Lecture 8

Lecture date: Sept. 14, 2007 *Scribe: Laura Derksen*

1 Exchangeable pairs

Recall that a pair of random variables is called exchangeable if (W, W') and (W', W) are equal in distribution.

During the last lecture we obtained an upper bound on the Wasserstein distance between such a W and a Gaussian random variable Z:

Theorem 1 *Let Z be a standard Gaussian random variable. If (W, W') is an exchangeable pair of r.v.'s, $\mathbf{E}(W' - W|W) = -\lambda W$ for some $0 < \lambda < 1$, and $\mathbf{E}(W^2) = 1$ (or $\mathbf{E}((W' - W)^2) = 2\lambda$), then*

$$Wass(W, Z) \leq \sqrt{\frac{2}{\pi} \mathrm{Var}\left(\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)\right)} + \frac{1}{3\lambda}\mathbf{E}(|W' - W|^3). \tag{1}$$

Intuitively, if $\mathbf{E}(W'-W|W) = -\lambda W$, $\mathbf{E}((W'-W)^2) = 2\lambda + o(\lambda)$, and $\mathbf{E}(|W'-W|^3) = o(\lambda)$ then $\mathrm{Wass}(W, Z) = o(1)$.

Usually the quantity $\frac{1}{2\lambda}(W' - W)^2$ is not concentrated. However, we will often have a σ-algebra $\mathcal{F}$ such that W is measurable with respect to $\mathcal{F}$ and

$$\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|\mathcal{F}\right)$$

is concentrated. By Jensen's inequality,

$$\mathrm{Var}\left(\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)\right) \leq \mathrm{Var}\left(\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|\mathcal{F}\right)\right).$$

2 Example: CLT for the scaled sum of i.i.d. random variables

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with mean 0 and variance 1. Let

$$W = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i.$$

As seen last lecture, we define W' as follows:

$$W' = \frac{1}{\sqrt{n}} \sum_{j \neq I} X_j + \frac{X_I}{\sqrt{n}},$$

where the index I is chosen uniformly at random from $\{1, 2, ..., n\}$ and X_I is independent from, and equal in distribution to the other X_i's.

Then $\mathbf{E}(W' - W|W) = -\frac{1}{n}W$ so $\lambda = \frac{1}{n}$. $\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)$ is hard to compute. However, we can write $\frac{1}{2\lambda}(W' - W)^2 = \frac{1}{2}(X_I' - X_I)^2$, and if $\mathcal{F}$ is $\sigma(X_1, X_2, ..., X_n)$ then

$$\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|\mathcal{F}\right) = \frac{1}{2} + \frac{1}{2n}\sum_{i=1}^{n} X_i^2,$$

which is concentrated.

3 Hoeffding combinatorial central limit theorem

Suppose $(a_{ij})_{i,j=1}^n$ is an array of numbers. Let π be a uniform random permutation of $\{1, 2, ..., n\}$. Let $W = \sum_{i=1}^n a_{i\pi(i)}$.

We would like to say something about how close $\frac{W - \mathbf{E}(W)}{\sqrt{\mathrm{Var}(W)}}$ is to the standard Gaussian distribution $N(0, 1)$.

Hoeffding's original proof involved a sequence of matrices $(a_{ij}^{(n)})_{i,j=1}^n$ and gave conditions for convergence to normality. The method of moments was used for the proof. The idea is to show that

$$\mathbf{E}\left(\left(\frac{W_n - \mathbf{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right)^k\right)$$

converges to 0 for k odd, and to $\frac{(2k)!}{2^k k!}$ for k even.

Bolthausen ('83 or '84) proved a Berry-Esseen bound for finite n using Stein's method.

