Investigations on Stochastic Calculus,
Statistics of Processes
Applied Statistics for Biology and Medical Research

Nicolas SAVY
Mathematics Institute of Toulouse - University of Toulouse III

Habilitation thesis defence

Toulouse, Wednesday June, 18th

Table of contents of the manuscript

Introduction

A- Stochastic Calculus and Statistics of Processes

1. Malliavin Calculus and Anticipative Integrals
2. A limit theorem for filtered Poisson Processes
3. Transportation inequality and Malliavin Calculus
4. Properties of estimators for some diffusion processes

B- Applied statistics for Biology and Medical Research

5. Models for patients' recruitment in clinical trials
6. Survival data analysis for prevention Randomized Controlled Trials
7. Elements for analysing mediation and evolution in Epidemiology
8. Various results in interaction with Biology

C- Conclusions and perspectives

Table of contents of the manuscript

Introduction

A- Stochastic Calculus and Statistics of Processes

1. Malliavin Calculus and Anticipative Integrals
2. A limit theorem for filtered Poisson Processes
3. Transportation inequality and Malliavin Calculus
4. Properties of estimators for some diffusion processes

B- Applied statistics for Biology and Medical Research

5. Models for patients' recruitment in clinical trials
6. Survival data analysis for prevention Randomized Controlled Trials
7. Elements for analysing mediation and evolution in Epidemiology
8. Various results in interaction with Biology

C- Conclusions and perspectives

Definition (Filtered Processes)

Class of stochastic processes defined by:

$$\left\{ X_t^K = \int_0^t K(t,s)\, \mathrm{d}X_s \; : \; t \in [0,T] \right\}$$

where

- X is the so-called underlying process
  - Brownian motion, Poisson process or Lévy process
- K is a deterministic kernel
  - Triangular ($K(t,s) = 0$ as soon as $s > t > 0$)

- Covers a wide range of classic stochastic processes (Fractional Brownian motion, Shot noise processes,...)
- Integrates correlation in increments in standard settings

Definition (Filtered Processes)

Class of stochastic processes defined by:

$$\left\{ X_t^K = \int_0^t K(t,s)\,\mathrm{d}X_s \;:\; t \in [0,T] \right\}$$

where

- X is the so-called underlying process
  - Brownian motion, Poisson process or Lévy process
- K is a deterministic kernel
  - Triangular ($K(t,s) = 0$ as soon as $s > t > 0$)

- Covers a wide range of classic stochastic processes (Fractional Brownian motion, Shot noise processes,...)
- Integrates correlation in increments in standard settings

Aims: Define an integral with respect to Filtered Processes

Problems: In general not a Martingale nor a Markov process

A solution: Define integrals by means of Malliavin Calculus

The ingredients (for a process X):

- $\mathcal{S}^X$ a dense subset of $\mathbb{L}^2(\Omega)$ w.r. to an inner product $<,>_X$
- D^X called stochastic gradient defined on $\mathcal{S}^X$ its domain $\mathbb{D}^{1,2,X}$

Definition

A process $u \in \text{Dom}(\delta^X)$ if there exists a constant $C(u)$ s.t.

$$\left|\mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right]\right| \leq C(u)\|F\|_{L^2(\Omega)} \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \qquad (1)$$

$F \to \langle \mathrm{D}^X F, u\rangle_H$ is continuous, there exists $\delta^X(u)$ such that:

$$\mathbb{E}\left[F\,\delta^X(u)\right] = \mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right] \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \qquad (2)$$

Aims: Define an integral with respect to Filtered Processes

Problems: In general not a Martingale nor a Markov process

A solution: Define integrals by means of Malliavin Calculus

The ingredients (for a process X):

- $\mathcal{S}^X$ a dense subset of $\mathbb{L}^2(\Omega)$ w.r. to an inner product $<,>_X$
- D^X called stochastic gradient defined on $\mathcal{S}^X$ its domain $\mathbb{D}^{1,2,X}$

Definition

A process $u \in \mathrm{Dom}(\delta^X)$ if there exists a constant $C(u)$ s.t.

$$\left|\mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right]\right| \leq C(u)\|F\|_{L^2(\Omega)} \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \tag{1}$$

$F \to \langle \mathrm{D}^X F, u\rangle_H$ is continuous, there exists $\delta^X(u)$ such that:

$$\mathbb{E}\left[F\,\delta^X(u)\right] = \mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right] \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \tag{2}$$

Aims: Define an integral with respect to Filtered Processes

Problems: In general not a Martingale nor a Markov process

A solution: Define integrals by means of Malliavin Calculus

The ingredients (for a process X):

- $\mathcal{S}^X$ a dense subset of $\mathbb{L}^2(\Omega)$ w.r. to an inner product $<,>_X$
- D^X called stochastic gradient defined on $\mathcal{S}^X$ its domain $\mathbb{D}^{1,2,X}$

Definition

A process $u \in \text{Dom}(\delta^X)$ if there exists a constant $C(u)$ s.t.

$$\left|\mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right]\right| \leq C(u)\|F\|_{L^2(\Omega)} \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \qquad (1)$$

$F \rightarrow \langle \mathrm{D}^X F, u\rangle_H$ is continuous, there exists $\delta^X(u)$ such that:

$$\mathbb{E}\left[F\,\delta^X(u)\right] = \mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right] \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \qquad (2)$$

Aims: Define an integral with respect to Filtered Processes

Problems: In general not a Martingale nor a Markov process

A solution: Define integrals by means of Malliavin Calculus

The ingredients (for a process X):

- $\mathcal{S}^X$ a dense subset of $\mathbb{L}^2(\Omega)$ w.r. to an inner product $<,>_X$
- D^X called stochastic gradient defined on $\mathcal{S}^X$ its domain $\mathbb{D}^{1,2,X}$

Definition

A process $u \in \mathrm{Dom}(\delta^X)$ if there exists a constant $C(u)$ s.t.

$$\left|\mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right]\right| \leq C(u)\|F\|_{L^2(\Omega)} \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \qquad (1)$$

$F \to \langle \mathrm{D}^X F, u\rangle_H$ is continuous, there exists $\delta^X(u)$ such that:

$$\mathbb{E}\left[F\,\delta^X(u)\right] = \mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right] \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \qquad (2)$$

Aims: Define an integral with respect to Filtered Processes

Problems: In general not a Martingale nor a Markov process

A solution: Define integrals by means of Malliavin Calculus

The ingredients (for a process X):

- $\mathcal{S}^X$ a dense subset of $\mathbb{L}^2(\Omega)$ w.r. to an inner product $<,>_X$
- D^X called stochastic gradient defined on $\mathcal{S}^X$ its domain $\mathbb{D}^{1,2,X}$

Definition

A process $u \in \mathrm{Dom}(\delta^X)$ if there exists a constant $C(u)$ s.t.

$$\left|\mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right]\right| \leq C(u)\|F\|_{L^2(\Omega)} \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \tag{1}$$

$F \to \langle \mathrm{D}^X F, u\rangle_H$ is continuous, there exists $\delta^X(u)$ such that:

$$\mathbb{E}\left[F\,\delta^X(u)\right] = \mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right] \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \tag{2}$$

Aims: Define an integral with respect to Filtered Processes

Problems: In general not a Martingale nor a Markov process

A solution: Define integrals by means of Malliavin Calculus

The ingredients (for a process X):

- $\mathcal{S}^X$ a dense subset of $\mathbb{L}^2(\Omega)$ w.r. to an inner product $<,>_X$
- D^X called stochastic gradient defined on $\mathcal{S}^X$ its domain $\mathbb{D}^{1,2,X}$

Definition

A process $u \in \mathrm{Dom}(\delta^X)$ if there exists a constant $C(u)$ s.t.

$$\left|\mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right]\right| \leq C(u)\|F\|_{L^2(\Omega)} \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \tag{1}$$

$F \to \langle \mathrm{D}^X F, u\rangle_H$ is continuous, there exists $\delta^X(u)$ such that:

$$\mathbb{E}\left[F\,\delta^X(u)\right] = \mathbb{E}\left[\langle \mathrm{D}^X F, u\rangle_H\right] \qquad \text{for any } F \in \mathbb{D}^{1,2,X} \tag{2}$$

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

This strategy is possible for X a
  - Brownian motion
  - filtered Brownian motion
  - standard Poisson process
  - marked Poisson process
  - Lévy process

- δ^X_C construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ^X_G construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and "true" stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

  This strategy is possible for X a
  - Brownian motion
  - filtered Brownian motion
  - standard Poisson process
  - marked Poisson process
  - Lévy process

- δ_C^X construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ_G^X construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and "true" stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

This strategy is possible for X a
  - Brownian motion
  - filtered Brownian motion
  - standard Poisson process
  - marked Poisson process
  - Lévy process

- δ^X_C construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ^X_G construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and "true" stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

  This strategy is possible for X a
  - Brownian motion
  - filtered Brownian motion
  - standard Poisson process
  - marked Poisson process
  - Lévy process

- δ^X_C construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ^X_G construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and "true" stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

  This strategy is possible for X a
  - Brownian motion (Nualart (1995))
  - filtered Brownian motion (Alòs, Mazet, Nualart (2001))
  - standard Poisson process (Nualart, Vives (1990))
  - marked Poisson process (Oksendal et al. (2004))
  - Lévy process (Solé, Utzet, Vives (2007))
- δ_C^X construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ_G^X construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and "true" stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

  This strategy is possible for X a
  - Brownian motion (Üstünel (1995))
  - filtered Brownian motion (Decreusefond Üstünel (1999))
  - standard Poisson process (Carlen, Pardoux (1990))
  - marked Poisson process (Decreusefond, **Savy** (2006))
  - Lévy process (Leòn, Solé, Utzet, Vives (2013))
- δ_C^X construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ_G^X construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and "true" stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

- Relevant choices of $(\mathcal{S}^X, D^X)$ allow to interpret δ^X as an integral: **Skohorod integral**
  - For deterministic u, $\delta^X(u)$ coincides with X-Wiener integral
  - For predictable u, $\delta^X(u)$ coincides with Itô integral

  This strategy is possible for X a
  - Brownian motion (Üstünel (1995))
  - filtered Brownian motion (Decreusefond Üstünel (1999))
  - standard Poisson process (Carlen, Pardoux (1990))
  - marked Poisson process (Decreusefond, **Savy** (2006))
  - Lévy process (Leòn, Solé, Utzet, Vives (2013))

- δ_C^X construct by means of
  - $\mathcal{S}^X$ comes from chaos expansion of a r.v.
  - D^X is the chaos annihilation operator
- δ_G^X construct by means of
  - $(\mathcal{S}^X, D^X)$ are cylindrical variables and ”true” stochastic gradient

No direct approach for filtered Poisson and filtered Lévy processes

First approach: Directly

Available only for filtered Brownian motion

Second approach: Use of an operator

One can construct an operator $\mathcal{K}^*$ such that

$$\mathcal{K}^*(\,\mathbb{I}_{[0,t]}) = K(t,\cdot)\;\mathbb{I}_{[0,t]} \tag{3}$$

One can construct integrals w.r.t. to X^K by means of integrals w.r.t. to X:

First approach: Directly

Available only for filtered Brownian motion

Second approach: Use of an operator

One can construct an operator $\mathcal{K}^*$ such that

$$\mathcal{K}^*(\mathbb{I}_{[0,t]}) = K(t,\cdot)\ \mathbb{I}_{[0,t]} \tag{3}$$

One can construct integrals w.r. to X^K by means of integrals w.r. to X:

First approach: Directly

Available only for filtered Brownian motion

Second approach: Use of an operator

One can construct an operator $\mathcal{K}^*$ such that

$$\mathcal{K}^*(\mathbb{I}_{[0,t]}) = K(t,\cdot)\ \mathbb{I}_{[0,t]} \qquad (3)$$

One can construct integrals w.r.t. to X^K by means of integrals w.r.t. to X:

$$\int_0^T Y_s\, dX_s^K \stackrel{def}{=} \int_0^T \mathcal{K}^*(Y)_s\, dX_s$$

First approach: Directly

Available only for filtered Brownian motion

Second approach: Use of an operator

One can construct an operator $\mathcal{K}^*$ such that

$$\mathcal{K}^*(\mathbb{I}_{[0,t]}) = K(t,\cdot)\ \mathbb{I}_{[0,t]} \tag{3}$$

One can construct integrals w.r.t. to X^K by means of integrals w.r.t. to X:

$$\int_0^T Y_s\, dX_s^K \stackrel{def}{=} \int_0^T \mathcal{K}^*(Y)_s\, dX_s$$

$$\int_0^T \mathbb{I}_{[0,t]}\, dX_s^K = \int_0^T \mathcal{K}^*(\mathbb{I}_{[0,t]})_s\, dX_s$$

First approach: Directly

Available only for filtered Brownian motion

Second approach: Use of an operator

One can construct an operator $\mathcal{K}^*$ such that

$$\mathcal{K}^*(\mathbb{I}_{[0,t]}) = K(t,\cdot)\ \mathbb{I}_{[0,t]} \tag{3}$$

One can construct integrals w.r.t. to X^K by means of integrals w.r.t. to X:

$$\int_0^T Y_s\, dX_s^K \stackrel{def}{=} \int_0^T \mathcal{K}^*(Y)_s\, dX_s$$

$$\int_0^T \mathbb{I}_{[0,t]}\, dX_s^K = \int_0^T \mathcal{K}^*(\mathbb{I}_{[0,t]})_s\, dX_s$$

$$X_t^K = \int_0^t K(t,s)\, dX_s$$

Third approach: Use of an integral operator - $\mathcal{S}$-transform

- In the Brownian setting $X \equiv B$
  - (Bender (2003))
- In the (marked) Poisson setting $X \equiv N$
  - (Bender, Marquart (2008))
- In Lévy setting $X \equiv L$
  - (**Savy**, Vives (2014))

Conclusions:
We can define many integrals for filtered Poisson processes

- Are these definitions equivalent ?
- How they behave with Lévy-Itô decomposition $L = B + J$?

$$\delta^L(u) = \delta^B(u) + \delta^J(u)$$

- Are the components $\delta^B(u)$ and $\delta^J(u)$ still independent ?

Third approach: Use of an integral operator - $\mathcal{S}$-transform

- In the Brownian setting $X \equiv B$
  - (Bender (2003))
- In the (marked) Poisson setting $X \equiv N$
  - (Bender, Marquart (2008))
- In Lévy setting $X \equiv L$
  - (**Savy**, Vives (2014))

Conclusions:
We can define many integrals for filtered Poisson processes

- Are these definitions equivalent ?
- How they behave with Lévy-Itô decomposition $L = B + J$?

$$\delta^L(u) = \delta^B(u) + \delta^J(u)$$

- Are the components $\delta^B(u)$ and $\delta^J(u)$ still independent ?

Third approach: Use of an integral operator - $\mathcal{S}$-transform

- In the Brownian setting $X \equiv B$
  - (Bender (2003))
- In the (marked) Poisson setting $X \equiv N$
  - (Bender, Marquart (2008))
- In Lévy setting $X \equiv L$
  - (**Savy**, Vives (2014))

Conclusions:
We can define many integrals for filtered Poisson processes

- Are these definitions equivalent ?
- How they behave with Lévy-Itô decomposition $L = B + J$?

$$\delta^L(u) = \delta^B(u) + \delta^J(u)$$

- Are the components $\delta^B(u)$ and $\delta^J(u)$ still independent ?

Third approach: Use of an integral operator - $\mathcal{S}$-transform

- In the Brownian setting $X \equiv B$
  - (Bender (2003))
- In the (marked) Poisson setting $X \equiv N$
  - (Bender, Marquart (2008))
- In Lévy setting $X \equiv L$
  - (**Savy**, Vives (2014))

Conclusions:

We can define many integrals for filtered Poisson processes

- Are these definitions equivalent ?
- How they behave with Lévy-Itô decomposition $L = B + J$?

$$\delta^L(u) = \delta^B(u) + \delta^J(u)$$

- Are the components $\delta^B(u)$ and $\delta^J(u)$ still independent ?

