

STATISTISCHE UNTERSTÜTZUNG BEI DER KREBSDIAGNOSTIK



Philipp Hermann & Milan Stehlik
Institut für Angewandte Statistik

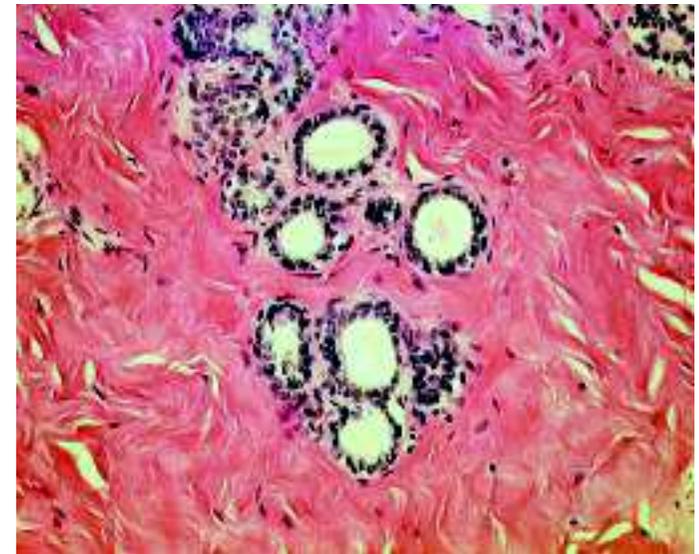
Idee: Entwickelt zwischen 1980 - 2000

- Prof. Mattfeldt ist mit folgendem Problem gekommen: er - als Histopathologe - soll entscheiden, welches Gewebe Krebs oder Mastopathie ist.
- Sein Wunsch war es, eine statistische Methode zu bekommen, die ihn bei dieser Unterscheidung unterstützt.
- Was ist Mastopathie?



Mastopathie

- Diabetische Mastopathien sind nicht-krebserregende Veränderungen / Läsionen in der Brust.
- Gründe für diese Veränderungen sind bisher unbekannt.
- Symptome sind festes, unregelmäßiges, leicht bewegliches, jedoch schmerzloses Brustgewebe.



Brustkrebs vs Mastopathie

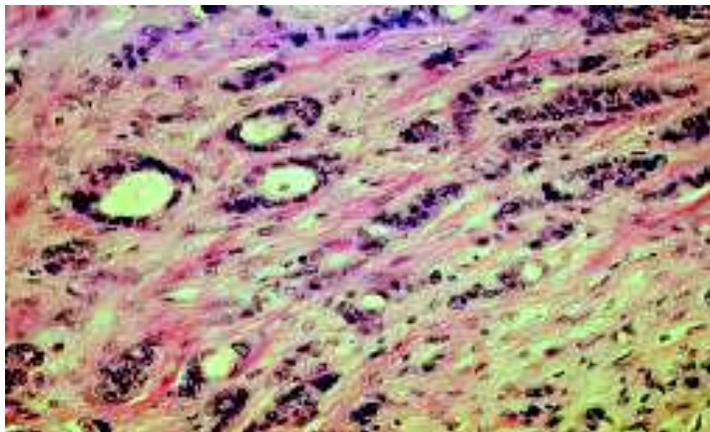
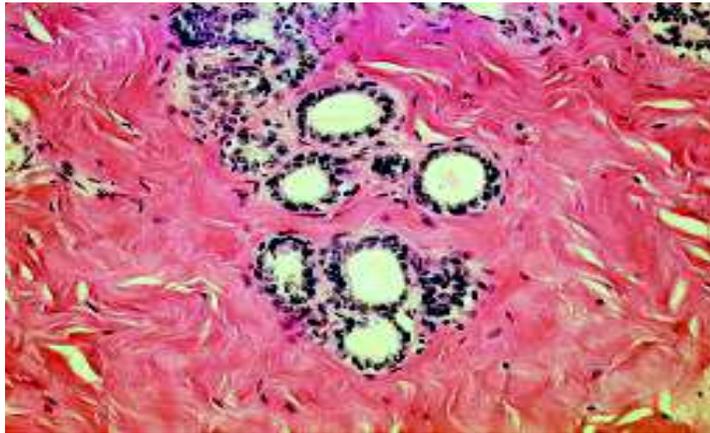


Abbildung: Mastopathie (oben) und Brustkrebs (unten) Zellproben.
Links unter Mikroskop - Rechts mit CCD-Kamera entwickelt.

Problem: Unterscheidung zwischen Krebs und Mastopathie

- Sehr altes Problem, da die geometrischen Eigenschaften ähnlich sind.
- Beide haben Höhlen/Löcher, aber sie sind in Folgendem unterschiedlich:
 1. Anzahl der Höhlen/Löcher
 2. Struktur der Höhlen/Löcher
 3. Platzierung der Höhlen/Löcher
- *Fraktale* Hypothese wurde von 2009 bis 2012 erforscht (Milan Stehlík).

Definition von Fraktalen

Definition: Fraktal (Mandelbrot, 1982)

Mathematische Objekte werden als **Fraktale** bezeichnet, falls:

- sie eine Art von Selbstähnlichkeit (welche unendlich oft wiederholt werden kann) besitzen.
- sie eine rekursive Struktur besitzen.
- deren Dimension keine natürliche Zahl ist.

Erinnerung: Flächen haben Dimension 2, Linien haben Dimension 1.

“Unsere” Fraktale haben eine Dimension zwischen 1 und 2.



Krebs und Mastopathie als Fraktale

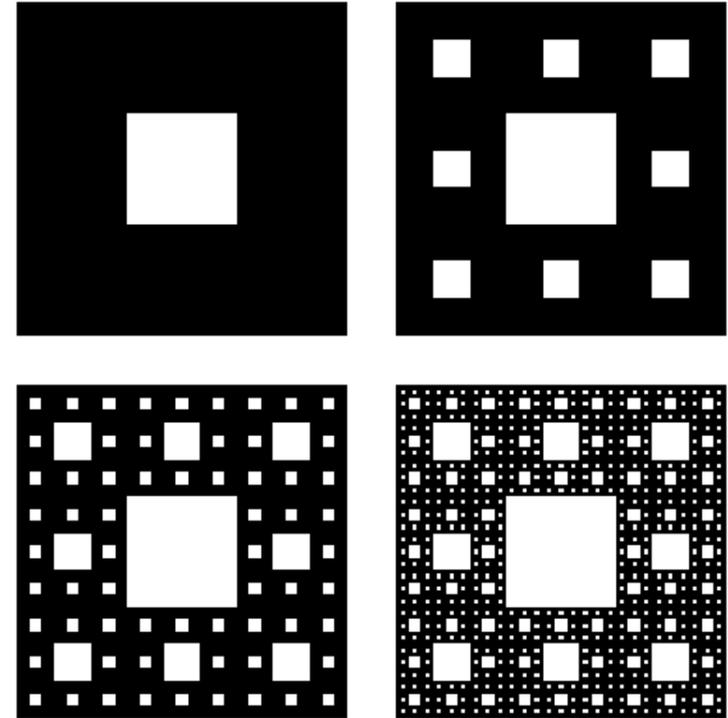
- Selbstähnlichkeiten sind die wichtigste Eigenschaft von *Fraktalen* (wie Brokkoli).
- Krebs und Mastopathie sind also selbstähnlich und können daher als Fraktale angesehen werden.
- Hypothese: “je höher die Dimension, desto abnormaler ist das Gewebe” (Tambasco, Eliasziw and Magliocco 2010).



