THE PALEY-ZYGMUND ARGUMENT AND THREE VARIATIONS

J. MICHAEL STEELE

Abstract. This is a class notes on techniques for proving "lower bounds." The leading example is the Paley-Zygmund argument. Closely related examples include the Chung-Erdös inequality and even Cantelli's inequality (which flips to an upper bound). The Cramer-Rao inequality rounds out the list.

1. Paley-Zygmund Argument

Consider a nonnegative random variable X. It is natural to let EX define a "unit of scale" and to look at probabilities such as $P(X \geq \theta EX)$ for $0 < \theta < 1$. In combinatorial problems it is often of importance to get a lower bound on this probability.

The classic way to proceed uses the Paley-Zygmund argument, which is also called the second moment method. One begins with the "tautological" identity determined by the scaled cut,

$$X = X\mathbb{1}(X < \theta EX) + X\mathbb{1}(X \geq \theta EX).$$

One then has the semi-automatic estimates

$$EX \leq \theta EX + E(X^2)^{1/2} P(X \geq \theta EX)^{1/2},$$

so, when we clear the expectations to the left and square, we have

$$\frac{(1-\theta)^2 (EX)^2}{E(X^2)} \leq P(X \geq \theta EX).$$

This simple inequality is at the heart of the "probabilistic method" which has been used by Erdös and others to solve some remarkable combinatorial problems.

Paley and Zygmund (1932) introduced this argument in a study of functions on the unit circle. They did not have probability in mind, but, after renomalization, any function on set with finite measure can be viewed as a random variable.

2. Chung-Erdös Inequality

Let $A_1, A_2, \ldots, A_n$ be events in a probability space. How can one get a lower bound on the probability of the event $B = \cup A_i$ that at least one of these events occurs? As in the Paley-Zygmund argument, one begins with a tautology,

$$\sum_{i=1}^{n} \mathbb{1}_{A_i} = \mathbb{1}_B \sum_{i=1}^{n} \mathbb{1}_{A_i}.$$

J. M. Steele: The Wharton School, Department of Statistics, Huntsman Hall 447, University of Pennsylvania, Philadelphia, PA 19104. Email address: `steele@wharton.upenn.edu`.

By Cauchy-Schwarz and squaring we then have

$$\{E(\sum_{i=1}^{n} \mathbb{1}_{A_i})\}^2 \leq P(B) E\left[\{\sum_{i=1}^{n} \mathbb{1}_{A_i}\}^2\right].$$

When we compute the expectations, we get the the lower bound

$$\frac{\sum_{i,j} P(A_i)P(A_j)}{\sum_{i,j} P(A_i \cap A_j)} \leq P\left(\cup_{i=1}^{n} A_i\right),$$

where the sums are over all pairs of integers $1 \leq i \leq n$ and $1 \leq j \leq n$.

This inequality is quite useful in arguments that refine the second Borel-Cantelli lemma, and this was the purpose for which it was introduced in Chung and Erdös (1952). Still, the argument has an inevitable quality to it, and it has been discovered many times, see e.g. Kochen and Stone (1964).

The effective use of the Chung-Erdös inequality often depends on a wise choice of the set of events to which it is applied. It is often applied to a blocks of events $A_j, A_{j+1}, \ldots, A_k$ of events in some infinite sequence of events $A_1, A_2, \ldots$, but sometimes it is applied to more sophisticated subsets of the sequence.

3. Cantelli's Inequality

The Cantelli inequality is a mild refinement of Chebyshev's inequality. To be forthright, it is not particularly useful. Nevertheless, the the proof of Cantelli inequality demonstrates an amusing variation of the tautology-plus-Cauchy-Schwarz argument.

Here we consider a random variable Y with $EY = 0$, we take $t > 0$ and note

$$0 \leq t = E(t - Y) \leq E[(t - Y)\mathbb{1}(t - Y \geq 0)].$$

By Cauchy-Schwarz we have

$$0 \leq t \leq \{E(t - Y)^2\}^{1/2} P(t \geq Y)^{1/2},$$

so, when we square and recall $EY = 0$, we have

$$t^2 \leq \{E(Y^2) + t^2\}\{1 - P(Y > t)\}.$$

Pure algebra then gives

$$P(Y > t) \leq \frac{E(Y^2)}{E(Y^2) + t^2}.$$

To make the comparison with Chebyshev's inequality, we take a random variable X and let $Y = X - EX$. The last inequality now gives

$$P(X > t + EX) \leq \frac{\operatorname{Var} X}{\operatorname{Var} X + t^2}.$$

This bound on the upper tail that is better than Chebyshev's inequality because of the extra summand $\operatorname{Var} X$ in the denominator.

4. Cramer-Rao Inequality

This inequality needs some terminology from mathematical statistics. We suppose that we have a family of densities $\{f_\theta(x) : \theta \in \Theta\}$, and we assume we have a function $\hat{\theta} : \mathbb{R} \to \mathbb{R}$ that we call an *esitimator*. We also suppose that this estimator is *unbiased* by which we mean that if X has the density f_θ then

$$E\hat{\theta}(X) = \theta \quad \text{or, equivalently} \int_{\mathbb{R}} \hat{\theta}(x) f_\theta(x)\, dx = \theta.$$

If we differentiate the last identity we have

$$\int_{\mathbb{R}} \hat{\theta}(x) \frac{d}{d\theta} f_\theta(x)\, dx = 1.$$

Just by the definition of the density function we for all θ that

$$\int_{\mathbb{R}} f_\theta(x)\, dx = 1.$$

and we can differentiate this to get

$$\int_{\mathbb{R}} \frac{d}{d\theta} f_\theta(x)\, dx = 0 = \int_{\mathbb{R}} \theta \frac{d}{d\theta} f_\theta(x)\, dx$$

If we the difference we get our desired tautology,

$$1 = \int_{\mathbb{R}} (\hat{\theta}(x) - \theta) \frac{d}{d\theta} f_\theta(x)\, dx = \int_{\mathbb{R}} (\hat{\theta}(x) - \theta) \frac{\frac{d}{d\theta} f_\theta(x)}{f_\theta(x)} f_\theta(x)\, dx,$$

and the Cauchy-Schwarz inequality gives us the bound,

$$1 \le \int_{\mathbb{R}} (\hat{\theta}(x) - \theta)^2 f_\theta(x)\, dx \int_{\mathbb{R}} \left\{ \frac{\frac{d}{d\theta} f_\theta(x)}{f_\theta(x)} \right\}^2 f_\theta(x)\, dx$$

The last integral has a name; it is called the expected Fisher information and it is denoted by $J(\theta)$. It is a bit like entropy and it is typically easy to calculate.

When we divide by $J(\theta)$ we get

$$\frac{1}{J(\theta)} \le \mathrm{Var}(\hat{\theta}(X)).$$

This is known as the Cramer-Rao inequality, and it tells us that no unbiased estimator can have a smaller variance than $1/J(\theta)$. This bound is the basis for a large part of what one knows about the efficiency of estimators. If you are working in the class of unbiased estimators and if you attain the lower bound $1/J(\theta)$, then you know that no one can ever beat you — however hard they try.

References

Chung, K. L. and Erdös, P. (1952), 'On the application of the Borel-Cantelli lemma', *Trans. Amer. Math. Soc.* **72**, 179–186.

Kochen, S. and Stone, C. (1964), 'A note on the Borel-Cantelli lemma', *Illinois J. Math.* **8**, 248–251.

Paley, R. and Zygmund, A. (1932), 'A note on analytic functions in the unit circle', *Proc. Camb. Phil. Soc.* **28**, 266–272.