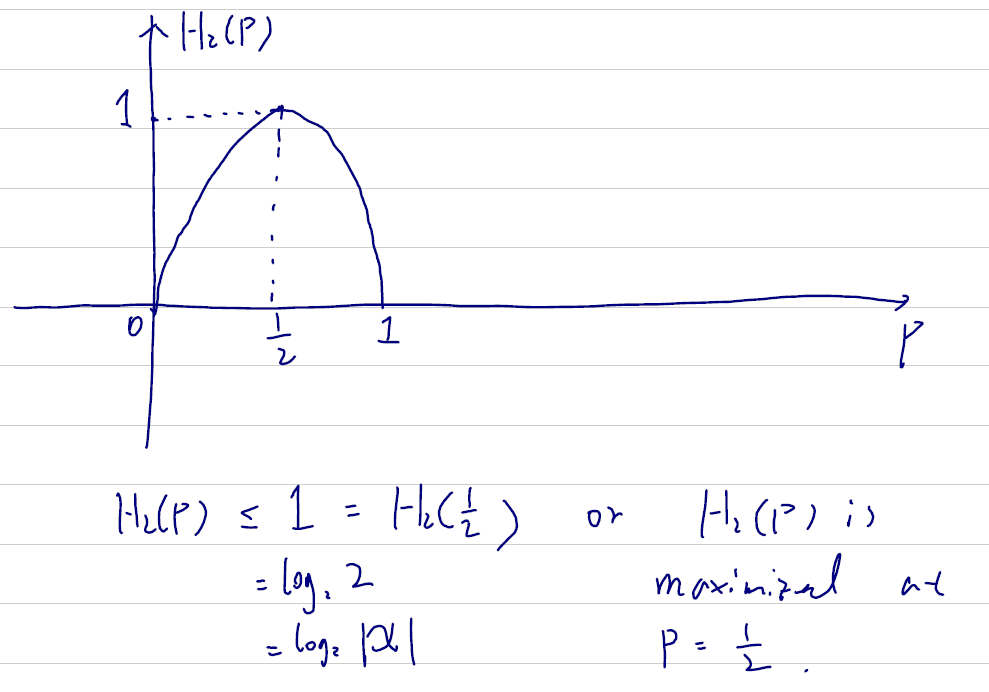
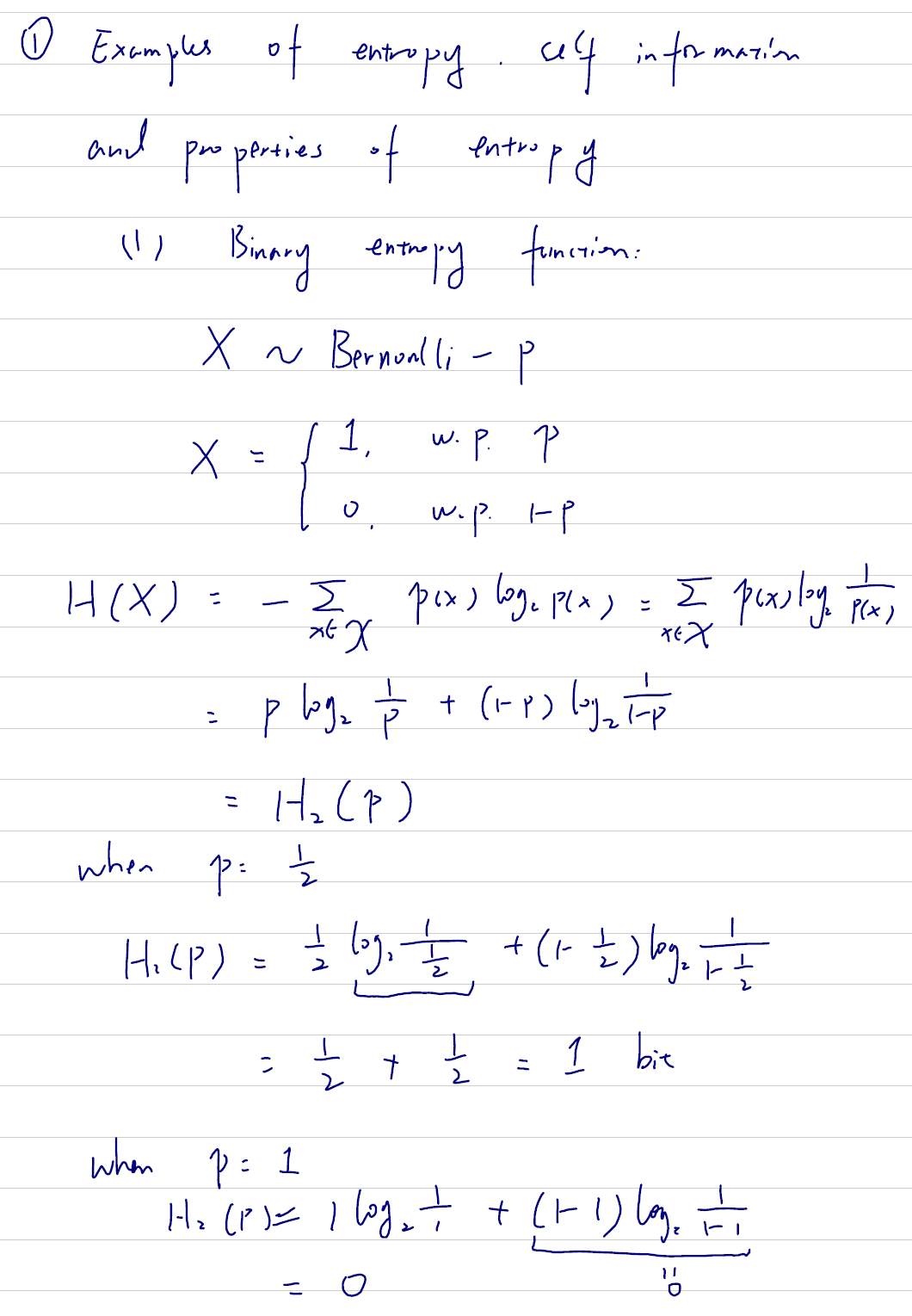
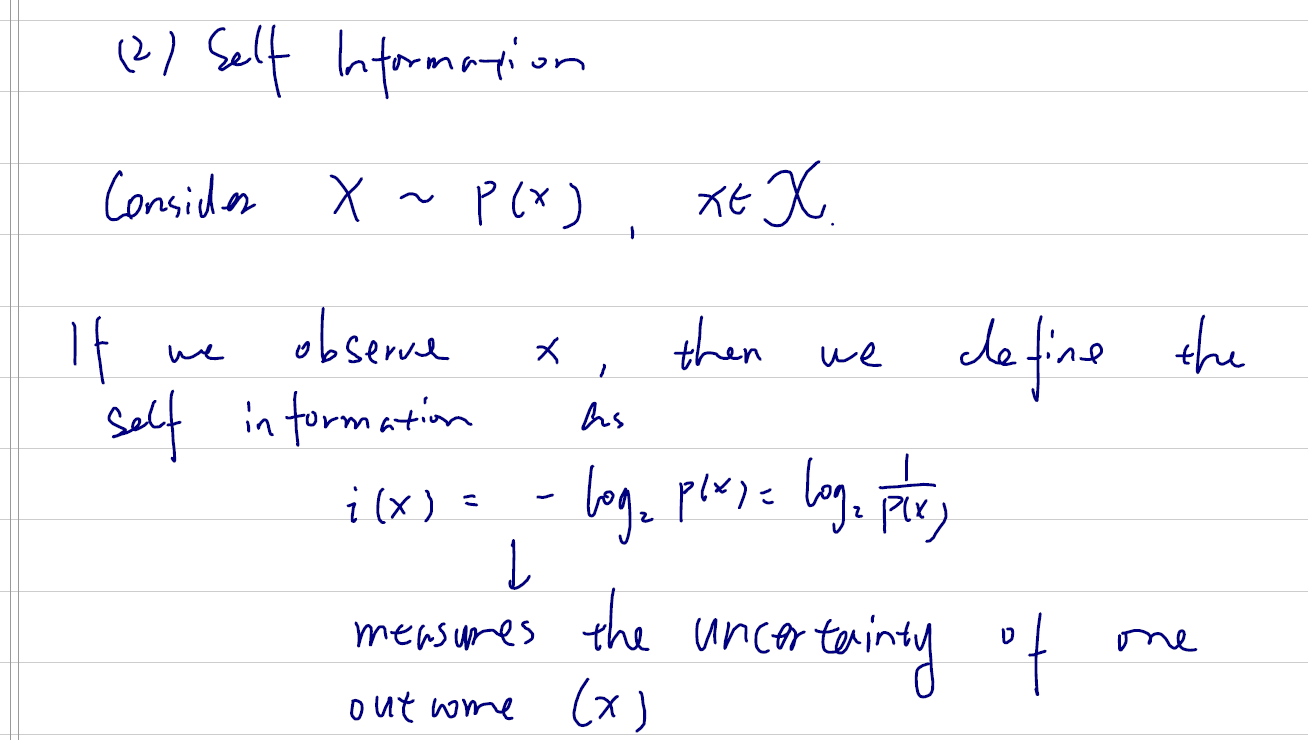


lecture 1 note and review slides

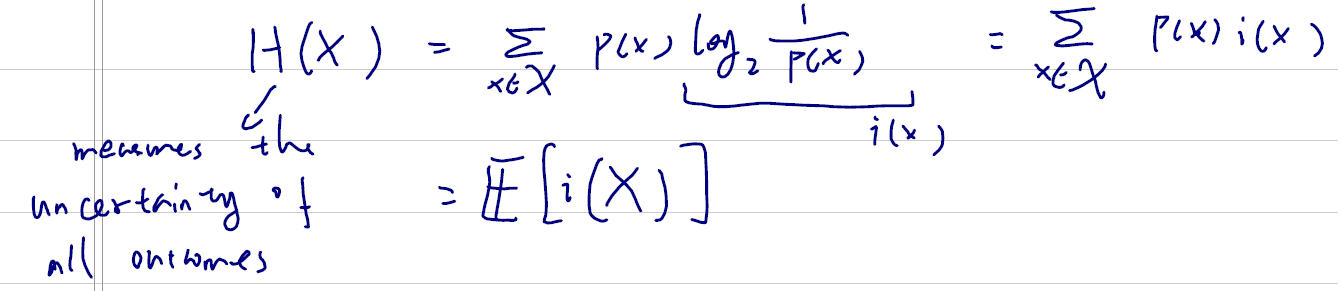
H(x) 本质上是指混乱程度 值越大不确定性越高

H2(x) 可以用来代表要用几个bit来表示所有信息

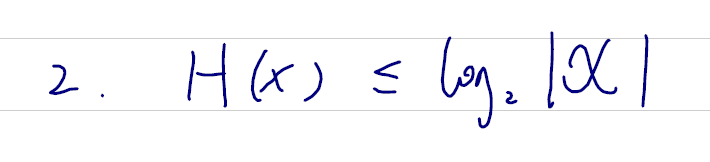
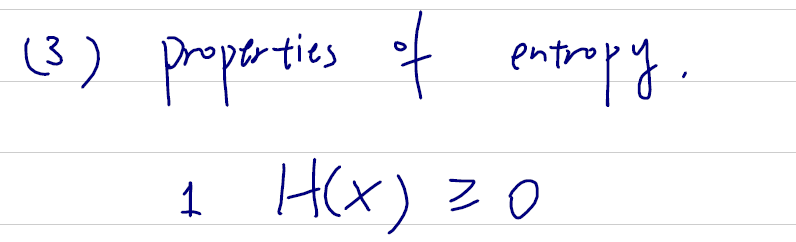


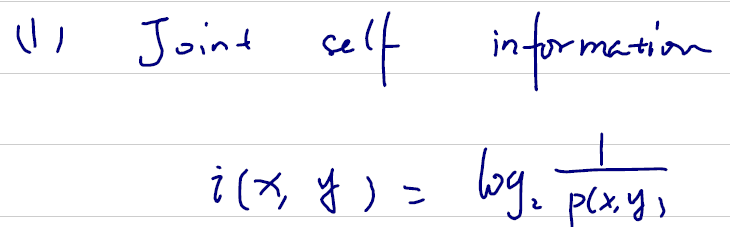
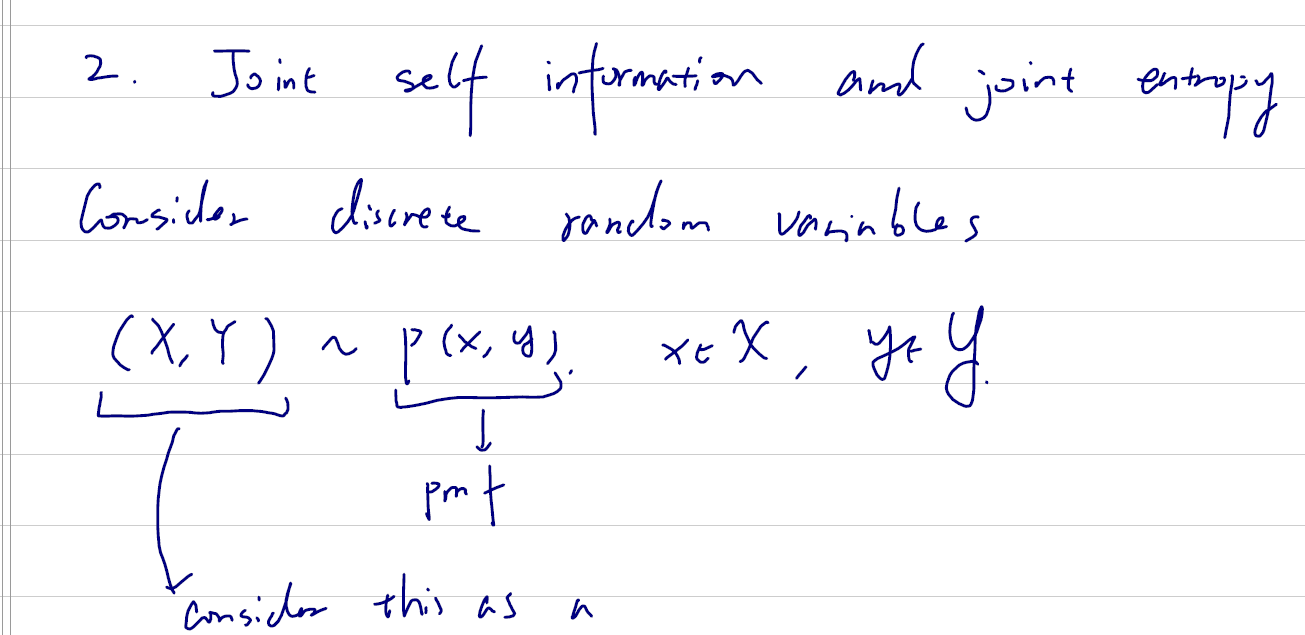


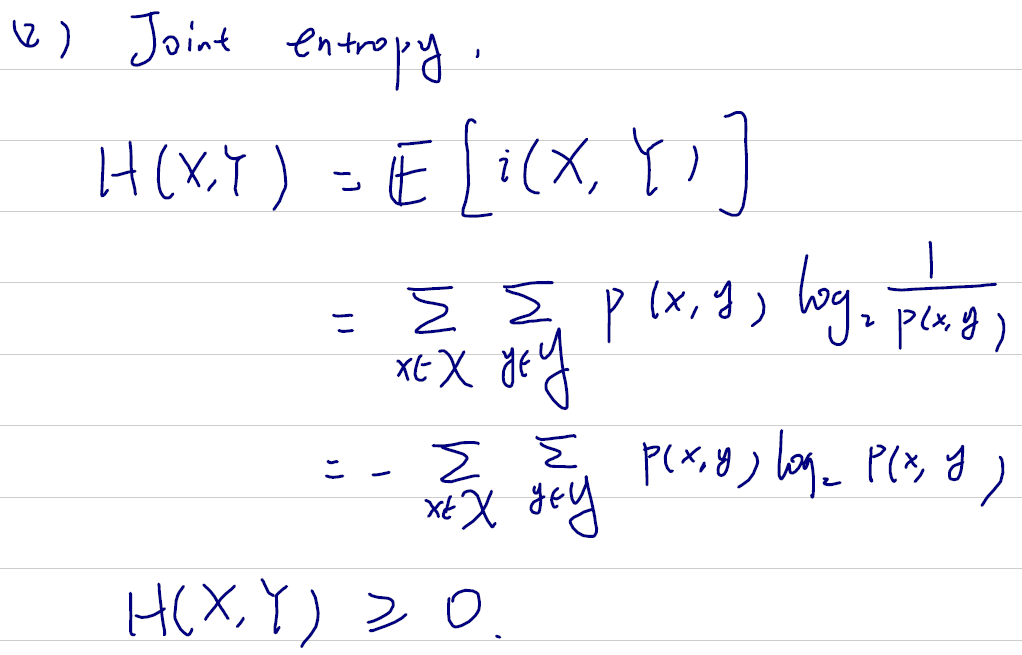
self information一个点的值越大 信息越多 因为低概率事件发生了 会带来很多信息

