# *A Note on Invariant Random Variables*

Jacek Cichoń and Marek Klonowski[†]

*Institute of Mathematics and Computer Science, Wrocław University of Technology, Poland*

In this paper we present a simple theory, based on the notion of group action on a set, which explains why processes of throwing random sets of points and throwing random lines are similar up to the second moment of counting functions connected with them. We also discuss other applications of this method and show how to calculate higher moments using the group acting on a set.

**Keywords:** group action, moments of random variable, random subset, projective space

## 1 Introduction

One method of improving safety of transmissions between simple sensing devices is to assign sets of cryptographic keys to them. Methods of distributions of such keys are called *key predistribution schema*. The basic probabilistic key predistribution schema from Eschenauer and Gligor (2002) can be described as follows: we have a pool $\mathcal{K}$ of cryptographic symmetric keys of cardinality $n$; each device $a$ obtains a randomly chosen subset $\mathcal{K}_a \subset \mathcal{K}$ of keys of cardinality $|\mathcal{K}_a| \approx \sqrt{n}$; due to Birthday Paradox $\mathcal{K}_a \cap \mathcal{K}_b \neq \emptyset$ with high probability; using any key $K \in \mathcal{K}_a \cap \mathcal{K}_b$ the two devices $a$ and $b$ can establish a secure connection. In order to control the probability of the event "$\mathcal{K}_a \cap \mathcal{K}_b \neq \emptyset$" one must carefully choose cardinalities of sets $\mathcal{K}_a$.

More advanced solutions use various kinds of combinatorial or geometric constructions. We can arrange the pool of keys $\mathcal{K}$ as a two dimensional space $V = (\mathbf{F}_p)^2$ over the field $\mathbf{F}_p$ and assign for each device $a$ a random line $\mathcal{K}_a$ in $V$. Then $\Pr[\mathcal{K}_a \cap \mathcal{K}_b \neq \emptyset] = \frac{1}{p}$.

A more interesting solution, based on finite projective geometries, was presented in Camtepe and Yener (2007). We fix a prime number $p$ and arrange the pool of keys $\mathcal{K}$ as the projective plane $\mathrm{PG}(2,p)$. This time sets $\mathcal{K}_a$ are lines in $\mathrm{PG}(2,p)$ and we get $\Pr[\mathcal{K}_a \cap \mathcal{K}_b \neq \emptyset] = 1$, since each of the two lines in $\mathrm{PG}(2,p)$ has a nonempty intersection.

There are a lot of variants of classical problems for each of the models described above. For example: we select independently random sets $\mathcal{K}_{a_1}, \ldots, \mathcal{K}_{a_k}$ and ask about cardinality of the set $\mathcal{K}_{a_1} \cup \cdots \cup \mathcal{K}_{a_k}$. The first case, with purely random subsets, is very closely related to the classical Coupon Collector Problem (see e.g. Gardy (1998), Flajolet et al. (1992)). During direct calculations of first two moments of these variables for all the above-mentioned models of keys generations we observed that first two moments are the same. The differences occur for the third moment. In this paper we want to explain this phenomenon.

[†]Emails: {`Jacek.Cichon, Marek.Klonowski`}`@pwr.wroc.pl`. Partially supported by Polish Ministry of Science and Higher Education, grant number N N206 1842 33.

If $q$ is a power of a prime then by $\mathbf{F}_q$ we denote the field with $q$ elements. If $V$ is a set and $n$ is a natural number then by $[H]^n$ we denote the family of all subsets of $H$ of cardinality $n$. The power set of $V$ is denoted by $\mathrm{P}(V)$. We denote by $\left\{ n \atop k \right\}$ Stirling numbers of the second type and by $\mathrm{s}(n,k)$ the signed Stirling numbers of the first kind. Finally, by $a^{\underline{k}}$ we denote the falling factorial, i.e. $a^{\underline{k}} = a(a-1)\cdots(a-(k-1))$. The expected value of a random variable $X$ is denoted by $\mathbf{E}\,(X)$. The indicator function of an event $A$ is denoted by $\mathbf{1}_A$.

