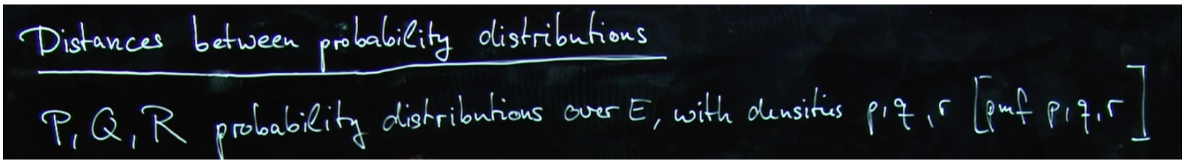
Recitation: Distance between probability distribution

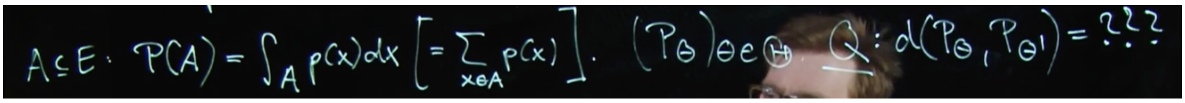
Recitation problem statement

1. For $p, q \in (0,1)$ and $n \geq 1$, compute the Kullback-Leibler divergence between $\mathsf{B}(n, p)$ and $\mathsf{B}(n, q)$.
2. Compute the Kullback-Leibler divergence between two Gaussian distributions that have the same variance.
3. Show that the Poisson distribution with parameter $1/n$ converges in total variation distance to the Dirac distribution at zero (i.e., the distribution of the random variable that is always equal to zero).