Sierpinski-Teppich (1916) als mathematisches Fraktal

■ Ausgangspunkt ist ein Quadrat:

1. Wir dividieren jedes Quadrat in neun gleich große Quadrate
2. Wir entfernen das mittlere (kleinere) Quadrat.
3. Wir wiederholen die Schritte 1 und 2 für jedes der neu entstandenen (noch vorhandenen) Quadrate.



■ *Box-Counting Dimension:*

$$D = \frac{\ln(N^r)}{\ln(3^r)} = \frac{\ln(8^4)}{\ln(3^4)} = 1.89$$

Abbildung: Entstehung eines Sierpinski-Teppichs

Animation eines Sierpinski-Teppichs

Brustkrebs und mathematische Fraktale

- Stehlík hat (mit Studenten) begonnen den Sierpinski-Teppich zu untersuchen.
- Warum Sierpinski-Teppich?
- *Weil er Dimension 1.89 hat → diese Dimension ist die Grenzdimension für die Mastopathie oder gesundes Gewebe.*
- Lebendes Gewebe hat Dimension 1.74 und Brustkrebs Dimension > 1.89 .
- *Sierpinski-Teppich ist jedoch zu regelmäßig im Vergleich zu Krebsgewebe.*

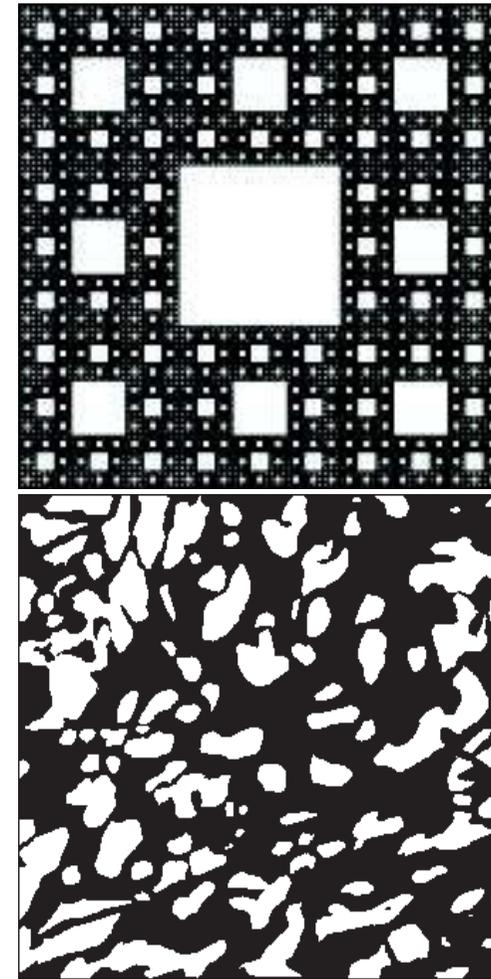
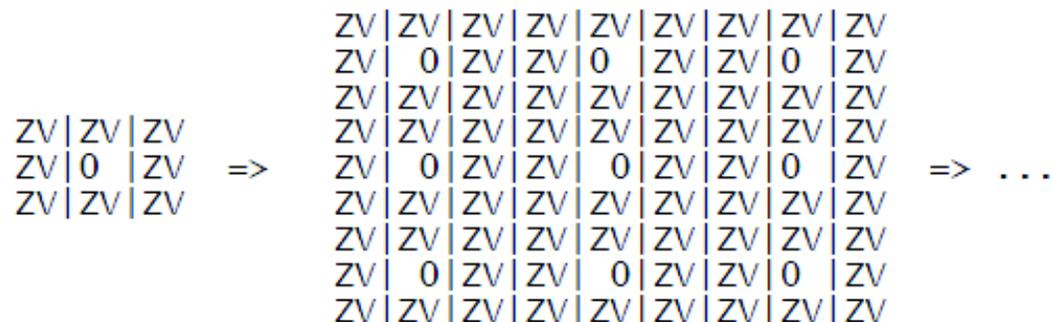


Abbildung: Sierpinski-Teppich (oben) und Krebsgewebe (unten).

Random Fractal Model: Instrument für die Unterscheidung

- Untersucht im Rahmen der Bachelor- sowie Masterarbeit von Philipp Hermann bei Milan Stehlik von 2012-2015.
- Zufallsvariablen (ZV) werden gemäß der zwei Gruppen gewählt.
- Validierungsmodelle wie **Boolean**, **Quermass Interaction Process**, **Hausdorff Measure**, **Wavelets** führen die Zufälligkeit in das Modell des (zu *regelmäßigen*) Sierpinski-Teppich ein.
- *Diese Methode verwendet Geometrie, Analysis, Algebra und Statistik.*
- Folglich können wir das unregelmäßige Verhalten annähernd nachstellen (auch für medizinische Zwecke sinnvoll).



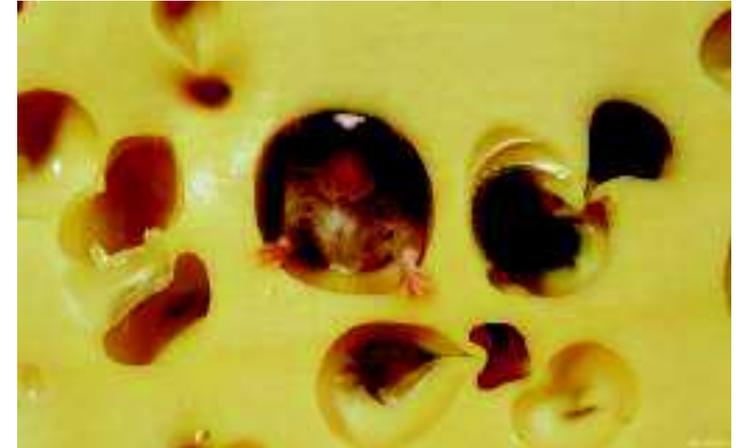
Wavelets: (Stehlik, Nicolis 2014-2015)

- Wavelets mischen verschiedene Formen von Fraktalen.
- Wavelet Methoden unterscheiden auch zwischen Mastopathie und Brustkrebs.



