The Depoissonization Quintet

Poisson–Mellin–Newton–Rice–Laplace

Brigitte Vallée
CNRS et Université de Caen

Séminaire Flajolet, IHP, 3 Octobre 2019

Plan of the talk

Two probabilistic models,

the Bernoulli model and the Poisson model.

with two tools,

the Poisson transform, the Poisson sequence.

Two "return" paths from the Poisson model to the Bernoulli model

- The first path : Depoissonization path with the Poisson transform.
- The second path : Rice path with the Poisson sequence.

Here, in this talk:

- Survey and comparison of the two paths
- Are they really different?
- It is true the Rice path be more restrictive to use?
- Study of the Rice path using the Laplace transform

I. General framework.

General framework.

Begin with (elementary) data

Consider algorithms which use as inputs finite sequences of data

If $\mathcal{X}$ is the set of data, then the set of inputs is $\mathcal{X}^\star = \bigcup_{n\geq 0} \mathcal{X}^n$

Context	(elementary) data	input	Study
source	a symbol from an alphabet	a (finite) word	entropy
text	an (infinite) word	a sequence of words	dictionary
geometry	a point	a sequence of points	convex hull

Probabilistic studies.

- The set $\mathcal{X}$ is endowed with probability $\mathbb{P}$
- The set $\mathcal{X}^N$ is endowed with probability $\mathbb{P}_{[N]}$

In cases (2) and (3), very often, the data are independently drawn with $\mathbb{P}$

Not in case (1) where the successive symbols may be strongly dependent.

Two probabilistic models.

Space of inputs := the set $\mathcal{X}^\star$ of the finite sequences $\boldsymbol{x}$ of elements of $\mathcal{X}$.
There are two main probabilistic models on the set $\mathcal{X}^\star$.

- The Bernoulli model $\mathcal{B}_n$, where the cardinality N of the sequence $\boldsymbol{x}$ is fixed to n (then $n \to \infty$);
  The Bernoulli model is more natural in algorithmics.

- The Poisson model $\mathcal{P}_z$ of parameter z, where the cardinality N of the sequence $\boldsymbol{x}$ is a random variable that follows a Poisson law of parameter z,
  $$\Pr[N = n] = e^{-z}\frac{z^n}{n!},$$
  (then $z \to \infty$). The Poisson model has nice probabilistic properties, notably independence properties $\Longrightarrow$ easier to deal with.

$\Longrightarrow$ A first study in the Poisson model,

followed with a return to the Bernoulli model

Average-case analysis of a cost R defined on $\mathcal{X}^\star$

$\mathcal{X}^\star :=$ set of the finite sequences of elements of $\mathcal{X}$

- Final aim : Study the sequence $n \mapsto f(n)$,
  $f(n) := \mathbb{E}_{[n]}[R] :=$ the expectation in the Bernoulli model $\mathcal{B}_n$
- Consider the expectation $\mathbb{E}_z[R]$ in the Poisson model $\mathcal{P}_z$

$$\mathbb{E}_z[R] = \sum_{n \geq 0} \mathbb{E}_z[R \mid N = n] \, \mathbb{P}_z[N = n]$$

$$= \sum_{n \geq 0} \mathbb{E}_{[n]}[R] \, \mathbb{P}_z[N = n] = e^{-z} \sum_{n \geq 0} f(n) \frac{z^n}{n!}$$

$\mathbb{E}_z[R]$ is the Poisson transform $P_f(z)$ of the sequence $n \mapsto f(n)$.

- With (properties of) the Poisson transform $P_f(z)$ of f
  return to (the asymptotics of) the sequence $n \mapsto f(n)$

The Poisson transform and the Poisson sequence

With a sequence $f : n \mapsto f(n)$ [the expectations in the $\mathcal{B}_n$ models],

we associate $$P_f(z) = e^{-z} \sum_{k \geq 0} f(k) \frac{z^k}{k!} = \sum_{k \geq 0} (-1)^k \frac{z^k}{k!} p(k)$$

- The series $P_f(z)$ is the Poisson transform of $n \mapsto f(n)$, also the expectation in the Poisson model $\mathcal{P}_z$
- The sequence $k \mapsto p(k)$ is the Poisson sequence of $n \mapsto f(n)$. It is denoted by $\Pi[f]$. The map Π is involutive.

