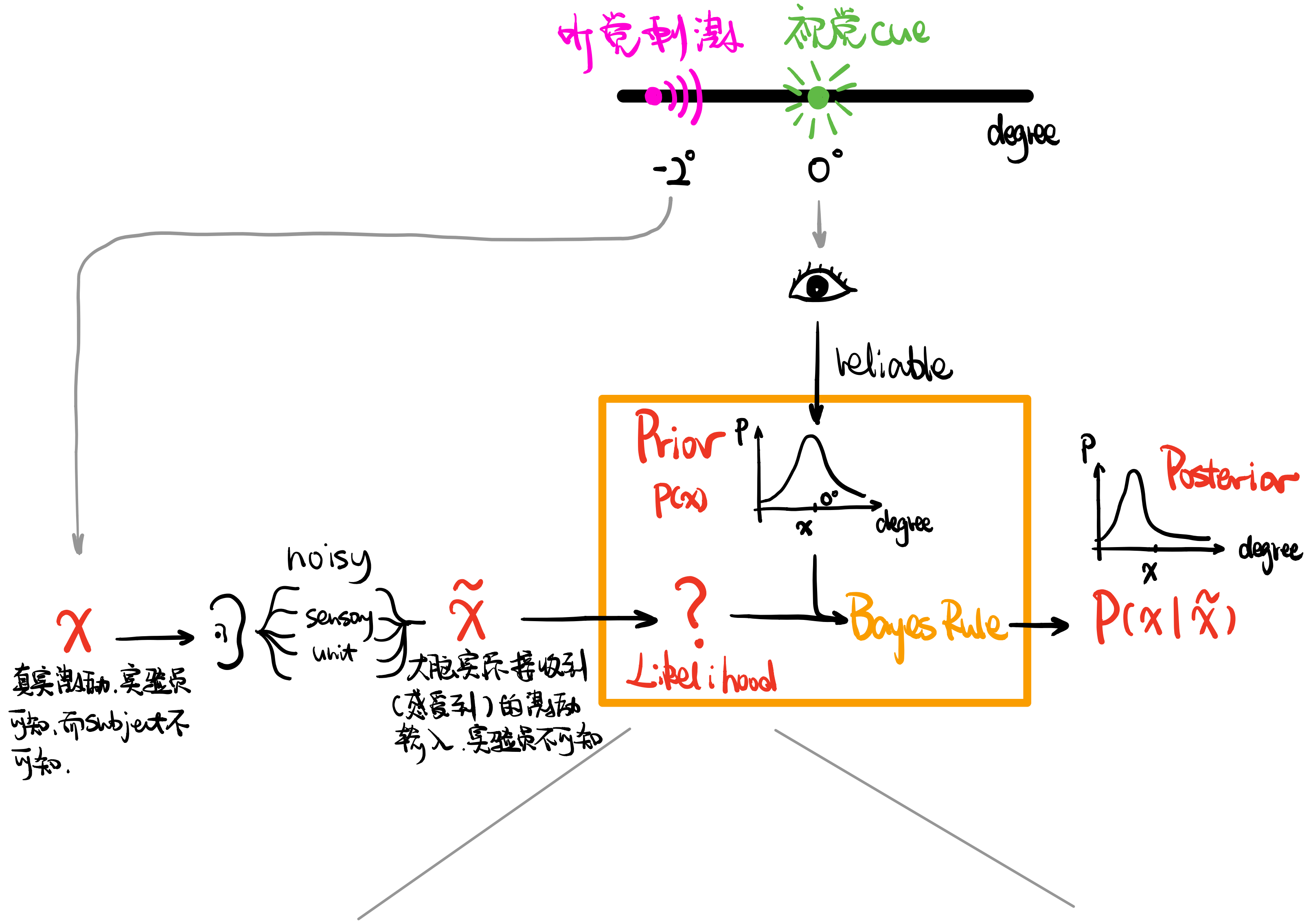
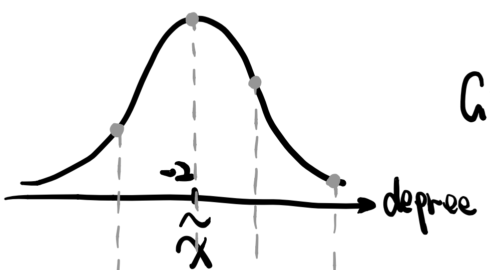


Likelihood在这个tutorial中，究竟是什么？

我理解的Likelihood是给定一个真实刺激 x，Subject听觉系统感受到的，真正输入大脑的 $\tilde{x}$ 的分布.

