

2011

Martingale Couplings and Bounds on Tails of Probability Distributions

Kyle Luh
Harvey Mudd College

Recommended Citation

Luh, Kyle, "Martingale Couplings and Bounds on Tails of Probability Distributions" (2011). *HMC Senior Theses*. 7.
https://scholarship.claremont.edu/hmc_theses/7

This Open Access Senior Thesis is brought to you for free and open access by the HMC Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in HMC Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.



Martingale Couplings and Bounds on Tails of Probability Distributions

Kyle Luh

Nicholas Pippenger, Advisor

Francis Su, Reader

May, 2011

HARVEY MUDD
COLLEGE

Department of Mathematics

Copyright © 2011 Kyle Luh.

The author grants Harvey Mudd College and the Claremont Colleges Library the nonexclusive right to make this work available for noncommercial, educational purposes, provided that this copyright statement appears on the reproduced materials and notice is given that the copying is by permission of the author. To disseminate otherwise or to republish requires written permission from the author.

Abstract

Wassily Hoeffding, in his 1963 paper, introduces a procedure to derive inequalities between distributions. This method relies on finding a martingale coupling between the two random variables. I have developed a construction that establishes such couplings in various urn models. I use this construction to prove the inequality between the hypergeometric and binomial random variables that appears in Hoeffding's paper. I have then used and extended my urn construction to create new inequalities.

Acknowledgments

I want to express my appreciation to Professor Nicholas Pippenger for advising me throughout the project. I would also like to thank Professor Francis Su for his helpful comments along the way. I would also like to express my gratitude to Claire Connelly for helping with the software issues I encountered.

Contents

Abstract	iii
Acknowledgments	v
1 Introduction	1
1.1 Stochastic Orders	1
1.2 The Hypergeometric Bound	2
1.3 Direction of Research	2
2 Background	3
2.1 Terminology	3
2.2 Sampling from a Finite Population	5
3 Research	7
4 History	11
4.1 When is a Random Variable More Variable than Another? . .	11
4.2 Converse of My Research	13
5 Future Work	15
Bibliography	17

List of Figures

2.1	Example of a convex function.	4
-----	---------------------------------------	---

Chapter 1

Introduction

Many realistic situations are much too intricate for exact calculations. In practice, one must resort to approximations and bounds. Establishing stochastic orders can lead to improved approximation and tighter bounds for various probabilistic models. Progress in this area can have ramifications in economics, statistical physics, queueing theory, reliability theory, epidemiology, and many other fields. (Muller and Stoyan, 2002)

1.1 Stochastic Orders

Stochastic orders are special cases of partial order relations of random variables.

Definition 1. A binary relation \preceq on a set S is a partial order if it satisfies the following conditions:

- Reflexivity: $x \preceq x$ for all $x \in S$;
- Transitivity: if $x \preceq y$ and $y \preceq z$ then $x \preceq z$;
- Antisymmetry: if $x \preceq y$ and $y \preceq x$ then $x = y$.

(Muller and Stoyan, 2002)

When S is the set of all distributions of real-valued random variables, the partial ordering is a stochastic order.

1.2 The Hypergeometric Bound

In Wassily Hoeffding's 1963 paper, he studies methods by which bounds on the tails of probability distributions can be derived. He uses a stochastic order to derive a bound on the hypergeometric distribution as a special case of sampling from a finite population. For a hypergeometric distribution where there are a total of N balls, M of which are deemed as a success if drawn. For n draws with k successes,

$$H(M, N, n, k) \leq \left(\left(\frac{p}{p+t} \right)^{p+t} \left(\frac{1-p}{1-p-t} \right)^{1-p-t} \right)^n, \quad (1.1)$$

where $k = (p+t)n$, $p = \frac{M}{N}$ for $t \geq 0$.

This was later proved by Chvátal (1979) in an ad-hoc method.

The *hypergeometric distribution* is a common model in applied probability theory used to model the drawing without replacement of discrete objects. For discrete probability distributions, it is often the case that exact numerical distributions are computationally intensive for large populations. Thus, rather than an exact solution, inequalities are often utilized to bound the possible values.

1.3 Direction of Research

The goal of my research at the outset was to provide a more accessible treatment of Hoeffding's method and then to construct a concrete model with which to prove Hoeffding's inequality. I later use this construction to apply Hoeffding's inequality to obtain new results.

