(iv) $F_{\mathbf{X}}$ is continuous from the right with respect to each of the coordinates, or any subset of them jointly, i.e., if $\mathbf{x}=(x_1, \ldots, x_n)$ and $\mathbf{z}_m=(z_{1m}, \ldots, z_{nm})$ such that $\mathbf{z}_m \downarrow \mathbf{x}$, then

$$F_{\mathbf{X}}(\mathbf{z}_m) \downarrow F_{\mathbf{X}}(\mathbf{x}).$$

(v) If $x_i \le x_i', i=1, \ldots, n$, then

$$F_{\mathbf{X}}(x_1, \ldots, x_n) \le F_{\mathbf{X}}(t_1, \ldots, t_n) \le F_{\mathbf{X}}(x_1', \ldots, x_n'),$$

where $t_i \in \{x_i, x_i'\}, i=1, 2, \ldots, n$. When $n=2$, we have

$$F_{X_1,X_2}(x_1, x_2) \le \left\{ \begin{array}{l} F_{X_1,X_2}(x_1, x_2') \\ F_{X_1,X_2}(x_1', x_2) \end{array} \right\} \le F_{X_1,X_2}(x_1', x_2').$$

(vi) If $x_1 \le x_1'$ and $x_2 \le x_2'$, then

$$P(x_1 < X_1 \le x_1', x_2 < X_2 \le x_2') = F_{X_1,X_2}(x_1', x_2') - F_{X_1,X_2}(x_1, x_2') - F_{X_1,X_2}(x_1', x_2) + F_{X_1,X_2}(x_1, x_2)$$

In particular, let $x_1' \uparrow \infty$ and $x_2' \uparrow \infty$, we get

$$P(x_1 < X_1 < \infty, x_2 < X_2 < \infty) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1,X_2}(x_1, x_2).$$

(vii) The joint cdf of $X_1, \ldots, X_k$, $k<n$, is

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = P(X_1 \le x_1, \ldots, X_k \le x_k) = P(X_1 \le x_1, \ldots, X_k \le x_k, -\infty < X_{k+1} < \infty, \ldots, -\infty < X_n < \infty) = \lim_{x_{k+1}, x_{k+2}, \cdots, x_n \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n).$$

In particular, the marginal cdf of X_1 is

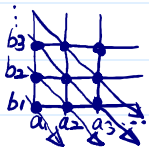
$$F_{X_1}(x) = P(X_1 \le x) = \lim_{x_2, x_3, \cdots, x_n \to \infty} F_{\mathbf{X}}(x, x_2, x_3, \ldots, x_n).$$

Theorem. A function $F_{\mathbf{X}}(x_1, \ldots, x_n)$ can be a joint cdf if $F_{\mathbf{X}}$ satisfies (i)-(v) in the previous theorem.

➢ Joint Probability Mass Function

- **Definition.** Suppose that $X_1, \ldots, X_n$ are discrete random variables. The joint pmf of $\mathbf{X}=(X_1, \ldots, X_n)$ is defined as

$$p_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n).$$



$P_{\mathbf{X}}: \mathbb{R}^n \to \mathbb{R}$

- **Theorem.** Suppose that $p_{\mathbf{X}}$ is a joint pmf. Then,

$X_1 \in \{a_1, a_2, \cdots\} \equiv \mathcal{X}_1$, $X_2 \in \{b_1, b_2, \cdots\} \equiv \mathcal{X}_2$, $\mathcal{X}_1 \times \mathcal{X}_2$ is countable

(a) $p_{\mathbf{X}}(x_1, \ldots, x_n) \geq 0$, for $-\infty < x_i < \infty$, $i = 1, \ldots, n$.

$P_{\mathbf{X}}(x) \leq 1$

(b) There exists a finite or countably infinite set $\mathcal{X} \subset \mathbb{R}^n$ such that $p_{\mathbf{X}}(x_1, \ldots, x_n) = 0$, for $(x_1, \ldots, x_n) \notin \mathcal{X}$.