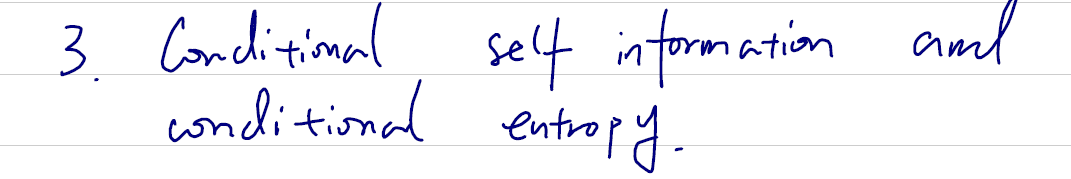


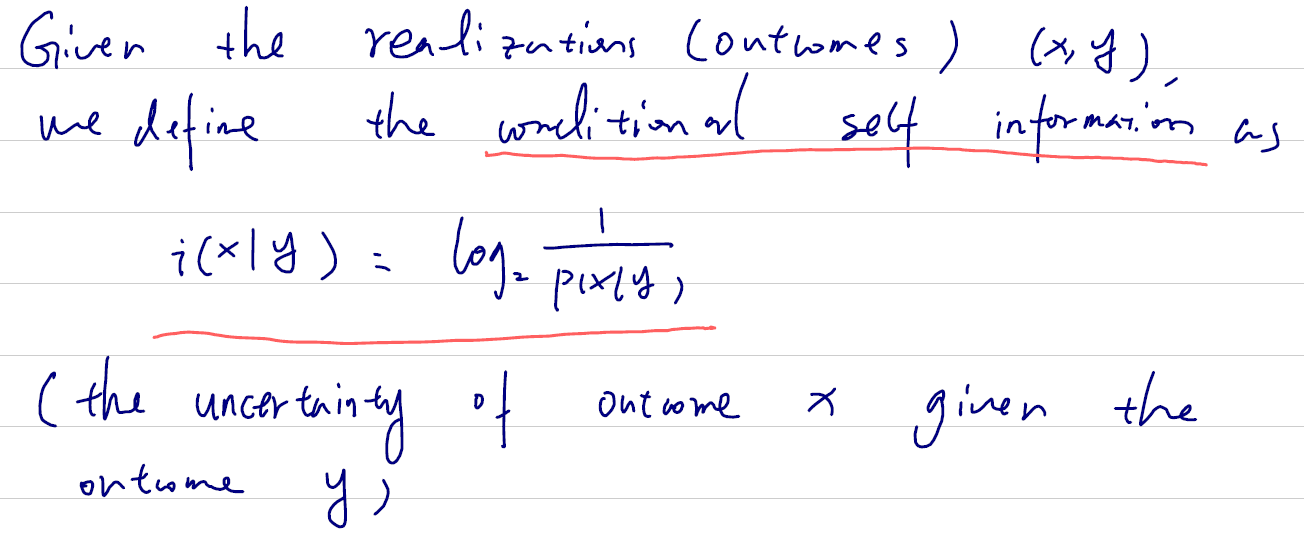
这里也说明H的概念是建立在一个 期望E[]上的

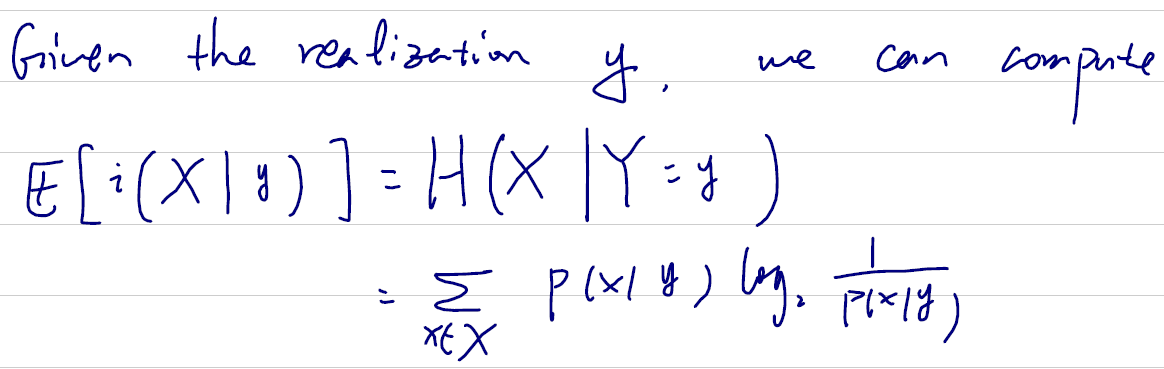


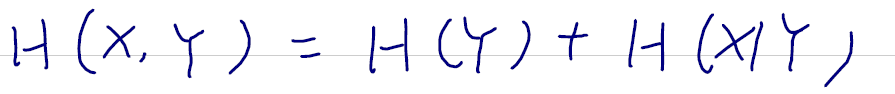


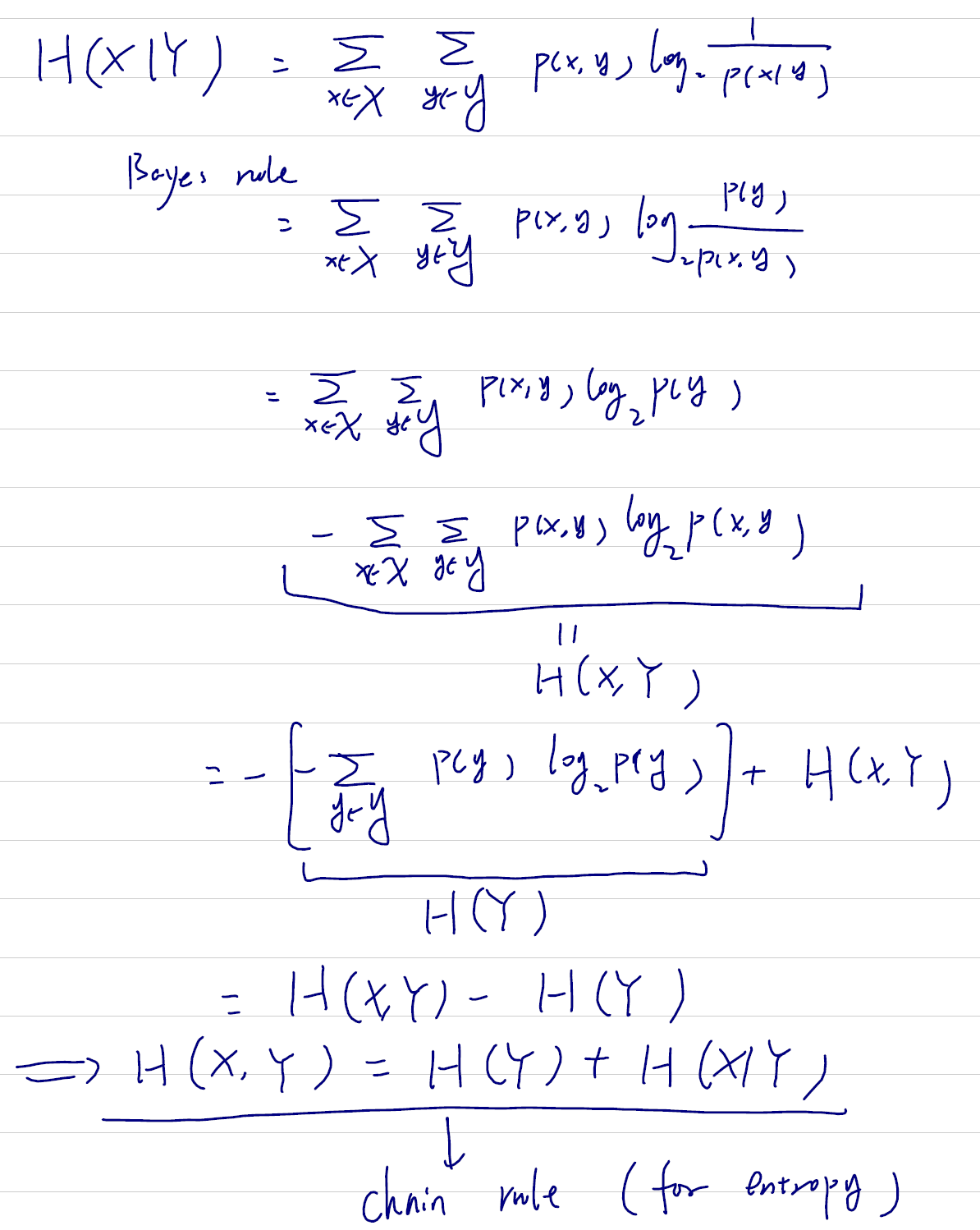


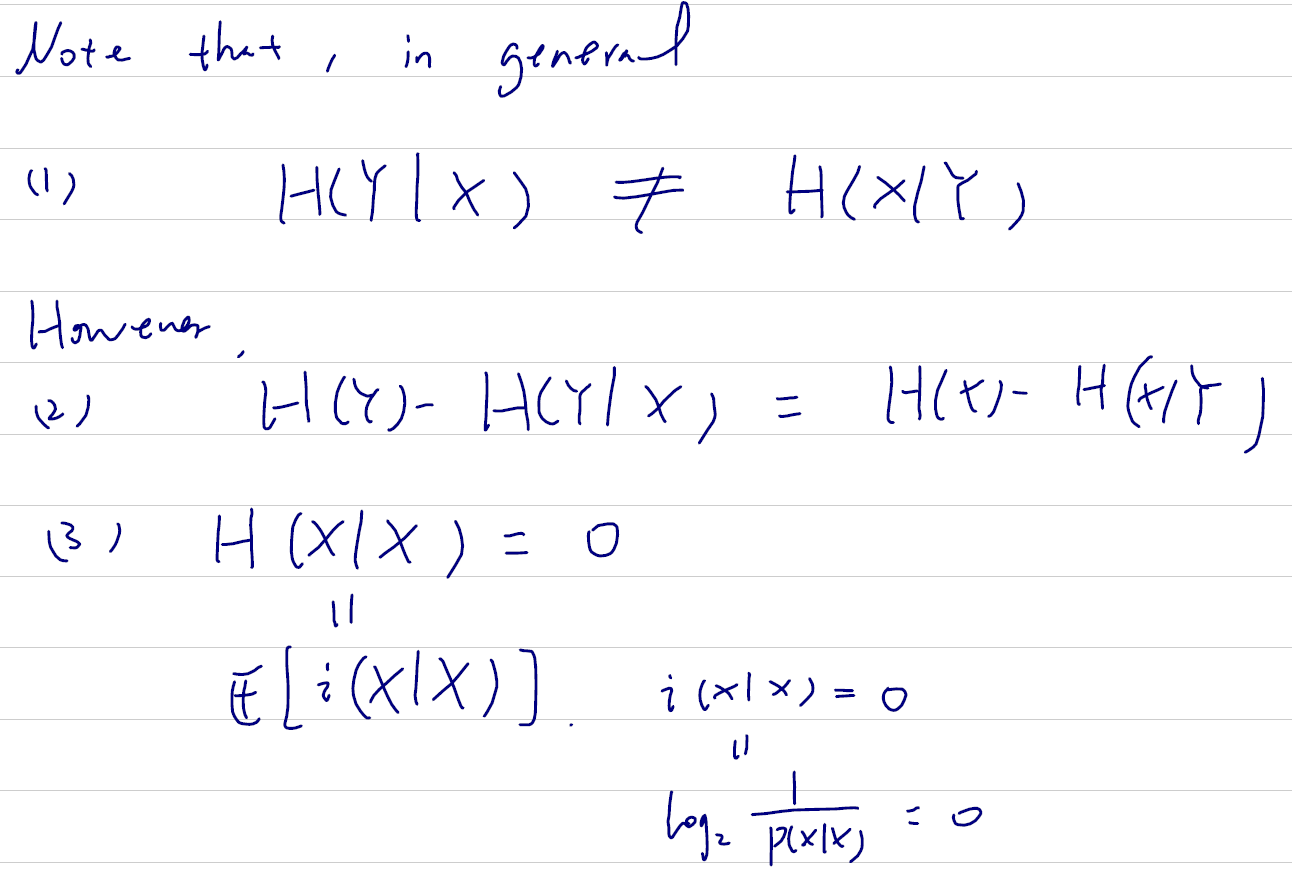


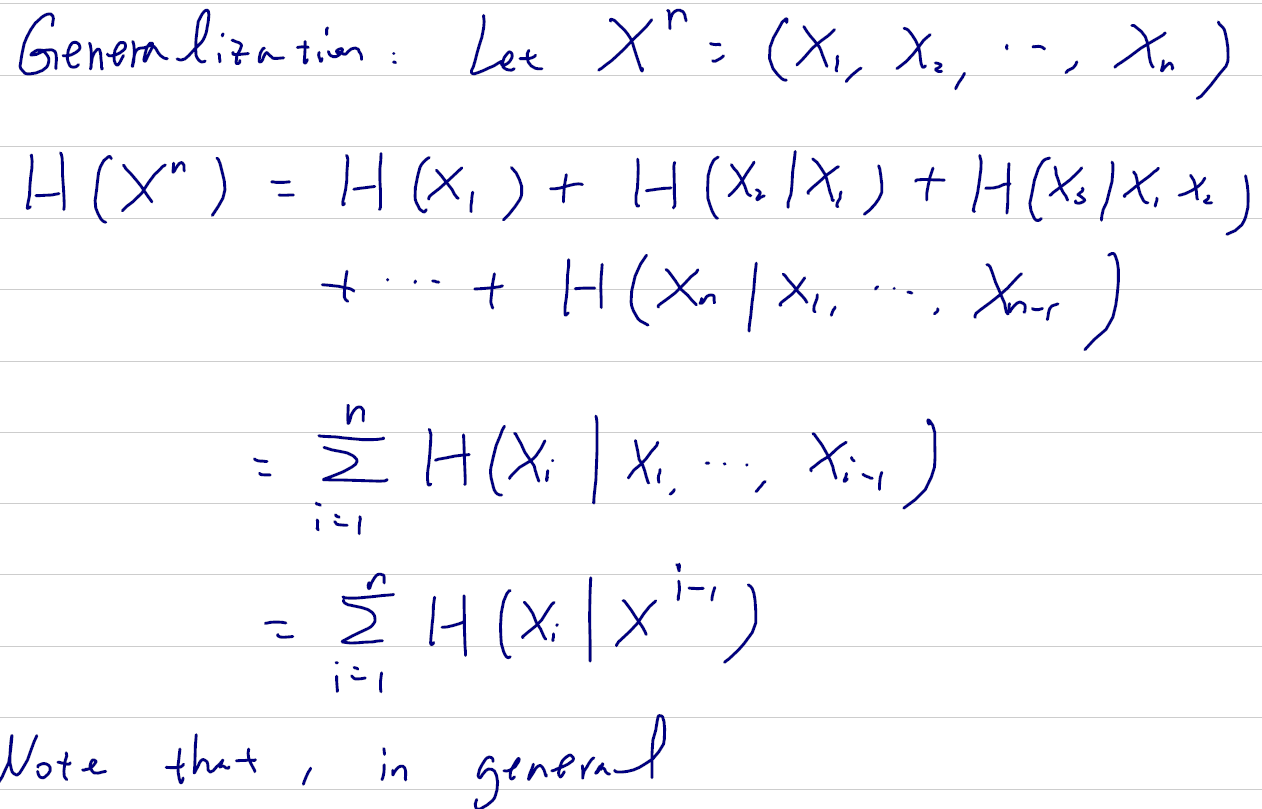










