Chair for Network Architectures and Services – Prof. Carle
Department of Computer Science
TU München

Network Analysis
Ch 3d) Probability Distributions, Tests, and Experimental Planning
IN2045

Dr. Alexander Klein
Dr. Nils Kammenhuber
Dr. Heiko Niedermayer
Prof. Dr.-Ing. Georg Carle

Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Some of today's slides/figures are borrowed from:
Oliver Rose
Averill Law, David Kelton

Technische Universität München

Probability Distributions

Probability Distributions in Scipy

- Scipy provides a large number of continuous and discrete distributions.

- Creating a distribution
  - RV = scipy.stats.DISTRIBUTION(PARAMETERS)
    - Parameters loc and scale typically define variable parts of distribution
    - In case of normal distribution loc is mean, scale standard deviation.
    - Example
      - rv = scipy.stats.norm(loc=1, scale=2) defines a distribution with mean 1 and stddev 2.

- Generating random numbers
  - The function rvs returns random number stream.
    - e.g. rv.rvs() (1 random number)
rv.rvs(100) (stream with 100 random numbers)

Probability Distributions in Scipy

- Other functions
  - pdf (Probability Density Function)
    - e.g. rv.pdf(-0.3)
  - cdf (Cumulative Distribution Function)
    - e.g. rv.cdf(-0.3)
  - ppf (Percent Point Funtion, Quantile)
    - e.g. rv.ppf(0.99)

- Plotting

```
x=np.linspace(rv.ppf(0.01),rv.ppf(0.99),100)
plt.plot(x,rv.pdf(x))
plt.show()
```

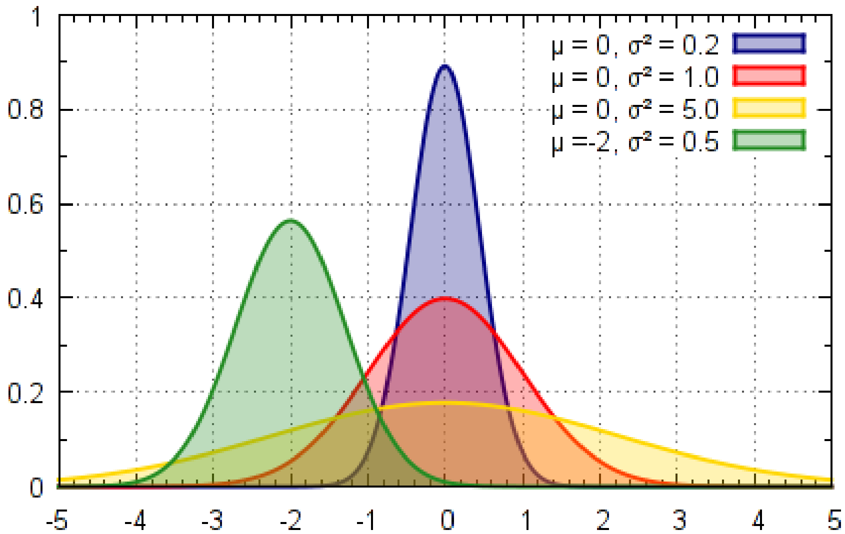
Random numbers - Continuous

- **Uniform distribution**: $RV\ \ X \sim U(a,b)$ (LK 8.3.1)
  - Density function: $f(x) = \frac{1}{b-a}, X \in [a;b]$
  - Range: $[a;b]$
  - Distribution function: $F(x) = \frac{x-a}{b-a}$
  - Expectation: $E(X) = \frac{a+b}{2}$
  - Variance: $VAR(X) = \frac{(b-a)^2}{12}$
  - Generation: $U \sim U(0,1), X = a + (b-a)U$

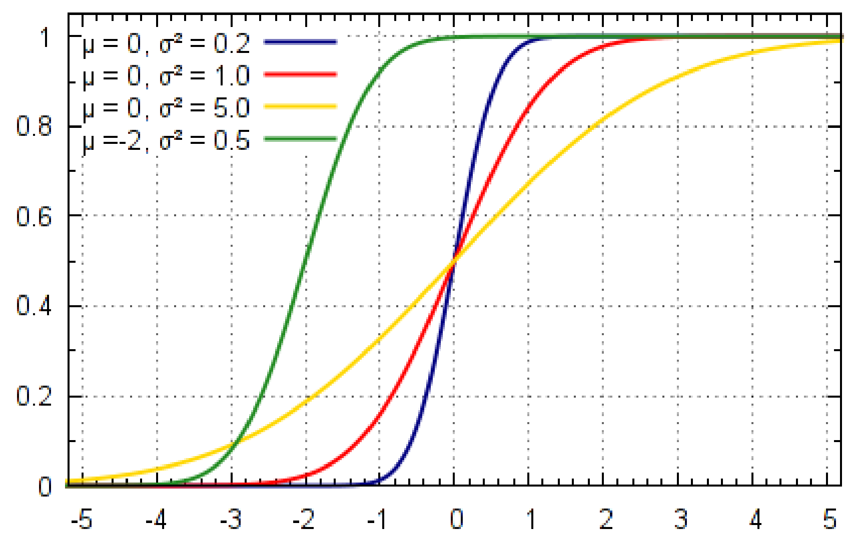
Random numbers - Continuous

- **Normal distribution(1/2):** $RV\ \ X \sim N(\mu, \sigma^2)$ (LK 8.3.6)
  - Density function: $f(x) = \dfrac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\left(\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right)}$
  - Distribution function: $F(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$
  - Range: $\left]-\infty;\infty\right[$
  - Mode: μ
  - Expectation: $E(X) = \mu$
  - Variance: $VAR(X) = \sigma^2$
  - Scalability: $X \sim N(0,1) \Rightarrow (\mu + \sigma X) \sim N(\mu, \sigma^2)$

- **Normal distribution(2/2):** $RV\ \ X \sim N(\mu, \sigma^2)$ (LK 8.3.6)



Probability Density Function



Cumulative Density Function

Lognormal distribution(1/2): $RV \;\; X \sim LN(\mu,\sigma^2)$ (LK 8.3.7)