Chapter 2

Background

Hoeffding proves his bound for the hypergeometric distribution and more general inequalities by creating a martingale coupling between random variables.

2.1 Terminology

Definition 2. A coupling between two random variables X and Y is a pair of dependent random variables \hat{X} and \hat{Y} such that $X \stackrel{d}{=} \hat{X}$ and $Y \stackrel{d}{=} \hat{Y}$ (Ross and Peköz, 2007).¹

Definition 3. A martingale coupling is a coupling (\hat{X}, \hat{Y}) such that

$$E[\hat{Y} | \hat{X} = k] = k.$$

An example of a martingale coupling is the situation in which we consider two particles performing a random walk. If we simply enforce the condition that one particle perform the exact same movements as the other, we have a martingale coupling. Notice that when we scrutinize the particles individually, their distributions have not changed.

The next step in Hoeffding's work then utilizes Jensen's inequality to deduce a bound.

Theorem 2.1 (Jensen's Inequality). *Given that the function $f(x)$ is convex in an interval I , then for positive numbers p_1, \dots, p_N such that $p_1 + \dots + p_N = 1$*

¹ $Y \stackrel{d}{=} \hat{Y}$ indicates that the distribution of Y is the same as that of \hat{Y} .

4 Background

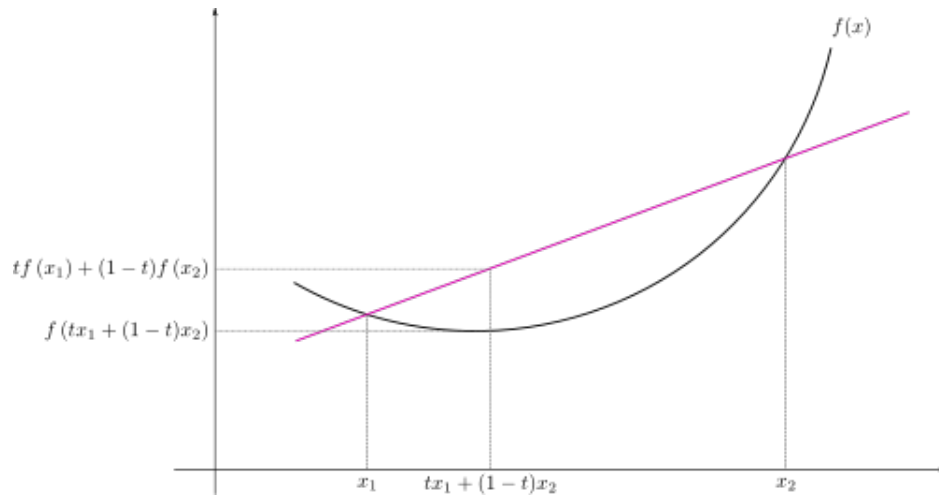


Figure 2.1 Example of a convex function.

and any numbers x_1, \dots, x_N in I ,

$$f\left(\sum_{i=1}^N p_i x_i\right) \leq \sum_{i=1}^N p_i f(x_i).$$

(Hoeffding, 1963)

Definition 4. A real-valued function defined on (a, b) is convex if

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

where $a < x_1 < b$, $a < x_2 < b$, $0 < t < 1$ (Rudin, 1964).

Figure 2.1 shows an example of a convex function.

Hoeffding uses Jensen's inequality in conjunction with martingale couplings to establish a bound for expectations of convex functions. I define this stochastic order as

Definition 5. A random variable X is at most as risky as a random variable Y , or $X \leq_{risk} Y$, if for any continuous, convex function f ,

$$E[f(X)] \leq E[f(Y)].$$

(Rothschild and Stiglitz, 1970)

This choice of terminology stems from the economists' application of this stochastic order to the assessment of risk.

2.2 Sampling from a Finite Population

Hoeffding proves his bound on the hypergeometric distribution by demonstrating that the bounds apply to the binomial distribution and that any tail-bound on the binomial must also apply to the hypergeometric.

