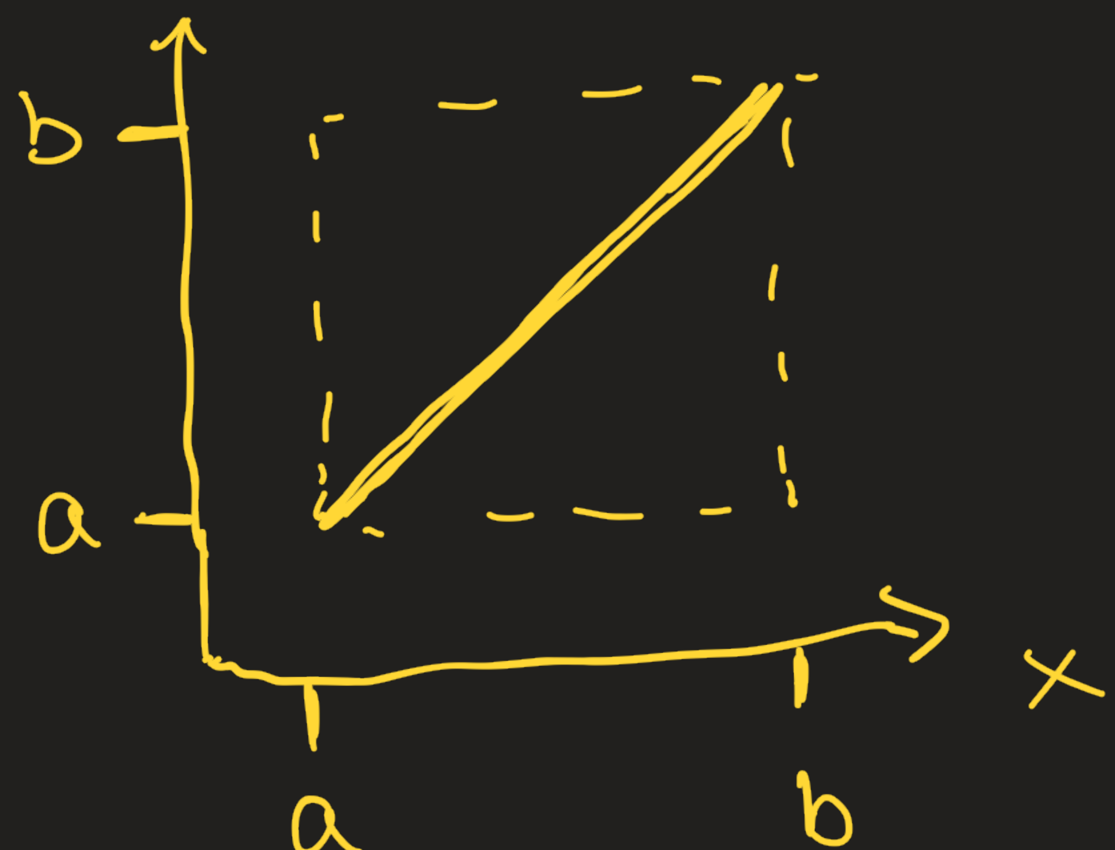
Lecture 4

Topics

- Independence & Conditioning
- Law of Large Numbers
- Bayes' rule & Law of Total Prob & Law of Total Expectation
- Ex: analyzing the performance of a rand alg.
- Properties of Gaussians
- Empirical Risk Minimization:
  how we use training data to fit ML models that work well

