**Note that** $P(Y_j/X_i) \neq P(X_i/Y_j)$
In fact, $P(Y_j/X_i)$ gives the prob of Yj given Xi is transmitted, while $P(X_i/Y_j)$ the prob of Xi given the Yj is received.

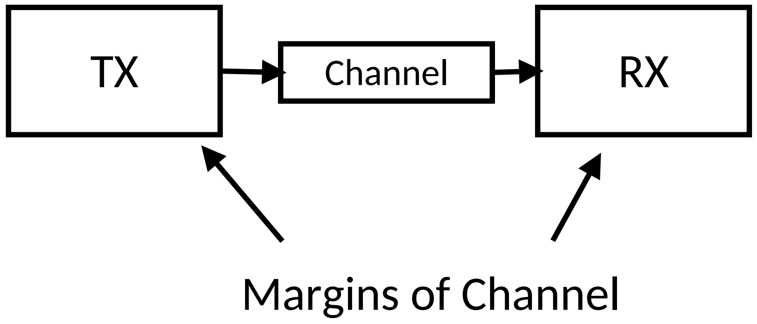
**Properties of I(Xi, Yi) :**
1. I(Xi, Yi) is symmetric i.e. I(Xi, Yi) = I(Yi, Xi)
2. **I(Yi, Xi) > 0**, if a posterior prob. > priori prob. Then Yi provides **+ve** information about Xi.
3**. I(Yi, Xi) = 0,** if a posterior prob. = priori prob. Then Yi provides no information about Xi.
4. **I(Yi, Xi) < 0,** if a posterior prob. = priori prob. Then Yi provides or adds ambiguity (fuzzy) to Xi.

**Marginal Entropy:**

A term usually used to denote both **source entropy H(X)** & **receiver entropy H(Y)**.

$$H(X) = -\sum_{i=1}^{n} P(X_i) \log_2 P(X_i) \qquad (bits/symbol)$$

$$H(Y) = -\sum_{j=1}^{m} P(Y_j) \log_2 P(Y_j) \qquad (bits/symbol)$$



The average amount of information associated with the pair (Xi, Yi) is called **joint (system) entropy**.*:*

$$H(X,Y) = H(XY) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 P\left(X_i, Y_j\right)$$

The average amount of information associated with the pair ($X_i$ / $Y_i$) & ($Y_i$ / $X_i$) are called conditional entropy.

$$\mathbf{H}(Y/X) = -\sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{P}(xi\ ,\ yj)\ log_2 \mathbf{P}(yj/xi)$$

$$= \text{Noise Entropy} \qquad \text{bits/symbol}$$

$$\mathbf{H}(X/Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{P}(xi\ ,\ yj)\ log_2 \mathbf{P}(xi/yj) = \text{Losses Entropy} \qquad \text{bits/symbol}$$

## <u>Transinformation (Average Mutual Information):</u>

It is the Average mutual information, this is statical average of all pairs I($x_i$, $y_i$) and it is measured by bits/symbol.

$$I(X,Y) = \sum_{j=1}^{m} \sum_{i=1}^{n} I\left(x_i, y_j\right) P\left(X_i, Y_j\right)$$

$$I(X,Y) = \sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 \frac{P\left(X_i/Y_j\right)}{P(X_i)}$$

$$I(X,Y) = \sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 \frac{P\left(Y_j/X_i\right)}{P(Y_j)}$$

**Example:** Show that H(X,Y) = H(X) + H(Y/X)

**Solution:**

$$H(X,Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 P\left(X_i, Y_j\right)$$

But **P(*XI* , *YJ*)** = **P(*XI* ).P( *YJ/XI*)**. Put this inside the log term only

$$H(X,Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 P\left(X_i\right) P(Y_j/X_i)$$

$$H(X,Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 P\left(X_i\right) - \sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 P(Y_j/Xi)$$

If we reverse the first sum *for i and j then* $\sum_{j=1}^{m} \mathbf{P(X}\boldsymbol{i}\mathbf{\ ,Y}\boldsymbol{j}\mathbf{)} = P\left(X_j\right)$

$$H(X,Y) = -\sum_{i=1}^{n} P\left(X_i\right) \log_2 P\left(X_i\right) - \sum_{j=1}^{m} \sum_{i=1}^{n} P\left(X_i, Y_j\right) \log_2 P(Y_j/Xi)$$

