

Network Analysis

IN2045

Chapter 3a - Statistics

Dr. Alexander Klein

Stephan Günther

Prof. Dr.-Ing. Georg Carle

Chair for Network Architectures and Services

Department of Computer Science

Technische Universität München

<http://www.net.in.tum.de>





Topics

- ❑ Random Variable
- ❑ Probability Space
- ❑ Discrete and Continuous RV
- ❑ Frequency Probability(Relative Häufigkeit)
- ❑ Distribution(discrete)
- ❑ Distribution Function(discrete)
- ❑ PDF,CDF
- ❑ Expectation/Mean, Mode,
- ❑ Standard Deviation, Variance, Coefficient of Variation
- ❑ p-percentile(quantile), Skewness, Scalability Issues(Addition)
- ❑ Covariance, Correlation, Autocorrelation
- ❑ Visualization of Correlation
- ❑ PP-Plot
- ❑ QQ-Plot
- ❑ Waiting Queue Examples



Classic definition of probability

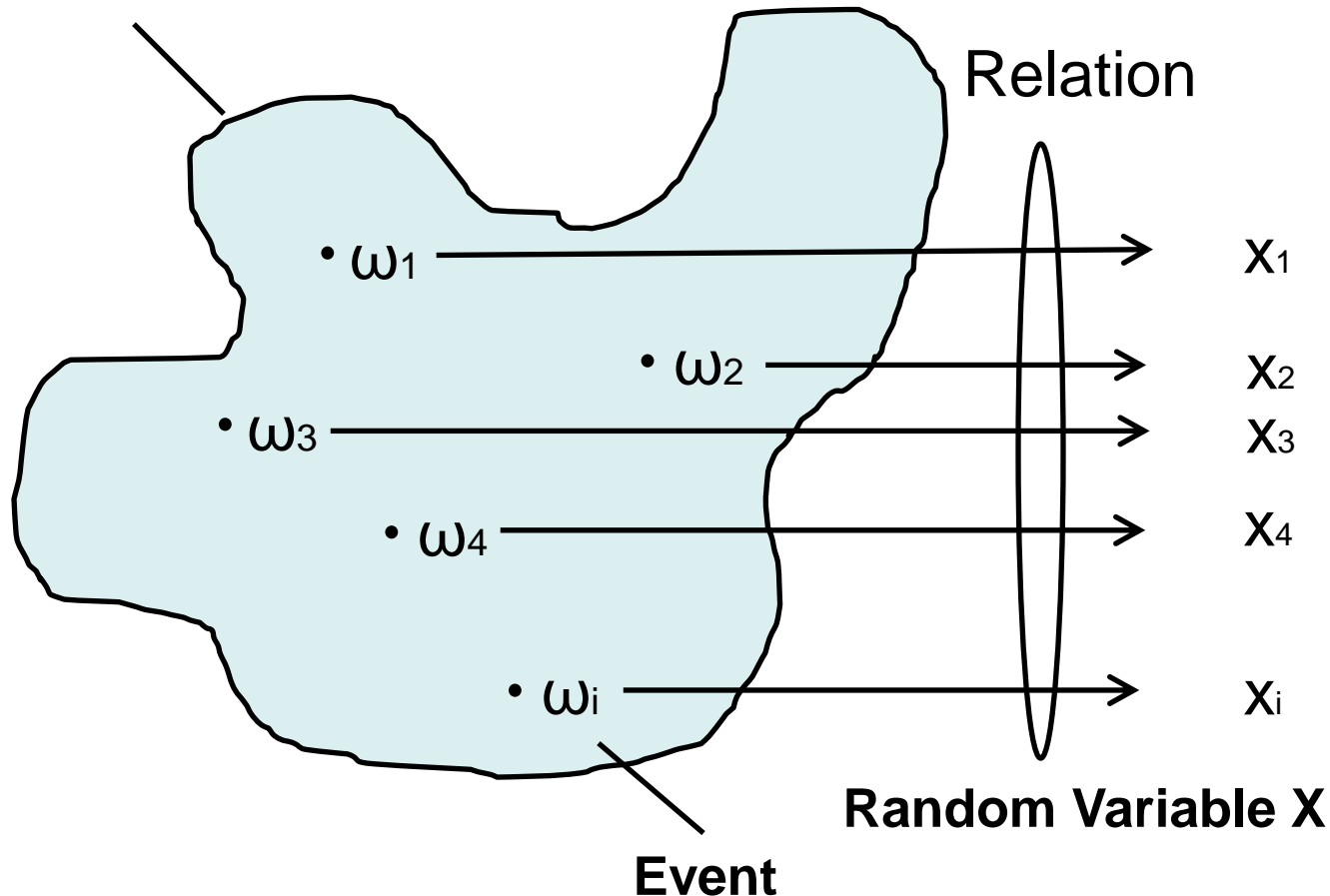
The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

- Pierre-Simon Laplace, A Philosophical Essay on Probabilities



□ Random Variable

Probability Space (Ereignisraum) $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \dots, \omega_i\}$





□ Discrete Random Variable:

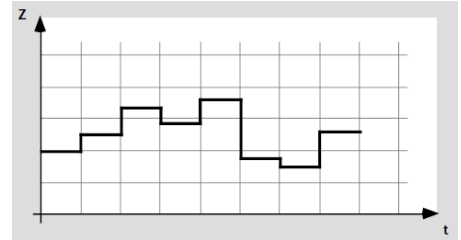


- Example: Flipping of a coin

- $\omega_1 = \{\text{head-0}\}, \omega_2 = \{\text{tail-1}\}$
- $X \in \{0, 1\}$

$$\Rightarrow \Omega = \{\omega_1, \omega_2\}$$

Countable



- Example: Rolling two dice



- $\omega_1 = \{2\}, \omega_2 = \{3\}, \dots, \omega_{11} = \{12\}$
- $X \in \{2, 3, 4, \dots, 12\}$

$$\Rightarrow \Omega = \{\omega_1, \omega_2, \dots, \omega_{11}\}$$

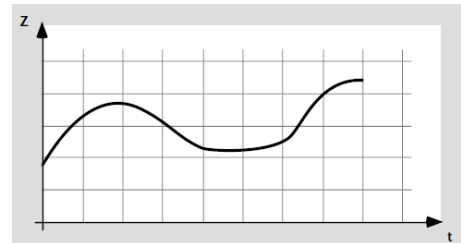
□ Continuous Random Variable:

- Example: Round Trip Time

- $T \in \{5\text{ms}, 200\text{ms}\}$
- $\omega_1 = \{t < 10\text{ms}\}, \omega_2 = \{10\text{ms} \leq t < 20\text{ms}\},$
 $\omega_3 = \{t \geq 20\text{ms}\} \Rightarrow \Omega = \{\omega_1, \omega_2\}$

- Example: Sensed Interference Level

Uncountable



Discrete or not discrete



□ Frequency Probability / Law of large numbers (Relative Häufigkeit)

- Number of random experiments
 - n total number of trials
 - X_i event or characteristic of the outcome
 - n_i number of trials where the event X_i occurred

$$h(X_i) = \frac{n_i}{n} \quad 0 \leq h(X_i) \leq 1 \quad \sum_i h(X_i) = 1 \quad \text{Vollständigkeitsrelation}$$

$$P(X_i) = \lim_{n \rightarrow \infty} \frac{n_i}{n} \quad 0 \leq P(X_i) \leq 1 \quad \sum_i P(X_i) = 1 \quad X_i \text{ disjoint}$$



□ Vollständiges Ereignissystem

$$P(Y) = \sum_{i=1}^N P(X_i)$$

□ Verbundereignis

$$P(X \cap Y) = P(X, Y) = P(Y, X)$$

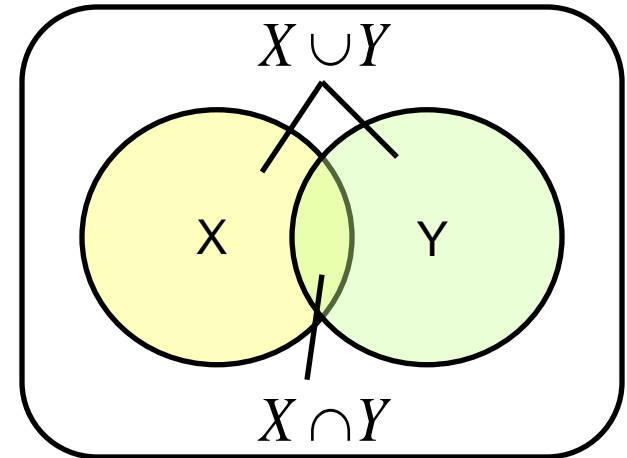
$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

□ Bedingte Wahrscheinlichkeit

$$P(X | Y) = \frac{P(X, Y)}{P(Y)} \quad \Rightarrow \quad P(X | Y) \geq P(X, Y)$$

□ Statistische Unabhängigkeit

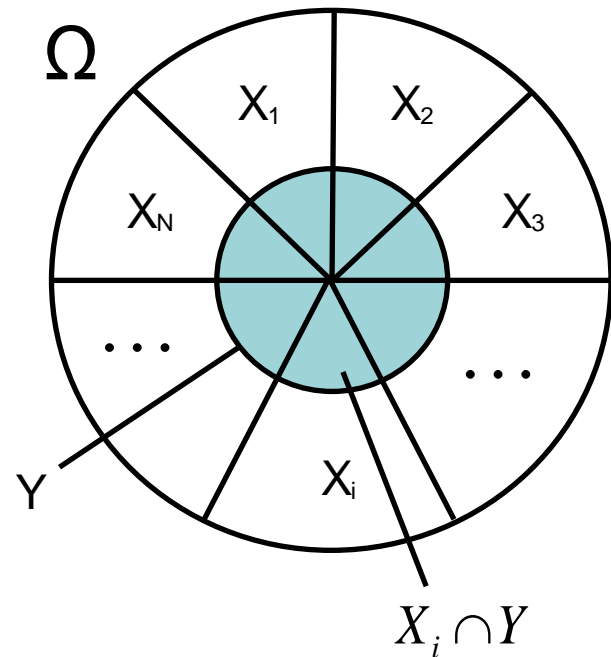
$$P(X | Y) = P(X) \quad \vee \quad P(X, Y) = P(X)P(Y)$$





□ Vollständiges Ereignissystem

- $$P(Y) = \sum_{i=1}^N P(X_i, Y)$$



□ Bayes Function

- $$P(X_i | Y) = \frac{P(Y | X_i) \cdot P(X_i)}{P(Y)} = \frac{P(Y | X_i) \cdot P(X_i)}{\sum_{k=1}^N P(Y | X_k) \cdot P(X_k)}$$



□ **Distribution** (Verteilung)

X – discrete random variable

- Function $x(i) = P(X = i)$, $i = 0, 1, 2, \dots, X_{\max}$ (Distribution)
– $x(i) \in [0, 1]$

$$-\sum_{i=0}^{X_{\max}} x(i) = 1 \quad (\text{Vollständigkeitsrelation})$$

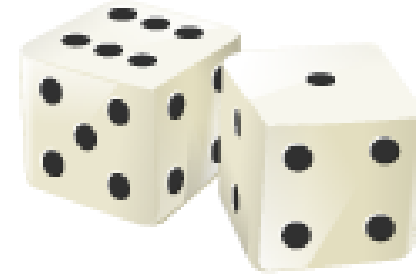
- Example:

Rolling two dice

- $\omega_1 = \{2\}, \omega_2 = \{3\}, \dots, \omega_{11} = \{12\} \Rightarrow \Omega = \{\omega_1, \omega_2, \dots, \omega_{11}\}$
- $X \in \{2, 3, 4, \dots, 12\}$



- Example: Throwing two dice



Sample Space (Ereignisraum)

