Journal of Classical Analysis
Volume 14, Number 2 (2019), 177–180 doi:10.7153/jca-2019-14-13

COUPON COLLECTOR'S PROBLEM WITH UNLIKE PROBABILITIES

TOSHIO NAKATA

Dedicated to Professor Izumi Kubo
on the occasion of his 80th birthday

Abstract. In this note we study the coupon collector's problem with unlike probabilities using majorization and a Schur concave function.

1. Introduction and results

The coupon collector's problem, an important example in a course on elementary probability, can be described as follows: Suppose that there are t types of coupons, and that each day a collector randomly gets a coupon with probability p_i corresponding to the ith coupon $(1 \leqslant i \leqslant t)$. How many days does the collector need to collect in order to have at least one of each type? There is a lot of literature about this problem (e.g. see [1, I.2 (b.11))], [2, §4], [3, Example II.3.11], [7]).

Let $X_{\mathbf{p}}$ be a random variable representing the number of days until all kinds of coupons have been collected, where $\mathbf{p} = (p_1, \dots, p_t)$ is a probability vector, namely

$$\sum_{i=1}^{t} p_i = 1 \quad \text{and} \quad p_i \geqslant 0.$$

Note that the explicit expectation form is known as

$$E(X_{\mathbf{p}}) = \sum_{q=0}^{t-1} (-1)^{t-1-q} \sum_{|J|=q} \frac{1}{1-P_J}, \tag{1}$$

where $P_J = \sum_{j \in J} p_j$ (see [2, equation (14b)]). Von Schelling [7, §2] writes the distribution in more generality and with a different notation. When p_i is equal to $\frac{1}{t}$, for all $1 \leqslant i \leqslant t$, the expectation is well-known,

$$E\left(X_{\left(\frac{1}{t},\dots,\frac{1}{t}\right)}\right) = t \sum_{i=1}^{t} \frac{1}{i} \sim t \log t.$$

Mathematics subject classification (2010): 60C05.
Keywords and phrases: Coupon collector's problem, majorization, Schur concave function.

Paper JCA-14-13

Moreover, the distribution is also well-known, that is,

$$P\left(X_{\left(\frac{1}{t},\dots,\frac{1}{t}\right)} \leqslant n\right) = \sum_{k=0}^{t} (-1)^k \binom{t}{k} \left(1-\frac{k}{t}\right)^n$$

(see [1, IV.2 equation (2.3)]), which is represented by

$$P\left(X_{\left(\frac{1}{t},\dots,\frac{1}{t}\right)} \leqslant n\right) = \frac{t!\left\{ {n \atop t} \right\}}{t^n}, \tag{2}$$

where $\left\{ {n \atop t} \right\}$ is Stirling numbers of the second kind (see [3, Example II.3.11]).

In this note, we show that $P(X_{\mathbf{p}} \leqslant n)$ is monotone with respect to $\mathbf{p}$ in the sense of *majorization*. The same setting for the Birthday problem was studied by Joag-Dev and Proschan [4].

Before stating the main theorem, we define a few terms and introduce some standard notations (see [6, 8]). For a probability vector $\mathbf{p} = (p_1, \dots, p_t)$, let $p_{[j]}$ be the jth largest value of $\{p_1, \dots, p_t\}$, that is, $p_{[1]} \geqslant p_{[2]} \geqslant \cdots \geqslant p_{[t]}$. A probability vector $\mathbf{p} = (p_1, p_2, \dots, p_t)$ is *majorized* by a probability vector $\mathbf{q} = (q_1, q_2, \dots, q_t)$, or $\mathbf{p} \prec \mathbf{q}$, if

$$\sum_{i=1}^{k} p_{[i]} \leqslant \sum_{j=1}^{k} q_{[j]},$$

for all $1 \leqslant k \leqslant t-1$. By definition it is easy to see that $\left(\frac{1}{t}, \dots, \frac{1}{t}\right) \prec \mathbf{p}$, for all $\mathbf{p}$. The symmetric function $f(\mathbf{p})$ defined on probability vectors is *Schur convex* (resp. *concave*) if $\mathbf{p} \prec \mathbf{q}$ implies $f(\mathbf{p}) \leqslant f(\mathbf{q})$ (resp. $f(\mathbf{p}) \geqslant f(\mathbf{q})$). Under the assumption of symmetry and differentiability of f, a necessary and sufficient condition for $f(p_1, \dots, p_n)$ to be Schur concave is

$$(p_1 - p_2)\left(\frac{\partial f}{\partial p_1} - \frac{\partial f}{\partial p_2}\right) \leqslant 0 \tag{3}$$

(see [6, p. 57]). Under these preliminaries we state the main theorem.

THEOREM 1. *The probability* $P(X_{\mathbf{p}} \leqslant n)$ *is a Schur concave function of* $\mathbf{p}$.

Now a random variable X is *stochastically smaller* than a random variable Y if $P(X > a) \leqslant P(Y > a)$, for all real a (see [5, Chap. IV.1.1]).

COROLLARY 1. *If* $\mathbf{p} \prec \mathbf{q}$ *then* $X_{\mathbf{p}}$ *is stochastically smaller than* $X_{\mathbf{q}}$. *In particular* $X_{\left(\frac{1}{t},\dots,\frac{1}{t}\right)}$ *is stochastically smaller than* $X_{\mathbf{p}}$, *for all* $\mathbf{p}$.

