Probability model $(\Omega, \mathcal{F}, Pr)$

$\mathcal{F}$ ← a collection of subsets of Ω
-sigma field-

Union Bound

$$Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} Pr(A_i)$$

} equality only holds if mutually exclusive
or
$A_i \cap A_j = \emptyset \ \forall i,j$

Conditional Probability

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

B ← new sample space

$Pr(B)$ ← normalization factor

Total Probability Theorem

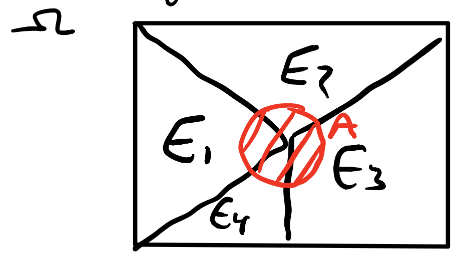
If $\{E_1, E_2, \ldots, E_k\}$ partition Ω
then

$$Pr(A) = \sum_{i=1}^{k} Pr(A \cap E_i)$$

$$= \underbrace{\sum_{i=1}^{k} Pr(A|E_i) Pr(E_i)}_{\text{generative models}}$$

This makes more sense since you know base event which occurred

<u>Example of Partition for k=4</u>



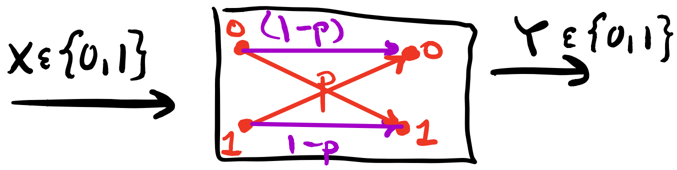
No intersections
$\bigcup_{i=1}^{k} E_i = \Omega$

Bayes Rule

Want to find "ground truth" that gave this observation.

$$\Pr[E_j \mid A] = \frac{\Pr[A \cap E_j]}{\Pr[A]} = \frac{\overbrace{\Pr[E_j]}^{\text{prior knowledge}}\ \overbrace{\Pr(A \mid E_j)}^{\text{generative model}}}{\underbrace{\sum_{i=1}^{k} \Pr[E_i]\Pr[A \mid E_i]}_{\Pr(A) \text{ by TPT}}}$$

Example: Binary Symmetric Channel



$\Pr(X=0) = \Pr(X=1) = \frac{1}{2}$

$\Pr(Y=0 \mid X=0] = 1-p$

$\Pr(Y=1 \mid X=0] = p$

If $Y=0$, what is $\hat{X}_{opt}$?

$\hat{X} = 0$ — wrong if $\hat{X}=1 \to Y=0$

$\Pr(\text{error}) = \Pr(X=1 \mid Y=0)$

or

$\hat{X} = 1$ — wrong if $\hat{X}=0 \to Y=1$

$\Pr(\text{error}) = \Pr(X=0 \mid Y=1)$

Want the answer which is LESS likely to be wrong?

$$\Pr(\text{error}) = \frac{\Pr(X=1)\Pr(Y=0 \mid X=1)}{\Pr(X=1)\Pr(Y=0 \mid X=1) + \Pr(X=0)\Pr(Y=1 \mid X=0)} = \frac{\frac{1}{2}p}{\frac{1}{2}p + \frac{1}{2}(1-p)}$$

Independence

Two events, A_1, A_2, are independent if $Pr(A_1 \cap A_2) = Pr(A_1)Pr(A_2)$

$$\Longrightarrow Pr[A_1 | A_2] = \frac{Pr[A_1 \cap A_2]}{Pr[A_2]} = \frac{Pr(A_1)Pr(A_2)}{Pr(A_2)}$$

Can be extended:

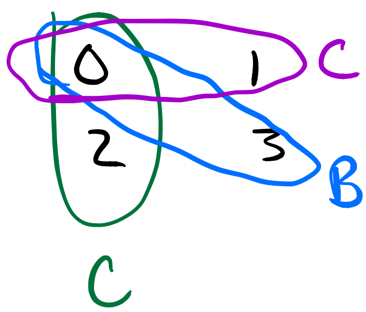
Events $\{A_1, A_2, \dots, A_n\}$ are independent if

$$Pr(A_1 \cap A_2 \cap \dots \cap A_k) = Pr(A_1)Pr(A_2)\cdots Pr(A_k) \quad \forall^{k \in [2,n]}_{k \in \mathbb{Z}}$$

Note: Pairwise independence $\not\Rightarrow$ independence!

Example: $\Omega = \{0,1,2,3\}$; all occur w/ equal probability

$A = \{0,1\}$ $\quad B = \{0,2\}$ $\quad C = \{0,3\}$



$Pr(AB) = Pr(\{0\}) = 1/4$

$Pr(BC) = Pr(\{0\}) = 1/4$

$Pr(AC) = Pr(\{0\}) = 1/4$

$Pr(ABC) = Pr(\{0\}) = 1/4$

BUT $Pr(A)Pr(B)Pr(C) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$!

