Poissonization and depoissonization

or how to enumerate the bones
of the fish without killing it

P. Jacquet
INRIA
AofA 2002, Stroebl, Austria

Plan

1. Definition of Poissonization
2. Definition of depoissonization
3. Depoissonization theorems
4. More sophisticated depoissonization theorems
5. Poissonization theorems

1 Definition of Poissonization

$\{a_n\}_{n\in N}$ a sequence of complex numbers such that $|a_n| = \exp(O(n))$.

The Poisson transform of sequence $\{a_n\}_{n\in N}$ is the generating function:

$$F(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!} e^{-z} \tag{1}$$

a_n	$F(z)$
1	1
n	z
$n(n-1)\cdots(n-k)$	z^{k+1}
$(-1)^n$	e^{-z}
α^n	$\exp((\alpha-1)z)$
$n!$	$e^{-z}/(1-z)$

Properties of Poissonization

The Poisson transform $G(z)$ is an entire function.

The Poisson transform fits well with combinatorial expressions.

Let a_n and b_n with respective Poisson transform $F(z)$ and $G(z)$.

sequence	Poisson transform
$\alpha^n a_n$	$F(\alpha z)e^{(\alpha-1)z}$
$\sum_k \binom{n}{k} a_k$	$F(z)e^z$
$\sum_k \binom{n}{k} a_k b_{n-k}$	$F(z)G(z)e^z$
a_{n+1}	$\frac{\partial}{\partial z}F(z) + F(z)$
a_{n+k}	$\sum_i \binom{k}{i} \frac{\partial^i}{\partial z^i} F(z)$

Poisson transform of binomial $\sum_k \binom{n}{k} p^k q^{n-k} a_k b_{n-k}$ with $p + q = 1$ is $F(pz)G(qz)$.

Probabilistic interpretation

Let $\mathbf{N}$ be an integer variable with Poisson distribution of parameter λ

$$P(\mathbf{N} = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

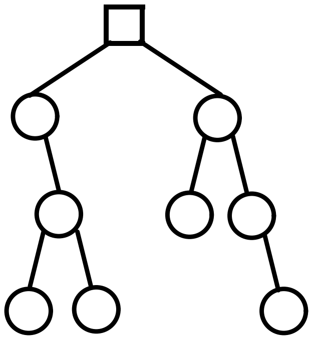
We have $\mathrm{E}[a_{\mathbf{N}}] = F(\lambda)$.

Generalization to complex values of Poisson parameter λ.

Poissonization was introduced by M. Kac (1949) on limited to real arguments.

Biased tries

Bourdon, Flajolet, Knessl, PJ, Régnier, Szpnkwsk, Vallée, ...



Average Size

$s_n = \mathrm{E}[S_n]$: $s_0 = s_1 = 1$ and
$s_n = 1 + \sum_k \binom{n}{k} p^k q^{n-k}(s_k + s_{n-k})$, Poisson transform:

$$S(z) = 1 + S(pz) + S(qz) - 2(1+z)e^{-z} \qquad (2)$$

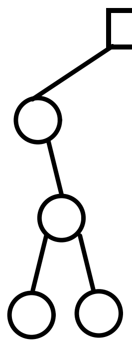
Average Depth

$d_n = n\mathrm{E}[D_n]$: $d_0 = 0$, $d_1 = 1$, and
$d_n = n + \sum_k \binom{n}{k} p^k q^{n-k}(d_k + d_{n-k})$, Poisson transform:

$$D(z) = z + D(pz) + D(qz) - ze^{-z} \qquad (3)$$

Reader Election

Fill, Janson, Mahmoud, Prodinger, ...



$\ell_n = \mathrm{E}[L_n]$. $\ell_0 = \ell_1 = 1$,
$\ell_n = 1 + 2^{-n}(\ell_n - 1) + 2^{-n} \sum_k \binom{n}{k} \ell_k$:

$$L(z) = 1 - (1+z)e^{-z} + (1 + e^{-z/2})L(\frac{z}{2}) \qquad (4)$$

Digital Search Tree

Kirschenhofer, PJ, Prodinger, Szpnkwsk, ...

Average External Path Length

$h_n = \mathrm{E}[H_n]$, $h_0 = 0$ and $h_1 = 1$:

$$\frac{\partial}{\partial z} H(z) + H(z) = H(pz) + H(qz) + z \qquad (5)$$

Many other important application of Poissonization

Divide and Conquer recurence, Derfel-Vogl (99)

Dynamic sources in Information Theory
Clément-Vallée (99)

Asymptotics

Mellin transform of Poisson **trie average size**, $S^*(s) = \int_0^\infty S(x)x^{s-1}ds$: $S^*(s) = -2\frac{(1+s)\Gamma(s)}{1-2^{1+s}}$

Simple pole at $s = -1$ translates into a $O(z)$ asymptotics for $z \to +\infty$:

$$S(z) = 2\frac{z}{\log 2}(1 + P_1(\log z)) + O(1), \qquad (6)$$

$P_1(x)$ is Mellin transform of Poisson **trie average depth** $D^*(s) = -\frac{s\Gamma(s)}{1-2^{1+s}}$

Double pole at $s = -1$ translates into a $O(z \log z)$ asymptotics for $z \to +\infty$:

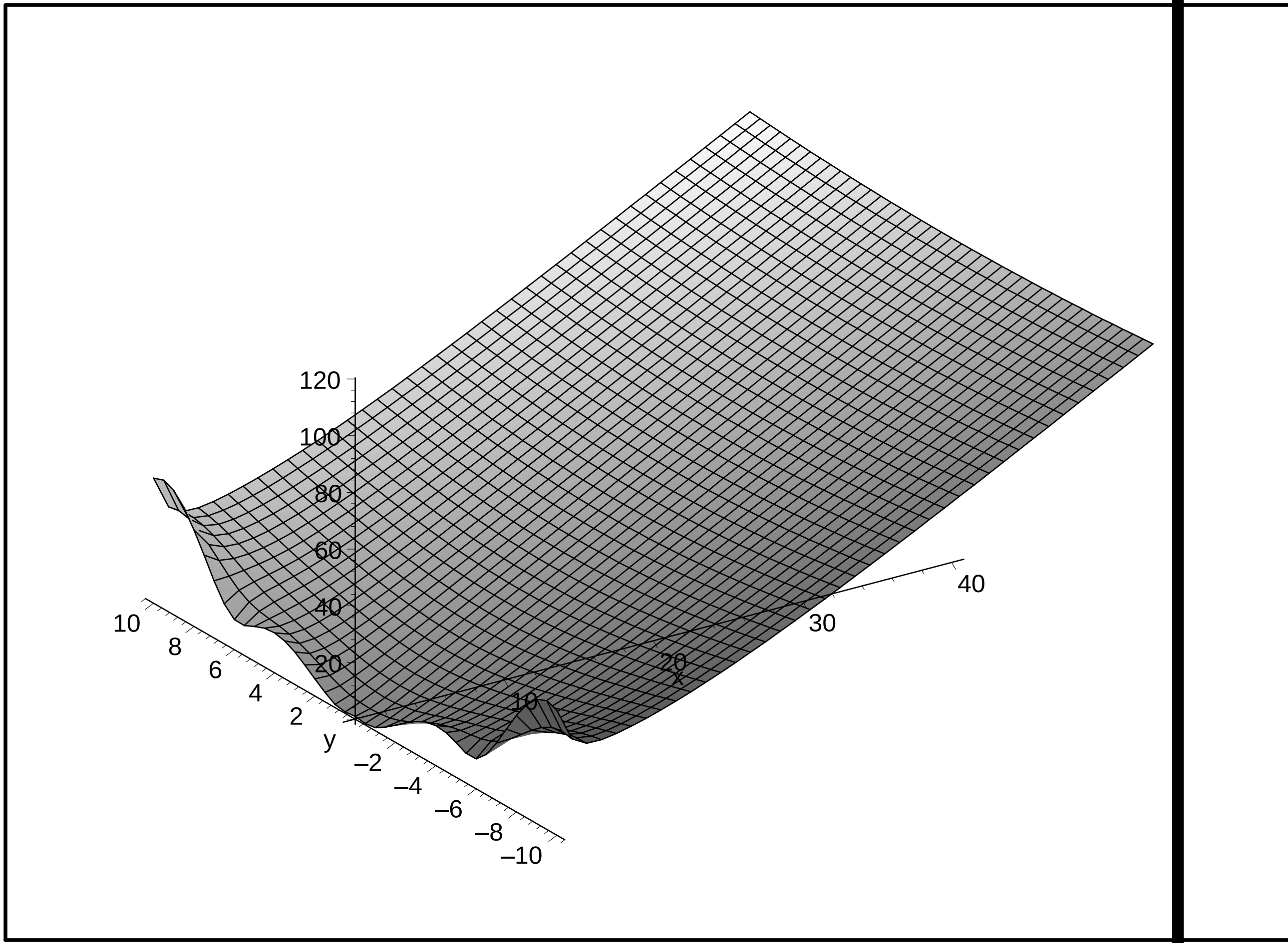
$$D(z) = \frac{z \log z}{\log 2}(1 + P_2(\log z)) + O(1) \qquad (7)$$

Reference on Mellin transform:
Dumas-Flajolet-Gourdon (95).

Asymptotics of **average reader election**

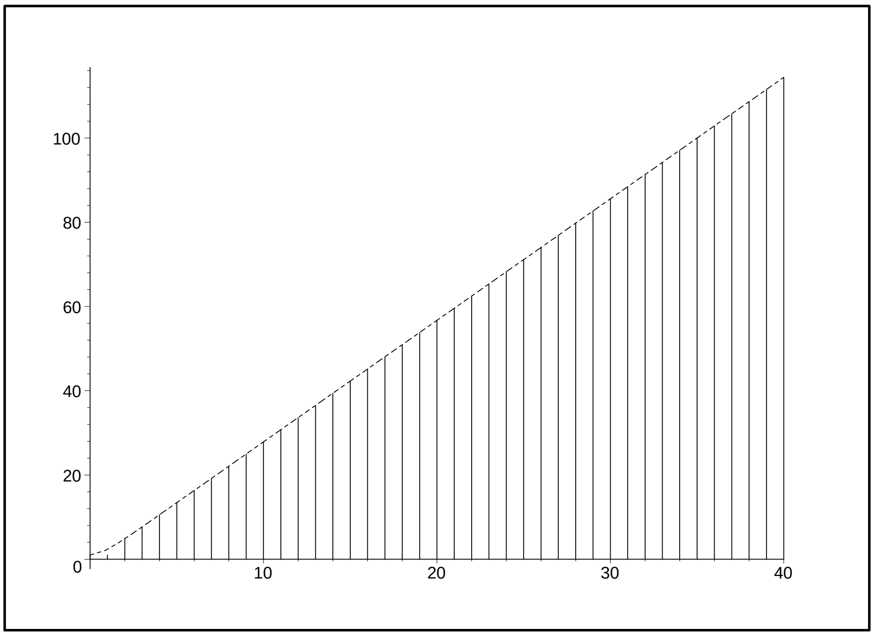
$$L(z) = \frac{\log z}{\log 2}(1 + P_3(\log z)) + O(1/z) \tag{8}$$

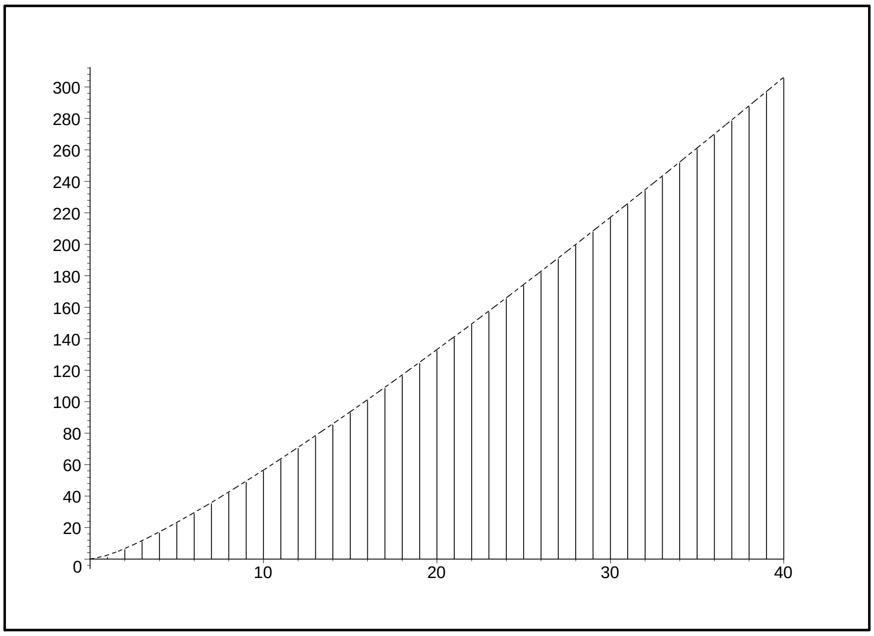
Remark The asymptotics hold when z is in a complex cone around R^+.



