Probab. Th. Rel. Fields 75, 149–157 (1987)

Probability Theory and Related Fields

A Generalization of Chernoff Inequality Via Stochastic Analysis*

Chii-Ruey Hwang and Shuenn-Jyi Sheu

Institute of Mathematics, Academia Sinica, Nankang, Taipei, Taiwan

Summary. Let μ be a probability measure on R^d with density $c(\exp(-2U(x))$, where $U \in C^2(R^d)$, $|\nabla U(x)|^2 - \Delta U(x) \to \infty$ and $U(x) \to \infty$ as $|x| \to \infty$. By using stochastic analysis and theorems in Schrödinger operators we have the following result: there exists a constant $c > 0$ such that

$$\operatorname{Var}_\mu f \leqq c\, E_\mu |\nabla f|^2 \tag{1}$$

for any $f \in L^1(\mu)$ with a well-defined distributional gradient ∇f. Under our condition, the operator $-\frac{1}{2}\Delta + \nabla U \cdot \nabla$ in $L^2(\mu)$ has discrete spectrum $0 = \lambda_1 < \lambda_2 = \ldots = \lambda_m < \lambda_{m+1} \leqq \ldots$ with corresponding eigenfunctions $\{\phi_n\}$ which form a C.O.N.S. (complete orthonormal system). If the R.H.S. of (1) is finite then equality holds iff $f = \sum_{i=1}^{m} b_i \phi_i$ for some $b_1, \ldots, b_m \in R$. Moreover, the constant c can be taken as $(2\lambda_2)^{-1}$.

When U is a quadratic form, (1) is the Chernoff inequality (Chernoff 1981; Chen 1982). The approach here can be generalized to infinite dimensional Gaussian measures, or the case with μ being a probability measure in a bounded domain of R^d or some discrete cases.

1. Introduction

Chernoff [4] proved that for a standard normal r.v. X the following inequality holds

$$\operatorname{Var} g(X) \leqq E[(g'(X))^2]$$

for all absolutely continuous g with finite $Eg^2(X)$. And the equality holds iff $g(x) = ax + b$ for some constants a, b. The proof involves expanding $g(x)$ in Hermite polynomials.

* This research was partially supported by the National Science Council of the Republic of China

Chen [3] used Schwartz inequality and martingale property to generalize Chernoff's result to multivariate normal case without the finite variance assumption.

A lemma in Geman and Hwang 1984 [6] proved a similar inequality

$$\operatorname{Var} g(X) \leqq M_U E|\nabla g(X)|^2$$

for X with the Gibbs density $c\exp(-2U(x))$ in a hypercube.

Cacoullos [2] studied the upper and lower bounds for normal, exponential, Poisson and binomial distributions. Due to the "boundary" condition in the last three distributions, extra terms are needed for the upper bounds. In the discrete case difference operator is used instead of the gradient.

Although the approach discussed in this article can be generalized to suitable discrete cases as well as continuous case in bounded domain, still we only consider the following set-up.

Motivated by the fact that the normal density is the stationary solution for the forward equation of Ornstein-Uhlenbeck process (also is the density of the invariant measure), the Gibbs probability measure μ in R^d can be consider as stationary measure of the diffusion process $x(\cdot)$:

$$dx(t) = -\nabla U(x(t))\,dt + dw(t),$$

where $w(\cdot)$ is a d-dimensional Brownian motion. This gives the connection of our problem to stochastic calculus. μ is given by

$$\frac{d\mu}{dx} = p(x) \equiv c\exp(-2U(x)) \tag{1.1}$$

with c a norming factor. In this paper, we will assume

(A) $U \in C^2(R^d)$, $|\nabla U(x)|^2 - \Delta U(x) \to \infty$ and $U(x) \to \infty$ as $|x| \to \infty$.

For a locally dx-integrable real valued function f, let ∇f denote its distributional gradient. Let $L = \frac{1}{2}\Delta - \nabla U \cdot \nabla$ and L^* the adjoint of L. Clearly $L^* p = 0$, i.e. $p(\cdot)$ is a stationary solution for the Brownian motion with drift $-\nabla U$. (L is closely related to the Schrödinger operator $\frac{1}{2}\Delta - \frac{1}{2}(|\nabla U(x)|^2 - \Delta U(x))$. The following are our main results.

