Coding multitype branching forests : application to the law of the total progeny.

> Loïc Chaumont Université d'Angers

> > Joint work with Rongli Liu, University of Nanjing

> > > ション ふゆ アメリア メリア しょうくしゃ

ANR MANEGE Marseille, 31/01/12 et 01/02/12

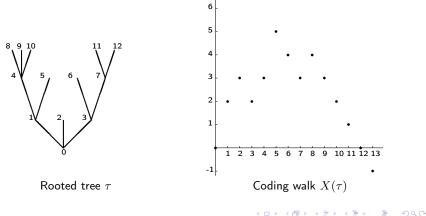
- $\blacktriangleright~\mu$ distribution on \mathbb{Z}_+ such that $\sum_{k=0}^\infty k\mu(k) \leq 1,~\mu(1) < 1$.
- τ branching (rooted) tree with offspring distribution μ .
- $O(\tau)$ total progeny of τ .
- ▶ $u_0, \ldots, u_{O(\tau)-1}$ vertices of τ ranked in the breadth first search order.

• $k_u(\tau)$ number of children of $u \in \tau$.

The genealogy of any tree τ is encoded through :

$$X_0 = 0$$
, $X_{n+1}(\tau) - X_n(\tau) = k_{u_n}(\tau) - 1$, $0 \le n \le O(\tau) - 1$.

 $(X_n)_{n\geq 0}$ downward skip free random walk with step distribution : $\mathbb{P}(X_1=i)=\mu(i+1).$



The law of the total progeny $O(\tau)$ follows from the identity :

$$O(\tau) = \inf\{n : X_n = -1\}$$

and the Ballot theorem :

$$P(T_1 = n) = \frac{1}{n} P(X_n = -1),$$

$$T_1 = \inf\{n : X_n = -1\}.$$

Theorem (Dwass, 1969)

The law of the total progeny of au is

$$\mathbb{P}_1(O(\tau) = n) = \frac{1}{n} \mu^{*n}(n-1).$$

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへで

The law of the total progeny $O(\tau)$ follows from the identity :

$$O(\tau) = \inf\{n : X_n = -1\}$$

and the Ballot theorem :

$$P(T_1 = n) = \frac{1}{n} P(X_n = -1),$$

$$T_1 = \inf\{n : X_n = -1\}.$$

Theorem (Dwass, 1969)

The law of the total progeny of τ is

$$\mathbb{P}_1(O(\tau) = n) = \frac{1}{n} \mu^{*n}(n-1).$$

▲□▶ ▲□▶ ▲三▶ ▲三▶ 三三 のへで



More generally, for any downward skipfree random walk (X_n) ,

$$P(T_k = n) = \frac{k}{n} P(X_n = -k),$$

where $T_k = \inf\{n : X_n = -k\}.$

► The law of the total progeny of the forest $\mathcal{F} = \{\tau_1, \dots, \tau_k\}$ is, $\mathbb{P}_k(O(\mathcal{F}) = n) = \frac{k}{-}\mu^{*n}(n-k)$.

▲□▶ ▲圖▶ ▲国▶ ▲国▶ ▲国 ● のへの



٠

More generally, for any downward skipfree random walk (X_n) ,

$$P(T_k = n) = \frac{k}{n} P(X_n = -k),$$

where $T_k = \inf\{n : X_n = -k\}.$

▶ The law of the total progeny of the forest $\mathcal{F} = \{\tau_1, \ldots, \tau_k\}$ is,

$$\mathbb{P}_k(O(\mathcal{F}) = n) = \frac{k}{n} \mu^{*n}(n-k) \,.$$

 μ_1 and μ_2 probabilities on $\mathbb{Z}_+ \times \mathbb{Z}_+$.

 $\mathbf{Z}_n := (Z_n^{(1)}, Z_n^{(2)})$, $n \ge 0$, 2-type branching process with progeny law (μ_1, μ_2) , such that $\mathbf{Z}_0 = (1, 0)$. Assume that

$$T := \inf\{n : \mathbf{Z}_n = 0\} < \infty, \quad \text{a.s.}$$

What is the joint law of

$$O_1 = \sum_{n=0}^{T} Z_n^{(1)} = \text{total number of individuals of type 1 at time } T$$

$$O_2 = \sum_{n=0}^{T} Z_n^{(2)} = \text{total number of individuals of type 2 at time } T ?$$

