A limit conditional theorem for random walks, Applications to IS

Conditioned random walks, Trondheim, 2012

Michel Broniatowski, Virgile Caron

LSTA, Université Paris 6, France

June 27th, 2012

$X_1^n := (X_1, .., X_n)$, X_i's i.i.d. on $\mathbb{R}^d$ with density p.
$f : \mathbb{R}^d \longrightarrow \mathbb{R}^s$
local condition

$$\frac{1}{n}\sum f(X_i) = a_n \tag{Local}$$

$(a_n)_{n\geq 1}$ is a convergent sequence.

Either $a_n \longrightarrow EX$ (LLN range) or $a_n \longrightarrow a \neq EX$ (LD).
Question What about the distribution of $(X_1, .., X_{k_n})$ given Local when

$$\begin{aligned} k_n/n &\longrightarrow 1 \quad \text{long runs} \\ n - k_n &\longrightarrow \infty \text{ not all} \end{aligned}$$

Notation $p\left(X_1^k = x_1^k \middle| \frac{1}{n}\sum f(X_i) = a_n\right) =: p_{a_n}\left(X_1^k = x_1^k\right)$.

Take $f(x) = x, d = s = 1, k = 1.$

$$\frac{1}{n}\sum X_i = a$$

Gibbs conditioning principle (local form) with a_n fixed $=a$.

$$\begin{aligned} \phi(t) &: = E \exp tX \\ m(t) &= \frac{d}{dt}\log\phi(t), \ \ s^2(t) = \frac{d^2}{dt^2}\log\phi(t), \mu_3(t) := \frac{d^3}{dt^3}\log\phi(t) \end{aligned}$$

t such that $m(t) = a$

$$\pi^a(x) := \frac{\exp tx}{\phi(t)}p(x)$$

Recall $E_{\pi^a}(X) = a$, $Var_{\pi^a}(X) = s^2(t)$. Then (extensions (Diaconis and Friedman (1988)$k = o(n)$, Dembo and Zeitouni (1995) $\limsup_n k/n < 1$)

$$\begin{aligned} \int |p_a - \pi^a|\, dx &\to 0 \text{ as } n \to \infty \\ \sup_{A\in\mathcal{B}(\mathbb{R})} P_a(A) - \Pi^a(A) &\to 0. \end{aligned}$$

Relevance of this question for IS?

$$
\begin{aligned}
p_A\left(X_1^k = x_1^k\right) \quad &: \quad = p\left(X_1^k = x_1^k \middle| \frac{1}{n}\sum f(X_i) \in A\right) \\
&\sim \quad \int_A g_u\left(X_1^k = x_1^k\right) p\left(\frac{1}{n}\sum f(X_i) = u \middle| \frac{1}{n}\sum f(X_i) \in A\right) \\
&= \quad : g_A\left(X_1^k = x_1^k\right).
\end{aligned}
$$

a mixture of the approximating densities with weigths charging A. (no dominating point in this approach).

If g_u is "good"

$$
p_u\left(X_1^k = x_1^k\right) = g_u\left(X_1^k = x_1^k\right)(1 + o(1))
$$

then we expect

$$
p_A\left(X_1^k = x_1^k\right) = g_A\left(X_1^k = x_1^k\right)(1 + o(1))
$$

which with k_n "close to n" is OK when sampling under g_A is possible.

A recursive approximation (turn to $f(x) = x, d = s = 1$ for convenience)

$$p_{a_n}\left(X_1^k = x_1^k\right) = \prod_{i=0}^{k-1} p\left(X_{i+1} = x_{i+1} \middle| S_1^n = na_n, X_1^i = x_1^i\right)$$