Asymptotics on coefficients

It seems that the coefficients have the same asymptotics than the Poisson transform: $s_n \approx S(n)$, and $d_n \approx D(n)$.



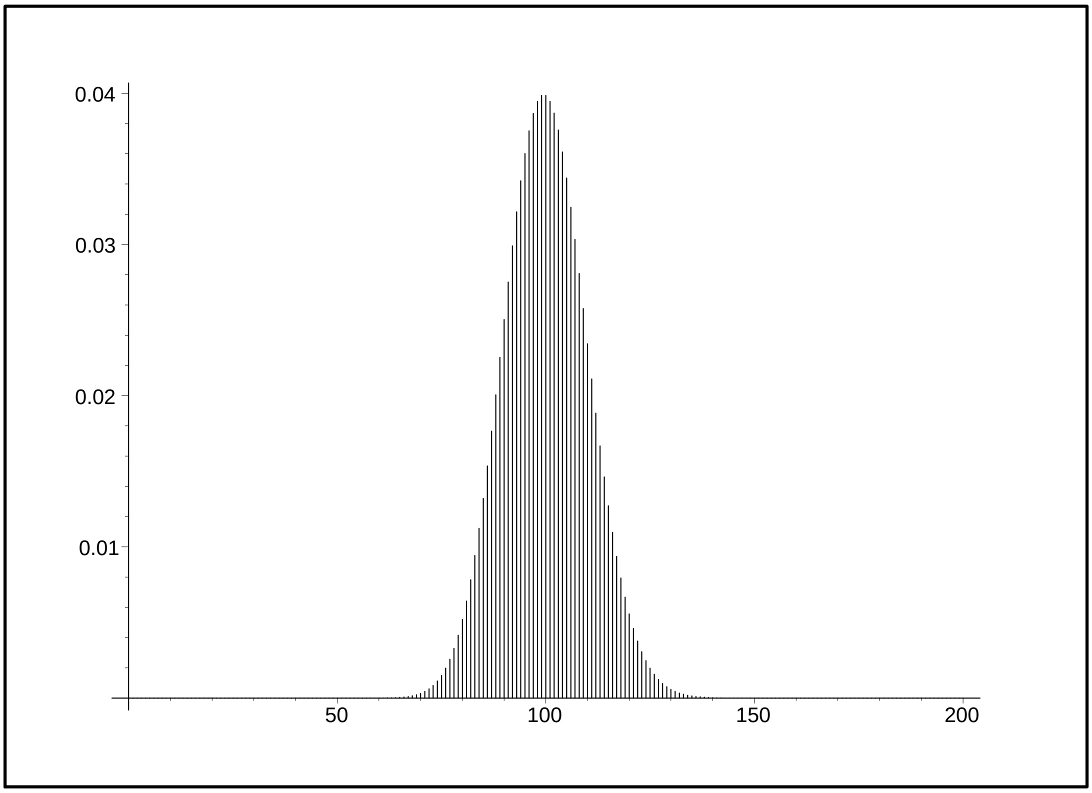


Asymptotics on coefficients

The sequence $\frac{z^n}{n!}e^{-z}$ sums to 1. For large $z \in R^+$, the sequence is close to a Gaussian law with mean z and std deviation $\sqrt{z}$:

$$\mathrm{E}[u^{\mathbf{N}}] = \exp(z(u-1))$$

$$e^{-\sqrt{z}t}\mathrm{E}[e^{\mathbf{N}t/\sqrt{z}}] = \exp(\frac{t^2}{2} + O(z^{-1/2}))$$



2 Definition of dePoissonization

From empiric measurement, when a_n has a smooth growth (e.g. polynomial) then $F(n) \approx a_n$.

