ROI Testing

Dominik Blank

Georg-August-Universität Göttingen

December 9, 2019

1. ROI Testing
   - The statistical model
   - Testing for the ROI
     - The statistical test
     - Statistical significance
     - Power of the test
     - Numerical results

2. Morphological operations
   - Opening and closing
   - Hypothesis testing and opening

ROI Testing

Let $M, N \in \mathbb{N}$ and $G = \{0, \dots, M-1\} \times \{0, \dots, N-1\}$. Assume we are given data

$$f(i,j) = c + v(i,j) + \varepsilon_{i,j}$$

- $(i,j) \in G$
- $c \in \mathbb{R}$ is constant
- $v : G \to \{0, \pm c\}$
- $\varepsilon_{m,n} \sim \mathcal{N}(0, \sigma^2)$ i.i.d. normal distributed random variables

Assumption 1: The image f contains a rectangular region of interest.

Assumption 2: The ROI has a checkerboard pattern.

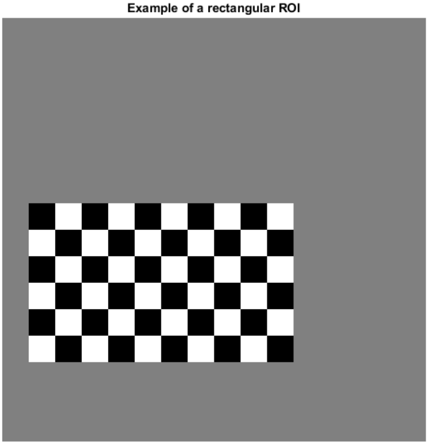


Figure: Example of a possible region of interest. ($M = 16$, $N = 16$)

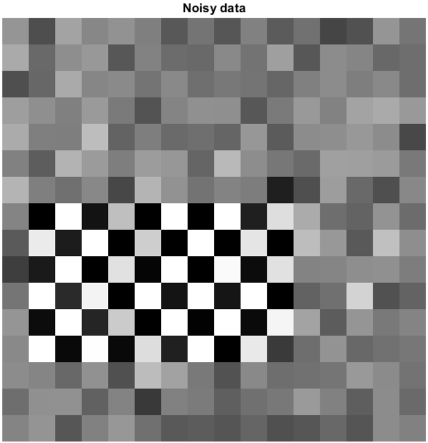


Figure: Same region of interest with noise added. ($\sigma = 30$)

Goal: Construct a statistical test for the region of interest.

For each pixel $(i, j) \in G$ we define four non-observable and four observable values

$$\begin{aligned} d_1^{\pm}(i,j) &= v(i \pm 1, j) - v(i,j) \\ d_2^{\pm}(i,j) &= v(i, j \pm 1) - v(i,j) \\ \tilde{d}_1^{\pm}(i,j) &= f(i \pm 1, j) - f(i,j) \\ \tilde{d}_2^{\pm}(i,j) &= f(i, j \pm 1) - f(i,j) \end{aligned}$$

and combine them to new values

$$\begin{aligned} d^{\pm}(i,j) &= \sqrt{d_1^{\pm}(i,j)^2 + d_2^{\pm}(i,j)^2} \\ \tilde{d}^{\pm}(i,j) &= \sqrt{\tilde{d}_1^{\pm}(i,j)^2 + \tilde{d}_2^{\pm}(i,j)^2} \end{aligned}$$

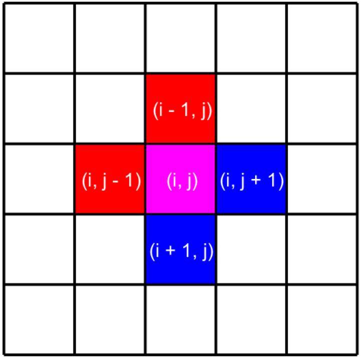


Figure: Pixel (i,j) with its neigbour pixels. The blue neighbours are used to calculate $\tilde{d}^{+}(i,j)$ and the red neighbours are used to calculate $\tilde{d}^{-}(i,j)$.