Theorem 1. *If* (A) *holds, then there exists a constant c such that*

$$\operatorname{Var}_\mu f \leqq c\, E_\mu |\nabla f|^2 \tag{1.2}$$

for any μ-integrable f for which ∇f exists. In fact, $-L$ has discrete spectrum $0 = \lambda_1 < \lambda_2 \leqq \lambda_3 \leqq \dots$ in $L^2(\mu)$ and c can be taken as $(2\lambda_2)^{-1}$. Moreover the equality holds iff f can be written as $\sum_{n=1}^{m} a_n \phi_n$, where ϕ_n is normalized eigenfunction of $-L$ corresponding to λ_n and m is the number such that $\lambda_2 = \dots = \lambda_m \neq \lambda_{m+1}$. Note that ϕ_1 is a constant function.

In particular, when $U(x) = \frac{1}{4}|x|^2$, we obtain the following result due to Chernoff [4] and Chen [3].

Corollary 2. *Let $X_1, \ldots, X_d$ be i.i.d. $N(0,1)$. Then*

$$\operatorname{Var} f(X_1, \ldots, X_d) \leqq E|\nabla f(X_1, \ldots, X_d)|^2,$$

the equality holds iff $f(x) = b_1 x_1 + \ldots + b_d x_d + a$ for some $b_1, \ldots, b_d$, $a \in R$.

In Sect. 2 we shall relate inequality (2) to the operator L and establish (2) for "nice" function by a Hilbert space argument.

In Sect. 3, certain properties of the process generated by L are used to show that (2) holds for general f.

The following notations will also be used: $C^m(R^d)$ (resp. $C^\infty(R^d)$) = set of real valued functions on R^d with continuous derivatives up to the m-th (resp. any) order.

$$C_0^\infty(R^d) = \{f \in C^\infty(R^d);\ f \text{ has compact support}\}.$$

∇f, Δf mean the gradient and Laplacian of f respectively.

$$L_0^2 = \{h\colon R^d \to R^d;\ |h|^2 \in L^2(\mu)\}.$$

Remarks. (i) Assumption (A) guarantees that $-L$ has nonnegative discrete spectrum and there is no explosion for the Brownian motion with drift $-\nabla U$.

(ii) We may consider the infinite dimensional Ornstein-Uhlenbeck operator L defined on $L^2(P^w)$, where $(W_0^r, B(W_0^r), P^w)$ is the r-dimensional Wiener space (see Ikeda and Watanabe [7], Chap. V, §7 for the notations). Then we have a similar inequality

$$\int f^2 \, dP^w \leqq \int \langle Df, Df \rangle_H \, dP^w,$$

if $\int f \, dP^w = 0$ and f is in the domain of L. The equality holds iff $f = \int_0^1 \alpha(s) \cdot dw(s)$, where $\alpha(s) = (\alpha_1(s), \ldots, \alpha_d(s))$, $\int_0^1 |\alpha(s)|^2 ds < \infty$ and $\int_0^1 \alpha(s) \cdot dw(s)$ is Ito's stochastic integral.

(iii) If (A) is not true, in general we don't have the inequality (2).

(iv) Borovkov and Utev 1983 [14] obtained some very interesting results by studying the quantity

$$R_\xi \equiv \sup_{g \in H_1(\xi)} \frac{\operatorname{Var} g(\xi)}{E(g'(\xi))^2},$$

where $H_1(\xi) = \{g \mid g\colon \mathbf{R} \to R,\ g \text{ is absolutely continuous},\ Dg(\xi) > 0,\ Eg^2(\xi) < \infty\}$ and ξ is a r.v. For examples: $R_\xi / \operatorname{Var} \xi = 1$ iff ξ is normal. If $R_\xi < \infty$ and $E(g'(\xi))^2 < \infty$ then $g \in L_2(\xi)$.

2. Preliminary Results

In the rest of the paper, for brevity, we shall omit the constant c in (1). We first show a relation which is important in our analysis.

