Large Deviations and Weak Convergence of Measures

with applications to Monte Carlo Estimators

Johan Ericsson

Master's Thesis Presentation
Stockholm University

June 13, 2024

Table of Contents

Motivation

- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and real valued random variable $X : \Omega \to \mathbb{R}$, with distribution $\mu := \mathbb{P} \circ X^{-1}$.

Motivation

- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and real valued random variable $X : \Omega \to \mathbb{R}$, with distribution $\mu := \mathbb{P} \circ X^{-1}$.
- In many applications we are interested in computing the expected value $\mathbb{E}[X]$

$$\theta = \mathbb{E}[X] = \int_{\Omega} X \, \mathrm{d}\mathbb{P} \, .$$

Motivation

- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and real valued random variable $X : \Omega \to \mathbb{R}$, with distribution $\mu := \mathbb{P} \circ X^{-1}$.
- In many applications we are interested in computing the expected value $\mathbb{E}[X]$

$$\theta = \mathbb{E}[X] = \int_{\Omega} X \,\mathrm{d}\mathbb{P}\,.$$

- In practice, it may not be possible to compute this integral and Monte Carlo (MC) methods are often used to simulate θ.

Monte Carlo

- Origins can be traced to the Manhattan Project (Von Neumann & Ulam)

Monte Carlo

- Origins can be traced to the Manhattan Project (Von Neumann & Ulam)
- The crude Monte Carlo (CMC) estimator of θ

$$\theta_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \approx \mathbb{E}[X].$$

- *Main idea behind MC-method:* Strong law of large numbers (SLLN)

Monte Carlo

- Origins can be traced to the Manhattan Project (Von Neumann & Ulam)
- The crude Monte Carlo (CMC) estimator of θ

$$\theta_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \approx \mathbb{E}[X].$$

- *Main idea behind MC-method:* Strong law of large numbers (SLLN) implies that $\theta_n(\omega) \to \theta$ almost surely.

Convergence rate of the CMC estimator

- SLLN implies that $\theta_n(\omega) \to \theta$ almost surely, but how large does n have to be?

Convergence rate of the CMC estimator

- SLLN implies that $\theta_n(\omega) \to \theta$ almost surely, but how large does n have to be?
- The convergence rate of the CMC estimator is $O(1/\sqrt{n})$.

Convergence rate of the CMC estimator

- SLLN implies that $\theta_n(\omega) \to \theta$ almost surely, but how large does n have to be?
- The convergence rate of the CMC estimator is $O(1/\sqrt{n})$.
- If we want $\mathbb{P}\big(|\theta_n - \theta| < \varepsilon|\theta|\big)$, then we need

$$n \gtrapprox \frac{z_{1-\alpha/2}^2 \sigma^2}{\varepsilon^2 \theta^2}$$

- The variance also controls the error!

Convergence rate of the CMC estimator

- SLLN implies that $\theta_n(\omega) \to \theta$ almost surely, but how large does n have to be?
- The convergence rate of the CMC estimator is $O(1/\sqrt{n})$.
- If we want $\mathbb{P}\big(|\theta_n - \theta| < \varepsilon|\theta|\big)$, then we need

$$n \gtrapprox \frac{z_{1-\alpha/2}^2 \sigma^2}{\varepsilon^2 \theta^2}$$

- The variance also controls the error!

Rare Events

- In many cases of interest θ is very small.

Rare Events

- In many cases of interest θ is very small.
- *Rare event* if $A \in \mathcal{F}$ satisfy $\mathbb{P}(A) \ll 1$ (less than10^{-3} or 10^{-4} in magnitude).

Rare Events

- In many cases of interest θ is very small.
- *Rare event* if $A \in \mathcal{F}$ satisfy $\mathbb{P}(A) \ll 1$ (less than10^{-3} or 10^{-4} in magnitude).
- The random variable $\mathbf{1_A}$ can be used to find the rare event probability:

Rare Events

- In many cases of interest θ is very small.
- *Rare event* if $A \in \mathcal{F}$ satisfy $\mathbb{P}(A) \ll 1$ (less than10^{-3} or 10^{-4} in magnitude).
- The random variable $\mathbf{1_A}$ can be used to find the rare event probability:

$$p := \mathbb{P}(A) = \int_A \mathrm{d}\mathbb{P} = \int_\Omega \mathbf{1}_A \, \mathrm{d}\mathbb{P} \, .$$

Rare Events

- In many cases of interest θ is very small.
- *Rare event* if $A \in \mathcal{F}$ satisfy $\mathbb{P}(A) \ll 1$ (less than10^{-3} or 10^{-4} in magnitude).
- The random variable $\mathbf{1_A}$ can be used to find the rare event probability:

$$p := \mathbb{P}(A) = \int_A \mathrm{d}\mathbb{P} = \int_\Omega \mathbf{1}_A \, \mathrm{d}\mathbb{P} \, .$$

