Estimators of tissue proportions from X-ray CT images

C.A. Glasbey
Biomathematics and Statistics Scotland
JCMF, King’s Buildings, Edinburgh, EH9 3JZ, Scotland

C.D. Robinson
Department of Mathematics and Statistics, University of Edinburgh
JCMF, King’s Buildings, Edinburgh, EH9 3JZ, Scotland

April 23, 2002

Abstract

Estimators are derived of tissue proportions from X-ray computed tomography (CT) images. These take into account that many pixels in such images are responses to mixtures of tissue types. The problem is motivated by an application involving estimation of sheep tissue weights. The standard estimator, a count of the number of pixels in a particular range of values, is compared with the maximum likelihood fit of a mixed-pixel distribution and a moment-based estimator. Both simulations and the application show the moment estimator to be best.

Key words: image analysis, maximum likelihood, mixed pixel distribution, moment estimator, threshold.

1 Introduction

The primary application of non-invasive medical imaging techniques such as X-ray computed tomography (CT), is diagnosis, by detection of physiological abnormalities. However, they also have the potential to be used quantitatively, to estimate tissue proportions, in human obesity studies, for example. Here, we describe a quantitative application involving estimation of sheep tissue proportions from X-ray CT, as part of a sheep breeding programme. However, the methodology is generic and could be used with other subjects, and potentially also with other imaging modalities. For a discussion of the use of medical imaging devices to estimate sheep composition, see Simm (1992).
One of the protocols used in the SAC-BioSS CT Unit in Edinburgh to estimate body composition is the Cavalieri method (Roberts et al., 1993), in which a series of equally-spaced CT cross-sectional images are obtained of each sheep, with the first image randomly positioned. The sum of the cross-sectional areas for a particular tissue type, multiplied by the interslice distance, gives an unbiased estimator of that tissue’s total volume. For example, Fig 1 shows every third image in a series for a 26-week old sheep. In CT, X-rays are projected through a subject from different directions, and a computer reconstructs an image of the spatial distribution of attenuation from the transmitted X-rays. For details of the filtered-backprojection algorithm, see, for example, Rosenfeld and Kak (1982, chapter 8) and Jain (1989, chapter 10). However, as we do not have access to the source data and are concerned to develop methods of relevance to many imaging modalities, we regard these images as the starting point for our analysis. The interslice distance was chosen so that approximately 18 images were obtained for each sheep: in order, from top-left to bottom-right, the images show cross-sections of the sheep’s legs, abdomen, chest and shoulders. We note that Roberts et al. (1993) found 18 slices to give reasonable prediction errors for the human body: 6% for bone, 2% for muscle and 9% for fat. X-ray attenuation is measured in Hounsfield units (Hu), which range from about −1000 to in excess of +1000, and for this particular instrument each pixel represents an area 2mm × 2mm in size. The display in Fig 1 has been scaled such that pixels with values less than −230Hu appear black and those greater than +170Hu appear as the lightest shade of grey. The lightest areas are bone, muscles and internal organs appear slightly lighter than fat tissue, and the U-shaped plastic cradle in which the sheep was lying can be seen. Also shown in Fig 1, superimposed in white on the CT image, are hand-drawn boundaries which segment the tissues of interest from internal organs and other areas of the images that are irrelevant for body composition, as described in the figure legend.