Let $f \in C_0^\infty(R^d)$. Integration by parts gives

$$\begin{aligned} \int |\nabla f|^2 e^{-2U} dx &= \int f \cdot (-2Lf) e^{-2U} dx \\ &\equiv \langle f, -2Lf \rangle, \end{aligned} \tag{2.1}$$

where

$$Lf = \tfrac{1}{2}\Delta f - \nabla U \cdot \nabla f$$

$\langle\ ,\ \rangle$ is the inner product in $L^2(\mu)$.

Operator L has the following properties.

Lemma 2.1. *$-L$ is essentially self-adjoint*[1]*, nonnegative definite on $C_0^\infty(R^d)$. And for $f \in D(L)$ = the domain of L in $L^2(\mu)$, we have $|\nabla f| \in L^2(\mu)$, moreover* (2.1) *holds.*

Proof. Let us consider the Schrödinger operator $\hat{L}$ defined on $L^2(dx)$[2] by

$$\hat{L}g = \tfrac{1}{2}\Delta g - Vg, \tag{2.2}$$

where

$$V = \tfrac{1}{2}|\nabla U|^2 - \tfrac{1}{2}\Delta U.$$

The operator $\hat{L}$ relates to L as shown by the following diagram:

$$\begin{array}{ccc} L^2(\mu) & \longleftrightarrow & L^2(dx) \\ f & \longleftrightarrow & fe^{-U} \\ {\scriptstyle L}\downarrow & & \downarrow{\scriptstyle \hat{L}} \\ Lf & \longleftrightarrow & (Lf)\cdot e^{-U} \end{array} \tag{2.3}$$

Since $\hat{L}$ is essentially self-adjoint on $C_0^\infty(R^d)$ by our condition on V (see [10], vol. II, Thm. X 29), so is L. This means that for $f \in D(L)$, there are $f_n \in C_0^\infty(R^d)$ so that

$$\begin{array}{l} f_n \to f \\ Lf_n \to Lf \end{array} \quad \text{in } L^2(\mu).$$

Then it is easy to see that (2.1) holds and $|\nabla f| \in L^2(\mu)$.

The following result is a weaker version of Theorem 1.

Theorem 2.2. *$-L$ has discrete spectrum $0 = \lambda_1 < \lambda_2 \leqq \lambda_3 \leqq \ldots \to \infty$ in $L^2(\mu)$. The corresponding normalized eigenfunctions $\{\phi_n\}$ form a C.O.N.S. Moreover, for $f \in D(L)$, and in particular for $f \in C_0^\infty(R^d)$, we have*

$$\int f^2 e^{-2U} dx \leqq c \int |\nabla f|^2 e^{-2U} dx \tag{2.4}$$

if

$$\int f e^{-2U} dx = 0. \tag{2.5}$$

The constant c can be taken to be $(2\lambda_2)^{-1}$. Equality holds in (2.4) *iff for some $a_2, \ldots, a_m \in R$*

$$f = \sum_{i=2}^{m} a_i \phi_i,$$

where m is the integer such that $\lambda_2 = \lambda_3 = \ldots = \lambda_m < \lambda_{m+1}$.

---

1 [10], vol. I

2 Although $\hat{L}$ (or L) is only densely defined on $L^2(dx)$ (or $L^2(\mu)$), for convenience we shall abuse the expression

Proof. $-\hat{L}$ is self-adjoint and nonnegative definite in $L^2(dx)$. By assumption (A), $V(x)\to\infty$ as $|x|\to\infty$, and Theorem XIII.67 on p. 249 in Reed and Simon [10] shows that $-\hat{L}$ has nondegenerate ground state and discrete spectrum and the eigenfunctions span $L^2(dx)$. From the diagram (2.3), the same holds for $-L$ in $L^2(\mu)$. Let $0=\lambda_1<\lambda_2\leqq\lambda_3\leqq\dots$ denote all the eigenvalues, counting the multiplicity, and $\{\phi_n\}$ the corresponding normalized eigenfunctions which form a C.O.N.S. in $L^2(\mu)$. Note that ϕ_1 is a constant function.

