Large deviations for trajectories of Feller processes

Richard Kraaij

Delft University of Technology

March 14, 2014

Cramers theorem

Theorem

Let X be a random variable on $\mathbb{R}$ such that for every $\lambda \in \mathbb{R}$: $\mathbb{E}[e^{\lambda X}] < \infty$. Then we have for a sequence $X^1, X^2, \dots$ of independent copies of X that

$$\mathbb{P}\left[\frac{1}{n}\sum_{i \leq n} X^i \approx v\right] \approx e^{-nI(v)}$$

where

$$I(v) = \sup_{\lambda}\left\{v\lambda - \log \mathbb{E}[e^{\lambda X}]\right\}.$$

The $\approx$ symbols should be interpreted as a lower bound:

$$\liminf_n \frac{1}{n} \log \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X^i \in A\right] \geq - \inf_{v \in A^\circ} I(v)$$

and a corresponding upper bound.

Example: Schilders theorem

Theorem

For every i, let $t \mapsto W^i(t)$, be a standard Brownian motion on $\mathbb{R}$, $W^i(0) = 0$. Then, we have for a continuous path $\omega \in C([0,\infty))$:

$$\mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n} W^i(t)\right\}_{t\geq 0} \approx \omega\right] \approx e^{-nI(\omega)},$$

where

$$I(\omega) = \begin{cases} \frac{1}{2}\int_0^\infty (\omega'(t))^2 \mathrm{d}t & \text{if } \omega \text{ is absolutely continuous} \\ \infty & \text{otherwise.} \end{cases}$$

Tilted measure

Theorem (Sanov)

Suppose we have independent realisations Y^i from a distribution μ on some complete separable metric space E. Then:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{\{Y^i\}} \approx \nu\right] \approx e^{-nH(\nu\,|\,\mu)},$$

where

$$H(\nu\,|\,\mu) := \begin{cases} \int \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu & \textit{if } \nu << \mu \\ \infty & \textit{otherwise.} \end{cases}$$

Tilted measure

Theorem (Sanov)

Suppose we have independent realisations Y^i from a distribution μ on some complete separable metric space E. Then:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{\{Y^i\}} \approx \nu\right] \approx e^{-nH(\nu\,|\,\mu)},$$

where

$$H(\nu\,|\,\mu) := \begin{cases} \int \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu & \text{if } \nu << \mu \\ \infty & \text{otherwise.} \end{cases}$$

Alternatively,

$$H(\nu\,|\,\mu) = \sup_{f\in C(E)} \left\{ \int f\mathrm{d}\nu - \log \int e^f \mathrm{d}\mu \right\}$$

Dawson-Gärtner theorem

Theorem

For every i, let $t \mapsto W^i(t)$, be a standard Brownian motion on $\mathbb{R}$, $W^i(0) = 0$. Then, we have for a path $\mu \in D_{\mathcal{P}(E)}([0,\infty))$:

$$\mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\delta_{\{W^i(t)\}}\right\}_{t\geq 0} \approx \mu\right] \approx e^{-nI(\mu)},$$

where,

$$I(\mu) = \begin{cases} \frac{1}{2}\int_0^\infty \|\dot{\mu}(t) - A^*\mu(t)\|^2_{\mu(t)} \,\mathrm{d}t & \text{if } \omega \text{ is absolutely continuous} \\ \infty & \text{otherwise,} \end{cases}$$

$Af = \frac{1}{2}f''$. The definition of the norm, is given by $Q : C_c^\infty(\mathbb{R}) \to C_c^\infty(\mathbb{R})$, $Qf = (f')^2$ and

$$\|\alpha\|_\nu = \sup_{f\in C_c^\infty} \langle f, \alpha\rangle - \frac{1}{2}\langle Qf, \nu\rangle.$$

Connection to mass transport

Use this result to study

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{\{W^i(t)\}}\approx\mu(t)\right]\approx e^{-nI_t(\mu(t)|\delta_0)},$$

where

$$I_t(\mu(t)\,|\,\delta_0) :=$$
$$\inf\{I(\nu)\;:\;s\mapsto\nu(t)\text{ absolutely continuous},\nu(0)=\delta_0,\nu(t)=\mu(t)\}.$$

Large deviations for the trajectory of a general Markov process

For every fixed time $t \geq 0$, the law of large numbers gives

$$\frac{1}{n} \sum_{i \leq n} \delta_{\{X^i(t)\}} \to \mu(t),$$

and there is a corresponding large deviation result(Sanov's theorem, next slide).

