Lecture 16

Lecture date: Oct 3, 2007 *Scribe: Anand Sarwate*

1 The general plan of attack

Recall the setup from the last lecture: we have iid ± 1 symmetric random variables $\varepsilon_1, \varepsilon_2, \ldots,$ and

$$S_n = \sum_{i=1}^{n} \varepsilon_i \tag{1}$$

is the simple random walk on the integers. Our goal is to construct a version of $(S_n)_{n \geq 0}$ and Brownian motion $(B_t)_{t \geq 0}$ on the same probability space such that for all n

$$\max_{i \leq k \leq n} |S_k - B_k| = O(\log n) \ . \tag{2}$$

A cartoon of the coupling is shown in Figure 1.

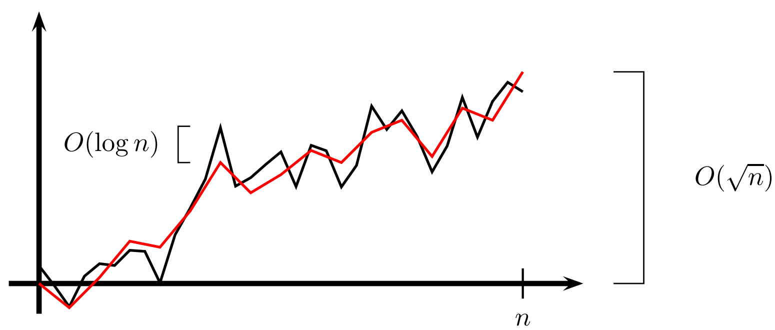

Figure 1: Coupling between the simple random walk and Brownian motion. At time n the variance is $\sqrt{n}$ but the maximum deviation between the two processes is at most $O(\log n)$.

The original KMT proof used an explicit construction of the coupling between the random walk and Brownian motion, whereas we rely on the Schauder-Tychonoff fixed point theorem. Our line of attack to construct the coupling is to couple a conditional process to the Brownian bridge. Figure 2 shows this coupling.

The induction hypothesis we will use is the following: given a possible value S_n, we can construct a random walk $S_0, S_1, \ldots, S_n$ with S_n having that value and a Brownian motion

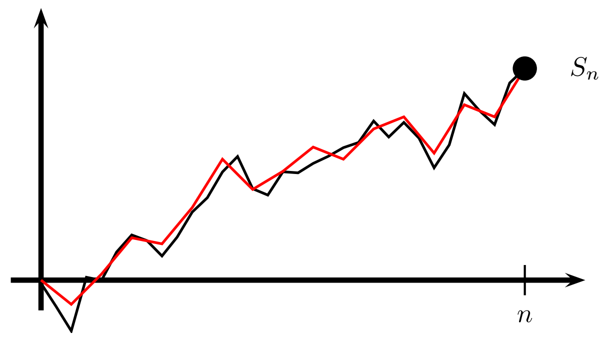

Figure 2: A tighter coupling between a simple random walk and the Brownian bridge, conditioned on $S_n = B_n$.

$(B_t)_{t\le n}$ conditioned to have $B_n = S_n$ such that for all $\lambda < \lambda_0$ we have

$$\mathbf{E}\exp\left(\lambda \max_{i\le n}|S_i - B_i|\right) \le \exp\left(C\log n + \frac{K\lambda^2 S_n^2}{n}\right) \ , \tag{3}$$

where C, K, and λ_0 must be chosen appropriately.

To see how to use this hypothesis, fix n and the value of S_n. Take $n/3 \le k \le 2n/3$ and assume that the induction hypothesis holds for all $n' < n$. The main step is to do a pointwise coupling of S_k with B_k such that the deviation is at most $O(1)$. Then we condition on a value for S_k and use the induction hypothesis to couple $(S_i)_{i\le k}$ with $(B_i)_{i\le k}$ and $(S_i)_{k\le i\le n}$ with $(B_i)_{k\le i\le n}$. Then by piecing it together you get that the result holds for n as well.

2 Pointwise coupling

The central question we have to answer is this : how do we couple something that is almost Gaussian to something that is exactly Gaussian such that we obtain exponentially decaying tails?

Suppose X is a random variable with density ρ, $\mathbf{E}\,X = 0$ and $\mathbf{E}\,X^2 < \infty$. Let

$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)} \ . \tag{4}$$

We know that for all well-behaved φ,

$$\mathbf{E}[X\varphi(X)] = \mathbf{E}[h(X)\varphi'(X)] \ . \tag{5}$$

The idea is that if $h(X) \approx \sigma^2$ with high probability, then X is approximately Gaussian with variance σ^2. Thus our objective is to construct a joint distribution on (X, Z) with marginals $X \sim \rho$ and $Z \sim \mathcal{N}(0, \sigma^2)$ such that the difference $X - Z$ is controlled by $h(X) - \sigma^2$.

