ASYMPTOTICS OF THE STIRLING NUMBERS OF THE SECOND KIND

Bruno Salvy & John Shackell

January 1998

We show how computer algebra can be used to compute automatically the asymptotics of the Bell numbers and of the average value and variance of the number of parts in a partition. This worksheet details the computations involved in Section 3 of our article *Symbolic Asymptotics: Multiseries of Inverse Functions*, available as INRIA Research Report #3264, 1997. The starting point for all these asymptotic expansions is the bivariate generating function of the Stirling numbers of the second kind:

```
> F:=exp(u*(exp(x)-1));
```

$$F := \mathbf{e}^{(u\,(\mathbf{e}^x - 1))}$$

```
> series(F,x);
```

$$1 + u\,x + \left(\frac{1}{2}u + \frac{1}{2}u^2\right)x^2 + \left(\frac{1}{6}u + \frac{1}{2}u^2 + \frac{1}{6}u^3\right)x^3 + \left(\frac{1}{24}u + \frac{7}{24}u^2 + \frac{1}{4}u^3 + \frac{1}{24}u^4\right)x^4 + \left(\frac{1}{120}u + \frac{1}{8}u^2 + \frac{5}{24}u^3 + \frac{1}{12}u^4 + \frac{1}{120}u^5\right)x^5 + \mathrm{O}(x^6)$$

The coefficient of $u^k\,x^n$ in this series is the Stirling number of the second kind $S_{n,\,k}$ divided by $n!$. These numbers count the number of partitions of the set $\{1,...,n\}$ in k parts.

For functions with a fast growth at infinity, like many functions related to F, the asymptotic behaviour of the Taylor coefficients can be computed by the saddle-point method. This method relies on using an integral along a closed contour in the complex plane chosen in such a way as to pass through a special point called the saddle-point. The integral is concentrated in the neighborhood of the saddle-point and good asymptotic estimates are obtained from a local expansion there. However, in these applications, the saddle-point is defined implicitly by an equation and only an asymptotic expansion of it is available. This introduces technical difficulties in the manipulation of the expansions. This worksheet explores how our algorithm for asymptotic inversion applies in such circumstances.

We first load our experimental implementation

```
> read `/export/salvy/Alcom/97/asympt_inv/load.mpl`;
```

Sufficient conditions for deriving an asymptotic expansion by the saddle-point method have been given by Harris and Schoenfeld. These conditions are satisfied by all the generating functions we consider in this worksheet. The following procedure (which can be skipped in a first reading) takes as input a function $\mathrm{f}(x)$, the associated saddle-point equation, an asymptotic scale (see below) and computes the asymptotic expansion of the nth Taylor coefficient of $\mathrm{f}(x)$ as n tends to infinity in terms of the saddle-point.

```
> harris_schoenfeld:=proc(f,x,R,nofR,nbterms,scale)local A, B, Bb,
  Cc, F, i, m, u, compo;A:=diff(log(f),x);for i to 2*nbterms do
  B[i]:=normal(x^i/i!*diff(A,[x$(i-1)]))
  od;Bb:=x*diff(B[1],x)/2;for i to 2*nbterms-2 do
  Cc[i]:=normal(-(B[i+2]+(-1)^i/(i+2)*B[1])/Bb) od;F[0]:=1;for i
  to nbterms-1 do
  F[i]:=(-1)^i/sqrt(Pi)*add(GAMMA(m+i+1/2)/m!*add(mul(Cc[u],u=comp
  o),compo=combinat[composition](2*i,m)),m=1..2*i)od;tower(subs([x
  =R,n=nofR],f/2/sqrt(Pi)/sqrt(Bb)/x^n*add(normal(F[i]/Bb^i),i=0..
  nbterms-1)),scale)end:
```

