9/5/13

① Probability Concepts

→ Sample Space (The space of events)

$\Omega = \{H, T\}$ (for a coin)
$= \{\text{spam}, \text{no-spam}\}$ (for email)
$= \{\text{car}, \text{boat}, \text{person}, \ldots\}$ for an image

→ Random Variable (Mapping from Sample Space to numbers)

Discrete $X : \Omega \rightarrow \{0, 1, \ldots, k\}$
$: \Omega \rightarrow \mathbb{R}$ (continuous)

→ Notation: X, Y random variables (capital) (e.g Y could be label we give to an email)

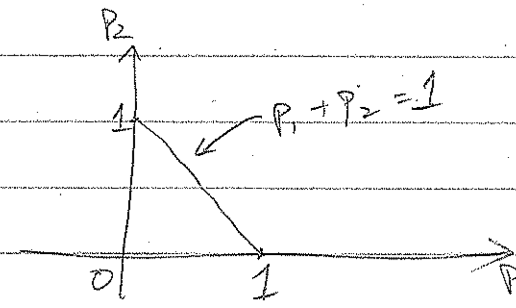
x, y: their states (e.g 0 or 1)

→ Probability Mass $\sum_{x=0} P(X = x) = 1$

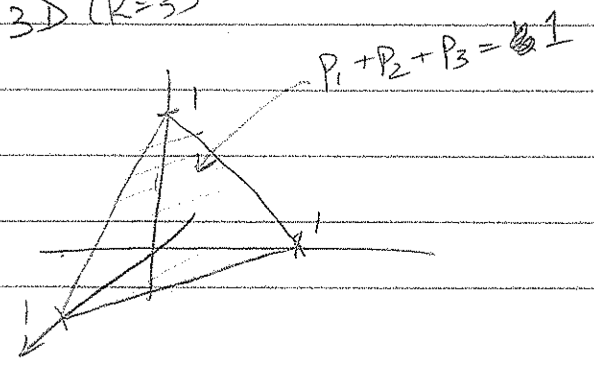
Sometimes it is useful to think of prob as a vector

$$\vec{P} = \begin{bmatrix} P_0 \\ \vdots \\ P_k \end{bmatrix} \qquad \left.\begin{array}{l} \vec{P} \in \mathbb{R}^k \\ \vec{P} \geq 0 \\ \sum p_i = 1 \end{array}\right\} \leftarrow \text{SIMPLEX}$$

2D $(k=2)$



3D $(k=3)$



For Continuous R.Vs Prob. Density Function

$$\int_x p(x)\,dx = 1$$

$p(x) \geq 0$

$p(x)$ can be ≥ 1

Same symbol p; Discrete or Continuous, clear from context.

→ Expectation of $f(X)$ ← some function of R.V. X

$$E_p[f(X)] = \sum_{x=0}^{k} f(x)p(x) \quad \text{(Discrete)}$$

Weighted

{Average value of $f(X)$}

$$= \int_x f(x)p(x)\,dx$$

Can think of $E[f(X)]$ as inner-product (or linear operation) for discrete RVs.

$$E[f(X)] = [f(0) \;---\; f(k)] \cdot \begin{bmatrix} p(0) \\ \vdots \\ p(k) \end{bmatrix} = \sum_{x=0}^{k} f(x)p(x)$$

→ Mean Value of X under p: set $f(X) = X$

$$\mu = E[X] = \sum_{x=0}^{k} x\,p(x)$$

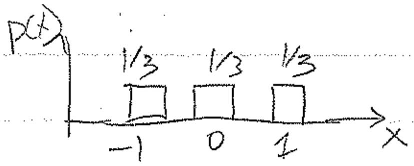
$= 3.5$ (for fair dice)

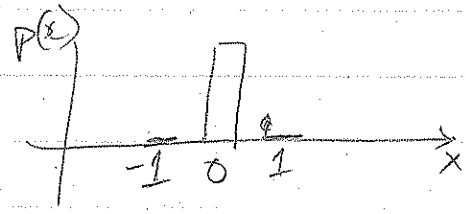
→ Variance: Estimate of "spread" around μ

$$f(X) = (X-\mu)^2$$

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

Example:





$\mu = 0$

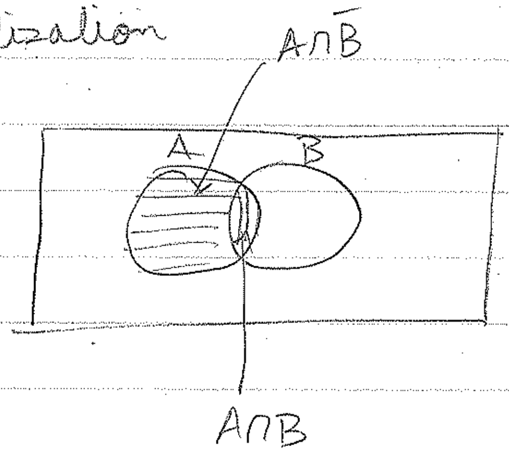
$$\mathrm{Var}(X) = E[X^2] = \frac{(-1)^2}{3} + \frac{0}{3} + \frac{1^2}{3} = \boxed{\frac{2}{3}}$$

$\mu = 0$

$$\mathrm{Var}(X) = 0 + 0 + 0 = \boxed{0}$$

No spread.

→ Marginalization



$$A = (A \cap B) \cup (A \cap \bar{B})$$

$$\Rightarrow p(A) = p(A\cap B) + p(A \cap \bar{B}) \underbrace{- p(\{A\cap B\} \cap \{A \cap \bar{B}\})}_{0}$$

$$\Rightarrow p(A) = p(A\cap B) + p(A\cap \bar{B})$$

→ Conditional Prob.

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

Chain Rule $P(X=x, Y=y) = P(Y=y \mid X=x)\,P(X=x)$

[Recursive Application]

$$P(X_1=x_1, X_2=x_2, \ldots, X_d=x_d, Y=y) = P(X_2=x_2, \ldots, X_d=x_d, Y=y \mid X_1=x_1) \cdot P(X_1=x_1)$$

$$= P(Y=y \mid X_1=x_1, \ldots, X_d=x_d) \cdot P(X_d=x_d \mid X_1=x_1, \ldots, X_{d-1}=x_{d-1}) \ldots P(X_2=x_2 \mid X_1=x_1)\, P(X_1=x_1)$$

→ Independence:

$$P(Y=y, X=x) = P(Y=y) \cdot P(X=x) \quad \boxed{\forall y, x}$$

Very Imp!

→ $\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_X)]$

where $\mu_X = E[X]$ $\mu_Y = E[Y]$

$$\text{Corr-Coeff}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

Bayes Rule

$$\underbrace{P(Y=y \mid X=x)}_{\text{Posterior}} = \frac{\overbrace{P(X=x \mid Y=y)}^{\text{Likelihood}}\ \overbrace{P(Y=y)}^{\text{Prior}}}{\underbrace{P(X=x)}_{\text{Evidence}}}$$