

# **Strategies for Efficient Computation of Multivariate Poisson Probabilities**

Panagiotis Tsiamirtzis and Dimitris Karlis

Department of Statistics

Athens University of Economics and Business

# Multivariate Poisson Distribution

## Genesis

Assume that  $Y_i$  are independent  $Poisson(\theta_i)$ ,  $i = 0, \dots, n$ . If

$$X_i = Y_0 + Y_i, \quad i = 1, \dots, n$$

then  $(X_1, \dots, X_n)$  will follow jointly a  $n$ -variate Poisson distribution ( $n$  denotes the dimension of the distribution).

## Properties

- Marginally each  $X_i \sim Poisson(\theta_0 + \theta_i)$ .
- Covariance:  $Cov(X_i, X_j) = \theta_0$ ,  $i, j = 1, \dots, n$  and  $i \neq j$

## Joint Probability Function

When  $\mathbf{X}$  is distributed as  $n$ -variate Poisson distribution, the joint probability function of  $(X_1 = x_1, \dots, X_n = x_n)$  is given by:

$$P_{\mathbf{X}}(x_1, \dots, x_n) = \exp\left(-\sum_{k=1}^n \theta_k\right) \prod_{k=1}^n \frac{\theta_k^{x_k}}{x_k!} \sum_{i=0}^s \prod_{j=1}^n \binom{x_j}{i} i! \left(\frac{\theta_0}{\prod_{m=1}^n \theta_m}\right)^i$$

where,  $s = \min\{x_1, \dots, x_n\}$ . Note that if  $s = 0$  then:

$$P_{\mathbf{X}}(x_1, \dots, x_n) = \exp(-\theta_0) \prod_{i=1}^n Po(x_i|\theta_i)$$

where,  $Po(x|\theta) = e^{-\theta}\theta^x/x!$  denotes the probability function of the simple Poisson distribution with parameter  $\theta$ .

## Computational Problems – Motivation

The main obstacle which limits the usage of multivariate Poisson distribution in practice, is the complexity of calculating the probability function.

Here are a few examples:

- Multivariate count data, where each observation has different parameters (due to offsets) and we need to calculate the probabilities for each observation (e.g. incidence of  $n$  different diseases in a series of counties).
- Iterative schemes (Newton Raphson for MLE, E.M. and MCMC) require to calculate the probabilities for each observation at every iteration!

Hence we need to find ways to speed up the calculation of the multivariate probabilities.

## Recurrence Relationships

Computation of probabilities can be accomplished via recursive schemes.

If we will call  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\mathbf{1} = (1, \dots, 1)$  and  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$

then:

$$x_i P_{\mathbf{X}}(\mathbf{x}) = \theta_i P_{\mathbf{X}}(\mathbf{x} - \mathbf{e}_i) + \theta_0 P_{\mathbf{X}}(\mathbf{x} - \mathbf{1}), \quad i = 1, \dots, n$$

Also:

$$P(X_1 = x_1, \dots, X_k = x_k, 0, \dots, 0) = P_{\mathbf{X}} \left( \mathbf{x} - \sum_{i=1}^k \mathbf{e}_i \right) \prod_{i=1}^k \frac{\theta_i}{x_i}$$

$k = 1, \dots, n - 1$  where the order of  $x_i$ 's and 0's can be interchanged to

cover all possible cases. Finally:

$$P_{\mathbf{X}}(0, \dots, 0) = \exp \left( - \sum_{i=0}^n \theta_i \right)$$

With the recurrence relationships, one can get the entire probability table.

This will require  $\prod_{i=1}^n (x_i + 1)$  probabilities to be estimated!

### Objective

Propose efficient strategies for calculating the probabilities based on the existing recurrence relationships.

## Bivariate Case

$$xP(x, y) = \theta_1 P(x - 1, y) + \theta_0 P(x - 1, y - 1)$$

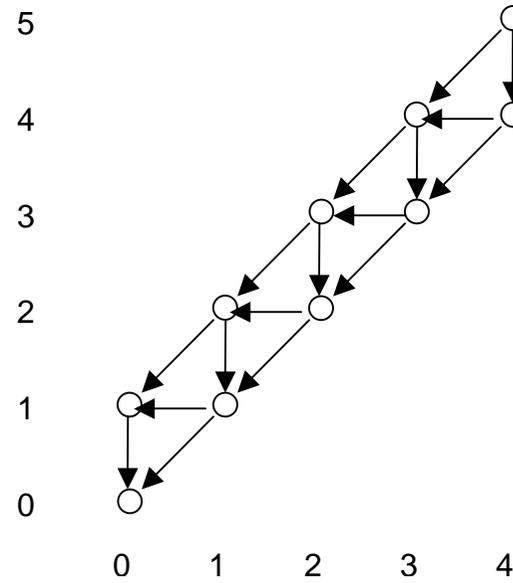
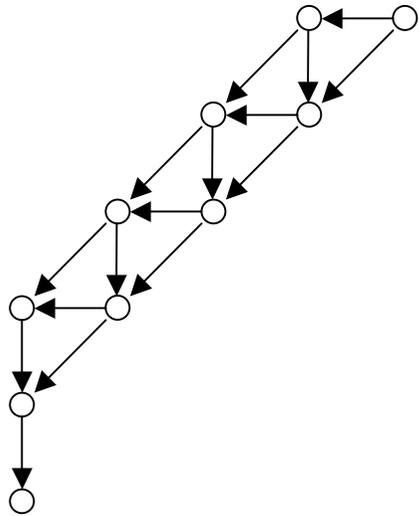
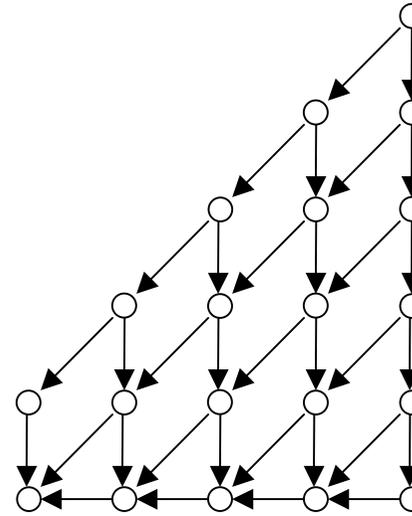
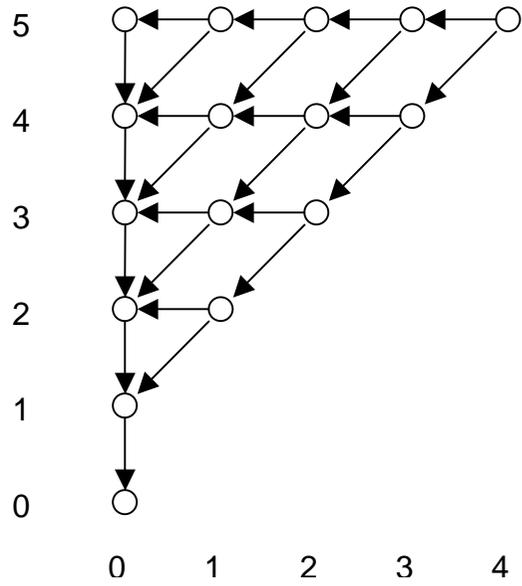
$$yP(x, y) = \theta_2 P(x, y - 1) + \theta_0 P(x - 1, y - 1)$$

