(\cdot)$ denotes the pdf of σ^2. For convenience, letting $x = y^2/2$ and $\lambda = \sigma^2$, one can rewrite (2) as

$$f(y) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{\lambda}} \exp\left(-\frac{x}{\lambda}\right) g(\lambda) d\lambda. \tag{3}$$

A number of forms for $g(\cdot)$ has been used in the literature. Lam [7], Lam and Goodman [8] and Lam [9] considered $g(\cdot)$ to have the exponential distribution. Lam [7], Teichroew [10] and Nadarajah and Kotz [11] considered $g(\cdot)$ to have the gamma distribution. The uniform distribution has also been used.

The aim of this note is to derive the most comprehensive list of forms for (3) by taking $g(\cdot)$ to belong to some sixteen flexible families. For each $g(\cdot)$, we derive the corresponding $f(\cdot)$ given by (3) as well as provide estimators of the associated parameters obtained by the method of moments and the method of maximum likelihood, see Sect. 2. An application of the derived distributions is illustrated in Sect. 3. It is shown that they are better models for the DCT coefficients than the Gaussian distribution given by (1).

The calculations of this note use several special functions, including the incomplete gamma function defined by

$$\gamma(a, x) = \int_0^x t^{a-1} \exp(-t) dt,$$

the modified Bessel function of the third kind defined by

$$K_\nu(x) = \frac{x^\nu \Gamma(1/2)}{2^\nu \Gamma(\nu + 1/2)} \int_1^\infty \exp(-xt)(t^2 - 1)^{\nu - 1/2} dt,$$

the generalized hypergeometric function defined by

$$ {}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \cdots (a_p)_k}{(b_1)_k (b_2)_k \cdots (b_q)_k} \frac{x^k}{k!}, $$

and, the Kummer function defined by

$$ \Psi(a, b; x) = \frac{\Gamma(1-b)}{\Gamma(1+a-b)} {}_1F_1(a; b; x) + \frac{\Gamma(b-1)}{\Gamma(a)} x^{1-b} {}_1F_1(1+a-b; 2-b; x), $$

where $(f)_k = f(f+1)\cdots(f+k-1)$ denotes the ascending factorial. The properties of these special functions can be found in Prudnikov et al. [12] and Gradshteyn and Ryzhik [13].

The estimation by the method of moments and the method of maximum likelihood requires the following notation: for a random sample $\lambda_1, \ldots, \lambda_n$ of λ, define

$$ \beta_1 = \frac{\{2\mu_1^3 - 3\mu_1\mu_2 + \mu_3\}^2}{\{\mu_2 - \mu_1^2\}^3} \tag{4} $$

and

$$ \beta_2 = \frac{-3\mu_1^4 + 6\mu_1^2\mu_2 - 4\mu_1\mu_3 + \mu_4}{\{\mu_2 - \mu_1^2\}^2} \tag{5} $$

as the sample skewness and sample kurtosis, respectively, where μ_j is the jth sample moment defined by $\mu_j = (1/n)\sum_{i=1}^n \lambda_i^j$ for $j = 1, 2, 3, 4$.

2 Models for Actual DCT Coefficient

In this section, we provide a collection of formulas for $f(\cdot)$ in (3) by taking $g(\cdot)$ to belong to sixteen flexible families. The estimators for the parameters of $f(\cdot)$ determined by the method of moments and the method of maximum likelihood are also given.

One Parameter Exponential Distribution: If g takes the form

$$ g(\lambda) = (1/\mu)\exp(-\lambda/\mu) $$

for $\lambda > 0$ and $\mu > 0$ then

$$ f(y) = \frac{\sqrt{2}x^{1/4}}{\sqrt{\pi}\mu^{3/4}} K_{1/2}\left(2\sqrt{\frac{x}{\mu}}\right). $$

Note that μ is the scale parameter. The moment estimator of μ is μ_1, the sample mean. This is also the maximum likelihood estimator of μ.

Two Parameter Gamma Distribution: If g takes the form

$$ g(\lambda) = \frac{\lambda^{\beta-1}\exp(-\lambda/\mu)}{\mu^\beta \Gamma(\beta)} $$

for $\lambda > 0$, $\beta > 0$ and $\mu > 0$ then

$$f(y) = \frac{\sqrt{2} x^{\beta/2-1/4}}{\sqrt{\pi}\Gamma(\beta)\mu^{(1/2+\beta)/2}} K_{\beta-1/2}\left(2\sqrt{\frac{x}{\mu}}\right). \tag{6}$$

Note that β and μ are the shape and scale parameters, respectively. The moment estimators of μ and β are $\mu_1\beta_1/4$ and $4/\beta_1$, respectively, where μ_1 is the sample mean and β_1 is the sample skewness given by (4). The maximum likelihood estimator of β is the solution of the equation:

$$n \log \mu_1 - n \log \beta + n\psi(\beta) = \sum_{i=1}^{n} \log \lambda_i,$$

where $\psi(x) = d \log \Gamma(x)/dx$ is the digamma function. The maximum likelihood estimator of μ is $\mu_1/\widehat{\beta}$.

One Parameter Half Logistic Distribution: If g takes the form

$$g(\lambda) = \frac{2\mu \exp(-\lambda\mu)}{\{1 + \exp(-\lambda\mu)\}^2}$$

for $\lambda > 0$ and $\mu > 0$ then

$$f(y) = \frac{2\sqrt{2}\mu^{3/4} x^{1/4}}{\sqrt{\pi}} \sum_{k=0}^{\infty} \binom{-2}{k} (k+1)^{-1/4} K_{1/2}\left(2\sqrt{\mu(k+1)x}\right).$$

Note that μ is the scale parameter. The moment estimator of μ is $(2\ln 2/\mu_1)^{1/3}$, where μ_1 is the sample mean. The maximum likelihood estimator of μ is the solution of the equation:

$$\frac{1}{\mu} + \frac{2}{n} \sum_{i=1}^{n} \frac{\lambda_i \exp(-\mu\lambda_i)}{1 + \exp(-\mu\lambda_i)} = \mu_1.$$

