Convolutions, variance and covariance

G. Palafox

November 15, 2020

1 Convolutions

A convolution is, in the more general sense, an operator between two functions which produces a third function. In this section, an application of convolutions of probability functions on finite groups is explored. In particular, convolutions are used to obtain the distribution of a random walk on a finite group. After the basic theory is presented, a practical example is given. Most of what is discussed in this section, and further topics regarding random walks on finite groups, can be found in the books of Steinberg [2012] or Diaconis [1988].

1.1 Random walks on finite groups

Let G be a finite group [Herstein, 1975]. A probability on G is a function $P : G \longrightarrow [0, 1]$ such that

$$\sum_{g\in G} P(g) = 1. \tag{1}$$

For a subset $A \subseteq G$, one defines $P(A) := \sum_{g\in A} P(g)$. Now, suppose P, Q are probabilities on G, and $X \sim P$, $Y \sim Q$ are chosen independently at random. What is the probability of $XY = g$ for some g? If $Y = h$, it must be that $X = gh^{-1}$ for XY to occur. Because of independence, the probability of this happening is $P(gh^{-1})Q(h)$. Summing the probabilities over all possible h gives us the probability of $XY = g$ equal to

$$\sum_{h\in G} P(gh^{-1})Q(h) = P * Q(g), \tag{2}$$

where $*$ denotes the convolution operator. Thus, if X, Y are independent and $X \sim P, Y \sim Q$, it follows that $XY \sim P * Q$. This in turn can be used to model a random walk on G as follows. Starting at the identity e of G, a walker chooses $X_1 \in G$ at random according to a probability P, and moves to X_1. Then, the walker chooses X_2 according to P and moves to X_2X_1. Following this process, the walker is choosing a sequence of independent, identically distributed random variables $X_1, X_2, \dots$ with common distribution P, landing at $X_kX_{k-1} \cdots X_1$ at step k. Let

$$\delta_g(h) = \begin{cases} 1 & \text{if } h = g; \\ 0 & \text{if } h \neq g. \end{cases} \tag{3}$$

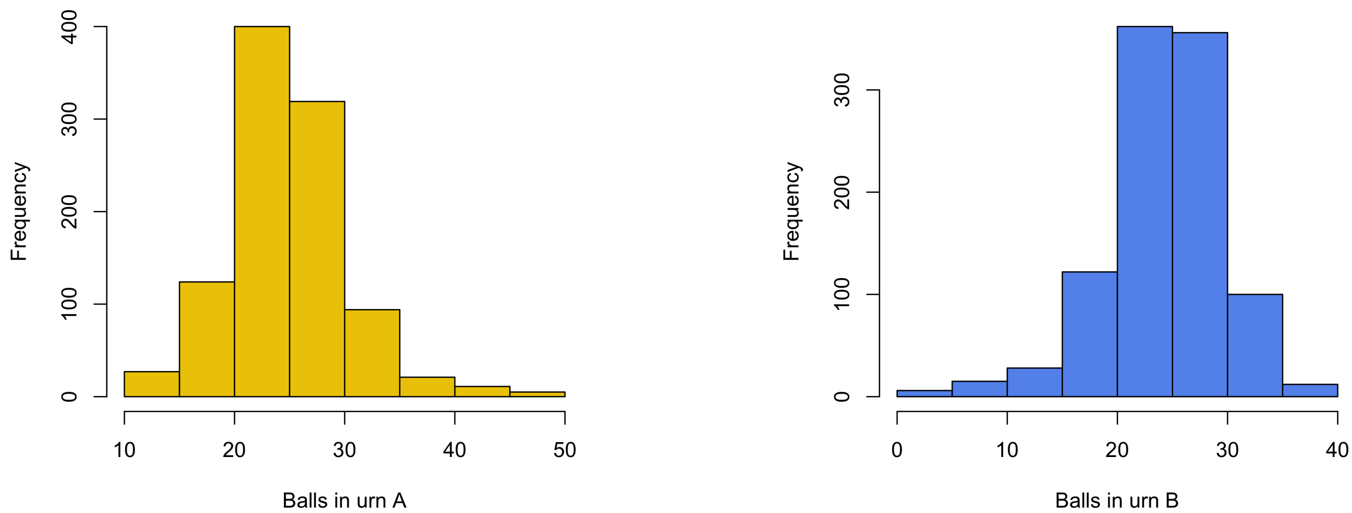
With this notation, one can let $Y_0 \sim \delta_e$ (so $Y_0 = e$ with probability 1), and $Y_k = X_kY_{k-1}$ for $k \geq 1$. The random variable Y_k gives the position of the walker in the k-th step, and by the preceding discussion, $Y_k \sim P^{*k}$, where P^{*k} denotes the convolution of P with itself k times.

