Ergodic Theorems: Suppose

$$\mathcal{L} = \frac{1}{2}\sum_{i,j} a_{i,j}(x)D_{i,j} + \sum_j b_j(x)D_j$$

is the generator of a process. Assume that $\{a_{i,j}\}$ are continuous and strictly positive definite, i.e. $\{a_{i,j}\}$ is positive definite for each $x \in R^d$. An invariant probability measure is one for which

$$\int p(t,x,A)\mu(dx) = \mu(A)$$

for every Borel set $A \subset R^d$. First μ must be absolutely continuous, i.e. $\mu(dy) = \phi(y)dy$ for some $\phi \in L_1(R^d)$. The proof depends on the fact that $p(t,x,dy)$ in fact has a density $p(t,x,y)dy$ for every $t > 0$, and $x \in R^d$. We would like to prove the following.

1. **Uniqueness.** $\mu(dx) = \phi(x)dx$ is unique. $\phi(x) > 0$ a.e.

2. **Law of Large Numbers.** With probability 1 with respect to any P_x,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(x(s))ds = \int f(y)\phi(y)dy$$

for any bounded measurable function $f(\cdot)$.

3. **Convergence to equilibrium.**

$$\lim_{t\to\infty} \int |p(t,x,y) - \phi(y)|dy = 0$$

4. **Central Limit Theorems.** If $\int f(y)\phi(y)dy = 0$, under additional assumptions

$$\frac{1}{\sqrt{t}}\int_0^t f(x(s))ds$$

has an asymptotic normal distribution for large t.

5. **Large Deviations from the Ergodic Theorem.** Study the large deviation behavior

$$P_x\left[\frac{1}{t}\int_0^t f(x(s))ds \in A\right]$$

Uniqueness: Uniqueness and ergodicity are related. Let μ be an invriant probability measure for the Markov process. This means that if we start with initial distribution μ the $P_\mu = \int P_x d\mu(x)$ is staionary process for $t \geq 0$ and can be easily extended as a stationary process for $-\infty < t < \infty$. If P_μ is not ergodic with respect to trnaslations, then there exists a trnaslation invariant set E on the space $C[(-\infty,\infty);R^d]$ of trajectories such that $0 < P_\mu(A) < 1$. We will show now that E can be chosen to be of the form $x(0) \in A$ for some Borel set. Note that any measurable set E can be approximated by sets measurable with respect to the σ-field $\sigma\{x(t) : |t| \le T\}$ and if it is translation invariant can be approximated by a set from $\sigma\{x(t) : T \le t \le 2T\}$ as well as from $\sigma\{x(t) : -2T \le t \le -T\}$. In particular it is in the past $\sigma\{x(t) : t \le -T\}$ as well as the future $\sigma\{x(t) : t \ge T\}$. But for a Markov process given the present $x(0)$ the past and the future are independent. This means given $x(0)$ the set E is indpendent of itself. Therefore $P[E|x(0)] = 0$ or 1. Now E can be identified with the set $\{x(\cdot) : P[E|x(0)] = 1\}$. If there are two invariant measures μ_1, μ_2, then clearly P_μ with $\mu = \frac{1}{2}[\mu_1 + \mu_2]$ can not be ergodic.

In particular if the invariant measure μ is not unique we can find a partition of $R^d = A \cup A^c$ such that the restrictions of μ to A and A^c are both invariant. This means for every $t > 0$, $p(t, x, A) = 1$ for a.e. $x \in A$ and $p(t, x, A^c) = 1$ for a.e $x \in A^c$.

