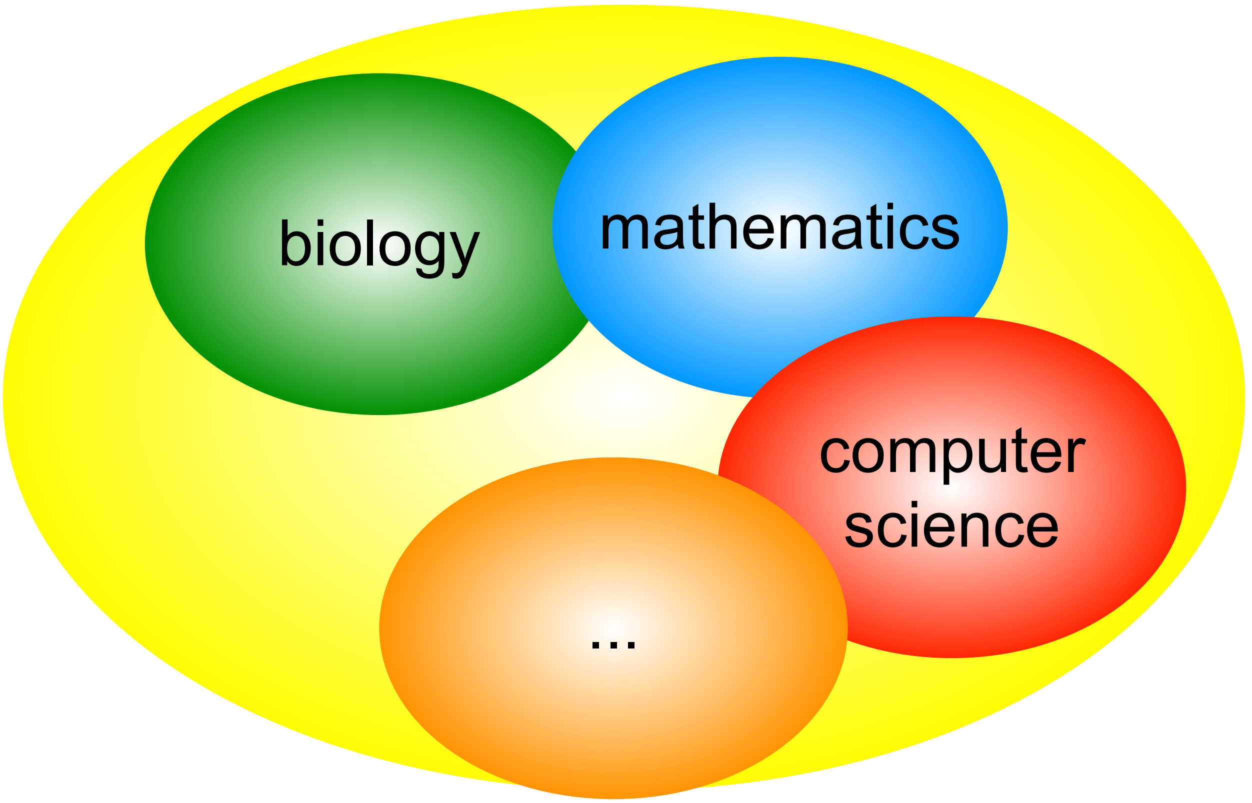
Centre for the Advanced Study
of Collective Behaviour

Universität
Konstanz

Probabilistic Modelling for Computer Scientists



Tatjana Petrov
February 2nd 2021,
Lecture 8: Statistics Refresher
(literature sources: M.Baron book, Blitzsten/Wang lecture notes)

Lecture Goals

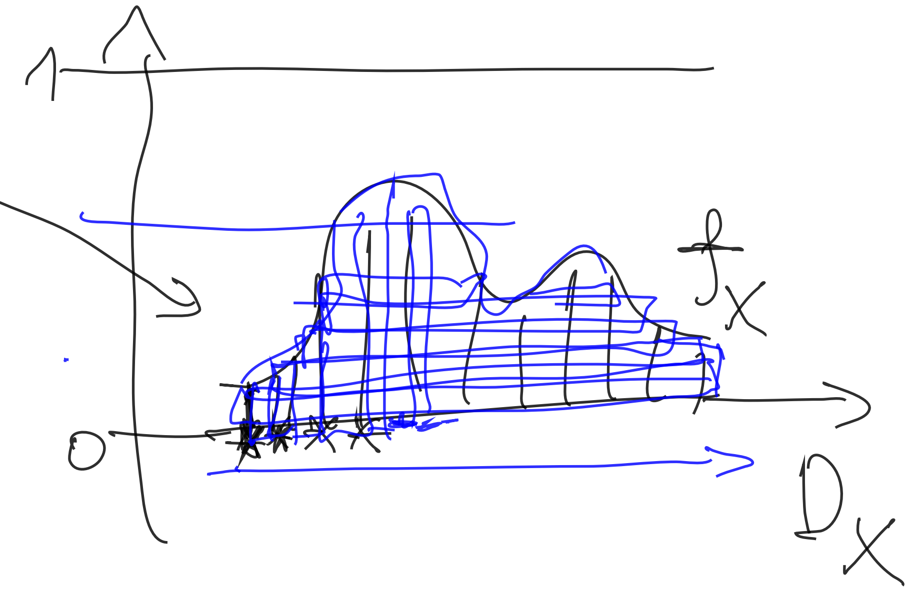
1. **Motivating example**

2. **Law of large numbers & central limit theorem**
    1. inequalities: Markov, Chebyshev
    2. LLN: proof for the weak version
    3. CLN: proof sketch and example

3. **Biochemical reaction networks**

Warm-up

Discrete	Continuous
$\mathbf{E}(X) = \sum_x x P(x)$	$\mathbf{E}(X) = \int x f(x) dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \int (x-\mu)^2 f(x) dx$ $= \int x^2 f(x) dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) P(x,y)$ $= \sum_x \sum_y (xy) P(x,y) - \mu_x \mu_y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \iint (x-\mu_X)(y-\mu_Y) f(x,y)\, dx\, dy$ $= \iint (xy) f(x,y)\, dx\, dy - \mu_x \mu_y$



