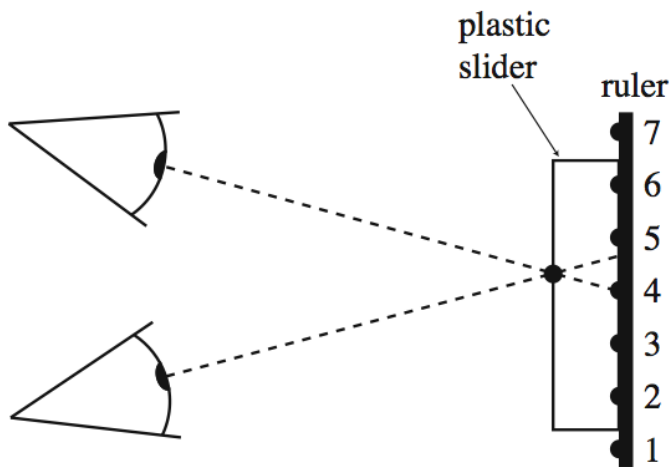
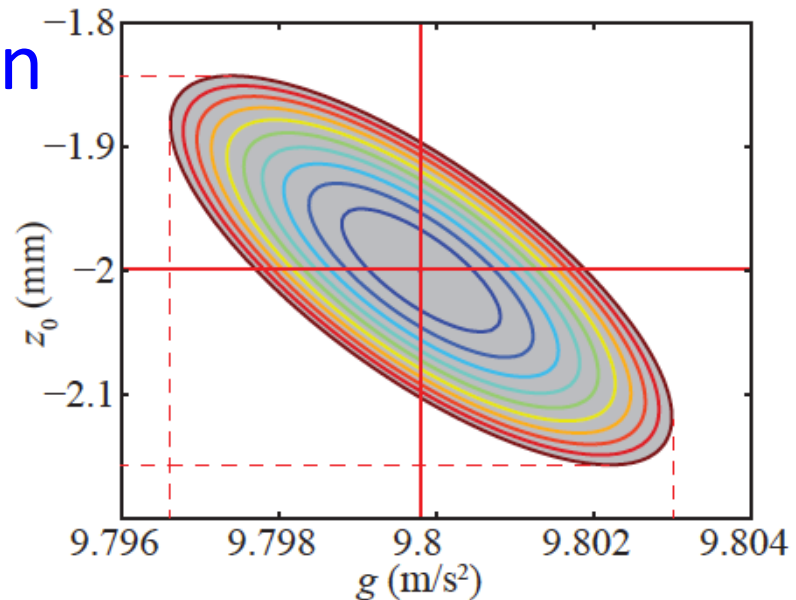


Einführung in die Datenanalyse



Thomas Ihn
HS 2014



Was ist Datenanalyse?