Rare Events

- The CMC estimator is in this context be given by

$$\theta_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad X_i \sim \mathrm{Be}(p),$$

Rare Events

- The CMC estimator is in this context be given by

$$\theta_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad X_i \sim \mathrm{Be}(p),$$

- The variance is given by the expression

$$\mathbb{V}[\mathbf{1}_A] = p(1-p).$$

Rare Events

- The CMC estimator is in this context be given by

$$\theta_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad X_i \sim \mathrm{Be}(p),$$

- The variance is given by the expression

$$\mathbb{V}[\mathbf{1}_A] = p(1-p).$$

- For rare events the variance and expectation are almost the same!

$$\frac{\mathbb{V}[\mathbf{1}_A]}{\mathbb{E}[\mathbf{1}_A]} = 1 - p, \quad n \gtrsim \frac{z_{1-\alpha/2}^2}{\varepsilon^2 p}$$

Rare Events and Large Deviations

- MC methods for rare events are computationally demanding.

Rare Events and Large Deviations

- MC methods for rare events are computationally demanding.
- Another approach: Large Deviations

Rare Events and Large Deviations

- MC methods for rare events are computationally demanding.
- Another approach: Large Deviations
- Used to approximate probabilities at an exponential scale.

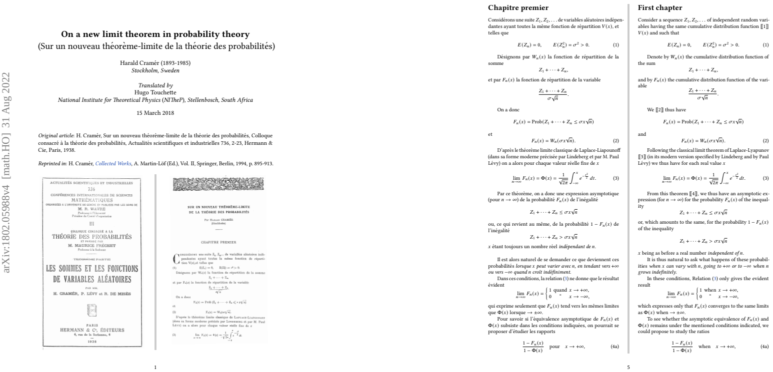
Rare Events and Large Deviations

- MC methods for rare events are computationally demanding.
- Another approach: Large Deviations
- Used to approximate probabilities at an exponential scale.
- First mathematical results in the theory of large deviations was published in 1938 [1], by Harald Cramér (actuary and affiliated with Stockholm University).
- Cramér's Motivation was insurance mathematics and ruin probabilites.

Rare Events and Large Deviations

- MC methods for rare events are computationally demanding.
- Another approach: Large Deviations
- Used to approximate probabilities at an exponential scale.
- First mathematical results in the theory of large deviations was published in 1938 [1], by Harald Cramér (actuary and affiliated with Stockholm University).
- Cramér's Motivation was insurance mathematics and ruin probabilites.
- S.R.S. Varadhan introduced the modern mathematical theory of large deviations. Seminal paper: [5] (Abel Prize for his contributions).

Large Deviations



arXiv:1802.05988v4 [math.HO] 31 Aug 2022

On a new limit theorem in probability theory
(Sur un nouveau théorème-limite de la théorie des probabilités)

Harald Cramér (1893-1985)
Stockholm, Sweden

Translated by
Hugo Touchette
National Institute for Theoretical Physics (NITheP), Stellenbosch, South Africa

15 March 2018

Original article: H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités, Colloque consacré à la théorie des probabilités, Actualités scientifiques et industrielles 736, 2-23, Hermann & Cie, Paris, 1938.

Reprinted in: H. Cramér, Collected Works, A. Martin-Löf (Ed.), Vol. II, Springer, Berlin, 1994, p. 895-913.

