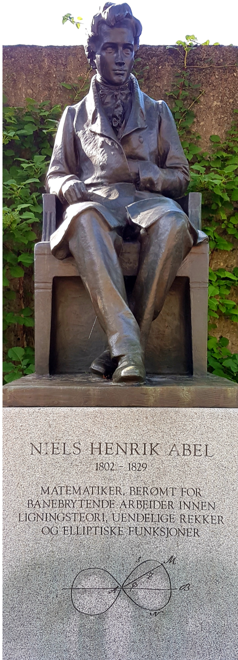
UiO : **Department of Mathematics**

University of Oslo

# Lecture 14. Markov Chain Monte Carlo Methods. The Exponential Distribution

STK2130 – Modellering av stokastiske prosesser

**Anton Yurchenko-Tytarenko**



# Markov Chain Monte Carlo Methods (MCMC)

Let $Z$ be a discrete random variable with a state space $\mathcal{S}$, and assume that:

$$\mathbb{P}(Z = i) = \pi_i = \frac{b_i}{B}, \quad i \in \mathcal{S}.$$

We assume that $b_i$ is known for all $i \in \mathcal{S}$. Since the probabilities must add up to 1, we obviously have:

$$\sum_{i \in \mathcal{S}} \frac{b_i}{B} = B^{-1} \sum_{i \in \mathcal{S}} b_i = 1.$$

Hence, it follows that the normalizing constant $B$ is given by:

$$B = \sum_{i \in \mathcal{S}} b_i.$$

Thus, in principle $B$ is known as well. However, if $|\mathcal{S}|$ is large, calculating $B$ may be a time-consuming task.

# Markov Chain Monte Carlo Methods (MCMC)

## Example

Let $T$ and $Z$ be two discrete random variables with state spaces $\mathcal{T}$ and $\mathcal{S}$ respectively. We assume that the marginal distribution of $Z$ and the conditional distribution of $T$ given $Z$ are known.

The conditional distribution of $Z$ given $T$ is then:

$$\mathbb{P}(Z = i \mid T = t) = \frac{P(Z = i)P(T = t \mid Z = i)}{\sum_{j \in \mathcal{S}} P(Z = j)P(T = t \mid Z = j)} = \frac{b_i(t)}{B(t)}, \quad i \in \mathcal{S}, t \in \mathcal{T},$$

where we have introduced:

$$b_i(t) = \mathbb{P}(Z = i)\mathbb{P}(T = t \mid Z = i), \quad i \in \mathcal{S}, t \in \mathcal{T},$$

$$B(t) = \sum_{j \in \mathcal{S}} \mathbb{P}(Z = j)\mathbb{P}(T = t \mid Z = j) = \mathbb{P}(T = t), \quad t \in \mathcal{T}.$$

If $|\mathcal{S}|$ is large, we may want to avoid calculating $B(t)$.

# Markov Chain Monte Carlo Methods (MCMC)

**Problem**

*Construct a Markov chain $\{X_n\}$ with state space $\mathcal{S}$ and stationary distribution equal to the distribution of $Z$.*

**SOLUTION (Hastings-Metropolis):** Let **Q** be **any** given irreducible Markov chain transition probability matrix on $\mathcal{S}$, and define:

$$\alpha_{i,j} = \min\left(\frac{b_j Q_{j,i}}{b_i Q_{i,j}}, 1\right), \quad i, j \in \mathcal{S}.$$

We then let the transition probability matrix of $\{X_n\}$, denoted **P**, be defined as follows:

$$P_{i,j} = Q_{i,j}\alpha_{i,j}, \quad i \neq j,$$
$$P_{i,i} = 1 - \sum_{j \neq i} Q_{i,j}\alpha_{i,j}, \quad i \in \mathcal{S}.$$

# Markov Chain Monte Carlo Methods (MCMC)

We then claim that $\{X_n\}$ is **time reversible** and has a stationary distribution equal to the distribution of $Z$. That is, $\pi_i = b_i/B$, for all $i \in \mathcal{S}$, and:

$$\pi_i P_{i,j} = \pi_j P_{j,i}, \quad \text{for all } i, j \in \mathcal{S}. \tag{1}$$