Goal: A large deviation result for the trajectory

$$t \mapsto \frac{1}{n} \sum_{i \leq n} \delta_{\{X^i(t)\}}$$

in the Skorokhod space of càdlàg measure valued trajectories $D_{\mathcal{P}(E)}([0,\infty))$.

Notation for general Feller processes

Let $t \mapsto X(t)$ be a time-homogeneous Feller process on a compact metric space (E, d), i.e. the semigroup of conditional expectations

$$S_t f(x) := \mathbb{E}[f(X(t)) \,|\, X(0) = x]$$

maps $C(E)$ onto $C(E)$. Furthermore, the semigroup is strongly continuous, i.e. the trajectories $t \mapsto S_t f$ are continuous in $(C(E), \|\cdot\|)$ for every $f \in C(E)$.

Notation for general Feller processes

Let $t \mapsto X(t)$ be a time-homogeneous Feller process on a compact metric space (E, d), i.e. the semigroup of conditional expectations

$$S_t f(x) := \mathbb{E}[f(X(t)) \,|\, X(0) = x]$$

maps $C(E)$ onto $C(E)$. Furthermore, the semigroup is strongly continuous, i.e. the trajectories $t \mapsto S_t f$ are continuous in $(C(E), \|\cdot\|)$ for every $f \in C(E)$.

We denote with $(A, \mathcal{D}(A))$ the generator of $\{S_t\}_{t \geq 0}$,

$$Af := \lim_{t \downarrow 0} \frac{S_t f - f}{t} \tag{2.1}$$

defined for

$$f \in \mathcal{D}(A) := \{f \in C(E) \,:\, \text{the limit in (2.1) exists}\}$$

Connection to large deviations

Lemma

We have the two times large deviation principle:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i\leq n}\delta_{\{X^i(0)\}}\approx\mu(0)\;,\;\frac{1}{n}\sum_{i\leq n}\delta_{\{X^i(t)\}}\approx\mu(t)\right]$$
$$\approx\exp\left\{-n\left(H(\mu(0)\,|\,\mathbb{P}_0)+I_t(\mu(t)\,|\,\mu(0))\right)\right\}$$

where

$$I_t(\mu(t)\,|\,\mu(0)):=\sup_{f\in C(E)}\{\langle f,\mu(t)\rangle-\langle V(t)f,\mu(0)\rangle\}$$

and $V(t)f=\log S(t)e^f$ *and the inner product is defined as* $\langle h,\nu\rangle=\int h\mathrm{d}\nu$.

Connection to large deviations

Lemma

We have the two times large deviation principle:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i\leq n}\delta_{\{X^i(0)\}}\approx\mu(0)\ ,\ \frac{1}{n}\sum_{i\leq n}\delta_{\{X^i(t)\}}\approx\mu(t)\right]$$
$$\approx\exp\left\{-n\left(H(\mu(0)\,|\,\mathbb{P}_0)+I_t(\mu(t)\,|\,\mu(0))\right)\right\}$$

where

$$I_t(\mu(t)\,|\,\mu(0)):=\sup_{f\in C(E)}\left\{\langle f,\mu(t)\rangle-\langle V(t)f,\mu(0)\rangle\right\}$$

and $V(t)f=\log S(t)e^f$ *and the inner product is defined as* $\langle h,\nu\rangle=\int h\mathrm{d}\nu$.