Quermass-Interaction (Stehlik-Mrkvicka 2010-12) I

- **Quermass interaction process:** ist eine zufällige Menge eines Raums \mathbf{X} , gegeben durch eine Vereinigung von interagierenden Kreisen mit zufälligen Radien und zufälligen Zentren.



Eine Konfiguration $\mathbf{x} = (x_1; \dots, x_n)$ von Kreisen x_1, \dots, x_n , aus \mathbf{X} wird durch die Dichte $f(\mathbf{x})$ beschrieben. Es bezieht sich auf die stationären zufälligen Kreise des **Boolschen Modells** mit Intensität der Kreiszentren und der Verteilung Q der Kreisradien.

$$f(\mathbf{x}) = \exp(\beta^T T(U\mathbf{x}))/c(\beta), \quad (1)$$

$T(U\mathbf{x}) = (A(U\mathbf{x}); L(U\mathbf{x}); c(U\mathbf{x}))$ sind geometrische Charakteristiken der Vereinigung $U\mathbf{x}$, wobei A das gesamte Volumen, L ist der Umfang und c eine Euler-Poincare Charakteristik ist.

Hausdorff Measure und genetische Algorithmen für Krebsforschung (Stehlik, Minarova 2009-12)

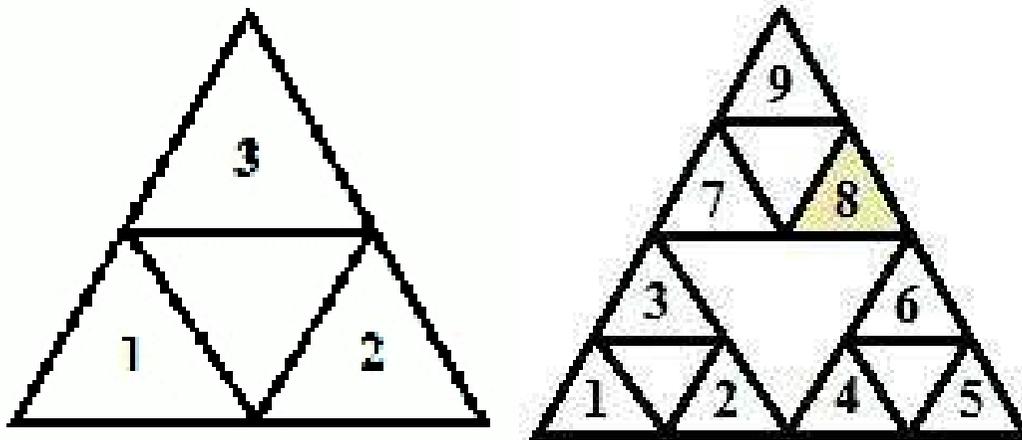


Abbildung: Erste und zweite Ramifikation (gegen Uhrzeigersinn)

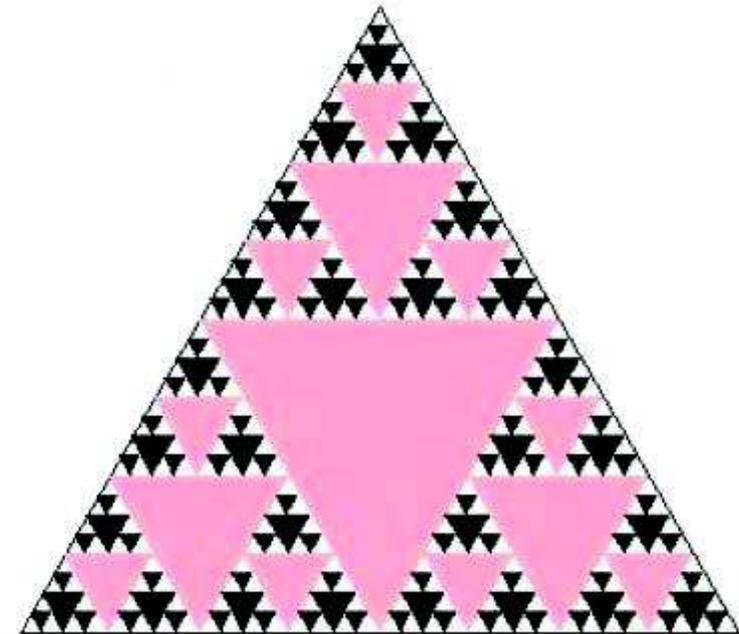


Abbildung: Sierpinski-Dreieck; 5. Ramifikation

$$S_1(x, y) = (x, y), S_2(x, y) = (x, y) + (l/2, 0), S_3(x, y) = (x, y) + (l/4, l\sqrt{3}/4). \quad (2)$$

Entwicklung der Methode

Boolsches Modell: (Mrkvicka and Mattfeldt 2011) ist jedoch nicht passend für die Modellierung von lebendem Gewebe.



Quermass-interaction process: (Stehlik and Mrkvicka 2010-2012)



Random Fractal Modell: (Stehlik and Hermann 2012-2015) Bernoulliverteilte Zufallsvariablen werden für jede (außer der jeweils mittleren) Box simuliert.