We note that, from PDE, we can obtain that $p(t, x, A)$ is a jointly continuous function of t, x for $t > 0$. In partcular $y : p(t, y, A) > \frac{1}{2}$ is an open set U, and for almost all $x \in A^c$

$$p(2t, x, A) = \int p(t, x, dy)p(t, y, A) \ge \frac{1}{2}p(t, x, U) = 0$$

Therefore to prove uniqueness it suffices to show that given $x \in R^d$ and any open ball B, $p(t, x, B) > 0$ for some $t > 0, x \in R^d$ any ball B. Suppose $P_x[x(t) \in B] = 0$ for all t. By Girsanov's theorem the same is true if we change the drift, and also if we rescale time. Therefore

$$Q^\epsilon_x[x(t) \in B] = 0$$

where Q^ϵ_x corresponds to

$$\mathcal{L}_\epsilon = \frac{\epsilon}{2}\sum_{i,j} a_{i,j}(x)D_{i,j} + \sum_j [\epsilon b_j(x) + c_j(x)]D_j$$

with $\{c_j(x)\}$ having compact support. If we let $\epsilon \to 0$ the same is true for the limit. But the limit is just an O.D.E and we can pick our coeffiicients!. This can not be true.

This proves the uniqueness of μ and the ergodicity of P_μ. It is not hard to see that from

$$\mu(A) = \int p(t, x, A)d\mu(x)$$

and the existence of a density $p(t, x, y)$ we can conclude that μ in fact has a density $\pi(x)$. The positivity of $\phi(x)$ again come from PDE. If $\int_A \phi(x)dx = 0$ for a set with positive Lebesgue meausre, then $p(t, x, A)$ would be 0 for all x and $t > 0$, and the PDE estimate shows the convergence of $p(t, x, A) \to \chi_A(x)$ in L_p as $t \to 0$. Another way of verifying uniqueness is by a direct argument. Let $\pi(x, y) = p(1, x, y)$. If μ is invariant then, it is invariant for $\pi(x, y)$ as well. Hence

$$\int \pi(x, y)dy\mu(dx) = \mu(dy)$$

This implies that $\mu(dx) = \phi(x)dx$ and

$$\int \pi(x, y)\phi(x)dx = \phi(y)$$

If there are two solutions let us denote their difference by $\psi(x)$. Then $\int \psi(x)dx = 0$,

$$\int \psi(x)\pi(x,y)dx = \psi(y)$$

and

$$\int |\psi(y)dy| = \int |\int \psi(x)\pi(x,y)dx|dy \leq \int\int |\psi(x)|\pi(x,y)dxdy = \int |\psi(x)|dx$$

In particular for almost all y,

$$|\int \psi(x)\pi(x,y)dx| = \int |\psi(x)|\pi(x,y)dx$$

This can not happen unless $\psi(x)$ is of one sign. But it has integral equal to 0. Hence $\psi = 0$.

Ergodic Theorem. Because of ergodicity we know that with probability 1 with respect to any P_μ,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(x(s))ds = \int f(y)\phi(y)dy$$

for any bounded measurable function $f(\cdot)$. On the other hand form the positivity of $p(t,x,y)$, for $t > 0$, $p(t,x,y)$ and μ are mutually absolutely continuous. By the Markov property if we ignore a small initial segment of time then the almost sure convergence w.r.t. P_μ gives a.e. convergence w.r.t P_x for any x.

Convergence to equilibrium. This is not much different from uniqueness. Let us denote by

$$\delta(t,x) = \int |p(t,x,y) - \phi(y)|dy$$

then

$$\begin{aligned}\delta(t+s,x) &= \int |\int [p(t,x,z') - \phi(z')]p(s,z',z)dz'|dz \\ &\leq \int |p(t,x,z') - \phi(z')|p(s,z',z)dzdz' \\ &= \delta(t,x,y)\end{aligned}$$

proving $\delta(t,x)$ is $\downarrow$ in t for fixed x. It is therefore sufficient to prove that $\delta(n,x) \to 0$ as $n \to \infty$. We have a Markov Chain with almost surely positive transition density $\pi(x,y) = p(1,x,y)$ satisfying, $x \to \pi(x,\cdot)$ is continuous as a map into $L_1(R^d)$ that admits an invariant probability density $\phi(x)$. We need to prove that for compact K,