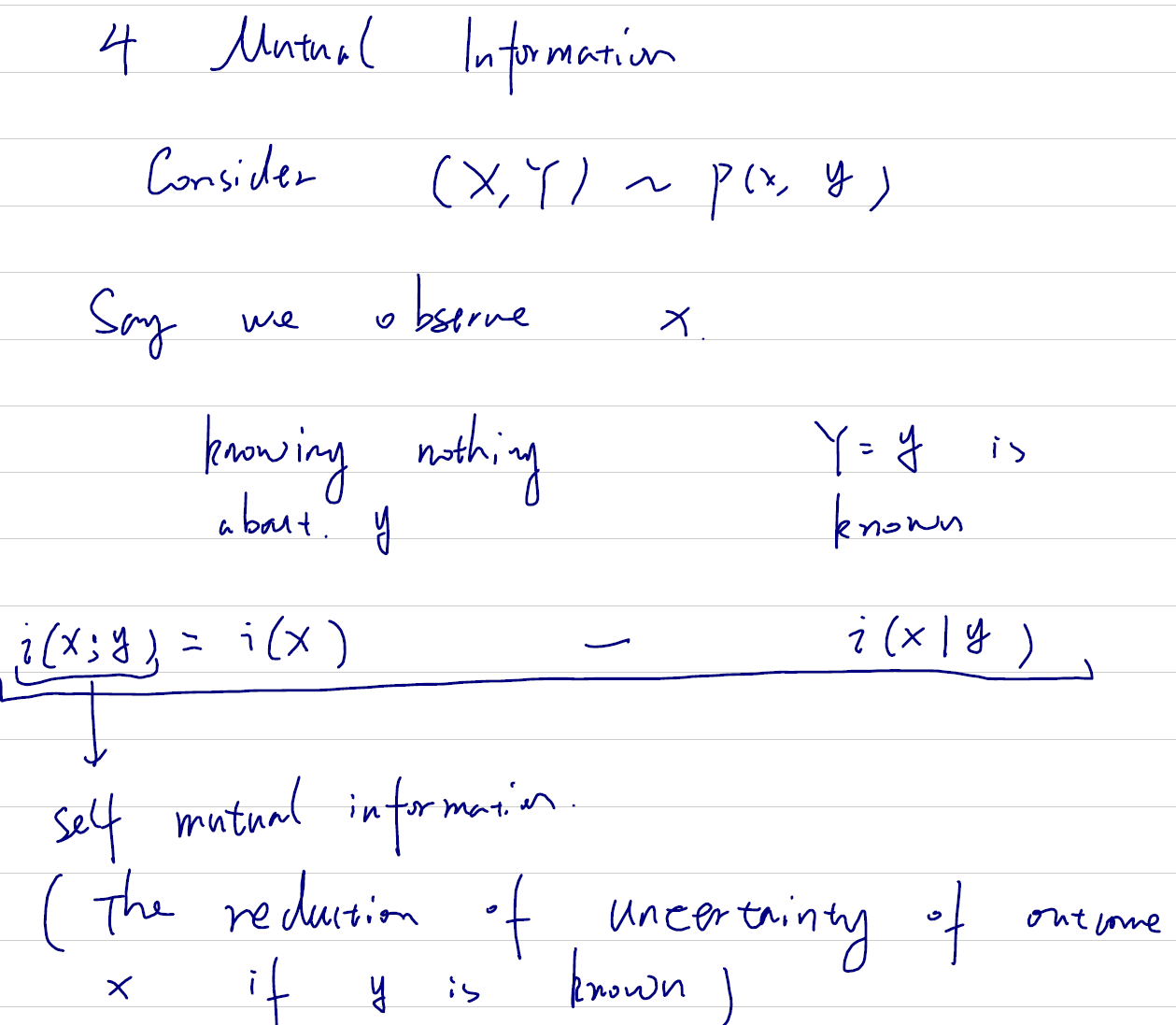


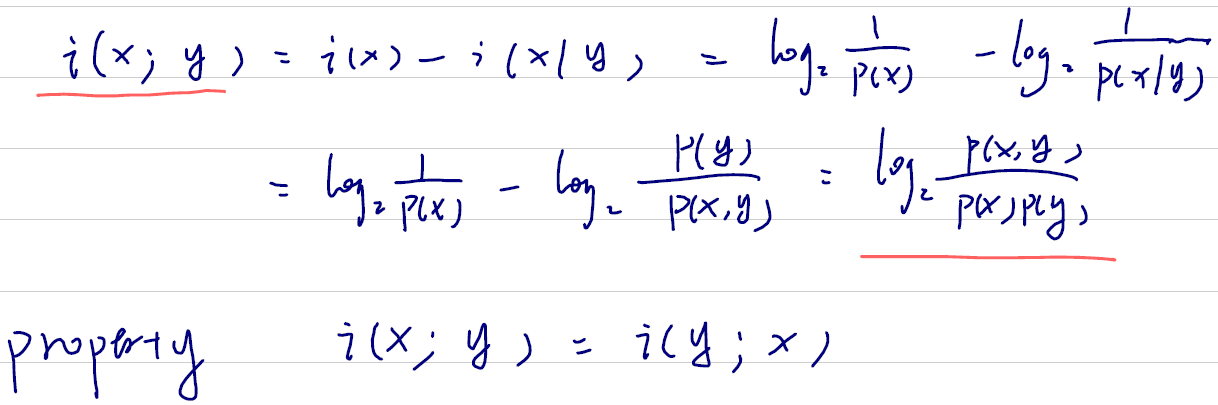
**Mutual information 和 entropy 类似都是 期望**

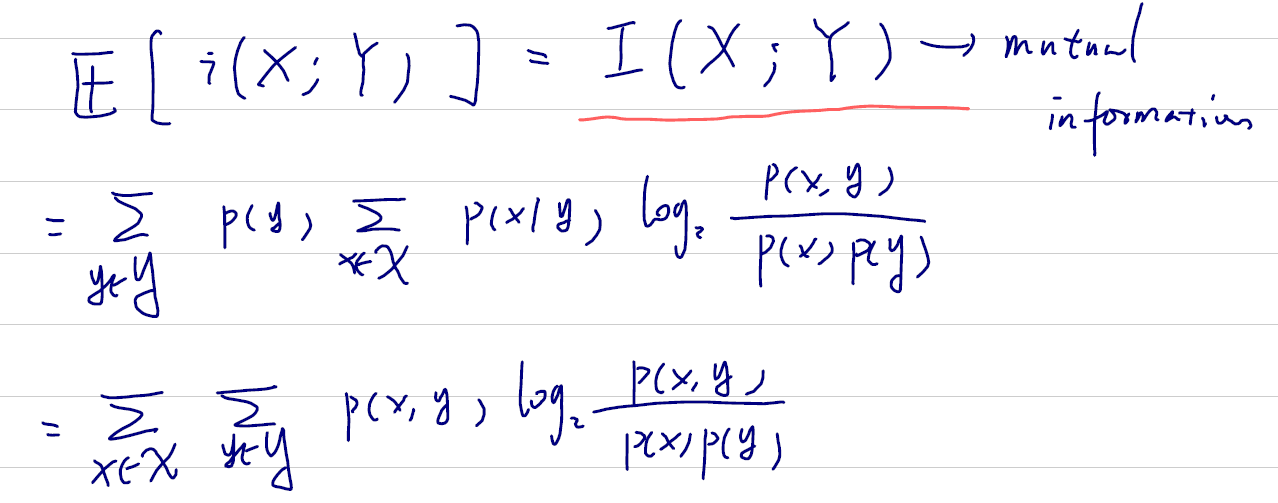
**entropy是uncertainty of one source (X)**

**mutual information is a reduction of uncertainties (两个或多个变量之间减少了多少uncertainty)** mutual entropy 计算的是几个事件同时发生时比分别发生时减少了多少信息

