Different distances

- What is the distance between two points?
- What is the distance between two lines?
- What is the distance between two curves?
- What is the distance between two matrices?
- What is the distance between two clusters?
- What is the distance between two words (strings)?
- What is the distance between two PDFs?
- What is the distance between two species?

What is Kullback_Leibler divergence?

- What is the distance between two PDFs?
- Answer: KLD (relative entropy) can be used
- How one PDF is different from other?

Consider two PDFs – p and q

x_1	x_2	x_3	…	…	x_n
$p(x_1)$	$p(x_2)$	$p(x_3)$	…	…	$p(x_n)$

x_1	x_2	x_3	…	…	x_n
$q(x_1)$	$q(x_2)$	$q(x_3)$	…	…	$q(x_n)$

Information content of p and q

$p(x_1)$	$p(x_2)$	$p(x_3)$	…	…	$p(x_n)$
$\text{Log}[1/p(x_1)]$	$\text{Log}\ [1/p(x_2)]$	$\text{Log}\ [1/p(x_3)]$	…	…	$\text{Log}\ [1/p(x_n)]$

$q(x_1)$	$q(x_2)$	$q(x_3)$	…	…	$q(x_n)$
$\text{Log}\ [1/q(x_1)]$	$\text{Log}\ [1/q(x_2)]$	$\text{Log}\ [1/q(x_3)]$	…	…	$\text{Log}\ [1/q(x_n)]$

Information difference between p and q

$\text{Log } [1/q(x_1)] - \text{Log } [1/p(x_1)] =$	$\text{Log } p(x_1) - \text{Log } q(x_1)$
$\text{Log } [1/q(x_2)] - \text{Log } [1/p(x_2)] =$	$\text{Log } p(x_2) - \text{Log } q(x_2)$
$\text{Log } [1/q(x_3)] - \text{Log } [1/p(x_3)] =$	$\text{Log } p(x_3) - \text{Log } q(x_3)$
…	…
$\text{Log } [1/q(x_n)] - \text{Log } [1/p(x_n)] =$	$\text{Log } p(x_n) - \text{Log } q(x_n)$

We have n number of differences
Make a single number – How?
expectation

Find the average information difference

Information difference	**Weighting factor**
Log $p(x_1)$ - Log $q(x_1)$	$p(x_1)$
Log $p(x_2)$ - Log $q(x_2)$	$p(x_2)$
Log $p(x_3)$ - Log $q(x_3)$	$p(x_3)$
…	…
Log $p(x_n)$ - Log $q(x_n)$	$p(x_n)$

$$we\ \ are\ \ going\ \ to\ \ sum\ \ (\log(p(x_i)) - \log(q(x_i)))$$

$$which\ is\ the\ \ weighing\ \ factor?\ \ \ p(x_i)$$

$$\sum_i p(x_i).(\log(p(x_i)) - \log(q(x_i)))$$

Relative entropy

- Aka Kullback–Leibler divergence
- Distance between two distributions
  - Measure of how one probability distribution diverges from another distribution

$$D_{KL}(P \| Q) = -\sum_x P(x) \log \frac{Q(x)}{P(x)}$$

$$= \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

Asymetrical

Let $\mathcal{X} = \{0, 1\}$ and consider two distributions p and q on $\mathcal{X}$. Let $p(0) = 1 - r$, $p(1) = r$, and let $q(0) = 1 - s$, $q(1) = s$. Then

$$D(p||q) = (1 - r)\log\frac{1 - r}{1 - s} + r\log\frac{r}{s}$$

and

$$D(q||p) = (1 - s)\log\frac{1 - s}{1 - r} + s\log\frac{s}{r}.$$

If $r = s$, then $D(p||q) = D(q||p) = 0$. If $r = \frac{3}{4}$, $s = \frac{1}{4}$, we can calculate

$$D(p||q) = \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{3}{4}} + \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{1}{4}} = 1 - \frac{1}{2}\log 3 = 0.2075 \text{ bit},$$

whereas

$$D(q||p) = \frac{3}{4}\log\frac{\frac{3}{4}}{\frac{1}{2}} + \frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{2}} = \frac{3}{4}\log 3 - 1 = 0.1887 \text{ bit}$$

J divergence – becomes distance

$$D_{KL}(P \| Q) = -\sum_x P(x) \log \frac{Q(x)}{P(x)} \qquad D_{KL}(Q \| P) = -\sum_x Q(x) \log \frac{P(x)}{Q(x)}$$