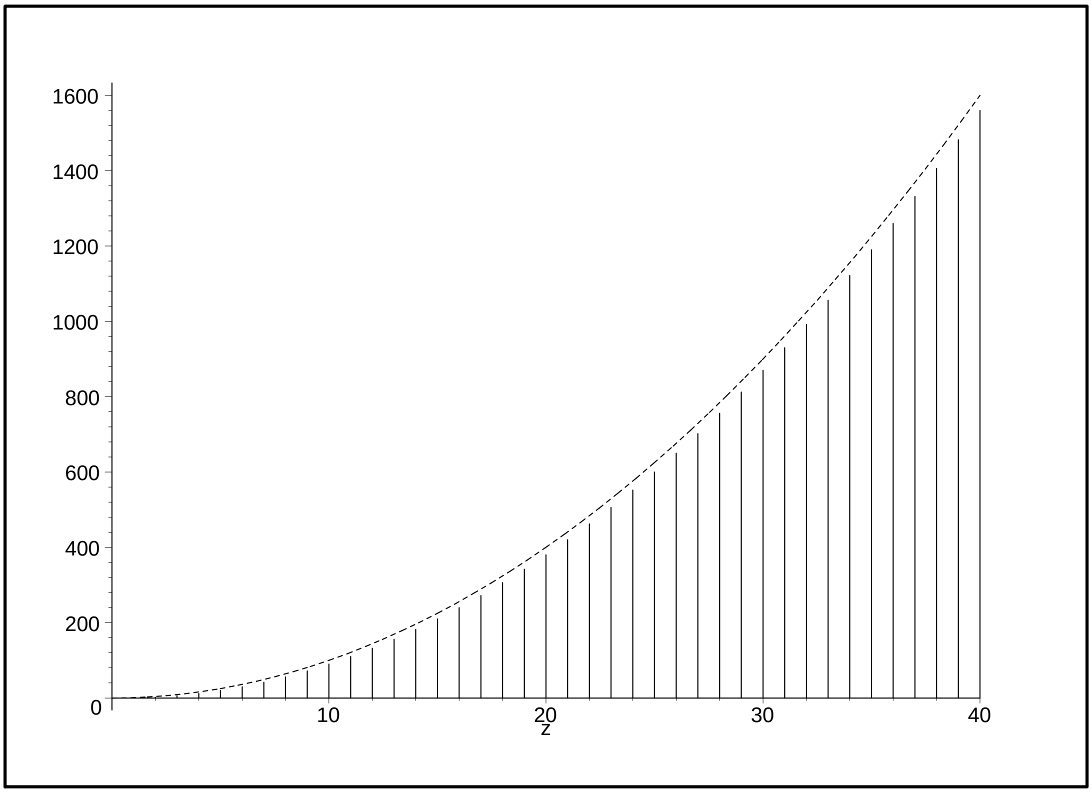


Figure 1: z^2 and its coefficients

The heuristic is not always true: $a_n = (-1)^n$, $F(z) = e^{-2z}$

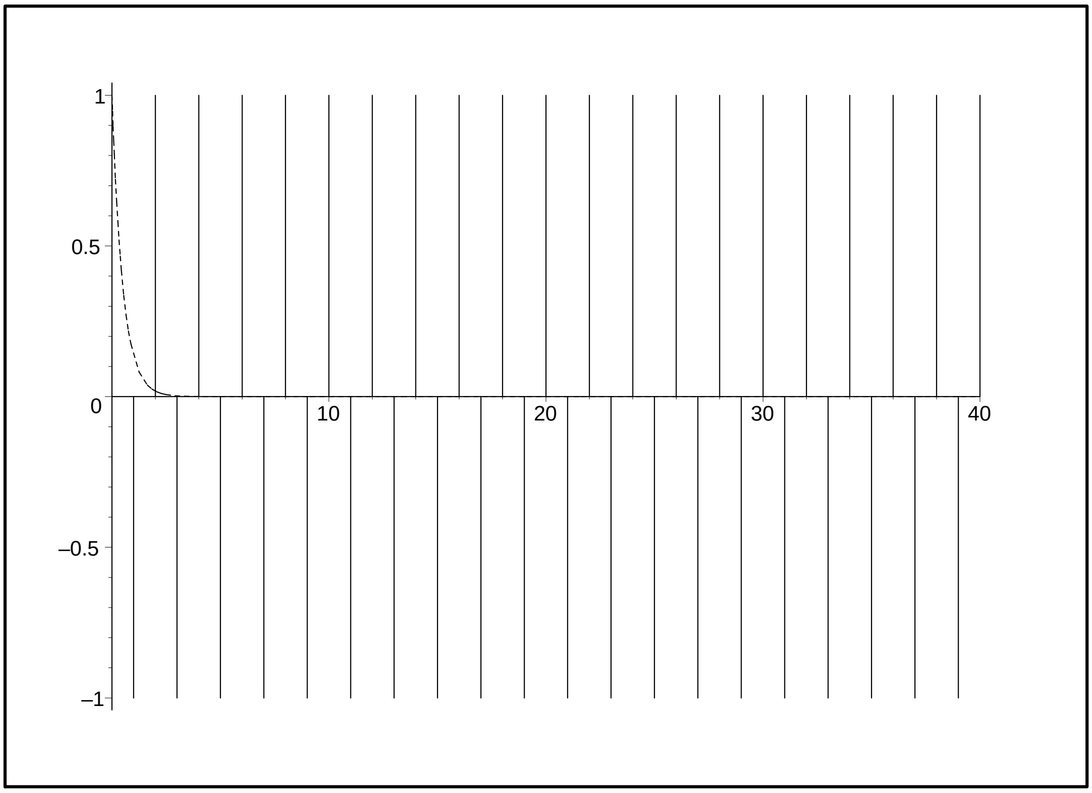


Figure 2: Poisson transform of $(-1)^n$

How to derive the asymptotics of the sequence from the asymptotics of the Poisson transform?

Or: how to enumerate the bones of the fish without removing the skin.

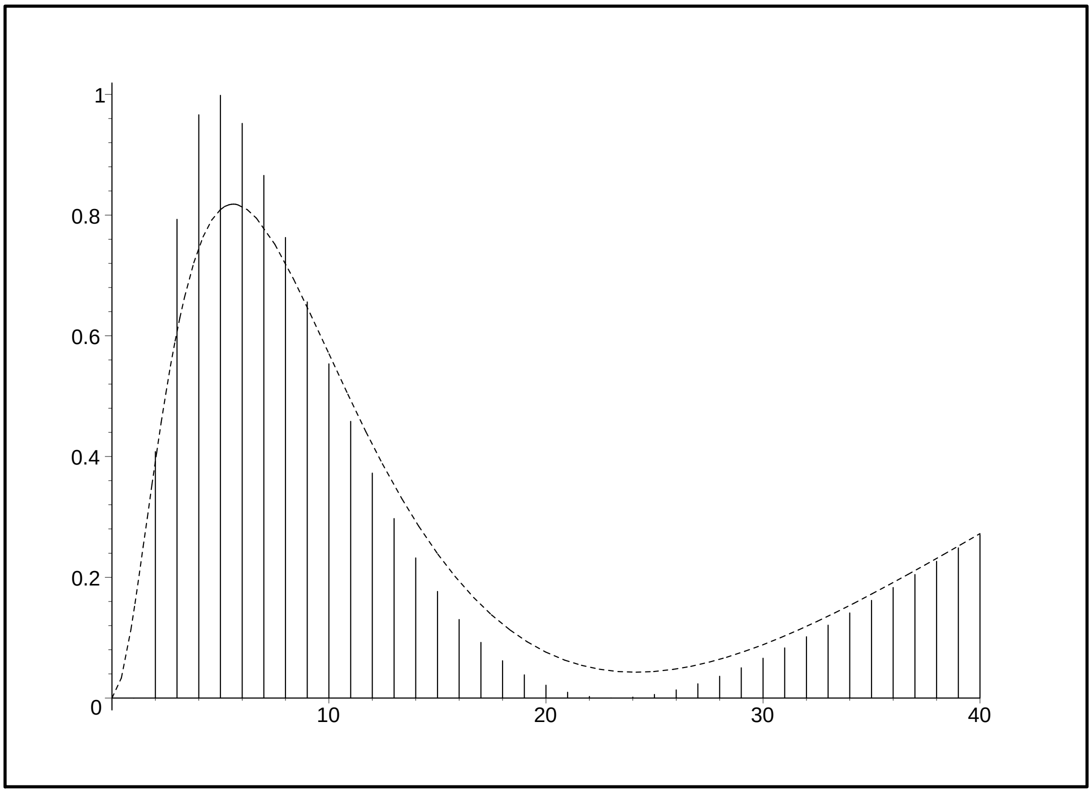


Figure 3: Poisson transform of $\sin^2(\log n)$

We need a Tauberian theorem.

