

---

# Verteilungen und Zufallsvariable

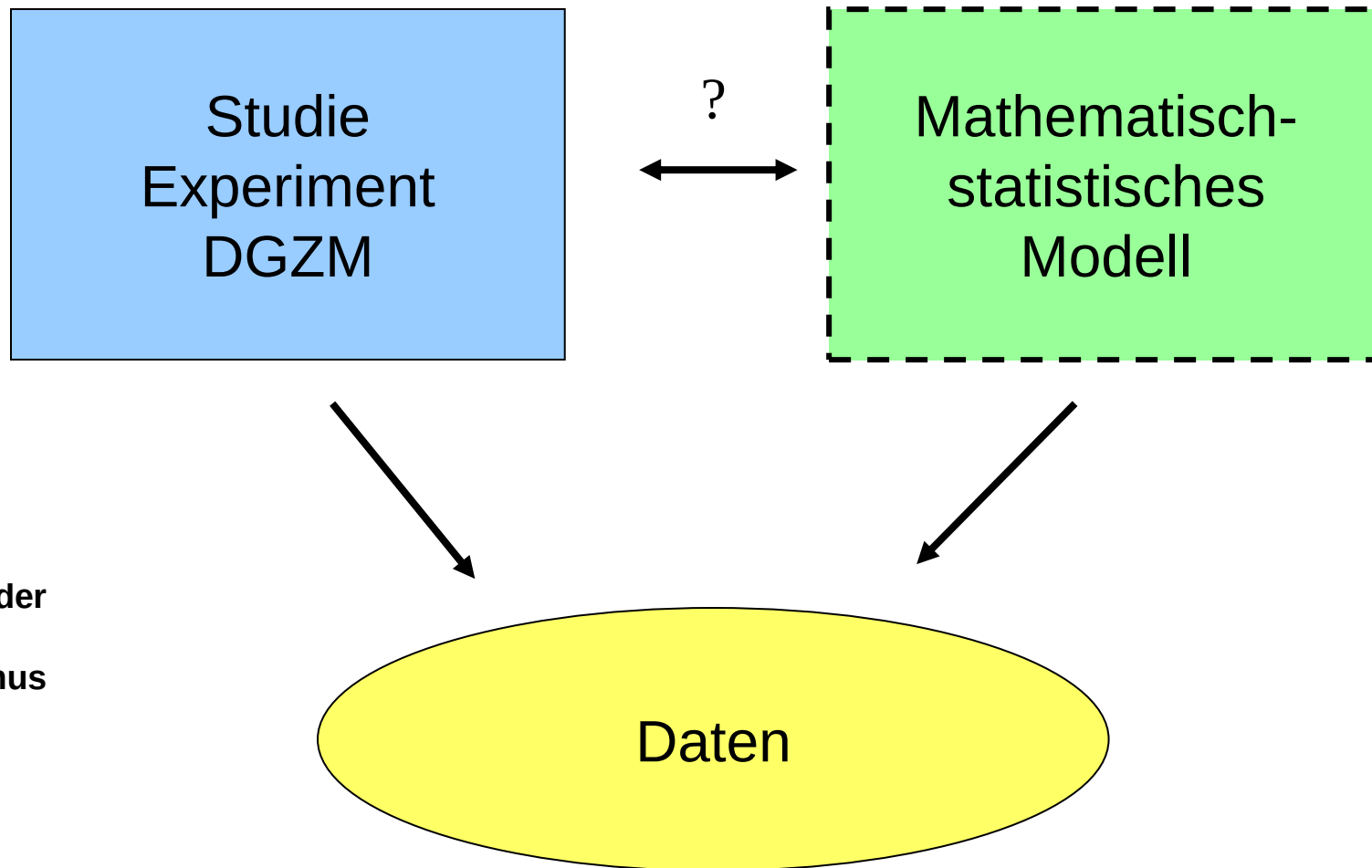
---

Dr. Dirk Hasenclever  
University of Leipzig

imise.

# Statistische Modellierung

---



DGZM=  
Daten  
generierender  
Zufalls-  
mechanismus

# Definition Wahrscheinlichkeit

---

**Definition 1.4:** Sei  $(\Omega, \mathfrak{S})$  ein meßbarer Raum und sei  $P$  eine Abbildung von  $\mathfrak{S}$  in  $\mathbb{R}$ .  $P$  heißt ein Wahrscheinlichkeitsmaß oder kurz eine Wahrscheinlichkeit, wenn gilt:

$$(I) \quad A \in \mathfrak{S} \Rightarrow P(A) \geq 0.$$

$$(II) \quad (A_n \in \mathfrak{S} \text{ für alle } n \in \mathbb{N} \text{ und } A_i \cap A_k = \emptyset \text{ für } i \neq k) \Rightarrow$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

$$(III) \quad P(\Omega) = 1.$$

$\Omega$  ~ Menge möglicher Elementarereignisse/Beobachtungen

$\mathfrak{S}$  ~ Mengenalgebra von Ereignissen (Teilmengen von  $\Omega$ )

# Statistisches Modell einer Messgröße

---

- Eine Messgröße hat einen **Träger(=“support”) = Menge beobachtbarer Messwerte.**
- Oft: Träger ist Teilmenge von  $\mathbb{R}$
- Eine Messgröße wird modelliert als eine **reelle Zufallsvariable  $X$ .**

# Diskrete Zufallsvariable

---

Sei die Messgröße **diskret**, d.h. der Träger endlich oder abzählbar.

- Eine **diskrete Zufallsvariable  $X$**  ist definiert durch die **Wahrscheinlichkeitsfunktion** auf den Elementarereignissen
  - **$pr(X=x)$  für alle  $x$  im Träger**
- Dabei ist  $pr$  eine Wahrscheinlichkeit im Sinne der Definition auf der von den Elementarereignissen im Träger erzeugten  $\sigma$ -Algebra.