Asymptotics of the Bell numbers

The Bell numbers B_n are the sum over k of the Stirling numbers $S_{n,\,k}$. Their generating function is

```
> f:=subs(u=1,F);
```

$$f := \mathbf{e}^{(\mathbf{e}^x - 1)}$$

Alternatively, since these numbers count the number of partitions of a set, this generating function and the numbers can be obtained using combstruct:

```
> G:={P=Set(Set(Atom,card>0))}:
> combstruct[gfsolve](G,labelled,x);
```

$$\{\,\mathrm{P}(x) = \mathbf{e}^{(\mathbf{e}^x - 1)}\,\}$$

Here are the first numbers:

```
> series(f,x,20);
```

$$1 + x + x^2 + \frac{5}{6}x^3 + \frac{5}{8}x^4 + \frac{13}{30}x^5 + \frac{203}{720}x^6 + \frac{877}{5040}x^7 + \frac{23}{224}x^8 + \frac{1007}{17280}x^9 + \frac{4639}{145152}x^{10} + \frac{22619}{1330560}x^{11} + \frac{4213597}{479001600}x^{12} + \frac{27644437}{6227020800}x^{13} + \frac{95449661}{43589145600}x^{14} + \frac{276591709}{261534873600}x^{15} + \frac{10480142147}{20922789888000}x^{16} + \frac{255755771}{1097800704000}x^{17} + \frac{97439543737}{914624815104000}x^{18} + \frac{5832742205057}{121645100408832000}x^{19} + \mathrm{O}(x^{20})$$

or, using combstruct

```
> seq(combstruct[count]([P,G,labelled],size=i),i=0..19);
```

1, 1, 2, 5, 15, 52, 203, 877, 4140, 21147, 115975, 678570, 4213597, 27644437, 190899322, 1382958545, 10480142147, 82864869804, 682076806159, 5832742205057

the factor $1/n!$ being due to the use of *exponential* generating functions.

The saddle-point equation we use is

```
> speq:=subs(x=R,diff(log(f),x)*x)-1=n;
```

$$speq := \mathbf{e}^R\,R - 1 = n$$

Then the asymptotic analysis of the Bell numbers is classically based on the asymptotic behaviour of the saddle-point R viewed as a function of n as n tends to infinity.

From the above equation our algorithm produces the asymptotic expansion of the saddle-point in two stages: first a sequence of ''exact'' information is computed by the procedure `Invert` which also creates the scale `scaleR` necessary for the expansions of functions of R:

```
> r1:=Invert(lhs(speq),R,'scaleR');
```

$$r1 := \left[\ln(R) + R + \ln\left(1 - \frac{1}{\mathbf{e}^R\,R}\right) - \left(\frac{1}{R} + \frac{1}{\ln(R)}\right)\ln(R)\,R,\ 1,\ \left(\frac{1}{R} + \frac{1}{\ln(R)}\right)\ln(R)\,R\right], \exp, \left[\ln\left(\left(\frac{1}{\mathbf{e}^R} + \frac{1}{R}\right)R\right), 2, R\right], [0, 2, R]$$

The last argument has been assigned the corresponding asymptotic scale:

```
> scaleR[list];
```

$$\left[\frac{1}{\ln(R)}, \frac{1}{R}, \frac{1}{\mathbf{e}^R}\right]$$

In a second stage, asymptotic information is extracted from the exact result `r1` and the scale necessary for the expansions of functions of n is created

```
> exp1:=multiseriesinverse([r1],scaleR,0,2,n,3,'scalen');
```

$$\mathrm{exp1} := \ln(n) - \ln(\ln(n)) + \frac{\ln(\ln(n))}{\ln(n)} + O\left(\frac{1}{\ln(n)^2}\right)$$

Here again the variable `scalen` holds the asymptotic scale related to the variable *n*:

```
> scalen[list];
```

$$\left[\frac{1}{\ln(\ln(\ln(n)))}, \frac{1}{\ln(\ln(n))}, \frac{1}{\ln(n)}\right]$$