# 2 Invariant Random Variables

Let $(G,\cdot)$ be a group. Let us recall (see e.g. Cameron (1996)) that an action of the group $G$ on the space $X$ is a binary function $G\times X\to X$, denoted as $(g,x)\mapsto g\cdot x$ such that $e\cdot x = x$ for all $x\in X$ and $g\cdot(h\cdot x) = (g\cdot h)\cdot x$ for all $g,h\in G$ and $x\in X$. This notion plays a very important role in finite combinatorics and is crucial in Pólya's counting theory (see deBruijn (1964)).

The action of $G$ on $X$ is called $n$-transitive if $X$ has at least $n$ elements and for any pairwise distinct $x_1$, ..., $x_n$ and pairwise distinct $y_1$, ..., $y_n$ elements from $X$ there is $g\in G$ such that $g\cdot x_k = y_k$ for all $1\leq k\leq n$. Notice that if the action of $G$ on $X$ is $n$-transitive and $1\leq r\leq n$, then the action is $r$-transitive, too.

Suppose that a group $(G,\cdot)$ acts on a space $V$. For subsets $A,B\subseteq V$ we define a relation

$$(A\sim_G B)\Leftrightarrow(\exists x\in G)(A = x\cdot B)$$

where $x\cdot B = \{x\cdot b : b\in B\}$. Clearly, $\sim_G$ is an equivalence relation on $\mathrm{P}(V)$.

**Definition 1** *Suppose that a group $G$ acts on a finite space $V$ and let $X$ be a random variable with values in $\mathrm{P}(V)$. Then $X$ is $G$-**invariant** if*

$$(\forall A,B\in \mathrm{P}(V))(A\sim_G B\Rightarrow \Pr[A = X] = \Pr[B = X])\ .$$

**Lemma 1** *Suppose that a group $(G,\cdot)$ acts on a finite space $V$, $a,b\subseteq V$, $a\sim_G b$ and that $X$ is a $G$-invariant random variable with values in $\mathrm{P}(V)$. Then $\Pr[a\subseteq X] = \Pr[b\subseteq X]$.*

**Proof:** Let us fix $x\in G$ such that $x\cdot a = b$. Then

$$\Pr[a\subseteq X] = \sum_A \Pr[a\subseteq A|X = A]\cdot\Pr[X = A] = \sum_{a\subseteq A}\Pr[X = A] = \sum_{x\cdot a\subseteq x\cdot A}\Pr[X = A] =$$
$$\sum_{b\subseteq B}\Pr[X = x^{-1}\cdot B] = \sum_{b\subseteq B}\Pr[X = B] = \ldots = \Pr[b\subseteq X]\ .$$

□

**Definition 2** *A random variable $X$ with values in $\mathrm{P}(V)$ is $r$-**homogeneous** if $|V|\geq r$ and for every two subsets $a,b$ of $V$ such that $|a| = |b|\leq r$ we have*

$$\Pr[a\subseteq X] = \Pr[b\subseteq X]\ .$$

**Lemma 1** *Suppose that a group $(G,\cdot)$ acts $r$-transitively on a finite space $V$ and that $X$ is a $G$-invariant random variable with values in the power set $\mathrm{P}(V)$. Then $X$ is $r$-homogeneous.*

**Proof:** If $G$ acts $r$-transitively on $V$, $1 \leq s \leq r$ then $G$ acts $s$-transitively on $V$, too. Hence if $a, b \subseteq V$ and $|a| = |b| \leq r$ then $a \sim_G b$, so the result follows from Lemma 1. $\square$

If $X$ is $r$-homogeneous random variable then we put

$$p(X, s) = \Pr[\{a_1, \ldots, a_s\} \subset X]$$

where $\{a_1, \ldots, a_s\}$ is an arbitrary subset of the space $V$ of pairwise distinct elements. The definition of $r$-homogeneous variables implies that number $p(X, s)$ is correctly defined, i.e. it does not depend on a particular choice of the set $\{a_1, \ldots, a_s\}$.