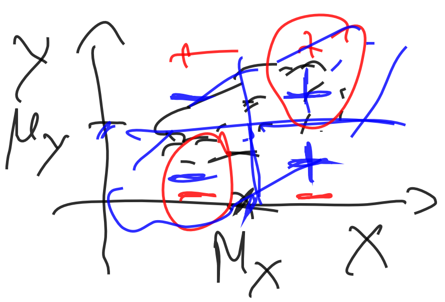
$$E(X) = \sum_{x \in D_X} x P(x) \qquad \mu := E(X)$$

$$Var(X) = E((X-\mu)^2) = \sum_{x\in D_X} (x-\mu)^2 P(x) = \sum_{x \in D_X} (x^2 - 2x\mu + \mu^2) P(x)$$

$$= \sum_{x\in D_X} x^2 P(x) - \sum_{x\in D_X} 2x\mu P(x) + \sum_{x \in D_X} \mu^2 P(x)$$

$$= E(X^2) - 2(E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2$$

$$= E(X^2) - \mu^2$$



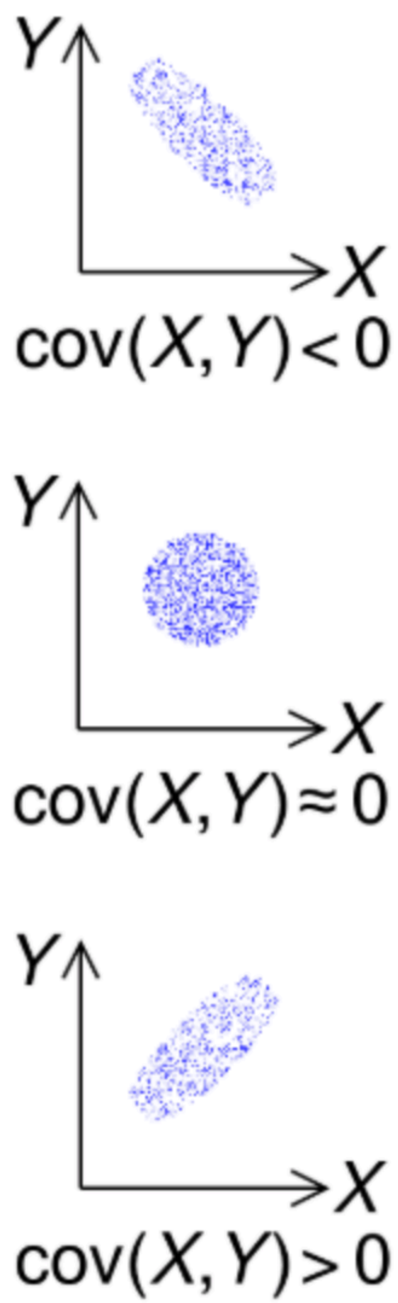
$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY) - E(X)E(Y)$$

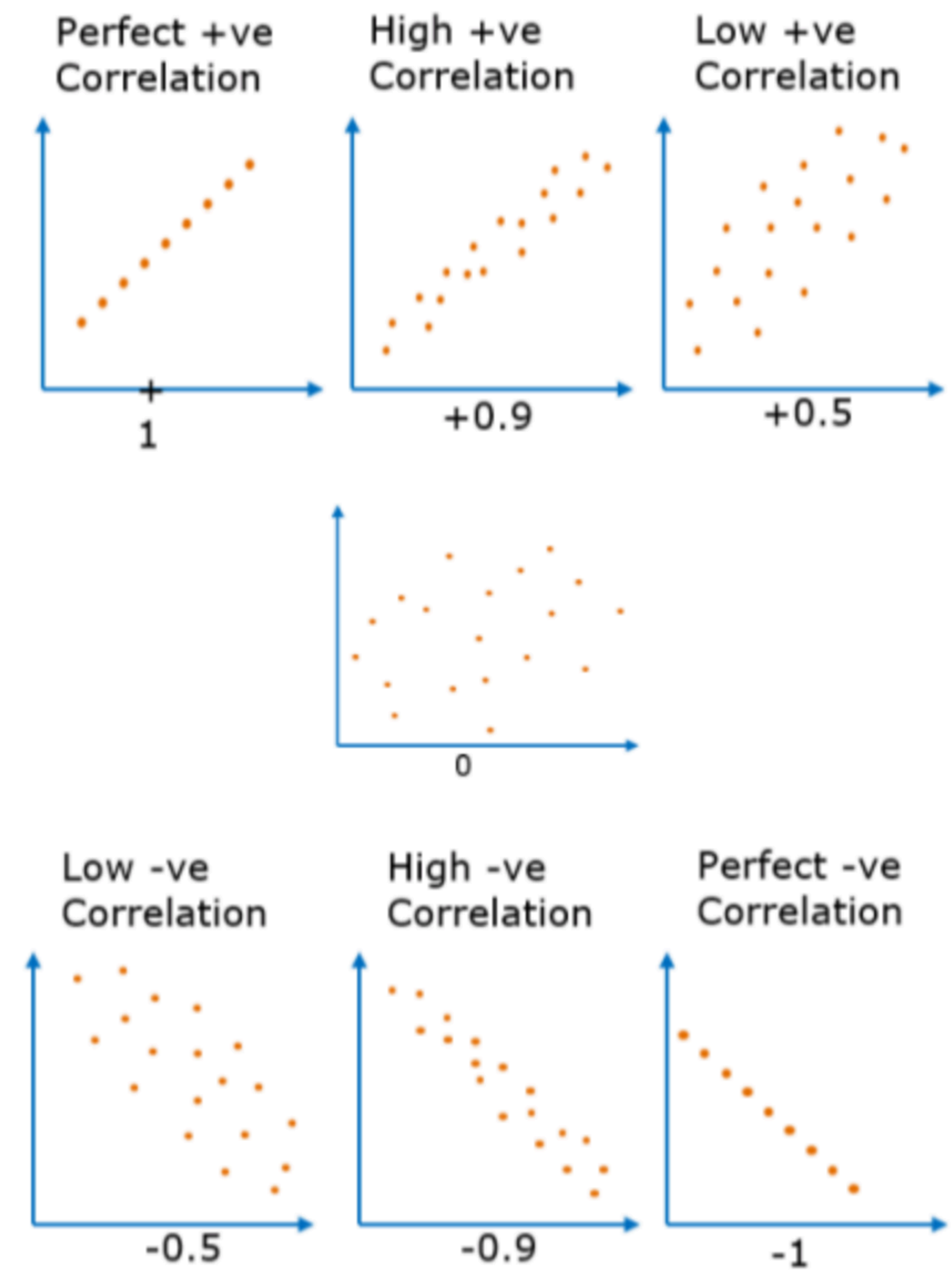
Warm-up

$$corr(X,Y) := \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$\sigma_X := \sqrt{Var(X)}$$

$$\sigma_Y := \sqrt{Var(Y)}$$





Warm-up

$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$ $\quad (10X + 2) : \begin{pmatrix} 13 & 14 & 15 & 16 & 17 & 18 \end{pmatrix}$

$c = 2$

Properties of variances and covariances

$$\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}(aX + bY, cZ + dW) = ac\,\mathrm{Cov}(X,Z) + ad\,\mathrm{Cov}(X,W) + bc\,\mathrm{Cov}(Y,Z) + bd\,\mathrm{Cov}(Y,W)$$