Let

$$A(x_1,x_2) = \begin{pmatrix} h(x_1) & \sigma\sqrt{h(x_1)} \\ \sigma\sqrt{h(x_1)} & \sigma^2 \end{pmatrix} . \tag{6}$$

Note that $A(x_1,x_2)$ does not depend on x_2, and that it is positive semidefinite. By Lemma 1 we can construct a probability measure μ on $\mathbb{R}^2$ such that if $(x_1,x_2)^T \sim \mu$ then

$$\mathbf{E}\left[\left\langle \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \nabla f \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right\rangle\right] = \mathbf{E}\left[\mathrm{Tr}\left(A(X_1,X_2)\,\mathrm{Hess}\, f(X_1,X_2)\right)\right] \tag{7}$$

for all suitable f. Rewriting this a bit:

$$\mathbf{E}\left[X_1 \frac{\partial f}{\partial X_1} + X_2 \frac{\partial f}{\partial X_2}\right] = \mathbf{E}\left[h(X_1)\frac{\partial^2 f}{\partial X_1^2} + 2\sigma\sqrt{h(X_1)}\frac{\partial^2 f}{\partial X_1 \partial X_2} + \sigma^2 \frac{\partial^2 f}{\partial X_2^2}\right] . \tag{8}$$

Now take $\varphi : \mathbb{R} \to \mathbb{R}$ such that $|\varphi(x)|$, $|x\varphi(x)|$ and $|h(x)\varphi'(x)|$ are uniformly bounded, and let Φ be an antiderivative of φ so that $\Phi' = \varphi$.

Consider $f(x_1,x_2) = \Phi(x_1)$. Then all the x_2 terms in (8) vanish, so

$$\mathbf{E}[X_1\varphi(X_1)] = \mathbf{E}[h(X_1)\varphi'(X_1)] , \tag{9}$$

and by the previous lemma $X_1 \sim \rho$. Similarly, taking $f(x_1,x_2) = \Phi(x_2)$, we get

$$\mathbf{E}[X_2\varphi(X_2)] = \sigma^2\,\mathbf{E}[\varphi'(X_2)] , \tag{10}$$

so $X_2 \sim \mathcal{N}(0,\sigma^2)$. Note that the off-diagonal terms in (6) vanish for these two choices of f. If we set those terms to 0 then X_1 and X_2 would be independent.

By an earlier Lemma, we now have the bound:

$$\begin{aligned}
\mathbf{E}\exp\left(\theta|X_1 - X_2|\right) &\leq 2\,\mathbf{E}\exp\left(2\theta^2 v_{12}(x_1,x_2)\right) && (11)\\
&= 2\,\mathbf{E}\exp\left(2\theta^2(a_{11}(x_1,x_2) + a_{22}(x_1,x_2) - 2a_{12}(x_1,x_2))\right) && (12)\\
&= 2\,\mathbf{E}\exp\left(2\theta^2(h(x_1) + \sigma^2 - 2\sigma\sqrt{h(x_1)})\right) && (13)\\
&= 2\,\mathbf{E}\exp\left(2\theta^2(\sqrt{h(x_1)} - \sigma)^2\right) . && (14)
\end{aligned}$$

Remark: In a sense, the choice of A in (6) is the tightest coupling possible. Consider the following alternate choice for A:

$$A(x_1,x_2) = \begin{pmatrix} h(x_1) & h(x_1)\wedge\sigma^2 \\ h(x_1)\wedge\sigma^2 & \sigma^2 \end{pmatrix} . \tag{15}$$

For this A we get $v_{12}(x_1,x_2) = |h(x_1) - \sigma^2|$. In coupling $h(S_n) = n + O(\sqrt{n})$, so $v_{12}(x_1,x_2) = O(\sqrt{n})$. This choice corresponds to the Skorohod embedding, which gives

$$\max_{1\leq k\leq n} |S_k - B_k| = O(n^{1/4}) . \tag{16}$$

This bound on the deviation is the best possible for summands with finite 4-th moment. For finite p-th moment we can get $O(n^{1/p})$. The assumptions in the KMT are that the moment generating function is finite in a neighborhood of 0, which gives a $O(\log n)$ deviation.

Why do we choose this particular function $h(\cdot)$? Suppose $X_1, X_2, \ldots, X_n$ are iid, distributed according to the density ρ, with mean 0 and unit variance, and define $h(\cdot)$ as in (4). Let

$$S = \sum_{i=1}^{n} X_i \ . \tag{17}$$

What is the function $h_S(\cdot)$ corresponding to the density of S? If we define as before $S_i = S - X_i$, we can calculate, using (5) :

$$\begin{aligned}
\mathbf{E}[S\varphi(S)] &= \sum_{i=1}^{n} \mathbf{E}\left[X_i \varphi\left(S_i + X_i\right)\right] &(18)\\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{E}\left[\varphi'\left(S_i + X_i\right) h(X_i)\right] &(19)\\
&= \mathbf{E}\left[\varphi'(S) \left(\sum_{i=1}^{n} h(X_i)\right)\right] \ . &(20)
\end{aligned}$$

Now, it is easy to check that $\mathbf{E}\, h(X_i) = \mathbf{E}\, X_i^2 = 1$. Therefore, using (5) again we see that

$$h_S(S) = \mathbf{E}\left[\sum_{i=1}^{n} h(X_i) \ \Big| \ S\right] = n + O(\sqrt{n}). \tag{21}$$

We now show the $O(1)$ bound on the coupling:

$$\begin{aligned}
(\sqrt{h_S(S)} - \sigma)^2 &= n\left(\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)^{1/2} - 1\right)^2 &(22)\\
&= n\left(1 + O\left(\frac{1}{\sqrt{n}}\right) + \ldots - 1\right)^2 &(23)\\
&= O(1) \ . &(24)
\end{aligned}$$