ACTUALITÉS SCIENTIFIQUES ET INDUSTRIELLES
736
CONFÉRENCES INTERNATIONALES DE SCIENCES MATHÉMATIQUES
III
THÉORIE DES PROBABILITÉS
LES SOMMES ET LES FONCTIONS DE VARIABLES ALÉATOIRES
H. CRAMÉR, P. LÉVY et R. DE MISES
PARIS
HERMANN & Cie, ÉDITEURS
1938

SUR UN NOUVEAU THÉORÈME-LIMITE DE LA THÉORIE DES PROBABILITÉS
Par Harald CRAMÉR (Stockholm)
CHAPITRE PREMIER

1

Chapitre premier

Considérons une suite $Z_1, Z_2, \ldots$ de variables aléatoires indépendantes ayant toutes la même fonction de répartition $V(x)$, et telles que

$$E(Z_n) = 0, \qquad E(Z_n^2) = \sigma^2 > 0. \tag{1}$$

Désignons par $W_n(x)$ la fonction de répartition de la somme

$$Z_1 + \cdots + Z_n,$$

et par $F_n(x)$ la fonction de répartition de la variable

$$\frac{Z_1 + \cdots + Z_n}{\sigma\sqrt{n}}.$$

On a donc

$$F_n(x) = \text{Prob}(Z_1 + \cdots + Z_n \leq \sigma x\sqrt{n})$$

et

$$F_n(x) = W_n(\sigma x\sqrt{n}). \tag{2}$$

D'après le théorème limite classique de Laplace-Liapounoff (dans sa forme moderne précisée par Lindeberg et par M. Paul Lévy) on a alors pour chaque valeur réelle fixe de x

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \tag{3}$$

Par ce théorème, on a donc une expression asymptotique (pour $n \to \infty$) de la probabilité $F_n(x)$ de l'inégalité

$$Z_1 + \cdots + Z_n \leq \sigma x\sqrt{n}$$

ou, ce qui revient au même, de la probabilité $1 - F_n(x)$ de l'inégalité

$$Z_1 + \cdots + Z_n > \sigma x\sqrt{n}$$

x étant toujours un nombre réel *indépendant de n*.

Il est alors naturel de se demander ce que deviennent ces probabilités *lorsque x peut varier avec n, en tendant vers $+\infty$ ou vers $-\infty$ quand n croît indéfiniment.*

Dans ces conditions, la relation (3) ne donne que le résultat évident

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 1 & \text{quand } x \to +\infty, \\ 0 & \text{''} \quad x \to -\infty, \end{cases}$$

qui exprime seulement que $F_n(x)$ tend vers les mêmes limites que $\Phi(x)$ lorsque $\to \pm\infty$.

Pour savoir si l'équivalence asymptotique de $F_n(x)$ et $\Phi(x)$ subsiste dans les conditions indiquées, on pourrait se proposer d'étudier les rapports

$$\frac{1 - F_n(x)}{1 - \Phi(x)} \quad \text{pour} \quad x \to +\infty, \tag{4a}$$

First chapter

Consider a sequence $Z_1, Z_2, \ldots$ of independent random variables having the same cumulative distribution function [[1]] $V(x)$ and such that

$$E(Z_n) = 0, \qquad E(Z_n^2) = \sigma^2 > 0. \tag{1}$$

Denote by $W_n(x)$ the cumulative distribution function of the sum

$$Z_1 + \cdots + Z_n,$$

and by $F_n(x)$ the cumulative distribution function of the variable

$$\frac{Z_1 + \cdots + Z_n}{\sigma\sqrt{n}}.$$

We [[2]] thus have

$$F_n(x) = \text{Prob}(Z_1 + \cdots + Z_n \leq \sigma x\sqrt{n})$$

and

$$F_n(x) = W_n(\sigma x\sqrt{n}). \tag{2}$$

Following the classical limit theorem of Laplace-Lyapunov [[3]] (in its modern version specified by Lindeberg and by Paul Lévy) we thus have for each real value x

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \tag{3}$$

From this theorem [[4]], we thus have an asymptotic expression (for $n \to \infty$) for the probability $F_n(x)$ of the inequality

$$Z_1 + \cdots + Z_n \leq \sigma x\sqrt{n}$$

or, which amounts to the same, for the probability $1 - F_n(x)$ of the inequality

$$Z_1 + \cdots + Z_n > \sigma x\sqrt{n}$$

x being as before a real number *independent of n*.

It is thus natural to ask what happens of these probabilities *when x can vary with n, going to $+\infty$ or to $-\infty$ when n grows indefinitely.*

In these conditions, Relation (3) only gives the evident result

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 1 & \text{when } x \to +\infty, \\ 0 & \text{''} \quad x \to -\infty, \end{cases}$$

which expresses only that $F_n(x)$ converges to the same limits as $\Phi(x)$ when $\to \pm\infty$.

To see whether the asymptotic equivalence of $F_n(x)$ and $\Phi(x)$ remains under the mentioned conditions indicated, we could propose to study the ratios

$$\frac{1 - F_n(x)}{1 - \Phi(x)} \quad \text{when} \quad x \to +\infty, \tag{4a}$$

5

Figure: Translation of Cramér's publication from French to English by Hugo Touchette [2]

Large Deviations

- A sequence (X_n) of i.i.d. random variables taking values in a Hausdorff topological space $\mathcal{X}$.
- In the theory of large deviations we want to find a rate function $I : \mathcal{X} \to [0, \infty]$, such that

$$\lim_{n\to\infty} \mathbb{P}(X_n \in A) \approx e^{-n \inf_{x\in A} I(x)}.$$

- Can be used to approximate rare event probabilites.

