- Formal definition convex function: $tf(a) + (t-1)f(b) \geq f(ta + (1-t)b)$, $\forall a, \forall b, t \in [0,1]$
- The value of Linearity of Expectation is that there is no need for the Random Variables to independent with respect to each other, this is the reason we try to create linear models so heavily. Proof:

$$\begin{aligned}
\mathbb{E}[aX + bY + c] &= \sum_{x,y} P(x,y)(ax + by + c) \\
&= \sum_{x,y} P(x,y)ax + \sum_{x,y} P(x,y)by + c\sum_{x,y} P(x,y) \\
&= a\sum_{x,y} P(y|x)P(x)x + b\sum_{x,y} P(x|y)P(y)y + c \\
&= a\sum_{x} P(x)x \underbrace{\sum_{y} P(y|x)}_{\text{sums to 1}} + b\sum_{y} P(y)y \underbrace{\sum_{x} P(x|y)}_{\text{sums to 1}} + c \\
&= a\mathbb{E}[x] + b\mathbb{E}[y] + c
\end{aligned} \tag{1}$$

- if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X][Y]$
- $\mathbf{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in X}(x - \mathbb{E}[X]^2)P(x)$

$$\begin{aligned}
\mathbf{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned} \tag{2}$$

- Scaling factor of Variance:

$$\begin{aligned}
&\mathbf{Var}[aX + b] = \\
&\quad = \mathbb{E}[(aX + b)^2] - (\mathbb{E}[aX + b])^2 \\
&\quad = \mathbb{E}[a^2X^2 + 2aXb + b^2] - (a\mathbb{E}[X] + b)^2 \\
&\quad = a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - (a^2\mathbb{E}[X]^2 + 2ab\mathbb{E}[X] + b^2) \\
&\quad = a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) \\
&\quad = a^2\mathbf{Var}[X]
\end{aligned} \tag{3}$$

- If X and Y are independent, then:

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] \tag{4}$$

- Since we like to work with easily interpretable units we will look at Standard Deviation as well:

$$\sigma[aX + b] = |a|\sigma[X] \tag{5}$$

- Covariance extends naturally from variance and gives information about the relationship between two random variables:

$$\mathbf{Cov}[X,Y] = \mathbb{E}[(X - E[X])(Y - \mathbb{E}(Y))] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{6}$$

$$\mathbf{Cov}[aX + u, bY + v] = ab\mathbf{Cov}[X,Y] \tag{7}$$

- If X, Y are independent then it is useful to note that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $\mathbf{Cov}[X,Y] = 0$

  Correlation is like "normalized" covariance:

$$\rho_{X,Y} = \mathbf{Corr}[X,Y] = \frac{\mathbf{Cov}[X,Y]}{\sqrt{\mathbf{Var}[X]}\sqrt{\mathbf{Var}[Y]}} \tag{8}$$

A Bernoilli Random Variable is represents one of the most fundamental probability distributions and is defined as:

$$p_x(k) = \begin{cases} p, & \text{if } k = 1 \\ (1-p), & \text{if } k = 0 \end{cases}$$

With a variance of:

$$\begin{aligned} \mathbb{E}[X] &= 1 \cdot p + 0 \cdot (1-p) \\ \mathbb{E}[X^2] &= 1^2 \cdot p + 0 \cdot (1-p) \\ \mathbf{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p) \end{aligned}$$