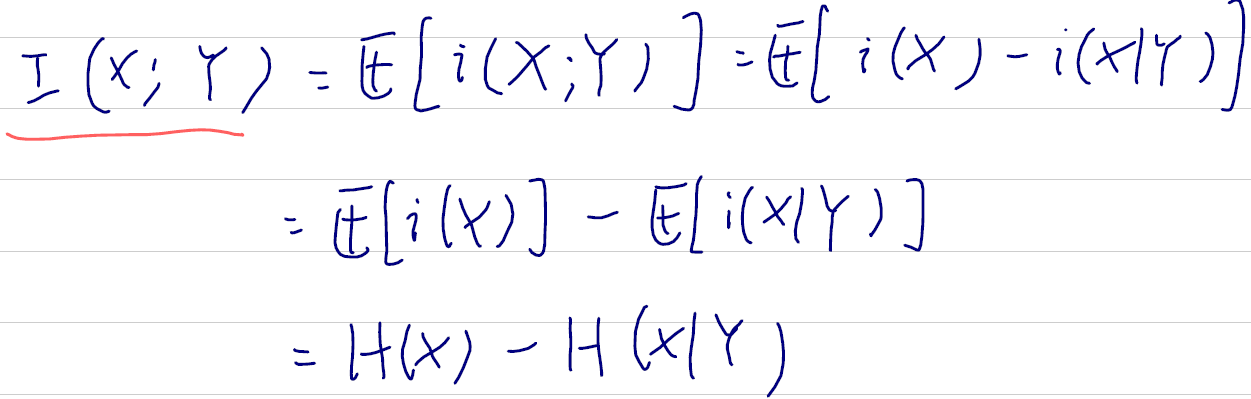
I的值越小说明事件之间越无关

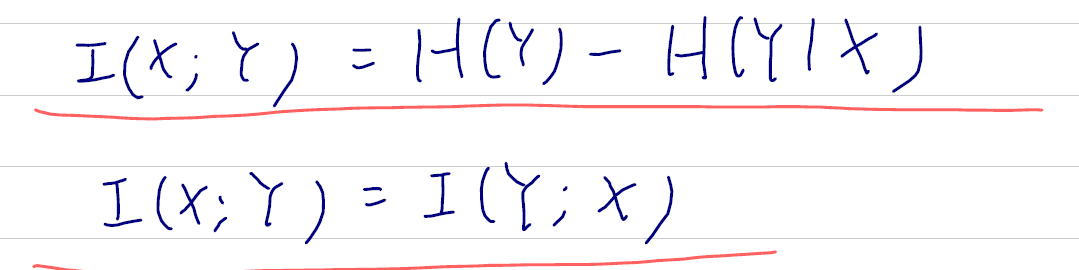


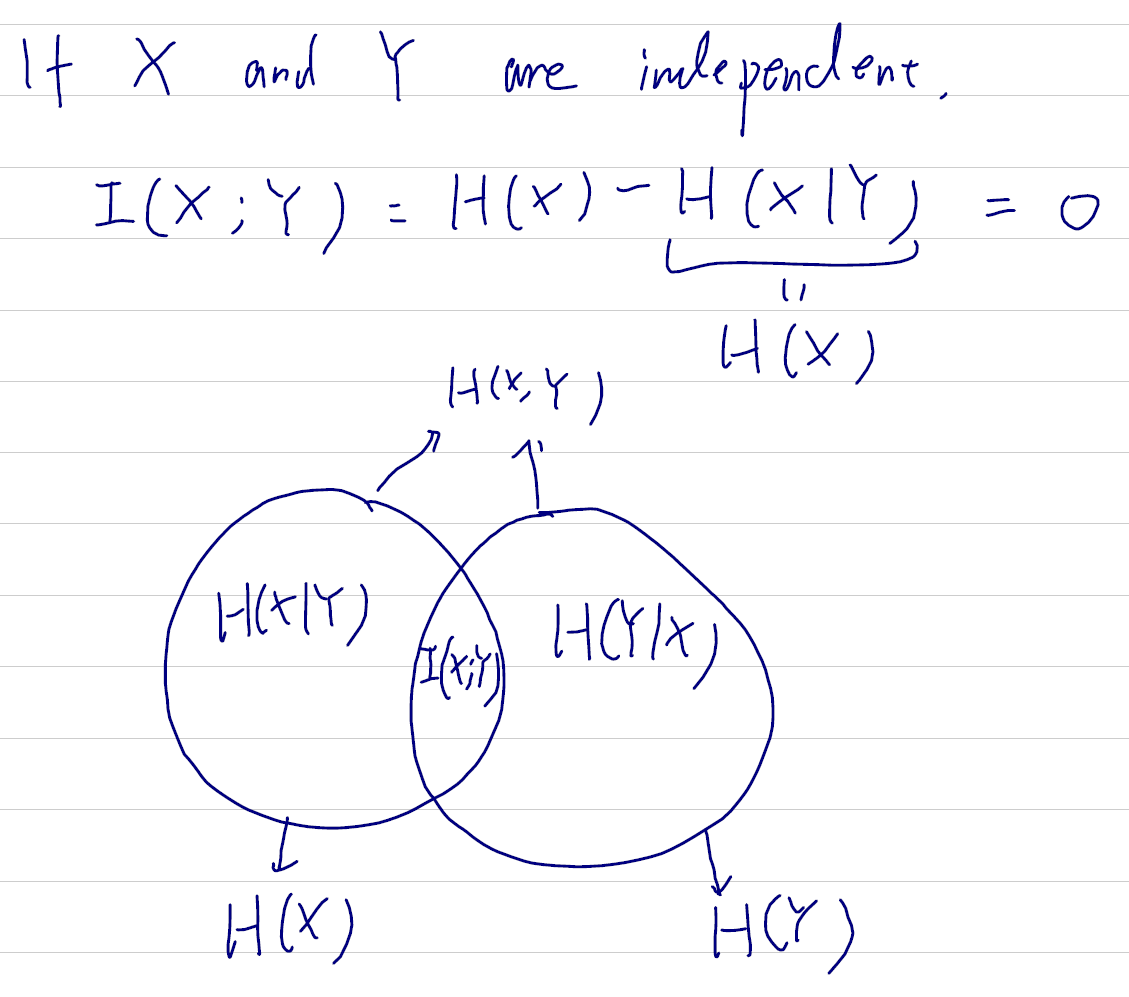


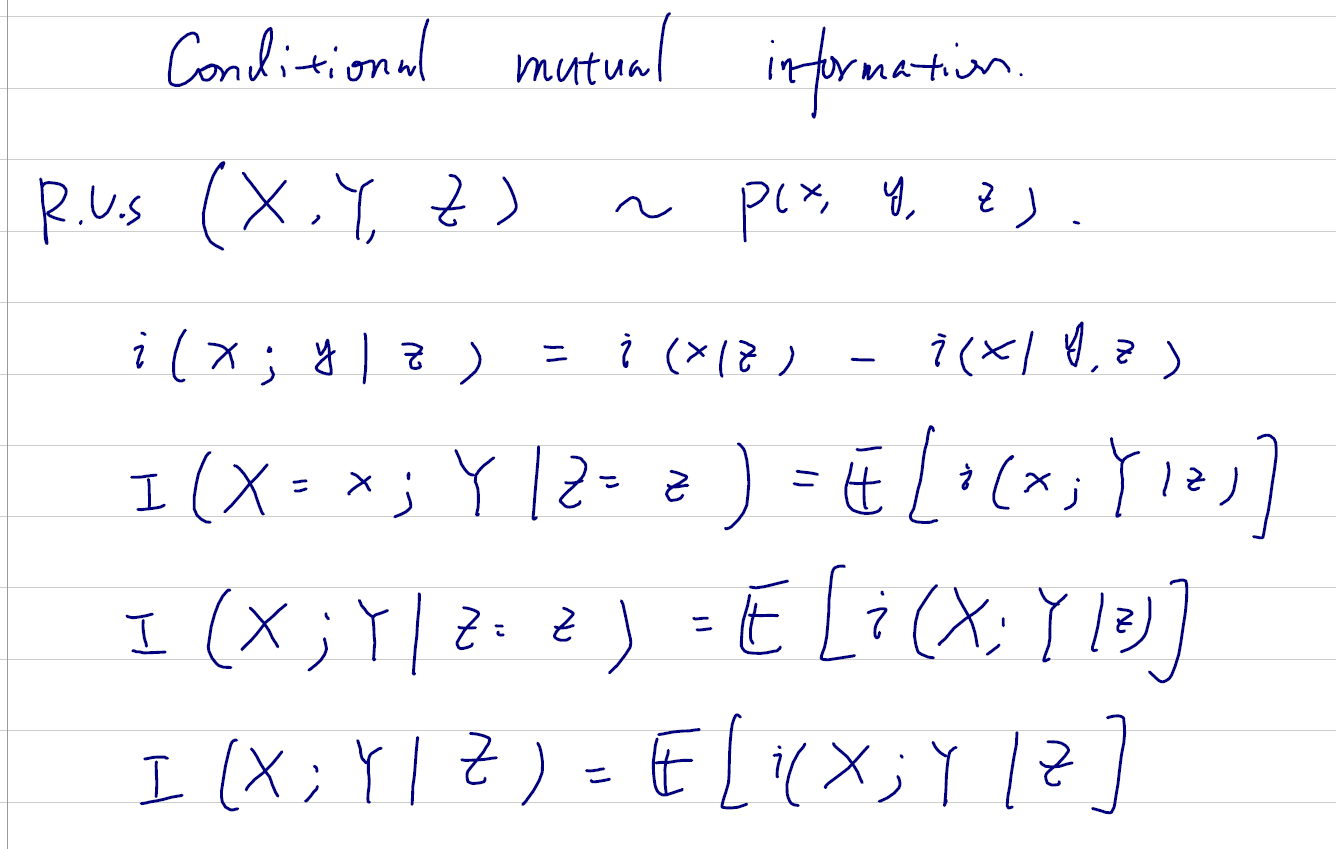


**mutual information是 self-mutual information的期望**



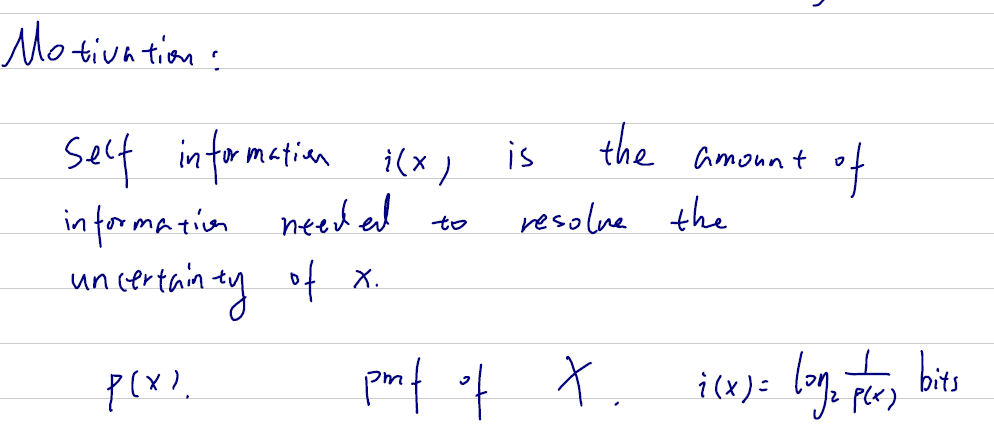




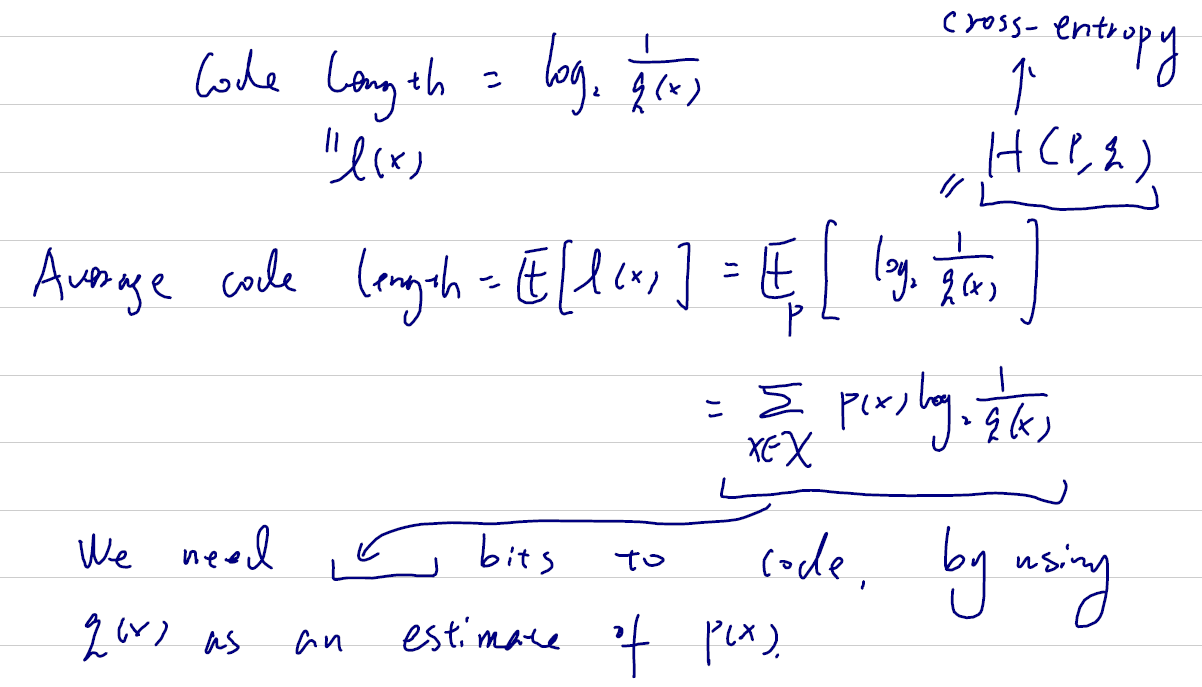


**lecture 2 note, same to lec2 slides**

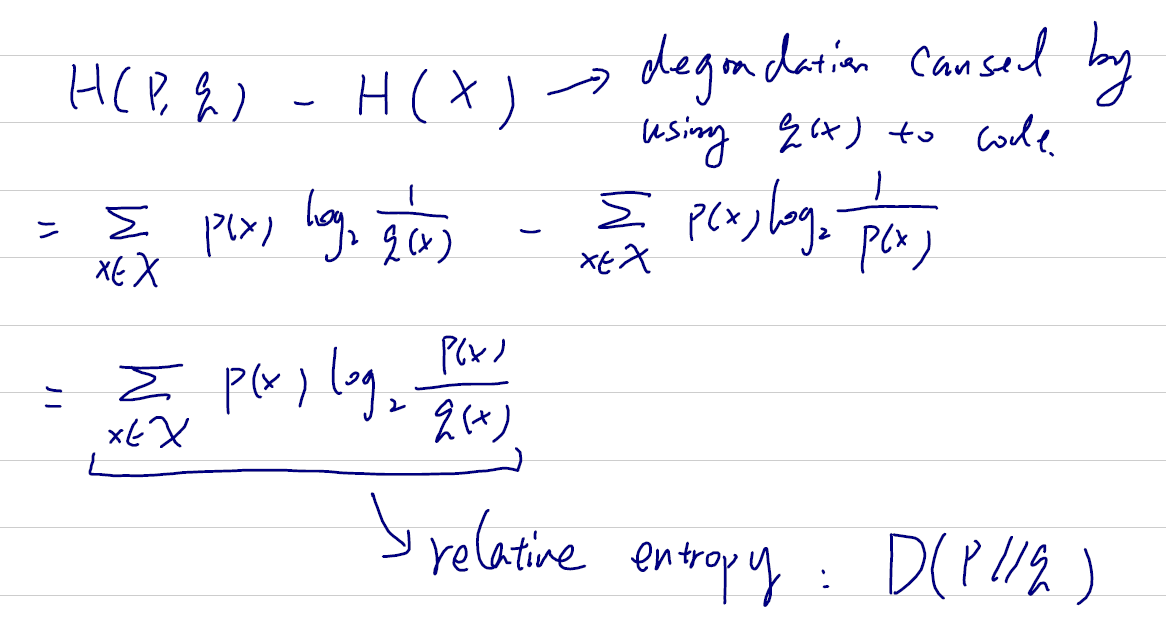
**relative entropy (KL divergence):**

 **i(x)代表x事件的信息要多少位来表示**

 **如果我们不知道p(x),让q(x)作p(x)的估计 (不一定准)**

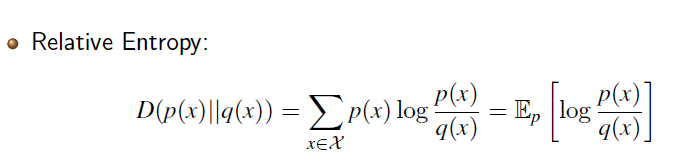
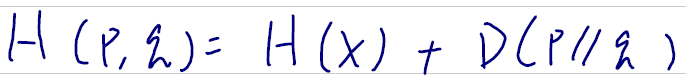


**这样得到的结果不一定准 于是我们要用一种方法或者标准来评判q(x)对p(x)的估计有多准**



**H(p,q) 是cross entropy 上方提到了，就是在正确的概率分布( p(x) )下**

**当前估计 q(x) 的信息(期望值)**

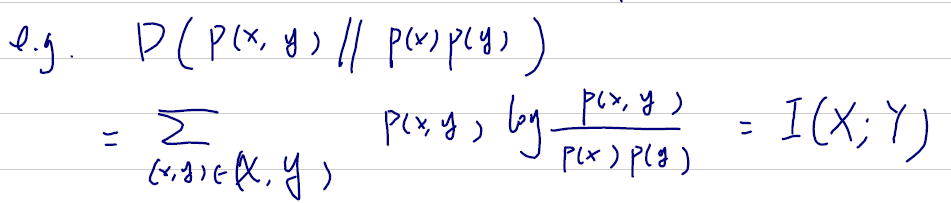


**to minimize cross-entropy we need to minimize D (relative entropy, KL-divergence)**

**H(X)是最优情况下所含的信息 它所需的bit数是最小的**

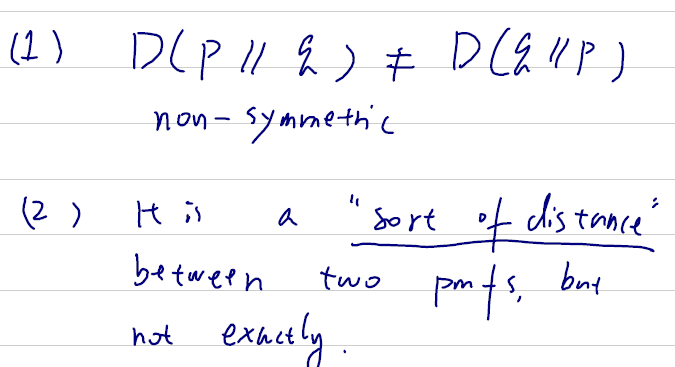
**在一个非最优的估计下(q(x)), 会导致信息的冗余 至于原因可以类比Haffmen编码**

**D(p//q)就是评价冗余信息的大小 值越大说明越不相近(相关)**

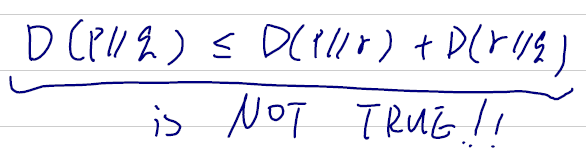


**mutual information is a particular case for relative entropy**

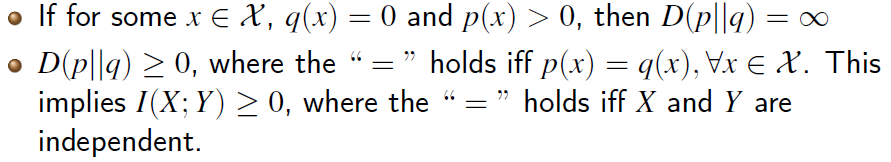
**relative entropy properties**

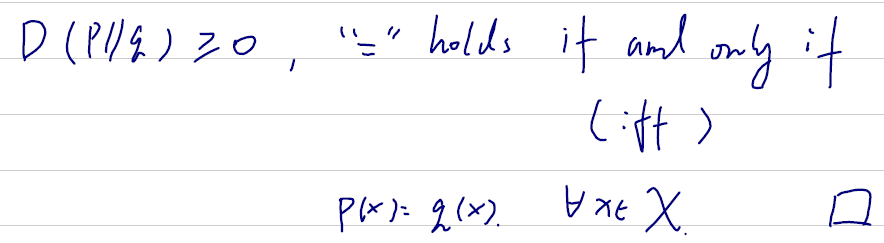
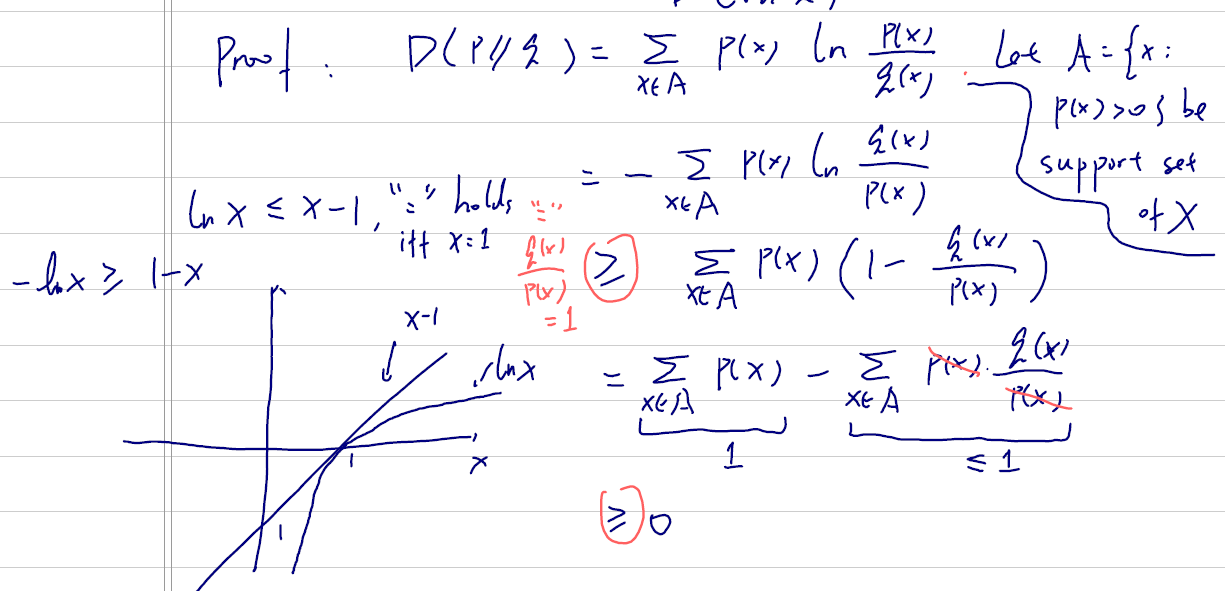