Results (**Savy**, Vives (2014))

Defined by the use of	$X = L$	$X = B$	$X = J$	LIP true	LIP Ind.
Intrinsic Chaos		$\delta_C^{B,K}$		-	-
Intrinsic $\mathcal{S}$-transform	$\delta_S^{L,K}$	$\delta_S^{B,K}$	$\delta_S^{J,K}$		
$\mathcal{K}^*$ and Chaos	$\delta_C^{L,\mathcal{K}^*}$	$\delta_C^{B,\mathcal{K}^*}$	$\delta_C^{J,\mathcal{K}^*}$		
$\mathcal{K}^*$ and $\mathcal{S}$-transform	$\delta_S^{L,\mathcal{K}^*}$	$\delta_S^{B,\mathcal{K}^*}$	$\delta_S^{J,\mathcal{K}^*}$		

Results (**Savy**, Vives (2014))

Defined by the use of	$X = L$	$X = B$	$X = J$	LIP true	LIP Ind.
Intrinsic Chaos		$\delta_C^{B,K}$		-	-
		$\parallel$			
Intrinsic $\mathcal{S}$-transform	$\delta_S^{L,K}$	$\delta_S^{B,K}$	$\delta_S^{J,K}$		
	$\parallel$	$\parallel$	$\parallel$		
$\mathcal{K}^*$ and Chaos	$\delta_C^{L,\mathcal{K}^*}$	$\delta_C^{B,\mathcal{K}^*}$	$\delta_C^{J,\mathcal{K}^*}$		
	$\parallel$	$\parallel$	$\parallel$		
$\mathcal{K}^*$ and $\mathcal{S}$-transform	$\delta_S^{L,\mathcal{K}^*}$	$\delta_S^{B,\mathcal{K}^*}$	$\delta_S^{J,\mathcal{K}^*}$		

Results (**Savy**, Vives (2014))

Defined by the use of	$X = L$	$X = B$	$X = J$	LIP true	LIP Ind.
Intrinsic Chaos		$\delta_C^{B,K}$		-	-
		$\parallel$			
Intrinsic $\mathcal{S}$-transform	$\delta_S^{L,K}$	$\delta_S^{B,K}$	$\delta_S^{J,K}$	Yes	
	$\parallel$	$\parallel$	$\parallel$		
$\mathcal{K}^*$ and Chaos	$\delta_C^{L,\mathcal{K}^*}$	$\delta_C^{B,\mathcal{K}^*}$	$\delta_C^{J,\mathcal{K}^*}$	Yes	
	$\parallel$	$\parallel$	$\parallel$		
$\mathcal{K}^*$ and $\mathcal{S}$-transform	$\delta_S^{L,\mathcal{K}^*}$	$\delta_S^{B,\mathcal{K}^*}$	$\delta_S^{J,\mathcal{K}^*}$	Yes	

INSTITUT de MATHEMATIQUES de TOULOUSE

Results (Savy, Vives (2014))

Defined by the use of	$X = L$	$X = B$	$X = J$	LIP true	LIP Ind.
Intrinsic Chaos		$\delta_C^{B,K}$		-	-
		$\shortparallel$			
Intrinsic $\mathcal{S}$-transform	$\delta_S^{L,K}$	$\delta_S^{B,K}$	$\delta_S^{J,K}$	Yes	No
	$\shortparallel$	$\shortparallel$	$\shortparallel$		
$\mathcal{K}^*$ and Chaos	$\delta_C^{L,\mathcal{K}^*}$	$\delta_C^{B,\mathcal{K}^*}$	$\delta_C^{J,\mathcal{K}^*}$	Yes	No
	$\shortparallel$	$\shortparallel$	$\shortparallel$		
$\mathcal{K}^*$ and $\mathcal{S}$-transform	$\delta_S^{L,\mathcal{K}^*}$	$\delta_S^{B,\mathcal{K}^*}$	$\delta_S^{J,\mathcal{K}^*}$	Yes	No

Results (**Savy**, Vives (2014))

Defined by the use of	$X = L$	$X = B$	$X = J$	LIP true	LIP Ind.
Intrinsic Chaos		$\delta_C^{B,K}$		-	-
		$\|\|$			
Intrinsic $\mathcal{S}$-transform	$\delta_S^{L,K}$	$\delta_S^{B,K}$	$\delta_S^{J,K}$	Yes	No
	$\|\|$	$\|\|$	$\|\|$		
$\mathcal{K}^*$ and Chaos	$\delta_C^{L,\mathcal{K}^*}$	$\delta_C^{B,\mathcal{K}^*}$	$\delta_C^{J,\mathcal{K}^*}$	Yes	No
	$\|\|$	$\|\|$	$\|\|$		
$\mathcal{K}^*$ and $\mathcal{S}$-transform	$\delta_S^{L,\mathcal{K}^*}$	$\delta_S^{B,\mathcal{K}^*}$	$\delta_S^{J,\mathcal{K}^*}$	Yes	No

Remark

What about integrals defined by "true" stochastic gradient ?

- *In the Brownian case* $\delta_G^B = \delta_C^B$
- *Even in standard Poisson setting* $\delta_G^J(u) \neq \delta_C^J(u)$

LDP and SLDP for O-U processes

- Consider a fractional Ornstein Uhlenbeck process:

$$\begin{cases} dX_t &= \theta X_t dt + dB_t^H \\ X_0 &= 0 \end{cases} \tag{4}$$

- Consider the **MLE of θ** associated to (4)

$$\widehat{\theta}_T = \frac{\text{"} \int_0^T X_t \, dX_t \text{"}}{\int_0^T X_t^2 \, dt} \tag{5}$$

Theorem

- ***Strong Law of Large Number:*** $\widehat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$
- ***Central Limit Theorem*** *in the stable case* ($\theta < 0$)

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} N(0, -2\theta)$$

- ***Central Limit Theorem*** *in the unstable case* ($\theta > 0$)

$$\exp(\theta T)(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} 2\theta C \quad \text{where } C \text{ is Cauchy}$$

- Consider a fractional Ornstein Uhlenbeck process:

$$\begin{cases} dX_t &= \theta X_t dt + dB_t^H \\ X_0 &= 0 \end{cases} \qquad (4)$$

- Consider the **MLE of θ** associated to (4)

$$\widehat{\theta}_T = \frac{\text{''} \int_0^T X_t \, dX_t\text{''}}{\int_0^T X_t^2 \, dt} \qquad (5)$$

Theorem

- ***Strong Law of Large Number***: $\widehat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$
- ***Central Limit Theorem*** in the stable case $(\theta < 0)$

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, -2\theta)$$

- ***Central Limit Theorem*** in the unstable case $(\theta > 0)$

$$\exp(\theta T)(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} 2\theta C \qquad \text{where } C \text{ is Cauchy}$$

- Consider a fractional Ornstein Uhlenbeck process:

$$\begin{cases} dX_t & = \theta X_t dt + dB_t^H \\ X_0 & = 0 \end{cases} \quad (4)$$

- Consider the **MLE of θ** associated to (4)

$$\widehat{\theta}_T = \frac{\text{"} \int_0^T X_t \, dX_t \text{"}}{\int_0^T X_t^2 \, dt} \quad (5)$$

Theorem

- ***Strong Law of Large Number***: $\widehat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$
- ***Central Limit Theorem*** *in the stable case* $(\theta < 0)$

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, -2\theta)$$

- ***Central Limit Theorem*** *in the unstable case* $(\theta > 0)$

$$\exp(\theta T)(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} 2\theta\mathcal{C} \qquad \text{where } \mathcal{C} \text{ is Cauchy}$$

- Consider a fractional Ornstein Uhlenbeck process:

$$\begin{cases} dX_t &= \theta X_t dt + dB_t^H \\ X_0 &= 0 \end{cases} \quad (4)$$

- Consider the **MLE of θ** associated to (4)

$$\widehat{\theta}_T = \frac{\text{"} \int_0^T X_t \, dX_t \text{"}}{\int_0^T X_t^2 \, dt} \quad (5)$$

Theorem

- ***Strong Law of Large Number***: $\widehat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$
- ***Central Limit Theorem*** *in the stable case* $(\theta < 0)$

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, -2\theta)$$

- ***Central Limit Theorem*** *in the unstable case* $(\theta > 0)$

$$\exp(\theta T)(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} 2\theta\mathcal{C} \qquad \text{where } \mathcal{C} \text{ is Cauchy}$$

- Consider a fractional Ornstein Uhlenbeck process:

$$\begin{cases} dX_t & = \theta X_t dt + dB_t^H \\ X_0 & = 0 \end{cases} \tag{4}$$

- Consider the **MLE of θ** associated to (4)

$$\widehat{\theta}_T = \frac{\text{"} \int_0^T X_t\, dX_t \text{"}}{\int_0^T X_t^2\, dt} \tag{5}$$

Theorem

- ***Strong Law of Large Number***: $\widehat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$
- ***Central Limit Theorem*** *in the stable case* ($\theta < 0$)

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, -2\theta)$$

- ***Central Limit Theorem*** *in the unstable case* ($\theta > 0$)

$$\exp(\theta T)(\widehat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathcal{L}} 2\theta\mathcal{C} \qquad \text{where } \mathcal{C} \text{ is Cauchy}$$

Asymptotic behaviour of $(\widehat{\theta}_T)$ in terms of **Large Deviation Principle**

Definition

- A family of random variables (Z_T) satisfies a **LDP** of good rate function I if there exists a function I l.s.c. from $\mathbb{R}$ to $[0, +\infty]$ s.t.

$$\lim_{T \to \infty} \frac{1}{T} \log \mathbb{P}\left[Z_T \geq c\right] = -I(c), \quad \text{for all } c \geq \mathbb{E}\left[Z_T\right]$$

- If I is regular and strictly convex then I expresses as the Fenchel-Legendre dual of the limit $\mathcal{L}$ of the log-Laplace transform of Z_T:

$$I(c) := \sup_{t \in \mathbb{R}} \left[ct - \mathcal{L}(t)\right].$$

Setting $(\theta < 0, H = \frac{1}{2})$

Setting $(\theta > 0, H = \frac{1}{2})$

Setting $(\theta < 0, H \neq \frac{1}{2})$

Asymptotic behaviour of $(\widehat{\theta}_T)$ in terms of **Large Deviation Principle**

Definition

- A family of random variables (Z_T) satisfies a **LDP** of good rate function I if there exists a function I l.s.c. from $\mathbb{R}$ to $[0, +\infty]$ s.t.

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}\left[Z_T \geq c\right] = -I(c), \quad \text{for all } c \geq \mathbb{E}\left[Z_T\right]$$

- If I is regular and strictly convex then I expresses as the Fenchel-Legendre dual of the limit $\mathcal{L}$ of the log-Laplace transform of Z_T:

$$I(c) := \sup_{t\in\mathbb{R}} \left[ct - \mathcal{L}(t)\right].$$

Setting $(\theta < 0, H = \frac{1}{2})$ (Bercu, Rouault (2002))

Setting $(\theta > 0, H = \frac{1}{2})$ (Bercu, Coutin, Savy (2012))

Setting $(\theta < 0, H \neq \frac{1}{2})$ (Bercu, Coutin, Savy (2011))

Asymptotic behaviour of $(\widehat{\theta}_T)$ in terms of **Large Deviation Principle**

Definition

- A family of random variables (Z_T) satisfies a **LDP** of good rate function I if there exists a function I l.s.c. from $\mathbb{R}$ to $[0, +\infty]$ s.t.

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}\left[Z_T \geq c\right] = -I(c), \quad \text{for all } c \geq \mathbb{E}\left[Z_T\right]$$

- If I is regular and strictly convex then I expresses as the Fenchel-Legendre dual of the limit $\mathcal{L}$ of the log-Laplace transform of Z_T:

$$I(c) := \sup_{t\in\mathbb{R}} \left[ct - \mathcal{L}(t)\right].$$

Setting $(\theta < 0, H = \frac{1}{2})$ (Bercu, Rouault (2002))

Setting $(\theta > 0, H = \frac{1}{2})$ (Bercu, Coutin, Savy (2012))

Setting $(\theta < 0, H \neq \frac{1}{2})$ (Bercu, Coutin, Savy (2011))

Asymptotic behaviour of $(\widehat{\theta}_T)$ in terms of **Large Deviation Principle**

Definition

- A family of random variables (Z_T) satisfies a **LDP** of good rate function I if there exists a function I l.s.c. from $\mathbb{R}$ to $[0, +\infty]$ s.t.

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}[Z_T \geq c] = -I(c), \quad \text{for all } c \geq \mathbb{E}[Z_T]$$

- If I is regular and strictly convex then I expresses as the Fenchel-Legendre dual of the limit $\mathcal{L}$ of the log-Laplace transform of Z_T:

$$I(c) := \sup_{t\in\mathbb{R}} [ct - \mathcal{L}(t)].$$

Setting $(\theta < 0, H = \frac{1}{2})$ (Bercu, Rouault (2002))

Setting $(\theta > 0, H = \frac{1}{2})$ (Bercu, Coutin, **Savy** (2012))

Setting $(\theta < 0, H \neq \frac{1}{2})$ (Bercu, Coutin, **Savy** (2011))

Asymptotic behaviour of $(\widehat{\theta}_T)$ in terms of **Large Deviation Principle**

Definition

- A family of random variables (Z_T) satisfies a **LDP** of good rate function I if there exists a function I l.s.c. from $\mathbb{R}$ to $[0, +\infty]$ s.t.

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}\left[Z_T \geq c\right] = -I(c), \quad \text{for all } c \geq \mathbb{E}\left[Z_T\right]$$

- If I is regular and strictly convex then I expresses as the Fenchel-Legendre dual of the limit $\mathcal{L}$ of the log-Laplace transform of Z_T:

$$I(c) := \sup_{t\in\mathbb{R}} \left[ct - \mathcal{L}(t)\right].$$

Setting $(\theta < 0, H = \frac{1}{2})$ (Bercu, Rouault (2002))

Setting $(\theta > 0, H = \frac{1}{2})$ (Bercu, Coutin, **Savy** (2012))

Setting $(\theta < 0, H \neq \frac{1}{2})$ (Bercu, Coutin, **Savy** (2011))

Asymptotic behaviour of $(\widehat{\theta}_T)$ in terms of **Large Deviation Principle**

Definition

- A family of random variables (Z_T) satisfies a **LDP** of good rate function I if there exists a function I l.s.c. from $\mathbb{R}$ to $[0, +\infty]$ s.t.

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}\left[Z_T \geq c\right] = -I(c), \quad \text{for all } c \geq \mathbb{E}\left[Z_T\right]$$

- If I is regular and strictly convex then I expresses as the Fenchel-Legendre dual of the limit $\mathcal{L}$ of the log-Laplace transform of Z_T:

$$I(c) := \sup_{t\in\mathbb{R}} \left[ct - \mathcal{L}(t)\right].$$

Setting $(\theta < 0, H = \frac{1}{2})$ (Bercu, Rouault (2002))

Setting $(\theta > 0, H = \frac{1}{2})$ (Bercu, Coutin, **Savy** (2012))

Setting $(\theta < 0, H \neq \frac{1}{2})$ (Bercu, Coutin, **Savy** (2011))

First, notice that, for any c,

$$\left\{\widehat{\theta}_T \leq c\right\} \Longleftrightarrow \left\{\frac{\text{"}\int_0^T X_t\, dX_t\text{"}}{\int_0^T X_t^2\, dt} \leq c\right\} \Longleftrightarrow \{Z_T(a,c) \leq 0\}$$

where

$$Z_T(a,c) = a\int_0^T X_t\, dX_t - ac\int_0^T X_t^2\, dt.$$

First, notice that, for any c,

$$\left\{\widehat{\theta}_T \leq c\right\} \Longleftrightarrow \left\{\frac{\text{“}\int_0^T X_t\, dX_t\text{”}}{\int_0^T X_t^2\, dt} \leq c\right\} \Longleftrightarrow \{Z_T(a,c) \leq 0\}$$

where

$$Z_T(a,c) = a\int_0^T X_t\, dX_t - ac\int_0^T X_t^2\, dt.$$

First, notice that, for any c,

$$\left\{\widehat{\theta}_T \leq c\right\} \Longleftrightarrow \left\{\frac{\text{"}\int_0^T X_t\, dX_t\text{"}}{\int_0^T X_t^2\, dt} \leq c\right\} \Longleftrightarrow \{Z_T(a,c) \leq 0\}$$

where

$$Z_T(a,c) = a\int_0^T Q_t^H\, dY_t^H - ac\int_0^T (Q_t^H)^2\, d < M^H >_t .$$

LDP for O-U processes

First, notice that, for any c,

$$\left\{\widehat{\theta}_T \leq c\right\} \Longleftrightarrow \left\{\frac{\text{''}\int_0^T X_t\, dX_t\text{''}}{\int_0^T X_t^2\, dt} \leq c\right\} \Longleftrightarrow \{Z_T(a,c) \leq 0\}$$

where

$$Z_T(a,c) = a\int_0^T Q_t^H\, dY_t^H - ac\int_0^T (Q_t^H)^2\, d < M^H >_t .$$

To establish an LDP, we have to study the Laplace transform

$$\mathcal{L}_T(a,c) = \frac{1}{T}\log\mathbb{E}\left[\exp(Z_T(a,c))\right]$$

especially the description of the domain Δ_c of the limit $\mathcal{L}$ of $\mathcal{L}_T$

First, notice that, for any c,

$$\left\{\widehat{\theta}_T \leq c\right\} \Longleftrightarrow \left\{ \frac{\text{''} \int_0^T X_t \, dX_t \text{''}}{\int_0^T X_t^2 \, dt} \leq c \right\} \Longleftrightarrow \{Z_T(a,c) \leq 0\}$$

where

$$Z_T(a,c) = a \int_0^T Q_t^H \, dY_t^H - ac \int_0^T (Q_t^H)^2 \, d < M^H >_t .$$

To establish an LDP, we have to study the Laplace transform

$$\mathcal{L}_T(a,c) = \frac{1}{T} \log \mathbb{E}\left[\exp(Z_T(a,c))\right]$$

especially the description of the domain Δ_c of the limit $\mathcal{L}$ of $\mathcal{L}_T$

$$I(c) = -\inf_{a \in \overline{\Delta_c}} \mathcal{L}(a) \tag{6}$$

and denote $a_c = \operatorname{argmin}_{a \in \overline{\Delta_c}} \mathcal{L}(a)$ (7)

