Bias in fMRI

Samuel Davenport

OxWaSP 2017

June 22, 2018

fMRI Model

1st level model

Suppose that we have a set of voxels: $\mathcal{V}$ eg the brain. For each subject, $j = 1, \ldots, m$, at each voxel v we have a vector of signal $\beta_j = \beta_j(v)$ where each entry corresponds to the signal under a certain stimulus condition.
We collect a vector of observations: Y_j and at each voxel we fit:

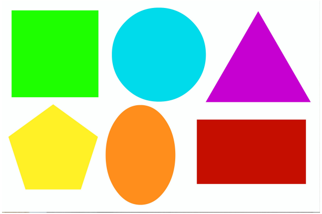
$$Y_j = X_j \beta_j + \epsilon_j.$$

This gives us a least squares estimate of β_j:

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T Y_j$$

Contrasts

We're often interested in the difference between stimulus conditions. Considering $c^T\beta_j$ for the contrast vector $c = [-1, 1, 0, 0, \dots, 0]$ allows us to identify the differences between the first two stimulus conditions. The data from the uk biobank presents subjects with faces in one stimulus condition and shapes in another.



2nd level model

We have a vector of contrasts: $\beta_c = [c^T\beta_1, \dots, c^T\beta_m]^T$. We would like to identify differences across groups of subjects, we fit the model:

$$\beta_c = X_g\beta_g + \eta$$

for some $m \times G$ group design matrix X_g and $G \times 1$ group difference vector β_g where G is the number of groups and noise η.

For $G = 1$, taking $X = 1_m$ to be a vector of ones, we have

$$\hat{\beta}_g = \frac{1}{m}\sum_j c^T\beta_j.$$

Two Sample Tests

In the case that $X = \begin{bmatrix} 1_{n_1} & \mathbf{0} \\ \mathbf{0} & 1_{n_2} \end{bmatrix}$, we have $\hat{\beta}_g = \left[\hat{\beta}_g^1 \;\; \hat{\beta}_g^2\right]^T$

$$\text{where } \hat{\beta}_g^1 = \frac{1}{n_1}\sum_{j=1}^{n_1} c^T\beta_j \text{ and } \hat{\beta}_g^2 = \frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2} c^T\beta_j.$$

Here what we're interested in is the difference between the group parameters β_g^1 and β_g^2. So we can use the difference $\hat{\beta}_g^1 - \hat{\beta}_g^2$ in order to test this.

Non-observability

However, $\hat{\beta}_c$ is not observable so we in practise use the estimate

$$\hat{\beta}_c := (c^T\hat{\beta}_1, \dots, c^T\hat{\beta}_n)$$

instead of

$$\beta_c = (c^T\beta_1, \dots, c^T\beta_n)$$

and do the regression

$$\hat{\beta}_c = X\beta + \eta + (\hat{\beta}_c - \beta_c) = X\beta + \epsilon$$

where $\epsilon = \eta + (\hat{\beta}_c - \beta_c)$. And we use $\frac{1}{m}\sum_j c^T\hat{\beta}_j$ for the one-sample statistic and $\frac{1}{n_1}\sum_{j=1}^{n_1} c^T\hat{\beta}_j - \frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2} c^T\hat{\beta}_j$ for the two-sample statistic.

Winner's Curse

Examples

Dice Example: Imagine you roll 10 fair dice and at random some of them show a 6. If you rolled them again would you expect them still to be 6?

