The Annals of Probability
1980, Vol. 8, No. 6, 1177–1182

LAPLACE'S METHOD REVISITED: WEAK CONVERGENCE OF PROBABILITY MEASURES[1]

BY CHII-RUEY HWANG

Brown University[2]

Let Q be a fixed probability on the Borel σ-field in R^n and H be an energy function continuous in R^n. A set N is related to H by $N = \{x \mid \inf_y H(y) = H(x)\}$. Laplace's method, which is interpreted as weak convergence of probabilities, is used to introduce a probability P on N. The general properties of P are studied. When N is a union of smooth compact manifolds and H satisfies some smooth conditions, P can be written in terms of the intrinsic measures on the highest dimensional mainfolds in N.

1. Introduction. Let Q be a fixed probability measure on $(R^n, \mathscr{B})$, where $\mathscr{B}$ is the Borel σ-algebra, and H be a continuous function in R^n with the following assumption

$$\text{(A1)} \qquad Q\{H(x) < a\} > 0 \text{ if } a > \inf_x H(x).$$

A set N is related to H by $N = \{x \mid H(x) = \inf_y H(y)\}$. To introduce a probability P on N, we appeal to Laplace's method which can be interpreted as weak convergence of probability measures. So in this article we shall study the limiting properties of $\{P_\theta \mid \theta > 0\}$ as $\theta \downarrow 0$, where P_θ is defined similarly to Laplace's method by

$$\text{(1.1)} \qquad \frac{dP_\theta}{dQ}(x) = \exp\left(-\frac{H(x)}{\theta}\right)\left[\int \exp\left(-\frac{H(x)}{\theta}\right) dQ(x)\right]^{-1}.$$

Usually H is called the energy function and θ the temperature parameter. If $P_\theta \to P$ weakly, then P should be the one we want.

$x = (x_1, \cdots, x_n)$ may be regarded as the state of some system and Q is a given probability associated with the system. Owing to some regularity constraint, x is confined to a subset N. For example, in statistical mechanics the constraint may be "conservation of energy", or, in pattern theory (Grenander [3], page 63) the constraint may be "equality of bond values". Of course, there are different ways to introduce a probability on N. In statistical mechanics, there is a natural way to do so on the constant energy surface N by using the gradient of the Hamiltonian H (Khinchin [5]). But this approach is not suitable here, since the gradient of H is identically zero on N. The method defined by (1.1) seems reasonable. Clearly P depends on H. To see what P looks like in a simple case, let us consider the following example: N is the unit sphere centered at 0 in R^n and $x_1, \cdots, x_n$ are i.i.d. $\mathscr{N}(0, 1)$. The function $H(x)$ is the square of the distance between x and N. Then all the assumptions in Theorem 3.1 are satisfied. It is easily seen that P concentrates on N and is the normalized surface area of N.

The necessary condition for $\{P_\theta\}$ being tight, i.e., H has a minimum, is provided by Proposition 2.1. Its corollary says that if there is a limiting probability measure then it concentrates on the minimal energy states. Propositions 2.2 and 2.3 give sufficient conditions for the tightness of $\{P_\theta\}$. To get the explicit expression for the limiting probability measure P, we assume $H \in C^3$ and Q has a continuous density. Theorem 2.1 corresponds to the classical Laplace's formula. The main result is Theorem 3.1, where P is written in terms of the intrinsic measures on the highest dimensional manifolds of the minimal energy states.

Received December 5, 1978.
[1] This research was supported by NSF Grant MCS 76-80762.
[2] Current address: Institute of Mathematics, Academia Sinica, Taipei, Taiwan, Republic of China.
AMS 1970 *subject classifications.* Primary 60B10; secondary 58C99.
Key words and phrases. Laplace's method, smooth manifold, weak convergence.

