Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

Universiteit Utrecht

[Faculty of Science
Information and Computing Sciences]

Algorithms for Decision Support

Simulation: Input analysis

- Until now:
  - Modelling
  - Simulation study
  - Validation

- Today we are going to look at stochastic variables

In this lecture:

- You learn about modelling uncertain input data, mostly by probability distributions

Stochastic input variables model:

- Variation
- Things that are uncertain from the viewpoint of the system

Examples stochastic input variables

- Production line
- Transportation planning at DHL
- Communication network
- Sensor (e.g. in Electronic Road Pricing)
- Military

Stochastic variables occur in simulation at different places:

1. Input data are modeled as stochastic variables
   - E.g time until arrival of next customer
2. Generate random variables
   - When you schedule a new Arrival event you have to generate a random number for the time delay
3. Analysis of results

Basics

- Experiment: process with uncertain outcome/result
- Stochastic variable represents the outcome of experiment
- Stochastic variable X
  - Discrete
  - Continuous

Discrete stochastic variable X

- Possible values $x_1, x_2, \ldots, x_n$

$$p_i = P(X = x_i)$$

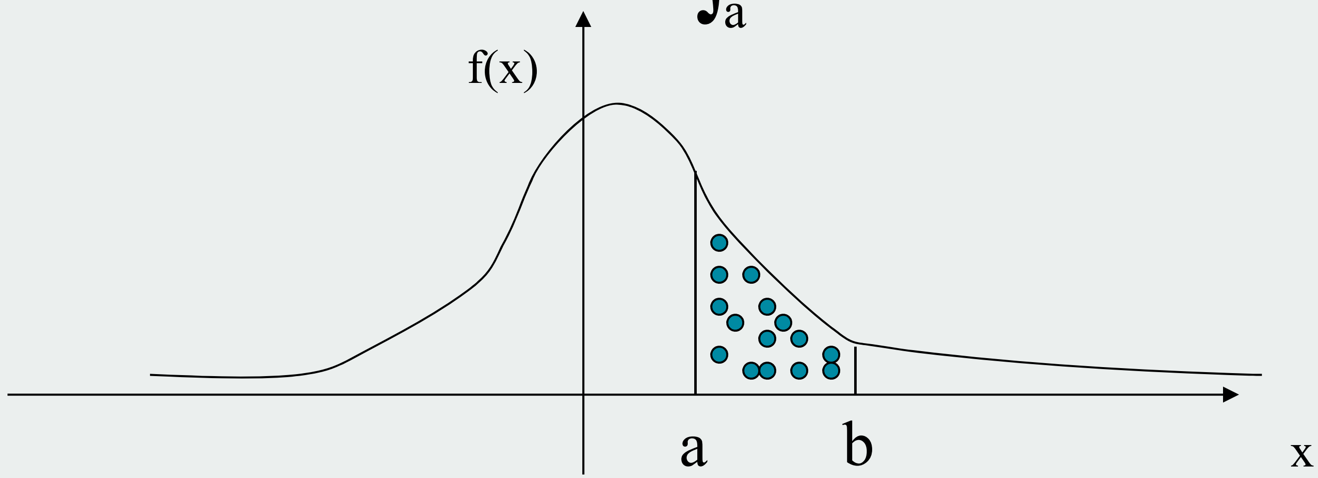
$$0 \le p_i \le 1, \sum_{i=1}^{n} p_i = 1$$

- Example:
  - Die
  - Flip a coin 4 times: X is the number of heads

Continuous stochastic variable X

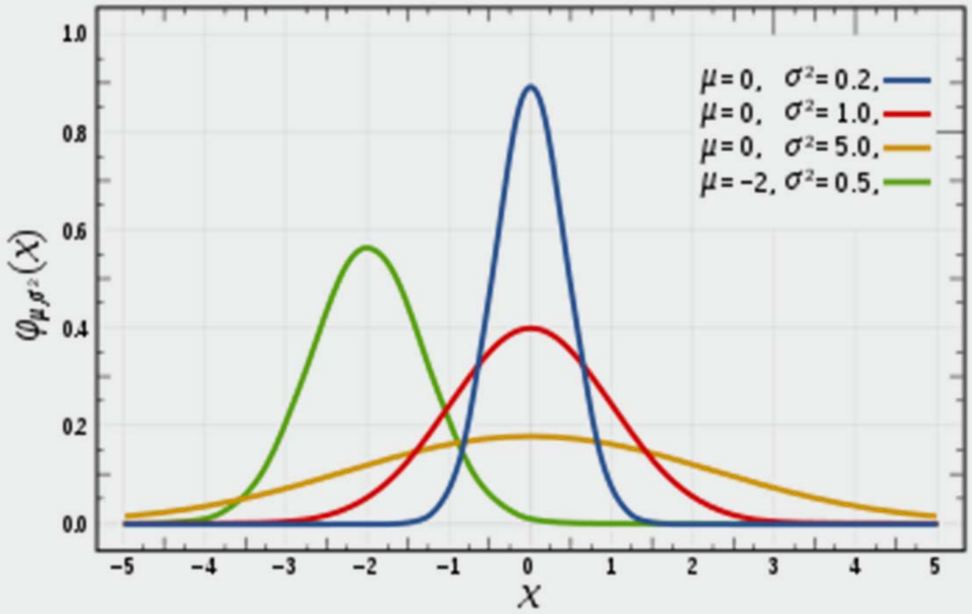
- Can take any value in an interval
- **Probability Density Function *f* (kansdichtheid)**
- Total surface under graph equals 1

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



Normal N(μ,σ²)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

Continuous stochastic variable X (2)

