

# Online Statistical Inference for Large-Scale Binary Images

Moo K. Chung<sup>1,2(✉)</sup>, Ying Ji Chuang<sup>2</sup>, and Hourii K. Vorperian<sup>2</sup>

<sup>1</sup> Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, USA

`mkchung@wisc.edu`

<sup>2</sup> Vocal Tract Development Laboratory, Waisman Center, University of Wisconsin, Madison, USA

**Abstract.** We present a unified online statistical framework for quantifying a collection of binary images. Since medical image segmentation is often done semi-automatically, the resulting binary images may be available in a sequential manner. Further, modern medical imaging datasets are too large to fit into a computer's memory. Thus, there is a need to develop an iterative analysis framework where the final statistical maps are updated sequentially each time a new image is added to the analysis. We propose a new algorithm for online statistical inference and apply to characterize mandible growth during the first two decades of life.

## 1 Introduction

The typical implementation of statistical inference in medical imaging requires that all the images are available in advance. That is the usual premise of existing medical image analysis tools such as Image J, SPM, SFL and AFNI. However, there are many situations where the entire imaging dataset is not available and parts of imaging data are obtained in a delayed sequential manner. This is a common problem in medical imaging, where not every subject is scanned and processed at the same time.

When the image size is large, it may not be possible to fit all of the imaging data in a computer's memory, making it necessary to perform the analysis by adding one image at a time in a sequential manner. In another situation, the imaging dataset may be so large that it is not practical to use all the images in the dataset but use a subset of the dataset. In this situation, we need to incrementally add stratified datasets one at a time to see if we are achieving reasonable statistical results. In all the above situations, we need a way to incrementally update the analysis result without repeatedly running the entire analysis whenever new images are added.

An *online algorithm* is one that processes its inputted data in a sequential manner [9]. Instead of processing the entire set of imaging data from the start, an online algorithm processes one image at a time. That way, we can bypass the memory requirement, reduce numerical instability and increase computational efficiency. Online algorithms and machine learning are both concerned

with problems of making decisions about the present based on knowledge of the past [3]. Thus, online algorithms are often encountered in machine learning literature but there is very limited number of studies in medical imaging possibly due to the lack of problem awareness. With the ever-increasing amount of large scale medical imaging databases such as Alzheimer’s disease neuroimaging initiative (ADNI) and human connectome project (HCP), the development of various online algorithm is warranted. In this study, we propose to develop online statistical inference procedures. The methods are then used in characterizing the mandible growth using binary mandible segmentations from CT.

## 2 Probabilistic Model of Binary Segmentation

Let  $p(x)$  be the probability of voxel  $x$  belonging to some region of interest (ROI)  $\mathcal{M}$ . Let  $\mathbf{1}_{\mathcal{M}}$  be an indicator function defined as  $\mathbf{1}_{\mathcal{M}}(x) = 1$  if  $x \in \mathcal{M}$  and 0 otherwise. We assume that the shape of  $\mathcal{M}$  is *random* (due to noise) and we associate it with probability  $p(x)$ :

$$P(x \in \mathcal{M}) = p(x), \quad P(x \notin \mathcal{M}) = 0.$$

The volume of  $\mathcal{M}$  given by  $vol(\mathcal{M}) = \int_{\mathbb{R}^3} \mathbf{1}_{\mathcal{M}}(x) dx$  is also random. The mean volume of  $\mathcal{M}$  is then

$$\mathbb{E} vol(\mathcal{M}) = \int_{\mathbb{R}^3} \mathbb{E} \mathbf{1}_{\mathcal{M}}(x) dx = \int_{\mathbb{R}^3} p(x) dx.$$

The integral of the probability map thus can be used as an estimate for the volume of ROI. Unfortunately, medical images often have holes and cavities that have to be patched topologically for accurate volume estimation (Fig. 1a). Such topological defects can be easily patched by Gaussian kernel smoothing without resorting to advanced topology correction methods (Fig. 1b) [5].