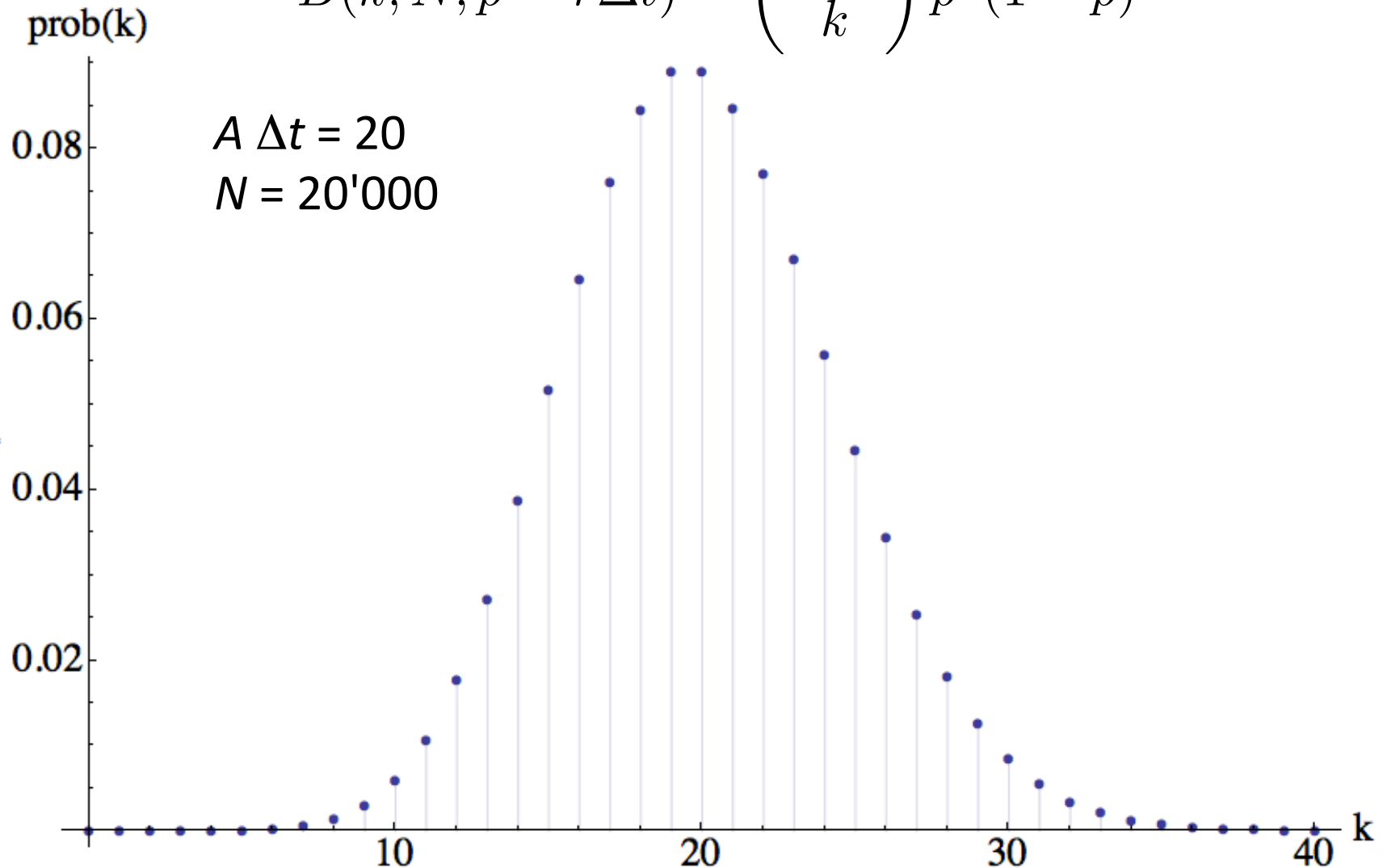
Datenanalyse ist die mathematische Form
des "gesunden Menschenverstands".

‘La théorie des probabilités n’est que le bon sens réduit au calcul.’