with the convention that  $P(x, y) = 0$ , if  $\min\{x, y\} < 0$ .

We have two recurrence relationships: Which one is preferable to be used?

The answer is to use both of them interchangeably! This does not generalize for the case of more variates!

# Calculating P(4,5)



## Trivariate Case

$$xP(x, y, z) = \theta_1 P(x - 1, y, z) + \theta_0 P(x - 1, y - 1, z - 1)$$

$$yP(x, y, z) = \theta_2 P(x, y - 1, z) + \theta_0 P(x - 1, y - 1, z - 1)$$

$$zP(x, y, z) = \theta_3 P(x, y, z - 1) + \theta_0 P(x - 1, y - 1, z - 1).$$

with the convention that  $P(x, y, z) = 0$ , if  $\min\{x, y, z\} < 0$ .

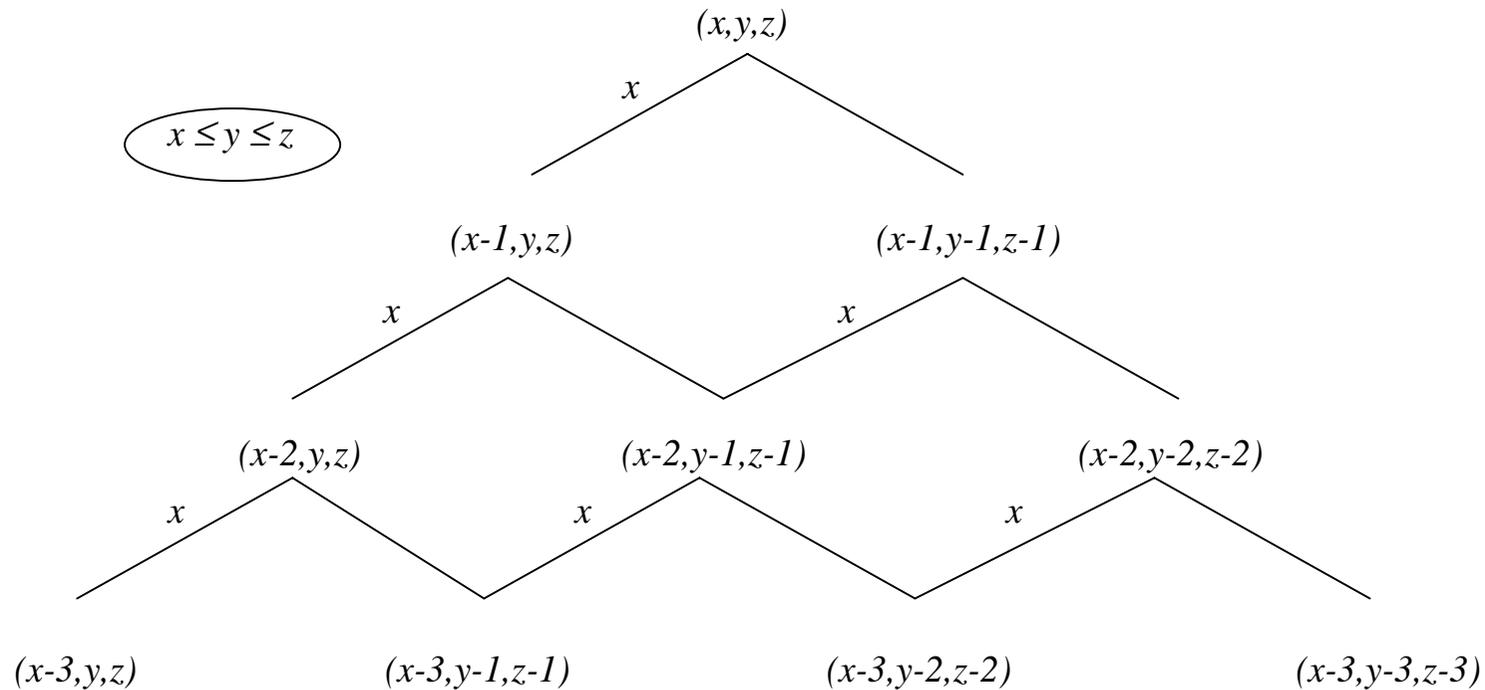
### Different strategies

- Use only one of them (but which one?) (the **Flat** algorithm)
- Use all of them in a particular order (the **Full** algorithm)

