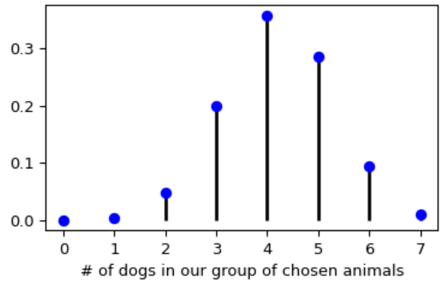
Preliminaries

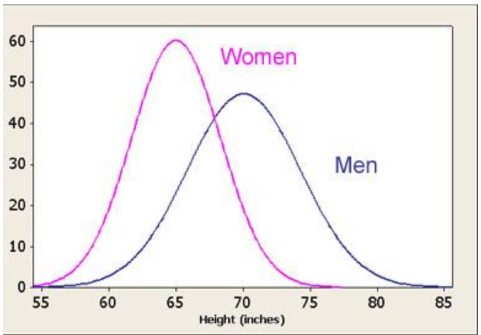
<u>Probability Concepts</u>

Expectation

- i.e. weighted mean = average
  - The mostly likely outcome
  - Expectation measure the central location of the distribution
- Probability distribution of a random variable X can be characterised by its
  - Probability mass function (PMF): $p_X(x)$ – if X is discrete
    - Discrete variable – number of dogs chosen
    - Assigns to each number of dogs the probability of it being chosen



  - Probability density function (PDF): $f_X(x)$ – if X is continuous
    - Continuous variable – person's exact height
    - The probability of the height being between 65 and 70 inches is the integral of the PDF from 65 to 70 (i.e. the area under the curve)



- For a random variable X with support S, the expectation E(X) is its mean μ_X
  - $$E(X) = \begin{cases} \sum_{x \in S} x p_X(x) & \text{if } X \text{ is discrete with pmf } p_X(x); \\ \int_{x \in S} x f_X(x) dx & \text{if } X \text{ is continuous with pdf } f_X(x) \end{cases}$$
    - Where the x before the p_x and f_x is the realised value of X
    - i.e. the expectation is a weighted average where the PMF/PDF provides the weight attributed to the realised value
  - Note: the support S is the set of all possible values of X
- Generalising – the expectation of g(X) is defined as
  - $$E(g(X)) = \begin{cases} \sum_{x \in S} g(x) p_X(x) & \text{if } X \text{ is discrete;} \\ \int_{x \in S} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$$
    - g is some function of the random variable

- Note: the g(X) before the $p_X(x)$ *and* $f_X(x)$ is the realization and the $p_X(x)$ and $f_X(x)$ are the weights

Moments

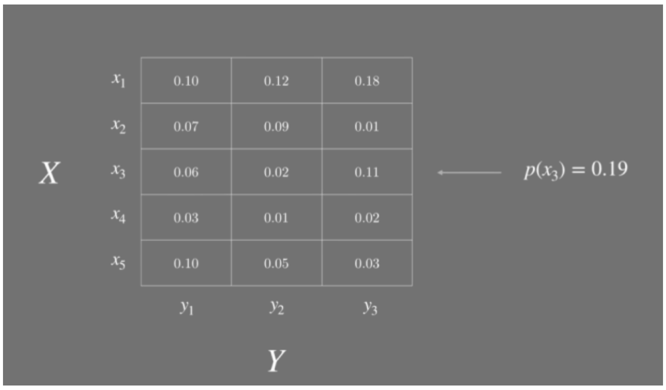
- First moment – equivalent to the mean
  - Discrete – $E[x] = \sum p_x(x)\, x$
  - Continuous – $E[x] = \int p_x(x)\, x\, dx$
- Second moment
  - Discrete – $E[x^2] = \sum p_x(x)\, x^2$
    - Variance = $\sum p_x(x)\, (x - E[x])^2 = E[x^2] - E[x]^2$
  - Continuous – $E[x^2] = \int p_x(x)\, x^2\, dx$
- n^{th} moment
  - Discrete – $E[x^n] = \sum p_x(x)\, x^n$
  - Continuous – $E[x^n] = \int p_x(x)\, x^n\, dx$

Variance, Skewness, Kurtosis and Covariance

- Variance – second central moment of X
  - Measures how close the values tend to be to the mean
    - i.e. how spread out the distribution is around the mean
  - $\sigma_X^2 \equiv Var(X) = E[(X - E(X))^2] = E[X^2] - E[X]^2$
  - Sample variance = $\hat{\sigma}_x^2 = \frac{\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^2}{T-1}$
- Skewness – tells us how skewed (positively or negatively) the distribution of X is
  - $Skew(X) = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$
    - Sample skewness = $\frac{\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^3}{T-1(\hat{\sigma}_x^3)}$
  - It is a unit-free measure (it is normalized) which means it is comparable against different types of random variables taking on different units of measurement
  - 3^{rd} central moment of X divided by the standard deviation^3
- Kurtosis – a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution
  - $Kur(X) = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$
    - Sample kurtosis = $\frac{\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^4}{T-1(\hat{\sigma}_x^4)}$
  - $Kur(X) = 3$ for a normal distribution
  - High (excess) kurtosis means heavy-tailed which means many outliers
  - It is a unit-free measure
  - 4^{th} central moment of X divided by the standard deviation^4
- Covariance – how closely related X and Y are in a *linear fashion*
  - $Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$