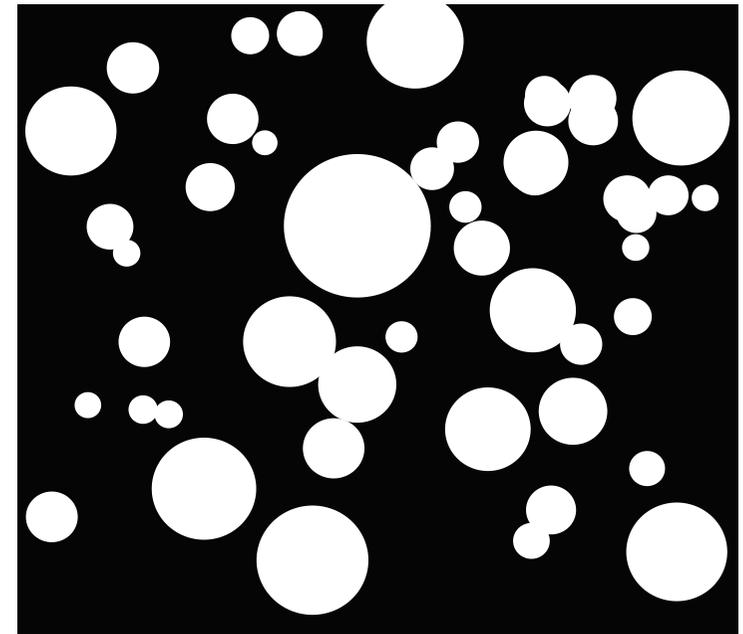


Abbildung:
Quermass-interaction process

Simulationsbeispiele

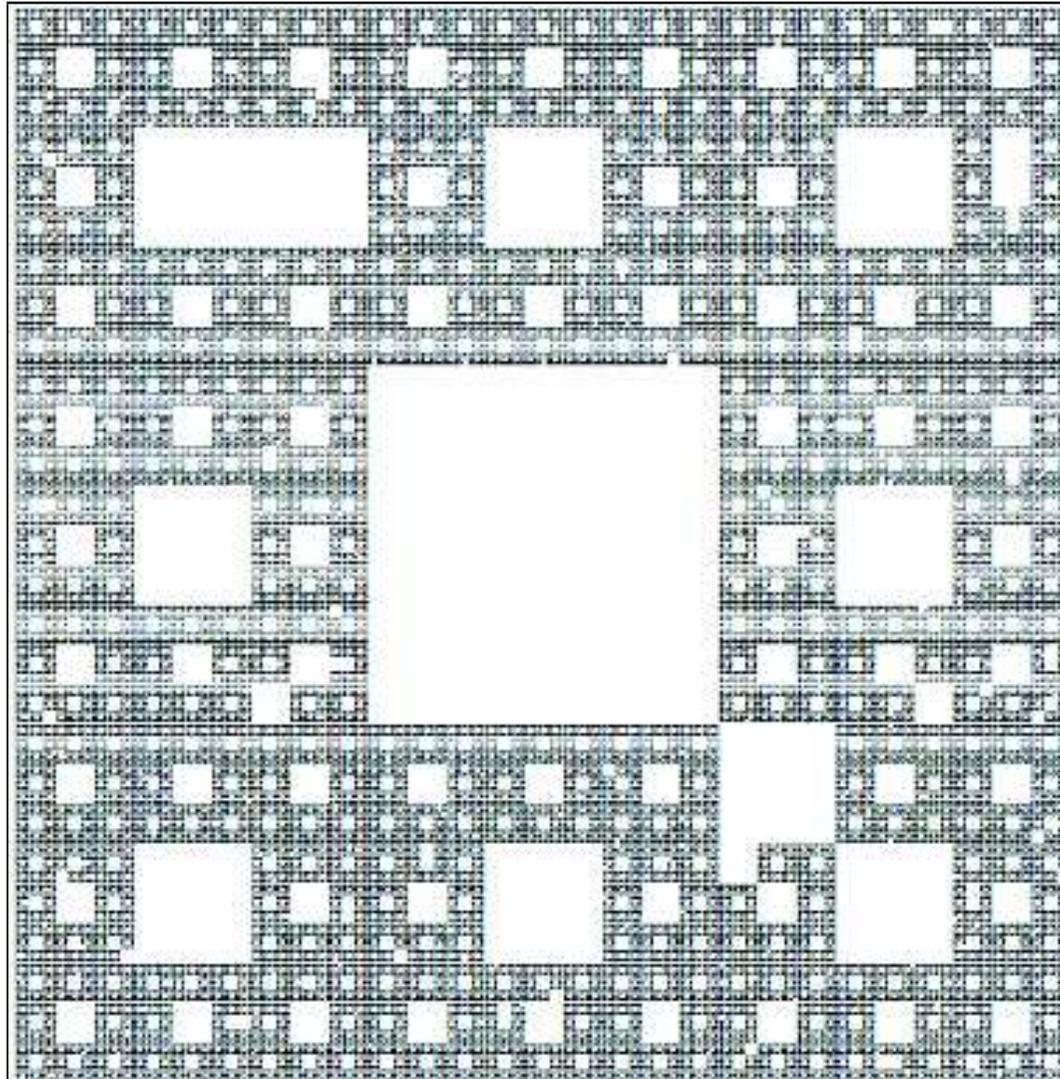


Abbildung: Beispiel mit $p = 0.01$

Simulationsbeispiele

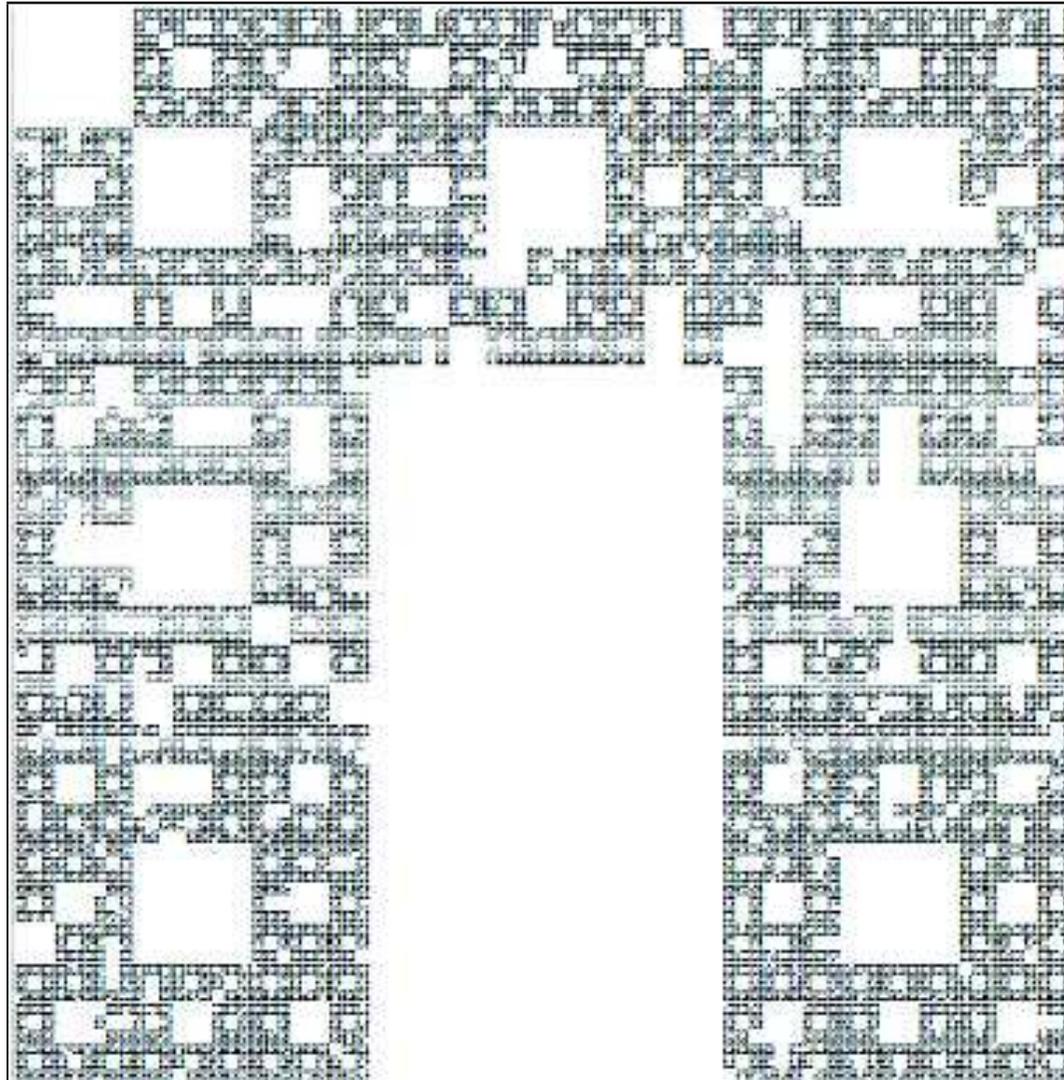


Abbildung: Beispiel mit $p = 0.05$