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$$

$$\rho(X,Y) = \rho(Y,X)$$

In particular,

$$\mathrm{Var}(aX + \cancel{b}) = a^2\,\mathrm{Var}(X)$$

$$\mathrm{Cov}(aX + \cancel{b}, cY + \cancel{d}) = ac\,\mathrm{Cov}(X,Y)$$

$$\rho(aX + b, cY + d) = \rho(X,Y)$$

For independent X and Y,

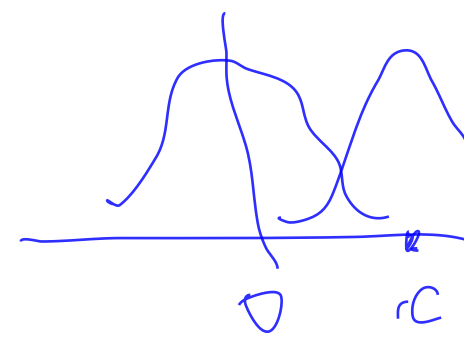
$$\mathrm{Cov}(X,Y) = 0$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

$a = b = 1$

$X \sim \mathcal{N}(0,1)$

$X + c$

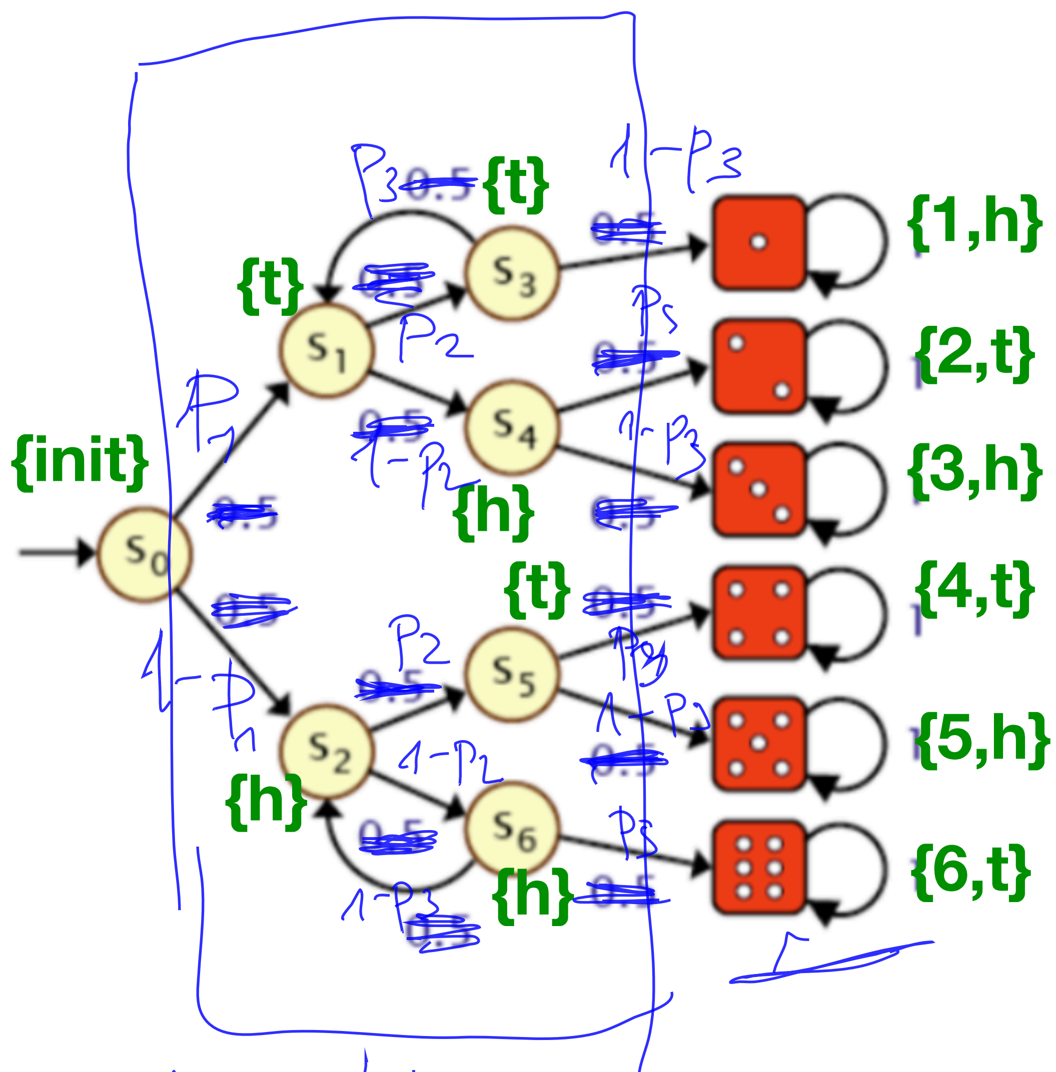


$a, b, c \in \mathbb{R}$

X, Y are RV's

Motivating example

Knuth's 6-sided die



N samples

1) statistical model-checking

Monte Carlo

$P(F\{3\}) \leq \frac{1}{6}$

P('yield 3 or 1 with max. 5 tails')

2) parameter inference

Is $p_1 = p_2 = p_3$ a fair coin?

1 2 3 4 5 6

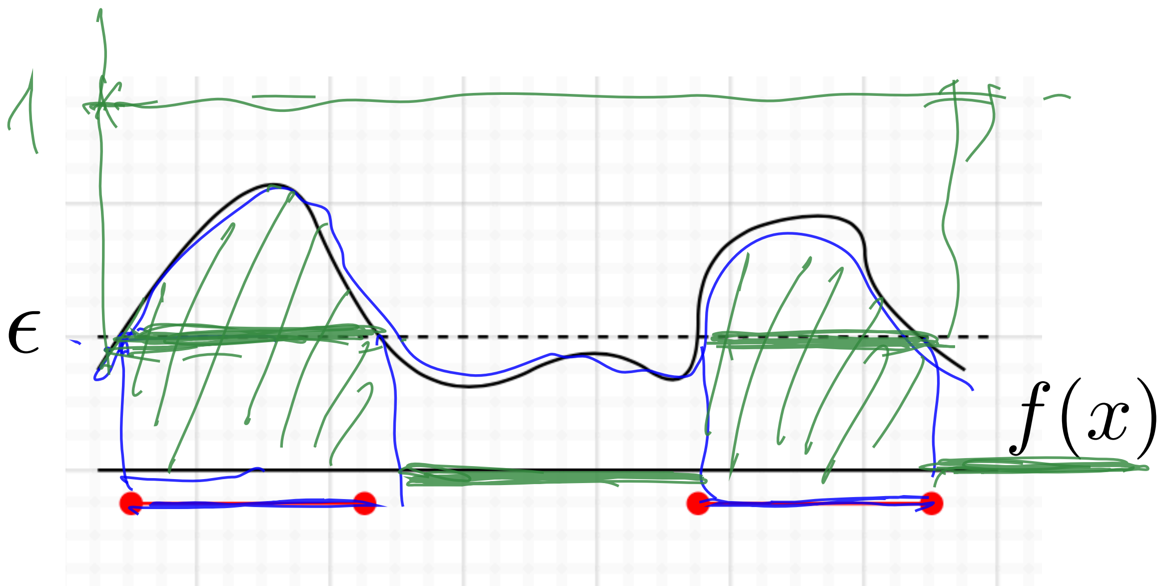
Markov Inequality

$$\mu(\{x \in X | f(x) \geq \epsilon\}) \leq \frac{\int_X f(x) d\mu}{\epsilon}$$ for all $\epsilon > 0$ and μ a non-negative measure on X

$E(X) := \int_X x \, dP_X$

Special case for probability measure:

$$\boxed{P(|X| \geq \epsilon) \leq \frac{E(X)}{\epsilon}}$$



Proof.