Showing that the inequality applies to the binomial random variable is shown in Hoeffding's paper (1963). We now examine the latter half of the procedure. The crucial theorem is

Theorem 2.2. *Let $X = \sum_{i=1}^n X_i$ where X_1, X_2, \dots, X_n denote a random sample with replacement from a population that consists of N values c_1, c_2, \dots, c_N . Let $Y = \sum_{i=1}^n Y_i$ where Y_1, Y_2, \dots, Y_n denote a sample without replacement from the same population. If $f(x)$ is a convex function, then*

$$E[f(Y)] \leq E[f(X)].$$

(Hoeffding, 1963)

The proof of this theorem, which can be found in Hoeffding's work (1963), relies on the existence of a martingale coupling between X and Y , and the application of Jensen's inequality. Hoeffding did not find an explicit coupling and his argument is convoluted.

Later, in my research, I use the concept of exchangeable random variables.

Definition 6. *The discrete random variables X_1, X_2, \dots, X_n are exchangeable if, for every permutation i_1, i_2, \dots, i_n of the integers $1, \dots, n$,*

$$P[X_{i_1} = x_1, X_{i_2} = x_2, \dots, X_{i_n} = x_n] = P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]$$

for all x_1, \dots, x_n . (Ross, 1994)

Chapter 3

Research

We now examine a concrete procedure that couples the distribution of sampling from a finite population without replacement to sampling with replacement. Consider the following situation. We have two urns. The first contains N black balls, and the second contains N white balls. Each black ball is given an associated numerical value. Now we begin drawing balls from the white urn. Each time we draw a white ball, we replace it with a black ball and record the value of the black ball. If we draw a black ball from the white urn, we record its value and replace it. We continue this process until all the balls in the white urn have been replaced with black balls. To simulate drawing with and without replacement from a finite population, we scrutinize a curtailed version of my procedure.

We would like to model a situation where $X = \sum_{i=1}^n X_i$ and X_1, X_2, \dots, X_n denote a random sample with replacement from a population that consists of N values c_1, c_2, \dots, c_N ; and $Y = \sum_{i=1}^n Y_i$ where Y_1, Y_2, \dots, Y_n denote a sample without replacement from the same population.

For the urn of N black balls, let the values of the balls be c_1, c_2, \dots, c_N . Now consider the point in our procedure where we have only drawn n times from the white urn. Let X be the sum of the values of the n draws. Next, we examine the point in time at which we have drawn n balls from the black urn. Let Y be the sum of the values of the n black balls.

Theorem 3.1. X, Y is a martingale:

$$E[Y|X = k] = k. \quad (3.1)$$

Proof. Consider replacing the white urn with an urn of N white balls that are numbered $1, \dots, N$. We examine the equivalent procedure of first drawing n black balls and then drawing n times from this numbered urn, replac-

ing with the drawn black balls. Let $c_{i_1}, c_{i_2}, \dots, c_{i_n}$ denote the values of the n black balls drawn. At the end of this process, we should have recorded a series of integers representing the numbered balls that were drawn. We must then map the n black ball values to one of the integers 1 through n . Notice that the values of the drawn black balls are exchangeable random variables so all $n!$ assignments are equally likely. The probability that any integer i is assigned to a value c_{i_j} is $\frac{1}{n}$. Thus, the expected value of any draw from the white urn will be $\sum_{j=1}^n \frac{c_{i_j}}{n}$. Since there are n draws from the white urn, the expected value of Y will be $n \sum_{j=1}^n \frac{c_{i_j}}{n} = \sum_{j=1}^n c_{i_j} = k$. \square

Next, I claim that my procedure faithfully replicates the situation of drawing with or without replacement from a finite population.

Theorem 3.2. *X has the distribution of drawing without replacement from a finite population and Y has the distribution of drawing with replacement from the same finite population.*

Proof. Both X and Y are defined on the same probability space. Clearly, the black balls are drawn without replacement and the urn is a finite population. Our procedure establishes a bijective map between the white and the black balls with all $(a + b)!$ bijective maps being equally likely. Thus, drawing with replacement from the white urn is equivalent to drawing with replacement from the black balls. Thus, both draws are from the same finite population. \square

The above procedure demonstrates that drawing from a finite population without replacement is at most as risky as drawing from the same finite population with replacement.

Notice that a special case of our procedure leads to the binomial and hypergeometric coupling that results in Hoeffding's hypergeometric inequality. If we consider the values of the black balls to be limited to 1 or 0 (this can represent red and blue balls by convention), then X will be a hypergeometric random variable and Y will be a binomial random variable. From now on, I present my research in this specific context since it improves notational convenience. However, the following theorems obey a similar generalization.

The impetus behind our next theorem comes from the intuitive observation that if we instead have ta red balls and tb blue balls and we let t approach infinity, then drawing from this urn will be indistinguishable from a binomial distribution $B(a + b, \frac{a}{a+b})$.

Theorem 3.3. $Y \sim H(a, a + b, n)$ is at most as risky as $X \sim H(ta, ta + tb, n)$ where $t \in \mathbb{Z}^+$.

Proof. We prove this statement by establishing a martingale coupling between the two random variables. Consider the following procedure:

1. We begin with an urn of a red balls and b blue balls.
2. Introduce an urn with $ta + tb$ white balls and associate t balls with a unique serial number so that there are $(a + b)$ different serial numbers each with t balls.
3. Draw a ball from the urn with the white balls.
 - If the ball is white, draw a colored ball from the first urn, record the color, and then find the $(t - 1)$ white balls with the same serial number and paint them the color of the colored ball drawn from the colored urn.
 - If the ball is colored, record its color and remove.
 - Continue this process until all the balls in the white urn have are colored.

Let X be the number of red balls drawn from the white urn from n draws, and let Y be the number of red balls drawn from the colored urn when there have been n draws from the colored urn. To see that this is a martingale coupling, we consider the situation where there have been k red balls drawn by the time n colored balls have been drawn from the colored urn. Using a similar numbering system in the proof of Theorem 3.2, we give each group of serial numbers drawn from the white urn a number from 1 to n . Recognizing that the balls drawn from the colored urn are exchangeable random variables, we see that the probability that any one of the integers 1 through n is $\frac{k}{n}$. Given there are n draws, the expected value of X given Y is k .

We can see that the final result is a hypergeometric distribution, $H(ta, ta + tb, n)$, by noticing that every ball in the colored urn with $(a + b)$ balls maps to t balls in the urn of white balls with the same serial number. All $(a + b)!$ maps are equally likely. Since we draw without replacement, the result is the hypergeometric $H(ta, ta + tb, n)$. \square

Following a similar line of intuition, we conjecture the following:

Conjecture 1. For, $p, q \in \mathbb{Z}^+$, if $p > q$ then $X \sim H(pa, pa + pb, n)$ is more variable than $Y \sim H(qa, qa + qb, n)$.

We now apply our construction to the Pòlya urn model. In this model, we begin with a red balls and b blue balls. At each step, we draw a random ball, record its color, and return it along with another ball of the same color. Let the random variable P be the number of red balls drawn in n draws. Let X denote the hypergeometric random variable with the same initial distribution in the urn.

Theorem 3.4. $X \leq_{risk} P$.

Proof. We first develop a martingale coupling. Consider an additional urn of $(a + b)$ white balls. My process is to draw a ball from the white urn, if it is white, we draw a ball from the colored urn, record its color, and then replace the white ball with the colored ball and another ball of the same color (from an external supply). Continue this process until all the balls in the white urn are colored. Now let P be the number of red balls drawn after n draws. Let X be the number of red balls drawn from the colored urn once n balls have been drawn.

The random variables X_i in the Pòlya urn model are exchangeable (Ross, 1994). Assuming k red balls were drawn from the colored urn by the time n were drawn, we again find the expected value of a draw from the white urn to be $\frac{k}{n}$. After n draws, the conditional expectation from the white urn will be k , which shows that we have a martingale coupling. Drawing from the white urn is identical to the Pòlya model because there is a bijective mapping from the colored urn to the white urn. \square

Chapter 4

History

The stochastic order described in Theorem 2.2 has been studied in the context of economic theory and particularly the assessment of risk (Rothschild and Stiglitz, 1970). The question that Rothschild and Stiglitz attempt to answer is when a random variable Y is more variable than another random variable X .

4.1 When is a Random Variable More Variable than Another?

In previous mathematical economic research, there are four main answers to this question. Y is more variable than X if

1. Y is distributionally equal to X plus Z where $E(Z|X) = 0$. In the language of probability, this is the Martingale coupling that we discussed.
2. For all concave functions U , $E[U(X)] \geq E[U(Y)]$.
3. Y has more weight in the tails than X .
4. Y has a greater variance than X .

The first three conditions lead to an identical partial ordering of distributions, so it is the equivalence of these relations that we investigate. The purpose of examining these different characterizations is that Machina and Pratt (1997) establish a correspondence between the martingale coupling of two random variables **(1)**, the convexity condition **(2)**, and what are known

as the integral conditions (3). Ultimately, the integral conditions are more convenient mathematical criteria with which to work. However, before we verify these relations, I will introduce the notation that will appear in the remainder of this section. X and Y are random variables, F and G are cumulative to the respective cumulative distribution functions, and f and g are the respective density functions (when they exist).

4.1.1 Mean Preserving Spreads

Let $h(x)$ be the piece-wise function defined as

$$h(x) = \begin{cases} \alpha \geq 0, & \text{if } a < x < a + t; \\ -\alpha \leq 0, & \text{if } a + d < x < a + d + t; \\ -\beta \leq 0, & \text{if } b < x < b + t; \\ \beta \geq 0, & \text{if } b + e < x < b + e + t; \\ 0 & \text{otherwise,} \end{cases}$$

where

$$0 \leq a + t \leq a + d \leq a + d + t \leq b \leq b + t \leq b + e \leq b + e + t \leq 1$$

and

$$\beta e = \alpha d.$$

It should be clear geometrically that

$$\int_0^1 h(x) dx = \int_0^1 xh(x) dx = 0.$$

Therefore, if $g = f + h$, then

$$\int_0^1 g(x) dx = \int_0^1 f(x) dx + \int_0^1 h(x) dx = 1$$

so the total probability is conserved, and

$$\int_0^1 xg(x) dx = \int_0^1 x(f(x) + h(x)) dx = \int_0^1 xf(x) dx,$$

which implies that the expected value of f is unchanged with the addition of h . Thus, an addition to f with the properties of h is known as a *mean preserving spread* (Rothschild and Stiglitz, 1970). This is an equivalent definition for a martingale coupling.

4.1.2 Integral Conditions

Rothschild and Stiglitz (1970) show that if F and G are related by a sequence of mean preserving spreads, then the following integral conditions hold:

$$\int_0^x (G(\omega) - F(\omega)) d\omega \geq 0 \text{ for all } x \in [0, 1), \quad (4.1)$$

$$\int_0^1 (G(\omega) - F(\omega)) d\omega = 0. \quad (4.2)$$

4.1.3 Convexity

Due to convexity, Rothschild and Stiglitz (1970) use Jensen's inequality that for all concave functions U , $E[U(X)] \geq E[U(Y)]$ to prove the equivalence of all three conditions. They prove that condition (1) implies condition (2), condition (2) implies condition (3), and condition (3) implies condition (1).

4.2 Converse of My Research

My research has mainly focused on developing martingale couplings to show that one random variable is at most as risky as another. However, the converse, that there is a martingale coupling for every pair of random variables that obey this stochastic order, has also been proven. (Muller and Stoyan, 2002)

Theorem 4.1 (Extension of Strassen's Theorem). *For random variables X and Y that have expectations, the following conditions are equivalent:*

- $X \leq_{risk} Y$.
- *There exist random variables \hat{X} and \hat{Y} with the same distributions as X and Y , such that $E[\hat{Y}|\hat{X}] = \hat{X}$. In other words, there exists a martingale coupling.*

(Muller and Stoyan, 2002)

Chapter 5

Future Work

I hope to prove Conjecture 1 by the use of my construction. From there, I foresee many more opportunities to apply this construction. From these orderings, many bounds can be tightened and new inequalities may be derived.

Bibliography

Chvátal, V. 1979. The tail of the hypergeometric distribution. *Discrete Mathematics* 25(3):285–287.

Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58(301):13–30.

Machina, M., and J. Pratt. 1997. Increasing risk: Some direct constructions. *Journal of Risk and Uncertainty* 14(2):103–127.

Muller, A., and D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons Inc.

Ross, S. 1994. *A First Course in Probability*. ProbabilityBookstore.com.

Ross, S.M., and E.A. Peköz. 2007. *A Second Course in Probability*. ProbabilityBookstore.com.

Rothschild, M., and J.E. Stiglitz. 1970. Increasing risk: I. a definition. *Journal of Economic theory* 2(3):225–243.

Rudin, W. 1964. *Principles of Mathematical Analysis*, vol. 1976. McGraw-Hill New York;