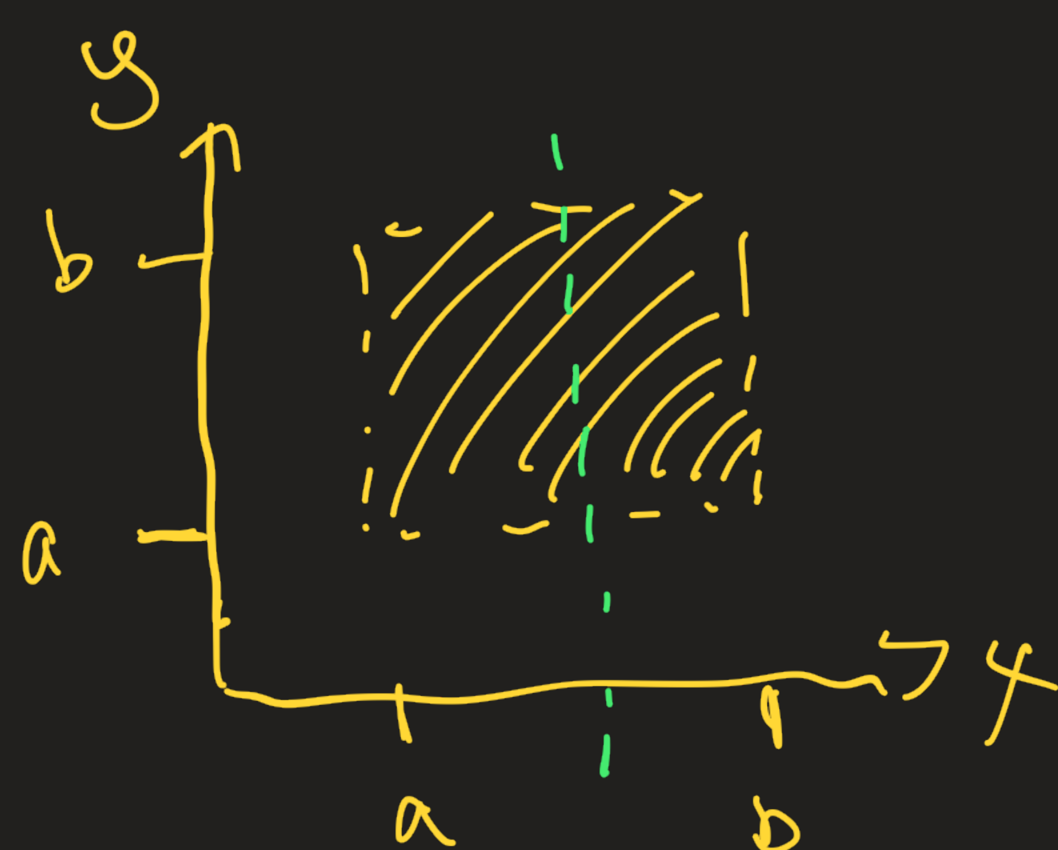
Independence & Conditioning

The extent to which X can tell us about Y depends on the joint dist $p(x,y)$



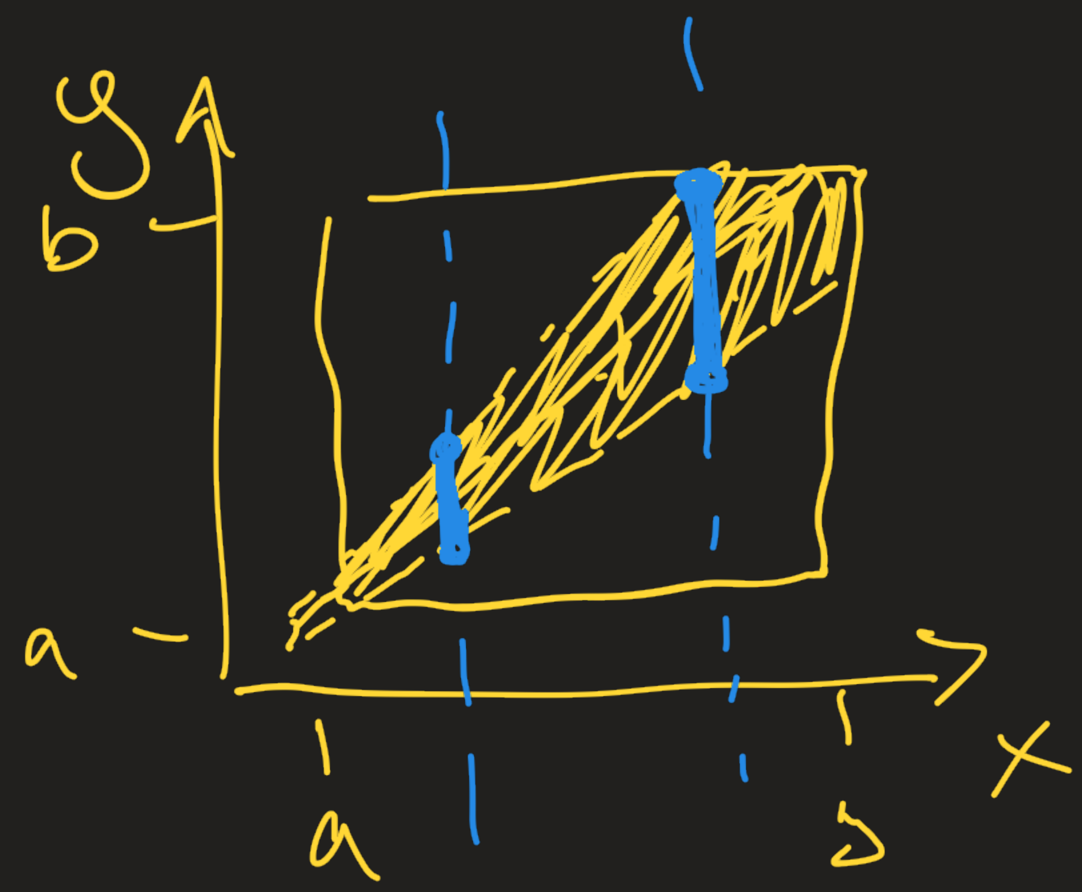
$$p(x,y) = \begin{cases} \frac{1}{b-a} & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

$\Rightarrow$ X completely determines Y



$$p(x,y) = \frac{1}{(b-a)^2} \quad \text{for } (x,y) \in [a,b]^2$$

$\Rightarrow$ X tells me nothing about Y



intermediate case:

knowing x tells me something about the value of y

In the completely independent case,

$$p(x,y) = \frac{1}{(b-a)^2} = \frac{1}{b-a} \cdot \frac{1}{b-a}$$

$$= p_x(x) \cdot p_y(y)$$

We say two random variables X & Y are indep iff

$$p(x,y) = p_x(x)\, p_y(y)$$

The joint pdf (pmf) is the product of the marginals. We write $X \perp\!\!\!\perp Y$