Now for any α (invariance of conditional densities under any tilting)

$$p\left(X_{i+1} = x_{i+1} \middle| S_1^n = na_n, X_1^i = x_1^i\right) = \pi^\alpha\left(X_{i+1} = x_{i+1} \middle| S_1^n = na_n, X_1^i = x_1^i\right)$$

$$\begin{aligned}
&\pi^{m_i}\left(X_{i+1}=x_{i+1}\middle|\, S_1^n=na_n, X_1^i=x_1^i\right)\\
=\;&\pi^{m_i}\left(X_{i+1}=x_{i+1}\right)\frac{\pi^{m_i}\left(S_{i+1}^n=na_n-s_1^i\right)}{\pi^{m_i}\left(S_i^n=na_n-s_1^{i-1}\right)}
\end{aligned}$$

with $s_1^i := x_1 + .. + x_i$, with m_i making the ratio simple to evaluate. A precise evaluation of the dominating terms in this latest expression is needed in order to handle the product in the joint density. Center, reduce, Edgeworth expansions with

$$m_i := m(t_i) := \frac{1}{n-i+1}\left(na_n - s_1^{i-1}\right)$$

. **Pb: the orders of magnitude of the x_i's in order to control the k_n products $\Longrightarrow$ under the conditional sampling** Develop some maximal inequalities. Use Edgeworth expansions, etc.

Instead of arbitrary x_i's consider Y_i's random under the conditional distribution **(typical paths).**
Order of magnitude:

$$\max_{1 \le i \le n} |Y_i| = O_{P_{a_n}}(\log n)$$

(not large, but $\to \infty$)
Theorem:

$$\begin{aligned} p_{a_n}\left(X_1^{k_n} = Y_1^{k_n}\right) &= g_{a_n}\left(Y_1^{k_n}\right)\left(1 + o_{P_{a_n}}(\delta_n)\right) \\ g_{a_n}\left(Y_1^{k_n}\right) &= p_{a_n}\left(X_1^{k_n} = Y_1^{k_n}\right)\left(1 + o_{G_{a_n}}(\delta_n)\right) \end{aligned}$$

For IS: this is OK (sample under g_u)
Remark: Implies

$$\sup_{B \in \mathcal{B}(\mathbb{R}^{k_n})} P_{a_n}(B) - G_{a_n}(B) \to 0.$$

$$g_t(y_1^k) := \prod_{i=0}^{k-1} g_i(\,y_{i+1}|\,y_1^i).$$

$$g_{i+1}(\,y_{i+1}|\,y_1^i) = C_i p(y_{i+1})\mathfrak{n}\left(\alpha\beta + a_n, \alpha, y_{i+1}\right)$$

where $\mathfrak{n}\left(\mu, \tau, x\right)$ is the normal density with mean μ and variance τ at x.
Here

$$\alpha = s_{i,n}^2\left(n - i - 1\right), \; \beta = t_{i,n} + \frac{\mu_3^{(i,n)}}{2s_{i,n}^4\left(n - i - 1\right)}$$

The terms in α and β depend on the past values and on the m.g.f. of X.
Remark: coincides with the exact gaussian conditional density in the gaussian case, with k up to n. For fixed k : coincides with the usual tilted+rate. When k_n large, the quadratic term in the $g_{i+1}(\,y_{i+1}|\,y_1^i)g_{i+1}(\,y_{i+1}|\,y_1^i)$ is dominant (reduces the variance). When conditioning on an average of the $f(X_i)'s$ a change in the formula.

Consequence

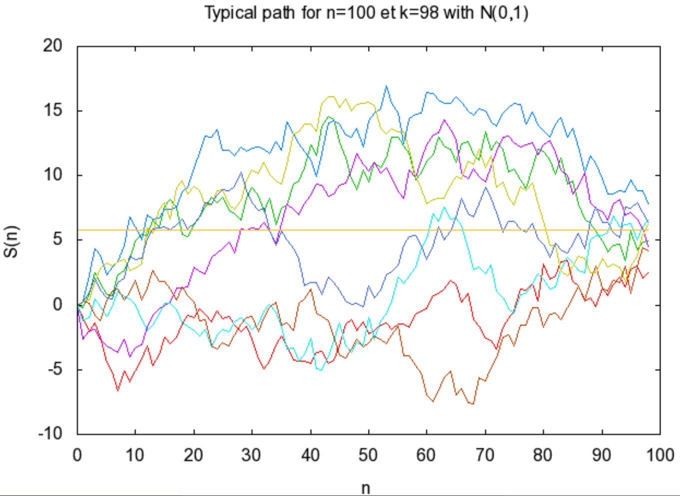
$$\begin{aligned} p_A\left(X_1^{k_n} = Y_1^{k_n}\right) &= g_A\left(Y_1^{k_n}\right)\left(1 + o_{P_A}\left(\delta_n\right)\right) \\ g_A\left(Y_1^{k_n}\right) &= p_A\left(X_1^{k_n} = Y_1^{k_n}\right)\left(1 + o_{Gg_A}\left(\delta_n\right)\right) \end{aligned}$$