- Special property of the lognormal distribution

  if $Y \sim N(\mu,\sigma^2) \Rightarrow e^Y \sim LN(\mu,\sigma^2)$

- Range: $[0,\infty)$

- Algorithm: Composition

  – $Y \sim N(\mu,\sigma^2) \Rightarrow X = e^Y$

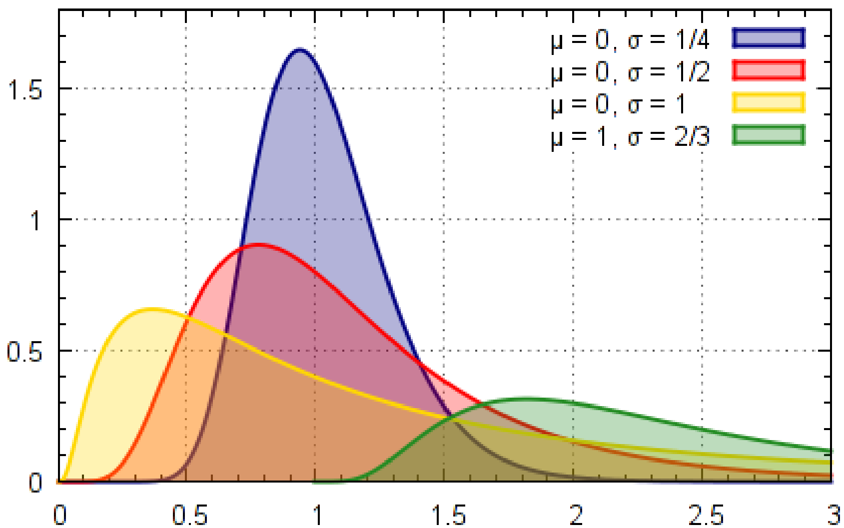
- Expectation: $E(X) = e^{\mu+\frac{\sigma^2}{2}}$

- Variance: $VAR(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

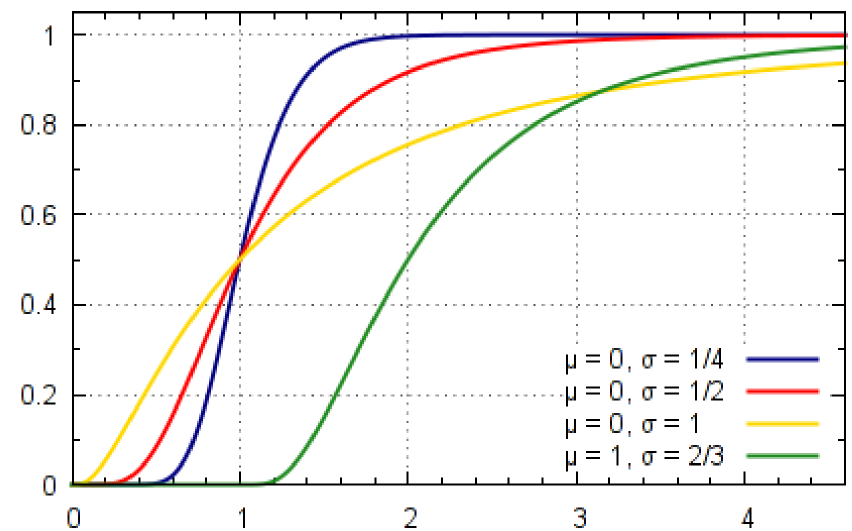
Note that μ and σ are NOT the mean and the variance of the lognormal distribution!

Random numbers

- Lognormal distribution(2/2): $RV\ \ X \sim LN(\mu,\sigma^2)$ (LK 8.3.7)



Probability Density Function



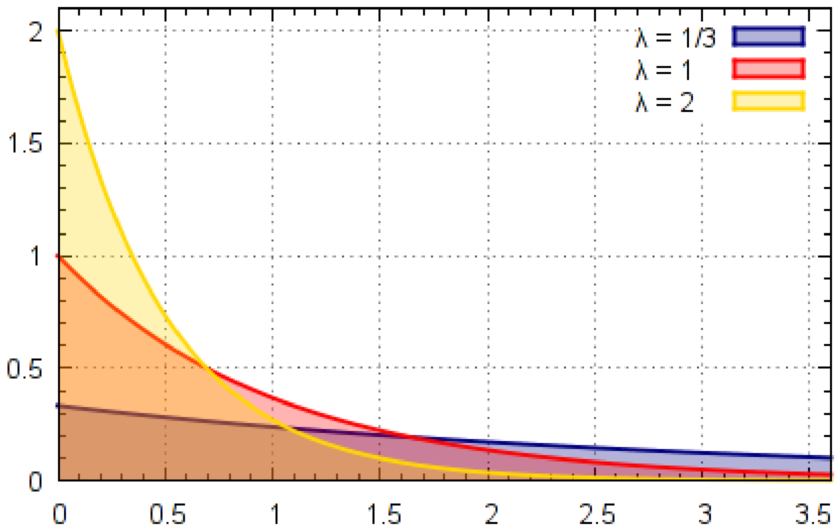
Cumulative Density Function

Random numbers

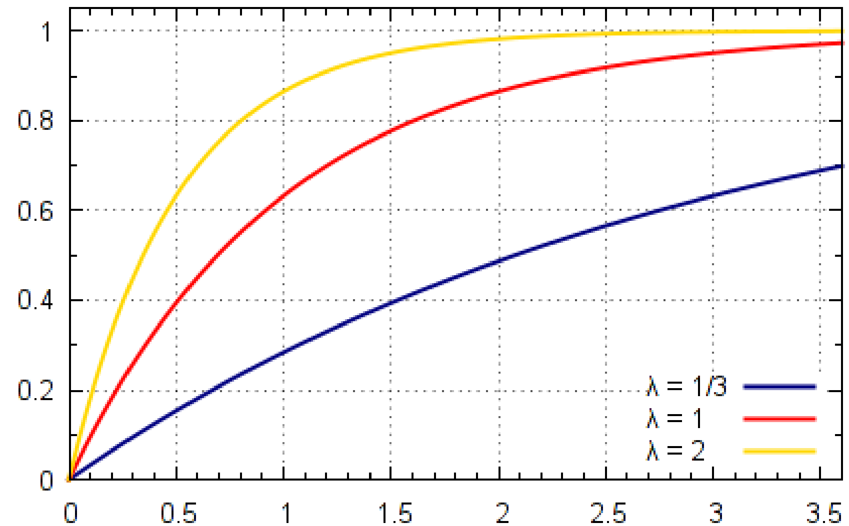
- **Exponential distribution(1/2):** $RV\ \ X \sim \exp(\lambda)$ (LK 8.3.2)
  - Density function: $f(x) = \lambda \cdot e^{-\lambda x} \quad \text{für} \quad x \geq 0$
  - Distribution function: $F(x) = 1 - e^{-\lambda x}$
  - Range: $[0, \infty[$ Mode: 0
  - Expectation: $E(X) = \frac{1}{\lambda}$
  - Variance: $VAR(X) = \frac{1}{\lambda^2}$
  - Coefficient of variation: $c_{Var} = 1$
  - Generation: Inversion $U \sim U(0,1), X = \frac{-\ln(U)}{\lambda}$

