The Bayesian Approach to Inverse Problems

Andrew M. Stuart,

Mathematics Institute
Warwick University
CV4 7AL, UK
e-mail: a.m.stuart@warwick.ac.uk

Abstract: These notes give various references to the literature that I will not, for reasons of brevity, give during the lectures themselves.

1. Gaussian Measures

A comprehensive reference for Gaussian measures is [Bog98]. This book works in the generality of locally convex topological spaces, and has material on Gaussian measure on Banach and Hilbert spaces as particular cases. A similar book, focussed on the setting in Banach spaces, is [Lif95]. For a simpler introduction to the Hilbert space setting see the relevant material in [DZ92].

For a discussion of Radon measures, and statement that a probability measure on a seprable Hilbert space, with the Borel sigma algebra, is Radon, see [Bog98], Theorem A3.11. The statement that any two Radon measures with equal Fourier transform coincide is Theorem A3.18 in [Bog98]. Theorem 1.2 from our notes is Theorem 2.3.1 in [Bog98]. Regularity of functions drawn from Gaussian measures (see our Theorem 1.4) is discussed in some detail in [DZ92].

2. Change of Measure from Gaussian

Inverse problems in differential equations, and their Bayesian formulation, is covered in detail in [Stu10]. The Radon-Nikodym Theorem 2.1 may be found in [Dud02], as can Lemma 2.2, which we state here in complete detail, to enable a briefer (imprecise) statement in the lectures. The particular form of the statement adopted here may be found in [HSV07].

Lemma 2.2 *Let μ, ν be probability measures on $S \times T$ where $(S, \mathbb{A})$ and $(T, \mathbb{B})$ are measurable spaces. Denote by (x, y) with $x \in S$ and $y \in T$ an element of $S \times T$. Assume that $\mu \ll \nu$ and that μ has Radon-Nikodym derivative ϕ with respect to ν. Assume further that the conditional distribution of $x|y$ under ν, denoted by $\nu^y(dx)$, exists. Then the conditional distribution of $x|y$ under μ, denoted $\mu^y(dx)$, exists and $\mu^y \ll \nu^y$. The Radon-Nikodym derivative is given by*

$$\frac{d\mu^y}{d\nu^y}(x) = \begin{cases} \frac{1}{c(y)}\phi(x, y), & \text{if } c(y) > 0, \text{ and} \\ 1 & \text{else} \end{cases} \tag{2.1}$$

with $c(y) = \int_S \phi(x, y)\, d\nu^y(x)$ for all $y \in T$.

3. μ−invariant SDEs

For an introduction to SDEs see [Oks03] for a mathematical viewpoint and [Gar85] for a more applied perspective. For discrete time Markov chains, proof of the exponential convergence rate for expectations, together with the law of large numbers, may be found in [MT93], Chapters 15 and 17 respectively. Use of this theory to study the continuous time processes arising in SDEs may be found in [MSH02], and Theorem 3.2 can be proved using this machinery. However we have motivated the ideas using a Fokker-Planck based approach, reminiscent of the analysis found in [Gar85]. Theorem 3.3 is proved in greater generality in [HSV07]. However ergodicity results for different SPDEs, in simpler settings, may be found in [DZ92].

4. $\mu-$invariant Markov Chains and MCMC

MCMC methods were introduced by solid state physicists in the seminal paper [MRTT53] and subsequently generalized to Metropolis-Hastings methods on $\mathbb{R}^n$ in [Has70]. The books [RC99, Liu01] contain overviews from the perspective of applied statistics. The formulation of Metropolis-Hastings methods on a general state space is given in [Tie98]; this is what we use to study our random walk methods on Hilbert space. Lemma 4.2 is proved in [BRS09] and the improved random walk of section 4.3 is introduced in [BS09, BRS09].