- **Cumulative Distribution Function *F* (verdelingsfunctie):**

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy$$

- **p-th percentile x_p:**

$$x_p \text{ such that :}$$

$$F(x_p) = P(X \le x_p) = \int_{-\infty}^{x} f(y)dy = p$$

Expected value (average) E(X)

- Let *X* and *Y* be stochastic variables

$$\mu = E(X) = \sum_i p_i x_i$$

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$E(cX) = cE(X)$$

$$E(X+Y) = E(X) + E(Y)$$

Variance, standard deviation

- Let X be a stochastic variable
- Variance

$$\sigma_X^2 = var(X) = E\left(\left(X - E(X)\right)^2\right) = E(X^2) - \left(E(X)\right)^2$$

- Standard deviation

$$\sigma_X = \sqrt{var(X)}$$

- Computation, let a and b be real numbers

$$var(aX + b) = a^2 var(X)$$

Exercise:
Variance standard die?
Variance die, 2,3,3,4,4,5?

Variance, covariance

- Let X and Y be stochastic variables

$$var(X+Y) = var(X) + var(Y) + 2\,cov(X,Y)$$

Variances cannot just be added

where

$$cov(X,Y) = E((X - E(X))(Y - E(Y)))$$

is the covariance of X and Y.

- Correlation

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_x\,\sigma_Y}$$

Independence

Two stochastic variables X and Y are independent if:

- Continuous

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B) \quad \forall \text{sets } A, B$$

- Discrete

$$P(X = x \text{ and } Y = y) = P(X = x)P(Y = y) \quad \forall x, y$$

Example:

One deck of cards, take 2 cards without putting back

X=#aces, Y=#kings, Dependent or independent?

Independence (2)

If X and Y independent stochastic variables:

$$E(XY) = E(X)E(Y)$$

$$\operatorname{cov}(X,Y) = E(X - E(X))(Y - E(Y)) = 0$$

$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$$

Input of simulation

1. Direct use of data: trace driven simulation
2. Empirical distribution
3. Theoretical probability distribution (see Law and Kelton tables 6.3 and 6.4 for an overview)

Theoretical distribution

- Given data $X_1,X_2,\ldots,X_n$ for a certain input entity of the system (e.g. interarrival times of customers)
- What probability distribution should we use to model the input entity?
- We assume the values X_i are Independent and Identically Distributed, so independent samples from the same distribution.

Uniform(a,b)

$$f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & \mathrm{x > b} \end{cases}$$

$$E(X) = \frac{a+b}{2}$$

$$\mathrm{var}(X) = \frac{(b-a)^2}{12}$$

Exponential(ß)

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad (x \geq 0)$$

$$F(x) = 1 - e^{-\frac{x}{\beta}} \quad (x \geq 0)$$

$$E(x) = \beta$$

$$\mathrm{var}(X) = \beta^2$$

Sometimes denoted with parameter $\lambda = {}^{1}\!/_{\beta}$

λ is the rate.

- Exp(ß)~gamma(1, ß)~weibull(1, ß)
- Memory-less

Exponential distribution is memory-less

- Suppose X follows an exponential distribution with $\mathrm{E(X)} = \beta$
- Probability that X is more than *t:*

$$P(X > t) = 1 - F(t) = 1 - \left(1 - e^{-\frac{t}{\beta}}\right) = e^{-\frac{t}{\beta}}$$

- Probability that X is more than s+*t* (so at least *t* larger than s) given that we know that it is at least *s*

$$P(X > s + t | X > s) = \frac{P(X>s+t \; and \; X>s)}{P(X>s)} = \frac{e^{-\frac{s+t}{\beta}}}{e^{-\frac{s}{\beta}}} = e^{-\frac{t}{\beta}}$$

- These are equal, so memory-less

Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

At the exam

- *You have to know the formulas of the uniform and exponential distribution by heart*

- *For the next distributions you have to know properties*
- *Formulas will be given if you need them.*

Gamma(α,β)

$$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} (x > 0)$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt (z > 0)$$

