Probability Review

- Events and Event spaces
- Random variables
- Joint probability distributions
  - Marginalization, conditioning, chain rule, Bayes Rule, law of total probability, etc.
- Structural properties
  - Independence, conditional independence
- Mean and Variance
- Examples

Sample space and Events

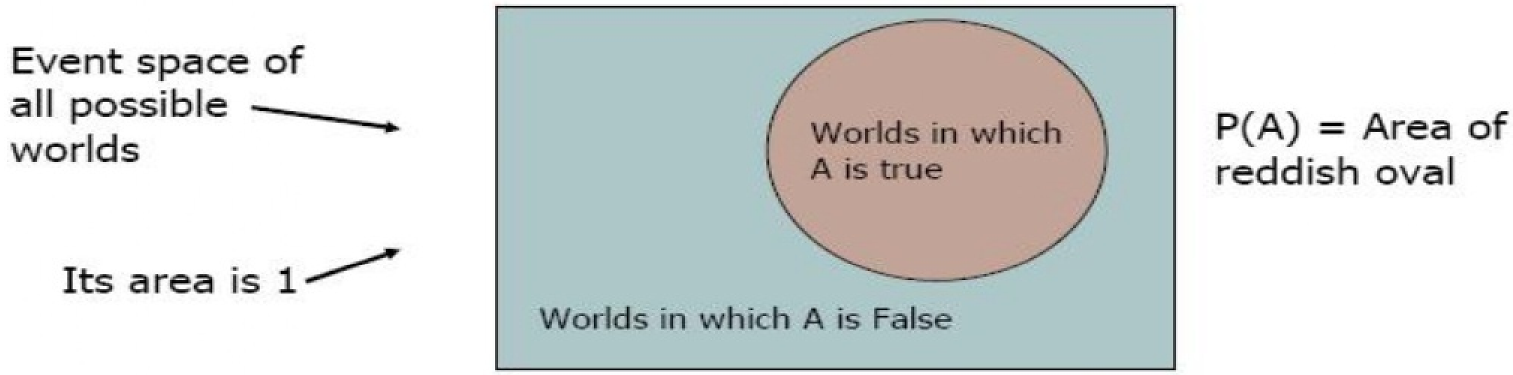
- : Sample Space, result of an experiment
  - If you toss a coin twice $\Omega = \{HH, HT, TH, TT\}$
- Event: a subset of Ω
  - First toss is head = {HH,HT}
- S: event space, a set of events:
  - Closed under finite union and complements
    - Entails other binary operation: union, diff, etc.
  - Contains the empty event and Ω

Probability Measure

- S :Defined over (E,S) so that
  - $P(\alpha) \geq 0$ for all α in S
  - $P(E) = 1$
  - If α, β are disjoint, then
    - $P(\alpha \cup \beta) = p(\alpha) + p(\beta)$
- We can deduce other axioms from the above ones
  - Ex: $P(\alpha \cup \beta)$ for non-disjoint event