**Lemma 2** *Suppose that $X, Y$ are $r$-homogeneous independent random variables with values in the finite space $V$ defined on the same probability space $\Omega$. Let $X^c(\omega) = V \setminus X(\omega)$ and $Z(\omega) = X(\omega) \cap Y(\omega)$. Then $X^c$ and $Z$ are $r$-homogeneous random variables.*

**Proof:** Let us fix a sequence $(a_1, \ldots, a_s)$ of pairwise different elements from $V$, where $1 \leq s \leq r$. Let $a = \{a_1, \ldots, a_s\}$. Then, using the Inclusion-Exclusion Principle, we get

$$\Pr[a \subseteq X^c] = 1 - \Pr[a_1 \in X \vee \ldots \vee a_s \in X] = \sum_{k=0}^{s} \binom{s}{k} (-1)^k p(X, k)$$

and

$$\Pr[a \subseteq Z] = \Pr[a \subseteq X \wedge a \subseteq Y] = \Pr[a \subseteq X] \cdot \Pr[a \subseteq Y] = p(X, s) \cdot p(Y, s)\ .$$

$\square$

Therefore the class of $r$-homogeneous random variables is closed under standard set theoretical finitary operations applied to independent variables. We will show that the first $r$ moments of $r$-homogeneous random variables determine the sequence $(p(X, k))_{k \leq r}$ and, conversely, that the sequence $(p(X, k))_{k \leq r}$ determines its first $r$ moments.

**Corollary 1** *Suppose that $X$ is $r$-homogeneous random variable with values in the power set $\mathrm{P}(V)$. Then*

$$\mathbf{E}\left(|X|^r\right) = \sum_{k=1}^{r} |V|^{\underline{k}} \left\{ {r \atop k} \right\} p(X, k)\ .$$

**Proof:** Notice that $|X| = \sum_{v \in V} \mathbf{1}_{v \in X}$. Therefore (see e.g. Flajolet and Sedgewick (2009), Chapter II, p. II.6)

$$\mathbf{E}\left(|X|^r\right) = \sum_{(x_1, \ldots, x_r) \in V^r} \Pr[\{x_1, \ldots, x_r\} \subseteq X] =$$

$$\sum_{k=1}^{r} \left\{ {r \atop k} \right\} k! \sum_{b \in [V]^k} \Pr[b \subseteq X] = \sum_{k=1}^{r} \binom{|V|}{k} k! \left\{ {r \atop k} \right\} p(X, k) = \sum_{k=1}^{r} |V|^{\underline{k}} \left\{ {r \atop k} \right\} p(X, k)\ .$$

$\square$

**Lemma 3** *Suppose that $X$ is a $r$-homogeneous random variable with values in the power set $\mathrm{P}(V)$. Then*

$$p(X,r) = \frac{1}{|V|^{\underline{r}}} \sum_{k=1}^{r} \mathrm{s}(r,k) \mathbf{E}\left(|X|^k\right) = \frac{\mathbf{E}\left(|X|^{\underline{r}}\right)}{|V|^{\underline{r}}} .$$

**Proof:** . Let $x_k = \mathbf{E}\left(|X|^r\right)$ and $y_k = |V|^{\underline{k}} p(X,k)$ for $k = 1,\ldots,r$. According to Corollary 1, these numbers satisfy the following system of linear equations:

$$x_k = \sum_{a=1}^{k} \begin{Bmatrix} k \\ a \end{Bmatrix} y_a \quad (k = 1,\ldots,r) ,$$

i.e. $(x_1,\ldots,x_k)^T = S \cdot (y_1,\ldots,y_k)^T$ where $S = (\left\{ {k \atop a} \right\})_{k,a=1,\ldots,r}$. Hence

$$(y_1,\ldots,y_k)^T = S^{-1} \cdot (x_1,\ldots,x_k)^T .$$

Recall that $S^{-1} = (\mathrm{s}(k,a))_{k,a=1,\ldots,r}$ (see e.g. Cameron (1996)), hence

$$p(X,r) = \frac{1}{|V|^{\underline{r}}} \sum_{k=1}^{r} \mathrm{s}(r,k) \mathbf{E}\left(|X|^k\right) .$$

The last equality follows from the formula $x^{\underline{r}} = \sum_{k=1}^{r} \mathrm{s}(r,k) x^k$. □

A direct application of the last theorem gives the following useful corollaries:

**Corollary 2** *Suppose that a random variable $X$ is $r$-homogeneous and that there exists $a$ such that $|X| \equiv a$. Then for each $b \le r$ we have*

$$p(X,b) = \frac{(a)^{\underline{b}}}{|V|^{\underline{b}}} .$$

**Corollary 3** *Suppose that $X$ is $1$-homogeneous random variable with values in the power set $\mathrm{P}(V)$. Let $a \in V$. Then*

$$\Pr[a \in X] = \frac{\mathbf{E}\left(|X|\right)}{|V|} .$$

**Corollary 4** *Suppose that $X$ is a $2$-homogeneous random variable with values in the power set $\mathrm{P}(V)$. Let $a, b \in V$ and $a \neq b$. Then*

$$\Pr[\{a,b\} \subseteq X] = \frac{\mathbf{E}\left(|X|^2\right) - \mathbf{E}\left(|X|\right)}{(|V|-1)|V|} .$$

# 3 Applications - I

Let us consider a 2-dimensional vector space $V$ of cardinality $p^2$, where $p$ is prime bigger than 2. Let $X_V$ be a random variable which randomly and uniformly chooses subsets of $V$ of cardinality $p$ and let $L_V$ be a random variable which randomly and uniformly chooses lines in $V$.

The group $\mathrm{Sym}(V)$ of all permutations of $V$ acts $r$-transitively for all $r \le p^2$ and the random variable $X_V$ is $\mathrm{Sym}(V)$-invariant. On the other hand the group Aff(V) of all invertible affine transformations acts 2-transitively on $V$ and $L_V$ is $\mathrm{Aff}(V)$-invariant. Notice that $|X_V| = |L_V| = p$, so from Corollary 2 we deduce that for each of the two different points $a, b$ from $V$ we have

1. $\Pr[a \in X_V] = \Pr[a \in L_V] = \frac{p^{\underline{1}}}{(p^2)^{\underline{1}}} = \frac{1}{p}$
2. $\Pr[\{a,b\} \subseteq X_V] = \Pr[\{a,b\} \subseteq L_V] = \frac{p^{\underline{2}}}{(p^2)^{\underline{2}}} = \frac{1}{p(p+1)}$

It can be easily checked that if $a,b,c$ are pairwise different then $\Pr[\{a,b,c\} \subseteq X_V] = \frac{-2+p}{p(-2-2p+p^2+p^3)}$ and $\Pr[\{a,b,c\} \subseteq X_V] \in \{0, \frac{1}{p(p+1)}\}$. The difference between random subsets and random lines lies, among others, in the fact that there are non-collinear triples on the plane.

Let us fix a prime number $p$ and let us consider the projective plane $H = \mathrm{PG}(2,p)$ over the field $\mathbf{F}_p$ (see e.g. Hirschfeld (1979), see also Fig. 1). Then $|H| = p^2+p+1$. Let $R_H$ be a random subset of $H$ of cardinality $p+1$ and let $P_H$ be a random line in $V$. Let us recall that each line in $H$ has $p+1$ points. The projective linear group $\mathrm{PGL}(H)$ acts 2-transitively on $H$. Therefore, as before, both random variables $R_H$ and $P_H$ are 2-homogeneous, so $p(R_H,1) = p(P_H,1) = \frac{1+p}{1+p+p^2}$ and $p(R_H,2) = p(P_H,2) = \frac{1}{1+p+p^2}$

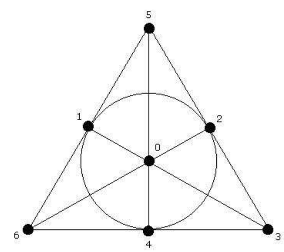


**Fig. 1:** The smallest possible projective plane $\mathrm{PG}(2,2)$ (Fano plane). It has 7 points and 7 lines.

# 4 Sums of Independent Invariant Random Variables

Let us fix a space $V$ and $r$-homogeneous random variable $X$ with values in $\mathrm{P}(V)$. Let $X_1, \ldots, X_k$ be independent copies of $X$ and $X^{(k)} = X_1 \cap \ldots \cap X_k$. From Theorem 2 we deduce that $X^{(k)}$ is $r$-homogeneous and