$\Gamma(z+1) = z\Gamma(z)$

$\Gamma(k+1) = k!, k$ positive integer

$\mathrm{E(X)} = \alpha\beta$

$\mathrm{var(X)} = \alpha\beta^2$

if $X_1 \sim gamma(\alpha_1, \beta), X_2 \sim gamma(\alpha_2, \beta)$

then $X_1 + X_2 \sim gamma(\alpha_1 + \alpha_2, \beta)$,

α shape parameter, β scale parameter

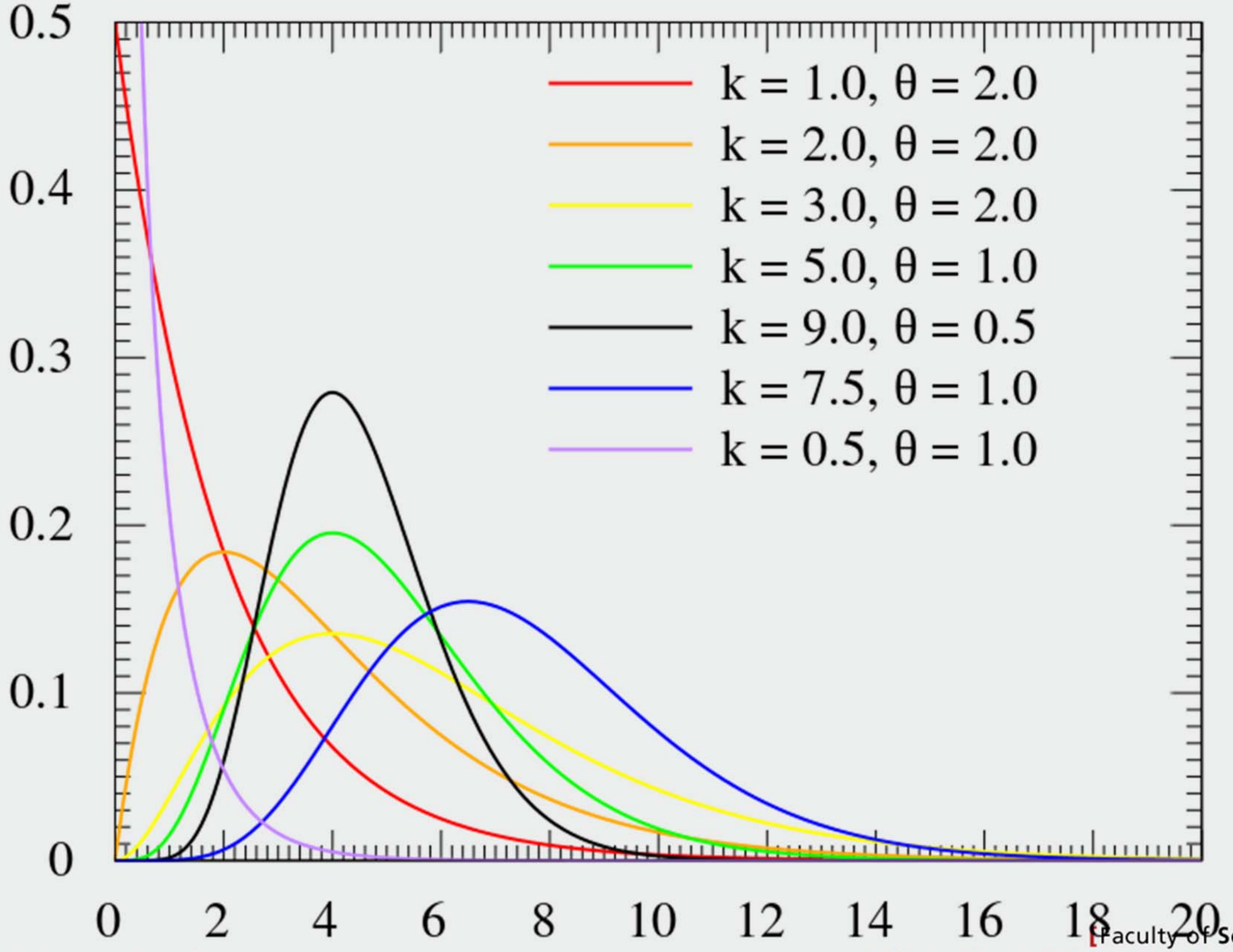
$\mathrm{gamma}(1, \beta) = \exp(\beta)$

$\mathrm{gamma}(k, \beta) = \text{k - Erlang}(\beta)$

$\mathrm{gamma}(k/2, 2) = \chi_k^2$

Exercise:
Variance Exp(p)?
Variance Gamma(k,p/k)?

Gamma(α (=k),β(=θ))



0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
k = 1.0, θ = 2.0
k = 2.0, θ = 2.0
k = 3.0, θ = 2.0
k = 5.0, θ = 1.0
k = 9.0, θ = 0.5
k = 7.5, θ = 1.0
k = 0.5, θ = 1.0

Normal N(μ,σ²)

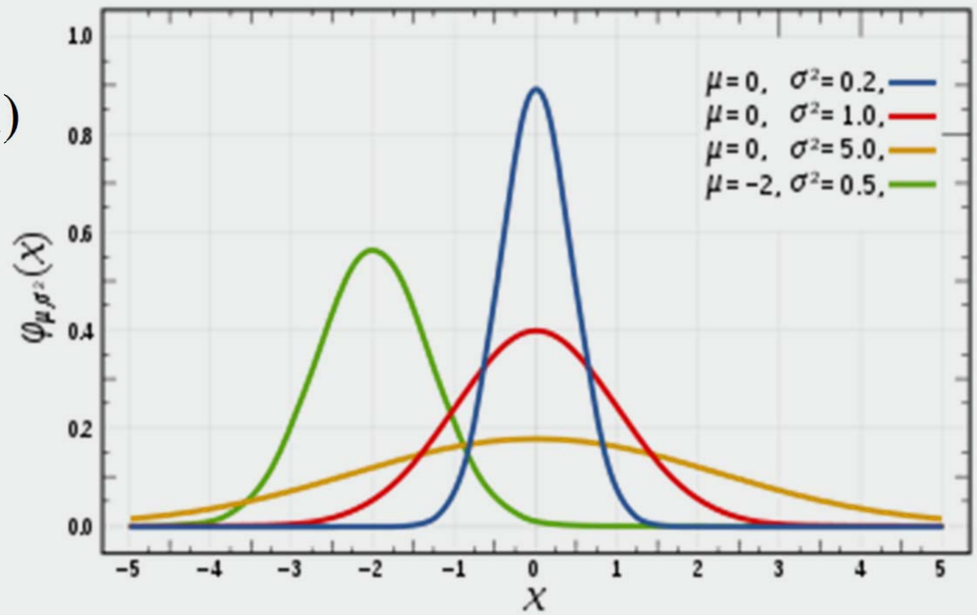
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E(X) = \mu$$

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

$N(0,1)$ standard normal

$$\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2) \Leftrightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$



Normal $N(\mu,\sigma^2)$

Exercise:

- The amount that a coffee machine puts in a cup is normally distributed with average $\mu = 170$ ml and standard deviation $\sigma = 4$. Coffee cups are 175 ml. What is (approximately) the probability of overflow?
- What should μ be such that the probability of overflow is 2%?

See statistical tables on course website.