Since (1) is trivially satisfied for $i = j$, we focus on the case where $i \neq j$, where (1) can be expressed as:

$$\frac{b_i}{B} Q_{i,j} \alpha_{i,j} = \frac{b_j}{B} Q_{j,i} \alpha_{j,i}, \quad i \neq j. \tag{2}$$

By eliminating $B$ from these equations and inserting the expression for $\alpha_{i,j}$ we get:

$$b_i Q_{i,j} \cdot \min\left(\frac{b_j Q_{j,i}}{b_i Q_{i,j}}, 1\right) = b_j Q_{j,i} \cdot \min\left(\frac{b_i Q_{i,j}}{b_j Q_{j,i}}, 1\right), \quad i \neq j. \tag{3}$$

# Markov Chain Monte Carlo Methods (MCMC)

**CASE 1:** $b_i Q_{i,j} \leq b_j Q_{j,i}$.
In this case $\alpha_{i,j} = 1$ while $\alpha_{j,i} = (b_i Q_{i,j})/(b_j Q_{j,i})$, and hence, (3) simplifies to:

$$b_i Q_{i,j} = b_j Q_{j,i} \cdot (b_i Q_{i,j})/(b_j Q_{j,i}), \quad i \neq j. \tag{4}$$

**CASE 2:** $b_i Q_{i,j} \geq b_j Q_{j,i}$.
In this case $\alpha_{i,j} = (b_j Q_{j,i})/(b_i Q_{i,j})$ while $\alpha_{j,i} = 1$, and hence, (3) simplifies to:

$$b_i Q_{i,j} \cdot (b_j Q_{j,i})/(b_i Q_{i,j}) = b_j Q_{j,i}, \quad i \neq j. \tag{5}$$

Since obviously both (4) and (5) hold true, we conclude that (1) holds true as well.

# Markov Chain Monte Carlo Methods (MCMC)

We recall that:

$$P_{i,j} = Q_{i,j}\alpha_{i,j}, \quad i \neq j,$$
$$P_{i,i} = 1 - \sum_{j \neq i} Q_{i,j}\alpha_{i,j}, \quad i \in \mathcal{S}.$$

Assume that $X_n = i$. Then $X_{n+1}$ can be generated using the following two-step Monte Carlo simulation procedure:

**STEP 1.** Generate a random variable $J$ with values in $\mathcal{S}$ such that $\mathbb{P}(J = j) = Q_{i,j}$, $j \in \mathcal{S}$.
**STEP 2.** Generate $K \in \{0, 1\}$ such that $\mathbb{P}(K = 1 \mid J = j) = \alpha_{i,j}$, and let:

$$X_{n+1} = K \cdot j + (1 - K) \cdot i$$

Thus, a transition from state $i$ to state $j$ where $i \neq j$ happens if and only if $J = j$ and $K = 1$. If not, the process stays in state $i$.

# Markov Chain Monte Carlo Methods (MCMC)

The Monte Carlo simulation procedure can be used to estimate some unknown parameter in the distribution of $Z$, e.g.:

$$\theta = \mathbb{E}[h(Z)] = \sum_{i \in \mathcal{S}} h(i)\mathbb{P}(Z = i),$$

where $h$ is some function of interest, and the normalizing constant $B$ of the distribution of $Z$ is too time-consuming to calculate.
By simulating the Markov chain $\{X_n\}$, having a stationary distribution which is equal to the distribution of $Z$, we may estimate $\theta$ by:

$$\hat{\theta}_n = \frac{1}{n} \sum_{m=1}^{n} h(X_m).$$

We know that $\hat{\theta}_n \to \theta$ when $n \to \infty$.

# Markov Chain Monte Carlo Methods (MCMC)

**Remark**

- $X_1, X_2, \ldots$ are not independent.
- The chain may converge slowly towards its stationary distribution.

Both these issues tend to have a negative effect on the convergence rate of the estimator $\hat{\theta}_n$.
If many of the $\alpha_{i,j}$-s are small, the Markov chain tends to get stuck for a long time before eventually transiting to another state. In such cases the estimator $\hat{\theta}_n$ will converge very slowly.