$$H(X,Y) = H(X) + H(Y/X)$$

**Homework: 1.** show that H(X,Y)=H(Y)+ H(X/Y). **2.** show that I(X,Y)=H(X)-H(X/Y) **3.** show that I(X,Y)=H(Y)-H(Y/X)

**Example:** Show that **I(X,Y)=H(X) – H(X/Y)** .

**Solution:** We know that

$$I(Xi\,,\,Yj) = \sum_{j=1}^{m}\sum_{i=1}^{n} P(xi\,,\,yj)\ log_2 \frac{P(xi\,/\,yj)}{P(x_i)}$$

$$= \sum_{j=1}^{m}\sum_{i=1}^{n} P(xi\,,\,yj)\ log_2 P(xi\,/\,yj) - \sum_{j=1}^{m}\sum_{i=1}^{n} P(xi\,,\,yj)\ log_2 P(x_i)$$

As before, we reverse the order of the 2nd sum for i and j then $\sum_{j=1}^{m}$ **P(*xi* , *yj*)** = **P(*xi* )** then:-

**I(X,Y)=H(X) – H(X/Y)** .

**Note** that above identity indicates that the Transinformation **I(X,Y)** is the average information gained at the **Rx** which is the difference between the information produced by the source H(x) and the information lost in the channel H(X/Y) [losses entropy] due to noise and jamming.

**Example:** Show that **I(X,Y)= 0** for extremely noisy channel?

**Solution:** For extremely noisy channel, Then $y_j$ gives no information about $x_i$ (the Rx can not decide anything about $x_i$ as if we transmit a deterministic signal $x_i$ but the Rx receives noise like signal $y_j$ that is completely has no correlation with $x_i$ .

**Then** $x_i$ and $y_j$ *are independent and*

**P(*xi* / *yj*)** $= P(x_i)$ *for all i and j then* $I(x_i\,,\,y_j) = \log_2 1 = 0$ for all i and j then

**I(X,Y)=** average of $I(x_i\,,\,y_j) = 0$ .

**Example:** The joint probability is given by

$$P(X_i, Y_j) = \begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.125 \\ 0.0625 & 0.0625 \end{bmatrix}$$

**Find :**

1. Marginal entropies 2. System Entropies
3. Noise and losses entropies 4. Mutual information between X1 and Y2
5. Transinformation and 6. Draw the channel model

**Solution:**

**1.** Marginal entropies (H(X) and H(Y))

$$P\left(X_j\right) = \sum_{j=1}^{2} P(X_i, Y_j) = [0.75 \quad 0.125 \quad 0.125]$$

$$P\left(Y_j\right) = \sum_{i=1}^{3} P(X_i, Y_j) = [0.5625 \quad 0.4375]$$

$$H(X) = -\sum_{i=1}^{3} P(X_i) \log_2 P(X_i) = \frac{1}{Ln\ (2)} [0.75\ Ln\ (0.75) + 2 \times 0.125\ Ln\ (0.125)\ ] = 1.06127\ bits/symbol$$

$$H(Y) = -\sum_{i=1}^{2} P(Y_i) \log_2 P(Y_i) = \frac{1}{Ln\ (2)} [0.5625\ Ln\ (0.5625) + 0.4375\ Ln\ (0.4375)\ ] = 0.9887\ bits/symbol$$

**2**. $$H(X,Y) = -\sum_{j=1}^{2} \sum_{i=1}^{3} P\left(X_i, Y_j\right) \log_2 P\left(X_i, Y_j\right)$$

$$= \frac{1}{Ln\ (2)} [0.5\ Ln\ (0.5) + 0.25 \ln(0.25) + 0.125\ Ln\ (0.125) + 2 \times 0.0625\ Ln\ (0.0625)] = 1.875\ bits/symbols$$

**Solution: Cont.** 3. Noise and losses entropies

H(Y/X)= H(X,Y) - H(X) = 1.875 – 1.06127 = 0.81373 bit/symbol. (Noise Entropy)
H(X/Y)= H(X,Y) - H(Y) = 1.875 – 0.9887 = 0.8863 bit/symbol . (Losses Entropy)

$$P(X\boldsymbol{i}, Y\boldsymbol{j}) = \begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.125 \\ 0.0625 & 0.0625 \end{bmatrix}$$

**4**. Mutual information between $X_1$ and $Y_2$

$$\mathbf{I}(\boldsymbol{X_1}, \boldsymbol{Y_2}) = Log_2 \frac{P(X_1/Y_2)}{P(X_1)} \quad \text{since } P(X1/Y2) = \frac{P(X1, Y2)}{P(Y2)} \text{ then}$$