Definitions of TV and KL

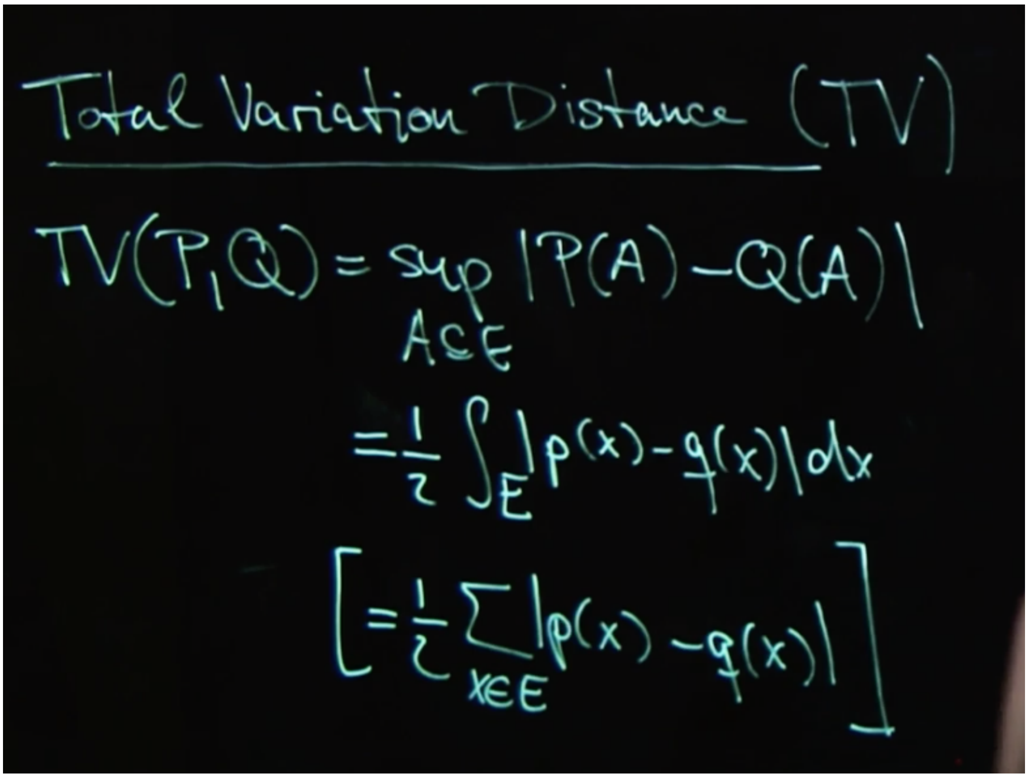


- dealing with continuous density functions or discrete mass functions



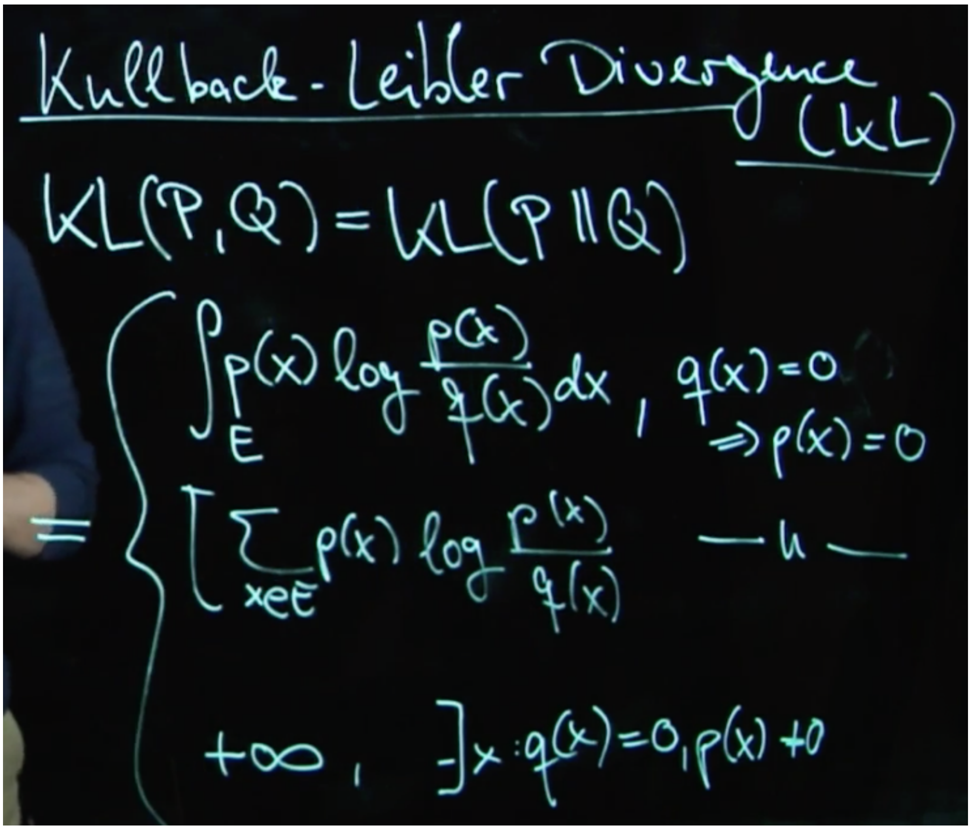
- first question, is theta identified (i.e. do 2 different values give different distributions)
- without this then there’s no way of knowing which theta we are approaching

Total Variance Distance (TV)