  $P(\alpha \cup \beta) = p(\alpha) + p(\beta) - p(\alpha \cap \beta)$

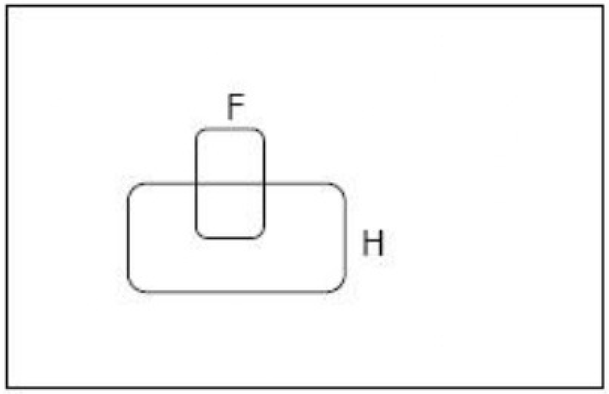
Visualization



- We can go on and define conditional probability, using the above visualization

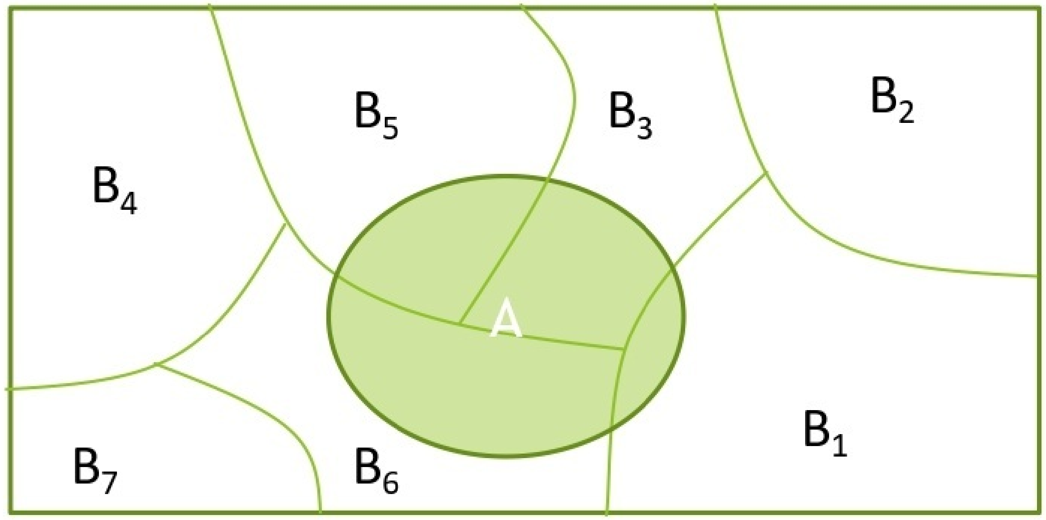
Conditional Probability

P(F|H) = Fraction of worlds in which H is true that also have F true



$$p(f \mid h) = \frac{p(F \cap H)}{p(H)}$$

Rule of total probability

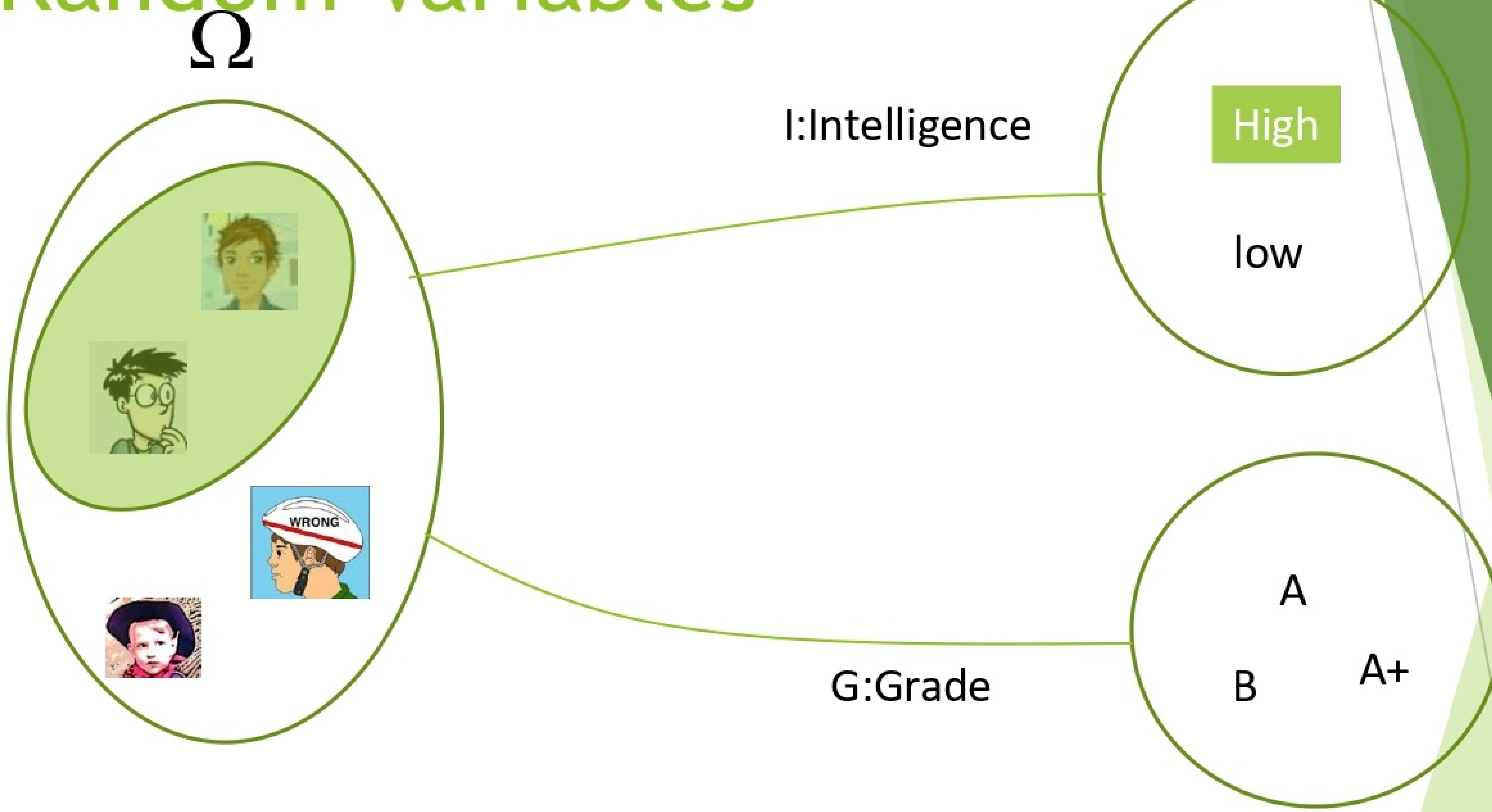


$$p(A) = \sum P(B_i)P(A \mid B_i)$$

From Events to Random Variable

- Almost all the semester we will be dealing with RV
- Concise way of specifying attributes of outcomes
- Modeling students (Grade and Intelligence):
  - $\Omega =$ all possible students
  - What are events
    - Grade_A = all students with grade A
    - Grade_B = all students with grade B
    - Intelligence_High = … with high intelligence
  - Very cumbersome
  - We need “functions” that maps from Ω to an attribute space.

Random Variables



P(I = high) = P({all students whose intelligence is high})

Discrete Random Variables