Lemma

- *The limit $\mathcal{L}$ of $\mathcal{L}_T$ is*

$$\mathcal{L}(a) = -\frac{1}{2}(a + \theta + \varphi(a))$$

with

$$\varphi(a) = \begin{cases} \sqrt{\theta^2 + 2ac} & \textbf{for } (\theta < 0) \\ -\sqrt{\theta^2 + 2ac} & \textbf{for } (\theta > 0) \end{cases}$$

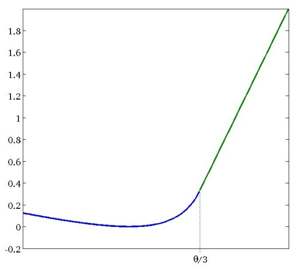
- *its domain Δ_c is*

$$\left\{ a \in \mathbb{R} \text{ s.t. } \theta^2 + 2ac > 0 \text{ and } \sqrt{\theta^2 + 2ac} > \max(a+\theta; -\delta_H(a+\theta)) \right\}$$

Theorem (Stable setting $\theta < 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\dfrac{(c-\theta)^2}{4c} & \text{if } c < \theta/3 \\ 2c - \theta & \text{if } c \geq \theta/3 \end{cases}$$



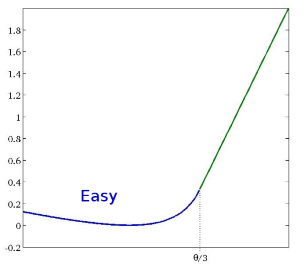
For $H = \frac{1}{2}$ (Florens-Landais, Pham (1999))

For $H \neq \frac{1}{2}$ (Bercu, Coutin, **Savy** (2011))

Theorem (Stable setting $\theta < 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\dfrac{(c-\theta)^2}{4c} & \text{if } c < \theta/3 \\ 2c - \theta & \text{if } c \geq \theta/3 \end{cases}$$



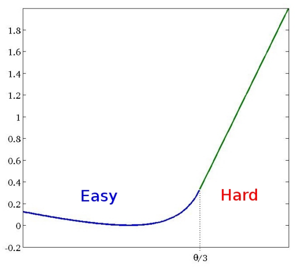
For $H = \frac{1}{2}$ (Florens-Landais, Pham (1999))

For $H \neq \frac{1}{2}$ (Bercu, Coutin, **Savy** (2011))

Theorem (Stable setting $\theta < 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\dfrac{(c-\theta)^2}{4c} & \text{if } c < \theta/3 \\ 2c - \theta & \text{if } c \geq \theta/3 \end{cases}$$



For $H = \frac{1}{2}$ (Florens-Landais, Pham (1999))

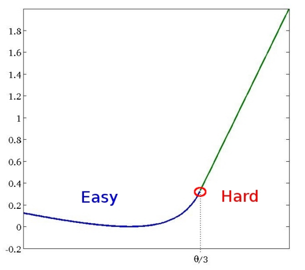
For $H \neq \frac{1}{2}$ (Bercu, Coutin, **Savy** (2011))

LDP for O-U processes

Theorem (Stable setting $\theta < 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\dfrac{(c-\theta)^2}{4c} & \text{if } c < \theta/3 \\ 2c - \theta & \text{if } c \geq \theta/3 \end{cases}$$



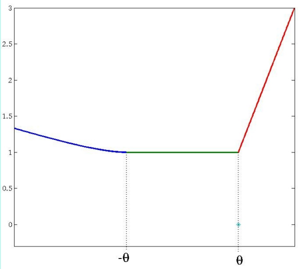
For $H = \frac{1}{2}$ (Florens-Landais, Pham (1999))

For $H \neq \frac{1}{2}$ (Bercu, Coutin, **Savy** (2011))

Theorem (Unstable setting $\theta > 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\frac{(c-\theta)^2}{4c} & \text{if } c \leq -\theta, \\ \theta & \text{if } |c| < \theta, \\ 0 & \text{if } c = \theta, \\ 2c - \theta & \text{if } c > \theta. \end{cases}$$

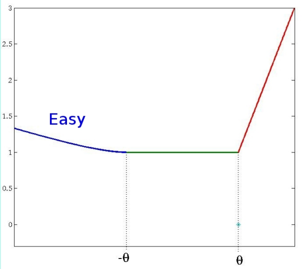


For $H = \frac{1}{2}$ (Bercu, Coutin, **Savy** (2012))

Theorem (Unstable setting $\theta > 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\frac{(c-\theta)^2}{4c} & \text{if } c \leq -\theta, \\ \theta & \text{if } |c| < \theta, \\ 0 & \text{if } c = \theta, \\ 2c - \theta & \text{if } c > \theta. \end{cases}$$

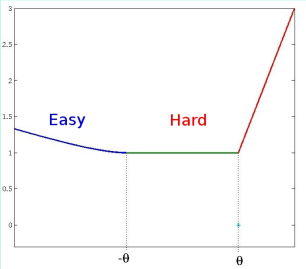


For $H = \frac{1}{2}$ (Bercu, Coutin, **Savy** (2012))

Theorem (Unstable setting $\theta > 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\frac{(c-\theta)^2}{4c} & \text{if } c \leq -\theta, \\ \theta & \text{if } |c| < \theta, \\ 0 & \text{if } c = \theta, \\ 2c - \theta & \text{if } c > \theta. \end{cases}$$

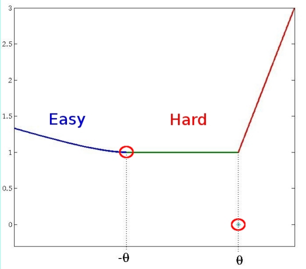


For $H = \frac{1}{2}$ (Bercu, Coutin, **Savy** (2012))

Theorem (Unstable setting $\theta > 0$)

The sequence $(\widehat{\theta}_T)$ satisfies an **LDP** *with rate function*

$$I(c) = \begin{cases} -\frac{(c-\theta)^2}{4c} & \text{if } c \leq -\theta, \\ \theta & \text{if } |c| < \theta, \\ 0 & \text{if } c = \theta, \\ 2c - \theta & \text{if } c > \theta. \end{cases}$$



For $H = \frac{1}{2}$ (Bercu, Coutin, **Savy** (2012))

To establish an SLDP we have to study deeper the behaviour of $\mathcal{L}$

- An expansion of $\mathcal{L}_T$ for any a in the interior of Δ_c

Lemma

For any a in the interior of Δ_c, we have the expansion :

$$\mathcal{L}_T(a) = \mathcal{L}(a) + \frac{1}{T}\mathcal{H}(a) + \frac{1}{T}\mathcal{R}_T(a)$$

where $\mathcal{H}(a) = -\frac{1}{2}\log\left(\frac{\varphi(a)-(a+\theta)}{2\varphi(a)}\right)$ and $\mathcal{R}_T(a)$ is a remainder term

- Analysis of the behaviour at a_c

Lemma

- ***Easy case:*** *no problem $a_c \in \Delta_c$*
- ***Hard case:*** *$a_c \notin \Delta_c$ but there is a family a_T such that $a_T \in \Delta_c$ for any T, $a_T \xrightarrow[T\to\infty]{} a_c$ and for T large enough,*

$$a_T = \sum_{k=0}^{p} \frac{a_k}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right).$$

To establish an SLDP we have to study deeper the behaviour of $\mathcal{L}$

- An expansion of $\mathcal{L}_T$ for any a in the interior of Δ_c

Lemma

For any a in the interior of Δ_c, we have the expansion :

$$\mathcal{L}_T(a) = \mathcal{L}(a) + \frac{1}{T}\mathcal{H}(a) + \frac{1}{T}\mathcal{R}_T(a)$$

where $\mathcal{H}(a) = -\frac{1}{2}\log\left(\frac{\varphi(a)-(a+\theta)}{2\varphi(a)}\right)$ and $\mathcal{R}_T(a)$ is a remainder term

- Analysis of the behaviour at a_c

Lemma

- ***Easy case:*** *no problem $a_c \in \Delta_c$*
- ***Hard case:*** *$a_c \notin \Delta_c$ but there is a family a_T such that $a_T \in \Delta_c$ for any T, $a_T \xrightarrow[T\to\infty]{} a_c$ and for T large enough,*

$$a_T = \sum_{k=0}^{p} \frac{a_k}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right).$$

To establish an SLDP we have to study deeper the behaviour of $\mathcal{L}$

- An expansion of $\mathcal{L}_T$ for any a in the interior of Δ_c

Lemma

For any a in the interior of Δ_c, we have the expansion :

$$\mathcal{L}_T(a) = \mathcal{L}(a) + \frac{1}{T}\mathcal{H}(a) + \frac{1}{T}\mathcal{R}_T(a)$$

where $\mathcal{H}(a) = -\frac{1}{2}\log\left(\frac{\varphi(a)-(a+\theta)}{2\varphi(a)}\right)$ *and* $\mathcal{R}_T(a)$ *is a remainder term*

- Analysis of the behaviour at a_c

Lemma

- ***Easy case:*** *no problem* $a_c \in \Delta_c$
- ***Hard case:*** $a_c \notin \Delta_c$ *but there is a family* a_T *such that* $a_T \in \Delta_c$ *for any* T, $a_T \xrightarrow[T\to\infty]{} a_c$ *and for* T *large enough,*

$$a_T = \sum_{k=0}^{p} \frac{a_k}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right).$$

To establish an SLDP we have to study deeper the behaviour of $\mathcal{L}$

- An expansion of $\mathcal{L}_T$ for any a in the interior of Δ_c

Lemma

For any a in the interior of Δ_c, we have the expansion :

$$\mathcal{L}_T(a) = \mathcal{L}(a) + \frac{1}{T}\mathcal{H}(a) + \frac{1}{T}\mathcal{R}_T(a)$$

where $\mathcal{H}(a) = -\frac{1}{2}\log\left(\frac{\varphi(a)-(a+\theta)}{2\varphi(a)}\right)$ *and* $\mathcal{R}_T(a)$ *is a remainder term*

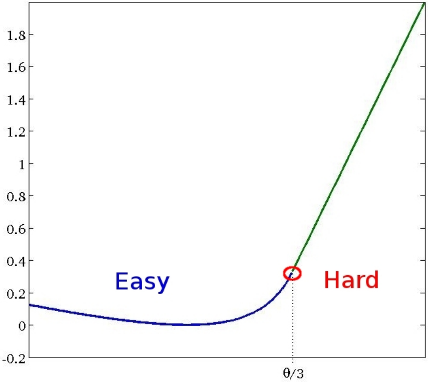
- Analysis of the behaviour at a_c

Lemma

- ***Easy case:*** *no problem* $a_c \in \Delta_c$
- ***Hard case:*** $a_c \notin \Delta_c$ *but there is a family* a_T *such that* $a_T \in \Delta_c$ *for any* T, $a_T \xrightarrow[T\to\infty]{} a_c$ *and for* T *large enough,*

$$a_T = \sum_{k=0}^{p} \frac{a_k}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right).$$

Stable Case



Theorem (Setting ($\theta < 0, H = \frac{1}{2}$) - (Bercu, Rouault (2002)))

The sequence $(\widehat{\theta}_T)$ satisfies an ***SLDP****.*

- ***For all*** $c < \theta/3$,

  it exists a sequence $(d_{c,k})$ such that, for any $p \geq 1$ and T large enough

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c) + K(c))}{\sigma_c t_c \sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{d_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

$$t_c = \frac{c^2 - \theta^2}{2c}, \quad \sigma_c^2 = -\frac{1}{2c}, \quad K(c) = -\frac{1}{2}\log\left(\frac{(c+\theta)(3c-\theta)}{4c^2}\right)$$

- ***For*** $c > \theta/3$ ***with*** $c \neq 0$,

  similar expansion holds

Theorem (Setting ($\theta < 0$, $H = \frac{1}{2}$) - (Bercu, Rouault (2002)))

The sequence $(\widehat{\theta}_T)$ satisfies an ***SLDP****.*

- ***For all*** $c < \theta/3$,

  it exists a sequence $(d_{c,k})$ such that, for any $p \geq 1$ and T large enough

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c) + K(c))}{\sigma_c t_c \sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{d_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

$$t_c = \frac{c^2 - \theta^2}{2c}, \quad \sigma_c^2 = -\frac{1}{2c}, \quad K(c) = -\frac{1}{2}\log\left(\frac{(c+\theta)(3c-\theta)}{4c^2}\right)$$

- ***For*** $c > \theta/3$ ***with*** $c \neq 0$,

  similar expansion holds

Theorem (Setting ($\theta < 0, H = \frac{1}{2}$) - (Bercu, Rouault (2002)))

- ***For*** $c = \theta/3$,

  it exists a sequence (d_k) *such that, for any* $p \geq 1$ *and* T *large enough*

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c))}{4\pi\tau_\theta}\frac{1}{T^{1/4}}\left[1+\sum_{k=1}^{2p}\frac{d_k}{(\sqrt{T})^k}+\mathcal{O}\left(\frac{1}{T^p\sqrt{T}}\right)\right]$$

  where $\tau_\theta = (-\theta/3)^{1/4}/\Gamma(1/4)$.

- ***For*** $c = 0$, *it exists a sequence* (b_k) *such that, for any* $p \geq 1$ *and* T *large enough*

$$\mathbb{P}\left[\widehat{\theta}_T \geq 0\right] = \frac{\exp(\theta T)}{\sqrt{\pi}\sqrt{-\theta}}\frac{1}{\sqrt{T}}\left[1+\sum_{k=1}^{p}\frac{b_k}{T^k}+\mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right].$$

Theorem (Setting ($\theta < 0, H = \frac{1}{2}$) - (Bercu, Rouault (2002)))

- ***For $c = \theta/3$,***

  it exists a sequence (d_k) such that, for any $p \geq 1$ and T large enough

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c))}{4\pi\tau_\theta} \frac{1}{T^{1/4}} \left[1 + \sum_{k=1}^{2p} \frac{d_k}{(\sqrt{T})^k} + \mathcal{O}\left(\frac{1}{T^p\sqrt{T}}\right)\right]$$

  where $\tau_\theta = (-\theta/3)^{1/4}/\Gamma(1/4)$.

- ***For $c = 0$,*** *it exists a sequence (b_k) such that, for any $p \geq 1$ and T large enough*

$$\mathbb{P}\left[\widehat{\theta}_T \geq 0\right] = \frac{\exp(\theta T)}{\sqrt{\pi}\sqrt{-\theta}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{b_k}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right].$$

INSTITUT de MATHEMATIQUES de TOULOUSE

Theorem (Setting ($\theta < 0$, $H \neq \frac{1}{2}$) - (Bercu, Coutin, Savy (2011)))

The sequence $(\widehat{\theta}_T)$ satisfies an ***SLDP****.*

- ***For all*** $c < \theta/3$,

  it exists a sequence $(b^H_{c,k})$ *such that, for any* $p > 0$ *and* T *large enough,*

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c) + K_H(c))}{\sigma_c t_c \sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{b^H_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

$$t_c = \frac{c^2 - \theta^2}{2c}, \quad \sigma_c^2 = -\frac{1}{2c}$$

$$K_H(c) = K(c) - \frac{1}{2}\log\left(1 + \frac{1 - \sin(\pi H)}{\sin(\pi H)} \frac{(c-\theta)^2}{4c^2}\right).$$

- ***For*** $c > \theta/3$ ***with*** $c \neq 0$,

  similar expansion holds.

Theorem (Setting ($\theta < 0, H \neq \frac{1}{2}$) - (Bercu, Coutin, Savy (2011)))

The sequence $(\widehat{\theta}_T)$ satisfies an ***SLDP****.*

- ***For all*** $c < \theta/3$,

  it exists a sequence $(b^H_{c,k})$ *such that, for any* $p > 0$ *and* T *large enough,*

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c)+K_H(c))}{\sigma_c t_c \sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{b^H_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

$$t_c = \frac{c^2 - \theta^2}{2c}, \quad \sigma_c^2 = -\frac{1}{2c}$$

$$K_H(c) = K(c) - \frac{1}{2}\log\left(1 + \frac{1-\sin(\pi H)}{\sin(\pi H)} \frac{(c-\theta)^2}{4c^2}\right).$$

- ***For*** $c > \theta/3$ ***with*** $c \neq 0$,

  similar expansion holds.