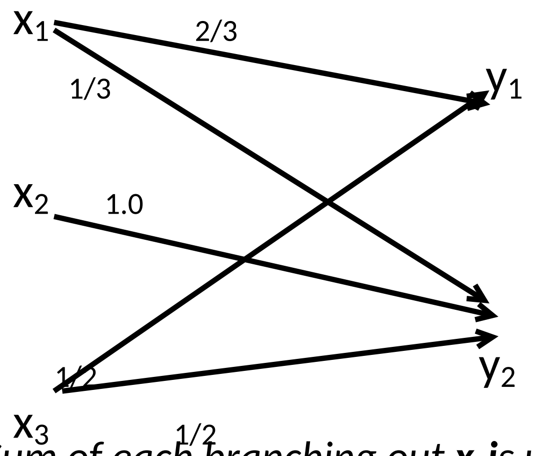
$$\mathbf{I}(\boldsymbol{X_1}, \boldsymbol{Y_2}) = Log_2 \frac{P(X_1, Y_2)}{P(X_1)P(Y_2)} = Log_2 \frac{0.25}{0.75 \times 0.4375} = -0.3923 \; bits \; . \; That\ means\ Y2 gives\ ambiguity\ about\ X1$$

**5.** Transinformation I(X,Y)=H(X)-H(X/Y) = 0.17497 bits/symbol

**6.** To draw a channel, we need to find $P(Y_j/X_i)$

$$P(Y\boldsymbol{j}/X\boldsymbol{i}) = \frac{P(X\boldsymbol{i}, Y\boldsymbol{j})}{P(X\boldsymbol{i})} = \begin{bmatrix} \frac{0.5}{0.75} & \frac{0.25}{0.75} \\ 0 & \frac{0.125}{0.125} \\ \frac{0.0625}{0.125} & \frac{0.0625}{0.125} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

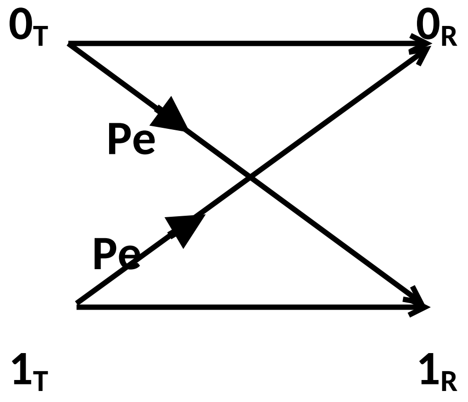
Sum of each row is unity




*Sum of each branching out* $\boldsymbol{x_i}$ *is unity*

**Example:** Find and plot the transinformation for a binary symmetric channel (BSC) shown if $P(0_T = P(1_T) = 1/2$.

**Solution:** We need to find I(X,Y)= H(Y)-H(Y/X). This BSC is a very well-known channel and Practical values for Pe $\ll 1$ . $0_T = x_1$ , $1_T = x_2$, $0_R = y_1$ and $1_R = y_2$



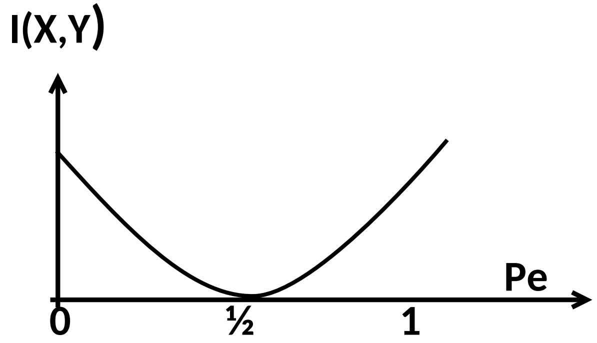

$$P(Y/X) = \begin{bmatrix} 1-Pe & Pe \\ Pe & 1-Pe \end{bmatrix}$$

$$P(X,Y) = \begin{bmatrix} \frac{1-Pe}{2} & \frac{Pe}{2} \\ \frac{Pe}{2} & \frac{1-Pe}{2} \end{bmatrix} \rightarrow P(Y) = [\tfrac{1}{2} \quad \tfrac{1}{2}]$$ and $H(Y) = H(Y)_{max} = \log_2 2 = 1$ bits

$$H(Y/X) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j)\, log_2 P(y_j/x_i)$$

$$= -[\{\tfrac{1-Pe}{2} \log 2\,(1-Pe)\}\times 2 + \{\tfrac{Pe}{2} \log 2\,(Pe)\}\times 2]$$