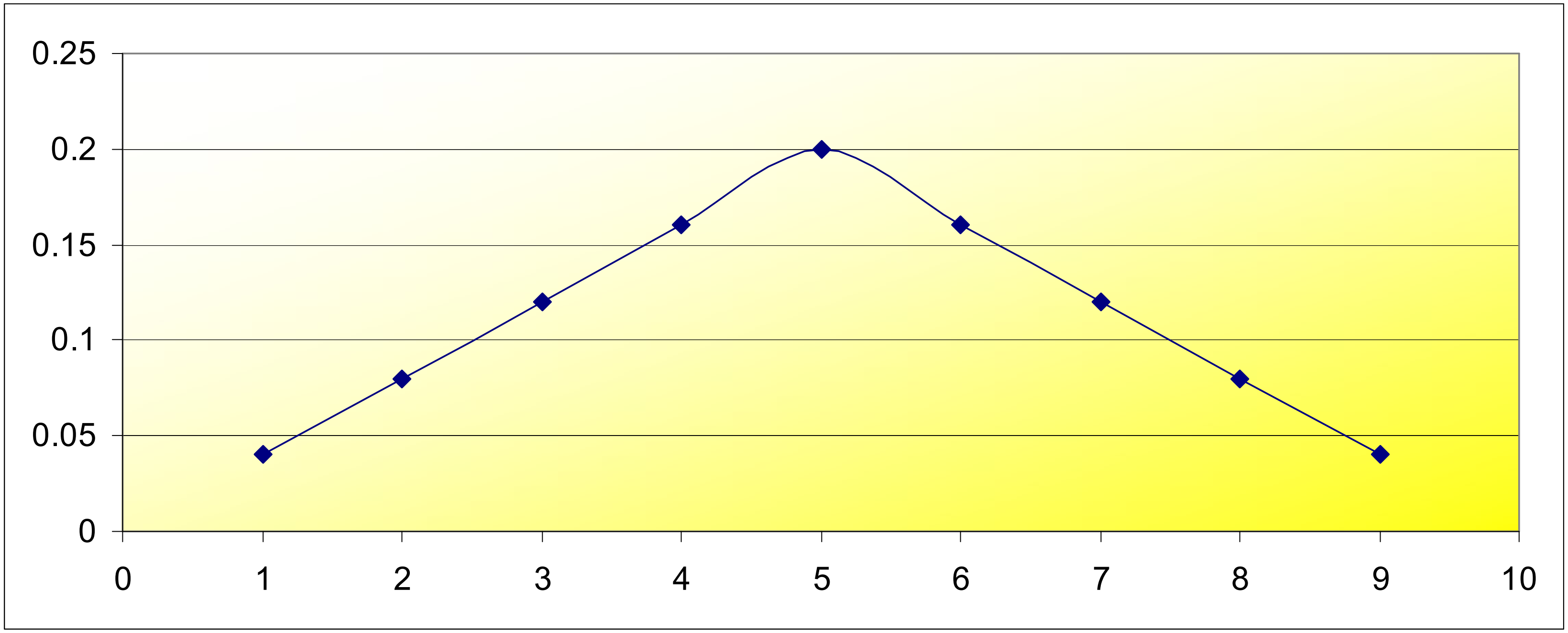
$$= \sum_x P(x) \log \frac{P(x)}{Q(x)} \qquad = \sum_x Q(x) \log \frac{Q(x)}{P(x)}$$

$D_{KL}(P||Q) \neq D_{KL}(Q||P)$ (in general)

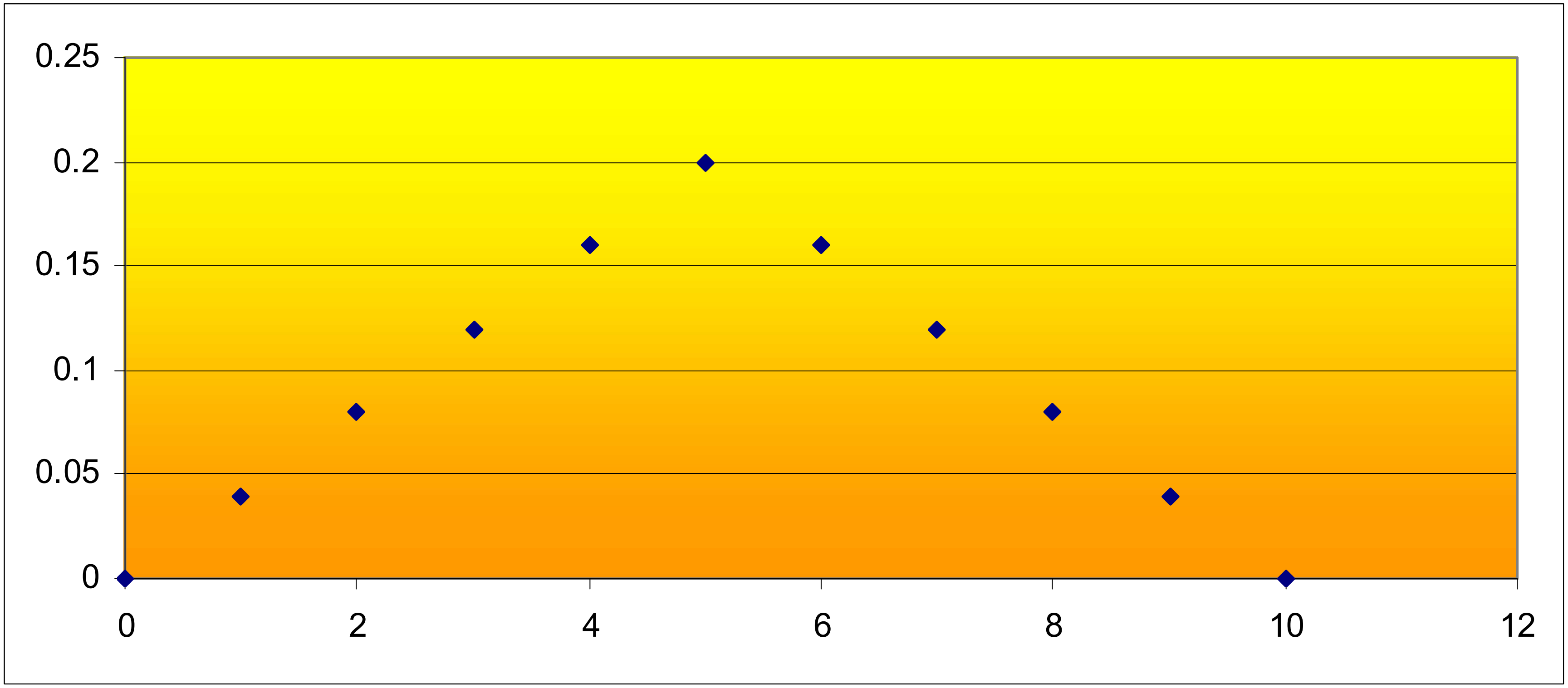
- We cannot call it as distance measure

$$\sqrt{\frac{1}{2}\left[D_{KL}(P \| Q)\right]^2 + \frac{1}{2}\left[D_{KL}(Q \| P)\right]^2} \quad : \ J \quad divergence$$