We assume the following, without loss of generality:

$$\sum_{j=1}^{n} a_{ij} = 0, \ \sum_{i=1}^{n} a_{ij} = 0 \text{ and } \frac{1}{n-1}\sum_{i,j=1}^{n} a_{ij}^2 = 1. \tag{2}$$

To see why this does not compromise generality, for an arbitrary $(a_i j)_{i,j=1}^n$ we define

$$a_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} a_{ij},$$

$$a_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij},$$

$$a_{\cdot\cdot} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij},$$

and

$$\tilde{a}_{ij} = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot}.$$

Now,

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{a}_{ij} &= \sum_{i=1}^{n} a_{ij} - \sum_{i=1}^{n} a_{i\cdot} - \sum_{i=1}^{n} a_{\cdot j} + \sum_{i=1}^{n} a_{\cdot\cdot} \\
&= \sum_{i=1}^{n} a_{ij} - \frac{1}{n}\sum_{i,j=1}^{n} a_{ij} - \sum_{i=1}^{n} a_{ij} + \frac{1}{n}\sum_{i,j=1}^{n} a_{ij} \\
&= 0.
\end{aligned}$$

Similarly, we can check that the other assumptions in (2) are satisfied by $(\tilde{a}_{ij})$.

We define

$$\tilde{W} = \sum_{i=1}^{n} \tilde{a}_{i\pi(i)} = \sum_{i=1}^{n} a_{i\pi(i)} - \sum_{i=1}^{n} a_{i\cdot} - \sum_{i=1}^{n} a_{\cdot\pi(i)} + na_{\cdot\cdot} = \sum_{i=1}^{n} a_{i\pi(i)} - na_{\cdot\cdot}.$$

It can easily be checked that

$$\frac{\tilde{W} - \mathbf{E}(\tilde{W})}{\sqrt{\mathrm{Var}(\tilde{W})}} = \frac{W - \mathbf{E}(W)}{\sqrt{\mathrm{Var}(W)}},$$

justifying (2).

We now return to our original problem, and assume (2). Then we have

$$\mathbf{E}(a_{i\pi(i)}) = \frac{1}{n}\sum_{j=1}^{n} a_{ij} = 0,$$

so $\mathbf{E}(W) = 0$. For the variance, we can write

$$\mathrm{Var}(W) = \sum_{i=1}^{n} \mathrm{Var}(a_{i\pi(i)}) + \sum_{i\neq j} \mathrm{Cov}(a_{i\pi(i)}, a_{j\pi(j)}).$$

First,

$$\mathrm{Var}(a_{i\pi(i)}) = \mathbf{E}(a_{i\pi(i)}^2) = \frac{1}{n}\sum_{j=1}^{n} a_{ij}^2,$$

so

$$\sum_{i=1}^{n} \mathrm{Var}(a_{i\pi(i)}) = \frac{1}{n} \sum_{i,j=1}^{n} a_{ij}^2.$$

Now we will calculate the covariance.

$$\begin{aligned}
\mathrm{Cov}(a_{i\pi(i)}, a_{j\pi(j)}) &= \mathbf{E}(a_{i\pi(i)} a_{j\pi(j)}) \\
&= \frac{1}{n-1} \sum_{k,l \neq k} a_{ik} a_{jl} \\
&= \frac{-1}{n(n-1)} \sum_{k} a_{ik} a_{jk}
\end{aligned}$$

where the last equality comes from the fact that $\sum_{l \neq k} a_{jl} = -a_{jk}$.

We now obtain

$$\begin{aligned}
\sum_{i \neq j} \mathrm{Cov}(a_{i\pi(i)}, a_{j\pi(j)}) &= \frac{-1}{n(n-1)} \sum_{i \neq j} \sum_{k} a_{ik} a_{jk} \\
&= \frac{1}{n(n-1)} \sum_{i,k} a_{ik}^2.
\end{aligned}$$

Combining the variance and covariance calculations above, and keeping (2) in mind, we obtain

$$\mathrm{Var}(W) = \frac{1}{n-1} \sum_{i,j=1}^{n} a_{ij}^2 = 1.$$

Next, we will create an exchangeable pair (π, π') by defining $\pi' = \pi \circ (I, J)$ and $W' = \sum_{i=1}^{n} a_{i\pi'(i)}$ where (I, J) is a uniformly random transposition.

To be continued in the next lecture.