— Pierre Simon, Marquis de Laplace

Binomialverteilung

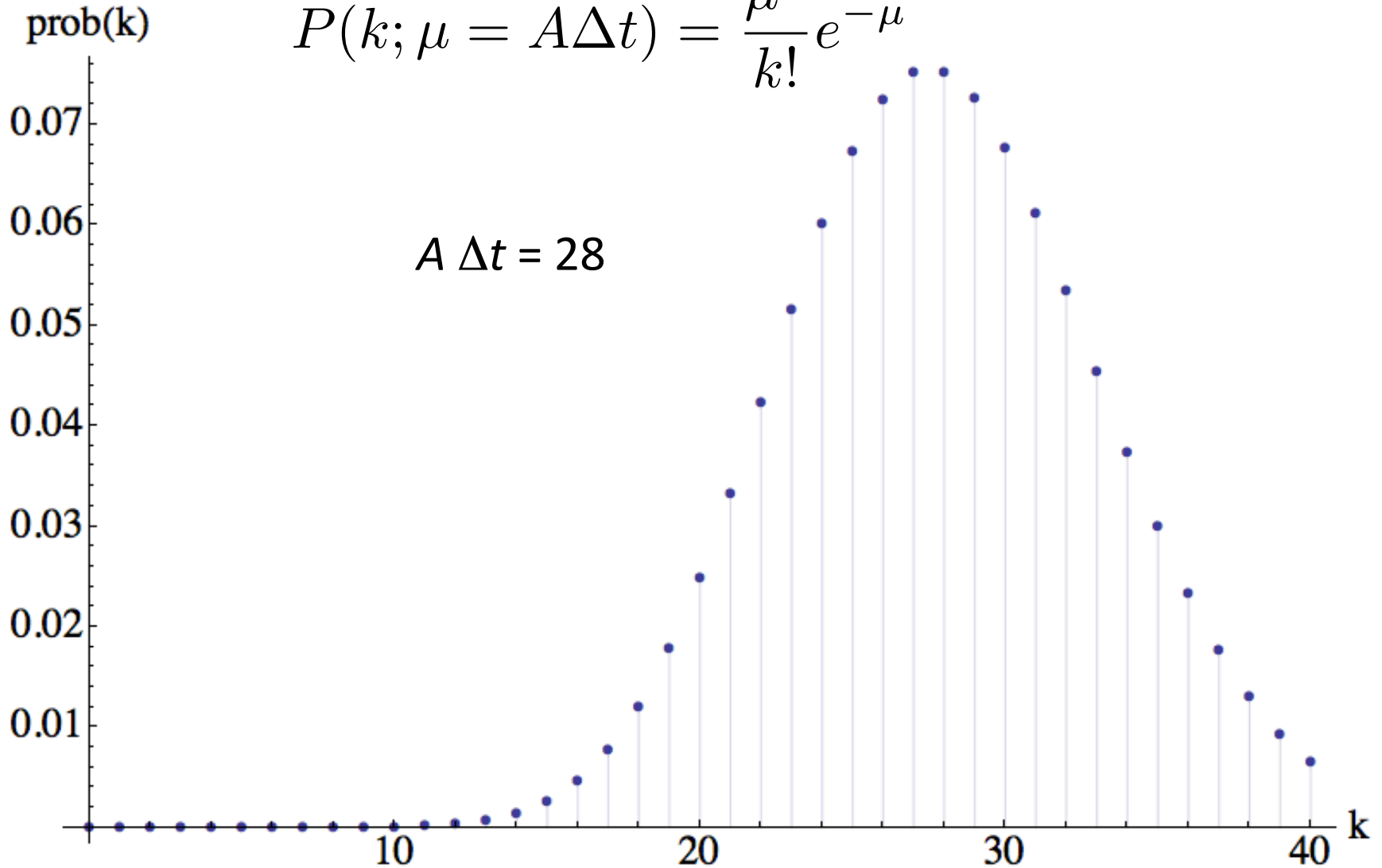
$$B(k; N, p = r\Delta t) = \binom{N}{k} p^k (1-p)^{N-k}$$



Poisson Verteilung

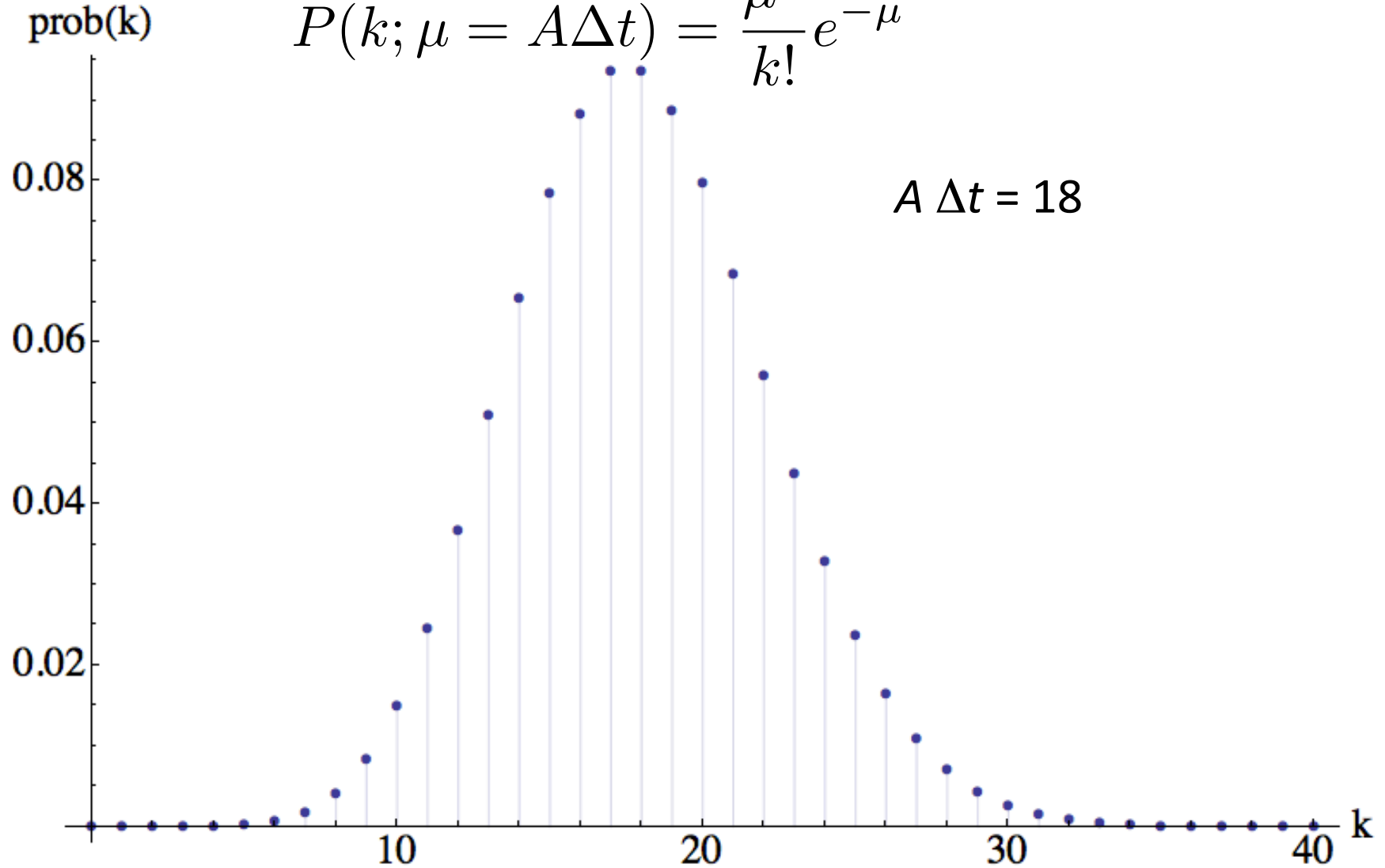
$$P(k; \mu = A\Delta t) = \frac{\mu^k}{k!} e^{-\mu}$$