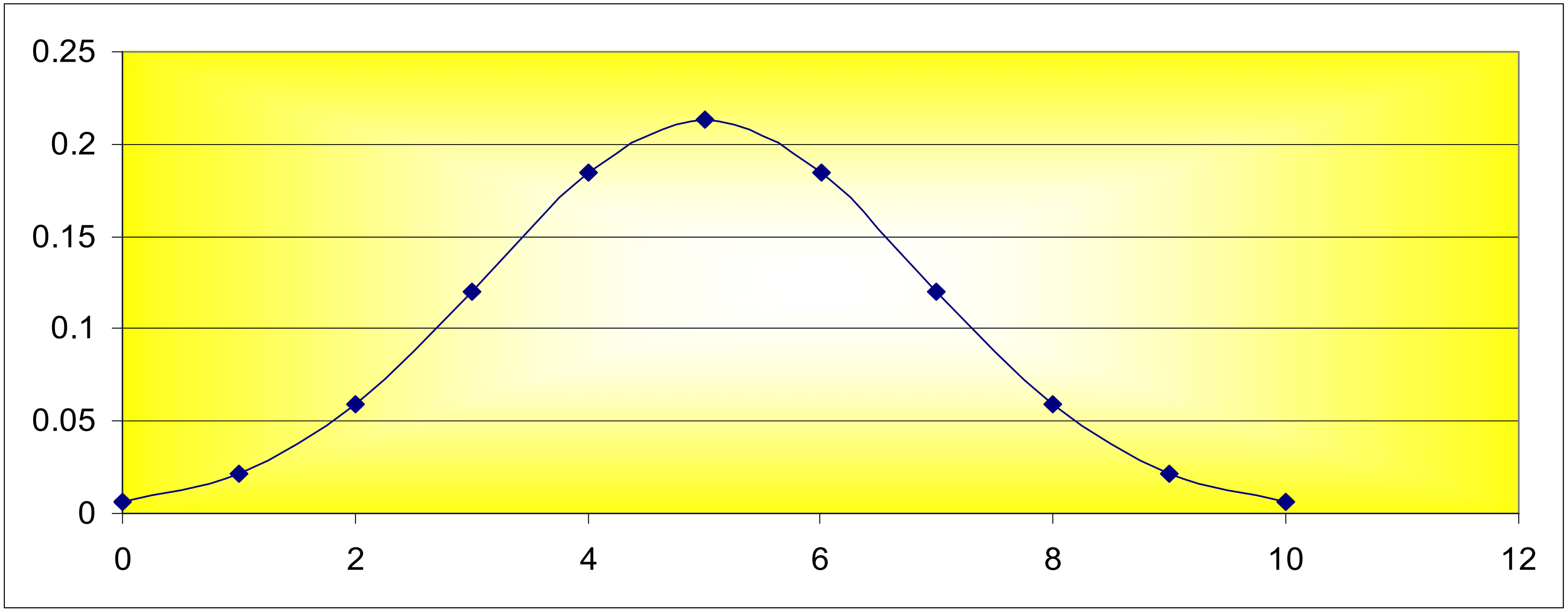
Triangular distribution



Some body assumes it is Gaussian

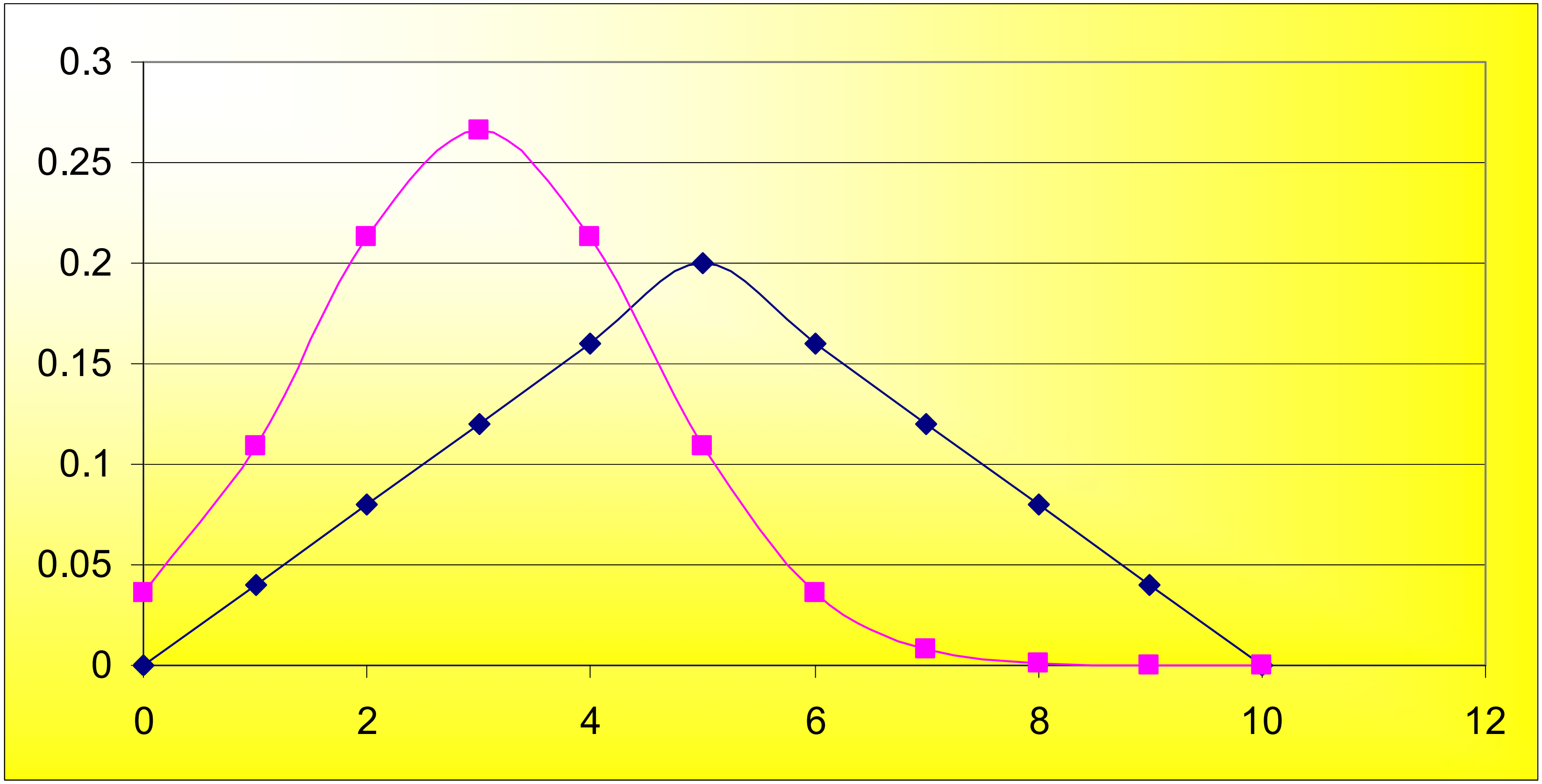


Use Gaussian to approximate the triangular (previous slide)