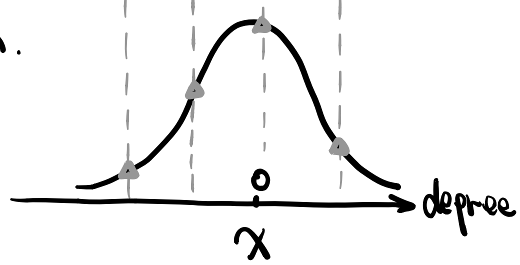
Given a specific stimuli $x=-2°$

$P(\tilde{x}|x=-2°)$



但，如果是这样，$P(\tilde{x}|x)$ 是无法与 $P(x)$ 进行tutorial中的pointwise multiplication来得到后验 $P(x|\tilde{x})$ 分布的.

$P(x)$



因为 $P(\tilde{x}|x=-2°)$ 是针对一个真实刺激，而 $P(x)$ 是对真实刺激在每个位置的先验估计

两者对不上号。再说，即使真求出了一个后验 $P(x|\hat{x})$，这个对 x 真实位置估计の分布是针对哪个特定的 $\hat{x}$？不要忘了你认为 $P(\hat{x}|x=-2^\circ)$ 中 $\hat{x}$ 是个分布啊！

所以说，Likelihood 并不是我理解の那样.

那如何理解？

我们先从"离散"的角度来考虑问题，假设听觉刺激只能出现在 -2° 0° 2° 这三个位置。由于视觉 cue 出现在 0°，所以 subject 认为声音刺激大概率出现在 0°。即 prior $p(x)$ 为，

x	-2°	0°	2°
$p(x)$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

当实验给出一个真实の声音激励 x，则必有一个通过听觉系统输入大脑的、确定的激励 $\hat{x}=-2^\circ$，基于 $\hat{x}$、prior，subject 得出对真实位置的后验估计.
固定的!!

x	-2°	0°	2°
$P(x\|\hat{x}=-2^\circ)$	?	?	?