# Markov Chain Monte Carlo Methods (MCMC)

For optimal performance, i.e., fast convergence, the matrix **Q** should ideally be chosen so that:

$$b_i Q_{i,j} = b_j Q_{j,i}, \quad \text{for all } i, j \in \mathcal{S}.$$

Then it follows that:

$$\alpha_{i,j} = \min\left(\frac{b_j Q_{j,i}}{b_i Q_{i,j}}, 1\right) = 1, \quad \text{for all } i, j \in \mathcal{S}.$$

Hence, $\mathbf{Q} = \mathbf{P}$, i.e., **Q** is itself the transition probability matrix of $\{X_n\}$.

Finding the optimal matrix **Q** implies finding a transition probability matrix with a stationary distribution which is equal to the distribution of $Z$. In real-life applications, this can be difficult.

Instead we may think of **Q** as our best guess, while the $\alpha_{i,j}$-s are correction factors which are used to generate a Markov chain with the correct stationary distribution.

# Gibbs sampling

Assume that $\mathbf{Z} = (Z_1, ..., Z_r)$ is a discrete random vector with values in $\mathcal{S}$ where:

$$\mathbb{P}(\mathbf{Z} = \mathbf{z}) = p(\mathbf{z}) = g(\mathbf{z})/B, \quad \text{for all } z \in \mathcal{S},$$

where the $g(\mathbf{z})$ is known for all $\mathbf{z} \in \mathcal{S}$ and $B$ is an unknown normalizing constant.

Utilization of the Gibbs sampler assumes that for any $i$ and values $x_j$, $j \neq i$, we can generate a random variable $Z$ having the probability mass function

$$\mathbb{P}(Z = z) = \mathbb{P}(Z_i = z \mid Z_j = z_j,\ j \neq i).$$

We then consider the first step of the Hastings-Metropolis algorithm, and assume that $X_n = \mathbf{z} = (z_1, ..., z_r)$. The candidate for the next state, $X_{n+1}$, is generated as follows:

1. Generate $K$ uniformly from the set $\{1, ..., r\}$.
2. For $K = k$, generate $Z_k = z$ conditional on $Z_i = z_i$, $i = 1, ..., (k-1), (k+1)..., r$.

The resulting candidate for the next state, denoted $\mathbf{y}$, is then:

$$\mathbf{y} = (z_1, ..., z_{k-1}, z, z_{k+1}, ..., z_r).$$

# Gibbs sampling

This implies that we have the following transition probabilities:

$$Q_{\mathbf{z},\mathbf{y}} = \frac{1}{r}\mathbb{P}(Z_k = z \mid Z_i = z_i,\ i \neq k)$$
$$= \frac{g(\mathbf{y})/B}{r \cdot \sum_{z_k} g(z_1, .., z_k, ...z_r)/B} = \frac{g(\mathbf{y})}{r \cdot \sum_{z_k} g(z_1, .., z_k, ...z_r)}$$

By the same type of argument, we also have:

$$Q_{\mathbf{y},\mathbf{z}} = \frac{g(\mathbf{z})}{r \cdot \sum_{z_k} g(\mathbf{z})}.$$

This implies that:

$$g(\mathbf{z})Q_{\mathbf{z},\mathbf{y}} = g(\mathbf{y})Q_{\mathbf{y},\mathbf{z}}, \quad \text{for all } \mathbf{z}, \mathbf{y} \in \mathcal{S}.$$

Hence, $\alpha_{\mathbf{z},\mathbf{y}} = 1$ for all $\mathbf{z}, \mathbf{y} \in \mathcal{S}$, and thus, $\mathbf{Q}$ is an optimal transition probability matrix.