Random numbers - Continuous

- **Exponential distribution(2/2):** $RV \;\; X \sim \exp(\lambda)$ (LK 8.3.2)



Probability Density Function



Cumulative Density Function

Pictures taken from Wikipedia

Random numbers - Discrete

- **Uniform (discrete) (1/2)** $RV\ \ X \sim DU(i,j)$ (LK 8.4.2)

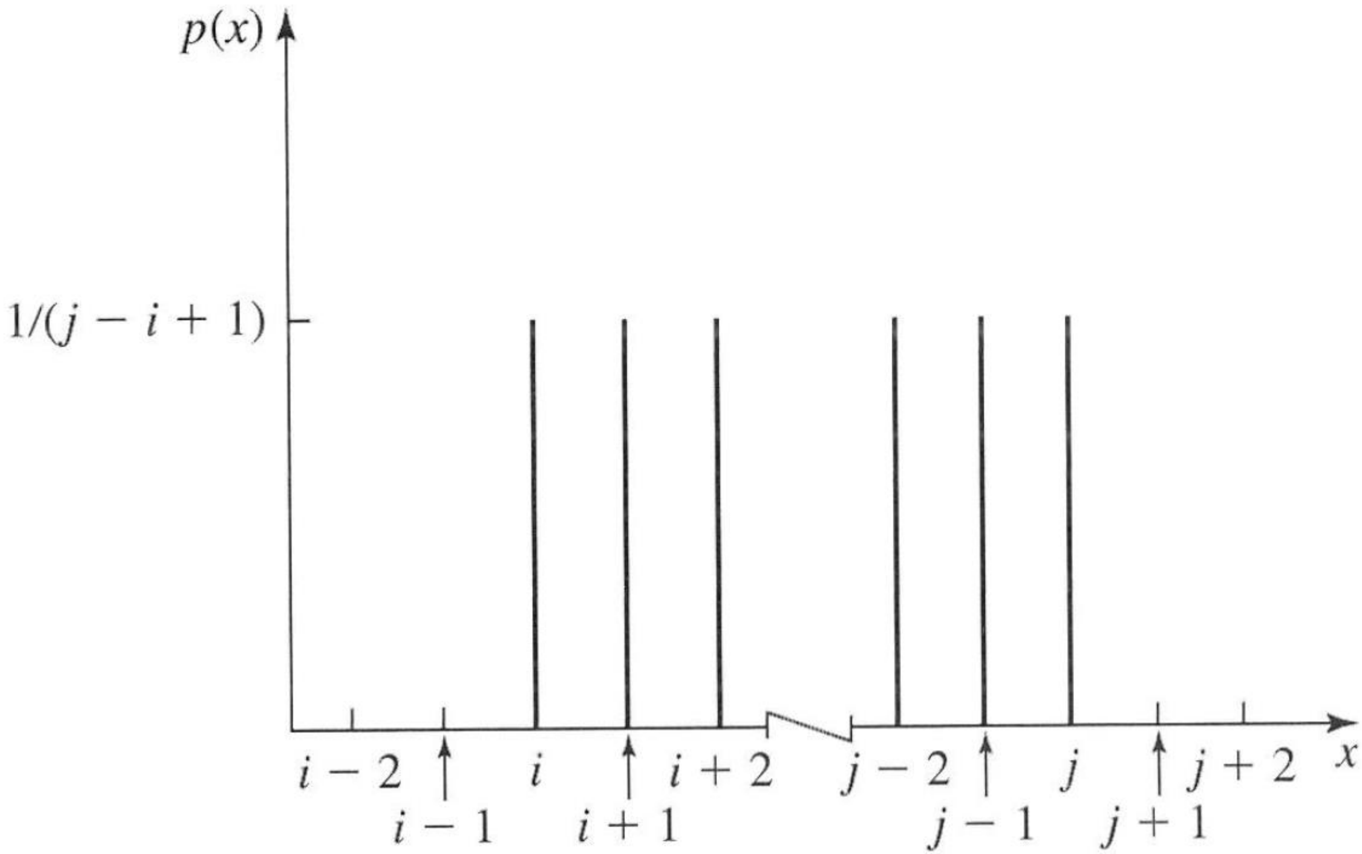
  - Distribution: $$p(k)=\begin{cases}\dfrac{1}{j-i+1} & if\ \ k\in\{i,i+1,i+2,\dots,j\}\\ 0 & Otherwise\end{cases}$$
  - Range: $i \le k \le j$
  - Expectation: $$E(X)=\frac{(i+j)}{2}$$
  - Variance: $$VAR(X)=\frac{(j-i+1)^2-1}{12}$$
  - Generation: Inversion

    $U \sim U(0,1) \qquad X = i + \lfloor (j-i+1)\cdot U \rfloor$

DU(0,1) and Bernoulli(0.5) distributions are the same

Random numbers - Discrete

- **Uniform (discrete) (2/2)** $RV \ \ X \sim DU(i, j)$ (LK 8.4.2)



Distribution

Random numbers - Discrete

- **Bernoulli (1/2)** $RV \;\; X \sim Bernoulli\;(p)$ (LK 8.4.1)
  - Example: Flipping a coin

  - Distribution:

$$p(k) = \begin{cases} 1-p & if \;\; k=0 \\ p & if \;\; k=1 \\ 0 & Otherwise \end{cases}$$

  - Range: $i \leq k \leq j$
  - Expectation: $E(X) = p$
  - Variance: $VAR(X) = p \cdot (1-p)$
  - Coefficient of variation: $c_{Var} = \sqrt{\frac{1-p}{n \cdot p}}$