问题是，夹在 $P(x)$、$P(x|\hat{x}=-2^\circ)$ 中间の是什么？
是

x	-2°	0°	2°
$P(\hat{x}=-2^\circ\|x)$			

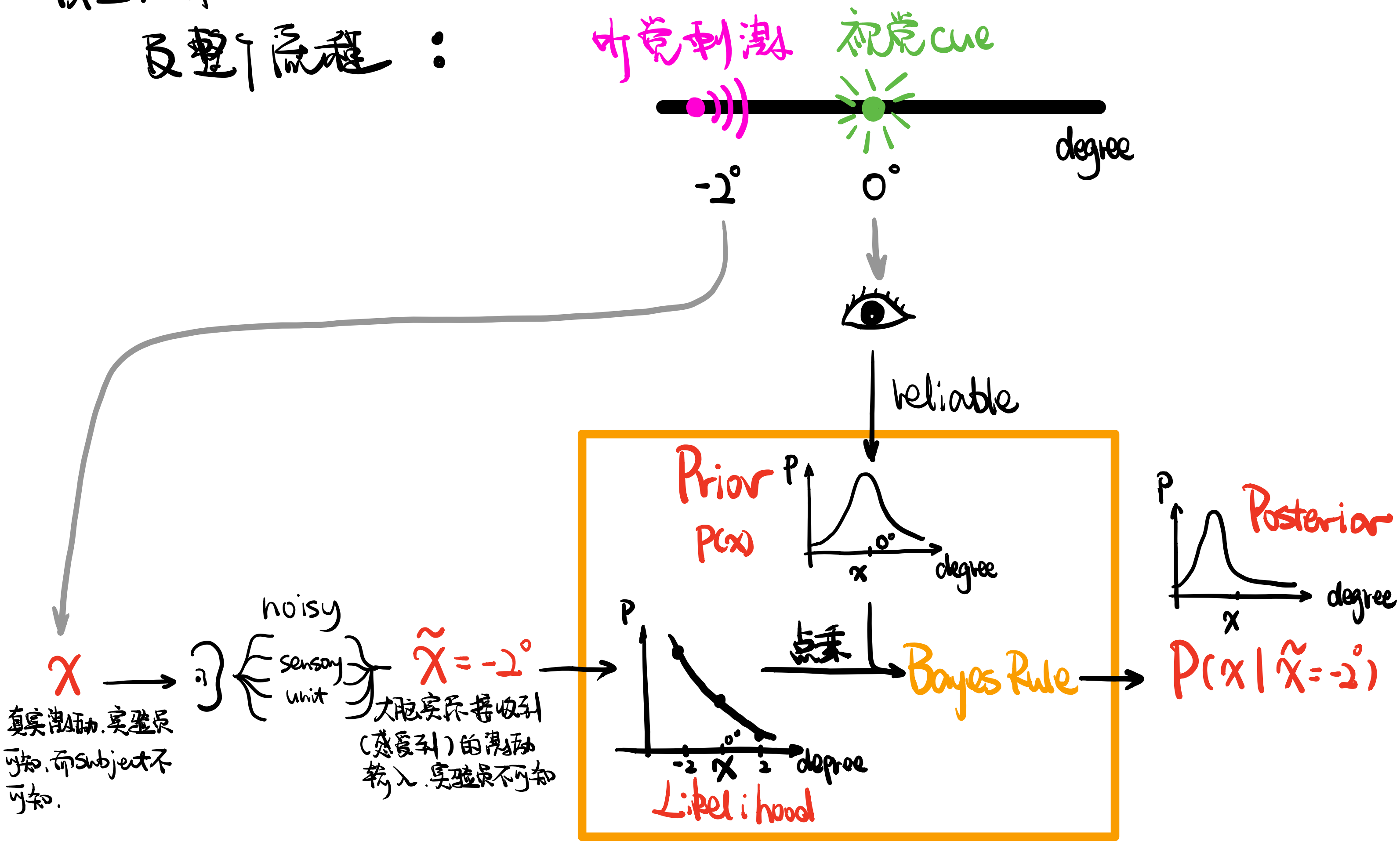
表示当真实激励取不同值时，大脑感知的输入为 $\hat{x}=-2^\circ$（一个特定情况）的概率.

x \ $\hat{x}$	-2°	0°	2°	
-2°	45次	11	5	$P(\hat{x}=-2^\circ\|x=-2^\circ)=\frac{45}{61}=0.74$
0°	10	60	8	$P(\hat{x}=-2^\circ\|x=0^\circ)=\frac{10}{78}=0.13$
2°	2	18	33	$P(\hat{x}=-2^\circ\|x=2^\circ)=\frac{2}{53}=0.04$

所以，整个计算流程为：

	x	-2°	0°	2°
Prior:	$p(x)$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
		pointwise multiplication		
Likelihood:	$P(\tilde{x}=2^\circ \mid x)$	0.74	0.13	0.04
Posterior:	$P(x \mid \tilde{x}=2^\circ)$	0.12	0.09	0.007
		normalization		