Two Parameter Inverse Gaussian Distribution: If g takes the form

$$g(\lambda) = \sqrt{\frac{\mu\phi}{2\pi}} \exp(\phi) \lambda^{-3/2} \exp\left\{-\frac{\phi}{2}\left(\frac{\lambda}{\mu} + \frac{\mu}{\lambda}\right)\right\}$$

for $\lambda > 0$, $\phi > 0$ and $\mu > 0$ then

$$f(y) = \frac{\sqrt{\phi} \exp(\phi)}{\pi} \left(1 + \frac{2x}{\mu\phi}\right)^{-1} K_{-1}\left(\sqrt{\phi\left(\phi + \frac{2x}{\mu}\right)}\right).$$

Note that both ϕ and μ are the scale parameters. The moment estimators of ϕ and μ are $9/\beta_1$ and μ_1, respectively, where μ_1 is the sample mean and β_1 is the sample skewness given by (4). The maximum likelihood estimators of μ and ϕ are the simultaneous solutions of the equations:

$$\frac{1}{2\mu} = \frac{\phi}{2n} \sum_{i=1}^{n} \frac{1}{\lambda_i} - \frac{\phi\mu_1}{2\mu^2}$$

and

$$1 + \frac{1}{2\phi} = \frac{\mu_1}{2\mu} + \frac{\mu}{2n} \sum_{i=1}^{n} \frac{1}{\lambda_i}.$$

Two Parameter Weibull Distribution: If g takes the form

$$g(\lambda) = \beta \lambda^{\beta-1} \mu^{-\beta} \exp\{-(\lambda/\mu)^{\beta}\}$$

for $\lambda > 0$, $\beta > 0$ and $\mu > 0$ then

$$f(y) = \frac{\beta x^{\beta-1/2}}{\mu^{\beta}\sqrt{2\pi}} \left[\sum_{j=0}^{q-1} \frac{(-A)^j}{j!} \Gamma(1/2 - \beta - \beta j) C_{1j} + \sum_{h=0}^{p-1} \frac{(-1)^h A^{(1/2-\beta+h)/\beta}}{h!\beta} \Gamma\left(\frac{\beta - 1/2 - h}{\beta}\right) C_{2h} \right]$$

provided that $\beta = p/q$ where $p \geq 1$ and $q \geq 1$ are co-prime integers,

$$C_{1j} = {}_1F_{p+q}(1; \Delta(p, 1/2 + \beta + \beta j), \Delta(q, 1 + j); z)$$

and

$$C_{2h} = {}_1F_{p+q}\left(1; \Delta\left(q, \frac{1/2 + h}{\beta}\right), \Delta(p, 1 + h); z\right).$$

Furthermore, $A = (x/\mu)^{\beta}$, $z = (-1)^{p+q} A^q / \{p^p q^q\}$ and $\Delta(k, a) = (a/k, (a+1)/k, \ldots, (a+k-1)/k)$. Note that β and μ are the shape and scale parameters, respectively. The moment estimator of β is the root of the equation

$$\sqrt{\beta_1} \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right\}^{3/2} = \Gamma\left(1 + \frac{3}{\beta}\right) - 3\Gamma\left(1 + \frac{1}{\beta}\right)\Gamma\left(1 + \frac{2}{\beta}\right) + 2\Gamma^3\left(1 + \frac{1}{\beta}\right),$$

where β_1 is the sample skewness given by (4). The moment estimator of μ is $\mu_1\{\Gamma(1 + 1/\beta)\}^{-1}$, where μ_1 is the sample mean. The maximum likelihood estimators of μ and β are the simultaneous solutions of the equations:

$$\frac{n}{\beta} + \sum_{i=1}^{n} \log \lambda_i = n \log \mu + \sum_{i=1}^{n} \left(\frac{\lambda_i}{\mu}\right)^{\beta} \log\left(\frac{\lambda_i}{\mu}\right)$$

and

$$\frac{n\beta}{\mu} = \beta \mu^{-\beta-1} \sum_{i=1}^{n} \lambda_i^{\beta}.$$

Three Parameter Stacy Distribution: If g takes the form

$$g(\lambda) = \frac{c\lambda^{c\gamma-1}\exp\{-(\lambda/\beta)^c\}}{\beta^{c\gamma}\Gamma(\gamma)}$$

for $\lambda > 0$, $c > 0$, $\gamma > 0$ and $\beta > 0$ then

$$f(y) = \frac{cx^{c\gamma-1/2}}{\beta^{c\gamma}\sqrt{2\pi}\Gamma(\gamma)}\left[\sum_{j=0}^{q-1}\frac{(-A)^j}{j!}\Gamma(1/2-c\gamma-cj)C_{3j}\right.$$
$$\left.+\sum_{h=0}^{p-1}\frac{(-1)^h A^{(1/2-c\gamma+h)/c}}{h!c}\Gamma\left(\frac{c\gamma-1/2-h}{c}\right)C_{4h}\right]$$

provided that $c = p/q$ where $p \geq 1$ and $q \geq 1$ are co-prime integers,

$$C_{3j} = {}_1F_{p+q}(1; \Delta(p, 1/2+c\gamma+cj), \Delta(q, 1+j); z)$$

and

$$C_{4h} = {}_1F_{p+q}\left(1; \Delta\left(q, 1-c+\frac{1/2+h}{c}\right), \Delta(p, 1+h); z\right).$$