For $f\in D(L)$, we have the expansion

$$f=\sum_{n=1}^{\infty} a_n\phi_n.$$

Assuming (2.5) we have $a_1=0$. (2.1) gives

$$\begin{aligned}\int|\nabla f|^2 e^{-2U}dx&=2\langle f,-Lf\rangle\\&=2\left\langle\sum_{n=2}^{\infty}a_n\phi_n,\sum_{n=2}^{\infty}\lambda_n a_n\phi_n\right\rangle\\&=2\sum_{n=2}^{\infty}\lambda_n a_n^2\\&\geqq 2\lambda_2\sum_{n=2}^{\infty}a_n^2\\&=2\lambda_2\int f^2e^{-2U}dx.\end{aligned}$$

This proves (2.4). The last assertion follows immediately.

3. Proofs of the Main Results

Compare Theorem 1 with Theorem 2.2, it remains to show that a function f satisfying $f\in L^1(\mu)$, $|\nabla f|\in L^2(\mu)$ can be approximated by functions in $D(L)$. We state this as a proposition.

Proposition 3.1. *Let $f\in L^1(\mu)$ and $|\nabla f|\in L^2(\mu)$. Then $f\in L^2(\mu)$. Furthermore, if*

$$f=\sum_{n=1}^{\infty} a_n\phi_n,$$

then

$$\nabla f=\sum_{n=1}^{\infty} a_n\nabla\phi_n \quad \text{in } L_0^2.$$

We also need some simple properties of $\{\phi_n\}$.

Lemma 3.2.

$$\int\nabla\phi_n\cdot\nabla\phi_k e^{-2U}dx=2\lambda_k\delta_{n,k},$$

where

$$\delta_{n,k}=\begin{cases}1 & \text{if } n=k,\\ 0 & \text{if } n\neq k.\end{cases}$$

Proof. By the argument as in the proof of Lemma 2.1, we have

$$\begin{aligned}\int \nabla\phi_n \cdot \nabla\phi_k e^{-2U} dx &= 2\int \phi_n \cdot (-L\phi_k) e^{-2U} dx \\ &= 2\lambda_k \int \phi_n \cdot \phi_k e^{-2U} dx \\ &= 2\lambda_k \delta_{n,k}.\end{aligned}$$

We now prove Theorem 1 and Corollary 2 by applying Proposition 3.1.

Proof of Theorem 1. We only need to show that $f \in L^1(\mu)$, $\nabla f \in L^2(\mu)$ and $\int f e^{-2U} dx = 0$ imply

$$\int f^2 e^{-2U} dx \leqq c \int |\nabla f|^2 e^{-2U} dx.$$

By Proposition 3.1, $f \in L^2(\mu)$ and

$$\nabla f = \sum_{n=2}^{\infty} a_n \nabla\phi_n$$

if $f = \sum_{n=2}^{\infty} a_n \phi_n$.

$$\begin{aligned}\int |\nabla f|^2 e^{-2U} dx &= \sum a_n a_k \int \nabla\phi_n \cdot \nabla\phi_k e^{-2U} dx \\ &= 2\sum_{n=2}^{\infty} \lambda_n a_n^2 \\ &\geqq 2\lambda_2 \sum_{n=2}^{\infty} a_n^2 \\ &= 2\lambda_2 \int f^2 e^{-2U} dx.\end{aligned}$$

This completes the proof of the theorem.

Proof of Corollary 2. Take $U(x) = \frac{1}{4}|x|^2$. By Theorem 1, it remains to show that $\lambda_2 = \lambda_3 = \ldots = \lambda_{d+1} = \frac{1}{2}$ and $\phi_2(x) = c_2 x_1, \ldots, \phi_{d+1}(x) = c_{d+1} x_d$. In fact, the generalized Hermite polynomials $H_{p_1, \ldots, p_d}(x)$ form a complete set of eigenfunctions for $-L = -\frac{1}{2}\Delta + \frac{1}{2} x \cdot \nabla$, where $H_{p_1, \ldots, p_d}(x)$ are defined by the relation

$$e^{a \cdot x - \frac{1}{2}|a|^2} = \sum a_1^{p_1} \ldots a_d^{p_d} H_{p_1, \ldots, p_d}(x)/p_1! \ldots p_d!$$
$$a = (a_1, \ldots, a_d) \in R^d.$$

Moreover,

$$L(H_{p_1, \ldots, p_d}) = -\frac{p_1 + \ldots + p_d}{2} H_{p_1, \ldots, p_d}.$$

Therefore

$$\lambda_2 = \ldots = \lambda_{d+1} = \tfrac{1}{2}$$
$$\phi_{i+1}(x) = c_{i+1} x_i, \qquad 1 \leqq i \leqq d.$$

We now prove Proposition 3.1 in two steps.