Figure 1: Some Dice

Examples

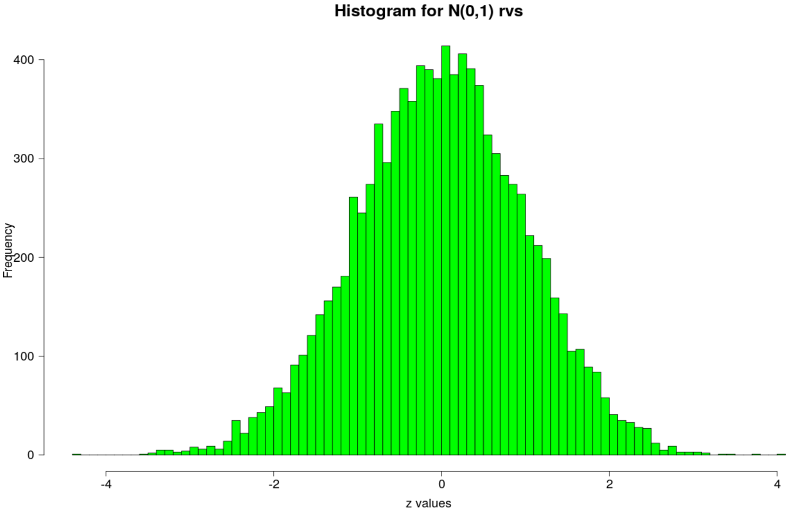
Dice Example: Imagine you roll 10 fair dice and at random some of them show a 6. If you rolled them again would you expect them still to be 6?

No you'd expect to obtain an average of:

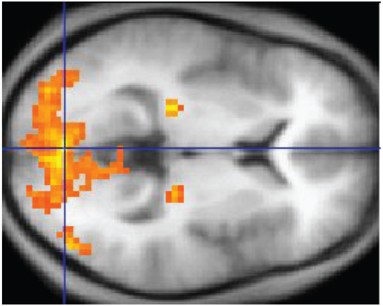
$$3.5 = (1 + 2 + 3 + 4 + 5 + 6)/6$$

Mean 0 example

Suppose for now that we have 10000 independent N(0,1) random variables. Then the largest are biased estimates for the true mean.



The Winner's Curse in fMRI



Choose significant voxels based on some statistic and its maxima.

(Vul et al., 2009)

Double dipping - circular inference

No activation

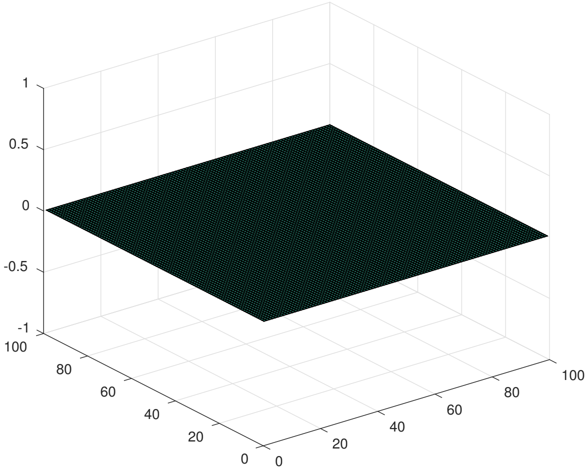


Figure 2: Zero true activation at each voxel.

Correlated Noise

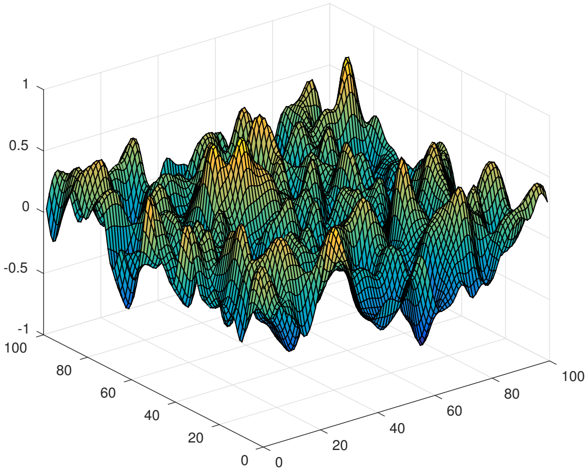


Figure 3: $Y(v) = \epsilon(v)$ where $\epsilon(v)$ is correlated gaussian noise with variance 1 that has been smoothed with an FWHM of 6.