- Random variables (RVs) which may take on only a **countable** number of **distinct** values
  - E.g. the total number of tails X you get if you flip 100 coins
- X is a RV with arity *k* if it can take on exactly one value out of $\{x_1, \ldots, x_k\}$
  - E.g. the possible values that X can take on are 0, 1, 2, …, 100

Probability of Discrete RV

- Probability mass function (pmf): $P(X = x_i)$
- Easy facts about pmf
  - $\Sigma_i P(X = x_i) = 1$
  - $P(X = x_i \cap X = x_j) = 0$ if $i \neq j$
  - $P(X = x_i \cup X = x_j) = P(X = x_i) + P(X = x_j)$ if $i \neq j$
  - $P(X = x_1 \cup X = x_2 \cup \ldots \cup X = x_k) = 1$

Common Distributions

- Uniform X*U*[1, …, *N*]
  - X takes values 1, 2, … *N*
  - P(X = *i*) = 1/*N*
  - E.g. picking balls of different colors from a box
- Binomial X *Bin*(*n*, *p*)
  - X takes values 0, 1, …, *n*
  - $$p(X=i)=\binom{n}{i}p^i(1-p)^{n-i}$$
  - E.g. coin flips

Continuous Random Variables

- Probability density function (pdf) instead of probability mass function (pmf)
- A pdf is any function $f(x)$ that describes the probability density in terms of the input variable x.

