首先，理解清楚信息矩阵的概念,根据wiki的解释，
In mathematical statistics, the Fisher information (sometimes simply called information[1]) is a way of measuring the amount of information that an **observable random variable X** carries about an **unknown parameter θ** of a distribution that models X.

Fisher Information 的公式表达

The variance of the score is defined to be the **Fisher information**:[5]

$$\mathcal{I}(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\middle|\theta\right] = \int \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta)\,dx,$$

关于这个公式，我们可以看出来，信息矩阵所针对的是未知参数**θ**，与x没有直接联系。此外，可以看出来，括号内部的部分期望为0。

$$\begin{aligned}
&\mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X;\theta)\middle|\theta\right] \\
&= \int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)} f(x;\theta)\,dx \\
&= \frac{\partial}{\partial\theta}\int f(x;\theta)\,dx \\
&= \frac{\partial}{\partial\theta}1 = 0.
\end{aligned}$$

在Agustinus博客中，它给出了向量的表达，以及对应离散的表达

We can then see it as an information. The covariance of score function above is the definition of Fisher Information. As we assume θ is a vector, the Fisher Information is in a matrix form, called Fisher Information Matrix:

$$\mathrm{F} = \mathop{\mathbb{E}}_{p(x|\theta)}\left[\nabla \log p(x|\theta)\,\nabla \log p(x|\theta)^{\mathrm{T}}\right].$$

However, usually our likelihood function is complicated and computing the expectation is intractable. We can approximate the expectation in F using empirical distribution $\hat{q}(x)$, which is given by our training data $X = \{x_1, x_2, \cdots, x_N\}$. In this form, F is called Empirical Fisher:

$$\mathrm{F} = \frac{1}{N}\sum_{i=1}^{N} \nabla \log p(x_i|\theta)\,\nabla \log p(x_i|\theta)^{\mathrm{T}}.$$

接下来我们来关注一下Hessian Matrix并寻找它和信息矩阵的关系。

首先关注Hessian matrix的定义
In mathematics, the Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a scalar-valued function, or scalar field. It describes the local curvature of a function of many variables.

还是根据Agustinus的博客，我们首先得到其概率分布对应的Hessian矩阵的表达。

$$
\begin{aligned}
\mathrm{H}_{\log p(x|\theta)} &= \mathrm{J}\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right) \\
&= \frac{\mathrm{H}_{p(x|\theta)}\, p(x|\theta) - \nabla p(x|\theta)\, \nabla p(x|\theta)^{\mathrm{T}}}{p(x|\theta)\, p(x|\theta)} \\
&= \frac{\mathrm{H}_{p(x|\theta)}\, p(x|\theta)}{p(x|\theta)\, p(x|\theta)} - \frac{\nabla p(x|\theta)\, \nabla p(x|\theta)^{\mathrm{T}}}{p(x|\theta)\, p(x|\theta)} \\
&= \frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)} - \left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)^{\mathrm{T}},
\end{aligned}
$$

然后，我们计算对应的期望

$$
\begin{aligned}
\mathop{\mathbb{E}}_{p(x|\theta)}\left[\mathrm{H}_{\log p(x|\theta)}\right] &= \mathop{\mathbb{E}}_{p(x|\theta)}\left[\frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)} - \left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)^{\mathrm{T}}\right] \\
&= \mathop{\mathbb{E}}_{p(x|\theta)}\left[\frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)}\right] - \mathop{\mathbb{E}}_{p(x|\theta)}\left[\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)\left(\frac{\nabla p(x|\theta)}{p(x|\theta)}\right)^{\mathrm{T}}\right] \\
&= \int \frac{\mathrm{H}_{p(x|\theta)}}{p(x|\theta)} p(x|\theta)\, \mathrm{d}x - \mathop{\mathbb{E}}_{p(x|\theta)}\left[\nabla \log p(x|\theta)\, \nabla \log p(x|\theta)^{\mathrm{T}}\right] \\
&= \mathrm{H}_{\int p(x|\theta)\, \mathrm{d}x} - \mathrm{F} \\
&= \mathrm{H}_1 - \mathrm{F} \\
&= -\mathrm{F}\,.
\end{aligned}
$$

请注意，H对**θ**的二阶导为0。
这样，得证，H的期望等于负的信息矩阵，而当H从期望变成固定的**θ**对应的信息矩阵时，E被去掉，H=-F。

Hessian和Covariance Matrix的关系

这个关系在提供的论文中，已经得到了证明，

Consider a Gaussian random vector $\boldsymbol{\theta}$ with mean $\boldsymbol{\theta}^{\star}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ so its joint probability density function (PDF) is given by:

$$p(\boldsymbol{\theta}) = (2\pi)^{-\frac{N_{\boldsymbol{\theta}}}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{\star})^{T} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^{\star})\right] \tag{A.1}$$

The objective function can be defined as its negative logarithm:

$$J(\boldsymbol{\theta}) \equiv -\ln p(\boldsymbol{\theta}) = \frac{N_{\boldsymbol{\theta}}}{2} \ln 2\pi + \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{\star})^{T} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^{\star}) \tag{A.2}$$

which is a quadratic function of the components in $\boldsymbol{\theta}$. By taking partial differentiations with respect to θ_l and $\theta_{l'}$, the (l, l') component of the Hessian matrix can be obtained:

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^{\star}) = \left.\frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_l \partial\theta_{l'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}} = (\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})^{(l,l')} \tag{A.3}$$

so the Hessian matrix is equal to the inverse of the covariance matrix:

$$\mathcal{H}(\boldsymbol{\theta}^{\star}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \tag{A.4}$$

A.4给出了H和协方差矩阵的关系，从这里可以看出，信息矩阵F=-H=-inv(cov)