Underlying Signal

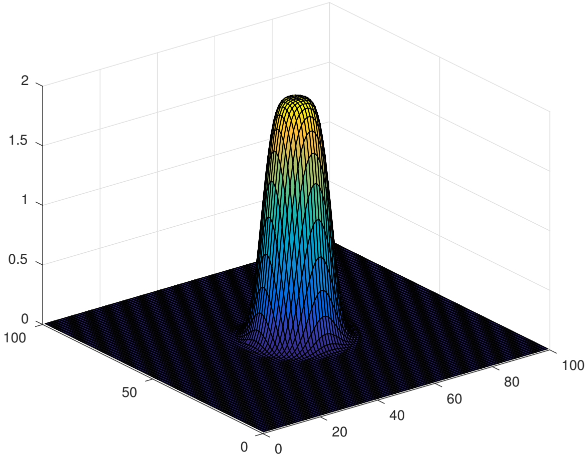


Figure 4: A cylindrical signal with maximum height 2. This is a map: $\mu : [1, 100] \times [1, 100] \to \mathbb{R}$.

Example Subject

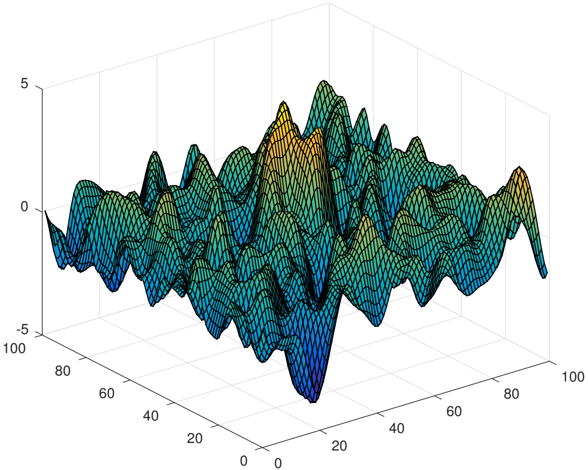
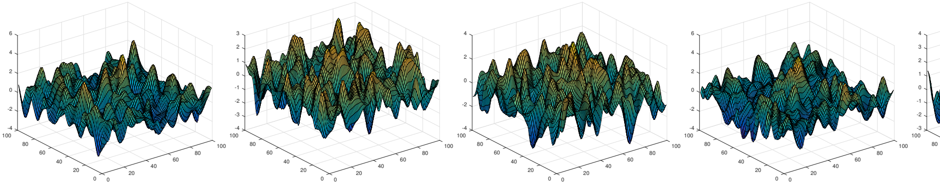


Figure 5: Signal of height 2 plus gaussian noise with variance 1 that has been smoothed with an FWHM of 6.

Many Subjects: the $c^T\hat{\beta}_i$ maps



The one-sample average: $\hat{\mu}$

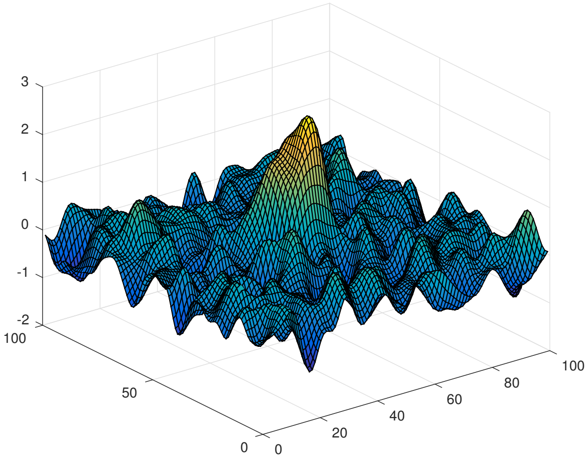


Figure 6: The average of the 9 subject maps. Notice how the maximum of this map is larger than 2.

Data-Splitting Approach

Split your subjects into two groups.

Data-Splitting Approach

Split your subjects into two groups.

Use half for significance and half for estimation of the effect size.

Solves the bias problem as have independence across subjects and since $\hat{\beta}_g$ is an unbiased estimate for β.

Issues: Less data to estimate so higher variance.

Ideally would like to have a method where you didn't have to sacrifice this data. Seems like magic but it is possible.!

Inference on the Mean

Spatial Maps

Suppose that we observe a noisy mean:

$$\hat{\mu}(v) = \mu(v) + \epsilon(v)$$

for some noise process ϵ and underlying mean μ which we wish to infer.