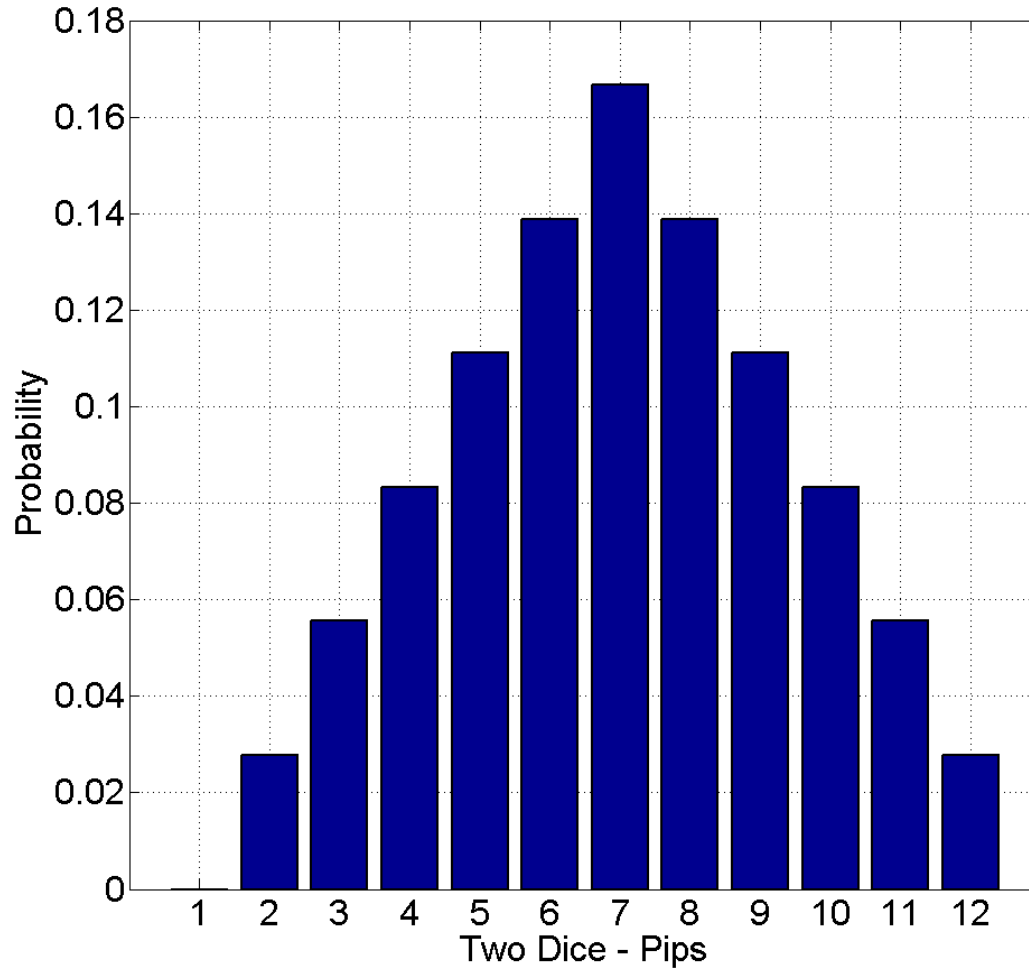
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)



- Example: Throwing two dice

Sample Space (Ereignisraum)

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)



Distribution

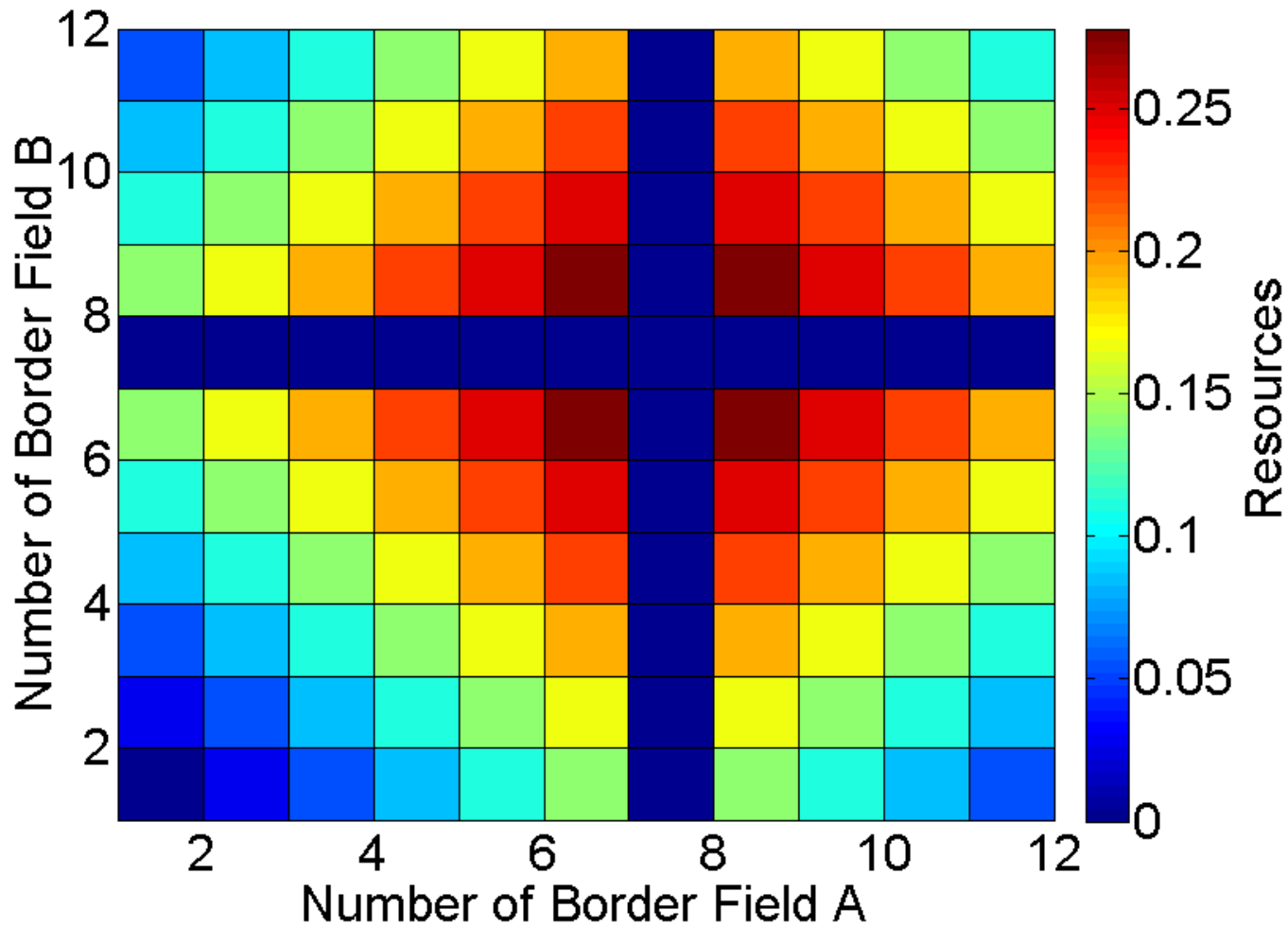


Palour Game: Die Siedler von Catan

- Rules:
 - Players are only allowed to build along borders of a field
 - Players roll two dice
 - If the sum of the dice corresponds to the number of the field, the player gets the resources from this field

- Question
 - Where is the best place for a building?





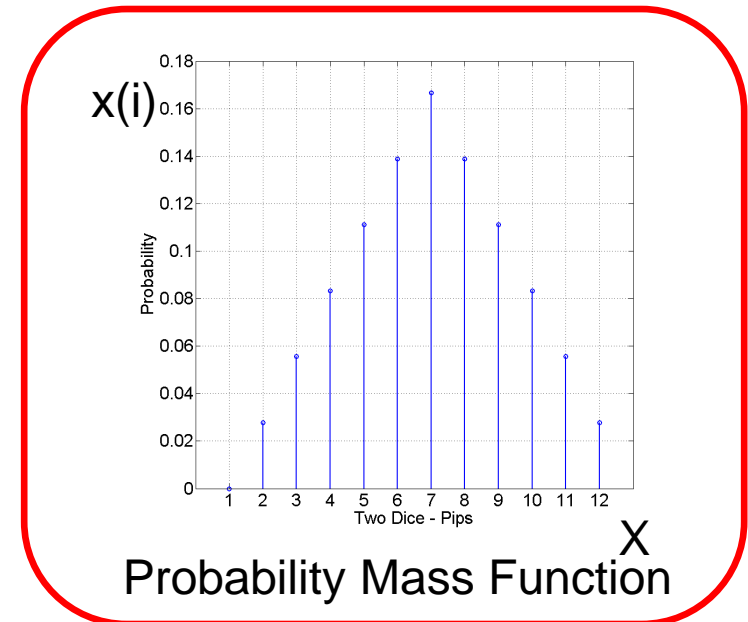


□ Probability Mass Function (Verteilung)

- Discrete random variable X
- i value of the random variable X
- $x(i)$ probability that the outcome of random variable X is i

- $x(i) = P\{X = i\}, \quad i = 0, 1, \dots, X_{\max} \quad (\text{Distribution})$

- $\sum_{i=0}^{X_{\max}} x(i) = 1 \quad (\text{Vollständigkeitsrelation})$

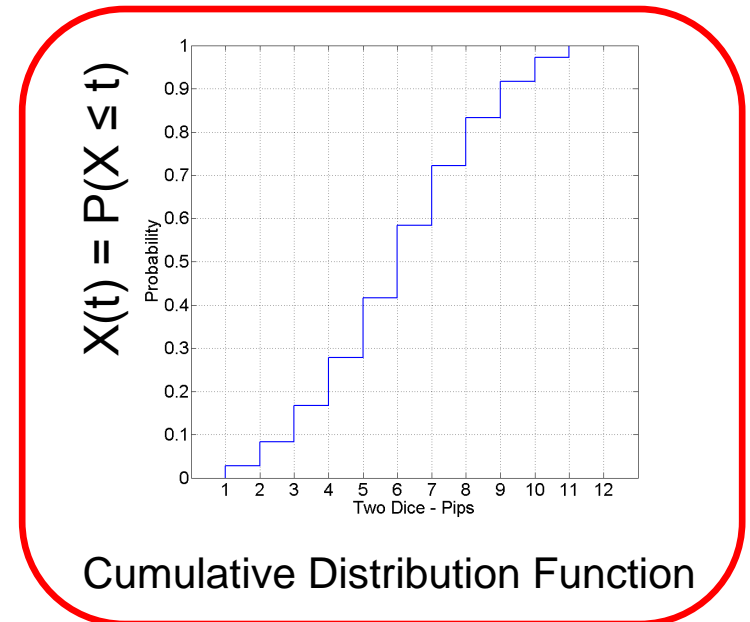




□ Cumulative Distribution Function (Verteilungsfunktion)

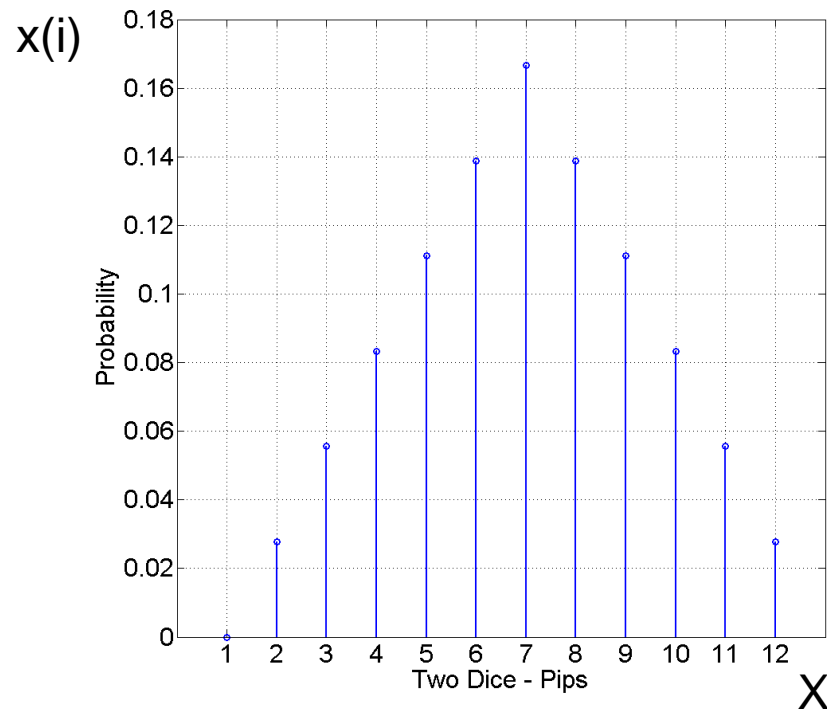
$$X(t) = P\{X \leq t\}$$

- $t_1 < t_2 \quad \Rightarrow \quad X(t_1) \leq X(t_2) \quad (\text{monotony})$
- $t_1 < t_2 \quad \Rightarrow \quad P\{t_1 < X \leq t_2\} = X(t_2) - X(t_1)$
- $X(-\infty) = 0 \quad \wedge \quad X(\infty) = 1$
- $X^c(t) = 1 - X(t) = P\{X > t\}$