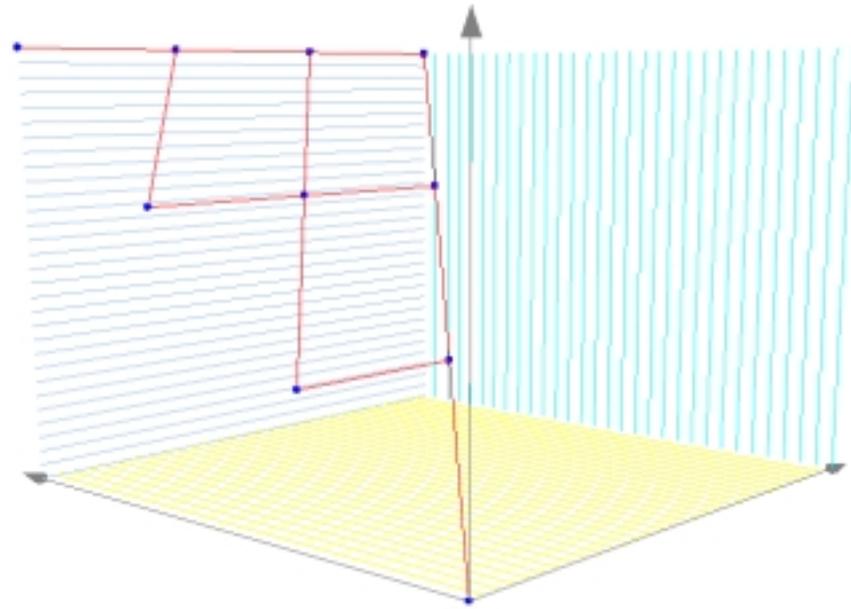
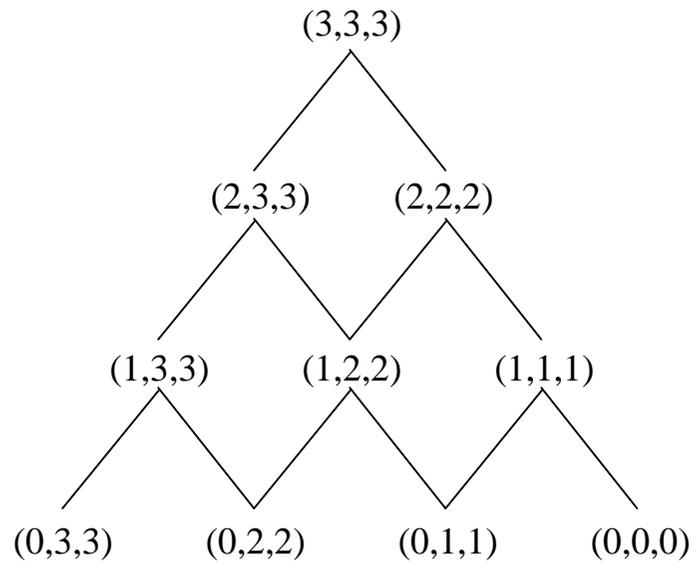
For both algorithms: if one of the coordinates becomes 0 then everything is simpler. So the idea is: how we can come closer to a 0 value with less steps?

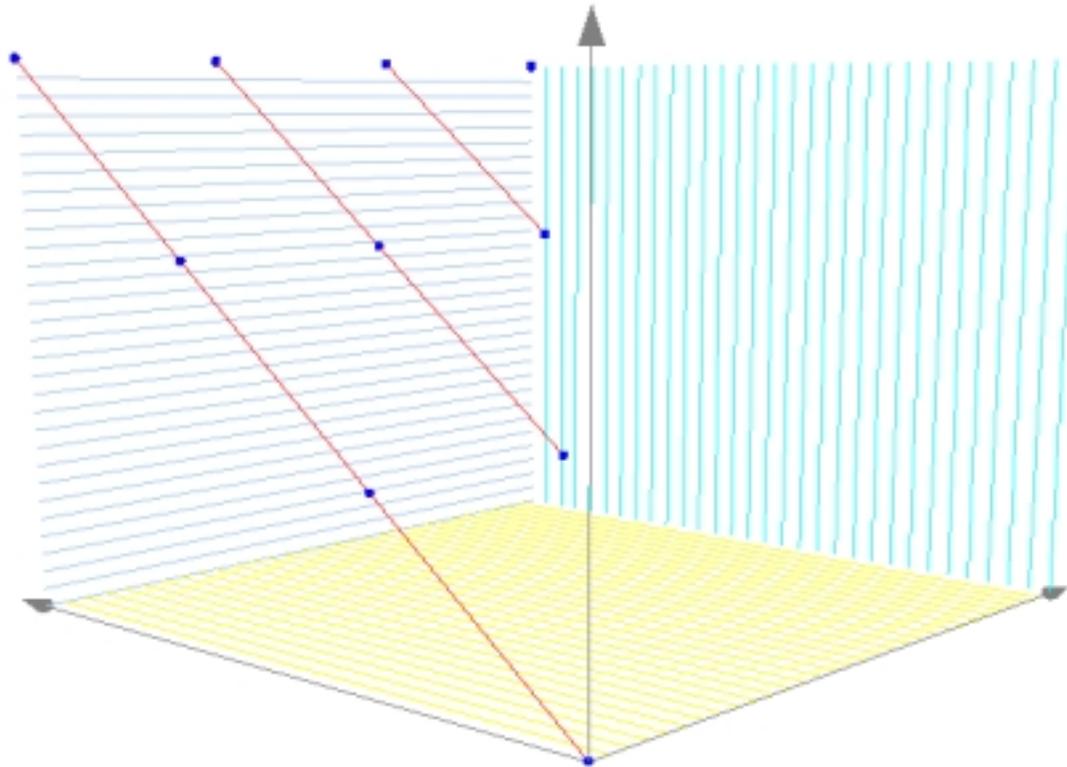
## FLAT ALGORITHM

According to this algorithm we use only the relationship, which involves the coordinate that moves us to the closer plane. Therefore in the case that  $x \leq y \leq z$ , we will use only the first one.