3 Depoissonization theorem

Basic (Version 0)

Theorem 1 *Let $F(z)$ be an entire function and a complex cone $\mathcal{S}_{\theta_0}$ with $\theta < \frac{\pi}{2}$ such that*

(i) $z \in \mathcal{S}_{\theta_0}$*:* $F(z) = O(z^\beta)$

(ii) $z \notin \mathcal{S}_{\theta_0}$*:* $F(z)e^z = O(e^{\alpha|z|})$*. for some* $\alpha < 1$*.*

Then $a_n = F(n) + O(n^{\beta-1/2})$*.*

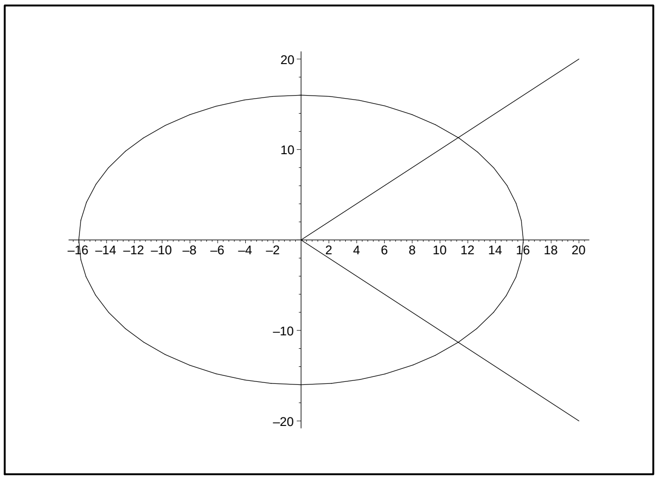
Case where condition (i) does not hold
$a_n = (1 + A)^n$ for some $A > 0$: $F(z) = e^{Az}$.

Case where condition (ii) does not hold
$a_n = (-1)^n$, $F(z) = e^{2z}$.

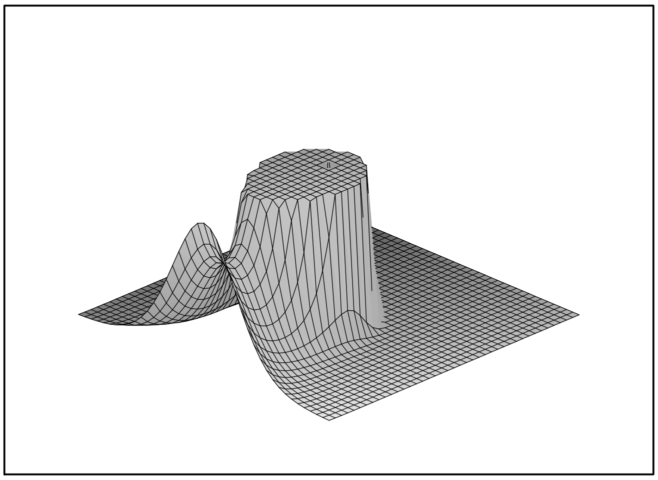
Historically this version is from PJ-Régnier (87-89)

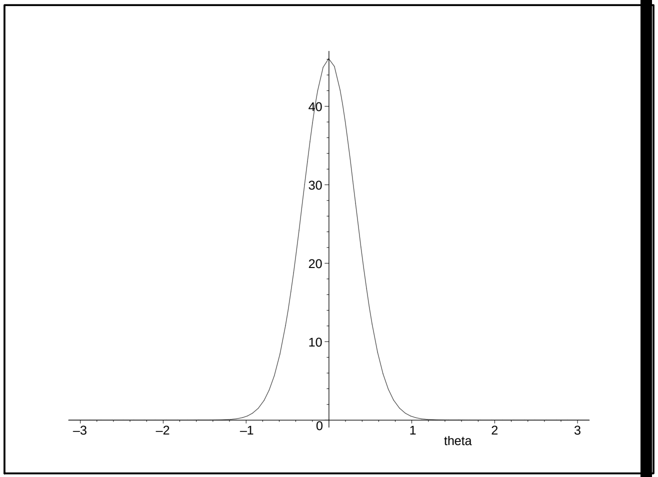
Proof of theorem

$$a_n = \frac{1}{2i\pi} \oint F(z) \frac{n! e^z}{z^{n+1}} dz$$



Integration path on $|z| = n$





Outside the cone $\mathcal{S}_{\theta_0}$

Contribution in $O(\frac{n!}{n^n}e^{\alpha n})$ exponentially small

inside the cone

$z = ne^{i\theta/\sqrt{n}}$: $\frac{n!e^z}{z^{n+1}} = \sqrt{2\pi n}\exp(-\theta^2/2 + O(n^{-1/2}\theta^3))$
asymptotically gaussian

Since $F'(z) = O(z^{\beta-1})$ in the cone:
$F(z) = F(n) + O(n^{\beta-1/2})$.

Contribution: $F(n) + O(n^{\beta-1/2})$. ■

DePoissonization theorem (version 1)

Theorem 2 *Under same condition of theorem 1, we have the expansion that holds for every fixed m:*

$$a_n = \sum_{i=0}^{m} \sum_{j=0}^{i+m} b_{ij} n^i F^{\langle j \rangle}(n) + O(n^{\beta - m - 1/2})$$

where $\sum_{ij} b_{ij} x^i y^j = \exp(x \log(1+y) - xy)$ *(Stirling numbers).*

Remark: Historically the expansion is due to Gonnet-Munro (84) but without the validity conditions.

Proof: Using expansion of $\frac{n! e^z}{z^n}$ and of $F(z)$ in the cone (since $F^{\langle k \rangle}(z) = O(z^{\beta - k})$);

using dominating convergence argument we prove that there is an expansion of a_n in $\sum b_{ij} n^i F^{\langle j \rangle}(n)$ for some b_{ij}.

Identification of the b_{ij} is done formally on a "probing" sequence of functions. ∎

Short-cut (but wrong) probing sequence:

$a_n = (1+y)^n,\ F(z) = e^{zy},\ F^{\langle k\rangle}(z) = y^k e^{zy}.$

$$\begin{aligned} a_n = &\quad \exp(n\log(1+y) - ny)e^{ny} \\ = &\quad \textstyle\sum_{ij} b_{ij} n^i y^j e^{ny} \\ = &\quad \textstyle\sum_{ij} b_{ij} n^i F^{\langle j\rangle}(n) \end{aligned}$$

i/j	0	1	2	3	4	5	6
0	1	0	0	0	0	0	0
1	0	0	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{4}$	$\frac{1}{5}$	$-\frac{1}{6}$
2	0	0	0	0	$\frac{1}{4}$	$-\frac{1}{6}$	$\frac{13}{72}$
3	0	0	0	0	0	0	$-\frac{1}{6}$

Application: the variance of tries

$q_n = \mathrm{E}[S_n^2]$, $V(z) = Q(z) - (S(z))^2$:

$$V(z) = V(pz)+V(qz)+(2S(z)-1+(1+z)e^{-z})(1+z)e^{-z} \tag{9}$$

$V(z) = C_1 z(1 + P_4(\log z)) + O(1) \quad (p = q = 1/2)$ and $Q(z) = O(z^2)$.

$$\begin{aligned} v_n &= q_n - (s_n)^2 \\ &= Q(n) - \frac{n}{2}Q^{\langle 2\rangle}(n) - (S(n) - \frac{n}{2}S^{\langle 2\rangle})^2 + O(1) \\ &= V(n) - \frac{n}{2}(S'(n))^2 + O(1). \end{aligned}$$

Term $n(S'(n))^2$ is of same order as term $V(n)$ and cannot be neglected: Poisson differs from Bernoulli on second moment (PJ-Régnier 87).