Probability of Continuous RV

- Properties of pdf
  - $$f(x) \geq 0, \forall x$$
  - $$\int_{-\infty}^{+\infty} f(x) = 1$$
- Actual probability can be obtained by taking the integral of pdf
  - E.g. the probability of X being between 0 and 1 is

Cumulative Distribution Function

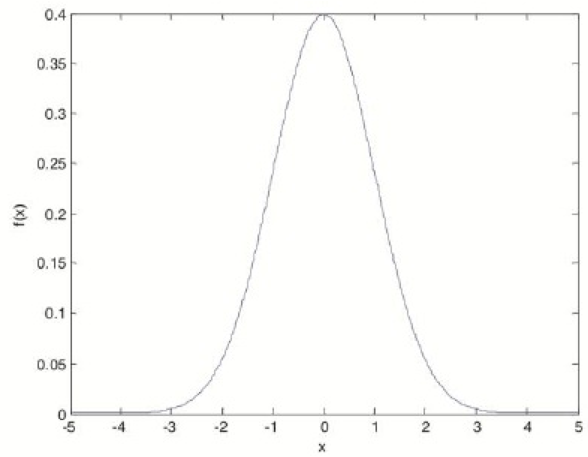
- $F_X(v) = P(X \leq v)$
- Discrete RVs
  - $F_X(v) = \Sigma_{vi} P(X = v_i)$
- Continuous RVs
  - $$F_X(v) = \int_{-\infty}^{v} f(x)dx$$
  - $$\frac{d}{dx} F_x(x) = f(x)$$

Common Distributions

- Normal X $N(\mu, \sigma^2)$

  - $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

  - E.g. the height of the entire population



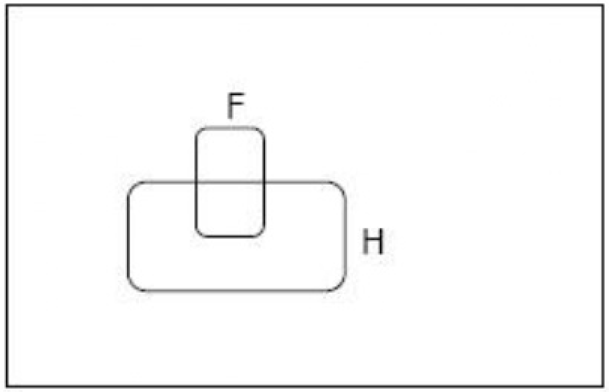
Conditional Probability

events

$$P(X = x \mid Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

But we will always write it this way:

$$P(x \mid y) = \frac{p(x, y)}{p(y)}$$



Bayes Rule

- We know that P(rain) = 0.5
  - If we also know that the grass is wet, then how this affects our belief about whether it rains or not?

$$P(rain \mid wet) = \frac{P(rain)P(wet \mid rain)}{P(wet)}$$

$$P(x \mid y) = \frac{P(x)P(y \mid x)}{P(y)}$$

- You can condition on more variables

$$P(x \mid y,z) = \frac{P(x \mid z)P(y \mid x,z)}{P(y \mid z)}$$

Independence

- X is independent of Y means that knowing Y does not change our belief about X.
  - $P(X|Y=y) = P(X)$
  - $P(X=x, Y=y) = P(X=x)\, P(Y=y)$
  - The above should hold for all x, y
  - It is symmetric and written as $X \perp Y$

Independence

- $X_1, \ldots, X_n$ are independent if and only if

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P\left(X_i \in A_i\right)$$

- If $X_1, \ldots, X_n$ are independent and identically distributed we say they are *iid* (or that they are a random sample) and we write

$$X_1, \ldots, X_n \sim \; P$$

Mean and Variance

$$\mu = E(\mathrm{X})$$

- Mean (Expectation): $E(\mathrm{X}) = \sum_{v_i} v_i \mathrm{P}(\mathrm{X} = v_i)$
  - Discrete RVs:

$$E(g(X)) = \sum_{v_i} g(v_i) P(X = v_i)$$

  - Continuous RVs:

$$E(\mathrm{X}) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Mean and Variance

- Variance:

$$Var(X) = E((X-\mu)^2)$$

$$Var(X) = E(X^2) - \mu^2$$

  - Discrete RVs:

$$V(\mathrm{X}) = \sum_{v_i} (v_i - \mu)^2 \, \mathrm{P}(\mathrm{X} = v_i)$$

  - Continuous RVs:

$$V(\mathrm{X}) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

- Covariance:

- Covariance:

$$Cov(X,Y) = E((X-\mu_x)(Y-\mu_y)) = E(XY) - \mu_x\mu_y$$

Mean and Variance

- Correlation:

$$\rho(X,Y) = Cov(X,Y)/\sigma_x\sigma_y$$

$$-1 \le \rho(X,Y) \le 1$$

Properties

- Mean
  - $E(X+Y) = E(X) + E(Y)$
  - $E(aX) = aE(X)$
  - If X and Y are independent, $E(XY) = E(X) \cdot E(Y)$
- Variance
  - $V(aX+b) = a^2 V(X)$
  - If X and Y are independent,

$$V(X+Y) = V(X) + V(Y)$$

Properties

- The conditional expectation of Y given X when the value of X = x is:

$$E(Y \mid X = x) = \int y * p(y \mid x)dy$$

- The Law of Total Expectation or Law of Iterated Expectation:

$$E(Y) = E[E(Y \mid X)] = \int E(Y \mid X = x)\, p_X(x)dx$$

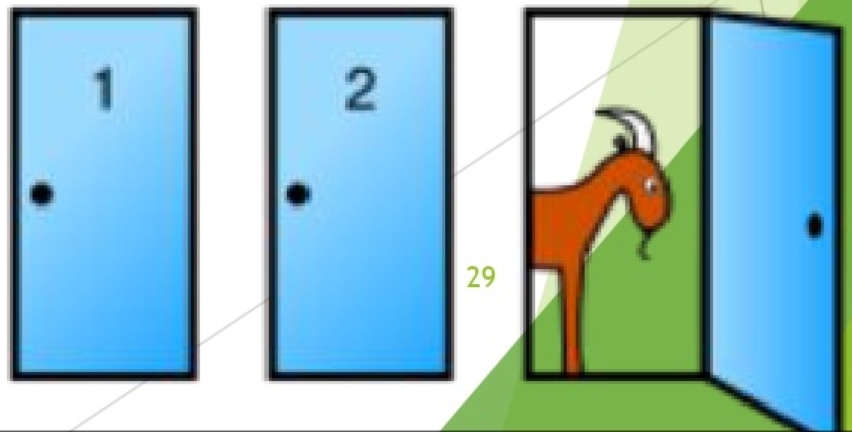
- The law of Total Variance:

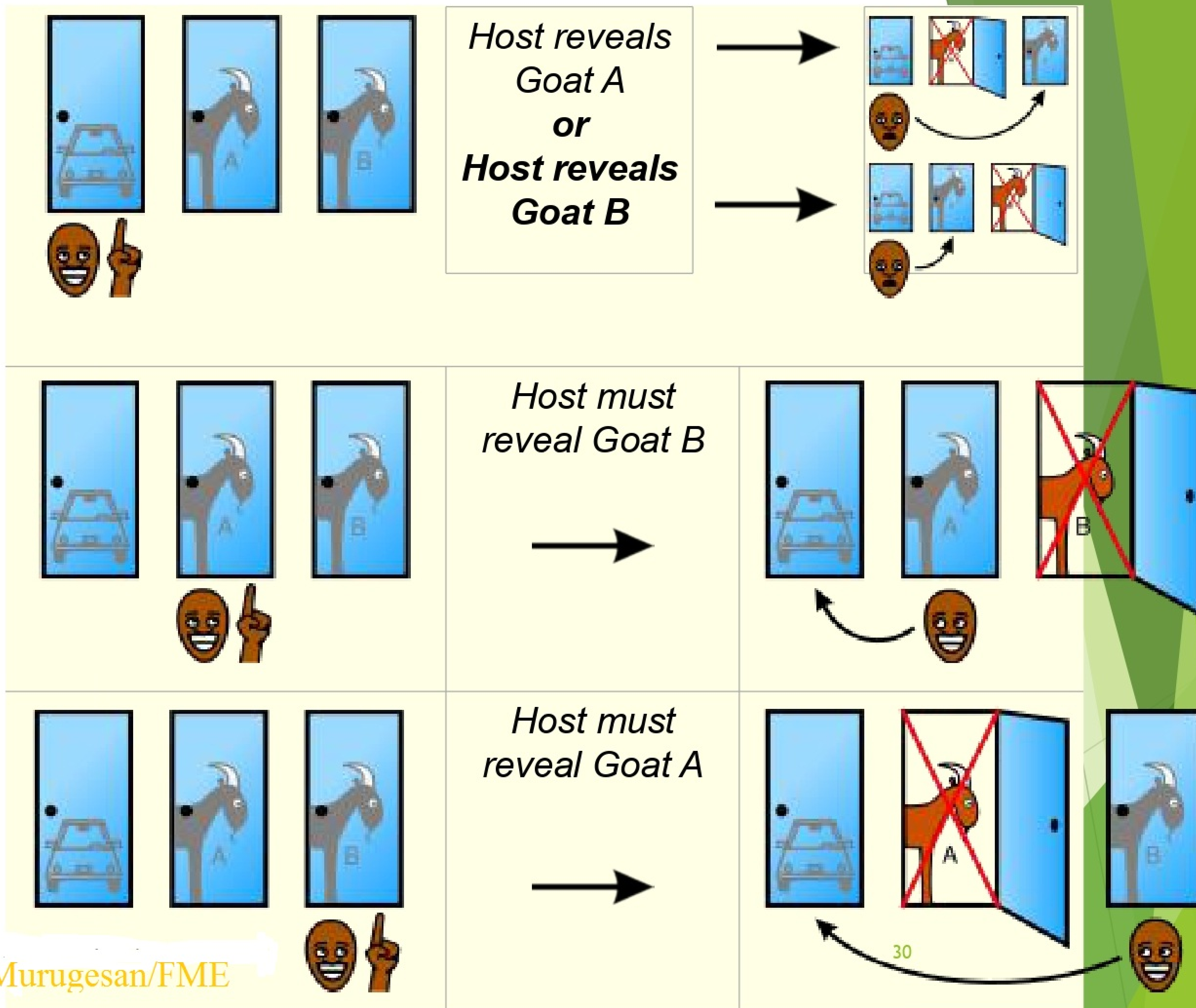
$$Var(Y) = Var\left[E(Y \mid X)\right] + E\left[Var(Y \mid X)\right]$$

Examples

Monty Hall Problem

- You're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- Do you want to pick door No. 2 instead?




Host reveals Goat A
or
Host reveals Goat B
Host must reveal Goat B
Host must reveal Goat A
A
B

Monty Hall Problem: Bayes Rule

- C_i : the car is behind door *i*, *i* = 1, 2, 3
- $P(C_i) = 1/3$
- H_{ij} : the host opens door *j* after you pick door *i*
- $$P\left(H_{ij} \middle| C_k\right) = \begin{cases} 0 & i = j \\ 0 & j = k \\ 1/2 & i = k \\ 1 & i \neq k, j \neq k \end{cases}$$

Monty Hall Problem: Bayes Rule cont.

- WLOG, *i*=1, *j*=3
- $$P\left(C_1 \middle| H_{13}\right) = \frac{P\left(H_{13} \middle| C_1\right) P\left(C_1\right)}{P\left(H_{13}\right)}$$
- $$P\left(H_{13} \middle| C_1\right) P\left(C_1\right) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

Monty Hall Problem: Bayes Rule cont.

υ

$$P(H_{13}) = P(H_{13}, C_1) + P(H_{13}, C_2) + P(H_{13}, C_3)$$
$$= P(H_{13} | C_1) P(C_1) + P(H_{13} | C_2) P(C_2)$$
$$= \frac{1}{6} + 1 \cdot \frac{1}{3}$$
$$= \frac{1}{2}$$

υ

$$P(C_1 | H_{13}) = \frac{1/6}{1/2} = \frac{1}{3}$$

Monty Hall Problem: Bayes Rule cont.

- $$P\left(C_1 \middle| H_{13}\right) = \frac{1/6}{1/2} = \frac{1}{3}$$
- $$P\left(C_2 \middle| H_{13}\right) = 1 - \frac{1}{3} = \frac{2}{3} > P\left(C_1 \middle| H_{13}\right)$$
- *You should switch!*