Consider a 3-dimensional Gaussian kernel  $K(x) = \frac{1}{(2\pi)^{3/2}} \exp(-\|x\|^2/2)$ , where  $\|\cdot\|$  is the Euclidean norm of  $x \in \mathbb{R}^3$ . The rescaled kernel  $K_t$  is defined as  $K_t(x) = K(x/t)/t^3$ . Gaussian kernel smoothing applied to the probability map  $p(x)$  is given by

$$K_t * p(x) = \int_{\mathbb{R}^3} K_t(x - y)p(y) dy.$$

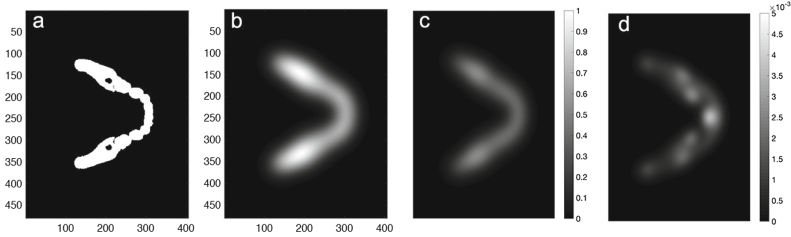
The volume estimate is invariant under smoothing:

$$\begin{aligned} \mathbb{E} vol(\mathcal{M}) &= \int_{\mathbb{R}^3} p(y) dy \\ &= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} K_t(x - y)p(y) dy dx = \int_{\mathbb{R}^3} K_t * p(x) dx. \end{aligned}$$

Here, we used the fact that Gaussian kernel is a probability density, i.e.,

$$\int_{\mathbb{R}^3} K_t(x, y) dx = 1$$

for any  $y \in \mathbb{R}^3$ . Thus, the smoothed probability map  $K_t * p(x)$  can be taken as a more robust probability map of whether a voxel belongs to a mandible and can be used as a response variable in modeling the growth of mandible.



**Fig. 1.** (a) A representative mandible binary segmentation that are affine registered to the template space. (b) Gaussian kernel smoothing of segmentation with bandwidth  $\sigma = 20$ . Smoothing can easily patch topological artifacts such as cavities and handles. The sample mean (c) and variance (d) of the smoothed maps computed using the online algorithm.

### 3 Online Algorithm for $t$ -Test

Given smoothed images  $x_1, \dots, x_m$ , an *online algorithm* for computing the sample mean image  $\mu_m$  is given by

$$\mu_m = \frac{1}{m} \sum_{i=1}^m x_i = \mu_{m-1} + \frac{1}{m}(x_m - \mu_{m-1}) \tag{1}$$

for any  $m \geq 1$ . The algorithm updates the previous mean  $\mu_{m-1}$  with new image  $x_m$ . This algorithm avoids accumulating large sums and tend to be numerically more stable [7].

An online algorithm for computing the sample variance map  $\sigma_m^2$  is algebraically involved [4, 11]. After lengthy derivation, it can be shown that

$$\sigma_m^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu_m)^2 = \frac{m-2}{m-1} \sigma_{m-1}^2 + \frac{1}{m} (x_m - \mu_{m-1})^2$$

for  $m \geq 2$ . The algorithm starts with the initial value  $\sigma_1^2 = 0$ . Figure 1 displays the results of mean and variance computation using the online algorithms.

For comparing a collection of images between groups, two-sample  $t$ -statistic can be used. Given measurements  $x_1, \dots, x_m \sim N(\mu^1, (\sigma^1)^2)$  in one group and  $y_1, \dots, y_n \sim N(\mu^1, (\sigma^2)^2)$  in the other group, the two-sample  $t$ -statistic for testing  $H_0 : \mu^1 = \mu^2$  at each voxel  $x$  is given by

$$T_{m,n}(x) = \frac{\mu_m^1 - \mu_n^2 - (\mu^1 - \mu^2)}{\sqrt{(\sigma^1)^2_m/m + (\sigma^2)^2_n/n}}, \tag{2}$$

where  $\mu_m^1, \mu_m^2, (\sigma^1)_m^2, (\sigma^2)_m^2$  are sample means and variances in each group estimated using the online algorithm.  $T_{m,n}$  is then sequentially computed as

$$T_{1,0} \rightarrow T_{2,0} \rightarrow \dots \rightarrow T_{m,0} \rightarrow T_{m,1} \rightarrow \dots \rightarrow T_{m,n}$$

in  $m + n$  steps.