Simulationsbeispiele

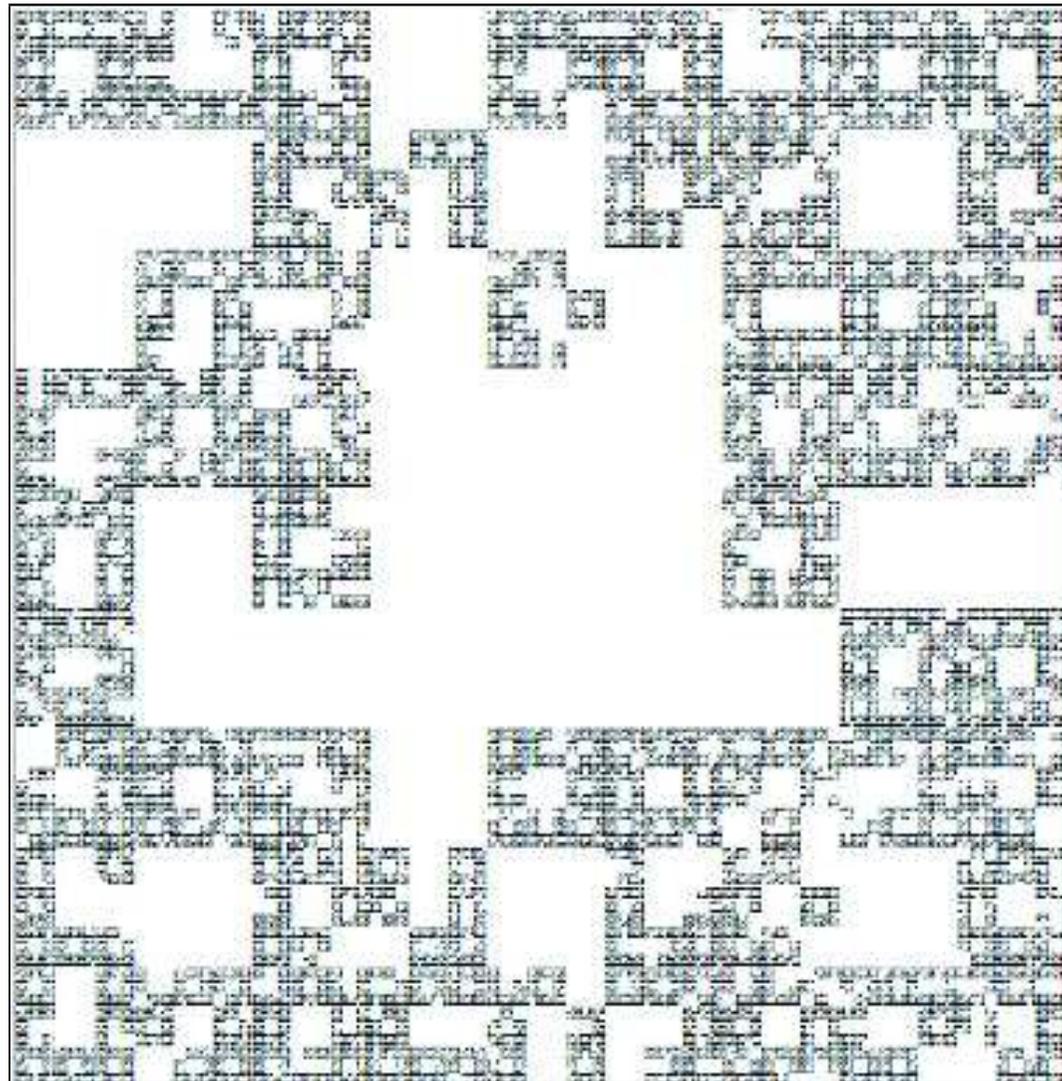


Abbildung: Beispiel mit $p = 0.1$

Simulationsbeispiele

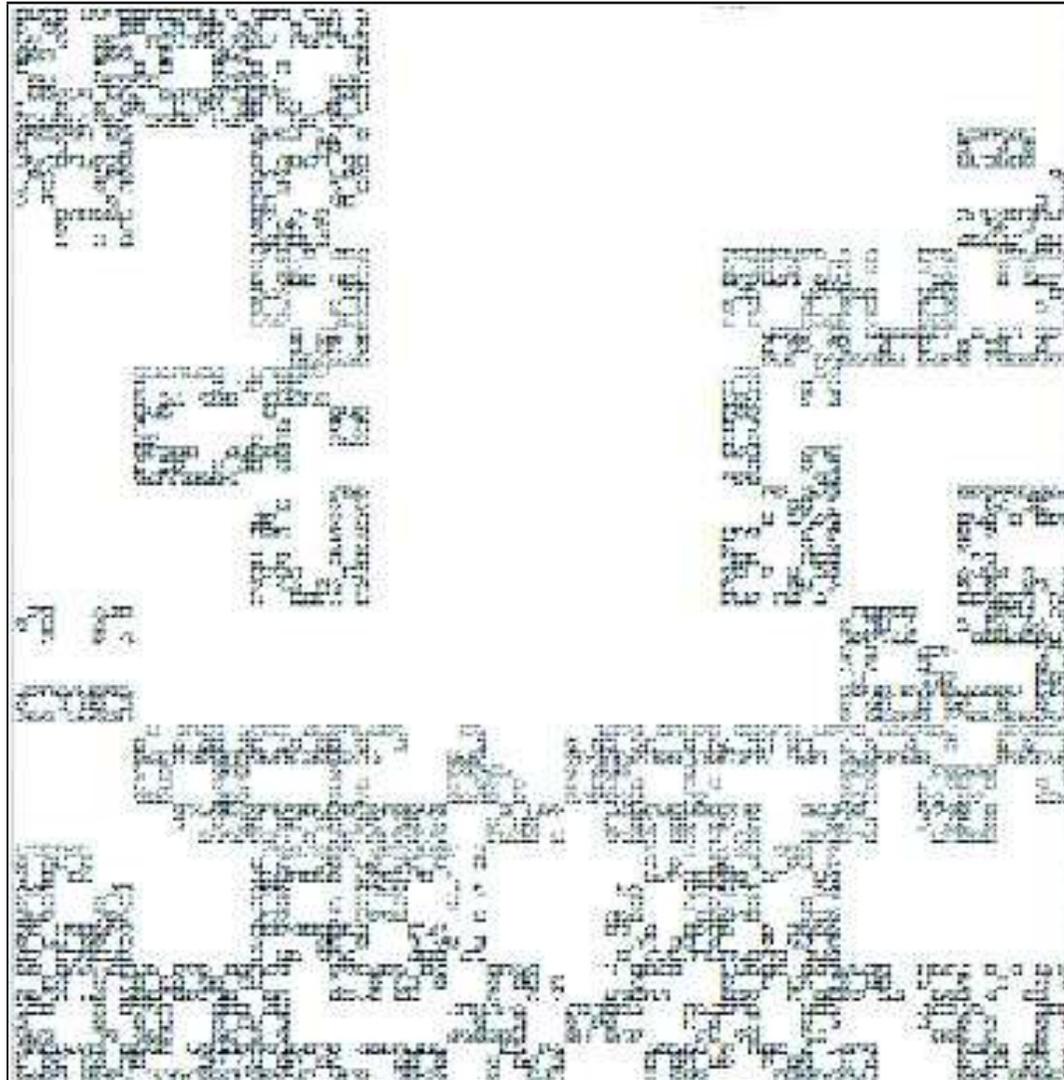


Abbildung: Beispiel mit $p = 0.2$

Simulationsbeispiele

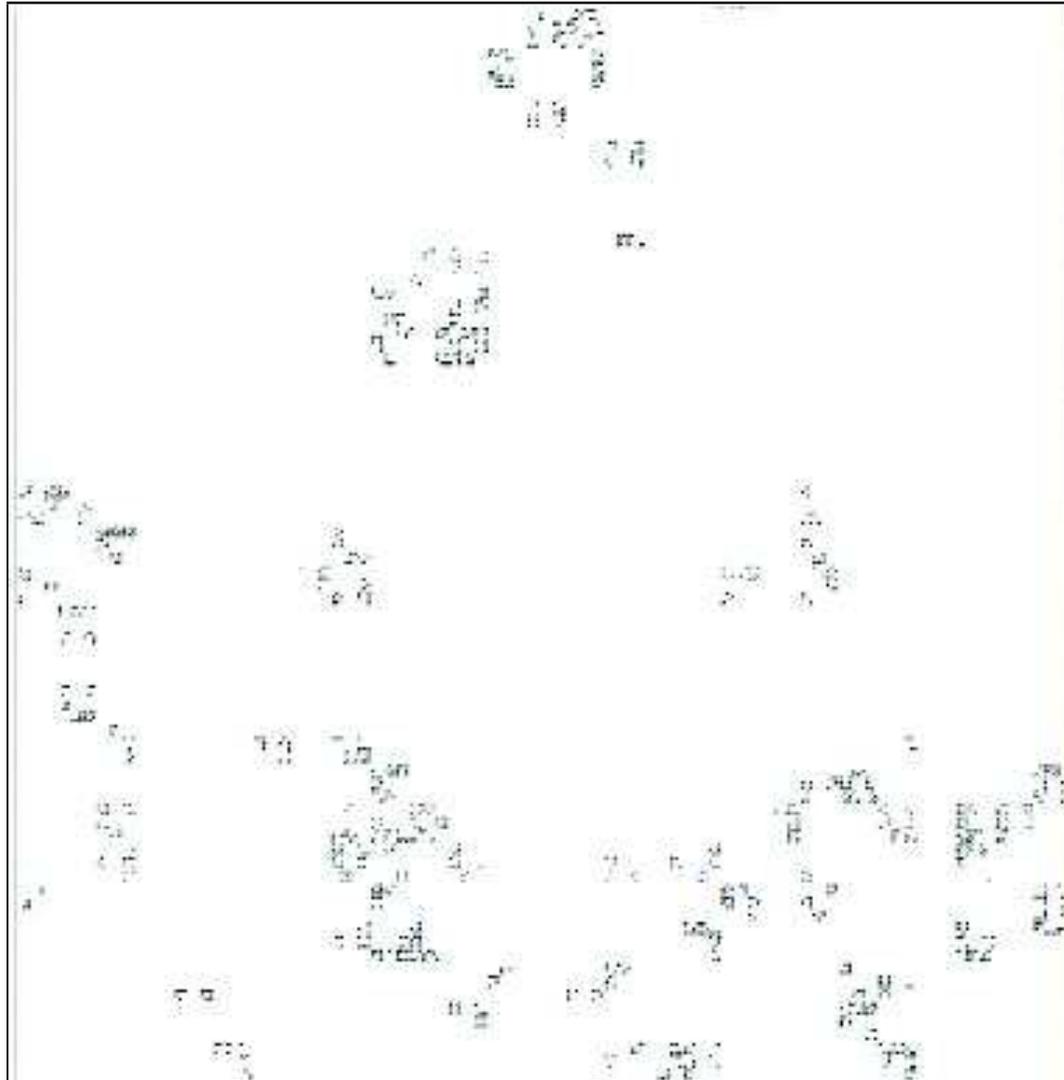


Abbildung: Beispiel mit $p = 0.5$

[p-p-p-q]-Modell: Erweiterung um einen Parameter

⇒ *bessere Annäherung an das Gewebe möglich.*

- Jeder Box wird eine Zufallsvariable zugeteilt.
- Wir definieren die Anzahl der Wiederholungen mit 4.
- Wir erweitern um den Parameter q (zusätzlich zu p).
- Daher ergibt sich das **[p-p-p-q]-Modell**



