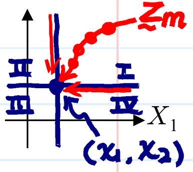
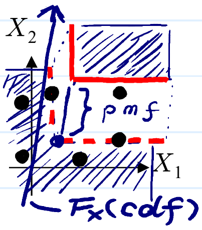
(3) in LNp.5-9 cf.





(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

$\downarrow x_1$ $\downarrow x_n$

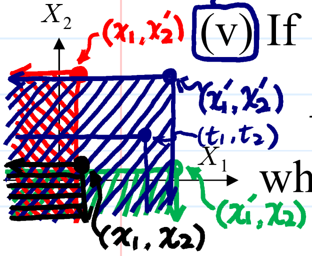
$z_{i_1} \geq z_{i_2} \geq \cdots \geq z_{im} \geq \cdots \geq x_i$
$i=1, \cdots, n$

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

(Ex)

- $(-\infty, \mathbf{z}_m] \equiv (-\infty, z_{1m}] \times \cdots \times (-\infty, z_{nm}]$
- $(-\infty, \mathbf{z}_m] \downarrow (-\infty, \mathbf{x}]$

non-decreasing (cf. (2) in LNp.5-9)



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

$(-\infty, x_1] \times \cdots \times (-\infty, x_n]$
$\subseteq (-\infty, t_1] \times \cdots \times (-\infty, t_n]$
$\subseteq (-\infty, x_1'] \times \cdots \times (-\infty, x_n']$

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

(Ex)

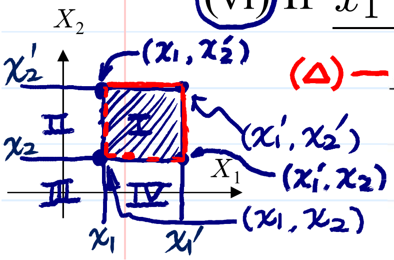
where $t_i \in \{x_i, x_i'\}$, $i=1, 2, \ldots, n$. When n=2, we have

can be generalized to $x_i \leq t_i \leq x_i'$

Q: how to generalize to any n? (□)

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

cf. (5) in LNp 5-10



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

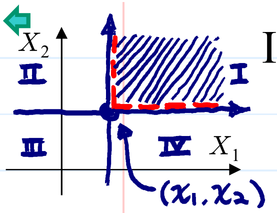
inclusion-exclusion identity (LNp. 3-12)

Blue ∩ (Red ∪ Green)^c
= Blue \ (Red ∪ Green)

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

$(x_1', x_2') \downarrow (x_1, x_2')$ $(x_1', x_2') \downarrow (x_1', x_2)$

e.g., $(x_1', x_2', x_3') \to (x_1, x_2', x_3'), (x_1', x_2, x_3'), (x_1', x_2', x_3)$



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

(Ex) What if $x_1 \downarrow -\infty$, $x_2 \downarrow -\infty$
: if $x_1 \downarrow -\infty$, $x_2' \uparrow \infty$
: if $x_1' \uparrow \infty$, $x_2 \downarrow -\infty$

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

marginal

can be any k r.v.'s

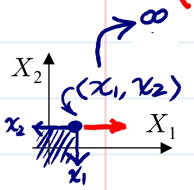
(vii) The joint cdf of $X_1, \ldots, X_k$, $k < n$, is

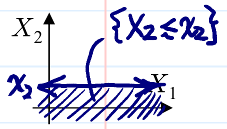
number of all r.v.'s

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

same event

(ii) in LNp.7-5 cf.





In particular, the marginal cdf of X_1 is

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

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

Sometimes, might call it marginal joint cdf or marginal joint distribution (abusing the terminologies)

joint ⇒ marginal

Theorem in LNp.5-11 cf.

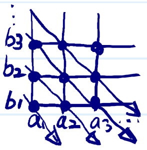
⊙ Theorem. A function $F_{\mathbf{X}}(x_1, \ldots, x_n)$ can be a joint cdf if $F_{\mathbf{X}}$ satisfies (i)-(v) in the previous theorem.