Theorem (Setting ($\theta < 0, H \neq \frac{1}{2}$) - (Bercu, Coutin, Savy (2011)))

- ***For $c = \theta/3$,***

  it exists a sequence (d_k^H) such that, for any $p \geq 1$ and T large enough,

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c))c_H}{4\pi\tau_\theta}\frac{1}{T^{1/4}}\left[1 + \sum_{k=1}^{2p}\frac{d_k^H}{(\sqrt{T})^k} + \mathcal{O}\left(\frac{1}{T^p\sqrt{T}}\right)\right]$$

$$\tau_\theta = (-\theta/3)^{1/4}/\Gamma(1/4), \qquad c_H = \sqrt{\sin(\pi H)}$$

- ***For $c = 0$,***

  it exists a sequence (b_k^H) such that, for any $p \geq 1$ and T large enough,

$$\mathbb{P}\left[\widehat{\theta}_T \geq 0\right] = \frac{\exp(\theta T)\sqrt{\sin(\pi H)}}{\sqrt{\pi}\sqrt{-\theta}}\frac{1}{\sqrt{T}}\left[1 + \sum_{k=1}^{p}\frac{b_k^H}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right].$$

Theorem (Setting ($\theta < 0, H \neq \frac{1}{2}$) - (Bercu, Coutin, Savy (2011)))

- ***For*** $c = \theta/3$,

  it exists a sequence (d_k^H) *such that, for any* $p \geq 1$ *and* T *large enough,*

$$\mathbb{P}\left[\widehat{\theta}_T \geq c\right] = \frac{\exp(-TI(c))c_H}{4\pi\tau_\theta}\frac{1}{T^{1/4}}\left[1 + \sum_{k=1}^{2p}\frac{d_k^H}{(\sqrt{T})^k} + \mathcal{O}\left(\frac{1}{T^p\sqrt{T}}\right)\right]$$

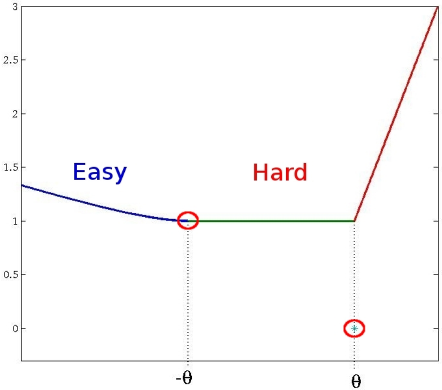
$$\tau_\theta = (-\theta/3)^{1/4}/\Gamma(1/4), \qquad c_H = \sqrt{\sin(\pi H)}$$

- ***For*** $c = 0$,

  it exists a sequence (b_k^H) *such that, for any* $p \geq 1$ *and* T *large enough,*

$$\mathbb{P}\left[\widehat{\theta}_T \geq 0\right] = \frac{\exp(\theta T)\sqrt{\sin(\pi H)}}{\sqrt{\pi}\sqrt{-\theta}}\frac{1}{\sqrt{T}}\left[1 + \sum_{k=1}^{p}\frac{b_k^H}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right].$$

Unstable Case



Theorem (Setting ($\theta > 0$, $H = \frac{1}{2}$) - (Bercu, Coutin, Savy (2012)))

*The sequence $(\widehat{\theta}_T)$ satisfies an **SLDP**.*

- ***For all $c < -\theta$,***

  there exists a sequence $(e_{c,k})$ such that, for any $p > 0$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \leq c\right] = \frac{\exp(-TI(c) + K(c))}{-a_c \sigma_c \sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{e_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

  where $a_c = \frac{c^2 - \theta^2}{2c}$, $\sigma_c^2 = -\frac{1}{2c}$ and $K(c)$ a constant

- ***For all $c > \theta$,***

  there exists a sequence $(f_{c,k})$ such that, for any $p > 0$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \geq c\right] = \frac{\exp(-TI(c) + K(c))}{a_c \sigma_c \sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{f_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

  where $a_c = 2(c - \theta)$, $\sigma_c^2 = \frac{c^2}{2(2c-\theta)^3}$ and $K(c)$ a constant

Theorem (Setting ($\theta > 0, H = \frac{1}{2}$) - (Bercu, Coutin, Savy (2012)))

The sequence $(\widehat{\theta}_T)$ satisfies an ***SLDP****.*

- ***For all*** $c < -\theta$,

  there exists a sequence $(e_{c,k})$ such that, for any $p > 0$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \leq c\right] = \frac{\exp(-TI(c) + K(c))}{-a_c\sigma_c\sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{e_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

  where $a_c = \frac{c^2-\theta^2}{2c}$, $\sigma_c^2 = -\frac{1}{2c}$ *and $K(c)$ a constant*

- ***For all*** $c > \theta$,

  there exists a sequence $(f_{c,k})$ such that, for any $p > 0$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \geq c\right] = \frac{\exp(-TI(c) + K(c))}{a_c\sigma_c\sqrt{2\pi}} \frac{1}{\sqrt{T}} \left[1 + \sum_{k=1}^{p} \frac{f_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

  where $a_c = 2(c-\theta)$, $\sigma_c^2 = \frac{c^2}{2(2c-\theta)^3}$ *and $K(c)$ a constant*

Theorem (Setting ($\theta > 0, H = \frac{1}{2}$) - (Bercu, Coutin, Savy (2012)))

- ***For all $|c| < \theta$ and $c \neq 0$,***

  there exists a sequence $(g_{c,k})$ such that, for any $p > 0$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \leq c\right] = \frac{\exp(-TI(c) + K(c))}{a_c \sigma_c \sqrt{2\pi}} \sqrt{T} \left[1 + \sum_{k=1}^{p} \frac{g_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

  where $a_c = \frac{\theta}{c+\theta}$, $\sigma_c^2 = \frac{c^2}{2\theta^3}$ and $K(c)$ a constant.

- ***For $c = -\theta$,***

  it exists a sequence (d_k) such that, for any $p \geq 1$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \leq -\theta\right] = \frac{\exp(-TI(c))\Gamma\left(\frac{1}{4}\right)}{2\sqrt{2\pi} a_\theta^{\frac{1}{2}} \sigma_\theta} T^{\frac{1}{4}} \left[1 + \sum_{k=1}^{2p} \frac{h_k}{(\sqrt{T})^k} + \mathcal{O}\left(\frac{1}{T^p \sqrt{T}}\right)\right]$$

  where $a_\theta = \sqrt{\theta}$ and $\sigma_\theta^2 = \frac{1}{2\theta}$.

Theorem (Setting ($\theta > 0$, $H = \frac{1}{2}$) - (Bercu, Coutin, Savy (2012)))

- ***For all $|c| < \theta$ and $c \neq 0$,***

  there exists a sequence $(g_{c,k})$ such that, for any $p > 0$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \leq c\right] = \frac{\exp(-TI(c) + K(c))}{a_c \sigma_c \sqrt{2\pi}} \sqrt{T} \left[1 + \sum_{k=1}^{p} \frac{g_{c,k}}{T^k} + \mathcal{O}\left(\frac{1}{T^{p+1}}\right)\right]$$

  where $a_c = \frac{\theta}{c+\theta}$, $\sigma_c^2 = \frac{c^2}{2\theta^3}$ and $K(c)$ a constant.

- ***For $c = -\theta$,***

  it exists a sequence (d_k) such that, for any $p \geq 1$ and T large enough,

$$\mathbb{P}\left[\hat{\theta}_T \leq -\theta\right] = \frac{\exp(-TI(c))\Gamma\left(\frac{1}{4}\right)}{2\sqrt{2}\pi a_\theta^{\frac{1}{2}} \sigma_\theta} T^{\frac{1}{4}} \left[1 + \sum_{k=1}^{2p} \frac{h_k}{(\sqrt{T})^k} + \mathcal{O}\left(\frac{1}{T^p \sqrt{T}}\right)\right]$$

  where $a_\theta = \sqrt{\theta}$ and $\sigma_\theta^2 = \frac{1}{2\theta}$.

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially:
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially:
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially:
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially:
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

What is recruitment period ?

- Clinical trials is one of the main elements of the marketing authorization of a new drug
- Such a request has to follow a protocol specifying
  - Patients inclusion and exclusion criteria
  - Statistic analysis plan especially:
    - which test is used
    - what are the type I and type II risks
    - **necessary sample size N**
- In order to recruit these N patients, several investigators centres are involved

Definition

The **recruitment period** is the duration between the initiation of the first of the C investigator centres and the instant $T(N)$ when the N patients are included.

- N is fixed but $T(N)$ is a random variable

Motivation of those investigations

- **Why a model of recruitment period ?**
  - The duration of the recruitment period is very **hard to control**
  - A clinical trial is **expensive**
    - $ 150.000.000: Average out-of-pocket clinical cost for each new drug
  - Pharma-Companies need tools to **be able** to decide:
    - to overpass the targeted duration of the trial T_R
    - stop the trial if it is too long

- **What a model of recruitment for ?**
  - To develop tools for the study the feasibility of a clinical trial
    - based on the estimation of $T(N)$ (punctually and by means of CI)
  - To Detect critical point in the recruitment
  - To define decision rules on the recruitment process to reach T_R
    - based on the estimation of the recruitment rate
    - based on the estimation of the number of centre to open

Motivation of those investigations

- **Why a model of recruitment period ?**
  - The duration of the recruitment period is very **hard to control**
  - A clinical trial is **expensive**
    - $ 150.000.000: Average out-of-pocket clinical cost for each new drug
  - Pharma-Companies need tools to **be able** to decide:
    - to overpass the targeted duration of the trial T_R
    - stop the trial if it is too long
- **What a model of recruitment for ?**
  - To develop tools for the study the feasibility of a clinical trial
    - based on the estimation of $T(N)$ (punctually and by means of CI)
  - To Detect critical point in the recruitment
  - To define decision rules on the recruitment process to reach T_R
    - based on the estimation of the recruitment rate
    - based on the estimation of the number of centre to open

- **How to model the recruitment period ?**
  - Analogy with queueing theory

Queueing theory		**Clinical research**
Storage capacity	$\longleftrightarrow$	target population or cohort
Server	$\longleftrightarrow$	None
Exit process	$\longleftrightarrow$	Drop-out patients
Entry process	$\longleftrightarrow$	Recruitment

- It is thus natural to model the recruitment period by means of **Poisson processes**.

- **How to model the recruitment period ?**
  - Analogy with queueing theory

Queueing theory		**Clinical research**
Storage capacity	$\longleftrightarrow$	target population or cohort
Server	$\longleftrightarrow$	None
Exit process	$\longleftrightarrow$	Drop-out patients
Entry process	$\longleftrightarrow$	Recruitment

- It is thus natural to model the recruitment period by means of **Poisson processes**.

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \mid \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

The recruitment process model

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i
  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process
  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration
  $\Longrightarrow$ **is the stopping time** $\inf \{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Consider a multicentric trial involving C investigator centres

- N: number of patients to be recruited
- T_R: expected duration of the trial
- $\mathcal{N}_i$: the recruitment process for centre i

  $\Longrightarrow$ **modelled by a PP of rate** λ_i
- $\mathcal{N}$: the global recruitment process

  $\Longrightarrow$ **modelled by a PP of rate** $\Lambda = \sum \lambda_i$
- $T(N)$: the recruitment duration

  $\Longrightarrow$ **is the stopping time** $\inf\{t \in \mathbb{R} \,|\, \mathcal{N}(t) \geq N\}$
- T_1 an interim time
- $\mathcal{F}_{T_1}$ denote the history of the process up to T_1

Theorem

- ***If λ is known (given by the investigator)*** *then*

  *The **feasibility of the trial** expresses by:*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right]$$

$$= 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt\right)^k e^{-\int_{T_1}^{T} (x_1+\ldots+x_C) dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \qquad (8)$$

  *The **expected duration** $\mathbb{E}[T_n]$ of the trial expresses by:*

$$\mathbb{E}\left[\inf_{t \in \mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \qquad (9)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (8) and (9)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  *The **feasibility of the trial** expresses by:*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right]$$

$$= 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt \right)^k e^{-\int_{T_1}^{T} (x_1 + \ldots + x_C) dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \qquad (8)$$

  *The **expected duration** $\mathbb{E}\left[T_n\right]$ of the trial expresses by:*

$$\mathbb{E}\left[\inf_{t \in \mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \qquad (9)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (8) and (9)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  The ***feasibility of the trial*** *expresses by:*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right]$$

$$= 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt \right)^k e^{-\int_{T_1}^{T} (x_1 + \ldots + x_C) dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \quad (8)$$

  The ***expected duration*** $\mathbb{E}[T_n]$ *of the trial expresses by:*

$$\mathbb{E}\left[\inf_{t \in \mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \quad (9)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (8) and (9)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  The ***feasibility of the trial*** *expresses by:*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right]$$

$$= 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt\right)^k e^{-\int_{T_1}^{T} (x_1+\ldots+x_C)dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \qquad (8)$$

  The ***expected duration*** $\mathbb{E}[T_n]$ *of the trial expresses by:*

$$\mathbb{E}\left[\inf_{t\in\mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \qquad (9)$$

  Involving $p_\lambda^{T_1}$ *the forward density of* λ.

- ***If λ is unknown*** *then*
  - $\hat{\lambda}$ *an estimation of* λ *from the data collected on* $[0, T_1]$
  - *Replace* λ *by* $\hat{\lambda}$ *in (8) and (9)*

Theorem

- ***If λ is known (given by the investigator)*** *then*

  *The **feasibility of the trial** expresses by:*

$$\mathbb{P}\left[\mathcal{N}(T_R) \geq N \mid \mathcal{F}_{T_1}\right]$$

$$= 1 - \sum_{k=0}^{N-N_1-1} \frac{1}{k!} \int_{\mathbb{R}^C} \left(\int_{T_1}^{T_R} (x_1 + \ldots + x_C) dt \right)^k e^{-\int_{T_1}^{T} (x_1+\ldots+x_C)dt} \prod_{i=1}^{C} p_\lambda^{T_1}(x_i)\, dx_i \qquad (8)$$

  *The **expected duration** $\mathbb{E}[T_n]$ of the trial expresses by:*

$$\mathbb{E}\left[\inf_{t\in\mathbb{R}} \{\mathcal{N}(t) \geq N\} \mid \mathcal{F}_{T_1}\right] = N \int_{\mathbb{R}^C} \frac{p_\lambda^{T_1}(x_1, \ldots, x_C)}{x_1 + \ldots + x_C} dx_1 \ldots dx_C \qquad (9)$$

  Involving $p_\lambda^{T_1}$ the forward density of λ.

- ***If λ is unknown*** *then*
  - *$\hat{\lambda}$ an estimation of λ from the data collected on $[0, T_1]$*
  - *Replace λ by $\hat{\lambda}$ in (8) and (9)*

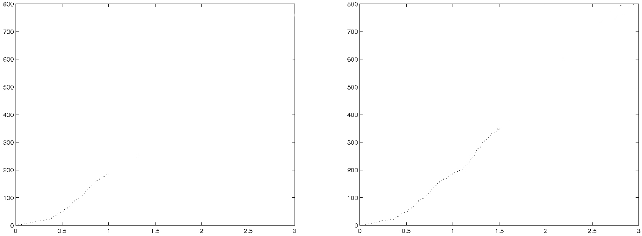


Figure: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots: Real data used to calibrate the model

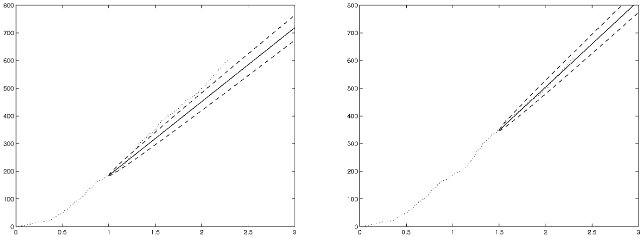


Figure: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots: Real data used to calibrate the model
- Solid line: estimated number of recruited patients
- Dotted line: Confidence Intervals

The recruitment process model

On-going studies

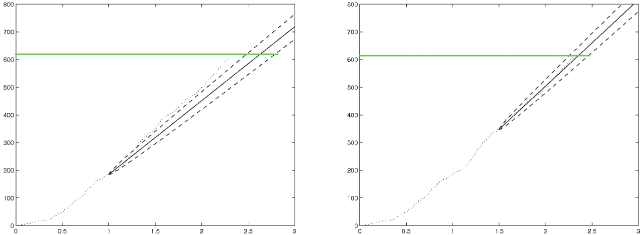
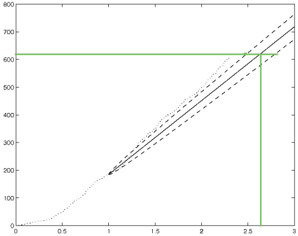


Figure: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots: Real data used to calibrate the model
- Solid line: estimated number of recruited patients
- Dotted line: Confidence Intervals



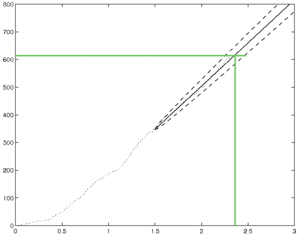


Figure: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots: Real data used to calibrate the model
- Solid line: estimated number of recruited patients
- Dotted line: Confidence Intervals

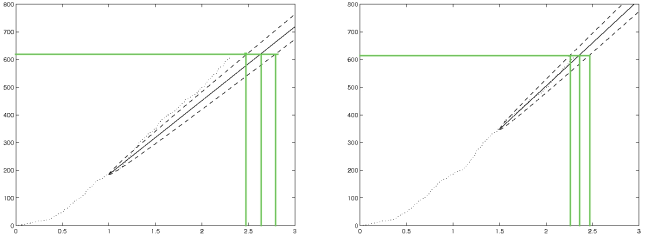


Figure: On going study at 1 year (on the left) and at 1.5 year (on the right)

- Dots: Real data used to calibrate the model
- Solid line: estimated number of recruited patients
- Dotted line: Confidence Intervals

Limit of this approach

Problem 1: If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

Limit of this approach

Problem 1: If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

Limit of this approach

Problem 1: If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

Limit of this approach

Problem 1: If p estimations are needed to describe $\mathcal{N}_i$, $C \cdot p$ estimation are needed to describe $\mathcal{N}$

When C large, this is not relevant

Problem 2 : If centre i has not recruited before T_1, then $\hat{\lambda}_i = 0$ and the model does not authorize centre i to recruit later

Empirical Bayesian model

Ones considers

$$(\lambda_1, \ldots, \lambda_C)$$

is a sample of size C distributed by a certain distribution $\mathcal{L}(\theta)$

Instead of estimate C values of λ, one estimates θ

- **Γ-Poisson model** (Anisimov, Fedorov (2007))
  - Rates are $\Gamma(\alpha, \beta)$ distributed.
  - Distribution of T is explicit.