2. General results.

PROPOSITION 2.1. *If H does not have a minimum then P_θ is not tight.*

PROOF. If there exists a sequence $P_\theta \to P$ weakly as $\theta \downarrow 0$, then we can choose $\{a_m\}$ such that $a_m \downarrow \inf_x H(x)$ and $P\{H(x) = a_m\} = 0$. Since H is continuous, $\partial\{H(x) \geq a_m\} \subseteq \{H(x) = a_m\}$. Then, for each m, $P(\partial\{H(x) \geq a_m\}) = 0$ and $P_\theta\{H(x) \geq a_m\} \to P\{H(x) \geq a_m\}$. But

$$P_\theta\{H(x) \geq a_m\} \leq \left[\int \exp\left(-\frac{(H(x) - a_m)}{\theta}\right) dQ\right]^{-1}$$
$$\leq \left[\int_{H(x)<a_m} \exp\left(-\frac{(H(x) - a_m)}{\theta}\right) dQ\right]^{-1}$$
$$\to 0 \text{ as } \theta \downarrow 0.$$

Hence, $P\{H(x) \geq a_m\} = 0\ \forall\ m$ which implies $\lim_{m\to\infty} P\{H(x) \geq a_m\} = P(R^n) = 0$, a contradiction. □

Because of Proposition 2.1, we assume

(A2) $\min_x H(x)$ exists and equals 0.

Let N be the set of all minimal energy states, then

COROLLARY 2.1. *If $P_\theta \to P$ weakly, then P concentrates on N.*

PROPOSITION 2.2. *If $Q(N) = m > 0$ then the limiting probability measure P exists and is uniformly distributed on N* w.r.t. Q.

PROOF. The density functions

$$f_\theta(x) \to f(x) = 0 \quad \text{if} \quad x \notin N$$
$$= \frac{1}{m} \quad \text{if} \quad x \in N.$$

If we define P by $(dP/dQ)(x) = f(x)$, then $P_\theta \to P$ weakly and P distributes uniformly on N.

The interesting case is when

(A3) $Q(N) = 0.$

PROPOSITION 2.3. *$\{P_\theta\}$ is tight, if*

(A4) *there exists $\epsilon > 0$ such that $\{H(x) \leq \epsilon\}$ is compact.*

PROOF.

$$P_\theta\{H(x) > \epsilon\} \leq \left(\int_{H(x)>\epsilon} \exp\left(-\frac{(H(x) - \epsilon)}{\theta}\right) dQ\right)^{-1}$$
$$\to 0 \text{ as } \theta \downarrow 0.$$

So for any $\delta > 0$ there exists $\theta_0 > 0$ such that $P_\theta\{H(x) \leq \epsilon\} \geq 1 - \delta$ for $\theta \leq \theta_0$.

Note that the compactness of N alone is not sufficient for the tightness of $\{P_\theta\}$. A counterexample can be found in (Hwang [4]).

Under assumptions (A1) to (A4) we shall study how P distributes on the minimal energy states N. Since $Q(N) = 0$, it seems that we cannot get enough information from Q unless Q is well connected to the analytic structure of R^n. Hence we make the following assumption

(A5) $H(x) \in C^3(R^n)$, $(dQ/d\mu)(x) = f(x)$, where μ is the Lebesgue measure on $(R^n, \mathscr{B})$ and f is chosen to be continuous.

THEOREM 2.1. *If* $N = \{x_1, \cdots, x_m\}$, $\det H''(x_i) \neq 0$ *for any i and there exists k such that* $f(x_k) > 0$, *then*

$$P\{x_i\} = \frac{f(x_i)(\det H''(x_i))^{-1/2}}{\sum_1^m f(x_l)(\det H''(x_l))^{-1/2}}.$$

PROOF. Choose a closed neighborhood A_i of x_i such that $x_l \notin A_i$ *if* $l \neq i$; then, by Laplace's formula

$$P_\theta(A_i) = \frac{\int_{Ai} \exp\left(-\frac{H(x)}{\theta}\right) f(x)\, dx}{\int_{R^n} \exp\left(-\frac{H(x)}{\theta}\right) f(x)\, dx}$$

$$\rightarrow \frac{f(x_i)(\det H''(x_i))^{-1/2}}{\sum_l f(x_l)(\det H''(x_l))^{-1/2}}. \qquad \square$$

3. N as the union of smooth manifolds. In addition to the previous assumptions, we also assume that each component of N is a smooth manifold (or C^3-manifold). These manifolds may be of different dimensions. We also assume that N has finitely many components. Will the limiting probability measure P concentrate on the highest dimensional manifolds? How do we write P in terms of a known measure in the manifolds? When θ is small enough, the major contribution is in a small neighborhood of N. Since the gradient of H at each point of N is 0, the implicit function theorem is not applicable here. In a small neighborhood of N we change the coordinate system to a polar coordinate system along N, then P can be written in terms of some intrinsic measures on the manifolds.