综上，明确一下likelihood
及整个流程：

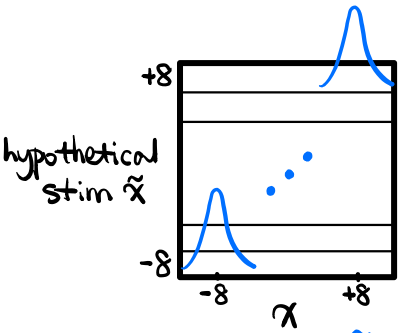


在一个trial中，一定会有一个固定的 $\tilde{x}$！

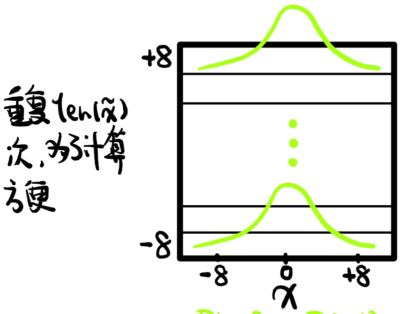
而likelihood $P(\tilde{x} \mid x)$ 表示 subject 具有的一个"能力"，即，他知道在不同真实刺激条件下，他能感受到这个 $\tilde{x}=-2^\circ$ 的概率。

Tutorial 3 的流程，由此，希望可以了解 Bayesian Stats 的概貌.

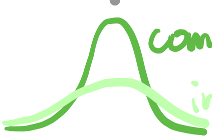
hypothetical_stim $\tilde{x}$：假设大脑接收到的 input -8 ~ 8，对于每一个 $\tilde{x}$，都会有一个 likelihood (一行)
.e.g., 大脑"认为"：我感受到 $\tilde{x}$=-8，不同真实激励 x 会使我感受到 $\tilde{x}$=-8 的概率.



Likelihood $P(\tilde{x}|x)$



prior p(x)



先验概率由两部分认知：

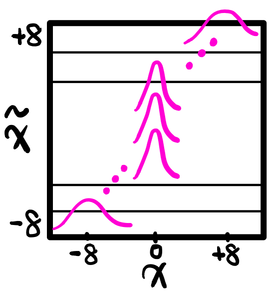
① common：声音的出现位置与视觉线索有关

② independent：两者无关.

两个先验分布以 P_{common} 为权重参数，进行累加.

点乘

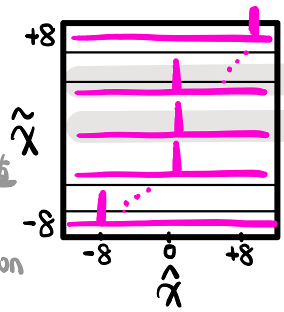
当大脑接收到 $\tilde{x}$ 作为输入时，其利用 Likelihood, Prior 得出 $\tilde{x}$ 对真实激励 x 的分布估计



posterior $p(x|\tilde{x})$

取分布均值对应 x 作为 decision

当大脑接收到 $\tilde{x}$ 作为输入时，其利用 Likelihood, Prior 对真实激励的预测 $\hat{x}$

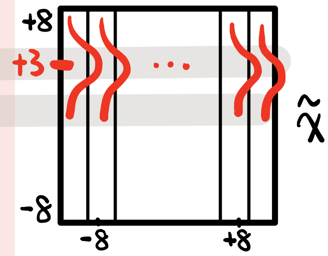


$P(\hat{x}|\tilde{x})$

实际问题：

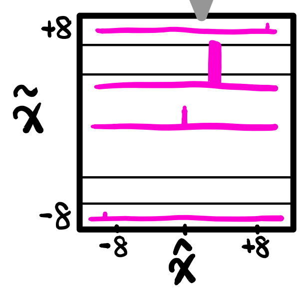
真实激励 x=+3

当给出一个真实声音激励 x=+3时，大脑接受到各种 $\tilde{x}$ 的概率.



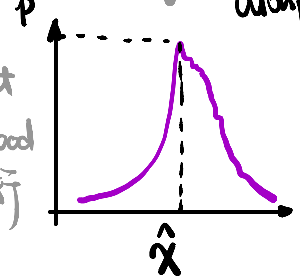
$P(\tilde{x}|x)$

根据在真实激励下，大脑感受各 $\tilde{x}$ 的概率，来对于各 $\tilde{x}$，对真实激励的预测 $\hat{x}$ 进行**加权**



marginalization along $\tilde{x}$

当声音激励为 x=3时，subject 基于 prior, likelihood 对真实位置进行预测.



这一部分就像是subject具有的能力一样.

① subject 有自己对声音激励位置的先验认知

② subject 大脑中,有基于每个可能 $\tilde{x}$ 的 likelihood

③ 大脑中有个贝叶斯过程.

