

# Statistikertreffen der ZHAW 2016

## Multivariate Wahrscheinlichkeitsverteilung und die Wahrnehmung des 'Zufalls'

*Dr. Yves-L. Grize*

Institut für Datenanalyse und Prozessdesign

Zürcher Hochschule für Angewandte Wissenschaften

[yves-laurent.grize@zhaw.ch](mailto:yves-laurent.grize@zhaw.ch)

Winterthur, 17. Mai 2016

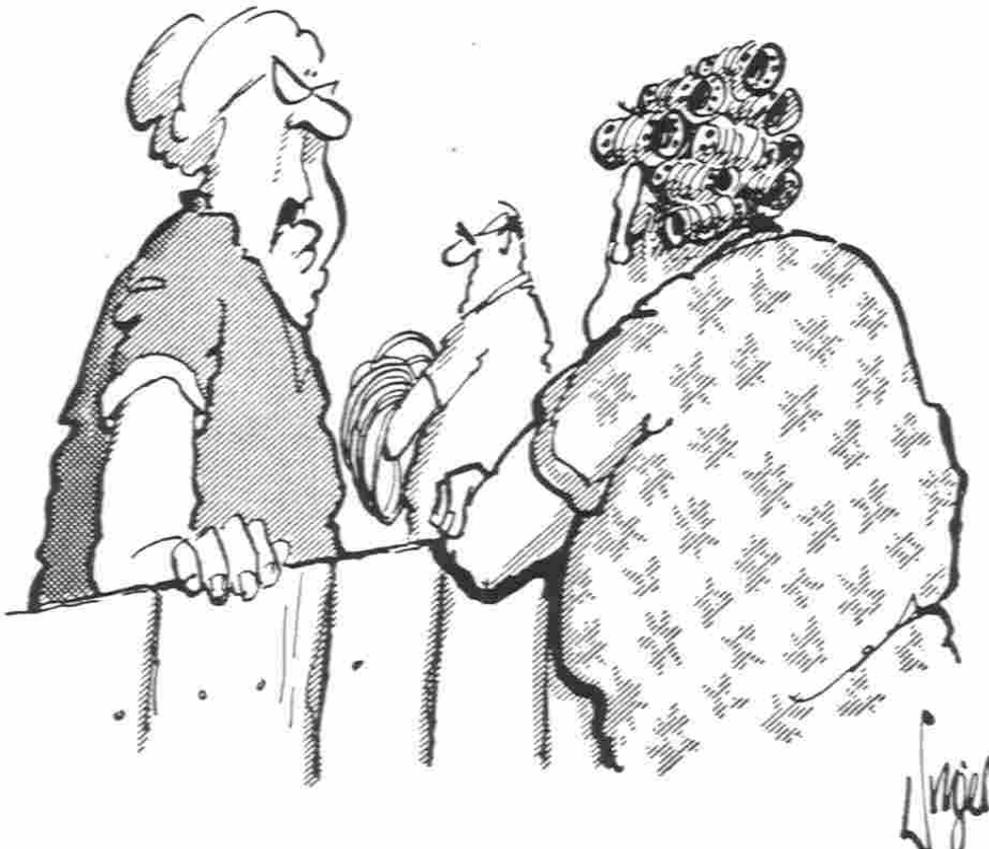
# Statistikertreffen der ZHAW 2016

Der Umgang mit Zufall und Unsicherheit wird durch den Begriff **Wahrscheinlichkeit** dominiert.

Das macht es nicht einfach...

- 1) Es gibt verschiedene Interpretationen davon
- 2) Das Handeln von "Wahrscheinlichkeiten" ist oft unintuitiv  
...

# Statistikertreffen der ZHAW 2016



**“The forecast said there’s a 50% chance of rain  
so genius is watering half the lawn.”**

# Statistikertreffen der ZHAW 2016

- 3) Und dazu noch ist multivariate Statistik schwieriger ...  
Es geht nämlich sowohl in der Verteilung der einzelnen Zufallsvariablen wie auch in ihrer gemeinsamen Verteilung.  
Die gemeinsame Verteilung ist was den Zusammenhang zwischen zwei Zufallsvariablen beschreibt.

Wie können wir alle diese Begriffe vermitteln?

# Statistikertreffen der ZHAW 2016

## Experimentieren ist immer gut!



Das Werfen einer Münze ist ein einfaches Zufallsexperiment, das in der Wahrscheinlichkeitstheorie und in der Statistik immer wieder benutzt wird. Die zufälligen Ergebnisse werden als Folge von Nullen und Einsen dargestellt, wobei z.B. 1 'Kopf' und 0 'Zahl' bedeutet.