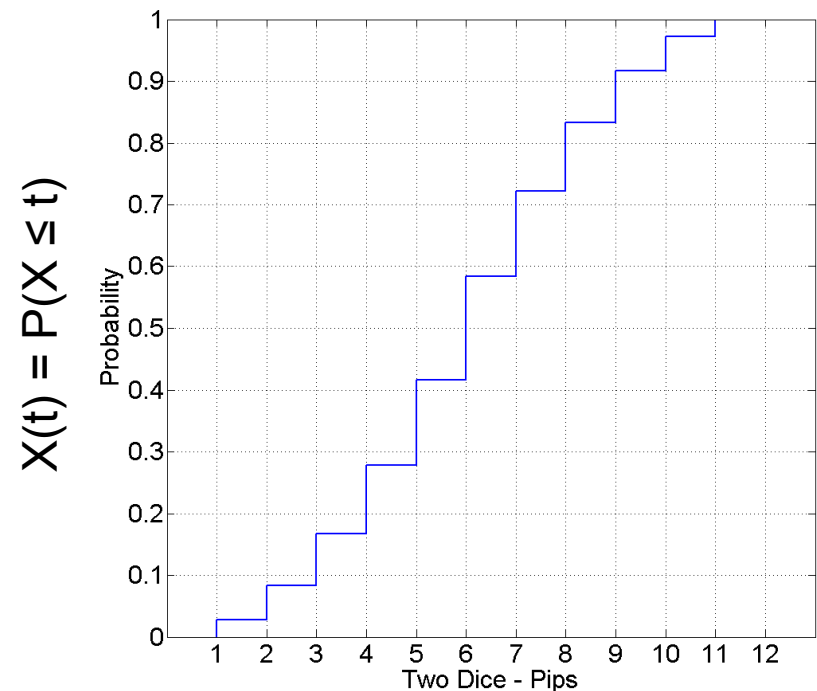




Difference between probability mass function and cumulative distribution function



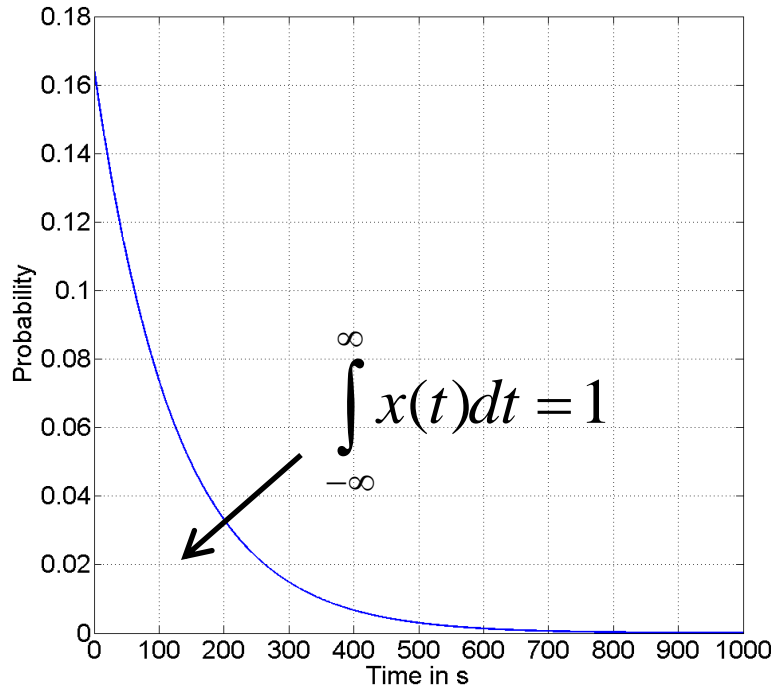
Probability Mass Function
(Verteilung)



Cumulative Distribution Function
(Verteilungsfunktion)

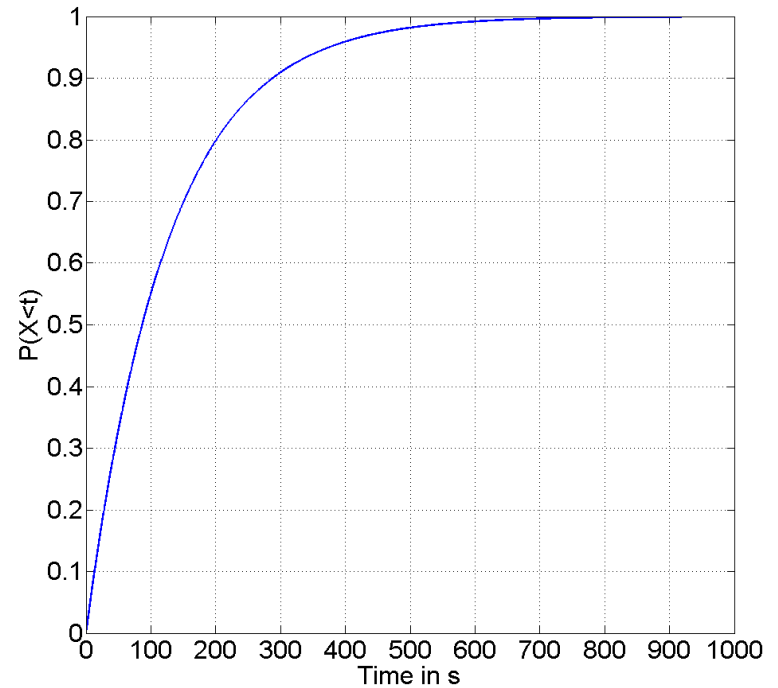


Continuous random variable



Probability Density Function
(Verteilungsdichtefunktion)

$$x(t) = \frac{d}{dt} X(t)$$



Cumulative Density Function

$$X(t) = \int_{-\infty}^t x(t)dt$$



□ **Expectation** (Erwartungswert)

- X : Probability density function
- $g(x)$: Function of random variable X

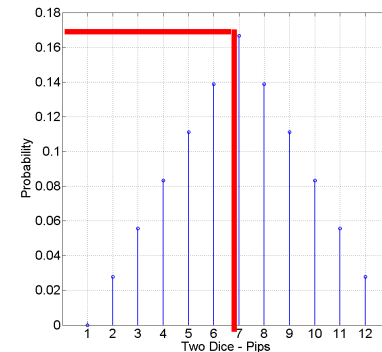
$$E[g(X)] = \int_{-\infty}^{\infty} g(t) \cdot x(t) dt$$

□ **Mean** (Mittelwert einer Zufallsvariablen)

$$m_1 = E[X] = \int_{-\infty}^{\infty} t \cdot x(t) dt$$

□ **Mode** (Outcome of the random variable with the highest probability)

$$c = \text{Max}(x(t))$$





- Gewöhnliche Momente einer Zufallsvariablen

- $g(X) = X^k \implies m_k = E[X^k] = \int_{-\infty}^{\infty} t^k \cdot x(t) dt, \quad k = 0, 1, 2, \dots$

- Central moment (Zentrales Moment)

- Variation of the random variable in respect to its mean

$$g(X) = (X - m_1)^k$$

$$\implies \mu_k = E[(X - m_1)^k] = \int_{-\infty}^{\infty} (t - m_1)^k \cdot x(t) dt, \quad k = 0, 1, 2, \dots$$

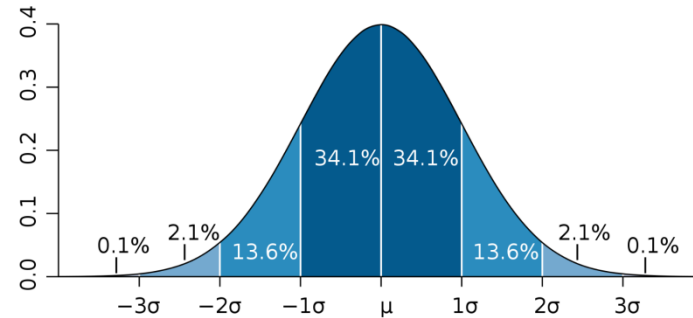
- Special Case (k=2):

$$\mu_2 = E[(X - m_1)^2] = \text{VAR}[X]$$



□ Standard deviation (Standardabweichung)

- $\sigma_X = \sqrt{\text{VAR}[X]}$



□ Coefficient of variation (Variationskoeffizient)

- $c_X = \frac{\sigma_X}{E[X]}, \quad E[X] > 0$

- The coefficient of variation is a normalized measure of dispersion of a probability distribution
- It is a dimensionless number which does not require knowledge of the mean of the distribution in order to describe the distribution

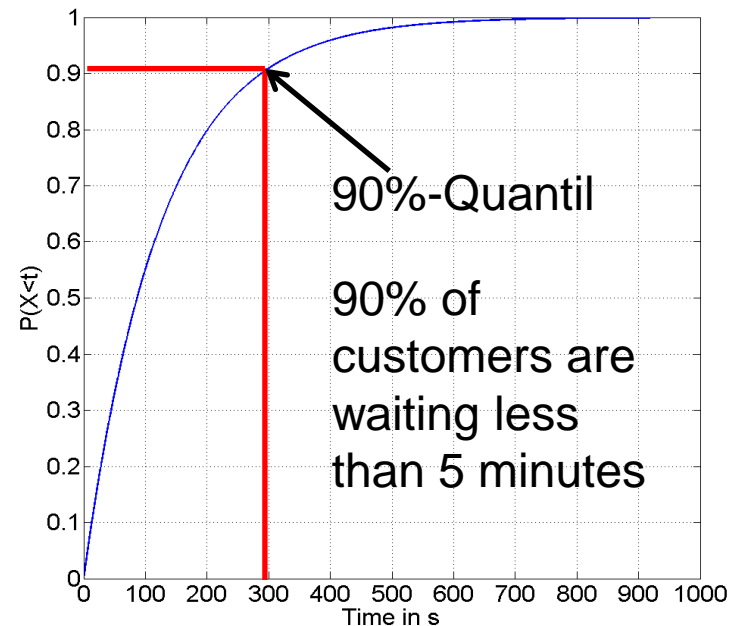
Picture taken from Wikipedia



□ **p-percentile t_p** (p-Quantil)

A percentile is the value of a variable below which a certain percent of observations fall

- VDF $F : R \rightarrow (0,1)$ (bijective)
- $F(x) = P(X < x) = p$
- $F^{-1}(p) = \inf\{x \in R : p \leq F(x)\}$
- Special Case:
 - Median 0.5-percentile
 - Upper percentile 0.75-percentile
 - Lower percentile 0.25-percentile
- Typical Use Case:
 - QoS in networks
(e.g. 99.9%-percentile of the delay)



Cumulative Density Function

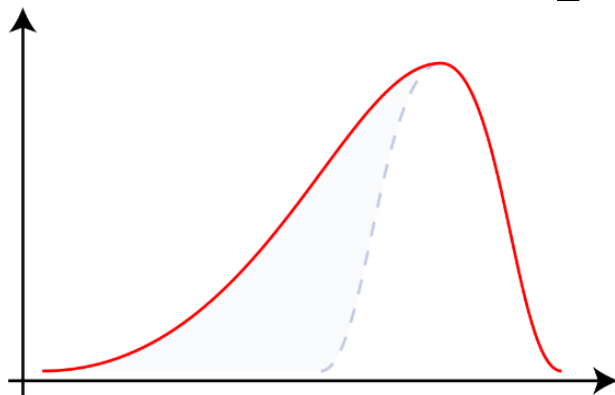


□ Skewness (Schiefe)

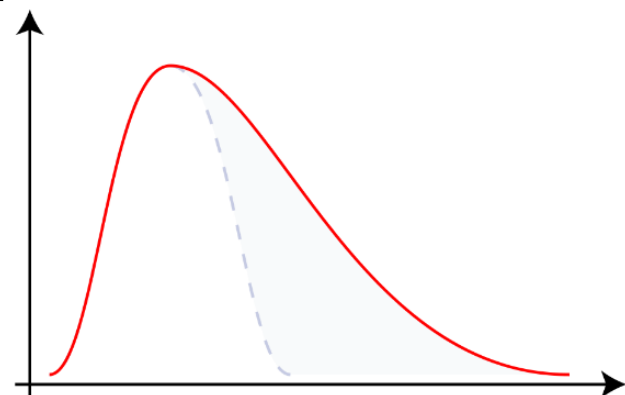
Skewness describes the asymmetry of a distribution

- $v < 0$: The left tail of the distribution is longer (linksschief)
=> Mass is concentrated in the right
- $v > 0$: The right tail of the distribution is longer (rechtsschief)
=> Mass is concentrated in the left

$$v_X = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$



Negative Skew



Positive Skew

Picture taken from Wikipedia



□ Scalability Issues

- Multiplication of a random variable X with a scalar s

- $Y = s \cdot X$

- $E[Y] = s \cdot E[X]$

- $VAR[Y] = s^2 \cdot VAR[X]$

- Addition of two random variables X and Y

- $Z = X + Y$

- $E[Z] = E[X] + E[Y]$

- $VAR[Z] = VAR[X] + VAR[Y]$ (only if X and Y independent)



□ Covariance

Covariance is a measure which describes how two variables change together

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X] \cdot E[Y]$$

- Special Case: $\text{Cov}(X, X) = \text{VAR}[X]$
- Other Characteristics:
 - $\text{Cov}(X, a) = 0$
 - $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
 - $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
 - $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$





□ Correlation function

Correlation function describes how two random variable tend to deviate from their expectation

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{VAR(X) \cdot VAR(Y)}}$$

▪ Characteristics:

- $Y = X$  $Cor(X, Y) = 1$ (Maximum positive)
- $Y = -X$  $Cor(X, Y) = -1$ (Maximum negative)
- $Cor(X, Y) > 0$ Both random variable tend to have either high or low values (difference to their expectation)
- $Cor(X, Y) < 0$ The random variables differ from each other such that one has high values while the other has low values and vice versa (difference to their expectation)



□ Autocorrelation (LK 4.9)

- Autocorrelation is the cross-correlation of a signal with itself. In the context of statistics it represents a metric for the similarity between observations of a stochastic process. From a mathematical point of view, autocorrelation can be regarded as a tool for finding repeating patterns of a stochastic process.

