9 Lecture 9

9.1 Martingales

Chernoff bounds are almost tight for most purposes for sums of independent 0-1 random variables, but they cannot be used for sums of dependent variables. In this case, if the dependency is of particular type, the Azuma-Hoeffding inequality provides a more general concentration bound.

Definition 8. A martingale is a sequence of random variables $X_0, X_1, \ldots,$ of bounded expectation such that for every $i \geq 0$,

$$\mathbb{E}[X_{i+1} \,|\, X_0, \ldots, X_i] = X_i.$$

More generally, a sequence of random variables $Z_0, Z_1, \ldots$ is a martingale with respect to the sequence $X_0, X_1, \ldots$ if for all $n \geq 0$ the following conditions hold:

- Z_n is a function of $X_0, \ldots, X_n$,
- $\mathbb{E}[|Z_n|] < \infty$,
- $\mathbb{E}[Z_{n+1} \,|\, X_0, \ldots, X_n] = Z_n$.

9.1.1 Gambler's fortune

Example (Gambler's fortune). A gambler plays a sequence of *fair* games. Let X_i be the amount that the gambler wins on the i-th game; this will be positive or negative with probability $1/2$. Let Z_i denote the gambler's total winnings immediately after the i-th game. Since the games are fair, $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[Z_{i+1} \,|\, X_0, \ldots, X_i] = Z_i + E[X_{i+1}] = Z_i$, which shows that every finite[9] sequence $Z_0, Z_1, \ldots, Z_n$ is a martingale. This a generalization of the simple random walk on a line, because each bet can be arbitrary; in particular, the gambler can use the past outcomes and any algorithm to calculate the amount of the next bet.

9.1.2 Balls and bins

Example (Balls and bins). Suppose that we throw m balls into n bins independently and uniformly at random. This is one of the most-studied random experiments and we usually ask questions about the expected maximum load or the expected number of empty bins.

Here we consider the expected number of empty bins. Let X_i be the random variable representing the bin into which the i-th ball falls. Let Y be a random variable representing the number of empty bins. Then the sequence of random variables

$$Z_i = \mathbb{E}[Y \,|\, X_1, \ldots, X_i]$$

is a martingale. Clearly Z_i is a function of the $X_1, \ldots, X_i$'s and has bounded expectation. Furthermore

$$\begin{aligned}
\mathbb{E}[Z_{i+1} \,|\, X_1, \ldots, X_i] &= \mathbb{E}[\mathbb{E}[Y \,|\, X_1, \ldots, X_i, X_{i+1}] \,|\, X_1, \ldots, X_i] \\
&= \mathbb{E}[Y \,|\, X_1, \ldots, X_i] \\
&= Z_i.
\end{aligned}$$

[9] An infinite sequence may have unbounded expectation, violating the second condition of the definition.

We can view Z_i as an estimate of Y after having observed the outcomes $X_1, \ldots, X_i$. At the beginning Z_0 is a crude estimate, simply the expectation of Y. As we add more balls to the bins, Z_i's give improved estimates of Y, and at the end we get the exact value $Z_m = Y$.

9.1.3 Doob martingales

Example (Doob martingales). The balls and bins example is a typical *Doob martingale*. In general, Doob martingales are processes in which we obtain a sequence of improved estimates of the value of a random variable as information about it is revealed progressively. More precisely, suppose that Y is a random variable that is a function of random variables $X_0, X_1, \ldots$. As we observe the sequence of random variables $X_0, \ldots, X_n$, we improve our estimates of Y. The sequence of the mean estimates

$$Z_t = \mathbb{E}[Y \mid X_0, \ldots, X_t],$$

form a martingale with respect to the sequence $X_0, \ldots, X_n$ (provided that the Z_t's are bounded). Indeed, when we argued that the balls and bins process is a martingale, we used no property of the experiment, therefore the following holds in general

$$\begin{aligned}\mathbb{E}[Z_{t+1} \mid X_0, \ldots, X_t] &= \mathbb{E}[\mathbb{E}[Y \mid X_0, \ldots, X_t, X_{t+1}] \mid X_0, \ldots, X_t]] \\ &= \mathbb{E}[Y \mid X_0, \ldots, X_t] \\ &= Z_t.\end{aligned}$$

9.2 Azuma-Hoeffding inequality

Note that the random variables of a martingale are not in general independent. However, the following general concentration bound holds for every martingale.