We now test the null hypothesis

$$H_0 : \min\{d^+(i,j), d^-(i,j)\} = 0$$

against the alternative hypothesis

$$H_1 : \min\{d^+(i,j), d^-(i,j)\} > 0$$

using the test statistic

$$T = \min\{\tilde{d}^+(i,j), \tilde{d}^-(i,j)\}$$

Question: How do we ensure a given statistical significance?

We observe that

$$\mathbb{P}(T \geq t \mid H_0) \leq \mathbb{P}(\tilde{d}^{\pm}(i,j) \geq t \mid d^{\pm}(i,j) = 0)$$

Thus, if we can find a threshold t_α, s.t. $\mathbb{P}(\tilde{d}^{\pm}(i,j) \geq t_\alpha \mid d^{\pm}(i,j) = 0) \leq \alpha$ for a given α, we have ensured a statistical significance of α.

Let $\varepsilon_{i,j}$ be fixed. Then

$$\begin{aligned}
&\mathbb{P}(\tilde{d}^{\pm}(i,j) \leq t \mid d^{\pm}(i,j) = 0) \\
&= \mathbb{P}\left((\varepsilon_{i\pm1,j} - \varepsilon_{i,j})^2 + (\varepsilon_{i,j\pm1} - \varepsilon_{i,j})^2 \leq t^2\right) \\
&= \mathbb{P}\left(\sqrt{\left(\frac{X_1}{\sigma}\right)^2 + \left(\frac{X_2}{\sigma}\right)^2} \leq \frac{t}{\sigma}\right)
\end{aligned}$$

with

$$\begin{aligned}
X_1 &= \varepsilon_{i\pm1,j} - \varepsilon_{i,j} \sim \mathcal{N}(-\varepsilon_{i,j}, \sigma^2) \\
X_2 &= \varepsilon_{i,j\pm1} - \varepsilon_{i,j} \sim \mathcal{N}(-\varepsilon_{i,j}, \sigma^2)
\end{aligned}$$

Then we can see, that the square root inside has non-central chi distribution with two degrees of freedom and non-centrality parameter

$$\lambda = \frac{\sqrt{2}|\varepsilon_{i,j}|}{\sigma}$$

So far we assumed $\varepsilon_{i,j}$ to be fixed, but it actually is a random variable itself with normal distribution: $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$. Thus, we have a compound probability distribution:

$$
\begin{aligned}
&\mathbb{P}(\tilde{d}^{\pm}(i,j) \leq t \mid d^{\pm}(i,j) = 0) \\
&= \int_0^{\frac{t}{\sigma}} \int_0^{\infty} \underbrace{x \exp\left(-\frac{x^2}{2} - \frac{\eta^2}{\sigma^2}\right) I_0\left(\frac{\sqrt{2}\eta}{\sigma} x\right)}_{\text{pdf of } \chi_2\left(\frac{\sqrt{2}\eta}{\sigma}\right) \text{ for fixed } \eta} \quad \underbrace{\frac{2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\eta^2}{2\sigma^2}\right)}_{\text{pdf of absolute value of } \mathcal{N}(0,\sigma^2)} \quad d\eta dx \\
&= \frac{1}{\sqrt{3}}\left(\frac{3}{2} - \frac{3}{2}\exp\left(-\frac{t^2}{3\sigma^2}\right) I_0\left(\frac{t^2}{6\sigma^2}\right)\right) - \sqrt{3} \\
&- \frac{2-\sqrt{3}}{2} Q_1\left(\frac{2-\sqrt{3}}{6}\sqrt{\frac{t}{\sigma}}, \frac{2+\sqrt{3}}{6}\sqrt{\frac{t}{\sigma}}\right) \\
&+ \frac{2+\sqrt{3}}{2} Q_1\left(\frac{2+\sqrt{3}}{6}\sqrt{\frac{t}{\sigma}}, \frac{2-\sqrt{3}}{6}\sqrt{\frac{t}{\sigma}}\right)
\end{aligned}
$$

With an easy program the threshold for a given α can be calculated. As can be seen from the CDF, it is enough to calculate the threshold t_α for $\sigma = 1$ and multiply the outcome by σ.

