Communication Papers of the of the 17th Conference on Computer Science and Intelligence Systems pp. 113–116
DOI: 10.15439/2022F221
ISSN 2300-5963 ACSIS, Vol. 32

Dirichlet's principle revisited: An inverse Dirichlet's principle definition and its bound estimation improvement using stochastic combinatorics

Lubomír Štěpánek
Department of Statistics and Probability
Faculty of Informatics and Statistics
Prague University of Economics and Business
nám. W. Churchilla 4, 130 67 Prague, Czech Republic
lubomir.stepanek@vse.cz

Filip Habarta
Department of Statistics and Probability
Faculty of Informatics and Statistics
Prague University of Economics and Business
nám. W. Churchilla 4, 130 67 Prague, Czech Republic
filip.habarta@vse.cz

Ivana Malá
Department of Statistics and Probability
Faculty of Informatics and Statistics
Prague University of Economics and Business
nám. W. Churchilla 4, 130 67 Prague, Czech Republic
malai@vse.cz

Luboš Marek
Department of Statistics and Probability
Faculty of Informatics and Statistics
Prague University of Economics and Business
nám. W. Churchilla 4, 130 67 Prague, Czech Republic
marek@vse.cz

***Abstract*—Dirichlet's principle, also known as a pigeonhole principle, claims that if $n \in \mathbb{N}$ item are put into $m \in \mathbb{N}$ containers, with $n > m$, then there is a container that contains more than one item. In this work, we focus rather on an inverse Dirichlet's principle (by switching items and containers), which is as follows: considering $n \in \mathbb{N}$ items put in $m \in \mathbb{N}$ containers, when $n < m$, then there is at least one container with no item inside. Furthermore, we refine Dirichlet's principle using discrete combinatorics within a probabilistic framework. Applying stochastic fashion on the principle, we derive the number of items n may be even greater than or equal to m, still very likely having one container without an item. The inverse definition of the problem rather than the original one may have some practical applications, particularly considering derived effective upper bound estimates for the items number, as demonstrated using some applied mini-studies.**

I. Introduction

WHILE the Dirichlet's principle is applied in various fields such as number theory and calculus [1], data compression [2], quantum mechanics [3] and many others, in this work, we focus only on the original definition in a pure discrete fashion, and mainly on a derived, inverse form. The original version of the Dirichlet's principle [4] claims that if there are more items than containers so that the items are put into the containers, then there is at least one container containing two or more items. Thus, due to the idea's relative simplicity, applications of the original Dirichlet's principle are usually limited to rather fancy problems coming from recreational mathematics.

Within the paper, going further, we refine the original Dirichlet's principle in an inverse form: when there are fewer items than containers and put inside them, there is at least one container with no items inside.

The estimate of a maximum number of items lower than the number of containers is relatively poor in a stochastic fashion, though. The inverse definition enables applying the pigeonhole principle to some real-world situations and helps solve them effectively. We demonstrate how the stochastic approach to inverse Dirichlet's principle, together with combinatorial calculations of probabilities, helps to get more effective estimates for selected problem parameters, as indicated above. To be more specific, we applied stochastic-based inverse Dirichlet's principle to a real-world situation called *an unoccupied doubleseat problem*, originated by the paper's authors, where an individual wants to know a probability of an unoccupied doubleseat in a row of doubleseats when booking their seat. The inverse Dirichlet's principle would return a relatively poor estimate of a maximum number of individuals who booked their seats so far, ensuring there should be at least one unoccupied doubleseat. However, the probabilistic approach can show a substantially significant probability of an unoccupied doubleseat even for more already booking individuals than estimated using the pigeonhole principle.

II. An inverse Dirichlet's principle definition and stochastic combinatorial approach to selected applications of the principle

A. An inverse Dirichlet's principle definition

Assuming there are $n \in \mathbb{N}$ items put into $m \in \mathbb{N}$ containers, with $n < m$, then there is for sure at least one container with no items inside it.