Large Deviations

- A sequence (X_n) of i.i.d. random variables taking values in a Hausdorff topological space $\mathcal{X}$.
- In the theory of large deviations we want to find a rate function $I : \mathcal{X} \to [0, \infty]$, such that

$$\lim_{n\to\infty} \mathbb{P}(X_n \in A) \approx e^{-n \inf_{x\in A} I(x)}.$$

- Can be used to approximate rare event probabilites.
- Can be used to analyze the convergence of MC estimators (substitute $X_n = \theta_n$).

Empirical Distributions

- Another interpretation of Monte Carlo estimators can be done through their empirical distributions.

Empirical Distributions

- Another interpretation of Monte Carlo estimators can be done through their empirical distributions.
- Let $X : \Omega \to \mathcal{X}$ be a random variable. If $f : \mathcal{X} \to \mathbb{R}$ is measurable, then $f(X)$ is a real valued random variable.

Empirical Distributions

- Another interpretation of Monte Carlo estimators can be done through their empirical distributions.
- Let $X : \Omega \to \mathcal{X}$ be a random variable. If $f : \mathcal{X} \to \mathbb{R}$ is measurable, then $f(X)$ is a real valued random variable.
- Let $\mathbf{M_1}(\mathcal{X})$ denote the collection of all probability measures on $\mathcal{X}$.

Empirical Distributions

- Another interpretation of Monte Carlo estimators can be done through their empirical distributions.
- Let $X : \Omega \to \mathcal{X}$ be a random variable. If $f : \mathcal{X} \to \mathbb{R}$ is measurable, then $f(X)$ is a real valued random variable.
- Let $\mathbf{M_1}(\mathcal{X})$ denote the collection of all probability measures on $\mathcal{X}$.

$$\delta : \Omega \to \mathbf{M_1}(\mathcal{X}), \quad \delta_{X(\omega)}(A) := \begin{cases} 1, & X(\omega) \in A, \\ 0, & X(\omega) \notin A, \end{cases}$$

Empirical Distributions

- Another interpretation of Monte Carlo estimators can be done through their empirical distributions.
- Let $X : \Omega \to \mathcal{X}$ be a random variable. If $f : \mathcal{X} \to \mathbb{R}$ is measurable, then $f(X)$ is a real valued random variable.
- Let $\mathbf{M_1}(\mathcal{X})$ denote the collection of all probability measures on $\mathcal{X}$.

$$\delta : \Omega \to \mathbf{M_1}(\mathcal{X}), \quad \delta_{X(\omega)}(A) := \begin{cases} 1, & X(\omega) \in A, \\ 0, & X(\omega) \notin A, \end{cases}$$

Empirical Distributions

- Empirical distribution of the CMC estimator:

$$\mathbf{L}_n(\omega) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}.$$

Empirical Distributions

- Empirical distribution of the CMC estimator:

$$\mathbf{L}_n(\omega) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}.$$

- Integrating over $\mathcal{X}$ with respect to the measure $\mathbf{L}_n(\omega)$:

$$\int_{\mathcal{X}} f \,\mathrm{d}\mathbf{L}_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \int_{\mathcal{X}} f \,\mathrm{d}\delta_{X_i(\omega)} = \frac{1}{n}\sum_{i=1}^{n} f\big(X_i(\omega)\big)$$

Empirical Distributions

- Empirical distribution of the CMC estimator:

$$\mathbf{L}_n(\omega) := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(\omega)}.$$

- Integrating over $\mathcal{X}$ with respect to the measure $\mathbf{L}_n(\omega)$:

$$\int_{\mathcal{X}} f \,\mathrm{d}\mathbf{L}_n(\omega) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{X}} f \,\mathrm{d}\delta_{X_i(\omega)} = \frac{1}{n}\sum_{i=1}^{n} f\big(X_i(\omega)\big)$$

- Does $\mathbf{L}_n$ converge to the distribution μ of $\mathcal{X}$ in $\mathbf{M}_1(\mathcal{X})$?

Topologies on $\mathbf{M}(\mathcal{X})$

- Convergence is a topological concept.

Topologies on $\mathbf{M}(\mathcal{X})$

- Convergence is a topological concept.
- The empirical distributions of some MC-estimators are not probability measures.