$$p(X^{(k)}, r) = (p(X,r))^k \ .$$

Let $F_k = |X^{(k)}|$. Then

$$(F_k)^r = \left( \sum_{x \in V} \mathbf{1}_{x \in X^{(k)}} \right)^r = \sum_{(x_1,\ldots,x_r) \in V^r} \mathbf{1}_{x_1 \in X^{(k)} \wedge \ldots \wedge x_r \in X^{(k)}} \ ,$$

therefore

$$\mathbf{E}\left((F_k)^r\right) = \sum_{(x_1,\ldots,x_r)\in V^r} \Pr[\{x_1,\ldots,x_r\} \subseteq X^{(k)}] = \sum_{(x_1,\ldots,x_r)\in V^r} (p(X,r))^k =$$
$$\sum_{l=1}^{r} \binom{|V|}{l} \left\{ {r \atop l} \right\} l! \cdot (p(X,l))^k = \sum_{l=1}^{r} \left\{ {r \atop l} \right\} \cdot |V|^{\underline{l}} \cdot (p(X,l))^k \ .$$

Using Corollary 3 we deduce that the number $\mathbf{E}\left((F_k)^r\right)$ depends only on numbers $r$, $|V|$, $\mathbf{E}(|X|)$, $\mathbf{E}\left(|X|^2\right), \ldots, \mathbf{E}(|X|^r)$ and $k$.

**Lemma 4** *For each $r \geq 1$ there is a function $\psi_r$ with the following property: if $X$ is an $r$-homogeneous random variable with values in $\mathrm{P}(V)$, $X_1, \ldots, X_k$ are independent copies of $X$ and $s_k = |X_1 \cup \ldots \cup X_k|$ then*

$$\mathbf{E}\left((s_k)^r\right) = \psi_r(k, |V|, (\mathbf{E}\left(|X|^j\right))_{j=1\ldots r}) \; .$$

**Proof:** Let $Y = X^c$ and $Y_i = V \setminus X_i$. Then, according to Theorem 2, $Y$ is $r$-homogeneous and $(Y_i)_{i=1,\ldots,k}$ are independent copies of $Y$. We put $g_k = |\bigcap_{i=1}^k Y_i|$ and observe that $s_k = |V| - g_k$. Next we have

$$(s_k)^r = (|V| - g_k)^r = \sum_{b=0}^{r} \binom{r}{b} (-1)^b (g_k)^b |V|^{r-b} \; .$$

The above discussion implies that for each $b \leq r$ the number $\mathbf{E}\left((g_k)^b\right)$ depends only on numbers $b$, $|V|$, $\mathbf{E}\left(|X|\right)$, $\mathbf{E}\left(|X|^2\right), \ldots, \mathbf{E}\left(|X|^r\right)$ and $k$. So the same holds for $\mathbf{E}\left((s_k)^r\right)$. $\square$

Let $X$ be a $r$-homogeneous set valued random variable with values in $\mathrm{P}(V)$ such that $|X| \equiv a$. Let $X_1 \ldots X_k$ be independent copies of $X$ and $S_k = X_1 \cup \ldots \cup X_k$. Let $s_k = |S_k|$ and $f_k = |V| - s_k$. The last theorem implies that $\mathbf{E}\left((s_k)^r\right)$ depends only on $|V|$, $r$, $k$ and $a$. Therefore in order to find $\mathbf{E}\left((s_k)^r\right)$ we may consider any random variable with a similar property. So let us consider the random variable $Y$ uniformly distributed over all subsets of $V$ of cardinality $a$. Then $Y$ is $t$-homogeneous for all $t \leq |V|$. Let $Y_1, \ldots, Y_k$ be independent copies of $Y$. Let $Y^{(k)} = Y_1 \cup \ldots \cup Y_k$ and $F^{(k)} = V \setminus Y^{(k)}$. Let $n = |V|$. It can see that $p(F^{(k)}, b) = ((n-b)^{\underline{a}} / n^{\underline{a}})^k$, so from Corollary 1 we deduce that

$$\mathbf{E}\left((f_k)^r\right) = \mathbf{E}\left(|F^{(k)}|^r\right) = \sum_{c=1}^{r} \left\{ {r \atop c} \right\} n^{\underline{c}} \left( \frac{(n-c)^{\underline{a}}}{n^{\underline{a}}} \right)^k$$

From this formula we can calculate all moments of $s_k$ of any order less or equal to $r$.