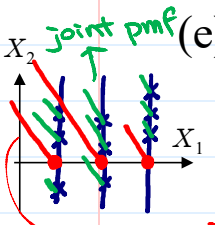
(c) $\sum_{\mathbf{x} \in \mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) = 1$, where $\mathbf{x} = (x_1, \ldots, x_n)$.

(d) For $A \subset \mathbb{R}^n$, $P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A \cap \mathcal{X}} p_{\mathbf{X}}(\mathbf{x})$.

number of all r.v.'s

(e) The joint pmf of $X_1, \ldots, X_k$, $k < n$, is

can be any k r.v.'s



$$p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$$
$$= P(X_1 = x_1, \ldots, X_k = x_k, -\infty < X_{k+1} < \infty, \ldots, -\infty < X_n < \infty) \quad \text{(same event)}$$
$$= \sum_{\substack{(x_1,\ldots,x_n) \in \mathcal{X} \\ -\infty < x_{k+1} < \infty, \ldots, -\infty < x_n < \infty}} p_{\mathbf{X}}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n).$$

($x_{k+1}, \ldots, x_n$: $-\infty$ to ∞)

joint ⇒ marginal

In particular, the marginal pmf of X_1 is

can be any one of the n r.v.'s

check the example in LNp.7-3

$$p_{X_1}(x) = P(X_1 = x)$$
$$= \sum_{\substack{(x, x_2, \ldots, x_n) \in \mathcal{X} \\ -\infty < x_2 < \infty, \ldots, -\infty < x_n < \infty}} p_{\mathbf{X}}(x, x_2, x_3, \ldots, x_n).$$

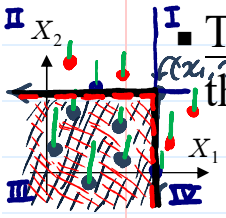
($x_2, x_3, \ldots, x_n$: $-\infty$ to ∞)

check the graph in LNp.7-8

- **Theorem.** A function $p_{\mathbf{X}}(x_1, \ldots, x_n)$ can be a joint pmf if $p_{\mathbf{X}}$ satisfies (a)-(c) in the previous theorem.

discrete

- **Theorem.** If $F_{\mathbf{X}}$ and $p_{\mathbf{X}}$ are the joint cdf and joint pmf of $\mathbf{X}$, then



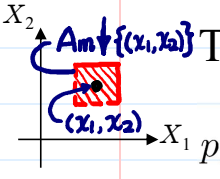
$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \sum_{\substack{(t_1,\ldots,t_n) \in \mathcal{X} \\ t_1 \leq x_1, \ldots, t_n \leq x_n}} p_{\mathbf{X}}(t_1, \ldots, t_n).$$

$P(X \in A)$

$A = (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$

Recall. $P(x) = F(x) - F(x-)$ in LNp.5-11

$P_{\mathbf{X}}(x_1, \cdots, x_n) \neq P(X_1 \leq x_1, \cdots, X_n \leq x_n) - P(X_1 < x_1, \cdots, X_n < x_n)$

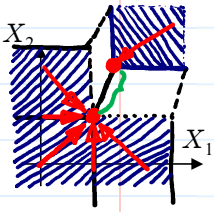


To derive $p_{\mathbf{X}}$ from $F_{\mathbf{X}}$, take $n=2$ to illustrate:

What if $n \geq 2$? Check (□) in LNp.7-6

$A_m \equiv$

$$p_{\mathbf{X}}(x_1, x_2) = \lim_{m \to \infty} P\left(x_1 - \frac{1}{m} < X_1 \leq x_1 + \frac{1}{m}, x_2 - \frac{1}{m} < X_2 \leq x_2 + \frac{1}{m}\right)$$



$$= \lim_{m \to \infty} \big[F_{\mathbf{X}}(x_1 + 1/m, x_2 + 1/m) - F_{\mathbf{X}}(x_1 + 1/m, x_2 - 1/m) - F_{\mathbf{X}}(x_1 - 1/m, x_2 + 1/m) + F_{\mathbf{X}}(x_1 - 1/m, x_2 - 1/m) \big]$$