$A \Delta t = 28$



Poisson Verteilung

$$P(k; \mu = A\Delta t) = \frac{\mu^k}{k!} e^{-\mu}$$



Grundregeln der Wahrscheinlichkeitsrechnung

Wir betrachten die *Aussagen*:

- G: ein Wähler wählt grün
- US: ein Wähler befürwortet die Umweltsteuer

Logische Negation:

- G: ein Wähler wählt nicht grün
- US: ein Wähler befürwortet die Umweltsteuer nicht

Grundregeln der Wahrscheinlichkeitsrechnung

Verbundwahrscheinlichkeit:

$\text{prob}(G,US)$: Wahrscheinlichkeit, dass ein Wähler grün wählt
und die Umweltsteuer befürwortet

Wahrscheinlichkeitstabelle:

	US	<u>US</u>	
G	$\text{prob}(G,US)$	$\text{prob}(G,\underline{US})$	
<u>G</u>	$\text{prob}(\underline{G},US)$	$\text{prob}(\underline{G},\underline{US})$	

Grundregeln der Wahrscheinlichkeitsrechnung

	US	<u>US</u>	
G	7.5%	2.5%	
<u>G</u>	60%	30%	

	US	<u>US</u>	
G	$\text{prob}(\text{G}, \text{US})$	$\text{prob}(\text{G}, \underline{\text{US}})$	
<u>G</u>	$\text{prob}(\underline{\text{G}}, \text{US})$	$\text{prob}(\underline{\text{G}}, \underline{\text{US}})$	

Grundregeln der Wahrscheinlichkeitsrechnung

	US	<u>US</u>	
G	7.5%	2.5%	
<u>G</u>	60%	30%	
	67.5%		

$$\text{prob}(\text{US}) = \text{prob}(\text{G}, \text{US}) + \text{prob}(\underline{\text{G}}, \text{US})$$

	US	<u>US</u>	
G	$\text{prob}(\text{G}, \text{US})$	$\text{prob}(\text{G}, \underline{\text{US}})$	
<u>G</u>	$\text{prob}(\underline{\text{G}}, \text{US})$	$\text{prob}(\underline{\text{G}}, \underline{\text{US}})$	
	$\text{prob}(\text{US})$		

Marginalisierungsregel

Grundregeln der Wahrscheinlichkeitsrechnung

	US	<u>US</u>	
G	7.5%	2.5%	10%
<u>G</u>	60%	30%	90%
	67.5%	32.5%	

$$\text{prob}(US) = \text{prob}(G, US) + \text{prob}(\underline{G}, US)$$

	US	<u>US</u>	
G	$\text{prob}(G, US)$	$\text{prob}(G, \underline{US})$	$\text{prob}(G)$
<u>G</u>	$\text{prob}(\underline{G}, US)$	$\text{prob}(\underline{G}, \underline{US})$	$\text{prob}(\underline{G})$
	$\text{prob}(US)$	$\text{prob}(\underline{US})$	