# 5 Applications - II

Let us fix once again a 2-dimensional vector space $V$ over the field $\mathbf{F}_p$. We consider two processes. In the first one we randomly and independently choose $k$ times subsets $X_1, \ldots, X_k$ of subsets of cardinality $p$. In the second one we randomly and independently choose $k$ times lines $L_1, \ldots, L_k$ in $V$. We finally put $X^{(k)} = X_1 \cup \ldots \cup X_k$ and $L^{(k)} = L_1 \cup \ldots \cup L_k$. We are interested in probabilistic properties of random variables $|X^{(k)}|$ and $|L^{(k)}|$. Let $X$ be a random variable uniformly distributed over all subsets of $V$ of cardinality $p$ and let $L$ be a random variable uniformly distributed over all lines in $V$. From discussion from Sec. 3 we know that both variables $X$ and $L$ are 2-homogeneous, therefore we may apply Theorem 4 and deduce that $\mathbf{E}\left(|X^{(k)}|\right) = \mathbf{E}\left(|L^{(k)}|\right)$ and $\mathbf{E}\left(|X^{(k)}|^2\right) = \mathbf{E}\left(|L^{(k)}|^2\right)$, i.e. that the first two moments of variables $|X^{(k)}|$ and $|L^{(k)}|$ are the same.

Almost the same discussion applies to projective spaces. Namely, let us fix a prime $p$ and consider the projective plane $PG(2,p)$. Let $X$ be a random variable uniformly distributed over all subsets of $PG(2,p)$ of cardinality $p+1$ and let $L$ be a random variable uniformly distributed over all lines in $PG(2,p)$. The group $\mathrm{PLG}(2,p)$ acts 2-transitively on $PG(2,p)$ and both variables $X$ and $L$ are $PG(2,p)$-invariant. Hence $X$ and $L$ are 2-homogeneous, so we may apply Theorem 4 and deduce that the first two moments of variables $|X^{(k)}|$ and $|L^{(k)}|$ are equal.

Let us consider the Steiner system $S(5,6,12)$ on the set $V$ of cardinality 12. The Mathieu group $M_{12}$ (see e.g. Rotman (1995)) of all permutations of $V$ preserving blocks of this system acts 5-transitively on $V$. Let us consider two processes: throwing random subsets of V of cardinality 6 and throwing blocks from this system. These two models of throwing sets have the same properties up to the fifth moment of their counting functions.

# 6 Beyond Homogeneity

Let $n \geq 3$, let us fix the cyclic group $\mathbb{C}_n$ and let us consider two processes. In the first one we choose randomly and independently sets $X_1, \ldots, X_k$ from $[\mathbb{C}_n]^2$ and in the second one we choose subsets $Y_1, \ldots, Y_k$ of the form $\{a, a+1\}$ (mod n). We put $X^{(k)} = \mathbb{C}_n \setminus (X_1 \cup \ldots \cup X_k)$, $Y^{(k)} = \mathbb{C}_n \setminus (Y_1 \cup \ldots \cup Y_k)$ and we want to calculate first two moments of variables $|X^{(k)}|$ and $|Y^{(k)}|$. The group $\mathbb{C}_n$ acts transitively on itself and both random variables are $\mathbb{C}_n$-invariant, so both variables are 1-homogeneous, so $\mathbf{E}\left(|X^{(k)}|\right) = \mathbf{E}\left(|Y^{(k)}|\right)$. The first model is well-known and is easy to calculate; we have $\mathbf{E}\left(|X^{(k)}|\right) = n(1-\frac{2}{n})^k$. The second moment of $|Y^{(k)}|$ can be calculated in the following way:

$$|Y^{(k)}|^2 = \left(\sum_{i=0}^{n-1} \mathbf{1}_{i \in Y^{(k)}}\right)^2 = \sum_{i=0}^{n-1} \mathbf{1}_{i \in Y^{(k)}} + \sum_{i \neq j} \mathbf{1}_{\{i,j\} \subseteq Y^{(k)}} \ ,$$

so

$$\mathbf{E}\left(|Y^{(k)}|^2\right) = \mathbf{E}\left(|Y^{(k)}|\right) + \sum_{i \neq j} \Pr[\{i,j\} \subseteq Y^{(k)}] = \mathbf{E}\left(|Y^{(k)}|\right) + \sum_{i \neq j} \Pr[\{i,j\} \subseteq Y_1^c]^k \ .$$