Theorem 17 (Azuma-Hoeffding inequality). *Let $X_0, X_1, \ldots, X_n$ be a martingale such that*

$$|X_i - X_{i-1}| \leq c_i.$$

Then for any $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(X_n - X_0 \geq \lambda) &\leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}\right), \textit{ and} \\ \mathbb{P}(X_n - X_0 \leq -\lambda) &\leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}\right).\end{aligned}$$

Proof. We will only prove the first inequality, as the proof of the second one is very similar. The proof is again an application of Markov's inequality to an appropriate random variable and it is similar to the proof of Chernoff's bounds.

To simplify the notation, we use the variables $Y_i = X_i - X_{i-1}$. The steps of the proof are

1. - We use the standard technique of Chernoff bounds and instead of bounding $\mathbb{P}(X_n - X_0 \geq \lambda)$, we bound $\mathbb{P}(e^{t(X_n - X_0)} \geq e^{\lambda t})$ using Markov's inequality

   $$\mathbb{P}(e^{t(X_n - X_0)} \geq e^{\lambda t}) \leq e^{-\lambda t}\, \mathbb{E}[e^{t(X_n - X_0)}].$$

   - From now on we focus on $\mathbb{E}[e^{t(X_n - X_0)}]$, which can be rewritten in terms of Y_i's instead of X_i's, as

   $$\mathbb{E}[e^{t(X_n - X_0)}] = \mathbb{E}\left[\prod_{i=1}^n e^{tY_i}\right],$$

   by telescoping, $X_n - X_0 = \sum_{i=1}^n (X_i - X_{i-1}) = \sum_{i=1}^n Y_i$.

2. At this point in the proof of the Chernoff bounds, we used the fact that the variables are independent and we rewrote the expectation of the product as a product of expectations. We cannot do this here because random variables Y_i are not independent. Instead, we consider the conditional expectation

$$\mathbb{E}\left[\prod_{i=1}^{n} e^{tY_i} \,|\, X_0, \ldots, X_{n-1}\right] = \left(\prod_{i=1}^{n-1} e^{tY_i}\right) \mathbb{E}\left[e^{tY_n} \,|\, X_0, \ldots, X_{n-1}\right],$$

because for fixed $X_0, \ldots, X_{n-1}$, all but the last factor in the product are constants and can be moved out of the expectation.

With this in mind, we turn our attention on finding an upper bound on $\mathbb{E}[e^{tY_i} \,|\, X_0, \ldots, X_{i-1}]$.

- We first observe that $\mathbb{E}[Y_i \,|\, X_0, \ldots, X_{i-1}] = 0$, by the martingale property:

$$\begin{aligned}\mathbb{E}[Y_i \,|\, X_0, \ldots, X_{i-1}] &= \mathbb{E}[X_i - X_{i-1} \,|\, X_0, \ldots, X_{i-1}] \\ &= \mathbb{E}[X_i \,|\, X_0, \ldots, X_{i-1}] - \mathbb{E}[X_{i-1} \,|\, X_0, \ldots, X_{i-1}] \\ &= X_{i-1} - X_{i-1} \\ &= 0\end{aligned}$$

- Using the premise that $|Y_i| \le c_i$, we bound

$$e^{tY_i} \le \beta_i + \gamma_i Y_i,$$

for $\beta_i = (e^{tc_i} + e^{-tc_i})/2 \le e^{(tc_i)^2/2}$, and $\gamma_i = (e^{tc_i} + e^{-tc_i})/(2c_i)$. To show this, rewrite Y_i as $Y_i = rc_i + (1-r)(-c_i)$, where $r = \frac{1+Y_i/c_i}{2} \in [0,1]$, and use the convexity of e^{tx} to get

$$\begin{aligned}e^{tY_i} &\le re^{tc_i} + (1-r)e^{-tc_i} \\ &= \frac{e^{tc_i} + e^{-tc_i}}{2} + Y_i\,\frac{e^{tc_i} - e^{-tc_i}}{2c_i} \\ &= \beta_i + \gamma_i Y_i.\end{aligned}$$

To bound β_i from above, use the fact that for every x: $e^x + e^{-x} \le 2e^{x^2/2}$.