### 4 Online Algorithm for Linear Regression

The online algorithm for linear regression is itself useful but additionally more useful in constructing an online algorithm for  $F$ -tests in the next section. Given data vector  $\mathbf{y}_{m-1} = (y_1, \dots, y_{m-1})'$  and design matrix  $Z_{m-1}$ , consider linear model

$$\mathbf{y}_{m-1} = Z_{m-1}\boldsymbol{\lambda}_{m-1} \tag{3}$$

with unknown parameter vector  $\boldsymbol{\lambda}_{m-1} = (\lambda_1, \lambda_2, \dots, \lambda_k)'$ . Multiplying  $Z'_{m-1}$  on the both sides we have

$$Z'_{m-1}\mathbf{y}_{m-1} = Z'_{m-1}Z_{m-1}\boldsymbol{\lambda}_{m-1} \tag{4}$$

Let  $W_{m-1} = Z'_{m-1}Z_{m-1}$ , which is a  $k \times k$  matrix. In most applications, there are substantially more data than the number of parameters, i.e.,  $m \gg k$ , and  $W_{m-1}$  is invertible. The least squares estimation (LSE) of  $\boldsymbol{\lambda}_{m-1}$  is given by

$$\boldsymbol{\lambda}_{m-1} = W_{m-1}^{-1}Z'_{m-1}\mathbf{y}_{m-1}.$$

When new data  $y_m$  is introduced to the linear model (3), the model is updated to

$$\begin{pmatrix} \mathbf{y}_{m-1} \\ y_m \end{pmatrix} = \begin{pmatrix} Z_{m-1} \\ z_m \end{pmatrix} \boldsymbol{\lambda}_m,$$

where  $z_m$  is  $1 \times k$  row vector. Subsequently, we have

$$W'_{m-1}\boldsymbol{\lambda}_{m-1} + z'_m y_m = (W_{m-1} + z'_m z_m)\boldsymbol{\lambda}_m.$$

Using Woodbury formula [6],

$$(W_{m-1} + z'_m z_m)^{-1} = W_{m-1}^{-1} - c_m W_{m-1}^{-1} z'_m,$$

where  $c_m = 1/(1 + z_m W_{m-1} z'_m)$  is scalar. Then we have the explicit online algorithm for updating the parameter vector:

$$\boldsymbol{\lambda}_m = (I - W_{m-1}^{-1} z'_m y_m - c_m W_{m-1}^{-1} z'_m W'_{m-1})\boldsymbol{\lambda}_{m-1} - c_m W_{m-1}^{-1} z'_m z'_m y_m, \tag{5}$$

where  $I$  is the identity matrix of size  $k \times k$ . Since the algorithm requires  $W_{m-1}$  to be invertible, the algorithm must start from

$$\boldsymbol{\lambda}_k \rightarrow \boldsymbol{\lambda}_{k+1} \rightarrow \dots \rightarrow \boldsymbol{\lambda}_m.$$

A similar online algorithm for fitting a general linear model (GLM) was introduced for real-time fMRI [2], where the Cholesky factorization was used to invert the covariance matrix in solving GLM. Our approach based on Woodbury formula does not require the factorization or inversion of matrices and thus more efficient.

## 5 Online Algorithm for $F$ -Test

Let  $y_i$  be the  $i$ -th image,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  to be the variables of interest and  $\mathbf{z}_i = (z_{i1}, \dots, z_{ik})'$  to be nuisance covariates corresponding to the  $i$ -th image. We assume there are  $m - 1$  images to start with. Consider a general linear model

$$\mathbf{y}_{m-1} = Z_{m-1}\boldsymbol{\lambda}_{m-1} + X_{m-1}\boldsymbol{\beta}_{m-1},$$