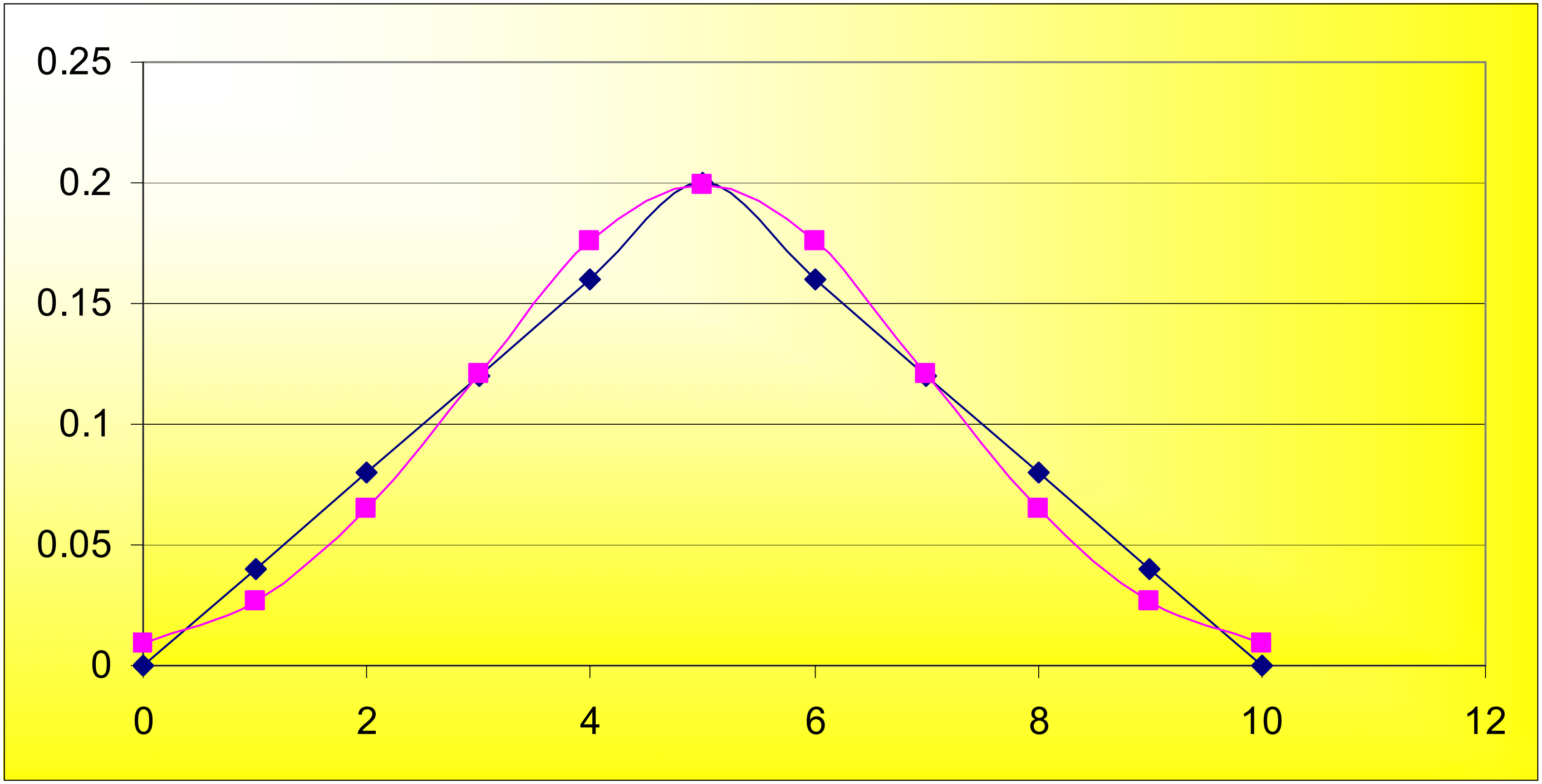


$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

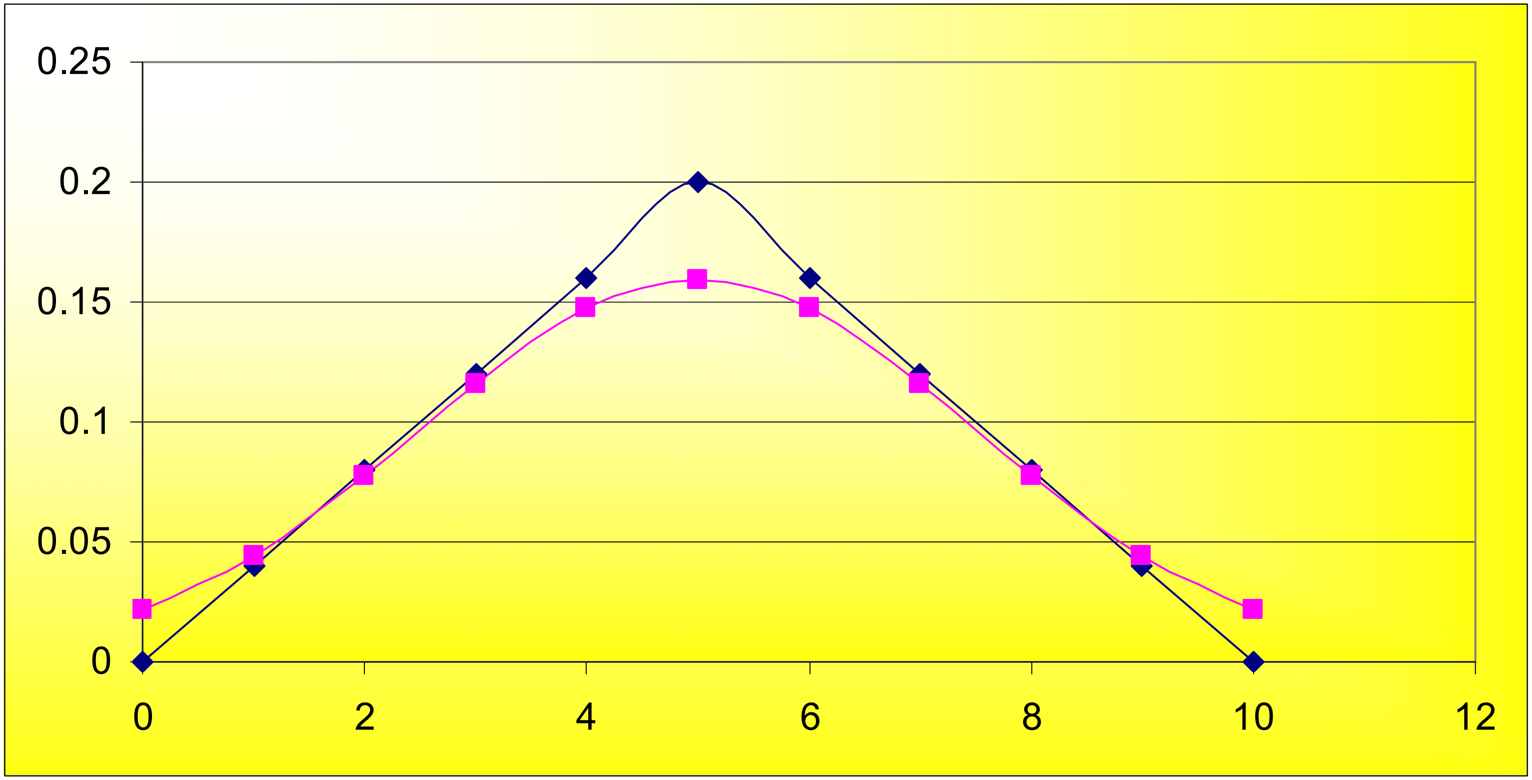