Definition:

- Correlation of two samples with distance k from a stochastic process X is given by:

$$\Rightarrow \text{Cor}(X, Y) \quad \text{with} \quad Y_i = X_{i+j}$$

Use case:

- Test of random number generators
- Evaluation of simulation results (c.f. Batch-Means)



Example:

00101110101001101100010011101010100011

00101110101001101100010011101010100011

Random

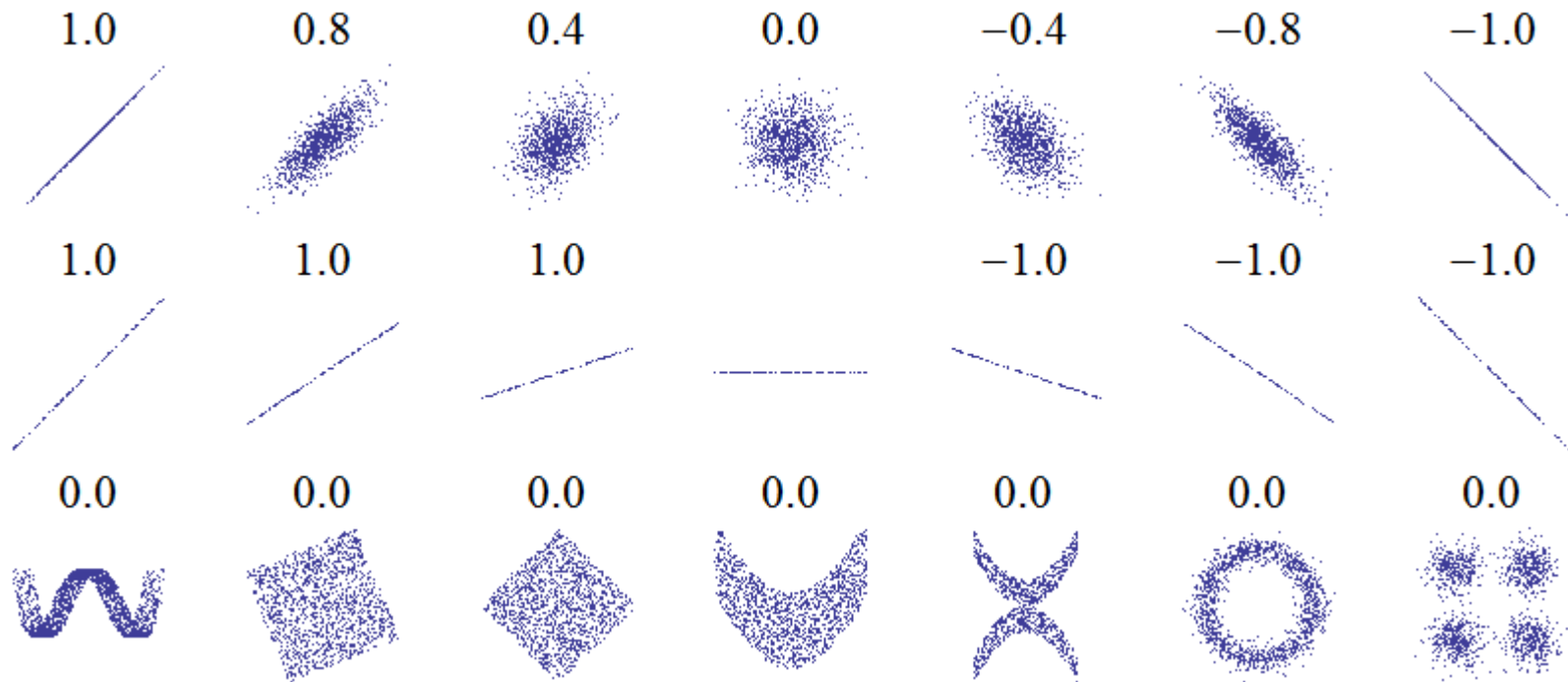


Autocorrelation Lag 4



□ Visualization of Correlation

Example: Two random variables X and Y are plotted against each other



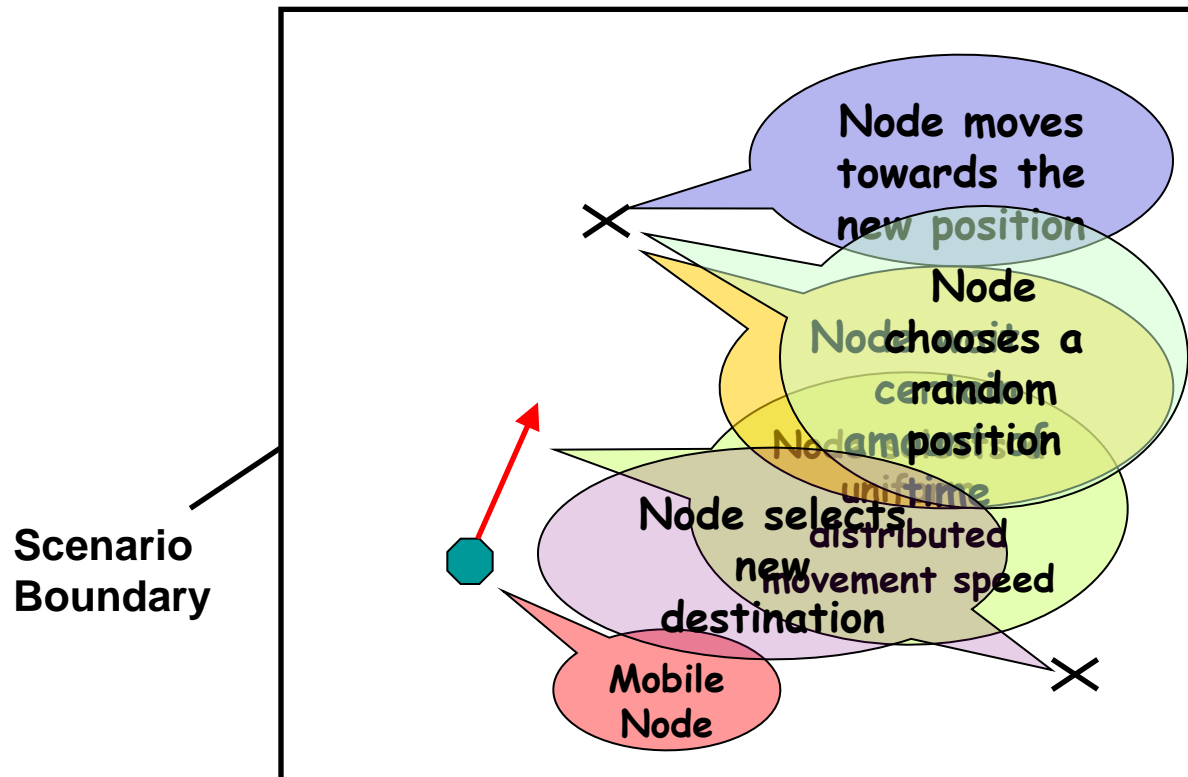
Picture taken from Wikipedia



□ Impact of correlation (1/2)

Example: Random Waypoint mobility model

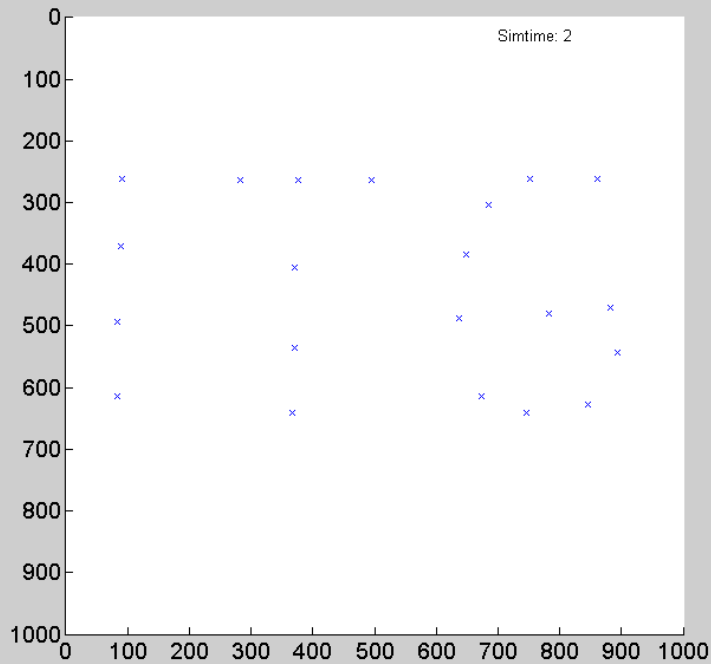
Algorithm



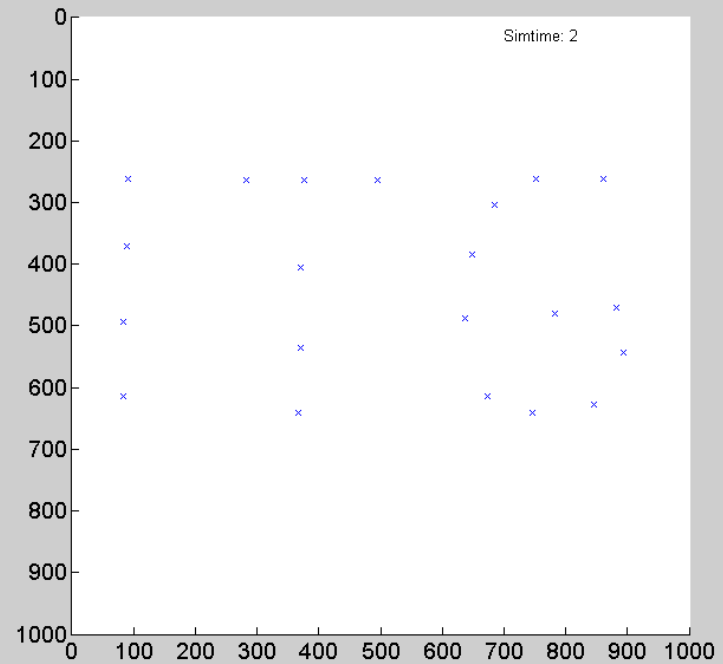


□ Impact of correlation (2/2)

Example: Random Waypoint mobility model



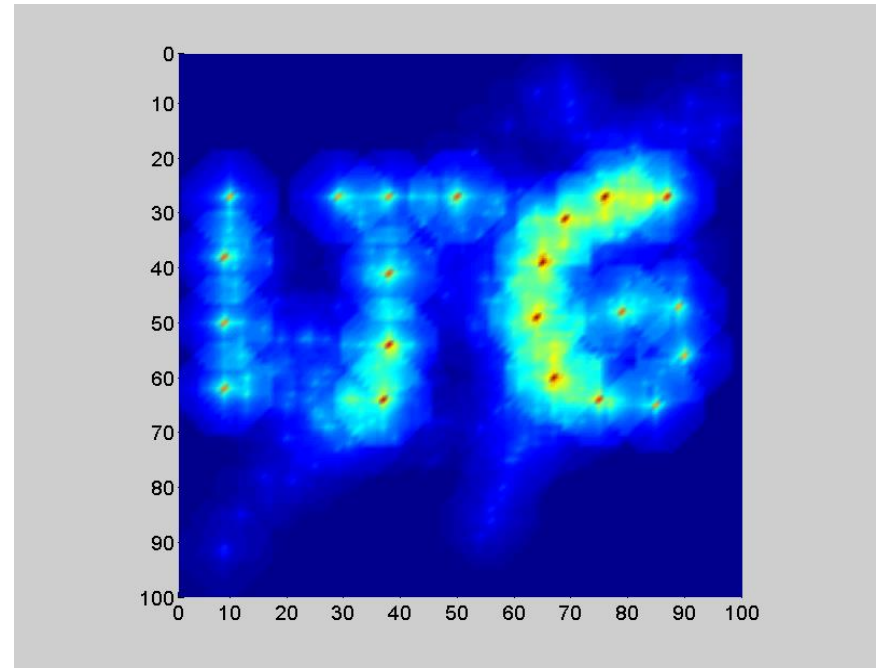
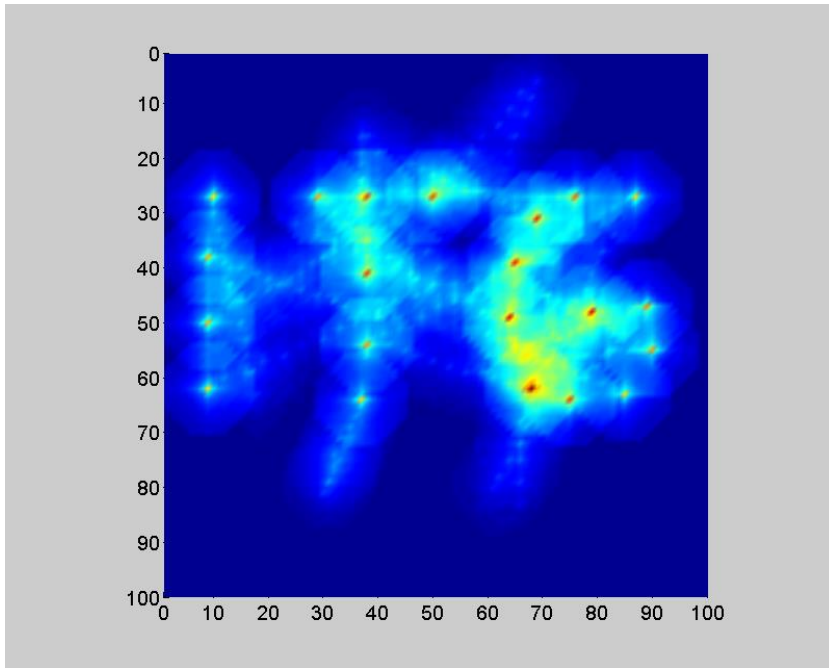
Uncorrelated next position selection