Define the mean matrix :

$$m_{ij} = \sum_{\mathbf{z} \in \mathbb{Z}^2_+} z_j \mu_i(\mathbf{z}), \quad i, j \in \{1, 2\}.$$

▶ $m_{12} > 0, 1 \ge m_{11} > 0$ and $m_{22} = m_{21} = 0$, (Bertoin, 2010) :

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2) = \frac{1}{n_1} \mu_1^{*n_1}(n_1 - 1, n_2), \quad n_1 \ge 1, n_2 \ge 0.$$

• $m_{12} > 0, 1 \ge m_{11}, m_{22} > 0$ but $m_{21} = 0$,

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2) = \frac{1}{n_1 n_2} \sum_{j=0}^{n_2} j \mu_1^{*n_1}(n_1 - 1, j) \mu_2^{*n_2}(0, n_2 - j).$$

・ ロ ト ・ 個 ト ・ ヨ ト ・ ヨ ・ つへで

In all the remaining cases the matrix $(m_{ij})_{i,j\in\{1,2\}}$ (or the process Z) is irreducible, i.e.

$$m_{12} > 0$$
 and $m_{21} > 0$.

Let ρ be the dominant eigenvalue (Perron-Frobenius).

Then,

$$\rho \leq 1 \Longleftrightarrow T := \inf\{n: \mathbf{Z}_n = 0\} < \infty \,, \quad \text{a.s.} \,.$$

The process is said to be critical ($\rho = 1$) or subcritical ($\rho < 1$).

- O_1 : total number of individuals of type 1.
- O_2 : total number of individuals of type 2.
- N_1 : total number of individuals of type 1 whose parent is of type 2.
- $N_2: {\rm total} \ {\rm number} \ {\rm of} \ {\rm individuals} \ {\rm of} \ {\rm type} \ 2 \ {\rm whose} \ {\rm parent} \ {\rm is} \ {\rm of} \ {\rm type} \ 1.$

Theorem

Assume that **Z** is irreducible and critical or subcritical and $\mathbf{Z}_0 = (1,0)$. Then for all $n_1 \ge 1$ $n_2 \ge 0$, $1 \le k_1 \le n_1$ and $0 \le k_2 \le n_2$,

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2) = \frac{k_2}{n_1 n_2} \mu_1^{*n_1}(n_1 - k_1, k_2) \mu_2^{*n_2}(k_1, n_2 - k_2).$$

- O_1 : total number of individuals of type 1.
- O_2 : total number of individuals of type 2.
- N_1 : total number of individuals of type 1 whose parent is of type 2.
- $N_2: {\rm total} \ {\rm number} \ {\rm of} \ {\rm individuals} \ {\rm of} \ {\rm type} \ 2 \ {\rm whose} \ {\rm parent} \ {\rm is} \ {\rm of} \ {\rm type} \ 1.$

Theorem

Assume that **Z** is irreducible and critical or subcritical and $\mathbf{Z}_0 = (1,0)$. Then for all $n_1 \ge 1$ $n_2 \ge 0$, $1 \le k_1 \le n_1$ and $0 \le k_2 \le n_2$,

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2) = \frac{k_2}{n_1 n_2} \mu_1^{*n_1} (n_1 - k_1, k_2) \mu_2^{*n_2} (k_1, n_2 - k_2).$$

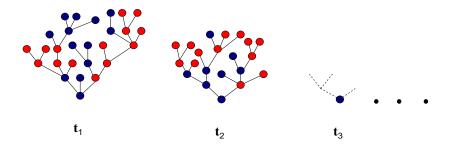
◆□▶ ◆□▶ ◆□▶ ◆□▶ ● ● ●

Define a 2-type forest,

$$\mathcal{F} = \left\{ \mathbf{t}_1, \mathbf{t}_2, \ldots \right\},\,$$

as an infinite sequence of independent 2-type rooted trees, with progeny law (μ_1, μ_2) .

- Each vertex $u \in \mathbf{t}_i$ is either of type 1 or type 2.
- The root of each tree is of type 1.
- Vertices of \mathcal{F} are ranked in the breadth first search order.