# The Exponential Distribution

**Definition**

A continuous random variable $X$ is said to have an **exponential distribution** with parameter $\lambda > 0$, denoted as $X \sim exp(\lambda)$, if its probability density function is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

If $X \sim exp(\lambda)$, then the **cumulative distribution function** of $X$ is given by:

$$F(x) = \mathbb{P}(X \leq x) = \int_0^x f(t)dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Moreover, the **survival function** of $X$ is given by:

$$\bar{F}(x) = \mathbb{P}(X > x) = 1 - F(x) = \begin{cases} e^{-\lambda x}, & x \geq 0, \\ 1, & x < 0. \end{cases}$$

# The Exponential Distribution

The exponential distribution is a special case of the **gamma distribution** with parameters $\alpha > 0$ and $\lambda > 0$, denoted as $X \sim Gamma(\alpha, \lambda)$ with probability density function:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $\Gamma(\alpha)$, defined for all $\alpha > 0$, is the gamma function given by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \Gamma(n) = (n-1)!, \quad n = 1, 2, ...$$

By substituting $u = \lambda x$ and $du = \lambda dx$, we find that:

$$\int_0^\infty f(x) dx = \frac{1}{\Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u} du = 1.$$

Thus, $f(x)$ is indeed a proper probability density.

# The Exponential Distribution

Assume that $X \sim exp(\lambda)$, and let $p > -1$. We then have:

$$\mathbb{E}[X^p] = \int_0^\infty x^p f(x)dx = \int_0^\infty \lambda x^p e^{-\lambda x} dx$$
$$= \frac{\Gamma(p+1)}{\lambda^p} \int_0^\infty \frac{\lambda^{p+1}}{\Gamma(p+1)} x^{(p+1)-1} e^{-\lambda x} dx$$
$$= \frac{Gamma(p+1)}{\lambda^p}.$$

In particular:

$$\mathbb{E}[X] = \frac{\Gamma(2)}{\lambda^1} = \frac{(2-1)!}{\lambda} = \frac{1}{\lambda}, \quad \mathbb{E}[X^2] = \frac{\Gamma(3)}{\lambda^2} = \frac{(3-1)!}{\lambda^2} = \frac{2}{\lambda^2},$$
$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

# The Exponential Distribution

Assume that $X \sim Gamma(\alpha, \lambda)$. Then the moment generating function of X is given by:

$$
\begin{aligned}
M_X(t) = \mathbb{E}[e^{tX}] &= \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\
&= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx \\
&= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx \\
&= \frac{\lambda^\alpha}{(\lambda-t)^\alpha},
\end{aligned}
$$

for all $t < \lambda$.
In particular, if $X \sim exp(\lambda)$, we have:

$$
M_X(t) = \frac{\lambda}{\lambda - t},
$$

for all $t < \lambda$.

# The Exponential Distribution

## Proposition 5.1

Assume that $X_1, ..., X_n$ are independent and $X_i \sim exp(\lambda)$, $i = 1, ..., n$, and let:

$$Y = X_1 + \cdots + X_n$$

Then $Y \sim Gamma(n, \lambda)$.

**PROOF:** Using moment generating functions we get:

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] = \mathbb{E}[e^{tX_1 + \cdots + tX_n}] = M_{X_1}(t) \cdots M_{X_n}(t) \\ &= \frac{\lambda}{\lambda - t} \cdots \frac{\lambda}{\lambda - t} = \frac{\lambda^n}{(\lambda - t)^n}. \end{aligned}$$

Hence, $Y \sim Gamma(n, \lambda)$.

# The Exponential Distribution

Proposition 5.1 is a special case of the following more general result:

## Proposition 5.1b

Assume that $X_1, \ldots, X_n$ are independent and $X_i \sim Gamma(\alpha_i, \lambda)$, $i = 1, ..., n$, and let:

$$Y = X_1 + \cdots + X_n.$$

Then $Y \sim Gamma(\alpha, \lambda)$, where $\alpha = \sum_{i=1}^{n} \alpha_i$.

**PROOF:** Using moment generating functions we get:

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{tX_1 + \cdots + tX_n}] = M_{X_1}(t) \cdots M_{X_n}(t)$$
$$= \frac{\lambda^{\alpha_1}}{(\lambda - t)^{\alpha_1}} \cdots \frac{\lambda^{\alpha_n}}{(\lambda - t)^{\alpha_n}} = \frac{\lambda^{\alpha}}{(\lambda - t)^{\alpha}}.$$

Hence, $Y \sim Gamma(\alpha, \lambda)$.

# Memoryless stochastic variables

A random variable $X$ is said to be memoryless if:

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X - t > s \mid X > t) = \mathbb{P}(X > s), \quad \text{for all } s, t \geq 0.$$

Thus, $X$ is memoryless if $(X - t) \mid (X > t)$ has the same distribution as $X$.