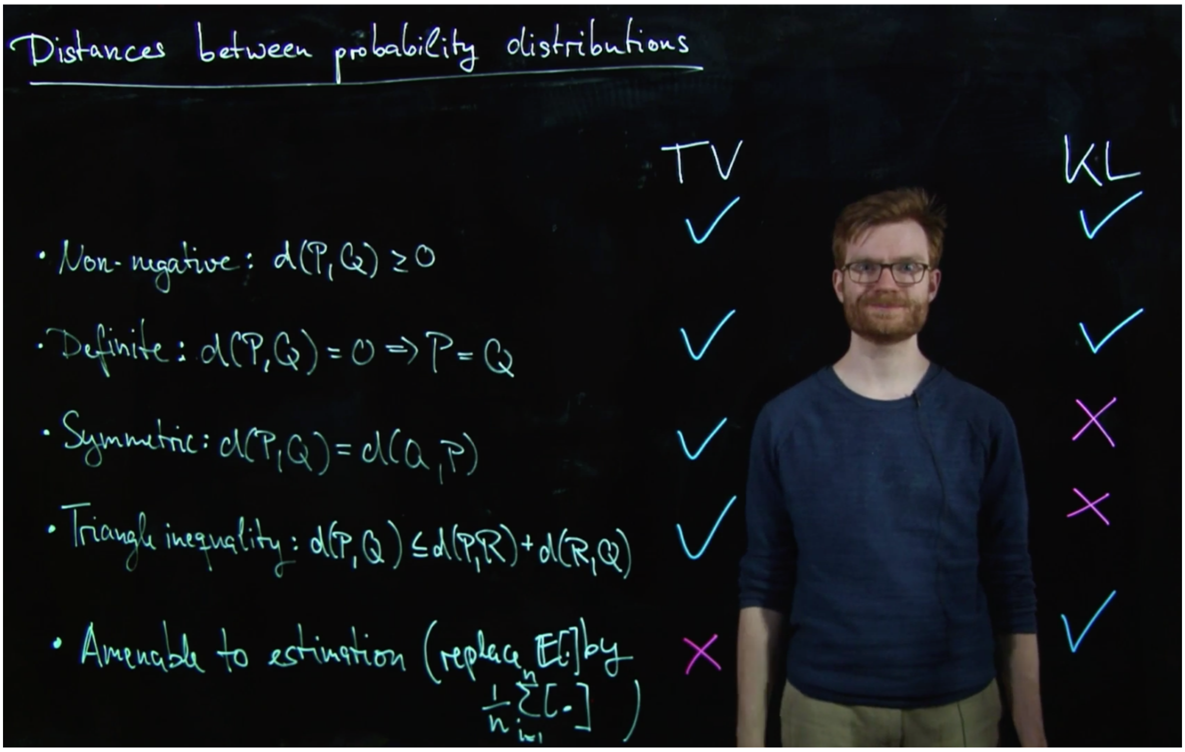
- sup is supremum (like maximum)
- taking the absolute value between the difference of two distributions
- integral continuous case, same discrete case

Kullback-Leibler Divergence



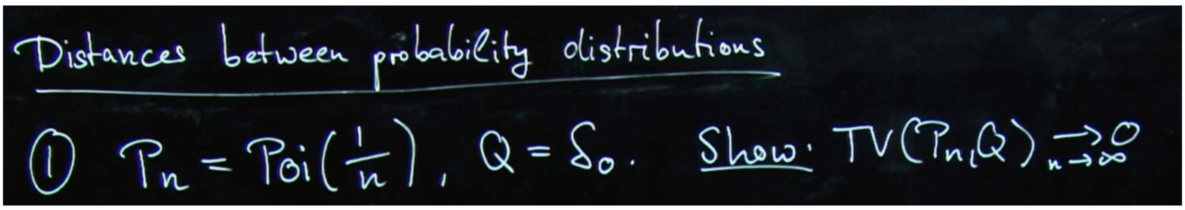
- called a divergence instead of a distance because it is not symmetric
- can’t swap around P and Q
- in integral we require p to be zero if q is zero
- if q is 0 when p is not then assign value of infinity
- -u- means same condition as above it holds for the discrete case

