J. Appl. Prob. **54,** 655–656 (2017)
*doi:*10.1017/jpr.2017.23

LETTER TO THE EDITOR

Dear Editor,

Minimum variance in the coupon collector's problem

We show that in the classical coupon collector's problem the number of coupons needed to complete the set has minimum variance when the drawing probabilities are equal. This solves a problem of Doumas and Papanicolaou (2012).

Consider the classical coupon collector's problem (Ross (2010)) with N different types of items. Items are drawn sequentially and each is of type k with probability p_k, $k = 1, \dots, N$, independently of previous items. Let $\boldsymbol{p} = (p_1, \dots, p_N)$ and let $T_N(\boldsymbol{p})$ denote the number of draws until we obtain at least one item of each type. While investigating the asymptotic behavior of the moments of T_N for general drawing probabilities, Doumas and Papanicolaou (2012) made the conjecture that for fixed N, the variance of T_N is minimized when $p_1 = \cdots = p_N = 1/N$. Here we prove this conjecture, which has attracted some attention (see Sendov and Shan (2015)).

It is well known that $T_N(\boldsymbol{p}) \le_{\mathrm{st}} T_N(\tilde{\boldsymbol{p}})$ if $\boldsymbol{p} \prec \tilde{\boldsymbol{p}}$, where '$\prec$' denotes majorization (see Marshall *et al.* (2009)). In other words, T_N becomes stochastically smaller when the drawing probabilities become more uniform. It follows that the mean and the tail probabilities of T_N are Schur-convex in $\boldsymbol{p}$ and are minimized when $\boldsymbol{p}$ is uniform. To deal with the variance, it is helpful to consider a more general problem where there are $N+1$ types of items, with drawing probabilities $(p_0, p_1, \dots, p_N)$ such that $\sum_{k=0}^{N} p_k = 1$. We are concerned, however, with only obtaining a complete set of N types, excluding the null type whose drawing probability is p_0 (see Anceaume *et al.* (2015) for related results on this problem). Let $T_N(p_0, \boldsymbol{p})$ denote the number of draws until we obtain at least one of each nonnull type.

Theorem 1. *It holds that* $\operatorname{var}(T_N(p_0, \boldsymbol{p}))$ *is minimized with* $\boldsymbol{p} = ((1-p_0)/N, \dots, (1-p_0)/N)$ *for fixed* $N = 1, 2, \dots$ *and* $p_0 \in [0, 1)$.

The conjecture of Doumas and Papanicolaou (2012) corresponds to the $p_0 = 0$ case. If the drawing probabilities for nonnull types are equal, the variance can be easily computed, that is,

$$\operatorname{var}\biggl(T_N\biggl(p_0, \frac{(1-p_0)\mathbf{1}_N}{N}\biggr)\biggr) = \frac{\alpha_N p_0 + \beta_N}{(1-p_0)^2}, \tag{1}$$

where

$$\mathbf{1}_N = (1, \dots, 1), \qquad \alpha_N = \sum_{j=1}^{N} \frac{N}{j}, \qquad \beta_N = \sum_{j=1}^{N-1} \frac{Nj}{(N-j)^2}.$$

Note that α_N and β_N can be written in terms of generalized harmonic numbers $H_N^{(r)} \equiv \sum_{j=1}^{N} 1/j^r$. That is,

$$\alpha_N = N H_N^{(1)}, \qquad \beta_N = N^2 H_N^{(2)} - N H_N^{(1)}.$$

Proof of Theorem 1. Let us use induction on N. The claim is trivially true for $N = 1$. Suppose that $N \ge 2$. Let X_1 denote the type of the first nonnull item and let G_1 denote the

Received 7 June 2016; revision received 17 October 2016.

number of draws until it is obtained. Then G_1 and X_1 are independent, G_1 has a geometric distribution with parameter $1 - p_0$, and $\mathbb{P}(X_1 = k) = p_k/(1 - p_0)$, $k = 1, \dots, N$. Moreover, G_1 and $T_N(p_0, \boldsymbol{p}) - G_1$ are independent and, conditional on $X_1 = k$, $T_N(p_0, \boldsymbol{p}) - G_1$ has the same distribution as $T_{N-1}(p_0 + p_k, \boldsymbol{p}_{-k})$, where $\boldsymbol{p}_{-k}$ denotes the vector $\boldsymbol{p}$ with the kth coordinate omitted. Thus,

$$\begin{aligned}
\operatorname{var}(T_N(p_0, \boldsymbol{p})) &= \operatorname{var}(G_1) + \operatorname{var}(T_N(p_0, \boldsymbol{p}) - G_1) \\
&\geq \frac{p_0}{(1-p_0)^2} + \sum_{k=1}^{N} \frac{p_k}{1-p_0} \operatorname{var}(T_{N-1}(p_0 + p_k, \boldsymbol{p}_{-k})) \\
&\geq \frac{p_0}{(1-p_0)^2} + \sum_{k=1}^{N} \frac{p_k}{1-p_0} \operatorname{var}\biggl(T_{N-1}\biggl(p_0 + p_k, \frac{(1-p_0-p_k)\mathbf{1}_{N-1}}{N-1}\biggr)\biggr) \\
&= \frac{p_0}{(1-p_0)^2} + \sum_{k=1}^{N} \biggl(\frac{p_k}{1-p_0}\biggr) \frac{\alpha_{N-1}(p_0+p_k) + \beta_{N-1}}{(1-p_0-p_k)^2}. \qquad (2)
\end{aligned}$$

where we have used $\operatorname{var}(Y) \geq \mathbb{E}\operatorname{var}(Y \mid X)$ in the first inequality, the induction hypothesis in the second inequality, and (1) in the last equality. For fixed $p_0 \in [0, 1)$, define $\phi(x) = x(\alpha_{N-1}(p_0 + x) + \beta_{N-1})/(1 - p_0 - x)^2$. One can verify that $\phi''(x) \geq 0$ for $x \in (0, 1 - p_0)$. Jensen's inequality yields $\sum_{k=1}^{N} \phi(p_k) \geq N\phi((1 - p_0)/N)$ subject to $\sum_{k=1}^{N} p_k = 1 - p_0$. Putting this minimal value in (2), and after some algebra, we have

$$\operatorname{var}(T_N(p_0, \boldsymbol{p})) \geq \frac{p_0}{(1-p_0)^2} + \frac{N}{1-p_0}\phi\biggl(\frac{1-p_0}{N}\biggr) = \frac{\alpha_N p_0 + \beta_N}{(1-p_0)^2}.$$

Since the right-hand side matches that of (1), we have shown that the equal probability case achieves minimum variance, as claimed. □

References

Anceaume, E., Busnel, Y. and Sericola, B. (2015). New results on a generalized coupon collector problem using Markov chains. *J. Appl. Prob.* **52,** 405–418.

Doumas, A. V. and Papanicolaou, V. G. (2012). The coupon collector's problem revisited: asymptotics of the variance. *Adv. Appl. Prob.* **44,** 166–195.

Marshall, A. W., Olkin, I. and Arnold, B. C. (2009). *Inequalities: Theory of Majorization and Its Applications*, 2nd edn. Springer, New York.

Ross, S. (2010). *Introduction to Probability Models*, 10th edn. Academic Press, Burlington, MA.

Sendov, H. and Shan, S. (2015). New representation theorems for completely monotone and Bernstein functions with convexity properties on their measures. *J. Theoret. Prob.* **28,** 1689–1725.

Yours sincerely,

Yaming Yu
Department of Statistics,
University of California,
CA 92697,
USA.
Email address: yamingy@uci.edu