4 Even more sophisticated dePoissonization theorems

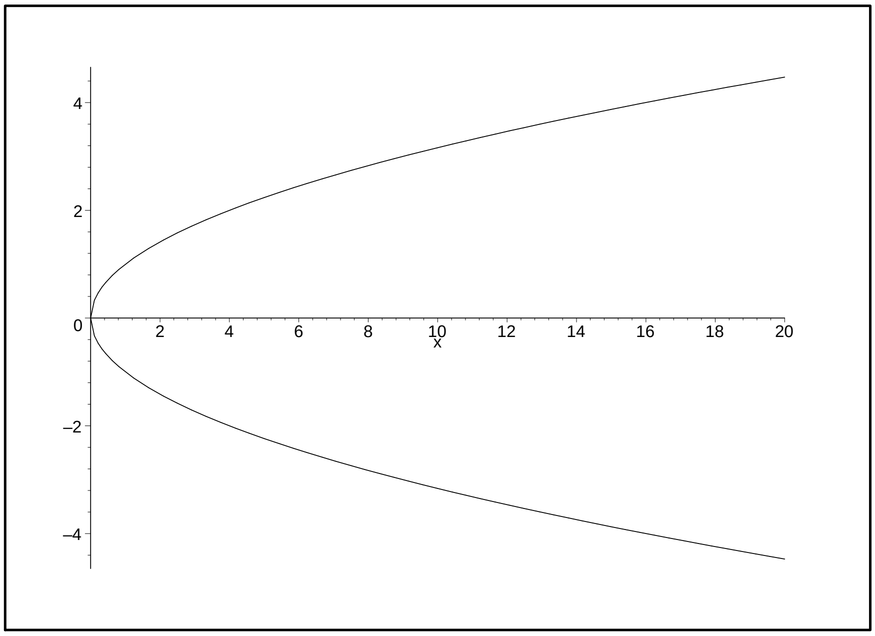
When the fish is harder to hold, or when there are too many fishes.

DePoissonization theorem (version 2)

Theorem 3 *Same conditions as in theorem 2, excepted that cone $\mathcal{S}_{\theta_0}$ is replaced by algebraic cone $\mathcal{C}(D,\delta) = \{z = x + iy,\ |y| < Dx^\delta\}$. Then*

$$a_n = \sum_{i=0}^{m} \sum_{j=0}^{i+m} b_{ij} n^i F^{\langle j \rangle}(n) + O\big(n^{\beta-(2\delta-1)m-1/2}\big)$$

The proof relies on the fact that $F^{\langle k \rangle} = O\big(z^{\beta-k\delta}\big)$ in the algebraic cone. ∎



Application: external path length in DST.

DePoissonization Theorem (diagonal DePo)

This is used for proving uniform error terms on bivariate Poisson transforms.

Theorem 4 *Let* $F_m(z) = \sum a_{n,m} \frac{z^n}{n!} e^z$ *be a sequence of Poisson transforms, if there exists a cone* $\mathcal{S}_{\theta_0}$ *and constant* B*,* $D > 0$ *and* $\alpha < 1$ *such that*

(i) $z \in \mathcal{S}_{\theta_0},\ |z| \in (n - Dn, n + Dn)$*:* $|F_n(z)| \leq Bn^{\beta}$

(ii) $z \notin \mathcal{S}_{\theta_0},\ |z| = n$*:* $|F_n(z)e^z| \leq e^{n-n^{\alpha}}$

Then

$$a_{n,n} = \sum_{i=0}^{m} \sum_{j=0}^{i+m} b_{ij} n^i F_n^{\langle j \rangle}(n) + O(n^{\beta-m-1/2})$$

DePoissonization Theorem (diagonal exponential DePo)

Theorem 5 *Sequence of Poisson transforms* $F_n(z)$. *There exists a cone* $\mathcal{S}_{\theta_0}$ *where* $\log(F_n(z))$ *exists. Let* $\beta \in (1/2, 2/3)$ *and* A, $B > 0$, $\alpha < 1$.

(i) $z \in \mathcal{S}_{\theta_0}$, $|z| \in (n - Dn, n + Dn)$: $|F_n(z)| \le Bn^\beta$

(ii) $z \notin \mathcal{S}_{\theta_0}$, $|z| = n$: $|F_n(z)e^z| \le e^{n - An^\alpha}$

then

$$a_{n,n} = F_n(n) \exp(-\frac{n}{2}(\frac{F'(n)}{F(n)})^2)(1 + O(n^{3\beta-2})) \quad (10)$$

Even more sophisticated theorems with combination of algebraic cones, diagonal and exponential dePoissonization, etc.

Applications

Depth distribution

$d_n(u) = n\mathrm{E}[u^{D_n}]$. $d_0(u) = 0$, $d_1(u) = u$,
$d_n(u) = u\sum_k \binom{n}{k} p^k (1-p)^{n-k}$:

$$D(z,u) = u(D(pz,u) + D(qz,u)) - (u-1)ze^{-z} \quad (11)$$

- Asymptotic analysis $D(z,e^t) = z^{s(t)}(1+o(1))$, with $p^{s(t)} + q^{s(t)} = e^{-t}$.
- By diagonal dePo: $D_n(e^t) = n^{s(t)}(1+o(1))$
- $s(t) = \frac{t}{h} + \frac{C_2 t^2}{2} + O(t^3)$, with $C_2 > 0$ when $p \neq q$, when $p = q = 1/2$ $C_2 = 0$.

Thus

$$e^{-t\sqrt{\log n}/h} D_n(e^{t/\sqrt{\log n}}) \to \exp(C_2 \frac{t^2}{2}) \quad (12)$$

Normal limiting distribution when $C_2 > 0$.

