Statistics 212: Lecture 24 (April 30, 2025)

Diffusion Sampling

Instructor: Mark Sellke

Scribe: Emma Finn

1 Recap of Last Class

Theorem 1.1 (Cameron–Martin). *Let μ be a deterministic function $\mu : [0,T] \to \mathbb{R}$ such that $\int_0^T \mu(t)^2 dt < \infty$. Denote by $\mathbb{P}$ the Wiener measure on $C([0,T];\mathbb{R})$ induced by a standard Brownian motion $(B_t)_{t\in[0,T]}$, and let $\mathbb{P}^\mu$ be the law of the process $B_t^\mu = B_t + \int_0^t \mu(s)\,ds$. Then $\mathbb{P}_\mu$ is the law of $B_t + \int_0^t \mu(s)ds$ and satisfies $\frac{d\mathbb{P}_\mu}{d\mathbb{P}}(B_{[0,T]}) = exp(\int_0^T \mu_s dB_s - \int_0^T \mu(s)^2/2ds)$. This is a local martingale.*

Corollary 1.2. *If $f(t) = \int_0^t \mu_s ds$ has an L^2 derivative then it follows that Law(B_t) and Law$(B_t + f_t)$ are mutually absolutely continuous.*

The next result is a converse. (Often the "Cameron-Martin theorem" means the conjunction of the corollary above and the theorem below.)

Theorem 1.3. *Conversely, if $f' \notin L^2$ then Law$(B_t + f_t)$ is singular with respect to Law(B_t), meaning that there exists a set $A \subseteq C([0,T])$ such that $\mathbb{P}_\mu(A) = 0$ and $\mathbb{P}_0(A) = 1$.*

Proof. (Sketch) The idea is to project on to a finite dimensional Fourier series subspace. We will work with perturbations of a Brownian bridge and assume $T = 1$ and $f(0) = f(1) = 0$.

Now consider the Fourier coefficients of a Brownian Bridge $\tilde{B}_t = (B_t - tB_1)$. From a previous homework, we have

$$\tilde{B}_t = \sum_{k=1}^{\infty} c_k \sin(k\pi t), \quad c_k \sim N\Big(0, \frac{2}{k^2\pi^2}\Big) \text{ independent.}$$

for some appropriate constant a. Likewise expand

$$f(t) = \sum_{k=1}^{\infty} d_k \sin(k\pi t),$$

so that

$$f'(t) = \sum_{k=1}^{\infty} k\pi\, d_k \cos(k\pi t)$$

and

$$\|f'\|_{L^2}^2 = \int_0^1 f'(t)^2 dt \propto \sum k^2 d_k^2.$$

Thus f has Fourier coefficients $d_1, d_2, d_3 \dots$ while f' has Fourier coefficients $d_1, 2d_2, 3d_3, \dots$. On the finite-dimensional projection onto modes $1, \dots, n$, notice that

$$\mathrm{Law}(c_1, \dots, c_k) \propto \exp(-\alpha \sum_{k \geq 1} k^2 c_k^2)$$

for $\alpha = \pi^2/2$. This sum defines an inner product

$$\langle \vec{c}, \vec{d} \rangle_* = \sum_k k^2 c_k d_k.$$

Then, notice that $\sum k^2 d_k^2 = \infty \iff ||\vec{d}||_*^2 = \langle \vec{d}, \vec{d} \rangle_* = \infty$. The idea is that shifting a centered Gaussian is an absolutely continuous perturbation exactly when the shift has bounded norm relative to the inner product defining the Gaussian.

To avoid infinite-dimensional technicalities, we can work in large finite dimension by truncating $\vec{c}$ to $(c_1, \dots, c_n)$. Thus define the truncated inner product

$$\langle \vec{c}, \vec{d} \rangle_{*,n} = \sum_{k=1}^{n} k^2 c_k d_k.$$

We will give events A_n which have high probability under the original Brownian bridge, but low probability under the shift by $\vec{d}$. A good choice of event A_n in Fourier space is

$$A_n = \{\vec{x} : \langle \vec{x}, \vec{d} \rangle_{*,n} \leq \|\vec{d}\|_{*,n}^2 / 2\}.$$