Defining the Bias

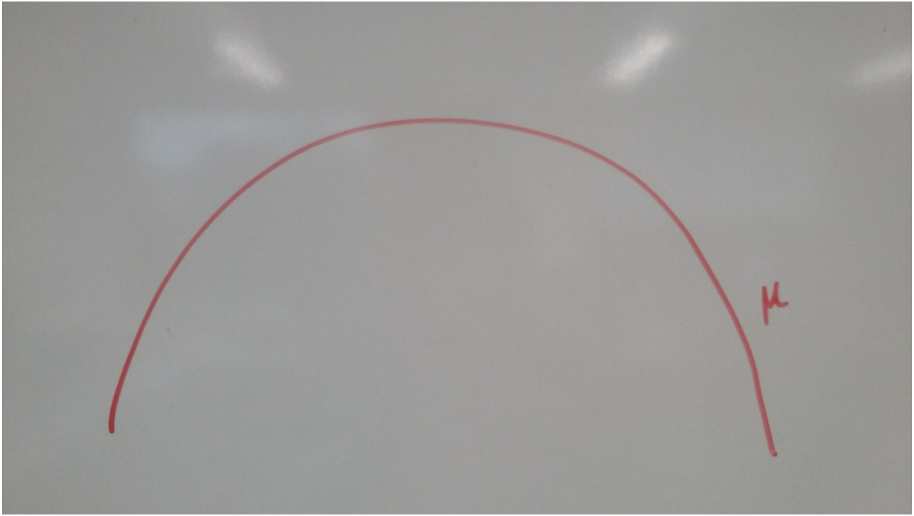


Figure 7: μ

Defining the Bias

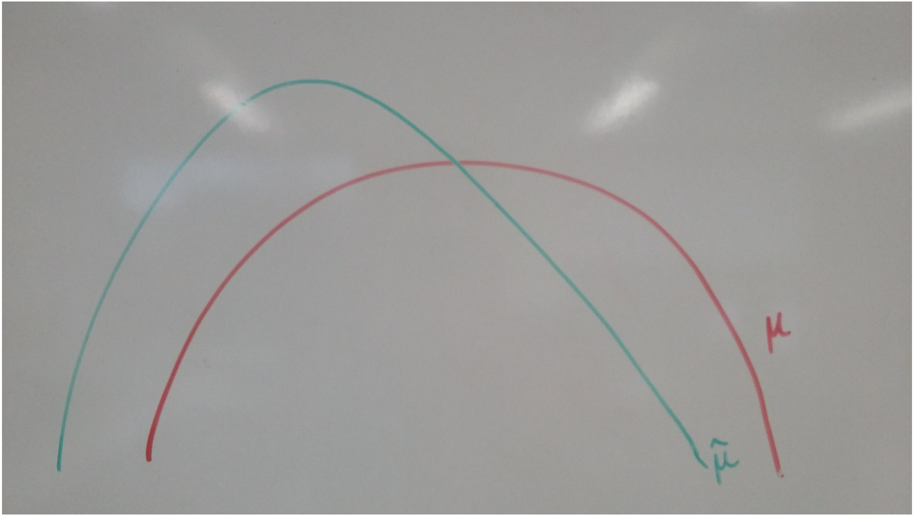


Figure 8: $\hat{\mu}$

Defining the Bias

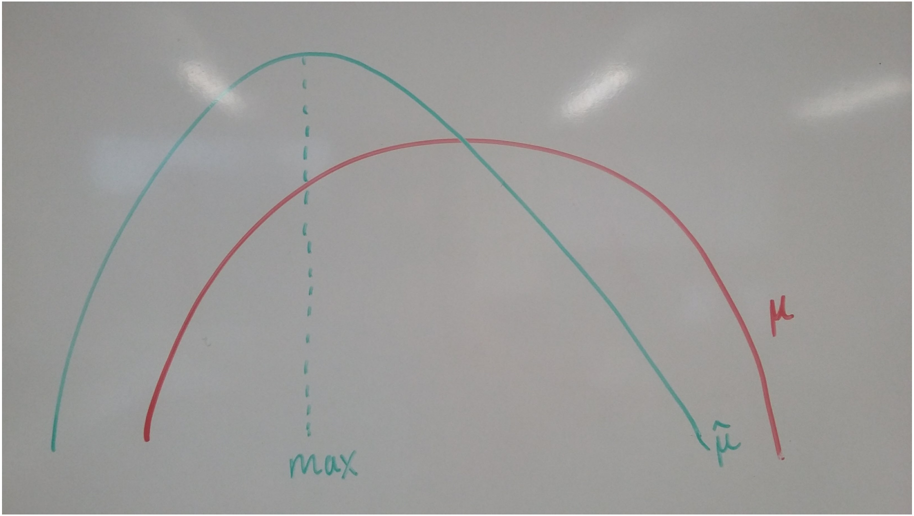