The procedure `harris_schoenfeld` written above is then able to compute the expansion of the Bell numbers in terms of the saddle-point *R* in the scale `scaleR` which has been obtained above

```
> d1:=harris_schoenfeld(f,x,R,lhs(speq),3,scaleR);
```

$$d1 := \frac{1}{2}\mathbf{e}^{(\mathbf{e}^R)}\,\mathbf{e}^{(-1)}\sqrt{2}\,R\left(1 - \frac{1}{24}\frac{2R^4 - 3R^3 - 18R + 2 - 20R^2}{(R+1)^3 R\,\mathbf{e}^R} + \frac{1}{1152}\frac{-695R^6 + 4R^8 - 156R^7 + 4 + 1972R^2 + 2916R^3 + 1092R^4 - 696R^5 - 72R}{(R+1)^6 R^2 (\mathbf{e}^R)^2}\right) \Big/ \left(\sqrt{\pi}\sqrt{R(\mathbf{e}^R R + \mathbf{e}^R)}\;\mathbf{e}^{(R\ln(R)\mathbf{e}^R)}\;\mathbf{e}^{((\mathbf{e}^R R - 1)\ln(R) - R\ln(R)\mathbf{e}^R + \ln(R))}\right)$$

The result follows upon substitution of *R* by its asymptotic expansion in terms of *n* (*exp1* above). Since the logarithm of this expansion is easier to handle, we first compute it:

```
> d2:=`tower/ln`(d1,scaleR):
> multiseries(d2,scaleR,0,3,nops(scaleR[list]),3);
```

$$(1 - \ln(R)R)\mathbf{e}^R - \frac{1}{2}R + \ln\left(\frac{1}{2}\frac{\mathbf{e}^{(-1)}\sqrt{2}}{\sqrt{\pi}}\right) - \frac{1}{2}\frac{1}{R} + O\left(\frac{1}{R^2}\right) + \frac{-\frac{1}{12} + \frac{3}{8}\frac{1}{R} - \frac{1}{24}\frac{1}{R^2} + O\left(\frac{1}{R^3}\right)}{\mathbf{e}^R}$$

Hence the result by substituting the expansion of *R* in terms of *n* in this last expansion:

```
> substitute(d2,scaleR,exp1,scalen,3);
```

$$\left(-\ln(\ln(n)) + \frac{\left(\frac{1}{\ln(\ln(n))} + 1\right)\ln(\ln(n))}{\ln(n)} + O\left(\frac{1}{\ln(n)^2}\right)\right)n - \frac{1}{2}\ln(n) + \frac{1}{2}\left(2\frac{\ln\left(\frac{1}{2}\frac{\mathbf{e}^{(-1)}\sqrt{2}}{\sqrt{\pi}}\right)}{\ln(\ln(n))} + 1\right)\ln(\ln(n)) - \frac{1}{2}\frac{\left(\frac{1}{\ln(\ln(n))} + 1\right)\ln(\ln(n))}{\ln(n)} + O\left(\frac{1}{\ln(n)^2}\right) + \frac{-\frac{1}{12}\ln(n) + \frac{1}{24}\left(2 + \frac{9}{\ln(\ln(n))}\right)\ln(\ln(n)) + O\left(\frac{1}{\ln(n)}\right)}{n}$$

The first part of this expansion is the classical expansion of $\ln(B_n/n!)$ which can be found in (de Bruijn 1981, p. 108). The remaining parts reflect the use of multiseries, a recent tool which makes it possible to deal with indefinite cancellation in a simple way.

Average number of parts in a partition

The technique demonstrated above also applies to the average number of parts in a partition, where an indefinite cancellation occurs and this exemplifies the use of multiple scales. The generating function of $\sum_k k\, S_{n,k}$ is

```
> g:=subs(u=1,diff(F,u));
```

$$g := (e^x - 1)\, e^{(e^x - 1)}$$

The average number of parts in a partition of a set of size n is the nth Taylor coefficient of g divided by the nth coefficient of f whose asymptotics we have just computed. The saddle-point equation for g is:

```
> speq2:=normal(subs(x=R,diff(log(g),x)*x)-1)=n;
```

$$speq2 := \frac{(e^R)^2 R - e^R + 1}{e^R - 1} = n$$

From this equation our algorithm produces the asymptotic expansion of the saddle-point in two stages as before:

```
> r2:=Invert(op(1,speq2),R):
> exp2:=multiseriesinverse([r2],scaleR,0,2,n,3,scalen);
```

$$\exp2 := \ln(n) - \ln(\ln(n)) + \frac{\ln(\ln(n))}{\ln(n)} + O\left(\frac{1}{\ln(n)^2}\right)$$