where  $Z_{m-1} = (z_{ij})$  is  $(m - 1) \times k$  design matrix,  $X_{m-1} = (x_{ij})$  is  $(m - 1) \times p$  design matrix.  $\boldsymbol{\lambda}_{m-1} = (\lambda_1, \dots, \lambda_k)'$  and  $\boldsymbol{\beta}_{m-1} = (\beta_1, \dots, \beta_p)'$  are unknown parameter vectors to be estimated at the  $(m - 1)$ -th iteration. Consider hypotheses

$$H_0 : \boldsymbol{\beta} = 0 \text{ vs. } H_1 : \boldsymbol{\beta} \neq 0.$$

The reduced null model when  $\boldsymbol{\beta} = 0$  is  $\mathbf{y}_{m-1} = Z_{m-1}\boldsymbol{\lambda}_{m-1}^0$ . The goodness-of-fit of the null model is measured by the sum of the squared errors (SSE):

$$\text{SSE}_{m-1}^0 = (\mathbf{y}_{m-1} - Z_{m-1}\boldsymbol{\lambda}_{m-1}^0)'(\mathbf{y}_{m-1} - Z_{m-1}\boldsymbol{\lambda}_{m-1}^0),$$

where  $\boldsymbol{\lambda}_{m-1}^0$  is estimated using the online algorithm (5). This provide the sequential update of SSE under  $H_0$ :

$$\text{SSE}_k^0 \rightarrow \text{SSE}_{k+1}^0 \rightarrow \dots \rightarrow \text{SSE}_m^0.$$

Similarly the fit of the alternate full model is measured by

$$\text{SSE}_{m-1}^1 = (\mathbf{y}_{m-1} - \mathbb{Z}_{m-1}\boldsymbol{\gamma}_{m-1}^1)'(\mathbf{y}_{m-1} - \mathbb{Z}_{m-1}\boldsymbol{\gamma}_{m-1}^1),$$

where  $\mathbb{Z}_{m-1} = [Z_{m-1} X_{m-1}]$  is the combined design matrix and

$$\boldsymbol{\gamma}_{m-1}^1 = \begin{pmatrix} \boldsymbol{\lambda}_{m-1}^1 \\ \boldsymbol{\beta}_{m-1}^1 \end{pmatrix}$$

is the combined parameter vector. Similarly SSE under  $H_1$  is given as

$$\text{SSE}_{k+p}^1 \rightarrow \text{SSE}_{k+1}^1 \rightarrow \dots \rightarrow \text{SSE}_m^1.$$

Under  $H_0$ , the test statistic at the  $m$ -th iteration  $f_m$  is given by

$$f_m = \frac{(\text{SSE}_0 - \text{SSE}_1)/p}{\text{SSE}_0/(m - p - k)} \sim F_{p, m-p-k}, \tag{6}$$

which is the  $F$ -statistic with  $p$  and  $m - p - k$  degrees of freedom.

## 6 Random Field Theory

Since the statistic maps are correlated over voxels, it is necessary to correct multiple comparisons using the random field theory [12], which is based on the expected Euler characteristic (EC) approach. Given statistic maps  $S$  such as  $t$ - or  $F$ -test maps, for sufficiently high threshold  $h$ , we have

$$P\left(\sup_{x \in \mathcal{M}} S(x) > h\right) = \sum_{d=0}^N \mu_d(\mathcal{M})\rho_d(h), \tag{7}$$

where  $\mu_d(\mathcal{M})$  is the  $d$ -th Minkowski functional or intrinsic volume of  $\mathcal{M}$ .  $\rho_d(h)$  is the EC-density of  $S$ . The explicit formulas for  $\mu_d$  and  $\rho_d$  are given in [12].