Insbesondere:

$$pr(X = x) \geq 0$$
$$\sum_{x \in \text{Träger}} pr(X = x) = 1$$

$$pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} pr(A_i), \text{ falls } A_i \text{ paarweise disjunkt}$$

# Bernoulli-Verteilung

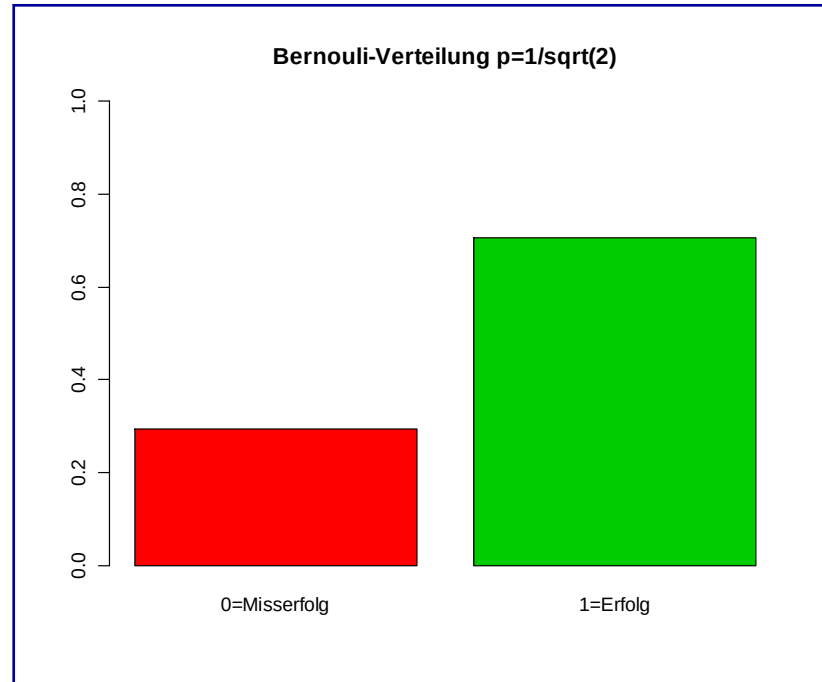
$$\text{Träger}(X) = \{0,1\}$$

$$\text{pr}(X = 1) = p$$

$$\text{pr}(X = 0) = 1 - p$$

$$E(X) = p \cdot 1 + (1 - p) \cdot 0 = p$$

$$\text{var}(X) = E((X - p)^2) = p \cdot (1 - p)^2 + (1 - p) \cdot (-p)^2 = p \cdot (1 - p)$$



# Verteilungen auf 1:N

---

- Eine Verteilung auf 1:N ist gegeben durch eine Wahrscheinlichkeitsfunktion

$$\text{pr}(X = i) = p_i, \quad i \text{ in } 1:N$$

- Die Wahrscheinlichkeitsfunktion ist bestimmt durch die Angabe von N reellen Zahlen

$$p_i \geq 0 \text{ mit der Nebenbedingung } \text{Summe } (p_i) = 1$$

- NB:  $\{ (p_1, \dots, p_N) \mid p_i \geq 0 \text{ und } \text{Summe } (p_i) = 1 \}$  nennt man auch den (N-1)-dimensionalen Simplex
- NB2: Wird in R als `sample(1:N, size=K, replace = T, prob=c(p1,...,pN ))` realisiert.

# Binomial-Verteilung

$X \sim \text{Binom}(N, p)$  falls  $X = \sum_{i=1}^N Y_i$

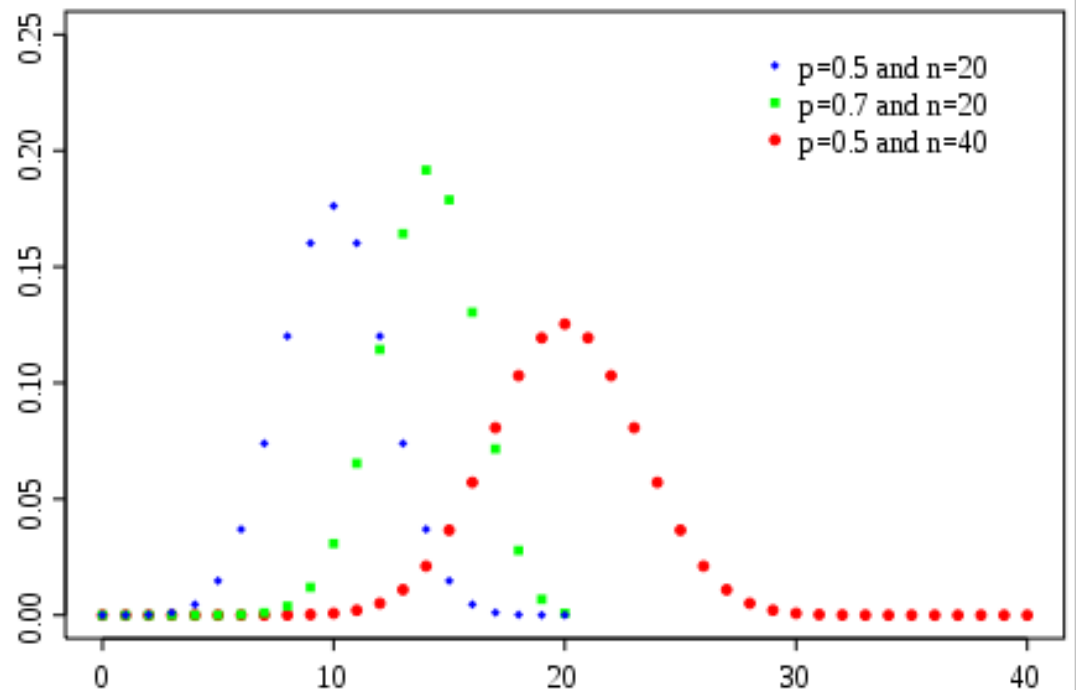
wobei  $Y_i \sim \text{Bernoulli}(p)$  und unabhängig.