Lemma 3.3. *Assume $f \in L^1(\mu)$ and $|\nabla f| \in L^2(\mu)$. Then*

$$f = \sum_{n=2}^{\infty} b_n \phi_n + g$$

with

$$g \in L^1(\mu), \quad L g = 0, \quad \nabla g \in L^2(\mu),$$

$$b_n = \frac{1}{2\lambda_2} \int \nabla f \cdot \nabla \phi_n e^{-2U} dx.$$

Lemma 3.4. *If* $g \in L^1(\mu)$, $|\nabla g| \in L^2(\mu)$ *and* $L g = 0$, *then* g *is a constant.*

Proof of Lemma 3.3. $\{\nabla \phi_n / \sqrt{2\lambda_n}\}$ is an orthonormal family in L_0^2. Let

$$c_n = \int \nabla f \cdot \nabla \phi_n / \sqrt{2\lambda_n}\, e^{-2U} dx.$$

Then

$$\sum |c_n|^2 \leqq \int |\nabla f|^2 e^{-2U} dx < \infty.$$

Let

$$f_0 = \sum_{n=2}^{\infty} b_n \phi_n$$

with

$$b_n = \frac{1}{\sqrt{2\lambda_2}} c_n.$$

It is easy to see by Lemma 3.2 that $\sum b_n \nabla \phi_n$ converges in L_0^2. Therefore

$$\nabla f_0 = \sum b_n \nabla \phi_n$$

is in L_0^2 and

$$\int \nabla (f - f_0) \cdot \nabla \phi_n e^{-2U} dx = 0.$$

Now for $\phi \in C_0^\infty(R^d)$, since $\nabla \phi = \sum a_n \nabla \phi_n$,

$$\begin{aligned} &\int \nabla (f - f_0) \cdot \nabla \phi\, e^{-2U} dx = 0 \\ &= \int (f - f_0) \cdot (-2L\phi) e^{-2U} dx. \end{aligned}$$

This implies that $g = f - f_0$ satisfies the properties stated in Lemma 3.3. Notice that $L g = 0$ in distributional sense as well as in classical sense by regularity results for the solution of elliptic equation. (See [13], part II, Chap. 1.)

Proof of Lemma 3.4. Recall that μ is the probability in R^d with density

$$p(x) = c\, e^{-2U(x)}.$$

We consider the diffusion with drift $-\nabla U$ and the corresponding stochastic differential equation

$$dx(t) = -\nabla U(x(t))\, dt + dw(t),$$

where $w(t)$ is a d-dimensional Brownian motion. Since

$$\begin{aligned} L e^{U(x)} &= \tfrac{1}{2} \Delta e^{U(x)} - \nabla U(x) \cdot \nabla e^{U(x)} \\ &= \tfrac{1}{2} (\Delta U(x) - |\nabla U(x)|^2)\, e^{U(x)} \\ &= -V(x)\, e^{U(x)}, \end{aligned}$$

by assumption (A), there exists $c > 0$ such that

$$L e^{U(x)} \leqq c\, e^{U(x)}, \quad x \in R^d.$$

Then, by a theorem on p. 191 of Varadhan [12], there is no explosion. Let $\{P_x\}$ denote the family of Markov processes on $\Omega = C\{[0,\infty), R^d\}$ indexed by the starting point x with the generator L, then μ is the invariant measure of $\{P_x\}$ ([12], p. 243).