④ 给出了基于不同 $\tilde{x}$ 输入的决策 $\hat{x}$

当实际给出一个激励时,可以得出对激励位置的估计.

Now that we have generated some data, we will attempt to recover the parameter $p_{independent}$ that was used to generate it.

- 用下图的整个流程，基于特定的 p_{common} or $p_{independent}$ 来生成一个虚拟参与者的数据：**(真实激励位置，预测位置)**
- 这些组数据是由参与者的先验参数 $p_{independent}$ 决定的。
- 假设参与者的大脑是一个贝叶斯过程，我们能否通过贝叶斯模型，利用参与者的这些行为数据，来估计出这个先验参数?

We have provided you with an incomplete function called `my_Bayes_model_mse()` that needs to be completed to perform the same computations you have performed in the previous exercises

- *就是上述的整个流程*

but over all the participant's trial, as opposed to a single trial.

- *但是这次不再是一个trial 了，而是所有的trial 一起算。*

The likelihood: has already been constructed; since it depends only on the hypothetical stimuli, it will not change.

- *也就是说，在贝叶斯统计中，likelihood 并不是要去优化的。因为遍历了hypothetical stimuli就可以知道了。*
- *但这对我来说很奇怪，不同人的感知，即likelihood，不应该是不同的吗?*

Prior: However, we will have to implement the prior matrix, since it depends on $p_{independent}$. We will therefore have to recompute the posterior, input and the marginal in order to get $p(\hat{x} \mid x)$.

- $p(\hat{x} \mid x)$ 是以 $p_{independent}$ 为变量的函数

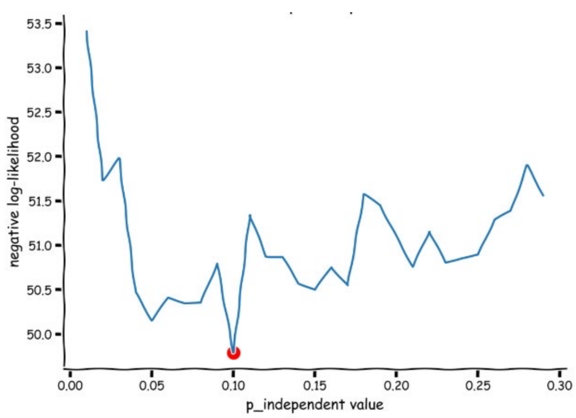
Optimization: Using $p(\hat{x} \mid x)$, we will then compute the negative log-likelihood for each trial and find the value of $p_{independent}$ that minimizes the negative log-likelihood (i.e. maximises the log-likelihood. See the model fitting tutorial from W1D3 for a refresher).