This expansion of this new saddle-point is the same as that of the previous case. But this does not mean that both saddle-points are equal. The difference between their asymptotic behaviours can be obtained by working in a different scale. Here is a refined estimate for the first saddle-point:

```
> sp1:=multiseriesinverse([r1],scaleR,2,1,n,3,'scalezeta');
```

$$sp1 := \%1 + \frac{1}{\%1\left(\frac{1}{\%1}+1\right)e^{\%1}} - \frac{1}{2}\,\frac{\frac{2}{\%1}+1}{\%1^2\left(\frac{1}{\%1}+1\right)^3 (e^{\%1})^2} + O\left(\frac{1}{(e^{\%1})^3}\right)$$

$$\%1 := \mathrm{RootOf}(\ln(\_Z) + \_Z - \ln(n))$$

```
> alias(zeta=%1):
> sp1;
```

$$\zeta + \frac{1}{\zeta\left(\frac{1}{\zeta}+1\right)e^{\zeta}} - \frac{1}{2}\,\frac{\frac{2}{\zeta}+1}{\zeta^2\left(\frac{1}{\zeta}+1\right)^3 (e^{\zeta})^2} + O\left(\frac{1}{(e^{\zeta})^3}\right)$$

and here is an estimate for the second saddle-point refined similarly

```
> sp2:=multiseriesinverse([r2],scaleR,2,1,n,3,scalezeta);
```

$$sp2 := \zeta + \frac{\frac{1}{\zeta} - 1}{\left(\frac{1}{\zeta}+1\right)e^{\zeta}} - \frac{1}{2}\,\frac{-\frac{1}{\zeta^2} + \frac{2}{\zeta} + 3 + \frac{4}{\zeta^3}}{\left(\frac{1}{\zeta}+1\right)^3 (e^{\zeta})^2} + O\left(\frac{1}{(e^{\zeta})^3}\right)$$

Hence the difference can be computed at this level of the asymptotic scale

```
> `multiseries/expand`(x1-x2,[x1,x2],[sp1,sp2],scalezeta,3);
```

$$\frac{1}{\left(\frac{1}{\zeta}+1\right)e^{\zeta}} + \frac{1}{2}\,\frac{\frac{2}{\zeta^3} - \frac{2}{\zeta^2} + \frac{2}{\zeta} + 3}{\left(\frac{1}{\zeta}+1\right)^3 (e^{\zeta})^2} + O\left(\frac{1}{(e^{\zeta})^3}\right)$$

or in terms of n

```
> substitute(SERIEStoseries(",'polynom'),scalezeta,exp1,scalen,
  3);
```

$$\frac{\ln(n) - \left(\frac{1}{\ln(\ln(n))} + 1\right)\ln(\ln(n)) + \mathrm{O}\left(\frac{1}{\ln(n)}\right)}{n} + \frac{\frac{3}{2}\ln(n)^2 - \frac{1}{2}\left(6 + \frac{7}{\ln(\ln(n))}\right)\ln(\ln(n))\ln(n) + \mathrm{O}(1)}{n^2}$$

Thus the saddle-points differ asymptotically by $\ln(n)/n$, which could not be detected from the asymptotic expansion in the original scale.

In order to compute the average number of parts in a partition, we have to divide the estimate for the nth coefficient of g by the estimate for the nth coefficient of f. Since these estimates are given in terms of different saddle-points, the computation is more easily done by obtaining both estimates in the same scale, which involves ζ, and then substituting the estimate for ζ in terms of n.

```
> estden:=substitute(d1,scaleR,sp1,scalezeta,3);
```