Topologies on $\mathbf{M}(\mathcal{X})$

- Convergence is a topological concept.
- The empirical distributions of some MC-estimators are not probability measures.

Three important spaces of measures:

1. $\mathbf{M}(\mathcal{X})$ finite signed measures on $\mathcal{X}$ (is a linear space)
2. $\mathbf{M}_+(\mathcal{X})$ nonnegative finite measures on $\mathcal{X}$
3. $\mathbf{M}_1(\mathcal{X})$ probability measures on $\mathcal{X}$

The τ-topology

- Let $f : \mathcal{X} \to \mathbb{R}$ be a bounded measurable function and $\mu \in \mathbf{M}(\mathcal{X})$. Then

$$\langle f, \mu \rangle = \int_{\mathcal{X}} f \, \mathrm{d}\mu \,,$$

is a dual pairing.

The τ-topology

- Let $f : \mathcal{X} \to \mathbb{R}$ be a bounded measurable function and $\mu \in \mathbf{M}(\mathcal{X})$. Then
$$\langle f, \mu \rangle = \int_{\mathcal{X}} f \, \mathrm{d}\mu \,,$$
is a dual pairing.
- The maps $\langle f, \cdot \rangle : \mathbf{M}(\mathcal{X}) \to \mathbb{R}$ generate a weak topology on $\mathbf{M}(\mathcal{X})$: the τ-topology

The τ-topology

- Let $f : \mathcal{X} \to \mathbb{R}$ be a bounded measurable function and $\mu \in \mathbf{M}(\mathcal{X})$. Then
$$\langle f, \mu \rangle = \int_{\mathcal{X}} f \, \mathrm{d}\mu \, ,$$
is a dual pairing.
- The maps $\langle f, \cdot \rangle : \mathbf{M}(\mathcal{X}) \to \mathbb{R}$ generate a weak topology on $\mathbf{M}(\mathcal{X})$: the τ-topology
- μ_α converges to μ in the τ-topology iff
$$\lim_{\alpha} \int_{\mathcal{X}} f \, \mathrm{d}\mu_\alpha = \int_{\mathcal{X}} f \, \mathrm{d}\mu \, ,$$
for every bounded measurable function f.

The topology of weak convergence

- The τ-topology does not capture any topological information of $\mathcal{X}$.

The topology of weak convergence

- The τ-topology does not capture any topological information of $\mathcal{X}$.
- If $\mathcal{X}$ is a metrizable space we can restrict the class of "test functions".

The topology of weak convergence

- The τ-topology does not capture any topological information of $\mathcal{X}$.
- If $\mathcal{X}$ is a metrizable space we can restrict the class of "test functions".

Definition

Let $\mathcal{X}$ be a metrizable space, then a net (μ_α) in $\mathbf{M}(\mathcal{X})$ *converges weakly* to $\mu \in \mathbf{M}(\mathcal{X})$ if

$$\lim_\alpha \int_\mathcal{X} f \, \mathrm{d}\mu_\alpha = \int_\mathcal{X} f \, \mathrm{d}\mu \,, \quad \text{for every } f \in C_b(\mathcal{X}).$$

The topology of weak convergence

- The τ-topology does not capture any topological information of $\mathcal{X}$.
- If $\mathcal{X}$ is a metrizable space we can restrict the class of "test functions".

Definition

Let $\mathcal{X}$ be a metrizable space, then a net (μ_α) in $\mathbf{M}(\mathcal{X})$ *converges weakly* to $\mu \in \mathbf{M}(\mathcal{X})$ if

$$\lim_\alpha \int_{\mathcal{X}} f \, \mathrm{d}\mu_\alpha = \int_{\mathcal{X}} f \, \mathrm{d}\mu \,, \quad \text{for every } f \in C_b(\mathcal{X}).$$

- Usually denoted $\mu_\alpha \implies \mu$.

The topology of weak convergence

- The τ-topology does not capture any topological information of $\mathcal{X}$.
- If $\mathcal{X}$ is a metrizable space we can restrict the class of "test functions".

Definition

Let $\mathcal{X}$ be a metrizable space, then a net (μ_α) in $\mathbf{M}(\mathcal{X})$ *converges weakly* to $\mu \in \mathbf{M}(\mathcal{X})$ if

$$\lim_\alpha \int_{\mathcal{X}} f \, \mathrm{d}\mu_\alpha = \int_{\mathcal{X}} f \, \mathrm{d}\mu \,, \quad \text{for every } f \in C_b(\mathcal{X}).$$

- Usually denoted $\mu_\alpha \implies \mu$.

The topology of weak convergence is weaker (has less open sets) than the τ-topology!

- Weak convergence of measures is often used in probability theory.