Important involutive binomial relation between $f(n)$ and $p(n)$

$$p(n) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} f(k), \quad \text{and} \quad f(n) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} p(k).$$

Description of the two possible paths.

Begin with a sequence $k \mapsto f(k)$ of polynomial growth,

consider its Poisson transform $P_f(z)$ and its Poisson sequence $\Pi[f] : n \mapsto p(n)$,

$$P_f(z) = e^{-z} \sum_{k \geq 0} f(k) \frac{z^k}{k!} = \sum_{n \geq 0} (-1)^n \frac{z^n}{n!} p(n)$$

Assume some "knowledge"

on the Poisson transform $P_f(z)$ or on the Poisson sequence $\Pi[f]$.

There are two paths for returning to the asymptotics of the initial sequence

- Depoissonisation method:
  - Deal with $P_f(z)$, find its asymptotics $(z \to \infty)$ [tools à la Mellin]
  - Compare the asymptotics of the sequence $f(n)$ $(n \to \infty)$ to the asymptotics of $P_f(n)$
- Rice method
  - Deal with the sequence $\Pi[f]$
  - and its analytic lifting ψ which exists [tools à la Mellin-Rice].
  - Return to the sequence $n \mapsto f(n)$ via the binomial formula which is tranfered into the Rice integral.

II. An instance of application: tries and toll functions.

Tries

A source $\mathcal{S}$ on a finite alphabet Σ $\quad \mathcal{X} := \Sigma^{\mathbb{N}}$

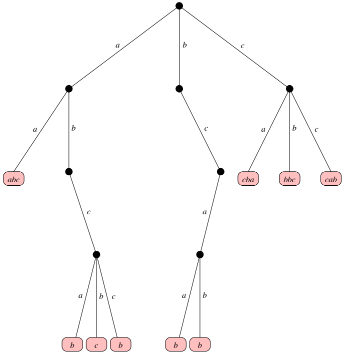
$\mathcal{X}^\star := \{$sequences of words of $\Sigma^{\mathbb{N}}$ produced by a source $\mathcal{S}\}$

The trie $\mathcal{T}(\boldsymbol{x})$ built on $\boldsymbol{x} \in \mathcal{X}^\star$ is a tree :

- If $|\boldsymbol{x}| = 0$, $\mathcal{T}(\boldsymbol{x}) = \emptyset$
- If $|\boldsymbol{x}| = 1$, $\mathcal{T}(\boldsymbol{x})$ is a leaf
- If $|\boldsymbol{x}| \geq 2$, then $\mathcal{T}(\boldsymbol{x})$ is formed with
  – an internal node o
  – and a sequence of tries $\mathcal{T}(\boldsymbol{x}_\sigma)$ for $\sigma \in \Sigma$

With $x := b(x) \cdot e(x) \in \Sigma^{\mathbb{N}}$, one associates

$x_\sigma = \epsilon \quad (b(x) \neq \sigma), \qquad x_\sigma := e(x) \quad (b(x) \neq \sigma)$



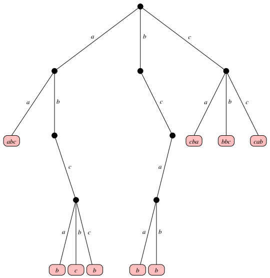
- Each internal node is labelled with a prefix $\boldsymbol{w}$,
- The associated subtrie deals with the words of $\boldsymbol{x}$ which begin with $\boldsymbol{w}$.

An instance of application: Toll functions and tries (II).

A sequence $n \mapsto f(n)$ with $\mathrm{val}(f) \geq 2$ plays the role of a toll function.