$$estden := \frac{\left(\frac{1}{2}\frac{\sqrt{2}\,\mathbf{e}^{(-1)}\sqrt{\frac{1}{\mathbf{e}^{\zeta}}}}{\sqrt{\pi}\sqrt{\zeta(\zeta+1)}} - \frac{1}{48}\frac{\left(\frac{2}{\zeta^4} + \frac{6}{\zeta^3} + \frac{16}{\zeta^2} + \frac{9}{\zeta} + 2\right)\mathbf{e}^{(-1)}\sqrt{2}\left(\frac{1}{\mathbf{e}^{\zeta}}\right)^{3/2}}{\sqrt{\pi}\sqrt{\zeta(\zeta+1)}\left(\frac{1}{\zeta}+1\right)^3} + \mathrm{O}\left(\left(\frac{1}{\mathbf{e}^{\zeta}}\right)^{5/2}\right)\right)\mathbf{e}^{(\mathbf{e}^{\zeta})}}{\mathbf{e}^{(\zeta\ln(\zeta)\mathbf{e}^{\zeta})}}$$

```
> n1:=harris_schoenfeld(g,x,R,op(1,speq2),3,scaleR):
> estnum:=substitute(SERIEStoseries(multiseries(n1,scaleR,2,1,n
  ops(scaleR[list]),3),'polynom'),scaleR,sp2,scalezeta,3);
```

$$estnum := \left(\frac{1}{2}\frac{\sqrt{2}\,\mathbf{e}^{(-1)}}{\sqrt{\pi}\sqrt{\zeta(\zeta+1)}\sqrt{\frac{1}{\mathbf{e}^{\zeta}}}} - \frac{1}{48}\frac{\left(\frac{2}{\zeta^4} + \frac{30}{\zeta^3} + \frac{76}{\zeta^2} + \frac{69}{\zeta} + 26\right)\sqrt{2}\,\mathbf{e}^{(-1)}\sqrt{\frac{1}{\mathbf{e}^{\zeta}}}}{\sqrt{\zeta(\zeta+1)}\left(\frac{1}{\zeta}+1\right)^3\sqrt{\pi}} + \mathrm{O}\left(\left(\frac{1}{\mathbf{e}^{\zeta}}\right)^{3/2}\right)\right)\mathbf{e}^{(\mathbf{e}^{\zeta})} \Big/ \mathbf{e}^{(\zeta\ln(\zeta)\mathbf{e}^{\zeta})}$$

Hence the ratio in ζ is

```
> ratio:=`multiseries/expand`(x2/x1,[x1,x2],[estden,estnum],sca
  lezeta,3);
```

$$ratio := \mathbf{e}^{\zeta} - \frac{1}{2}\frac{\frac{2}{\zeta^2} + \frac{3}{\zeta} + 2}{\left(\frac{1}{\zeta}+1\right)^2} + \mathrm{O}\left(\frac{1}{\mathbf{e}^{\zeta}}\right)$$

And here is the final result in n: the average number of parts in a partition of n elements:

```
> substitute(SERIEStoseries(ratio,'polynom'),scalezeta,exp1,sca
len,3);
```

$$\left(\frac{1}{\ln(n)}+\frac{\ln(\ln(n))}{\ln(n)^2}+O\left(\frac{1}{\ln(n)^3}\right)\right)n-1+\frac{1}{2}\frac{1}{\ln(n)}-\frac{1}{2}\frac{\ln(\ln(n))\left(-1+\frac{2}{\ln(\ln(n))}\right)}{\ln(n)^2}+O\left(\frac{1}{\ln(n)^3}\right)$$

This result is classical. However, it is worth noting that classical references give at best the first term of the asymptotic behaviour (n/log n) and one of them is wrong by a factor of e=exp(1). This gives an idea of the difficulty of performing such computations by hand.

Variance

Computation of the variance leads to further cancellation. It is obtained from the generating function

```
> h:=subs(u=1,diff(u*diff(F,u),u));
```

$$h:=(\mathbf{e}^x-1)\,\mathbf{e}^{(\mathbf{e}^x-1)}+(\mathbf{e}^x-1)^2\,\mathbf{e}^{(\mathbf{e}^x-1)}$$