Let $s(x) := \begin{cases} 0, & f(x) < \epsilon \\ \epsilon, & f(x) \geq \epsilon \end{cases}$

It is true that $f(x) \geq s(x) \geq 0$.

$$\int_{D_X} f(x) d\mu \geq \int_{D_X} s(x) d\mu = \epsilon \mu(\{x \in X \mid f(x) \geq \epsilon\})$$

Example 1.
In a group of 100 people, is it possible that at least 50~~%~~ are older than twice the average age?

$$P(A > \underbrace{2E(A)}_{\epsilon}) \leq \frac{E(A)}{2E(A)} = \frac{1}{2}$$

Example 2.
How many people at most have income greater than five times average?

$$P(I \geq \underbrace{5E(I)}_{\epsilon}) \leq \frac{\cancel{E(I)}}{5\cancel{E(I)}} = 0.2$$

Chebyshev Inequality

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

$$P(|X-\mu| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$ for all $\epsilon > 0$ and P a prob. measure on X

$\epsilon = k \cdot \sqrt{Var(X)}$

$\mu(\{x \mid f(x) \geq \varepsilon\}) \leq \frac{\int f(x) d\mu}{\varepsilon}$

$f(x) := (X-\mu)^2$

$P((X-\mu)^2 \geq \varepsilon^2) \leq \frac{\overbrace{\int (X-\mu)^2 x d\mu}^{Var(X)}}{\varepsilon^2}$

$\parallel$

$P(|X-\mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$

$P(X \in \mu \pm \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$

Example. Suppose we randomly select a journal article from a source with an average of 1000 words per article, with a standard deviation of 200 words. Show that the probability that it has between 600 and 1400 words (i.e. within *k* = 2 standard deviations of the mean) must be at least 75%.

By Chebyshev's inequality,

~~P(X not in (1000+-400))<=200^2/400^2=1/4~~

$E(W) = \mu_W = 1000$

$\sigma_W = 200$

$P(W \notin \mu_W \pm \underbrace{2\sigma_W}_{k}) \leq \frac{\sigma_W^2}{(2\sigma_W)^2} = \frac{1}{4}$

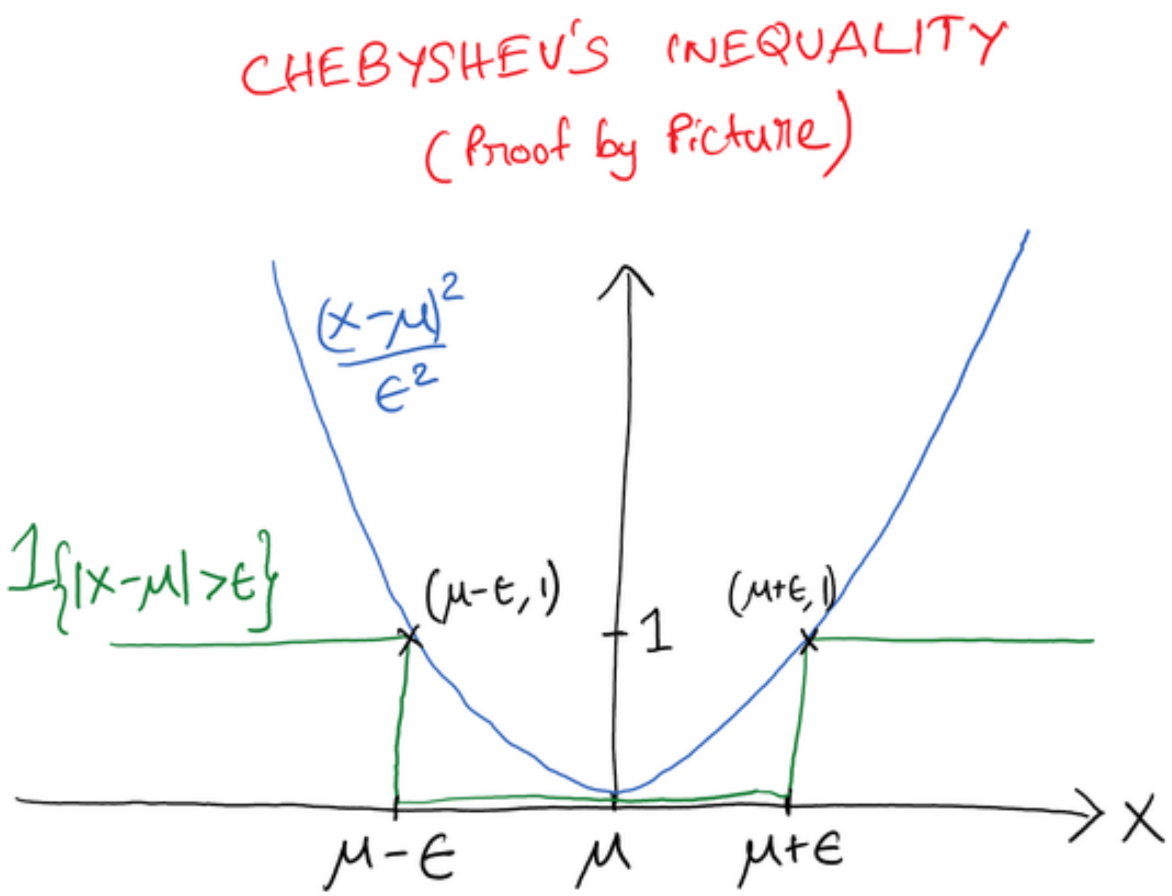
$\Downarrow$

$P(W \in \mu_W \pm 2\sigma_W) \geq \frac{3}{4}$

$P(W \in [600, 1400]) \geq .75$

Chebyshev Inequality

$$P(|X-\mu| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2} \quad \text{for all } \epsilon > 0 \text{ and } P \text{ a prob. measure on } X$$



$$\mathbb{P}[|X-\mu|>\epsilon] = \mathbb{E}\left[\mathbb{1}_{\{|X-\mu|>\epsilon\}}\right] \leq \mathbb{E}\left[\frac{(X-\mu)^2}{\epsilon^2}\right] = \frac{Var(X)}{\epsilon^2}$$

Strong and Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be i.i.d. RV with mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ sample mean of the first n RV's.