Furthermore, $A = (x/\beta)^c$, $z = (-1)^{p+q}A^q/\{p^p q^q\}$ and $\Delta(k, a) = (a/k, (a+1)/k, \ldots, (a+k-1)/k)$. Note that c and γ are the shape parameters while β is the scale parameter. The moment estimators of c and γ are the solutions of the equations

$$\sqrt{\beta_1}\left[\frac{\Gamma(\gamma+2/c)}{\Gamma(\gamma)} - \left\{\frac{\Gamma(\gamma+1/c)}{\Gamma(\gamma)}\right\}^2\right]^{3/2}$$
$$= 2\left\{\frac{\Gamma(\gamma+1/c)}{\Gamma(\gamma)}\right\}^3 - 3\frac{\Gamma(\gamma+1/c)\Gamma(\gamma+2/c)}{\Gamma^2(\gamma)} + \frac{\Gamma(\gamma+3/c)}{\Gamma(\gamma)}$$

and

$$\beta_2\left[\frac{\Gamma(\gamma+2/c)}{\Gamma(\gamma)} - \left\{\frac{\Gamma(\gamma+1/c)}{\Gamma(\gamma)}\right\}^2\right]^2$$
$$= -3\left\{\frac{\Gamma(\gamma+1/c)}{\Gamma(\gamma)}\right\}^4 + 6\frac{\Gamma^2(\gamma+1/c)\Gamma(\gamma+2/c)}{\Gamma^3(\gamma)} - 4\frac{\Gamma(\gamma+1/c)\Gamma(\gamma+3/c)}{\Gamma^2(\gamma)}$$
$$+ \frac{\Gamma(\gamma+4/c)}{\Gamma(\gamma)},$$

where β_1 and β_2 are the sample skewness and the sample kurtosis given by (4) and (5), respectively. The moment estimator of β is $\mu_1\Gamma(\gamma)/\Gamma(\gamma+1/c)$, where μ_1 is the sample mean. The maximum likelihood estimators of c, γ and β are the simultaneous solutions of the equations:

$$\frac{n}{c} + \gamma\sum_{i=1}^{n}\log\lambda_i = n\gamma\log\beta + \sum_{i=1}^{n}\left(\frac{\lambda_i}{\beta}\right)^c\log\left(\frac{\lambda_i}{\beta}\right),$$

$$c\sum_{i=1}^{n}\log\lambda_i = nc\log\beta + n\psi(\gamma)$$

and

$$c\beta^{-c-1}\sum_{i=1}^{n}\lambda_i^c = \frac{nc\gamma}{\beta}.$$

One Parameter Half Gaussian Distribution: If g takes the form

$$g(\lambda) = \frac{2}{\sqrt{2\pi}\,\mu}\exp\left(-\frac{\lambda^2}{2\mu^2}\right)$$

for $\lambda > 0$ and $\mu > 0$ then

$$f(y) = \frac{x^{3/2}}{\pi\mu}\left[\Gamma(-1/2)\,{}_1F_3(1;\Delta(2,3/2),\Delta(1,1+j);z)\right.$$
$$\left.+\sum_{h=0}^{1}\frac{(-1)^h A^{(-1/2+h)/2}}{2h!}\Gamma\left(\frac{1/2-h}{2}\right){}_1F_3\left(1;\Delta\left(1,\frac{3/2+h}{2}\right),\Delta(2,1+h);z\right)\right],$$

where $A = x^2/(2\mu^2)$, $z = -A/4$ and $\Delta(k,a) = (a/k,(a+1)/k,\ldots,(a+k-1)/k)$. Note that μ is the scale parameter. The moment estimator of μ is $\mu_1\sqrt{\pi/2}$, where μ_1 is the sample mean. The maximum likelihood estimator of μ is $\sqrt{\mu_2}$, where μ_2 is the second sample moment.

Two Parameter Fréchet Distribution: If g takes the form

$$g(\lambda) = \frac{k\theta^k}{\lambda^{k+1}}\exp\left\{-\left(\frac{\theta}{\lambda}\right)^k\right\}$$

for $\lambda > 0$, $k > 0$ and $\theta > 0$ then

$$f(y) = \frac{k\theta^k\sqrt{x}}{\sqrt{2\pi}}\sum_{j=0}^{q-1}\frac{(-A)^j}{j!}\Gamma(-1/2+kj)\,{}_{p+1}F_q(1;\Delta(p,-1/2+kj),\Delta(q,1+j);(-1)^q z)$$

provided that $0 < k < 1$ and $k = p/q$ where $p \geq 1$ and $q \geq 1$ are co-prime integers, $A = (x/\beta)^k$, $z = (-1)^{p+q}A^q/\{p^p q^q\}$ and $\Delta(k,a) = (a/k,(a+1)/k,\ldots,(a+k-1)/k)$. On the other hand, if $k > 1$ then

$$f(y) = \frac{\theta^k\sqrt{x}}{\sqrt{2\pi}}\sum_{h=0}^{p-1}\frac{(-1)^h A^{(1/2-h)/k}}{h!}\Gamma\left(\frac{h-1/2}{k}\right)C_{5h},$$

where

$$C_{5h} = {}_{q+1}F_p\left(1;\Delta\left(q,\frac{h-1/2}{k}\right),\Delta(p,1+h);\frac{(-1)^p}{z}\right).$$

The moment estimator of k is the root of the equation

$$\sqrt{\beta_1}\left\{\Gamma\left(1-\frac{2}{k}\right)-\Gamma^2\left(1-\frac{1}{k}\right)\right\}^{3/2}=2\Gamma^3\left(1-\frac{1}{k}\right)-3\Gamma\left(1-\frac{1}{k}\right)\Gamma\left(1-\frac{2}{k}\right)+\Gamma\left(1-\frac{3}{k}\right),$$

where β_1 is the sample skewness given by (4). Note that k and θ are the shape and scale parameters, respectively. The moment estimator of θ is $\mu_1/\Gamma(1-1/k)$, where μ_1 is the sample mean. The maximum likelihood estimators of k and θ are the simultaneous solutions of the equations:

$$\sum_{i=1}^{n}\left(\frac{\theta}{\lambda_i}\right)^k\log\left(\frac{\theta}{\lambda_i}\right)=\frac{n}{k}+n\log\theta$$

and

$$k\theta^{k-1}\sum_{i=1}^{n}\lambda_i^{-k}=\frac{nk}{\theta}.$$

Two Parameter Pareto Distribution: If g takes the form

$$g(\lambda)=\frac{ak^a}{\lambda^{a+1}}$$

for $\lambda > k$ and $a > 0$ then

$$f(y)=\frac{ak^a}{\sqrt{2\pi}x^{a+1/2}}\gamma\left(a+1/2,\frac{x}{k}\right).$$

The moment estimator of a is the root of the equation

$$\sqrt{\beta_1}\sqrt{a}(a-3)=2(a+1)\sqrt{a-2},$$

where β_1 is the sample skewness given by (4). Note that a and k are the shape and scale parameters, respectively. The moment estimator of k is $\mu_1(a-1)/a$, where μ_1 is the sample mean. The maximum likelihood estimators of k and a are $\min\lambda_i$ and $\{\log\min\lambda_i-(1/n)\sum_{i=1}^{n}\log\lambda_i\}^{-1}$, respectively.