$$\lim_{n\to\infty} \sup_{x\in K} \int |\pi^n(x,y) - \phi(y)|dy$$

The basic idea here is coupling. Let us take two copies with transition probability

$$\begin{aligned}\widehat{\pi}(x_1, x_2; dy_1, dy_2) = &\int \min\{\pi(x_1, y)\pi(x_2, y)\}\delta(y - y_1) \times \delta(y - y_2)dy \\ &+ \int [\pi(x_1, y) - \min\{\pi(x_1, y)\pi(x_2, y)\}]\delta(y - y_1)dy \\ &\times \int [\pi(x_2, y) \min\{\pi(x_1, y)\pi(x_2, y)\}]\delta(y - y_2)dy\end{aligned}$$

In words, if $g(y) = \min\min\{\pi(x_1, y)\pi(x_2, y)\}$ the two components jump on to the diagonal $y = y_1 = y_2$ with density $g(y)dy$ and jump indpendently with the remaining probability. Once they are on the diagonal they stay on the diagonal. It is not hard to see that each component evolves according to π. Therefore

$$\int |\pi^n(x_1, y) - \pi^n(x_2, y)|dy \leq \widehat{\pi}^n(x_1, x_2; \{y_1 \neq y_2\}) = \delta_n(x_1, x_2)$$

and if $r_n(x_1, x_2, dy_1, dy_2)$ is the part of $\widehat{\pi}^n$ away from the diagonal

$$\delta_{n+1}(x_1, x_2) = \delta_n(x_1, x_2) - 2\int \min\{\pi(y_1, y), \pi(y_2, y)\}r_n(x_1, x_2, dy_1, dy_2)dy$$

Since $\delta_n \geq 0$ we must have

$$\int \min\{\pi(y_1, y), \pi(y_2, y)\}r_n(x_1, x_2, dy_1, dy_2)dy \to 0$$

as $n \to \infty$. Since $\pi(x, y) > 0$ a.e, this is possible only if

$$\int_{\substack{|y_1| \leq \ell \\ |y_2| \leq \ell}} r_n(x_1, x_2, dy_1, dy_2) \to 0$$

On the other hand if the total mass of r_n does not go to 0, we must have

$$\lim_{\ell \to \infty} \liminf_{n \to \infty} r_n(x_1, x_2, \{|y_1| \geq \ell \cup |y_2| \geq \ell\}) \geq c > 0$$

Since

$$r_n(x_1, x_2, \{|y_1| \geq \ell \cup |y_2| \geq \ell\}) \leq \pi^n(x_1, |y_1| \geq \ell) + \pi^n(x_2, |y_2| \geq \ell)$$

this contradicts the ergodic theorem

$$\frac{1}{n}\sum_{j=1}^{n} \pi^j(x, A) \to \mu(A)$$

Central Limit Theorem. Consider

$$A(t) = \int_0^t f(x(s)ds$$

If $\int f(x)d\mu(x) = 0$, one can expect to prove the C.L.T that states

$$\frac{A(t)}{\sqrt{t}} \simeq N(0, \sigma^2)$$

The proof depends on ability to solve

$$\mathcal{L}u = -f$$

with a nice u. This is the Fredholm alternative. Then by Itô's formula

$$u(x(t)) - u(x) = \int_0^t \sigma(x(s))(\nabla u)(x(s))d\beta(s) - \int_0^t f(x(s))ds$$

Therefore

$$A(t) = \int_0^t \sigma(x(s))(\nabla u)(x(s))d\beta(s) - u(x(t)) - u(x)$$

Dividing by $\sqrt{t}$ if we assume that $\frac{u(x(t))}{\sqrt{t}} \to 0$ as $t \to \infty$, we need only analyse

$$\frac{1}{\sqrt{t}} \int_0^t \sigma(x(s))(\nabla u)(x(s))d\beta(s) = \widehat{\beta}(\frac{1}{t} \int_0^t < a(x(s))\nabla u(x(s)), \nabla u(x(s)) > ds)$$