Question: *What happens the the sample mean when n gets large?*

strong LLN $\overline{X}_n \to \mu$, as $n \to \infty$ with probability 1.

$\to$ *(almost sure convergence)*

weak LLN $P(|\overline{X}_n - \mu| > \epsilon) \to 0$ for any ϵ, as $n \to \infty$.

$\to$ *(convergence in distribution)*

Strong and Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be i.i.d. RV with mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ sample mean of the first n RV's.

Question: *What happens the the sample mean when n gets large?*

strong LLN $\overline{X}_n \to \mu$, as $n \to \infty$ with probability 1.

(almost sure convergence)

weak LLN $P(|\overline{X}_n - \mu| > \epsilon) \to 0$ for any ϵ, as $n \to \infty$.

(convergence in distribution)

Example. Suppose that X~Bern(p). Then, the sample mean of the first n trials tends to p, as n goes to infinity, with probability 1.

Strong and Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be i.i.d. RV with mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ sample mean of the first n RV's.

Question: *What happens the the sample mean when n gets large?*

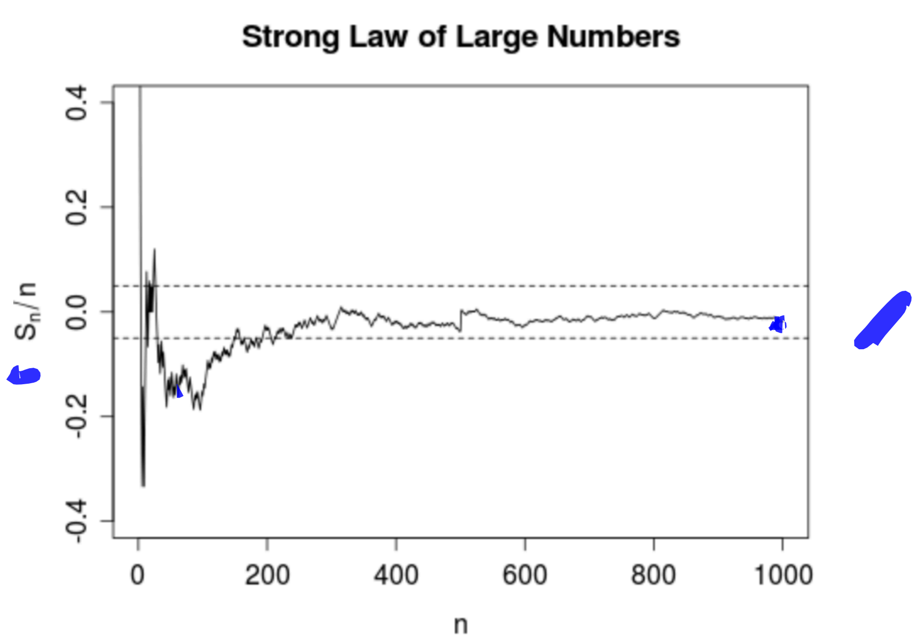
strong LLN $\overline{X}_n \to \mu$, as $n \to \infty$ with probability 1.

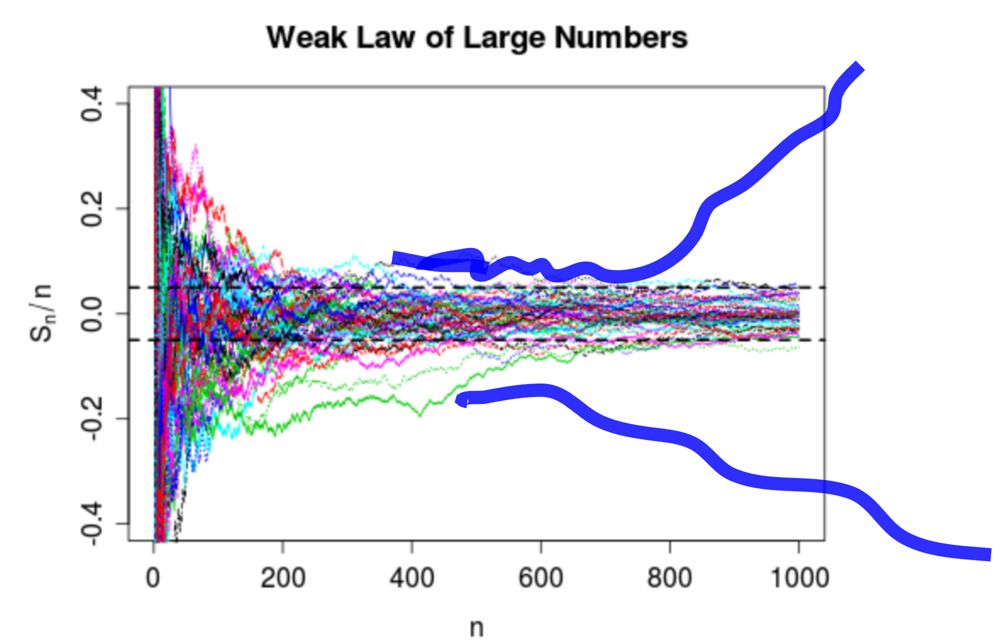
(almost sure convergence)

weak LLN $P(|\overline{X}_n - \mu| > \epsilon) \to 0$ for any ϵ, as $n \to \infty$.

(convergence in distribution)

Example. Suppose that X~Bern(p). Then, the sample mean of the first n trials tends to p, as n goes to infinity, with probability 1.





Strong and Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be i.i.d. RV with mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ sample mean of the first n RV's.

Question: *What happens the the sample mean when n gets large?*

strong LLN

$\overline{X}_n \to \mu$, as $n \to \infty$ with probability 1.

(almost sure convergence)

weak LLN

$P(|\overline{X}_n - \mu| > \epsilon) \to 0$ for any ϵ, as $n \to \infty$.

(convergence in distribution)

Example. Suppose that X~Bern(p). Then, the sample mean of the first n trials tends to p, as n goes to infinity, with probability 1.

Example. Suppose that X~Poisson(lambda). Then, since Poisson(lambda_1)+Poisson(\lambda_2) = Poisson(\lambda_1+\lambda_2), we have that

$$lim_{N\to\infty} \frac{Poisson(N\lambda)}{N} = \underline{\lambda}$$

$$\sum_n Poi(\lambda) = \underline{Poi(n\lambda)}$$

Remark: This will be useful in context of scaling copy numbers to concentrations in biochemical reaction networks.

$$P(|X-\mu| > \varepsilon) \le \frac{Var(X)}{\varepsilon^2}$$

Strong and Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be i.i.d. RV with mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ sample mean of the first n RV's.

Question: *What happens the the sample mean when n gets large?*

strong LLN

$\overline{X}_n \to \mu$, as $n \to \infty$ with probability 1.