Wir möchten:

1. einerseits die Wahrnehmung des Zufalls diskutieren,
2. und andererseits eine multivariate Verteilung visualisieren.

# Statistikertreffen der ZHAW 2016

## Ein interessantes statistisches Experiment (nach A. Gelman)

Zwei zufällige 0-1 Folgen von je 100 Ziffern werden durch die Kursteilnehmer erzeugt:

- eine durch Werfen einer Münze,
- und die andere durch reine Erfindung.

Der Referent ist bei der Erzeugung der Folgen nicht dabei!  
Nachträglich soll aber der Referent versuchen, die künstliche Folge zu identifizieren.

# Statistikertreffen der ZHAW 2016

## Beispiel zweier solchen Folgen:

00111000110010000100

01000101001100010100

00100010001000000001

11101001100011110100

00110010101100001111

01110100011000110111

11001100010101100100

10001001011011011100

10001000000011111001

01100100010010000100

Wie kann man die zufällige Folge leicht identifizieren?

# Statistikertreffen der ZHAW 2016

## Wir betrachten folgende bivariate Verteilung:

Wir werfen eine Münze 100 Mal. Danach beobachten wir in der resultierenden Folge von 0's und 1's die folgenden Grössen:

$X$  = Anzahl von Wechsel von 0 auf 1 oder 1 auf 0

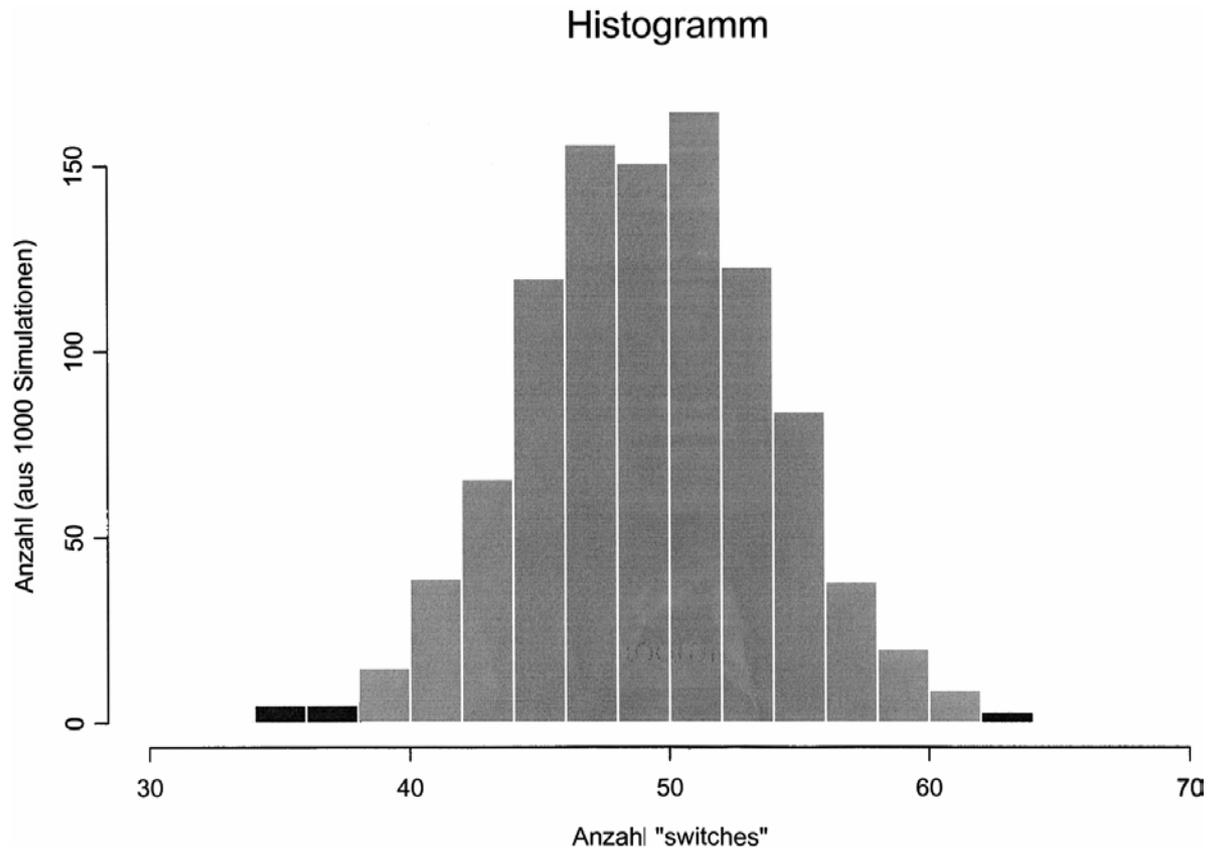
$Y$  = Länge der längsten konstanten Strecke von 0's oder von 1's

$X$  und  $Y$  sind zwei Zufallsvariablen.

