# Connexions and the Gumbel Distribution

Myron Hlynka

Department of Mathematics and Statistics
University of Windsor
Windsor, ON, Canada.

October, 2016

# Outline

- Definition of Gumbel Distribution
- Riemann zeta function
- proof that 0=1
- coupon collector's problem
- Integer partitions

# Definition of Gumbel Distribution

DEFINITION: $X$ has a Gumbel distribution (max) if the pdf of $X$ is $f(x) = e^{-x}e^{-e^{-x}}$.

## Proposition

*(a) The cdf (max) is* $F(x) = e^{-e^{-x}}$*.*
*(b)* $E(X) = \gamma$*, where* $\gamma$ *is Euler's constant,*
$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} - \ln n\right)$
*(c) Suppose that* $(X_1, X_2, \ldots)$ *is a sequence of independent random variables, each with the standard exponential distribution. The distribution of* $Y_n = max\{X_1, X_2, \ldots, X_n\} - ln(n)$ *converges to the standard Gumbel distribution as* $n \to \infty$*.*

# Definition of Laplace Transform

DEFINITION: The Laplace transform $L(s)$ of a pdf $f(x)$ with positive support is given by

$$L_X(s) = \int_0^\infty e^{-sx} f(x) dx$$

where $s > 0$.

# Catastrophe Process

THEOREM: Let $X$ be a r.v. with positive support and with pdf $f(x)$. Let $Y$ be a r.v. independent of $X$, such that $Y \sim$ exponential with rate $s$. Then

$$L_X(s) = P(X < Y).$$

# Catastrophe Process

THEOREM: Let $X$ be a r.v. with positive support and with pdf $f(x)$. Let $Y$ be a r.v. independent of $X$, such that $Y \sim$ exponential with rate $s$. Then

$$L_X(s) = P(X < Y).$$

- The exponential random variable $Y$ is called the catastrophe.

# Catastrophe Process

THEOREM: Let $X$ be a r.v. with positive support and with pdf $f(x)$. Let $Y$ be a r.v. independent of $X$, such that $Y \sim$ exponential with rate $s$. Then

$$L_X(s) = P(X < Y).$$

- The exponential random variable $Y$ is called the catastrophe.
- The Laplace transform of a p.d.f of a random variable $X$ is the probability that $X$ occurs before the catastrophe.

# Catastrophe Process

THEOREM: Let $X$ be a r.v. with positive support and with pdf $f(x)$. Let $Y$ be a r.v. independent of $X$, such that $Y \sim$ exponential with rate $s$. Then

$$L_X(s) = P(X < Y).$$

- The exponential random variable $Y$ is called the catastrophe.
- The Laplace transform of a p.d.f of a random variable $X$ is the probability that $X$ occurs before the catastrophe.
- More precisely, the Laplace transform of a probability density function $f(x)$ of a random variable $X$ can be interpreted as the probability that $X$ precedes a catastrophe where the time to the catastrophe is an exponentially distributed random variable $Y$ with rate $s$, independent of $X$.

# Laplace Transform of max(exp)

Let $X_1, \ldots, X_n$ be iid exponential r.v with common rate $\lambda = 1$.
Let $M = max X_i$.
Then $L_M(s) = P(M < Y)$ where $Y$ is exp rate $s$, so

$$L(s) = (\frac{n}{n+s})(\frac{n-1}{n-1+s})\ldots(\frac{1}{1+s})$$

# Laplace Transform of max(exp)

Let $X_1, \ldots, X_n$ be iid exponential r.v with common rate $\lambda = 1$.
Let $M = max X_i$.
Then $L_M(s) = P(M < Y)$ where $Y$ is exp rate $s$, so
$$L(s) = (\frac{n}{n+s})(\frac{n-1}{n-1+s})\ldots(\frac{1}{1+s})$$

But $prod(LT) = LT(sum)$ so $M = M_1 + \cdots + M_n$ where $M_i$ are indep with $M_i$ exp with rates $n, n-1, \ldots, 1$ respectively.

# $1 + 1/2 + 1/3 + \dots$

$E(M) = E(M_1) + E(M_2) + \cdots + E(M_n) = 1/n + 1/(n-1) + \cdots + 1$
so if $X$ is Gumbel, then
$E(X) = lim(E(M) - ln(n))$
$= lim((1 + 1/2 + 1/3 + ... + 1/n) - ln(n)) = \gamma$

# More on $1 + 1/2 + 1/3 + \ldots$

(a) $1 + 1/2 + 1/3 + ... + 1/n$ is the sum of the first $n$ terms of the harmonic series (which diverges).
(b) Although $1 + 1/2 + 1/3 + ...$ diverges, the partial sums (other than the first one) are never equal to an integer.
(b) The Riemann zeta function is $\zeta(s) = \sum_{n=0}^{\infty} \frac{1}{n^s}$.