Träger =  $\{0, 1, 2, \dots, N\}$

$$\text{pr}(X=k) = \binom{N}{k} p^k (1-p)^{N-k}$$

$$E(X) = N \cdot p$$

$$\text{var}(X) = N \cdot p \cdot (1-p)$$



[http://en.wikipedia.org/wiki/Binomial\\_distribution](http://en.wikipedia.org/wiki/Binomial_distribution)

# Konstruktion neuer ZV aus bereits definierten ZV

---

- Gegeben  $X_1, \dots, X_k$  Zufallsvariable.
- Sei  $g$  eine Funktion, die auf dem Mengenprodukt  $\text{Träger}(X_1) \times \dots \times \text{Träger}(X_k)$  definiert ist.
- Dann definiert  $Z := g(X_1, \dots, X_k)$  eine neue ZV mit  $\text{Träger}(Z) = g(\text{Träger}(X_1) \times \dots \times \text{Träger}(X_k))$
- Bsp: Binomialverteilung, Würfel mit zwei Würfeln,...

# Mischung zweier Bernoulli Verteilungen

---

- $X \sim \text{Bernoulli}(p_X)$  ;  $Y \sim \text{Bernoulli}(p_Y)$ ;  $Q \sim \text{Bernoulli}(q)$   
(unabhängig)

X falls  $Q = 0$

- Neue ZV:  $X_Q Y =$

Y falls  $Q = 1$

$$\begin{aligned}\text{pr}(X_Q Y = 1) &= \text{pr}(X=1) \text{pr}(Q=0) + \text{pr}(Y=1) \text{pr}(Q=1) \\ &= p_X \cdot (1-q) + p_Y \cdot q\end{aligned}$$

Also:  $X_Q Y \sim \text{Bernoulli}(p_X \cdot (1-q) + p_Y \cdot q)$

Bernoulli Mischungen  
kann man nicht  
als Mischungen erkennen –  
denn sie sind wieder  
Bernoulli ZV

# Poisson Verteilung

$X \sim \text{Poisson}(\lambda)$  mit  $\lambda > 0$  reell

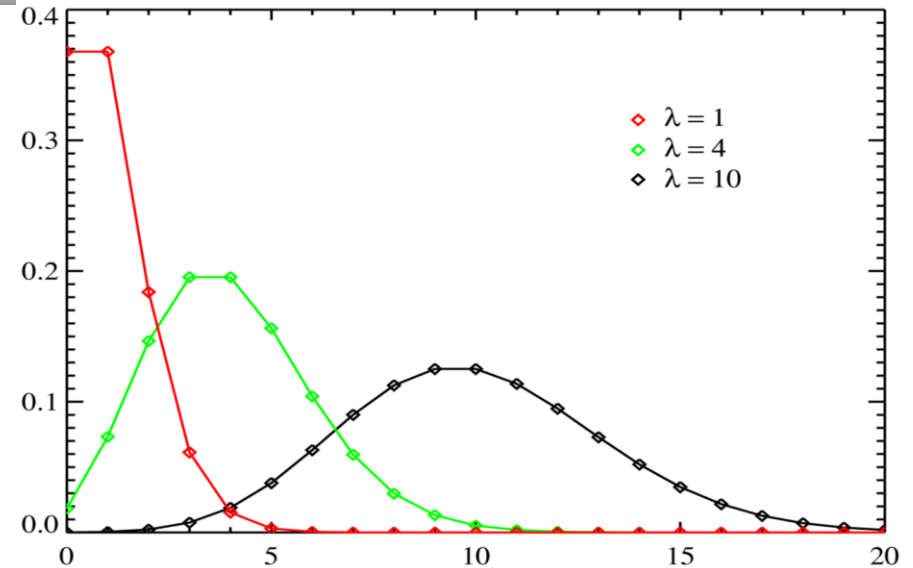
$\text{Träger}(X) = \mathbb{N}_0 = \{0, 1, 2, \dots\}$

$$\text{pr}(X = k) = \frac{\lambda^k \exp(-\lambda)}{k!}$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k \exp(-\lambda)}{k!} = \exp(-\lambda) \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(-\lambda) \exp(\lambda) = \exp(0) = 1$$

$$E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k \exp(-\lambda)}{k!} = \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1} \exp(-\lambda)}{(k-1)!} = \lambda$$

$$\text{var}(X) = \lambda$$



# Kumulierte Verteilungsfunktion einer diskreten ZV

---

Sei  $X$  eine diskrete ZV mit ordinalem Träger:

$F_X(x) = \text{pr}(X \leq x)$  heißt kumulative Verteilungsfunktion.

Zusammenhang mit der Wahrscheinlichkeitsfunktion:

$$F_X(x) = \sum_{k \leq x} \text{pr}(X = k)$$

Umgekehrt: sei  $x'$  der Vorgänger von  $x$

$$\text{pr}(X = x) = F_X(x) - F_X(x')$$

# Stetige Zufallsvariable

---

Träger (X) ist  $\mathbb{R}$  oder ein Intervall in  $\mathbb{R}$ .

Die Verteilung wird gegeben durch die Verteilungsfunktion

$$F_X(x) = \text{pr}(X \leq x)$$

Die zugehörige  $\sigma$ -Algebra wird erzeugt von den Intervallen  $[a, b]$ .