It is easy to see that A_n has probability tending to 1 under the Brownian bridge ($\mathbb{P}_0(A_n) \to 1$), and tending to 0 under the shifted measure ($\mathbb{P}_\mu(A_n) \to 0$), if $\|\vec{d}\|_{*,n}^2 \to \infty$ (which is equivalent to $\|\vec{d}\|_*^2 = \infty$). (Just note that the 1-dimensional projection $\langle \vec{x}, \vec{d} \rangle_{*,n}$ has distribution $N(0, \|\vec{d}\|_{*,n}^2)$ under $\mathbb{P}_0$ and $N(\|\vec{d}\|_{*,n}^2, \|\vec{d}\|_{*,n}^2)$ under $\mathbb{P}_\mu$.)

To get singularity, we need a single event A instead of a sequence A_n. For this, let n_k be an increasing sequence of integers such that A_{n_k} has probability at least $\mathbb{P}_0(A_{n_k}) \geq 1 - 2^{-k}$ under the Brownian bridge, and at most $\mathbb{P}_\mu(A_{n_k}) \leq 2^{-k}$ under the shift. Then let A be the event that all but finitely many of the events $(A_{n_k})_{k \geq 1}$ occur. Borel-Cantelli shows $\mathbb{P}_0(A) = 1$ and $\mathbb{P}_\mu(A) = 0$, as desired. □

2 Diffusion!

2.1 Sequential Sampling and Polya's Urn

Suppose my goal is to sample from some probability measure on $\mu \sim P([0,1])$. We define the law μ_{seq} on infinite sequences $(X_1, X_2, \dots)$ by first taking $p \sim \mu$, and then we have $X_1, X_2, \dots \overset{iid}{\sim} \mathrm{Bern}(p)$, which are in fact correlated since p is unknown.

Claim: If I can sample form this sequence measure I can also sample from μ, since $\lim_{T \to \infty} \frac{X_1 + \dots + X_T}{T} = p$ almost surely. Now we can sample μ_{seq} without first choosing p. Steps:

- Sample $X_1 \sim \mathrm{Law}_{\mu_{\text{seq}}}(X_1) \sim \mathrm{Bern}(\mathbb{E}^\mu[p])$
- Sample $X_2 \sim \mathrm{Law}_{\mu_{\text{seq}}}(X_2 | \mathscr{F}_1) \sim \mathrm{Bern}(\mathbb{E}^\mu[p | \mathscr{F}_1])$
- ...
- Sample $X_k \sim \mathrm{Law}_{\mu_{\text{seq}}}(X_k | \mathscr{F}_{k-1})$
- ...

Where we define $\mathscr{F}_{k-1} = \sigma(X_1, \dots X_{k-1})$ We have μ_{all} for the law of $(p, X_1, X_2, \dots)$ and we have $\mathrm{Law}_{\mu_{all}}(X_1, X_2, \dots | p) \overset{iid}{\sim} \mathrm{Bern}(p)$. Then $\mathrm{Law}_{\mu_{all}}(X_1, \dots) = mu_{\text{seq}} = \int \mathrm{Bern}(p)^{\otimes \infty} d\mu(p)$

Slogan: To sample I just need to be able to compute conditional means. The idea is to do this for a LOT of our X_is in this way and then we have that $p \approx \frac{X_1 + X_2 + \dots + X_T}{T}$

2.1.1 Polya's Urn

Consider an urn that starts with 1 black ball and 1 white ball. At each step, choose 1 ball uniformly, replace it with a copy. Suppose we have b black balls and w white balls in the urn then $P(\text{next ball is black}) = \frac{b+1}{b+w+2}$ in the next step. This is identical to sequential sampling as above with $\mu \sim \text{Unif}(0,1)$. In particular, let $X_i = 1$ if the ith ball is black and 0 if it is white. We begin by sampling $X_1 \sim \text{Bern}(\mathbb{E}^\mu[p]) = \text{Bern}(\frac{1}{2})$. Suppose we pick a black ball, next. Then $X_1 = 1$, and so $\mathbb{E}(p|X_1) = 2/3$. Then, we sample $X_2 \sim \text{Bern}(\mathbb{E}^\mu[p|\mathscr{F}_1]) = \text{Bern}(2/3)$. We can continue in the way described above. If k of the $X_1 \dots, X_n$ are black (ie = 1). Then we have that $\text{Law}(p|\mathscr{F}_n) \propto p^k(1-p)^{n-k}$, which means it has a Beta distribution. By beta properties we have $\mathbb{E}(p|\mathscr{F}_n) = \frac{k+1}{n+2}$.