Mean = 3; std = 1.5



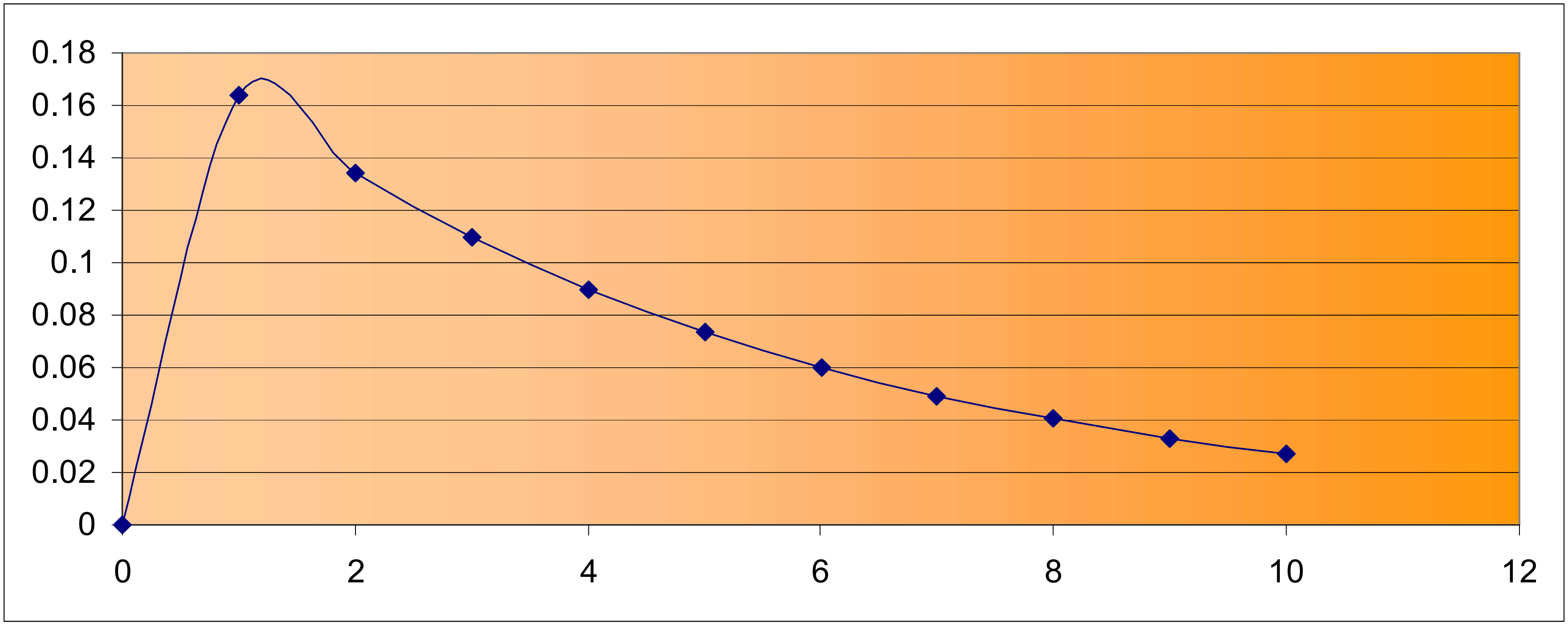
Mean = 5; std = 2



Mean = 5; std = 2.5

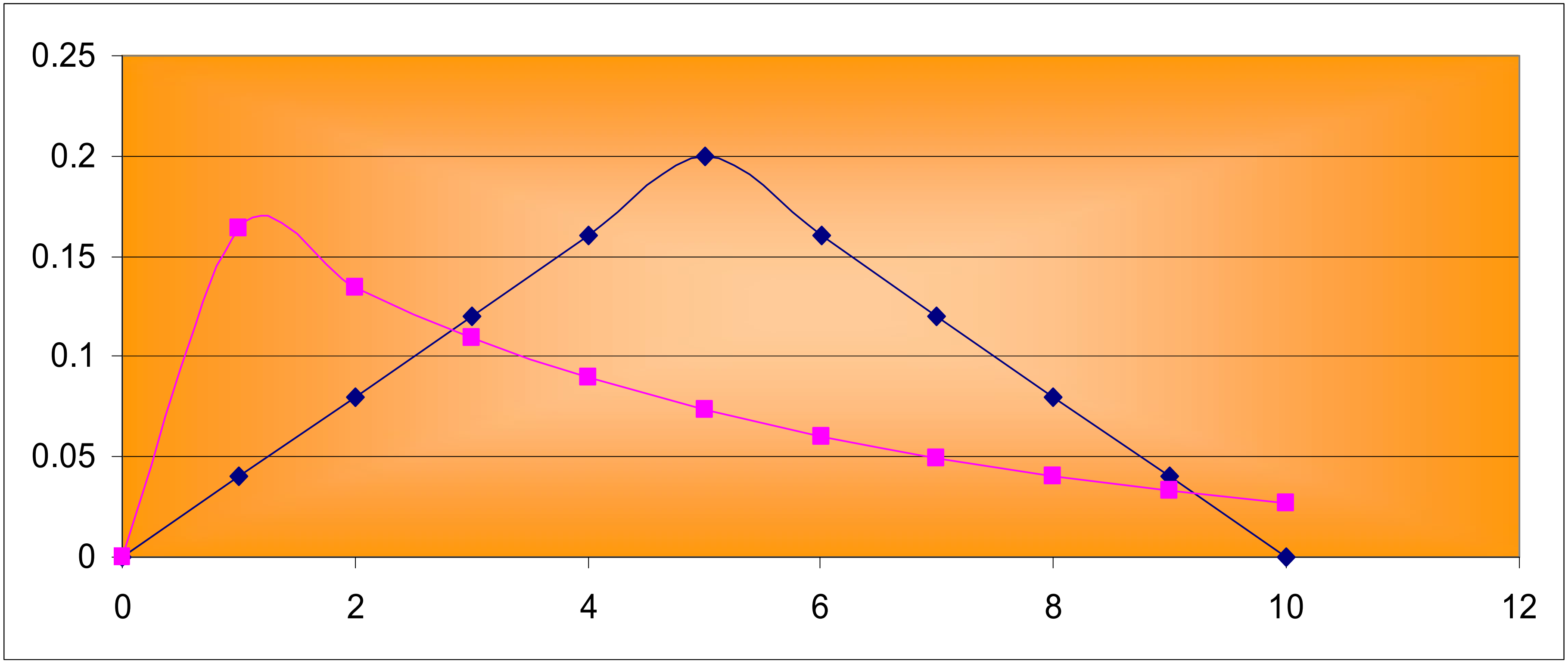