One way of estimating tissue proportions is by first classifying each pixel separately. However, this is unnecessary if we only want to know overall proportions (Santiago and Gage, 1995); we can simply analyse the histogram of pixel values from all the images of each sheep. We discuss this issue further in §5. Fig 2 shows such histograms for 100 sheep, grouped by age, including the one displayed in Fig 1. Note, these histograms are only for the areas segmented by the hand-drawn boundaries, and exclude internal organs, etc. These sheep were part of a calibration experiment and, after imaging, they were slaughtered and dissected, so we know their true body compositions (to within dissection errors). In Fig 2, we have restricted the range of Hounsfield values to focus on values relevant to tissues of muscle and fat. The smaller peak on the left of most histograms is produced by fat tissue, and the larger peak on the right is the result of muscle. Pure tissues of either type have only a narrow range of pixel values: the many pixels falling between these peaks are caused by mixed pixels, termed ‘mixels’ by Choi et al. (1991), which are part muscle and part fat. Currently in the SAC-BioSS CT Unit, proportions are estimated simply by thresholding at the midpoint between the two peaks in the histogram. Here we compare this method with two alternatives. We derive the estimators in §2, and conduct a simulation study to compare them in §3. Then, in §4 we apply them to the sheep data in Fig 2.
Figure 1: A series of equally-spaced X-ray CT images of a 26-week old sheep (every third image in a Cavalieri sample). Also shown in white are hand-drawn boundaries which segment the tissues of interest: for the top-left and bottom-right images, these are the encircled regions, and for the remaining images they are the areas enclosed between the two boundaries.
Figure 2: Histogram of pixel values, for a restricted range of Hounsfield values, for tissues of interest of sheep aged (a) 14 weeks, (b) 18 weeks, (c) 22 weeks, (d) 26 weeks (including the sheep shown in Fig 1).
2 Estimators

We consider a model-based approach for estimating tissue proportions in §2.1 and a moment-based approach in §2.2. However, we first give some underlying theory and consider the simplest method, the use of a threshold midway between the mean pixel values for pure fat and pure muscle.

Let \( I(u) \) be a binary indicator function that denotes the tissue type at location \( u \) in a cross-section through a body, where \( u \) is a continuous variable in 2-D, although for notational brevity we denote it by a single letter. So, in our application,

\[
I(u) = \begin{cases} 
1 & \text{if location } u \text{ is fat} \\
0 & \text{if location } u \text{ is muscle.}
\end{cases}
\]

Suppose that the signal (X-ray attenuation in our application) at location \( u \), denoted \( X(u) \), is independently distributed, conditional on \( I(u) \), with mean \( \mu_{I(u)} \), and with a variance \( \nu_{I(u)}^2 \) due to the intrinsic variability in biological tissue. We cannot observe \( X \), but instead, our image data are

\[
y_i = \int w(i-u)X(u) \; du + e_i \quad \text{for all integer pairs } i.
\]

Here, \( i \) denotes the 2-D pixel index, \( w \) is the point spread function of the process by which the data are obtained, with \( \int w \; du = 1 \), and \( e_i \) is an observation error, with mean 0, variance \( \nu_e^2 \). In our application, the ‘process’ is an X-ray CT machine together with a reconstruction algorithm, and the assumptions of spatial invariance and linearity are only an approximation. (It is a potential source of confusion that, in visual displays of images, pixels are shown as square or rectangular boxes; because pixels are better thought of as point samples of blurred signals, with a point spread function that is usually not unity in a displayed square and zero elsewhere.) It follows that \( y_i \) is distributed with mean \( \{p_i\mu_1 + (1-p_i)\mu_0\} \), variance \( \{p_i\sigma_1^2 + (1-p_i)\sigma_0^2\} \), where

\[
p_i = \int w(i-u)I(u) \; du
\]

denotes the (unobserved) proportion of tissue 1 represented in pixel \( i \) and \( \sigma_i^2 = \nu_I^2 \times \int w^2 \; du + \nu_e^2 \).

We note that both Choi et al. (1991) and Kitamoto and Takagi (1999) made the inappropriate assumption that the value of a mixed pixel is a weighted average of the values of pure pixels, and therefore that it has a variance of \( \{p_i^2\sigma_1^2 + (1-p_i^2)\sigma_0^2\} \). From now on we will need \( \sigma_i^2 \), but not \( \nu_I^2 \) or \( \nu_e^2 \), so it is not necessary to distinguish between the sources of variation contributing to \( y_i \). Further, in reality it is likely that \( X \), conditional on \( I \), will be spatially correlated, resulting in variances which are scaled versions of these, but again it is neither possible nor necessary to take this into account. Although, for simplicity, we have described a 2-D case, the above is equally applicable in 3-D, which is important as X-ray CT cross-sections are of finite thickness.