Type $1 = \bullet$

Type $2 = \bullet$

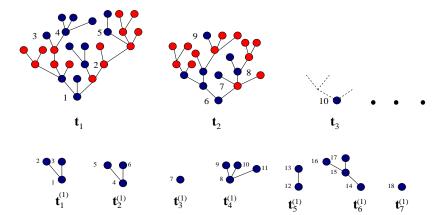
Ordering vertices of type 1:

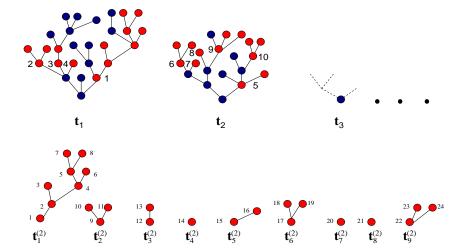
Subtrees of type 1 are ranked according to the breadth first search order of their roots in the forest :

 $\mathbf{t}_{1}^{(1)}, \mathbf{t}_{2}^{(1)}, \dots, \mathbf{t}_{n}^{(1)}, \dots$

► Then vertices u_i⁽¹⁾,..., u_j⁽¹⁾ of t_n⁽¹⁾ are ranked according to the 'local' breadth first search order of t_n⁽¹⁾:

$$\underbrace{u_0^{(1)}, \dots, u_{i_1-1}^{(1)}}_{\mathbf{t}_1^{(1)}}, \underbrace{u_{i_1}^{(1)}, \dots, u_{i_1+i_2-1}^{(1)}}_{\mathbf{t}_2^{(1)}}, \dots$$





Let $k_i(u)$ be the number of children of type i of the vertex u. Then define the integer valued chains $X = (X^{(1)}, X^{(2)})$ and $Y = (Y^{(1)}, Y^{(2)})$ by :

$$X_{n+1}^{(1)} - X_n^{(1)} = k_1(u_n^{(1)}) - 1 \qquad Y_{n+1}^{(1)} - Y_n^{(1)} = k_1(u_n^{(2)})$$
$$X_{n+1}^{(2)} - X_n^{(2)} = k_2(u_n^{(1)}) \qquad Y_{n+1}^{(2)} - Y_n^{(2)} = k_2(u_n^{(2)}) - 1.$$

Proposition

The chains X and Y are independent random walks in $\mathbb{Z} \times \mathbb{Z}_+$ and $\mathbb{Z}_+ \times \mathbb{Z}$, respectively, with step distributions :

$$P(X_1 = (i, j)) = \mu_1(i+1, j), \quad P(Y_1 = (i, j)) = \mu_2(i, j+1).$$

うして ふゆう ふほう ふほう しょうくのう

Define

$$T_k = \inf\{n : X_n^{(1)} = -k\}$$
 $S_k = \inf\{n : Y_n^{(2)} = -k\}.$

Then,

- ► X⁽²⁾(T_k) is the number of subtrees of type 2 encountered when k subtrees of type 1 have been visited,
- ► Y⁽¹⁾(S_k) is the number of subtrees of type 1 encountered when k subtrees of type 2 have been visited.

Therefore, if k_i , i = 1, 2 is the total number of subtrees of type i in the first tree \mathbf{t}_1 of the 2-type forest \mathcal{F} , then

$$\begin{cases} k_2 = X^{(2)}(T_{k_1}) \\ k_1 = 1 + Y^{(1)}(S_{k_2}) \end{cases}$$

Define

$$T_k = \inf\{n : X_n^{(1)} = -k\}$$
 $S_k = \inf\{n : Y_n^{(2)} = -k\}.$

Then,

- ► X⁽²⁾(T_k) is the number of subtrees of type 2 encountered when k subtrees of type 1 have been visited,
- ► Y⁽¹⁾(S_k) is the number of subtrees of type 1 encountered when k subtrees of type 2 have been visited.

Therefore, if k_i , i = 1, 2 is the total number of subtrees of type i in the first tree t_1 of the 2-type forest \mathcal{F} , then

$$\begin{cases} k_2 = X^{(2)}(T_{k_1}) \\ k_1 = 1 + Y^{(1)}(S_{k_2}). \end{cases}$$

Let (k_1, k_2) be the smallest solution of

(S)
$$\begin{cases} k_2 = X^{(2)}(T_{k_1}) \\ k_1 = 1 + Y^{(1)}(S_{k_2}). \end{cases}$$

Proposition

- k_i , i = 1, 2 is the total number of subtrees of type i in t_1 .
- T_{k_1} is the total number of individuals of type 1 in t_1 .
- \triangleright S_{k_2} is the total number of individuals of type 2 in \mathbf{t}_1 .
- \mathbf{t}_1 is encoded by the two 2-dimensional chains :

$$[(X_n^{(1)}, X_n^{(2)}), 0 \le n \le T_{k_1}] [(Y_n^{(1)}, Y_n^{(2)}), 0 \le n \le S_{k_2}].$$

Recall that :

- O_1 : total number of individuals of type 1.
- O_2 : total number of individuals of type 2.
- ▶ N_1 : total number of individuals of type 1 whose parent is of type 2.
- \blacktriangleright N₂ : total number of individuals of type 2 whose parent is of type 1.