Note that if $X$ is the lifetime of some unit, $(X - t)$ is the remaining lifetime given that the unit has survived up to the time $t$.

If $X \sim exp(\lambda)$, we have:

$$\begin{aligned}\mathbb{P}(X > s + t \mid X > t) &= \frac{\mathbb{P}(\{X > s + t\} \cap \{X > t\})}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s).\end{aligned}$$

Hence, we conclude that $X$ is memoryless.

# Memoryless stochastic variables

The memoryless property:

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s), \quad \text{for all } s, t \geq 0.$$

is equivalent to the following:

$$\mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t), \quad \text{for all } s, t \geq 0.$$

Since $\bar{F}(x) = \mathbb{P}(X > x)$, this property can also be written as:

$$\bar{F}(s + t) = \bar{F}(s)\bar{F}(t), \quad \text{for all } s, t \geq 0.$$

We now show that the exponential distribution is essentially the only distribution with this property.

# Memoryless stochastic variables

## Proposition

Let $X$ be a random variable and let $\bar{F}(x) = \mathbb{P}(X > x)$ be such that:

$$\bar{F}(x+y) = \bar{F}(x) \cdot \bar{F}(y), \quad \text{for all } x, y \geq 0. \tag{6}$$

Denote

$$\lambda := -\log(\bar{F}(1)) > 0. \tag{7}$$

Then $X \sim exp(\lambda)$.

**PROOF:** We first note that by (7), it follows that:

$$0 < \bar{F}(1) = e^{-\lambda} < 1. \tag{8}$$

Secondly we note that since cumulative distribution functions always are right-continuous, it follows that $\bar{F} = 1 - F$ is right-continuous as well.

# Memoryless stochastic variables

By repeated use of (6) it follows that for $n, m \in \mathbb{N}^+$, we have:

$$\bar{F}\left(\frac{m}{n}\right) = \bar{F}\left(\frac{1}{n} + \cdots + \frac{1}{n}\right) = \bar{F}^m\left(\frac{1}{n}\right), \tag{9}$$

where the sum contains $m$ terms. In particular, by letting $m = n$, we get:

$$\bar{F}(1) = \bar{F}\left(\frac{n}{n}\right) = \bar{F}^n\left(\frac{1}{n}\right). \tag{10}$$

Alternatively, (10) can be written as:

$$\bar{F}\left(\frac{1}{n}\right) = (\bar{F}(1))^{1/n}. \tag{11}$$

By (8) and that $\bar{F}$ is right-continuous, (11) implies that:

$$\bar{F}(0) = \lim_{n\to\infty} \bar{F}\left(\frac{1}{n}\right) = \lim_{n\to\infty} (\bar{F}(1))^{1/n} = 1.$$

Hence, since $\bar{F}$ must be non-increasing, $\bar{F}(x) = 1$ for all $x \leq 0$.

# Memoryless stochastic variables

We now combine (9) and (11), and get:

$$\bar{F}\left(\frac{m}{n}\right) = \bar{F}^m\left(\frac{1}{n}\right) = \bar{F}(1)^{m/n}, \quad \text{for all } m, n \in \mathbb{N}^+.$$

Thus, since $\bar{F}(1) = e^{-\lambda}$, we have proved that:

$$\bar{F}(q) = \bar{F}(1)^q = e^{-\lambda q}, \quad \text{for all } q \in \mathbb{Q}^+.$$

Now, let $x \in \mathbb{R}^+$. Since the set $\mathbb{Q}^+$ is dense in $\mathbb{R}^+$, there exists a decreasing sequence $\{q_r\} \subset \mathbb{Q}^+$ such that:

$$\lim_{r\to\infty} q_r = x.$$

Since $\bar{F}$ is right-continuous, this implies that:

$$\bar{F}(x) = \lim_{r\to\infty} \bar{F}(q_r) = \lim_{r\to\infty} e^{-\lambda q_r} = e^{-\lambda x}.$$

Hence, we conclude that $X \sim exp(\lambda)$.

# Example 5.2

## Problem

The amount of time one spends in a bank, denoted $X$, is exponentially distributed with mean ten minutes. That is, $X \sim exp(\lambda) = exp\left(\frac{1}{10}\right)$.

