

# The maximum value in complete geometric words

Margaret Archibald<sup>1</sup>    Arnold Knopfmacher<sup>2</sup>

<sup>1</sup>Department of Mathematics and Applied Mathematics  
University of Cape town

<sup>2</sup>School of Mathematics  
University of the Witwatersrand

GASCom 2012

# Outline

- 1 Introduction
  - The Basic Problem
  - Previous Research
  - The Problem in More Detail
- 2 Method
  - Notation
  - Previous Results
  - Techniques
- 3 Results
  - Expectation
  - Notation
  - Variance
  - Notes

# Introduction

## The Basic Problem

- Terminology
  - **Geometric words:** Samples of  $n$  geometric random variables  $(\Gamma_1, \Gamma_2, \dots, \Gamma_n)$  where  $\mathbb{P}\{\Gamma_j = i\} = pq^{i-1}$ , for  $1 \leq j \leq n$ , with  $p + q = 1$ .
  - **Complete geometric words:** Every value between 1 and the maximum must occur at least once.
- Examples:
  - 112131 is complete
  - 211421 is not complete
  - (232242 is not complete but is considered to be “gap-free”).
- Goal:
  - Find mean and variance of the maximum for complete words.

# Introduction

## The Basic Problem

- Terminology
  - **Geometric words:** Samples of  $n$  geometric random variables  $(\Gamma_1, \Gamma_2, \dots, \Gamma_n)$  where  $\mathbb{P}\{\Gamma_j = i\} = pq^{i-1}$ , for  $1 \leq j \leq n$ , with  $p + q = 1$ .
  - **Complete geometric words:** Every value between 1 and the maximum must occur at least once.
- Examples:
  - 112131 is complete
  - 211421 is not complete
  - (232242 is not complete but is considered to be “gap-free”).
- Goal:
  - Find mean and variance of the maximum for complete words.

# Introduction

## The Basic Problem

- Terminology
  - **Geometric words:** Samples of  $n$  geometric random variables  $(\Gamma_1, \Gamma_2, \dots, \Gamma_n)$  where  $\mathbb{P}\{\Gamma_j = i\} = pq^{i-1}$ , for  $1 \leq j \leq n$ , with  $p + q = 1$ .
  - **Complete geometric words:** Every value between 1 and the maximum must occur at least once.
- Examples:
  - 112131 is complete
  - 211421 is not complete
  - (232242 is not complete but is considered to be “gap-free”).
- Goal:
  - Find mean and variance of the maximum for complete words.

# Introduction

## Previous Research

- *Yet another application of a binomial recurrence. Order Statistics,*
  - W. Szpankowski and V. Rego, 1990
- *Distinctness of compositions of an integer: a probabilistic analysis,*
  - P. Hitczenko and G. Louchard, 2004
- *Gap-free compositions and gap-free samples of geometric random variables,*
  - P. Hitczenko and A. Knopfmacher, 2005
- *Gaps in samples of geometric random variables,*
  - W. Goh and P. Hitczenko, 2007
- *On gaps and unoccupied urns in sequences of geometrically distributed random variables,*
  - G. Louchard and H. Prodinger, 2008.

# Introduction

## The Problem in More Detail

The coin-flipping game (each coin represents an independent variable).

- Round 1:  $n$  coins are flipped
- Round 2: all coins now showing “heads” will be flipped again.
- Round 3: all coins now showing “heads” will be flipped again.
- etc...
- End of game: when all coins are showing “tails”.
- Note: At no point can we flip all heads.

# Introduction

## The Problem in More Detail

The coin-flipping game (each coin represents an independent variable).

- Round 1:  $n$  coins are flipped
- Round 2: all coins now showing “heads” will be flipped again.
- Round 3: all coins now showing “heads” will be flipped again.
- etc...
- End of game: when all coins are showing “tails”.
- Note: At no point can we flip all heads.



# Introduction

## The Problem in More Detail

The coin-flipping game (each coin represents an independent variable).