A *proof*, built by contradiction, is clear and as follows. Suppose that each of $m \in \mathbb{N}$ containers contains at least one of all $n \in \mathbb{N}$ items inside, with $n < m$. Then there is minimally

$m \cdot 1 = m$ items, which is in contradiction with the assumption of $n < m$ items. $\square$

We apply the inverse version of Dirichlet's principle to a real-world inspired problem called *an unoccupied doubleseat problem*, where an individual wants to know how likely there is an unoccupied doubleseat in a row of k of them (*containers*) when n individuals (*items*) randomly occupied the seats before him. Moreover, we show that Dirichlet's principle can be replaced by a stochastic approach, estimating the same output as the inverse pigeonhole principle, i. e. a maximum number of items so that there is (likely) an unoccupied container[1].

B. An unoccupied doubleseat problem

Let us assume we have a row of $k \in \mathbb{N}$ doubleseats[2] as in Fig. 1 and $n \in \mathbb{N} \cup \{0\}$ individuals who have randomly occupied n of the seats, one seat per individual, with $0 \leq n \leq 2k$. If there is a newcomer who would like to sit down on one of the seats not occupied so far, what is the probability that there is at least one unoccupied[3] doubleseat in the row[4]?

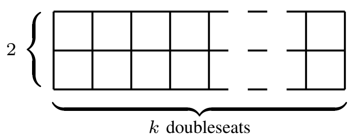

Fig. 1. A row of $k \in \mathbb{N}$ doubleseats, used in the *unoccupied doubleseat problem*.

1) A solution by the inverse Dirichlet's principle: The pigeonhole principle is limited to offering only a discrete solution. When the newcomer wants to know how likely there are one or more unoccupied doubleseats, assuming there are $2k$ seats in total, arranged as k doubleseats (*containers*), and n seats are already occupied (*items*). Applying the Dirichlet's principle, when there are $n < k$ individuals taking the seats, there must be at least one unoccupied doubleseat. Otherwise, if there are $n \geq k$ individuals occupying the seats, no one can assure whether there are one or more free doubleseats. For $n > 2k - 2$ sitting individuals, there is for sure no unoccupied doubleseat.

2) A solution by a combinatorial stochastic approach: Let us mark as p a probability there is at least one unoccupied doubleseat, assuming a row of k doubleseats and n randomly sitting individuals at the moment, $0 \leq n \leq 2k$. When $0 \leq n \leq k - 1$, then obviously, as claimed above, $p = 1$. For $k \leq n \leq 2k$, we get

$$\begin{aligned} p &= P(\geq 1 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\ &= 1 - P(0 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k). \end{aligned} \tag{1}$$

[1]Somewhat similar problem, bur using rather the original Dirichlet's principle, not the inverse one, is so-called birthday problem, [5].

[2]For example, in a bus, a train, or a plane.

[3]I. e., both seats of such a doubleseat are unoccupied.

[4]Newcomers, particularly when alone, prefer to sit down on an unoccupied doubleseat to not sitting next to someone other.

The term $P(0 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k)$ in formula (1) could be estimated as follows – since the first n individuals take their seats randomly, we may assume

$$\begin{aligned} p &= P(\geq 1 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\ &= 1 - P(0 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\ &= 1 - \frac{N}{M}. \end{aligned} \tag{2}$$

While the denominator M of the term $1 - p$ is straightforward, since a number of all ways how $2k$ seats can by taken by n individuals is equal to $M = \binom{2k}{n}$, the numerator N is tricky.