Random numbers - Discrete

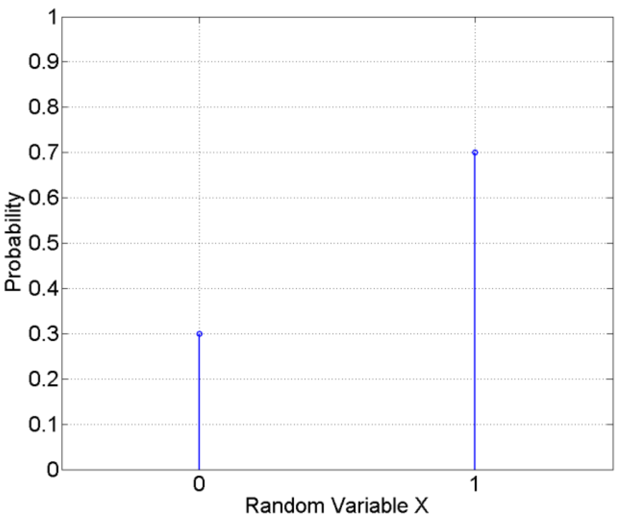
- **Bernoulli (2/2)** $RV\ X \sim Bernoulli\ (p)$ (LK 8.4.1)

  - Mode: 0 or 1 (depends on the definition of the outcome)
  - Generation: Inversion $U \sim U(0,1)$

$$X = \begin{cases} 0 \text{ if } U < p \\ 1 \text{ Otherwise} \end{cases}$$

  - Distribution

    $Bernoulli\ (0.3)$



Random numbers - Discrete

- **N-Bernoulli (1/2)** $RV\ \ X \sim Bernoulli\ (n, p)$ (LK 8.4.4)

  - Example: Flipping a coin n times

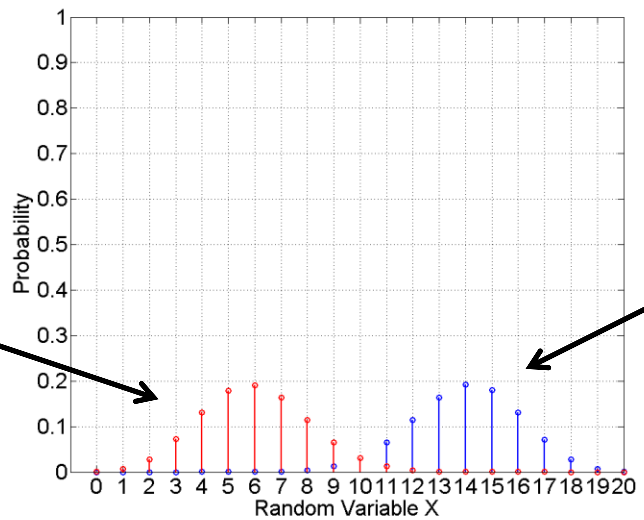
  - Distribution: $p(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad 0 \le k \le n$
  - Range: $0 \le k \le n$
  - Expectation: $E(X) = np$
  - Variance: $VAR(X) = n \cdot p \cdot (1-p)$
  - Coefficient of variation: $c_{Var} = \sqrt{\dfrac{1-p}{n \cdot p}}$

Random numbers - Discrete

- **N-Bernoulli (2/2)** $RV \;\; X \sim Bernoulli \;\; (n, p)$ (LK 8.4.4)
  - Mode: 0 or 1 (depends on the definition of the outcome)
  - Generation: Composition

$$Bernoulli \;\; (n, p) \approx \sum_{0 \le i < n} Bernoulli \;\; (p)$$

  - Distribution



$Bernoulli \;\; (20,0.3)$

$Bernoulli \;\; (20,0.7)$

Random numbers - Discrete

- **Geom (1/2)** $RV\ \ X \sim Geom\,(p)$ (LK 8.4.5)

  - Example: Number of unsuccessful Bernoulli – Experiments until a successful outcome (e.g. number of retransmissions)
  - Distribution: $p(x) = p \cdot (1-p)^{x}$
  - Distribution function: $F(x) = 1-(1-p)^{\lfloor x \rfloor +1}$
  - Expectation: $E(X) = \dfrac{1-p}{p}$
  - Variance: $VAR(X) = \dfrac{1-p}{p^{2}}$
  - Coefficient of variation: $c_{Var} = \sqrt{\dfrac{1}{1-p}}$

Random numbers - Discrete

- **Geom (2/2)** $RV\ \ X \sim Geom\,(p)$ (LK 8.4.5)

  - Mode: 0

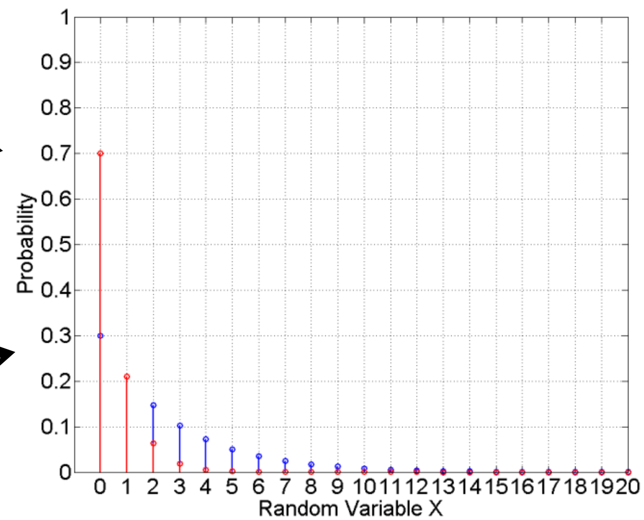
  - Generation: Inversion $U \sim U(0,1)$

$$X = \left\lfloor \frac{\ln(U)}{\ln(1-p)} \right\rfloor$$

  - Distribution

$Geom\,(0.7)$ →

$Geom\,(0.3)$ →



$$p(0) = p$$

Random numbers - Discrete

- **Poisson(1/3)** $RV \;\; X \sim Poisson\;(\lambda)$ (LK 6.2.4)

  - Example: Number of events that occur in an interval of time when the events are occurring at a constant rate (number of items in a batch of random size)

  - Distribution: $$p(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \qquad if \;\; x \in \{0,1,2,\ldots\}$$

  - Distribution function: $$F(x) = \begin{cases} e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & if \;\; x \geq 0 \\ 0 & if \;\; x < 0 \end{cases}$$