## Calculating $P(3,3,3)$ using the Flat Algorithm

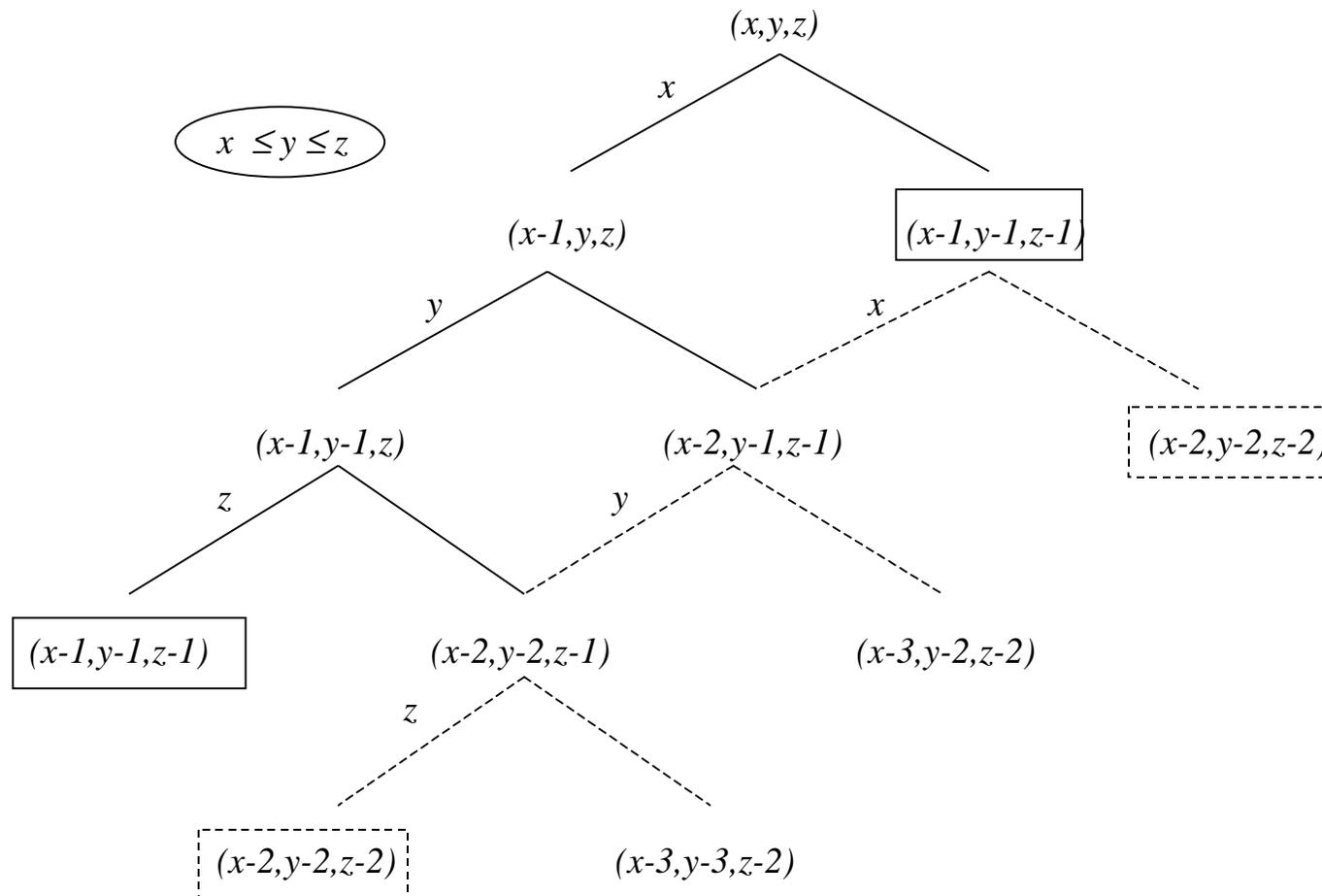




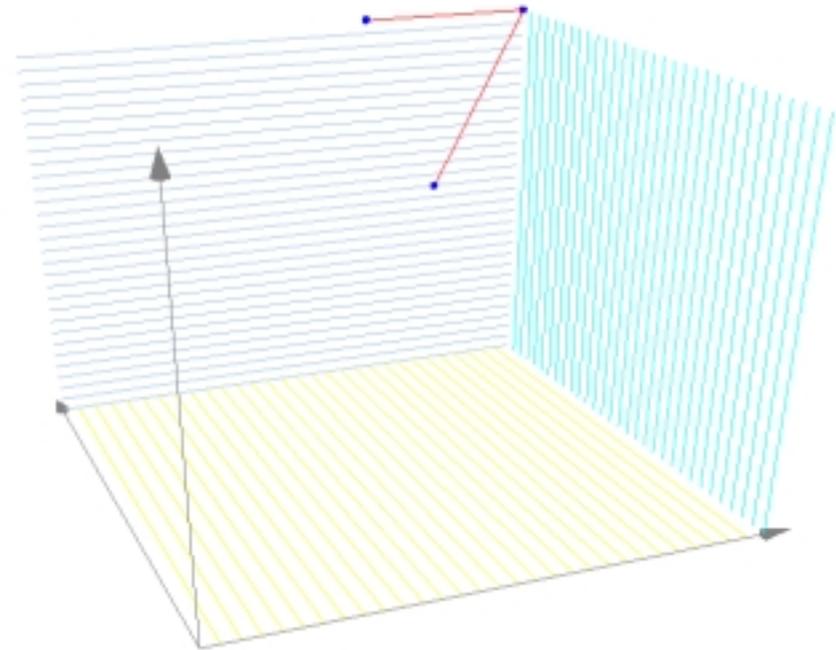
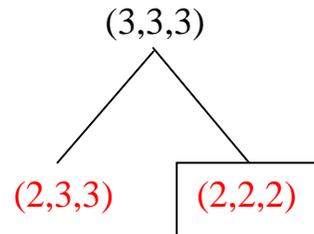
Using the Flat algorithm we are moving along a plane until the minimum coordinate becomes zero.