Durch  $F_X(x)$  ist die Wahrscheinlichkeitsfunktion bereits bestimmt:

$$\text{pr}(X \in [a, b]) = F_X(b) - F_X(a)$$

Ist  $F_X$  differenzierbar so heißt  $f_X(x) := \frac{dF_X}{dx}$  die Dichtefunktion von X

und es gilt:

$$\text{pr}(X \in [a, b]) = \int_a^b f_X(x) dx$$

# Normalverteilung

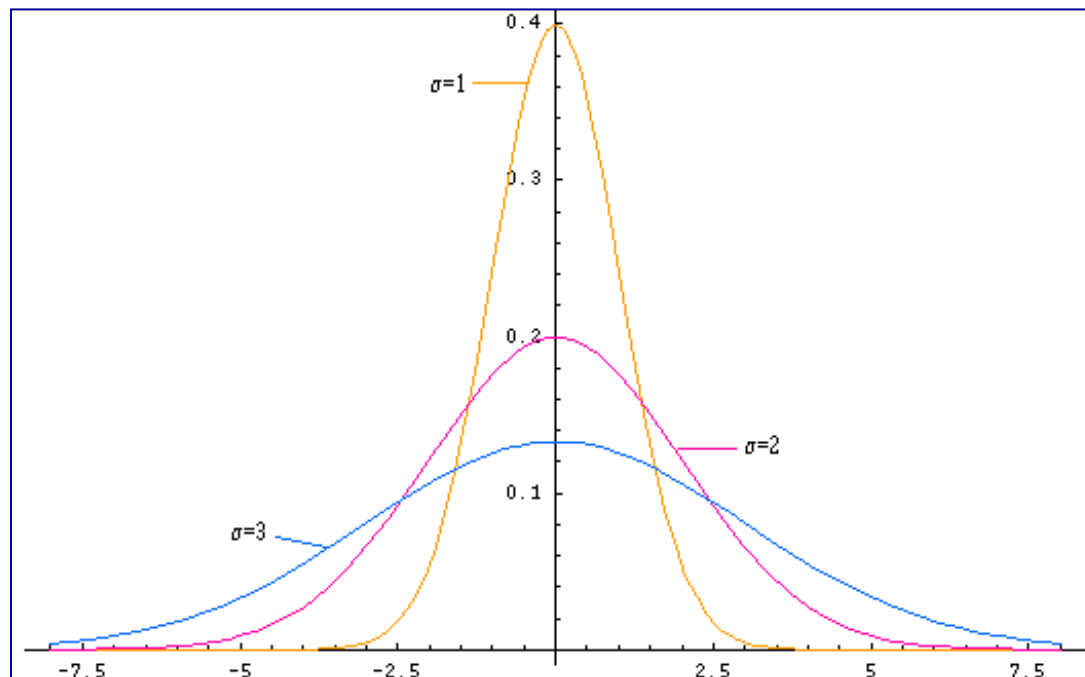
$$\text{Träger}(X) = \mathbf{R}$$

$$X \sim N(\mu, \sigma^2)$$

$$\text{Dichte } f_X(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$E(X) = \mu$$

$$\text{var}(X) = \sigma^2$$



# Exponential Verteilung

Dichtefunktion:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

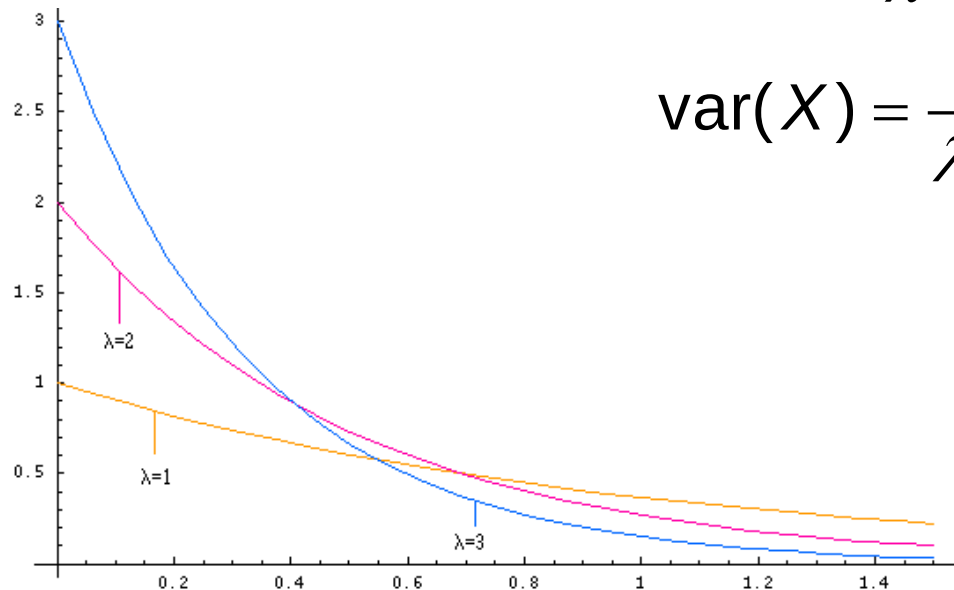
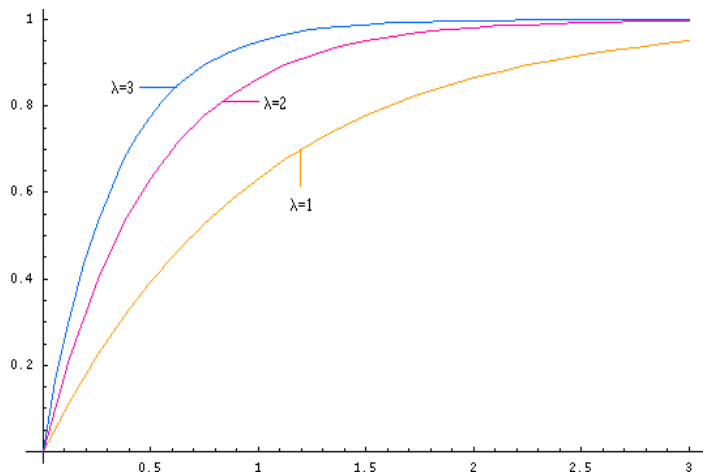
Verteilungsfunktion:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