- **Π-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Rates are Pareto-(x_m, k_p) distributed.
  - 20% of centres recruit 80% of patients.
  - Distribution of T is no more explicit (Monte Carlo Simulation).

- **$\mathcal{U}\Gamma$-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Centre opening date are unknown and uniformly distributed

- **Γ-Poisson model** (Anisimov, Fedorov (2007))
  - Rates are $\Gamma(\alpha, \beta)$ distributed.
  - Distribution of T is explicit.
- **Π-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Rates are Pareto-(x_m, k_p) distributed.
  - 20% of centres recruit 80% of patients.
  - Distribution of T is no more explicit (Monte Carlo Simulation).
- **$\mathcal{U}\Gamma$-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Centre opening date are unknown and uniformly distributed

- **Γ-Poisson model** (Anisimov, Fedorov (2007))
  - Rates are $\Gamma(\alpha, \beta)$ distributed.
  - Distribution of T is explicit.
- **Π-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Rates are Pareto-(x_m, k_p) distributed.
  - 20% of centres recruit 80% of patients.
  - Distribution of T is no more explicit (Monte Carlo Simulation).
- **$\mathcal{U}\Gamma$-Poisson model** (Mijoule, Savy and **Savy** (2012))
  - Centre opening date are unknown and uniformly distributed

- **Objectives:**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies:** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

- **Objectives:**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies:** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

- **Objectives:**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies:** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

- **Objectives:**
  - $N = 610$ patients
  - $T_R = 3$ years
  - $C_R = 77$ investigators centres
- **On-going studies:** after 1 year, after 1.5 year and after 2 years
- **The estimated duration** of the trial

The model	Time 1	Time 1.5	Time 2
Constant intensity	3.30	2.63	2.44
Γ-Poisson model	3.31	2.63	2.44
Π-Poisson model	2.63	2.39	2.36
$\mathcal{U}\Gamma$-Poisson model	2.60	2.34	2.36

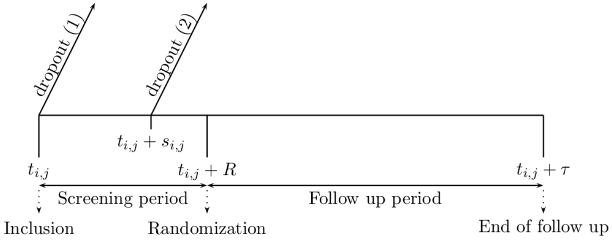
- **Effective duration** of the trial : **2.31 years**
  - The end of the trial was predicted with an error of **15 days, 10 mouths** before the expected date
  - **56 centres** would be enough for ending in 3 years.

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))

- **Drop-out at the inclusion**
  modelled by a probability p_i in centre i
  $(p_1, \ldots, p_C)$ sample having a beta distribution
- **Drop-out during the screening period**
  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i
  $(\theta_1, \ldots, \theta_C)$ sample having a gamma distribution

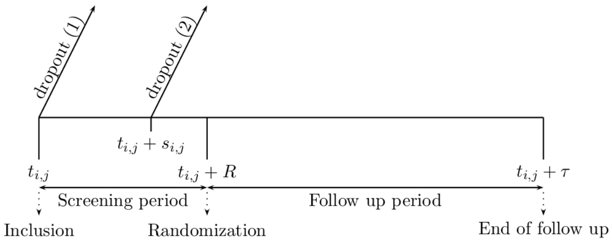
INSTITUT de MATHEMATIQUES de TOULOUSE

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))



- **Drop-out at the inclusion**
  modelled by a probability p_i in centre i
  $(p_1, \ldots, p_C)$ sample having a beta distribution
- **Drop-out during the screening period**
  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i
  $(\theta_1, \ldots, \theta_C)$ sample having a gamma distribution

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))



- **Drop-out at the inclusion**

  modelled by a probability p_i in centre i

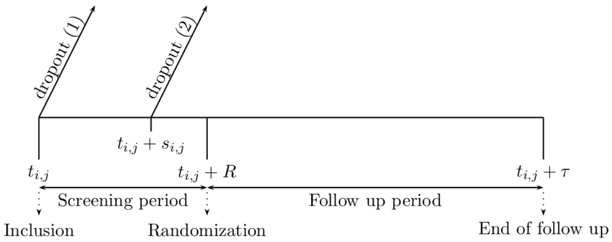
  $(p_1, \dots, p_C)$ sample having a beta distribution

- **Drop-out during the screening period**

  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i

  $(\theta_1, \dots, \theta_C)$ sample having a gamma distribution

Models investigated in (Anisimov, Mijoule, **Savy** (in progress))



- **Drop-out at the inclusion**

  modelled by a probability p_i in centre i

  $(p_1, \dots, p_C)$ sample having a beta distribution

- **Drop-out during the screening period**

  modelled $s_{i,j}$ modelled by an exponential distribution of intensity θ_i

  $(\theta_1, \dots, \theta_C)$ sample having a gamma distribution

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes:

$$\mathcal{L}_1(\alpha, \beta, p) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(p)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes:

$$\mathcal{L}_1(\alpha, \beta, \psi_1, \psi_2) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(\psi_1, \psi_2)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes:

$$\mathcal{L}_1(\alpha, \beta, \psi_1, \psi_2) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(\psi_1, \psi_2)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

- The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson.
- Drop-out process is directed by p a constant or $B(\psi_1, \psi_2)$.
- T_1 is an interim time.
  - τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
  - n_i number of **recruited patients** for centre i up to T_1
  - r_i number of **randomized patients** for centre i up to T_1

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i), 1 \leq i \leq C\}$, the log-likelihood function writes:

$$\mathcal{L}_1(\alpha, \beta, \psi_1, \psi_2) = \mathcal{L}_{1,1}(\alpha, \beta) + \mathcal{L}_{1,2}(\psi_1, \psi_2)$$

- *Notice the separation of the log-likelihood function (processes independent)*
- *$\mathcal{L}_{1,1}$ and $\mathcal{L}_{2,2}$ are explicit functions allowing optimisation.*

The recruitment dynamic is $\Gamma(\alpha, \beta)$-Poisson. T_1 is an interim time.

- τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
- n_i number of **recruited patients** for centre i up to T_1
- r_i number of **randomized patients** for centre i up to T_1
- ν_i number of **patients entered in screening period** for centre i in the interval $[T_1 - R, T_1]$

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i, \nu_i),\ 1 \leq i \leq C\}$, the predicted process of the number of randomized patients in centre i, $\{\widehat{\mathcal{R}}^i(t),\ t \geq T_1 + R\}$, expenses as

$$\widehat{\mathcal{R}}_i(t) = r_i + \mathrm{Bin}(\nu_i, \widehat{p}) + \Pi_{\widehat{p}\,\widehat{\lambda}_i}(t - T_1 - R).$$

$$\widehat{p} = \left(\sum_{i=1}^{C} n_i\right)^{-1} \sum_{i=1}^{C} r_i \quad \text{and} \quad \widehat{\lambda}_i = \mathrm{Ga}(\widehat{\alpha} + n_i, \widehat{\beta} + \tau_i)$$

The recruitment dynamic is $\Gamma(\alpha,\beta)$-Poisson. T_1 is an interim time.

- τ_i the **duration of activity** of centre i up to T_1 (assume $\tau_i \geq R$)
- n_i number of **recruited patients** for centre i up to T_1
- r_i number of **randomized patients** for centre i up to T_1
- ν_i number of **patients entered in screening period** for centre i in the interval $[T_1 - R, T_1]$

Theorem ((Anisimov, Mijoule, **Savy** (in progress)))

Given data $\{(n_i, r_i, \tau_i, \nu_i),\ 1 \leq i \leq C\}$, the predicted process of the number of randomized patients in centre i, $\{\widehat{\mathcal{R}}^i(t),\ t \geq T_1 + R\}$, expenses as

$$\widehat{\mathcal{R}}_i(t) = r_i + \mathrm{Bin}(\nu_i, \widehat{p}_i) + \Pi_{\widehat{p}_i\,\widehat{\lambda}_i}(t - T_1 - R).$$

$$\widehat{p}_i = \mathrm{Beta}(\widehat{\psi}_1 + k_i, \widehat{\psi}_2 + n_i - k_i), \quad \textit{and} \quad \widehat{\lambda}_i = \mathrm{Ga}(\widehat{\alpha} + n_i, \widehat{\beta} + \tau_i)$$

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i
- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i
  - $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$
- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i
  - $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$
  - $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$
    - g_i is a triangular function $g_i(t,s) = 0$ when $t \leq s$
    - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i
  - $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$
- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i
  - $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$
  - $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0\leq T_i^j\leq t} g_i(t, T_i^j)$
    - g_i is a triangular function $g_i(t,s)=0$ when $t\leq s$
    - T_i^j are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t,s)=0$ when $t\leq s$
  - T_j^i are randomization time of the patient j by centre i

Improvement of patient recruitment' models

An additive model for cost

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i \lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i \mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1 - p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i \mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0 \leq T_j^i \leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t, s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i
  - $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i
  - $\Rightarrow$ modelled by a **PP($(1 - p_i)\lambda_i$)**
  - $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$
  - $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0 \le T_j^i \le t} g_i(t, T_j^i)$
    - g_i is a triangular function $g_i(t,s) = 0$ when $t \le s$
    - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t,s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

Consider a clinical trial such that for centre i,

- The inclusion process $\mathcal{N}_i$ is modelled by a **PP(λ_i)**
- The probability for a patient to be screening failure is p_i

- $\mathcal{F}_i(t)$: the number of screening failure at time t for center i

  $\Rightarrow$ modelled by a **PP($p_i\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{F}_i(t)$: $J_i\mathcal{F}_i(t)$

- $\mathcal{R}_i(t)$ the number of randomized patients at time t for center i

  $\Rightarrow$ modelled by a **PP($(1-p_i)\lambda_i$)**

  $\Rightarrow$ cost proportional to $\mathcal{R}_i(t)$: $K_i\mathcal{R}_i(t)$

  $\Rightarrow$ cost depend of the duration of the follow-up: $\sum_{0\leq T_j^i\leq t} g_i(t, T_j^i)$

  - g_i is a triangular function $g_i(t, s) = 0$ when $t \leq s$
  - T_j^i are randomization time of the patient j by centre i

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i:

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \sum_{0 \le T_j^i \le t} g_i(t, T_j^i) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \ge 0} \{\mathcal{R}(t) \ge N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N), s)\, d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i:

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \int_0^t g_i(t,s) d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s) d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i:

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \int_0^t g_i(t,s) d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}[\mathcal{C}(T(N))]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N), s) d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i:

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \int_0^t g_i(t,s) d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s) d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i:

$$\mathcal{C}_i(t) = J_i\mathcal{F}_i(t) + K_i\mathcal{R}_i(t) + \int_0^t g_i(t,s)d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t\geq 0}\{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s)d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

In (Mijoule, Minois, Anisimov, **Savy** (forthcoming 2014)) ones considers the additive model for the cost generated by centre i:

$$\mathcal{C}_i(t) = J_i \mathcal{F}_i(t) + K_i \mathcal{R}_i(t) + \int_0^t g_i(t,s) d\mathcal{R}_i(s) + \underbrace{F_i + G_i\, t}_{\text{independent of patients}}$$

- The duration of the trial is the stopping time

$$T(N) = \inf_{t \geq 0} \{\mathcal{R}(t) \geq N\}$$

- The total cost of the trial is thus $\mathcal{C}(T(N)) = \sum_{i=1}^{C} \mathcal{C}_i(T(N))$
- In order to compute $\mathcal{C} = \mathbb{E}\left[\mathcal{C}(T(N))\right]$ we have to compute

$$\mathbb{E}\left[\int_0^{T(N)} g_i(T(N),s) d\mathcal{R}_i(s)\right].$$

It is not possible to use martingale arguments to compute such an expression

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume $(\lambda_i)_{1\leq i\leq C}$ and $(p_i)_{1\leq i\leq C}$* ***are known***

  $\Longrightarrow$ ***we have an explicit expression of $\mathcal{C}$***

- *Assume $\lambda_i \sim \Gamma(\alpha,\beta)$ and $p_i \sim B(\psi_1,\psi_2)$*
- *Consider an interim time T_1, and consider that the i-th centre has*
  - *screened n_i patients*
  - *randomized r_i patients*
- *Given (n_i, r_i) the posterior distribution of*
  - *the rate is $\lambda_i \sim \Gamma(\alpha+n_i, \beta+T_1)$*
  - *the probability of screening failure is $p_i \sim B(\psi_1+r_i, \psi_2+n_i-r_i)$*

  $\Longrightarrow$ ***we can compute $\mathcal{C}$ by means of Monte Carlo simulation***

INSTITUT de MATHEMATIQUES de TOULOUSE

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume $(\lambda_i)_{1\leq i\leq C}$ and $(p_i)_{1\leq i\leq C}$* ***are known***

  $\Longrightarrow$ ***we have an explicit expression of $\mathcal{C}$***

- *Assume $\lambda_i \sim \Gamma(\alpha,\beta)$ and $p_i \sim B(\psi_1,\psi_2)$*
- *Consider an interim time T_1, and consider that the i-th centre has*
  - *screened n_i patients*
  - *randomized r_i patients*
- *Given (n_i, r_i) the posterior distribution of*
  - *the rate is $\lambda_i \sim \Gamma(\alpha+n_i, \beta+T_1)$*
  - *the probability of screening failure is $p_i \sim B(\psi_1+r_i, \psi_2+n_i-r_i)$*

  $\Longrightarrow$ ***we can compute $\mathcal{C}$ by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

  $\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha,\beta)$ *and* $p_i \sim B(\psi_1,\psi_2)$
- *Consider an interim time* T_1*, and consider that the i-th centre has*
  - *screened* n_i *patients*
  - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
  - *the rate is* $\lambda_i \sim \Gamma(\alpha+n_i, \beta+T_1)$
  - *the probability of screening failure is* $p_i \sim B(\psi_1+r_i, \psi_2+n_i-r_i)$

$\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

  $\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha, \beta)$ *and* $p_i \sim B(\psi_1, \psi_2)$
- *Consider an interim time* T_1*, and consider that the i-th centre has*
  - *screened* n_i *patients*
  - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
  - *the rate is* $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$
  - *the probability of screening failure is* $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$

  $\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Habilitation thesis defence
Stochastic Calculus
Anticipative Integrals and Malliavin Calculus
LDP and SLDP for Ornstein-Uhlenbeck processes
Applied Statistics for Medical Research
Recruitment modelling
Survival data analysis
Conclusions and Perspectives

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

  $\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha,\beta)$ *and* $p_i \sim B(\psi_1,\psi_2)$
- *Consider an interim time* T_1, *and consider that the i-th centre has*
  - *screened* n_i *patients*
  - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
  - *the rate is* $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$
  - *the probability of screening failure is* $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$

  $\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Theorem ((Mijoule, Minois, Anisimov, **Savy** (2014)))

- *Assume* $(\lambda_i)_{1\leq i\leq C}$ *and* $(p_i)_{1\leq i\leq C}$ ***are known***

  $\Longrightarrow$ ***we have an explicit expression of*** $\mathcal{C}$

- *Assume* $\lambda_i \sim \Gamma(\alpha,\beta)$ *and* $p_i \sim B(\psi_1,\psi_2)$
- *Consider an interim time* T_1*, and consider that the* i*-th centre has*
  - *screened* n_i *patients*
  - *randomized* r_i *patients*
- *Given* (n_i, r_i) *the posterior distribution of*
  - *the rate is* $\lambda_i \sim \Gamma(\alpha + n_i, \beta + T_1)$
  - *the probability of screening failure is* $p_i \sim B(\psi_1 + r_i, \psi_2 + n_i - r_i)$

  $\Longrightarrow$ ***we can compute*** $\mathcal{C}$ ***by means of Monte Carlo simulation***

Assume the closure of centre j, denote

- $T^j(N)$ the duration of the trial without centre j
- $\mathcal{C}^j(t)$ the cost of the trial at time t without centre j

- By means of Monte Carlo simulation we are able to evaluate the variation of cost due to centre j closure:

$$\Delta\mathcal{C}_j = \mathbb{E}\left[\mathcal{C}(T(N)) - \mathcal{C}^j(T^j(N))\right]$$

- Consider $(\Delta\mathcal{C}_j, T^j(N))$ to decide on the closure of centre j.

