

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 $B(n, p)$ and $B(n, q)$.
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

Distances between probability distributions

P, Q, R probability distributions over E , with densities p, q, r [pmf p, q, r]

- dealing with continuous density functions or discrete mass functions

$A \subseteq E : P(A) = \int_A p(x) dx \left[= \sum_{x \in A} p(x) \right]. (P_\theta)_{\theta \in \Theta} \quad Q : d(P_\theta, P_{\theta'}) = ? ? ?$

- 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)

Total Variation Distance (TV)

$$TV(P, Q) = \sup_{A \subseteq E} |P(A) - Q(A)|$$

$$= \frac{1}{2} \int_E |P(x) - Q(x)| dx$$

$$\left[= \frac{1}{2} \sum_{x \in E} |P(x) - Q(x)| \right]$$

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

Kullback-Leibler Divergence

Kullback-Leibler Divergence (KL)

$$KL(P, Q) = KL(P \parallel Q)$$

$$= \begin{cases} \int_E P(x) \log \frac{P(x)}{Q(x)} dx, & Q(x) = 0 \Rightarrow P(x) = 0 \\ \left[\sum_{x \in E} P(x) \log \frac{P(x)}{Q(x)} \right] - h \\ +\infty, &]x : Q(x) = 0, P(x) > 0 \end{cases}$$

- 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

| <u>Distances between probability distributions</u> | | |
|--|----|----|
| | TV | KL |
| • Non-negative: $d(P, Q) \geq 0$ | ✓ | ✓ |
| • Definite: $d(P, Q) = 0 \Rightarrow P = Q$ | ✓ | ✓ |
| • Symmetric: $d(P, Q) = d(Q, P)$ | ✓ | ✗ |
| • Triangle inequality: $d(P, Q) \leq d(P, R) + d(R, Q)$ | ✓ | ✗ |
| • Amenable to estimation (replace $E[\cdot]$ by $\frac{1}{n} \sum_{i=1}^n [\cdot]$) | ✗ | ✓ |



- 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

| <u>Distances between probability distributions</u> | |
|---|--|
| ① $P_n = \text{Poi}\left(\frac{1}{n}\right), Q = \delta_0.$ | Show: $\text{TV}(P_n, Q) \xrightarrow{n \rightarrow \infty} 0$ |

- P is poisson and Q is delta function at 0

$T_n = \text{Poi}(\bar{n})$, $\bar{n} = \infty$. Show $\text{TV}(P_n, Q) \rightarrow 0$

[$\text{Poi}(\lambda)$ has pmf $p_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k=0, 1, 2, \dots$; $q(k) = \begin{cases} 1, & k=0 \\ 0, & \text{otherwise} \end{cases}$]

Proof: $\text{TV}(P_n, Q) = \frac{1}{2} \sum_{k=0}^{\infty} |p_{1/n}(k) - q(k)| = \frac{1}{2} \sum_{k=0}^{\infty} \left| \frac{(\frac{1}{n})^k}{k!} e^{-1/n} - S_0(k) \right|$

 $= \frac{1}{2} \left| \frac{(\frac{1}{n})^0}{0!} e^{-1/n} - 1 \right| + \frac{1}{2} \sum_{k \geq 1} \left| \frac{(\frac{1}{n})^k}{k!} e^{-1/n} \right|$

$$\begin{aligned} &= \frac{1}{2} \left| \underbrace{\frac{(\frac{1}{n})^0}{0!} e^{-1/n}}_{=1} - 1 \right| + \underbrace{\frac{1}{2} \sum_{k \geq 1} \frac{(\frac{1}{n})^k}{k!} e^{-1/n}}_{\substack{\rightarrow 0 \\ n \rightarrow \infty}} \xrightarrow{n \rightarrow \infty} 0 \\ &= P_n(\{1, 2, \dots\}) \\ &= 1 - P_n(\{0\}) \\ &= 1 - \frac{(\frac{1}{n})^0}{0!} e^{-1/n} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

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

KL between binomial distributions

Distances between probability distributions

② $P = \text{Bin}(n, p)$, $Q = \text{Bin}(n, q)$, $p, q \in (0, 1)$, $f(p, k) = \binom{n}{k} p^k (1-p)^{n-k}$

$\text{KL}(P \parallel Q) = \sum_{k=0}^n f(p, k) \log \frac{f(p, k)}{f(q, k)} = \sum_{k=0}^n f(p, k) \log \left[\frac{\binom{n}{k} p^k (1-p)^{n-k}}{\cancel{\binom{n}{k}} q^k (1-q)^{n-k}} \right]$

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

log simplifies easier later

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

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

Note:

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

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

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

$$\begin{aligned}
 q \rightarrow 0 : KL(P, Q) &\rightarrow \infty \\
 q = 0, p \in (0, 1) : KL(P, Q) &= \infty \\
 q \rightarrow 1; p \rightarrow 0; p \rightarrow 1 &???
 \end{aligned}$$

KL between Gaussian distributions

Distances between probability distributions

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

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

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

$$= -\frac{1}{2}(x^2 - 2xa + a^2) + \frac{1}{2}(x^2 - 2xb + b^2)$$

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

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

$$= \boxed{\frac{1}{2}(a-b)^2}$$

- 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