Four Parameter Two Sided Power Distribution: If g takes the form

$$g(\lambda)=\begin{cases}\frac{p}{b-a}\left(\frac{\lambda-a}{m-a}\right)^{p-1}, & \text{if } a\le\lambda\le m,\\ \frac{p}{b-a}\left(\frac{b-\lambda}{b-m}\right)^{p-1}, & \text{if } m\le\lambda\le b\end{cases}$$

for $0<a\le\lambda\le b<\infty$, $0<a\le m\le b<\infty$ and $p>0$ then

$$f(y)=\frac{p}{(b-a)\sqrt{2\pi}}\left[(m-a)^{1-p}\sum_{k=0}^{p-1}\binom{p-1}{k}(-a)^k x^{p-1/2-k}C_{6k}+(b-m)^{1-p}\sum_{k=0}^{p-1}\binom{p-1}{k}(-b)^k x^{p-1/2-k}C_{7k}\right],$$

where

$$C_{6k} = \gamma\left(\frac{1}{2} - p + k, \frac{x}{a}\right) - \gamma\left(\frac{1}{2} - p + k, \frac{x}{m}\right)$$

and

$$C_{7k} = \gamma\left(\frac{1}{2} - p + k, \frac{x}{m}\right) - \gamma\left(\frac{1}{2} - p + k, \frac{x}{b}\right).$$

Note that p is the shape parameter while a, b and m are the scale parameters. The moment estimators of p and $\theta = (m-a)/(b-a)$ are the solutions of the equations

$$\sqrt{\beta_1}\left\{e_2 - e_1^2\right\}^{3/2} = 2e_1^3 - 3e_1e_2 + e_3$$

and

$$\beta_2\left\{e_2 - e_1^2\right\}^2 = -3e_1^4 + 6e_1^2e_2 - 4e_1e_3 + e_4,$$

where

$$e_k = \frac{p\theta^{k+1}}{p+k} - \sum_{i=0}^{k}\binom{k}{k-i}\frac{p(\theta-1)^{i+1}}{p+i},$$

where β_1 and β_2 are the sample skewness and the sample kurtosis given by (4) and (5), respectively. For the maximum likelihood estimators, see van Dorp and Kotz [14, 15].

Two Parameter Beta Distribution: If g takes the form

$$g(\lambda) = \frac{\lambda^{a-1}(1-\lambda)^{b-1}}{B(a,b)}$$

for $0 < \lambda < 1$, $a > 0$ and $b > 0$ then

$$f(y) = \frac{\Gamma(b)\exp(-x)}{\sqrt{2\pi}\,B(a,b)}\Psi(b, 3/2 - a; x).$$

Note that both a and b are the shape parameters. The moment estimators of a and b are the solutions of the equations

$$\sqrt{\beta_1}\sqrt{ab}(a+b+2) = 2(b-a)\sqrt{a+b+1}$$

and

$$\beta_2 ab(a+b+2)(a+b+3) = 6\{a^3 - a^2(2b-1) + b^2(b+1) - 2ab(b+2)\},$$

where β_1 and β_2 are the sample skewness and the sample kurtosis given by (4) and (5), respectively. The maximum likelihood estimators of a and b are the simultaneous solutions of the equations:

$$\sum_{i=1}^{n}\log\lambda_i = n\psi(a) - n\psi(a+b)$$

and

$$\sum_{i=1}^{n} \log(1-\lambda_i) = n\psi(b) - n\psi(a+b).$$

Two Parameter Inverted Beta Distribution: If g takes the form

$$g(\lambda) = \frac{\lambda^{\gamma-1}}{B(\gamma,\beta)(1+\lambda)^{\gamma+\beta}}$$

for $\lambda > 0$, $\gamma > 0$ and $\beta > 0$ then

$$f(y) = \frac{\Gamma(1/2+\beta)}{\sqrt{2\pi}\,B(\gamma,\beta)} \Psi(1/2+\beta, 3/2-\gamma; x).$$

Note that both γ and β are the shape parameters. The moment estimators of β and γ are the solutions of the equations

$$\begin{aligned}\sqrt{\beta_1}\left\{\gamma(\gamma+\beta-1)\right\}^{3/2} &= 2\gamma^3(\beta-2)^{3/2} - 3\gamma^2(\gamma+1)(\beta-1)\sqrt{\beta-2} \\ &\quad + \gamma(\gamma+1)(\gamma+2)(\beta-1)^2\sqrt{\beta-2}(\beta-3)^{-1}\end{aligned}$$

and

$$\begin{aligned}\beta_2\gamma^2(\gamma+\beta-1)^2 &= -3\gamma^4(\beta-2)^2 + 6\gamma^3(\gamma+1)(\beta-1)(\beta-2) \\ &\quad - 4\gamma^2(\gamma+1)(\gamma+2)(\beta-1)^2(\beta-2)(\beta-3)^{-1} \\ &\quad + \gamma(\gamma+1)(\gamma+2)(\gamma+3)(\beta-1)^3(\beta-2)(\beta-3)^{-1}(\beta-4)^{-1},\end{aligned}$$

where β_1 and β_2 are the sample skewness and the sample kurtosis given by (4) and (5), respectively. The maximum likelihood estimators of β and γ are the simultaneous solutions of the equations:

$$\sum_{i=1}^{n} \log\lambda_i - \sum_{i=1}^{n} \log(1-\lambda_i) = n\psi(\gamma) - n\psi(\gamma+\beta)$$

and

$$\sum_{i=1}^{n} \log(1-\lambda_i) = n\psi(\gamma+\beta) - n\psi(\beta).$$

Two Parameter Lomax Distribution: If g takes the form

$$g(\lambda) = \frac{ac^a}{(c+\lambda)^{a+1}}$$

for $\lambda > 0$, $a > 0$ and $c > 0$ then

$$f(y) = \frac{a\Gamma(1/2+a)}{\sqrt{2c\pi}} \Psi(1/2+a, 1/2; x/c).$$

The moment estimator of a is the root of the equation

$$\sqrt{\beta_1} a^{3/2} = 2(a-2)^{3/2} - 6(a-1)\sqrt{a-2} + 6(a-1)^2\sqrt{a-2}(a-3)^{-1},$$

where β_1 is the sample skewness given by (4). Note that a and c are the shape and scale parameters, respectively. The moment estimator of c is $(a-1)\mu_1$, where μ_1 is the sample mean. The maximum likelihood estimators of a and c are the simultaneous solutions of the equations:

$$\sum_{i=1}^{n} \log(c + \lambda_i) = \frac{n}{a} + n \log c$$

and

$$(a+1)\sum_{i=1}^{n} \frac{1}{c+\lambda_i} = \frac{na}{c}.$$

Two Parameter Generalized Pareto Distribution: If g takes the form

$$g(\lambda) = \frac{1}{k}\left(1 - \frac{c\lambda}{k}\right)^{1/c-1}$$

for $\lambda > 0$, $-\infty < c < \infty$ and $k > 0$ then

$$f(y) = \begin{cases} \frac{\Gamma(1/2-1/c)(-c)^{-1/2}}{\sqrt{2k\pi}} \Psi\left(\frac{1}{2} - \frac{1}{c}, \frac{1}{2}; -\frac{xc}{k}\right), & \text{if } c \le 0, \\ \frac{\Gamma(1/c)c^{-1/2}}{\sqrt{2k\pi}} \exp\left(-\frac{cx}{k}\right)\Psi\left(\frac{1}{c}, \frac{1}{2}; \frac{xc}{k}\right), & \text{if } c > 0. \end{cases}$$

Note that c and k are the shape and scale parameters, respectively. The moment estimator of c is the root of the equation

$$\sqrt{\beta_1}\left\{-\frac{1}{c^3} B\left(-\frac{1}{c} - 2, 3\right) - \frac{1}{c^4} B^2\left(-\frac{1}{c} - 1, 2\right)\right\}^{3/2}$$
$$= \frac{2}{c^6} B^3\left(-\frac{1}{c} - 1, 2\right) + \frac{3}{c^5} B\left(-\frac{1}{c} - 1, 2\right) B\left(-\frac{1}{c} - 2, 3\right) + \frac{1}{c^4} B\left(-\frac{1}{c} - 3, 4\right)$$

for $c \le 0$ or

$$\sqrt{\beta_1}\left\{-\frac{1}{c^3} B\left(-\frac{1}{c} - 2, 3\right) - \frac{1}{c^4} B^2\left(-\frac{1}{c} - 1, 2\right)\right\}^{2}$$
$$= \frac{2}{c^6} B^3\left(2, \frac{1}{c}\right) + \frac{3}{c^5} B\left(2, \frac{1}{c}\right) B\left(3, \frac{1}{c}\right) + \frac{1}{c^4} B\left(4, \frac{1}{c}\right)$$

for $c > 0$, where β_1 is the sample skewness given by (4). The moment estimator of k is given by

$$k = \begin{cases} c^2 \mu_1 \left\{B\left(-\frac{1}{c} - 1, 2\right)\right\}^{-1}, & \text{if } c \le 0, \\ c^2 \mu_1 \left\{B\left(2, \frac{1}{c}\right)\right\}^{-1}, & \text{if } c > 0, \end{cases}$$

where μ_1 is the sample mean. The maximum likelihood estimators of c and k are the simultaneous solutions of the equations:

$$(c-1)\sum_{i=1}^{n}\frac{\lambda_i}{k^2}\left(1-\frac{c\lambda_i}{k}\right)^{-1}=\frac{n}{k}$$

and

$$c(c-1)\sum_{i=1}^{n}\frac{\lambda_i}{k}\left(1-\frac{c\lambda_i}{k}\right)^{-1}=\sum_{i=1}^{n}\log\left(1-\frac{c\lambda_i}{k}\right).$$

Two Parameter Burr III Distribution: If g takes the form

$$g(\lambda)=kc\lambda^{-c-1}(1+\lambda^{-c})^{-k-1}$$

for $\lambda>0$, $c>0$ and $k>0$ then one can write

$$f(y)=\frac{kc}{\sqrt{2\pi}}I,$$

where

$$I=\int_0^\infty\frac{w^{c-1/2}\exp(-wx)}{(1+w^c)^{k+1}}dw.$$

This integral cannot be reduced to an explicit form for general c. It can be reduced for particular values of c. For instance, if $c=2$ (no physical motivation) then

$$\begin{aligned}I&=x^{2k-1/2}\Gamma(1/2-2k)\,{}_1F_2\left(k+1;k+1-\frac{1}{4},-\frac{3}{4};-\frac{x^2}{4}\right)\\&+\frac{1}{2}B\left(k-\frac{1}{4},\frac{5}{4}\right){}_1F_2\left(\frac{5}{4};\frac{1}{2},\frac{1}{4}-k;-\frac{x^2}{4}\right)\\&-\frac{x}{2}B\left(k-\frac{3}{4},\frac{7}{4}\right){}_1F_2\left(\frac{7}{4};\frac{3}{2},\frac{7}{4}-k;-\frac{x^2}{4}\right).\end{aligned}$$