$$\text{var}(X) = \frac{1}{\lambda^2}$$

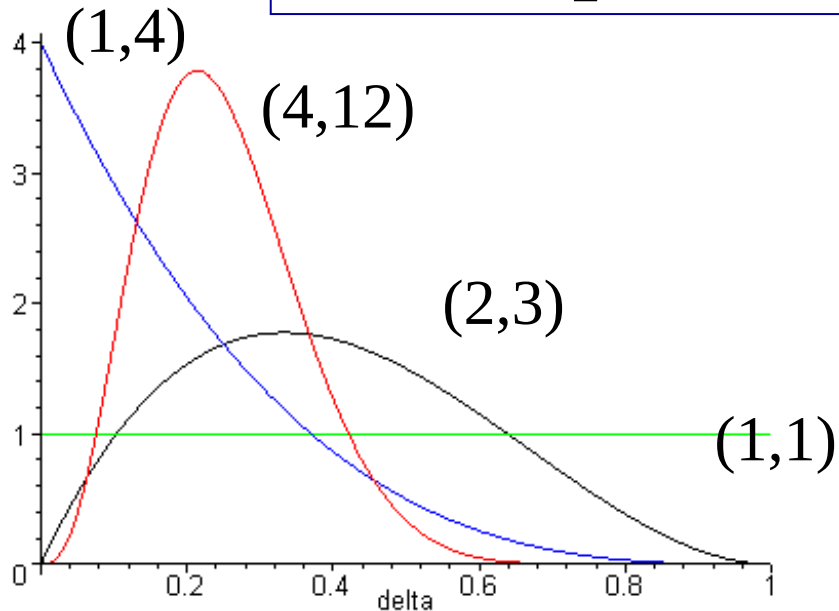
Träger= $[0, \infty)$



# Beta-Verteilung

$$\text{Beta}(a,b): f(p | a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}$$

$$\left[ \Gamma(a) = \int_0^{\infty} \exp(-x) x^{a-1} dx; \quad \Gamma(n+1) = n! \quad \forall n \right]$$



$$E(X) = \frac{a}{a+b}; \quad \text{var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Träger=[0,1]

# Kumulierte Verteilungsfunktion und Dichte

---

- Kumulierte Verteilungsfunktion:  $F_X(x) = \text{pr}(X \leq x)$ 
  - Monoton steigend:  $x > y$  folgt  $\text{pr}(X \leq x) \geq \text{pr}(X \leq y)$   
–(Aber nicht unbedingt streng!)
  - Variiert zwischen 0 und 1.
  
- Dichte ist Ableitung  $f_X(x) = F_X(x)'$ 
  - $f_X(x) \geq 0$  (nicht negativ)
  - Integral über Träger( $f_X(x)$   $dx$ ) = 1 (Normiert)

# Transformation von Dichten

---

Sei  $X$  eine reelle Zufallsvariable,  
Sei  $h$  eine streng monoton steigende,  
differenzierbare Funktion mit Umkehrfunktion  $h^{-1}$

Definiere  $Y := h(X)$ , also  $X = h^{-1}(Y)$

$F_Y(y) = \text{pr}(Y \leq y) = \text{pr}(X \leq h^{-1}(y)) = F_X(h^{-1}(y))$ , da  $h$  streng monoton

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(h^{-1}(y)) \quad \text{also}$$

$$f_Y(y) = f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y) \quad \text{nach Kettenregel}$$

# Transformation von Dichten (Bsp. Lognormalverteilung)

---

Sei  $X \sim N(\mu, \sigma^2)$

$Y := \exp(X)$  heißt dann log-normalverteilt, da  $\log(Y)$  normal.  
(Viele Blutwerte sind in der Praxis lognormalverteilt.)

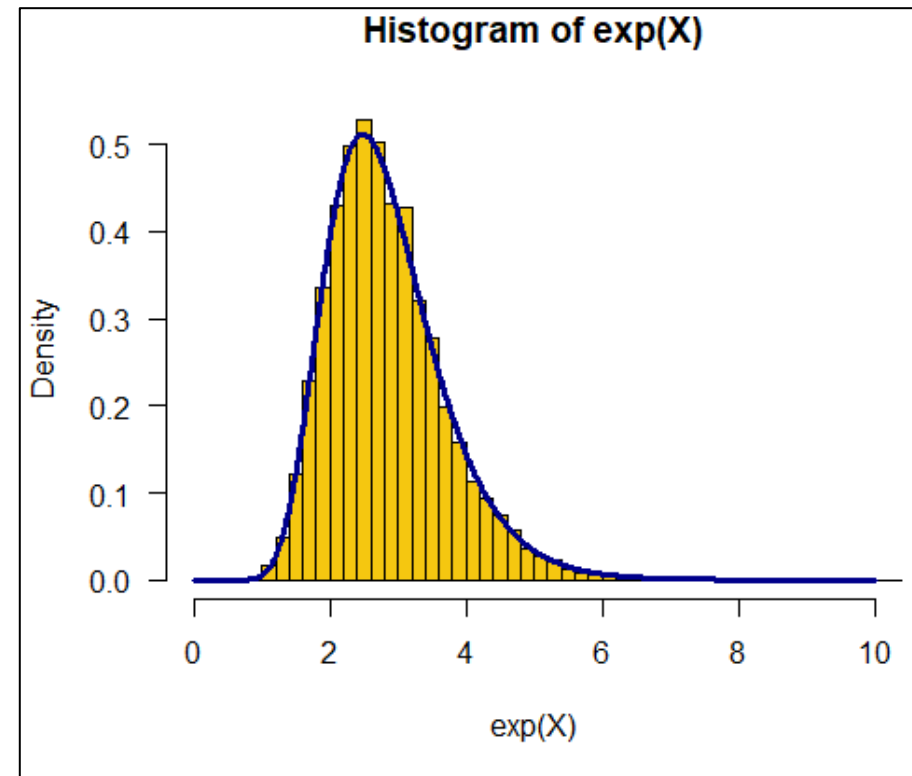
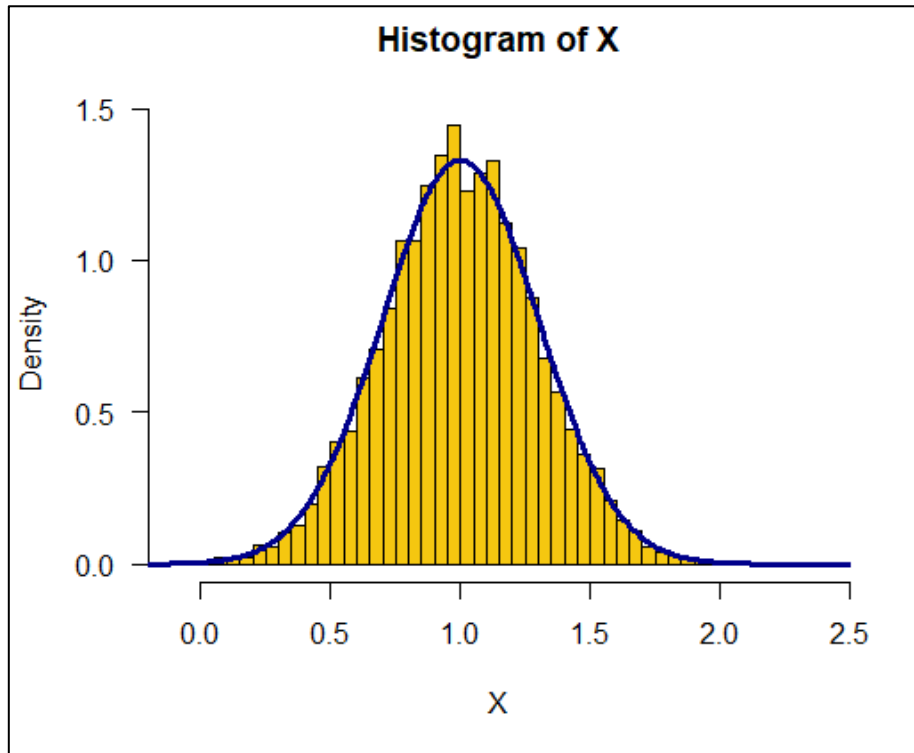
Umkehrfunktion:  $X = \log(Y)$