➢ 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

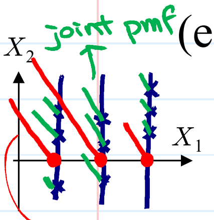
$p_\mathbf{X}: \mathbb{R}^n \to \mathbb{R}$

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



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

- **Theorem.** Suppose that $p_\mathbf{X}$ is a joint pmf. Then,
  (a) $p_\mathbf{X}(x_1, \ldots, x_n) \geq 0$, for $-\infty < x_i < \infty$, $i = 1, \ldots, n$.
  (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



$$\begin{aligned} p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) &= P(X_1 = x_1, \ldots, X_k = x_k) \\ &= P(\boxed{X_1 = x_1, \ldots, X_k = x_k,} \\ &\qquad \boxed{-\infty < X_{k+1} < \infty, \ldots, -\infty < X_n < \infty}) \\ &= \sum_{\substack{(x_1,\ldots,x_n) \in \mathcal{X} \\ \boxed{-\infty < x_{k+1} < \infty}, \ldots, -\infty < x_n < \infty}} p_\mathbf{X}(x_1, \ldots, x_k, \underset{-\infty}{\overset{\infty}{x_{k+1}}}, \ldots, \underset{-\infty}{\overset{\infty}{x_n}}). \end{aligned}$$

same event.

joint ⇒ marginal

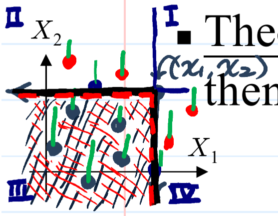
In particular, the marginal pmf of X_1 is

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

check the example in LNp.7-3

$$\begin{aligned} p_{X_1}(x) &= P(X_1 = x) \\ &= \sum_{\substack{(x,x_2,\ldots,x_n) \in \mathcal{X} \\ \boxed{-\infty < x_2 < \infty}, \ldots, -\infty < x_n < \infty}} p_\mathbf{X}(x, \underset{-\infty}{\overset{\infty}{x_2}}, \underset{-\infty}{\overset{\infty}{x_3}}, \ldots, \underset{-\infty}{\overset{\infty}{x_n}}). \end{aligned}$$

- **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.



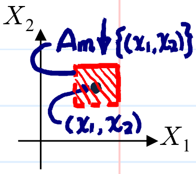
- **Theorem.** If $F_\mathbf{X}$ and $p_\mathbf{X}$ are the joint cdf and joint pmf of $\overset{\text{discrete}}{\mathbf{X}}$, then

$$\underset{P(X \in A)}{\overset{\shortparallel}{F_\mathbf{X}(x_1, \ldots, x_n)}} = \sum_{\substack{(t_1,\ldots,t_n) \in \mathcal{X} \\ \boxed{t_1 \leq x_1}, \ldots, t_n \leq x_n}} p_\mathbf{X}(t_1, \ldots, t_n).$$

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

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

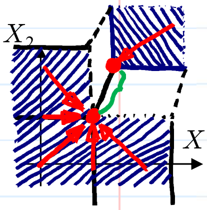
$p_\mathbf{X}(x_1, \cdots, x_n) \overset{?}{=} P(X_1 \leq x_1, \cdots, X_n \leq x_n) - P(X_1 < x_1, \cdots, X_n < x_n)$



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

What if $n > 2$? Check (□) in LNp.7-6

$$p_\mathbf{X}(x_1, x_2) = \lim_{m \to \infty} P\Big(\underbrace{x_1 - \frac{1}{m} < X_1 \leq x_1 + \frac{1}{m}, x_2 - \frac{1}{m} < X_2 \leq x_2 + \frac{1}{m}}_{A_m \equiv}\Big)$$



by (△) in LNp.7-6

$$\begin{aligned} &= \lim_{m \to \infty} \Big[F_\mathbf{X}(x_1 + 1/m, \overset{\to \{X_1 \leq x_1, X_2 \leq x_2\}}{x_2 + 1/m}) - F_\mathbf{X}(x_1 + 1/m, \overset{\to \{X_1 \leq x_1, X_2 \underline{<} x_2\}}{x_2 - 1/m}) \\ &\qquad - F_\mathbf{X}(x_1 - 1/m, \overset{\to \{X_1 \underline{<} x_1, X_2 \leq x_2\}}{x_2 + 1/m}) + F_\mathbf{X}(x_1 - 1/m, \overset{\to \{X_1 \underline{<} x_1, X_2 \underline{<} x_2\}}{x_2 - 1/m}) \Big] \\ &= F_\mathbf{X}(x_1, x_2) - F_\mathbf{X}(x_1, x_2-) - F_\mathbf{X}(x_1-, x_2) + F_\mathbf{X}(x_1-, x_2-) \end{aligned}$$