Random Variables

Given $(\Omega, \mathcal{F}, Pr)$ a random variable is a <u>function</u>
$X: \Omega \to \mathbb{R}$ such that $\forall x \in \mathbb{R}$, $\{\omega \mid X(\omega) \leq x\} \in \mathcal{F} \subseteq \Omega$

Example: Throw two dice: d_1, d_2

$\Omega = \{(1,1), \ldots, (6,6)\}$ w/ each event occuring w/ Pr of $1/36$

Define $X = \max(d_1, d_2) - \min(d_1, d_2)$

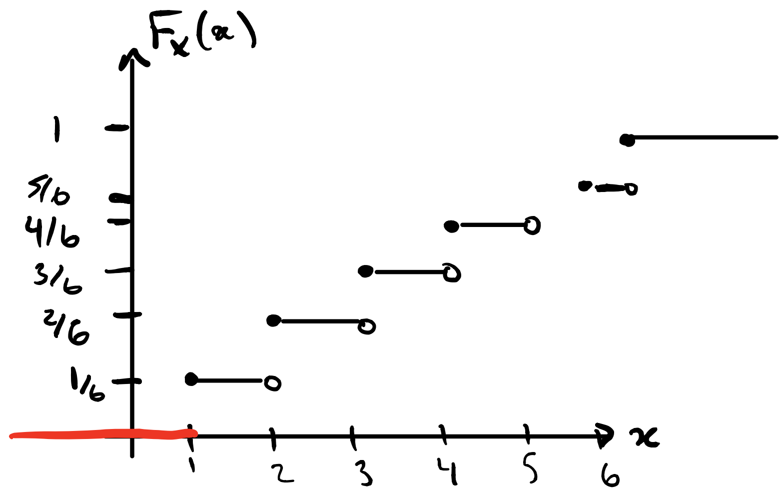
$$Pr(X=3) = Pr(\{\omega \mid X(\omega)=3\})$$
$$= Pr(\{(1,4), (2,5), (3,6), (4,1), (5,2), (6,3)\})$$
$$= \frac{1}{6} \leftarrow \frac{6 \text{ outcomes}}{36 \text{ total outcomes}}$$

Cumulative Distribution Function

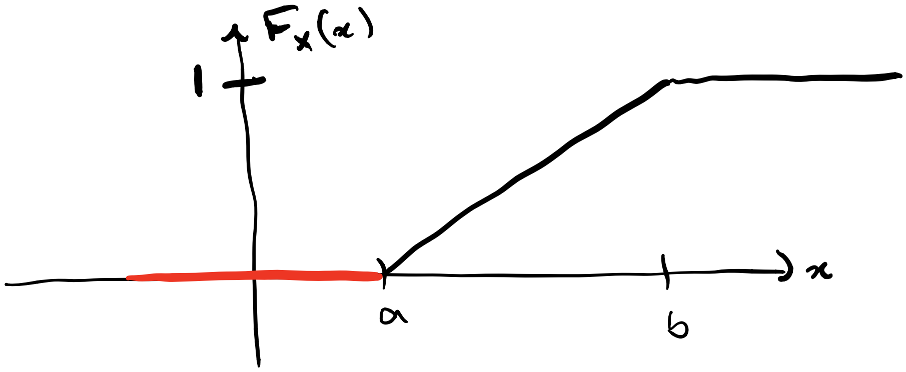
The CDF is defined as

$$F_X(x) = Pr(X \leq x) \quad \forall x \in \mathbb{R}$$
$$= Pr(\{\omega \mid X(\omega) \leq x\})$$

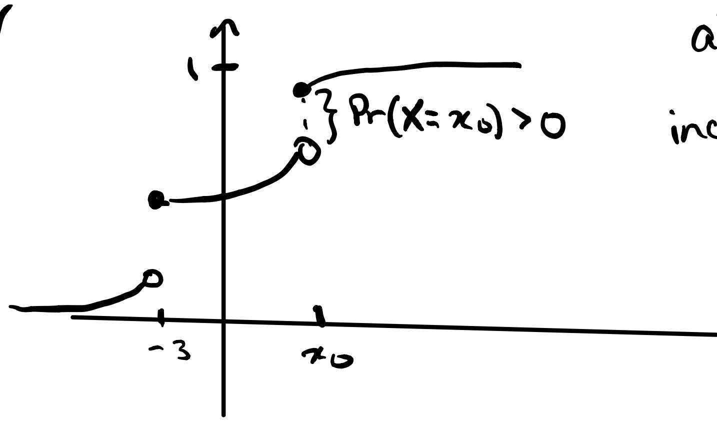
Throw a die



Uniform



Mixed RV



always monotonically increasing

Probability Mass Function

For a discrete r.v.

$$P_X(x) = Pr(X = x)$$

i.e Die