$$f_Y(y) = f_X(\log(y)) \cdot \frac{d}{dy} \log(y)$$

!


$$f_Y(y) dy = \frac{1}{y} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right) dy$$

# Transformation von Dichten (Bsp. Lognormalverteilung)



```
X<-rnorm(10000, 1, .3)
hist(X, nclass=40, prob=T, las=1, col=7)
yy<-(-1000:1000)/100
lines(yy, dnorm(yy, 1, .3), lwd=3, col="darkblue")
hist(exp(X), nclass=40, prob=T, las=1, col=7, xlim=c(0,10))
hh<-function(x){ dnorm(log(x), 1, .3)/x}yy<-(1:1000)/100
lines(yy, hh(yy), lwd=3, col="darkblue" )
```

## Transformation von Dichten (Bsp. Quadrat der Gleichverteilung)

---

Sei  $X \sim \text{uniform}([a,b])$  ;  $b > a \geq 0$

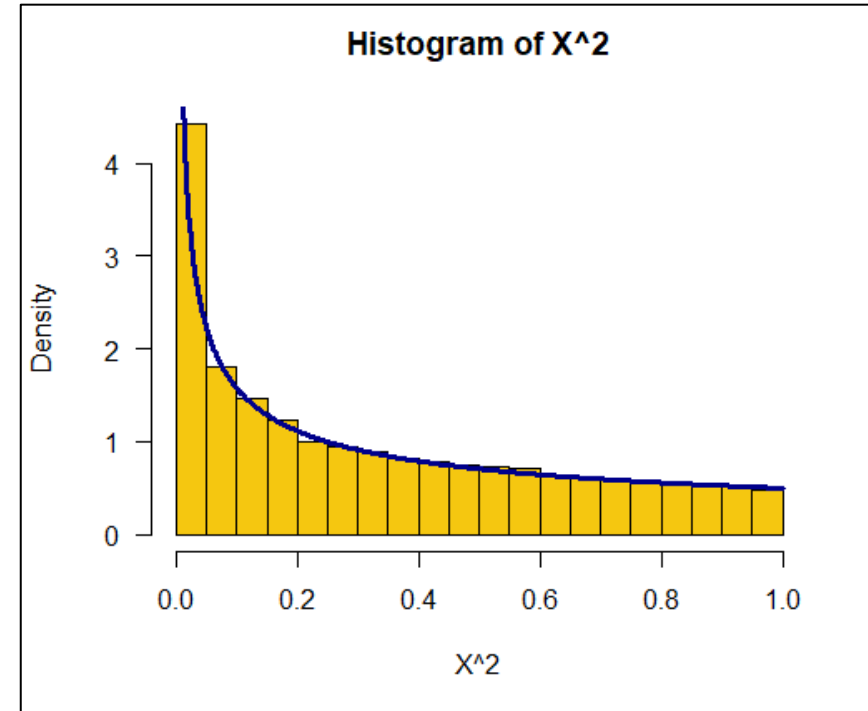
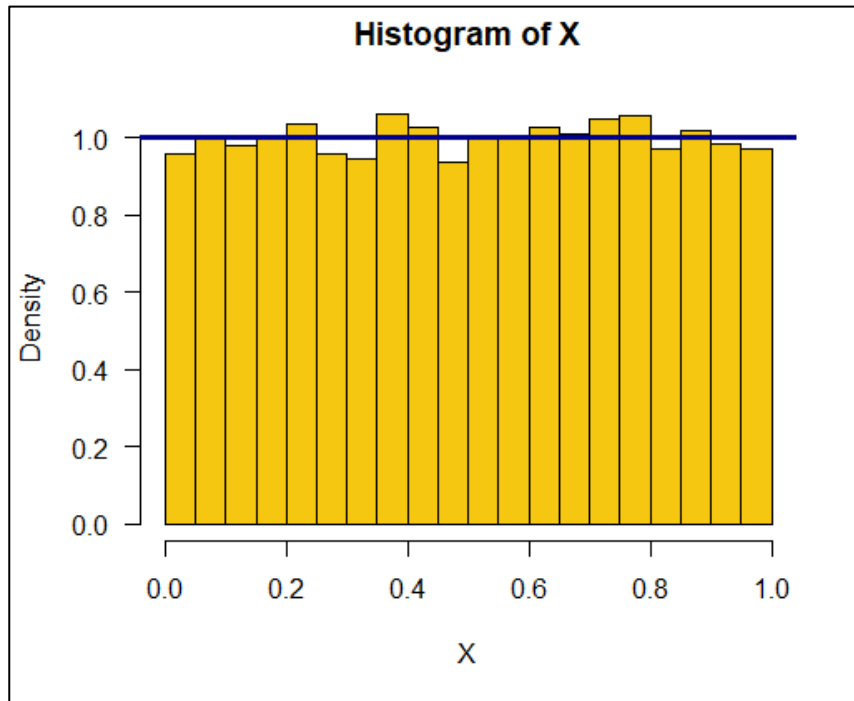
$$f_X(x) = \frac{1}{(b-a)}$$