Let M be a k-dimensional compact smooth manifold in R^n. Then, by the *tubular neighborhood theorem* (Milnor [6], page 115), there exists a tubular neighborhood $T(\epsilon)$ of M such that for any z in $T(\epsilon)$, z can be written as $m + v$, where m is a point on M and $v \perp M$ at m with $|v| < \epsilon$. The map $z \to (m, v)$ is a diffeomorphism.

In order to relate $dx = dx_1 \cdots dx_n$ to a natural intrinsic measure on M in a tubular neighborhood $T(\epsilon)$, we need some facts from differential geometry. Let $m = m(u^1, \cdots, u^k)$ be local coordinates and $\mathscr{N}(1), \cdots \mathscr{N}(n-k)$ be normalized smooth normal vectors of M. For any $z \in T(\epsilon)$

$$z = m(u^1, \cdots, u^k) + t_1 \mathscr{N}(1) + \cdots + t_{n-k} \mathscr{N}(n-k),$$

where $|(t_1, \cdots, t_{n-k})| < \epsilon$. Then,

$$dx_1 \cdots dx_n = |\delta_\alpha^\beta + \sum_i t_i G_\alpha^\beta(i)|\, dt_1 \cdots dt_{n-k}\, d\mathscr{M},$$

with the following notations:

δ_α^β : Kronecker's notation,

$G_\alpha^\beta(i)$: $\dfrac{\partial N(i)}{\partial u^\alpha} = \sum_\beta G_\alpha^\beta(i) \dfrac{\partial m}{\partial u^\beta} + \bar{\mathscr{N}},$

where $\bar{\mathscr{N}}$ is a linear combination of $\mathscr{N}(i)$'s (Weyl [7]),

$\mathscr{M}$: a natural intrinsic measure on M (Boothby [2] Chapter 6),

$|\delta_\alpha^\beta + \sum_i t_i G_\alpha^\beta(i)|$: the abbreviation of $|\det[\delta_\alpha^\beta + \sum_i t_i G_\alpha^\beta(i)]_{\alpha,\beta}|$.

THEOREM 3.1. *Assume that N has finitely many components and each component is a compact smooth manifold. The energy function H and probability Q satisfy* (A1) *to* (A5). *If the density f in* (A5) *is not identically zero on the highest dimensional manifolds and* det $\partial^2 H/\partial t^2(u) \neq 0$ *for* $u \in N$, *then the limiting probability measure concentrates on the highest dimensional manifolds and can be written as*:

$$\frac{dP}{d\mathcal{M}}(u) = \frac{f(u)\left(\det \frac{\partial^2 H}{\partial t^2}(u)\right)^{-1/2}}{\int_N f(u)\left(\det \frac{\partial^2 H}{\partial t^2}(u)\right)^{-1/2} d\mathcal{M}}$$

where $\mathcal{M}$ *is the sum of intrinsic measures on the highest dimensional manifolds.*

PROOF. Let $\{M_l\}_1^q$ be the components of N and g be a bounded continuous function from R^n to R. Consider

$$\int_{R^n} \exp\left(\frac{-H(z)}{\theta}\right) f(z)g(z)\, dz. \tag{3.1}$$

The difference between

$$\sum_{l=1}^q \int_{T_l(\epsilon)} \exp\left(\frac{-H(z)}{\theta}\right) f(z)g(z)\, dz$$

and (3.1) is exponentially small, where $T_l(\epsilon)$ is an ϵ-tabular neighborhood of M_l, $T_l(\epsilon) \cap T_d(\epsilon) = \phi$ if $l \neq d$, and $T_l(\epsilon)$ is chosen closed.