- Combine the above to get

$$\begin{aligned}\mathbb{E}[e^{tY_i} \,|\, X_0, \ldots, X_{i-1}] &\le \mathbb{E}[\beta_i + \gamma_i Y_i \,|\, X_0, \ldots, X_{i-1}] \\ &= \beta_i \le e^{(tc_i)^2/2}.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^{n} e^{tY_i}\right] = \mathbb{E}\left[(\prod_{i=1}^{n-1} e^{tY_i})\, e^{tY_n} \,|\, X_0, \ldots, X_{n-1}\right] &= (\prod_{i=1}^{n-1} e^{tY_i})\; \mathbb{E}[e^{tY_n} \,|\, X_0, \ldots, X_{n-1}] \\ &\le (\prod_{i=1}^{n-1} e^{tY_i})\, e^{(tc_n)^2/2}.\end{aligned}$$

3. We now take expectations on both sides to get rid of the conditional expectation,

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^{n} e^{tY_i}\right] &= \mathbb{E}\left[\mathbb{E}\left[(\prod_{i=1}^{n-1} e^{tY_i}) e^{tY_n} \,|\, X_0, \ldots, X_{n-1}\right]\right] \\ &\le \mathbb{E}\left[\prod_{i=1}^{n-1} e^{tY_i}\right] e^{(tc_n)^2/2}.\end{aligned}$$

4. Using standard techniques we can now finish the proof.

   - By induction, $\mathbb{E}[\prod_{i=1}^{n} e^{tY_i}] \leq \prod_{i=1}^{n} e^{(tc_i)^2/2} = e^{t^2 \sum_{i=1}^{n} c_i^2/2}$
   - Therefore $\mathbb{P}(e^{t(X_n - X_0)} \geq e^{\lambda t}) \leq e^{-\lambda t} e^{t^2 \sum_{i=1}^{n} c_i^2/2}$
   - Set $t = \lambda / \sum_{i=1}^{n} c_i^2$ to minimise the above expression and get the bound of the theorem.

Step 2 in the proof is crucial because, using conditionals, it bounds the product of the random variables $\prod_{i=1}^{n-1} e^{tY_i}$ and e^{tY_n}, although the two variables are not in general independent. □

9.2.1 Gambler's fortune, concentration of gains

Consider again the case of a gambler who plays a sequence of *fair* games. We have seen that if Z_i denotes the gambler's total winnings immediately after the i-th game, the sequence $Z_0, Z_1, \ldots, Z_n$ is a martingale. Suppose that the gambler has a very sophisticated algorithm to decide the amount that he bets every day that takes into account past bets and outcomes. Since the games are fair, the expected gain Z_n is 0.

Are the winnings concentrated around the mean value? Not in general; consider for example, the case in which the gambler puts higher and higher bets. Suppose now that there is a bound on the size of bets, for example suppose that the bets are at most 10 pounds. By the Azuma-Hoeffding inequality, the final winnings are concentrated with high probability in $[-k, k]$, where $k = O(\sqrt{n \log n})$.

9.2.2 Sources

- Mitzenmacher and Upfal, Sections 12.1, 12.4
- James Lee, lecture7.pdf (page 1)

10 Lecture 10

10.1 Applications of the Azuma-Hoeffding inequality

10.1.1 Chromatic number of random graphs

Random graphs are very useful models. The standard model is the $G_{n,p}$ model in which we create a graph of n nodes and edges selected independently, each with probability p. There are other models, such as the rich-get-richer models that try to model social networks, random geometric graphs, and others.