Joint, Marginal and Conditional Distributions

- Joint probability distribution of 2 random variables X and Y is characterised by
  - Joint pmf $p_{X,Y}\ (x, y)$ if discrete
    - i.e. P($X = x, Y = y$)
  - Joint pdf $f_{X,Y}\ (x, y)$ if continuous
- The marginal distribution of X is characterised by
  - $p_X(x) = \sum_{y \epsilon S_Y} p_{X,Y}(x, y)$ if discrete
    - i.e. obtain the PMF of X and Y and sum over all possible values of y
    - i.e. obtain all the values of $X = x$ for all values of y and sum
    - S_Y is the support of Y



X	y_1	y_2	y_3
x_1	0.10	0.12	0.18
x_2	0.07	0.09	0.01
x_3	0.06	0.02	0.11
x_4	0.03	0.01	0.02
x_5	0.10	0.05	0.03

  - $f_X(x) = \int_{y \epsilon S_Y} f_{X,Y}(x, y) dy$ if continuous
    - S_Y is the support of Y
- Conditional distribution of Y given X is characterised by
  - i.e. the probability of y occurring given x has occurred
  - Conditional pmf $p_{Y|X}(y|x) = \frac{Joint\ PMF(x,y)}{PMF(x)} = \frac{p_{X,Y}(x,y)}{p_X(x)}$ if discrete
  - Conditional pdf $f_{Y|X}(y|x) = \frac{Joint\ PDF(x,y)}{PDF(x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}$ if continuous

Conditional Expectation and Law of Iterated Expectations

- Conditional expectation of Y given X is
  - $$E(Y|X = x) = \begin{cases} \sum_{y \in S_Y} y p_{Y|X}(y|x) & \text{if discrete;} \\ \int_{y \in S_Y} y f_{Y|X}(y|x) dy & \text{if continuous.} \end{cases}$$
    - Weight the realised values of y by the conditional distribution function of y given x
- Generalising – conditional expectation of h(Y) given X is
  - $$E(h(Y)|X = x) = \begin{cases} \sum_{y \in S_Y} h(y) p_{Y|X}(y|x) & \text{if discrete;} \\ \int_{y \in S_Y} h(y) f_{Y|X}(y|x) dy & \text{if continuous.} \end{cases}$$
    - h is some function
- Law of iterated expectations – if the mean of Y is finite, $E(Y) = E[E(Y|X)]$
  - i.e. the average of Y is equivalent to the average of the conditional expectation of Y|X

- Holds if X and Y are random variables
- Note: you can introduce any additional expectation within an existing conditional expectation *provided that* the conditioning set in the inner layer of expectation contains the outer layer of the conditioning set
  - i.e. the condition in the outside expectation is part of the condition in the insider set
  - i.e. inner conditioning set contains the outer conditioning set
  - E.g. in E[E($\varepsilon_t | \mathcal{F}_{t-1}$)|$\mathcal{F}_{t-2}$], $\mathcal{F}_{t-1}$ contains $\mathcal{F}_{t-2}$

Independence

- Independence – X and Y do not have a relationship with each other
- X and Y are independent if the joint probability of X and Y is equal to the product of the marginal probabilities of X and Y
  - i.e. $P(X \in A, Y \in B) = P(X \in A)\ P(Y \in B)$ for all subsets A, B in $\mathbb{R}$
- Using the joint, conditional and marginal PMF/PDF, for all $x \in S_X, y \in S_Y$, x and y are independent if the following holds
  - $p_{X,Y}(x,y) = p_X(x)p_Y(y)\ \ OR\ \ p_{Y|X}(y|x) = p_Y(y)$ if discrete
  - $f_{X,Y}(x,y) = f_X(x)f_Y(y)\ \ OR\ f_{Y|X}(y|x) = f_Y(y)$ if continuous
- $X_1, X_2, \dots, X_n$ are jointly independent if $P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdot \dots \cdot P(X_n \in A_n)$ for all subsets $A_1, \dots, A_n$ in $\mathbb{R}$
  - Note: it is much easier to disprove the joint independence of 2 variables

Independence vs. Zero Correlation

- Definition of uncorrelation – $Cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = 0$
  - If X and Y are uncorrelated, then $E(XY) = E(X)E(Y)$
- Independence implies zero correlation, but not vice versa
  - i.e. zero correlation is a weaker assumption
- To prove X and Y are independent requires that $E(g(X)h(Y)) = E(g(X))E(h(Y))$ for all functions g(.) and h(.)
  - This is a stronger requirement than zero correlation
  - $E(XY) = E(X)E(Y)$ is a special version of the above as it sets g(.) and h(.) as identity functions (i.e. the function just equals a constant)
- Special case in which independence implies zero correlation and vice versa
  - If *(X, Y)* follows a bivariate normal distribution with $Corr(X,Y) = 0$, then X and Y are independent