With the toll f, associate the cost R defined on $\mathcal{X}^\star$

$$R(\boldsymbol{x}) := \sum_{\boldsymbol{w} \in \Sigma^\star} f(N_{\boldsymbol{w}}(\boldsymbol{x})),$$



- $N_{\boldsymbol{w}}(\boldsymbol{x})$ is the number of words of $\boldsymbol{x}$ which begin with the prefix $\boldsymbol{w}$
- $f(N_{\boldsymbol{w}}(\boldsymbol{x}))$ is the toll "paid"

  by the subtrie whose root is labelled by $\boldsymbol{w}$

$f(k) = 1 \Longrightarrow R(\boldsymbol{x})$ is the number of internal nodes of $\mathcal{T}(\boldsymbol{x})$

$f(k) = k \Longrightarrow R(\boldsymbol{x})$ is the external path length of $\mathcal{T}(\boldsymbol{x})$

Another instance (less classical) : $f(k) = k \log k \Longrightarrow$ .....

$R(\boldsymbol{x})$ is the number of symbol comparisons performed by QuickSort on $\boldsymbol{x}$.

What is the mean value of the cost $R(\boldsymbol{x})$ when $\boldsymbol{x} \in \mathcal{X}^n$?

An instance of application: Toll functions and tries (III).

Study $r(n) :=$ the mean value of $R := \sum_{\boldsymbol{w}\in\Sigma^\star} f(N_{\boldsymbol{w}})$ in the $\mathcal{B}_n$ model

The toll f gives its Poisson transform $P_f(z)$ and/or its sequence $\Pi[f]$

$$P_f(z) = e^{-z} \sum_{n\geq 2} f(n) \frac{z^n}{n!} = \sum_{n\geq 2} (-1)^n p_f(n) \frac{z^n}{n!}$$

The same for the cost R, with the sequence $r(n)$,

$$P_r(z) = e^{-z} \sum_{n\geq 2} r(n) \frac{z^n}{n!} = \sum_{n\geq 2} (-1)^n p_r(n) \frac{z^n}{n!}$$

For $\boldsymbol{w} \in \Sigma^\star$, $\pi_{\boldsymbol{w}} :=$ the probability that a word begins with $\boldsymbol{w}$

N follows $\mathcal{P}_z \implies N_{\boldsymbol{w}}$ follows $\mathcal{P}_{z\cdot\pi_{\boldsymbol{w}}} \implies \mathbb{E}_z[f(N_{\boldsymbol{w}})] = P_f(z\,\pi_{\boldsymbol{w}})$

$$P_r(z) = \sum_{\boldsymbol{w}\in\Sigma^\star} P_f(z\,\pi_{\boldsymbol{w}}) \qquad p_r(n) = \left[\sum_{\boldsymbol{w}\in\Sigma^\star} \pi_{\boldsymbol{w}}^n\right] p_f(n)$$

Sequence $f(n) \implies$ A good knowledge of $P_f(z)$ and/or $p_f(n)$

With the source $\mathcal{S}$ ands its $\pi_w \implies$ A good knowledge of $P_r(z)$ and/or $p_r(n)$

How to return to $r(n)$?

II – The Depoissonization path

Main contributors

- Haymann [1956]
- Jacquet and Szpankowski [1998] (two papers), Jacquet [2014]
- Hwang-Fuchs-Zacharovas [2010]

The Depoissonization path deals with the Poisson transform $P_f(z)$. It

- compares $f(n)$ and $P_f(n)$ with the Poisson–Charlier expansion

$$f(n) := \sum_{j \geq 0} \frac{P_f^{(j)}(n)}{j!} \tau_j(n), \qquad \text{with } \tau_j(n) := n![z^n]\left((z-n)^j e^z\right)$$

- uses properties of the Mellin transform $P_f^\star$ for the asymptotics of P_f
- needs depoissonization sufficient conditions $\mathcal{JS}$ on $P_f(z)$,
  for truncating the Poisson-Charlier expansion
- obtains the asymptotics of $f(n)$.
- better understands the $\mathcal{JS}$ conditions:

Theorem [Jacquet-Szpankowski] The two conditions are equivalent

- $P_f(z)$ satisfies the conditions $\mathcal{JS}$
- There is an analytical lifting $\varphi(z)$ for the sequence f of polynomial growth in a horizontal cone of angle θ_0 for some $\theta_0 > 0$.