Note that pmf ⟷ pdf, $\Sigma \longleftrightarrow \int$

Joint Probability Density Function

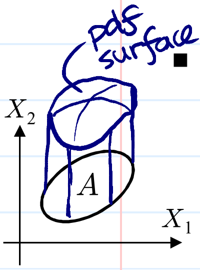
$X:\Omega \to \mathbb{R}^n$, $f_X:\mathbb{R}^n \to \mathbb{R}$

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

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



- **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 $A \subset \mathbb{R}^n$,

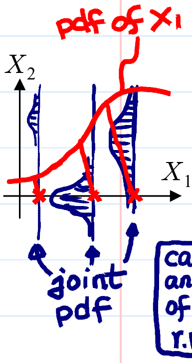
In practice, could be difficult to identify A for the event of interest

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

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 k r.v.'s in $\mathbf{X}$



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

by (0) in LNp 7-7

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_k, \underset{-\infty}{\overset{\infty}{x_{k+1}}}, \ldots, \underset{-\infty}{\overset{\infty}{x_n}})\, dx_{k+1} \cdots dx_n.$$

$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} \Big(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, \cdots t_k, t_{k+1}, \cdots t_n)\, dt_1 \cdots dt_n$

In particular, the marginal pdf of X_1 is

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

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

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

continuous

$P(X \in A) = \int \cdots \int_A f_X(t_1, \cdots, t_n)\, dt_1 \cdots dt_n$

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

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

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

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

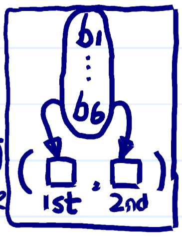
- Examples.

  ➢ 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

① ② ② ③ ③ ③
b_1 b_2 b_3 b_4 b_5 b_6

cf. black & white balls



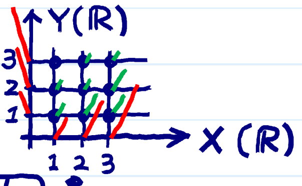
$\Omega = \{(b_1, b_2), (b_1, b_3), \cdots, (b_6, b_5)\}$

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

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

jointly distributed r.v.'s

Let X = label on the 1st ball drawn,
Y = label on the 2nd ball drawn.



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

discrete

joint pmf → $p(x, y)$; $p_Y(y)$ ← marginal pmf of Y

$p(x, y)$		X			$p_Y(y)$
		1	2	3	
Y	1	^{0} 0	^{1} 2/30	^{2} 3/30	1/6
	2	^{1} 2/30	^{0} 2/30	^{1} 6/30	2/6
	3	^{2} 3/30	^{1} 6/30	^{0} 6/30	3/6
marginal pmf of X → $p_X(x)$		1/6	2/6	3/6	

$P(X=3, Y=3)$
$= P(X=3)P(Y=3|X=3)$
$= \frac{3}{6} \times \frac{2}{5} = \frac{6}{30}$ ← #Ω

$Z = |X-Y|$
$Z : \Omega \to \mathbb{R}$
$P(Z=1) = ?$

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

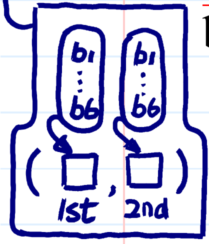
- **Q**: $P(|X-Y|=1)$=?? ← need joint dist

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

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

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

$\#\Omega = 36$



$P(|X-Y|=1)$
$= \frac{4}{9}$

$p(x, y)$		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

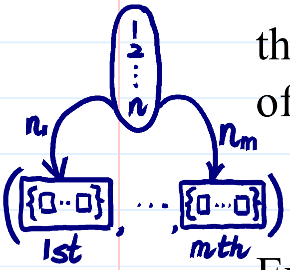
➢ Multinomial Distribution $\xleftrightarrow{\text{c.f.}}$ binomial distribution