LogNormal LN(μ,σ²)

$$f(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

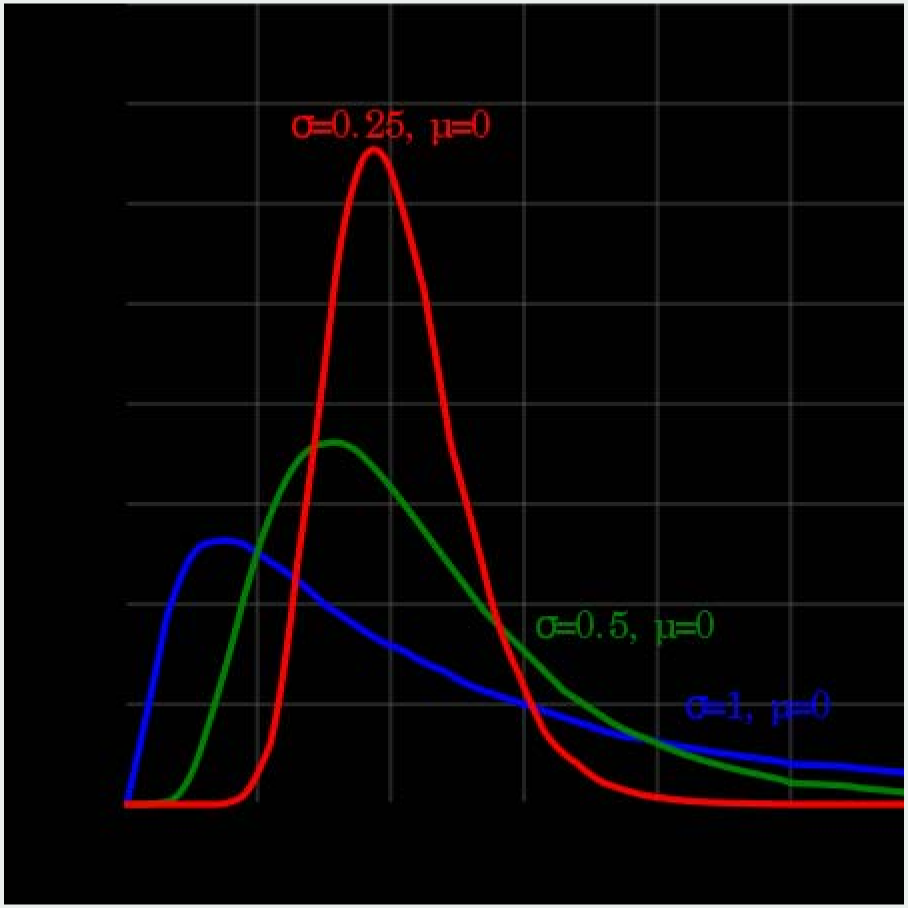
$\sigma > 0$ shape parameter, e^{μ} scale parameter

$E(X) = \mathrm{e}^{\mu + \sigma^2/2}$

$\mathrm{var}(X) = \mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2} - 1)$

$X \sim LN(\mu, \sigma^2) \Leftrightarrow X = e^{Y}$ with $Y \sim N(\mu, \sigma^2)$

LogNormal LN(μ,σ²): density function for μ=0



Discrete distributions

- Binomial(n,p):

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Geometric(p):

$$p(k) = P(X = k) = p(1-p)^k$$

Discrete distributions: Poisson(λ)

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \ (k = 0,1, \dots)$$

$$E(X) = \lambda; \quad var(X) = \lambda$$

- Let $Y_1, Y_2, \dots$ be independent and have an exponential distribution with rate λ, i.e. expected value $\frac{1}{\lambda}$, e.g. $Y_1, Y_2, \dots$ are interarrival times
- Then $\max\{i | \sum_{j=1}^{i} Y_j \leq 1\}$, i.e. the number of Y's that fit in 1 unit, i.e. the number of arrivals in 1 time period has the Poisson(λ) distribution.

Universiteit Utrecht

[Faculty of Science
Information and Computing Sciences]

Poisson process

- The arrival process $Y_1, Y_2, \ldots$ is called Poisson process with intensity λ.
- Suppose we have generated the number of arrivals from in a given time interval I by drawing this number from the Poisson distribution.
  - Then each individual arrival time is from the uniform distribution on that interval.

Background Poisson process

- Poisson distribution is `limit' of binomial distribution
  - Law of rare events

- Exponential inter arrival times result in Poisson distribution for number of arrivals per time unit
  - Memory-less

Poisson proces Suppose 100 students independently decide on going to the Super, each with probability $\frac{1}{10}$. The number of students that visits the Super follows a binomial distribution with $n = 100$ and $p = \frac{1}{10}$. The expected value equals 10.

n	p	$E(X)$	Distribution
100	$\frac{1}{10}$	10	Binomial
1000	$\frac{1}{100}$	10	Binomial
10000	$\frac{1}{1000}$	10	Binomial
100000	$\frac{1}{10000}$	10	Binomial
∞	0	10	Poisson with $\lambda = np = 10$

If we increase n and decrease p in such a way that we still have $np = 10$, the number of students that visit the Super, still follows the binomila distribution with $E(X) = 10$. If $n \to \infty$ and $p \to 0$ in such a way that $np = 10$, then the number of students that visit the Super follow a Poisson distribution with $E(X) = 10$. So, the Poisson Distribution is the limit of the binomial distribution. This is called the Law of rare events.

Suppose we have exponential inter-arrival times with average $\frac{1}{\mu}$ and intensity μ. We divide 1 time period into n small time periods of length $\frac{1}{n}$.