III – The Rice path

Main contributors

- Norlünd, Norlünd-Rice [1929, 1954]
- Flajolet and Sedgewick [1995]

For a sequence $n \mapsto f(n)$ of polynomial growth,
the Rice path deals with the Poisson sequence $\Pi[f]$.

- It proves the existence of an analytical lifting ψ of the sequence $\Pi[f]$

$$\psi(s) = P_f^\star(-s)/\Gamma(-s)$$

  with the (direct) Mellin transform and Newton interpolation.
  without any other condition on the sequence $n \mapsto f(n)$.

- If ψ is of polynomial growth on the right (for $\Re s \to \infty$),
  the binomial relation is transfered into a Rice integral expression

$$f(n) = \frac{1}{2i\pi} \int_{a-i\infty}^{a+i\infty} L_n(s) \cdot \psi(s)\, ds$$

  with the Rice kernel $\quad L_n(s) = \dfrac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n+1-s)} = B(n+1, -s)\,.$

- If ψ is "tame" [meromorphic and of polynomial growth] on the left,
  a shifting to the left of the integral provides the asymptotics of $f(n)$.

$$\psi(s) = P_f^\star(-s)/\Gamma(-s)$$

What are the sufficient conditions on the sequence f

which entail the tameness and polynomial growth of ψ?

Not well studied ! Are they true in any practical situation ?

The main object of this talk.

Meromorphy is often easy to ensure, the poles are easy to find...

And polynomial growth?

- Holds (very often) for $P_f^\star(s)$. But with the factor $1/\Gamma(s)$?
- Sometimes, in tries problems, the factor $\Gamma(s)$ appears in $P_f^\star(s)$.

But what about other sequences, for instance $f(k) = k \log k$

- where the depoissonization path can be used.
- Is the Rice path useful in this case?
- Is it true that the Rice path is useful only for very specific cases?

We now answer these questions with the Laplace (inverse) transform.....

A new (easy) tool in the Rice path: Canonical sequence.

Definition. For a non zero real sequence $n \mapsto f(n)$

$$\mathtt{deg}(f) := \inf\{c \mid f(k) = O(k^c)\} \qquad \mathtt{val}(f) := \min\{k \mid f(k) \neq 0\}\,.$$

The sequence f is reduced if $\mathtt{val}(f) = 0$ and $\mathtt{deg}(f) < -1$.

Reduced sequences are important in analyses à la Mellin.

With f, we associate a canonical sequence $\sigma(f)$ that is reduced, as follows:

Main tool: The shift T and its iterates T^ℓ,

$$T[f](n) = \frac{f(n+1)}{n+1}, \qquad T^\ell[f](n) = \frac{f(n+\ell)}{(n+1)\ldots(n+\ell)}$$

Choose the smallest integer $\ell > d+1$, the sequence $\sigma(f)$ defined as

$$\sigma(f)(n) := T^\ell[f](n), \quad \text{for } n \geq 1, \qquad \sigma(f)(0) = 1$$

has degree $c := d - \ell < -1$ and valuation 0.

Π anticommutes with $T \implies$ close relation between $\Pi[\sigma(f)]$ and $\Pi[f]$.

Strategy: we first deal with $\sigma(f)$ and $\Pi[\sigma(f)]$, then return to the initial f.

IV – The Rice–Laplace approach.

Consider a sequence F of polynomial growth (with Φ and Ψ)

and its canonical sequence f (with φ and ψ)

Two general steps on the canonical sequence f, with φ and ψ.

