# Honors 1: Undergraduate Math Lab*

Peter Sarnak†, Steven Miller‡, Alex Barnett§

September $11^{\text{th}}$, 2002

Courant Institute of Mathematical Sciences
New York University
New York, NY

**Abstract**

Quotient groups. Basic probability theory for random walk. Lecture by Steven Miller; notes by Alex Barnett and Steven Miller.

## 1 Quotient groups

Say we have a finite Abelian group $G$ (this means for all $x, y \in G$, $xy = yx$) of order $m$ which has a subgroup $H$ of order $r$. We will use multiplication as our group operation. Recall the *coset* of an element $g \in G$ is defined as the set of elements $gH = g\{h_1, h_2, \cdots, h_r\}$. Since $G$ is Abelian (commutative) then $gH = Hg$ and we will make no distinction between left and right cosets here.

The *quotient group* (or *factor group*), symbolized by $G/H$, is the group formed from the cosets of all elements $g \in G$. We treat each coset $g_iH$ as an element, and define the multiplication operation as usual as $g_iHg_jH$. Why do we need $G$ to be Abelian? The reason is we can then analyze $g_iHg_jH$, seeing that it equals $g_ig_jHH$. We will analyze this further when we prove that the set of cosets is a group.

---

*Homepage: `http://www.math.nyu.edu/~millerj/`

†E-mail: `sarnak@math.princeton.edu`

‡E-mail: `millerj@cims.nyu.edu` or `sjmiller@math.princeton.edu`

§E-mail: `barnett@nmr.mgh.harvard.edu`

There are several important facts to note. First, if $G$ is not Abelian, then the set of cosets might not be a group. Second, recall we proved the coset decomposition rule: given a finite group $G$ (with $n$ elements) and a subgroup $H$ (with $r$ elements) then there exist elements $g_1$ through $g_k$ such that

$$G = \bigcup_{i=1}^{k} g_i H. \tag{1}$$

The choices for the $g_i$'s is clearly not unique. If $g_1$ through $g_k$ work, so do $g_1 h_1$ through $g_k h_k$, where $h_i$ is any element of $H$. Recall this was proved by showing any two cosets are either distinct or identical.

We will show below that, for $G$ Abelian, the set of cosets is a group. Note, however, that while it might at first appear that there are many different ways to write the coset group, they really are the same. For example, the cosets $gH$ and $gh_1h_2^4h_3H$ are equal. This is similar to looking at integers mod $n$; mod 12, the integers 5, $-7$ and 19 are all equal, even though they look different.

We now prove that the set of cosets is a group (for $G$ Abelian).

**Closure.** By commutivity $g_iHg_jH = g_ig_jHH$. What is "$HH$"? Just the set of all $r^2$ possible combinations of elements of $H$. By closure, and the existence of the identity, this just gives $H$ again (recall no element in a group can appear more than once—duplicates are removed). Therefore $g_iHg_jH = g_ig_jH$. Now, as $G$ is a group and is closed, $g_ig_j \in G$. Thus, there is a $\alpha$ such that $g_ig_j \in g_\alpha H$ (as $G = \bigcup_{\beta=1}^{k} g_\beta H$. Therefore, there is an $h \in H$ such that $g_ig_j = g_\alpha h$, which implies $g_ig_jH = g_\alpha hH = g_\alpha H$. Thus, the set of cosets is closed under coset multiplication. Note, however, that while the coset $g_ig_jH$ is in our set of cosets, it may be written differently.

**Identity.** If $e$ is identity of $G$, then $eHg_iH = g_iH$ and $g_iHeH = g_iH$, so $eH$ is the identity of this quotient group.

**Associativity.** Since as you may have noticed, the quotient group elements behave just like those of $G$, associativity follows from that of $G$.

**Inverse.** It is easy to guess $g^{-1}H$ is the inverse of $gH$. Check it: $g^{-1}HgH = g^{-1}gH = eH =$ identity, also true the other way round of course by commutativity. Unfortunately, $g^{-1}H$ might not be listed as one of our cosets! Thus, we must be a little more careful. Fortunately, as $g^{-1} \in G = \bigcup_{\beta=1}^{k} g_\beta H$, there is an $\alpha$ such that $g^{-1} \in g_\alpha H$. Then, there is an $h \in H$ with $g^{-1} = g_\alpha h$. Thus, $g^{-1} = g_\alpha hH = g_\alpha H$, and direct calculation will show that the coset $g_\alpha H$ is the inverse (under coset multiplication) of $gH$.

# 2 Random walk and discrete probability

Each step in a random walk is a random event. We first study a single random event, then the combination of two random events, then multiply repeated events.

For other introductory probability theory see
`http://engineering.uow.edu.au/Courses/Stats/File24.html`

## 2.1 Probability distribution for a single event, mean, variance

Our single event consists of one of a set of choices happening. The choices are labelled by $i = 1 \cdots N$, which are exclusive (no more than one can happen), and complete (no less than one can happen). For instance, for a single coin toss,

$$\begin{aligned} i &= 1 \qquad\qquad H,\ \text{heads}, \\ i &= 2 \qquad\qquad T,\ \text{tails}. \end{aligned}$$

We take the choice $i$ as a *random variable*, meaning all we know is a probability $p_i \geq 0$ that each choice can happen. $p = 0$ means it never happens, $p = 1$ means it always happens, and most things are somewhere in between (the unbiased coin has $p_i = \frac{1}{2}$, $\forall i$, ignoring the small probabilities of the coin landing on its edge or quantum tunneling through the table.)

The set of $\{p_i\}$ we call the *probability distribution*. Completeness implies

$$\sum_i p_i = 1. \tag{2}$$

Note the abbreviation $\sum_i$ for $\sum_{i=1}^N$.

We have some quantity $f$ which has a value $f_i$ for each choice $i$. For instance, $f$ could be the number of dots on each face $i$ of a die, in which case $f_i = i$. *Mean* and *variance* are ways to characterize (summarize) the distribution over $f$.

**Mean.** The mean or *expectation value* (expected value) of $f$ over the distribution is

$$\bar{f} \equiv E[f] = \sum_i p_i f_i. \tag{3}$$

This is just a weighted average of $f$. Check it for the (unweighted) dice. $p_i = \frac{1}{6}$, $\forall i$. You should get $\bar{f} = \frac{7}{2}$.

**Variance.** The variance is the square of the *standard deviation* $\sigma_f$:

$$\sigma_f^2 \equiv \text{var}[f] \;=\; \sum_i p_i (f_i - \bar{f})^2. \tag{4}$$

Crudely $\sigma_f$ gives the width of the distribution in $f$. From the definition you can see $\sigma_f$ is the *root mean square* (rms) deviation from the mean. Expanding out the square, and using earlier results,

$$\begin{aligned} \text{var}[f] &= \sum_i p_i f_i^2 - 2\bar{f} \cdot \sum_i p_i f_i + \bar{f}^2 \cdot \sum_i p_i \\ &= \sum_i p_i f_i^2 - \bar{f}^2 \\ &= E[f^2] - E[f]^2. \end{aligned} \tag{5}$$

This is a very useful formula. The first term is often known as the *second moment* of $f$. The *first moment* is just the mean. The $m^{th}$ *moment* is defined as $\sum_i p_i f_i^m$.

A quick comment about units. If $f_i$ and $f$ are measured in feet (for example), the variance (which gives information on how spread out $f_i$ is) has units feet-squared. If someone asks how tall the people in our class are, one would answer about $5\frac{1}{2}$ feet. If one is further pressed to give a range for the heights of our class, one might say $5\frac{1}{2}$ feet, plus or minus $\frac{1}{4}$ of a foot. One would not give the error range in feet-squared! Thus, in measuring error it is the square-root of the variance that comes into play. Note that if $f_i$ is in feet, the variance is in feet-squared, and the square-root of the variance is in feet.

## 2.2 Multiple events

Consider two random events. Let the first have choices $i = 1 \cdots N$ (with probabilities $p_1$ through $p_N$); let the second event have choices $j = 1 \cdots M$ (with probabilities $q_1$ through $q_M$). Then there are $NM$ choices (possibilities) for the combined event, which we could label by $k \equiv ij$. Since $k$ is also a random variable, it has probability $r_k = r_{ij}$ (you could think of this as a matrix in $i, j$). If the two events are *independent* (also called *uncorrelated*), then

$$r_{ij} \;=\; p_i q_j, \;\; \forall ij \qquad\qquad \text{independence.} \tag{6}$$

(In other words, the matrix is separable in the $i$ and $j$ directions). Many events we study will be independent. For example, if you flip a fair coin twice, the result of the first flip has no effect on the result of the second flip. Or if you roll a fair die, et cetera.

As before we have a quantity $f_i$ associated with each choice $i$ for the first event. For the second event we have a (in general different) quantity $g_j$ for its choice $j$. We want to learn about the sum of these two quantities,

$$s \equiv f + g. \tag{7}$$

Note that for each combined choice $ij$, this quantity $s$ has value $s_{ij} = f_i + g_j$. How is $s$ is distributed? We will show that *independence* of the two events implies a simple law giving the mean and variance of $s$ in term of those of $f$ and $g$.

We compute the mean of $s$, following Eq. 3, except now we are summing over all combined possibilities $ij$,

$$\begin{aligned} E[s] &= \sum_{ij} r_{ij} s_{ij} = \sum_{ij} p_i q_j (f_i + g_j) = \sum_i p_i f_i \cdot \sum_j q_j + \sum_i p_i \cdot \sum_j q_j g_j \\ &= \bar{f} \cdot 1 + 1 \cdot \bar{g} = \bar{f} + \bar{g}. \end{aligned} \tag{8}$$

So, the means *add*.

We isolate this important fact:

**Lemma 2.1** *For independent events, the mean of a sum is the sum of the means. Equivalently, the sum of the expected values is the expected value of the sums. Thus, for any independent events $A$ and $B$, $E[A + B] = E[A] + E[B]$.*

What if we multiply $A$ by a constant $c$? For example, consider outcomes $A_i$ with probabilities $p_i$. The mean $\bar{A} = E[A]$. What is the mean of the new event with outcomes $cA_i$ occurring with probabilities $p_i$?

Well,

$$\begin{aligned} E[cA] &= \sum_i p_i c A_i \\ &= c \sum_i p_i A_i \\ &= cE[A]. \end{aligned} \tag{9}$$

We have therefore shown

**Lemma 2.2** *The mean of a multiple is the multiple of the mean. Equivalently, the expected value of a multiple is the multiple of the expected value. Thus,* $E[cA] = cE[A]$.

We now calculate the variance of a sum, using the above results. For the variance of $s = f + g$ we can use Eq. 5 to get

$$\begin{aligned} \mathrm{var}[s] &= E[s^2] - E[s]^2 \\ &= E[(f+g)^2] - (E[f+g])^2 \\ &= E[(f+g)^2] - (E[f]+E[g])^2 \text{ by Lemma 2.1} \\ &= E[(f+g)^2] - (\bar{f}+\bar{g})^2 \\ &= E[f^2] + 2E[fg] + E[g^2] - (\bar{f}+\bar{g})^2. \end{aligned} \tag{10}$$

We justify the last step as follows:

$$\begin{aligned} E[(f+g)^2] &= E[f^2 + 2fg + g^2] \\ &= E[(f^2 + 2fg) + g^2] \\ &= E[(f^2 + 2fg)] + E[g^2] \text{ by Lemma 2.1} \\ &= E[f^2] + E[2fg] + E[g^2] \text{ by Lemma 2.1} \\ &= E[f^2] + 2E[fg] + E[g^2] \text{ by Lemma 2.2.} \end{aligned} \tag{11}$$

Using independence again we can factorize $E[fg] = \sum_{ij} r_{ij} f_i g_j = \sum_i p_i f_i \cdot \sum_j q_j g_j = E[f]E[g] = \bar{f}\bar{g}$. Elegantly, this term is responsible for cancelling the cross-term in $(\bar{f}+\bar{g})^2$, and collecting the remaining terms leaves

$$\mathrm{var}[s] = \mathrm{var}[f] + \mathrm{var}[g] \tag{12}$$

So, the variances *also* add.