1.1.1 Ehrenfest's urn

For a concrete example, the following is given. Suppose there are two urns, A and B, and n balls. At time 0, all balls are in urn A. At each step, one ball is chosen uniformly at random, and moved to the other urn. Denote by $\mathbb{Z}/2\mathbb{Z}$ the group of integers modulo 2. The state at step t can be encoded in a vector $v = (c_1, \dots, c_n) \in (\mathbb{Z}/2\mathbb{Z})^n$, where $c_i = 0$ if and only if ball i is in urn A. Denote by e_i the element of $(\mathbb{Z}/2\mathbb{Z})^n$ having a 1 in its i-th coordinate, and 0 in the rest. If at an arbitrary time the state of the process is encoded by a vector v as previously described, changing ball i to a different urn consists of adding e_i to the state vector v. That is, the state changes to $v + e_i$. Thus, starting at identity $e = (0, 0, \dots, 0)$ (all balls in urn A), the process of interchanging balls between the urns corresponds to a random walk on $(\mathbb{Z}/2\mathbb{Z})^n$ driven by probability

$$P(g) = \begin{cases} 1/n & \text{if } g \in \{e_1, e_2, \dots, e_n\}; \\ 0 & \text{else.} \end{cases} \tag{4}$$

A thousand steps of this process, with fifty balls total, was simulated with R [R Core Team, 2020] on a Jupyter notebook [Kluyver et al., 2016]. The average amount of balls in urns A and B where 25.25 and 24.75 respectively, which suggests



(a) Distribution of balls in urn A.

(b) Distribution of balls in urn B.

Figure 1: Histogram for ball distribution in a thousand steps of an Erhenfest process with fifty balls.

the urns are balanced on average. Histograms showing the distribution of the balls in the urns are shown in Figure 1. Additionally, using ImageMagick [The ImageMagick Development Team], a GIF animation[1] was created for twenty steps of this process with six balls total.

2 Goodness of fit

In this section, we perform a goodness of fit test to determine whether the degree distribution of Pennsylvania's road network [Leskovec and Krevl, 2014] follows a Poisson distribution. The network has $|V| = 1,088,092$ nodes, and an average degree of $\lambda_d = 2.834$. A χ^2 goodness of fit test is performed to see if the road "is random". Given the large size of the road, if nodes were connected to each other at random, a Poisson degree distribution would be expected [Newman, 2018]. To perform the test, first the nodes of degree $k = 1, 2, \ldots, 14$ are counted. Then, the expected Poisson distribution is obtained computationally, generating $|V|$ numbers following a $\mathrm{Poiss}(\lambda_d)$ distribution. The expected and observed degrees are displayed in Table 1. Writing O_k for the observed nodes with degree k, and E_k for the expected nodes with degree k, a statistic

$$\sum_{k=1}^{14} = \frac{(O_i - E_i)^2}{E_i} = 672,672.24. \tag{5}$$

is obtained. The corresponding p-value is 0, so it is concluded that the degree distribution is not Poisson.