$\{X_1 \leq x_1, X_2 \leq x_2\}$, $\{X_1 \leq x_1, X_2 < x_2\}$, $\{X_1 < x_1, X_2 \leq x_2\}$, $\{X_1 < x_1, X_2 < x_2\}$

by (Δ) in LNp.7-6

$$= F_{\mathbf{X}}(x_1, x_2) - F_{\mathbf{X}}(x_1, x_2-) - F_{\mathbf{X}}(x_1-, x_2) + F_{\mathbf{X}}(x_1-, x_2-)$$

Joint Probability Density Function

$X:\Omega\to\mathbb{R}^n$, $f_X:\mathbb{R}^n\to\mathbb{R}$

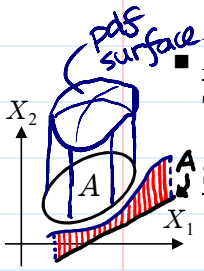
Note that pmf↔pdf, $\Sigma\leftrightarrow\int$

- **Definition.** A function $f_{\mathbf{X}}(x_1, \ldots, x_n)$ can be a joint pdf if

(1) $f_{\mathbf{X}}(x_1, \ldots, x_n)\geq 0$, for $-\infty<x_i<\infty$, $i=1, \ldots, n$.

(2) $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1\cdots dx_n = 1$. ($\int\cdots\int_{\mathbb{R}^n}$)

If $\dim(A)<n$, $P(\mathbf{X}\in A)=0$



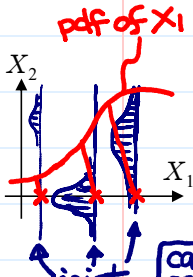
- **Definition.** Suppose that $X_1, \ldots, X_n$ are continuous r.v.'s. The joint pdf of $\mathbf{X}=(X_1, \ldots, X_n)$ is a function $f_{\mathbf{X}}(x_1, \ldots, x_n)$ satisfying (1) and (2) above, and for any event $A\subset\mathbb{R}^n$,

$$P(\mathbf{X}\in A) = \int\cdots\int_A f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1\cdots dx_n.$$

In practice, could be difficult to identify A for the event of interest. (multiple integration)

joint ⇒ marginal

- **Theorem.** Suppose that $f_{\mathbf{X}}$ is the joint pdf of $\mathbf{X}=(X_1, \ldots, X_n)$. Then, the joint pdf of $X_1, \ldots, X_k$, $k<n$, is (can be any R r.v.'s in $\mathbf{X}$)



$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) \overset{\text{by (*) in LNp.7-11}}{=} \frac{\partial^k}{\partial x_1\cdots\partial x_k} F_{X_1,\ldots,X_k} \overset{(\square)}{=} \frac{\partial^k}{\partial x_1\cdots\partial x_k}\left(\lim_{\substack{x_{k+1}\to\infty\\ \vdots\\ x_n\to\infty}} F_{X_1,\ldots,X_k,X_{k+1},\ldots,X_n}\right)$$

(by (□) in LNp.7-7)

$$\overset{\text{check (A) in LNp.7-11}}{=} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)\, dx_{k+1}\cdots dx_n.$$

($x_{k+1}, \ldots, x_n$: $-\infty\to\infty$)

$\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(t_1,\ldots,t_k,t_{k+1},\ldots,t_n)\,dt_1\cdots dt_n$

In particular, the marginal pdf of X_1 is (can be any one of the n r.v.'s)

$$f_{X_1}(x) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x, x_2, \ldots, x_n)\, dx_2\cdots dx_n.$$

($x_2, \ldots, x_n$: $-\infty\to\infty$)

- **Theorem.** If $F_{\mathbf{X}}$ and $f_{\mathbf{X}}$ are the joint cdf and joint pdf of $\mathbf{X}$, then (continuous)

($P(\mathbf{X}\in A)=\int\cdots\int_A f_{\mathbf{X}}(t_1,\cdots,t_n)\,dt_1\cdots dt_n$, $A=(-\infty,x_1]\times\cdots\times(-\infty,x_n]$.)