for any "thick" $A \in \mathcal{B}\left(\mathbb{R}\right)$ and

$$\sup_{B \in \mathcal{B}(\mathbb{R}^{k_n})} P_A(B) - G_A(B) \to 0.$$

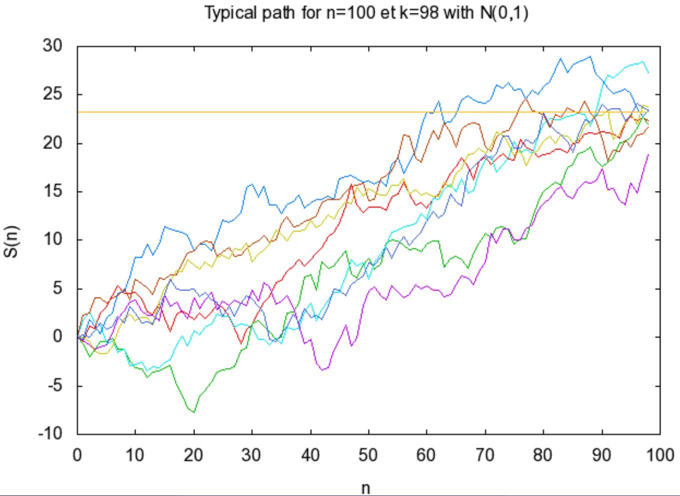
Trajectories

Normal Case, CLT



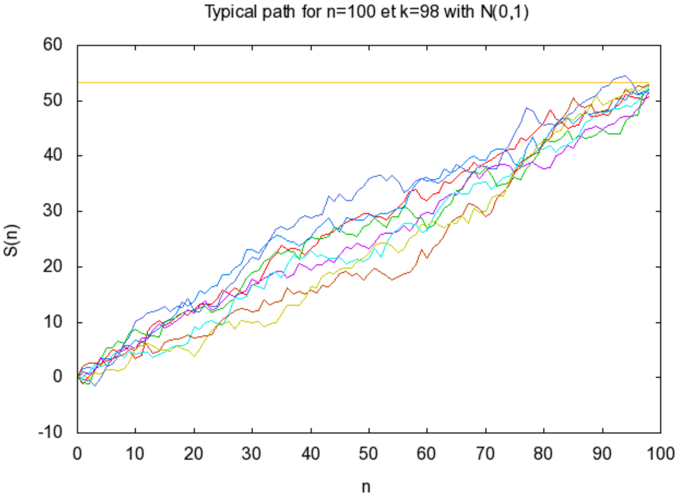
Trajectories

Normal Case, MD



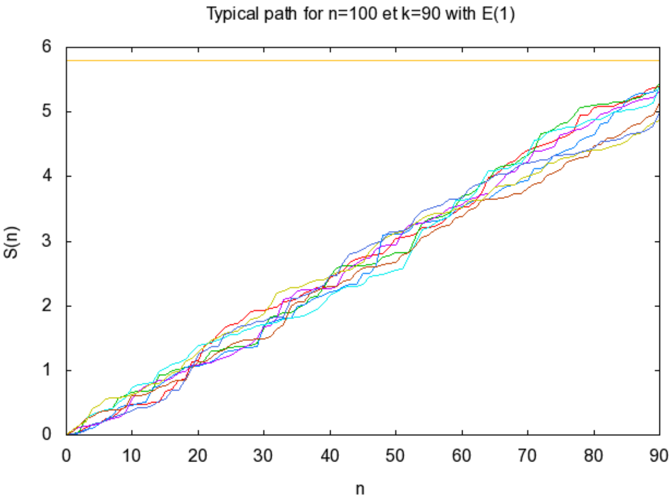
Trajectories

Normal Case, LDP

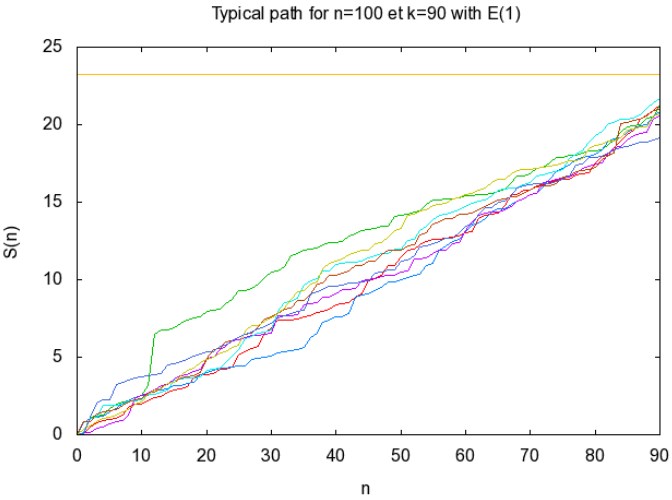


Trajectories