(almost sure convergence)

weak LLN

$P(|\overline{X}_n - \mu| > \epsilon) \to 0$ for any ϵ, as $n \to \infty$.

(convergence in distribution)

Proof of weak LLN.

$$P(|\overline{X}_n - \mu| > \varepsilon) \le \frac{Var(\overline{X}_n)}{\varepsilon^2} = \frac{Var\left(\frac{X_1+\cdots+X_n}{n}\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}Var(X_1+\cdots+X_n)}{\varepsilon^2} = \frac{\sum Var(X_i)}{n^2\varepsilon^2} = \frac{n\,Var(X)}{n^2\varepsilon^2} \to 0 \text{ as } n \to \infty$$

Question

$$n^{i}(\overline{X}_n - \mu)$$

$i = 0$

$i = 1$

$i = \frac{1}{2}$

Central Limit Theorem

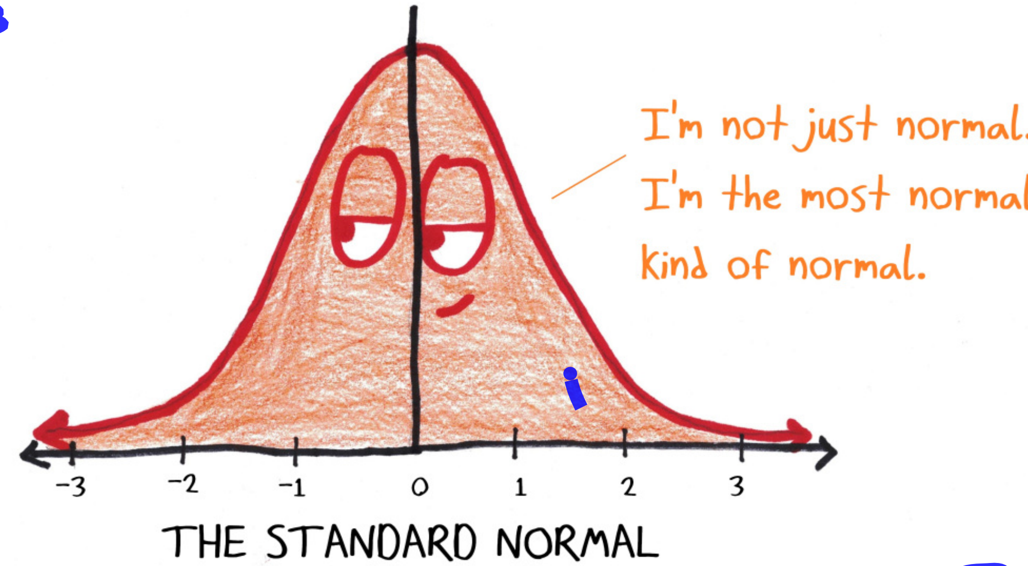
$\frac{X_1 + \dots + X_n}{n}$

Central Limit Theorem

As $n \to \infty$, the CDF of $\left(n^{1/2}\frac{\bar{X}_n - \mu}{\sigma}\right)$ converges to that of $\mathcal{N}(0,1)$.

Equivalently, $$\frac{\sum_{j=1}^{n} X_j}{\sqrt{n}\sigma} \to \mathcal{N}(0,1)$$

Example. Suppose we randomly select a journal article from a source with an average of 1000 words per article, with a standard deviation of 200 words. We can then infer that the probability that it has between 600 and 1400 words (i.e. within *k* = 2 standard deviations of the mean) must be at least 75%, because there is no more than 1/4 chance to be outside that range, by Chebyshev's inequality. **But if we additionally know that the distribution is normal, we can say there is a 75% chance the word count is between 770 and 1230 (which is an even tighter bound).**



$P(X \in \mu \pm 2\sigma) \geq 0.75$

$(600, 1400)$

(CLT) $P(X \in (770, 1230)) \geq 0.75$

$W \sim \mathcal{N}(1000, 200)$

$\frac{W - \mu_W}{\sigma_W} \sim \mathcal{N}(0,1)$

$Z \sim \mathcal{N}(0,1)$

Central Limit Theorem: Proof

$$\frac{\sum_{j=1}^{n} X_j}{\sqrt{n}\sigma} \rightarrow \mathcal{N}(0,1)$$

Proof. We will prove the CLT assuming that the MGF $M(t)$ of the X_j exists (note that we have been assuming all along that the first two moments exist). We will show that the MGFs converge, which will imply that the CDFs converge (however, we will not show this fact).

Lecture Goals

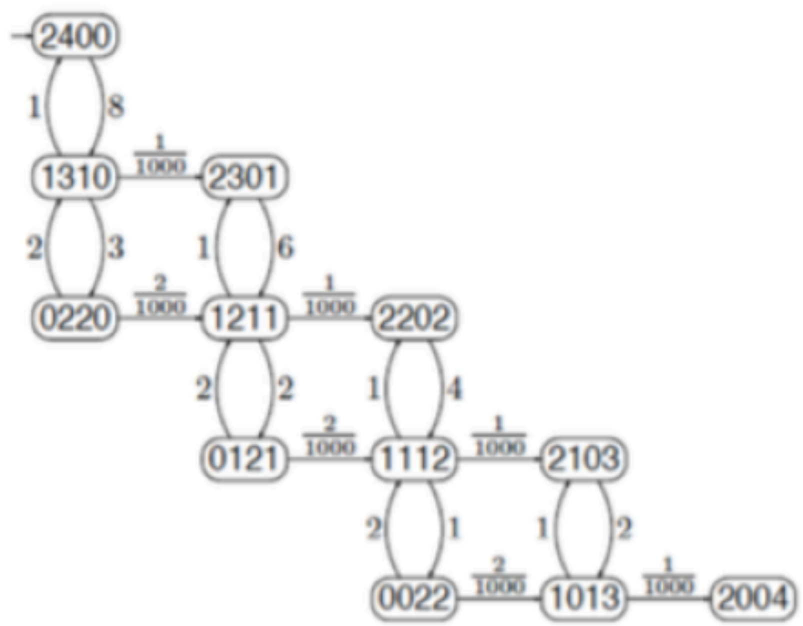
1. **Motivating example**

2. **Law of large numbers & central limit theorem**
   1. inequalities: Markov, Chebyshev
   2. LLN: proof for the weak version
   3. CLN: proof sketch and example

3. **Biochemical reaction networks**

Relationships between stoc. and det. models [Kurtz, 1980]

Enzyme-catalyzed substrate conversion as a CTMC



States:	*init*	*goal*
enzymes	2	2
substrates	4	0
complex	0	0
products	0	4

Transitions: $E + S \underset{1}{\overset{1}{\rightleftharpoons}} C \xrightarrow{0.001} E + P$

e.g., $(x_E, x_S, x_C, x_P) \xrightarrow{0.001 \cdot x_C} (x_E + 1, x_S, x_C - 1, x_P + 1)$ for $x_C > 0$