Comparing properties of TV and KL

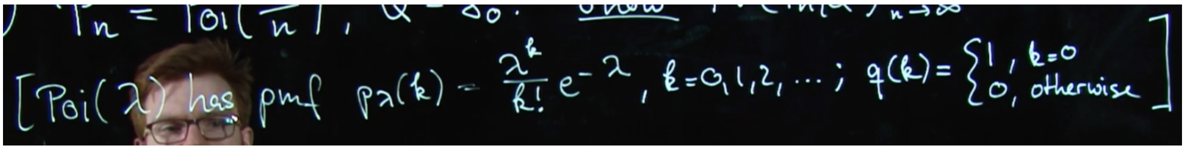


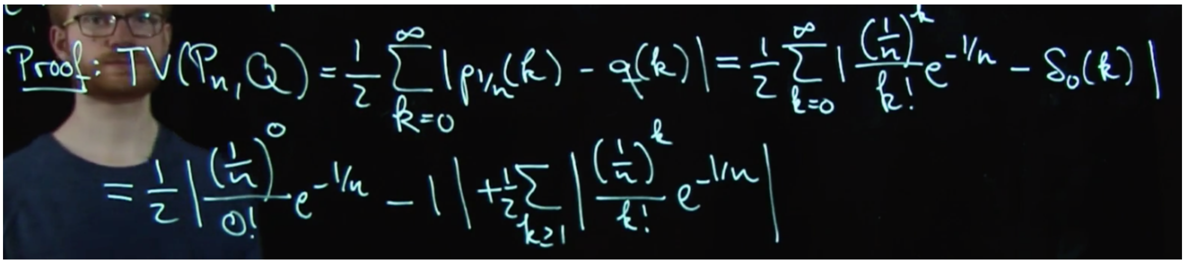
- called distances as they are non-negative
- definite: if distance is 0 then the distributions are equal
- triangle inequality: taking distances between 2 distributions can also be achieved by taking the distances between them and a third distribution
- KL divergence appears to be worse due to triangle inequality and not being symmetric, but it is amenable to estimation (replace expectation by sample average)
-

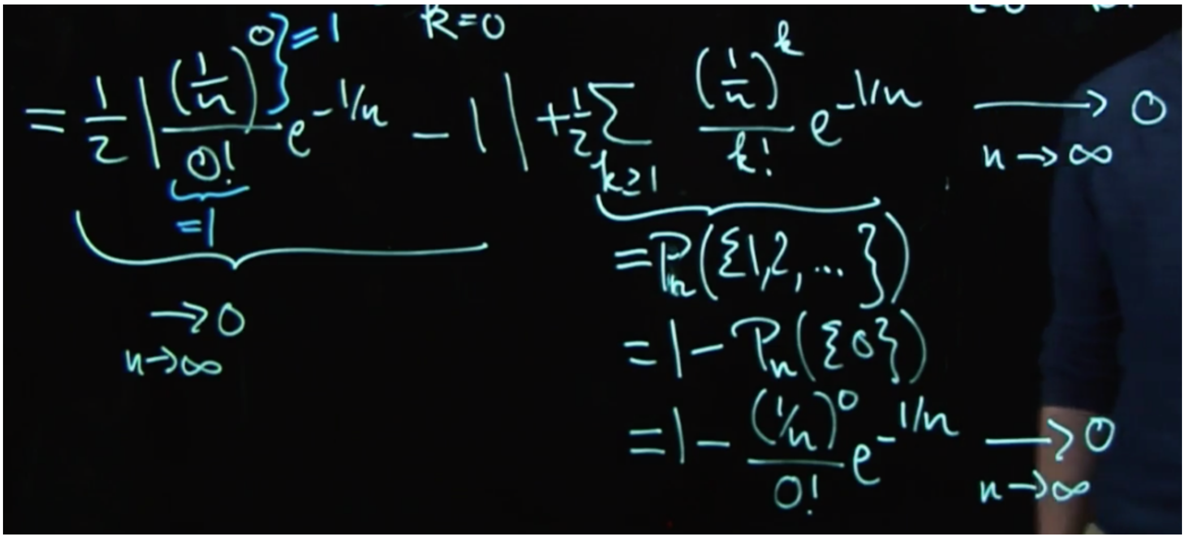
Convergence of Poisson distributions to the delta function