Correlated next position selection



□ Mobility Example





- ❑ Visual comparison of different distributions
 - ❑ **Quantile-Quantile Plot**
 - ❑ **Probability-Probability Plot**

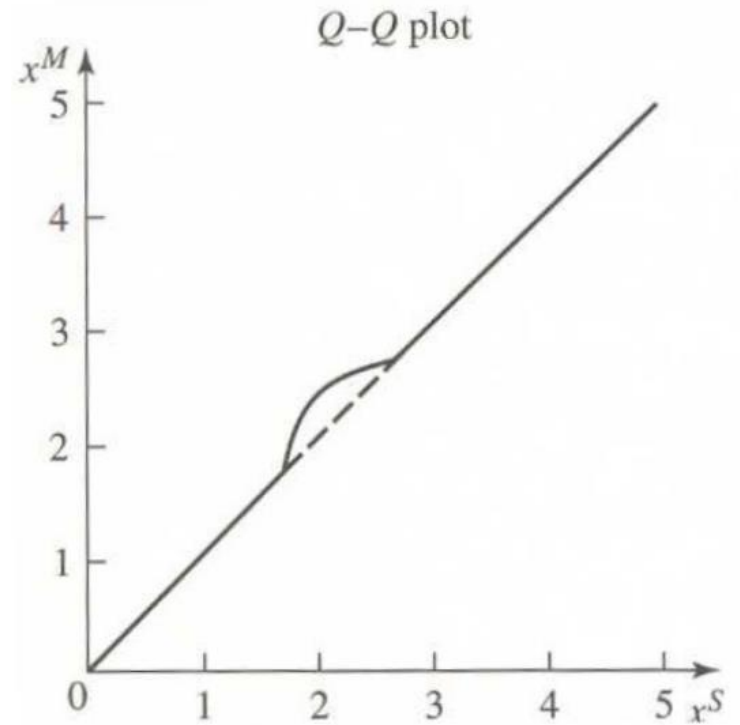
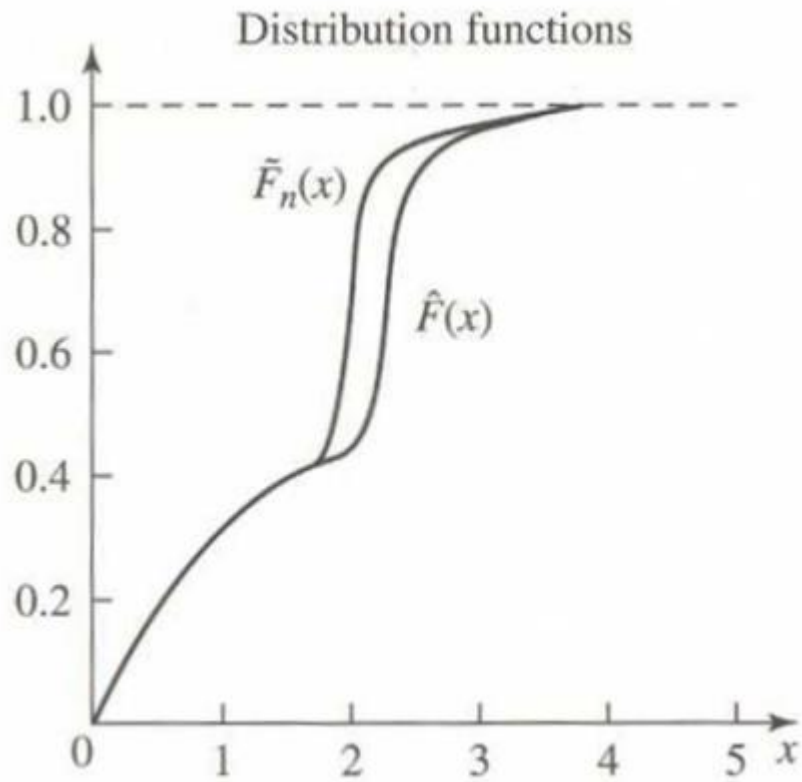


Quantile-Quantile plots (QQ plots)

- Usage: Compare two distributions against each other
 - Usually: Measurement distribution vs. theoretical distribution – do the measurements fit an assumed underlying theoretical model?
 - Also possible: Measurement distribution vs. other measurement distribution – are the two measurement runs really from the same population, or is there variation between the two?
- How it works:
 - Determine 1% quantile, 2% quantile, ..., 100% quantile for distributions
 - Plot 1% quantile vs. 1% quantile, 2% quantile vs 2% quantile, etc.
 - Not restricted to percentiles – usually, each of the n data points from the measurement is taken as its own $1/n$ quantile
- How to read:
 - If everything is located along the line $x=y$ then the two distributions are very similar
 - QQ plots amplify discrepancies near the “tail” of the distributions
- Warning about scales:
 - Plot program often automatically assign X and Y different scales
 - Straight line indicates: choice of distribution OK, but parameters don't fit

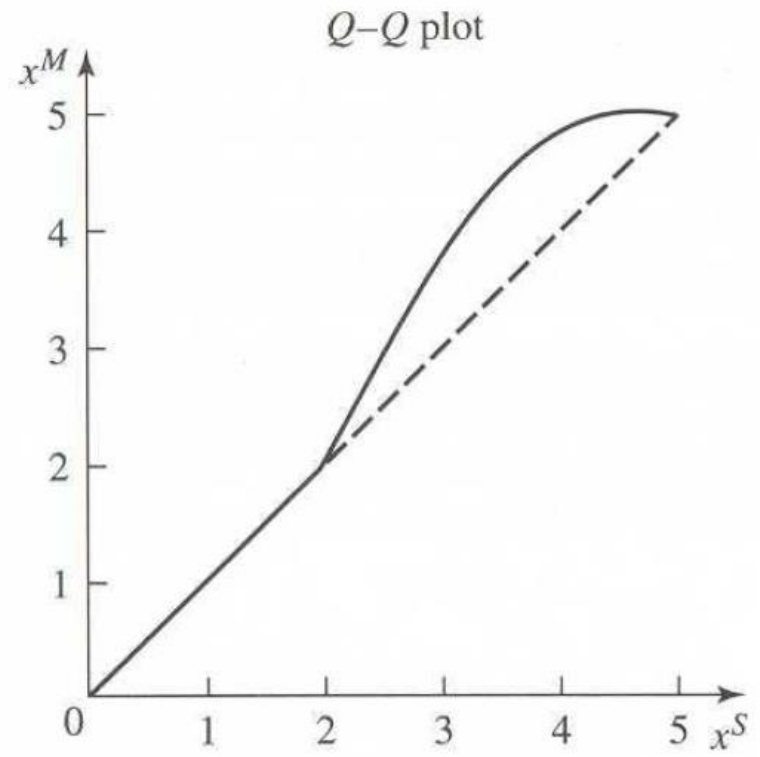
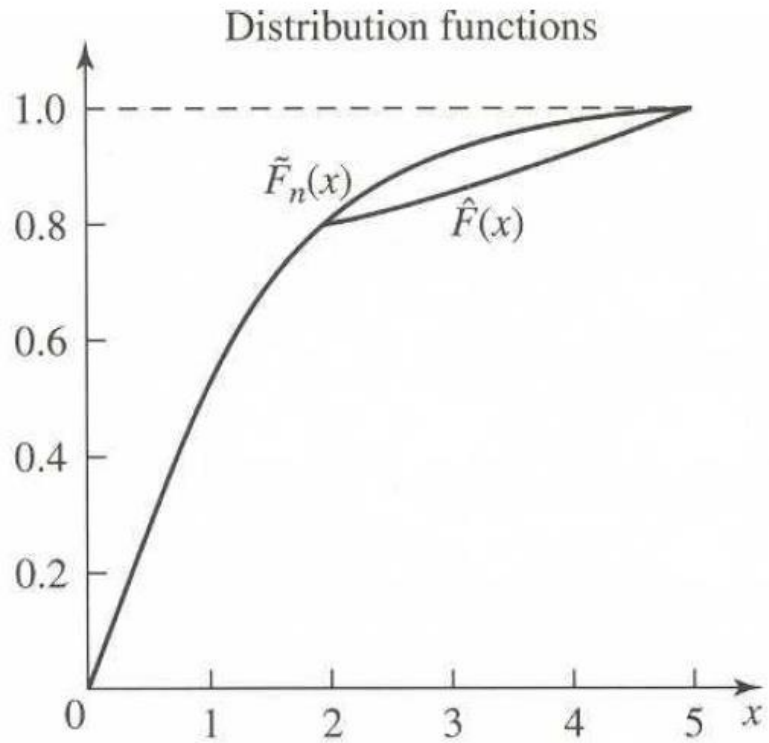


□ QQ Plot





□ QQ Plot





Probability-Probability plots (PP plots)

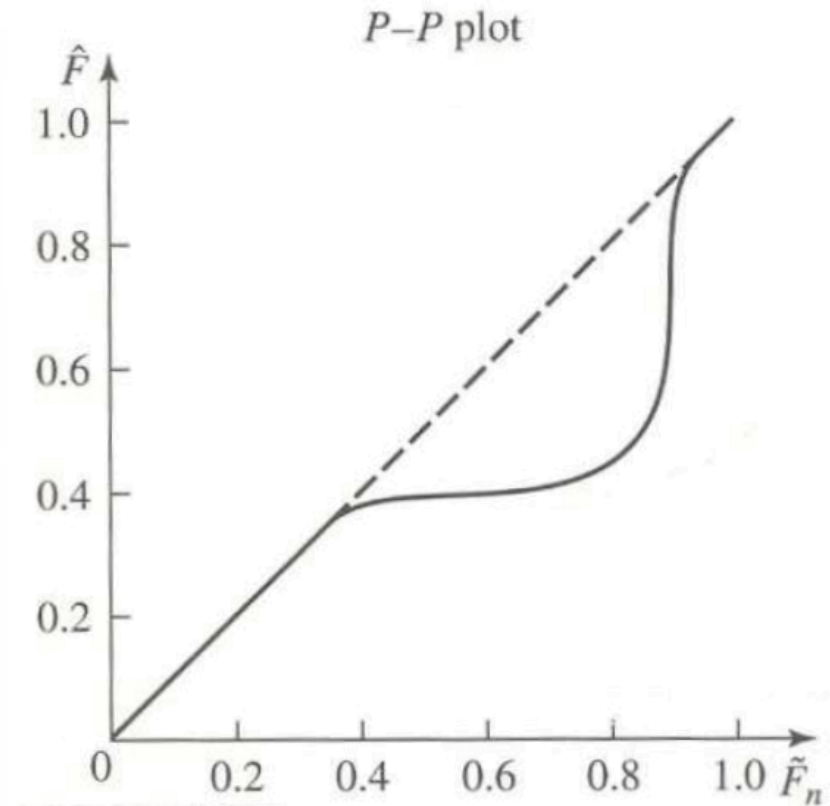
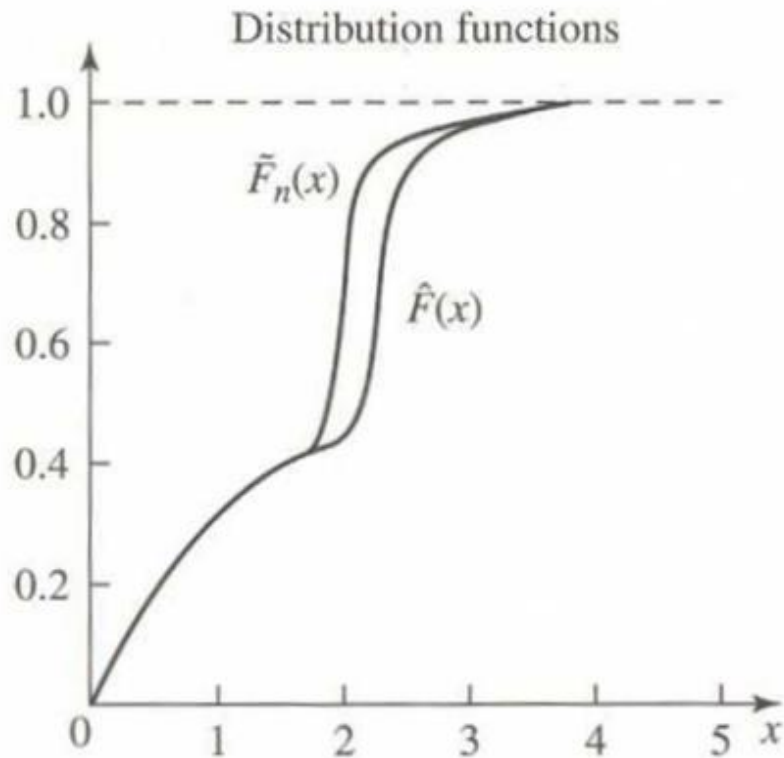
- ❑ Very similar to QQ plot
- ❑ QQ plot is more common, though