Assuming there is no unoccupied doubleseat, each of k doubleseats is occupied by one or two of n individuals. Thus $n-k$ doubleseats must be fully occupied, while $k-(n-k) = 2k-n$ doubleseats are occupied only by one individual. The number of ways $n-k$ doubleseats of k in total are fully occupied, is equal to $\binom{k}{n-k}$, while the number of ways $2k-n$ doubleseats are occupied only by one individual is 2^{2k-n}. In total, the numerator N of the $1-p$ term, i. e. number of ways how one or both seats per each of k doubleseats are taken by n individuals, is equal to $N = \binom{k}{n-k} \cdot 2^{2k-n}$. Putting things together, we improve formula (2) as

$$\begin{aligned} p &= P(\geq 1 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\ &= 1 - P(0 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\ &= 1 - \frac{N}{M} = \\ &= 1 - \frac{\binom{k}{n-k} \cdot 2^{2k-n}}{\binom{2k}{n}}. \end{aligned} \tag{3}$$

What is worth mentioning is the probability p, i. e., there are one or more unoccupied doubleseats for n already sitting individuals, is close to 1.0 even for values $n > k - 1$, as we can see in Table I and Fig. 2.

TABLE I
MAXIMUM NUMBERS n OF ALREADY SITTING INDIVIDUALS THAT STILL ENSURE THERE IS ONE OR MORE UNOCCUPIED DOUBLESEAT WITH THE PROBABILITY p FOR GIVEN TOTAL NUMBER k OF DOUBLESEATS.

$k = 10$		$k = 20$		$k = 30$	
p	n	p	n	p	n
1.00	9	1.00	19	1.00	29
> 0.99	10	> 0.99	24	> 0.99	40
> 0.95	11	> 0.95	26	> 0.95	43
> 0.90	12	> 0.90	28	> 0.90	45
> 0.80	13	> 0.80	29	> 0.80	47

3) A lower bound for the probability estimate coming from the combinatorial stochastic approach: Considering formula (3), one may try to derive a lower bound for the probability there are one or more unoccupied doubleseats in a row of $k \in \mathbb{N}$ doubleseats, assuming $n \in \mathbb{N} \cup \{0\}$ seats are already occupied. However, before the lower bound construction, we introduce some lemmas to be applied.

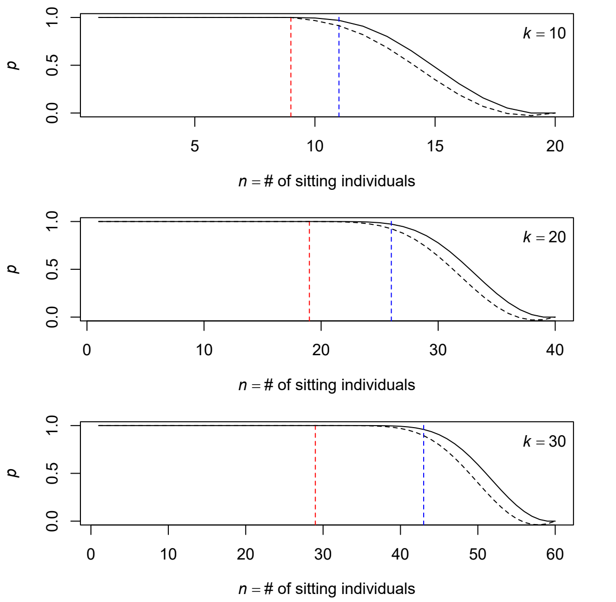

Fig. 2. Maximum numbers n of already sitting individuals that still ensure there is one or more unoccupied doubleseats with the probability p for given total number k of doubleseats. There are estimates of maximum n ensuring ≥ 1 unoccupied doubleseat by the inverse Dirichlet's principle (the red dashed line) and by the stochastic refining of the Dirichlet's principle (the blue dashed line, with $p > 0.95$). The black dashed line stands for the lower bound of the probability estimate.

Lemma. Assuming real numbers $0 < d < r \leq s$, a fraction $\frac{r}{s}$ is greater than or equal to a fraction $\frac{r-d}{s-d}$, i. e. $\frac{r}{s} \geq \frac{r-d}{s-d}$.