- P is poisson and Q is delta function at 0

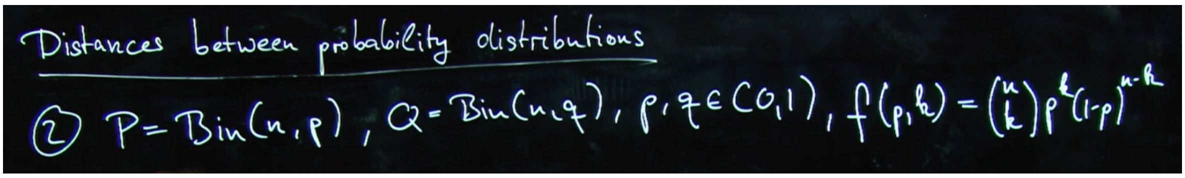


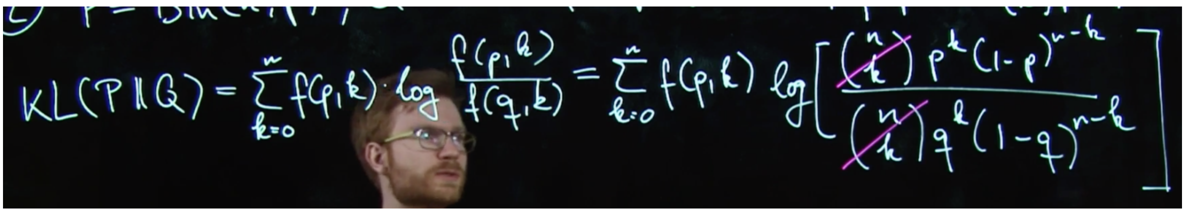




- evaluating at k=0,
- can remove the absolute value because they are all positive numbers
-
-

KL between binomial distributions





- just plugging in the log functions at the moment as the function before the

- log simplifies easier later
-

$$= \sum_{k=0}^{n} f(p,k)\left[\log\left(\left(\frac{p}{q}\right)^{k}\right) + \log\left(\left(\frac{1-p}{1-q}\right)^{n-k}\right)\right]$$

$$= \sum_{k=0}^{n} f(p,k)\left[k\log\left(\frac{p}{q}\right) + (n-k)\log\left(\frac{1-p}{1-q}\right)\right]$$

- the terms inside the logs are just numbers so can be pulled out

Note:

$$\boxed{\begin{array}{l} X \sim Bin(n,p) \\ E[X] = n \cdot p \end{array}}$$

plug in expectation of binomial for pmf divided by k, so the f(p,k) and k is replaced by n.p

$$= \log\left(\frac{p}{q}\right) \cdot np + \log\left(\frac{1-p}{1-q}\right) \cdot (n-np)$$

$$q \to 0 : KL(P,Q) \to \infty$$
$$q = 0, p \in (0,1) : KL(P,Q) = \infty$$
$$q \to 1;\ p \to 0;\ p \to 1\ ???$$

KL between Gaussian distributions

Distances between probability distributions

(3) $P = \mathcal{N}(a,1)$, $Q = \mathcal{N}(b,1)$, $a, b \in \mathbb{R}$; $f_{a,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\cdot(x-a)^2}$

$$KL(P\|Q) = \int_{\mathbb{R}} f_{a,1}(x) \log\left[\frac{f_{a,1}(x)}{f_{b,1}(x)}\right] dx = \int_{\mathbb{R}} f_{a,1}(x) \log\left[\frac{\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}\cdot(x-a)^2}}{\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}\cdot(x-b)^2}}\right] dx$$

$$= \int_{\mathbb{R}} f_{a,1}(x) \cdot \cancel{\log e}^{\underbrace{-\frac{1}{2}(x-a)^2 + \frac{1}{2}(x-b)^2}_{= -\frac{1}{2}(x^2 - 2xa + a^2) + \frac{1}{2}(x^2 - 2xb + b^2)}} dx \quad = \int_{\mathbb{R}} f_{a,1}(x) \cdot \left[x(a-b) - \frac{1}{2}a^2 + \frac{1}{2}b^2\right] dx$$

- again was useful to leave the first density function alone initially

$$= (a-b) \underbrace{\int_{\mathbb{R}} x f_{a,1}(x) dx}_{=a} + \left(-\frac{1}{2}a^2 + \frac{1}{2}b^2\right) \underbrace{\int_{\mathbb{R}} f_{a,1}(x) dx}_{=1} = \underbrace{(a-b)a}_{=a^2 - ba} - \frac{1}{2}a^2 + \frac{1}{2}b^2 = \frac{1}{2}(a^2 - 2ab + b^2) \boxed{= \frac{1}{2}(a-b)^2}$$

- with 2 Gaussians with variance of 1 we get back the exact difference between their two means
- =a because that is the mean and =1 because it is a probability density function