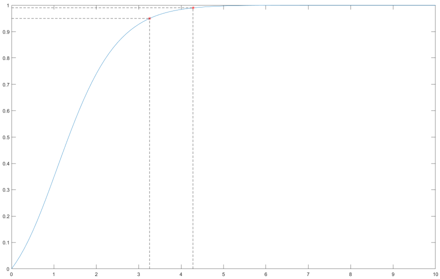


Figure: Cumulative distribution function of $\mathbb{P}(\tilde{d}^{\pm}(i,j) \leq t \mid d^{\pm}(i,j) = 0)$ with thresholds for $\alpha = 0.05$ and $\alpha = 0.01$. ($t_{0.05} = 3.2554$, $t_{0.01} = 4.2791$, $\sigma = 1$)

We are also interested in bounds for the power of this test.
Using results and notations from the previous sections, we get the upper bound

$$1-\beta = \mathbb{P}(T \geq t \mid H_1) \leq \mathbb{P}(\tilde{d}^+ \geq t \mid d^+ = \sqrt{8}c)$$

On the other hand, we get the lower bound

$$1-\beta = \mathbb{P}(T \geq t \mid H_1) \geq 1 - 2 \cdot \mathbb{P}(\tilde{d}^+ \leq t \mid d^+ = c)$$

Thus we can conclude, that

$$1-\beta \in \left[\max\left\{1 - 2 \cdot \mathbb{P}(\tilde{d}^+ \leq t \mid d^+ = c), 0\right\}, \mathbb{P}(\tilde{d}^+ \geq t \mid d^+ = \sqrt{8}c)\right]$$

In the case of a grayscale image, we assume $c = 127.5$. For $t_{0.05} = 3.2554$ and 10.000.000 simulations of the noise terms, we get the following empirical bounds dependent on the standard deviation σ.

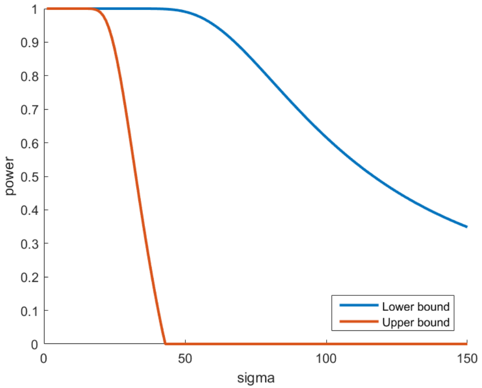


Figure: For $\alpha = 0.05$ this graph shows the lower and upper bounds for the power of the test for $\sigma \in \{1, 2, \ldots, 150\}$.

Since $d^{\pm}$ is the sum of dependent squared normal distributed random variables, it is not a simple Chi distributed random variable. If it were, we would get

$$1 - \beta \in \left[\max\left\{1 - 2 \cdot \left(1 - Q_1\left(\frac{c_{bg}}{\sqrt{2}\sigma}, \frac{t}{\sqrt{2}\sigma}\right)\right), 0\right\}, Q_1\left(\frac{2c_{bg}}{\sigma}, \frac{t}{\sqrt{2}\sigma}\right)\right]$$

In the independent case we would take $t = 2\sigma\sqrt{-\log(\alpha)}$ and get the following bounds for the power dependet on σ. It is remarkable, how similar these bounds are to the empirical bounds.