Recall HW 5:

\# reachable states for 2 enzymes and 4 substrates is 12

\# reachable states for 20 enzymes and 40 substrates is O(1000)

\# reachable states for 200 enzymes an 400 substrates O(100000)

Can we approximate the dynamics by a deterministic system, when all species are highly abundant?

$$E + S \rightarrow_{k1} C, \qquad C \rightarrow_{k2} E + S, \qquad C \rightarrow_{k3} E + P$$

Stochastic model:

$$E + S \rightarrow_{c1} C$$
$$C \rightarrow_{c2} E + S$$
$$C \rightarrow_{c3} E + P$$

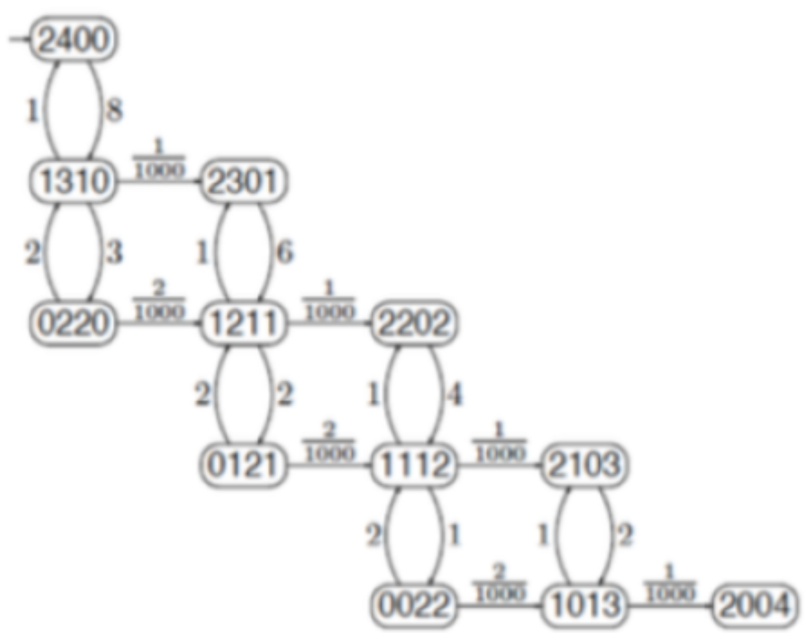
Species units: molecules
Rates: molecules per second
State change: stochastic

Deterministic model:

$$E + S \rightarrow_{k1} C$$
$$C \rightarrow_{k2} E + S$$
$$C \rightarrow_{k3} E + P$$

Species units: concentrations (per mol per l)
Rate units: mols per second
State change: deterministic

CTMC:



Set of ODE's:

$$\dot{x}_E(t) = -k_1 x_E(t) x_S(t) + k_2 x_C(t) + k_3 x_C(t)$$
$$\dot{x}_S(t) = -k_1 x_E(t) x_S(t) + k_2 x_C(t)$$

(conservation laws)

$$x_C(t) := x_E(0) - x_E(t)$$
$$x_P(t) := x_S(0) + x_P(0) - x_S(t) - x_C(t)$$

When can we approximate the dynamics of the CTMC by a deterministic system?

$$E + S \rightarrow_{k1} C, \qquad C \rightarrow_{k2} E + S, \qquad C \rightarrow_{k3} E + P$$

Stochastic model:

$$E + S \rightarrow_{c1} C$$
$$C \rightarrow_{c2} E + S$$
$$C \rightarrow_{c3} E + P$$

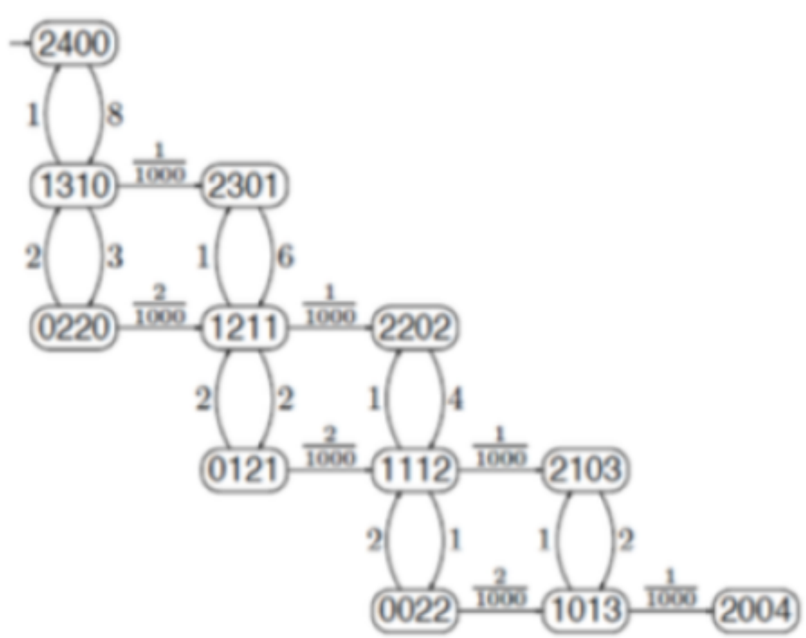
Species units: molecules
Rates: molecules per second
State change: stochastic

Deterministic model:

$$E + S \rightarrow_{k1} C$$
$$C \rightarrow_{k2} E + S$$
$$C \rightarrow_{k3} E + P$$

Species units: concentrations (per mol per l)
Rate units: mols per second
State change: deterministic

CTMC:



Set of ODE's:

$$\dot{x}_E(t) = -k_1 x_E(t) x_S(t) + k_2 x_C(t) + k_3 x_C(t)$$
$$\dot{x}_S(t) = -k_1 x_E(t) x_S(t) + k_2 x_C(t)$$

(conservation laws)

$$x_C(t) := x_E(0) - x_E(t)$$
$$x_P(t) := x_S(0) + x_P(0) - x_S(t) - x_C(t)$$

Volume

$$X[molecules] = x[mols\ per\ liter] \cdot \boxed{n_A[molecules\ in\ mol] \cdot \Omega[cell\ volume\ in\ l]}$$

e.g. $X[molecules] = 10^{-5}\frac{mol}{l} \cdot 6.023 \cdot 10^{23}\frac{1}{mol} \cdot 10^{-15} l = 6023$

$$E + S \rightarrow_{k1} C, \qquad C \rightarrow_{k2} E + S, \qquad C \rightarrow_{k3} E + P$$

Stochastic model:

$$E + S \rightarrow_{c1} C$$
$$C \rightarrow_{c2} E + S$$
$$C \rightarrow_{c3} E + P$$