# Poem on Riemann

Where are the zeros of zeta of s?
...
https://www.math.upenn.edu/ pemantle/songs/zeta.new
Poem by Tom Apostol on the Riemann hypothesis.

# Proof that 0=1

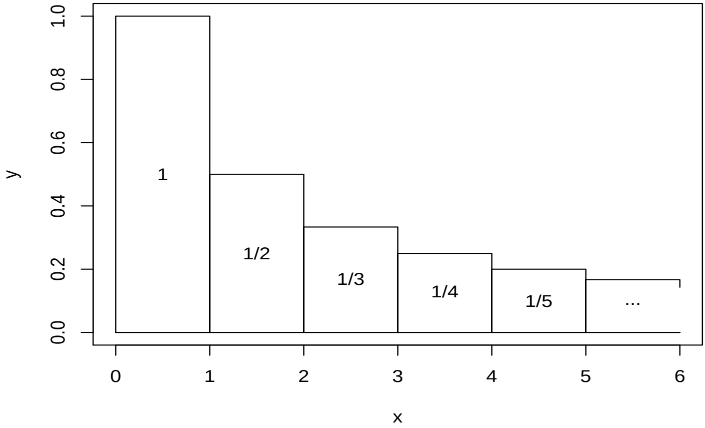
We next give a fallacious proof that $0 = 1$ based on a geometric consideration of the harmonic series

$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$

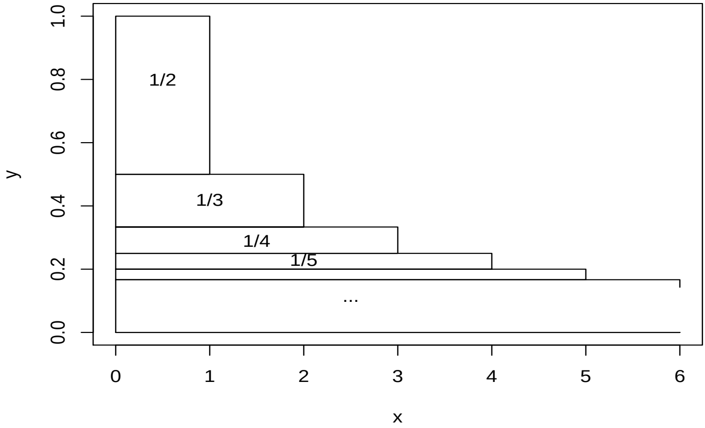
Let $1 + S = 1 + (\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots)$

The area $1 + S$ appears in the following plot with vertical strips, followed by another with horizontal strips.

# 0=1



# 0=1



From the diagram , we have that

$$1 + S = Area(Figure1) = Area(Figure2) = \frac{1}{2} + \frac{1}{3} + \cdots = S$$

From the diagram , we have that

$$1 + S = Area(Figure1) = Area(Figure2) = \frac{1}{2} + \frac{1}{3} + \cdots = S$$

so $1 = 0$.

From the diagram , we have that

$$1 + S = Area(Figure1) = Area(Figure2) = \frac{1}{2} + \frac{1}{3} + \cdots = S$$

so $1 = 0$.

Since this is false, it follows that the harmonic series diverges.

# coupon collector's problem

Let $\{X_i | i \in N\}$ be iid random variables uniformly distributed on $\{1, 2, 3, .., N\}$. Let $T_N$ be the smallest index $n$ such that $\{1, ..., N\} \subset \{X_1, \ldots, X_n\}$ , i.e. the least number of trials such that all $N$ coupons have been obtained. Then $\dfrac{T_N - Nlog(N)}{N}$ converges to the Gumbel distribution.

-from http://math.stackexchange.com/questions/563797

# Coupon collector

There are N different coupons to be collected, uniformly distributed, one per package.
What is the expected number of packages that must be opened in order to collect all N coupons?
SOLUTION:
The first package, with prob 1, gives a new coupon. The 2nd package, with prob (n-1)/1 gives a new coupon. So expected number of extra steps to get second coupon is the reciprocal, n/(n-1).
Similarly get n/(n-2) extra steps for 3rd coupon.
Expected total number of steps is $1 + n/(n-1) + n/(n-2) + \cdots + n/1$
$E(number) = n(1 + 1/2 + 1/3 + \cdots + 1/n)$

# Partitions and R

Let $p(n)$ be the number of partitions of $n$.
e.g.
$5 = 5, 4+1, 3+2, 3+1+1, 2+2+1, 2+1+1+1, 1+1+1+1+1$
so $p(5) = 7$
Let $p_N(n)$ be the number of partitions of $n$ with all parts $\leq N$
e.g. $p_3(5) = 5$
The following is from J.Roccia, P. Leboeuf.
$\frac{p_N(n)}{p(n)} \approx e^{-e^{-g}}$ where $g = \sqrt{\pi^2/(6n)}N + 1/2 ln(\pi^2/(6n))$ and
$N = o(n^{1/3})$

From Wagner

Theorem 2. Let $R_{r,n}$ be the largest integer in a random partition of $n$ that occurs at least $r$ times. Then the normalised random variable $n^{-1/2}(R_{r,n} - \frac{\sqrt{6n}}{2\pi r} logn)$ tends to a Gumbel distribution with mean $\frac{\gamma - log(\pi r/\sqrt{6})}{\pi r/\sqrt{6}}$ and variance $1/r^2$.