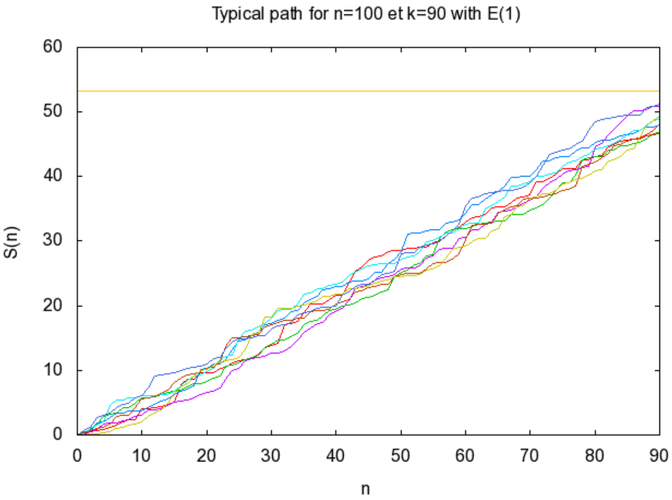
Exponential Case, CLT



Trajectories
Exponential Case, MD

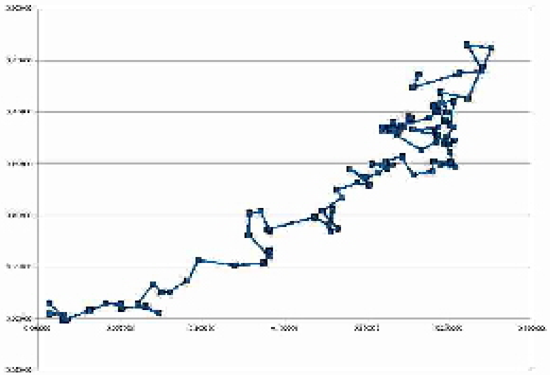


Trajectories
Exponential Case, LD



Trajectories

Normal Case, R2, MD



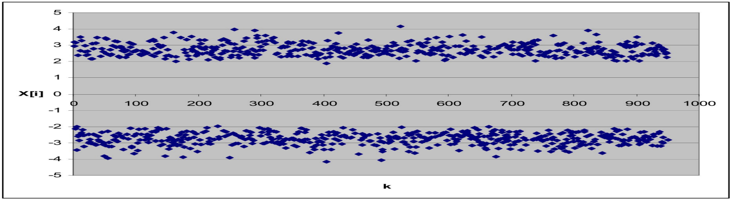
$f:\mathbb{R}^d \to \mathbb{R}$. $\mathbf{X}_1, ..., \mathbf{X}_n$ i.i.d. $f(\mathbf{X}_1)$ light tails. a_n large (high level, $a_n >> Ef(\mathbf{X}_1)$). **Draw** $x : f(x)$ **of order** a_n.