In contrast to others (Choi et al., 1991, for example) our interest is not in individual pixels, but rather in the total volume of tissue 1, i.e., \( \int I \; du \), where integration is over those parts of the image that we are interested in, such as the regions segmented by the white boundaries in Fig 1. Consider

\[
N\bar{p} = \sum_i p_i = \sum_i \left( \int w(i-u)I(u) \; du \right) = \int \left( \sum_i w(i-u) \right) I(u) \; du,
\]
where $N$ is the number of pixels in the summation and \( \bar{p} \) is the sample mean of \( p \). Therefore, \( N\bar{p} \) is equivalent to \( \int I \ du \) if and only if \( \sum_i w(i - u) \equiv 1 \) for all \( u \): a property that is true for some but not all point spread functions. For example, if \( w \) is a Gaussian density function, then the approximation is good provided the standard deviation exceeds 0.5. However, even when \( \int I \ du \) and \( N\bar{p} \) are not equivalent, the latter is still an unbiased estimator of the former provided that the pixel sampling frame can be assumed to have been randomly positioned.

We consider the use of a threshold, \( t \), to estimate tissue proportions. We classify each pixel as follows:

\[
\hat{p}_i = \begin{cases} 
1 & \text{if } y_i \leq t \\
0 & \text{otherwise},
\end{cases}
\]

and estimate \( \int I \ du \) using \( \sum \hat{p} \). (For simplicity, we assume throughout that \( \mu_1 < \mu_0 \).) In particular, we use a threshold midway between the means for the two tissue types, and the estimator is derivable from the histogram of pixel values, as

\[
\hat{F}_{\text{threshold}} = \sum_{y \leq t} n_y \quad \text{where } t = \frac{\mu_1 + \mu_0}{2},
\]

where \( n_y \) denotes the number of pixels taking value \( y \). For a review of this and other threshold methods, within the broader context of image analysis, see Glasbey and Horgan (1995). We assume throughout that \( (\mu_1, \mu_0, \sigma_1^2, \sigma_0^2) \) are known. If they are not, then although it would appear, from Fig 1(a) in particular, that it would not always be possible to estimate them from the histogram of a single sheep, it would be possible to estimate common values from all data, as we do in §4.

### 2.1 Model-based

Glasbey and Robinson (1999) found that the point spread function, \( w \), for the CT machine could be approximated by an isotropic, bivariate, Gaussian density function with standard deviation \( \tau = 0.4 \) pixel units (\( \approx 0.8 \text{mm} \)). To illustrate, Fig 3(a) shows a subimage of the sheep cradle, with mixed pixels along its outer boundary, which has been approximated by the arc of a circle. It is reasonable to assume that both air and cradle have constant X-ray attenuation values, so the value of a pixel will simply be a function of the distance to the edge of the cradle. Fig 3(b) shows pixel values plotted against this distance, with the values corresponding to a Gaussian point spread function superimposed, from which we see that the agreement is very good.

If we assume that boundaries between fat and muscle tissues are smooth and that a negligible number of pixels are affected by more than one boundary, then the proportion of fat in a mixed pixel \( (p) \) is related to \( D \), the perpendicular, signed distance to the nearest boundary, by

\[
p = \Phi \left( \frac{D}{\tau} \right),
\]

where \( \Phi \) is the standard Gaussian integral. Glasbey and Robinson (1999) also showed that, conditional on \( D \) being close to zero, it is reasonable to assume that \( D \) is uniformly distributed.
If $D$ is not close to zero then its precise distributional form has no bearing on the distribution of $p$, which will in any case be close to zero or one. Therefore, we can assume that $D$ is uniformly distributed over the range $\tau L$ to $\tau U$, and the probability density function for $p$ is

$$f(p) = \frac{\sqrt{2\pi}}{U - L} e^{[\Phi^{-1}(p)]^2/2}, \quad \text{for} \quad \Phi(L) \leq p \leq \Phi(U), \quad (2)$$

which they termed the mixed pixel distribution. An illustration is given in Fig 4. We see that it is an asymmetric U-shaped distribution, which has the same qualitative shape as the beta distribution proposed by Kitamoto and Takagi (1999). However, their choice, and the uniform distributions used by Santiago and Gage (1995) and Laidlaw et al. (1998), were chosen empirically, whereas our choice is well-founded via a particular point spread function. As X-ray CT cross-sections are of finite thickness, we also have to consider the third dimension. However, result (2) still holds, provided that either tissue boundaries are approximately orthogonal to the imaging plane or the point spread function is also Gaussian in the third dimension, not necessarily with the same standard deviation. Both these seem to be reasonable working assumptions.