# Generating Random Partitions and R

```
a=rbinom(9,1,.5)
a
1 0 1 0 0 1 1 1 0
b=which(a==1)
b
1 3 6 7 8
d=c(0,b)
d
0 1 3 6 7 8
e=c(b,10)
e
1 3 6 7 8 10
f=sort(e-d)
f
1 1 1 2 2 3
```

```
g=as.data.frame(table(f))
g
f Freq
1 1 3
2 2 2
3 3 1
h=g[,2]
h
3 2 1
k=factorial(sum(h))/prod(factorial(h))
k
60
```

```
q=c();for(i in 1:100000)
{ a=rbinom(19,1,.5);b=which(a==1);d=c(0,b);e=c(b,20);
f=sort(e-d);g=as.data.frame(table(f));h=g[,2];
k=factorial(sum(h))/prod(factorial(h));
if(runif(1)<1/k) {q=c(q,1+sum(a))}}; hist(q)
```

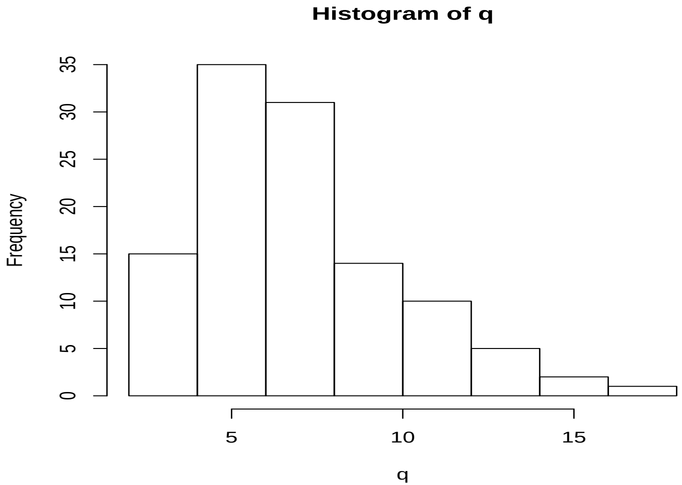



Figure: histogram of number of parts

# A better algorithm

```
n=20
x=exp(-pi/sqrt(6*n))
for(i in 1:1000) {z=rgeom(n,x);if (sum((1:n)*z)==n){print(z)}}
3 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
.
```

Source: Fristedt, The Structure of Random Partitions of Large Integer (1993).

# Partitions and Maple

Countng partitions is easy with generating functions.

$f(n) := \prod_{k=1}^{n} \left(1 + \sum_{i=1}^{\text{floor}(n/k)} x^{ki}\right)$

f(75);

simplify(%);

expand(%);

$... + 8118264x^{75} + 7089500x^{74} + ...7 * x^5 + 5 * x^4 + 3 * x^3 + 2 * x^2 + x + 1$

.

Thus $p(75) = 8117264$

.

We can also obtain partitions restricted to parts less than or equal to $m$ for $m < n$. $f(n) := \prod_{k=1}^{m} \left(1 + \sum_{i=1}^{\text{floor}(n/k)} x^{ki}\right)$

References:
1. A.Comtet, S.N. Majumdar, S.Ouvry, S.Sabhapandit. Integer partitions and exclusion statistics: Limit shapes and the largest part of Young diagrams. J. Stat. Mech. (2007)
2. J.Roccia, P. Leboeuf. Level density of a Fermi gas and integer partitions: a Gumbel-like finite-size correction. Phys.Rev.,2010
3. D.Ralaivaosaona. A phase transition in the distribution of the length of integer partitions. DMTCS proc. AQ, 2012, 265282
4. P.Erdos and J.Lehner. The distribution of the number of summands in the partitions of a positive integer. Duke Math. J., 8:335345, 1941.
5. S. Wagner. Limit Distributions of smallest gap and largest repeated part in integer partitions. The Ramanujan Journal 25/2 (2011), 229-246.
6. Stephen DeSalvo and James Y. Zhao. Random Sampling of Contingency Tables via Probabilistic Divide-and-Conquer. Tech report. 2016

# The End

The End
The End
The End