For

$$P\left(\text{all the } \mathbf{X}_i's \approx a_n \middle| \frac{1}{n}\sum \mathbf{X}_i > a_n\right) \to 1$$

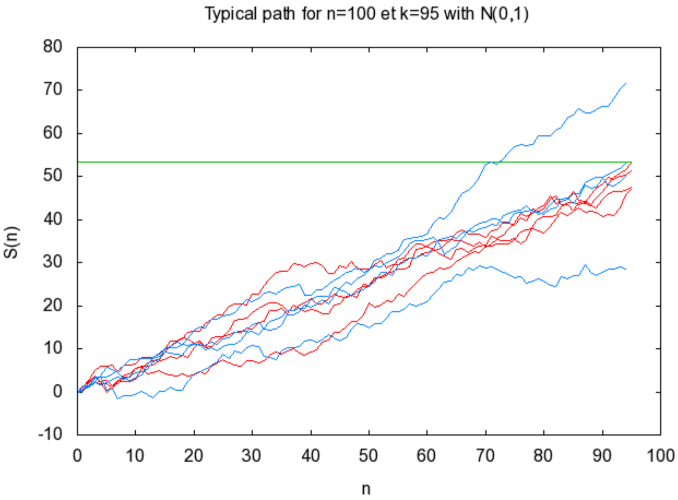
$f(x) = x^2$, $\mathcal{L}(\mathbf{X})$ =Symmetric Weibull shape parameter 2, $a_n = 10$, $n = 1000$

No more Gibbs equivalent in this asymptotics (B-Cao, 2012 Arxiv)



Trajectories

Blue:Tilted, Red:Adaptative Method

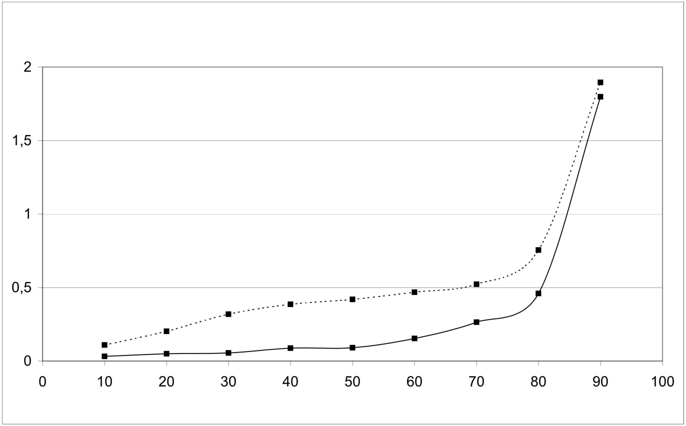


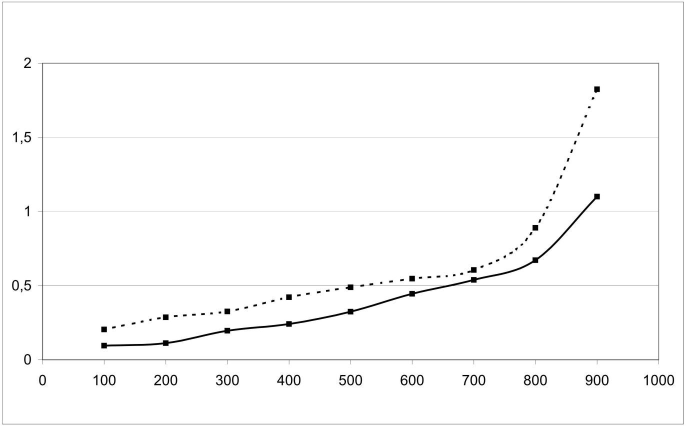
How long are good conditioned sampled runs?
An empirical benchmark: the relative error as a function of k_n

$$RE(k) := E_{G_u} \frac{\left| p_u(Y_1^k) - g_u(Y_1^k) \right|}{p_u(Y_1^k)}$$

to be estimated.