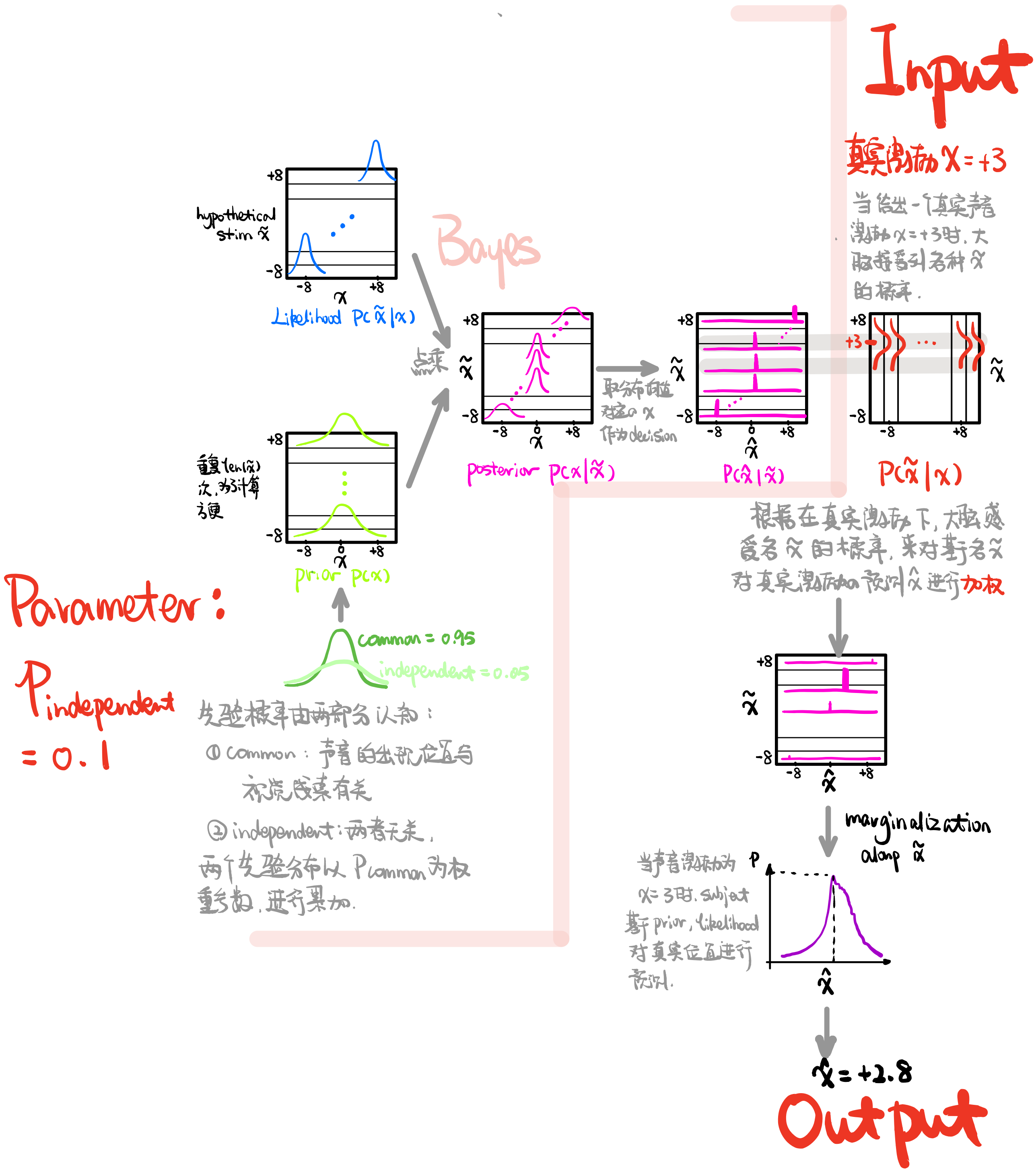
- *优化出一个 $p_{independent}$，使得-LL 最小即 使得贝叶斯模型最能fit 参与者的行为数据。可见，说了这么多，贝叶斯方法就是在估计参与者的先验知识！！*

In this experiment, we assume that trials are independent from one another. This is a common assumption--and it's often even true! It allows us to define negative log-likelihood as:

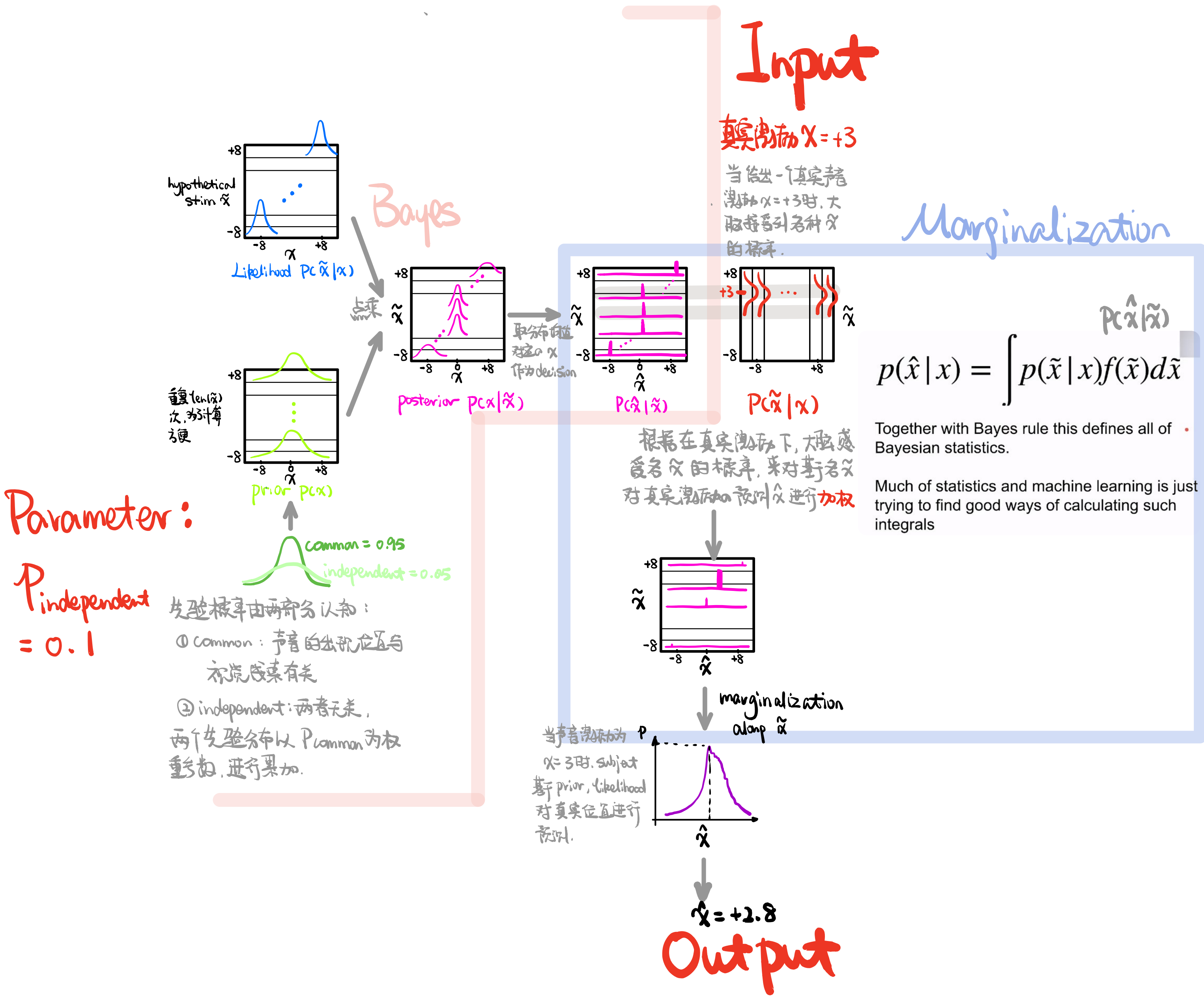
$$-LL = -\sum_i \log p(\hat{x}_i \mid x_i)$$

where $\hat{x}_i$ is the participant's response for trial i, with presented stimulus x_i





Input
真实刺激 x = +3
当给出一个真实声音刺激 x = +3 时，大脑接受到各种 x̃ 的概率。
Bayes
hypothetical stim x̃
Likelihood P(x̃|x)
点乘
重复 len(x̃) 次，为了计算方便
prior p(x)
posterior p(x|x̃)
取分布均值对应的 x 作为 decision
P(x̂|x̃)
P(x̃|x)
根据在真实刺激下，大脑感受各 x̃ 的概率，来对基于各 x̃ 对真实刺激的预测 x̂ 进行加权
Parameter:
P independent = 0.1
common = 0.95
independent = 0.05
先验概率由两部分认知：
① common：声音的出现位置与视觉刺激有关
② independent：两者无关，
两个先验分布以 P common 为权重参数，进行累加。
marginalization along x̃
当声音刺激为 x = 3 时，subject 基于 prior，likelihood 对真实位置进行预测。
x̂ = +2.8
Output



$$p(\hat{x}\,|\,x) = \int p(\tilde{x}\,|\,x) f(\tilde{x}) d\tilde{x}$$

Together with Bayes rule this defines all of Bayesian statistics.

Much of statistics and machine learning is just trying to find good ways of calculating such integrals