## 7 Application

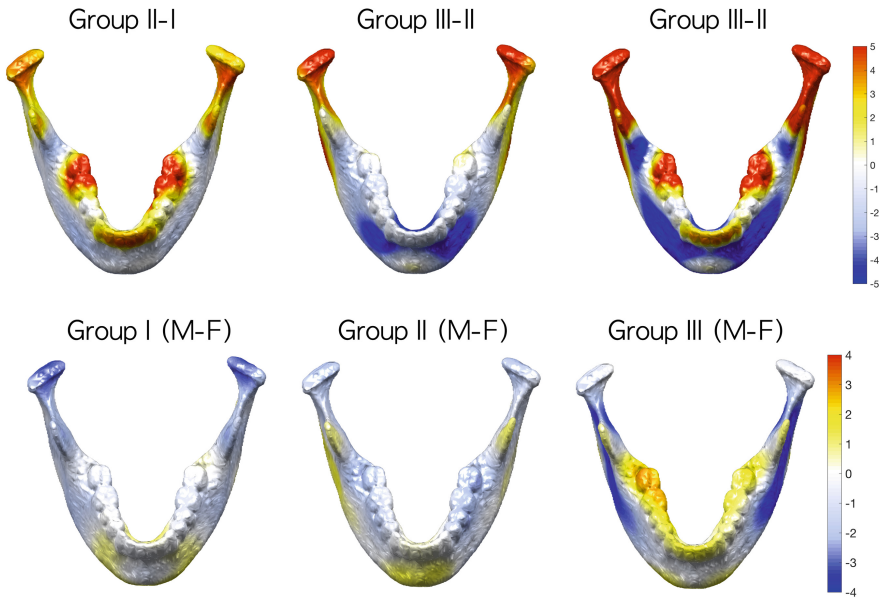
*Subjects.* The dataset consisted of 290 typically developing individuals ranging in age from birth to 20 years old. Only CT images showing the full mandible without any motion or any other artifacts were selected though minimal dental artifacts were tolerated. The age distribution of the subjects is  $9.66 \pm 6.34$  years. The minimum age was 0.17 years and maximum age was 19.92 years. A total of 160 male and 130 female subjects were divided into 3 groups. Group I (age below 7) contained 130 subjects. Group II (between 7 and 13) contained 48 subjects. Group III (between 13 and 20) contained 112 subjects. The main biological question of interest was whether there were localized regions of growth between these age groups. The same grouping was used in the previous study [5].

*Image preprocessing.* CT images were visually inspected and determined to capture the whole mandible geometry. The mandibles in CT were semi-automatically segmented using an in-house processing pipeline that involves image intensity thresholding using the Analyze software package (AnalyzeDirect, Inc., Overland Park, KS). Each of the processed mandibles were examined visually and edited manually by raters. The segmented binary images were then affine registered to the mandible labeled as F226-15-04-002-M (Fig. 1). The mandible F226-15-04-002-M served as the template. Due to the lack of existing prior map in the field, we simply used the normalized binary segmentation results as the probability map  $p(x)$ .

CT images are inherently noisy due to errors associated with image acquisition. Compounding the image acquisition errors, there are errors caused by image registration and semiautomatic segmentation. So it is necessary to smooth out the affine registered and semiautomatic segmented images. We smoothed the binary images with Gaussian kernel with bandwidth  $\sigma = 20$  voxels (Fig. 1). Since CT image resolution is 0.35 mm, 20 voxel wide bandwidth is equivalent to 7 mm. The bandwidth was chosen to reflect the size of missing teeth and cavities. Any smaller filter size will not mask large missing teeth and cavity. The average of all 290 smoothed binary images was computed and used as the final template. For visualization, the statistical maps are projected onto the surface of template.

*Age effects.* We performed the  $t$ -test to assess age effects between the groups. The resulting  $t$ -statistic maps are displayed in Fig. 2-top. Voxels above or below  $\pm 4.41$  were considered significant in the  $t$ -statistic between age groups I and II at the 0.05 level after the multiple comparisons correction. Similarly for other age group comparisons, voxels above or below  $\pm 4.43$  (between II and III) or 4.37 (between I and III) were considered significant at the 0.05 level. These regions are colored dark red or dark blue. The dark red regions show positive growth (bone deposition) and dark blue regions show negative growth (bone resorption). The findings are consistent with previous studies based on 2D surface deformation [5] and landmarks [10].