Figure 9: The location of the maximum of $\hat{\mu}$

Defining the Bias

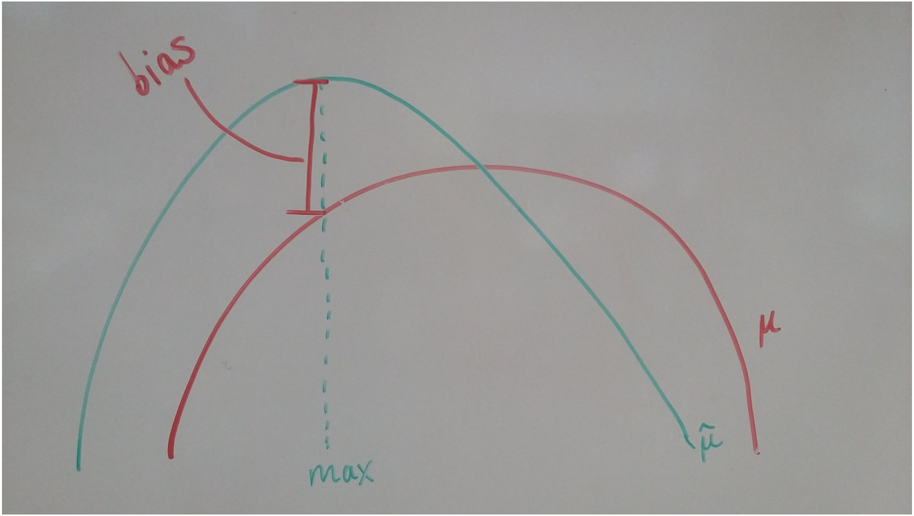


Figure 10: The bias that randomly arises from choosing the maximum.