Note that both c and k are the shape parameters. The moment estimators of c and k are the solutions of the equations

$$\begin{aligned}&\sqrt{\beta_1}\left\{kB\left(1-\frac{2}{c},k+\frac{2}{c}\right)-k^2B^2\left(1-\frac{1}{c},k+\frac{1}{c}\right)\right\}^{3/2}\\&=2k^3B^3\left(1-\frac{1}{c},k+\frac{1}{c}\right)-3k^2B\left(1-\frac{1}{c},k+\frac{1}{c}\right)B\left(1-\frac{2}{c},k+\frac{2}{c}\right)\\&\quad+kB\left(1-\frac{3}{c},k+\frac{3}{c}\right)\end{aligned}$$

and

$$\beta_2 \left\{ kB\left(1-\frac{2}{c}, k+\frac{2}{c}\right) - k^2 B^2\left(1-\frac{1}{c}, k+\frac{1}{c}\right) \right\}^2$$
$$= -3k^4 B^4\left(1-\frac{1}{c}, k+\frac{1}{c}\right) + 6k^3 B^2\left(1-\frac{1}{c}, k+\frac{1}{c}\right) B\left(1-\frac{2}{c}, k+\frac{2}{c}\right)$$
$$- 4k^2 B\left(1-\frac{1}{c}, k+\frac{1}{c}\right) B\left(1-\frac{3}{c}, k+\frac{3}{c}\right) + kB\left(1-\frac{4}{c}, k+\frac{4}{c}\right),$$

where β_1 and β_2 are the sample skewness and the sample kurtosis given by (4) and (5), respectively. The maximum likelihood estimators of k and c are the simultaneous solutions of the equations:

$$\sum_{i=1}^{n} \log\left(1+\lambda_i^{-c}\right) = \frac{n}{k}$$

and

$$\sum_{i=1}^{n} \log \lambda_i - (k+1) \sum_{i=1}^{n} \frac{\lambda_i^{-c} \log \lambda_i}{1+\lambda_i^{-c}} = \frac{n}{c}.$$

Two Parameter Burr XII Distribution: If g takes the form

$$g(\lambda) = kc\lambda^{c-1}(1+\lambda^c)^{-k-1}$$

for $\lambda > 0$, $c > 0$ and $k > 0$ then one can write

$$f(y) = \frac{kc}{\sqrt{2\pi}} I,$$

where

$$I = \int_0^\infty \frac{w^{kc-1/2} \exp(-wx)}{(1+w^c)^{k+1}} dw.$$

Again, this integral cannot be reduced to an explicit form for general c. For particular values like $c = 2$ (no physical motivation) one can reduce the integral to

$$I = x^{3/2} \Gamma(-3/2) {}_1F_2\left(k+1; \frac{7}{4}, \frac{5}{4}; -\frac{x^2}{4}\right)$$
$$+ \frac{1}{2} B\left(\frac{3}{4}, \frac{1/2+2k}{2}\right) {}_1F_2\left(\frac{1/2+2k}{2}; \frac{1}{2}, \frac{1}{4}; -\frac{x^2}{4}\right)$$
$$- \frac{x}{2} B\left(\frac{1}{4}, \frac{2k+3/2}{2}\right) {}_1F_2\left(\frac{2k+3/2}{2}; \frac{3}{2}, \frac{3}{4}; -\frac{x^2}{4}\right).$$

Note that both c and k are the shape parameters. The moment estimators of c and k are the solutions of the equations

$$\sqrt{\beta_1}\left\{kB\left(k-\frac{2}{c},1+\frac{2}{c}\right)-k^2B^2\left(k-\frac{1}{c},1+\frac{1}{c}\right)\right\}^{3/2}$$
$$=2k^3B^3\left(k-\frac{1}{c},1+\frac{1}{c}\right)-3k^2B\left(k-\frac{1}{c},1+\frac{1}{c}\right)B\left(k-\frac{2}{c},1+\frac{2}{c}\right)$$
$$+kB\left(k-\frac{3}{c},1+\frac{3}{c}\right)$$

and

$$\beta_2\left\{kB\left(k-\frac{2}{c},1+\frac{2}{c}\right)-k^2B^2\left(k-\frac{1}{c},1+\frac{1}{c}\right)\right\}^{2}$$
$$=-3k^4B^4\left(k-\frac{1}{c},1+\frac{1}{c}\right)+6k^3B^2\left(k-\frac{1}{c},1+\frac{1}{c}\right)B\left(k-\frac{2}{c},1+\frac{2}{c}\right)$$
$$-4k^2B\left(k-\frac{1}{c},1+\frac{1}{c}\right)B\left(k-\frac{3}{c},1+\frac{3}{c}\right)+kB\left(k-\frac{4}{c},1+\frac{4}{c}\right),$$

where β_1 and β_2 are the sample skewness and the sample kurtosis given by (4) and (5), respectively. The maximum likelihood estimators of k and c are the simultaneous solutions of the equations:

$$\sum_{i=1}^{n}\log\left(1+\lambda_i^c\right)=\frac{n}{k}$$

and

$$\sum_{i=1}^{n}\log\lambda_i-(k+1)\sum_{i=1}^{n}\frac{\lambda_i^c\log\lambda_i}{1+\lambda_i^c}=\frac{n}{c}.$$

3 Application

As mentioned in Sect. 1, a popular modal for the DCT coefficients of images such as text documents is the Gaussian distribution. Here, we show that the distributions derived in Sect. 2 are better models for the DCT coefficients whether they follow the Gaussian distribution or not. To show this, we simulated 100 samples each of size 10 from each of the following distributions:

(1) the standard Gaussian distribution (Model 1);
(2) the Student's t distribution with degrees of freedom $\nu = 1$ (Model 2);
(3) the standard logistic distribution given by the pdf $f(x) = \exp(-x)/\{1+\exp(-x)\}^2$ for $-\infty < x < \infty$ (Model 3);
(4) the standard Laplace distribution given by the pdf $f(x) = (1/2)\exp(- \mid x \mid)$ for $-\infty < x < \infty$ (Model 4).