(*) check (□) in LNp.7-10

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \ldots, t_n)\, dt_1\cdots dt_n, \text{ and}$$

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1\cdots\partial x_n} F_{\mathbf{X}}(x_1, \ldots, x_n).$$

at the continuity points of $f_{\mathbf{X}}$.

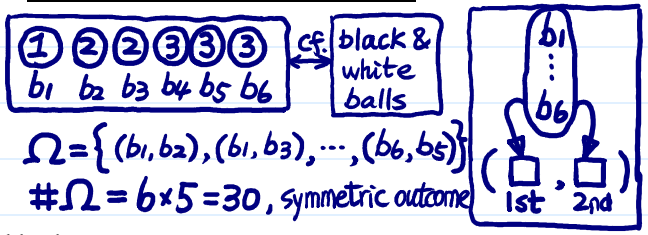
(△) $\int\cdots\int_B f_{X_1,\ldots,X_k}(x_1,\ldots,x_k)\,dx_1\cdots dx_k \overset{?}{=} P((X_1,\ldots,X_k)\in B)$ ($B\subset\mathbb{R}^k$)

$= P((X_1,\ldots,X_k)\in B \text{ and } (X_{k+1},\ldots,X_n)\in\mathbb{R}^{n-k})$

$= \int\cdots\int_B\int\cdots\int_{\mathbb{R}^{n-k}} f_{\mathbf{X}}(x_1,\ldots,x_k,x_{k+1},\ldots,x_n)\,dx$

- Examples.

➢ Experiment. Two balls are drawn without replacement from a box with one ball labeled 1, two balls labeled 2, three balls labeled 3. (6 balls)



$\Omega=\{(b_1,b_2),(b_1,b_3),\cdots,(b_6,b_5)\}$

$\#\Omega = 6\times 5 = 30$, symmetric outcome

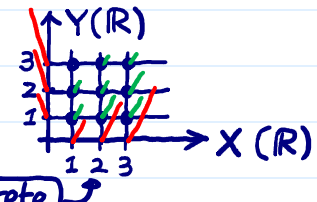
$X:\Omega\to\mathbb{R}$, $Y:\Omega\to\mathbb{R}$

Let

X = label on the 1st ball drawn,

Y = label on the 2nd ball drawn.

(jointly distributed r.v.'s)



- The joint pmf and marginal pmfs of (X, Y) are (discrete)

joint pmf

$p(x,y)$		X			$p_Y(y)$
		1	2	3	
Y	1	0	2/30	3/30	1/6
	2	2/30	2/30	6/30	2/6
	3	3/30	6/30	6/30	3/6
$p_X(x)$		1/6	2/6	3/6	

marginal pmf of Y

marginal pmf of X

{{Y=1},{Y=2},{Y=3}}: a partition of Ω.
{{X=1},{X=2},{X=3}}: a partition of Ω.

P(X=3,Y=3) = P(X=3)P(Y=3|X=3) = 3/6 × 2/5 = 6/30 ← #Ω

Q: The balls are drawn without replacement. Why do X (from 1st ball) and Y (from 2nd ball) have same marginal distributions?

Z=|X-Y|, Z:Ω→ℝ, P(Z=1)=?

- **Q**: $P(|X-Y|=1)=??$ ← need joint dist.

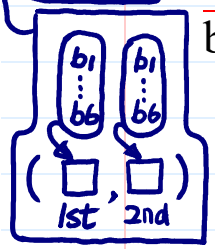
transformation of (X,Y)

$|X(\omega)-Y(\omega)|, \omega\in\Omega$

$P(|X-Y|=1) = P(X=1, Y=2) + P(X=2, Y=1) + P(X=2, Y=3) + P(X=3, Y=2) = 8/15.$

Law of Total Probability: $P(X=x)=\sum_{y=1}^{3} P(X=x|Y=y)P(Y=y) = \sum_{y=1}^{3} P(X=x, Y=y)$

#Ω=36



P(|X-Y|=1) = 4/9

- **Q**: What are the joint pmf and marginal pmfs of (X, Y) if the balls are drawn with replacement (LNp. 4-6)?