Size distribution

$s_n(u) = \mathrm{E}[u^{S_n}]$, $s_0(u) = s_1(u) = u$,
$s_n = u \sum_k \binom{n}{k} p^k q^{n-k} s_k s_{n-k}$:

$$S(z,u) = uS(pz,u)S(qz,u) - (u^2-1)u(1+z)e^{-z} \quad (13)$$

- Asymptotic analysis
  $\log(S(z,e^t)) = S(z)t + V(z)\frac{t^2}{2} + O(zt^3)$,
- Diagonal dePo: $F_n(z) = e^{-ts_n/\sqrt{v_n}} S(z, e^{t/\sqrt{v_n}})$,
- when $z = O(n)$: $\log F_n(z) =$
  $\exp(\frac{S(z)-s_n}{\sqrt{v_n}} + \frac{V(z)}{v_n}t^2/2 + o(n^{-1/2})) = O(n^{1/2})$;
- Diagonal exponential dePo:
  $e^{-ts_n/\sqrt{v_n}} s_n(e^{t/\sqrt{v_n}}) =$
  $\exp(\log F_n(n) - \frac{n}{2}((\log F_n(n))')^2)(1 + O(n^{-1/2}))$.

Since $\log F_n(n) = -\frac{v(n)}{2v_n}t^2 + O(n^{-1/2})$ and $(\log F_n(n))' = \frac{s'(n)}{\sqrt{v_n}}t + O(n^{-1})$ then

$\log F_n(n) = \frac{v(n)-n(s'(n))^2/2}{v_n}t^2/2 + o(n^{-1/2}) = \frac{t^2}{2} + o(n^{-1/2})$, the distribution is asymptotically normal.

This fish was hard!

Digital Search Trees

Path Length

$h_n(u) = \mathrm{E}[u^{H_n}]$, $h_0(u) = 1$,
$h_{n+1}(u) = u^n \sum_k \binom{n}{k} p^k q^{n-k}$.

$$\frac{\partial}{\partial z}\mathcal{H}(z,u) = \mathcal{H}(puz,u)\mathcal{H}(quz,u)\ . \qquad (14)$$

with $\mathcal{H}(z,u) = H(z,u)e^z$ (exponential PGF).

(Ugly) asymptotic analysis provides
$\log \mathcal{H}(z,e^t) = O(z^\kappa(t))$ with $(pe^t)^{\kappa(t)} + (qe^t)^{\kappa(t)} = 1$.
See retarded equation by Drmota.

Similar to tries size but harder fish, requires polynomial cone in order to prove limiting normal distribution.

Used to get redundancy of Ziv-Lempel compression algorithm (Louchard, Szpnkwsk)

5 Poissonization theorems

Or how to imagine the skin of the fish when only the bones remain

Poissonization Theorem Part 1

Theorem 6 *Let $g(z)$ be an analytic continuation of sequence a_n: $a_n = g(n)$, and $F(z)$ the Poisson transform of a_n. Let a cone $\mathcal{S}_{\theta_0}0$ and a parameter β such that*

(i) $z \in \mathcal{S}_{\theta_0}0$*:* $g(z) = O(z^\beta)$

then for all integer m and $z \in \mathcal{S}_{\theta_0}0$:

$$F(z) = g(z) + \sum_{i=1}^{m} \sum_{j=i+1}^{m} a_{ij} z^i g^{\langle j \rangle}(z) + O(z^{\beta - m}) \quad (15)$$

with $\sum_{ij} a_{ij} x^i y^j = \exp(x(e^y - 1) - xy)$. ($a_{ij}$ second associate Stirling numbers).

Easy proof based on asymptotics on Poisson coefficients.

As (wrong) short-cut $g(x) = e^{xy}$,

$$\begin{aligned} F(z) &= \exp((e^x - 1)z) = \exp(z(e^x - 1) - xz)e^{xz} \\ &= \sum_{ij} a_{ij} z^i x^j e^{xz} \\ &= \sum_{ij} a_{ij} z^i g^{\langle j \rangle}(z) \end{aligned}$$

Poissonization theorem Part 2

Theorem 7 *Let $g(z)$ be an analytic continuation of sequence a_n: $a_n = g(n)$, and $F(z)$ the Poisson transform of a_n. Let a cone $\mathcal{S}_{\theta_0}0$ and a parameter β such that*

(i) $z \in \mathcal{S}_{\theta_0}0$: $g(z) = O(z^\beta)$

Then for all cone $\mathcal{S}_{\theta_0}$ $(\theta > 0)$ there exists $A > 0$ and $\alpha < 1$ such that

$$z \notin \mathcal{S}_{\theta_0} \Rightarrow |F(z)e^z| \leq Ae^{\alpha|z|}$$

Proof: We refer to Laplace transform $g^*(s) = \int_0^\infty g(x)e^{-xs}dx$ exists for $\Re(s) > 0$. Using reverse Laplace $g(n) = \int_{\mathcal{L}} g^*(s)e^{ns}ds$ we have the identity:

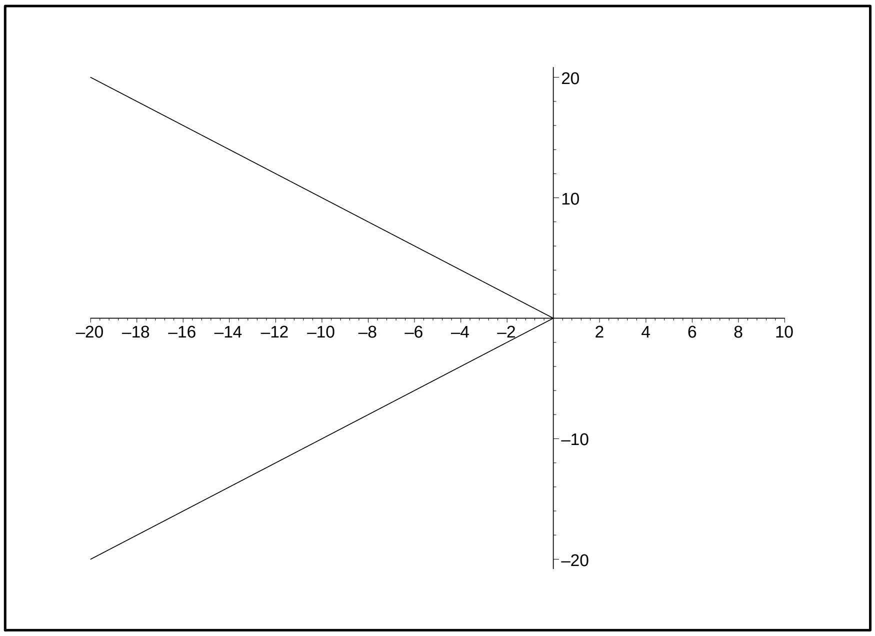
$$F(z)e^z = \int_{\mathcal{L}} g^*(s)e^{e^s z}ds \tag{16}$$

where $\mathcal{L}$ is a piece linear curve that parallel the boundary of the definition domain of $g^*(s)$, e.g. $(\varepsilon - i\infty, \varepsilon + i\infty)$.