Species units: molecules
Rates: molecules per second
State change: stochastic

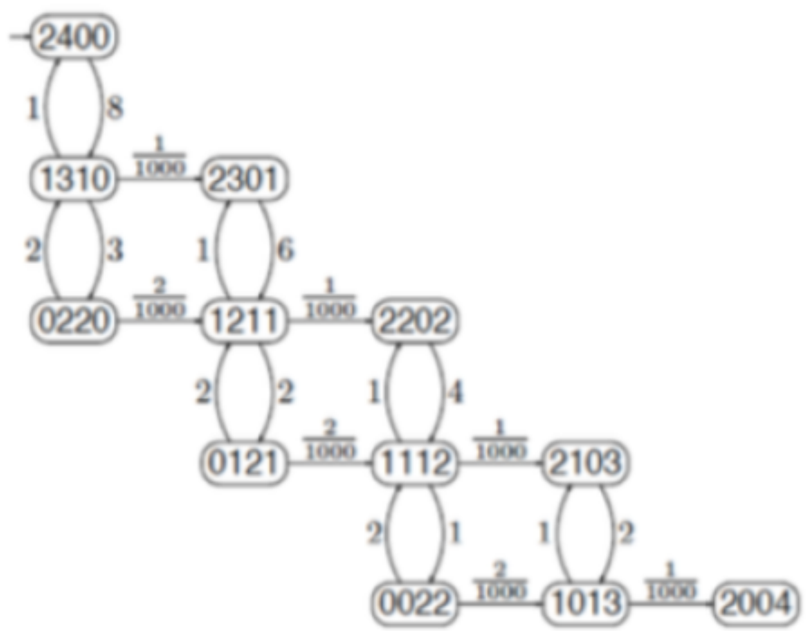
scale with volume

Deterministic model:

$$E + S \rightarrow_{k1} C$$
$$C \rightarrow_{k2} E + S$$
$$C \rightarrow_{k3} E + P$$

Species units: concentrations (per mol per l)
Rate units: mols per second
State change: deterministic

CTMC:



Set of ODE's:

$$\dot{x}_E(t) = -k_1 x_E(t) x_S(t) + k_2 x_C(t) + k_3 x_C(t)$$
$$\dot{x}_S(t) = -k_1 x_E(t) x_S(t) + k_2 x_C(t)$$

(conservation laws)

$$x_C(t) := x_E(0) - x_E(t)$$
$$x_P(t) := x_S(0) + x_P(0) - x_S(t) - x_C(t)$$

Volume

$$X[molecules] = x[mols\ per\ liter] \cdot \boxed{n_A[molecules\ in\ mol] \cdot \Omega[cell\ volume\ in\ l]}$$

e.g. $X[molecules] = 10^{-5}\dfrac{mol}{l} \cdot 6.023 \cdot 10^{23}\dfrac{1}{mol} \cdot 10^{-15} l = 6023$

$$E + S \to_{k1} C, \qquad C \to_{k2} E + S, \qquad C \to_{k3} E + P$$

Stochastic model:

$$E + S \to_{c1} C$$
$$C \to_{c2} E + S$$
$$C \to_{c3} E + P$$

Species units: molecules
Rates: molecules per second
State change: stochastic

scale with volume

$$x(t) := \frac{X(t)}{Volume}$$

$$\lambda_i^D(x(t)) \approx \frac{\lambda_i(X(t))}{Volume}$$

Deterministic model:

$$E + S \to_{k1} C$$
$$C \to_{k2} E + S$$
$$C \to_{k3} E + P$$

Species units: concentrations (per mol per l)
Rate units: mols per second
State change: deterministic

CTMC written as random time-change model:

$$\begin{aligned}(X_E, X_C, X_C, X_P)(t) = &(X_E, X_S, X_C, X_P)(0) \\ &+ \boxed{R_1(t)(-1,-1,0,0)} \\ &+ R_2(t)(1,1,-1,0) \\ &+ R_3(t)(1,0,-1,1)\end{aligned}$$

$$R_i(t) = Poisson\left(\int_0^t \lambda_i((X_E, X_C, X_C, X_P)(s))ds\right)$$

(how many times a particular reaction fired until time t - a non-homogeneous Poisson process)

Set of ODE's:

$$\dot{x}_E(t) = \boxed{-k_1 x_E(t) x_S(t)} + k_2 x_C(t) + k_3 x_C(t)$$
$$\dot{x}_S(t) = \boxed{-k_1 x_E(t) x_S(t)} + k_2 x_C(t)$$

(conservation laws)

$$x_C(t) := x_E(0) - x_E(t)$$
$$x_P(t) := x_S(0) + x_P(0) - x_S(t) - x_C(t)$$

$$E + S \rightarrow_{k1} C, \qquad C \rightarrow_{k2} E + S, \qquad C \rightarrow_{k3} E + P$$

Stochastic model:

$$E + S \rightarrow_{c1} C$$
$$C \rightarrow_{c2} E + S$$
$$C \rightarrow_{c3} E + P$$

Species units: molecules
Rates: molecules per second
State change: stochastic

scale with volume

$$x(t) := \frac{X(t)}{Volume}$$

$$\lambda_i^D(x(t)) \approx \frac{\lambda_i(X(t))}{Volume}$$

Deterministic model:

$$E + S \rightarrow_{k1} C$$
$$C \rightarrow_{k2} E + S$$
$$C \rightarrow_{k3} E + P$$

Species units: concentrations (per mol per l)
Rate units: mols per second
State change: deterministic

CTMC written as random time-change model:

$$\begin{aligned}(X_E, X_C, X_C, X_P)(t) = &(X_E, X_S, X_C, X_P)(0)\\ &+ \boxed{R_1(t)(-1,-1,0,0)}\\ &+ R_2(t)(1,1,-1,0)\\ &+ R_3(t)(1,0,-1,1)\end{aligned}$$

$$R_i(t) = Poisson\left(\int_0^t \lambda_i((X_E, X_C, X_C, X_P)(s))ds\right)$$

(how many times a particular reaction fired until time t - a non-homogeneous Poisson process)

Set of ODE's:

$$\dot{x}_E(t) = \boxed{-k_1 x_E(t) x_S(t)} + k_2 x_C(t) + k_3 x_C(t)$$
$$\dot{x}_S(t) = \boxed{-k_1 x_E(t) x_S(t)} + k_2 x_C(t)$$

$$\begin{aligned}\frac{R_1(t)}{V} &= \frac{1}{V} Poisson\left(\int_0^t \lambda_1((X_E, X_C, X_C, X_P)(s))ds\right)\\ &= \frac{1}{V} Poisson\left(V \int_0^t \lambda_1^D((x_E, x_C, x_C, x_P)(s))ds\right)\\ &\rightarrow \int_0^t \lambda_1^D((x_E, x_C, x_C, x_P)(s)), \text{ by LLN, when } V \rightarrow \infty\\ &= \int_0^t k_1 x_E(s) x_S(s) ds\end{aligned}$$

Next Lectures

1. Parameter inference
2. Statistical model checking
3. Outlook