*Sex effects.* Within each group, we tested the significance of sexual dimorphism by performing the two-sample  $t$ -test between males and females. The resulting  $t$ -statistic maps are displayed in Fig. 2-bottom. Any region above or below  $\pm 4.37$ ,



**Fig. 2.** Top:  $t$ -stat. maps showing mandible growth. The elongation of mandible is shown between Groups II and III, and I and III. The condyle regions show prominent growth in Group III- I comparison. At the same time, the elongation is shown as negative growth (dark blue). Bottom:  $t$ -stat. maps (male - female) showing sex differences in each age group. There were no significant sex differences in groups I and II. However, pubertal and post-pubertal sex difference are evident in group III that starts at age 13.

4.89 and 4.50 (for group I, II and III respectively) were considered significant at 0.05 level after the multiple comparisons correction. In group I and II, there is no gender differences. In group III, the statistical significance is localized in the regions between Condyle and Gonion in the both sides. Such findings are consistent with general findings on sexual dimorphism that become evident during puberty.

## 8 Discussion

The image processing and analysis somewhat resembles the voxel-based morphometry (VBM) widely used in modeling the gray and white matter tissue probability maps in structural brain magnetic resonance imaging studies [1]. VBM does not necessarily require very accurate nonlinear registration. The shape difference is implicitly encoded in tissue density maps. If perfect registration is done, the tissue density maps will be identical across subjects and we will not detect any difference. Thus, in our study, only affine registration is used. Previously we used a diffeomorphic surface shape model to a similar dataset [5], where we also obtained getting the similar pattern of wide spread growth in almost identical places.

In VBM, the posterior probability map is estimated using the prior probability map. However, there is no such prior map in mandibles CT studies yet. Our 290 subject average probability map is distributed as a potential prior map for other Bayesian shape modeling [8]: <http://www.stat.wisc.edu/~mchung/VBM>.

**Acknowledgements.** This work was supported by NIH Research Grants R01 DC6282, P-30 HD03352, UL1TR000427 and R01 EB022856.

## References

1. Ashburner, J., Friston, K.: Why voxel-based morphometry should be used. *NeuroImage* **14**, 1238–1243 (2001)
2. Bagarinao, E., Nakai, T., Tanaka, Y.: Real-time functional MRI: development and emerging applications. *Magn. Reson. Med. Sci.* **5**, 157–165 (2006)
3. Blum, A.: On-line algorithms in machine learning. In: Fiat, A., Woeginger, G.J. (eds.) *Online Algorithms*. LNCS, vol. 1442, pp. 306–325. Springer, Heidelberg (1998). doi:[10.1007/BFb0029575](https://doi.org/10.1007/BFb0029575)
4. Chan, T.F., Golub, G.H., LeVeque, R.J.: Algorithms for computing the sample variance: analysis and recommendations. *Am. Stat.* **37**, 242–247 (1983)
5. Chung, M.K., Qiu, A., Seo, S., Vorperian, H.K.: Unified heat kernel regression for diffusion, kernel smoothing and wavelets on manifolds and its application to mandible growth modeling in CT images. *Med. Image Anal.* **22**, 63–76 (2015)
6. Deng, C.Y.: A generalization of the Sherman-Morrison-Woodbury formula. *Appl. Math. Lett.* **24**, 1561–1564 (2011)
7. Finch, T.: Incremental calculation of weighted mean and variance. University of Cambridge, 4:11–4:15 (2009)
8. Jaakkola, T., Jordan, M.: A variational approach to bayesian logistic regression models and their extensions. In: *Sixth International Workshop on Artificial Intelligence and Statistics*, vol. 82 (1997)
9. Karp, R.M.: On-line algorithms versus off-line algorithms: how much is it worth to know the future? In: *IFIP Congress*, vol. 12, pp. 416–429 (1992)
10. Kelly, M.P., Vorperian, H.K., Wang, Y., Tillman, K.K., Werner, H.M., Chung, M.K., Gentry, L.R.: Characterizing mandibular growth using three-dimensional imaging techniques and anatomic landmarks. *Arch. Oral Biol.* **77**, 27–38 (2017)
11. Knuth, D.: *The Art of Computing, Volume 2: Seminumerical Algorithms*. Addison-Wesley, Boston (1981)
12. Worsley, K.J., Cao, J., Paus, T., Petrides, M., Evans, A.C.: Applications of random field theory to functional connectivity. *Hum. Brain Mapp.* **6**, 364–367 (1998)