Sind sie unabhängig? Wie hängen zusammen ab?

# Statistikertreffen der ZHAW 2016

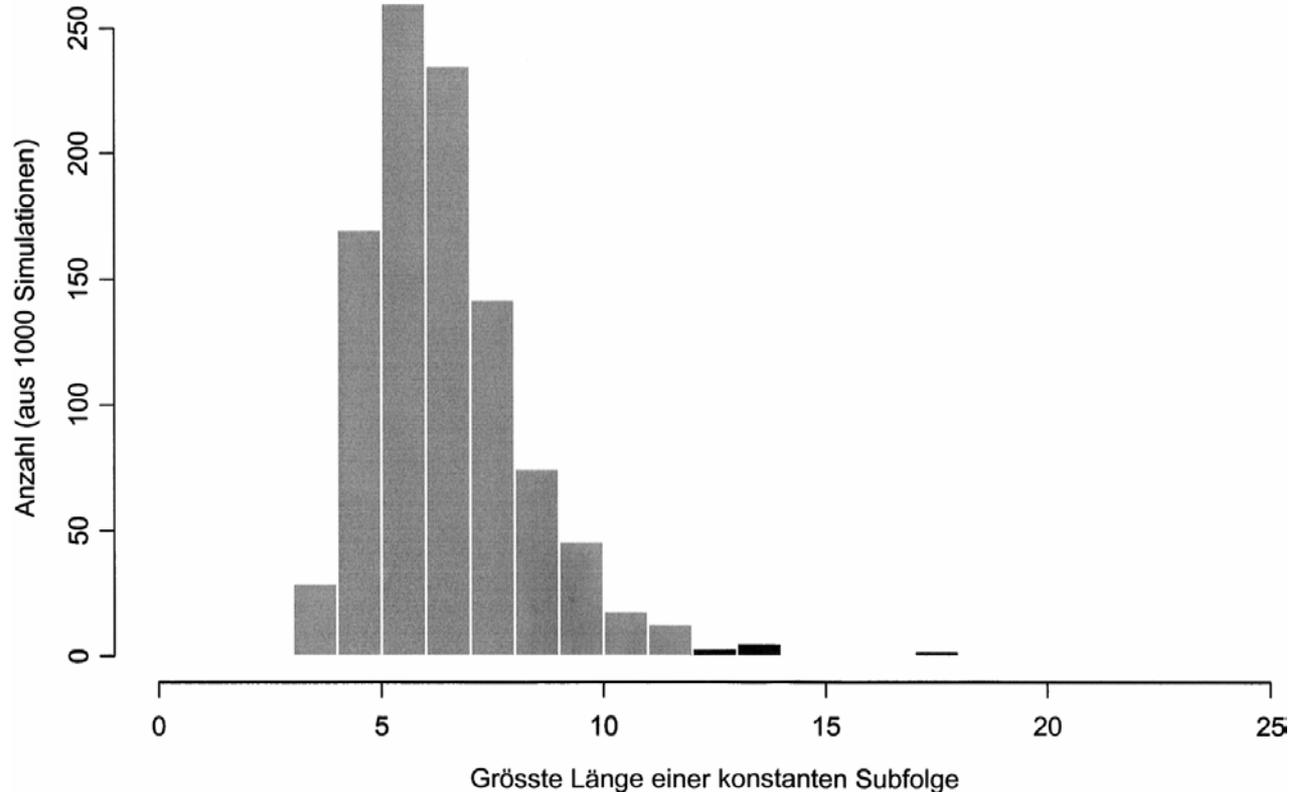
## Randverteilung von X



# Statistikertreffen der ZHAW 2016

## Randverteilung von