Proof. Since $r \leq s$, it is also $rd \leq sd$ and $-rd \geq -sd$, and, eventually, $rs - rd \geq rs - sd$. Thus, if $rs - rd \geq rs - sd$, it is also $r(s-d) \geq s(r-d)$, and since $r - d > 0$ and $s - d > 0$, it is also $\frac{r}{s} \geq \frac{r-d}{s-d}$. $\square$

Lemma. Assuming real numbers $0 < d_i < r \leq s$ for $\forall i \in \{1, 2, \ldots, m\}$, it is

$$\prod_{i=1}^{m} \frac{(r-d_i)(r+d_i)}{(s-d_i)(s+d_i)} \leq \left(\frac{r}{s}\right)^{2m}. \tag{4}$$

Proof. Considering the previous lemma, it is obviously $\frac{r^2-d_i^2}{s^2-d_i^2} \leq \frac{r^2}{s^2}$ for $\forall i \in \{1, 2, \ldots, m\}$, thus, it is also $\prod_{i=1}^{m} \frac{r^2-d_i^2}{s^2-d_i^2} \leq \left(\frac{r^2}{s^2}\right)^m = \left(\frac{r}{s}\right)^{2m}$, and, finally, by reformulation, it is $\prod_{i=1}^{m} \frac{(r-d_i)(r+d_i)}{(s-d_i)(s+d_i)} \leq \left(\frac{r}{s}\right)^{2m}$. $\square$

Let us now derive a lower bound for the probability in formula (3). We get

$$\begin{aligned}
p &= 1 - P(0 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\
&= 1 - \frac{\binom{k}{n-k} \cdot 2^{2k-n}}{\binom{2k}{n}} = \\
&= 1 - \frac{\frac{k!}{(n-k)!(2k-n)!}}{\frac{(2k)!}{n!(2k-n)!}} \cdot 2^{2k-n} = \\
&= 1 - \frac{n!k!}{(n-k)!(2k)!} \cdot 2^{2k-n} = \\
&= 1 - \frac{n!}{(n-k)!} \frac{k!}{(2k)!} \cdot 2^{2k-n} = \\
&= 1 - \underbrace{\frac{n(n-1)\cdots(n-k+1)}{2k(2k-1)\cdots(2k-k+1)}}_{k \text{ terms}} \cdot 2^{2k-n} = \\
&= 1 - \frac{(\eta+\lambda)(\eta+\lambda-1)\cdots\eta\cdots(\eta-\lambda+1)(\eta-\lambda)}{(\kappa+\lambda)(\kappa+\lambda-1)\cdots\kappa\cdots(\kappa-\lambda+1)(\kappa-\lambda)} \cdot 2^{2k-n},
\end{aligned}$$

where $\eta = n - \frac{k-1}{2}$, $\kappa = 2k - \frac{k-1}{2}$ and $\lambda = \frac{k-1}{2}$. Applying the formula (4), we get

$$\prod_{i=0}^{(k-1)/2} \frac{(\eta - (\lambda - i))(\eta + (\lambda - i))}{(\kappa - (\lambda - i))(\kappa + (\lambda - i))} \leq \left(\frac{\eta}{\kappa}\right)^k,$$

so,

$$\begin{aligned}
p &= 1 - P(0 \text{ unoccupied doubleseat} \mid k \leq n \leq 2k) = \\
&= 1 - \frac{(\eta+\lambda)(\eta+\lambda-1)\cdots\eta\cdots(\eta+\lambda-1)(\eta+\lambda)}{(\kappa+\lambda)(\kappa+\lambda-1)\cdots\kappa\cdots(\kappa+\lambda-1)(\kappa+\lambda)} \cdot 2^{2k-n} \geq \\
&\geq 1 - \left(\frac{\eta}{\kappa}\right)^k = \\
&= 1 - \left(\frac{n - \frac{k-1}{2}}{2k - \frac{k-1}{2}}\right)^k \cdot 2^{2k-n},
\end{aligned} \tag{5}$$