Note: $V(t)$ is a semigroup and $Hf=e^{-f}Ae^f$ is its generator:

$$\frac{\partial}{\partial t}V(t)f\Big|\quad=\frac{AS(t)e^f}{S(t)e^f}\Big|\quad=e^{-f}Ae^f$$

Application of Sanov's theorem on Markov processes

The Markov process X corresponds to a measure $\mathbb{P}$ on $D_E([0,\infty))$. Pick some $g \in C(E)$ such that $e^g \in \mathcal{D}(A)$. Suppose that we define the function

$$G(X) := \exp\left\{ g(X(t)) - g(X(0)) - \int_0^t Hg(X(s))\mathrm{d}s \right\},$$

on the complete separable metric space $D_E([0,t])$, where $Hg = e^{-g}Ae^g$, then by Sanov's theorem we need to consider $\mathbb{Q}$, defined by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = G.$$

Theorem (Palmowski and Rolski, 2002)

The measure $\mathbb{Q}$, *corresponds to a Markov process generated by the generator* A^g, *defined by* $A^g f = e^{-g} A(fe^g) - (e^{-g} f)Ae^g$.

Theorem (Palmowski and Rolski, 2002)

The measure $\mathbb{Q}$, corresponds to a Markov process generated by the generator A^g, defined by $A^g f = e^{-g}A(fe^g) - (e^{-g}f)Ae^g$.

Example: Jump process

If X is a jump process generated by
$Af(x) = \sum_y r(x,y)\,[f(y) - f(x)]$, then
$A^g f(x) = \sum_y r(x,y)e^{g(y)-g(x)}\,[f(y) - f(x)]$ and
$Hf(x) = \sum_y r(x,y)\left[e^{f(y)-f(x)} - 1\right]$.

Theorem (Palmowski and Rolski, 2002)

The measure $\mathbb{Q}$, *corresponds to a Markov process generated by the generator* A^g, *defined by* $A^g f = e^{-g} A(fe^g) - (e^{-g} f) Ae^g$.

Example: Jump process

If X is a jump process generated by
$Af(x) = \sum_y r(x,y) [f(y) - f(x)]$, then
$A^g f(x) = \sum_y r(x,y) e^{g(y)-g(x)} [f(y) - f(x)]$ and
$Hf(x) = \sum_y r(x,y) \left[e^{f(y)-f(x)} - 1\right]$.

Example: Brownian motion

If X is a standard Brownian motion generated by $Af(x) = \frac{1}{2} f''(x)$, then $A^g f(x) = \frac{1}{2} f''(x) + f'(x) g'(x)$ and
$Hf(x) = Af(x) + \frac{1}{2}(f'(x))^2$.

Weakly continuous trajectory of measures

As $\mathcal{M}(E) = (C(E), \|\cdot\|)^*$, we can denote $\langle f, \mu \rangle = \int f \mathrm{d}\mu$ as the pairing between a space and its dual space.

Let $\mu(t)$ denote the law of $X(t)$. As

$\langle f, \mu(t) \rangle = \mathbb{E}[f(X(t))] = \mathbb{E}[S_t f(X(0))] = \langle S_t f, \mu(0) \rangle = \langle f, S_t^* \mu(0) \rangle,$

it follows that $\mu(t) = S_t^* \mu(0)$. Because $t \mapsto S_t f$ is continuous for every f, the trajectory

$$t \mapsto \mu(t) = S_t^* \mu(0)$$

is weakly continuous. In other words $t \mapsto \mu(t)$ is an element in $C_{\mathcal{P}(E)}([0, \infty))$

Problems with weak differentiability of paths

We would like to have a concept of deriving a path of measures.

$$\frac{\partial}{\partial t}\mu(t) := \lim_{h \downarrow 0} \frac{\mu(t+h) - \mu(t)}{h}.$$

Problems with weak differentiability of paths

We would like to have a concept of deriving a path of measures.

$$\frac{\partial}{\partial t}\mu(t) := \lim_{h \downarrow 0} \frac{\mu(t+h) - \mu(t)}{h}.$$

Note that

$$\frac{\mu(t+h) - \mu(t)}{h} = \frac{S_h^* \mu(t) - \mu(t)}{h}.$$

So for $f \in \mathcal{D}(A)$, we find

$$\lim_{h \downarrow 0} \frac{\langle f, \mu(t+h) \rangle - \langle f, \mu(t) \rangle}{t} = \langle Af, \mu(t) \rangle \quad (=? \ \langle f, A^* \mu(t) \rangle \ ).$$

A core for a generator

Definition

A linear subspace $D \subset \mathcal{D}(A)$ is called a core for A if it is dense in $\mathcal{D}(A)$ for the norm $\|f\|_A := \|f\| + \|Af\|$.