- ❑ Difference to QQ plot:
 - QQ plot compares [quantiles of] two distributions:
1% quantile vs. 1% quantile, etc.
 - Graphically: the y axes of the cumulative density distribution functions are plotted against each other
 - PP plot compares probabilities of two distributions
 - Graphically: the y axes of the probability density functions are plotted against each other

- ❑ How to read:
 - Basically the same as QQ plot
 - PP plots highlight differences near the centers of the distributions
(whereas QQ plots highlight differences near the ends of the distributions)



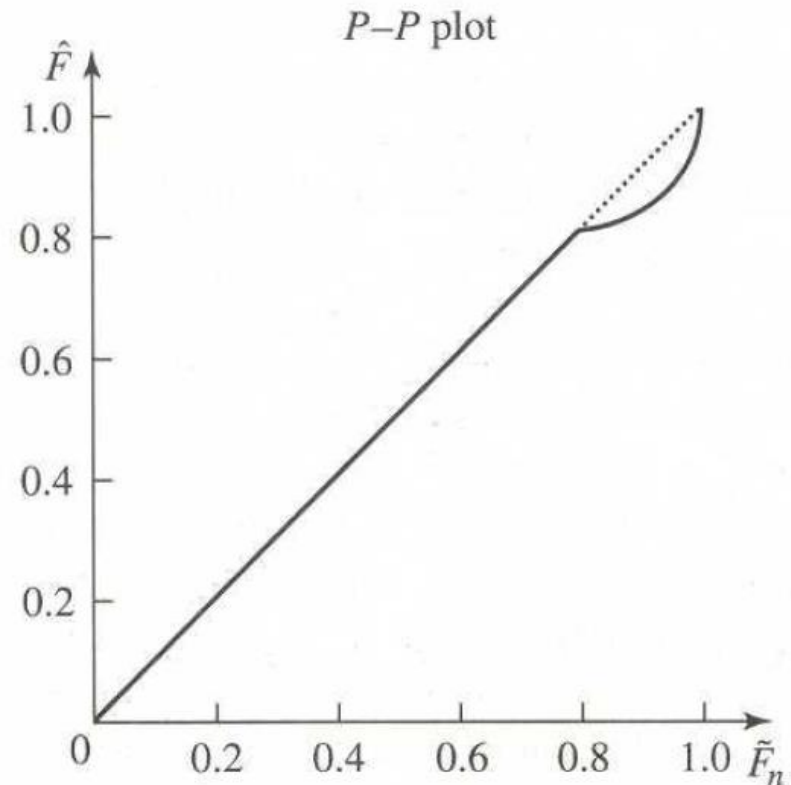
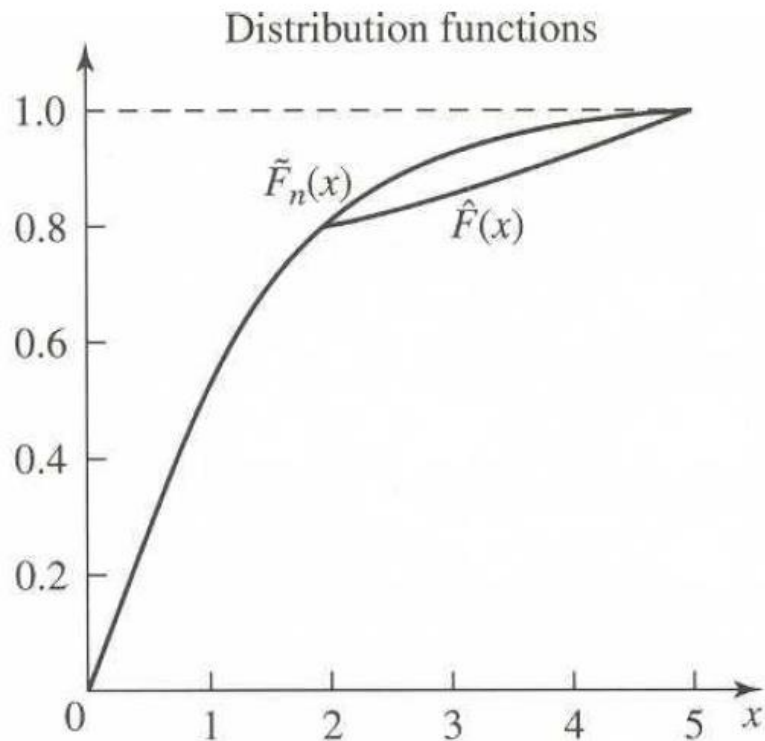
□ PP Plot



The difference between the “middles” of $\hat{F}(x)$ and $\tilde{F}_n(x)$ amplified by the P - P plot.



□ PP Plot



The difference between the right tails of $\hat{F}(x)$ and $\tilde{F}_n(x)$ amplified by the $Q-Q$ plot.



- Monty Hall Problem – (also known as the goat problem)
 - American game show „Let’s make a deal“ adopted in Germany „Geh auf Ganze“



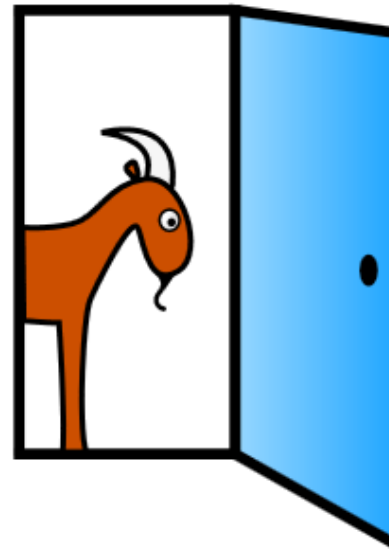


Statistics Fundamentals

- Game rules:
 - Behind one door is a price
 - Behind the other doors is the goat / Zonk (It is assumed that the candidate is not interested in neither the goat nor the Zonk)
 - Candidate may choose one door
 - Game master will open one door after the decision of the candidate and will offer the candidate the choice to choose a different door.



Should the candidate change his/her decision?





- Definition: RV $Z=i$: „Zonk/Goat is behind door i“
 - $P(Z=i)=1/n$ (Laplace)
- Definition: RV $C=i$: „Candidate has chosen door i“
 - $P(C=i)=1/n$ (Laplace)
- Z and C are independent
 - $P(Z=i \wedge C=i)=P(Z=i) \cdot P(C=i)$
- Definition: RV $O=i$: „Door i was opened“
 - $P(O=i|Z=i)=0$ „The winning door was not opened“
 - $P(Z=i \wedge O=i) = P(O=i|Z=i) \cdot P(Z=i) = 0$ (Bayes)
 - $P(Z=i \wedge O \neq i) = P(Z=i) - P(Z=i \wedge O=i) = P(Z=i)$ (Totale Wahrscheinlichkeit)
- $P(O=i|C=i)=0$ „The selected door will NOT be opened“
 - $P(C=i \wedge O=i) = P(O=i|C=i) \cdot P(C=i) = 0$ (Bayes)
 - $P(C=i \wedge O \neq i) = P(C=i) - P(C=i \wedge O=i) = P(C=i)$ (Totale Wahrscheinlichkeit)



- Win probability if the player does not change his selection

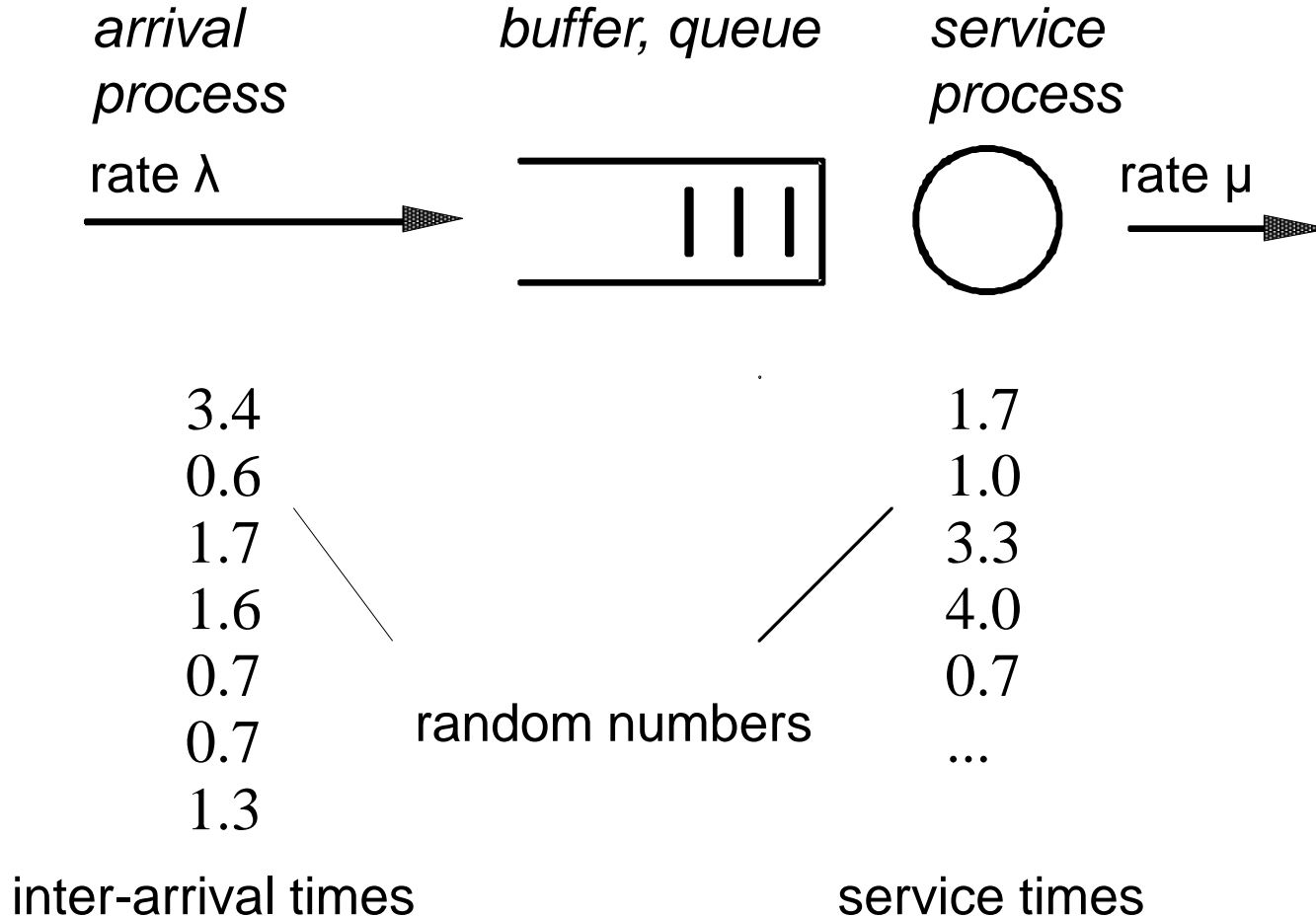
$$\begin{aligned} P(Z = i | C = i \wedge O \neq i) &= \frac{P(Z = i \wedge (C = i \wedge O \neq i))}{P(C = i \wedge O \neq i)} \\ &= \frac{P(Z = i \wedge C = i)}{P(Z = i)} = \frac{\frac{1}{n} \cdot \frac{1}{n}}{\frac{1}{n}} = \frac{1}{n} \end{aligned}$$

- Win probability if player changes his/her selection

$$\begin{aligned} P(Z = i | C \neq i \wedge O \neq i) &= \frac{1 - P(Z = i \wedge (C = i \wedge O \neq i))}{n - 2} \\ &= \frac{1 - \frac{1}{n}}{n - 2} = \frac{n - 1}{n \cdot (n - 2)} \end{aligned}$$



Waiting Queue Theory





What are we talking about... and why?