which is the lower bound of the probability p. Counting up all arithmetic operations in formulas (3) and (5), we get asymptotic time complexity $\Theta(8k - n)$ for the precisely calculated probability p while only $\Theta(3k - n + 7)$ for the probability lower bound. Moreover, the lower bound formula (5) minimizes the risk of over- or underfloating due to avoiding terms with combinatorial coefficients of $\binom{a}{b}$ type. Checking the Fig. 2, we see the probability lower bound (the black dashed line) is quite effective with limited possibility to be improved.

C. *An unoccupied l-seat problem*

Now, we generalize the *unoccupied doubleseat problem* in terms of changing the doubleseats to l-seats for $l \in \mathbb{N}$. Let us assume we have a row of $k \in \mathbb{N}$ consecutive l-seats as in Fig. 3 and $n \in \mathbb{N} \cup \{0\}$ individuals who randomly sitting on n of the seats, one seat per individual, with $0 \leq n \leq lk$. What is the probability there is an unoccupied[5] l-seat?

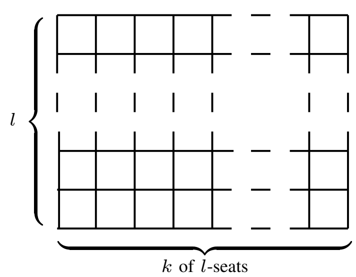

Fig. 3. A row of $k \in \mathbb{N}$ consecutive l-seats, used in the *unoccupied l-seat problem.*

[5] I. e., all l seats of such an l-seat are unoccupied.

1) A solution by the inverse Dirichlet's principle: Obviously, by applying the pigeonhole principle, assuming there are k of l-seats (*containers*), and n seats are already occupied (of lk in total) (*items*), if and only if there are $n < k$ sitting individuals, there is for sure one or more unoccupied l-seats. On the other hand, for $n > lk - l$ sitting individuals, there is surely no unoccupied l-seat.

2) A solution by a combinatorial stochastic approach: Again, by reformulation the probability p estimating there is at least one unoccupied l-seat using formula (1), assuming k of l-seats and n randomly sitting individuals at the moment, $k \leq n \leq lk$ (as far as for $0 \leq n \leq k-1$ is the problem solved using the Dirichlet's principle), we get

$$\begin{aligned} p &= P(\geq 1 \text{ unoccupied } l\text{-seat} \mid k \leq n \leq lk) = \\ &= 1 - P(0 \text{ unoccupied } l\text{-seat} \mid k \leq n \leq lk) = \\ &= 1 - \frac{N}{M} = 1 - \frac{N}{\binom{lk}{n}}, \end{aligned} \tag{6}$$

where the denominator M is how lk seats can be taken by n individuals, i. e. $M = \binom{lk}{n}$; however, the numerator N is analytically nontrivial and is calculated exhaustively using Algorithm 1 with asymptotic time complexity roughly $\Theta\left(2n(2l + k)\right)$.

Algorithm 1: Calculating a total number N of ways all k of l-seats are occupied by n individuals, so that there is at least one sitting individual on each l-seat

Data: $k \in \mathbb{N}$ of l-seats for $l \in \mathbb{N}$, n individuals with $k \leq n \leq lk$

Result: A total number N of ways all k of l-seats are occupied by n individuals, so that there is at least one sitting individual on each l-seat

$N = 0$;
for $\forall\ [d_1, d_2, \ldots, d_k] \in \mathbb{N}^k : \sum_{i=1}^{k} d_i = n$ **do**
  if $\forall i \in \{1, 2, \ldots, k\} : 1 \leq d_i \leq l$ **then**
    $N \leftarrow N + \prod_{i=1}^{k} \binom{l}{d_i}$
  end
end

Checking the Table II and Fig. 4, compared to Dirichlet's principle results, the stochastic approach still returns larger estimates of the maximum number n of sitting individuals ensuring (likely) an unoccupied l-seat; however, stochastically-based maximum n's seem closer to Dirichlet-based estimates for the l-seats problem than for the doubleseat one.