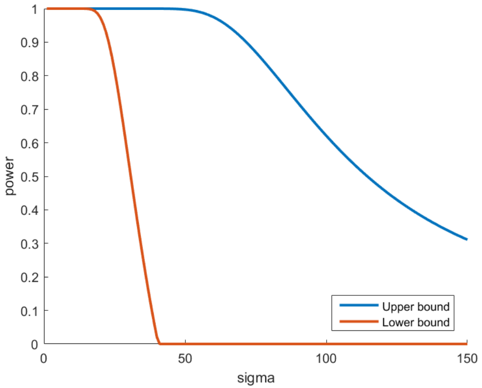


Figure: For $\alpha = 0.05$ this graph shows the lower and upper bounds for the power of the test for $\sigma \in \{1, 2, \ldots, 150\}$, if the terms of $\tilde{d}^{\pm}$ were independent.

Morphological operations

Reminder: We always consider structuring elements with odd side length.

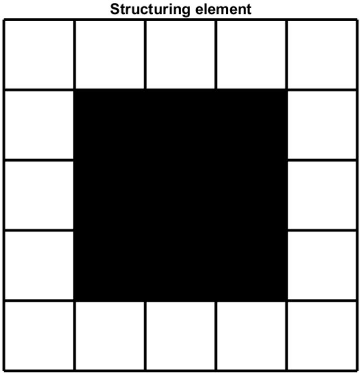
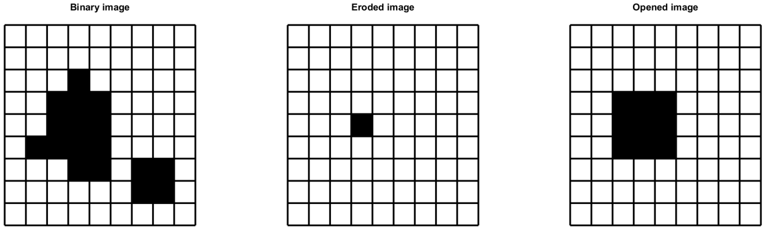
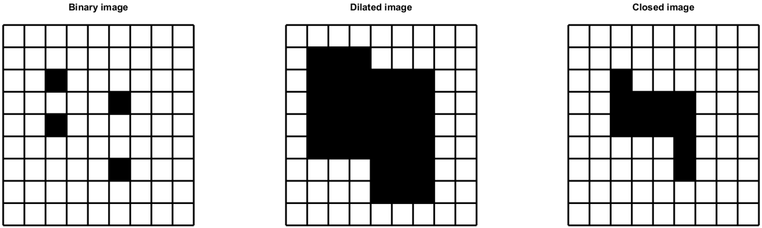


Figure: A 3×3 structuring element.



Example of a binary image (black boxes represent 1). The second image is the erosion of the image by a 3×3 structuring element. The third image is the dilation of the erosion, i.e. the opening of the image.



Example of a binary image (black boxes represent 1). The second image is the dilation of the image by a 3×3 structuring element. The third image is the erosion of the dilation, i.e. the closing of the image.

Question: What is the effect of opening and closing on statistical significance and power?

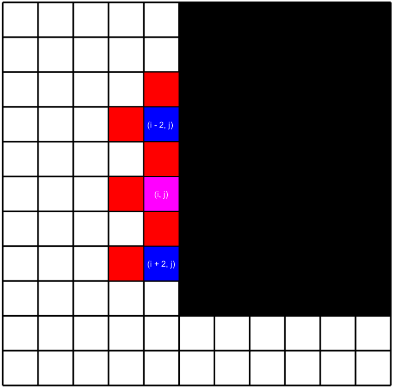


Figure: If we consider only every second pixel in the column, these will be independent.

We will now take a look at the effect of opening on the significance level.

Theorem

Let f be an image that contains a rectangular ROI. Assume that we are given a binarized image f_{bin} with

$$\mathbb{P}(f_{bin}(i,j) = 1 \mid H_0(i,j)) \leq \alpha$$

where $H_0(i,j)$ denotes the null hypothesis for the pixel (i,j).
Let $k \in \mathbb{N}$ be odd and B be a square structuring element with side length k. Then the following inequality holds:

$$\mathbb{P}((f_{bin} \circ B)(i,j) = 1 \mid H_0(i,j)) \leq k^2 \alpha^{\frac{k-1}{2}}$$