Note that if $X \perp\!\!\!\perp Y$ then

$$\mathbb{E} XY = (\mathbb{E}X)(\mathbb{E}Y)$$

Prf

$$\mathbb{E}(XY) = \int xy\, p(x,y)\, dx\, dy$$

$$= \int xy\, p_x(x)\, p_y(y)\, dx\, dy$$

$$= \left[\int x\, p_x(x)\, dx\right]\left[\int y\, p_y(y)\, dy\right]$$

$$= (\mathbb{E}X)(\mathbb{E}Y)$$

Very important: if f & g are functions then $f(X) \perp\!\!\!\perp g(Y)$

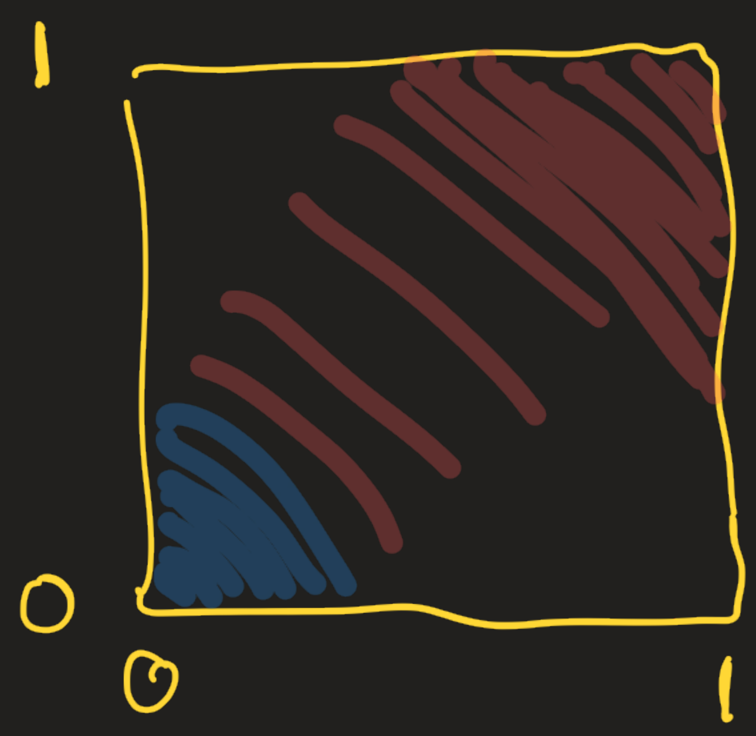
$$\Rightarrow \mathbb{E}\left[f(X)g(Y)\right] = \left[\mathbb{E}f(X)\right]\left[\mathbb{E}g(Y)\right]$$

In general we can quantify exactly what X tells us about Y using the conditional pdf (pmf)

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} \quad (\text{assume } p_X(x) \neq 0)$$

relative prob. of Y taking on the value y if we know that $X = x$

Ex $p(x,y) = \frac{6}{7}(x+y)^2$ on $[0,1]^2$



← more likely to sample (x,y) when x and y are both large

Compute $p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}$

Compute $p_X(x) = \int_0^1 p(x,y)\,dy = \frac{2}{7}(3x^2+3x+1)$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{3(x+y)^2}{3x^2+3x+1}$$

Ex: $x = \frac{1}{2}$

$$p_{Y|X}\left(y \mid x = \tfrac{1}{2}\right) = \frac{12}{13}\left(y + \frac{1}{2}\right)^2$$

Note that if $X \perp\!\!\!\perp Y$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{p_X(x)\,p_Y(y)}{p_X(x)} = p_Y(y)$$

$$\Rightarrow \boxed{p_{Y|X}(y|x) = p_Y(y)}$$

We can define conditional pdfs for mult. r.v.s. e.g.

$$p(z|x,y) = \frac{p(x,y,z)}{p_{x,y}(x,y)}$$

Have chain rules for conditional pdfs

$$p(y|x) = \frac{p(x,y)}{p(x)} \Rightarrow p(x,y) = p(y|x)\,p(x)$$

$$p(x,y,z) = p(z|x,y)\,p(x,y)$$

$$= p(z|x,y)\,p(y|x)\,p(x)$$

Independence, Variance, Law of Large Numbers

If $X \sim \mathbb{P}$ and we have $X_i \sim \mathbb{P}$ where X_i are independent the we have

- means: $\mathbb{E} X_i = \mathbb{E} X$
- variances: $\mathrm{Var}(X_i) = \mathrm{Var}(X)$

$X \perp\!\!\!\perp Y$: $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Prf

$$\mathrm{Var}(X+Y) = \mathbb{E}\left(X+Y - \mathbb{E}(X+Y)\right)^2$$

$$\mathrm{Var}(X+Y) = \mathbb{E}\left(X+Y-\mathbb{E}(X+Y)\right)^2$$

$$= \mathbb{E}\left((X-\mathbb{E}X) + \mathbb{E}(Y-\mathbb{E}Y)\right)^2$$

$$= \mathbb{E}\left[(X-\mathbb{E}X)^2 + 2(X-\mathbb{E}X)(Y-\mathbb{E}Y) + (Y-\mathbb{E}Y)^2\right]$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