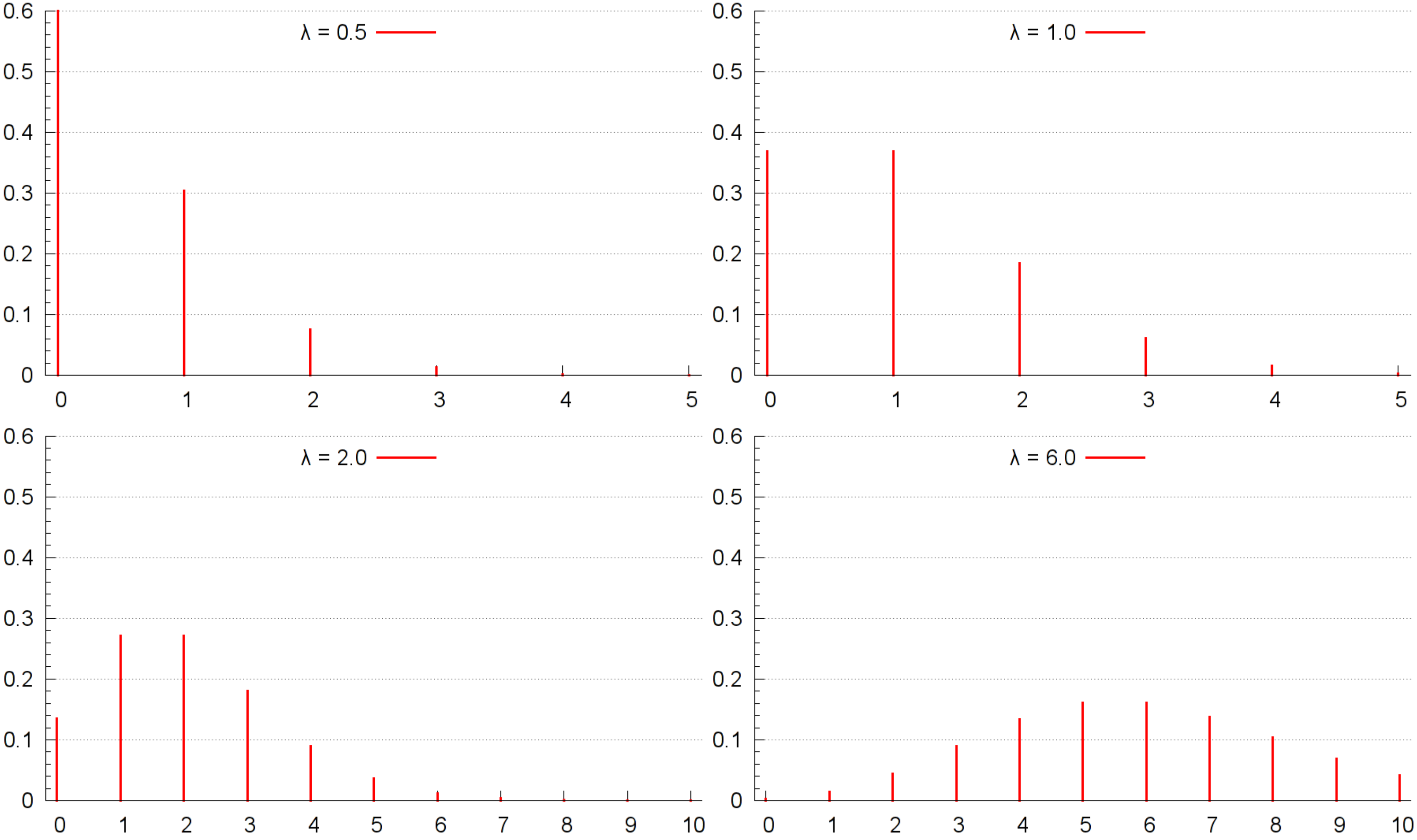
  - Parameter: $\lambda > 0$

Random numbers - Discrete

- **Poisson(2/3)** $RV \;\; X \sim Poisson\;(\lambda)$ (LK 6.2.4)
  - Range: $\{0,1,2,3,\ldots\}$
  - Expectation: $E(X) = \lambda$
  - Variance: $VAR(X) = \lambda$
  - Coefficient of variation: $c_{Var} = \frac{1}{\sqrt{\lambda}}$
  - Mode $\begin{cases} \lambda \cap \lambda - 1 & \lambda \text{ is an integer} \\ \lfloor \lambda \rfloor & \text{otherwise} \end{cases}$
  - Special characteristics:
    - $x = 0$ ⇒ exponential distribution (time interval between two consecutive events)
    - Number of events until a certain point in time is Poisson distributed
    - Period of time until n events have occurred is Erlang distributed

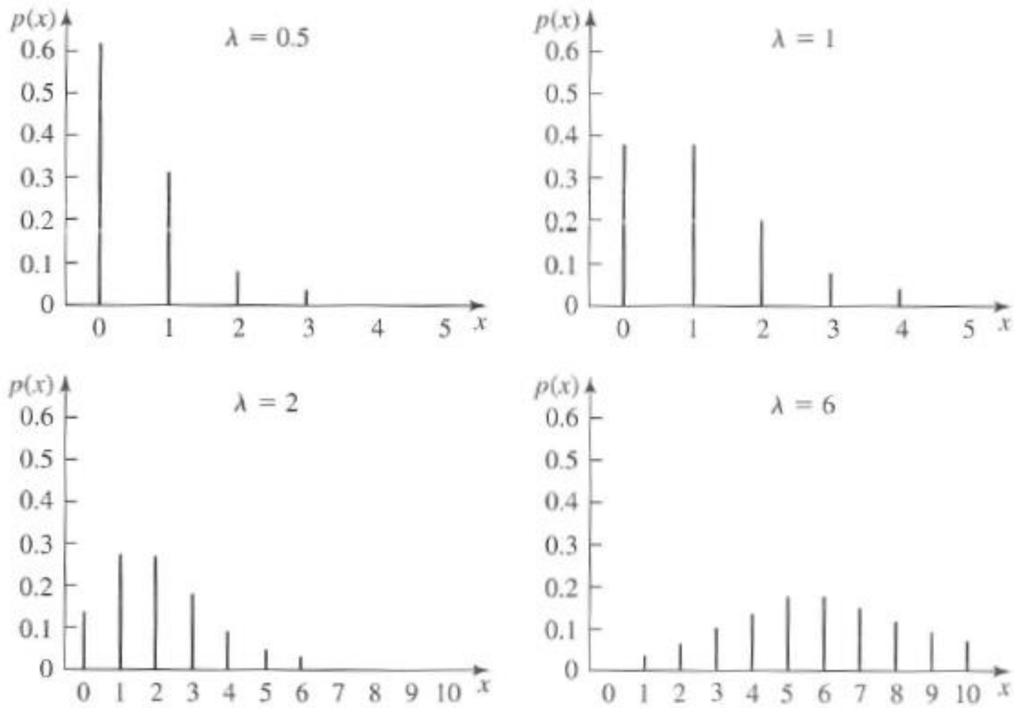
Random numbers - Discrete

- **Poisson(3/3)** $RV\ \ X \sim Poisson\ (\lambda)$ (LK 6.2.4)



Random numbers - Discrete

- **Poisson(3/3)** $RV \;\; X \sim Poisson\ (\lambda)$ (LK 6.2.4)



Picture taken from LK, p.309

Statistical Tests

Statistical tests

- Scenario: Given a set of measurements, we want to check if they conform to a distribution; say: U(0,1)
- Graphs are nice indicators, but not really tangible: “How straight is that line?” etc.
- We want clearer things: Numbers or yes/no decisions
- Statistical tests can do the trick, but…
  - Warning #1: Tests only can tell if measurements do not fit a particular distribution—i.e., no “yes, it fits” proof!
  - Warning #2: The result is never absolutely certain, there is always an error margin.
  - Warning #3: Usually, the input must be ‘iid’:
    - Independent
    - Identically distributed
  - ⇒You never get a ‘proof’, not even with an error margin!