Marginalisierungsregel

Grundregeln der Wahrscheinlichkeitsrechnung

	US	<u>US</u>	
G	7.5%	2.5%	10%
<u>G</u>	60%	30%	90%
	67.5%	32.5%	

$$\text{prob}(\text{US} | \text{G}) = \text{prob}(\text{G}, \text{US}) / \text{prob}(\text{G})$$

	US	<u>US</u>	
G	$\text{prob}(\text{G}, \text{US})$	$\text{prob}(\text{G}, \underline{\text{US}})$	$\text{prob}(\text{G})$
<u>G</u>	$\text{prob}(\underline{\text{G}}, \text{US})$	$\text{prob}(\underline{\text{G}}, \underline{\text{US}})$	$\text{prob}(\underline{\text{G}})$
	$\text{prob}(\text{US})$	$\text{prob}(\underline{\text{US}})$	

bedingte Wahrscheinlichkeit

Grundregeln der Wahrscheinlichkeitsrechnung

	US	<u>US</u>	
G	7.5%	2.5%	10%
<u>G</u>	60%	30%	90%
	67.5%	32.5%	100%

$$\text{prob}(US) + \text{prob}(\underline{US}) = 1$$

	US	<u>US</u>	
G	$\text{prob}(G, US)$	$\text{prob}(G, \underline{US})$	$\text{prob}(G)$
<u>G</u>	$\text{prob}(\underline{G}, US)$	$\text{prob}(\underline{G}, \underline{US})$	$\text{prob}(\underline{G})$
	$\text{prob}(US)$	$\text{prob}(\underline{US})$	1

Normierung von Wahrscheinlichkeitsverteilungen

Grundregeln der Wahrscheinlichkeitsrechnung

Produktregel

$$\text{prob}(G,US) = \text{prob}(G)\text{prob}(US | G) = \text{prob}(US)\text{prob}(G | US)$$

Grundregeln der Wahrscheinlichkeitsrechnung

Produktregel

$$\text{prob}(G,US) = \text{prob}(G)\text{prob}(US | G) = \text{prob}(US)\text{prob}(G | US)$$

Bayes'sches Theorem

Produktregel nach $\text{prob}(US | G)$ auflösen:

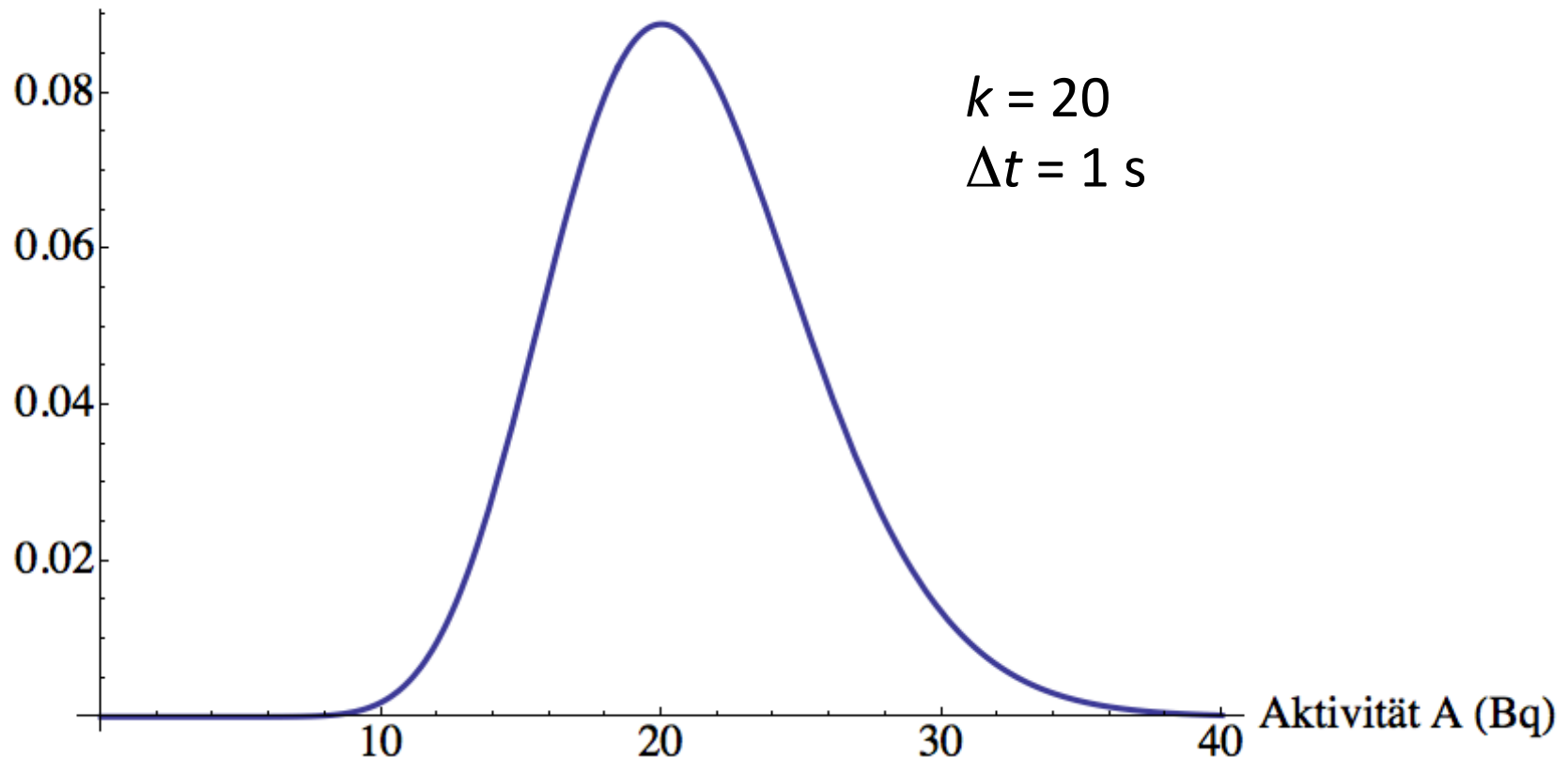
$$\text{prob}(US | G) = \frac{\text{prob}(US)\text{prob}(G | US)}{\text{prob}(G)}$$

Wahrscheinlichkeitsdichte für die Aktivität der Quelle

$$\text{pdf}(A|k, \Delta t, M, I) = \Delta t \frac{(A\Delta t)^k}{k!} e^{-A\Delta t} = \Gamma(A; k + 1, 1/\Delta t)$$

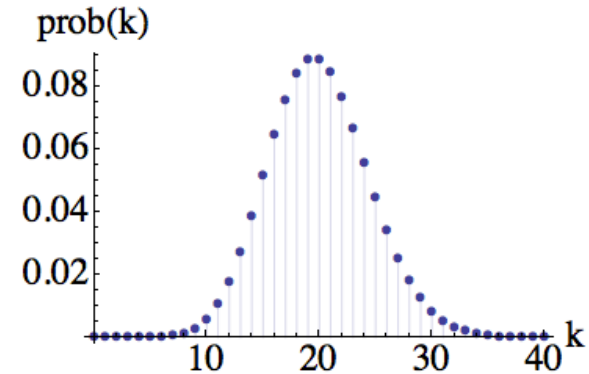
Gamma-Verteilung

pdf(A|k,Δt,M,I)



Zusammenfassung

- Wahrscheinlichkeitsverteilungen:
 - Binomialverteilung
 - Poissonverteilung
 - Gammaverteilung
- Grundregeln und Begriffe der Wahrscheinlichkeitsrechnung:
 - bedingte Wahrscheinlichkeit
 - Marginalisierungsregel
 - Produktregel
 - Bayes'sches Theorem
- Grundbegriffe der Datenanalyse
 - Modell und Hintergrundinformation
 - Parameterabschätzung mit Hilfe des Bayes'schen Theorems
 - bedingte Wahrscheinlichkeiten invertieren



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Literaturempfehlungen

Zwei **dünne** Bücher für Praktiker:

- D.S. Sivia, Data Analysis: A Bayesian Tutorial, Oxford University Press
- H.J.C. Berendsen: A Student's Guide to Data and Error Analysis, Cambridge University Press

Mehr über Wahrscheinlichkeitstheorie:

- E.T. Jaynes, Probability Theory, The Logic of Science, Cambridge University Press