- Simple queue model:
 - Customers arrive at random times
 - Execution unit serves customers (random duration)
 - Only one customer at a time; others need to queue

- Standard example

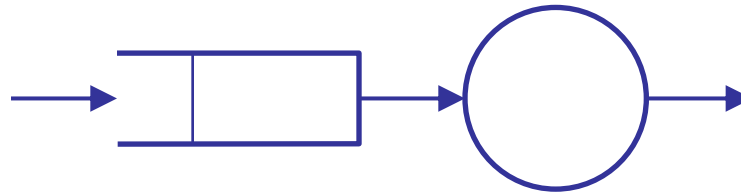
- Give deeper understanding of important aspects, e.g.
 - Random distributions (input)
 - Measurements, time series (output)
 - ...



Queuing model: Input and output

□ Input:

- (Inter-)arrival times of customers (usually random)
- Job durations (usually random)



□ Direct output:

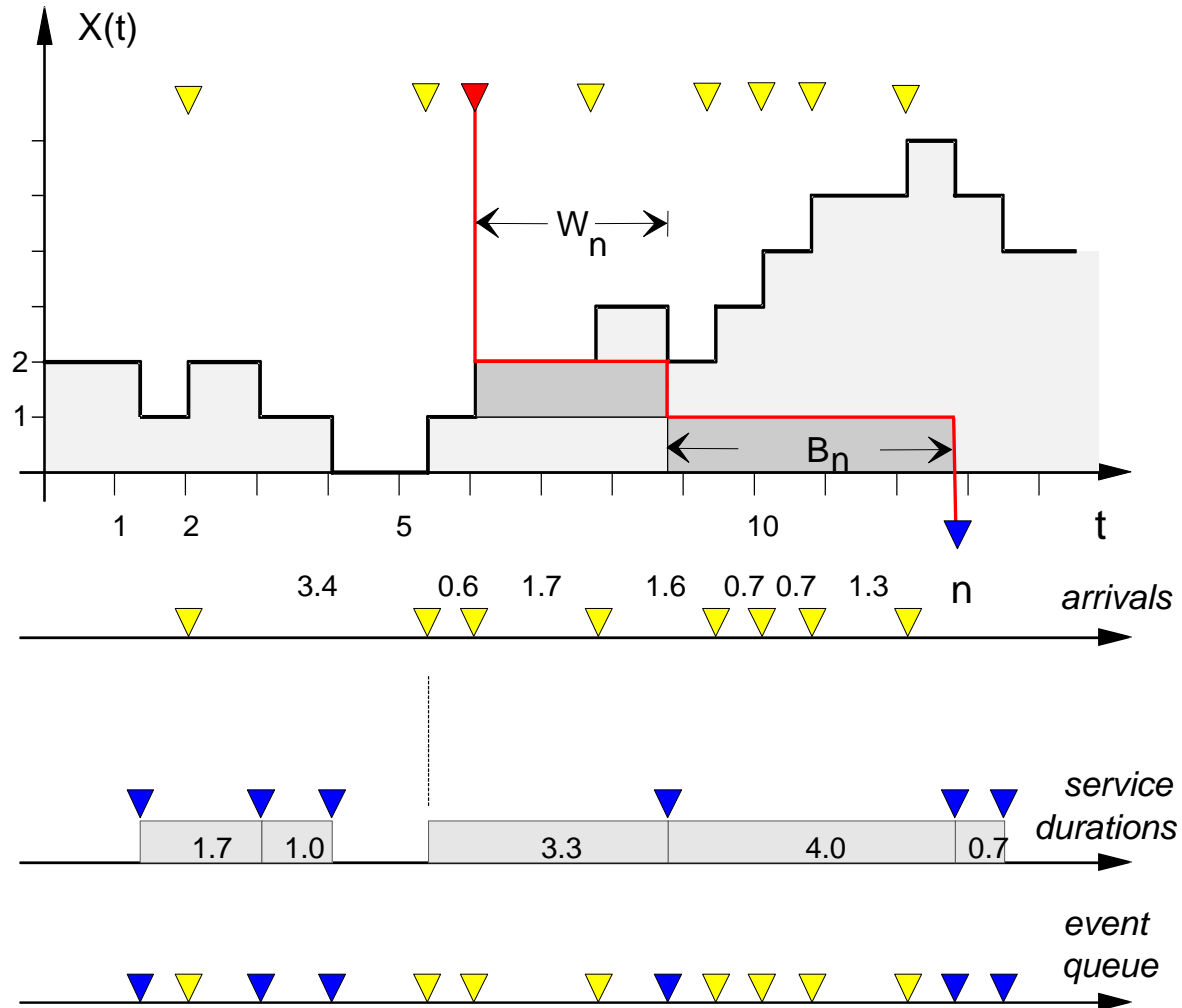
- Departure times of customers

□ Indirect output:

- Inter-arrival times for departure times of customers
- Queue length
- Waiting time in the queue
- Load of service unit (how often idle, how often working)



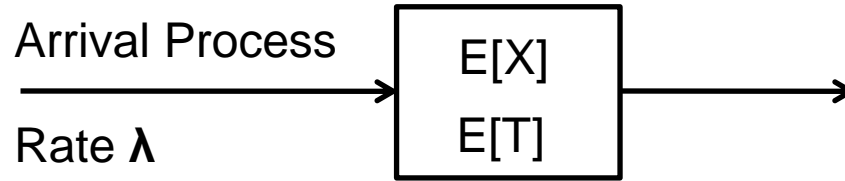
Little Theorem





Little Theorem

- λ : average arrival rate
- $E[X]$: average number of packets in the system
- $E[T]$: average retention time of packets in the system



$$t_o \rightarrow \infty$$

$$\bar{T} = \frac{1}{N} \sum_{i=1}^N T_i \approx \frac{1}{N} \int_0^{t_o} X(t) dt$$

$$\lambda = \lim_{t_o \rightarrow \infty} \bar{\lambda} = \lim_{t_o \rightarrow \infty} \frac{N}{t_o}$$

$$\bar{X} = \frac{1}{t_o} \int_0^{t_o} X(t) dt \implies \bar{X} \approx \frac{N}{t_o} E[T]$$

$$E[T] = \lim_{t_o \rightarrow \infty} \bar{T} = \lim_{t_o \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N T_i$$

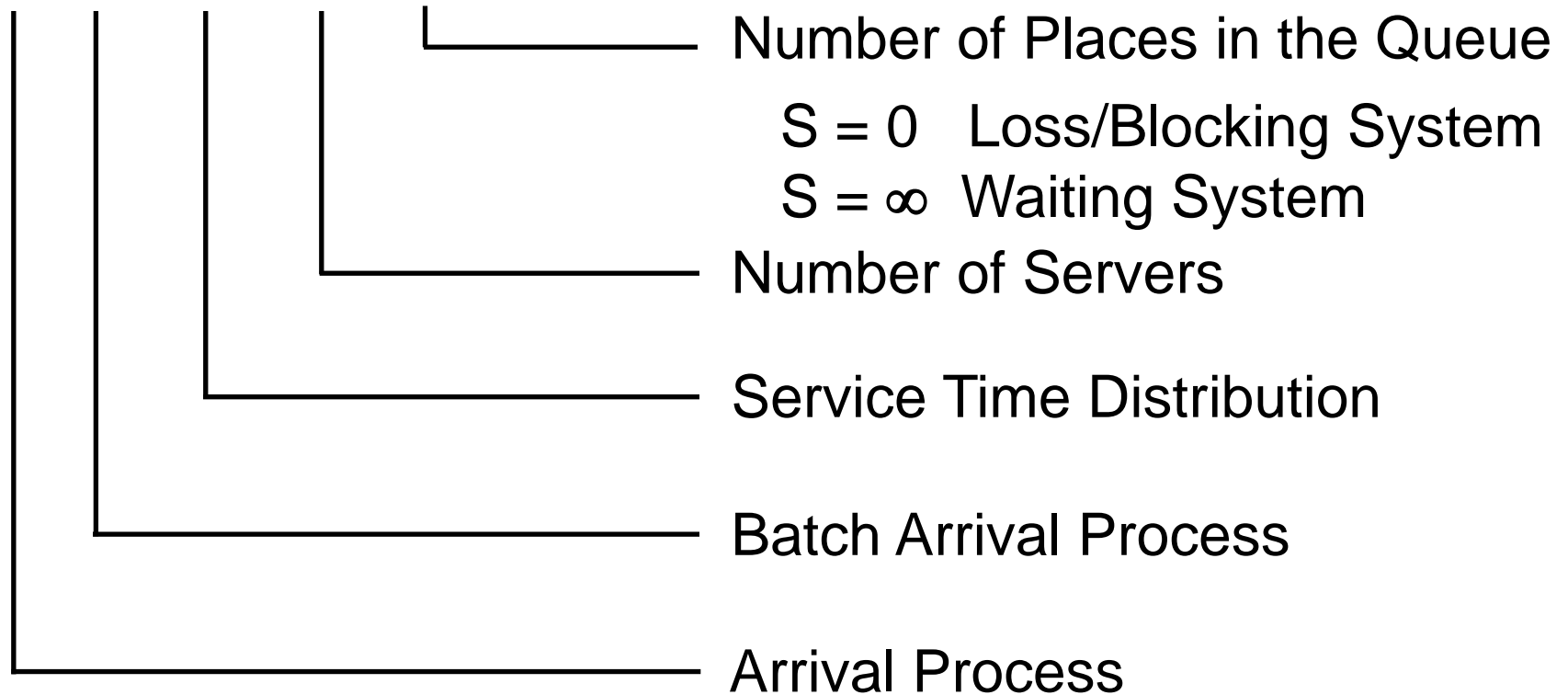
$$\bar{\lambda} \approx \frac{N}{t_o} \implies \bar{\lambda} \cdot \bar{T} \approx \bar{X}$$

$$E[X] = \lim_{t_o \rightarrow \infty} \bar{X} = \lim_{t_o \rightarrow \infty} \frac{1}{t_o} \int_0^{t_o} X(t) dt$$



□ Kendall Notation

$GI^{[x]} / GI / n - S$





□ Queuing Discipline

- FIFO / FCFS First In First Out / First Come First Served
- LIFO / LCFS Last In First Out / Last Come First Served
- SIRO Service In Random
- PNPN Priority-based Service
- EDF Earliest Deadline First

□ Distributions

- M Markovian Exponential Service Time
- D Degenerate Distribution A deterministic service time
- E_k Erlang Distribution Erlang k distribution
- GI General distribution General independent
- H_k Hyper exponential Hyper k distribution



□ System Characteristics

- Average customer waiting time
- Average processing time of a customer
- Average retention time of a customer
- Average number of customers in the queue
- Customer blocking probability
- Utilization of the system / individual processing units





□ Questions:

- How does the number of service units affect the system?
- What impact has a higher variance of the arrival and/or service process on the performance of the system?
- Which system has a higher utilization? One with an unlimited number of waiting slots or one with a limited number?
- Which system has a lower retention time? One with many slow serving units or one with a single but fast serving unit?
- How does the queuing strategy (FIFO, LIFO, EDF) affect the average waiting time and the waiting time distribution?



□ Exercise

- System A: D / D / 1 - ∞
 - Arrival rate $\lambda = 1 / s$
 - Service rate $\mu = [1;10] / s$

- System B: M / M / 1 - ∞
 - Arrival rate $\lambda = 1 / s$
 - Service rate $\mu = [1;10] / s$

- System C: M / M / 20 - ∞
 - Arrival rate $\lambda = 10 / s$
 - Service rate $\mu = 1 / s$

- System D: M / M / 1 - ∞
 - Arrival rate $\lambda = 10 / s$
 - Service rate $\mu = 20 / s$

- What is the maximum (meaningful) utilization of the system?
- Which system performs better?
- What impact does the utilization have on the system?

- Which system performs better?
- Would you prefer a single fast processing unit instead of multiple slow processing units?