1. What is the probability that a customer will spend more than fifteen minutes in the bank?
2. What is the probability that a customer will spend more than fifteen minutes in the bank given that she is still in the bank after ten minutes?

**SOLUTION:**

1. $$\mathbb{P}(X > 15) = e^{-15\lambda} = e^{-15/10} \approx 0.223.$$
2. $$\mathbb{P}(X > 15 \mid X > 10) = e^{-(15-10)\lambda} = e^{-5/10} \approx 0.607.$$

# The Exponential Distribution

Assume that $X_1$, $X_2$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, 2$. We want to calculate the probability of the event that $X_1 < X_2$.
The 2-dimensional random variable $(X_1, X_2)$ has distribution with density

$$f(x_1, x_2) = \lambda_1 \lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 x_2)} \mathbb{1}_{\{x_1 > 0, x_2 > 0\}},$$

therefore

$$\begin{aligned}
\mathbb{P}(X_1 < X_2) &= \int \int_{x_1 < x_2} f(x_1, x_2) dx_1 dx_2 = \lambda_1 \lambda_2 \int_0^\infty \int_0^{x_2} e^{-(\lambda_1 x_1 + \lambda_2 x_2)} dx_1 dx_2 \\
&= \lambda_1 \lambda_2 \int_0^\infty e^{-\lambda_2 x_2} \left( \int_0^{x_2} e^{-\lambda_1 x_1} dx_1 \right) dx_2 = \lambda_1 \lambda_2 \int_0^\infty e^{-\lambda_2 x_2} \frac{1}{\lambda_1} \left( 1 - e^{-\lambda_1 x_2} \right) dx_2 \\
&= \lambda_2 \int_0^\infty e^{-\lambda_2 x_2} dx_2 - \lambda_2 \int_0^\infty e^{-(\lambda_1 + \lambda_2) x_2} dx_2 = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

# The Exponential Distribution

Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, ..., n$.

$$\mathbb{P}(\min_{1 \le i \le n} X_i > x) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i > x\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i > x) = \prod_{i=1}^{n} e^{-\lambda_i x} = e^{-(\sum_{i=1}^{n} \lambda_i) x}.$$

Thus, we have shown that $\min_{1 \le i \le n} X_i \sim exp\left(\sum_{i=1}^{n} \lambda_i\right)$.

# The Exponential Distribution

The following result combines the two previous results.
Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, ..., n$. We want to calculate the probability that $X_i$ is the smallest of all the variables, i.e., that $X_i < X_j$ for all $j \neq i$.

$$\begin{aligned}\mathbb{P}\left(X_i < X_j \text{ for all } j \neq i\right) &= \mathbb{P}(X_i < \min_{j \neq i} X_j) \\ &= \frac{\lambda_i}{\lambda_i + \sum_{j \neq i} \lambda_j} \\ &= \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}\end{aligned}$$

# The Exponential Distribution

## Proposition 5.2

Assume that $X_1, \ldots, X_n$ are independent and that $X_i \sim exp(\lambda_i)$, $i = 1, ..., n$. Then mini $X_i \sim exp(\sum_{i=1}^{n} \lambda_i)$. Moreover, $\min_i X_i$ and the rank order of $X_1, \ldots, X_n$ are independent.

**PROOF:** Since the exponential distribution is memoryless, we get that:

$$\mathbb{P}(X_{i_1} < \cdots < X_{i_n} \mid \min_{1 \le i \le n} X_i > t)$$

$$= \mathbb{P}\left(X_{i_1} < \cdots < X_{i_n} \,\middle|\, \cap_{i=1}^{n} X_i > t\right)$$

$$= \mathbb{P}\left(X_{i_1} - t < \cdots < X_{i_n} - t \,\middle|\, \cap_{i=1}^{n} X_i > t\right)$$

$$= \mathbb{P}(X_{i_1} < \cdots < X_{i_n}).$$

UiO : Department of Mathematics
University of Oslo

Anton Yurchenko-Tytarenko

# Lecture 14. Markov Chain Monte Carlo Methods. The Exponential Distribution

STK2130 – Modellering av stokastiske prosesser