$Y := X^2$  hat Träger  $[a^2, b^2]$

Umkehrfunktion:  $X = Y^{1/2}$

$$\begin{aligned} f_Y(y) &= f_X(y^{1/2}) \cdot \frac{d}{dy} y^{1/2} \\ &= \frac{1}{(b-a)} \cdot \frac{1}{2y^{1/2}} \end{aligned}$$

# Transformation von Dichten (Bsp. Quadrat der Gleichverteilung)



```
a=0;b=1
X <- runif(10000,a,b)
hist(X, nclass=20, prob=T, las=1 , col=7)
abline(h=1, lwd=3,col="darkblue")
hist(X^2, nclass=20, prob=T, las=1 , col=7)
hh<-function(x){ 1/2/sqrt(x) }
yy<(1:1000)/1000
lines(yy,hh(yy),lwd=3,col="darkblue" )
```

# Erwartungswert

---

Sei  $X$  eine diskrete Zufallsvariable:

Der Erwartungswert von  $X$  ist definiert als:

$$E(X) = \sum_{x \in \text{Träger}(X)} x \cdot \text{pr}(X=x)$$

Sei  $X$  eine stetige Zufallsvariable:

Der Erwartungswert von  $X$  ist definiert als:

$$E(X) = \int_{\text{Träger}(X)} x \cdot f_X(x) dx$$

Der Erwartungswert einer Funktion  $g(X)$  ist definiert als:

$$E(g(X)) = \int_{\text{Träger}(X)} g(x) \cdot f_X(x) dx$$

# Eigenschaften des Erwartungswerts

---

Der Erwartungswert ist linear :

a)  $E(\lambda \cdot X) = \lambda \cdot E(X)$  für ein Skalar  $\lambda$

b)  $E(X + Y) = E(X) + E(Y)$  für beliebige Zufallsvariable

Beweis nächste Vorlesung

Sind  $X$  und  $Y$  *unabhängig*, so gilt außerdem :

c)  $E(X \cdot Y) = E(X) \cdot E(Y)$

# Varianz

---

Sei  $X$  eine Zufallsvariable.

Die Varianz ist definiert als  
$$\text{var}(X) := E( (X - E(X))^2 )$$

Vorsicht: Es gibt Verteilungen,  
für welche die Varianz nicht existiert.  
(z.B. Cauchy-Verteilung)

# Eigenschaften der Varianz I

---

Verschiebungssatz:

$$\text{var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

Beweis mit Eigenschaften des Erwartungswerts:

$$\begin{aligned} E((X - E(X))^2) &= \\ E(X^2 - 2XE(X) + E(X)^2) &= \\ E(X^2) - E(2XE(X)) + E(X)^2 &= \\ E(X^2) - 2E(X)E(X) + E(X)^2 \end{aligned}$$

Ausmultiplizieren  
Additivität von E  
Linearität von E

# Eigenschaften der Varianz II

---

Lineare Transformation:

$$\text{var}(a \cdot X + b) = a^2 \cdot \text{var}(X)$$

Beweis mit Eigenschaften des Erwartungswerts:

$$\begin{aligned} \text{var}(aX + b) &= E((aX+b - E(aX+b))^2) = && \text{Definition} \\ &= E((aX+b - a(E(X) + b))^2) = && \text{Linearität von E} \\ &= E(a^2 \cdot (X - E(X))^2) = a^2 \cdot \text{var}(X) && \text{Vereinfachen und Auskammern} \end{aligned}$$

# Eigenschaften der Varianz III

---

Varianz von Summer und Differenz:

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \pm 2 \cdot \text{cov}(X, Y)$$

Insbesondere falls X,Y unabhängig:

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y)$$

Achtung: auch bei einer Differenz addieren sich die Varianzen.

cov(X,Y) korrigiert bei Abhängigkeit.

# Kovarianz

---

$$\text{cov}(X, Y) := E((X - E(X)) \cdot (Y - E(Y)))$$

Eigenschaften:

- a)  $\text{cov}(X, Y) = \text{cov}(Y, X)$  (Symmetrie)
- b)  $\text{cov}(a + X, Y) = \text{cov}(X, Y)$  (translationsinvariant)
- c)  $\text{cov}(\alpha X + \beta Z, Y) = \alpha \cdot \text{cov}(X, Y) + \beta \cdot \text{cov}(Z, Y)$  (linear)

Folgen direkt aus Eigenschaften des Erwartungswerts.

# MARKOFFsche Ungleichung

---

Satz: Sei  $X \geq 0$  eine reelle nicht negative Zufallsvariable mit Dichte  $f_X(x) dx$  für die  $E(X)$  existiert:

$$pr( X \geq \lambda \cdot E(X) ) \leq \frac{1}{\lambda} \quad (\forall \lambda > 0)$$

Beweis:

$$E(X) = \int_0^{\infty} x \cdot f_X(x) dx \geq \int_{\lambda \cdot E(X)}^{\infty} x \cdot f_X(x) dx \geq$$

$$\lambda \cdot E(X) \cdot \int_{\lambda \cdot E(X)}^{\infty} f_X(x) dx \geq$$

$$\lambda \cdot E(X) \cdot pr( X \geq \lambda \cdot E(X) )$$

# TSCHEBYSCHEFFsche Ungleichung

---

Satz: Sei  $X$  eine reelle Zufallsvariable.  
 $E(X)$  und  $\text{var}(X)$  existiere.