Assume the closure of centre j, denote

- $T^j(N)$ the duration of the trial without centre j
- $\mathcal{C}^j(t)$ the cost of the trial at time t without centre j

- By means of Monte Carlo simulation we are able to evaluate the variation of cost due to centre j closure:

$$\Delta\mathcal{C}_j = \mathbb{E}\left[\mathcal{C}(T(N)) - \mathcal{C}^j(T^j(N))\right]$$

- Consider $(\Delta\mathcal{C}_j, T^j(N))$ to decide on the closure of centre j.

Assume the closure of centre j, denote

- $T^j(N)$ the duration of the trial without centre j
- $\mathcal{C}^j(t)$ the cost of the trial at time t without centre j

- By means of Monte Carlo simulation we are able to evaluate the variation of cost due to centre j closure:

$$\Delta\mathcal{C}_j = \mathbb{E}\left[\mathcal{C}(T(N)) - \mathcal{C}^j(T^j(N))\right]$$

- Consider $(\Delta\mathcal{C}_j, T^j(N))$ to decide on the closure of centre j.

- The GuidAge Trial
  - Double blind controlled randomized trial
  - 240 mg per day of de Ginkgo Biloba versus placebo
  - G.B. appears **to delay the conversion** to dementia of Alzheimer type
  - Primary endpoint: conversion to Alzheimer disease (time to event)

- **Statistical Analysis Plan:** Logrank test

  Conclusion: P-value 0.3044

  No Significant effect of the treatment

- **Re-analysis:** Fleming-Harrington's test ($q = 3$)

  Conclusion: P-value 0.0041

  Significant effect of the treatment

- **Is logrank test relevant for such a prevention study ?**

- D: the time to event random variable
  - F: the distribution function associated to D
  - $S = 1 - F$: the survival function associated to D
  - λ: the risk function associated to D
- Subjects may be right-censored by C independent of D.
  - $X_i = D_i \wedge C_i$: observed data
  - $\delta_i = \mathbb{I}_{\{D_i \leq C_i\}}$: censoring indicator
- $N_n(t)$: Number of events observed at time t:

$$N_n(t) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq t, \delta_i = 1\}}$$

- $Y_n(t)$: Number of at risk subjects at time t:

$$Y_n(t) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i \geq t\}}$$

We aim to test:

$$\mathcal{H}_0 : S^P(t) = S^T(t),\ \forall t$$
$$\mathcal{H}_1 : S^P(t) \neq S^T(t)$$

S^P and S^T are the survival functions associated respectively to the Placebo arm and Treatment arm.

We aim to test:

$$\mathcal{H}_0 : \lambda^P(t) = \lambda^T(t),\ \forall t$$
$$\mathcal{H}_1 : \lambda^P(t) \neq \lambda^T(t)$$

S^P and S^T are the survival functions associated respectively to the Placebo arm and Treatment arm.

We aim to test:

$$\mathcal{H}_0 : \lambda^P(t) = \lambda^T(t),\ \forall t$$
$$\mathcal{H}_1 : \lambda^P(t) \neq \lambda^T(t)$$

S^P and S^T are the survival functions associated respectively to the Placebo arm and Treatment arm.

Logrank test defined as:

$$\mathrm{LR}_{W_n}(t) = \int_0^t \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y^P_{n_P}(s) Y^T_{n_T}(s)}{Y^P_{n_P}(s) + Y^T_{n_T}(s)} \left[\frac{dN^P_{n_P}(s)}{Y^P_{n_P}(s)} - \frac{dN^T_{n_T}(s)}{Y^T_{n_T}(s)} \right]$$

We aim to test:

$$\mathcal{H}_0 : \lambda^P(t) = \lambda^T(t),\ \forall t$$
$$\mathcal{H}_1 : \lambda^P(t) \neq \lambda^T(t)$$

S^P and S^T are the survival functions associated respectively to the Placebo arm and Treatment arm.

Weighted Logrank test defined as:

$$\mathrm{LR}_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y^P_{n_P}(s) Y^T_{n_T}(s)}{Y^P_{n_P}(s) + Y^T_{n_T}(s)} \left[\frac{dN^P_{n_P}(s)}{Y^P_{n_P}(s)} - \frac{dN^T_{n_T}(s)}{Y^T_{n_T}(s)} \right]$$

W_n is an adapted, positive, predictable process

We aim to test:

$$\mathcal{H}_0 : \lambda^P(t) = \lambda^T(t),\ \forall t$$
$$\mathcal{H}_1 : \lambda^P(t) \neq \lambda^T(t)$$

S^P and S^T are the survival functions associated respectively to the Placebo arm and Treatment arm.

Weighted Logrank test defined as:

$$\mathrm{LR}_{W_n}(t) = \int_0^t W_n(s) \sqrt{\frac{n_P + n_T}{n_P n_T}} \frac{Y^P_{n_P}(s) Y^T_{n_T}(s)}{Y^P_{n_P}(s) + Y^T_{n_T}(s)} \left[\frac{dN^P_{n_P}(s)}{Y^P_{n_P}(s)} - \frac{dN^T_{n_T}(s)}{Y^T_{n_T}(s)} \right]$$

W_n is an adapted, positive, predictable process

The $\mathcal{H}_1$ assumption detected by the test depends on the weight

- **Constant piecewise weight**

$$W(t) = \begin{cases} 0 & \text{if} \quad t < t^* \\ 1 & \text{if} \quad t > t^* \end{cases}$$

Easy to interpret: t^* beginning of effect

- **Fleming Harrington weight** for late effect detection

$$W_n(t) = (1 - \widehat{S}_n(t))^q$$

where $\widehat{S}_n$ is the Kaplan-Meier estimator of S under the $\mathcal{H}_0$

Classical test but hard to interpret: what is the role of q ?

- Application of Fleming-Harrington's test in a Clinical trial setting
  - Wa have to choose a value for parameter q
  - We have to compute the necessary sample size

  (Garès, Andrieu, Dupuy, **Savy** (submitted to JRSS-C 2014))

- Comparison of CPWL test and Fleming Harrington's test
  - Comparison of performances
  - Bridge between the parameters of each test

  (Garès, Andrieu, Dupuy, **Savy** (Forthcoming EJS 2014))

- Introduction of a versatile test with "expert prior"
  - Maximum between logrank and Fleming Harrington tests
  - A computation procedure for sample size

  (Garès, Andrieu, Dupuy, **Savy** (in review for Stat. in Med. 2014))

Problems investigated

- Application of Fleming-Harrington's test in a Clinical trial setting
  - Wa have to choose a value for parameter q
  - We have to compute the necessary sample size

  (Garès, Andrieu, Dupuy, **Savy** (submitted to JRSS-C 2014))

- Comparison of CPWL test and Fleming Harrington's test
  - Comparison of performances
  - Bridge between the parameters of each test

  (Garès, Andrieu, Dupuy, **Savy** (Forthcoming EJS 2014))

- Introduction of a versatile test with "expert prior"
  - Maximum between logrank and Fleming Harrington tests
  - A computation procedure for sample size

  (Garès, Andrieu, Dupuy, **Savy** (in review for Stat. in Med. 2014))

- Application of Fleming-Harrington's test in a Clinical trial setting
  - Wa have to choose a value for parameter q
  - We have to compute the necessary sample size

    (Garès, Andrieu, Dupuy, **Savy** (submitted to JRSS-C 2014))
- Comparison of CPWL test and Fleming Harrington's test
  - Comparison of performances
  - Bridge between the parameters of each test

    (Garès, Andrieu, Dupuy, **Savy** (Forthcoming EJS 2014))
- Introduction of a versatile test with "expert prior"
  - Maximum between logrank and Fleming Harrington tests
  - A computation procedure for sample size

    (Garès, Andrieu, Dupuy, **Savy** (in review for Stat. in Med. 2014))

- **Objectives:**
  - To evaluate the performance of a test
  - To compare tests

- **Strategy:**
  - Make use of the **Asymptotic (Relative) Efficiency** (Van der Vaardt (1998))
  - There exists several notions of ARE
  - Asymptotic normality
    $\Longrightarrow$ Pitman's ARE more convenient

- **Consequences:**
  - Identification of the assumptions under which the test is **optimal**
  - Allow us to perform simulations studies

- **Objectives:**
  - To evaluate the performance of a test
  - To compare tests
- **Strategy:**
  - Make use of the **Asymptotic (Relative) Efficiency** (Van der Vaardt (1998))
  - There exists several notions of ARE
  - Asymptotic normality
    $\Longrightarrow$ Pitman's ARE more convenient
- **Consequences:**
  - Identification of the assumptions under which the test is **optimal**
  - Allow us to perform simulations studies

Strategy of these investigations

- **Objectives:**
  - To evaluate the performance of a test
  - To compare tests
- **Strategy:**
  - Make use of the **Asymptotic (Relative) Efficiency** (Van der Vaardt (1998))
  - There exists several notions of ARE
  - Asymptotic normality
    $\Longrightarrow$ Pitman's ARE more convenient
- **Consequences:**
  - Identification of the assumptions under which the test is **optimal**
  - Allow us to perform simulations studies

One tests the following assumptions

$$\begin{cases} \mathcal{H}_0 & : \lambda^T = \lambda^P = \lambda_{\theta_0}, \\ \mathcal{H}_1 & : \lambda^T = \lambda_{\theta^T} \quad \text{and} \quad \lambda^P = \lambda_{\theta^P} \end{cases}$$

Theorem

- ***Under*** $\mathcal{H}_0$ ***:***

$$\mathrm{LR}_{W_n} \xrightarrow[n \to +\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_0$$

  - $\mathbb{G}_0$ *centred Gaussian process of covariance function* $\Sigma_0(w, \lambda_{\theta_0})$

- ***Under*** $\mathcal{H}_1$ ***:***

$$\mathrm{LR}_{W_n} - \sqrt{n}\,\mu_{(\theta^P, \theta^T)} \xrightarrow[n \to +\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_1$$

  - $\mathbb{G}_1$ *centred Gaussian process of covariance function* $\Sigma_1(w, \lambda_{\theta^P}, \lambda_{\theta^T})$
  - $\mu_{(\theta^P, \theta^T)}$ *is function of* w, λ_{θ^P} *and* λ_{θ^T}

INSTITUT de MATHEMATIQUES de TOULOUSE

One tests the following assumptions

$$\begin{cases} \mathcal{H}_0 & : \lambda^T = \lambda^P = \lambda_{\theta_0}, \\ \mathcal{H}_1 & : \lambda^T = \lambda_{\theta^T} \quad \text{and} \quad \lambda^P = \lambda_{\theta^P} \end{cases}$$

Theorem

- ***Under*** $\mathcal{H}_0$ ***:***

$$\mathrm{LR}_{W_n} \xrightarrow[n \to +\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_0$$

  - $\mathbb{G}_0$ *centred Gaussian process of covariance function* $\Sigma_0(w, \lambda_{\theta_0})$

- ***Under*** $\mathcal{H}_1$ ***:***

$$\mathrm{LR}_{W_n} - \sqrt{n}\,\mu_{(\theta^P,\theta^T)} \xrightarrow[n \to +\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_1$$

  - $\mathbb{G}_1$ *centred Gaussian process of covariance function* $\Sigma_1(w, \lambda_{\theta^P}, \lambda_{\theta^T})$
  - $\mu_{(\theta^P,\theta^T)}$ *is function of* w, λ_{θ^P} *and* λ_{θ^T}

Asymptotic normality of WLT

One tests the following assumptions

$$\begin{cases} \mathcal{H}_0 & : \lambda^T = \lambda^P = \lambda_{\theta_0}, \\ \mathcal{H}_1 & : \lambda^T = \lambda_{\theta^T} \quad \text{and} \quad \lambda^P = \lambda_{\theta^P} \end{cases}$$

Theorem

- ***Under*** $\mathcal{H}_0$ ***:***

$$\mathrm{LR}_{W_n} \xrightarrow[n\to+\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_0$$

  - $\mathbb{G}_0$ *centred Gaussian process of covariance function* $\Sigma_0(w, \lambda_{\theta_0})$
- ***Under*** $\mathcal{H}_1$ ***:***

$$\mathrm{LR}_{W_n} - \sqrt{n}\,\mu_{(\theta^P,\theta^T)} \xrightarrow[n\to+\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_1$$

  - $\mathbb{G}_1$ *centred Gaussian process of covariance function* $\Sigma_1(w, \lambda_{\theta^P}, \lambda_{\theta^T})$
  - $\mu_{(\theta^P,\theta^T)}$ *is function of* w, λ_{θ^P} *and* λ_{θ^T}

One tests the following assumptions

$$\begin{cases} \mathcal{H}_0 & : \lambda^T = \lambda^P = \lambda_{\theta_0}, \\ \mathcal{H}_1 & : \lambda^T = \lambda_{\theta^T} \quad \text{and} \quad \lambda^P = \lambda_{\theta^P} \end{cases}$$

Theorem

- ***Under*** $\mathcal{H}_0$***:***

$$\mathrm{LR}_{W_n} \xrightarrow[n\to+\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_0$$

  - $\mathbb{G}_0$ *centred Gaussian process of covariance function* $\Sigma_0(w, \lambda_{\theta_0})$
- ***Under*** $\mathcal{H}_1$ ***:***

$$\mathrm{LR}_{W_n} - \sqrt{n}\,\mu_{(\theta^P,\theta^T)} \xrightarrow[n\to+\infty]{\mathcal{L}(\mathbb{D})} \mathbb{G}_1$$

  - $\mathbb{G}_1$ *centred Gaussian process of covariance function* $\Sigma_1(w, \lambda_{\theta^P}, \lambda_{\theta^T})$
  - $\mu_{(\theta^P,\theta^T)}$ *is function of* w, λ_{θ^P} *and* λ_{θ^T}

The idea is to consider the assumptions

$$\begin{cases} \mathcal{H}_0 & : \ F^T = F^P = F_{\theta_0}, \\ \mathcal{H}_1 & : \ F^T = F_{\theta^T_{n_T}} \quad \text{and} \quad F^P = F_{\theta^P_{n_P}} \end{cases}$$

in such a way that

- **The class of assumptions is sufficiently wide**

$$F_\theta(t) = \Psi(g(t) + \theta), \qquad \theta \in \Theta$$

- Ψ a distribution function with continuous second derivative
- g an increasing differentiable function

- **Choosing**

$$\theta_{n_P}^{P} = \theta_0 + c\sqrt{\frac{n_T}{n_P(n_P+n_T)}} \quad \text{and} \quad \theta_{n_T}^{T} = \theta_0 - c\sqrt{\frac{n_P}{n_T(n_P+n_T)}}$$

the singularity vanishes:

$$\sqrt{n}\,\mu_{(\theta_{n_P}^{P},\theta_{n_T}^{T})} \xrightarrow[n\to\infty]{a.s.} \mu_{\theta_0}$$

- **The efficiency of the test** can be measure by means of Pitman's **Asymptotic Efficiency**

$$AE = \frac{(\mu_{\theta_0}(\tau))^2}{\Sigma_0(\tau,\tau)}$$

- **Asymptotic Efficiency** depends on
  - the weight
  - the pattern of the assumptions

- **Choosing**

$$\theta^P_{n_P} = \theta_0 + c\sqrt{\frac{n_T}{n_P(n_P+n_T)}} \quad \text{and} \quad \theta^T_{n_T} = \theta_0 - c\sqrt{\frac{n_P}{n_T(n_P+n_T)}}$$

the singularity vanishes:

$$\sqrt{n}\,\mu_{(\theta^P_{n_P},\theta^T_{n_T})} \xrightarrow[n\to\infty]{a.s.} \mu_{\theta_0}$$

- **The efficiency of the test** can be measure by means of Pitman's **Asymptotic Efficiency**

$$AE = \frac{(\mu_{\theta_0}(\tau))^2}{\Sigma_0(\tau,\tau)}$$

- **Asymptotic Efficiency** depends on
  - the weight
  - the pattern of the assumptions

- **Choosing**

$$\theta^P_{n_P} = \theta_0 + c\sqrt{\frac{n_T}{n_P(n_P+n_T)}} \quad \textit{and} \quad \theta^T_{n_T} = \theta_0 - c\sqrt{\frac{n_P}{n_T(n_P+n_T)}}$$

the singularity vanishes:

$$\sqrt{n}\,\mu_{(\theta^P_{n_P},\theta^T_{n_T})} \xrightarrow[n\to\infty]{a.s.} \mu_{\theta_0}$$

- **The efficiency of the test** can be measure by means of Pitman's **Asymptotic Efficiency**

$$AE = \frac{(\mu_{\theta_0}(\tau))^2}{\Sigma_0(\tau,\tau)}$$

- **Asymptotic Efficiency** depends on
  - the weight
  - the pattern of the assumptions

- **Choosing**

$$\theta^P_{n_P} = \theta_0 + c\sqrt{\frac{n_T}{n_P(n_P+n_T)}} \quad \text{and} \quad \theta^T_{n_T} = \theta_0 - c\sqrt{\frac{n_P}{n_T(n_P+n_T)}}$$

the singularity vanishes:

$$\sqrt{n}\,\mu_{(\theta^P_{n_P},\theta^T_{n_T})} \xrightarrow[n\to\infty]{a.s.} \mu_{\theta_0}$$

