Large deviations for the weighted empirical measures of importance sampling

(Sanov's theorem for importance sampling)

Pierre Nyquist

Department of Mathematics
KTH Stockholm

Third Northern Triangular Seminar
April 11, 2011

Work in progress with Henrik Hult, KTH

Introduction

Problem setting

Let X be a random variable, distribution F, taking values in some space $\mathcal{X}$. Consider the task of computing $\Phi(F)$ for some functional Φ:

- Expectation: $\Phi_f(F) = \int f dF =: F(f)$, for some $f : \mathcal{X} \mapsto \mathcal{R}$,
- Quantile: $\Phi_q(F) = F^{-1}(q) = \inf\{x : F((x,\infty)) \leq q\}$, $q \in (0,1)$,
- L-statistic: $\Phi(F) = \int_0^1 \phi(q) F^{-1}(q) dq$.

When explicit computation is impossible, turn to simulation.

Introduction

Standard Monte Carlo

- Take an i.i.d. sample $X_1, ..., X_n$ from F and construct the empirical measure

$$\mathbf{F}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

  The Monte Carlo estimate of $\Phi(F)$ is $\Phi(\mathbf{F}_n)$ (plug-in estimate).
- Monte Carlo may require a large sample size, e.g., for rare events or extreme quantiles.
- Importance sampling a way to (possibly) reduce sample size.

Introduction

Importance sampling (IS)

- Take an i.i.d. sample $X_1, ..., X_n$ from the *sampling distribution* G, $F \ll G$.
- Weight function $w := \frac{dF}{dG}$.
- Construct the weighted empirical measure

$$\mathbf{G}_n^w = \frac{1}{n}\sum_{i=1}^{n} w(X_i)\delta_{X_i}.$$

- Yields the importance sampling estiamte $\Phi(\mathbf{G}_n^w)$.

Performance of simulation algorithms

Introduction

- Performance of IS determined by the choice of sampling distribution G.
- Evaluated in terms of $\mathrm{Var}(\Phi(\mathbf{G}_n^w))$ in the unbiased case.
- Biased case more complicated, e.g., empirical process theory [5].

Performance of simulation algorithms

Main idea

- Use large deviation results for the empirical measures to quantify the performance of importance sampling algorithms.
- Relate performance to the rate function associated with a large deviations principle.

Large deviations

Cramér's and Sanov's theorems

- Cramér's theorem: The empirical mean of i.i.d. $\mathcal{R}$-valued random variables satisfies the LDP with rate function

$$\Lambda^*(s) = \sup_{\theta \in \mathcal{R}} \{\theta s - \Lambda(\theta)\},$$

where $\Lambda(\theta) = \log \int \exp\{\theta x\} dF(x)$.

- Sanov's theorem: The empirical measure of i.i.d. random variables with common distribution F satisfies the LDP with rate function

$$\mathcal{H}(G \mid F) = \int \log \frac{dG}{dF} dG,$$

i.e., the relative entropy w.r.t. F.

Performance of Monte Carlo

Estimation of an expectation

- For $\Phi(\mathbf{F}_n) = \mathbf{F}_n(f)$, Cramér's theorem implies an upper bound

$$\limsup_n \frac{1}{n} \log \mathbf{P}(|\ \mathbf{F}_n(f) - F(f)\ | > \epsilon) \leq - \inf_{x \in B(F(f),\epsilon)^c} \Lambda^*(x).$$

- Suggests, for n sufficiently large,

$$\mathbf{P}(\mathbf{F}_n(f) \in B(F(f),\epsilon)^c) \approx \exp\{-n \inf_{x \in B(F(f),\epsilon)^c} \Lambda^*(x)\}.$$

- Sample size needed for an upper bound δ on the probability:

$$n \approx \frac{1}{\inf_{x \in B(F(f),\epsilon)^c} \Lambda^*(x)} (-\log \delta).$$

Performance of Monte Carlo

Estimation for a general functional

- For general functionals Φ, want to consider the probability of $\mathbf{F}_n$ being close to F.
- Sanov's theorem provides an LDP for the empirical measures $\mathbf{F}_n$ of Monte Carlo.
- Let, e.g., $A_\epsilon = \{G \in \mathcal{M}_1 : \mid \Phi(G) - \Phi(F) \mid > \epsilon\}$. By Sanov's,

$$\limsup_n \frac{1}{n} \log \mathbf{P}(\mathbf{F}_n \in A_\epsilon) \leq - \inf_{G \in \bar{A}_\epsilon} \mathcal{H}(G \mid F).$$

Performance of IS

Application of large deviation results

Suppose the weighted empirical measures of IS satisfy an LDP.