Fix l, and consider

$$\int_{T_l(\epsilon)} \exp\left(\frac{-H(z)}{\theta}\right) f(z)g(z)\, dz$$

$$= \int_{M_l}\int_{|t|\le\epsilon} \exp\left(\frac{-H(t,u)}{\theta}\right) f(t,u)g(t,u)\,|\delta_\alpha^\beta + \textstyle\sum_i t_i G_\alpha^\beta(i)|\, dt_1 \cdots dt_{n-k_l} d\mathcal{M}_l$$

where $\mathcal{M}_l$ is the intrinsic measure on M_l and k_l is the dimension of M_l. For each fixed u, by Laplace formula

$$\frac{\int_{|t|\le\epsilon} \exp\left(\frac{-H(t,u)}{\theta}\right) f(t,u)g(t,u)\,|\delta_\alpha^\beta + \sum_i t_i G_\alpha^\beta(i)|\, dt_1 \cdots dt_{n-k_l}}{(2\pi\theta)^{(n-k_l)/2^2}}$$

$$\to f(0,u)g(0,u)\left(\det \frac{\partial^2 H}{\partial t^2}(0,u)\right)^{-1/2}, \tag{3.2}$$

$$H(t,u) = \frac{1}{2} < \frac{\partial^2 H}{\partial t^2}(0,u)t, t> + \frac{1}{6}\frac{\partial^3 H}{\partial t^3}(\bar{t},u)(t)$$

where $\bar{t} \in$ segment $(0, t)$ and

$$\frac{\partial^3 H}{\partial t^3}(\bar{t},u)(t) = \sum_{i=1}^{k_l}\sum_{j=1}^{k_l}\sum_{k=1}^{k_l} \frac{\partial^3 H}{\partial t_i \partial t_j \partial t_k}(\bar{t},u) t_k t_j t_i.$$

Let $\lambda(u)$ be the minimal eigenvalues of $(\partial^2 H/\partial t^2)(0,u)$. Because $(\partial^2 H/\partial t^2)(0,u)$ is positive definite and M_l is compact, $\min_{u\in M_l} \lambda(u) = \lambda > 0$. Choose $0 < 2\delta_l < \lambda$, then $H(t,u) \ge \delta_l |t|^2 + \frac{1}{6}(\partial^3 H/\partial t^3)(\bar{t},u)(t)$.

Let

$$B = \max_{ijk}\max_{|t|\le\epsilon; u\in M}\left|\frac{\partial^3 H}{\partial t_i \partial t_j \partial t_k}(t, u)\right|,$$

then $B < \infty$. We can choose ϵ_l small enough such that

$$\tfrac{1}{2}\,\delta_l |t|^2 - \tfrac{1}{6}\textstyle\sum_{ijk} B|t_i t_j t_k| \ge 0 \text{ for any } |t| \le \epsilon_l.$$

Then for any $|t| \le \epsilon_l$ we have $H(t, u) \ge \frac{1}{2}\,\delta_l |t|^2$. So we can replace ϵ by ϵ_l and still have the same result in (3.2).

Let

$$A(\theta, l) = \int_{u\in M_l}\int_{|t|\le\epsilon_l} \exp\left(\frac{-H(t, u)}{\theta}\right) f(t, u)g(t, u)\,|\delta_\alpha^\beta + \textstyle\sum_i t_i G_\alpha^\beta(i)|\,dt_1 \cdots dt_{n-k_l}\, d\mathscr{M}_l,$$

$$\bar{A}(\theta, l) = \int_{u\in M_l}\int_{|t|\le\epsilon_l} \exp\left(\frac{-H(t, u)}{\theta}\right) f(t, u)\,|\delta_\alpha^\beta + \textstyle\sum_i t_i G_\alpha^\beta(i)|\,dt_1 \cdots dt_{n-k_l}\, d\mathscr{M}_l.$$

Because $0 \le \exp(-H(t, u)/\theta) \le \exp(-\delta_l |t|^2/2\theta)$, $\mathscr{M}_l(M_l) < \infty$, by (3.2) and the dominated convergence theorem, we have

$$\frac{A(\theta, l)}{(2\pi\theta)^{(n-k_l)/2}} \to \int_{u\in M_l} f(0, u)g(0, u)\left(\det\frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} d\mathscr{M}_l, \tag{3.3}$$

$$\frac{\bar{A}(\theta, l)}{(2\pi\theta)^{(n-k_l)/2}} \to \int_{u\in M_l} f(0, u)\left(\det\frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} d\mathscr{M}_l. \tag{3.4}$$