A core for a generator

Definition

A linear subspace $D \subset \mathcal{D}(A)$ is called a core for A if it is dense in $\mathcal{D}(A)$ for the norm $\|f\|_A := \|f\| + \|Af\|$.

Lemma

A linear subspace $D \subset \mathcal{D}(A)$ that is dense in $C(E)$ and that is invariant under the semigroup $S(t)$ is a core for A.

The Conditions

Condition

We have a core $D \subset C(E)$ for A that is equipped with some topology τ_D such that

1. *D is an algebra, and if $f \in D$, then $e^f \in D$.*
2. *(D, τ_D) is a separable barrelled locally convex Hausdorff space.*
3. *The topology τ_D is finer than the sup norm topology restricted to D.*
4. *For every $g \in D$, the generator $A^g : (D, \tau_D) \to (C(E), \|\cdot\|)$ is continuous.*
5. *There is a symmetric neighbourhood $\mathcal{N}$ of 0 in (D, τ_D) such that*
$$\sup_{f \in \mathcal{N}} \|Hf\| = \sup_{f \in \mathcal{N}} \left\| e^{-f} A e^{f} \right\| \leq 1,$$
and for every $c > 0$
$$\sup_{f \in c\mathcal{N}} \|Hf\| < \infty.$$

Absolutely continuous paths of measures 1

Definition (D^*-absolutely continuous paths)

A path $s \mapsto \mu(s)$ is D^*-absolutely continuous if there exists a measurable path $s \mapsto u(s)$ in D^*, such that for every $f \in D$ and $t \geq 0$

$$\int_0^t |\langle f, u(s)\rangle| \mathrm{d}s < \infty$$

and if for every $f \in D$, we have that

$$\lim_{h \to 0} \frac{\langle f, \mu(t+h)\rangle - \langle f, \mu(t)\rangle}{h} = \langle f, u(t)\rangle$$

for almost every time t. We denote $u(t) = \dot{\mu}(t)$.

Absolutely continuous paths of measures 2

We obtain that a path is in $\mathcal{AC}$ if we have for every f and $t \geq 0$ that

$$\begin{aligned}\langle f, \mu(t)\rangle - \langle f, \mu(0)\rangle &= \int_0^t \langle f, \dot{\mu}(s)\rangle \mathrm{d}s \\ &= \langle f, \int_0^t \dot{\mu}(s)\mathrm{d}s\rangle\end{aligned}$$

where we use the barrelledness of D to define the integral in the last line.

The Lagrangian

We define the 'Lagrangian' $\mathcal{L} : \mathcal{P}(E) \times D^* \to [0, \infty]$, by setting $\mathcal{L}(\mu, u) = \sup_{f \in D} \{\langle f, u\rangle - \langle Hf, \mu\rangle\}$.

Clearly, $\mathcal{L}$ is non-negative and lower semi-continuous with respect to the weak and weak* topologies.