Abbildung: Gewebebild mit CCD-Kamera von Brustkrebs

Simulationsbeispiele für [p-p-p-q]-Modell



Abbildung: Beispiel mit $p = q = 0.1$

Simulationsbeispiele für [p-p-p-q]-Modell



Abbildung: Beispiel mit $p = 0.1, q = 0.6$

Simulationsbeispiele für [p-p-p-q]-Modell

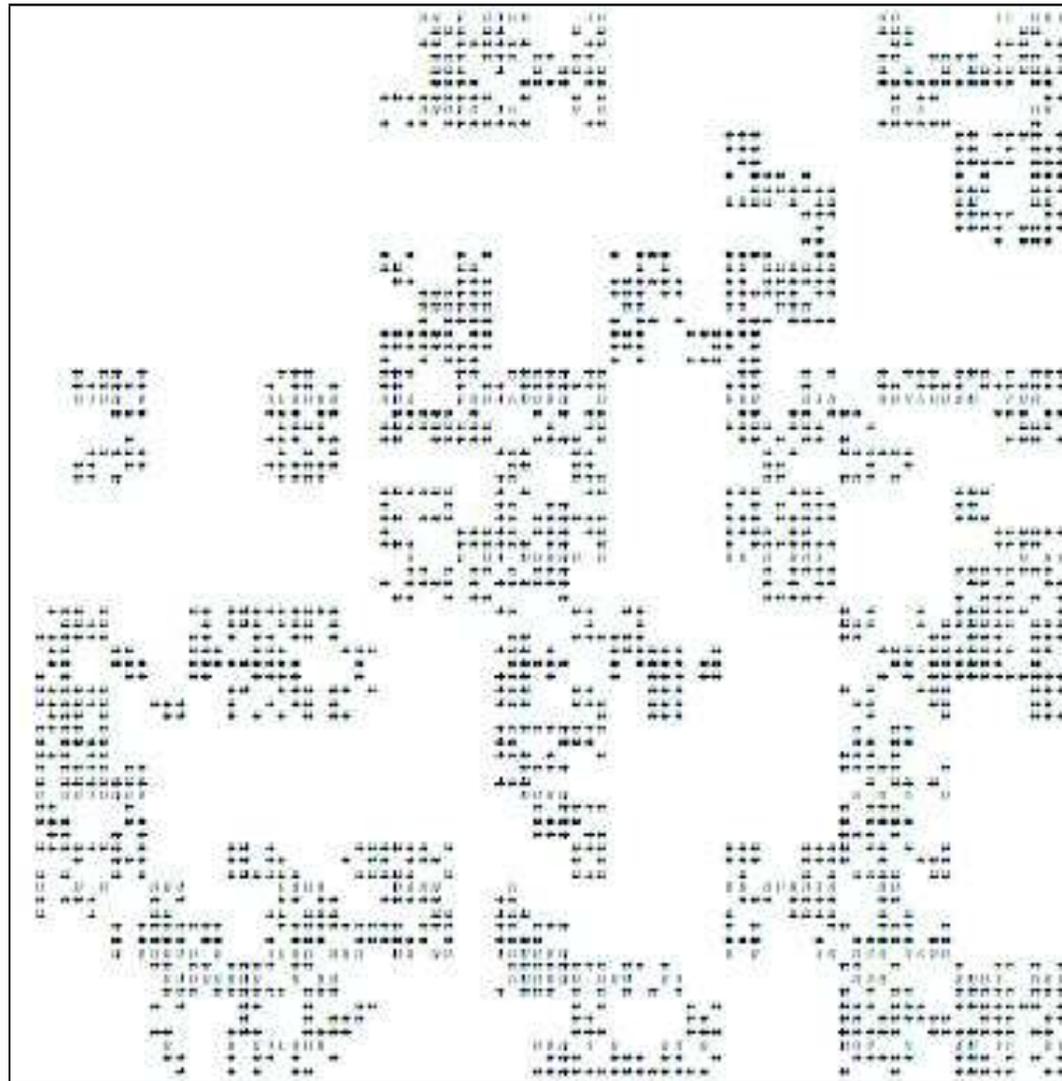
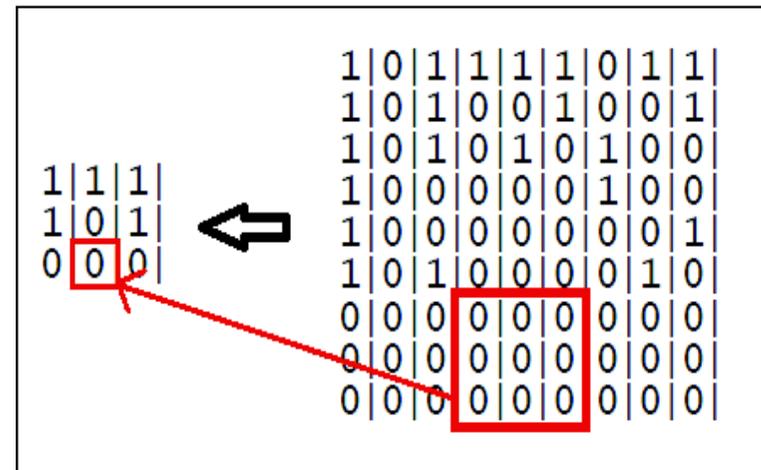
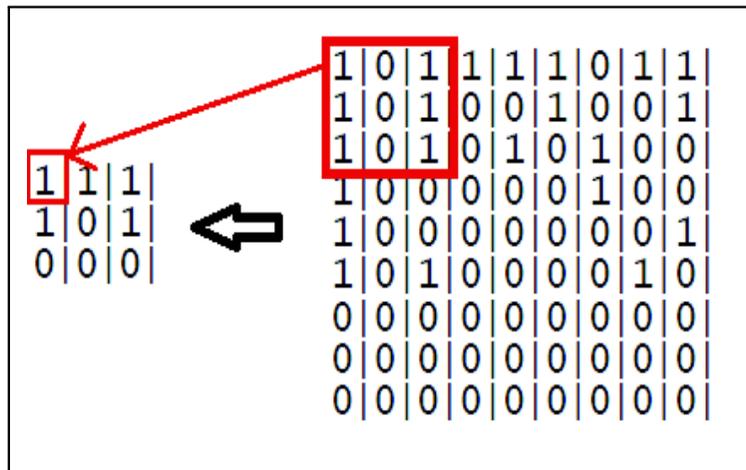


Abbildung: Beispiel mit $p = 0.3, q = 0.3$

Parameterschätzung

Wir zeigen hier wie das Bild entsteht

- Rückschlüsse aus Simulationsschritten ziehen:



- Befülle die Zelle der kleineren Matrix (m_{r-1}) mit 1 falls zumindest einmal 1 in der Untermatrix der Größe 3×3 der größeren Matrix (m_r) vorkommt. Falls nicht setze diese auf 0.
- Schätzungsgleichungen für p und q sind daher:

$$\hat{q} = \frac{\# 0 \text{ of } m_{r=4} - \# 0 \text{ of } m_{r=3} \cdot 9}{\# 1 \text{ of } m_{r=3} \cdot 9} \quad \hat{p} = \frac{\# 0 \text{ of } m_{r=3} - \# 0 \text{ of } m_{r=2} \cdot 9}{\# 1 \text{ of } m_{r=2} \cdot 9}$$

wobei # Anzahlen/Häufigkeiten, m die Matrix und r den Schritt anzeigt.