- **The efficiency of the test** can be measure by means of Pitman's **Asymptotic Efficiency**

$$AE = \frac{(\mu_{\theta_0}(\tau))^2}{\Sigma_0(\tau,\tau)}$$

- **Asymptotic Efficiency** depends on
  - the weight
  - the pattern of the assumptions

- **Choosing**

$$\theta^P_{n_P} = \theta_0 + c\sqrt{\frac{n_T}{n_P(n_P+n_T)}} \quad \text{and} \quad \theta^T_{n_T} = \theta_0 - c\sqrt{\frac{n_P}{n_T(n_P+n_T)}}$$

the singularity vanishes:

$$\sqrt{n}\,\mu_{(\theta^P_{n_P},\theta^T_{n_T})} \xrightarrow[n\to\infty]{a.s.} \mu_{\theta_0}(w, \lambda_{\theta_0})$$

- **The efficiency of the test** can be measure by means of Pitman's **Asymptotic Efficiency**

$$AE = \frac{(\mu_{\theta_0}(\tau))^2}{\Sigma_0(\tau,\tau)}$$

- **Asymptotic Efficiency** depends on
  - the weight
  - the pattern of the assumptions

- **Choosing**

$$\theta^P_{n_P} = \theta_0 + c\sqrt{\frac{n_T}{n_P(n_P+n_T)}} \quad \text{and} \quad \theta^T_{n_T} = \theta_0 - c\sqrt{\frac{n_P}{n_T(n_P+n_T)}}$$

the singularity vanishes:

$$\sqrt{n}\,\mu_{(\theta^P_{n_P},\theta^T_{n_T})} \xrightarrow[n\to\infty]{a.s.} \mu_{\theta_0}(w, \lambda_{\theta_0})$$

- **The efficiency of the test** can be measure by means of Pitman's **Asymptotic Efficiency**

$$AE = \frac{(\mu_{\theta_0}(\tau))^2}{\Sigma_0(\tau,\tau)}$$

- **Asymptotic Efficiency** depends on
  - the weight
  - the pattern of the assumptions

Theorem ((Garès, Dupuy, Andrieu, **Savy** (JRSS-C 2014)))

Given a

- *a shift* Δ
- *a constant* $\lambda_0 > 0$
- *a parameter* $q > 0$

there exists a function $\Gamma^q(\cdot, \lambda_0, \Delta)$ *such that the Fleming-Harrington test* $\mathrm{FH}(q)$ *has* ***maximum Asymptotic Efficiency*** *to test*

$$\begin{cases} \mathcal{H}_0 & : \ \lambda^P = \lambda_0, \\ \mathcal{H}_1 & : \ \lambda^T = \lambda_0 \, \Gamma^q(\cdot, \lambda_0, \Delta) \end{cases} \tag{10}$$

Study the performance of $\mathrm{FH}(q)$ thanks to simulation study

Theorem ((Garès, Dupuy, Andrieu, **Savy** (JRSS-C 2014)))

Given a

- *a shift* Δ
- *a constant* $\lambda_0 > 0$
- *a parameter* $q > 0$

there exists a function $\Gamma^q(\cdot, \lambda_0, \Delta)$ *such that the Fleming-Harrington test* $\mathrm{FH}(q)$ *has* ***maximum Asymptotic Efficiency*** *to test*

$$\begin{cases} \mathcal{H}_0 & : \lambda^P = \lambda_0, \\ \mathcal{H}_1 & : \lambda^T = \lambda_0 \, \Gamma^q(\cdot, \lambda_0, \Delta) \end{cases} \tag{10}$$

Study the performance of $\mathrm{FH}(q)$ **thanks to simulation study**

- **Parameters (usually given by investigator)**
  - the sample size n
  - $c = S^P(\tau)$
  - $S^T(\tau)$ or discrepancy rate $r = \frac{S^T(\tau) - S^P(\tau)}{1 - S^P(\tau)}$
- **The data in the placebo group**
  - Simulated from an exponential distribution with parameter $\lambda_0 > 0$
  - λ_0 is given by the desired proportion of censored data:

$$\lambda_0 = -\frac{\ln(S^P(\tau))}{\tau}$$

- **The data in the treatment group**
  - Fix $q_S > 0$
  - Compute $\Delta(q_S)$
  - Simulate data from the hazard function

$$\lambda^T(t) = \lambda_0 \, \Gamma^q(\cdot, \lambda_0, \Delta(q_S))$$

- **Such a data set denoted $\mathcal{S}_1(q_S, n, r, c)$ is optimal for $\mathrm{FH}(q_S)$**

- **Parameters (usually given by investigator)**
  - the sample size n
  - $c = S^P(\tau)$
  - $S^T(\tau)$ or discrepancy rate $r = \frac{S^T(\tau) - S^P(\tau)}{1 - S^P(\tau)}$
- **The data in the placebo group**
  - Simulated from an exponential distribution with parameter $\lambda_0 > 0$
  - λ_0 is given by the desired proportion of censored data:

$$\lambda_0 = -\frac{\ln(S^P(\tau))}{\tau}$$

- **The data in the treatment group**
  - Fix $q_S > 0$
  - Compute $\Delta(q_S)$
  - Simulate data from the hazard function

$$\lambda^T(t) = \lambda_0 \, \Gamma^q(\cdot, \lambda_0, \Delta(q_S))$$

- **Such a data set denoted $\mathcal{S}_1(q_S, n, r, c)$ is optimal for** $\mathrm{FH}(q_S)$

- **Parameters (usually given by investigator)**
  - the sample size n
  - $c = S^P(\tau)$
  - $S^T(\tau)$ or discrepancy rate $r = \frac{S^T(\tau) - S^P(\tau)}{1 - S^P(\tau)}$
- **The data in the placebo group**
  - Simulated from an exponential distribution with parameter $\lambda_0 > 0$
  - λ_0 is given by the desired proportion of censored data:

$$\lambda_0 = -\frac{\ln(S^P(\tau))}{\tau}$$

- **The data in the treatment group**
  - Fix $q_S > 0$
  - Compute $\Delta(q_S)$
  - Simulate data from the hazard function

$$\lambda^T(t) = \lambda_0 \, \Gamma^q(\cdot, \lambda_0, \Delta(q_S))$$

- **Such a data set denoted $\mathcal{S}_1(q_S, n, r, c)$ is optimal for FH(q_S)**

- **Parameters (usually given by investigator)**
  - the sample size n
  - $c = S^P(\tau)$
  - $S^T(\tau)$ or discrepancy rate $r = \frac{S^T(\tau) - S^P(\tau)}{1 - S^P(\tau)}$
- **The data in the placebo group**
  - Simulated from an exponential distribution with parameter $\lambda_0 > 0$
  - λ_0 is given by the desired proportion of censored data:

$$\lambda_0 = -\frac{\ln(S^P(\tau))}{\tau}$$

- **The data in the treatment group**
  - Fix $q_S > 0$
  - Compute $\Delta(q_S)$
  - Simulate data from the hazard function

$$\lambda^T(t) = \lambda_0 \, \Gamma^q(\cdot, \lambda_0, \Delta(q_S))$$

- **Such a data set denoted $\mathcal{S}_1(q_S, n, r, c)$ is optimal for $\mathrm{FH}(q_S)$**

- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

q_S	Logrank	$q_T = 1$	$q_T = 2$	$q_T = 3$	$q_T = 4$
0	**0.640**	0.534	0.420	0.349	0.294
1	0.620	**0.743**	0.713	0.670	0.632
2	0.609	0.845	**0.877**	0.871	0.853
3	0.593	0.873	0.912	**0.914**	**0.914**
4	0.587	0.887	0.940	0.957	**0.961**
5	0.588	0.910	0.962	0.974	**0.980**

Table: Empirical power of FH(q_T) under scenarios $\mathcal{S}_1(q_S, 2000, 0.2, 0.8)$

Main result (Garès, Dupuy, Andrieu, **Savy** (JRSS-C 2014))

- No solution for choosing q
- Fleming Harrington's test exhibits little sensitivity to the value of q

- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

q_S	Logrank	$q_T = 1$	$q_T = 2$	$q_T = 3$	$q_T = 4$
0	**0.640**	0.534	0.420	0.349	0.294
1	0.620	**0.743**	0.713	0.670	0.632
2	0.609	0.845	**0.877**	0.871	0.853
3	0.593	0.873	0.912	**0.914**	**0.914**
4	0.587	0.887	0.940	0.957	**0.961**
5	0.588	0.910	0.962	0.974	**0.980**

Table: Empirical power of FH(q_T) under scenarios $\mathcal{S}_1(q_S, 2000, 0.2, 0.8)$

Main result (Garès, Dupuy, Andrieu, **Savy** (JRSS-C 2014))

- No solution for choosing q
- Fleming Harrington's test exhibits little sensitivity to the value of q

- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

q_S	Logrank	$q_T = 1$	$q_T = 2$	$q_T = 3$	$q_T = 4$
0	**0.640**	0.534	0.420	0.349	0.294
1	0.620	**0.743**	0.713	0.670	0.632
2	0.609	0.845	**0.877**	0.871	0.853
3	0.593	0.873	0.912	**0.914**	**0.914**
4	0.587	0.887	0.940	0.957	**0.961**
5	0.588	0.910	0.962	0.974	**0.980**

Table: Empirical power of FH(q_T) under scenarios $\mathcal{S}_1(q_S, 2000, 0.2, 0.8)$

Main result (Garès, Dupuy, Andrieu, **Savy** (JRSS-C 2014))

- No solution for choosing q
- Fleming Harrington's test exhibits little sensitivity to the value of q

Habilitation thesis defence
Stochastic Calculus
Anticipative Integrals and Malliavin Calculus
LDP and SLDP for Ornstein-Uhlenbeck processes
Applied Statistics for Medical Research
Recruitment modelling
Survival data analysis
Conclusions and Perspectives

- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

q_S	Logrank	$q_T = 1$	$q_T = 2$	$q_T = 3$	$q_T = 4$
0	**0.640**	0.534	0.420	0.349	0.294
1	0.620	**0.743**	0.713	0.670	0.632
2	0.609	0.845	**0.877**	0.871	0.853
3	0.593	0.873	0.912	**0.914**	**0.914**
4	0.587	0.887	0.940	0.957	**0.961**
5	0.588	0.910	0.962	0.974	**0.980**

Table: Empirical power of FH(q_T) under scenarios $\mathcal{S}_1(q_S, 2000, 0.2, 0.8)$

Main result (Garès, Dupuy, Andrieu, **Savy** (JRSS-C 2014))

- No solution for choosing q
- Fleming Harrington's test exhibits little sensitivity to the value of q

Comparison Fleming Harrington and CPW logrank tests

- CPWL depends on a parameter t^*
- This parameter has a concrete reality (beginning of the effect)

The strategy:

- Find the assumptions under which CPWL(t^*) is optimal

- Fix a value for t_S^*
- Generate 2000 data sets $\mathcal{S}_2(t_S^*, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

- Fix a value for q_S
- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of CPWL(t^*)
- Compute the empirical power of the test

- CPWL depends on a parameter t^*
- This parameter has a concrete reality (beginning of the effect)

The strategy:

- Find the assumptions under which CPWL(t^*) is optimal

- Fix a value for t^*_S
- Generate 2000 data sets $\mathcal{S}_2(t^*_S, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

- Fix a value for q_S
- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of CPWL(t^*)
- Compute the empirical power of the test

Comparison Fleming Harrington and CPW logrank tests

- CPWL depends on a parameter t^*
- This parameter has a concrete reality (beginning of the effect)

The strategy:

- Find the assumptions under which CPWL(t^*) is optimal

- Fix a value for t^*_S
- Generate 2000 data sets $\mathcal{S}_2(t^*_S, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

- Fix a value for q_S
- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of CPWL(t^*)
- Compute the empirical power of the test

Comparison Fleming Harrington and CPW logrank tests

- CPWL depends on a parameter t^*
- This parameter has a concrete reality (beginning of the effect)

The strategy:

- Find the assumptions under which CPWL(t^*) is optimal

- Fix a value for t_S^*
- Generate 2000 data sets $\mathcal{S}_2(t_S^*, n, r, c)$
- Analyse this data set by means of Fleming-Harrington q_T
- Compute the empirical power of the test

- Fix a value for q_S
- Generate 2000 data sets $\mathcal{S}_1(q_S, n, r, c)$
- Analyse this data set by means of CPWL(t^*)
- Compute the empirical power of the test

Main result (Garès, Dupuy, Andrieu, **Savy** (EJS 2014))

- CPWL(t^*) test exhibits little sensitivity to the value of t^*
- FH(q) is less sensitive to q than the CPWL(t^*) is to t^*
- Given t^*, it is possible to find the value of $q(t^*)$ which maximizes

$$\mathrm{ARE}(\mathrm{FH}(q(t^*)), \mathrm{CPWL}(t^*))$$

CPWL(t^*)	$t^* =$	0.2	0.4	0.6	0.8
FH($q(t^*)$)	$q(t^*) =$	0.5	1.2	2.4	5.9
FH(q)	$q =$	1	2	3	4
CPWL($t^*(q)$)	$t^*(q) =$	0.3	0.5	0.6	0.7

Table: Correspondence between q and t^*.

A versatile test

- Fleming-Harrington's test is a good test when late effects exist
- Logrank's test is a good test when effects are constant in time

  Investigators do not want to bet on a situation rather than another

- In (Garès, Dupuy, Andrieu, **Savy** (SIM 2014)), we introduce MWL statistics

$$\mathrm{MWL}^{\vec{q}}(t) = \max_{i=1,\dots,m}\left(\left|\frac{\mathrm{FH}^{q_i}(t)}{\widehat{\sigma}_{q_i}(t)}\right|\right)$$

- For $i = 1, \dots, m$, assume given p_i the probability that late effect of "type q_i" occurs (**expert a priori**)
- We investigate its performances for testing

$$\begin{cases} \mathcal{H}_0 & : F^T = F^P = F, \\ \mathcal{H}_1 & : \cup_{i=1}^{m} \{F^T = \Psi^{q_i}(g + \theta^T(i)) \text{ and } F^P = \Psi^{q_i}(g + \theta^P(i))\} \end{cases}$$

A versatile test

- Fleming-Harrington's test is a good test when late effects exist
- Logrank's test is a good test when effects are constant in time

  Investigators do not want to bet on a situation rather than another

- In (Garès, Dupuy, Andrieu, **Savy** (SIM 2014)), we introduce MWL statistics

$$\mathrm{MWL}^{\vec{q}}(t) = \max_{i=1,\ldots,m}\left(\left|\frac{\mathrm{FH}^{q_i}(t)}{\widehat{\sigma}_{q_i}(t)}\right|\right)$$

- For $i = 1, \ldots, m$, assume given p_i the probability that late effect of "type q_i" occurs (**expert a priori**)
- We investigate its performances for testing

$$\begin{cases} \mathcal{H}_0 & : \; F^T = F^P = F, \\ \mathcal{H}_1 & : \; \cup_{i=1}^m \{F^T = \Psi^{q_i}(g+\theta^T(i)) \text{ and } F^P = \Psi^{q_i}(g+\theta^P(i))\} \end{cases}$$

A versatile test

- Fleming-Harrington's test is a good test when late effects exist
- Logrank's test is a good test when effects are constant in time

  Investigators do not want to bet on a situation rather than another

- In (Garès, Dupuy, Andrieu, **Savy** (SIM 2014)), we introduce MWL statistics

$$\mathrm{MWL}^{\vec{q}}(t) = \max_{i=1,\ldots,m} \left(\left| \frac{\mathrm{FH}^{q_i}(t)}{\widehat{\sigma}_{q_i}(t)} \right| \right)$$

- For $i = 1, \ldots, m$, assume given p_i the probability that late effect of "type q_i" occurs (**expert a priori**)
- We investigate its performances for testing

$$\begin{cases} \mathcal{H}_0 & : \ F^T = F^P = F, \\ \mathcal{H}_1 & : \ \cup_{i=1}^{m} \{F^T = \Psi^{q_i}(g + \theta^T(i)) \text{ and } F^P = \Psi^{q_i}(g + \theta^P(i))\} \end{cases}$$

A versatile test

- Fleming-Harrington's test is a good test when late effects exist
- Logrank's test is a good test when effects are constant in time

  Investigators do not want to bet on a situation rather than another

- In (Garès, Dupuy, Andrieu, **Savy** (SIM 2014)), we introduce MWL statistics

$$\mathrm{MWL}^{\vec{q}}(t) = \max_{i=1,\dots,m}\left(\left|\frac{\mathrm{FH}^{q_i}(t)}{\widehat{\sigma}_{q_i}(t)}\right|\right)$$

- For $i = 1, \dots, m$, assume given p_i the probability that late effect of "type q_i" occurs (**expert a priori**)
- We investigate its performances for testing

$$\begin{cases} \mathcal{H}_0 & : F^T = F^P = F, \\ \mathcal{H}_1 & : \cup_{i=1}^{m} \{F^T = \Psi^{q_i}(g + \theta^T(i)) \text{ and } F^P = \Psi^{q_i}(g + \theta^P(i))\} \end{cases}$$

A versatile test

- Fleming-Harrington's test is a good test when late effects exist
- Logrank's test is a good test when effects are constant in time