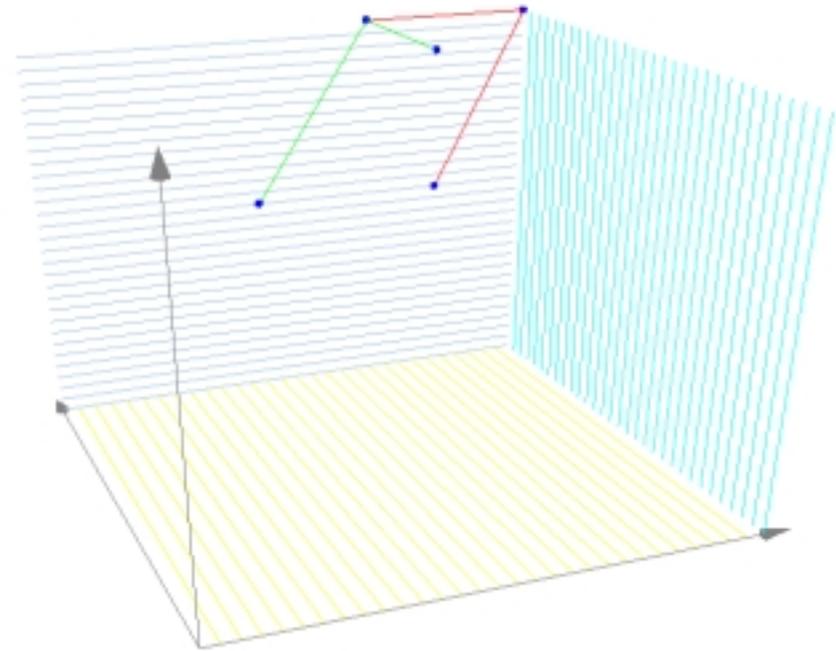
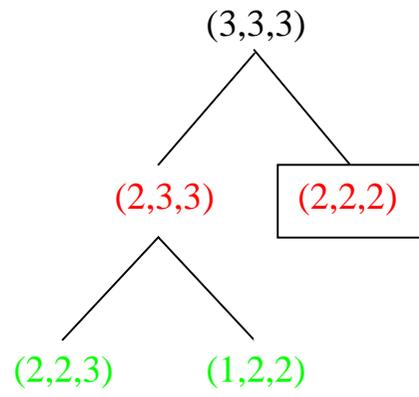
## FULL ALGORITHM

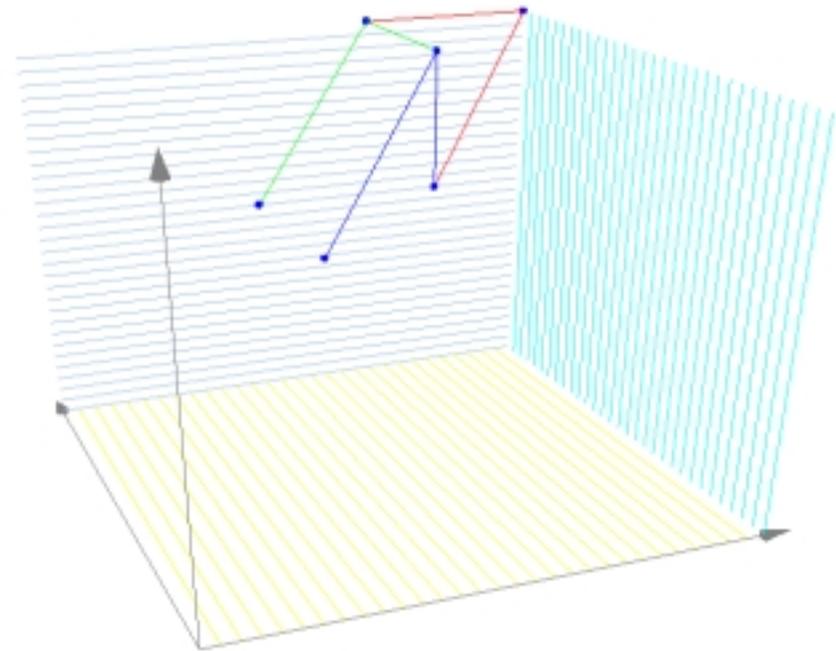
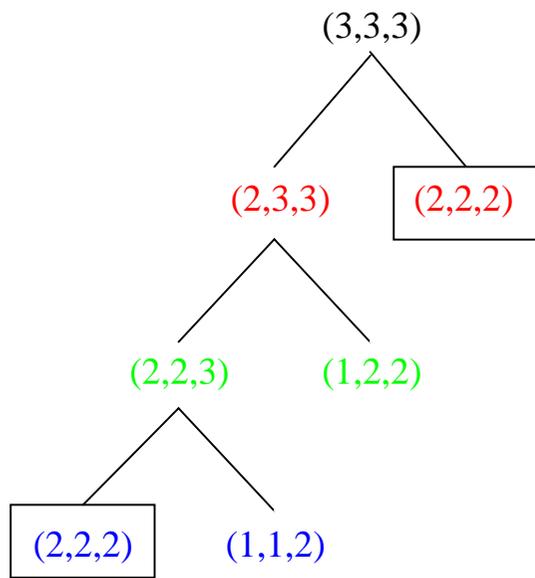
According to this algorithm we use all the recurrence relationships starting with the one which moves us parallel to the minimum coordinate axis, continue with the break up that involves the second order coordinate and finally use the one which involves the maximum. More specifically, if  $x \leq y \leq z$ , we have:

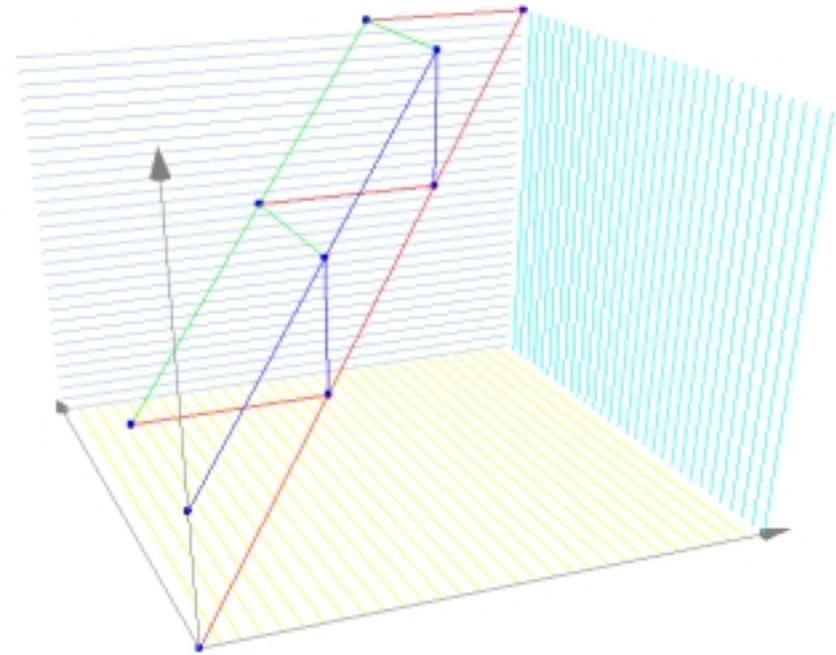
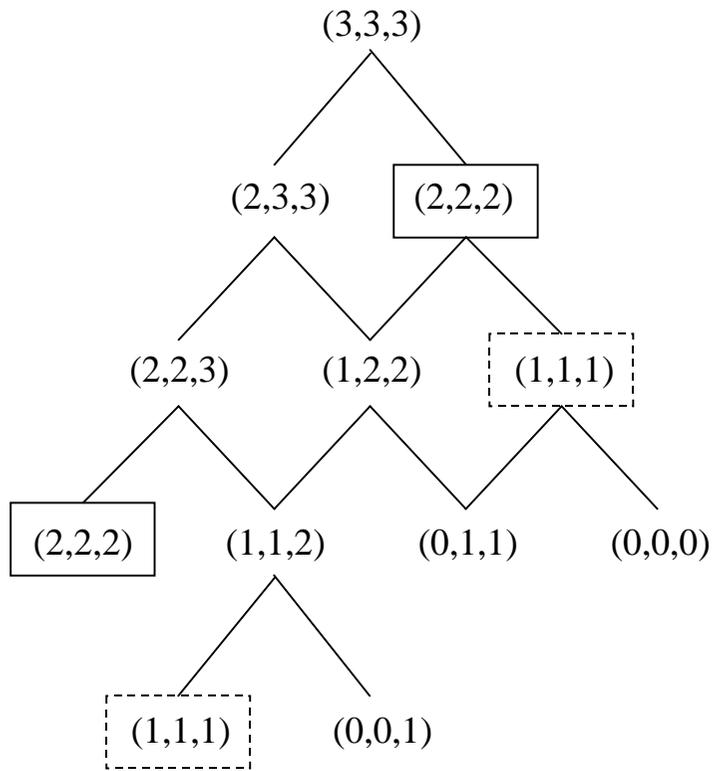


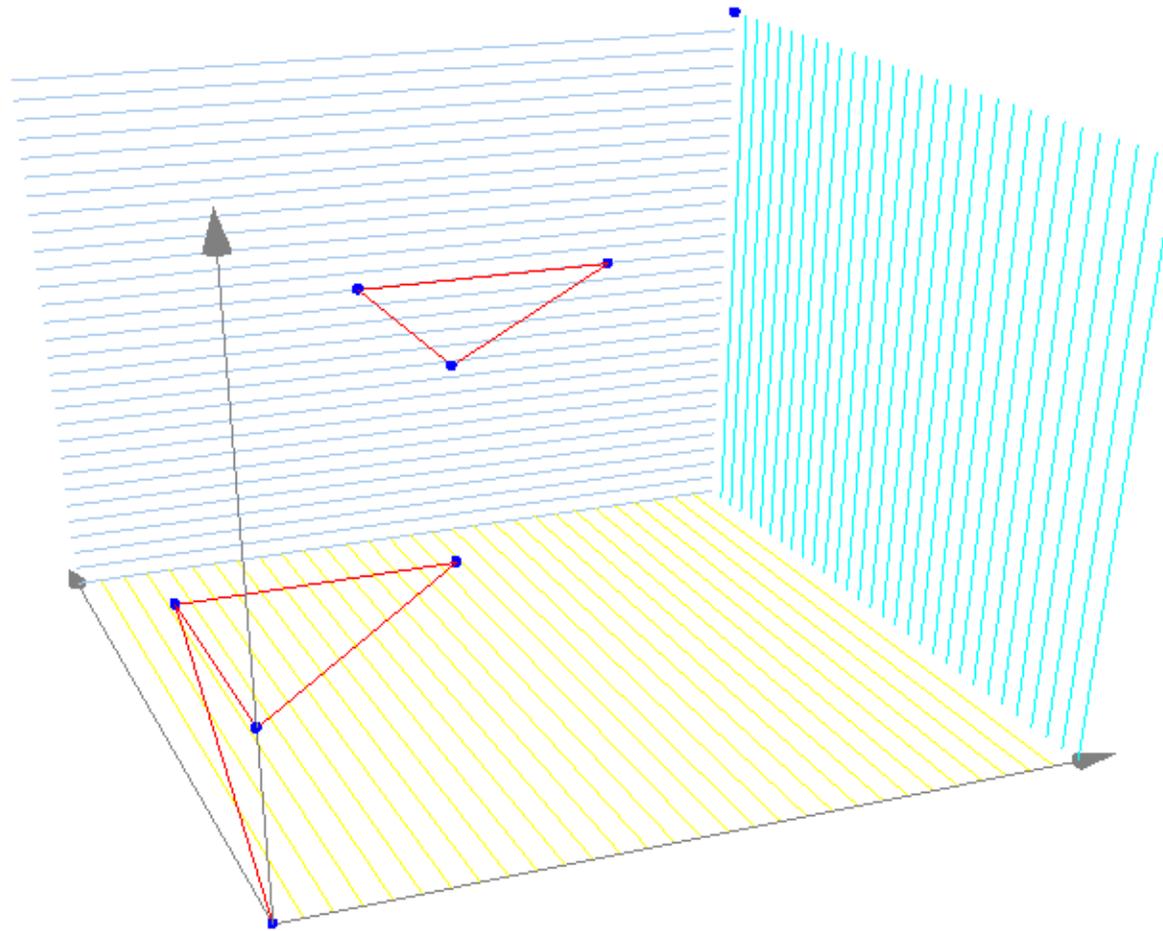
# Calculating $P(3,3,3)$ using the Full Algorithm