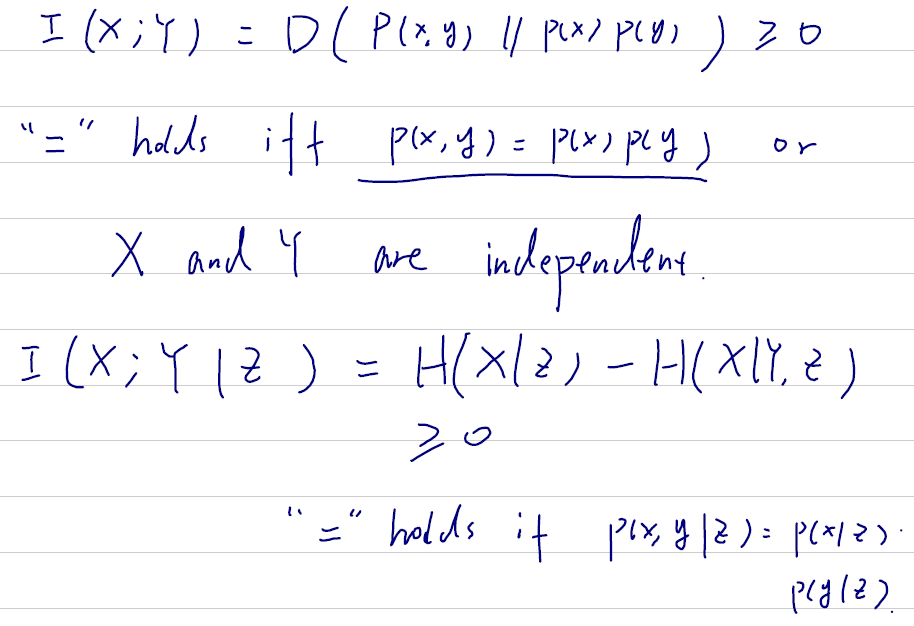
**Why not exactly?**



**(3) & (4):**

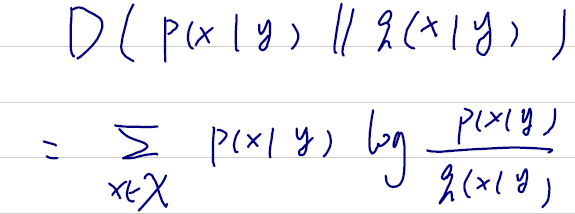




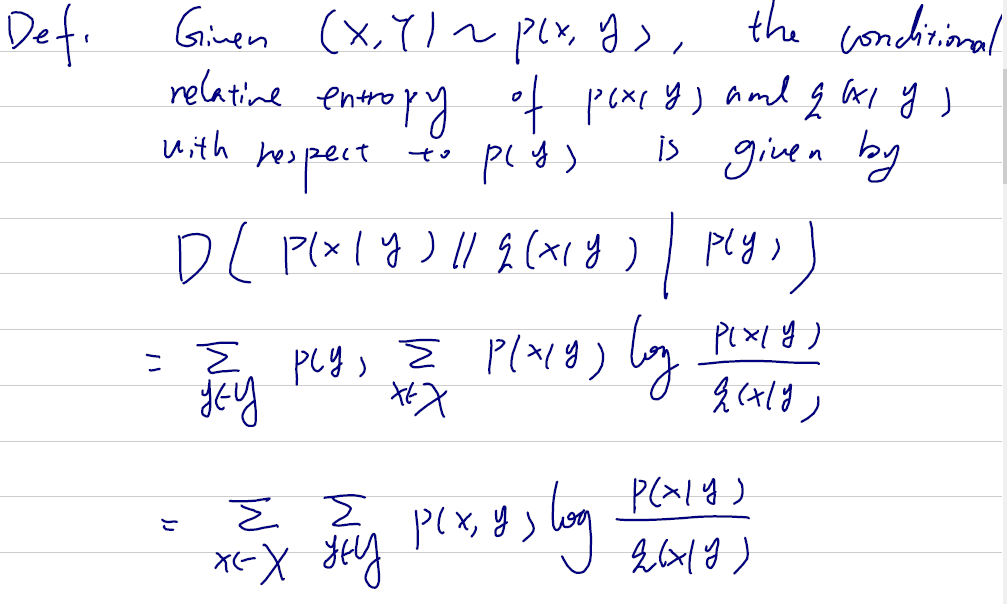


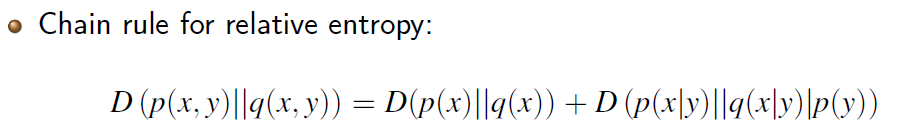
**conditional relative entropy:**

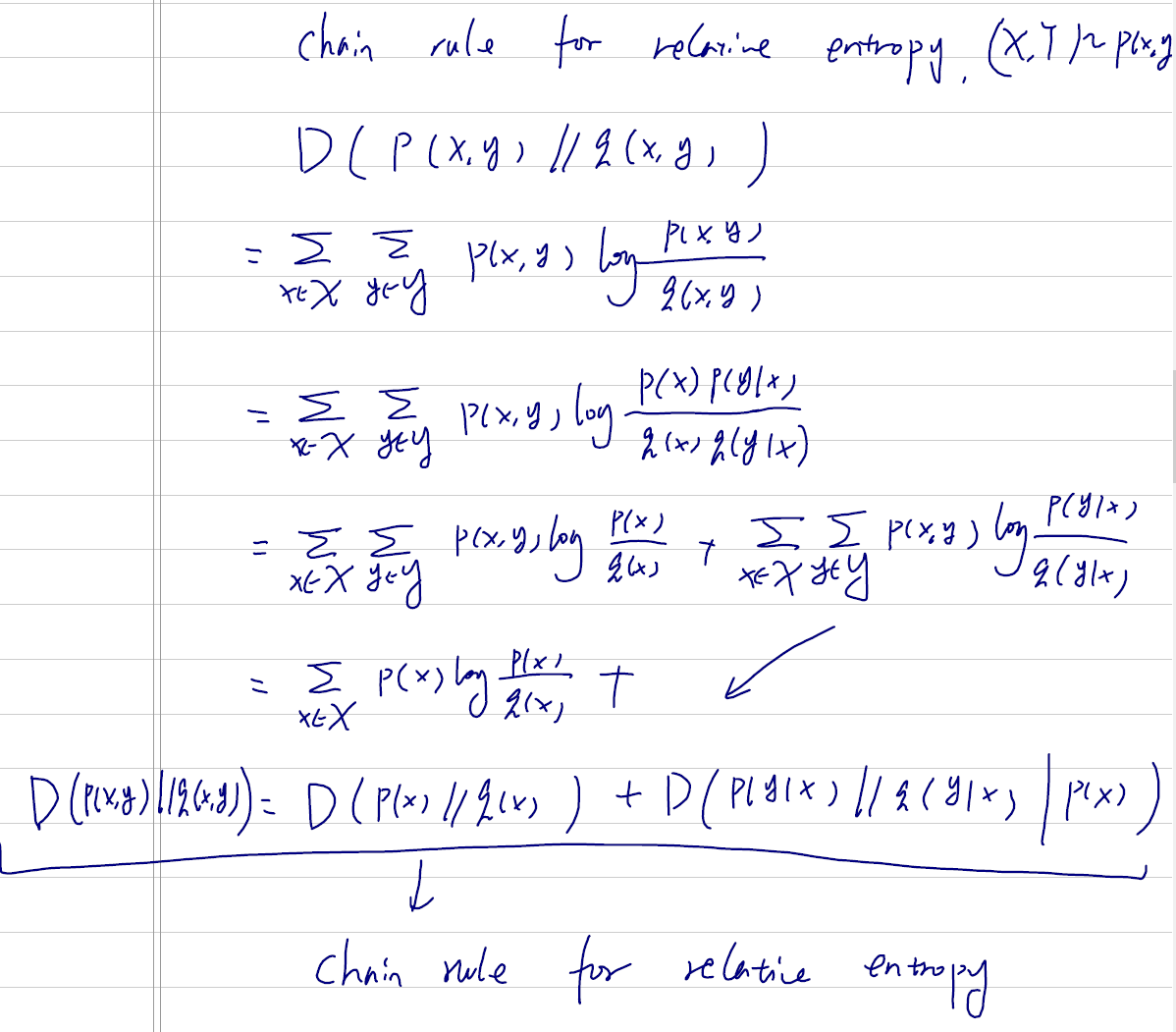
**单一 y :**

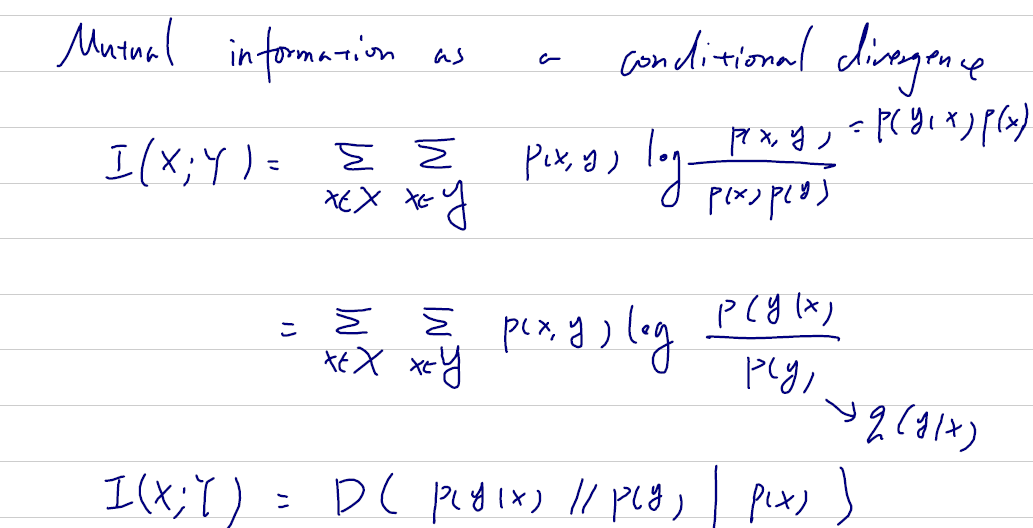
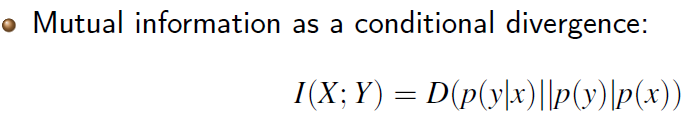


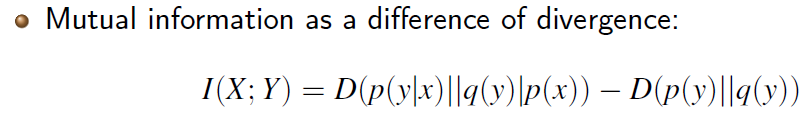
**y 是某个分布时：**

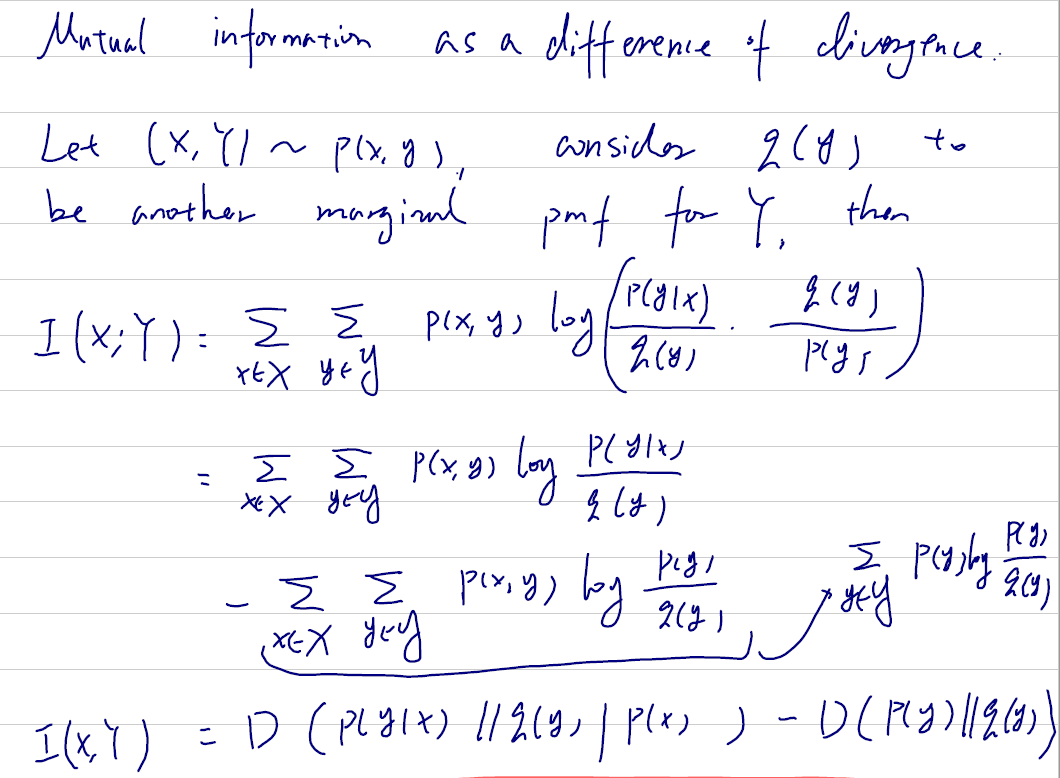


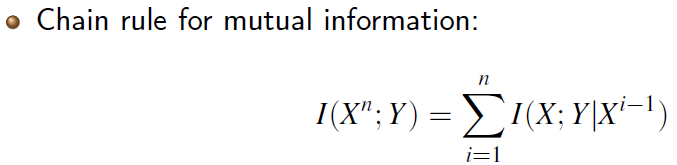


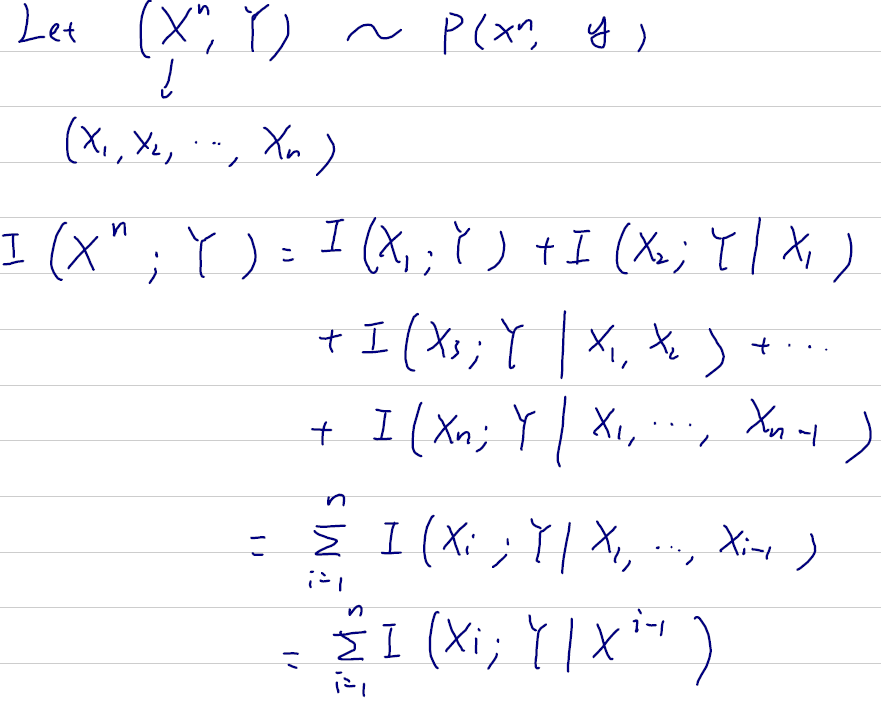


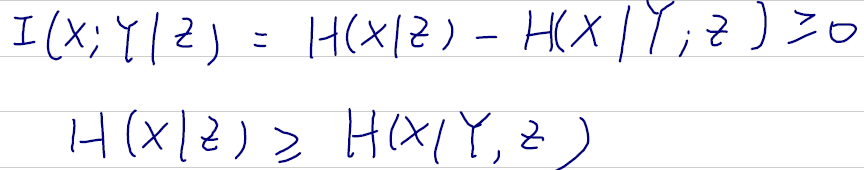


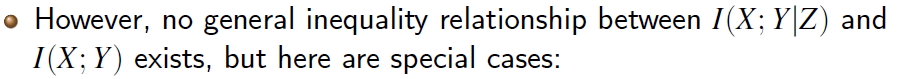


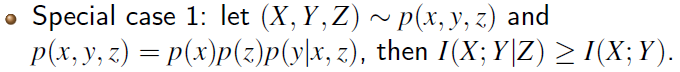


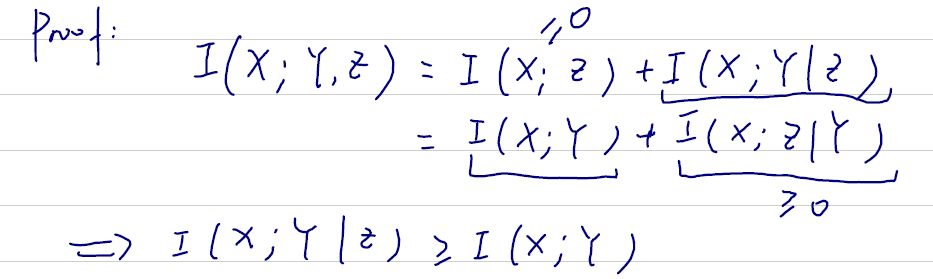




more info will reduce uncertainty





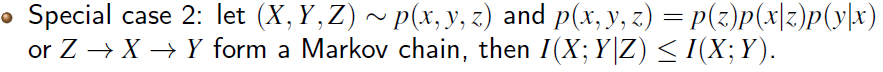


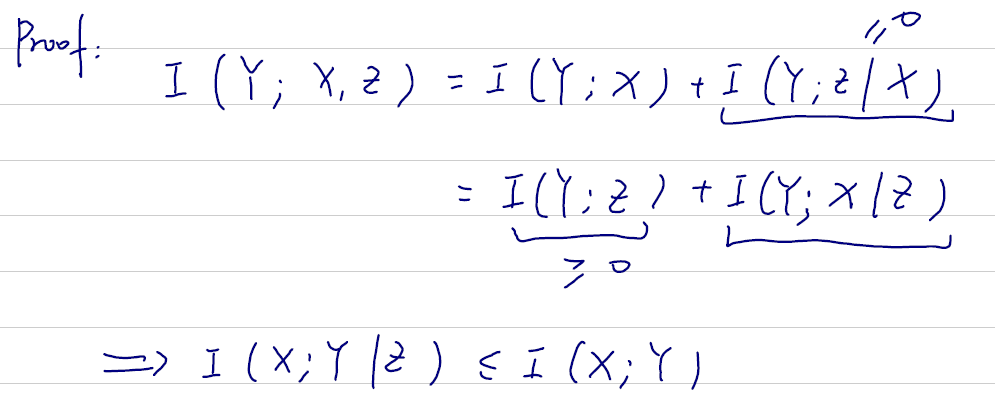
mutual entropy 计算的是几个事件同时发生时比分别发生时减少了多少信息