- Weak convergence of measures is often used in probability theory.
- In Ch. 3 of the thesis we study the the τ-topology and the topology of weak convergence.

- Weak convergence of measures is often used in probability theory.
- In Ch. 3 of the thesis we study the the τ-topology and the topology of weak convergence.
- Extend many results from $\mathbf{M_1}(\mathcal{X})$ to $\mathbf{M}_+(\mathcal{X})$.

- Weak convergence of measures is often used in probability theory.
- In Ch. 3 of the thesis we study the the τ-topology and the topology of weak convergence.
- Extend many results from $\mathbf{M_1}(\mathcal{X})$ to $\mathbf{M}_+(\mathcal{X})$.
- Builds upon the work of Varadarajan in [4].

Convergence of MC-estimators

Varadarajan proved in 1958 the empirical distributions of the CMC-estimator converges weakly [3].

Convergence of MC-estimators

Varadarajan proved in 1958 the empirical distributions of the CMC-estimator converges weakly [3].

Theorem

Let $\mathcal{X}$ be a separable metrizable space and (X_i) a sequence of i.i.d. random variables taking values in $\mathcal{X}$ with law μ. Then the empirical distributions $\mathbf{L}_n$ converge weakly to μ almost surely, i.e.

$$\mathbb{P}(\{\omega \in \Omega \ : \ \mathbf{L}_n(\omega) \implies \mu\} = 1.$$

We extend this result to the empirical distribution of the importance sampling estimator.

Large Deviations

- Let $\mathcal{X}$ be a topological space and $\mathcal{B}$ a σ-algebra on $\mathcal{X}$ that contain all open sets.

Large Deviations

- Let $\mathcal{X}$ be a topological space and $\mathcal{B}$ a σ-algebra on $\mathcal{X}$ that contain all open sets.
- A function $f : \mathcal{X} \to [0, \infty]$ is said to be a *good rate function* if it is lower semicontinuous with compact level sets

$$\{x \in \mathcal{X} \ : \ f(x) \leq t\}, \quad t \in [0\infty].$$

Large Deviations

- Let $\mathcal{X}$ be a topological space and $\mathcal{B}$ a σ-algebra on $\mathcal{X}$ that contain all open sets.
- A function $f : \mathcal{X} \to [0, \infty]$ is said to be a *good rate function* if it is lower semicontinuous with compact level sets

$$\{x \in \mathcal{X} \ : \ f(x) \leq t\}, \quad t \in [0\infty].$$

A sequence of probability measures (μ_n) the *large deviation principle (LDP)* with rate function I if

$$-\inf_U I \leq \liminf_{n\to 0} \frac{1}{n} \log\left[\mu_n(U)\right]$$

for every open set U, and

$$-\inf_C I \geq \limsup_{n\to 0} \frac{1}{n} \log\left[\mu_n(C)\right]$$

for every closed set C.

Large Deviations

Main steps for proving that (μ_n) satisfies the LDP.

Large Deviations

Main steps for proving that (μ_n) satisfies the LDP.

1. Prove that (μ_n) satisfy the LDP lower bound:

$$-\inf_U I \leq \liminf_{n\to 0} \frac{1}{n} \log\left[\mu_n(U)\right]$$

for every open set U.

Main steps for proving that (μ_n) satisfies the LDP.

1. Prove that (μ_n) satisfy the LDP lower bound:

$$-\inf_U I \le \liminf_{n\to 0} \frac{1}{n} \log\left[\mu_n(U)\right]$$

   for every open set U.

2. Prove that (μ_n) satisfy LDP upper bound:

$$-\inf_C I \ge \limsup_{n\to 0} \frac{1}{n} \log\left[\mu_n(C)\right]$$

   for every closed set C.

Large Deviations

Main steps for proving that (μ_n) satisfies the LDP.

1. Prove that (μ_n) satisfy the LDP lower bound:

$$-\inf_U I \leq \liminf_{n\to 0} \frac{1}{n} \log\left[\mu_n(U)\right]$$

   for every open set U.

2. Prove that (μ_n) satisfy LDP upper bound:

$$-\inf_C I \geq \limsup_{n\to 0} \frac{1}{n} \log\left[\mu_n(C)\right]$$

   for every closed set C.

3. Identification of the rate function.

Large Deviations

- If $\mathcal{X}$ is a regular topological space, then the rate function I is unique.

Large Deviations

- If $\mathcal{X}$ is a regular topological space, then the rate function I is unique.
- The LDP is unique!
- The upper bound can be hard to prove...