We make a further assumption that, conditional on $p$, the pixel value $(y)$ is normally distributed, so

$$f(y|p) = \frac{e^{-\{(y-p\mu_1-(1-p)\mu_0)^2/2\{p\sigma_1^2+(1-p)\sigma_0^2\}\}}}{\sqrt{2\pi \{p\sigma_1^2+(1-p)\sigma_0^2\}}}.$$
Figure 4: An illustration of the mixed pixel distribution (2) when \((L, U) = (-2.5, 4)\).

It follows that the probability density function of pixel values is

\[
    f(y) = \int_{\Phi(L)}^{\Phi(U)} f(y | p) f(p) dp,
\]

which is analytically intractable, but can be evaluated numerically.

To use this model to estimate tissue proportions, we assume that \((\mu_1, \mu_0, \sigma_1^2, \sigma_0^2)\) are known, and estimate \((L, U)\) by numerically maximising the log-likelihood

\[
    \sum_y n_y \log f(y).
\]

Then, we estimate the fat area from

\[
    \hat{F}_{\text{model}} = N \int_{\Phi(L)}^{\Phi(U)} p f(p) dp,
\]

by numerical integration.

2.2 Moment-based

We have had to make several assumptions to obtain the model-based estimator, some of which may not be wholly appropriate for images such as Fig 1. In particular, the assumption that a pixel is only near at most one tissue boundary fails when there are many thin layers of tissue. Therefore, it is of interest to explore a method with minimal assumptions.
If we make no assumptions about the point spread function, \( w \), or the distributions of pixel values, \( f(y) \), except to assume that
\[
E(y|p) = p\mu_1 + (1-p)\mu_0,
\]
then
\[
\hat{F}_{\text{moment}} = N \frac{\bar{y} - \mu_0}{\mu_1 - \mu_0}
\]
is an unbiased estimator of \( N\bar{y} \). The unbiasedness follows because
\[
E(\hat{F}_{\text{moment}} \mid \bar{y}) = N \frac{E[E(y|p) \mid \bar{y}] - \mu_0}{\mu_1 - \mu_0}
\]
\[
= N \frac{E[p\bar{y} \mu_1 + (1 - E[p\bar{y}])\mu_0 - \mu_0}{\mu_1 - \mu_0}
\]
\[
= N\bar{y}.
\]

It is possible to compute the variance of this estimator. Using results in Stuart et al. (1999, p. 29)
\[
\text{var}(\hat{F}_{\text{moment}} \mid \bar{y}) = E[\text{var}(\hat{F}_{\text{moment}} \mid p) \mid \bar{y}] + \text{var}[E(\hat{F}_{\text{moment}} \mid p) \mid \bar{y}]
\]
\[
= E \left[ N \frac{p\sigma^2_m + (1-p)\sigma^2_m}{(\mu_1 - \mu_0)^2} \right] \bar{y} + 0
\]
\[
= N \frac{\sigma^2_m + (\sigma^2_1 - \sigma^2_0)\bar{y}}{(\mu_1 - \mu_0)^2}.
\]
The variance could be reduced by replacing \( y \) by \( y' \), where
\[
y' = \begin{cases} 
\mu_1 & \text{if } y < \mu_1 \\
y & \text{if } \mu_1 \leq y \leq \mu_0 \\
\mu_0 & \text{if } \mu_0 < y
\end{cases}
\]
However, the resulting estimator is not unbiased and performed badly in a simulation study similar to that described on §4, so we have preferred to simply use \( y \).