(S)
$$\begin{cases} k_2 = X^{(2)}(T_{k_1}) \\ k_1 = 1 + Y^{(1)}(S_{k_2}). \end{cases}$$

Then,

$$\begin{split} \mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2) = \\ P(T_{k_1} = n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S).) \end{split}$$

Let $(U_k, 0 \le k \le k_1)$ and $(V_k, 0 \le k \le k_2)$ be independent, integer valued, nondecreasing, with $U_0 = V_0 = 0$ and with cyclically exchangeable increments.

$$(S_{U,V}) \begin{cases} k_1 = r_1 + V_{k_2} \\ k_2 = r_2 + U_{k_1} \end{cases}$$

Theorem (Bivariate ballot Theorem)

Assume that $(S_{U,V})$ admits a solution a.s., then

$$\begin{split} P(\ (k_1,k_2) \text{ is the smallest solution of } (S).) &= \\ \frac{k_1r_2 + k_2r_1 - r_1r_2}{k_1k_2} P(U_{k_1} = k_2 - r_2, V_{k_2} = k_1 - r_1) \,. \end{split}$$

Apply the biveriate ballot Theorem to $r_1=1,\,r_2=0,$ and to $U_k=X^{(2)}(T_k)$ and $V_k=Y^{(1)}(S_k),$

$$\begin{split} \mathbb{P}_{(1,0)}(O_1 &= n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2) \\ P(T_{k_1} &= n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S). \\ &= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(S_{k_2}) = k_1 - 1, X^{(2)}(T_{k_1}) = k_2) \\ &= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(n_2) = k_1 - 1, X^{(2)}(n_1) = k_2) \\ &= \frac{1}{k_1} P(T_{k_1} = n_1, X_{n_1}^{(2)} = k_2) \mathbb{P}(S_{k_2} = n_2, Y_{n_2}^{(1)} = k_1 - 1) \end{split}$$

◆□▶ ◆圖▶ ◆필▶ ◆필▶ ─ ヨ ─ のへで

Apply the biveriate ballot Theorem to $r_1=1,\,r_2=0,$ and to $U_k=X^{(2)}(T_k)$ and $V_k=Y^{(1)}(S_k),$

$$\begin{split} &\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2) \\ &P(T_{k_1} = n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S).) \\ &= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(S_{k_2}) = k_1 - 1, X^{(2)}(T_{k_1}) = k_2) \\ &= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(n_2) = k_1 - 1, X^{(2)}(n_1) = k_2) \\ &= \frac{1}{k_1} P(T_{k_1} = n_1, X_{n_1}^{(2)} = k_2) \mathbb{P}(S_{k_2} = n_2, Y_{n_2}^{(1)} = k_1 - 1) \\ &= \frac{k_2}{n_1 n_2} \mu_1^{*n_1} (n_1 - k_1, k_2) \mu_2^{*n_2} (k_1, n_2 - k_2) \,. \end{split}$$

▲□▶ ▲□▶ ▲□▶ ▲□▶ = 三 のへで

Apply the biveriate ballot Theorem to $r_1=1,\,r_2=0,$ and to $U_k=X^{(2)}(T_k)$ and $V_k=Y^{(1)}(S_k),$

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2)$$

 $P(T_{k_1} = n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S).)$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(S_{k_2}) = k_1 - 1, X^{(2)}(T_{k_1}) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(n_2) = k_1 - 1, X^{(2)}(n_1) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, X_{n_1}^{(2)} = k_2) \mathbb{P}(S_{k_2} = n_2, Y_{n_2}^{(1)} = k_1 - 1)$$

$$= \frac{k_2}{n_1 n_2} \mu_1^{*n_1} (n_1 - k_1, k_2) \mu_2^{*n_2} (k_1, n_2 - k_2) \,.$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで

Apply the biveriate ballot Theorem to $r_1=1$, $r_2=0$, and to $U_k=X^{(2)}(T_k)$ and $V_k=Y^{(1)}(S_k)$,