$p(x,y)$		X			$p_Y(y)$
		1	2	3	
Y	1	1/36	2/36	3/36	1/6
	2	2/36	4/36	6/36	2/6
	3	3/36	6/36	9/36	3/6
$p_X(x)$		1/6	2/6	3/6	

same marginal pmf as without replacement

same marginal pmf as without replacement

$P(x,y)=p_X(x)p_Y(y)$ → independent

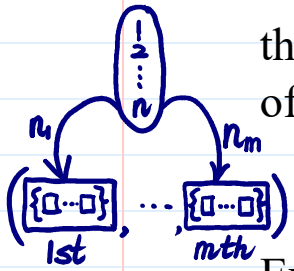
➢ Multinomial Distribution ⟷ c.f. binomial distribution

- Recall. Partitions
  - If $n\geq 1$ and $n_1, \ldots, n_m \geq 0$ are integers for which

$$n_1 + \cdots + n_m = n,$$

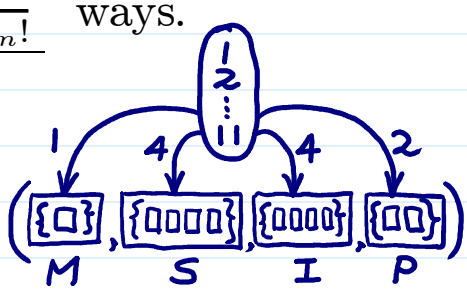
then a set of n elements may be partitioned into m subsets of sizes $n_1, \ldots, n_m$ in

$$\binom{n}{n_1,\cdots,n_m} = \frac{n!}{n_1!\times\cdots\times n_m!} \text{ ways.}$$



  - Example (LNp.2-8) : MISSISSIPPI (1 2 3 4 ⋯ 11)

$$\binom{11}{4,1,2,4} = \frac{11!}{4!1!2!4!}.$$



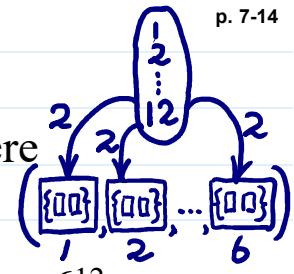
- Example (Die Rolling) ⟷ c.f. Coin Tossing
  - **Q**: If a balanced (6-sided) die is rolled 12 times, P(each face appears twice)=??

各種結果出現次數

相同實驗，不斷重覆

  - Sample space of rolling the die once (basic experiment): $\Omega_0 = \{1, 2, 3, 4, 5, 6\}$. ← symmetric outcomes

- The sample space for the 12 trials is
$$\Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^{12}$$
An outcome $\omega \in \Omega$ is $\omega = (i_1, i_2, \ldots, i_{12})$, where $1 \le i_1, \ldots, i_{12} \le 6$.



symmetric outcomes ∵ balanced die

- There are 6^{12} possible outcomes in Ω, i.e., $\#\Omega = 6^{12}$.
- Among all possible outcomes, there are $\binom{12}{2,2,2,2,2,2} = \frac{12!}{(2!)^6}$ of which each face appears twice.
- $P(\text{each face appears twice}) = \frac{12!}{(2!)^6}\left(\frac{1}{6}\right)^{12} = \frac{1}{\#\Omega}$

In Ω_0, $P(\{1\}) = \cdots = P(\{6\}) = 1/6$

- Generalization.
  - Consider a basic experiment which can result in one of m types of outcomes. Denote its sample space as
$$\Omega_0 = \{1, 2, \ldots, m\}$$ ← may not be symmetric
Let $p_i = P(\text{outcome } i \text{ appears in a basic experiment})$,
then, (i) $p_1, \ldots, p_m \ge 0$, and
(ii) $p_1 + \cdots + p_m = 1$.

