我们先来一个相对简单的例子。假设我们有两个随机变量x1,x2，各自服从一个高斯分布$N_1(\mu_1,\sigma_1^2),N_2(\mu_2,\sigma_2^2)$，那么这两个分布的KL散度该怎么计算呢?

我们知道

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

那么KL(p1,p2)就等于

$$\begin{aligned}
&\int p_1(x)log\frac{p_1(x)}{p_2(x)}dx\\
&=\int p_1(x)(logp_1(x)dx-logp_2(x))dx=\int p_1(x)*(log\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{\frac{(x-\mu_1)^2}{2\sigma_1^2}}-log\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}})dx\\
&=\int p_1(x)*(-\frac{1}{2}log2\pi-log\sigma_1-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{1}{2}log2\pi+log\sigma_2+\frac{(x-\mu_2)^2}{2\sigma_2^2})dx\\
&=\int p_1(x)(log\frac{\sigma_2}{\sigma_1}+[\frac{(x-\mu_2)^2}{2\sigma_2^2}-\frac{(x-\mu_1)^2}{2\sigma_1^2}])dx\\
&=\int(log\frac{\sigma_2}{\sigma_1})p_1(x)dx+\int(\frac{(x-\mu_2)^2}{2\sigma_2^2})p_1(x)dx-\int(\frac{(x-\mu_1)^2}{2\sigma_1^2})p_1(x)dx\\
&=log\frac{\sigma_2}{\sigma_1}+\frac{1}{2\sigma_2^2}\int((x-\mu_2)^2)p_1(x)dx-\frac{1}{2\sigma_1^2}\int((x-\mu_1)^2)p_1(x)dx
\end{aligned}$$

（更新）到这里停一下，有童鞋问这里右边最后一项的化简，这时候积分符号里面的东西是不看着很熟悉？没错，就是我们常见的方差嘛，于是括号内外一约分，就得到了最终的结果——$\frac{1}{2}$。

好，继续。

$$\begin{aligned}
&=log\frac{\sigma_2}{\sigma_1}+\frac{1}{2\sigma_2^2}\int((x-\mu_2)^2)p_1(x)dx-\frac{1}{2}\\
&=log\frac{\sigma_2}{\sigma_1}+\frac{1}{2\sigma_2^2}\int((x-\mu_1+\mu_1-\mu_2)^2)p_1(x)dx-\frac{1}{2}\\
&=log\frac{\sigma_2}{\sigma_1}+\frac{1}{2\sigma_2^2}[\int(x-\mu_1)^2p_1(x)dx+\int(\mu_1-\mu_2)^2p_1(x)dx+2\int(x-\mu_1)(\mu_1-\mu_2)]p_1(x)dx-\frac{1}{2}\\
&=log\frac{\sigma_2}{\sigma_1}+\frac{1}{2\sigma_2^2}[\int(x-\mu_1)^2p_1(x)dx+(\mu_1-\mu_2)^2]-\frac{1}{2}
\end{aligned}$$

$$=log\frac{\sigma_2}{\sigma_1}+\frac{\sigma_1^2+(\mu_1-\mu_2)^2}{2\sigma_2^2}-\frac{1}{2}$$

得出来了，我们假设N2是一个正态分布，也就是说$\mu_2=0,\sigma_2^2=1$那么N1长成什么样子能够让KL散度尽可能地小呢?

也就是说$KL(\mu_1,\sigma_1)=-log\sigma_1+\frac{\sigma_1^2+\mu_1^2}{2}-\frac{1}{2}$。

我们用“肉眼”看一下就能猜测到当$\mu_1=0,\sigma_1=1$时，KL散度最小。从公式中可以看出，如果μ_1偏离了0，那么KL散度一定会变大。而方差的变化则有些不同：

当σ_1大于1时，$\frac{1}{2}\sigma_1^2$将越变越大，而$-log\sigma_1$越变越小；

当σ_1小于1时，$\frac{1}{2}\sigma_1^2$将越变越小，而$-log\sigma_1$越变越大；

那么哪边的力量更强大呢？我们可以作图出来：

```
import numpy as np
import matplotlib.pyplot as plt
x = np.linspace(0.5,2,100)
y = -np.log(x)+x*x/2-0.5
plt.plot(x,y)
plt.show()
```

从图中可以看出