Use Weibull to approximate the triangular

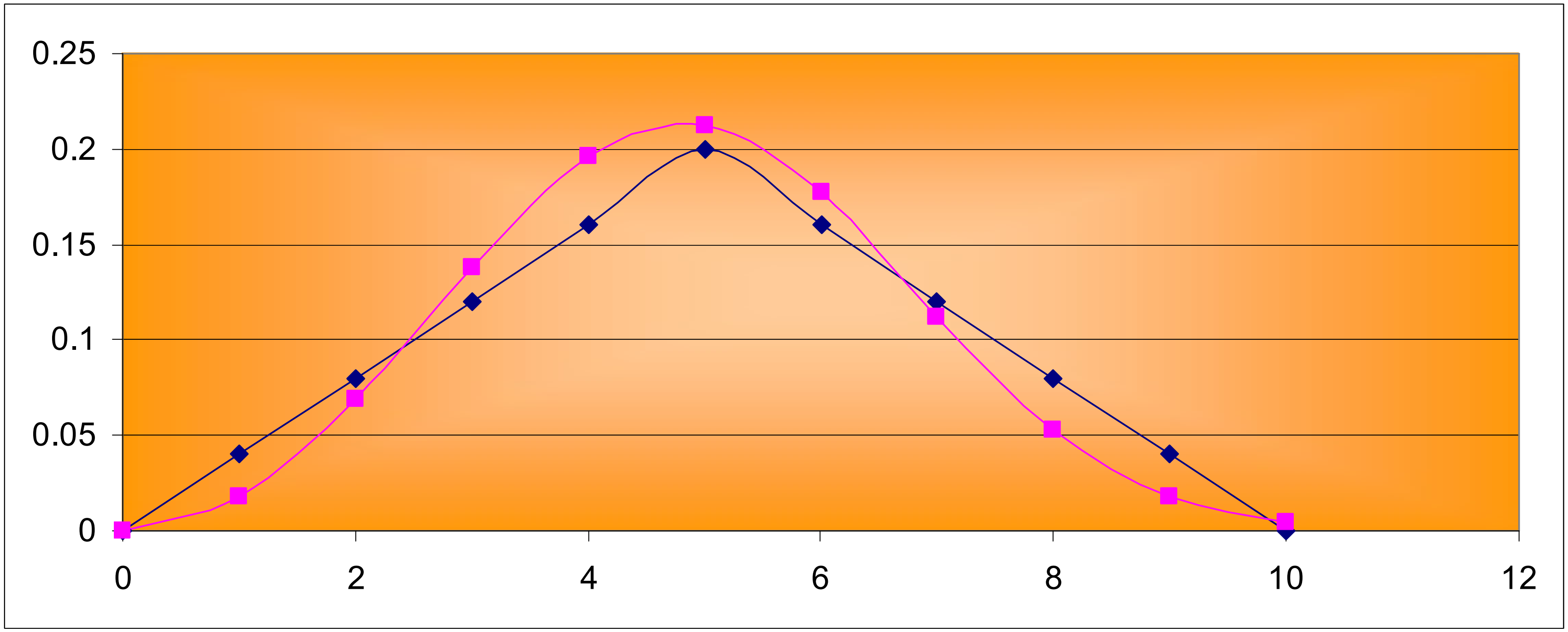


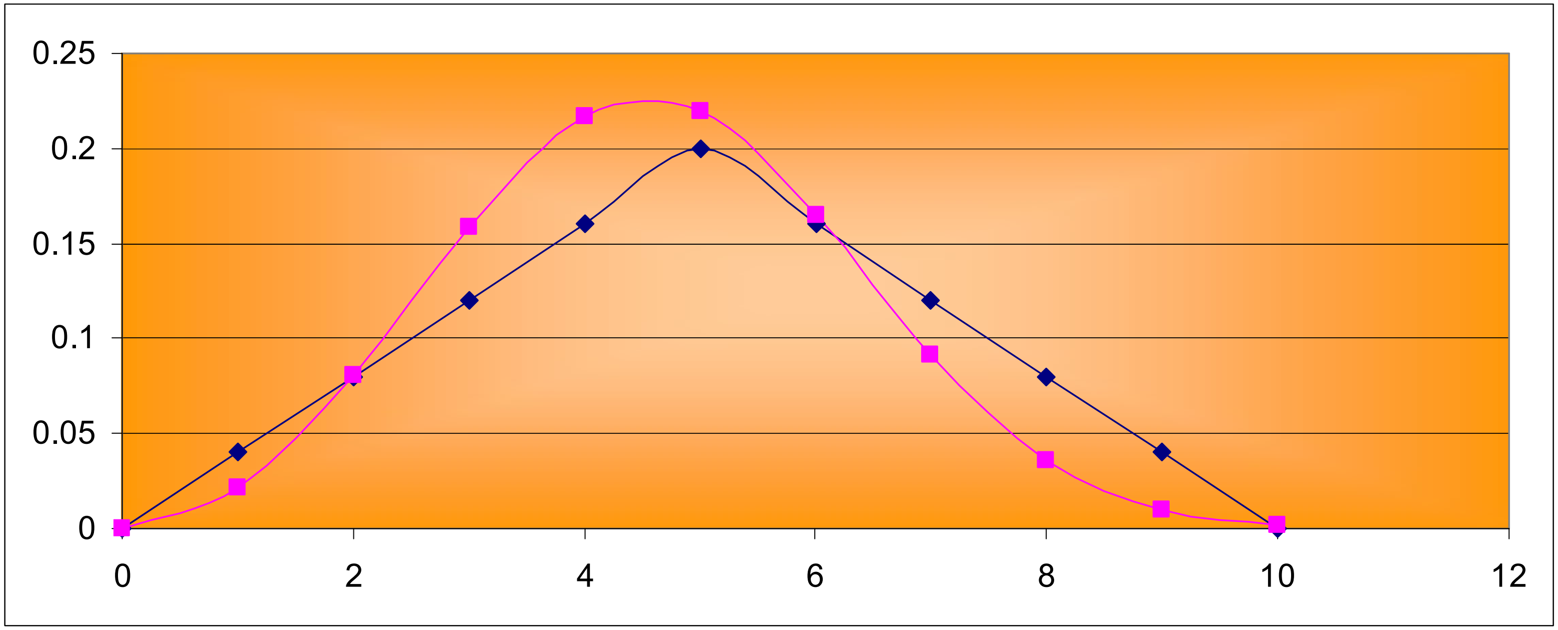
$$\left(\frac{k}{\lambda}\right) \cdot \left(\frac{x}{\lambda}\right)^{(k-1)} . e^{\left(-\left(\frac{x}{\lambda}\right)^{k}\right)} \quad for \quad x >= 0$$

