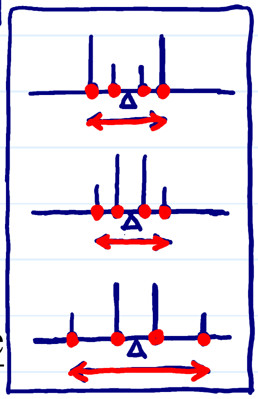
- Example (Committees, LNp.5-6) distribution

$(x-\mu)^2$	$x-\mu$	x	x^2	$f(x)$	$xf(x)$	$(x-\mu)^2 f(x)$	$x^2 f(x)$
4	-2	0	0	5/210	0/210	20/210	0/210
1	-1	1	1	50/210	50/210	50/210	50/210
0	0	2	4	100/210	200/210	0/210	400/210
1	1	3	9	50/210	150/210	50/210	450/210
4	2	4	16	5/210	20/210	20/210	80/210
		Totals		1	2 $E(X)$	2/3 $Var(X)$	14/3 $E(X^2)$

deviation of x from μ

So, $\mu = 2$, $\sigma^2 = 2/3$, and $\sigma = \sqrt{2/3}$

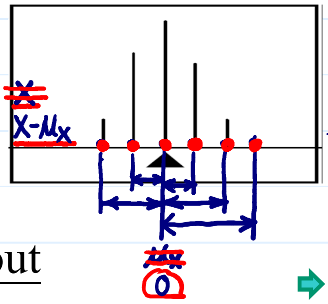


- Note. X (random) cf. distribution (fixed)
  - μ_X and σ_X^2 only depends on f_X. They are *fixed* constants, not *random* numbers.
  - If X has units, then μ_X and σ_X have the same unit as X, and variance has unit squared.

acting like a scale of a ruler
→ standard deviation
→ high probability within $\mu \pm 2\sigma$,
$\sigma^2 = 1$, ± 2
$\sigma^2 = 100$, ± 20

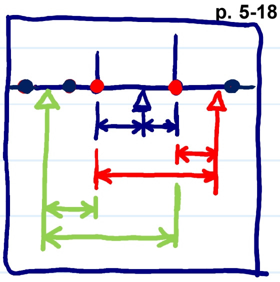
➢ Intuitive Interpretation of Variance → $E[(X-\mu_X)^2]$

- Variance is the weighted average value of the squared deviation of X from μ_X. 權重 $f_X(x)$
- Variance is related to how the pmf is spread out



➢ Some properties of variance. $E[(X-\mu_X)^2] = Var(X) \geq 0$

- The variance of a r.v. is always non-negative
- The only r.v. with variance equal to zero is a r.v. which can only take on a single value (μ_X).



It's actually not random

prove for discrete r.v.'s (but it's also true for other types of r.v.'s)

$\sum_{x \in \mathcal{X}} (x-\mu_X)^2 f_X(x) = 0 \iff P_X(X=\mu_X) = 1.$

$\iff (x-\mu_X)^2 f_X(x) = 0, \forall x \in \mathcal{X}$

$\iff (x-\mu_X)^2 = 0$, for x s.t. $f_X(x) > 0$

pmf 1 degenerate
$c = \mu_X$

➢ Theorem. For $a, b \in \mathbb{R}$, $Var(aX+b) = a^2\, Var(X)$ (gone) $\Rightarrow$ $\boxed{\sigma_{aX+b} = |a|\sigma_X}$

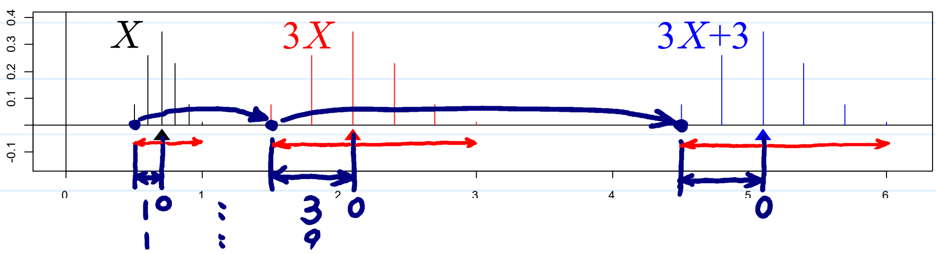
fixed constants

cf. Thm in LNp.5-16

proof. Let $Y = aX + b$, then $E(Y) = a \cdot \mu_X + b \equiv \mu_Y$.

$$
\begin{aligned}
Var(Y) &= E(Y-\mu_Y)^2 = E[(aX+\cancel{b}) - (a\mu_X + \cancel{b})]^2 \\
&= E[a^2(X-\mu_X)^2] = a^2 E(X-\mu_X)^2 = a^2 Var(X)
\end{aligned}
$$