  - Repeat the basic experiment n times. Then, the sample space for the n trials is
$$\Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^n$$
$\omega \in \Omega$, ω: n-dim vector $= (\omega_1, \cdots, \omega_n)$, $\omega_i \in \Omega_0 = \{1, \ldots, m\}$

discrete r.v.

Let X_i = # of trials with outcome i, $i = 1, \ldots, m$, $X_i(\omega)$

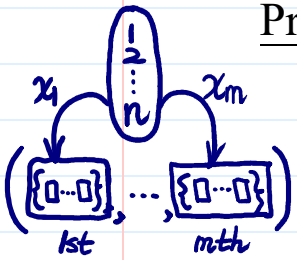
Then, (i) $X_1, \ldots, X_m : \Omega \to \mathbb{R}$, and
(ii) $X_1 + \cdots + X_m = n$.

jointly distributed $\mathbf{X} = (X_1, \cdots, X_m) : \Omega \to \mathbb{R}^m$

  - The joint pmf of $X_1, \ldots, X_m$ is
$$p_{\mathbf{X}}(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m) = \binom{n}{x_1, \cdots, x_m} p_1^{x_1} \times \cdots \times p_m^{x_m}.$$
for $x_1, \ldots, x_m \ge 0$ and $x_1 + \cdots + x_m = n$.

integers; range $\mathcal{X}$

Proof. The probability of any sequence with x_i i's is



ω

$$p_1^{x_1} \times \cdots \times p_m^{x_m},$$
and there are
$$\binom{n}{x_1, \cdots, x_m}$$
such sequences. ← they are symmetric outcomes.

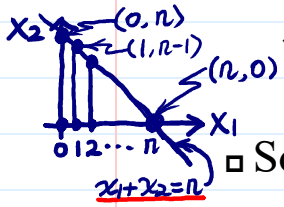
Q: $\sum_{x \in \mathcal{X}} p_{\mathbf{X}}(x) \neq 1$

Note. Not saying all ω's in Ω are symmetric

- The distribution of a random vector $\mathbf{X}=(X_1, \dots, X_m)$ with the above joint pmf is called the *multinomial* distribution with parameters n, m, and $p_1, \dots, p_m$, denoted by Multinomial$(n, m, p_1, \dots, p_m)$.
  - The multinomial distribution is called after the Multinomial Theorem:

$$(a_1 + \cdots + a_m)^n = \sum_{x_i \in \{0,\dots,n\};\ i=1,\dots,m;\ x_1+\cdots+x_m=n} \binom{n}{x_1, \cdots, x_m} a_1^{x_1} \times \cdots \times a_m^{x_m} \quad (*)$$

$\sum_{x \in \mathcal{X}} P_X(x) = (p_1+\cdots+p_m)^n = 1$

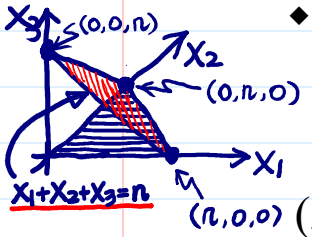


  - It is a generalization of the binomial distribution from 2 types of outcomes to m types of outcomes.

binomial: head, tail; multinomial

- Some Properties.

$m=2$, (X_1, X_2): binomial & $X_2 = n - X_1$



  - Because $X_i = n - (X_1+\cdots+X_{i-1}+X_{i+1}+\cdots+X_m)$, and
$p_i = 1 - (p_1+\cdots+p_{i-1}+p_{i+1}+\cdots+p_m)$,
WLOG, we can write $\in \mathbb{R}^m$, but the dimension of its $\mathcal{X}$ = $m-1$
$(X_1,\dots, X_{m-1}, X_m) \to (X_1,\dots, X_{m-1}, n-(X_1+\cdots+X_{m-1}))$