λ = 5; k = 1



$\lambda = 5.5; k = 3$





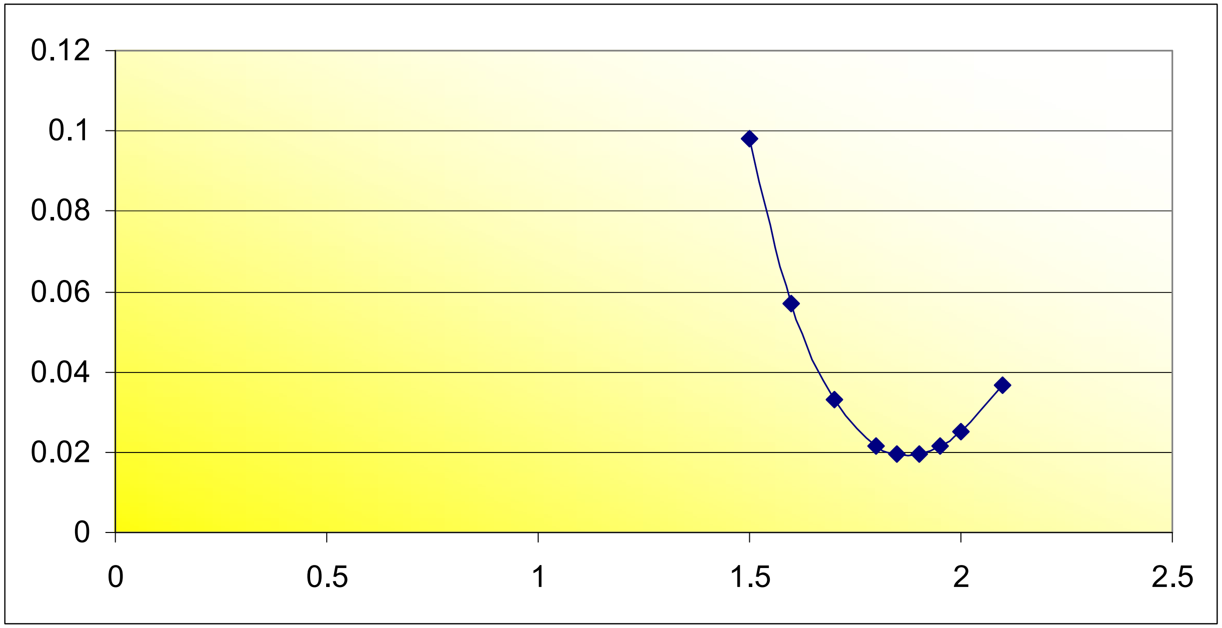
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12

Which is good?

- Kullback–Leibler divergence/distance answers this

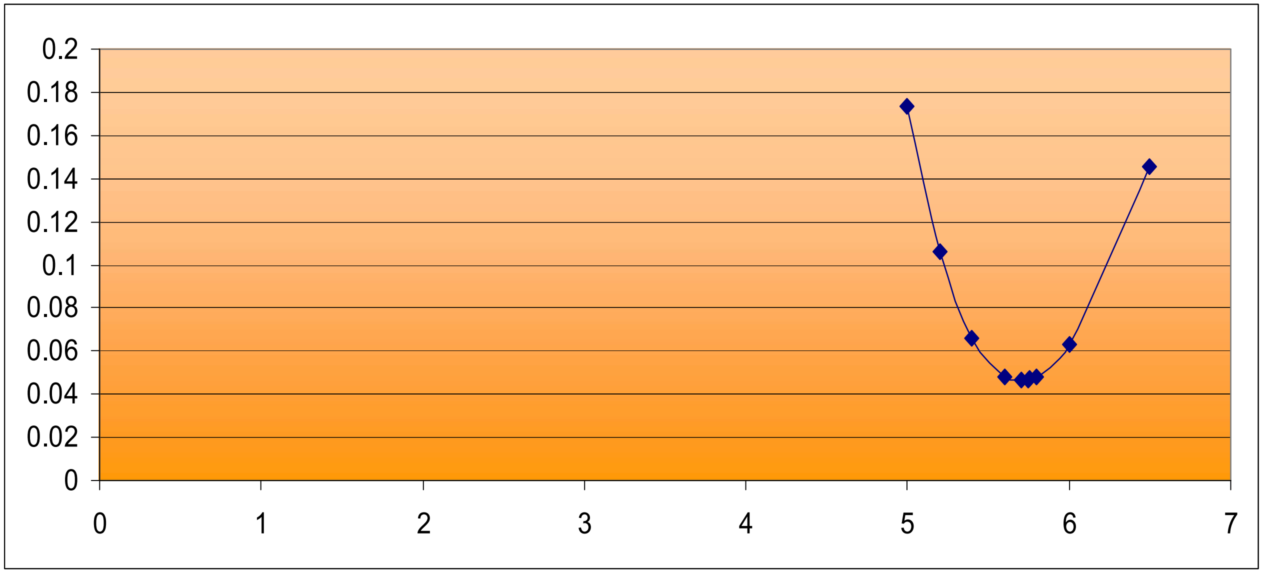
KLD versus Std (Gaussian)

Std	KLD
1.5	0.098153
1.6	0.057095
1.7	0.021518
1.8	0.019576
1.85	0.019729
1.9	0.021716
2	0.025309
2.1	0.036549
2.5	0.116225



KLD versus Lambda (Weibull)

lamda	Kldistance
5	0.173221
5.2	0.105958
5.4	0.066108
5.6	0.048314
5.7	0.046326
5.75	0.046856
5.755	0.046961
5.8	0.048332
6	0.062778
6.5	0.145433



KLD of two matrices-e.g. 1

- Consider the matrices A and B
- The matrix elements are from the set {0, 1, 2, 3, 4}
- Find the $E(\varepsilon^2)$ where ε is the error computed using A-B
- Find KLD(A||B) and KLD(B||A)

$$A = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}, \quad and \quad B = \begin{bmatrix} 1 & 4 & 3 \\ 1 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

KLD of two matrices-e.g. 2

- Consider the matrices A and B
- The matrix elements are from the set {0, 1, 2, 3, 4}
- Find KLD(A||B) and KLD(B||A)

$$A = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}, \quad and \quad B = \begin{bmatrix} 1 & 4 & 3 \\ 1 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

KLD of two matrices-e.g. 2

- Consider the matrices A and B
- The matrix elements are from the set {0, 1, 2, 3, 4}
- Find the $E(X^2)$
- Find KLD(A||B) and KLD(B||A)

$$A = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}, \quad and \quad B = \begin{bmatrix} 2 & 0 & 4 \\ 3 & 1 & 1 \end{bmatrix}$$

Thought on KLD

$$D_{KL}(P||Q) = -\sum P(x)\, log \frac{Q(x)}{P(x)}$$

$$= \sum P(x)\, log \frac{P(x)}{Q(x)}$$

$$= \sum P(x) \log P(x) - \sum P(x) \log\, Q(x)$$

$$= -Shannon's\ entropy + Cross\ \ entropy$$

Entropy, Relative entropy & cross entropy

Relative entropy = Cross entropy – Shannon's entropy

$$D_{KL}(P||Q) = H(P,Q) - H(P)$$