3 Theoretical results

To conclude this work, two theorems concerning variance and covariance are presented. Theorem 1 was tested computationally for random integers a, b, c, d between one and two hundred, and a pair of one thousand number pseudo-random vectors X, Y, where $X \sim \mathrm{Unif}(0,1), Y \sim \mathrm{Exp}(0.5)$; $X \sim N(0,1), Y \sim \mathrm{Exp}(0.5)$; and $X \sim \mathrm{Geom}(0.33), Y \sim \mathrm{Poiss}(1)$. In a thousand repetitions, the equality always held. Theorem 2 was also verified computationally with the same pair of vectors X, Y, and it too held true in each of the one thousand repetitions.

Theorem 1. *For constants a, b, c, d and random variables X, Y,*

$$\mathrm{Cov}\,(aX + b, cY + d) = ac\,\mathrm{Cov}\,(X, Y)\,. \tag{6}$$

[1] The notebook with the code for all the simulations in this report, as well the GIF animation, can be found in the Github Repository: https://github.com/palafox794/AppliedProbabilityModels/tree/master/Assignment11.

Table 1: Observed and expected degree of nodes.

Degree	Expected	Observed
1	245,208	188,317
2	256,577	90,740
3	243,159	532,686
4	171,512	267,256
5	97,772	7,759
6	45,727	1,237
7	18,825	80
8	6,526	13
9	2,039	4
10	550	0
11	147	0
12	41	0
13	7	0
14	2	0

Proof. By definition, it is seen that

$$\begin{aligned}
\mathrm{Cov}\,(aX+b, cY+d) &= \mathbb{E}[(aX+b)(cY+d)] - \mathbb{E}[aX+b]\,\mathbb{E}[cY+d] && (7)\\
&= \mathbb{E}[acXY + adX + bcY + bd] - (a\,\mathbb{E}[x] + b)(c\,\mathbb{E}[y] + d) && (8)\\
&= ac\,\mathbb{E}[XY] + ad\,\mathbb{E}[X] + bc\,\mathbb{E}[Y] + bd - ac\,\mathbb{E}[X]\,\mathbb{E}[Y] - ad\,\mathbb{E}[X] - bc\,\mathbb{E}[Y] - bd && (9)\\
&= ac(\,\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]) && (10)\\
&= ac\,\mathrm{Cov}\,(X,Y)\,. && (11)
\end{aligned}$$

□

Theorem 2. *For random variables X, Y,*

$$\mathrm{Var}\,(X+Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) + 2\,\mathrm{Cov}\,(X,Y)\,. \tag{12}$$

Proof. Since $\mathrm{Var}\,(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, it follows that

$$\begin{aligned}
\mathrm{Var}\,(X+Y) &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 && (13)\\
&= \mathbb{E}[X^2 + 2XY + Y^2] - (\,\mathbb{E}[X] + \mathbb{E}[Y])^2 && (14)\\
&= \mathbb{E}[X^2] + 2\,\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\,\mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[Y]^2 && (15)\\
&= \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) + 2(\,\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]) && (16)\\
&= \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) + 2\,\mathrm{Cov}\,(X,Y)\,. && (17)
\end{aligned}$$

□

References

P. Diaconis. *Group representations in probability and statistics.* Lecture notes-monograph series. Institute of Mathematical Statistics, 1988. ISBN 978-0-940600-14-0.

I. N. Herstein. *Topics in algebra.* Xerox College Pub, 2nd edition, 1975. ISBN 978-0-536-01090-2.

T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*, page 87. IOS Press, 2016.

J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

M. Newman. *Networks*, volume 1. Oxford University Press, Oct 2018. ISBN 978-0-19-880509-0. doi: 10.1093/oso/9780198805090.001.0001. URL https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198805090.001.0001/oso-9780198805090.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

B. Steinberg. *Representation theory of finite groups: an introductory approach*. Universitext. Springer, 2012. ISBN 978-1-4614-0775-1.

The ImageMagick Development Team. Imagemagick. URL https://imagemagick.org.