Useful Formulae and Definitions:

From here on, we let A and B be two random variables on two potentially different sets of outcomes.

- Joint Probability: The probability that a realization of the random variables A and B happen simultaneously. This is often denoted via $\mathbb{P}(A, B)$ or $\mathbb{P}(A \cap B)$.

  Ex. The probability that you draw a card from a deck that has both the number 7 and the color red. Here A describes the probability of getting a color, and B the probability of getting a particular value.

  $$\mathbb{P}(A = \text{red} \cap B = 7) = \frac{2}{52}$$

- The product rule of probability (not calculus): A useful way of calculating either joint or conditional probabilities.

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

  Ex. The same as above but calculated differently. First, $\mathbb{P}(B = 7) = \frac{4}{52}$ (there are 4 sevens). Next, suppose a seven is drawn, then the probability of it being red is $\mathbb{P}(A = \text{red}|B = 7) = \frac{1}{2}$ since half of the sevens are of that color. Putting this into the product rule:

  $$\mathbb{P}(A = \text{red} \cap B = 7) = \mathbb{P}(A = \text{red}|B = 7)\mathbb{P}(B = 7) = \frac{1}{2} * \frac{4}{52} = \frac{2}{52}$$

- The chain rule of probability (also not calculus): Let C be a third random variable on a potentially different set of outcomes from A and B. The chain rule allows us to extend the product rule to more random variables.

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

EFY: Derive the chain rule via the product rule.

- Marginals and the rule of total probability (sum rule): The marginal of A for the joint distribution $\mathbb{P}(A, B)$ is just $\mathbb{P}(A)$. How is this calculated if you know $\mathbb{P}(A, B)$? One way is with the rule of total probability.

  $$\mathbb{P}(A) = \sum_{b\in\Omega} \mathbb{P}(A, B = b) = \sum_{b\in\Omega} \mathbb{P}(A|B = b)\mathbb{P}(B = b)$$

  Ex. Using the cards example, find the marginal for $\mathbb{P}(B)$ and calculate $\mathbb{P}(B = 7)$. There are two values that A can take, red and black. Using the sum rule

  $$\mathbb{P}(B = 7) = \mathbb{P}(B = 7|A = \text{red})\mathbb{P}(A = \text{red}) + \mathbb{P}(B = 7|A = \text{black})\mathbb{P}(A = \text{black}) = \frac{1}{13}$$

- Bayes' Rule: For information and example please see slideshow.

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

- Naive Bayes' Assumption for calculating class conditionals: The features of A are conditionally independent given B. Suppose A is a joint distribution of say A_1, A_2, A_3, this assumption allows us to calculate $\mathbb{P}(A = [A_1, A_2, A_3]|B)$ as

  $$\mathbb{P}(A = [A_1, A_2, A_3]|B) = \mathbb{P}(A_1|B)\mathbb{P}(A_2|B)\mathbb{P}(A_3|B)$$

  More generally:

  $$\mathbb{P}(A_{[1:n]}|B) = \prod_{i=1}^{n} \mathbb{P}(A_i|B)$$

  - **This is often an unrealistic situation!** Can you come up with an example for which conditional independence fails?