- Joint distributions: usually consist of more than one RV, and usually it's nice to visualize what happens via a table.

  In essence, given two random variables X, Y and their joint distribution $P(X, Y)$, then the marginal distribution of X can be written as:

  $$P(X = k) = \sum_j P(X = k \cap Y = j)$$

  in other words, if we are looking for the distribution of X, then we sum over all the possible values of Y while holding the value of $X = k$ constant.

  In terms of the table formulation, this basically corresponds to going to the row where $X = k$, and summing over all the probabilities that exist down that row/column.

- Independence of joint distributions: defined basically where

  $$P(X = a \cap Y = b) = P(X = a)P(Y = b)$$

  the logic basically goes that because you can write them as a product, then they must be independent.

- Indicator variables: strategy basically involves changing a complex RV X into simpler ones using Bernoulli variables. Basically, we can think of splitting X into n indicator variables:

  $$X = I_1 + I_2 + \cdots + I_n = \sum_{i=1}^{n} I_i$$

  This is useful when calculating the expectation:

  $$E(X) = E(I_1 + I_2 + \cdots + I_n) = E(I_1) + \cdots + E(I_n)$$

  Now, if each indicator is a Bernoulli variable, then they are defined to succeed with probability p_i (in the most general case we don't require I_i have the same distributions), so their expectation value is:

  $$E(I_i) = 1 \cdot p_i + 0 \cdot (1 - p_i) = P(I_i = 1) = p_i$$

  Therefore, we can write:

  $$E(X) = \sum_{i=1}^{n} P(I_i = 1) = \sum_{i=1}^{n} p_i$$

- Example with balls and bins: consider m balls into n bins, and define X to be the number of empty bins. What is $E(X)$? Computing the actual distribution of X (which we would need if we were to use the standard formula for expectation) would be a nightmare, since there are so many cases we need to consider. However, we now split X up into indicator variables:

  $$X = I_1 + \cdots + I_n$$

  where each I_i is an indicator whose value is 1 when bin i is empty, and 0 otherwise. Then, now let's focus on a particular bin, say WLOG we choose bin 1. Then, this simplifies a lot! Using the formula we derived in the previous bullet,

  $$E(I_1) = P(I_1 = 1) = P(\text{bin 1 is empty}) = \left(1 - \frac{1}{n}\right)^m$$

  If we multiply this by the n bins, then we have (we multiply because the bins are indistinguishable from each other)

  $$E(X) = nE(I_1) = n\left(1 - \frac{1}{n}\right)^m$$

  apparently, according to the notes, this tends towrads $\frac{n}{e}$, n't need to worry about that.