  Investigators do not want to bet on a situation rather than another

- In (Garès, Dupuy, Andrieu, **Savy** (SIM 2014)), we introduce MWL statistics

$$\mathrm{MWL}^{\vec{q}}(t) = \max_{i=1,\dots,m}\left(\left|\frac{\mathrm{FH}^{q_i}(t)}{\widehat{\sigma}_{q_i}(t)}\right|\right)$$

- For $i = 1, \dots, m$, assume given p_i the probability that late effect of "type q_i" occurs (**expert a priori**)
- We investigate its performances for testing

$$\begin{cases} \mathcal{H}_0 & : \; F^T = F^P = F, \\ \mathcal{H}_1 & : \; \cup_{i=1}^m \{F^T = \Psi^{q_i}(g + \theta^T(i)) \text{ and } F^P = \Psi^{q_i}(g + \theta^P(i))\} \end{cases}$$

Main result (Garès, Dupuy, Andrieu, **Savy** (SIM 2014))

- Computation of asymptotic distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$
- Procedure for NSS computing
- **Good power** even when far from the optimal assumption

q_S	LR	FH^1	FH^2	FH^3	FH^4	FH^5
0	**0.629**	0.526	0.416	0.334	0.289	0.256
1	0.625	**0.756**	0.744	0.702	0.655	0.611
2	0.609	0.839	**0.864**	0.863	0.850	0.835
3	0.623	0.869	0.919	**0.925**	0.922	0.910
4	0.626	0.891	0.943	0.959	0.961	**0.963**
5	0.608	0.911	0.963	0.976	0.978	**0.982**
q_s		MWL^1	MWL^2	MWL^3	MWL^4	MWL^5
0		**0.620**	0.606	0.589	0.584	0.582
1		0.729	**0.731**	0.720	0.692	0.679
2		0.797	**0.828**	0.826	0.816	0.801
3		0.833	0.881	**0.897**	0.896	0.888
4		0.864	0.923	0.936	**0.946**	0.945
5		0.880	0.947	0.959	0.967	0.968

where $MWL^q = MWL^{(0,q)}$ with $p(q) = \frac{1}{2}$

An "omnibus" test

Main result (Garès, Dupuy, Andrieu, **Savy** (SIM 2014))

- Computation of asymptotic distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$
- Procedure for NSS computing
- **Good power** even when far from the optimal assumption

q_S	LR	FH^1	FH^2	FH^3	FH^4	FH^5
0	**0.629**	0.526	0.416	0.334	0.289	0.256
1	0.625	**0.756**	0.744	0.702	0.655	0.611
2	0.609	0.839	**0.864**	0.863	0.850	0.835
3	0.623	0.869	0.919	**0.925**	0.922	0.910
4	0.626	0.891	0.943	0.959	0.961	**0.963**
5	0.608	0.911	0.963	0.976	0.978	**0.982**

q_s		MWL^1	MWL^2	MWL^3	MWL^4	MWL^5
0		**0.620**	0.606	0.589	0.584	0.582
1		0.729	**0.731**	0.720	0.692	0.679
2		0.797	**0.828**	0.826	0.816	0.801
3		0.833	0.881	**0.897**	0.896	0.888
4		0.864	0.923	0.936	**0.946**	0.945
5		0.880	0.947	0.959	0.967	0.968

where $MWL^q = MWL^{(0,q)}$ with $p(q) = \frac{1}{2}$

Main result (Garès, Dupuy, Andrieu, **Savy** (SIM 2014))

- Computation of asymptotic distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$
- Procedure for NSS computing
- **Good power** even when far from the optimal assumption

q_S	LR	FH^1	FH^2	FH^3	FH^4	FH^5
0	**0.629**	0.526	0.416	0.334	0.289	0.256
1	0.625	**0.756**	0.744	0.702	0.655	0.611
2	0.609	0.839	**0.864**	0.863	0.850	0.835
3	0.623	0.869	0.919	**0.925**	0.922	0.910
4	0.626	0.891	0.943	0.959	0.961	**0.963**
5	0.608	0.911	0.963	0.976	0.978	**0.982**
q_s		MWL^1	MWL^2	MWL^3	MWL^4	MWL^5
0		**0.620**	0.606	0.589	0.584	0.582
1		0.729	**0.731**	0.720	0.692	0.679
2		0.797	**0.828**	0.826	0.816	0.801
3		0.833	0.881	**0.897**	0.896	0.888
4		0.864	0.923	0.936	**0.946**	0.945
5		0.880	0.947	0.959	0.967	0.968

where $\mathrm{MWL}^q = \mathrm{MWL}^{(0,q)}$ with $p(q) = \frac{1}{2}$

An "omnibus" test

Main result (Garès, Dupuy, Andrieu, **Savy** (SIM 2014))

- Computation of asymptotic distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$
- Procedure for NSS computing
- **Good power** even when far from the optimal assumption

q_S	LR	FH^1	FH^2	FH^3	FH^4	FH^5
0	**0.629**	0.526	0.416	0.334	0.289	0.256
1	0.625	**0.756**	0.744	0.702	0.655	0.611
2	0.609	0.839	**0.864**	0.863	0.850	0.835
3	0.623	0.869	0.919	**0.925**	0.922	0.910
4	0.626	0.891	0.943	0.959	0.961	**0.963**
5	0.608	0.911	0.963	0.976	0.978	**0.982**
q_s		MWL^1	MWL^2	MWL^3	MWL^4	MWL^5
0		**0.620**	0.606	0.589	0.584	0.582
1		0.729	**0.731**	0.720	0.692	0.679
2		0.797	**0.828**	0.826	0.816	0.801
3		0.833	0.881	**0.897**	0.896	0.888
4		0.864	0.923	0.936	**0.946**	0.945
5		0.880	0.947	0.959	0.967	0.968

where $\mathrm{MWL}^q = \mathrm{MWL}^{(0,q)}$ with $p(q) = \frac{1}{2}$

- Works in progress and perspectives are directed by
  - The funded projects and by projects tenders
    - **SMPR** - Principal Investigator - IRESP (plan cancer) 2013-2015
    - **IBISS** - Investigator workpackage "statistics" - ANR 2013 - 2016
    - **IMPACTISS** - Partner - IRESP 2014-2017 (submitted)
    - **SCT** - Project in maturation
  - The Ph-D students' works
    - **Nathan Minois (2013-2016)**
      Patients recruitment modelling
    - **Fabrice Billy (2013-2016)**
      Survival data analysis
  - The will to continue to share my time between projects on **Applied Statistics for Medical Research** and problems in **Stochastic Calculus**

- Works in progress and perspectives are directed by
  - The funded projects and by projects tenders
    - **SMPR** - Principal Investigator - IRESP (plan cancer) 2013-2015
    - **IBISS** - Investigator workpackage "statistics" - ANR 2013 - 2016
    - **IMPACTISS** - Partner - IRESP 2014-2017 (submitted)
    - **SCT** - Project in maturation
  - The Ph-D students' works
    - **Nathan Minois (2013-2016)**
      Patients recruitment modelling
    - **Fabrice Billy (2013-2016)**
      Survival data analysis
  - The will to continue to share my time between projects on **Applied Statistics for Medical Research** and problems in **Stochastic Calculus**

- Works in progress and perspectives are directed by
  - The funded projects and by projects tenders
    - **SMPR** - Principal Investigator - IRESP (plan cancer) 2013-2015
    - **IBISS** - Investigator workpackage "statistics" - ANR 2013 - 2016
    - **IMPACTISS** - Partner - IRESP 2014-2017 (submitted)
    - **SCT** - Project in maturation
  - The Ph-D students' works
    - **Nathan Minois (2013-2016)**
      Patients recruitment modelling
    - **Fabrice Billy (2013-2016)**
      Survival data analysis
  - The will to continue to share my time between projects on **Applied Statistics for Medical Research** and problems in **Stochastic Calculus**

- **Perspectives in Stochastic calculus**
  - What are the classes of stopping times which yields to class of processes ?
  - What is a filtered Brownian motion in local time scale ?
- **Perspectives in survival data analysis**
  - Extend the idea of expert prior
  - Explore the non-proportional hazard situation in competing risk setting
    - **Fabrice Billy Webe (Ph.D. student (2013-2016))**
    - Jean-Yves Dauxois (INSA - co-advisor)
    - John O'Quigley (LSTA - Paris VI)
    - Cyrille Delpierre (INSERM 1027 - Team 5)

- **Perspectives in Stochastic calculus**
  - What are the classes of stopping times which yields to class of processes ?
  - What is a filtered Brownian motion in local time scale ?
- **Perspectives in survival data analysis**
  - Extend the idea of expert prior
  - Explore the non-proportional hazard situation in competing risk setting
    - **Fabrice Billy Webe (Ph.D. student (2013-2016))**
    - Jean-Yves Dauxois (INSA - co-advisor)
    - John O'Quigley (LSTA - Paris VI)
    - Cyrille Delpierre (INSERM 1027 - Team 5)

- **Perspectives in Stochastic calculus**
  - What are the classes of stopping times which yields to class of processes ?
  - What is a filtered Brownian motion in local time scale ?

- **Perspectives in survival data analysis**
  - Extend the idea of expert prior
  - Explore the non-proportional hazard situation in competing risk setting
    - **Fabrice Billy Webe (Ph.D. student (2013-2016))**
    - Jean-Yves Dauxois (INSA - co-advisor)
    - John O'Quigley (LSTA - Paris VI)
    - Cyrille Delpierre (INSERM 1027 - Team 5)

- **Perspectives in Stochastic calculus**
  - What are the classes of stopping times which yields to class of processes ?
  - What is a filtered Brownian motion in local time scale ?

- **Perspectives in survival data analysis**
  - Extend the idea of expert prior
  - Explore the non-proportional hazard situation in competing risk setting
    - **Fabrice Billy Webe (Ph.D. student (2013-2016))**
    - Jean-Yves Dauxois (INSA - co-advisor)
    - John O'Quigley (LSTA - Paris VI)
    - Cyrille Delpierre (INSERM 1027 - Team 5)

- **Perspectives in Stochastic calculus**
  - What are the classes of stopping times which yields to class of processes ?
  - What is a filtered Brownian motion in local time scale ?
- **Perspectives in survival data analysis**
  - Extend the idea of expert prior
  - Explore the non-proportional hazard situation in competing risk setting
    - **Fabrice Billy Webe (Ph.D. student (2013-2016))**
    - Jean-Yves Dauxois (INSA - co-advisor)
    - John O'Quigley (LSTA - Paris VI)
    - Cyrille Delpierre (INSERM 1027 - Team 5)

- **Statistical Models for Patients Recruitment**
  Principal investigator (IRESP - 2013-2015)
  - Mixing Recruitment dynamic and longitudinal studies
  - Develop technique to include covariates in estimations procedure
    - Nathan Minois (Ph.D. student (2013-2016))
    - Sandrine Andrieu (INSERM 1027 - Team 1 - co-advisor)
    - Vladimir Anisimov (Quintiles)
    - Valérie Lauwers (INSERM 1027)
    - Guillaume Mijoule
    - Stéphanie Savy (INSERM 1027)

- **Biological Incorporation of Social Health Inequalities**
  Workpackage 4 (ANR - 2013-2015)
  - How to measure a mediated effects ?
  - How to manage missing data in databases ?
    - Thierry Lang (INSERM 1027 - Team 5)
    - Cyrille Delpierre (INSERM 1027 - Team 5)
    - Benoît Lepage (INSERM 1027 - Team 5)
    - Chloé Dimeglio (INSERM 1027 - Team 5)

- **Statistical Models for Patients Recruitment**
  Principal investigator (IRESP - 2013-2015)
  - Mixing Recruitment dynamic and longitudinal studies
  - Develop technique to include covariates in estimations procedure
    - **Nathan Minois (Ph.D. student (2013-2016))**
    - Sandrine Andrieu (INSERM 1027 - Team 1 - co-advisor)
    - Vladimir Anisimov (Quintiles)
    - Valérie Lauwers (INSERM 1027)
    - Guillaume Mijoule
    - Stéphanie Savy (INSERM 1027)

- **Biological Incorporation of Social Health Inequalities**
  Workpackage 4 (ANR - 2013-2015)
  - How to measure a mediated effects ?
  - How to manage missing data in databases ?
    - Thierry Lang (INSERM 1027 - Team 5)
    - Cyrille Delpierre (INSERM 1027 - Team 5)
    - Benoît Lepage (INSERM 1027 - Team 5)
    - Chloé Dimeglio (INSERM 1027 - Team 5)

- **Statistical Models for Patients Recruitment**
  Principal investigator (IRESP - 2013-2015)
  - Mixing Recruitment dynamic and longitudinal studies
  - Develop technique to include covariates in estimations procedure
    - **Nathan Minois (Ph.D. student (2013-2016))**
    - Sandrine Andrieu (INSERM 1027 - Team 1 - co-advisor)
    - Vladimir Anisimov (Quintiles)
    - Valérie Lauwers (INSERM 1027)
    - Guillaume Mijoule
    - Stéphanie Savy (INSERM 1027)

- **Biological Incorporation of Social Health Inequalities**
  Workpackage 4 (ANR - 2013-2015)
  - How to measure a mediated effects ?
  - How to manage missing data in databases ?
    - Thierry Lang (INSERM 1027 - Team 5)
    - Cyrille Delpierre (INSERM 1027 - Team 5)
    - Benoît Lepage (INSERM 1027 - Team 5)
    - Chloé Dimeglio (INSERM 1027 - Team 5)

- **Statistical Models for Patients Recruitment**
  Principal investigator (IRESP - 2013-2015)
  - Mixing Recruitment dynamic and longitudinal studies
  - Develop technique to include covariates in estimations procedure
    - **Nathan Minois (Ph.D. student (2013-2016))**
    - Sandrine Andrieu (INSERM 1027 - Team 1 - co-advisor)
    - Vladimir Anisimov (Quintiles)
    - Valérie Lauwers (INSERM 1027)
    - Guillaume Mijoule
    - Stéphanie Savy (INSERM 1027)

- **Biological Incorporation of Social Health Inequalities**
  Workpackage 4 (ANR - 2013-2015)
  - How to measure a mediated effects ?
  - How to manage missing data in databases ?
    - Thierry Lang (INSERM 1027 - Team 5)
    - Cyrille Delpierre (INSERM 1027 - Team 5)
    - Benoît Lepage (INSERM 1027 - Team 5)
    - Chloé Dimeglio (INSERM 1027 - Team 5)

- **Statistical Models for Patients Recruitment**
  Principal investigator (IRESP - 2013-2015)
  - Mixing Recruitment dynamic and longitudinal studies
  - Develop technique to include covariates in estimations procedure
    - **Nathan Minois (Ph.D. student (2013-2016))**
    - Sandrine Andrieu (INSERM 1027 - Team 1 - co-advisor)
    - Vladimir Anisimov (Quintiles)
    - Valérie Lauwers (INSERM 1027)
    - Guillaume Mijoule
    - Stéphanie Savy (INSERM 1027)

- **Biological Incorporation of Social Health Inequalities**
  Workpackage 4 (ANR - 2013-2015)
  - How to measure a mediated effects ?
  - How to manage missing data in databases ?
    - Thierry Lang (INSERM 1027 - Team 5)
    - Cyrille Delpierre (INSERM 1027 - Team 5)
    - Benoît Lepage (INSERM 1027 - Team 5)
    - Chloé Dimeglio (INSERM 1027 - Team 5)

- **Statistical Models for Patients Recruitment**
  Principal investigator (IRESP - 2013-2015)
  - Mixing Recruitment dynamic and longitudinal studies
  - Develop technique to include covariates in estimations procedure
    - **Nathan Minois (Ph.D. student (2013-2016))**
    - Sandrine Andrieu (INSERM 1027 - Team 1 - co-advisor)
    - Vladimir Anisimov (Quintiles)
    - Valérie Lauwers (INSERM 1027)
    - Guillaume Mijoule
    - Stéphanie Savy (INSERM 1027)
- **Biological Incorporation of Social Health Inequalities**
  Workpackage 4 (ANR - 2013-2015)
  - How to measure a mediated effects ?
  - How to manage missing data in databases ?
    - Thierry Lang (INSERM 1027 - Team 5)
    - Cyrille Delpierre (INSERM 1027 - Team 5)
    - Benoît Lepage (INSERM 1027 - Team 5)
    - Chloé Dimeglio (INSERM 1027 - Team 5)

- **Simulated Clinical Trials**
  - How to model the whole clinical trial ?
  - including dose/responds, drop-out, side-effect, recruitment...
  - Workshop "Clinical Trials Simulation"
    - Institut of Mathematics of Toulouse
    - Spring 2015
    - Everybody is welcome...

- **Simulated Clinical Trials**
  - How to model the whole clinical trial ?
  - including dose/responds, drop-out, side-effect, recruitment...
  - Workshop "Clinical Trials Simulation"
    - Institut of Mathematics of Toulouse
    - Spring 2015
    - Everybody is welcome...

- **Simulated Clinical Trials**
  - How to model the whole clinical trial ?
  - including dose/responds, drop-out, side-effect, recruitment...

  - Workshop "Clinical Trials Simulation"
    - Institut of Mathematics of Toulouse
    - Spring 2015
    - Everybody is welcome...

Thank you for your attention...