Estimating the Bias from $\hat{\mu}$

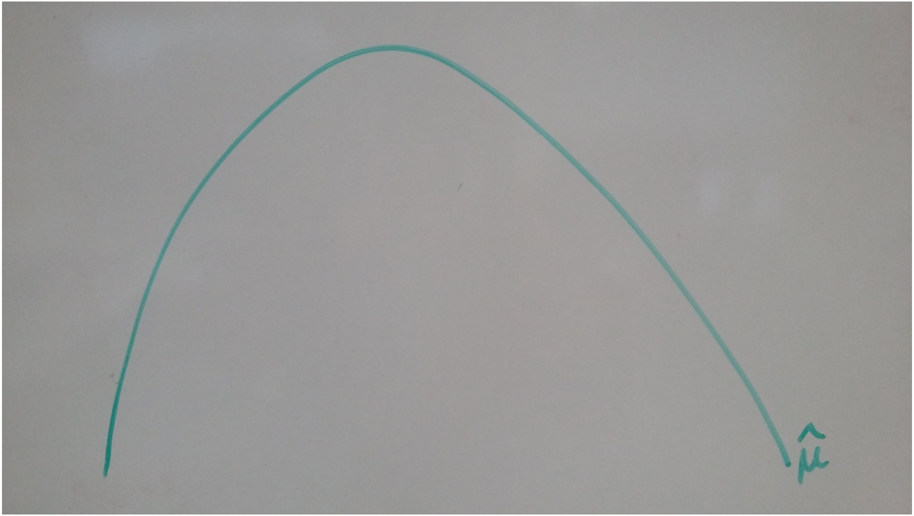


Figure 11: $\hat{\mu}$

Estimating the Bias from $\hat{\mu}$

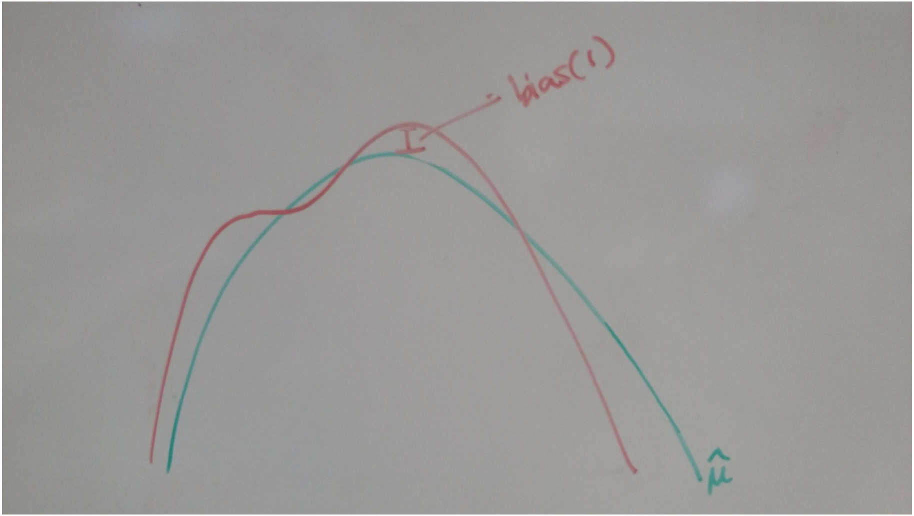


Figure 12: One iteration of the bias.

Estimating the Bias from $\hat{\mu}$

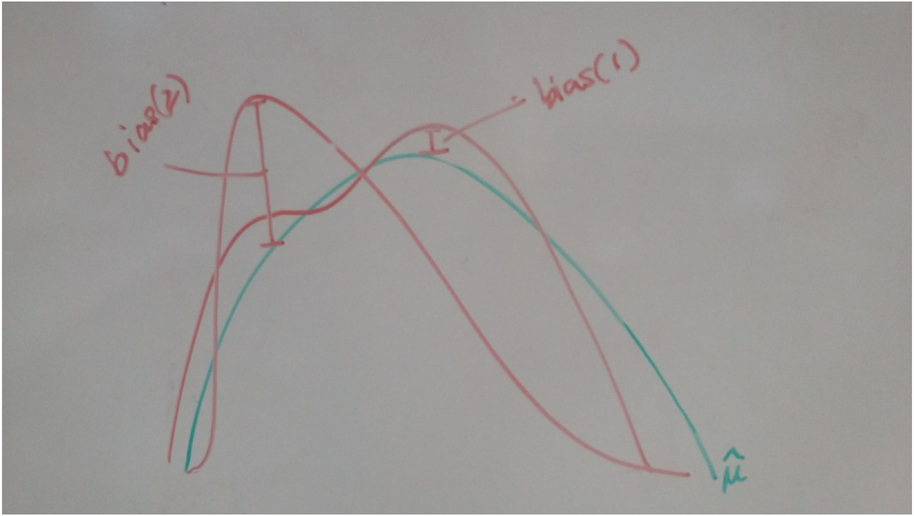


Figure 13: A second iteration of the bias.

Selective Inference Algorithm

Algorithm

Suppose we observe $\hat{\mu} = \mu + \epsilon$ for and wish to infer μ.