Using the Full algorithm we are trying to stay as close as possible to the main diagonal by “traveling” to all dimensions. This results a triangle, which moves towards the closest plane.

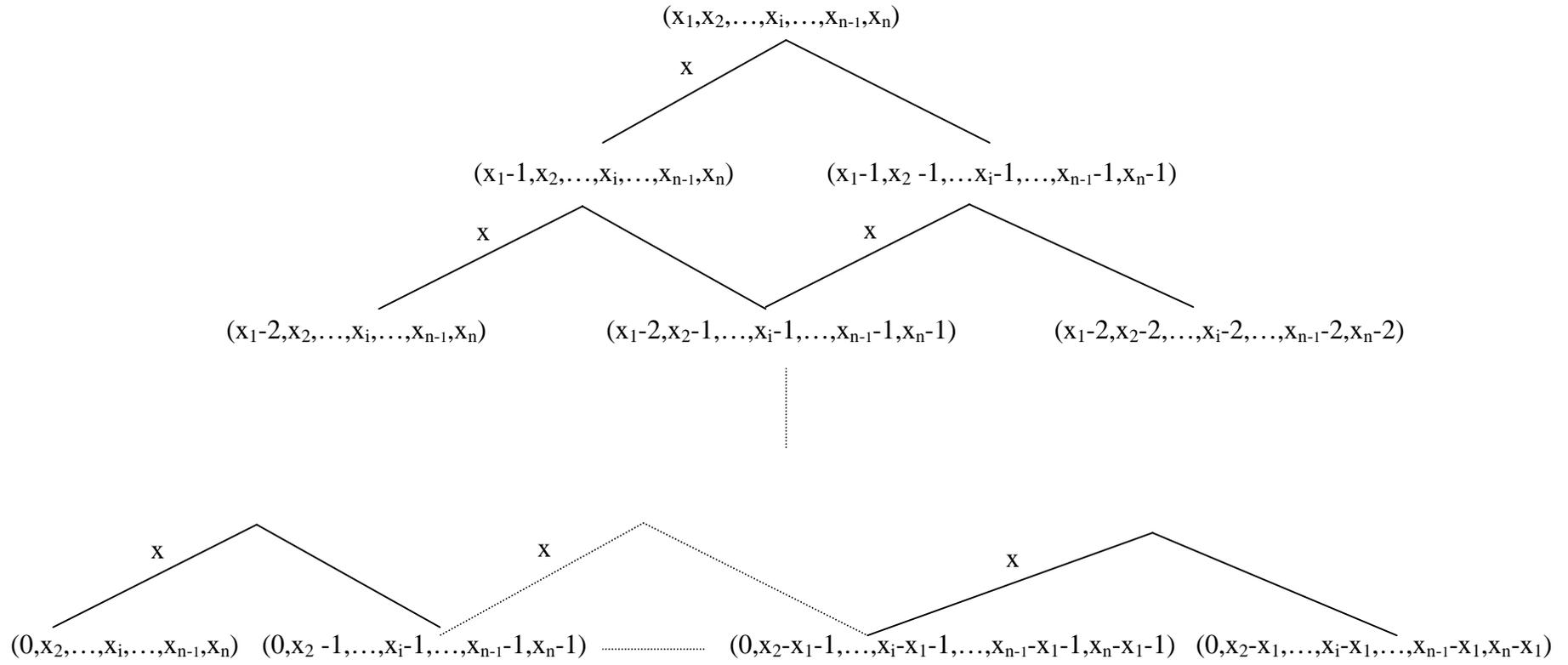
## Multivariate Case

The generalization of the algorithms introduced in the trivariate case is:

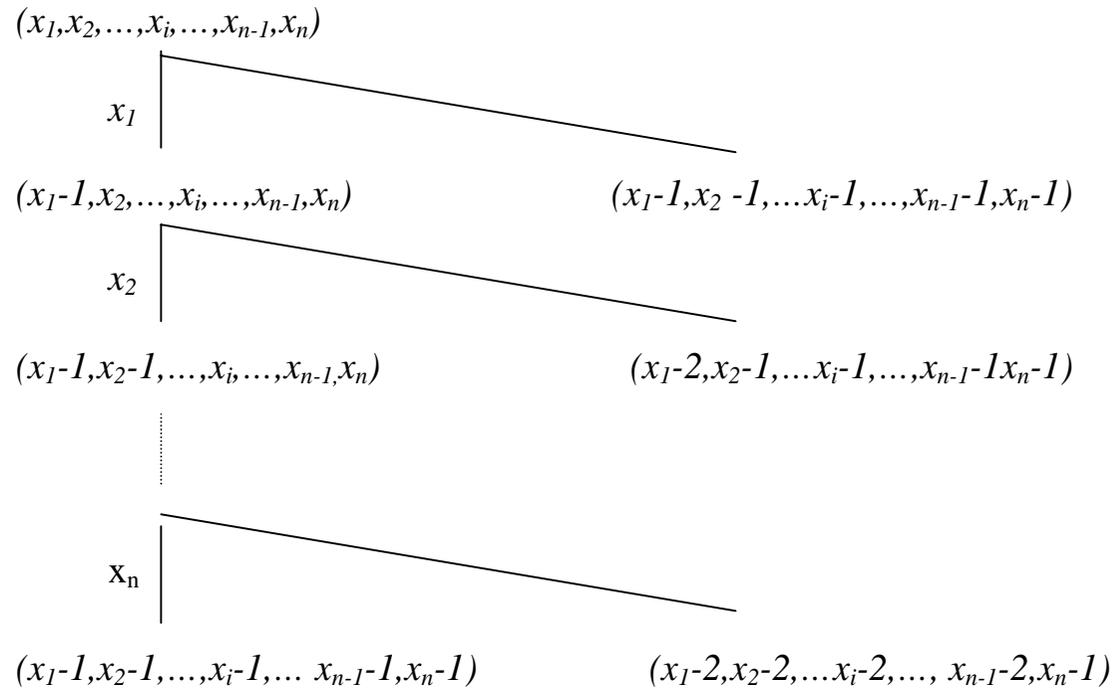
- **Flat Algorithm:** Uses only the recurrence relationship that involves the minimum coordinate.
- **Full Algorithm:** Uses all the recurrence relationships. The order which we use them is indicated by the magnitude of each coordinate (we start breaking up using the relationship that involves the minimum coordinate and we end up with the one that involves the maximum).

Both algorithms terminate when all the points (we ended up) have the minimum coordinate equal 0.

## The Flat Algorithm in the Multivariate Case



## The Full Algorithm in the Multivariate Case



## Comparison of the Algorithms

Which one of the two proposed algorithms is preferable? There is a trade-off between the dimension and the minimum coordinate:

### Lemma

When we have to calculate the  $n$ -variate Poisson probability  $P_{\mathbf{X}}(\mathbf{x})$  then:

- (i) If  $\min\{x_i\} < 2n - 3$ , use the Flat algorithm
- (ii) If  $\min\{x_i\} > 2n - 3$ , use the Full algorithm
- (iii) If  $\min\{x_i\} = 2n - 3$ , it is indifferent which algorithm you will use.

(Note that one may combine the two algorithms for creating more efficient algorithms).

## Comparison of the Algorithms II

In the following table we presents the value of the minimum coordinate for a range of different dimensions  $n$ , in order the flat algorithm to be more efficient.

$n$	2	3	4	5	8	10	12	15	20
$\min\{x_i\}$	0	2	4	6	12	16	20	26	36

For  $n = 2$  the flat algorithm is always inferior, while for  $n = 3$  only if  $\min\{x_i\} = 2$  the flat algorithm is preferable. For higher dimensions the flat algorithm is preferable (for instance if  $n = 10$  then  $\min\{x_i\}$  needs to be higher than 16 in order to prefer the full algorithm).

The total number of points need to be calculated under different configurations can be seen in the following table:

			Flat	Full	number				Flat	Full	number
dim	min	max	Algor.	Algor.	of points	dim	min	max	Algor.	Algor.	of points
2	0	5	✓	✓	6	5	0	5	✓	✓	6
2	0	10	✓	✓	11	5	0	10	✓	✓	11
2	5	10		✓	16	5	5	10	✓		26
2	5	15		✓	21	5	5	15	✓		31
2	10	15		✓	26	5	10	15		✓	56
2	10	20		✓	31	5	10	20		✓	61
3	0	5	✓	✓	6	10	0	5	✓	✓	6
3	0	10	✓	✓	11	10	0	10	✓	✓	11
3	5	10		✓	21	10	5	10	✓		26
3	5	15		✓	26	10	5	15	✓		31
3	10	15		✓	36	10	10	15	✓		71
3	10	20		✓	41	10	10	20	✓		76

## Extensions

### Other Families of Bivariate Distributions

Panjer's univariate family denoted by  $K \sim \mathcal{R}(\alpha, \beta)$  is defined as:

$$P(K = k) = \left( \alpha + \frac{\beta}{k} \right) P(K = k - 1), \quad k \geq 1, \quad \alpha, \beta \geq 0$$

If  $R_j \sim \mathcal{R}(\alpha_j, \beta_j)$ ,  $j = 0, 1, 2$  mutually independent, and:

$$X = R_0 + R_1$$

$$Y = R_0 + R_2$$

then  $(X, Y)$  belong to Hesselager's bivariate family of distributions.

The resulting bivariate distribution has recurrence relationships of the form:

$$\begin{aligned}
 P(x, y) &= \left( \alpha_0 + \frac{\beta_0}{x} \right) P(x-1, y-1) + \left( \alpha_1 + \frac{\beta_1}{x} \right) P(x-1, y) \\
 &+ \left( \alpha_0 \alpha_1 + \frac{\alpha_1 \beta_0 + \alpha_0 \beta_1}{x} \right) P(x-2, y-1), \quad x \geq 1
 \end{aligned}$$

$$\begin{aligned}
 P(x, y) &= \left( \alpha_0 + \frac{\beta_0}{y} \right) P(x-1, y-1) + \left( \alpha_2 + \frac{\beta_2}{y} \right) P(x, y-1) \\
 &+ \left( \alpha_0 \alpha_2 + \frac{\alpha_2 \beta_0 + \alpha_0 \beta_2}{y} \right) P(x-1, y-2), \quad y \geq 1.
 \end{aligned}$$

respectively, assuming that  $P(x, y) = 0$  if  $\min\{x, y\} < 0$ .

## Other Extensions

- Multivariate Poisson distribution with more covariance structure
- Other multivariate distributions with more complicated recurrence relationships
- Calculation of moments via recurrence relationships (useful for actuarial purposes)