I的值越小说明事件之间越无关

这里x,z都会影响y 给定z时uncertainty变小

所以有z时减少的信息多





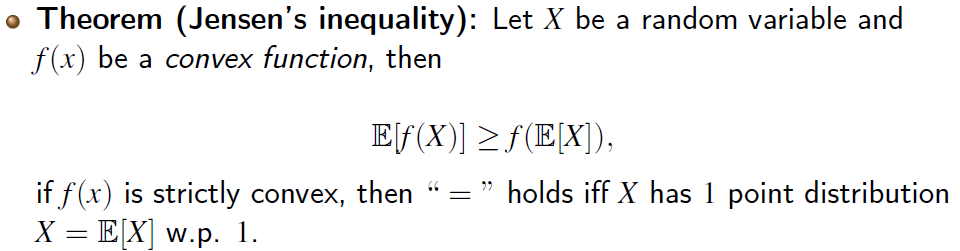
证明的第一行等于0是因为x给定时 z和y是independent的 z改变并不会改变y

应该有>=的证明 因为y只与x有关

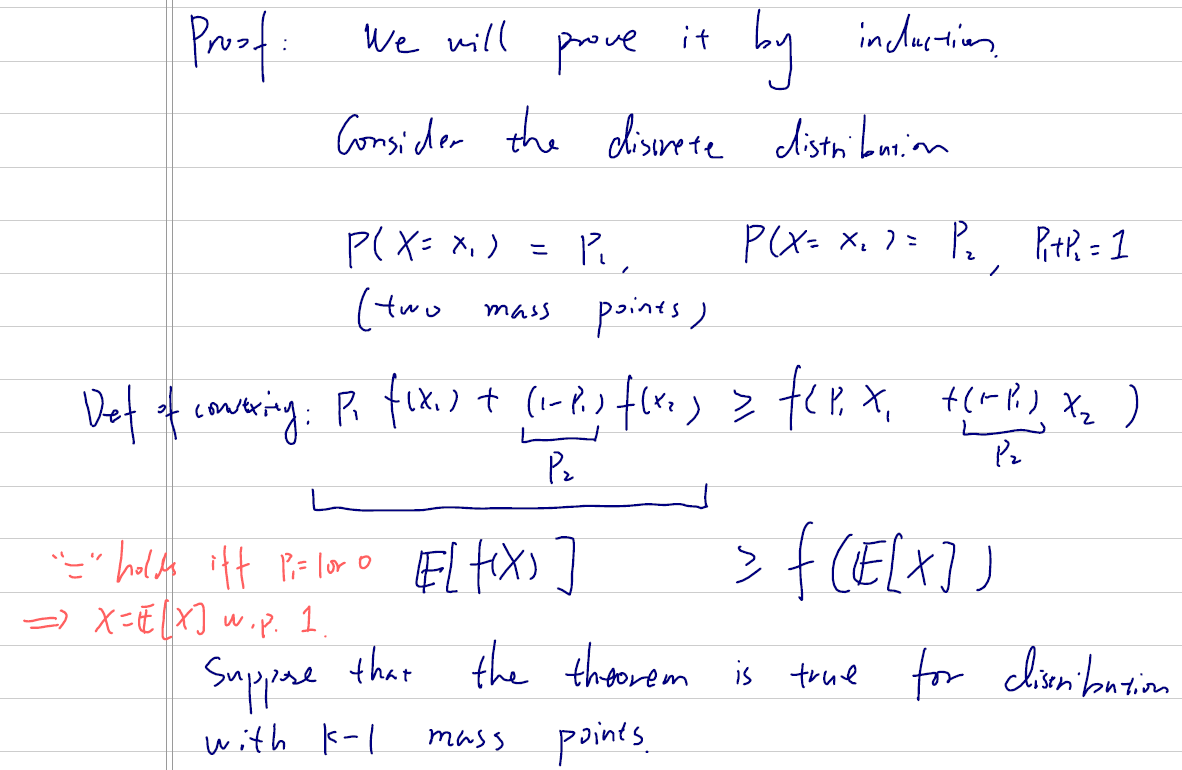
lecture 3 note and slides

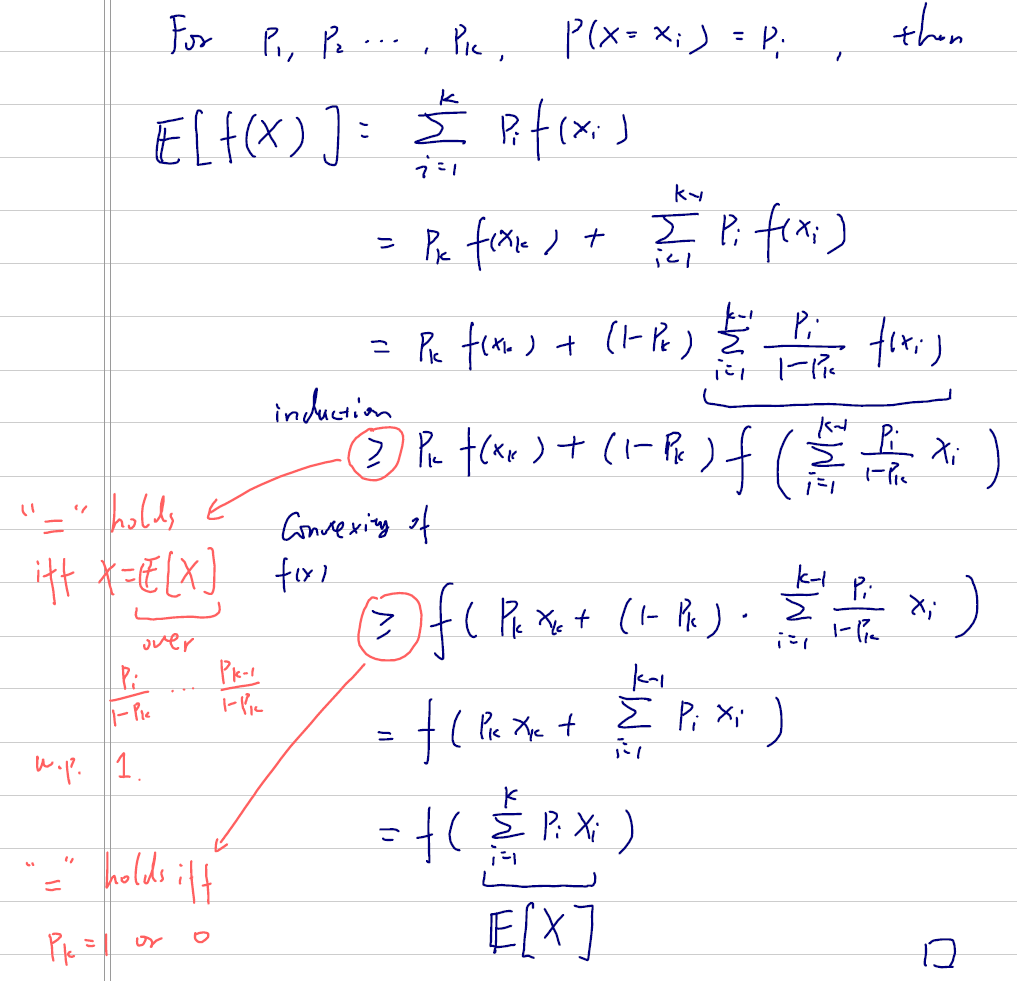


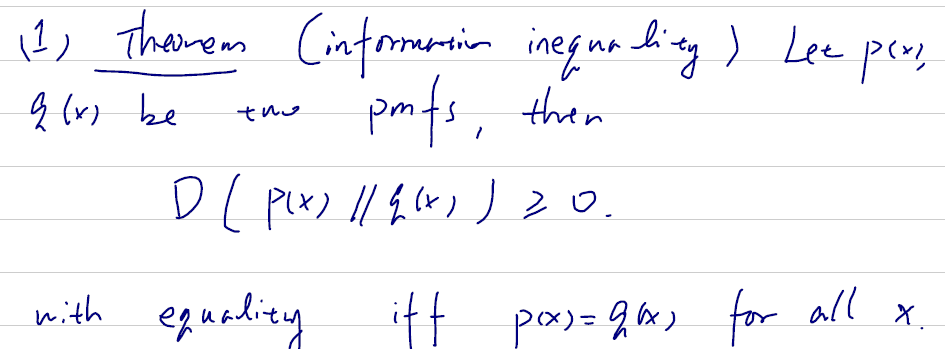
Jensen’s inequality



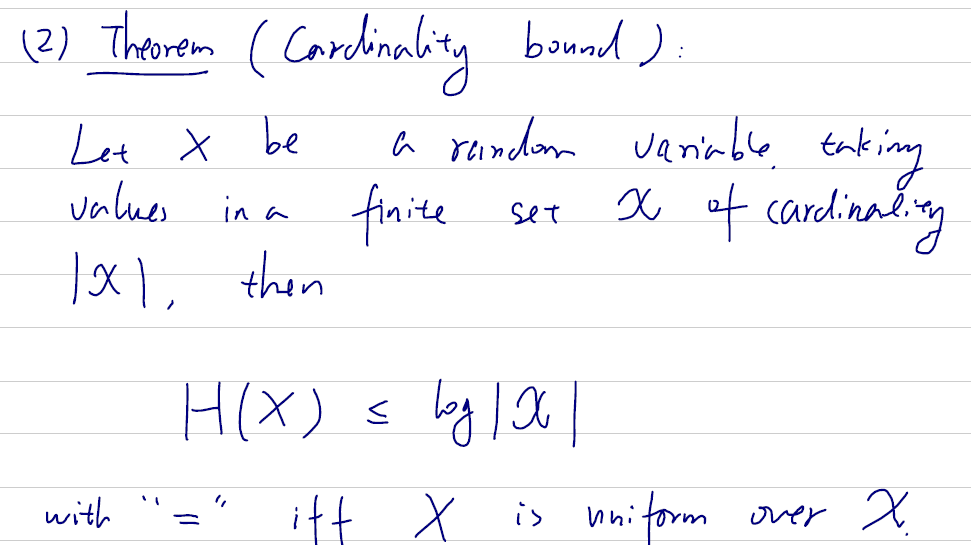
inductin:归纳法证明

离散情况下成立

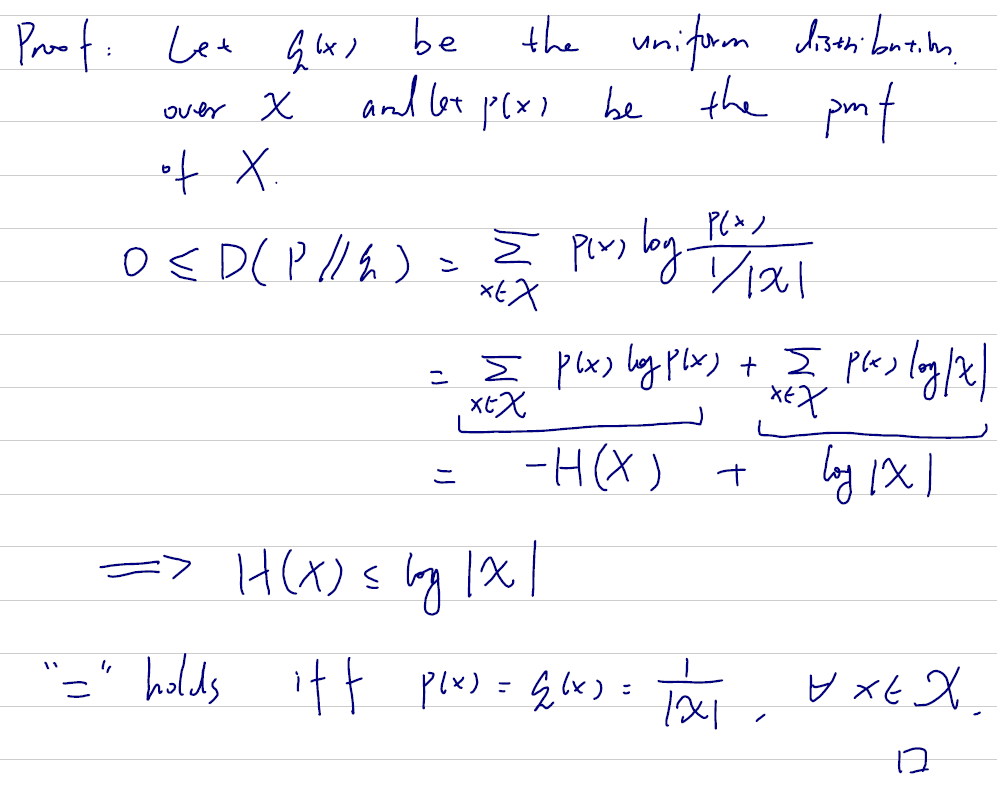


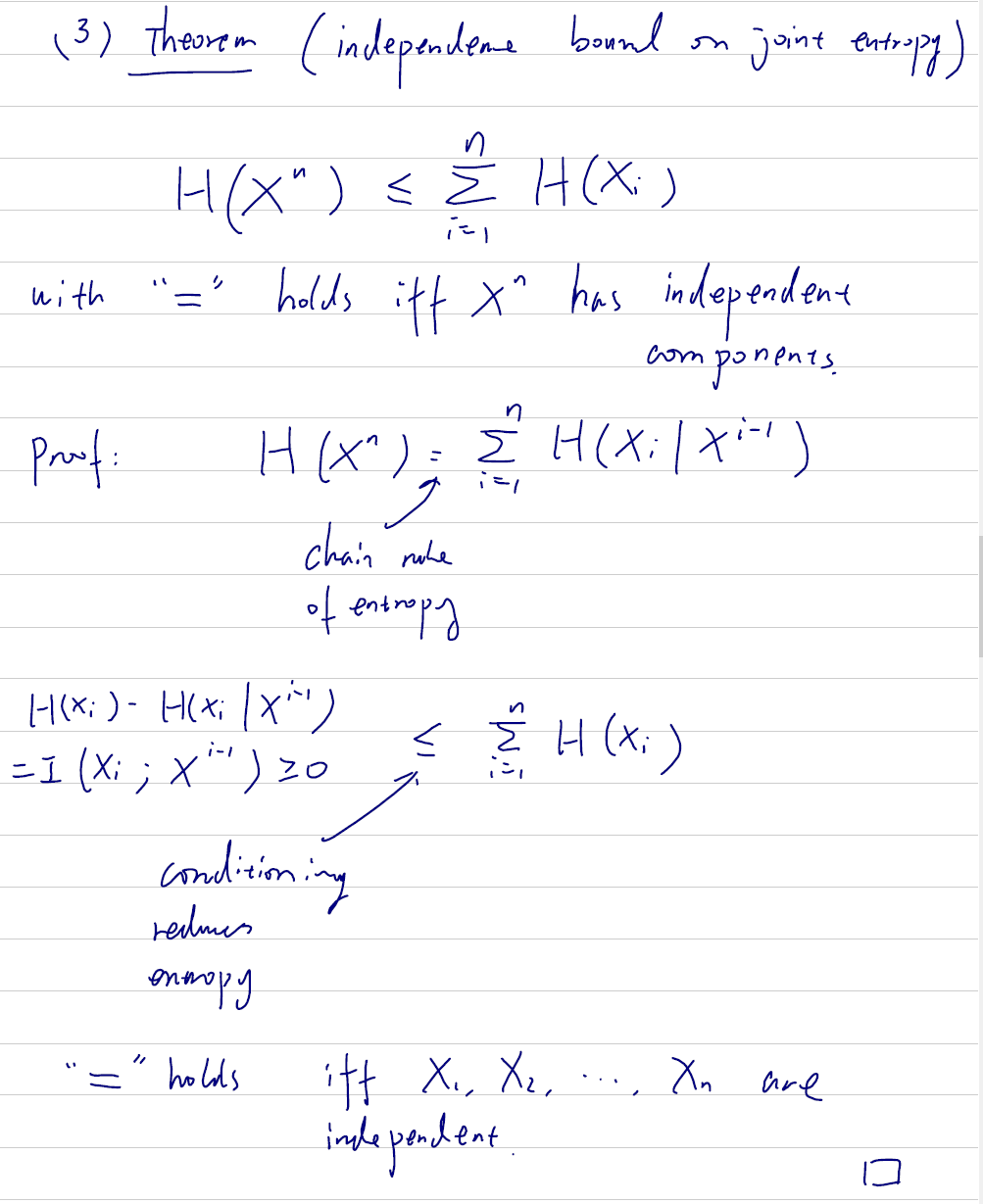