For each of the $100 \times 4 = 400$ samples, we fitted the distribution given by (6) as well as the standard Gaussian distribution. The method of maximum likelihood described in Sect. 2 was used. We computed $2(\log L_2 - \log L_1)$ for each fit, where L_1 and L_2 denote the maximized likelihoods for the two distributions. Figure $k+1$ shows the box plot of the values of $2(\log L_2 - \log L_1)$ for the 100 samples from Model k, $k = 1, 2, 3, 4$.

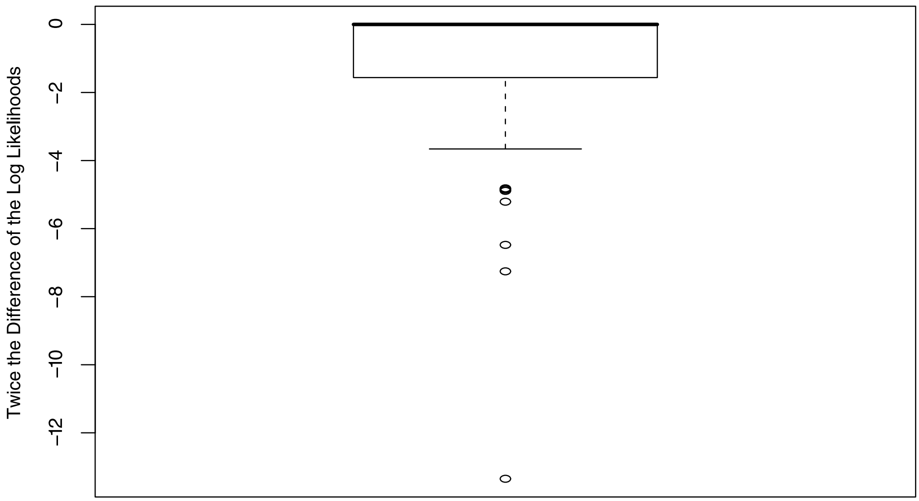

Fig. 2 Box plot of the values of $2(\log L_2 - \log L_1)$ for the 100 simulated samples each of size 10 from the standard Gaussian distribution

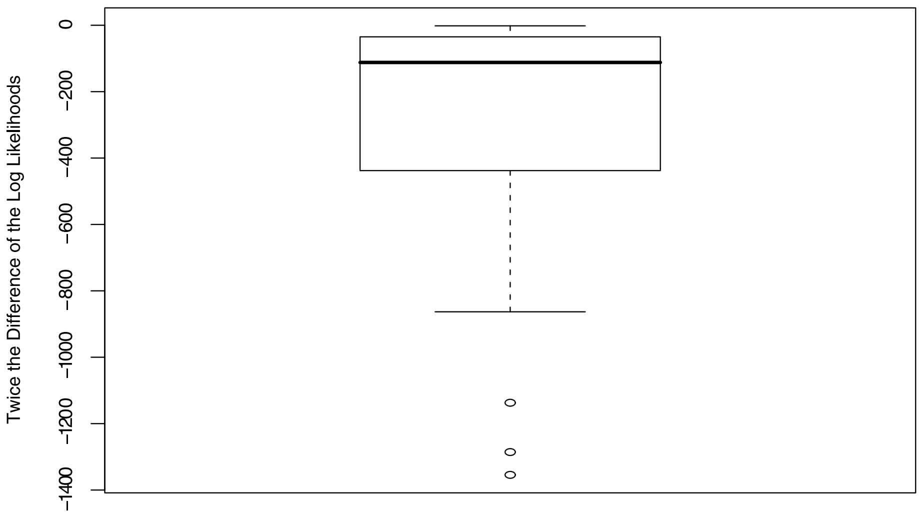

Fig. 3 Box plot of the values of $2(\log L_2 - \log L_1)$ for the 100 simulated samples each of size 10 from the Student's t distribution with degrees of freedom $\nu = 1$

A box plot is a convenient way of graphically depicting numerical data through their five–number summaries: the smallest observation, lower quartile, median, upper quartile, and the largest observation (from the bottom to top). Observations which are considered "outliers" or "extreme observations" are indicated by open dots. The box plot was invented by the American statistician John Tukey. We refer the readers to Tukey [16] for details.

The box plots in Figs. 2, 3, 4 and 5 show that the distribution of $2(\log L_2 - \log L_1)$ lies entirely below zero. In other words, the likelihood L_1 for the fit of (6) is always greater than the likelihood L_2 for the fit of the standard Gaussian distribution. So, one can infer that the

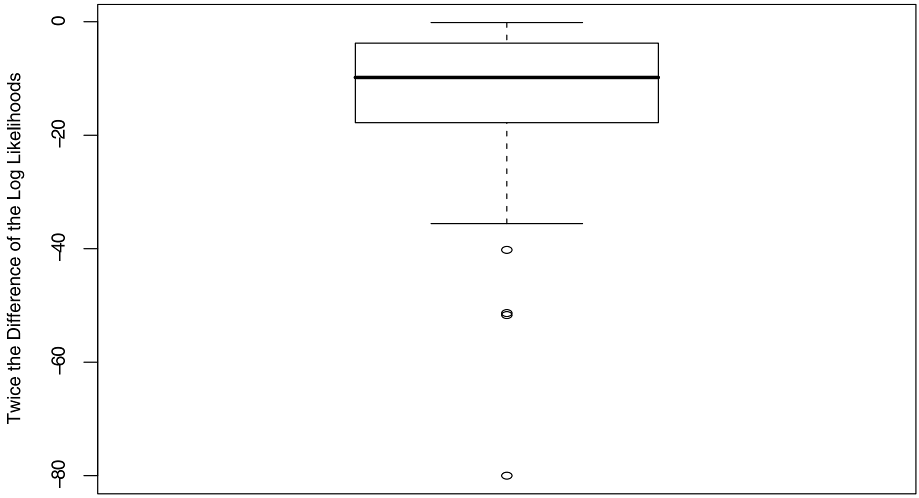

Fig. 4 Box plot of the values of $2(\log L_2 - \log L_1)$ for the 100 simulated samples each of size 10 from the standard logistic distribution