χ^2 test (Pearson, 1900)

- Input:
  - Series of n measurements $X_1 \dots X_n$
  - A distribution function f (the 'theoretical function')
- Measurements will be tested against the distribution
  - ~formal comparison of a histogram with the density function of the theoretical function
- Null hypothesis H0:
  The X_i are IID random variables with distribution function f

χ^2 test: How it works

- Divide [0…1] into k equal-size intervals
- Count how many X_i fall into which interval (histogram):
  N_j := number of X_i in j-th interval $[a_{j-1} \ldots a_j[$
- Calculate how many X_i would fall into the j-th interval if they were sampled from the theoretical distribution:

$$p_j := \int_{a_{j-1}}^{a_j} f(x)dx$$

(*f*: density of theor. dist.)

- Calculate squared normalised difference between the observed and the expected:

$$\chi^2 := \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$$

- Obviously, if χ^2 is "too large", the differences are too large, and we must reject the null hypothesis
- But what is "too large"?

χ^2 test: Using the χ^2 distribution

- The χ^2 distribution
  - A test distribution
  - Parameter: degrees of freedom (short df)
  - χ^2(k–1 df) = Γ(½(k–1) , 2) (gamma distribution)
  - Mathematically: The sum of n independent squared normal distributions
- Compare the calculated χ^2 against the χ^2 distribution
  - If we use k intervals, then χ^2 is distributed corresponding to the χ^2 distribution with k–1 df
  - Let $\chi^2_{k-1,1-\alpha}$ be the (1–α) quantile of the distribution
  - α is called the confidence level
  - Reject H0 if $\chi^2 > \chi^2_{k-1,1-\alpha}$ (i.e., the X_i do not follow the theoretical distribution function)

χ^2 test and degrees of freedom

- χ^2 test can be used to test against *any* distribution
- Easy in our case: We know the parameters of the theoretical distribution f —it's U(0,1)
- Different in the general case:
  - For example, we may know it's N(μ, σ) (normal distribution) but we know neither μ nor σ
  - Fitting a distribution: Find parameters for f that make f fit the measurements X_i best
  - Topic of a later lecture
- Theoretically:
  Have to estimate m parameters
  ⇒ Also have to take $\chi^2_{k-m-1,1-\alpha}$ into account
- Practically:
  m≤2 and large k
  ⇒ Don't care…

χ^2: which parameters?

- How many intervals (k)?
  - A difficult problem for the general case
  - Warning: A smaller or a greater k may change the outcome of the test!
  - As a general rule, use k>100
  - As a general rule, make the intervals equal-sized
  - As another general rule, make sure that $\forall j$: $np_j \geq 5$ (i.e., have enough samples that we expect to have at least 5 samples in each interval)
- ⇒ As a general rule, you need a lot of measurements!
- What confidence level?
  - At most α=0.10 (almost too much);
    typical values: 0.001, 0.01, 0.05 [, and 0.10]
  - The smaller, the better confidence in the test result

Alternatives to test

- Kolmogorov–Smirnov test (KS test)
  - Another very popular test
  - Advantages:
    - No grouping into intervals required
    - Valid for any sample size, not only for large n
    - More powerful than χ^2 for a number of distributions
  - Disadvantages:
    - Applicability more limited than χ^2
    - Difficult to apply to discrete data
    - If distribution needs to be fitted (unknown parameters), then K-S works only for a number of distributions
- Anderson–Darling test (A–D test)
  - Higher power than K-S for some distributions
- …a lot of other tests
  - Rule of thumb: The less more specialised the test, the higher its power compared to other tests – but the less generally applicable

Statistical Tests

- So far, we've seen the χ2 distribution fitting test and the Kolmogorov-Smirnov test (KS)
- Both test if a given set of measurements is consistent with a theoretical distribution
  - Note the wording: „Consistent with“, but not „comes from“
- There are many, many other statistical tests for many, many other applications

Statistical Tests = Hypothesis Tests

- We would like to „prove“ some statement, based on statistical calculations
  Examples:
  - Measurements x_i are consistent with a normal distribution
  - The mean of the measurements xi is greater than 5
- Call this statement our 'work hypothesis' or 'alternative hypothesis' (Arbeitshypothese) H_A
- Formulate the contrary: null hypothesis H_0
- H_A and H_0 need to be:
  - Exclusive: Either H_A is true or H_0 is true
  - Exhaustive: All possible results will satisfy one of the two

Test Statistic

- Hope to find statistical evidence that H_0 is highly improbable
- Mathematically:
  - Input data = x_i (...rather arbitrary label)
  - Calculate a so-called test statistic: $TS(x_i)$
  - Usually: If test statistic is above some threshold, then refuse H_0
  - Test statistic depends on specific test
  - Threshold depends on specific test and on desired accuracy

Test Accuracy: Error Types

- As mentioned before: No test can give a 100% guarantee – we're talking about statistics here, and statistics always deals with the unknown
- Differentiate between two types of errors:

	Test rejects H_0	Test accepts H_0
In reality, H_0 is false	Correct decision	Type II error, β error, false negative
In reality, H_0 is true	Type I error, α error, false positive	Correct decision (albeit not the one that we wanted in most cases…)

Error types explained by example (1/2)

- Suppose you have developed a medical drug. Development has cost an enormous amount of money. Now you want to test if the drug is harmful to your patients
- Type I error (α error)
  - Probability that people get harmed
  - Can cost lives: Invest a lot of effort to avoid it.
- Type II error (β error)
  - Probability that you reject a drug that is actually perfectly safe
  - Can waste money: Unpleasant, but more acceptable.

