An Introduction to the Probabilistic Method for Combinatorics in Isabelle/HOL

Chelsea Edmonds

Department of Computer Science and Technology, University of Cambridge
cle47@cl.cam.ac.uk

Keywords: Formalised Mathematics, Isabelle/HOL, Combinatorics, Probabilistic Method

The formalisation of mathematics continues to gain interest at a rapid pace. This growth is motivated by not only the need for correctness, but also the many potential benefits to mathematical research due to developments in automation, search, and natural language technology. There are now considerable formal libraries available across many modern proof assistants, including Isabelle/HOL, Lean, and Coq.

Traditionally, formalisations have focused on particularly important or valuable theorems. However, there are numerous, often intuitive techniques which mathematicians use on a regular basis that, if formalised separately, could become powerful tools in their own right. This is common in the field of combinatorics, which draws on a mixed-bag of techniques from other areas of mathematics. From a formalisation perspective, this offers several interesting challenges. For example, it requires combining formal libraries from different areas of mathematics and translating intuitive proof ideas and patterns to the more concrete representation required by the formal environment.

While formal combinatorics libraries have been growing over the last few years, there are still few formalisations which use *the probabilistic method*: a collection of powerful, commonly used tools from probability which are critical to proofs on the existence and construction of combinatorial structures. As such, the formalisation of these techniques would create a platform for significant future formalisations of research level mathematics.

The known existing formalisations of the probabilistic method [3, 4], are focused on proving two foundational results, not on the techniques itself. In this work, I turn the focus to the basic methodology and techniques, in order to be able to more easily prove multiple results. This work uncovered both some unexpected challenges for working with probability in Isabelle/HOL, as well as successes in both formalisation methodology, and verified theorems.

In this presentation, I'll focus on the formalisation of the basic techniques to prove properties on hypergraphs, a relatively simple and accessible area of combinatorics. I'll introduce hypergraphs through their formal definitions, completed using locales, Isabelle's module system. This builds on the success of the locale approach for other combinatorial structures [2]. Similarly, the probability concepts required are relatively simple and will be introduced through their pre-existing formal definitions in the Isabelle/HOL library.

I'll then present my work on formalising and applying the the basic probabilistic method in Isabelle/HOL, based of pen and paper results from the core textbook by Alon and Spencer [1]. The basic method follows a rough five step process on paper: (1) introduce randomness to a finite problem domain; (2) randomly construct on object from that domain; (3) identify the desired properties, or properties to avoid, i.e. "bad events"; (4) show the constructed object has the desired properties with probability strictly greater than 0, or has the "bad events" with probability strictly less than 1; then (5) obtain an object from the space with (or without) the desired properties.

Step (1) presents the first challenge of probability in a formal environment: the need to formally define the probability space which is rarely done on paper. I'll present my solution using locales to minimise duplication of this definition across related proofs. Once the probability space is defined, step (2) is usually relatively simple. Step (3) is often best done separately as definitions and lemmas on the property of a structure. This makes the overall library more modular, and avoids complicating the probabilistic proof. On paper, step (4) usually requires the use of probability lemmas, which usually focus on various bounds. I'll present the formalisation of several general lemmas to help reason on such bounds, which further tie into step (5). Specifically, I formalised basic bounds such as the complete independence bound, as well as the first formalisation of the Lovasz Locale lemma: a fundamental result in probability in the context of combinatorial results.

To conclude, I'll demonstrate an example of applying this basic method to prove an important result on the existence of hypergraphs with 2-vertex colourings. This proof itself is straightforward given the formal technique infrastructure now in place. I anticipate this methodology will be applicable to formalise many other results in combinatorics.

Acknowledgements

This work was funded by a Cambridge Australia Poynton International Scholarship, and a Cambridge Department of Computer Science and Technology Studentship. Thanks also to my supervisor, Prof. Lawrence C. Paulson (supported by the ALEXANDRIA project grant) for the useful discussions on this research.

References

[1] Noga Alon and Joel H. Spencer. *The Probabilistic Method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, Hoboken, N.J, 4th edition edition, 2016.

[2] Chelsea Edmonds and Lawrence C. Paulson. A modular first formalisation of combinatorial design theory. In Fairouz Kamareddine and Claudio Sacerdoti Coen, editors, *Intelligent Computer Mathematics*, pages 3–18. Springer, 2021.

[3] Lars Hupel. Properties of Random Graphs – Subgraph Containment. *Isabelle Archive of Formal Proofs.* `http://isa-afp.org/entries/Random_Graph_Subgraph_Threshold.html`, Formal Proof Development.

[4] Lars Noschinski. Proof Pearl: A Probabilistic Proof for the Girth-Chromatic Number Theorem. In *Interactive Theorem Proving. ITP 2012.*, volume 7406 of *Lecture Notes in Computer Science.* Springer Berlin Heidelberg.