The corresponding saddle-point equation is

```
> speq3:=subs(x=R,normal(diff(log(h),x)*x))-1=n;
```

$$speq3:=\frac{(-1+\mathbf{e}^R+(\mathbf{e}^R)^2)\,R}{\mathbf{e}^R-1}-1=n$$

The expansion of the saddle-point in terms of ζ is obtained as before

```
> r3:=Invert(op(1,speq3),R);
```

$$r3:=\left[\ln(R)+R+\ln\left(-\frac{-\frac{1}{(\mathbf{e}^R)^2}+\frac{1}{\mathbf{e}^R}+1-\frac{1}{\mathbf{e}^R\,R}+\frac{1}{(\mathbf{e}^R)^2\,R}}{-1+\frac{1}{\mathbf{e}^R}}\right)-\left(\frac{1}{R}+\frac{1}{\ln(R)}\right)\ln(R)\,R,1,\left(\frac{1}{R}+\frac{1}{\ln(R)}\right)\ln(R)\,R\right],\exp,\left[\ln\left(\left(\frac{1}{\mathbf{e}^R}+\frac{1}{R}\right)R\right),2,R\right],[0,2,R]$$

```
> sp3:=multiseriesinverse([r3],scaleR,2,1,n,3,scalezeta);
```

$$sp3:=\zeta+\frac{\frac{1}{\zeta}-2}{\left(\frac{1}{\zeta}+1\right)\mathbf{e}^\zeta}-\frac{3}{2}\frac{\frac{2}{\zeta^3}-\frac{3}{\zeta^2}+2}{\left(\frac{1}{\zeta}+1\right)^3(\mathbf{e}^\zeta)^2}+O\left(\frac{1}{(\mathbf{e}^\zeta)^3}\right)$$

This leads to the following estimate for the nth coefficient of h, first in terms of R and then in terms of ζ by substituting R by its asymptotic expansion in terms of ζ:

```
> n2:=harris_schoenfeld(h,x,R,op(1,speq3),3,scaleR):
> estnum2:=substitute(SERIEStoseries(multiseries(n2,scaleR,2,1,
nops(scaleR[list]),3),'polynom'),scaleR,sp3,scalezeta,3);
```

$estnum2:=$

$$\left(\frac{1}{2}\frac{\sqrt{2}\,\mathbf{e}^{(-1)}}{\sqrt{\pi}\,\sqrt{\zeta\,(\zeta+1)}\left(\frac{1}{\mathbf{e}^{\zeta}}\right)^{3/2}}-\frac{1}{48}\frac{\left(\frac{2}{\zeta^4}+\frac{30}{\zeta^3}+\frac{88}{\zeta^2}+\frac{105}{\zeta}+50\right)\mathbf{e}^{(-1)}\sqrt{2}}{\left(\frac{1}{\zeta}+1\right)^3\sqrt{\pi}\,\sqrt{\zeta\,(\zeta+1)}\,\sqrt{\frac{1}{\mathbf{e}^{\zeta}}}}+\mathrm{O}\left(\sqrt{\frac{1}{\mathbf{e}^{\zeta}}}\right)\right)$$
$$\mathbf{e}^{(\mathbf{e}^{\zeta})}\Big/\mathbf{e}^{(\zeta\ln(\zeta)\,\mathbf{e}^{\zeta})}$$

From there we get the variance in terms of ζ

```
> finalres:=`multiseries/expand`(x3/x1-(x2/x1)^2,[x1,x2,x3],[es
  tden,estnum,estnum2],scalezeta,3);
```

$$\mathit{finalres} := \frac{\mathbf{e}^{\zeta}}{\zeta\left(\frac{1}{\zeta}+1\right)}+\mathrm{O}(1)$$

and in terms of n

```
> substitute(SERIEStoseries(",'polynom'),scalezeta,exp1,scalen,
  3);
```

$$\left(\frac{1}{\ln(n)^2}-\frac{\left(-2+\frac{1}{\ln(\ln(n))}\right)\ln(\ln(n))}{\ln(n)^3}+\mathrm{O}\left(\frac{1}{\ln(n)^4}\right)\right)n$$

Again, only the first term is given in the literature. Our method can compute as many terms as desired, as well as the expansions of higher moments.

Conclusion

Multiseries are specially useful for saddle-point asymptotics that arise frequently in combinatorial applications. Delicate expansions (that lead to errors in the technical literature) can be handled automatically. Such a process is specially useful for moment computations that involve minute saddle-point perturbations.