Error types explained by example (2/2)

- Suppose you have developed a new network protocol. By applying a statistical test to the output of some network simulations, you hope to show that the protocol increases network performance (=H_A).
- Type I error (α error)
  - Probability that you claim that the protocol is great, whereas it is actually rubbish
  - If you do not specify your α error, or if it is too large (i.e., your confidence level is too low), then nobody will believe your results!
    - But also beware that you can achieve any confidence level given a study on the basis of non-representative scenarios with enough sample values!
- Type II error (β error)
  - Probability that you wrongly assume that your great protocol does not help anything
  - Presumably interesting to you, but the reader of your paper does not care about the risk that you might have failed detecting the performance increase: Obviously, you did not fail, since otherwise the paper would not have been written…

Balancing error types

- Problem:
  - Reducing one error increases the other and vice versa. Damn.
  - Only solution to reduce both: Increase the sample size. Usually a superlinear factor (e.g., to reduce one error by 1/2 while keeping the other constant, we must increase sample size by 4)
- In the majority of the cases, keeping the *α* error low is more important
  - α = 5% has been accepted for years (although there has been some criticism), 1% is better, 0.1% is extremely good
  - β = 10% or 20% is usually acceptable; but usually, it is not calculated
  - Do not choose α too small if there are only few samples: Small sample size and small α both will increase β to unacceptable values – then you would almost always accept the null hypothesis and thus (wrongly) reject your work hypothesis

Error types: summary

- Usually, Type-1 errors (α errors) are the more serious ones
- In order to minimise one type of error (e.g., Type 1 error), you only have the choice between…:
  - Increasing the Type 2 error
  - Increasing the sample size
  - Picking a different statistical test that has better error properties

An „Alternative“: Significance Tests

P-value (R. A. Fisher): How likely is the result to happen?

- Test statistic is a dependent random variable that follows a specific distribution (test distribution, e.g., Student's t distribution or χ^2 distribution) if the null hypothesis holds
- Using the theoretical distribution, calculate the probability that our measurements attain our given values or even more extreme values if the null hypothesis holds:
  - This is defined as the p value
  - Note that the p value itself is uniformly distributed in [0...1] if the null hypothesis holds, and it is near 0 if it does not hold.
- Refuse H0 if this seems unlikely: i.e., refuse if $p \leq \alpha$
- In other words: Our threshold for the test statistic is the point where its distribution „has no meat“, i.e., the p value gets too low

We have two types of tests?

- In theory, distinguish:
  - Hypothesis test that we just explained:
    Fix an α, calculate the test statistic and accept or reject the null hypothesis
  - Fisher's probability test:
    For the given data, calculate the p value for the null hypothesis, and decide how likely the null hypothesis is
- In practice, combine both!
  - p value is more expressive
  - Fixed α is more commonly known/accepted; often allows better comparisons to other studies

How to combine both types of a test?

- With modern statistical programs, this is possible – in most cases, it is even done automatically!
- Good practice:
  - Tell the reader your p value (especially if null hypothesis sounds quite likely!)
  - Traditionally, the p value is judged with star symbols within braces:
    - [***] means: $P \leq 0.1\%$
    - [**] means: $0.1\% < P \leq 1\%$
    - [*] means: $1\% < P \leq 5\%$
- If possible, calculate the p value and derive statements about α

  e.g.: „The null hypothesis could be refused at a confidence level of α=0.5, but not at a confidence level of α=0.1"

Experimental Planning

Comparing two alternative systems

- Comparison of two systems:
  Is there a difference in value for a given response variable?

  e.g., difference in achieved network throughput

- Test criterion:
  1. Calculate difference between the two response variables
  2. This difference is statistically significant if its confidence interval (CI) does not contain 0

  e.g.: CI (throughputTCP Reno – throughputTCP Vegas) $\not\ni$ 0

  → We can assume that the difference in throughput which the two congestion control algorithms TCP Reno and TCP Vegas achieve is statistically significant

Is this enough?

- Good: Very simple
- Bad: Quite restricted applicability
  - Only should be applied if the response has the same variance for the two levels – not often the case
    - Better: Modified or Welch two-sided t confidence intervals
  - Calculating the confidence interval for the response differences only can tell us if two levels of one factor make a difference
  - What if we want to analyse more than two levels for a given factor?
    - E.g., TCP Reno vs. TCP Vegas vs. TCP Cubic: 3 levels
  - What if we have more than one factor?
    - E.g., TCP congestion control algorithm, TCP window size, network delay, link bandwidth: 4 factors

Other issues with respect to testing and studies

- Publication Bias
  - Only positive examples are publised.
  - Given 1 positive example, 19 negative, having this is related to the chance to meet a ***p-value*** of 5 percent.
  - Consequence
    - Decline effect: Effect of treat or network protocol decreases over repetitions and for larger subsequent studies.

Why compare system alternatives?

- Goals:
  - Better understanding of system
  - Better control of system
  - Better performance of system
  - ***Make a decision!***

- Methods:
  - Try out in different simulated environments
    - Try out different workloads with different characteristics
    - Try out different network topologies
  - Try out with different system parameters

Linear model and regression

- Have n samples $x_{1...n}$ and $y_{1...n}$ of two random variables x and y
- y is 'not really' a random variable:
  it's also dependent on x
- Linear model: $y = a \cdot x + b + e$
  - a: slope
  - b: intercept
  - e: error
- Idea: Chose a and b such that e is minimised
  - Calculate sum of squared errors:

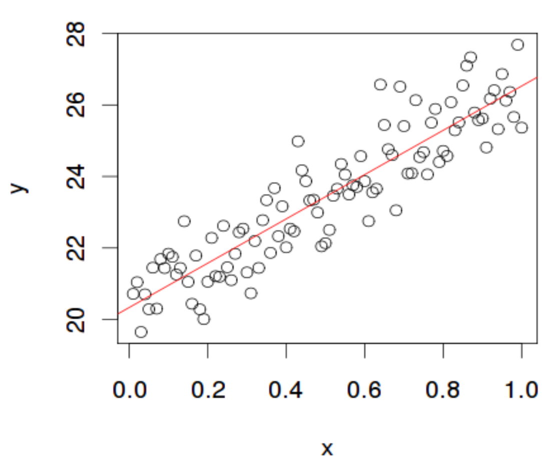
⇨ Minimise Sum of Squared Errors (SSE)

Calculating a and b

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - mean(x))(y_i - mean(y))}{\frac{1}{n}\sum_{i=1}^{n}(x_i - mean(x))^2} = \frac{Cov(x,y)}{Var(x)}$$

- N.B.: different, but equivalent formulae in literature (you can omit dividing by n-1 in var and cov)
- Usually built into statistical programs

- Graphical interpretation:
  Fit a straight line that goes through the points in the (x,y) scatterplot
  - b: intercept (Achsenabschnitt)
  - a: slope (Steigung)



Are we actually allowed to apply regression?