$$= -[(1-Pe) \log 2\,(1-Pe) + Pe \log 2(Pe)]$$

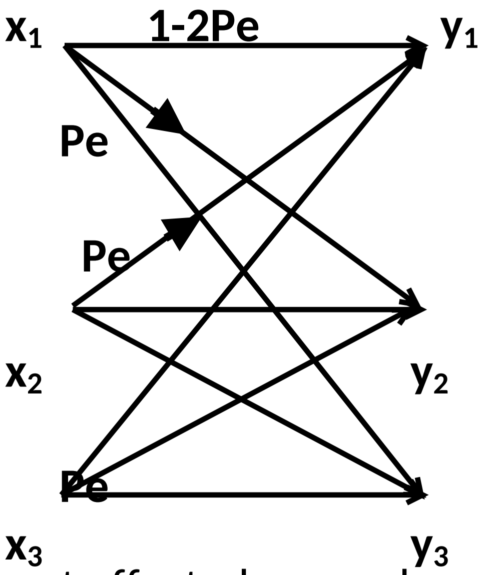



$$I(X,Y) = H(Y) - H(Y/X) = 1 + [(1-Pe) \log 2\,(1-Pe) + Pe \log 2\,(Pe)]$$

**H.W.** A BSC has $P(Y/X) = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$ if $I(0_T)$=3 bits find the system and losses entropies?

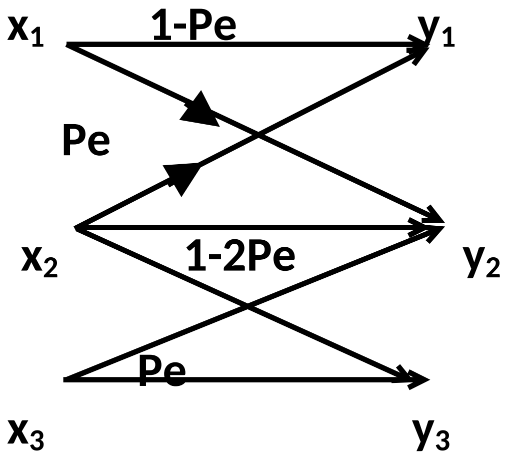
# Ternary Symmetric Channel (TSC)

This has the transitional prob:




$$p(Y/X)=\begin{array}{c|ccc|} & y1 & y2 & y3 \\ x1 & 1-2pe & pe & pe \\ x2 & pe & 1-2pe & pe \\ x3 & pe & pe & 1-2pe \end{array}$$

This TSC is symmetric but not very practical since practically $x_1$ and $x_3$ do not affected so much as $x_2$. In fact the interference between $x_1$ and $x_3$ is much less than the interference between $x_1$ & $x_2$ or $x_2$ & $x_3$. Hence, the more practical but nonsymmetric channel has the conditional prob:




$$p(Y/X)=\begin{array}{c|ccc|} & y1 & y2 & y3 \\ x1 & 1-pe & pe & 0 \\ x2 & pe & 1-2pe & pe \\ x3 & 0 & pe & 1-pe \end{array}$$

Where $x_1$ interfere with $x_2$ exactly the same as interference between $x_2$ and $x_3$, but $x_1$ and $x_3$ are not interfered.

**1-lossless channel:** This has only one nonzero element in each column of the transitional matrix **p(Y/X).** As an example

$$p(Y/X)=\begin{array}{c} \\ x1 \\ x2 \\ x3 \end{array}\begin{bmatrix} y1 & y2 & y3 & y4 & y5 \\ 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This channel has **H(X/Y)=0** and **I(X,Y)=H(X)** with zero losses entropy. (draw the channel model of this channel)

**2-Determinstic channel:** This has only one nonzero element in each row of the transitional matrix p(Y/X). As an example:

$$p(Y/X)=\begin{array}{c} \\ x1 \\ x2 \\ x3 \\ x4 \\ x5 \end{array}\begin{bmatrix} y1 & y2 & y3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This has **H(Y/X)=0** and **I(X,Y)=H(Y)** with zero noise entropy. (draw the channel model of this channel).

**3-Noiseless channel:** This has only one nonzero element in each row and column of the transitional matrix p(Y/X), i.e. it is an identity matrix. As an example: This has **H(X/Y)=H(Y/X)=0**, and **I(X,Y)=H(X)=H(Y).** (draw the channel model of this channel).

$$p(Y/X)=\begin{array}{c} \\ x1 \\ x2 \\ x3 \end{array}\begin{bmatrix} y1 & y2 & y3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that noiseless channel is a losses and deterministic channel