CS174 Fall 98: Lecture Note 6

Alistair Sinclair, September 1998; based on earlier notes by Manuel Blum/Douglas Young.

Unbalancing Lights

This topic is based on pp. 20–21 (Chapter 2, Section 5) of Alon & Spencer, attached.

One line of this argument (see ($*$) on page 21) needs additional explanation, which we now provide. This will also introduce another important new concept.

The Normal Distribution and the Central Limit Theorem

Consider an experiment and an associated r.v. with expectation $\mathrm{E}(X) = \mu$. Suppose we repeat the experiment n times, independently. Let the observed values of the r.v. be $X_1, X_2, \ldots, X_n$, and define $S_n = X_1 + X_2 + \cdots + X_n$. The sample average of our observations is $A_n = \frac{1}{n}S_n$.

What would you expect to happen to the sample average as $n \to \infty$? You would (I hope) expect it to get progressively closer to μ. (Since the sample average is a r.v., the phrase "get progressively closer to" has to be interpreted a little carefully: what we really mean is that $\Pr[|A_n - \mu| > \epsilon] \to 0$ as $n \to \infty$, for all $\epsilon > 0$.)

The above corresponds to our intuition that the behavior of an experiment repeated many times looks more and more like the expected behavior of a single experiment. It is commonly referred to as the Law of Large Numbers.

If we also know the variance $\mathrm{Var}(X) = \sigma^2$ (which we assume is finite), then we can make a much stronger statement: namely, the distribution of the sample average as $n \to \infty$ looks like a bell-shaped curve centered at μ, whose width is determined by σ. This bell shape is defined by the Normal distribution, and the result is called the Central Limit Theorem (CLT).

Definition: The (standard) Normal distribution is the continuous distribution with density function $\mathrm{n}(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$. Thus, if a r.v. Z is Normal, then its distribution is given by

$$\Pr[Z \le \beta] = \mathcal{N}(\beta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\beta} \mathrm{e}^{-x^2/2}dx. \qquad \Box$$

Note that $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}dx = 1$, so $\mathrm{n}(x)$ is a well-defined density function.

Note: The Normal distribution is the only piece of non-discrete probability we will need in this course. For more details, see (e.g.) Chapter VII of Feller.

Ex: Verify that, if Z has the Normal distribution, then $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$. $\Box$

The Normal distribution is a symmetric, bell-shaped curve centered at 0. Its height at 0 is about 0.4, and the peak is sharp: 50% of the mass is contained in the interval $[-0.67, 0.67]$, and 99.7% in the interval $[-3, 3]$. (See attached graphs from Feller.)

Now let's go back to our sample average $A_n = \frac{1}{n}S_n$. Clearly, $\mathrm{E}(A_n) = \frac{1}{n}\sum_i \mathrm{E}(X_i) = \mu$. And it is not hard to see that $\mathrm{Var}(A_n) = \frac{1}{n}\sigma^2$. (This last fact is a good exercise for you.)

The CLT says that, as $n \to \infty$, the distribution of $\frac{A_n - \mu}{\sigma/\sqrt{n}}$ approaches the Normal distribution. In other words, the discrepancy between A_n and its mean, μ, measured in units of the standard deviation $\frac{\sigma}{\sqrt{n}}$, is asymptotically normal.

Theorem (Central Limit Theorem): In the above notation, for every fixed β,

$$\Pr\left[A_n - \mu < \beta \tfrac{\sigma}{\sqrt{n}}\right] \to \mathcal{N}(\beta) \quad \text{as } n \to \infty. \qquad \Box$$

A proof of the CLT is beyond the scope of this course. However, let's use the CLT to justify line $(*)$ in Alon & Spencer.

In this situation we have $S_n = X_1 + \cdots + X_n$, where the X_i are independent with $\Pr[X_i = +1] = \Pr[X_i = -1] = \frac{1}{2}$. Thus $\mu = \mathrm{E}\,(X_i) = 0$ and $\sigma^2 = \mathrm{Var}\,(X_i) = 1$.

The CLT tells us that $\frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n}{\sqrt{n}}$ is asymptotically Normal. This means that as $n \to \infty$ the density of $\frac{|S_n|}{\sqrt{n}}$ approaches $2 \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ for $x \geq 0$. (Of course, it is zero for $x < 0$.)

So we have

$$\frac{\mathrm{E}\,(|S_n|)}{\sqrt{n}} \to \sqrt{\frac{2}{\pi}} \int_0^\infty x\mathrm{e}^{-x^2/2} dx = \sqrt{\frac{2}{\pi}}. \qquad (**)$$

Now, line $(*)$ claims that $\mathrm{E}\,(|S_n|) = \left(\sqrt{\frac{2}{\pi}} + o(1)\right)\sqrt{n}$. The term "$o(1)$" is just notation for any function $f(n)$ which tends to 0 as $n \to \infty$. (More generally, $o(g(n))$ stands for any function $f(n)$ such that $\frac{f(n)}{g(n)} \to 0$ as $n \to \infty$.) So line $(**)$ is in fact equivalent to line $(*)$ in Alon & Spencer, and we are done.