- Check correlation coefficient for linearity.

- Warning:
  - The residuals e (as in y = a·x + b + e) must be normally distributed!
  - Exploit the central limit theorem: Calculate averages of multiple independent simulation runs with the same factor level
  - Check that it looks normal: QQ plots or some normality test

Regression and experiment planning

- In our nomenclature: y = response, x = factor level
- Regression can tell us how much the factor influences the response. Answers questions like:
  - Does it make sense to explore further factor levels in a given direction?
  - Does it make sense to check factor levels in between?
- Good:
  - We now can have multiple factor levels
- Bad:
  - We still have only one factor
  - It must be linearly proportional
  - The residuals must be normally distributed
    (but that constraint won't go away with ANOVA either)

Nonlinear Regression 1/2

- Often, the relationship between x and y is not linear
- Solution: Try to find a suitable transformation
  - Let y be the simulation outcome (response)
  - Then apply the model $y^* = a \cdot x + b + e$
    where $y^* = f(y)$
  - Transformation function f can be, for example:
    - Logarithm
    - Exponential
    - Square root
    - Square
    - Some other polynomial (usually quadratic or cubic)
    - Logistic function (logistic regression)
    - Inverse (1/x)
    - …

Nonlinear Regression 2/2

- Which transformation function is the right one?
  - Careful consideration of the system: You have to think!
  - Check if the y* are normally distributed – the y are probably not normally distributed in this case
- QQ plots can help
- Admittedly, a matter of experience
- Warning:
  - Overfitting, arbitrary curve fitting: “Just try around with some transformations and pick the one that matches best” – no, try to avoid that!
  - A correlation can be coincidence
  - Correlation does not imply causation
  - Example: Decreasing number of pirates leads to increasing global temperatures (Church of the Flying Spaghetti Monster)
  - Again: First think about the system, then postulate a meaningful transformation

ANOVA

- Short for ‘ANalysis Of VAriances’
  - Historical term
  - Explained in next slides
- Be careful: “variance analysis” is a more general term!
  Often, that term describes a slightly different analysis:
  - Calculate variances of the responses for different levels of one (or several) factors
  - Analyse statistically if the variances are the same
  - Very similar to ANOVA, but slightly different!

ANOVA Terminology

- **factor**: input variable (e.g., TCP window size), condition, structural assumption (e.g., TCP congestion control algorithm)
- **level:** one factor value that is used in our experiments
- **response:** system parameter of interest that depends on given set of factors (e.g., achieved TCP throughput)
- **run**: evaluation of response for a given set of factor values
  - i.e., the analysed simulation result
  - There will (should!) be multiple runs

Remember:

- In simulation experiments, responses vary for runs of the same factor values due to random effects.
- In experiments, the same is true due to system variation (other users, etc.).
- Therefore: several runs / measurements have to be performed!

ANOVA Nomenclature

- Factor has a levels (‘treatments’ for historical reasons: ANOVA was developed in pharmaceutical research)
- Each level is replicated/observed *n* times
- Data:

level	*replication*		
	1	$\cdots$	*n*
1	y_{11}	$\cdots$	y_{1n}
$\vdots$	$\vdots$		$\vdots$
a	y_{a1}	$\cdots$	y_{an}

- Question we want to answer:
  - Is there an effect of factor levels on system responses?
  - If so: how much?

ANOVA and experiment planning

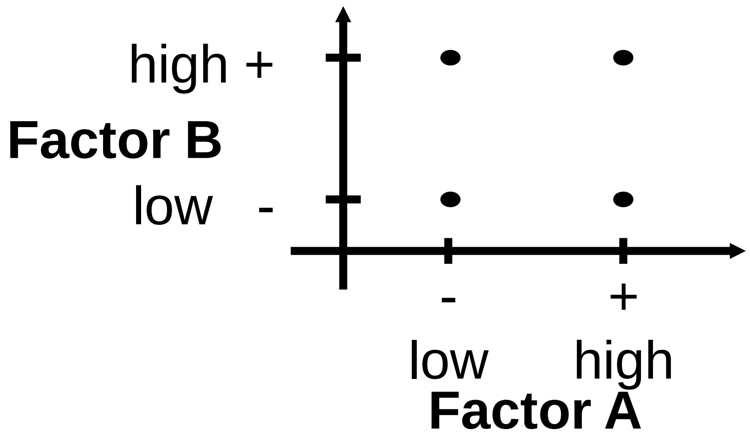
- Usually many factors
  - Example: TCP window size, TCP congestion control algorithm, network bandwidth, network delay,
    packet loss rate
- Which factor combinations should we try out? – ANOVA can give answers to these questions:
  - Which factors are interesting factors (i.e., have much influence), so we should try out more levels for them?
  - Which factors have interesting interactions, so we should try out more factor level combinations for them?
  - Which factors, which interactions can be left out?
- Structuring the experiments like this is called factorial design
  - Of course, not limited to simulation experiments

- Warning:
  - **It is not sufficient to vary one parameter at a time!**
  - **Parameters may interact (see next slides)**

2k factorial designs

- Example: 2 factors, i.e., a 2^2 design

- 4 design points:



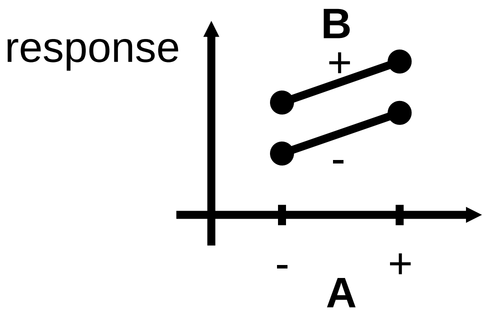
- Design matrix:

Run	Factor A	Factor B	Response
1	-	-	r_1
2	+	-	r_2
3	-	+	r_3
4	+	+	r_4

2k factorial designs

- Interaction of factors A and B: Is there a difference in the changes of the response if A is changed while B is kept either on level ‘+’ or ‘–’?

- no interaction, i.e.
  no (or small) difference in changes:



- interaction, difference in changes: