Faculty of Engeneering and Natural Sciences

JOHANNES KEPLER UNIVERSITY LINZ

Statistics in neuroscience (with modelling here and there)

Massimiliano Tamborrino
massimiliano.tamborrino@jku.at
www.jku.at/stochastik/tamborrino

Institute for Stochastics
Johannes Kepler University Linz

Outline

A good quote on Statistics

* *"The best thing about being a statistician is that you get to play in everyone else's backyard."* (John Tukey, Bell Labs, Princeton University).

A good quote on Statistics

* *"The best thing about being a statistician is that you get to play in everyone else's backyard."* (John Tukey, Bell Labs, Princeton University).

Statisticians apply statistical thinking and methods to a wide variety of scientific, social, and business endeavours in such areas as:

- astronomy;
- chemistry (e.g. problem of estimating folding rates for some protein having a coordinate switching between the folded and unfolded state);
- biology (including physiology, **neuroscience**, etc.);
- bioinformatics (e.g. patter recognition, see on Tuesday);
- economics/finance;
- engineering;
- genetics (detection of genetic variations);
- mathematics;
- medicine;
- **all kind of recognition;**
- physics (you've seen it yesterday);
- **psychology** (visual attention).

What is Statistics?

Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances. (Davidian, M. and Louis, T. A., 10.1126/science.1218685).

Before starting... Let us focus on:

- **Are our data/experiments good?**
- **What question do we want to answer?**

- **Which model do we want to consider?**

  \+

  \-

What is Statistics?

Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances. (Davidian, M. and Louis, T. A., 10.1126/science.1218685).

Before starting... Let us focus on:

- **Are our data/experiments good?**
- **What question do we want to answer?**
  - *An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem*. John Tukey.
  - *There are no routine statistical questions, only questionable statistical routines*. Sir D. R. Cox.
- **Which model do we want to consider?**

  \+

  \-

What is Statistics?

Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances. (Davidian, M. and Louis, T. A., 10.1126/science.1218685).

Before starting... Let us focus on:

- **Are our data/experiments good?**
- **What question do we want to answer?**
  - *An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem*. John Tukey.
  - *There are no routine statistical questions, only questionable statistical routines*. Sir D. R. Cox.
- **Which model do we want to consider?**
  \+ *All models are wrong, but some are useful*. G. E. P. Box.
  \- *Statisticians, like artists, have the bad habit of falling in love with their models*. G. E. P. Box.

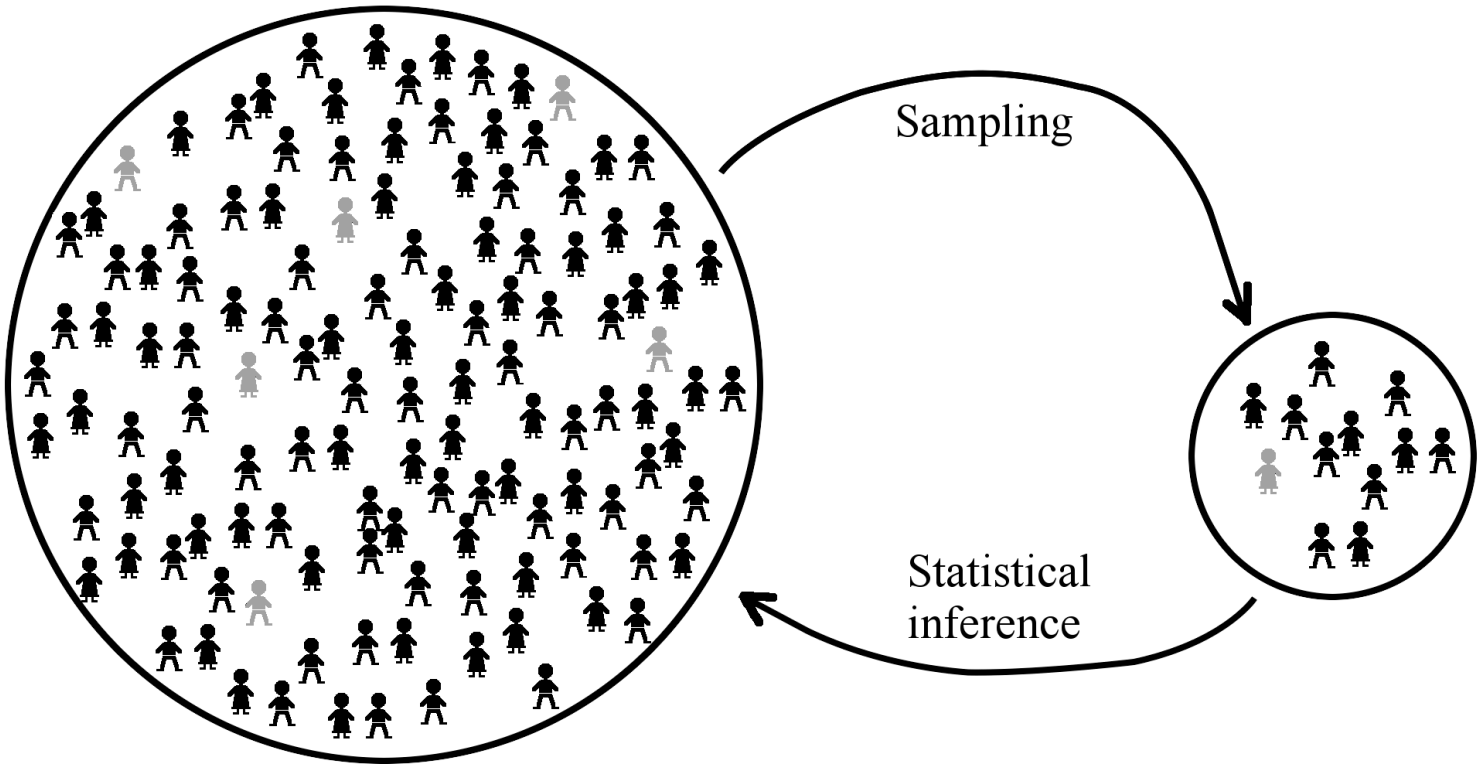
JOHANNES KEPLER UNIVERSITY LINZ

Sampling and statistical inference

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. John Tukey.

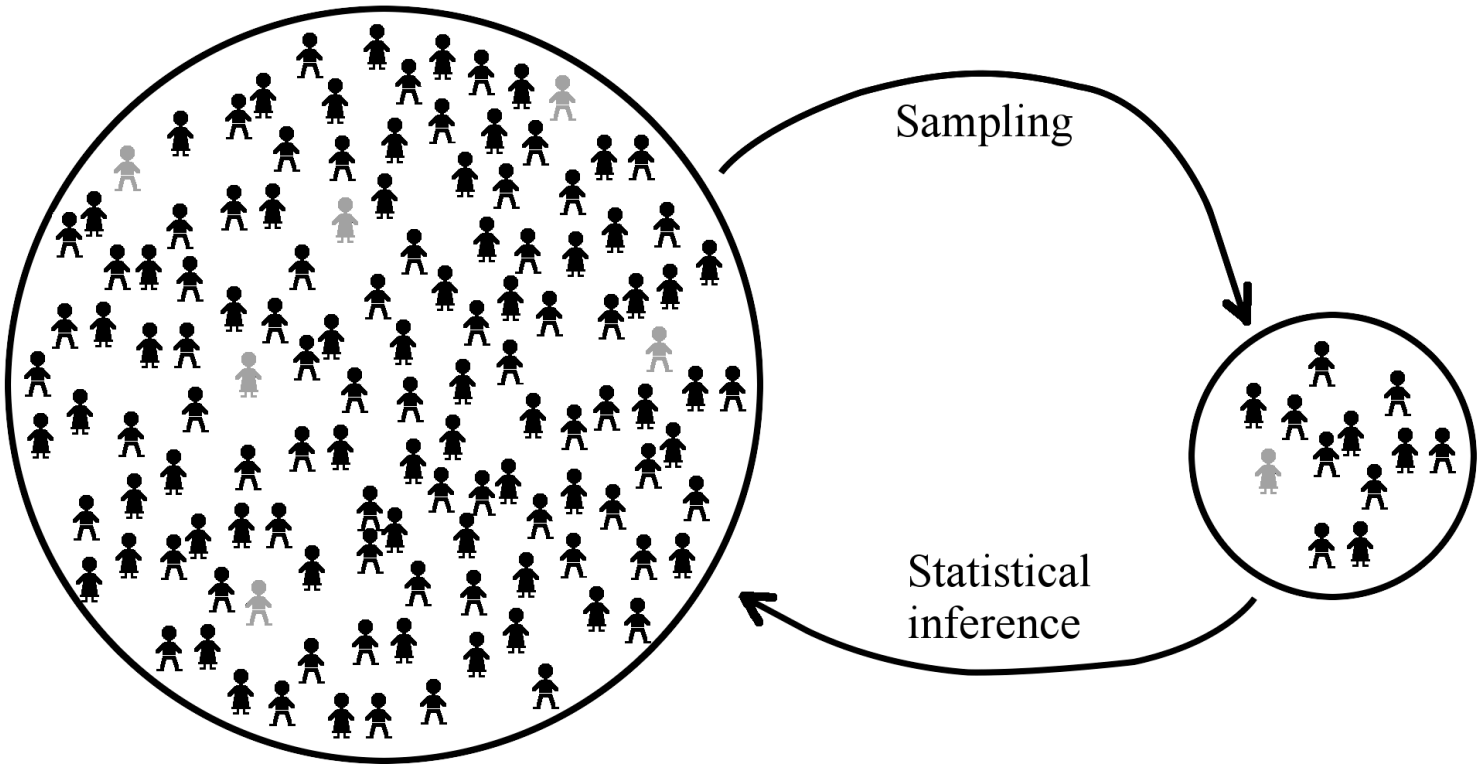
Sampling and statistical inference

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. John Tukey.



Sampling and statistical inference

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. John Tukey.



All generalizations are false, including this one. Mark Twain.

The importance of good sampling/data

An election will be held next week and, by polling a sample of the voting population, we are trying to predict whether the Republican or Democratic candidate will prevail. Which of the following methods of selection is likely to yield a representative sample?

(a) Poll all people of voting age attending a basketball game.

(b) Obtain a copy of the voter registration list, randomly choose 100 names, and question them.

(c) Use the results of a television call-inpoll, in which the station asked its listeners to call in and name their choice.

(d) Randomly choose 100 names from the telephone directory and call these people.

The importance of good sampling/data

An election will be held next week and, by polling a sample of the voting population, we are trying to predict whether the Republican or Democratic candidate will prevail. Which of the following methods of selection is likely to yield a representative sample?

(a) Poll all people of voting age attending a basketball game.

(b) Obtain a copy of the voter registration list, randomly choose 100 names, and question them.

(c) Use the results of a television call-inpoll, in which the station asked its listeners to call in and name their choice.

(d) Randomly choose 100 names from the telephone directory and call these people.

The approach (d) led to a disastrous prediction in the 1936 presidential election, in which Franklin Roosevelt defeated Alfred Landon by a landslide. A Landon victory had been predicted by the Literary Digest. The magazine based its prediction on the preferences of a sample of voters chosen from lists of cars and telephone owners.

Why the Literary Digests prediction was so far off?

A further quote on Statistics and data

* *If you torture the data enough, nature will always confess*. Ronald Coase.

A further quote on Statistics and data

* *If you torture the data enough, nature will always confess*. Ronald Coase.

Two doctors are planning to try out a new treatment on patients, both male and female. The 1st doctor writes:

	Male		Female	
Treatment	Standard	New	Standard	New
Dead:	950	9000	5000	5
Alive:	50(5%)	1000 (10%)	5000 (50%)	95(95%)

**On Simpson's Paradox and the Sure-Thing Principle*, Coling R. Blyth, J. Am. Stat. Assoc., 67 (388), 364–366 (on KUSSS)

A further quote on Statistics and data

* *If you torture the data enough, nature will always confess*. Ronald Coase.

Two doctors are planning to try out a new treatment on patients, both male and female. The 1st doctor writes:

	Male		Female	
Treatment	Standard	New	Standard	New
Dead:	950	9000	5000	5
Alive:	50(5%)	1000 (10%)	5000 (50%)	95(95%)

The 2nd doctor, sum up the data in the following table:

Treatment	Standard	New
Dead:	5950	9005
Alive:	5050 (46%)	1095 (11%)

**On Simpson's Paradox and the Sure-Thing Principle*, Coling R. Blyth, J. Am. Stat. Assoc., 67 (388), 364–366 (on KUSSS)

Sample space, outcome, events, probability

Sample space Ω: set of all possible outcomes (i.e. results) of an experiment.
A possible **outcome** ω is an element of the sample space, i.e. $\omega \in \Omega$.
An **event** A is a subset of the sample space Ω, i.e. $A \subset \Omega$.

* Rules for probabilities are for statistics what arithmetics is for mathematics.

Some famous results from probability theory:
Let A and B be two events in Ω, with $\mathbb{P}(B) > 0$.

* **Conditional probability** for A given B: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

* **Multiplication rule**. $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

***Bayes' Theorem**: $\mathbb{P}(A|B) =$ $\Rightarrow$ we can compute $\mathbb{P}(A|B)$ from $\mathbb{P}(B|A)$.

* **Law of total probability**. If $\{B_1, \ldots, B_n\}$ is a finite partition of Ω, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

* A, B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
Theorem: A, B independent $\iff \mathbb{P}(A|B) =$.

Sample space, outcome, events, probability

Sample space Ω: set of all possible outcomes (i.e. results) of an experiment.
A possible **outcome** ω is an element of the sample space, i.e. $\omega \in \Omega$.
An **event** A is a subset of the sample space Ω, i.e. $A \subset \Omega$.

* Rules for probabilities are for statistics what arithmetics is for mathematics.

Some famous results from probability theory:
Let A and B be two events in Ω, with $\mathbb{P}(B) > 0$.

* **Conditional probability** for A given B: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

* **Multiplication rule**. $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

***Bayes' Theorem**: $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \Rightarrow$ we can compute $\mathbb{P}(A|B)$ from $\mathbb{P}(B|A)$.

* **Law of total probability**. If $\{B_1, \ldots, B_n\}$ is a finite partition of Ω, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

* A, B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
Theorem: A, B independent $\iff \mathbb{P}(A|B) =$.

Sample space, outcome, events, probability

Sample space Ω: set of all possible outcomes (i.e. results) of an experiment.
A possible **outcome** ω is an element of the sample space, i.e. $\omega \in \Omega$.
An **event** A is a subset of the sample space Ω, i.e. $A \subset \Omega$.

* Rules for probabilities are for statistics what arithmetics is for mathematics.

Some famous results from probability theory:
Let A and B be two events in Ω, with $\mathbb{P}(B) > 0$.

* **Conditional probability** for A given B: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

* **Multiplication rule**. $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

***Bayes' Theorem**: $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \Rightarrow$ we can compute $\mathbb{P}(A|B)$ from $\mathbb{P}(B|A)$.

* **Law of total probability**. If $\{B_1, \ldots, B_n\}$ is a finite partition of Ω, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

* A, B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
Theorem: A, B independent $\iff \mathbb{P}(A|B) = \mathbb{P}(A)$.

Let's go back to our medical issue

	Male (M)		Female (F)	
Treatment	Standard (S)	New (N)	Standard (S)	New (N)
Dead (D):	950	9000	5000	5
Alive (A):	50(5%)	1000 (10%)	5000 (50%)	95(95%)

Treatment	Standard (S)	New (N)
Dead (D):	5950	9005
Alive (A):	5050 (46%)	1095 (11%)

We have:

$$\mathbb{P}(A|N,M) = 0.1 \quad > \quad \mathbb{P}(A|S,M) = 0.05$$
$$\mathbb{P}(A|N,F) = 0.95 \quad > \quad \mathbb{P}(A|S,F) = 0.50$$

Let's go back to our medical issue

	Male (M)		Female (F)	
Treatment	Standard (S)	New (N)	Standard (S)	New (N)
Dead (D):	950	9000	5000	5
Alive (A):	50(5%)	1000 (10%)	5000 (50%)	95(95%)

Treatment	Standard (S)	New (N)
Dead (D):	5950	9005
Alive (A):	5050 (46%)	1095 (11%)

We have:

$$\begin{aligned} \mathbb{P}(A|N,M) = 0.1 &> \mathbb{P}(A|S,M) = 0.05 \\ \mathbb{P}(A|N,F) = 0.95 &> \mathbb{P}(A|S,F) = 0.50 \\ \text{However} \quad \mathbb{P}(A|N) = 0.1 &< \mathbb{P}(A|S) = 0.46. \end{aligned}$$

Let's go back to our medical issue

	Male (M)		Female (F)	
Treatment	Standard (S)	New (N)	Standard (S)	New (N)
Dead (D):	950	9000	5000	5
Alive (A):	50(5%)	1000 (10%)	5000 (50%)	95(95%)

Treatment	Standard (S)	New (N)
Dead (D):	5950	9005
Alive (A):	5050 (46%)	1095 (11%)

We have:

$$\begin{aligned} \mathbb{P}(A|N,M) = 0.1 &> \mathbb{P}(A|S,M) = 0.05 \\ \mathbb{P}(A|N,F) = 0.95 &> \mathbb{P}(A|S,F) = 0.50 \\ \text{However} \quad \mathbb{P}(A|N) = 0.1 &< \mathbb{P}(A|S) = 0.46. \end{aligned}$$

How can that be possible?

Let's go back to our medical issue

	Male (M)		Female (F)	
Treatment	Standard (S)	New (N)	Standard (S)	New (N)
Dead (D):	950	9000	5000	5
Alive (A):	50(5%)	1000 (10%)	5000 (50%)	95(95%)

Treatment	Standard (S)	New (N)
Dead (D):	5950	9005
Alive (A):	5050 (46%)	1095 (11%)

We have:

$$\begin{aligned} \mathbb{P}(A|N,M) = 0.1 &> \mathbb{P}(A|S,M) = 0.05 \\ \mathbb{P}(A|N,F) = 0.95 &> \mathbb{P}(A|S,F) = 0.50 \\ \text{However} \quad \mathbb{P}(A|N) = 0.1 &< \mathbb{P}(A|S) = 0.46. \end{aligned}$$

How can that be possible? Look at this table

	Standard	New
Male	1000 (9%)	10000 (99%)
Female	10000 (91%)	100 (1%)

and think about conditional probability!

Random variables. Sample space. Distribution

We use probability theory to describe phenomena involving randomness.

The sample space Ω is the set of all possible outcomes (i.e. results) ω of an experiment.

A random/stochastic variable X is the random outcome of an 'experiment'.

E: set containing the possible values assumed by X. Discrete rv: $E = \mathbb{N}$ or subset (finite/infinite). Continuous rv: $E = \mathbb{R}$ or (infinite) subset.

The distribution of X describes the probability of an outcome or ranges of outcomes. To define it, we need to specify the probability of each outcome.

Examples when rolling dice:
* X: rv recording the outcome when rolling a die. Ω? E?
* X: rv counting the ones when rolling 2 dice. Ω?E?

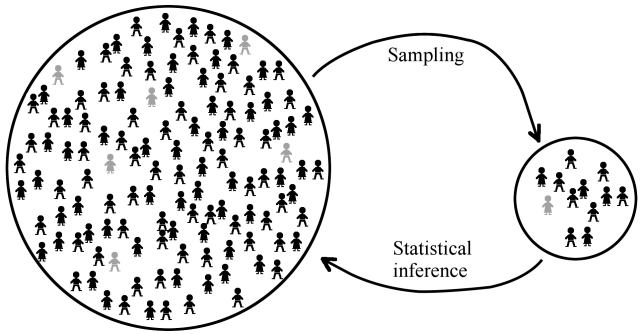
Statistical model

We run an experiment and we get some data.

1. We perform a statistical analysis, e.g. estimate some parameter, run some tests, etc.
2. We derive our conclusions **based on our data.**

New data $\Rightarrow$ New statistical analysis $\Rightarrow$ New conclusion ?

How much can we rely on our conclusions/estimates?



* Data (or observations) $\mathbf{x} = (x_1, \ldots, x_n)$: outcome of an experiment. $\mathbf{x}$ is a value (realization) of a **random variable** $\mathbf{X} = (X_1, \ldots, X_n)$.
* Idea: use $\mathbf{x}$ to obtain info about the distribution (why?) of $\mathbf{X}$.
* Statistical Model: set $\mathcal{P}$ of possible probability measures P_θ parametrized by a parameter θ, that is

$$\mathcal{P} = \{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\},$$

with Θ: parameter space.

Samples, statistical models and statistics

*Identifiability: For $\theta_1, \theta_2 \in \Theta$, if $\theta_1 \neq \theta_2 \Rightarrow \mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. [e.g. FPTs for Wiener process]

*Sample $\mathbf{X} = (X_1, \ldots, X_n)$: collection of independent rvs X_i with distribution $P_{i,\boldsymbol{\theta}} \Rightarrow$ Statistical model:

*iid sample $\mathbf{X}$: collection of iid rvs where all X_i have the same distribution $P_{i,\boldsymbol{\theta}} = P_{1,\boldsymbol{\theta}} \Rightarrow$ Statistical model:

Obs $p(\mathbf{x}; \theta) := p_{\boldsymbol{\theta}}(\mathbf{x}) \overset{indep.}{=} \prod_{i=1}^n p_{i,\boldsymbol{\theta}}(x_i) \overset{ident.distr}{=} \prod_{i=1}^n p_{1,\boldsymbol{\theta}}(x_i)$.

*A statistic is a function of the random sample which *compresses the data*, i.e. it defines a *data reduction*/*data summary*. Very well known statistics:
Sample mean $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$; *Sample variance* $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, minimum and maximum.

Mathematical statements derived from statistical methods are exact under the chosen model but their validity in practice depends on how well the model reflects the problem! $\Rightarrow$ All models are true and good BEFORE validating them on data. $\Rightarrow$ Model validation (Does the model fit the data?) is **CRUCIAL**.

JKU JOHANNES KEPLER UNIVERSITY LINZ

Samples, statistical models and statistics

*Identifiability: For $\theta_1, \theta_2 \in \Theta$, if $\theta_1 \neq \theta_2 \Rightarrow \mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. [e.g. FPTs for Wiener process]

*Sample $\mathbf{X} = (X_1, \ldots, X_n)$: collection of independent rvs X_i with distribution $P_{i,\boldsymbol{\theta}} \Rightarrow$ Statistical model: $\mathcal{P} = P_{1,\boldsymbol{\theta}} \times \cdots P_{n,\boldsymbol{\theta}}$.

*iid sample $\mathbf{X}$: collection of iid rvs where all X_i have the same distribution $P_{i,\boldsymbol{\theta}} = P_{1,\boldsymbol{\theta}} \Rightarrow$ Statistical model:

Obs $p(\mathbf{x}; \theta) := p_{\boldsymbol{\theta}}(\mathbf{x}) \overset{indep.}{=} \prod_{i=1}^n p_{i,\boldsymbol{\theta}}(x_i) \overset{ident.distr}{=} \prod_{i=1}^n p_{1,\boldsymbol{\theta}}(x_i)$.

*A statistic is a function of the random sample which *compresses the data*, i.e. it defines a *data reduction/data summary*. Very well known statistics:
Sample mean $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$; *Sample variance* $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, minimum and maximum.

Mathematical statements derived from statistical methods are exact under the chosen model but their validity in practice depends on how well the model reflects the problem! $\Rightarrow$ All models are true and good BEFORE validating them on data. $\Rightarrow$ Model validation (Does the model fit the data?) is **CRUCIAL**.

JKU JOHANNES KEPLER UNIVERSITY LINZ

Samples, statistical models and statistics

*Identifiability: For $\theta_1, \theta_2 \in \Theta$, if $\theta_1 \neq \theta_2 \Rightarrow \mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. [e.g. FPTs for Wiener process]

*Sample $\mathbf{X} = (X_1, \ldots, X_n)$: collection of independent rvs X_i with distribution $P_{i,\boldsymbol{\theta}} \Rightarrow$ Statistical model: $\mathcal{P} = P_{1,\boldsymbol{\theta}} \times \cdots P_{n,\boldsymbol{\theta}}$.

*iid sample $\mathbf{X}$: collection of iid rvs where all X_i have the same distribution $P_{i,\boldsymbol{\theta}} = P_{1,\boldsymbol{\theta}} \Rightarrow$ Statistical model: $\mathcal{P} = P_{1,\boldsymbol{\theta}}^{\otimes n}$.

Obs $p(\mathbf{x}; \theta) := p_{\boldsymbol{\theta}}(\mathbf{x}) \overset{indep.}{=} \prod_{i=1}^n p_{i,\boldsymbol{\theta}}(x_i) \overset{ident.distr}{=} \prod_{i=1}^n p_{1,\boldsymbol{\theta}}(x_i)$.

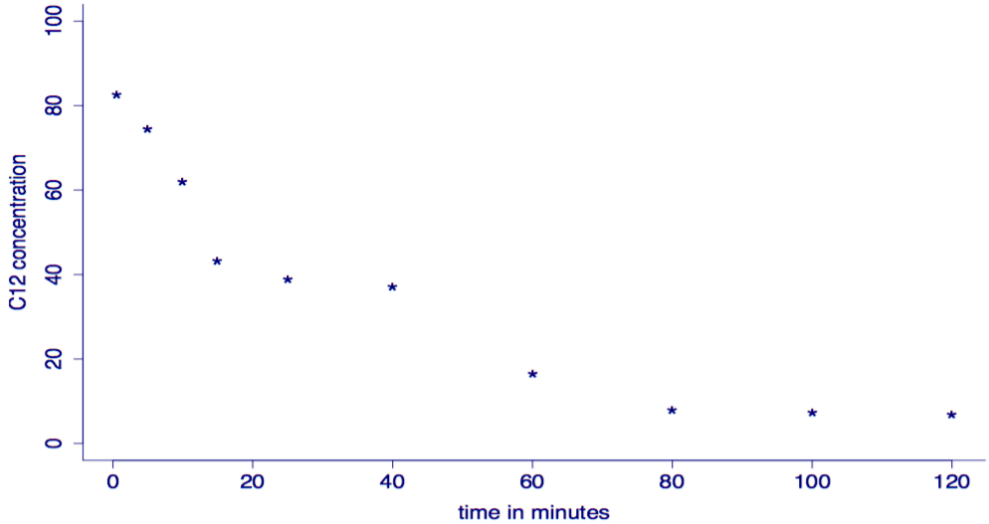
*A statistic is a function of the random sample which *compresses the data*, i.e. it defines a *data reduction/data summary*. Very well known statistics:
Sample mean $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$; *Sample variance* $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, minimum and maximum.

Mathematical statements derived from statistical methods are exact under the chosen model but their validity in practice depends on how well the model reflects the problem! $\Rightarrow$ All models are true and good BEFORE validating them on data. $\Rightarrow$ Model validation (Does the model fit the data?) is **CRUCIAL**.

JKU JOHANNES KEPLER UNIVERSITY LINZ

Example: Deterministic vs Stochastic models.

Experimental data of dodecanedioic acid (C12) in rat livers

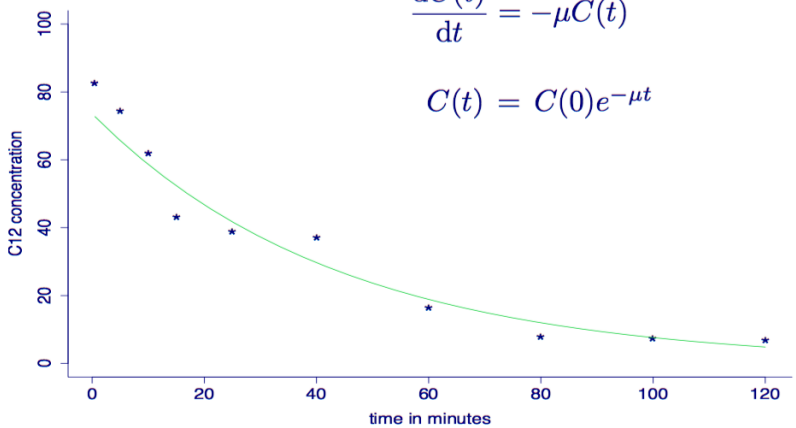


Data from: S. Ditlevsen, A. De Gaetano. *Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers*. Bulletin of Mathematical Biology, 67 (3), 547–561, 2005.

Assuming exponential decay

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = -\mu C(t)$$

$$C(t) = C(0)e^{-\mu t}$$

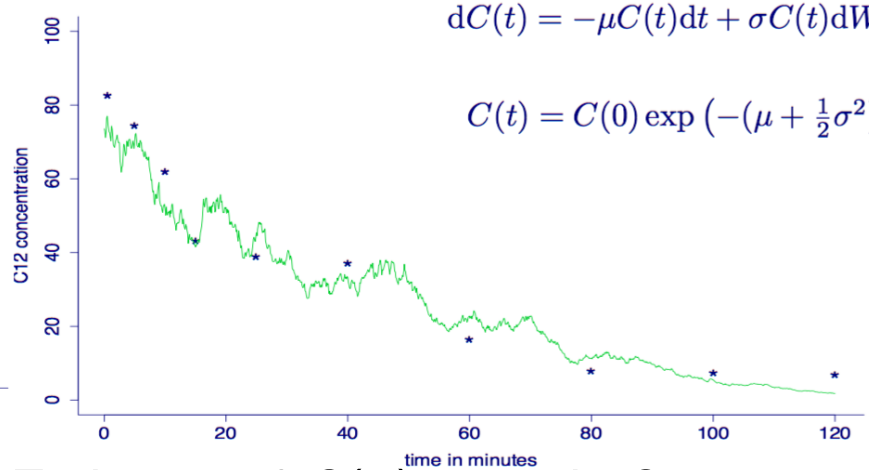


Estimate of $C(0)$ and μ?

Assuming exponential decay with noise

$$\mathrm{d}C(t) = -\mu C(t)\mathrm{d}t + \sigma C(t)\mathrm{d}W(t)$$

$$C(t) = C(0)\exp\left(-(\mu + \tfrac{1}{2}\sigma^2)t + \sigma W(t)\right)$$



Estimate of $C(0)$, μ and σ?

Estimation in fully (discretely) observed diffusion processes

* Diffusion process: Stochastic process obtained as solution of a SDE (e.g. using Ito's formula).

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

X, b and W d-dimensional, σ: $d \times d$ matrix.

State space: $D \subseteq \mathbb{R}^d$.

For $d = 1$, $\quad D = (l, r), -\infty \leq l < r \leq \infty$.

Data: $X_{t_1}, \ldots, X_{t_n}, t_i = \Delta i, i = 1, \ldots, n$.

Prop: A diffusion process is a Markov process.

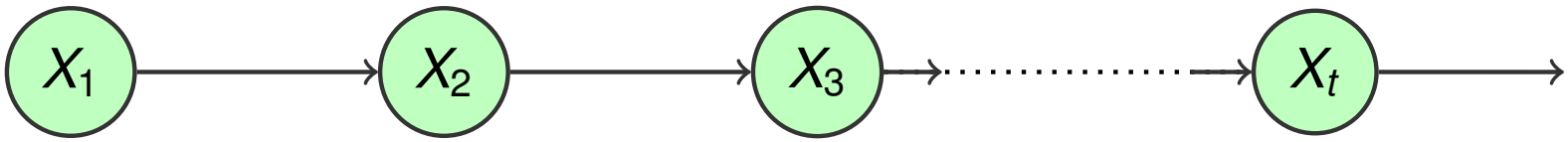
Markov process (Hints)

* For simplicity, let us consider discrete time stochastic process $\{X_t : t \in \mathbb{N}\}$.

* Finite *state-space* $S = \{1, \ldots, m\}$ such that $X_t = j \in S$ for all $t \in \mathbb{N}$.

* **Markov property**: Define $\mathbf{X}^{(t)} = (X_1, X_2, \ldots, X_t)$ the history till time t. Then

$$\mathbb{P}(X_{t+1} \mid \mathbf{X}^{(t)}) = \mathbb{P}(X_{t+1} \mid X_t, X_{t-1}, \ldots, X_1) = \mathbb{P}(X_{t+1} \mid X_t).$$



* A **Markov chain** (MC) is a stochastic process with the property that the future states are independent of the past states **given the present state**.

Maximum likelihood estimation (MLE).

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

*Data: $X_{t_1}, \ldots, X_{t_n}, \quad t_1 < \cdots < t_n$.

∗Likelihood function: $L_n(\theta) = p(X_{t_1}, \ldots, X_{t_n}; \theta)$.

(What is a likelihood function? Meaning? Interpretation?)

* Maximizing the likelihood yields the maximum likelihood estimator.

* Note: It is easier to maximize the log-likelihood $l_n(\theta) = \log L_n(\theta)$ (it has the same maximum).

* For simplicity: assume $\theta \in \mathbb{R}$, i.e. only one parameter unknown (similar for $\theta \in \Theta \subseteq \mathbb{R}^p$). Derive $\hat{\theta}$ as follows:

1. Compute the likelihood function $L_n(\theta)$;
2. Compute the log-likelihood $l_n(\theta) = \log L(\theta)$;
3. Compute the score function(s) $\partial_\theta l_n(\theta)$;
4. Estimator $\hat{\theta}$ such that $\partial_\theta l_n(\theta) = 0$.

How to compute the likelihood $L_n(\theta)$?

The "correct" likelihood function

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p.$$

Data: $X_{t_1}, \ldots, X_{t_n}, \quad t_1 < \ldots < t_n$.

We have the "correct"(under the assumed model) likelihood-function:

$$L_n(\theta) = p(X_{t_0}, X_{t_1}, \ldots, X_{t_n};\theta).$$

Applying Bayes' theorem several times we get

$$\begin{aligned} L_n(\theta) &\overset{\text{Bayes' Th.}}{=} p(X_{t_n}|X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}};\theta) \times p(X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}};\theta) \\ &= \text{Continuing applying Bayes' theorem this way} \\ &= p(X_{t_n}|X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}};\theta) \times p(X_{t_{n-1}}|X_{t_0}, X_{t_1}, \ldots, X_{t_{n-2}};\theta) \times \cdots \\ &\quad \cdots \times p(X_{t_2}|X_{t_0}, X_{t_1};\theta) \times p(X_{t_1}|X_{t_0};\theta)p(X_{t_0};\theta) \end{aligned}$$

An important property of our observations, which they inherit from the diffusion process: they are a Markov process. Thus:

$$p(X_{t_n}|X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}};\theta) = p(X_{t_n}|X_{t_{n-1}};\theta)$$

and therefore

$$\begin{aligned} L_n(\theta) &= p(X_{t_n}|X_{t_{n-1}};\theta) \times \cdots p(X_{t_1}|X_{t_0};\theta)p(X_{t_0};\theta) = \prod_{i=1}^{n} p(X_{t_i}|X_{t_{i-1}};\theta)p(X_{t_0};\theta). \\ \overset{\text{log-lik}}{\Rightarrow} l_n(\theta) &= \log p(X_{t_0};\theta) + \sum_{i=1}^{n} \log p(X_{t_i}|X_{t_{i-1}};\theta) \\ \overset{\text{Score function}}{\Rightarrow} \partial_\theta l_n(\theta) &= [\partial_\theta p(X_{t_0};\theta)]/p(X_{t_0};\theta) + \sum_{i=1}^{n} [\partial_\theta p(X_{t_i}|X_{t_{i-1}};\theta)]/p(X_{t_i}|X_{t_{i-1}};\theta). \end{aligned}$$

Likelihood function: $L_n(\theta) = \prod_{i=1}^{n} p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) p(X_{t_0}; \theta), \qquad \Delta_i := t_i - t_{i-1}$

with $p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) = p(X_{t_i} | X_{t_{i-1}}; \theta)$ transition density, i.e. probability density function of the conditional distribution of $X_{t+\Delta}$ give that $X_t = x$. (Also conditional density of $X_{t+s+\Delta}$ given $X_{t+s} = x$ since the process is time homogeneous).

* If $p(\Delta, x, y; \theta)$ is

known $\Rightarrow$ MLE as before, either analytically or numerically, i.e.
1. compute the log-likelihood $l_n(\theta)$. 2. solve $\partial_\theta l_n(\theta) = 0$.

unknown (or nasty) What to do?

1st Approach Approximate $p(\Delta, x, y; \theta)$ with something we like, e.g. normal transition density.
$\Rightarrow$ Derive the approximate likelihood function.
$\Rightarrow$ Derive an estimator as before, setting the score function to zero.

2nd Approach Consider other suitable function instead of the likelihood (e.g. Least squares estimation).

Alternative: Unknown quantities can be also approximated by simulations!

Example: Ornstein-Uhlenbeck (OU) process

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t$$

where $\beta > 0, \alpha \in \mathbb{R}, \sigma > 0$ and $X_0 = x_0$.

Solution (Using Ito's formula):

$$X_t = \alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s.$$

Sum of deterministic terms and an integral of a deterministic function with respect to a Wiener process with normally distributed increments $\Rightarrow$ The distribution is normal.

The conditional expectation is

$$\mathbb{E}[X_t | X_0 = x_0] = \mathbb{E}\left[\alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s\right] =$$

The conditional variance is

$$\mathrm{Var}(X_t | X_0 = x_0) = \mathbb{E}\left[\left(\sigma \int_0^t e^{-\beta(t-s)} dW_s\right)^2\right] \overset{\text{Ito's isom.}}{=}$$

Thus, $(X_t | X_0 = x_0) \sim N\left(\alpha + (x_0 - \alpha)e^{-\beta t}, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$.

Asymptotically $X_t \sim$ (or always if $X_0 \sim N(\alpha, \frac{\sigma^2}{2\beta})$).

Example: Ornstein-Uhlenbeck (OU) process

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t$$

where $\beta > 0, \alpha \in \mathbb{R}, \sigma > 0$ and $X_0 = x_0$.

Solution (Using Ito's formula):

$$X_t = \alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s.$$

Sum of deterministic terms and an integral of a deterministic function with respect to a Wiener process with normally distributed increments
$\Rightarrow$ The distribution is normal.

The conditional expectation is

$$\mathbb{E}[X_t | X_0 = x_0] = \mathbb{E}\left[\alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s\right] = \alpha + (x_0 - \alpha)e^{-\beta t}.$$

The conditional variance is

$$\mathrm{Var}(X_t | X_0 = x_0) = \mathbb{E}\left[\left(\sigma \int_0^t e^{-\beta(t-s)} dW_s\right)^2\right] \overset{\text{Ito's isom.}}{=}$$

Thus, $(X_t | X_0 = x_0) \sim N\left(\alpha + (x_0 - \alpha)e^{-\beta t}, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$.

Asymptotically $X_t \sim$ (or always if $X_0 \sim N(\alpha, \frac{\sigma^2}{2\beta})$).

Example: Ornstein-Uhlenbeck (OU) process

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t$$

where $\beta > 0, \alpha \in \mathbb{R}, \sigma > 0$ and $X_0 = x_0$.

Solution (Using Ito's formula):

$$X_t = \alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s.$$

Sum of deterministic terms and an integral of a deterministic function with respect to a Wiener process with normally distributed increments
$\Rightarrow$ The distribution is normal.

The conditional expectation is

$$\mathbb{E}[X_t | X_0 = x_0] = \mathbb{E}\left[\alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s\right] = \alpha + (x_0 - \alpha)e^{-\beta t}.$$

The conditional variance is

$$\mathrm{Var}(X_t | X_0 = x_0) = \mathbb{E}\left[\left(\sigma \int_0^t e^{-\beta(t-s)} dW_s\right)^2\right] \overset{\text{Ito's isom.}}{=} \sigma^2 \mathbb{E}\left[\int_0^t e^{-2\beta(t-s)} ds\right] = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$$

Thus, $(X_t | X_0 = x_0) \sim N\left(\alpha + (x_0 - \alpha)e^{-\beta t}, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$.

Asymptotically $X_t \sim$ (or always if $X_0 \sim N(\alpha, \frac{\sigma^2}{2\beta})$).

Example: Ornstein-Uhlenbeck (OU) process

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t$$

where $\beta > 0, \alpha \in \mathbb{R}, \sigma > 0$ and $X_0 = x_0$.

Solution (Using Ito's formula):

$$X_t = \alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s.$$

Sum of deterministic terms and an integral of a deterministic function with respect to a Wiener process with normally distributed increments
$\Rightarrow$ The distribution is normal.

The conditional expectation is

$$\mathbb{E}[X_t | X_0 = x_0] = \mathbb{E}\left[\alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s\right] = \alpha + (x_0 - \alpha)e^{-\beta t}.$$

The conditional variance is

$$\mathrm{Var}(X_t | X_0 = x_0) = \mathbb{E}\left[\left(\sigma \int_0^t e^{-\beta(t-s)} dW_s\right)^2\right] \overset{\text{Ito's isom.}}{=} \sigma^2 \mathbb{E}\left[\int_0^t e^{-2\beta(t-s)} ds\right] = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$$

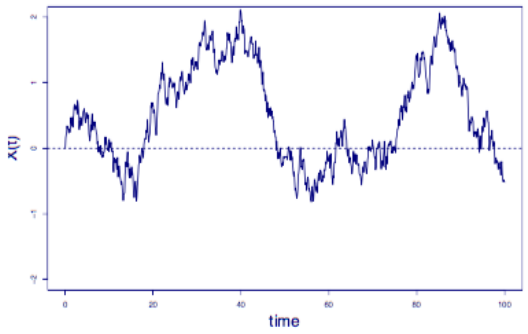
Thus, $(X_t | X_0 = x_0) \sim N\left(\alpha + (x_0 - \alpha)e^{-\beta t}, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$.

Asymptotically $X_t \sim N(\alpha, \frac{\sigma^2}{2\beta})$ (or always if $X_0 \sim N(\alpha, \frac{\sigma^2}{2\beta})$).

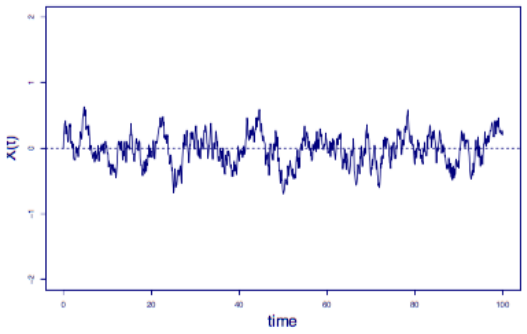
JKU JOHANNES KEPLER UNIVERSITY LINZ

Parameter interpretation in the OU process

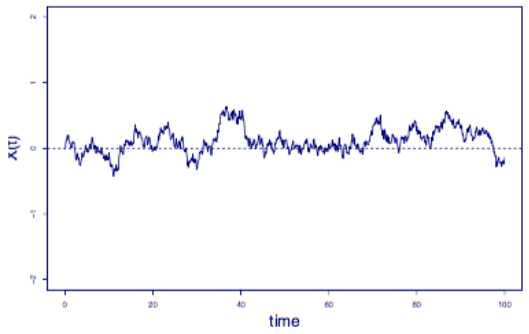
Asymptotically $X_t \sim N(\alpha, \frac{\sigma^2}{2\beta})$.



$\beta = 0.01, \sigma = 1$



$\beta = 0.1, \sigma = 1$



$\beta = 0.01, \sigma = 0.5$

β: how "strongly" the system reacts to perturbation.

(the "decay rate"or "growth-rate").

σ^2: the variation or the size of the noise.

α: the asymptotic mean.

Example: Least squares estimation for OU.

* Consider the OU process (for simplicity with only one parameter):

$$dX_t = -\theta X_t + W_t, \qquad (\text{i.e. } \beta = \theta, \alpha = 0, \sigma = 1).$$

* Data: $X_{t_1}, \ldots, X_{t_n}$ with $\Delta = t_i - t_{i-1} \Rightarrow$ Data: $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$.

* For these parameter values, we saw before that

$$\begin{aligned} (X_{t_i}|X_{t_{i-1}}) &= (X_{i\Delta}|X_{(i-1)\Delta}) \sim N\left(X_{(i-1)\Delta}e^{-\theta\Delta}, \frac{1-e^{-2\theta\Delta}}{2\theta}\right) \\ \Rightarrow \qquad & \mathbb{E}_\theta(X_{i\Delta}|X_{(i-1)\Delta}) = xe^{\theta\Delta} \qquad \mathrm{Var}_\theta(X_{i\Delta}|X_{(i-1)\Delta}) = \frac{1-e^{-2\theta\Delta}}{2\theta}. \end{aligned}$$

Idea: Approximate $X_{i\Delta}$ with its conditional mean.

$\Rightarrow$ We can find an estimator for θ by **minimizing**

$$K_n(\theta) = \sum_{i=1}^{n} \left(X_{i\Delta} - e^{\theta\Delta}X_{(i-1)\Delta}\right)^2$$

$\Rightarrow$ Least squares estimation or minimum contrast estimation.

Solve $\frac{d}{d\theta}K_n(\theta) = 0$ gives us

$$\hat{\theta} = -\frac{1}{\Delta}\log\left(\frac{\sum_{i=1}^n X_{(i-1)\Delta}X_{i\Delta}}{\sum_{i=1}^n X_{(i-1)\Delta}^2}\right),$$

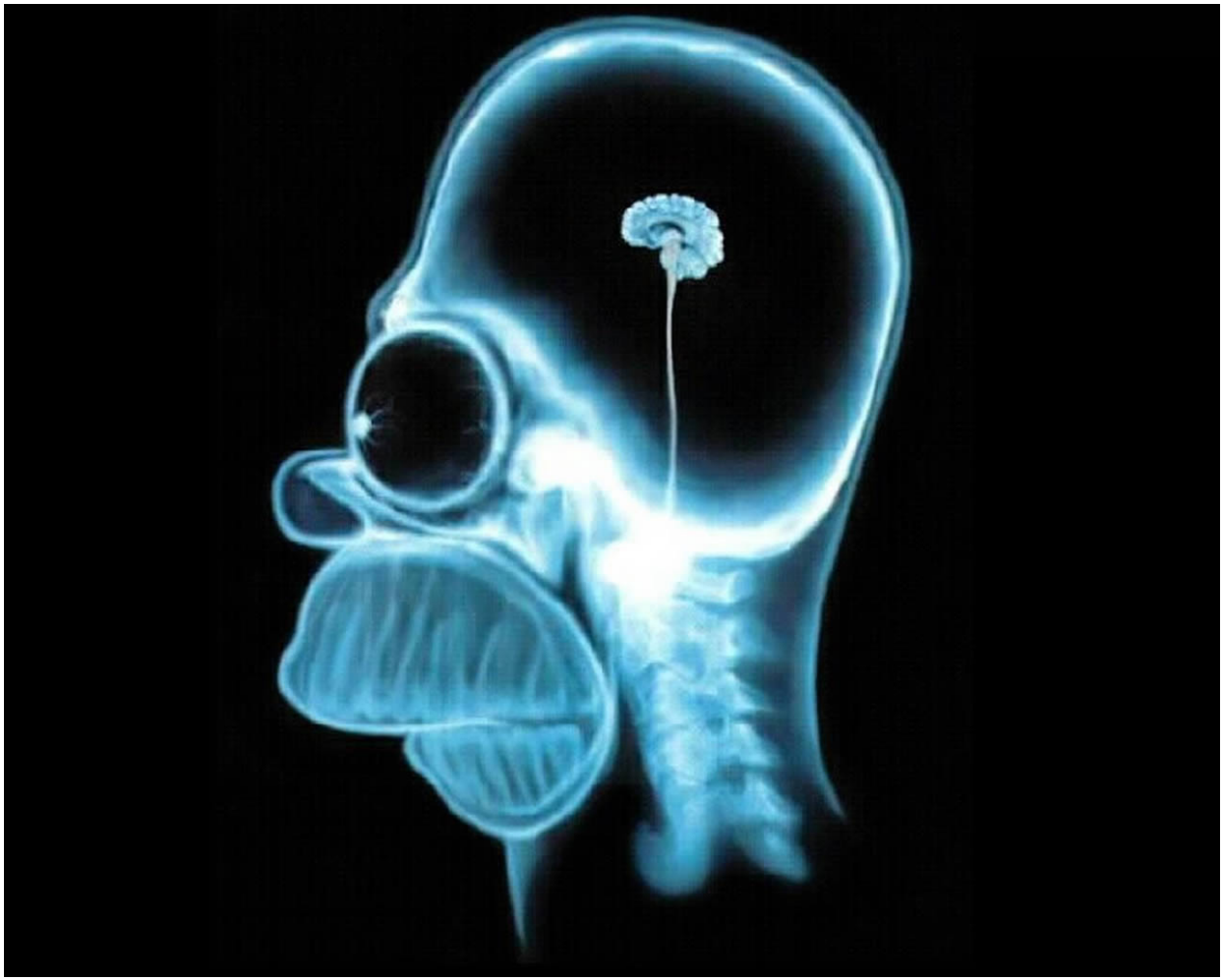
provided that $\sum_{i=1}^n X_{(i-1)\Delta}X_{i\Delta} > 0$.

* **Note**: Since $p(X_{i\Delta}|X_{(i-1)\Delta};\theta)$ is known, one could also perform MLE.

How does a human brain look like?

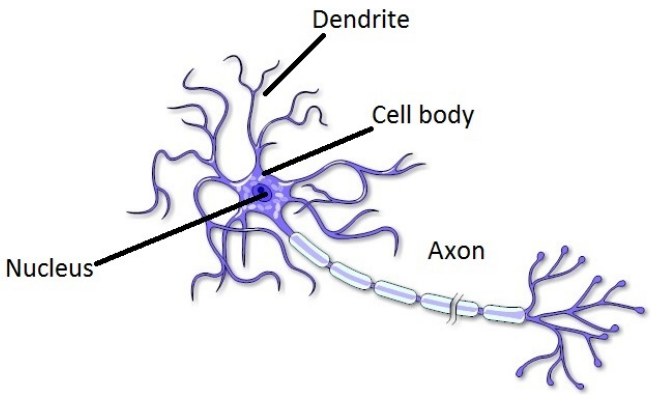
What is going on in our brain?
How does a single neuron look like?

How does a human brain look like?



What is going on in our brain?
How does a single neuron look like?

Structure of a neuron and neuron activity



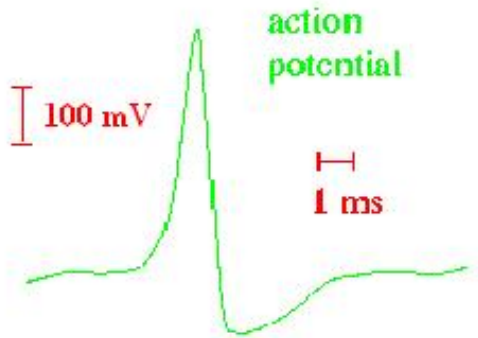
Three functionally distinct parts:

1. dendrites: input device collecting signals
2. soma (cell body): central processing unit
3. axon: output device

Connections between neurons are due to synapses (presynaptic, postsynaptic): chemical neurotrasmitters

* Two different types of measurements of the neuronal activity:
  1. Intracellular recordings. Type of data: sampling from the membrane potential V_t.
  2. Extracellular recordings. Type of data: spike times.

Action Potential or Spike

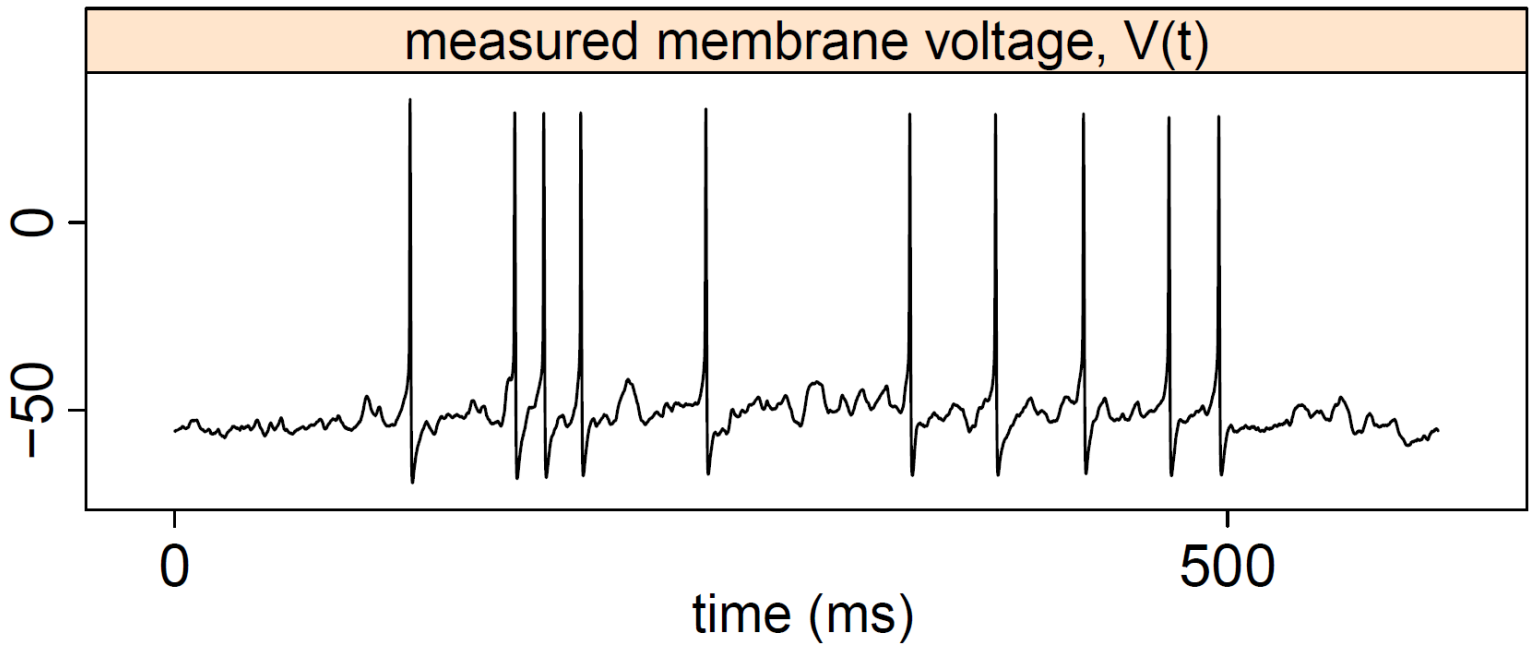


An action potential (or spike) is a short voltage pulse of 1-2 ms of duration and around $100mV$ of amplitude. The action potentials of a specific neuron have a characteristic shape.

Intracellular recording

* Intracellular recording: Difference between the membrane potential (voltage) V_t internal to the cell and the surrounding. In absence of inputs, it spontaneously decays toward an initial voltage $V(0)$ of about $-65mV$.

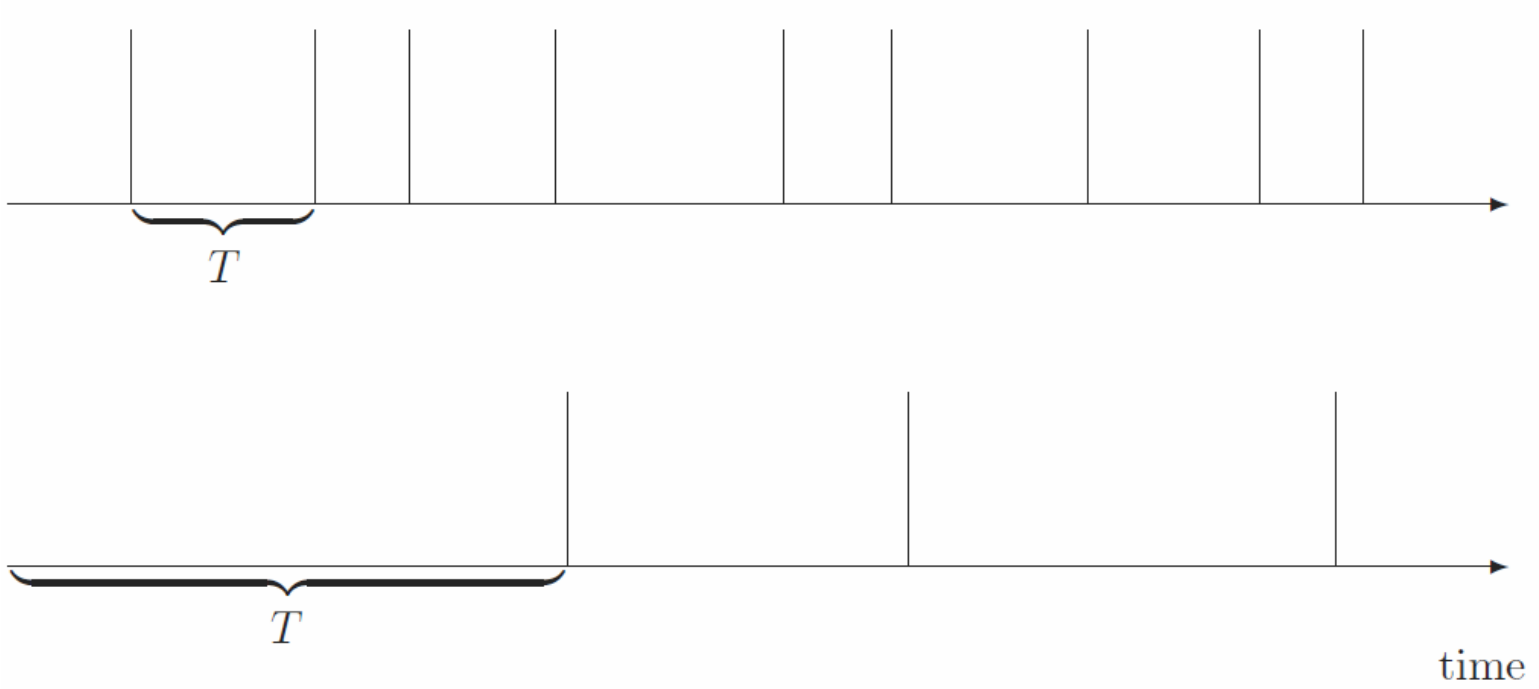
* Data: $V_{t_1}, \ldots, V_{t_n}$ from the membrane potential V_t.



Observations of the membrane potential in a spinal motoneuron of an adult red-eared turtle during 600 ms measured every 0.1 ms. Data from Berg Laboratory (Berg et al.2007)

Extracellular recording (Spike train)

We observe the spikes.



Two alternative approaches:

- **Rate coding** $\Rightarrow$ Working with point processes.
- **Temporal coding** $\Rightarrow$ Working with Interspike-Intervals (ISIs).

Relevant questions at a single-neuron level

Relevant questions at a single-neuron level

- How to model the single neuronal behaviour?
  - Models driven by ODEs (Lapique, Hodgkin-Huxley, FitzHugh-Nagumo, Morris Lecar);
  - Models driven by SDEs (Jump processes, LIF models, FN/ML models with noise);
  - Models driven by SDEs with adaptations (MAT models).

Relevant questions at a single-neuron level

- How to model the single neuronal behaviour?
  - Models driven by ODEs (Lapique, Hodgkin-Huxley, FitzHugh-Nagumo, Morris Lecar);
  - Models driven by SDEs (Jump processes, LIF models, FN/ML models with noise);
  - Models driven by SDEs with adaptations (MAT models).
- How to model/reproduce/predict spikes?
  - HH type of models: firing through stable limit cycles
  - LIF models: hitting times to a (“artificial”) threshold

Relevant questions at a single-neuron level

- How to model the single neuronal behaviour?
  - Models driven by ODEs (Lapique, Hodgkin-Huxley, FitzHugh-Nagumo, Morris Lecar);
  - Models driven by SDEs (Jump processes, LIF models, FN/ML models with noise);
  - Models driven by SDEs with adaptations (MAT models).
- How to model/reproduce/predict spikes?
  - HH type of models: firing through stable limit cycles
  - LIF models: hitting times to a ("artificial") threshold
- Are single neuron models any useful? Why?
  - Investigate the effect of external stimuli
  - Investigate relationships between inputs and output signal $\Rightarrow$ Information measures (entropy, Fisher information)
  - Investigate the role of noise in neural coding (**noise may provide a better decoding!**)

Relevant questions at a single-neuron level

- How to model the single neuronal behaviour?
  - Models driven by ODEs (Lapique, Hodgkin-Huxley, FitzHugh-Nagumo, Morris Lecar);
  - Models driven by SDEs (Jump processes, LIF models, FN/ML models with noise);
  - Models driven by SDEs with adaptations (MAT models).
- How to model/reproduce/predict spikes?
  - HH type of models: firing through stable limit cycles
  - LIF models: hitting times to a ("artificial") threshold
- Are single neuron models any useful? Why?
  - Investigate the effect of external stimuli
  - Investigate relationships between inputs and output signal $\Rightarrow$ Information measures (entropy, Fisher information)
  - Investigate the role of noise in neural coding (**noise may provide a better decoding!**)
- What information can we extract from data?

Hodgkin-Huxley model & friends

Detailed model taking opening/closing of ion channels into account. A HH model is a system of 4 ODEs (V_t: membrane voltage) given by

$$C_m \frac{dV_t}{dt} = -g_K n^4 (V_t - E_K) - g_{Na} m^3 h (V_t - E_{Na}) - g_L (V_t - E_L) + I$$

Gating variables:

$$\dot{m} = \alpha_m(V_t)(1-m) - \beta_m(V_t)m$$

$$\dot{h} = \alpha_h(V_t)(1-h) - \beta_h(V_t)h$$

$$\dot{n} = \alpha_n(V_t)(1-n) - \beta_n(V_t)n$$

A.L. Hodgkin and A. F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, J. Physiol., 117(4): 500-544, 1952. $\Rightarrow$ Nobel Prize in Physiology or Medicine in 1963.

Important simplified versions: Morris-Lecar model, Fitzhugh-Nagumo model.

* These models create a feedback system capable of producing **spikes** (through limit cycle).

Inference from intracellular recording

(for conductance/current based models as Hodgkin-Huxley & friends)

* Data: Intracellular recording $V_{t_1}, \ldots, V_{t_n}$.
* Goal: Estimation of parameters appearing in both V_t and W_t.
* Unrealistic scenario: Assume W_t to be observed.
    * Why do we do that?
    * How do we do that? MLE.
* Realistic scenario:

S. Ditlevsen, A. Samson. Parameter estimation in neuronal stochastic differential equation models from intracellular recordings of membrane potentials in single neurons: a Review. J. Soc. Franc. Stat., 2015

Inference from intracellular recording

(for conductance/current based models as Hodgkin-Huxley & friends)

* Data: Intracellular recording $V_{t_1}, \ldots, V_{t_n}$.

* Goal: Estimation of parameters appearing in both V_t and W_t.

* Unrealistic scenario: Assume W_t to be observed.
    * Why do we do that?
    * How do we do that? MLE.

* Realistic scenario: W_t is not observed!

$\Rightarrow$ **Estimation from partially observed models**:
    * If lucky: Hidden-Markov models.
    * If not: particle filter for (non)-autonomous systems, stochastic approximation methods, EM algorithm (with particle filters), etc.

S. Ditlevsen, A. Samson. Parameter estimation in neuronal stochastic differential equation models from intracellular recordings of membrane potentials in single neurons: a Review. J. Soc. Franc. Stat., 2015

Hidden Markov models (Hints)

Define the history up to time t:

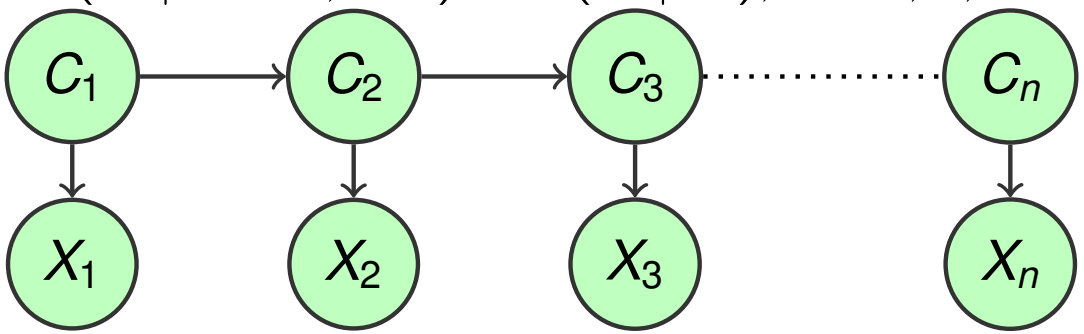
$$\mathbf{C}^{(t)} = (C_1, C_2, \ldots, C_t) \qquad \mathbf{X}^{(t)} = (X_1, X_2, \ldots, X_t)$$

The model consists of two parts:
Unobserved **parameter process**: $\{C_t : t \in \mathbb{N}\}$ satisfying the Markov property:

$$P(C_t \mid \mathbf{C}^{(t-1)}) = P(C_t \mid C_{t-1}), t = 2, 3, \ldots$$

Observed **state-dependent process**: $\{X_t : t \in \mathbb{N}\}$ satisfying conditional independence given the state of the MC:

$$P(X_t \mid \mathbf{X}^{(t-1)}, \mathbf{C}^{(t)}) = P(X_t \mid C_t), t = 1, 2, \ldots$$



(Some of the) Goals when dealing with HMMs

* Forecasting for X_{t+n} given $\mathbf{X}^{(t)}$. * Local decoding of C_t given $\mathbf{X}^{(t)}$.
* Global decoding of $(C_1, \ldots, C_t)$ given $\mathbf{X}^{(t)}$.

JKU JOHANNES KEPLER UNIVERSITY LINZ

Common method: EM algorithm

* Iterative method to perform MLE with missing data.

* Useful when the complete data likelihood is easier to maximized then the observed data likelihood.

Observed data: $X^{(N)} : (X_1, \ldots, X_N)$.

Unobserved (missing, hidden) data: $C^{(N)} : (C_1, \ldots, C_N)$.

Complete data: $(X^{(N)}, C^{(N)})$.

Observed data likelihood: $p(X_1, \ldots, X_N; \theta)$.

Complete data likelihood (CDL) : $p(X_1, C_1, \ldots, X_N, C_N; \theta)$.

Idea: Choose an initial value for the parameter θ.

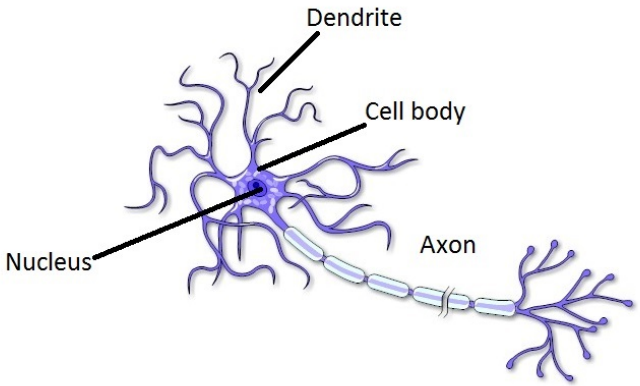
E-step: Conditional expectations of the missing data given the observations and the current value of θ.

M-step: Update θ by maximizing the CDL with respect to θ, with the missing data replaced by their conditional expectations, calculated in the E-step.

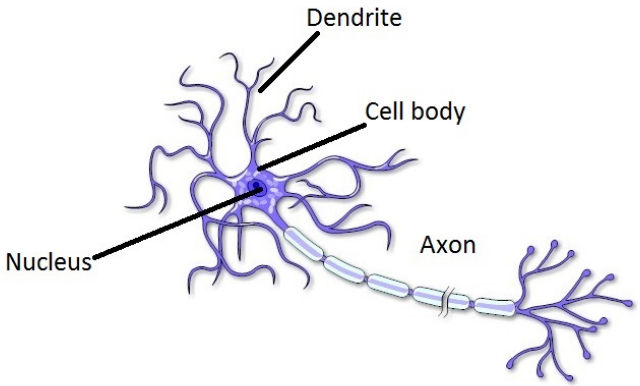
Repeat until convergence, e.g. change in θ is less than some chosen value. $\Rightarrow$ the obtained $\hat{\theta}$ is then a stationary point of the likelihood - not necessarily a global maximum.

JOHANNES KEPLER UNIVERSITY LINZ

Leaky Integrate-and-Fire (LIF) models

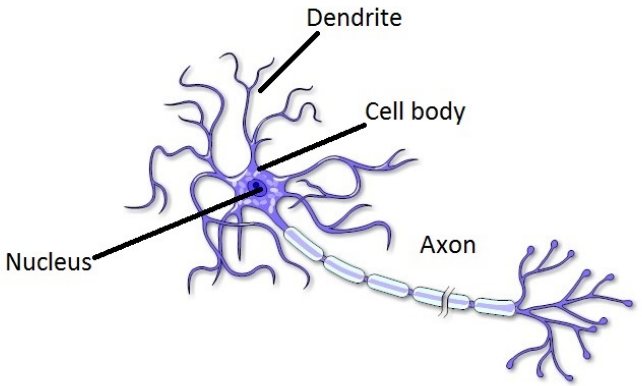


Leaky Integrate-and-Fire (LIF) models



$\Longrightarrow$ ●

Leaky Integrate-and-Fire (LIF) models



$\Longrightarrow$ ●

The model:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW(t); \quad X_0 = x_0$$

X_t: membrane potential at time t after a spike.

x_0: initial voltage membrane potential at time t after a spike.

A spike is produced when the membrane voltage X_t exceed a firing threshold $S_t > X_0 = 0$.

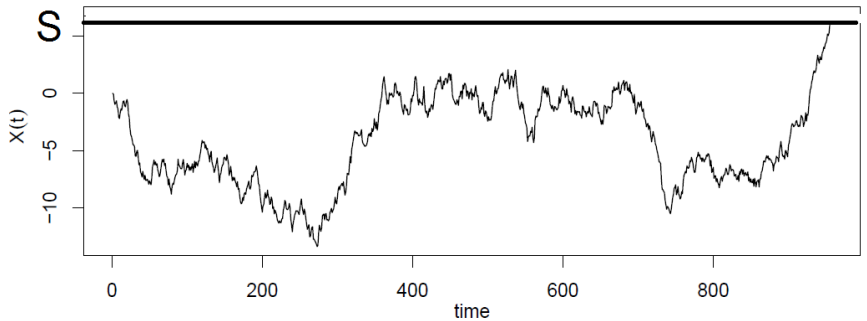
After firing, the process is reset to x_0. The interspike interval T is identified with the first passage time (FPT) of the threshold:

$$T = \inf\{t > 0 : X_t \geq S_t\}$$

KEY PROPERTY: **Interspike intervals are iid!** (Is this realistic??)

Inference from intracellular recording (for LIF)

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW(t); \qquad X_0 = x_0, \qquad \theta \in \Theta \subseteq \mathbb{R}^p.$$



* Data: Intracellular recording $X_{t_1}, \ldots, X_{t_n}$.
* Scenario: Fully (discretely) observed process.
* Goal: Estimation of θ.

* This is NOT the same scenario we saw before: the presence of the threshold changes the underlying transition density!

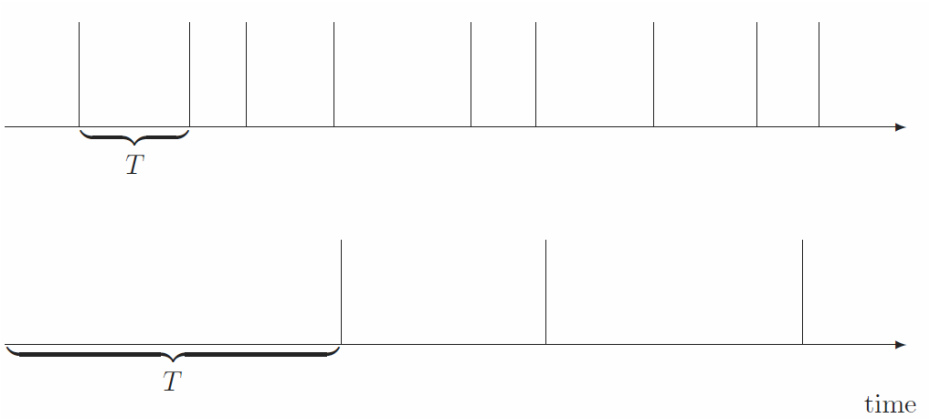
$\Rightarrow$ Perform MLE from the previous transition free density introduce **biases** in the estimators!

$\Rightarrow$ We should consider the transition density of X_t in presence of absorbing boundary.

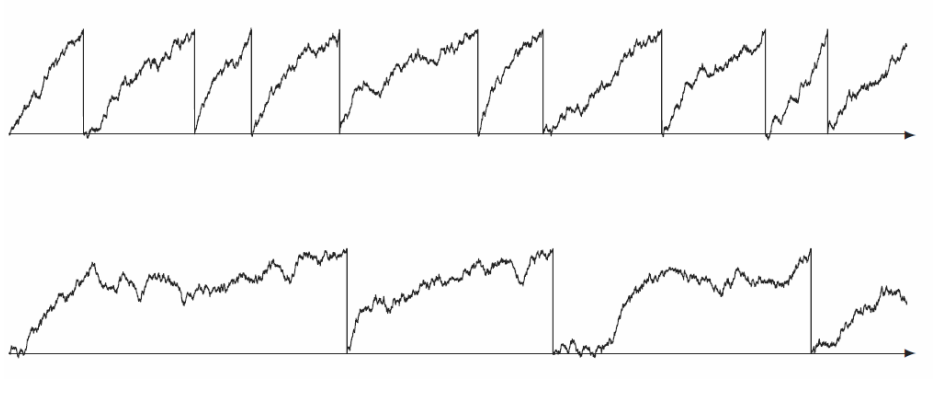
Issue: This density is often unknown! What to do?
* Numerical methods to calculate it.
* Monte-Carlo simulation based methods.

Inference from extracellular recordings (for LIF)

Simplest (and most common) hypothesis:
constant threshold $S_t = S > X_0$ (Is it realistic?)



Underlying processes



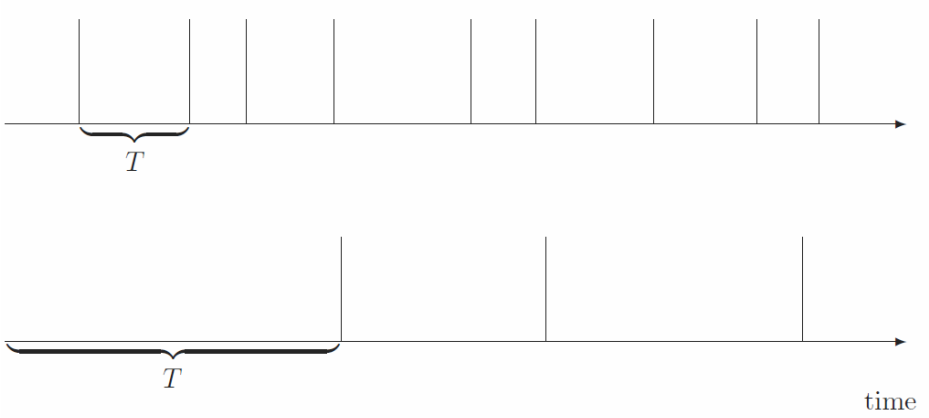
Note: There is only information on the time scale, nothing on the scale of X_t.

Interspike intervals are assumed to be iid (Is that realistic??).

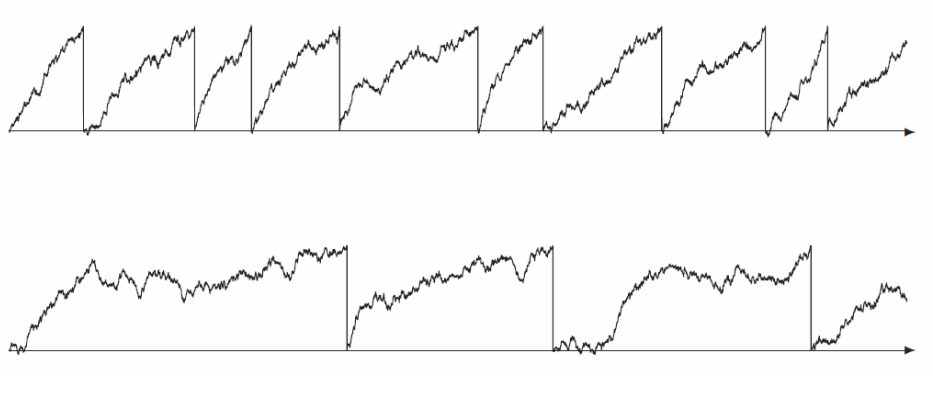
Common approaches: Maximum likelihood estimation ($p_T(t;\theta)$ is rarely known $\Rightarrow$ we compute it numerically or via simulations), Moment estimation, Laplace estimator of T, etc.

Inference from extracellular recordings (for LIF)

Simplest (and most common) hypothesis:
constant threshold $S_t = S > X_0$ (Is it realistic?)



Underlying processes



Note: There is only information on the time scale, nothing on the scale of X_t. $\Rightarrow$ Something is not identifiable in the model from these data $\Rightarrow$ something (typically the threshold S) has to be assumed known!

Interspike intervals are assumed to be iid (Is that realistic??).

Common approaches: Maximum likelihood estimation ($p_T(t;\theta)$ is rarely known $\Rightarrow$ we compute it numerically or via simulations), Moment estimation, Laplace estimator of T, etc.

Example: Brownian motion with drift

Simplest model (known in neuroscience as Perfect-Integrator)

$$dX_t = \mu dt + \sigma dW_t, \qquad \mu > 0, \sigma > 0; \qquad X_0 = 0 < S$$

The FPT distribution of T is known (one of the few cases!): T is Inverse Gaussian with scale parameter S/μ and shape parameter S^2/σ^2, with mean, variance and density:

$$T \sim IG\left(\frac{S}{\mu}, \frac{S^2}{\sigma^2}\right), \qquad \mathbb{E}[T] = \frac{S}{\mu}, \qquad \mathrm{Var}(T) = \frac{S\sigma^2}{\mu^2},$$

$$p_T(t;\theta) = \frac{S}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(S-\mu t)^2}{2\sigma^2 t}\right) \tag{1}$$

Note: **Identifiability issue** ($\tilde{S} = aS, \tilde{\mu} = a\mu, \tilde{\sigma}^2 = a^2\sigma^2, a > 0$ yield the same distribution!) $\Rightarrow$ We cannot estimate S, μ, σ^2 at the same time $\Rightarrow$ We assume S to be known.

$\Rightarrow$ Parameters to estimate: μ and σ^2.

Data: $T_1, \ldots, T_n$ FPTs of X_t through S. Important: iid rvs!

Possible approaches: MLE and Moment Estimation

MAXIMUM LIKELIHOOD ESTIMATORS

$$L_n(\theta) \overset{iiddata}{=} \prod_{i=1}^{n} p_T(T_i;\theta) \overset{(1)}{=} \prod_{i=1}^{n} \left(\frac{S}{\sqrt{2\pi\sigma^2 T_i^3}} \right) \exp\left(-\frac{(S-\mu T_i)^2}{2\sigma^2 T_i} \right)$$

$$\Rightarrow l_n(\theta) = \sum_{i=1}^{n} \log p_T(T_i) = \sum_{i=1}^{n} \log \left(\frac{S}{\sqrt{2\pi\sigma^2 T_i^3}} \right) - \sum_{i=1}^{n} \frac{(S-\mu T_i)^2}{2\sigma^2 T_i}$$

Score functions:

$$\Rightarrow \quad \partial_\mu l_n(\theta) = \sum_{i=1}^{n} \frac{(S-\mu T_i)}{\sigma^2}; \qquad \partial_{\sigma^2} l_n(\theta) = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(S-\mu T_i)^2}{2(\sigma^2)^2 T_i}$$

Solving $\partial_\mu l_n(\theta) = 0$ and $\partial_{\sigma^2} l_n(\theta) = 0$ in μ and σ^2, yields

$$\hat{\mu}_{\text{MLE}} = \frac{S}{\bar{T}}; \qquad \hat{\sigma}^2_{\text{MLE}} = S^2 \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} - \frac{1}{\bar{T}} \right), \qquad \text{with} \quad \bar{T} = \frac{1}{n} \sum_{i=1}^{n} T_i \quad \text{(sample mean)}.$$

MOMENT ESTIMATORS

Idea: Obtain estimators by equalizing true moments and sample moments. $\Rightarrow$ The moment estimators $\hat{\mu}_{ME}$ and $\hat{\sigma}^2_{ME}$ are obtained by solving a system of 2 equations in 2 unknown (true moments = sample moments):

$$\mathbb{E}[T] = \frac{S}{\mu} = \bar{T}; \qquad \text{Var}(T) = \frac{S\sigma^2}{\mu^3} = \frac{1}{n-1} \sum_{i=1}^{n} (T_i - \bar{T})^2$$

$$\Rightarrow \quad \hat{\mu}_{\text{ME}} = \frac{S}{\bar{T}} = \hat{\mu}_{\text{MLE}} \qquad \hat{\sigma}^2_{\text{ME}} = \frac{\hat{\mu}^3}{S} \frac{1}{n-1} \sum_{i=1}^{n} (T_i - \bar{T})^2 \neq \sigma^2_{\text{MLE}}.$$

JOHANNES KEPLER UNIVERSITY LINZ

Next time you deal with data

Important initial questions

1. Are the data “good”enough? Are they iid?
2. What question (do I want to)/can I answer from those data?
3. What can be a good model for the available dataset?

Once the previous questions have been answered, new questions arise:

1. How can I estimate the relevant parameters?
   $\Rightarrow$ Develop reliable statistical methods.
2. How much can I trust my conclusions?
   $\Rightarrow$ Use simulations to get simulated/asymptotic errors/confidence intervals for the true parameters.
3. Is the chosen model able to fit the data? $\Rightarrow$ Model validation.
4. How should I choose between two possible competing models?
   $\Rightarrow$ Model selection, e.g. Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) .