and by the ergodic theorem this converges to $\widehat{\beta}(\sigma^2) \simeq N(0, \sigma^2)$ with

$$\sigma^2 = \int < a(x)(\nabla u)(x), (\nabla u)(x) > d\mu(x)$$

One can work with approximate solutions of $\mathcal{L}u = -f$ namely

$$\lambda u_\lambda - \mathcal{L}u_\lambda = f$$

and reduce the CLT to the behavior of u_λ as $\lambda \to 0$. One needs

$$\lambda \int \|u_\lambda\|^2 d\mu \to 0$$

and

$$\int < a(x)(\nabla u_\lambda(x) - g(x)), (\nabla u_\lambda(x) - g(x)) > d\mu_(x) \to 0$$

for some g. One then applies Itô's formula to $u_\lambda(x)$ and chooses $\lambda = t^{-1}$. The errors are easily controlled.

To see this we need to control

$$X_\lambda(t) = \frac{1}{\sqrt{t}} \int_0^t \lambda u_\lambda(x(s))ds$$

Starting from the equlibrium μ

$$\|X_\lambda(t)\|_2 \le \lambda\sqrt{t}\|u_\lambda\|_2 \le \sqrt{\lambda}\|u_\lambda\|_2 \to 0$$

As for the stochastic integral part we have the CLT valid for

$$\frac{M_\lambda(t)}{\sqrt{t}} = \frac{1}{\sqrt{t}}\int_0^t \sigma(x(s))(\nabla u_\lambda)(x(s))d\beta(s)$$

The difference between between $M_{\lambda_1}(t)$ and $M_{\lambda_2}(t)$ is easily estimated by

$$\begin{aligned}&\frac{1}{t}E\Big[[M_{\lambda_1}(t) - M_{\lambda_2}(t)]^2\Big] \\ &= \int < a(x)(\nabla u_{\lambda_1}(x) - \nabla u_{\lambda_2}(x)), (\nabla u_{\lambda_1}(x) - \nabla u_{\lambda_2}(x)) > d\mu_(x) \to 0\end{aligned}$$

A special situation is when $\mathcal{L}$ is self adjoint with respect to μ. Then one defines the Dirichlet norm by

$$\|u\|_1^2 = \int < a(x)(\nabla u)(x), (\nabla u)(x) > d\mu(x)$$

and the dual norm

$$\|f\|_{-1} = \sup_u \frac{\int f(x)u(x)d\mu}{\|u\|_1}$$

The finiteness of $\|f\|_{-1}$ implies both the conditions needed for CLT and this is easily seen by the spectral theorem. If $k(d\sigma)$ is the spectral measure of f with respect to $\mathcal{L}$, then the assumption is

$$\int_{0+}^\infty \frac{1}{\sigma}dk(\sigma) < \infty$$

The conditions to be proved are

$$\lim_{\lambda\to 0}\int_{0+}^\infty \frac{\lambda}{(\lambda+\sigma)^2}k(d\sigma) = 0$$

and

$$\lim_{\substack{\lambda_1\to 0\\ \lambda_2\to 0}}\int_{0+}^\infty |\frac{1}{(\lambda_1+\sigma)} - \frac{1}{(\lambda_2+\sigma)}|^2\sigma dk(\sigma) = 0$$

Both are consequences of the dominated convergence theorem.

Large Deviations. We wish to estimate

$$P_x[\frac{1}{t}\int_0^t f(x(s))ds \ge \ell]$$

where $\ell > \int f(x)d\mu(x)$. Does this decay exponentially? If so what is the rate? We will do one example. Brownian motion on the circle. The operator is $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$ acting on periodic functions of period 1. The invariant measure is just Lebesgue measure. $\phi(x) \equiv 1$.