3 Comparison of estimators

We used a simulation study to compare the three estimators developed in §2. Images were simulated of size 60 \( \times \) 60, to approximate the number of pixels in cross-sectional images in Fig 1. A set of seven randomly-positioned parallel lines were simulated which crossed each image, with an inter-line spacing of at least 2 pixel units, to represent alternate layers of fat and muscle tissue, with \((\mu_1, \mu_0, \sigma^2_1, \sigma^2_0) = (-120, 50, 600, 1000)\). Fig 5(a) shows one such image. By choosing seven boundaries, the number of mixed pixels matched that in typical sheep CT images, and with widths of at least 2 pixel units, the model assumptions in §2.1 were satisfied. True tissue areas were computed using standard formulae from coordinate geometry. We also
simulated images with four thin layers of tissue, of width 0.25, 0.5 or 1 pixel units, to represent some of the thin layers of tissue evident in Fig 1. Fig 5(b) shows one image for width = 0.5 pixel units.

The methods of estimation given by (1), (5), and (6) were compared on the basis of the root mean square error (r.m.s.e.) of estimated fat area, averaged over 100 independent simulations. If, in the $j$th simulation, the true fat area is $F_j$ and the estimated area using a particular method is $\hat{F}_j$, then

$$r.m.s.e. = \sqrt{\frac{1}{100} \sum_{j=1}^{100} (F_j - \hat{F}_j)^2}.$$  \hspace{1cm} (8)

Results are given in Table 1, expressed as % of total area. We see that, if widths > 2, then, as we might expect, the model-based method performs best, with a r.m.s.e. of 0.165%. However, in the presence of narrow tissue layers, this method performs less well, because then the assumption no longer holds that a pixel is near only one boundary, and the model is no longer valid. The moment-based method is more robust, and consistently produces r.m.s.e.s of 0.27%.

We can decompose the mean square error into a squared bias term and a variance term, so that

$$r.m.s.e. = \sqrt{(\hat{F} - \tilde{F})^2 + \frac{1}{100} \sum_{j=1}^{100} ((F_j - \hat{F}) - (\hat{F}_j - \tilde{F}))^2} \hspace{1cm} \text{where} \hspace{1cm} \hat{F} = \frac{1}{100} \sum_{j=1}^{100} F_j$$

and $\tilde{F}$ is similarly defined. These separate terms are also given in Table 1, from which we see that changing the widths of tissue layers affects the bias but not the variance of the estimators. From (7), with the proportion of fat being approximately 50%, the standard deviation of the moment estimator is 0.28%, which agrees well with the simulated results.
<table>
<thead>
<tr>
<th>method</th>
<th>widths of tissue layers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt; 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>root mean square error</th>
<th>threshold</th>
<th>model</th>
<th>moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.378</td>
<td>0.165</td>
<td>0.273</td>
</tr>
<tr>
<td></td>
<td>0.495</td>
<td>0.221</td>
<td>0.272</td>
</tr>
<tr>
<td></td>
<td>1.754</td>
<td>0.628</td>
<td>0.272</td>
</tr>
<tr>
<td></td>
<td>1.572</td>
<td>0.896</td>
<td>0.274</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>bias</th>
<th>threshold</th>
<th>model</th>
<th>moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.315</td>
<td>0.024</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>0.448</td>
<td>0.121</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>1.741</td>
<td>0.606</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>1.564</td>
<td>0.885</td>
<td>0.027</td>
</tr>
</tbody>
</table>

| standard deviation | threshold | model | moment |
|                   | 0.209     | 0.163 | 0.272  |
|                   | 0.211     | 0.184 | 0.271  |
|                   | 0.216     | 0.163 | 0.270  |
|                   | 0.156     | 0.140 | 0.273  |

Table 1: Root mean square errors, biases and standard deviations of fat area estimation, expressed as % of total area, for 3 estimators applied to 100 simulations for each of 4 tissue widths.