Y

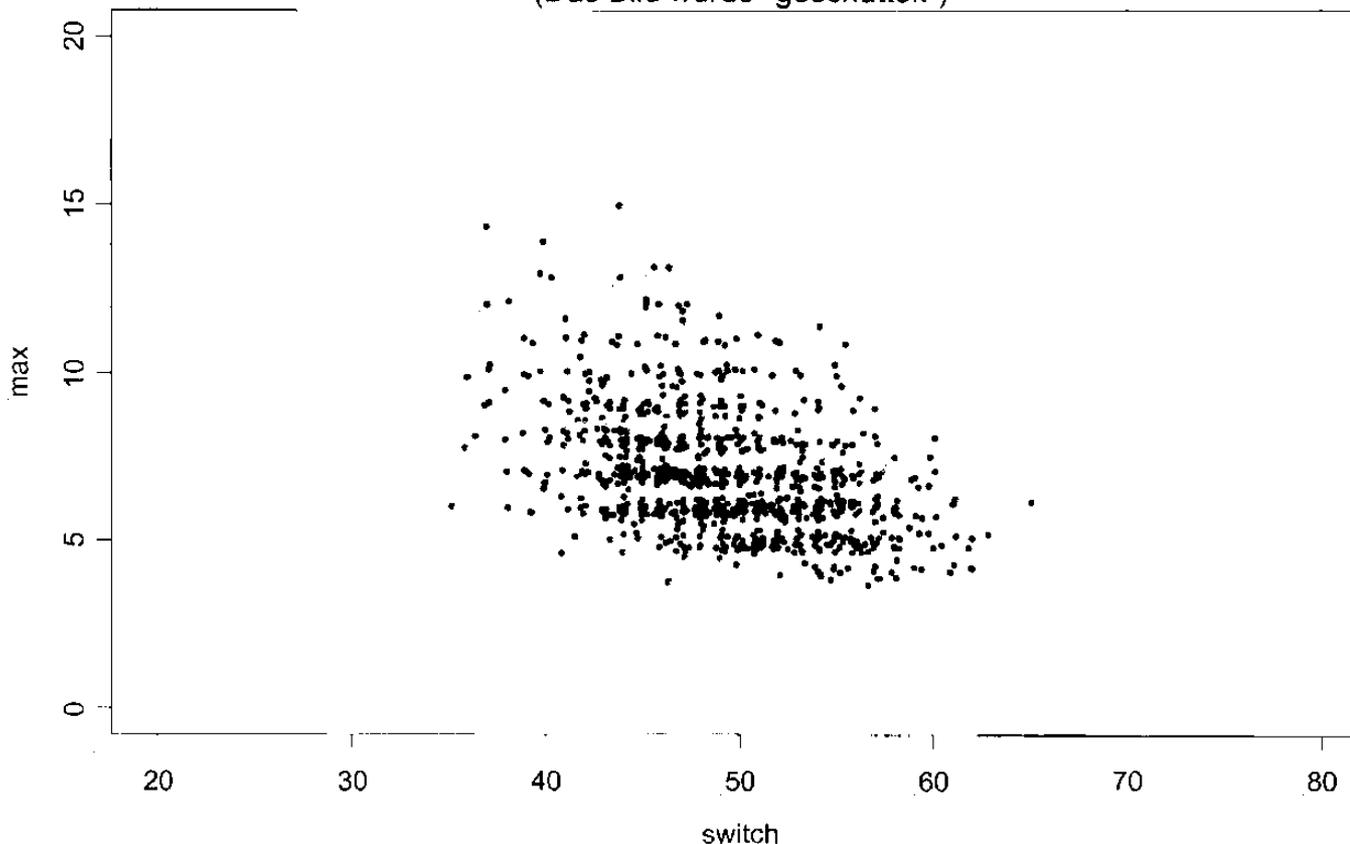
Histogramm



# Statistikertreffen der ZHAW 2016

## Gemeinsame Verteilung von X und Y

Simulation von 1000 0-1-Folgen der Länge 100  
(Das Bild wurde "geschüttelt")