□ Exercise

- System E: $M / M / 10 - \infty$
 - Arrival rate $\lambda = 9 / s$
 - Service rate $\mu = 1 / s$

- System F: $M / M / 100 - \infty$
 - Arrival rate $\lambda = 90 / s$
 - Service rate $\mu = 1 / s$

- System G: $M / D / 1 - \infty$
 - Arrival rate $\lambda = 1 / s$
 - Service rate $\mu = 1 / 0.7 / s$

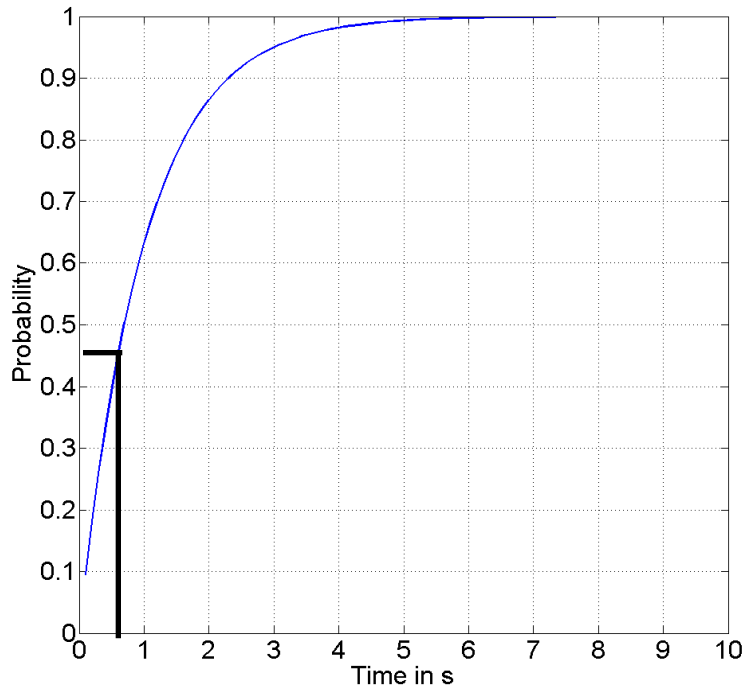
- System H: $D / M / 1 - \infty$
 - Arrival rate $\lambda = 1 / s$
 - Service rate $\mu = 1 / 0.7 / s$

- What is the maximum (meaningful) utilization of the system?
- Which system performs better?

- Which system performs better?
- Which system has a shorter avg waiting time?



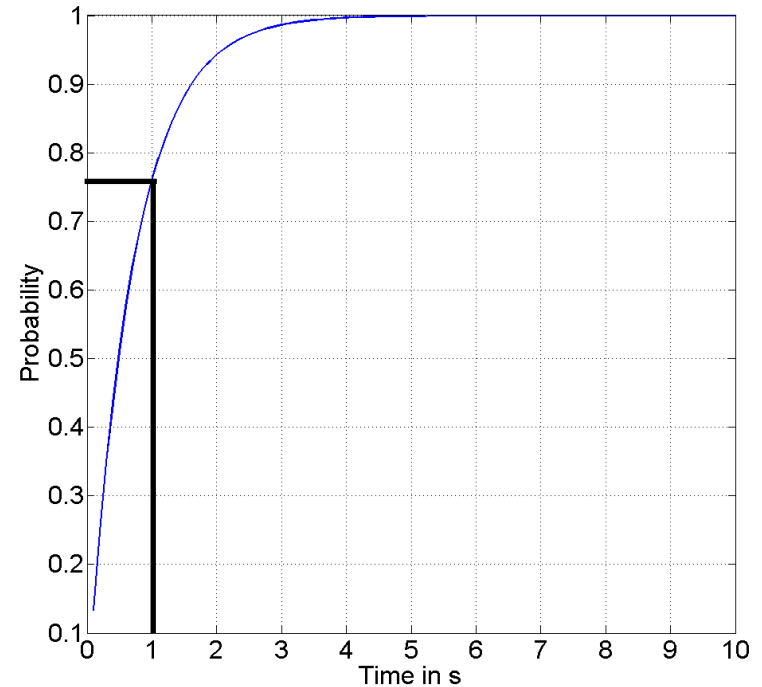
□ System G: M / D / 1 - ∞



$$\lambda = 1$$

$$P(T > 0.7s) = 1 - P(T \leq 0.7s) = 0.55$$

□ System H: D / M / 1 - ∞



$$\mu = 1 / 0.7 = \sim 1.43$$

$$P(T < 1s) = 0.75$$



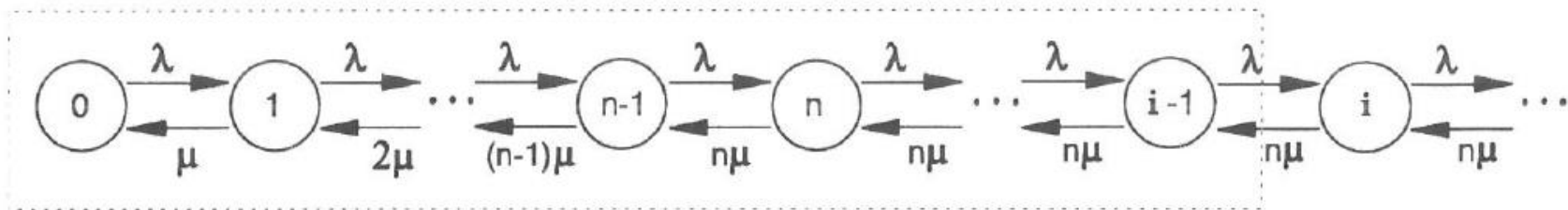
□ M / M / n - ∞

- System of equations

$$\lambda x(i-1) = i\mu x(i), \quad i = 1, 2, 3, \dots, n,$$

$$\lambda x(i-1) = n\mu x(i), \quad i = n+1, \dots$$

$$\sum_{i=0}^{\infty} x(i) = 1$$



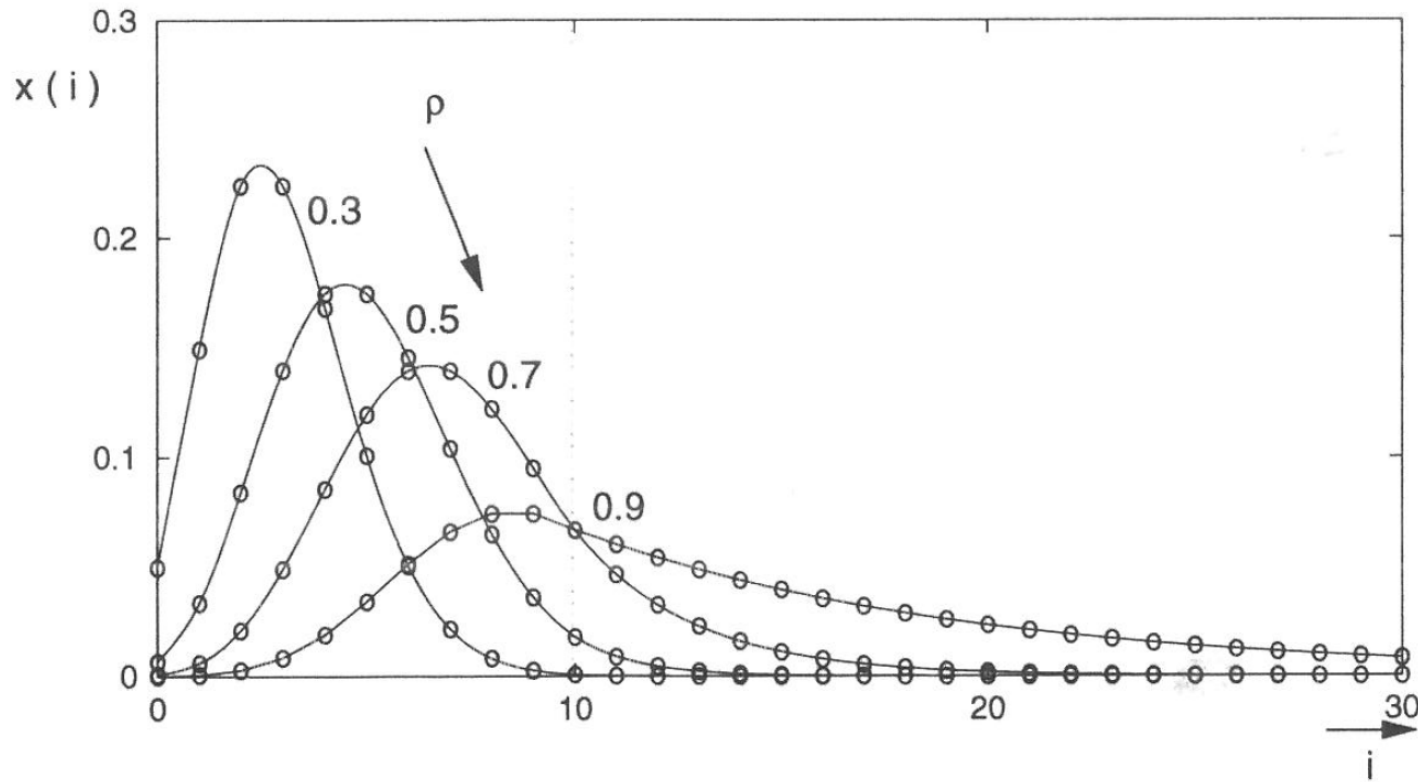
Makrozustand S

State Transition Diagram - M / M / n - ∞

Picture taken from Tran-Gia, Analytische Leistungsbewertung verteilter Systeme, p. 98



□ M / M / 10 - ∞



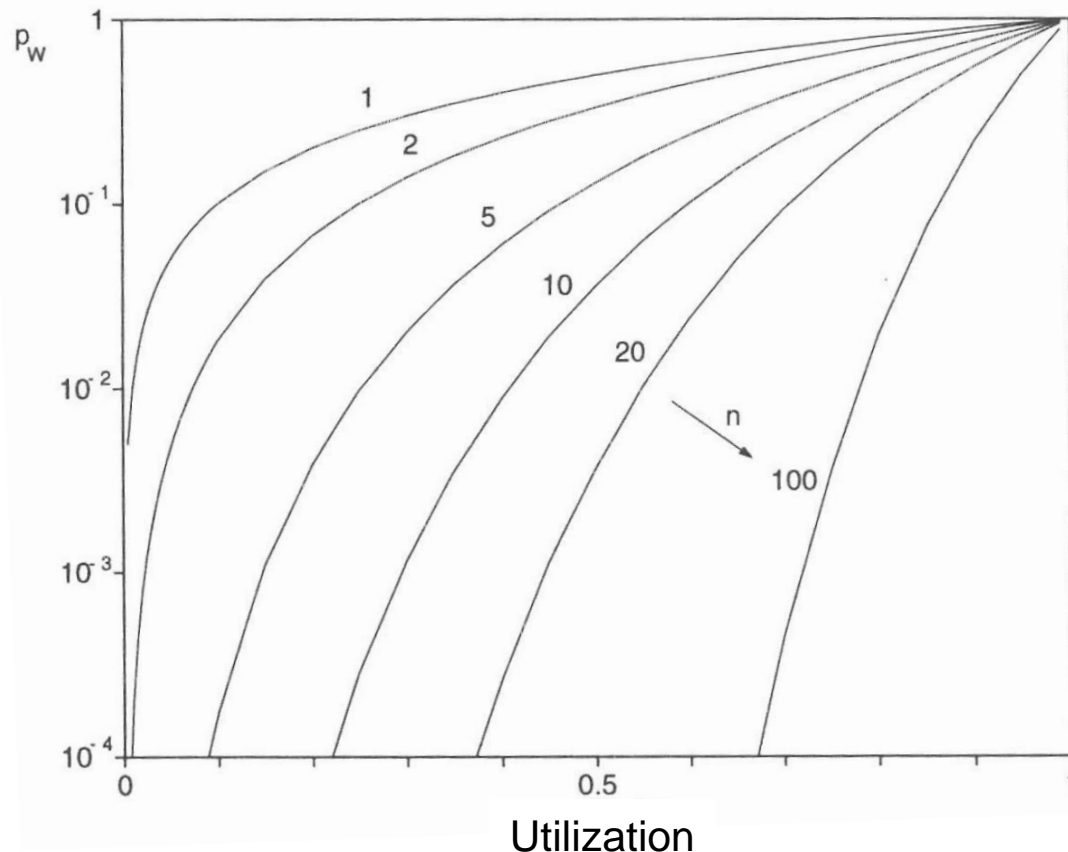
State Distribution - M / M / 10 - ∞

Picture taken from Tran-Gia, Analytische Leistungsbewertung verteilter Systeme, p. 99



□ $M / M / 10 - \infty$

- The waiting probability decreases with an increasing number of processing units (assuming constant utilization)



- n = number of processing units
- ρ_w - Waiting probability

Picture taken from Tran-Gia, Analytische Leistungsbewertung verteilter Systeme, p. 100