Remark: When $A = (a, \infty)$, the same indicators for k_n


2
1,5
1
0,5
0
0
10
20
30
40
50
60
70
80
90
100


2
1,5
1
0,5
0
0
100
200
300
400
500
600
700
800
900
1000

Evaluate $k = k_n$

- Simulate u in A following

$$L\left(\frac{1}{n}\sum f\left(X_i\right)\middle|\frac{1}{n}\sum f\left(X_i\right)\in A\right)$$

(Approximate the distribution (ex: LDP$\Longrightarrow$use Petrov, etc), or Metropolis-Hastings)

Evaluate $k = k_n$

- Simulate u in A following

$$L\left(\frac{1}{n}\sum f\left(X_i\right)\middle|\,\frac{1}{n}\sum f\left(X_i\right)\in A\right)$$

(Approximate the distribution (ex: LDP$\Longrightarrow$use Petrov, etc), or Metropolis-Hastings)

- Simulate Y_1^k with density $g_u \sim p\left(.\middle|\,\frac{1}{n}\sum f\left(X_i\right)=u\right).$

Evaluate $k = k_n$

- Simulate u in A following
$$L\left(\frac{1}{n}\sum f\left(X_i\right)\middle|\frac{1}{n}\sum f\left(X_i\right)\in A\right)$$
(Approximate the distribution (ex: LDP$\Longrightarrow$use Petrov, etc), or Metropolis-Hastings)
- Simulate Y_1^k with density $g_u \sim p\left(.\middle|\frac{1}{n}\sum f\left(X_i\right)=u\right)$.
- Simulate Y_{k+1}^n with tilted density at point m_k

Evaluate $k = k_n$

- Simulate u in A following

$$L\left(\frac{1}{n}\sum f(X_i) \middle| \frac{1}{n}\sum f(X_i) \in A\right)$$

(Approximate the distribution (ex: LDP$\Longrightarrow$use Petrov, etc), or Metropolis-Hastings)

- Simulate Y_1^k with density $g_u \sim p\left(.\middle| \frac{1}{n}\sum f(X_i) = u\right)$.
- Simulate Y_{k+1}^n with tilted density at point m_k
- Evaluate the IS ratio

$$\frac{\prod_{i=1}^{n} p(Y_i)}{g_u\left(Y_1^k\right)\prod_{i=k+1}^{n}\pi^{m_k}(Y_i)} 1_A\left(\frac{1}{n}\sum f(Y_i)\right)$$

Evaluate $k = k_n$

- Simulate u in A following
$$L\left(\frac{1}{n}\sum f\left(X_i\right)\middle|\frac{1}{n}\sum f\left(X_i\right)\in A\right)$$
(Approximate the distribution (ex: LDP$\Longrightarrow$use Petrov, etc), or Metropolis-Hastings)
- Simulate Y_1^k with density $g_u \sim p\left(.\middle|\frac{1}{n}\sum f\left(X_i\right)=u\right)$.
- Simulate Y_{k+1}^n with tilted density at point m_k
- Evaluate the IS ratio
$$\frac{\prod_{i=1}^{n} p\left(Y_i\right)}{g_u\left(Y_1^k\right)\prod_{i=k+1}^{n}\pi^{m_k}\left(Y_i\right)}1_A\left(\frac{1}{n}\sum f\left(Y_i\right)\right)$$
- Repeat from top L times

Evaluate $k = k_n$

- Simulate u in A following
$$L\left(\frac{1}{n}\sum f\left(X_i\right)\middle|\frac{1}{n}\sum f\left(X_i\right)\in A\right)$$
(Approximate the distribution (ex: LDP$\Longrightarrow$use Petrov, etc), or Metropolis-Hastings)
- Simulate Y_1^k with density $g_u \sim p\left(.\middle|\frac{1}{n}\sum f\left(X_i\right)=u\right)$.
- Simulate Y_{k+1}^n with tilted density at point m_k
- Evaluate the IS ratio
$$\frac{\displaystyle\prod_{i=1}^{n} p\left(Y_i\right)}{g_u\left(Y_1^k\right)\displaystyle\prod_{i=k+1}^{n}\pi^{m_k}\left(Y_i\right)} 1_A\left(\frac{1}{n}\sum f\left(Y_i\right)\right)$$
- Repeat from top L times
- Average $\Longrightarrow \widetilde{P_n}$