This whole equivalence story means that the sequence $(b, w, w, b, b, b, w, \dots) \sim \mu_{seq}$ for $\mu \sim \text{Unif}(0,1)$. The limiting fraction of b is $\text{Unif}(0,1)$.

2.2 Diffusion Sampling from the Perspective of Stochastic Localization

We'd now like to apply this discretized approach to continuous distributions.
Goal Sample $X \sim \mu \in P(\mathbb{R}^d)$ where μ is compactly supported etc. This might be computationally hard when d is big (unlike the Polya urn example). Instead of Bernoulli observations, use Gaussian noisy observations and we'll take a small step limit.

Consider a small ϵ signal to noise ratio per step. Given the unkown X, observe

$$w_1 \sim \epsilon X + Z_1$$

$$w_2 \sim \epsilon X + Z_2$$

where $Z_i \overset{iid}{\sim} N(0,1)$. We apply the same sequential sampling procedure. Here $\mu_{all} = Law(x, w_1, w_2, \dots)$ while $\mu_{seq} = Law(w_1, w_2, \dots)$ Now we can sample μ_{seq} without first choosing X. We generate one sample as follows:

- Sample $W_1 \sim \epsilon\mu * N(0,\epsilon)$
- Sample $W_2 \sim \epsilon Law(x|\mathscr{F}_1) * N(0,\epsilon)$
- ...
- Sample $W_k \sim Law(X|\mathscr{F}_k) * N(0, \epsilon I_d)$
- ...

As before $X \approx \frac{W_1+W_2+\cdots+W_T}{T\epsilon}$

As written, this is tricky to compute because we have to sample each W_k from a convolution. But you can pass to the continuous time $\epsilon \to 0$ limit and things get nicer. Here for the "known x" process we keep track of $y_{k\epsilon} = \sum_{j=1}^k W_j \sim k\epsilon x + N(0, k\epsilon)$. Hence $W_k \mid X \sim \mathcal{N}\big(\epsilon X, \epsilon Id_d\big)$.

After T steps one recovers $\widehat{X}_T = \frac{1}{T\epsilon}\sum_{k=1}^T W_k \to X$ a.s.

In the limit, $\epsilon \to 0$, first notice that conditionally on X, you get $y_t = w_t|X \sim BM$ with drift X and $dy_t = dB_t + X dt$.

...

First, recall the fact that $\mu_{path} = \text{Law}(Z_t)$

$$dZ_t = dB_t + m_t dt$$

and from the Cameron Martin Theorem, we have $m_t = \mathbb{E}^{\mu_{all}}[X|Y_t = Z_t] = \frac{\int x e^{\langle X, Z_t\rangle(-tx^2/2)} d\mu(x)}{\int e^{\langle X, Z_t\rangle(-tx^2/2)} d\mu(x)}$

Q: Why are Y_t and Z_t equal in distribution?
Consider the process y_t relative to its natural filtration $\mathscr{G}_t$, where (X unknown). Then we have that $y_t - \int_0^t \mathbb{E}(X|\mathscr{G}_s]ds$ is a $\mathscr{G}_t$ martingale. In fact M_t is Brownian Motion with respect to $\mathscr{G}_{\sqcup}$. To see why, note first that Dubins-Schwartz tells us M_t is a time change of Brownian motion. We know that the time change must be trivial, since if there were non-trivial time change, then the quadratic variation "clock" $\lim_{\delta \downarrow 0} \sum (X_{(k+1)\delta} - X_{k\delta})^2$ would be wrong.

2.3 Standard Diffusion

One can write down the corresponding partial differential equation for the evolving density, or compare to the usual "score-based" diffusion samplers by re-parameterizing time and variance (see references).

2.4 References

For more on diffusions along the lines of the above approach, see Sampling, Diffusions, and Stochastic Localization by Andrea Montinari (`https://arxiv.org/abs/2305.10690`). (Many many **many** other good diffusion references as well.)