Anwendungsbeispiel: 40 deutsche Patientinnen

- Jeweils ca. 10 Bilder für jede der 40 Patientinnen vorhanden (20 Brustkrebs und 20 Mastopathie).
- Schätze die Parameter \hat{p} und \hat{q} für jedes Bild und berechne den Mittelwert für jedes Individuum.
- *Heterogenität zwischen den Mittelwerten je Patientin.*
- *Signifikante Unterschiede zwischen den zwei Gruppen erkennbar.*
- Simuliere Gewebe für beide Gruppen mit den ermittelten Parametern und vergleiche auf Unterschiede.
 - Mastopathie: $\bar{\hat{p}} = 0.0533$; $\bar{\hat{q}} = 0.0191$
 - Brustkrebs: $\bar{\hat{p}} = 0.1428$; $\bar{\hat{q}} = 0.0554$

Anwendungsbeispiel: 40 deutsche Patientinnen

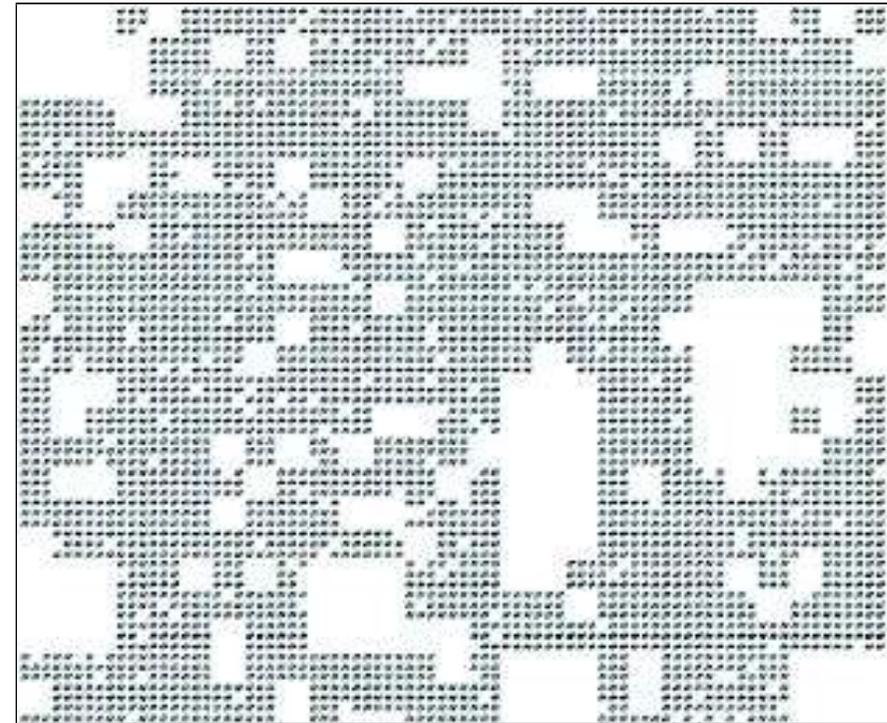
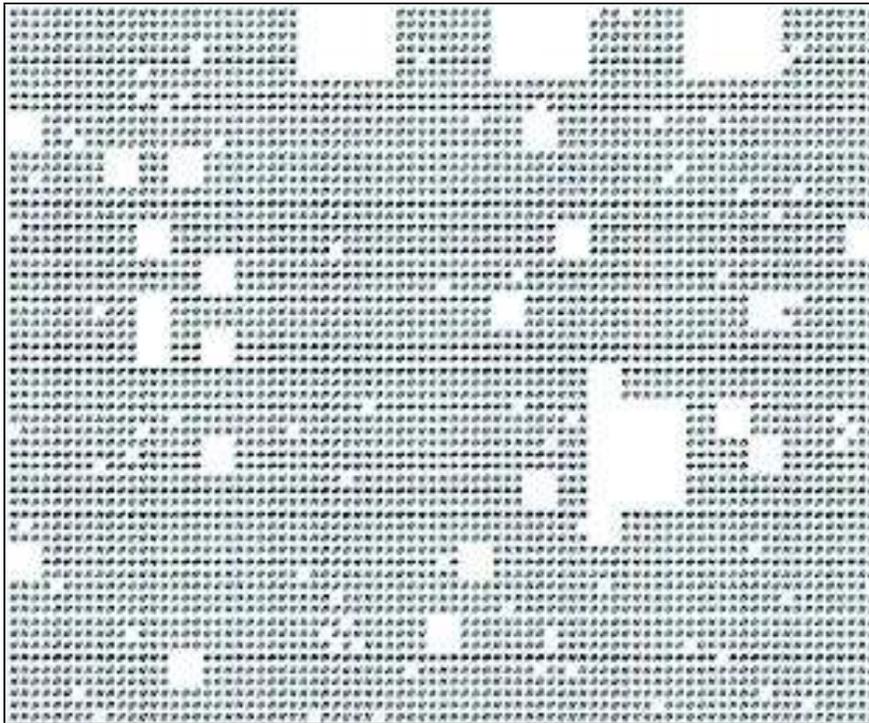


Abbildung: Simulation von Mastopathiegewebe (links) und Krebsgewebe (rechts) mit den geschätzten Mittelwerten von p und q .

Zusammenfassung

- Wir haben ein Modell entwickelt um zwischen Mastopathie und Krebs zu unterscheiden.
- Dieses Modell ist das bis jetzt beste Modell in dieser Forschungsrichtung.
- *Das Modell besteht sowohl aus dem theoretischen Teil als auch aus Computer Implementationen.*
- Diese Resultate wurden in *Statistics in Medicine* veröffentlicht.
- Es gibt großes Interesse an diesen Resultaten von medizinischen Experten (z.B.: Ärzte und Histopathologe auch in Hinsicht auf “personalized medicine”).

Wir bedanken uns bei

Ihnen für Ihre Aufmerksamkeit!

- Milan's Frau Silvia für ihre Unterstützung, Zeit und Diskussionen.
- Philipp's Familie für deren Anmerkungen und die Unterstützung.
- Tomas Mrkvicka, Orietta Nicolis, Maria Minarova, Torsten Mattfeldt und Fabian Wartner für die Zusammenarbeit.
- den KollegInnen am IFAS.