PropertiesStandard IS (i.i.d. replications under the tilted at dominating point). In LDP, for

$$\frac{1}{n}\sum f(X_i) > a$$

$$Var_{\pi^a}\widehat{P_n} \text{ proportional to } \sqrt{n} \text{ (e.g. Sadowsky and Bucklew)}$$
$$\widehat{Var_{g_A}}\widetilde{P_n} \text{ proportional to } \sqrt{n-k} \text{ (optimal on the } k_n \text{ first summands}$$

$\widetilde{P_n}$ has a small asymptotic variability when evaluated on classes of subsets of $\mathbb{R}^n$ whose probability goes to 1 under the sampling g_A .
Defaults: many steps, time run (obviously not worth for standard cases)
Accuracy (no dominating point)?
Compare with other methods in quasi-standard cases

$X_1, ..., X_{100}$ where X_1 has a normal distribution $N(0.05, 1)$ and let

$$\mathcal{E}_{100} := \left\{ x_1^{100} : \frac{|x_1 + ... + x_{100}|}{100} > 0.28 \right\}$$

$$P_{100} = P\left((X_1, ..., X_{100}) \in \mathcal{E}_{100}\right) = 0.01120.$$

simple **disymmetric case.** The standard i.i.d. IS scheme introduces the dominating point $a = 0.28$ and the family of i.i.d. tilted r.v's with common $N(a, 1)$ distribution. The resulting estimator of P_{100} is $0, 01074$ (with $L = 1000$), indicating that the event $S_{1,100}/100 < -0.28$ is ignored in the evaluation of P_{100}. Also the hit rate is of order 50%. It can also be seen that $S_1^{100}/100 < -0.28$ is never visited through the procedure.

C**omparison with the cross entropy method**. The sampling distribution is chosen as a normal one with variance 1, as adapted to this situation; the mean is estimated recursively through Kullback minimisation. When the initialisation mean is close to 0.28 then the performance is similar to the classical IS scheme, since the successive means keep close to 0.28; at the contrary when it is defined close to -0.28 the sequence of sampling distributions tend to concentrate around $N(-.28, 1)$ and the resulting estimate produces a relative error of order 100%. Indeed it is roughly $\left|\left(10^{-4} - 10^{-2}\right)/10^{-2}\right|$ since $P_{\mathcal{E}_{100}^-}$ ˜10^{-4} where $\mathcal{E}_{100}^- := \left\{x_1^{100} : \frac{x_1 + \ldots + x_{100}}{100} < -0.28\right\}$.

Rôle of point conditioning in stats:
sufficiency

$$p_\theta\left(\left.x_1^n\right|t\left(x_1^n\right)\right) \text{ independent upon } \theta$$

Rao-Blackwell: $S(X_1^n)$ estimator of θ. $t(x_1^n)$ any statistics

$$MSE\left(E\left(\left.S(X_1^n)\right|t(X_1^n)\right)\right) \leq MSE\left(S(X_1^n)\right)$$

Optimality when t is sufficient (Lehman-Scheffé) **PB: estimate** $E\left(\left.S(X_1^n)\right|T(X_1^n)\right)$.

Observe x_1^n, and $t\left(x_1^n\right)$; in exponential families $t\left(x_1^n\right) = \frac{1}{n}\sum t(x_i) = t_{obs}$ (converges under θ_0).

Choose k_n

Simulate according to $g_{t_{obs}}(.) \sim p_\theta\left(.\middle| \frac{1}{n}\sum t(X_i) = \frac{1}{n}\sum t(x_i)\right)$ (ind upon θ)

Estimate $E\left(S(X_1^n)\middle| t(X_1^n) = t\left(x_1^n\right)\right)$ averaging the values of S on the k_n realizations under $g_{t_{obs}}$

Remark t sufficient for $g_{t_{obs}}(.)$, so put any θ in the definition of $g_{t_{obs}}$