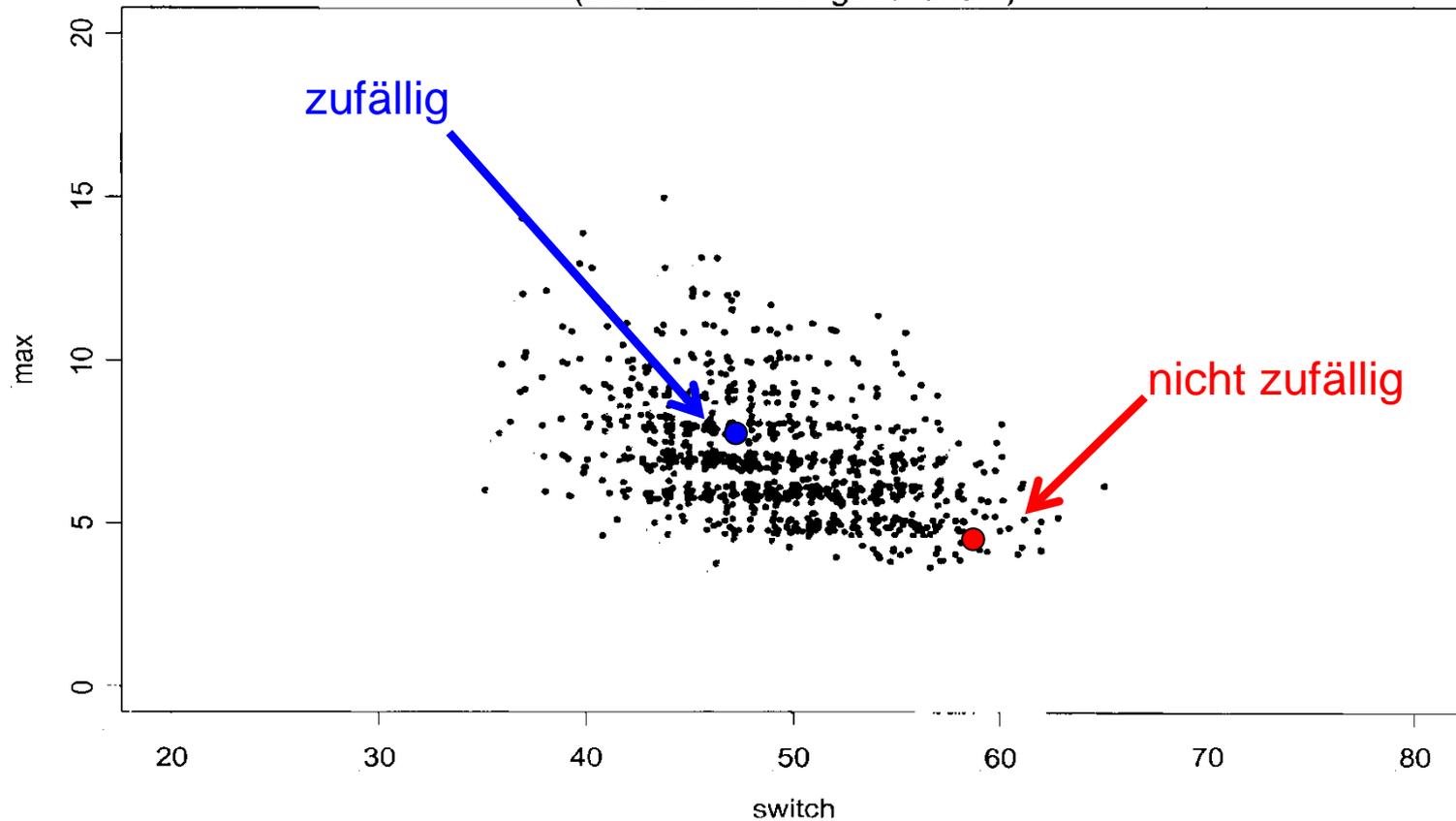


Bemerkung:  
X und Y wurden in  
der Grafik  
'geschüttelt'  
(R-Funktion  
jitter)

Die Punkte  
streuen nicht  
symmetrisch, es  
gibt eine  
Abhängigkeit  
zwischen X und Y.

# Statistikertreffen der ZHAW 2016

Simulation von 1000 0-1-Folgen der Länge 100  
(Das Bild wurde "geschüttelt")



# Statistikertreffen der ZHAW 2016

## Ablauf des Experiments

- Die zufällige Sequenz wird vom Referent identifiziert!
- Für beide Folgen werde die Werte von  $X$  und  $Y$  berechnet.
- Die gemeinsame Verteilung wird gezeigt bzw. simuliert.
- Die Realisierungen von  $X$  und  $Y$  werden in die Grafik eingetragen.
  
- Über das Vorhanden von «Mustern» in zufälligen Folgen kann auch diskutiert werden!
  
- Insbesondere kann aufmerksam gemacht werden, warum kleine Stichprobe «nicht-repräsentativ» sein können.

# Statistikertreffen der ZHAW 2016

## Zusammenfassung

- Das Beispiel zeigt, wie schlecht eine zufällige Reihenfolge von 0 und 1 wahrgenommen wird
- Das Beispiel erklärt, was eine bivariate Verteilung ist.
- Das Beispiel zeigt auch, wie ein statistische Entscheidung zustande kommt, auch mit der Risiko einer Fehlentscheidung.