Compactness of level sets

Lemma

The set $\{\mathcal{L} \le c\}$ *is weak* compact.*

Compactness of level sets

Lemma

The set $\{\mathcal{L} \leq c\}$ *is weak* compact.*

Theorem

For each $M > 0$ *and time* $T \geq 0$*,*

$$\left\{(\mu, u) \in C_{\mathcal{P}(E)}([0,t]) \times \mathcal{U} \ : \ \int_0^T \mathcal{L}(\mu(s), u(s))\mathrm{d}s \leq M \, \mu \in \mathcal{AC}, \dot{\mu} = u\right\}$$

is a compact subset of $C_{\mathcal{P}(E)}([0,t]) \times \mathcal{U}$*.*

$\mathcal{U}$ is the space of measurable maps $u : [0,t] \to D^*$. We say that a net u^α converges to u, where $u^\alpha, u \in \mathcal{U}$ if for every $t \geq 0$ and $f \in D$

$$\int_0^t |\langle f, u^\alpha(s)\rangle - \langle f, u(s)\rangle| \, \mathrm{d}s \to 0.$$

'proof' of the lemma

Condition (5): There is a symmetric neighbourhood of 0 in (D, τ_D): $\mathcal{N} \subset D$ such that

$$\sup_{f \in \mathcal{N}} \|Hf\| = \sup_{f \in \mathcal{N}} \left\| e^{-f} A e^{f} \right\| \leq 1,$$

Theorem (Bourbaki-Alaoglu)

Let $\mathcal{N}$ be a neighbourhood of 0 in D, then the set

$$\{u \ : \ |\langle f, u \rangle| \leq 1 \text{ for all } f \in \mathcal{N}\} \subset D^*$$

is weak compact.*

proof of compactness of level sets.

Let $f \in \mathcal{N}$ from condition (5) and let $u \in \{\mathcal{L} \leq c\}$.

$$\begin{aligned} |\langle f, u \rangle| &\leq \langle Hf, \mu \rangle \vee \langle H(-f), \mu \rangle + \mathcal{L}(\mu, u) \\ &\leq 1 + c \end{aligned}$$

Rewriting $Vf = \log S(t)e^f$

Lemma

Under the main condition, we can write:

$$\langle V(t)f, \mu(0)\rangle = \sup_{\substack{\nu \in \mathcal{AC} \\ \nu(0)=\mu(0)}} \left\{ \langle f, \nu(t)\rangle - \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) \mathrm{d}s \right\}$$

Main idea:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle V(t)f, \mu(t)\rangle|_{t=0} = \sup_{u \in D^*} \{\langle f, u\rangle - \mathcal{L}(\mu(0), u)\} = \langle Hf, \mu(0)\rangle$$

Rewriting the conditional rate function

As a result:

$$
\begin{aligned}
I_t(\mu(t) \mid \mu(0)) &:= \sup_{f \in C(E)} \{\langle f, \mu(t)\rangle - \langle V(t)f, \mu(0)\rangle\} \\
&= \sup_{f} \inf_{\substack{\nu \in \mathcal{AC} \\ \nu(0)=\mu(0)}} \langle f, \mu(t)\rangle - \langle f, \nu(t)\rangle + \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s \\
&= \inf_{\substack{\nu \in \mathcal{AC} \\ \nu(0)=\mu(0), \nu(t)=\mu(t)}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s
\end{aligned}
$$

Large deviation principle on the path space

Theorem

The trajectory

$$t \mapsto \frac{1}{n} \sum_{i \leq n} \delta_{\{X^i(t)\}}$$

satisfies the large deviation principle on $D_{\mathcal{P}(E)}([0,\infty))$ *with rate function*

$$I(\mu) = \begin{cases} H(\mu(0) \,|\, \mathbb{P}_0) + \int_0^\infty \mathcal{L}(\mu(s), \dot{\mu}(s)) \mathrm{d}s & \text{if } \mu \in \mathcal{AC} \\ \infty & \text{otherwise.} \end{cases}$$

Closing comments

1. The theorem also holds for locally compact spaces and generalises the Dawson-Gärtner theorem.
2. The extension to Polish spaces is work in progress.
3. It is hard to obtain an explicit formula for $\mathcal{L}$. However, the the pair $\mathcal{L}, H$ resembles the Lagrangian-Hamiltonian formalism in analytical mechanics, so it gives some intuition.
4. For example: even though this is not rigorous, the solutions to the Euler-Lagrange equation for $\mathcal{L}$ correspond to the Doob-transforms of the Markov process.