to proof it, need to use jensen’s ineq

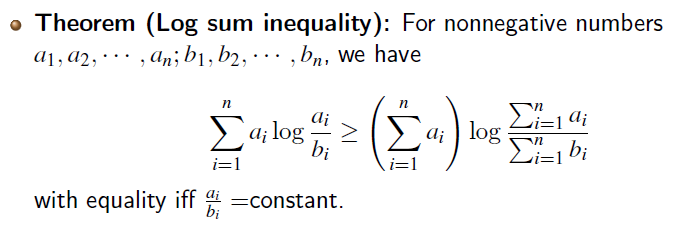


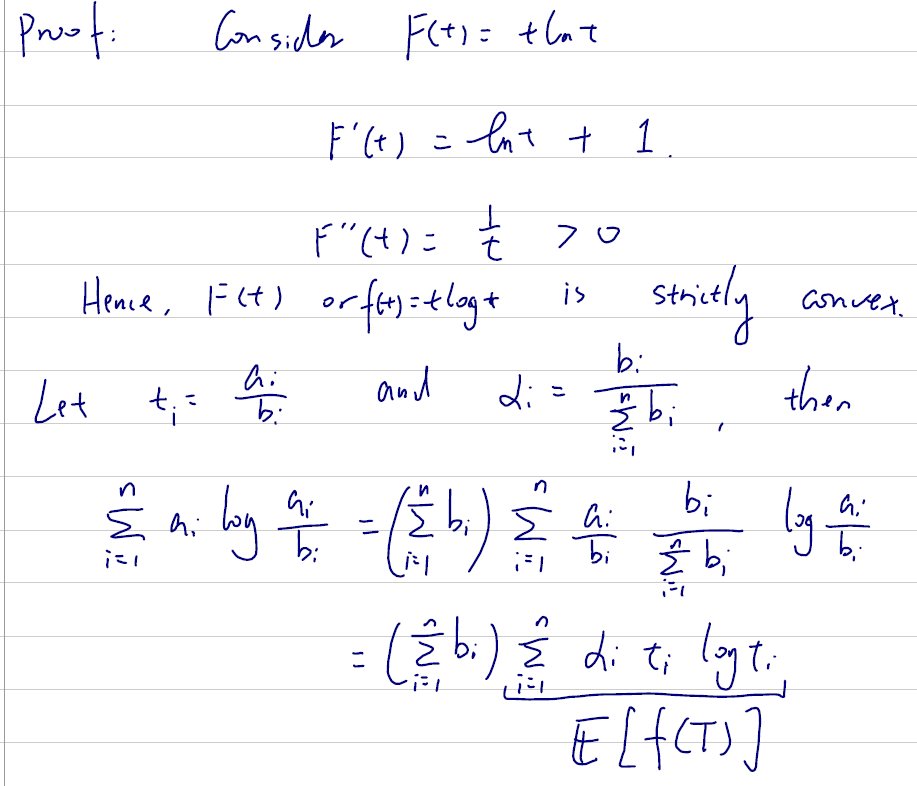
|X| 指 X set 里的元素个数

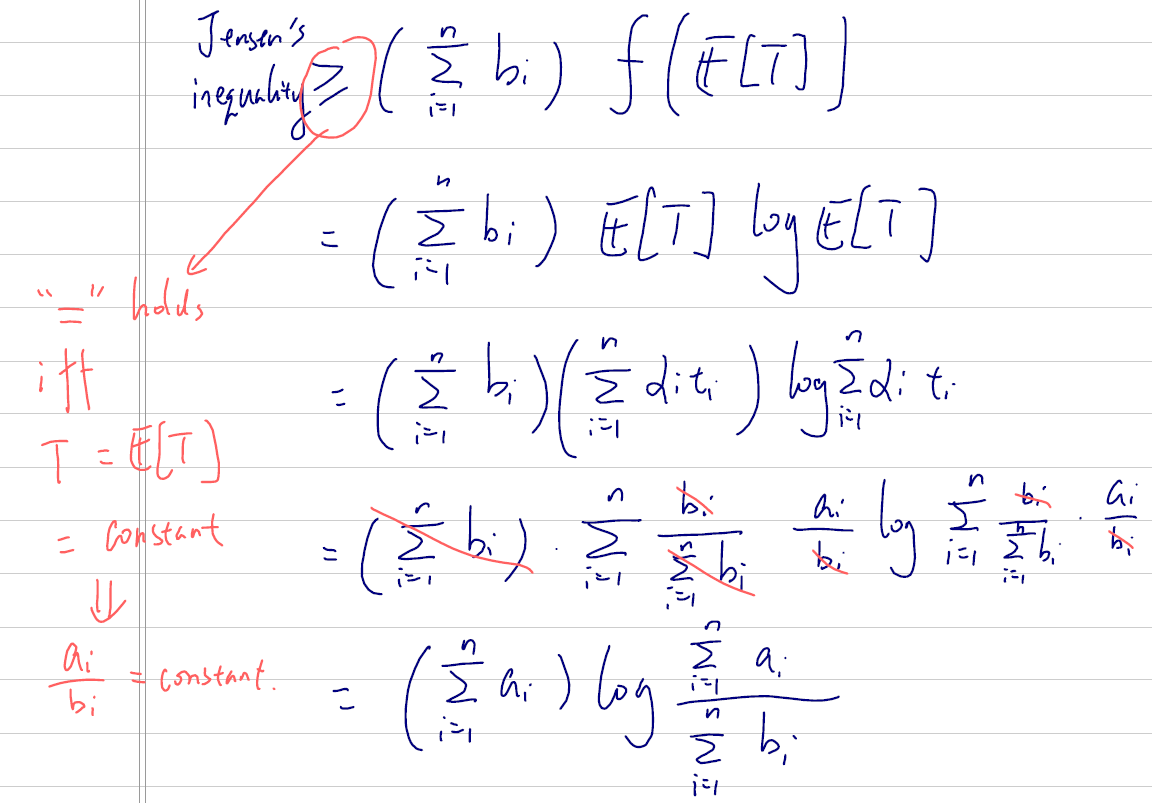




Log suminequality

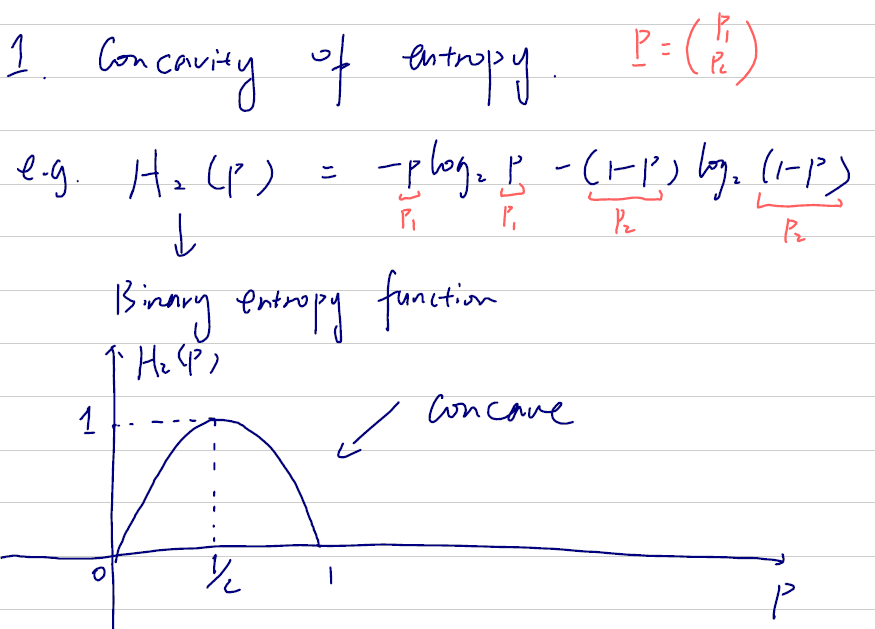




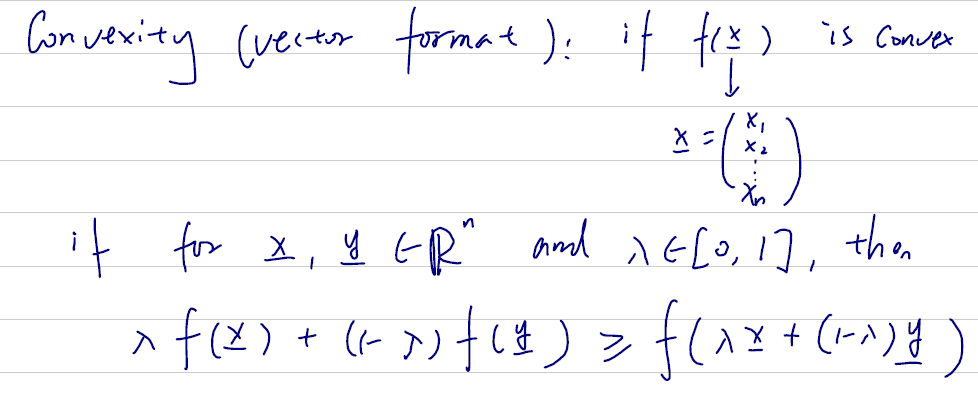


the convexity/concavity of information quantities

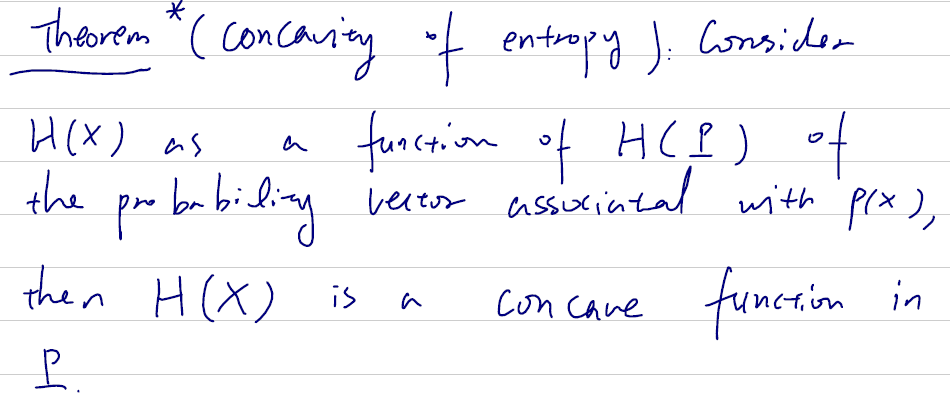
domain是一维情况



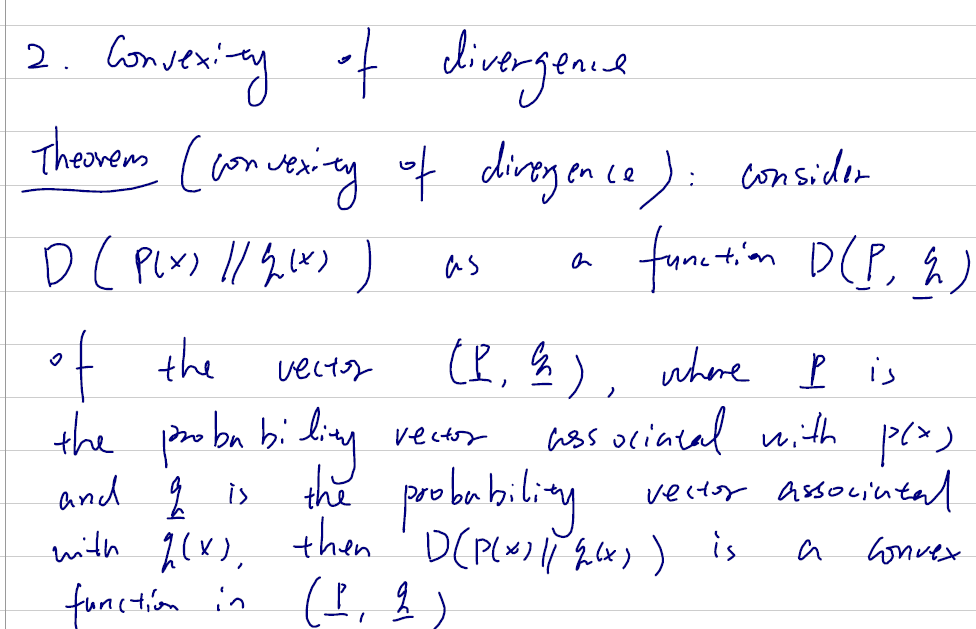
domain是n维情况:



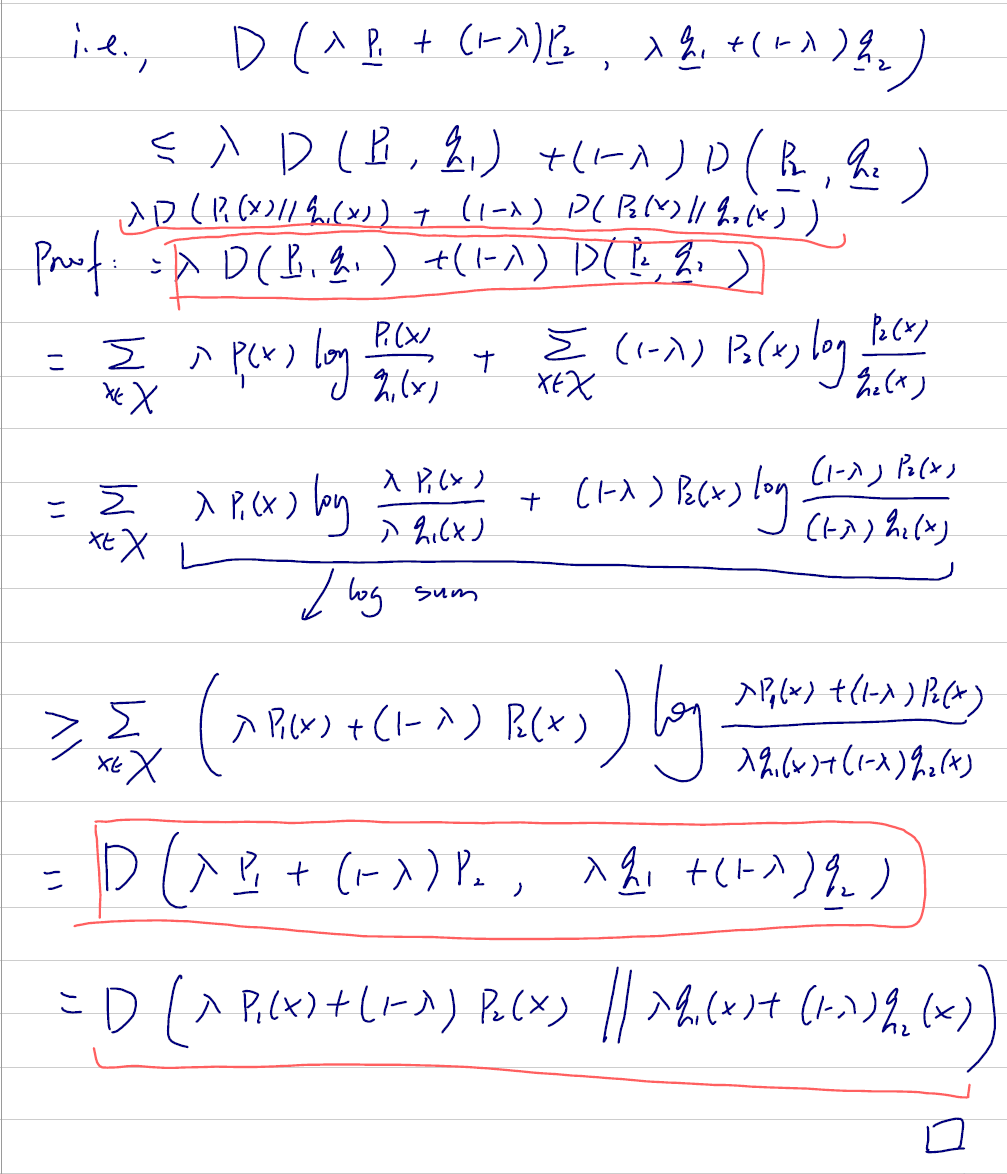
entropy 一定是 concave



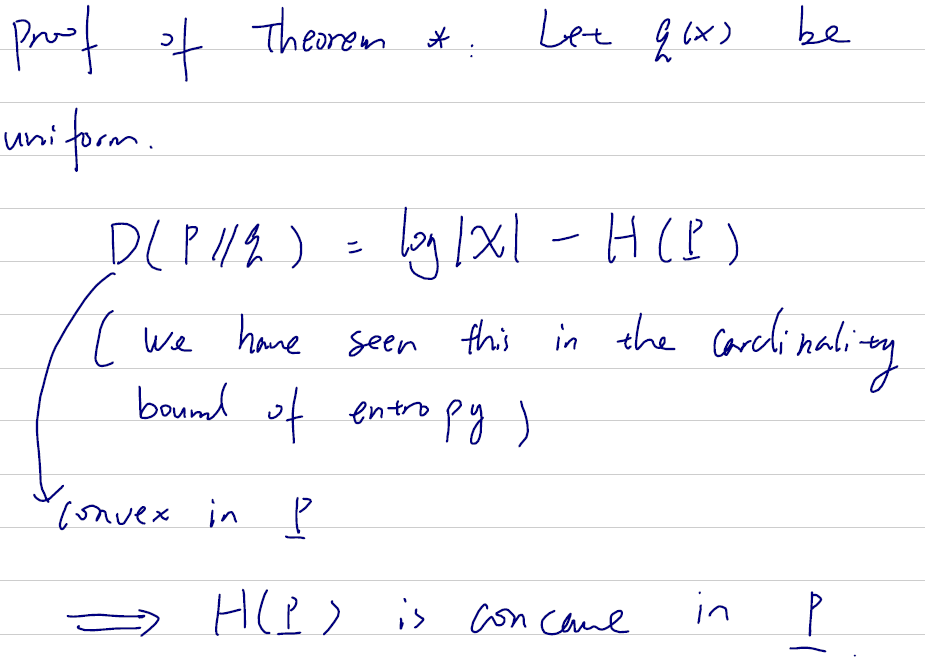
divergence 一定是 convex



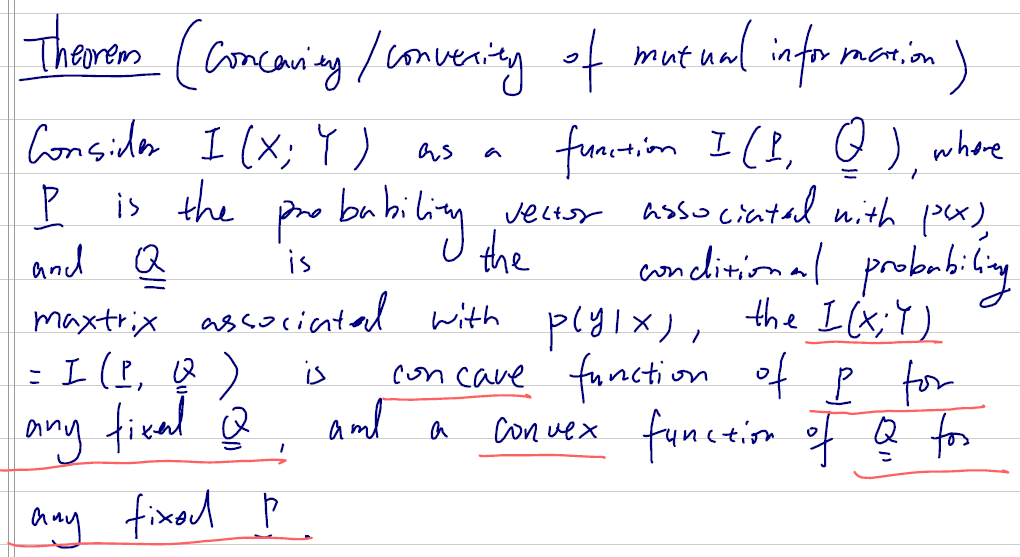
proof of divergence convexity

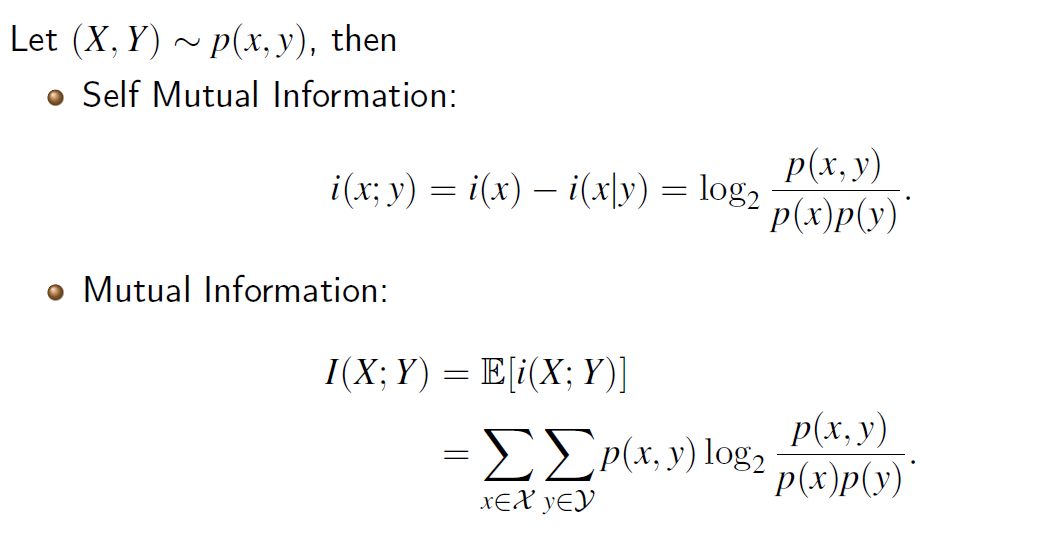


proof of entropy concavity



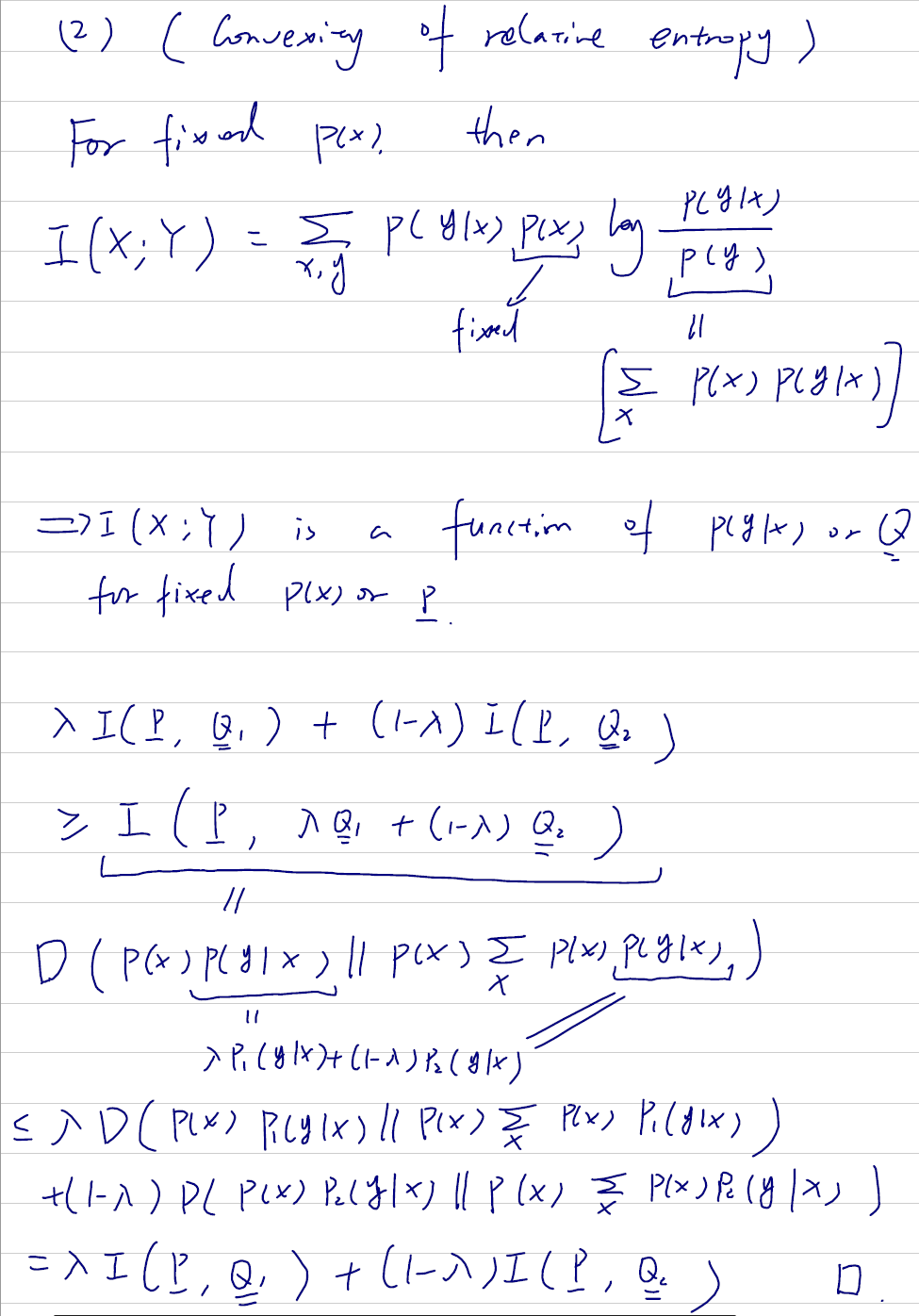
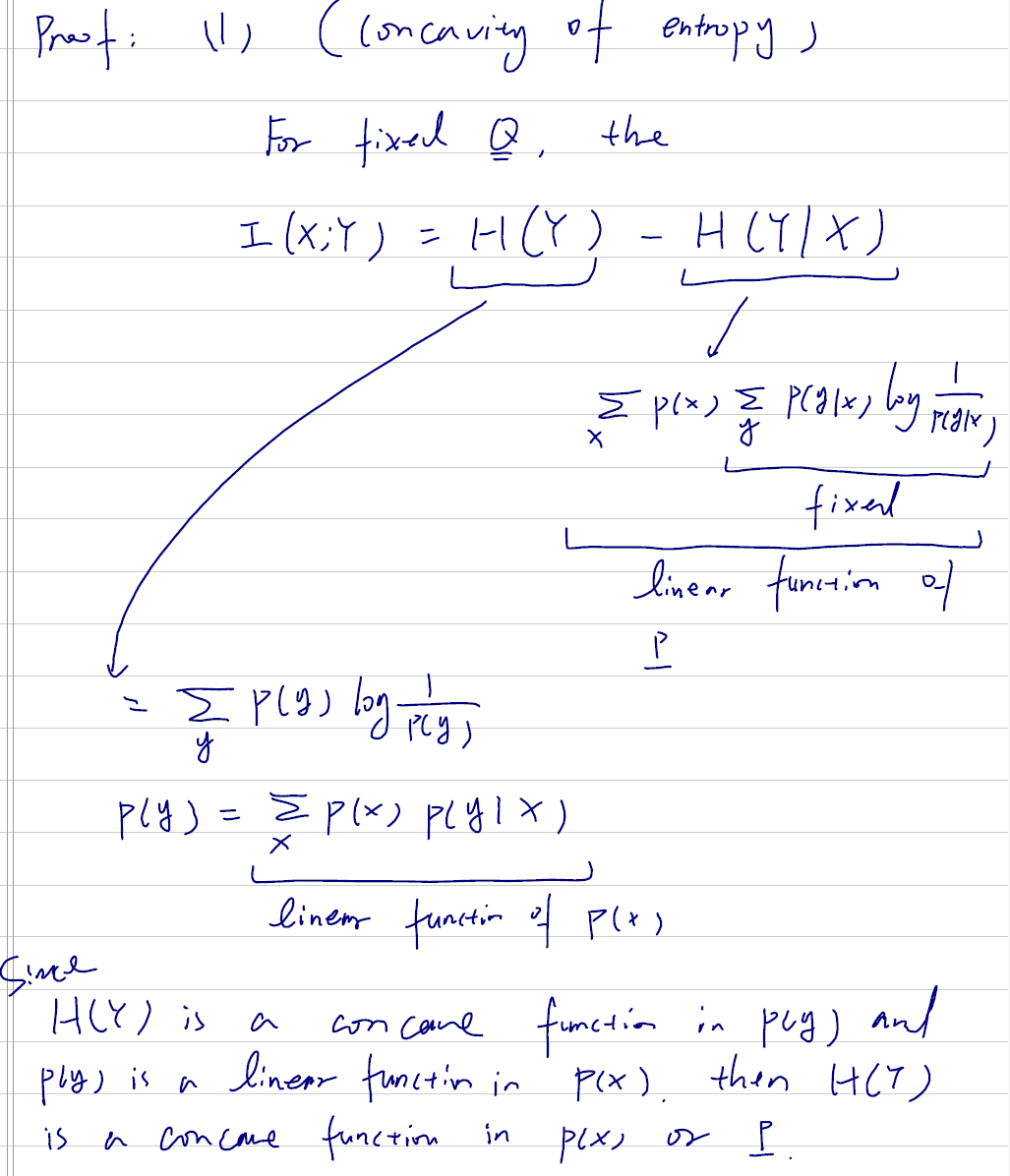
concavity/convexity of mutual information







Y确定X可变时是个concave, X确定 Y可变时是个convex



Lec4 note and slides