– When φ is of polynomial growth on an half-plane, we obtain an integral expression for $\psi(s)$ that involves the Inverse Laplace transform $\widehat{\varphi}$ of φ

– With a convenient expansion of $\widehat{\varphi}(u)$ at $u = 0$,

this ensures the tameness of ψ and thus the tameness of Ψ

A third step

– Find sufficient conditions on the initial sequence F

which entail such a convenient expansion for $\widehat{\varphi}(u)$ at $u = 0$.

The Rice-Laplace approach (I). A new integral expression for $\psi(s)$.

$$P_f(z) = e^{-z} \sum_{n \geq 0} f(n) \frac{z^n}{n!} = \sum_{n \geq 0} (-1)^n p(n) \frac{z^n}{n!}$$

Proposition. Consider a sequence $f : n \mapsto f(n)$ which

(a) admits an analytic lifting φ on $\Re s > a$ with $a \in]-1, 0[$,

(b) with the estimate $\varphi(s) = O(|s+1|^c)$ there with $c < -1$.

Then,

(i) the function φ admits an inverse Laplace transform

$$\widehat{\varphi}(u) := \frac{1}{2i\pi} \int_{\Re s = b} \varphi(s) e^{su} ds \qquad (b \in]a, 1])$$

on $]0, +\infty[$ of exponential decreasing as $u \to +\infty$

(ii) The analytical lifting ψ of the sequence $\Pi[f]$ exists on $\Re s > -1$ and is expressed as an integral on the real line,

$$\psi(s) = \int_0^{+\infty} \widehat{\varphi}(u) \cdot (1 - e^{-u})^s du$$

The Rice–Laplace approach (II).

$$\psi(s) = \int_0^{+\infty} \widehat{\varphi}(u)\,(1-e^{-u})^s du = \int_0^{+\infty} \widehat{\varphi}(u)\,u^s \left[\frac{1-e^{-u}}{u}\right]^s du$$

- $\psi(s)$ is closely related to the Mellin transform $s \mapsto \widehat{\varphi}^\star(s+1)$.
- Tameness of ψ is related to the expansion of $u \mapsto \widehat{\varphi}(u)$ at $u=0$.

We now just need a theorem which "transfers" the behaviour of φ into the behaviour of its inverse Laplace transform $\widehat{\varphi}(u)$ (at $u=0$).

We do not know the existence of such a general transfer theorem.
This is why we exhibit a class of sequences F where such a transfer holds

The Rice–Laplace approach (III).
A particular class of interest : Basic sequences.

Definition. Consider a pair (d, b) with a real d and an integer $b \geq 0$.
A sequence F is *basic* with pair (d, b) if it admits on some halfplane $\Re s > a \geq 0$, an analytic extension of the form

$$\Phi(s) = s^d \log^b(s)\ W(1/s), \text{ with } W \text{ analytic at } 0, \text{ and } W(0) \neq 0.$$

Main Theorem. Consider a sequence F which admits on some halfplane $\Re s > a \geq 0$, an analytic extension of the form

$$\Phi(s) = s^d \log^b(s)\ W(1/s), \text{ with } W \text{ analytic at } 0, \text{ and } W(0) \neq 0.$$

Then, the analytic continuation $\Psi(s)$ of the $\Pi[F]$ sequence is of polynomial growth on any halfplane $\Re s \geq a > d$ and is tame at $s = d$.

Then, the Rice method can be applied in this case....

V. Comparison between the two paths
Depoissonization and Rice-Laplace

Remind that the Depoissonization path deals with sequences whose analytic lifting is of polynomial growth inside horizontal cones.

We have described a precise framework for the Rice method,
the Rice–Laplace method, that deals with sequences,
whose analytic lifting admits a precise asymptotic expansion on half-planes.

– it exhibits an integral representation of the analytical lifting ψ of $\Pi[F]$;
– it proves the tameness of Ψ ;

Then, we have validated the Rice path
– in a similar framework to the Depoissonization framework;
– it is of more restricted use, but also easier to deal with.