5. MCMC and Diffusion Limits

The idea of deriving diffusion limts for MCMC methods was systematically developed in the papers [RGG97, RR98, RR01]. Those papers concerned target measures with a product structure (independence amongst coordinates) and as a consequence it is possible to prove a diffusion limit for a single component of the vector in the target space. Such product measures are not typical of applications, where correlations are to be expected. Nonetheless the intuition from these papers is very informative for the measures μ of interest to us. This is because $\mu \ll \mu_0$ where the Gaussian measure μ_0 *does* have a product structure in appropriate coordinates. Thus although our limit theorems are to an SDE in the Hilbert space H, and are hence infinite dimensional, many of the basic mechanisms identified in the product case play a key role. In particular the scaling of δ with N to obtain an order one acceptance probability, and the approximate form of the accetpance probability $\alpha = 1 \wedge \exp(Z_\ell)$ with Z_ℓ a Gaussian $N(-\ell^2, 2\ell^2)$.

The idea of using a continous time interpolant of the discrete time Markov chain is developed in the simple context of Euler approximation of SDEs in [Mao97, HMS03]. The invariance principle for the noise process W^N uses a result from the paper [Ber86]. The SPDE limit theorem is built on these ideas and may be found in [MPS11], and on the arxiv at [MPS10]. A related diffusion limit theorem, but for a Langevin based proposal, may be found at [PST11].

Acknowledgements AMS is grateful to EPSRC, ERC and ONR for financial support.

References

[Ber86] E. Berger. Asymptotic behaviour of a class of stochastic approximation procedures. *Probability theory and related fields*, 71(4):517–552, 1986.

[Bog98] V.I. Bogachev. *Gaussian Meausures.* American Mathematical Society, 1998.

[BRS09] A. Beskos, G. O. Roberts, and A. M. Stuart. Optimal scalings for local Metropolis-Hastings chains on non-product targets in high dimensions. *Ann. Appl. Prob.*, 19:863–898, 2009.

[BS09] A. Beskos and A.M. Stuart. MCMC methods for sampling function space. In *Invited Lectures, Sixth International Congress on Industrial and Applied Mathematics, ICIAM07, Editors Rolf Jeltsch and Gerhard Wanner*, pages 337–364. European Mathematical Society, 2009.

[Dud02] R.M. Dudley. *Real Analysis and Probability.* Cambridge University Press, Cambridge, 2002.

[DZ92] G. DaPrato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1992.

[Gar85] C. W. Gardiner. *Handbook of stochastic methods.* Springer-Verlag, Berlin, second edition, 1985. For physics, chemistry and the natural sciences.

[Has70] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[HMS03] D.J. Higham, X. Mao, and A.M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis*, 40(3):1041–1063, 2003.

[HSV07] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling, part II: The nonlinear case. *Annals of Applied Probability*, 17:1657–1706, 2007.

[Lif95] M.A. Lifshits. *Gaussian Random Functions*, volume 322 of *Mathematics and its Applications.* Kluwer, Dordrecht, 1995.

[Liu01] Jun S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, 2001.

[Mao97] Xuerong Mao. *Stochastic differential equations and their applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997.

[MPS10] J.C. Mattingly, N.S. Pillai, and A.M. Stuart. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Arxiv preprint arXiv:1003.4306*, 2010.

[MPS11] J.C. Mattingly, N.S. Pillai, and A.M. Stuart. SPDE Limits of the Random Walk Metropolis Algorithm in High Dimensions. *Ann. Appl. Prob*, 2011.

[MRTT53] N. Metropolis, R.W. Rosenbluth, M.N. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[MSH02] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[MT93] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer, 1993.

[Oks03] B. Oksendal. *Stochastic differential equations*. Universitext. Springer, sixth edition, 2003. An introduction with applications.

[PST11] N.S. Pillai, A.M. Stuart, and A.H. Thiery. Optimal Scaling and Diffusion Limits for the Langevin Algorithm in High Dimensions. *Arxiv preprint arXiv:1103.0542*, 2011.

[RC99] C.P. Robert and G.C. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag, 1999.

[RGG97] G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.*, 7:110–120, 1997.

[RR98] G.O. Roberts and J. Rosenthal. Optimal scaling of discrete approximations to langevin diffusions. *J. Roy. Stat. Soc. B*, 60:255–268, 1998.

[RR01] G.O. Roberts and J. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16:351–367, 2001.

[Stu10] A.M. Stuart. Inverse problems: a Bayesian approach. *Acta Numerica*, 19, 2010.

[Tie98] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.