We now have

$$P(\text{arrival in interval}) = F(\frac{1}{n}) = (\text{memory less}) = 1 - e^{-\frac{\mu}{n}} = 1-(1-\frac{\mu}{n}+\frac{1}{2}(\frac{\mu}{n})^2-\ldots) \approx \frac{\mu}{n}.$$

Consequently

$$P(\text{no arrival in interval}) \approx 1 - \frac{\mu}{n}$$

Observe that in the above we have used that:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Now the number of arrival follows a binomial distribution with n and $p = \frac{\mu}{n}$ and hence the expected number of arrivals equals $np = \mu$.

Now suppose $n \to \infty$, then $p \to 0$. This models the situation where a very large number of people *individually* make a decision on going to the Super and decide to go the Super with a very small probability. By the law of rare events the number of arrivals follows a Poisson distribution with $\lambda = np = \mu$. Therefore the Poisson proces is a good model for the arrival of customers from the outside world.

Probability distribution: overview

- Continuous:
  - Uniform: first guess
  - Exponential: inter arrival times
  - Gamma: time to complete task
  - Weibull: time to complete task, time to failure
  - Normal: errors of various types, sum of large number of other quantities
  - LogNormal: time to complete task, estimate in the absence of data, time until maintenance, income
- Discrete:
  - Binomial: number of successes
  - Geometric: time until first success
  - Poisson(λ): gives number of arrival per time period when inter arrival times are exponentially distributed with parameter 1/λ.

Different distributions: example

- Single server queue
- Exponential interarrival times: avg 1 minute
- Historic data: 98 service times

Service time distribution	**Avg delay**	**Avg queue length**	**% delays >= 15**
Exponential	4.356	4.363	4.7
Gamma	2.849	2.845	1
Weibull	2.687	2.692	0.7
Lognormal	4.816	4.825	5.8
Normal	3.308	3.309	1.7

Which distribution?

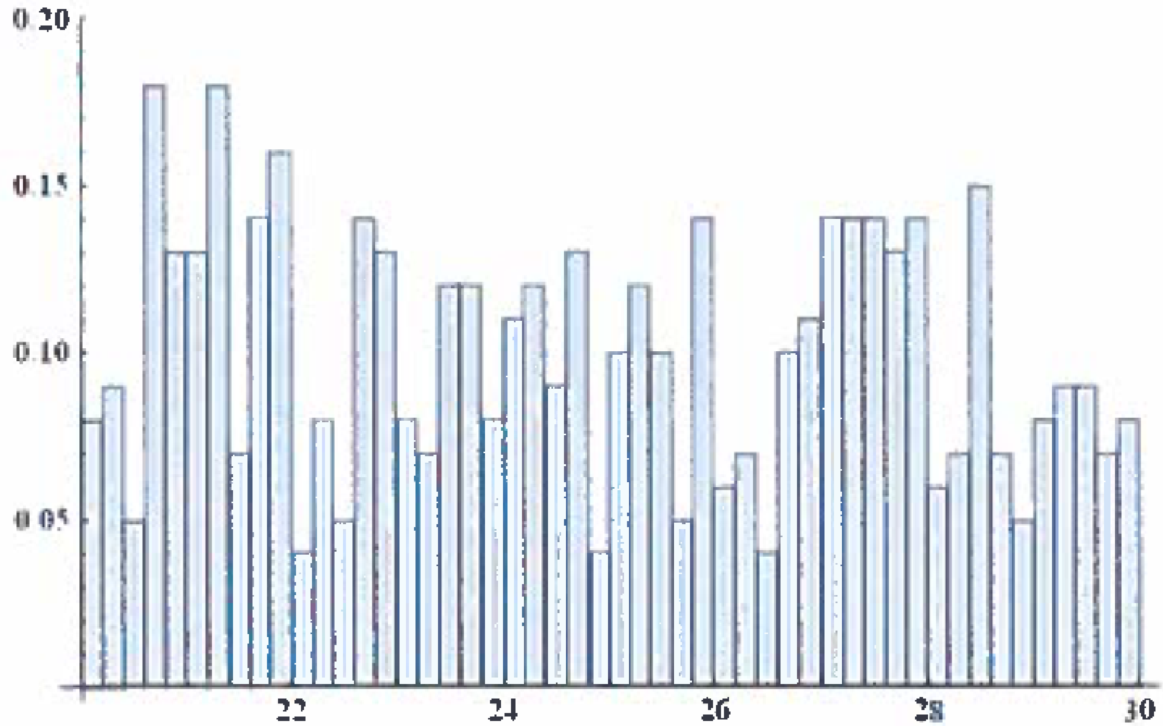


Figure 4.9 Probability density histogram with 50 bins of PF processing times data

Which distribution?

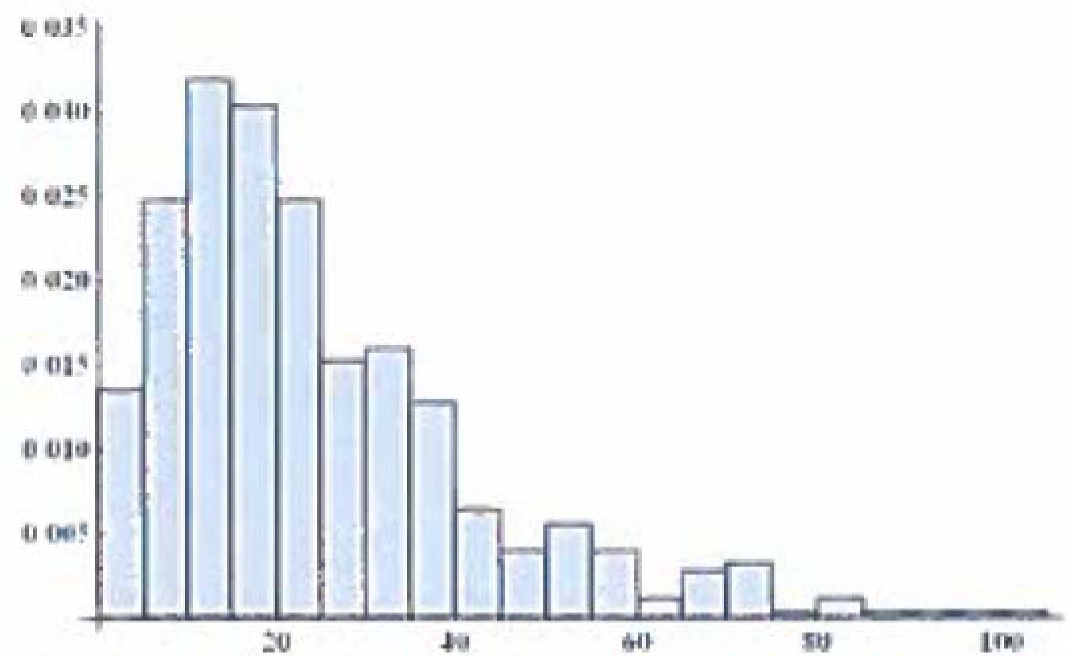


Figure 4.6 Probability density histogram of provided DC processing time measurements

Which distribution?

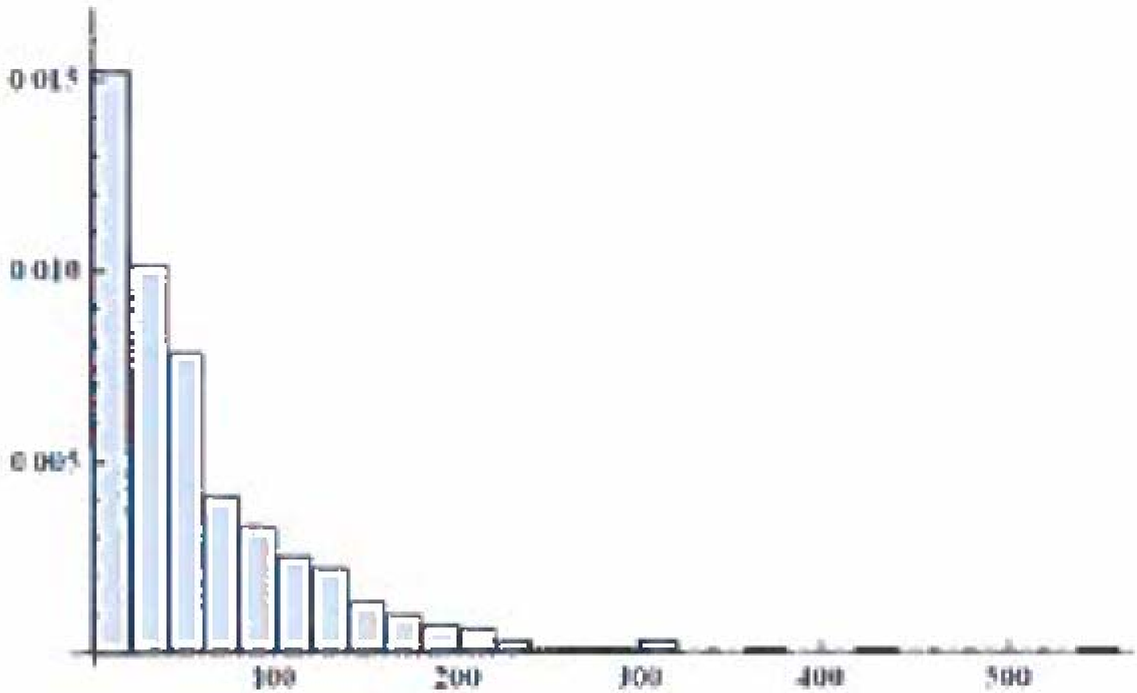


Figure 4.1 Probability density diagram of the provided IM Processing time measurements

Empirical distribution

- Given observations $X_1 <, \dots, < X_n$. Find probability distribution that directly follows from these observations.
- Discrete or continuous possible
- Discrete:
  - each observation probability 1/n.
- Example:
  - Time to complete assignment: 2,3,5,8,16,20
  - Discrete: only these values, each with p = 1/6
  - If you also want to generate values like 4.5, use continuous CDF.

Empirical distribution

- Continuous:
  - linear interpolation,
  - interval $[X_1, X_n]$

$$F(X_i) = P(X \leq X_i) = \frac{i-1}{n-1}$$

$$F(x) = \begin{cases} 0 & x < X_1 \\ \frac{i-1}{n-1} + \frac{x-X_i}{(X_{i+1}-X_i)(n-1)} & X_i \leq x < X_{i+1} \\ 1 & x \geq X_n \end{cases}$$

$$f(x) = \begin{cases} 0 & x < X_1 \\ \frac{1}{(X_{i+1}-X_i)(n-1)} & X_i \leq x < X_{i+1} \\ 0 & x \geq X_n \end{cases}$$

Input of simulation

1. Direct use of data: trace driven simulation
2. Empirical distribution
3. Theoretical probability distribution (see Law and Kelton tables 6.3 and 6.4 for an overview)

Advantages, disadvantages.

Input of simulation: advantages

1. Direct use of data: trace driven simulation
   - valid, few data, no modeling difficulties
2. Empirical distribution:
   - Given range, may have irregularities
3. Theoretical probability distribution (see Law and Kelton tables 6.3 and 6.4 for an overview)
   - smooth, compact, easy to change to run another scenario, no bound on the range, physical or theoretical reason

Wrap-up

- Simulation need stochastic variables for input entities that are subject to uncertainty
  - Interarrival times of customers
  - Time until machine breakdown
- Probability distributions are the best way to model these things:
  - E.g. interarrival times from exponential distribution
  - When your simulation program does: `schedule new arrival`

    You have to generate a random number from the exponential distribution to put the event in the event list with the right time-stamp

Fitting a distribution

- When you do a simulation study
- You hope to have a collection of data for the entity you need to model
- *Question: What probability distribution should I use?*
- **Fitting a distribution!**

Fitting a distribution

- Observations $X_1,\ldots,X_n$
- Finding a Cumulative Distribution Function F that models the observations is called **Fitting**

Goodness-of-fit

- Quality if the fit
- Can we assume that $X_1,\ldots,X_n$ really have distribution F ?

Fitting a distribution

- Clean the data
- Make a histogram
- Select a type of distribution
  - Visual inspection
  - Fitting software
- Estimate the parameters
- Evaluate the goodness of fit:
  - Heuristically
  - Statistically

Fitting a distribution

- Available software:
  - Expert fit (free for small number of data elements)
  - MATLAB
  - R

If fitting software suggests a distribution, it will be one with many parameters.

Fitting a distribution: Estimate the parameters

- $X_1,\ldots,X_n$ samples of Independent Identically Distributed (IID) stochastic variables
  - Sample mean $$\overline{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$
  - Sample variance $$S^2(n) = \frac{\sum_{i=1}^{n}(X_i - \overline{X}(n))^2}{n-1}$$
  - *These are unbiased estimators!*

Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

Evaluate goodness of fit heuristically: Q-Q plot

- Q-Q plot
  - Assume $X_1 \leq X_2 \leq \ldots \leq X_n$

If observations X_i are from distribution F

then we should have $\mathrm{P}(\mathrm{X} \leq \mathrm{X_i}) \approx \frac{i-0.5}{n}$ (where P is computed from F)

$\Leftrightarrow \mathrm{F}(\mathrm{X_i}) \approx \frac{i-0.5}{n} \Leftrightarrow X_i$ is percentile $\frac{i-0.5}{n}$

$\Leftrightarrow \mathrm{X_i} \approx \mathrm{F}^{-1}(\frac{i-0.5}{n})$

Draw $(X_i, F^{-1}(\frac{i-0.5}{n}))$

~ straight line `y=x'

Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

Goodness of fit heuristically: Q-Q plot

Draw $(X_i, F^{-1}(\frac{i-0.5}{n}))$

$F^{-1}(\frac{i-0.5}{n})$:

what would the distribution give?

My data : X_i

Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

Goodness of fit heuristically: Q-Q plot example

- F: U[0,1]
  Data points 0.11; 0.29; 0.52; 0.69; 0.93
- F: U[0,1]
  Data points 0.1; 0.11;1 0.3; 0.4; 0.9

Evaluate goodness of fit: statistical tests

- Hypothesis H_0:
  - the observations $X_1, X_2, \ldots, X_n$ follow distribution *F*
- Do we accept or reject this hypothesis?

- Chi-squared test
- Kolmogorov-Smirnov test

- Can be performed with R

Statistical testing in general: Example

- Suppose we investigate the average number of hours gaming per week for CS students and we do not want collect numbers from all students.
- Our hypothesis is H0: avg = 25
- Suppose we sample five persons and find: 24, 24,24, 26, 26 (*avg* 24.8)
- Compute average from sample and ask: \`Do we believe H0?’
- Believe H0 if avg is close enough to 25
  - How close is close enough?
  - The above sample looks close enough, but how about: 20, 22, 23, 24, 26 (*avg* 23)
  - Can statistics help to decide?

Statistical hypothesis testing

1. Formulate hypothesis H_0
   - For goodness-of-fit tests the observations $X_1, X_2, \ldots, X_n$ follow distribution F
2. Choose type of test
3. Determine significance α and decision rule
4. Compute test statistic and take decision

Errors

		Case	
		H_0 true	**H_0 false**
Decision	**Accept H_0**	OK Probability 1-α *Confidence level*	Type 2
	Reject H_0	Type 1 Probability α *Significance*	OK

Chi-squared test

H_0: the observations $X_1, X_2, \ldots, X_n$ follow distribution F

$$[a_0, a_1), [a_1, a_2), \ldots, [a_{K-1}, a_K)$$

Observed number:

$$\mathrm{N_i} = \# X_j \text{ in } [a_i, a_{i+1})$$

Expected number according to probability distribution:

$$E_i = np_i = n \int_{a_i}^{a_{i+1}} f(x)dx$$

$$\text{Test statistic}: \mathrm{X}^2 = \sum_{i=0}^{K-1} \frac{(N_i - E_i)^2}{E_i}$$

Interval choice:

- No $E_i < 1$
- No more than 20% has $N_i < 5$

Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

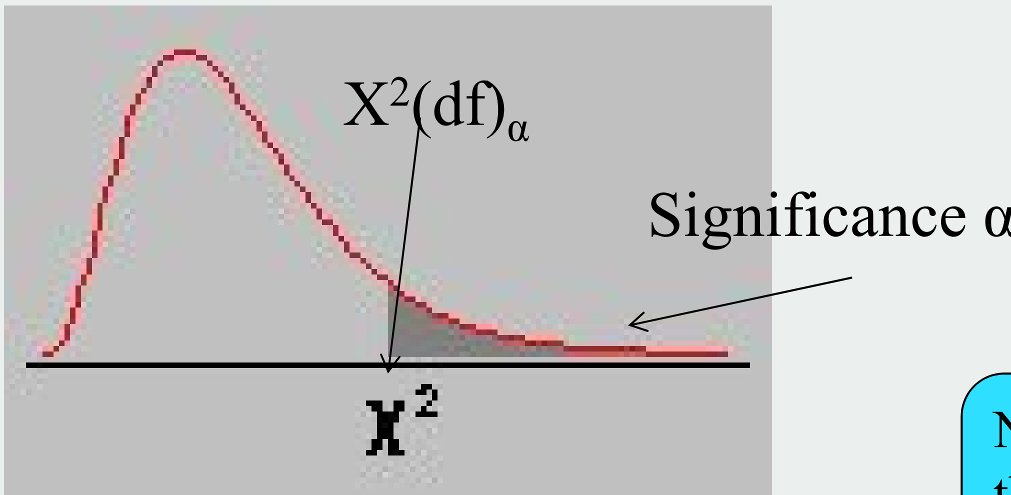
Chi-squared test (2)