Fact: if $X \perp\!\!\!\perp Y \Rightarrow \mathrm{Cov}(X,Y) = 0$

$$\Downarrow$$

$f(X) \perp\!\!\!\perp g(Y)$ where $f(X) = X - \mathbb{E}X$, $g(Y) = Y - \mathbb{E}Y$

$$\Downarrow$$

$$\mathbb{E}[f(X)g(Y)] = \mathrm{Cov}(X,Y) = [\mathbb{E}f(X)][\mathbb{E}g(Y)] = 0$$

$\Rightarrow \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

also $\underline{\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)}$ for any $c \in \mathbb{R}$

$$\begin{aligned}\mathrm{Var}(cX) &= \mathbb{E}(cX - \mathbb{E}(cX))^2 \\ &= c^2\,\mathbb{E}(X - \mathbb{E}X)^2 \\ &= c^2\,\mathrm{Var}(X)\end{aligned}$$

if X_i are independent and $X_i \sim \mathbb{P}$

look at $\underbrace{\mathrm{Var}\left(\frac{1}{n}\sum X_i\right)} = \frac{1}{n^2}\mathrm{Var}\left(\sum X_i\right)$

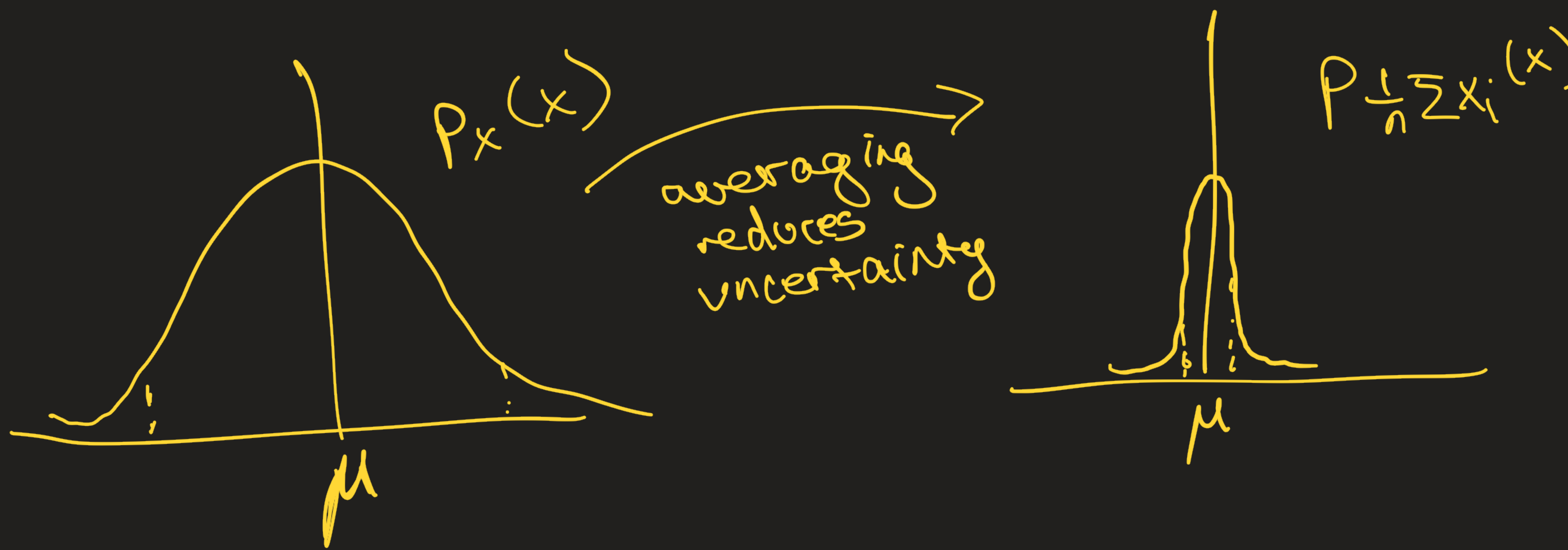
$$= \frac{1}{n^2}\underbrace{\sum \mathrm{Var}(X_i)}_{n\mathrm{Var}(X)} = \frac{1}{n}\mathrm{Var}(X)$$