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

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

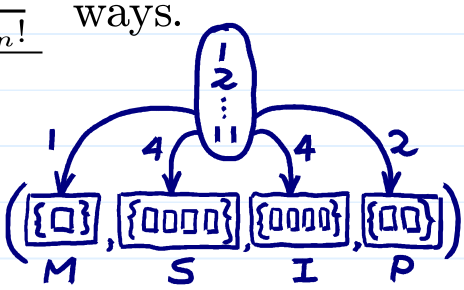
then a set of n elements may be partitioned into m subsets of sizes $n_1, \ldots, n_m$ in

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



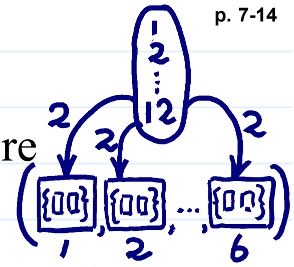
  - ▫ Example (LNp.2-8) : MISSISSIPPI
    1 2 3 4 · · · 11

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



- Example (Die Rolling) $\xleftrightarrow{\text{c.f.}}$ Coin Tossing
  - ▫ **Q**: If a balanced (6-sided) die is rolled 12 times, ← 相同实驗，不斷重覆
    各种結果出現次數 → P(each face appears twice)=??
  - ▫ 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.

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

- $P(\text{each face appears twice}) = \frac{12!}{(2!)^6} \underbrace{\left(\frac{1}{6}\right)^{12}}_{= \frac{1}{\#\Omega}}$

- 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

$$\Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^n$$

→ $\boxed{\omega \in \Omega, \omega: n\text{-dim vector}}$

discrete r.v.

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

(ω)

Then, (i) $X_1, \ldots, X_m : \Omega \to \mathbb{R}$, and $\mathbf{X} = (X_1, \ldots, X_m) : \Omega \to \mathbb{R}^m$

jointly distributed

(ii) $X_1 + \cdots + X_m = n$.

- The joint pmf of $X_1, \ldots, X_m$ is

$$\begin{aligned} p_{\mathbf{X}}(x_1, \ldots, x_m) &= P(X_1 = x_1, \ldots, X_m = x_m) \\ &= \binom{n}{x_1, \ldots, x_m} p_1^{x_1} \times \cdots \times p_m^{x_m}. \end{aligned}$$

integers

for $x_1, \ldots, x_m \ge 0$ and $x_1 + \cdots + x_m = n$. ← $\boxed{\text{range } \mathcal{X}}$

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

(ω)

$$p_1^{x_1} \times \cdots \times p_m^{x_m},$$

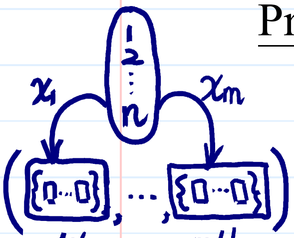
$\boxed{Q: \sum_{\mathbf{x} \in \mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) \overset{?}{=} 1}$

and there are

$$\binom{n}{x_1, \ldots, x_m}$$

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

such sequences. ← they are symmetric outcomes.



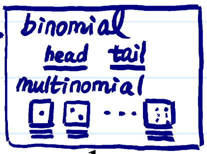
- The distribution of a random vector $\mathbf{X}=(X_1, \ldots, X_m)$ with the above joint pmf is called the *multinomial* distribution with parameters n, m, and $p_1, \ldots, p_m$, denoted by Multinomial$(n, m, p_1, \ldots, p_m)$.
  - The multinomial distribution is called after the Multinomial Theorem:

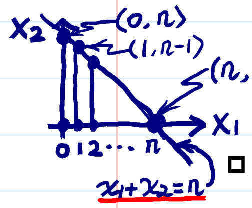
$$\sum_{x \in \mathcal{X}} p_{\mathbf{X}}(x) = (p_1+\cdots+p_m)^n = 1$$

$$(a_1 + \cdots + a_m)^n = \sum_{\substack{x_i \in \{0,\ldots,n\};\ i=1,\ldots,m \\ x_1+\cdots+x_m=n}} \binom{n}{x_1, \cdots, x_m} \overset{p_1}{a_1^{x_1}} \times \cdots \times \overset{p_m}{a_m^{x_m}} \quad (*)$$

$\mathcal{X}$

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





- 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 $\mathcal{X} = m-1$

$$(X_1,\ldots, X_{m-1}, X_m) \rightarrow (X_1,\ldots, X_{m-1}, n-(X_1+\cdots+X_{m-1}))$$