- Let $A_\epsilon \subset \mathcal{M}$ be some set that relates to the accuracy of the estimate $\Phi(\mathbf{G}_n^{wf})$.
- The LDP implies, for sufficiently large n,

$$\mathbf{P}(\mathbf{G}_n^{wf} \in A_\epsilon) \approx \exp\{-n \inf_{\nu \in \bar{A}_\epsilon} I(\nu)\}.$$

- With δ the desired bound for the probability, we obtain

$$n \approx \frac{1}{\inf_{\nu \in \bar{A}_\epsilon} I(\nu)}(-\log \delta).$$

Performance of IS

- For $\mathbf{G}_n^w(f)$ Cramér's theorem is applicable, yielding an asymptotic upper bound on the error probability.
- Sanov's theorem not applicable for the weighted empirical measures $\mathbf{G}_n^w$.
- Need an LDP for $\mathbf{G}_n^w$ in order to quantify the notion of the weighted empirical measures being close to F.

LDP for the weighted empirical measures of IS

- Suffices to have the weighted empirical measures $\mathbf{G}_n^w$ close fo F in the region that largely determines $\Phi(F)$.
- Let f be an F-integrable function characterizing the importance of different regions of the space $\mathcal{X}$. Want

$$\mathbf{G}_n^{wf} = \frac{1}{n}\sum_{i=1}^{n} w(X_i)f(X_i)\delta_{X_i},$$

to be close to F^f, where F^f is defined as

$$F^f(g) = \int g(x)f(x)dF(x),$$

for each bounded, measurable function g.

LDP for the weighted empirical measures of IS

Laplace principle

Theorem Let F, G and f be given as above, with $F \ll G$ on the support of f. Suppose that $\int e^{wf} dG < \infty$. Then, for any bounded, continuous $h : \mathcal{M} \mapsto \mathcal{R}$,

$$\lim_n \frac{1}{n} \log \mathbb{E}[e^{-nh(\mathbf{G}_n^{wf})}] = - \inf_{\nu \in \mathcal{M}} \{h(\nu) + I(\nu)\}.$$

- $\Gamma = \{Q \in \mathcal{M}_1 : \mathcal{H}(Q \mid G) < \infty, Q(wf) < \infty\}$.
- $\Psi : \Gamma \mapsto \mathcal{M}$ the mapping, for each bounded, measurable g,

$$\Psi(G; g) = \int g(x) w(x) f(x) dG(x).$$

- $I(\nu) = \inf\{\mathcal{H}(Q \mid G) : \Psi(Q) = \nu, Q \in \Gamma\}$.

Application

Comparison of Monte Carlo and importance sampling

Possible ways to use the derived result to quantify performance of simulation algorithms:

- Comparison of Monte Carlo and importance sampling in terms of the rate of decay of the error probability.
- Compare the sample size n, expressed in the true quantity $\Phi(F)$, needed for Monte Carlo and importance sampling respectively to reach the desired accuracy.
- Larger rate, i.e., $\inf_{x \in A_\epsilon} I(x)$, suggests improved performance.

LDP for the weighted empirical measures

Idea of proof

- Relies on the weak convergence approach to large deviations[1].
- Identify $W_n = -\frac{1}{n}\log \mathbb{E}[\exp\{(-nh(\mathbf{G}_n^{wf})\}]$ with the total cost of a stochastic control problem and derive a representation formula.
- The Laplace principle upper bound

$$\lim_n \frac{1}{n}\log \mathbb{E}[e^{-nh(\mathbf{G}_n^{wf})}] \leq -\inf_{\nu\in\mathcal{M}}\{h(\nu)+I(\nu)\},$$

  requires the most work compared to the case of ordinary empirical measures (Sanov's theorem).

[1] Dupuis and Ellis (1997)

Summary

- Proposed a way to use the rate function of large deviation results to quantify the performance of importance sampling algorithms.
- Derived a Laplace principle for the weighted empirical measures of IS.

References I

[1] ASMUSSEN, S., AND GLYNN, P. W.
Stochastic simulation: Algorithms and analysis.
Springer, 2007.

[2] BUCKLEW, J. A.
Introduction to rare event simulation.
Springer, 2004.

[3] DEMBO, A., AND ZEITOUNI, O.
Large deviations techniques and applications, 2nd ed.
Springer, 1998.

[4] DUPUIS, P., AND ELLIS, R. S.
A weak convergence approach to the theory of large deviations, second ed.
John Wiley & Sons, Inc., 1997.

[5] HULT, H., AND SVENSSON, J.
Efficient calculation of risk measures by importance sampling.
Preprint (2009).