Take-away: if X_i are independent & identically distributed ($X_i \sim \mathbb{P}$)

then

$$\mathrm{Var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\mathrm{Var}(X)$$

$$\mathbb{E}\,\frac{1}{n}\sum X_i = \underline{\mathbb{E}X}$$



The fact that $\mathrm{Var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\mathrm{Var}(X)$ has massive implications for ML. Makes ML possible: b/c implies the law of large numbers:

$$\frac{1}{n}\sum X_i \xrightarrow[\text{as } n\to\infty]{} \mathbb{E}X$$

"Flip a biased coin n times & average to get a good est of the bias"

LLN $\Rightarrow$ $$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \mathbb{E}f(X)$$ for any function f

Indep & Cond prob for events

A indep of B if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

similarly we define

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

Bayes' rule

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B)$$
$$= \mathbb{P}(B|A)\mathbb{P}(A)$$

$$\Rightarrow \boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}}$$

Bayes' rule

notice if A & B are indep

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

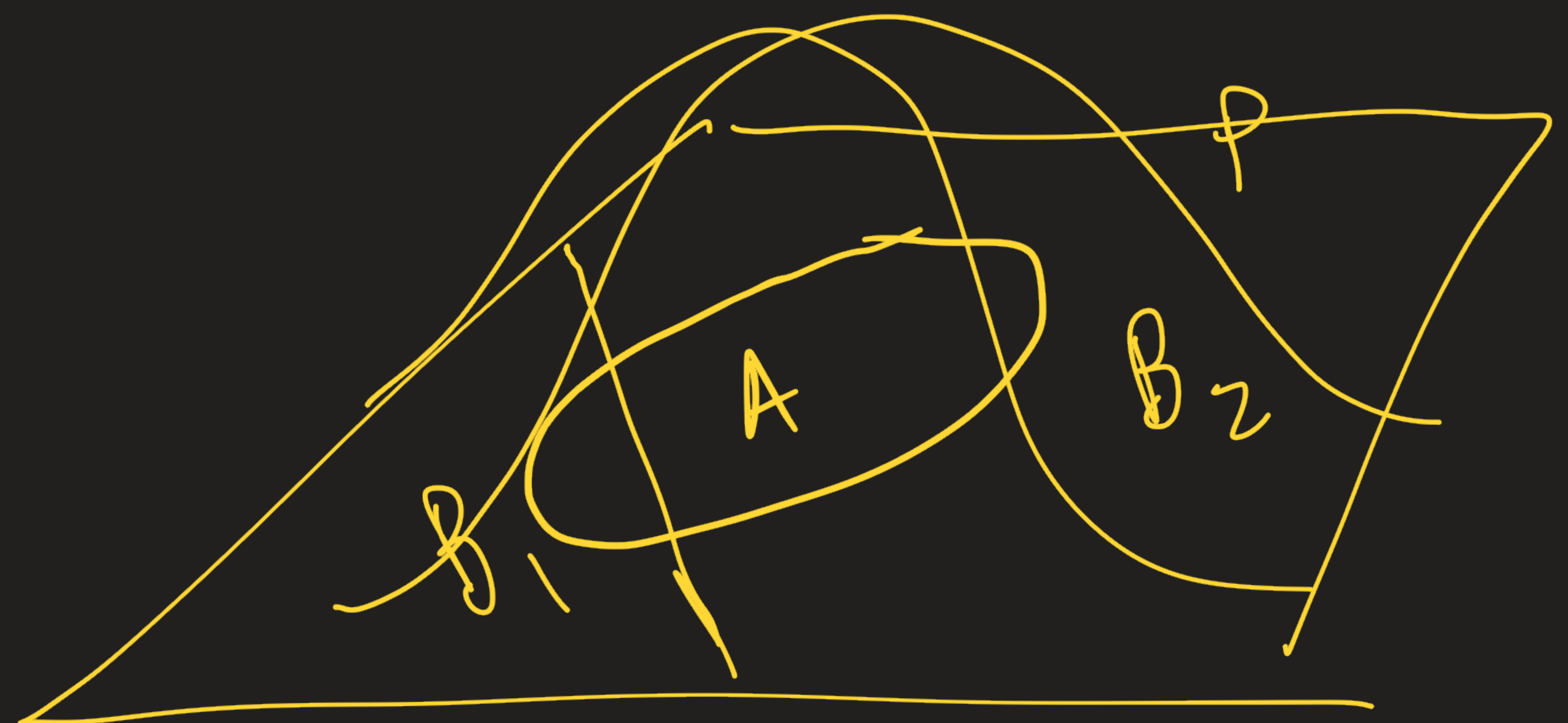
Law of Total Prob.