$a > 1$
$0 < a < 1$
$a < 0$

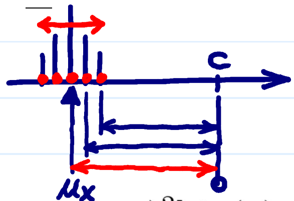


- **Theorem.** If X is a (discrete) r.v. with mean μ_X, then for any $c \in \mathbb{R}$, fixed

mean square error = Var + bias²

$$E[(X-c)^2] = \sigma_X^2 + (c-\mu_X)^2. \quad (*)$$

deviation of X from C



proof.

$$E[(X-c)^2] = E[(\underbrace{X}_{\text{random}} - \underbrace{\mu_X + \mu_X - c}_{\text{fixed}})^2] = \sum_{x\in\mathcal{X}} [(x-\mu_X+\mu_X-c)^2] f_X(x)$$

$$= \sum_{x\in\mathcal{X}} [(x-\mu_X)^2 + 2(x-\mu_X)(\mu_X-c) + (\mu_X-c)^2] f_X(x)$$

$$= \underbrace{\sum_{x\in\mathcal{X}} (x-\mu_X)^2 f_X(x)}_{\sigma_X^2} + 2(\mu_X-c)\underbrace{\sum_{x\in\mathcal{X}} (x-\mu_X) f_X(x)}_{0 \text{ (LNp.5-14)}} + (\mu_X-c)^2 \underbrace{\sum_{x\in\mathcal{X}} f_X(x)}_{1}$$

a 2nd-order polynomial of c → $g(c)$

- **Corollary.** $E[(X-c)^2]$ is minimized by letting $c=\mu_X$; and the minimum value is σ_X^2. ← (proof. It's clear by (*))

1. alternative definition of mean
2. useful for prediction

- **Corollary.** $\sigma_X^2 = E(X^2) - (E(X))^2$. ← (proof. Let the c in (*) be zero)

useful for calculating σ_X^2

(Recall: $E(X^2) = \sum_{x\in\mathcal{X}} x^2 f_X(x)$.)

  - Example (Committees, LNp.5-17). $Var(X)=14/3-2^2=2/3$.

- $E(X^n)$ is often called the n^{th} *moment* of X → Variance = (2nd moment) − (1st moment)²

a fixed constant; 動差 → Recall. mgf in LNp.5-5; $\boxed{E(X^2)}$; $E(X)$

❖ **Reading**: textbook, Sec 4.3, 4.4, 4.5

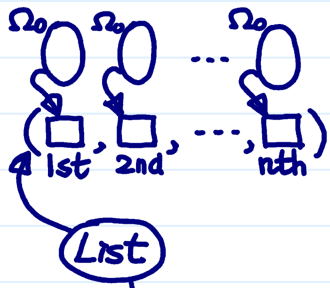
Some Commonly Used Discrete Distributions

- Bernoulli and binomial Distributions

伯努利; 二項

$\mathcal{F}_0 = \{E : E \subset \Omega_0\}$; cf.

- Experiment: A basic experiment with sample space Ω_0 (and p.m. P_0) is repeated n times.



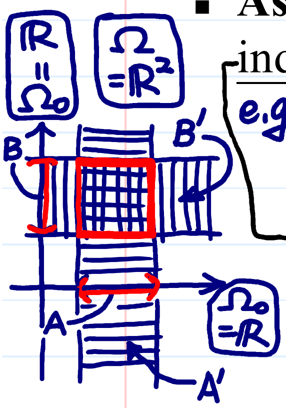
In general, P_0 is not enough to uniquely determine P.

  - Example. (a) Sampling with replacement
    (b) Coin Tossing
    (c) Roulette
  - The sample space for the n trials is

$$P \rightarrow \Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^n \qquad \Omega = \{\omega = (\omega_1, \cdots, \omega_n) \mid \omega_i \in \Omega_0\}$$

like a vector space

  - **Assume** that events depending on different trials are independent ($\subset \Omega$)



e.g. ith trial, $A \subset \Omega_0 \Rightarrow A' = \Omega_0 \times \cdots \times \Omega_0 \times A \times \Omega_0 \times \cdots \times \Omega_0 \subset \Omega$ (ith trial)

jth trial, $B \subset \Omega_0 \Rightarrow B' = \Omega_0 \times \cdots \times \Omega_0 \times B \times \Omega_0 \times \cdots \times \Omega_0 \subset \Omega$ (jth trial)

$(i \neq j)$ $\quad P(A' \cap B') = P(A')P(B') = P_0(A)P_0(B)$ ← P_0 defines P

e.g. $1 \le i_1 < i_2 < \cdots < i_k \le n$

A′ : formed by i_1, i_3, i_5 trials
B′ : formed by i_2, i_6 trials
⋮
mutually independent.