- Round 1:  $n$  coins are flipped
- Round 2: all coins now showing “heads” will be flipped again.
- Round 3: all coins now showing “heads” will be flipped again.
- etc...
- End of game: when all coins are showing “tails”.
- Note: At no point can we flip all heads.

# Introduction

## The Problem in More Detail

The coin-flipping game (each coin represents an independent variable).

- Round 1:  $n$  coins are flipped
- Round 2: all coins now showing “heads” will be flipped again.
- Round 3: all coins now showing “heads” will be flipped again.
- etc...
- End of game: when all coins are showing “tails”.
- Note: At no point can we flip all heads.

# Introduction

## The Problem in More Detail

The coin-flipping game (each coin represents an independent variable).

- Round 1:  $n$  coins are flipped
- Round 2: all coins now showing “heads” will be flipped again.
- Round 3: all coins now showing “heads” will be flipped again.
- etc...
- End of game: when all coins are showing “tails”.
- Note: At no point can we flip all heads.

# Introduction

## The Problem in More Detail

Alternative model:

- Label urns  $1, 2, \dots$  and throw  $n$  balls.
- The probability that a ball falls into urn  $j$  is  $pq^{j-1}$ .
- In this case we would not allow any empty urns.

# Introduction

## The Problem in More Detail

Alternative model:

- Label urns  $1, 2, \dots$  and throw  $n$  balls.
- The probability that a ball falls into urn  $j$  is  $pq^{j-1}$ .
- In this case we would not allow any empty urns.

# Method

## Notation

- $q = \text{“heads”}$ ,  $p = \text{“tails”}$ .

Using this notation, we have that the probability of having  $k$  heads after the first round is

$$\binom{n}{k} q^k p^{n-k}.$$

- $p = 1 - q$
- $Q := 1/q$
- $L := \log Q (= \ln Q)$

# Method

## More Notation

- $\chi_k := \frac{2k\pi i}{L}$ , where  $k \in \mathbb{Z}$  and  $k \neq 0$
- $H_k := \sum_{i=1}^k \frac{1}{i}$  denotes the  $k$ th harmonic number.
- $\gamma := 0.5772\dots$  denotes Euler's constant
- $p_n$  = the probability that geometric word is complete
- $M_n$  = average maximum value in a complete geometric word
- $V_n$  = second factorial moment of the maximum value in a complete geometric word
- $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$

# Method

## Previous Results

Let  $p_n$  be the probability that a geometric word is complete. Then, Hitczenko and Knopfmacher (2005) found:

- when  $p = q = 1/2$ ,

$$p_n = \frac{1}{2}$$

for every  $n \geq 1$

- otherwise for  $0 < q < 1$ ,

$$p_n \sim 1 - \frac{1}{L} \sum_{j \geq 1} p_j \frac{q^j}{j} - \delta_P(\log_Q n)$$

as  $n \rightarrow \infty$ , where

$$\delta_P(x) := \frac{1}{L} \sum_{k \neq 0} T^*(\chi_k) e^{2k\pi i x} \quad \text{with} \quad T^*(\chi_k) = \sum_{j=1}^{\infty} p_j \frac{q^j}{j!} \Gamma(\chi_k + j).$$



# Method

## Previous Results

Let  $p_n$  be the probability that a geometric word is complete. Then, Hitczenko and Knopfmacher (2005) found:

- when  $p = q = 1/2$ ,

$$p_n = \frac{1}{2}$$

for every  $n \geq 1$

- otherwise for  $0 < q < 1$ ,

$$p_n \sim 1 - \frac{1}{L} \sum_{j \geq 1} p_j \frac{q^j}{j} - \delta_P(\log_Q n)$$

as  $n \rightarrow \infty$ , where

$$\delta_P(x) := \frac{1}{L} \sum_{k \neq 0} T^*(\chi_k) e^{2k\pi i x} \quad \text{with} \quad T^*(\chi_k) = \sum_{j=1}^{\infty} p_j \frac{q^j}{j!} \Gamma(\chi_k + j).$$

# Method

## Previous Results

Let  $p_n$  be the probability that a geometric word is complete. Then, Hitczenko and Knopfmacher (2005) found:

- when  $p = q = 1/2$ ,

$$p_n = \frac{1}{2}$$

for every  $n \geq 1$

- otherwise for  $0 < q < 1$ ,

$$p_n \sim 1 - \frac{1}{L} \sum_{j \geq 1} p_j \frac{q^j}{j} - \delta_P(\log_Q n)$$

as  $n \rightarrow \infty$ , where

$$\delta_P(x) := \frac{1}{L} \sum_{k \neq 0} T^*(\chi_k) e^{2k\pi i x} \quad \text{with} \quad T^*(\chi_k) = \sum_{j=1}^{\infty} p_j \frac{q^j}{j!} \Gamma(\chi_k + j).$$

# Method

## Techniques

- Recursions,
  - e.g.  $M_n = p_n + \sum_{k=0}^{n-1} \binom{n}{k} q^k p^{n-k} M_k$  for  $n \geq 2$  with  $M_0 = 0$  and  $M_1 = p_1 = p$ .
- Generating functions,
  - e.g.  $F_n(u) = \sum_{k=1}^n p_{n,k} u^k$  where  $p_{n,k}$  denotes the probability that a word of length  $n$  is complete and has maximum value  $k$ .
- Functional equations
  - e.g.  $F_n(u) = u \sum_{k=0}^{n-1} \binom{n}{k} q^k p^{n-k} F_k(u)$  with  $F_0(u) = 1$ ,  $F_1(u) = pu$ .

# Method

## Techniques

- Recursions,
  - e.g.  $M_n = p_n + \sum_{k=0}^{n-1} \binom{n}{k} q^k p^{n-k} M_k$  for  $n \geq 2$  with  $M_0 = 0$  and  $M_1 = p_1 = p$ .
- Generating functions,
  - e.g.  $F_n(u) = \sum_{k=1}^n p_{n,k} u^k$  where  $p_{n,k}$  denotes the probability that a word of length  $n$  is complete and has maximum value  $k$ .
- Functional equations
  - e.g.  $F_n(u) = u \sum_{k=0}^{n-1} \binom{n}{k} q^k p^{n-k} F_k(u)$  with  $F_0(u) = 1$ ,  $F_1(u) = pu$ .

# Method

## Techniques

- Recursions,
  - e.g.  $M_n = p_n + \sum_{k=0}^{n-1} \binom{n}{k} q^k p^{n-k} M_k$  for  $n \geq 2$  with  $M_0 = 0$  and  $M_1 = p_1 = p$ .
- Generating functions,
  - e.g.  $F_n(u) = \sum_{k=1}^n p_{n,k} u^k$  where  $p_{n,k}$  denotes the probability that a word of length  $n$  is complete and has maximum value  $k$ .
- Functional equations
  - e.g.  $F_n(u) = u \sum_{k=0}^{n-1} \binom{n}{k} q^k p^{n-k} F_k(u)$  with  $F_0(u) = 1$ ,  $F_1(u) = pu$ .

# Method

## Techniques

- Poissonisation
  - e.g. define  $H(z) := M(z)e^{-z}$
- Mellin transforms

$$f^*(s) = \mathcal{M}(f(z), s) := \int_0^\infty f(x)x^{s-1} dx$$

- e.g.  $\mathcal{M}(e^{-z} - 1, s) = \Gamma(s)$
- Mellin inversion formula  $H(x) = \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} H^*(s)x^{-s} ds$ 
  - e.g. the asymptotic calculations follow from residue analysis:

$$H(x) \sim - \sum_{k \in \mathbb{Z}} \text{Res}_{s=\chi_k} (H^*(s)x^{-s})$$

where  $\chi_k := \frac{2k\pi i}{L}$

# Method

## Techniques

- Poissonisation
  - e.g. define  $H(z) := M(z)e^{-z}$
- Mellin transforms

$$f^*(s) = \mathcal{M}(f(z), s) := \int_0^\infty f(x)x^{s-1}dx$$

- e.g.  $\mathcal{M}(e^{-z} - 1, s) = \Gamma(s)$
- Mellin inversion formula  $H(x) = \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} H^*(s)x^{-s}ds$ 
  - e.g. the asymptotic calculations follow from residue analysis:

$$H(x) \sim - \sum_{k \in \mathbb{Z}} \text{Res}_{s=\chi_k} (H^*(s)x^{-s})$$

where  $\chi_k := \frac{2k\pi i}{L}$

# Method

## Techniques

- Poissonisation
  - e.g. define  $H(z) := M(z)e^{-z}$
- Mellin transforms

$$f^*(s) = \mathcal{M}(f(z), s) := \int_0^\infty f(x)x^{s-1} dx$$

- e.g.  $\mathcal{M}(e^{-z} - 1, s) = \Gamma(s)$
- Mellin inversion formula  $H(x) = \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} H^*(s)x^{-s} ds$ 
  - e.g. the asymptotic calculations follow from residue analysis:

$$H(x) \sim - \sum_{k \in \mathbb{Z}} \text{Res}_{s=\chi_k} (H^*(s)x^{-s})$$

where  $\chi_k := \frac{2k\pi i}{L}$



# Results

## Expectation

### Theorem

*The average value of the maximum in a complete word is*

$$\frac{M_n}{p_n} \sim \log_Q n + 1 + \frac{1}{L p_n} \left( \gamma - \frac{L}{2} - A + \frac{B'}{L} + L \delta_M(\log_Q n) \right),$$

*where*

$$\delta_M(x) := -\frac{1}{L} \sum_{k \neq 0} e^{2k\pi i \log_Q n} \left( T_2^*(\chi_k) - \frac{1}{L} T^*{}'(\chi_k) \right).$$

# Results

## Notation

- $A = \sum_{n \geq 1} M_n \frac{q^n}{n}$
- $A' = \sum_{n \geq 1} M_n \frac{q^n \Gamma'(n)}{n!}$
- $B = \sum_{n \geq 1} p_n \frac{q^n}{n}$
- $B' = \sum_{n \geq 1} p_n \frac{q^n \Gamma'(n)}{n!}$
- $B'' = \sum_{n \geq 1} p_n \frac{q^n \Gamma''(n)}{n!}$
- $V = \sum_{n \geq 1} V_n \frac{q^n}{n}$
- $T_2^*(\chi_k) = \Gamma(\chi_k) + \sum_{j=1}^{\infty} M_j \frac{q^j}{j!} \Gamma(\chi_k + j)$
- $T^{*'}(\chi_k) = \sum_{j=1}^{\infty} p_j \frac{q^j}{j!} \Gamma'(\chi_k + j)$

# Results

## Variance

### Theorem

*The variance of the maximum in a complete sample of geometric random variables is given by*

$$\begin{aligned} & \frac{1}{12L^2(B-L)^2} \left( -12A^2L^2 + 12ABL^2 + 24AB'L - 24AL^3 + 24\gamma AL^2 \right. \\ & \quad - 24A'BL + 24A'L^2 + 12BB'' - 4BL^3 + 12BL^2V \\ & \quad + 12\gamma BL^2 - 2\pi^2BL - 12\gamma^2BL - 12B''L - 12B'^2 \\ & \quad \left. + 12B'L^2 - 24\gamma B'L + L^4 - 12L^3V + 2\pi^2L^2 \right) \\ & \quad + \delta_*(\log_Q n). \end{aligned}$$

*where  $\delta_*(\log_Q n)$  is a fluctuating function of small magnitude.*

# Results

## Notes

### Notes

- For  $p = \frac{1}{2}$  (so  $p_n = \frac{1}{2}$ ) the average value is

$$2M_n \sim \log_2 n + \frac{1}{2} + \frac{\gamma}{L} - \frac{2A}{L} + \delta_M(\log_2 n).$$

Thus AveMax in a general word exceeds AveMax in a complete word by roughly  $\frac{2A}{L} \approx 1.2032$ .

- For  $p = 1/2$ , the variance constant is 1.00723.

# Thank you