We must calculate the second term manually. Let us recall that (Lemma 1) if $a \sim_G b$ and a random variable $X$ is $G$-invariant, then $\Pr[a \subseteq X] = \Pr[b \subseteq X]$. So, in our case it is enough to consider pairs of the form $(1,j)$, where $j \neq 1$. Note that if $j \notin \{0,2\}$ then $\Pr[\{1,j\} \subseteq Y_1^c] = 0$ and that $\Pr[\{1,2\} \subseteq Y_1^c] = (n-3)/n$, so finally we get

$$\mathbf{E}\left(|Y^{(k)}|^2\right) = n\left(1 - \frac{2}{n}\right)^k + 2n\left(1 - \frac{3}{n}\right)^k .$$

This observation can be generalized. Suppose that we are analyzing a set valued $G$-invariant random variable and suppose that $G$ acts $(r-1)$-transitively and that we know all moments $\mathbf{E}\left(|X|^i\right)$ for $i < r$. Then

$$\mathbf{E}\left(|X|^r\right) = \mathbf{E}\left(\left(\sum_{x \in V} \mathbf{1}_{x \in X}\right)^r\right) = f(\mathbf{E}\left(|X|\right), \ldots, \mathbf{E}\left(|X|^{r-1}\right)) + \\ \sum \{\Pr[\{a_1, \ldots, a_r\} \subseteq X] : (a_1, \ldots, a_r) \in \mathrm{Diff}(V, r)\}$$

where

$$\mathrm{Diff}(V, r) = \{(a_1, \ldots, a_r) \in V^r : \bigwedge_{i \neq j} (a_i \neq a_j)\}$$

and $f$ is a function easy to calculate(i). Notice that the relation $\sim_G$ splits the set $\mathrm{Diff}(V,r)$ into disjoint classes and if $a \sim_Q b$ then $\Pr[a \subseteq X] = \Pr[b \subseteq X]$. In typical cases the are only few equivalence classes, so the calculations are easy.

Let us consider, for example, the process of throwing random lines on finite plane $(\mathbf{F}_p)^2$. Note that there are $p^2$ points in this space. Let $L_i$ be the $i$-th line chosen and $C^{(k)}$ be the set of points not covered by any of first $k$ lines. In order to calculate the first two moments we may replace lines by subsets of cardinality $p$ and we get $\mathbf{E}\left(|C^{(k)}|\right) = p^2\left(1-\frac{1}{p}\right)^k$ and $\mathbf{E}\left(|C^{k)}|^2\right) = p^2\left(1-\frac{1}{p}\right)^k + p^2(p^2-1)\left(1-\frac{2p+1}{p(p+1)}\right)^k$. The first interesting moment is the third one. Namely, we have

$$\mathbf{E}\left(|C_k|^3\right) = \sum_{x,y,z}\Pr[\{x,y,z\} \subseteq L_1^c \cap \ldots \cap L_k^c] = \sum_{x,y,z}\Pr[\{x,y,z\} \subseteq L_1^c]^k$$

$$\sum_x \Pr[\{x\} \subseteq L_1^c]^k + 3\sum_{x \neq y}\Pr[\{x,y\} \subseteq L_1^c]^k + \sum_{(x,y,z)\in \mathrm{Diff}(V,3)}\Pr[\{x,y,z\} \subset L_1^c]^k$$

In our case there are only two equivalence classes: collinear triples and non-collinear triples. There are no lines containing non-collinear triples, and for each collinear triple there is only one line containing it; there are $p^2(p^2-1)(p-2)$ collinear triples; there are $p^2+p$ lines; so the last factor reduces to $p^2(p^2-1)(p-2)(1-\frac{1}{p(p+1)})^k$. After some simplifications we get the following formula

$$p^2\left(1-\frac{2}{p}\right)^k + 3p^2(p^2-1)\left(1-\frac{1+2p}{p+p^2}\right)^k + p^2(1+p+p^2+p^3)\left(1-\frac{1}{p+p^2}\right)^k$$

for $\mathbf{E}\left(|C^{(k)}|^3\right)$. A similar calculation for throwing random sets from $[(\mathbf{F}_p)^2]^p$ gives us the formula for the third moment $\mathbf{E}\left(|G^{(k)}|^3\right)$ of the number of non-marked points after throwing $k$ sets:

$$p^2\left(1-\frac{2}{p}\right)^k + 3p^2(p^2-1)\left(1-\frac{1+2p}{p+p^2}\right)^k + p^2(p^2-1)(p^2-2)\left(1-\frac{-2-3p+3p^2}{p(p^2-2)}\right)^k .$$