One way to estimate the rate is to try and compute

$$\lim_{t\to\infty} \frac{1}{t} \log E_x \left[\sigma \int_0^t f(x(s))ds\right] = F(\sigma)$$

Then by Tchebechev's inequality

$$P_x[\frac{1}{t}\int_0^t f(x(s))ds \geq \ell] \leq e^{-\sigma\ell t + tF(\sigma) + o(t)}$$

and

$$\limsup \frac{1}{t} \log P_x[\frac{1}{t}\int_0^t f(x(s))ds \geq \ell] \leq F(\sigma) - \ell\sigma$$

Optimizing over $\sigma > 0$,

$$\limsup \frac{1}{t} \log P_x[\frac{1}{t}\int_0^t f(x(s))ds \geq \ell] \leq -\sup_{\sigma>0}[\ell\sigma - F(\sigma)]$$

It is not hard to see that $F'(0) = \int f(x)dx$, and therefore for $\ell > \int f(x)dx$,

$$\sup_{\sigma>0}[\ell\sigma - F(\sigma)] = \sup_{\sigma}[\ell\sigma - F(\sigma)]$$

The same is true in the other side. The large deviation rate function can be then

$$I(\ell) = \sup_{\sigma}[\ell\sigma - F(\sigma)]$$

What is $F(\sigma)$? We can look at Feynman-Kac formula and

$$u(t,x) = E_x\left[\sigma \int_0^t f(x(s))ds\right]$$

solves

$$u_t = \frac{1}{2}u_{xx} + \sigma f u; u(0,x) = 1$$

If we denote the eigen values and eigen functions of $\frac{1}{2}u_{xx} + \sigma f u$ by $\{\lambda_j, \phi_j\}$, then

$$u(t,x) = \sum_j \phi_j(x) e^{t\lambda_j} \int \phi_j(y)dy$$

The eigenvalues tend to $-\infty$ and the one that counts will be the largest one, the principal eigen value. Note that $\phi_1(y) > 0$. Hence

$$F(\sigma) = \lambda_1(\sigma f) = \sup_{\|g\|_2=1}[\sigma \int f(x)g^2(x)dx - \frac{1}{2}\int |g_x|^2 dx]$$

We can rewrite this as

$$F(\sigma) = \lambda_1(\sigma f) = \sup_{\substack{g\geq 0\\ \int g(x)dx=1}} \left[\sigma \int f(x)g(x)dx - \frac{1}{8}\int \frac{|g_x|^2}{g}dx\right]$$

This has an interpretation. Instead of $\frac{1}{2}\frac{d^2}{dx^2}$ modify with a drift

$$\mathcal{L}_g = \frac{1}{2}\frac{d^2}{dx^2} + \frac{g_x}{2g}\frac{d}{dx}$$

This has invariant density g. Now the erigodic theorem will produce g averages. The process then will behave like a process with invariant density g. Call this process $Q_{g,x}$.

$$P_x[A] = \int_A \frac{dP_x}{dQ_{g,x}} dQ_{g,x}$$

Our set A will have $Q_{g,x}$ measure 1. We need to concentrate on how small the derivative $\frac{dP_x}{dQ_{g,x}}$ can be on A. By Girsanov formula

$$\log \frac{dQ_{g,x}}{dP_x} = \int \frac{g_x}{2g}(x(s))dx(s) - \frac{1}{8}\int \frac{g_x^2(x(s))}{g^2(x(s))}ds$$

The ergodic theorem under $Q_{g,x}$ will produce a limit

$$\frac{1}{t}\log \frac{dQ_{g,x}}{dP_x} \to \frac{1}{4}\int \frac{g_x^2}{g}dx - \frac{1}{8}\int \frac{g_x^2}{g}dx = \frac{1}{8}\int \frac{g_x^2}{g}dx$$

giving with the help of Jensen's inequality a lower bound of

$$\exp\big[-\frac{t}{8}\int \frac{g_x^2}{g}dx\big]$$

for a Brownian Motion to behave like a process with invariant density g for a duration t. This provides for our original estimate

$$I(\ell) = \inf_{\substack{g\geq 0, \int g\,dx=1\\ \int f(x)g(x)dx=\ell}} \big[\frac{1}{8}\int \frac{g_x^2}{g}dx\big]$$