We now take the special case when the second event is identical to the first. That is, $M = N$, $q_i = p_i$, $g_i = f_i$, $\forall i$. In this case the above shows that $\bar{s} = 2\bar{f}$ and $\mathrm{var}[s] = 2\mathrm{var}[f]$.

We can repeat the above combination law for successively repeated events. Such events are called *identical independently distributed* (iid) events. Suppose we have $K$ repetitions of iid events, with total $s \equiv f + g + \cdots + z$, we can just repeatedly apply the above rules to get

$$\bar{s} = K\,\bar{f} \tag{13}$$

$$\mathrm{var}[s] = K\,\mathrm{var}[f]. \tag{14}$$

The mean and variance are not the only characteristics of the distribution that add like in this way. Amazingly, there is an infinite sequence of special combinations of the higher moments, called *cumulants*, which add just like this. The mean and variance are just the first two cumulants.

### 2.3 Simplest random walk

We now have the tools to characterize a random walk. We choose $N = 2$ and $p_1 = p_2 = \frac{1}{2}$, just as with the coin toss, and define the "step" displacements $f_1 = +1$ and $f_2 = -1$. This corresponds to a drunkard taking (uncorrelated) steps of unit length along the integer line.

We use Eqs. 3 and 5 to evaluate the mean and variance of a single step event.

$$\bar{f} = +\frac{1}{2} - \frac{1}{2} = 0 \tag{15}$$

$$\mathrm{var}[f] = \frac{1}{2}.(1)^2 + \frac{1}{2}.(-1)^2 - (0)^2 = 1. \tag{16}$$

For $K$ steps, starting from the origin, we have the final displacement of $s$, the sum of all the steps, using the formulae above,

$$\bar{s} = K \cdot 0 = 0 \tag{17}$$

$$\mathrm{var}[s] = K \cdot 1 = K. \tag{18}$$

So the standard deviation, *i.e.* the width of the distribution of $s$, has value

$$\sigma_s \equiv \sqrt{\mathrm{var}[s]} = \sqrt{K}. \tag{19}$$

This is our first vital fact about all but the most pathological [1] random walks: the distribution has width which scales like $K^{1/2}$. This means that a *typical* distance from the origin is $\sqrt{K}$. This is called a *diffusion process* and is very common in the real world.

Again, remember that if the person walks in feet, the variance (which is a measure of how much the distribution spreads out) will be in feet-squared. By taking the square-root we again have units of feet.

[1] Random walks which do *not* exhibit this power law have infinite variance and exhibit *anomalous diffusion*. For more on this, see M. Bazant's excellent course at `http://www-math.mit.edu/~bazant/teach/18.325`

## 2.4 Central Limit Theorem

The Central Limit Theorem (CLT) states that the distribution on $s$ tends to a *Gaussian distribution*,

$$p(s) \approx \mathcal{N}(\mu, \sigma^2) \;\equiv\; \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}} \tag{20}$$

with mean $\mu = \bar{s}$ and variance $\sigma^2 = \mathrm{var}[s]$ as given above, as $K \to \infty$. This is a very common "bell curve" with width $\sigma$ centered about $\mu$. We have not defined $p(s)$ very rigorously—it is simply the probability of being displaced $s$ from the origin at the end of the $K$-step random walk. An exact formula for $p(s)$ involves counting all the ways that $\pm 1$ can be added $K$ times to get exactly $s$. We will postpone this, and the proof of CLT, for next time.

Remarkably the CLT applies to *any* $N$ with any discrete step distribution $\{p_i\}$ and any step displacements $\{f_i\}$. It also applies to the case of continous-valued steps with distribution $p(f)$ along the real line. However a criterion for validity is always the *finiteness of the second moment* of $p(f)$.