Let A be an event and $B_1, \dots, B_n$ be events that partition Ω

$$
\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) \\
&= \mathbb{P}(A \cap (B_1 \cup \dots \cup B_n)) \\
&= \mathbb{P}((A \cap B_1) \cup \dots \cup (A \cap B_n)) \\
&= \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)
\end{aligned}
$$

$$\boxed{\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}$$

Law of total prob.



$$\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\,\mathbb{P}(B_2)$$

Conditional Expectation

Define conditional expectation

$$\to \mathbb{E}\left[Y \mid X = x\right] = \int y\, p(y|x)\, dy$$

fact:

$$\underline{\mathbb{E}_X}\left[\mathbb{E}\left[Y \mid X\right]\right] = \mathbb{E}Y$$

$$
\begin{aligned}
\mathbb{E}Y &= \int y\, p_Y(y)\, dy \\
&= \int y \left[\int_x p(x, y)\, dx \right] dy \\
&= \int y \left[\int_x p(y|x)\, p(x)\, dx \right] dy \\
&= \int_x p(x) \underbrace{\left[\int_y y\, p(y|x)\, dy \right]}_{\mathbb{E}[Y|X=x]} dx
\end{aligned}
$$

$$\boxed{\mathbb{E}\, Y} = \int p(x)\, \mathbb{E}[Y \mid X = x]\, dx$$

$$\boxed{= \mathbb{E}_X\left[\mathbb{E}[Y \mid X = x]\right]}$$

called the "tower property"

Our first randomized alg!

- Given a biased coin w/ an unknown bias and we want to use it to simulate draws from a fair coin
- Application of conditioning & independence

Von Neumann (VA):

1. Sample X & Y i.i.d. from tossing the biased coin
2. if $X \neq Y$
   return X
   otherwise:
   return VA()

Claim $\mathbb{P}(VA() = H) = \frac{1}{2}$

$$
\begin{aligned}
h &= \mathbb{P}(VA() = H) \\
&= \mathbb{P}\big((X \neq Y \text{ and } X = H) \text{ or } (X = Y \text{ and } VA() = H)\big) \\
&= \mathbb{P}(X \neq Y \text{ and } X = H) + \mathbb{P}(X = Y \text{ and } VA() = H) \\
&= \mathbb{P}(X = H \text{ and } Y = T) + \mathbb{P}(X = Y)\,\mathbb{P}(VA() = H)
\end{aligned}
$$

$$\Rightarrow \quad h = \mathbb{P}(X=H)\,\mathbb{P}(Y=T)$$
$$+\ \mathbb{P}(X=Y)\,h$$

notice $\quad \mathbb{P}(X=Y) = \mathbb{P}(X=Y=H)$
$$+\ \mathbb{P}(X=Y=T)$$
$$= p^2 + (1-p)^2$$

$$\Rightarrow \quad h = p(1-p) + \{p^2 + (1-p)^2\}\,h$$

$\Rightarrow$ solve $\quad h = \dfrac{p-p^2}{1-\{p^2+(1-p)^2\}} = \dfrac{1}{2}$