$$\text{pr}( |X - E(X)| > \lambda \sqrt{\text{var}(X)} ) < \frac{1}{\lambda^2}$$

Beweis: MARKOV Ungleichung für

$X' := (X - E(X))^2$  und  $\lambda' = \lambda^2$  ergibt:

$$\text{pr}( (X - E(X))^2 \geq \lambda' \cdot E((X - E(X))^2) ) \leq \frac{1}{\lambda'}$$

# BERNOULLI'sches Gesetz der großen Zahl

Betrachte Folge von Häufigkeitsschätzern  $Y_n$  :

$$pr\left(Y_n = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k} \quad 0 < p < 1 \text{ und } k \in \{0, \dots, n\}$$

Satz:  $\forall \varepsilon > 0$  gilt  $\lim_{n \rightarrow \infty} pr(|Y_n - p| > \varepsilon) = 0$

Beweis: Sei  $X_n := Y_n - p$ .  $E(X_n) = 0$ .

$$\text{var}(X_n) = \text{var}(Y_n - p) = \text{var}(Y_n) = \frac{1}{n^2} \cdot n \cdot p(1-p) = \frac{p(1-p)}{n}$$

Tschebyscheff-Ungleichung:

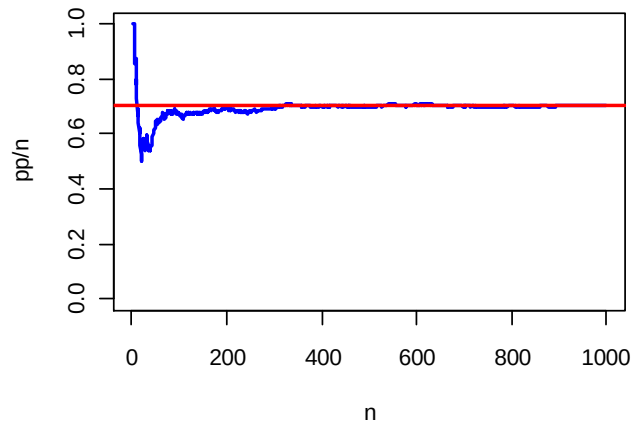
$$pr(|Y_n - p| > \lambda \sqrt{\frac{p(1-p)}{n}}) < \frac{1}{\lambda^2}$$

Setze  $\lambda := \varepsilon \sqrt{\frac{n}{p(1-p)}}$  dann gilt:

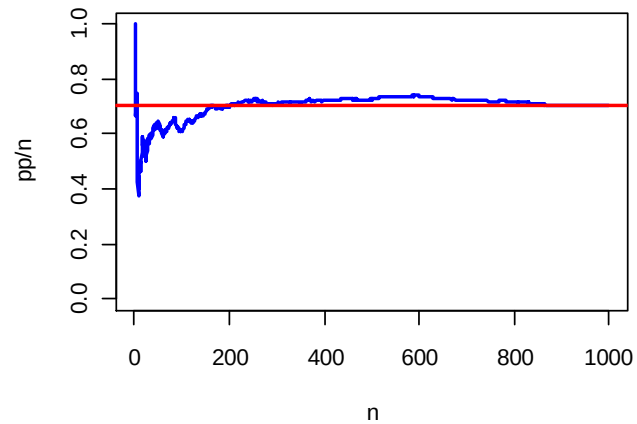
$$pr(|Y_n - p| > \varepsilon) < \frac{p(1-p)}{n \cdot \varepsilon^2} \rightarrow 0 \text{ für } n \rightarrow \infty$$

# Simulationen

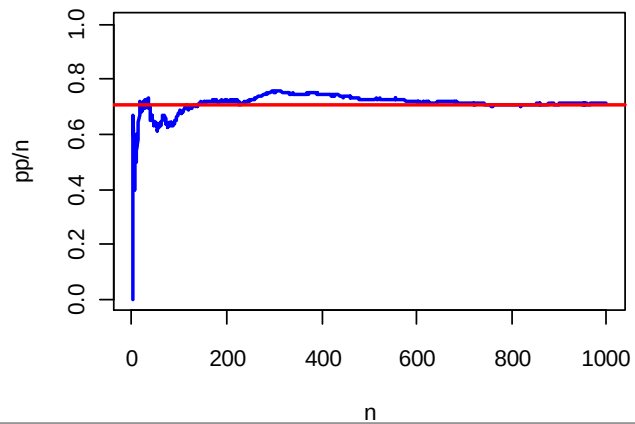
Gesetz der Grossen Zahl



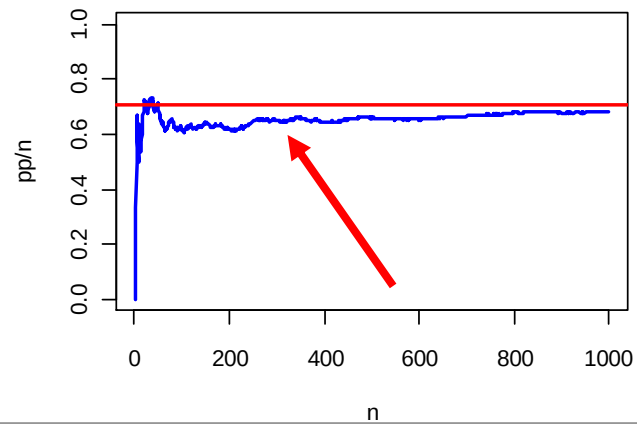
Gesetz der Grossen Zahl



Gesetz der Grossen Zahl



Gesetz der Grossen Zahl



# Bedeutung des BERNOULLIsches Gesetz der großen Zahl

---

- Dieser Satz zeigt **innerhalb der rein mathematischen Wahrscheinlichkeitstheorie**, dass die mathematische Formalisierung des intuitiven Wahrscheinlichkeitsbegriffs erklären kann, warum **Häufigkeiten bei steigender Wiederholungszahl des einschlägigen Zufallsexperiments stabil werden.**