In order to satisfy our curiosity we compared third moment of random variables $|C^{(k)}|$ and $|G^{(k)}|$ for the plane $\mathbf{F}_{29}^2$ (see Fig. 2). Clearly both moments tend to 0 when $k$ tends to infinity but we see that the rates of convergence are very different.

# 7 Conclusion

Throwing random sets of points and throwing random lines are very similar, at least up to first two moments of their counting functions. This holds both for classical finite planes and for finite projective planes. This phenomenon arises from the fact that both classes of thrown objects are invariant under some groups acting 2-transitively on the underlying space.

Our method reduces some probabilistic considerations to investigation of a simpler process of throwing random sets. For the analysis of the last one we can use standard combinatorial analysis tools.

# Acknowledgements

We would like to thank referees for their helpful comments.

(i) More precisely: $f(x_1,\ldots,x_{r-1}) = -\sum_{k=1}^{r-1} \mathrm{s}(r,k)\cdot x_k$

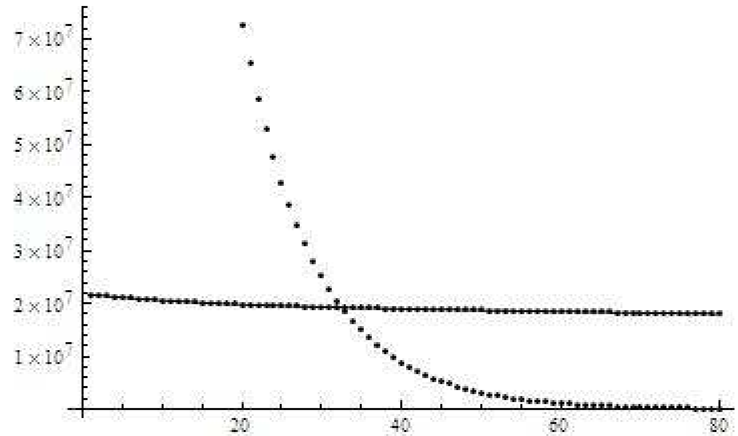


**Fig. 2:** Third moments of random variables $|C^{(k)}|$ (joined dots) and $|G^{(k)}|$ for $\mathbf{F}_{29}^2$.

# References

P. J. Cameron. *Combinatorics: Topics, Techniques, Algorithms*. Cambridge University Press, 1996.

S. A. Camtepe and B. Yener. Combinatorial design of key distribution mechanisms for wireless sensor networks. *IEEE/ACM Transactions on Networking*, 15(2), 2007.

N. G. deBruijn. Polya's theory of counting. In E. F. Beckenbach, editor, *Applied Combinatorial Mathematics*. Wiley, New York, 1964.

L. Eschenauer and V. D. Gligor. A key management scheme for distributed sensor networks. In *9th ACM Conference on Computer and Communication Security (CCS'2002)*, pages 41–47. ACM, 2002.

P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, New York, NY, USA, 2009. ISBN 0521898064, 9780521898065.

P. Flajolet, D. Gardy, and L. Thimonier. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Appl. Math.*, 39(3):207–229, 1992. ISSN 0166-218X. doi: http://dx.doi.org/10.1016/0166-218X(92)90177-C.

D. Gardy. Occupancy urn models in the analysis of algorithms. Preprint, 1998.

J. Hirschfeld. *Projective Geometries over Finite Fields*. Oxford Mathematical Monographs. Clarendon Press, Oxford, 1979.

J. J. Rotman. *An Introduction to the Theory of Groups*. Graduate Texts in Mathematics, vol. 148. Springer-Verlag, 1995.