Let $\max_{1\le l\le q} k_l = m$,

$$\frac{\displaystyle\int_{R^n} \exp\left(\frac{-H(z)}{\theta}\right) f(z)g(z)\,dz}{\displaystyle\int_{R^n} \exp\left(\frac{-H(z)}{\theta}\right) f(z)\,dz}$$

$$\approx \frac{\sum_l A(\theta, l)}{\sum_l \bar{A}(\theta, l)} = \frac{\sum_l A(\theta, l)(2\pi\theta)^{-(n-m)/2}}{\sum_l \bar{A}(\theta, l)(2\pi\theta)^{-(n-m)/2}}$$

$$= \frac{\sum_l A(\theta, l)(2\pi\theta)^{-(n-k_l)/2}(2\pi\theta)^{(m-k_l)/2}}{\sum_l \bar{A}(\theta, l))(2\pi\theta)^{-(n-k_l)/2}(2\pi\theta)^{(m-k_l)/2}}$$

$$\to \frac{\sum_{k_l=m}\displaystyle\int_{M_l} f(0, u)g(0, u)\left(\det\frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} d\mathscr{M}_l}{\sum_{k_l=m}\displaystyle\int_{M_l} f(0, u)\left(\det\frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} d\mathscr{M}_l}, \tag{3.5}$$

by using (3.3), (3.4) and $(2\pi\theta)^{m-k_l} \to 0$ if $k_l \ne m$. Let $\mathscr{M} = \sum_{k_l=m} \mathscr{M}_l$, i.e., $\mathscr{M}(S) = \sum_{k_l=m} \mathscr{M}_l(S \cap M_l)$ where S is a Borel set in N then (3.5) becomes

$$\frac{\displaystyle\int_N f(0, u)g(0, u)\left(\det\frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} d\mathscr{M}}{\displaystyle\int_N f(0, u)\left(\det\frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} d\mathscr{M}}. \tag{3.6}$$

We can regard $\mathscr{M}$ as a measure on $(R^n, \mathscr{B})$ by considering $\mathscr{M}(B) = \mathscr{M}(B \cap N)$. If we define

$$k(z) = \left(\det \frac{\partial^2 H}{\partial t^2}(0, u)\right)^{-1/2} \quad \text{if} \quad z = u \in N,$$
$$= 0 \qquad \text{otherwise,}$$

then (3.6) becomes

$$\int_{R^n} g(z) \left[\frac{f(z)k(z)}{\int_{R^n} f(z)k(z)\, d\mathscr{M}} \right] d\mathscr{M}.$$

If P is defined by

$$\frac{dP}{d\mathscr{M}}(z) = \frac{f(z)k(z)}{\int_{R^n} f(z)k(z)\, d\mathscr{M}}$$

then $P_\theta \to P$ weakly. Clearly P concentrates on N. And there is no ambiguity in writing

$$\frac{dP}{d\mathscr{M}}(u) = \frac{f(u)\left(\det \frac{\partial^2 H}{\partial t^2}(u)\right)^{-1/2}}{\int_N f(u)\left(\det \frac{\partial^2 H}{\partial t^2}(u)\right)^{-1/2} d\mathscr{M}} \quad \text{for } u \in N.$$

The assertion is therefore proved. □

REMARKS.

1. Propositions 2.1 to 2.3 are true for arbitrary complete separable metric space.
2. There are some interesting results when Q is Gaussian in some Hilbert space and H is a quadratic form which usually violates the assumption A4 (Hwang [4]).

Acknowledgment. This is a part of my thesis. I would like to express my deepest gratitude to my advisor Professor Ulf Grenander for his guidance and encouragement.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley & Sons, New York.
[2] BOOTHBY, W. (1975). *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press.
[3] GRENANDER, U. (1976). *Pattern Synthesis, Lectures in Pattern Theory Vol. I.* Springer-Verlag.
[4] HWANG, C. R. (1978). Frozen patterns and minimal energy states. Thesis, Div. of Appl. Math., Brown Univ.
[5] KHINCHIN, A. I. (1957). *Mathematical Foundations of Statistical Mechanics.* Dover, New York.
[6] MILNOR J. W. and STASHEFF, J. D. (1974). *Characteristic Classes.* Princeton Univ. Press.
[7] WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–474.

INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI, TAIWAN
REPUBLIC OF CHINA