Proof. By Theorem 1, if $\mathbf{p} \prec \mathbf{q}$ we have $P(X_{\mathbf{p}} > n) \leqslant P(X_{\mathbf{q}} > n)$, for all n. Since $\left(\frac{1}{t}, \dots, \frac{1}{t}\right) \prec \mathbf{p}$, for all $\mathbf{p}$, we obtain the desired result. □

Thus *it is harder to collect all kinds of coupons if there is some bias for the probability of appearance of coupons.*

COROLLARY 2. *The expectation* $E(X_{\mathbf{p}})$ *is a Schur convex function of* $\mathbf{p}$. *In particular* $E\left(X_{\left(\frac{1}{t},\dots,\frac{1}{t}\right)}\right) \leqslant E(X_{\mathbf{p}})$, *for all* $\mathbf{p}$.

Proof. By Corollary 1, if $\mathbf{p} \prec \mathbf{q}$ we have

$$E(X_{\mathbf{p}}) = \sum_{n=0}^{\infty} P(X_{\mathbf{p}} > n) \leqslant \sum_{n=0}^{\infty} P(X_{\mathbf{q}} > n) = E(X_{\mathbf{q}}),$$

which implies Schur convexity. □

Note that by virtue of Corollary 1 we can prove Corollary 2 without applying directly the criterion equation (3).

2. Proof of Theorem 1

For convenience, letting $f(\mathbf{p}) = P(X_{\mathbf{p}} \leqslant n)$ we have

$$f(\mathbf{p}) = \sum_{\{i_1,\ldots,i_n\}=\{1,\ldots,t\}} p_{i_1}\cdots p_{i_n},$$

where the summation runs through all $(i_1,\ldots,i_n)$ such that there exists a surjection $g:\{1,\ldots,n\}\to\{1,\ldots,t\}$ satisfying $g(k)=i_k$, for $1\leqslant k\leqslant n$. Accordingly we have

$$f(\mathbf{p}) = \sum_{(l_1,\ldots,l_t)} \binom{n}{l_1;\cdots;l_t} p_1^{l_1}p_2^{l_2}\cdots p_t^{l_t} = n! \sum_{(l_1,\ldots,l_t)} \frac{p_1^{l_1}}{l_1!}\cdot\frac{p_2^{l_2}}{l_2!}\cdots\frac{p_t^{l_t}}{l_t!},$$

the sum being taken over $l_k \geqslant 1$ and $\sum_{k=1}^{t} l_k = n$. Note that because of

$$\sum_{(l_1,\ldots,l_t)} \binom{n}{l_1;\cdots;l_t} = t!\left\{ n \atop t \right\},$$

we can confirm equation (2) if $p_1=\cdots=p_t=1/t$ (see [3, II.3.4]). Hence we have

$$f(\mathbf{p}) = n![z^n]\prod_{i=1}^{t}(e^{zp_i}-1),$$

where $[z^n]A(z)$ denotes the coefficient of z^n for $A(z)$. Because f is symmetric and differentiable with respect to $p_1,\ldots,p_t$, we check equation (3) to show the Schur concavity of $f(\mathbf{p})$. Since

$$\frac{\partial f}{\partial p_1} = n![z^n]ze^{zp_1}\prod_{i=2}^{t}(e^{zp_i}-1),$$

we have

$$\frac{\partial f}{\partial p_1} - \frac{\partial f}{\partial p_2} = n![z^{n-1}](e^{zp_2}-e^{zp_1})\prod_{i=3}^{t}(e^{zp_i}-1) = (p_2-p_1)h(\mathbf{p}),$$

where $h(\mathbf{p})$ is some positive function. Hence

$$(p_1-p_2)\left(\frac{\partial f}{\partial p_1} - \frac{\partial f}{\partial p_2}\right) = -(p_1-p_2)^2 h(\mathbf{p}) \leqslant 0$$

yields that $f(\mathbf{p})$ is Schur concave. □

Acknowledgement. The author would like to thank the editor of JCA and the reviewer for valuable suggestion and comments.

REFERENCES

[1] W. FELLER, *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd ed., Wiley, New York, 1968.
[2] P. FLAJOLET, D. GARDY AND L. THIMONIER, *Birthday paradox, coupon collectors, caching algorithms and self-organizing search*, Discr. Appl. Math. **39**, (1992), 207–229.
[3] P. FLAJOLET AND R. SEDGEWICK, *Analytic Combinatorics*, Cambridge University Press, 2008.
[4] K. JOAG-DEV AND F. PROSCHAN, *Birthday Problem with Unlike Probabilities*, Amer. Math. Monthly, **99**, (1992), 10–12.
[5] T. LINDVALL, *Lectures on the coupling method*, Dover Publications, Inc., Mineola, NY, 2002.
[6] A. MARSHALL AND I. OLKIN, *Inequalities: Theory of majorization and its applications*, Academic Press, New York-London, 1979.
[7] H. VON SCHELLING *Coupon collecting for unequal probabilities*, Amer. Math. Monthly, **61**, (1954), 306–311.
[8] J. STEELE, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge UP, 2004.

(Received March 26, 2019)

Toshio Nakata
Department of Mathematics
University of Teacher Education Fukuoka
Akama-Bunkyomachi, Munakata, Fukuoka, 811-4192, Japan
e-mail: nakata@fukuoka-edu.ac.jp