Large Deviations

- If $\mathcal{X}$ is a regular topological space, then the rate function I is unique.
- The LDP is unique!
- The upper bound can be hard to prove...
- (μ_n) satisfies *a weak large deviation principle* with rate function I if it satisfies the lower bound and

$$-\inf_K I \geq \limsup_{n\to\infty} \frac{1}{n} \log\left[\mu_n(K)\right]$$

for every compact set $K \subset \mathcal{X}$.

- If $\mathcal{X}$ is a regular topological space, then the rate function I is unique.
- The LDP is unique!
- The upper bound can be hard to prove...
- (μ_n) satisfies *a weak large deviation principle* with rate function I if it satisfies the lower bound and

$$-\inf_K I \geq \limsup_{n\to\infty} \frac{1}{n} \log \left[\mu_n(K)\right]$$

for every compact set $K \subset \mathcal{X}$.

- Easier to prove a weak LDP. Can go from weak to full LDP by proving that the sequence (μ_n) is exponentially tight.

Large Deviations of empirical means

- Consider (X_i) i.i.d. random variables with distribution μ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in the topological linear space $\mathcal{X}$.

Large Deviations of empirical means

- Consider (X_i) i.i.d. random variables with distribution μ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in the topological linear space $\mathcal{X}$.
- The empirical means $\mathbf{S}_n : \Omega \to \mathcal{X}$, defined by

$$\mathbf{S}_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} X_i(\omega).$$

- The distributions are given by $\mu_n = \mathbb{P} \circ \mathbf{S}_n{}^{-1}$.

Cramér's Theorem

Theorem (Weak Cramér's Theorem)

The sequence (μ_n) of distributions of the empirical means satisfy a weak large deviation principle with a convex rate function $I = \Lambda^*$, and

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x\in A} \Lambda^*(x),$$

for every convex and open $A \subset \mathcal{X}$.

Here Λ^* is the Legendre-Fenchel transform of

$$\Lambda(\lambda) = \Lambda_\mu := (\lambda) = \log \mathbb{E}\left[e^{\langle \lambda, X\rangle}\right] = \log \left[\int_{\mathcal{X}} e^{\langle \lambda, d\rangle}\, \mathrm{d}\mu(x)\right].$$

Large Deviations and CMC

- If we want

$$\mathbb{P}\Big(|\theta_n - \theta| < \varepsilon|\theta| \Big) > 1 - \alpha.$$

Large Deviations and CMC

- If we want

$$\mathbb{P}\Big(|\theta_n - \theta| < \varepsilon|\theta| \Big) > 1 - \alpha.$$

- Let

$$R_\varepsilon := B\left(\theta, \varepsilon|\theta|\right), \quad A_\varepsilon = R_\varepsilon^c,$$

Large Deviations and CMC

- If we want

$$\mathbb{P}\Big(|\theta_n - \theta| < \varepsilon|\theta| \Big) > 1 - \alpha.$$

- Let

$$R_\varepsilon := B\left(\theta, \varepsilon|\theta|\right), \quad A_\varepsilon = R_\varepsilon^c,$$

then we get

$$\mathbb{P}(\theta_n \in A_\varepsilon) \leq \alpha.$$

Large Deviations and CMC

- If we want

$$\mathbb{P}\Big(|\theta_n - \theta| < \varepsilon|\theta| \Big) > 1 - \alpha.$$

- Let

$$R_\varepsilon := B\left(\theta, \varepsilon|\theta|\right), \quad A_\varepsilon = R_\varepsilon^c,$$

then we get

$$\mathbb{P}(\theta_n \in A_\varepsilon) \leq \alpha.$$

- For large n we can interpret this as

$$\mu_n(A_\varepsilon) \lessapprox e^{-nI(A_\varepsilon)}.$$

- To achieve the desired precision with confidence $1 - \alpha$ we want

$$e^{-nI(A_\varepsilon)} \leq \alpha.$$

Large Deviations and CMC

- To achieve the desired precision with confidence $1 - \alpha$ we want

$$e^{-nI(A_\varepsilon)} \leq \alpha.$$

- Taking logarithms of both sides we get

$$nI(A_\varepsilon) \geq \log(\alpha)$$

Large Deviations and CMC

- To achieve the desired precision with confidence $1 - \alpha$ we want

$$e^{-nI(A_\varepsilon)} \leq \alpha.$$

- Taking logarithms of both sides we get

$$nI(A_\varepsilon) \geq \log(\alpha)$$

- Rearranging yields:

$$n \gtrapprox \frac{\log(\alpha)}{I(A_\varepsilon)}.$$

Example with $X \sim N(\theta, \sigma^2)$

- Let $\theta = \mathbb{E}[X]$ for $X \sim N(\theta, \sigma^2)$. Then the rate function is given by

$$I(x) = \frac{(x-\theta)^2}{2\sigma^2},$$

Example with $X \sim N(\theta, \sigma^2)$

- Let $\theta = \mathbb{E}[X]$ for $X \sim N(\theta, \sigma^2)$. Then the rate function is given by

$$I(x) = \frac{(x-\theta)^2}{2\sigma^2},$$

- and

$$I(A_\varepsilon) = \inf\{I(x) : |x-\theta| > \varepsilon|\theta|\} = \frac{\varepsilon^2\theta^2}{2\sigma^2}$$