Algorithm 1 Parametric Bootstrap Bias Calculation

Input: $\hat{\mu}$ and some number of bootstrap iterations: N.
for $n = 1, \dots, N$ **do**
    Generate a normal smooth noise process: ϵ_b and let $\hat{\mu}_n = \hat{\mu} + \epsilon_n$.
    Find the location of the maximum of $\hat{\mu}_n$ and let $\hat{\mu}_n^{max}$ be its value. Let $v_{\max}$ be a 3D vector of the coordinates of this maxima such that $\hat{\mu}_n(v_{\max}) = \hat{\mu}_n^{max}$.
    Let the bias estimate be $B_n(\hat{\mu}) = \hat{\mu}_n(v_{\max}) - \hat{\mu}(v_{\max})$.
end for
Calculate $\hat{\delta} := \frac{1}{N}\sum_{n=1}^{N} B_n(\hat{\mu})$.
end for
return $\hat{\mu}^{max} - \hat{\delta}$.

Algorithm

Algorithm 2 Non-Parametric Bootstrap Bias Calculation

```
Input: Contrast images: Z_1 = c^T β_1, ..., Z_m = c^T β_m and some number of bootstrap iterations: N.
Let z = (1/m) Σ_{j=1}^m Z_j.
for n = 1, ..., N do
    Simulate Z_1^*, ..., Z_m^* independently with replacement from Z_1, ..., Z_m.
    Let y = (1/m) Σ_{j=1}^m Z_j^*.
    Find the maximum of y and let y_max be its value. Let i be a 3D vector of the coordinates of this maxima such that y(i) = y_max.
    Let the bias be B_n = y(i) − z(i).
end for
Calculate δ̂ := (1/N) Σ_{n=1}^N B_n.
end for
return z_max − δ̂.
```

Have a look at some of the code. Run: Run: dispres('mean', 50).

Figures illustrating the results.

The bias from the maximum.

One of the criticisms of Independent splitting is that it can only look locally, whereas the bootstrap can do more than that.

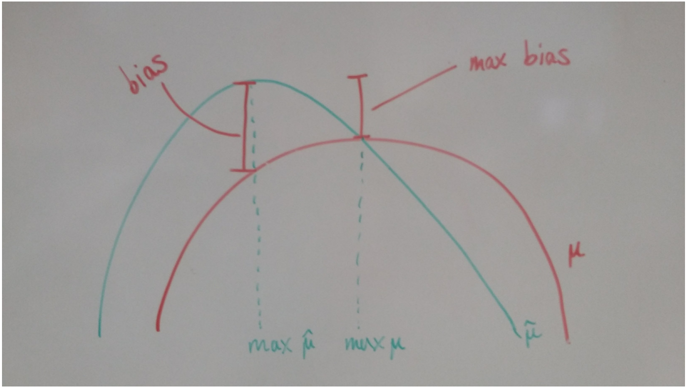


Figure 14: Should probably use the max bias instead. This can easily be changed in the algorithm. At each step calculating the bias: C_n where we have used C to allow us to distinguish from the previous bias estimates.

Theoretical Bias under the parametric bootstrap.

Suppose that we knew the true mean: μ, then if we know how the process ϵ is generated then by iterating we have a consistent estimator.
In the max case we have that C_n is a random variable with $\mathbb{E}C_n = c$.
Then our estimate is $\hat{\mu}(m_C) - \frac{1}{n}\sum_{n=1}^{N} C_n$.
Then

$$\mathbb{E}\left[\hat{\mu}(m_C) - \frac{1}{n}\sum_{n=1}^{N} C_n\right] = \mathbb{E}[\hat{\mu}(m_C) - \mu(m_C) + \mu(m_C)] - c$$

$$= c + \mathbb{E}[\mu(m_C)] - c = \mu(m_C),$$

as m_C is fixed.
In non-max case B_n is a random variable with $\mathbb{E}B_n = b$. Our estimate is: $\hat{\mu}(m) - \frac{1}{n}\sum_{n=1}^{N} B_n$

$$\mathbb{E}\left[\hat{\mu}(m) - \frac{1}{n}\sum_{n=1}^{N} B_n - \mu(m)\right] = \mathbb{E}[\hat{\mu}(m) - \mu(m)] - \frac{1}{n}\sum_{n=1}^{N} B_n = b - b = 0$$

where m is the location of the maximum of $\hat{\mu}$.

If you could simulate from μ, then could take n realizations and obtain a consistent estimator of the bias.

That's all folks.

Figure 15: Questions? :)

Bibliography