Let g satisfy the condition of the lemma, $g \in C^2(R^d)$ (see [13], part II, Chap. 1, § 1.3). Ito's formula gives

$$g(x(t)) - g(x(0)) = \int_0^t \nabla g(x(s)) \cdot dw(s). \tag{3.1}$$

Since

$$\begin{aligned} E_\mu E_x \int_0^t |\nabla g(x(s))|^2\, ds &= E^{P_\mu} \int_0^t E^{P_\mu}[|\nabla g(x(s))|^2 \,|\, x(0)] \\ &= \int_0^t E^{P_\mu}[|\nabla g(x(s))|^2] \\ &= t\, E_\mu[|\nabla g|^2] < \infty, \end{aligned}$$

where P_μ is the probability distribution of $x(\cdot)$ with initial distribution μ, there exists $M \subset R^d$ with $\mu(M) = 1$ such that whenever $x \in M$

$$E_x \int_0^t |\nabla g(x(s))|^2\, ds < \infty \qquad \forall t > 0. \tag{3.2}$$

Now fix $x \in M$, (3.1), (3.2) imply $g(x(\cdot))$ is a square integrable martingale and

$$E_x[g(x(s))] = g(x).$$

Then

$$E_x[|g(x(s))|] \geqq |g(x)| \qquad \forall x \in M. \tag{3.3}$$

Integrate both sides of (3.3) with respect to μ:

$$E_\mu[|g|] = E_\mu E_x[|g(x(t))|] \geqq E_\mu[|g|].$$

We must have

$$E_x[|g(x(t))|] = |g(x)|, \qquad \forall x \in M,\ \forall t > 0. \tag{3.4}$$

Let $h_r(x) = |g(x)| \wedge r$ for $r > 0$. By a theorem on p. 251 of [12],

$$\begin{aligned} |g(x)| &= E_x[|g(x(t))|] \\ &\geqq E_x[h_r(x(t))] \to E_\mu h_r \qquad \text{as } t \to \infty. \end{aligned}$$

Hence $|g(x)| \geqq E_\mu[|g|]$. Proceed as the proof of (3.4), and we have

$$|g(x)| = E_\mu[|g|], \qquad \forall x \in M. \tag{3.5}$$

Then (3.5) holds for all x in R^d by the continuity of $g(\cdot)$, therefore, g is a constant.

Addendum. After this paper was typed, the authors learned that Chen [15] had made some interesting generalization in another direction to a related inequality involving both the upper and lower bounds for infinitely divisible distributions in $\mathbf{R}^d$. In view of Theorem 2 in [14], it is clear that an extra term has to be added to the R.H.S. of (2) in this case.

References

1. Adams, R.A.: Sobolev spaces. New York: Academic Press 1975
2. Cacoullos, T.: On upper and lower bounds for variance of a function of a random variable. Ann. Probab. **10**, 799–809 (1982)
3. Chen, L.H.: An inequality for the multivariate normal distribution. J. Multivariate Anal. **12**, 306–315 (1982)
4. Chernoff, H.: A note on an inequality involving the normal distribution. Ann. Probab. **9**, 533–535 (1981)
5. Dynkin, E.B.: Markov processes, vol. II. Berlin, Heidelberg, New York: Springer 1965
6. Geman, S., Hwang, C.-R.: Stochastic optimization by simulated annealing. In preparation (1984)
7. Ikeda, N., Watanabe, S.: Stochastic differential equations and diffusion processes. Amsterdam: North Holland 1981
8. Kallianpur, G.: Stochastic filtering theory. Berlin, Heidelberg, New York: Springer 1980
9. Meyer, P.A.: Note sur les processus d'Ornstein-Uhlenbeck. Lect. Notes Math. **920**. Berlin, Heidelberg, New York: Springer 1981
10. Reed, R., Simon, B.: Method of modern mathematical physics, vols. I, II, IV. New York: Academic Press 1978
11. Stroock, D.W.: The Malliavin calculus, a functional analytic approach. J. Funct. Anal. **44**, 212–257 (1981)
12. Varadhan, S.R.S.: Lectures on diffusion problems and partial differential equations. Tata Institute (1980). Berlin, Heidelberg, New York: Springer 1980
13. Bers, L., John, F., Schechter, M.: Partial differential equation. New York: Interscience Publisher (1957)
14. Borovkov, A.A., Utev, S.A.: On an inequality and a related characterization of the normal distribution. Theor. Probab. Appl. **28**, 219–228 (1983)
15. Chen, L.H.Y.: Poincaré-type inequalities via stochastic integral. Probab. Th. Rel. Fields, submitted (1984)

Received March 12, 1984