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2)$$

 $P(T_{k_1} = n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S).)$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(S_{k_2}) = k_1 - 1, X^{(2)}(T_{k_1}) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(n_2) = k_1 - 1, X^{(2)}(n_1) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, X_{n_1}^{(2)} = k_2) \mathbb{P}(S_{k_2} = n_2, Y_{n_2}^{(1)} = k_1 - 1)$$

 $= \frac{k_2}{n_1 n_2} \mu_1^{*n_1} (n_1 - k_1, k_2) \mu_2^{*n_2} (k_1, n_2 - k_2) \,.$

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

Apply the biveriate ballot Theorem to $r_1=1,\,r_2=0,$ and to $U_k=X^{(2)}(T_k)$ and $V_k=Y^{(1)}(S_k),$

 $\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2)$

 $P(T_{k_1} = n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S).)$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(S_{k_2}) = k_1 - 1, X^{(2)}(T_{k_1}) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(n_2) = k_1 - 1, X^{(2)}(n_1) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, X_{n_1}^{(2)} = k_2) \mathbb{P}(S_{k_2} = n_2, Y_{n_2}^{(1)} = k_1 - 1)$$

$$= \frac{k_2}{n_1 n_2} \mu_1^{*n_1} (n_1 - k_1, k_2) \mu_2^{*n_2} (k_1, n_2 - k_2) \,.$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで

Apply the biveriate ballot Theorem to $r_1=1$, $r_2=0$, and to $U_k=X^{(2)}(T_k)$ and $V_k=Y^{(1)}(S_k)$,

$$\mathbb{P}_{(1,0)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - 1, N_2 = k_2)$$

 $P(T_{k_1} = n_1, S_{k_2} = n_2 \text{ and } (k_1, k_2) \text{ is the smallest solution of } (S).)$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(S_{k_2}) = k_1 - 1, X^{(2)}(T_{k_1}) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, S_{k_2} = n_2, Y^{(1)}(n_2) = k_1 - 1, X^{(2)}(n_1) = k_2)$$

$$= \frac{1}{k_1} P(T_{k_1} = n_1, X_{n_1}^{(2)} = k_2) \mathbb{P}(S_{k_2} = n_2, Y_{n_2}^{(1)} = k_1 - 1)$$

$$= \frac{k_2}{n_1 n_2} \mu_1^{*n_1} (n_1 - k_1, k_2) \mu_2^{*n_2} (k_1, n_2 - k_2) \,.$$

◆□▶ ◆圖▶ ◆臣▶ ◆臣▶ ─ 臣 ─ のへで

More generally, when $\mathbf{Z}_0 = (r_1, r_2)$:

Theorem

Assume that Z is irreducible and critical or subcritical and $Z_0 = (r_1, r_2)$. Then for all $n_1 \ge r_1$ $n_2 \ge r_2$, $r_1 \le k_1 \le n_1$ and $r_2 \le k_2 \le n_2$,

$$\mathbb{P}_{(r_1,r_2)}(O_1 = n_1, O_2 = n_2, N_1 = k_1 - r_1, N_2 = k_2 - r_2) = \frac{r_1k_2 + r_2k_1 - r_1r_2}{n_1n_2} \mu_1^{*n_1}(n_1 - k_1, k_2) \mu_2^{*n_2}(k_1, n_2 - k_2).$$

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● ● ●

Three types :

• A_{ij} = number of individuals of type j whose parent is of type i.

Theorem

Assume that **Z** is irreducible and critical or subcritical and $\mathbf{Z}_0 = (r_1, r_2, r_3)$. Then for all $n_j \ge 1$ and $0 \le k_{ij} \le n_j$, j = 1, 2, 3,

$$\mathbb{P}(O_1 = n_1, O_2 = n_2, O_3 = n_3, A_{ij} = k_{ij}, i = 1, 2, 3, i \neq j) = (n_1 n_2 n_3)^{-1} \{ r_1[(r_3 + k_{12})k_{23} + (k_{23} + r_2 + k_{12})(r_3 + k_{13})] + k_{21}[r_3 k_{32} + (k_{23} + r_3 + k_{13})r_2] + k_{31}[r_2 k_{23} + (k_{32} + r_2 + k_{12})r_3] \} \times \mu_1^{*n_1}(n_1 - k_{21} - k_{31} - r_1, k_{12}, k_{13}) \times \mu_2^{*n_2}(k_{21}, n_2 - k_{12} - k_{32} - r_2, k_{23}) \times \mu_3^{*n_3}(k_{31}, k_{32}, n_3 - k_{13} - k_{23} - r_3) .$$