Let's consider the chromatic number $\chi(G)$ of a random graph G in the $G_{n,p}$ model. The chromatic number of a graph is the minimum number of colors needed in order to color the nodes of the graph in such a way that no adjacent nodes have the same color. It is NP-hard to compute the chromatic number of a graph.

Interesting questions:

- What is the expected chromatic number $\mathbb{E}[\chi(G)]$?
- Is the chromatic number concentrated around its expected value?

Interestingly, we can answer the second question without knowing $\mathbb{E}[\chi(G)]$. To do this we consider a *node exposure martingale.* Let G_i denote the subgraph with nodes $\{1, \ldots, i\}$ and consider the Doob martingale

$$Z_i = \mathbb{E}[\chi(G) \mid G_1, \ldots, G_i].$$

Note that $Z_0 = \mathbb{E}[\chi(G)]$.

Clearly, by exposing a new node cannot change the expected chromatic number by more than 1, that is

$$|Z_{i+1} - Z_i| \leq 1.$$

We can then apply the Azuma-Hoeffding inequality to get

$$\mathbb{P}(Z_n - Z_0 \geq \lambda\sqrt{n}) \leq \exp(-\lambda^2/2).$$

This shows that the chromatic number is with high probability within $O(\sqrt{n \log n})$ from its expected value[10].

10.1.2 Pattern matching

Consider a random string $X = (X_1, \ldots, X_n)$ in which the characters are selected independently and uniformly at random from a fixed alphabet Σ, with $|\Sigma| = s$. Let $p = (p_1, \ldots, p_k)$ be a fixed string (pattern). Let F be the number of occurrences of p in X.

- What is the expected number $\mathbb{E}[F]$? The probability that the pattern appears in positions $i+1, \ldots, i+k$ is exactly $1/s^k$. By linearity of expectation, the expected number of occurrences of p in X is $(n-k+1)/s^k$.

[10] In fact, much sharper concentration bounds are known.

We will show that the number of occurrences of p in X is highly concentrated. Consider the Doob martingale

$$Z_i = \mathbb{E}[F \mid X_1, \ldots, X_i],$$

with $Z_0 = \mathbb{E}[F]$ and $Z_n = F$. The important observation is that

$$|Z_{i+1} - Z_i| \leq k,$$

because every character can participate in at most k occurrences of the pattern.

We can apply the Azuma-Hoeffding inequality to getting

$$\mathbb{P}(|F - \mathbb{E}[F]| \geq \lambda k \sqrt{n}) \leq 2 \exp(-\lambda^2 k^2 n/(2nk^2) = 2e^{-\lambda^2}/2.$$

10.1.3 Balls and bins - number of empty bins

We consider again the balls and bins experiment. Suppose that we throw $m = n$ balls into n bins and tet's consider the expected number of empty bins. Let X_i be the random variable representing the bin into which the i-th ball falls. Let Y be a random variable representing the number of empty bins and consider the Doob martingale

$$Z_i = \mathbb{E}[Y \mid X_1, \ldots, X_i],$$

with $Z_0 = \mathbb{E}[Y]$ and $Z_n = Y$.

Since each ball cannot change the expectation by more than 1, we have

$$|Z_{i+1} - Z_i| \leq 1.$$

Applying the Azuma-Hoeffding inequality, we again see that the number of empty bins is highly concentrated around its mean value. More precisely,

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq \epsilon n) \leq 2e^{-\epsilon^2 n/2}.$$

But what is the expected number $\mathbb{E}[Y]$? The probability that a bin is empty is $(1 - 1/n)^n$, so by linearity of expectation, the expected number of empty bins is

$$n\left(1 - \frac{1}{n}\right)^n \approx \frac{n}{e}.$$

10.1.4 Sources

- Mitzenmacher and Upfal, Sections 12.1, 12.4
- James Lee, lecture7.pdf (page 1)