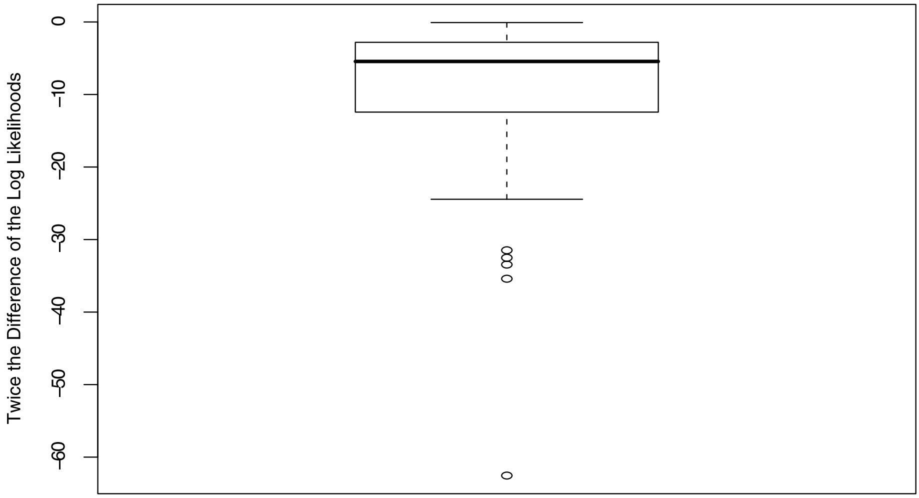

Fig. 5 Box plot of the values of $2(\log L_2 - \log L_1)$ for the 100 simulated samples each of size 10 from the standard Laplace distribution

model given by (6) performs better than the Gaussian distribution when the DCT coefficients are in fact Gaussian distributed, see Fig. 2. More importantly, when the DCT coefficients are not Gaussian distributed the model given by (6) performs better when compared to the Gaussian model, see Figs. 3–5. The results were similar when the other distributions derived in Sect. 2 were used and for larger sample sizes of 100, 1000 and 10,000. Hence, the derived distributions in Sect. 2 provide versatile models whether the DCT coefficients are Gaussian distributed or not.

4 Conclusions

We have derived sixteen flexible models for the distribution of the actual DCT coefficient given by (3) and the corresponding estimation procedures by the method of moments and the method of maximum likelihood. We have established the superior performance of the derived models over the Gaussian distribution given by (1). The models can be useful for the DCT coefficients of large text documents or images of that type (see Lam [7] and Lam and Goodman [8]).

An extension of the work is to consider the generalized Gaussian distribution in place of the Gaussian distribution, i.e. replace (1) by

$$f(x) = \frac{1}{2\beta\Gamma(1+\alpha)} \exp\left(-\left|\frac{x}{\beta}\right|^{1/\alpha}\right)$$

for $-\infty < x < \infty$, where $\beta > 0$ is the scale parameter and $\alpha > 0$ is the shape parameter. In this case, the actual DCT coefficient distribution can be given by

$$f(x) = \frac{1}{2\Gamma(1+\alpha)} \int_0^\infty \frac{1}{\beta} \exp\left(-\left|\frac{x}{\beta}\right|^{1/\alpha}\right) g(\beta) d\beta. \tag{7}$$

Expressions for (7) similar to those in Sect. 2 can be derived by taking $g(\cdot)$ to belong to the sixteen families. We hope to present these expressions and show their applicability in a future paper.

Acknowledgements The author would like to thank the Editor and the two referees for carefully reading the paper and for their comments which greatly improved the paper.

References

1. Ahmed, N., Natarajan, T., Rao, K.R.: Discrete cosine transform. IEEE Trans. Comput. **23**, 90–93 (1974)
2. Wang, Z., Hunt, B.: The discrete *W* transform. Appl. Math. Comput. **16**, 19–48 (1985)
3. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice Hall, New York (1989)
4. Rao, K.R., Yip, P.: Discrete Cosine Transform: Algorithms, Advantages, Applications. Academic Press, Boston (1990)
5. Duhamel, P., Vetterli, M.: Fast Fourier transforms: a tutorial review and a state of the art. Signal Process. **19**, 259–299 (1990)
6. Reininger, R.C., Gibson, J.D.: Distribution of the two–dimensional DCT coefficients for images. IEEE Trans. Commun. **31**, 835–839 (1983)
7. Lam, E.Y.: Analysis of the DCT coefficient distributions for document coding. IEEE Signal Process. Lett. **11**, 97–100 (2004)
8. Lam, E.Y., Goodman, J.W.: A mathematical analysis of the DCT coefficient distributions for images. IEEE Trans. Image Process. **9**, 1661–1666 (2000)
9. Lam, E.Y.: Statistical modelling of the wavelet coefficients with different bases and decomposition levels. IEE Proc. Vis. Image Signal Process. **151**, 203–206 (2004)
10. Teichroew, D.: The mixture of normal distributions with different variances. Ann. Math. Stat. **28**, 510–512 (1957)
11. Nadarajah, S., Kotz, S.: On the DCT coefficient distributions. IEEE Signal Process. Lett. **13**, 601–603 (2006)
12. Prudnikov, A.P., Brychkov, Y.A., Marichev, O.I.: Integrals and Series, vol. 1. Gordon and Breach Science, Amsterdam (1986)
13. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series, and Products, 6th edn. Academic Press, San Diego (2000)
14. van Dorp, J.R., Kotz, S.: A novel extension of the triangular distribution and its parameter estimation. J. R. Stat. Soc. Ser. D—Stat. **51**, 63–79 (2002)
15. van Dorp, J.R., Kotz, S.: The standard two-sided power distribution and its properties: with applications in financial engineering. Am. Stat. **56**, 90–99 (2002)
16. Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley, Reading (1977)