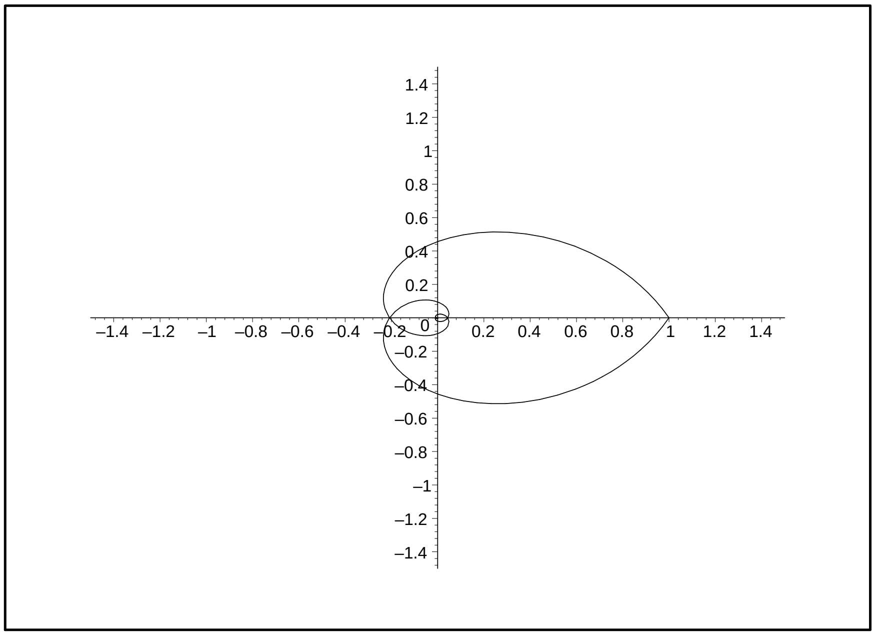
Lemma 1 *The Laplace transform can be extended on cone* $\mathcal{S}_{\theta_0+\pi/2}$.

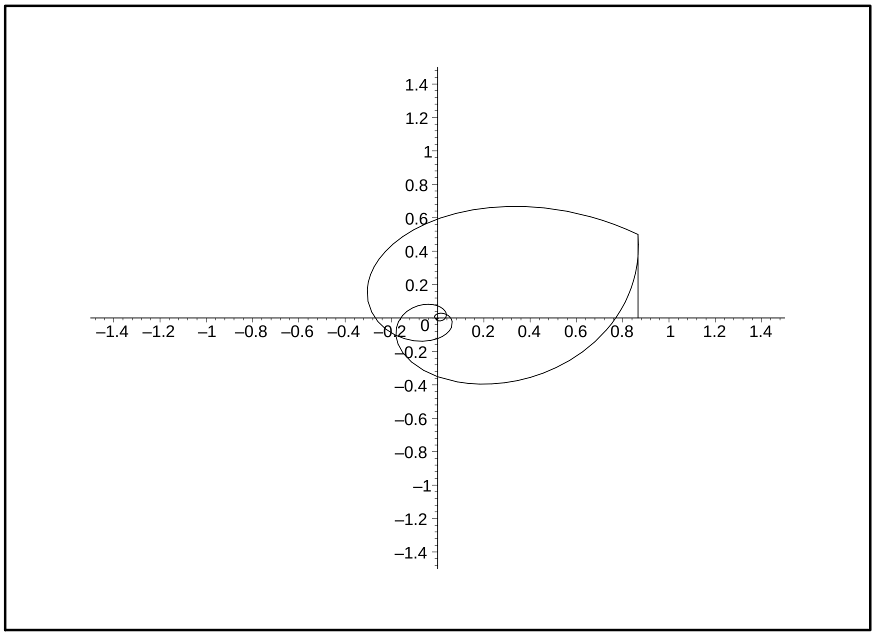
Proof The integral $\int_0^\infty g(e^{i\theta}x)e^{-xs}dx$ is defined for $Re(s) > 0$ and is equal to $e^{-i\theta}g^*(e^{-i\theta}s)$. ■

One takes for $\mathcal{L}$ a curve close to the border of $\mathcal{S}_{\theta_0+\pi/2}$:



Lemma 2 *Given $\theta > 0$ we have*
$\max_{s\in\mathcal{L}} \Re(e^{s+i\theta}) < 1.$





End of proof for Poissonization

Let $z \notin \mathcal{S}_{\theta_0}$ for a given $\theta > 0$, $(\arg(z) > \theta)$.

Let $A = \int_{\mathcal{L}} |g^*(s) ds|$ and $\alpha = \max_{s \in \mathcal{L}} \Re(e^{s+i\theta}) < 1$

$$
\begin{aligned}
|F(z)e^z| &\leq \int |g^*(s) ds| \exp(\Re(e^{s+i\theta})|z|) \\
&\leq A \exp(\alpha |z|)
\end{aligned}
$$

$\blacksquare$

Evaluation of binomial entropy

Sequence $h_n^B = -\sum_k \binom{n}{k} p^k q^{n-k} \log(\binom{n}{k} p^k q^{n-k})$.

- We have $h_n^B = -\log n! + hn + b_n$ with $b_n = \sum_k \binom{n}{k} p^k q^{n-k} (\log k! + \log(n-k)!)$.
- Poisson transform of b_n is $F(pz) + F(qz)$, $F(z)$ being Poisson transform of $\log n!$.
- $\log n!$ has analytic expansion $\log \Gamma(z+1) = (z+\frac{1}{2}) \log z - \frac{1}{2} z + \log \sqrt{2\pi} + O(1/z)$.
- Poissonization: $F(z) = (z+\frac{1}{2}) \log z - \frac{1}{2} z + \log \sqrt{2\pi} + C_3 + O(1/z)$, for some computable constant C_3.
- DePo of $F(pz) + F(qz)$ leads to

$$h_n = \frac{1}{2} \log n + \frac{1}{2} + \log \sqrt{2\pi pq} + 2C_3 + O(1/n). \tag{17}$$

Most material of this talk in

Article: "Analytical depoissonization and its application" PJ-Szpnkwsk (98).

Book: *Average Case Analysis of Algorithms on Sequences* Szpankowski (00).