III. Conclusion Remarks

Refining the Dirichlet's principle using a stochastic combinatorial approach enables us to improve estimates of the upper number n of items randomly placed in k of l-containers, still likely ensuring there is at least one l-container with no item in it. On a more practical note, whenever someone books a seat in a row of l-seats, even if the number of already booked seats

TABLE II
Maximum numbers n of already sitting individuals that still ensure there is one or more unoccupied l-seats with the probability p for given total number k of l-seats.

$k = 6$, $l = 3$		$k = 9$, $l = 3$	
p	n	p	n
1.00	5	1.00	8
> 0.99	5	> 0.99	9
> 0.95	6	> 0.95	10
> 0.90	6	> 0.90	11
> 0.80	7	> 0.80	12

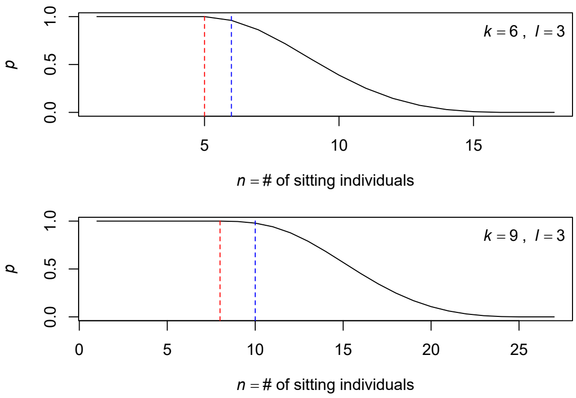

Fig. 4. Maximum numbers n of already sitting individuals that still ensure there is one or more unoccupied l-seats with the probability p for given total number k of l-seats. There are estimates of maximum n ensuring ≥ 1 unoccupied l-seat by the inverse Dirichlet's principle (the red dashed line) and by the stochastic refining of the Dirichlet's principle (the blue dashed line, with $p > 0.95$).

is higher than the number of the l-seats, there is still a high probability of an unoccupied l-seat, unexpectedly according to the discrete inverse Dirichlet's principle, though.

IV. Acknowledgement

This paper is supported by the grant OP VVV IGA/A, CZ.02.2.69/0.0/0.0/19_073/0016936 by the Internal Grant Agency of the Prague University of Economics and Business.

References

[1] Lynn Loomis and Shlomo Sternberg. "Potential theory in $\mathbb{E}^n$". In: *Advanced Calculus*. World Scientific, Mar. 2014, pp. 474–508. DOI: 10.1142/9789814583947_0013.

[2] Giridhar D. Mandyam, Nasir U. Ahmed, and Neeraj Magotra. "DCT-based scheme for lossless image compression". In: *SPIE Proceedings*. SPIE, Apr. 1995. DOI: 10.1117/12.206386.

[3] Yakir Aharonov, Fabrizio Colombo, Sandu Popescu, et al. "Quantum violation of the pigeonhole principle and the nature of quantum correlations". In: *Proceedings of the National Academy of Sciences* 113.3 (Jan. 2016), pp. 532–535. DOI: 10.1073/pnas.1522411112.

[4] Benoit Rittaud and Albrecht Heeffer. "The Pigeonhole Principle, Two Centuries Before Dirichlet". In: *The Mathematical Intelligencer* 36.2 (Aug. 2013), pp. 27–29. DOI: 10.1007/s00283-013-9389-1.

[5] Mario Cortina Borja and John Haigh. "The birthday problem". In: *Significance* 4.3 (Aug. 2007), pp. 124–127. DOI: 10.1111/j.1740-9713.2007.00246.x.