Example with $X \sim N(\theta, \sigma^2)$

- Let $\theta = \mathbb{E}[X]$ for $X \sim N(\theta, \sigma^2)$. Then the rate function is given by

$$I(x) = \frac{(x-\theta)^2}{2\sigma^2},$$

- and

$$I(A_\varepsilon) = \inf\{I(x) : |x-\theta| > \varepsilon|\theta|\} = \frac{\varepsilon^2\theta^2}{2\sigma^2}$$

- Gives the LDP bound for the sample size:

$$n \gtrapprox \frac{\log(\alpha)2\sigma^2}{\varepsilon^2\theta^2}.$$

Example with $X \sim N(\theta, \sigma^2)$

- Let $\theta = \mathbb{E}[X]$ for $X \sim N(\theta, \sigma^2)$. Then the rate function is given by

$$I(x) = \frac{(x-\theta)^2}{2\sigma^2},$$

- and

$$I(A_\varepsilon) = \inf\{I(x) : |x-\theta| > \varepsilon|\theta|\} = \frac{\varepsilon^2\theta^2}{2\sigma^2}$$

- Gives the LDP bound for the sample size:

$$n \gtrapprox \frac{\log(\alpha)2\sigma^2}{\varepsilon^2\theta^2}.$$

- Compare with introduction:

$$n \gtrapprox \frac{z_{1-\alpha/2}^2\sigma^2}{\varepsilon^2\theta^2}.$$

Convexity and Large Deviations

Definition

Let $f : \mathcal{X} \to [-\infty, \infty]$, then the *Legendre-Fenchel transform*of f is the function $f : \mathcal{X}^* \to [-\infty, \infty]$ defined by

$$\begin{aligned} f^*(\lambda) &= \sup\{\langle \lambda, x \rangle - f(x) \ : \ x \in \mathcal{X}\} \\ &= -\inf\{f(x) - \langle \lambda, x \rangle \ : \ x \in \mathcal{X}\}. \end{aligned}$$

Convexity and Large Deviations

Definition

Let $f : \mathcal{X} \to [-\infty, \infty]$, then the *Legendre-Fenchel transform*of f is the function $f : \mathcal{X}^* \to [-\infty, \infty]$ defined by

$$\begin{aligned} f^*(\lambda) &= \sup\{\langle \lambda, x\rangle - f(x) \ : \ x \in \mathcal{X}\} \\ &= -\inf\{f(x) - \langle \lambda, x\rangle \ : \ x \in \mathcal{X}\}. \end{aligned}$$

Theorem (Biconjugate Theorem)

Let $f : \mathcal{X} \to (-\infty, \infty]$ not be identically ∞, then $f = f^{**}$ if and only if f is convex and lower semicontinuous.

What we have not covered

- Sanov's Theorem and LDP for the distributions of the empirical distributions.

What we have not covered

- Sanov's Theorem and LDP for the distributions of the empirical distributions.
- Projective Limits.

What we have not covered

- Sanov's Theorem and LDP for the distributions of the empirical distributions.
- Projective Limits.
- Weak Convergence Approach.

What we have not covered

- Sanov's Theorem and LDP for the distributions of the empirical distributions.
- Projective Limits.
- Weak Convergence Approach.

Questions

Feel free to ask any questions!

References I

[1] Harald Cramér. "Sur un nouveau theoreme-limite de la theorie des probabilities". In: *Scientifiques et Industrielles* 736 (1938), pp. 5–23.

[2] Harald Cramér and Hugo Touchette. *On a new limit theorem in probability theory (Translation of 'Sur un nouveau théorème-limite de la théorie des probabilités')*. 2022. arXiv: `1802.05988 [math.HO]`.

[3] Veeravalli S Varadarajan. "On the convergence of sample probability distributions". In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 19.1/2 (1958), pp. 23–26.

[4] Veeravalli S Varadarajan. "Weak convergence of measures on separable metric spaces". In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 19.1/2 (1958), pp. 15–22.

References II

[5] S. R. S. Varadhan. "Asymptotic probabilities and differential equations". In: *Communications on Pure and Applied Mathematics* 19.3 (1966), pp. 261–286. DOI: https://doi.org/10.1002/cpa.3160190303. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160190303. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160190303.