4 Application

We assume that \((\mu_1, \mu_0, \sigma_1^2, \sigma_0^2)\) have common values for all 100 sheep in the calibration experiment. This is consistent with Fig 2. We estimated these parameters, together with \((L,U)\), by maximising the log-likelihood, given by (4), for the histogram of pixel values from all sheep aged 26 weeks. Fig 6 shows the data and the fit, which can be seen to be impressive, though not perfect, and would be rejected by a formal test of goodness-of-fit, as would any other model for such a large dataset. (The regularly-spaced spikes in the histogram are due to rounding errors in the CT machine during the construction of the image.)

We applied the three methods of estimation given by (1), (5), and (6) to the 100 histograms in Fig 2. We then multiplied the estimated fat areas by the interslice distance for each sheep and by a scaling constant to convert volumes to weights, to obtain estimates of fat weight. We observe that there will be some spill-over at the boundaries in Fig 1, with a small loss of fat and muscle tissues and a small gain of organ and skin tissues. However, these are small effects, which to some extent will cancel each other out. They could be dealt with if deemed sufficiently important but, for simplicity, we have ignored them. Table 2 summarises the results, in terms of r.m.s.e.s as % of total weight, for each age class and overall. We again see that the moment method gives the best results, though the differences between the model and moment estimators are not great. Errors are much larger than in Table 1. A large proportion of the remaining variability is intrinsic, due primarily to the Cavalieri sampling method, but also to dissection errors. However, the gains achieved over the thresholding method mean that, to achieve a specified level of precision, fewer images would be needed, leading to time savings in scanning and manually interpreting images.
Figure 6: *Distribution of pixel values in 26-week animals: (---) observed count, (---) maximum likelihood fit of distribution (3).*

<table>
<thead>
<tr>
<th>method</th>
<th>age (weeks)</th>
<th>14</th>
<th>18</th>
<th>22</th>
<th>26</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>threshold</td>
<td></td>
<td>6.44</td>
<td>5.35</td>
<td>3.55</td>
<td>3.50</td>
<td>4.63</td>
</tr>
<tr>
<td>model</td>
<td></td>
<td>3.95</td>
<td>3.96</td>
<td>3.65</td>
<td>3.29</td>
<td>3.64</td>
</tr>
<tr>
<td>moment</td>
<td></td>
<td>3.91</td>
<td>4.01</td>
<td>3.44</td>
<td>3.18</td>
<td>3.56</td>
</tr>
</tbody>
</table>

Table 2: *Root mean square errors of fat weight, expressed as % of total weight, using 3 estimators for data from 100 sheep.*

5 Discussion

Estimators of tissue volumes in X-ray CT images have been derived which take into account that many pixels in CT images are responses to mixtures of tissue types. One approach is to obtain the maximum likelihood fit of a recently-developed mixed-pixel distribution. Alternatives are thresholding and a moment-based estimator. Simulations and a real application show the moment estimator to be best. In this particular application, improvements in precision were quite small, when compared with the variability due to Cavalieri sampling. However, the methodology is generic and has the potential to be used with 3-D images, other imaging modalities and subject domains where gains may be greater.

An alternative approach would have been to first classify each pixel separately, to sub-pixel resolution (Jensen and Anastassiou, 1995; Hitchcock and Glasbey, 1997; Gavin and Jennison, 1997). However, as we only want to know overall volumes, this is unnecessarily complicated and computationally intensive. Also, it would require stronger assumptions to be made in order
to model the tissues. Therefore, the results are likely to be less robust than those we have developed.

Further work is needed to develop efficient estimators when more than two tissue types are present or pure-pixel distributions are unknown. It may be necessary to take into account a pixel’s spatial context, i.e. values of neighbouring pixels. Methods have already been developed for cases in remote sensing where the image is multivariate (Foody and Cox, 1994; Drake et al., 1999; Faraklioti and Petrou, 2000).

Acknowledgements

We thank Chris Theobald and Mark Young for advice on this work, which was funded by the Ministry of Agriculture, Fisheries and Food, the Meat and Livestock Commission and the Scottish Executive Environment and Rural Affairs Department as part of the LINK Sustainable Livestock Production Programme (Project LK 0607), and by Engineering and Physical Sciences Research Council funding for the second author.

References