- Suppose H_0 is true
- Suppose we draw N values $X_1, X_2, \ldots, X_n$ from the probablity distribution F
- Then we can compute $\mathrm{X}^2 = \sum_{i=0}^{K-1} \frac{(N_i - E_i)^2}{E_i}$
- Now X^2 is a stochastic variable
- Statistical theory learns us X^2 follows a chi-squared-distribution with df=K-1
- We expect it to be small, but it may by coincidence attain a large value.
- However, at some point we do not believe H_0 anymore.

Universiteit Utrecht

[Faculty of **Science**
Information and Computing Sciences]

If H0 is true then X² ~ chi-squared-distribution with df=K-1



www.statsoft.com

Now, the probability that we reject H_0 while it is true is only α

Accept H_0 if $X^2 \leq X^2_\alpha(df)$
Reject otherwise

Example

- Are the following values from U[0,1]
  a) 0.07; 0.15; 0.24; 0.31; 0.42; 0.51; 0.55; 0.65; 0.73; 0.76; 0.85; 0.97
  b) 0.07; 0.08; 0.15; 0.18; 0.51; 0.52; 0.53; 0.58; 0.64; 0.68; 0.74; 0.95

- Use intervals [0;0.25), [0.25;0.5), [0.5;0.75), [0.75;1] and α=5%.

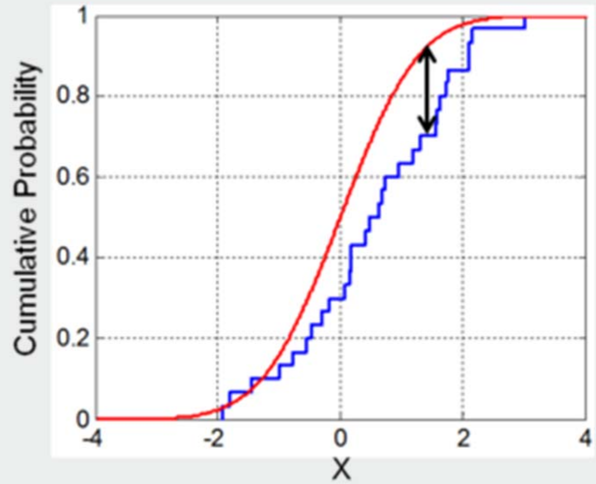
Statistical table:
http://www.statsoft.com/Textbook/Distribution-Tables#chi

Kolmogorov-Smirnov test

H_0 : Do $X_1 < X_2 < .... < X_n$ follow distribution F?

Empirical distribution $F_n(x) = \dfrac{i}{n}$ for $X_i \leq x < X_{i+1}$

$D_n^* = \sup_x | F(x) - F_n(x) |$



Universiteit Utrecht

[Faculty of Science
Information and Computing Sciences]

Kolmogorov-Smirnov test (2)

$D_n^* = \sup_x |F(x) - F_n(x)|$

If H_0 is true the following holds

$$H(t) = \lim_{n\to\infty} P\left(\sqrt{n}D_n^* \leq t\right) = 1 - \sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2t^2}$$

$H(t)$ defines a probability distribution

After this lecture

- You know different methods for generating stochastic input variables for simulation
- You know different probability distributions
- You are able to fit a probability distribution to input data

Covariance, correlation

Given 2 stochastic variables X and Y :

covariance: $\mathrm{cov}(X,Y) = E((X - E(X))(Y - E(Y)))$

correlation : $\rho_{\mathrm{XY}} = \dfrac{\mathrm{cov}(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}$

Example 1: Die
- X = score
- Y =7 – score of same die

Example 2
- X score Dutch coin
- Y score Italian coin
- Where, head = 1, tail = 0

Independence

Two stochastic variables X and Y are independent if:

- Continuous

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B) \quad \forall \text{sets } A, B$$

- Discrete

$$P(X = x \text{ and } Y = y) = P(X = x)P(Y = y) \quad \forall x, y$$

Independence (2)

If X and Y independent stochastic variables:

$$\mathrm{E}(\mathrm{XY}) = \mathrm{E}(\mathrm{X})\mathrm{E}(\mathrm{Y})$$

$$\mathrm{cov}(\mathrm{X}, \mathrm{Y}) = 0$$

$$\mathrm{var}(\mathrm{X} + \mathrm{Y}) = \mathrm{var}(\mathrm{X}) + \mathrm{var}(\mathrm{Y})$$

Wrap-up

- Previous lecture we studied some well-known probability distribution.
- How do you know which probability distribution you should use?
  - From data
  - Sometimes from theory
- This lecture:
  - How to find out if you choose the correct distribution?

Wrap-up

- Simulation need stochastic variables for input entities that are subject to uncertainty
  - Interarrival times of customers
  - Time until machine breakdown
- Probability distributions are the best way to model these things:
  - E.g. interarrival times from exponential distribution
  - When your simulation program does: `schedule new arrival`

    You have to generate a random number from the exponential distribution to put the event in the event list with the right time-stamp