# BASICS OF BAYESIAN STATISTICS

The Bayes theorem states that

$$p(b|a) = \frac{p(a|b) \cdot p(b)}{p(a)}$$

PROOF Using the laws of probability we have $p(a,b) = p(a|b)\,p(b) = p(b|a)\,p(a)$

$$\Rightarrow \quad p(b|a) = \frac{p(a|b)\,p(b)}{p(a)}$$

NOTE $p(a)$ is called marginal likelihood or evidence and can be written as

$$p(a) = \int p(a|b)\,p(b)\,db$$

EXAMPLE

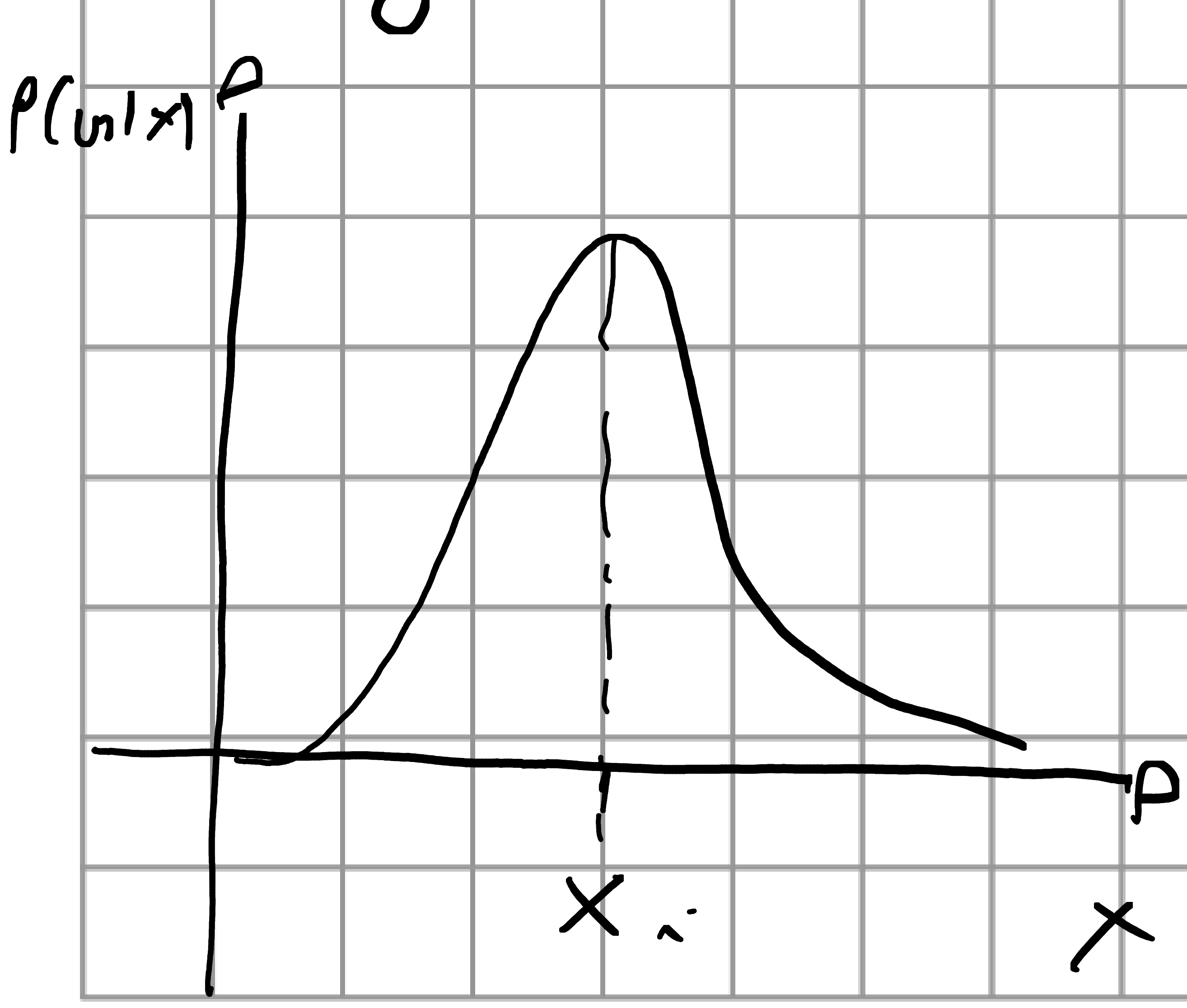
We want to estimate the mean of a Gaussian distribution. The first step is to write the our likelihood

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

If we observe one value of $x_i$, then we have that the posterior is

$$P(\mu \mid x_i) = \frac{\mathcal{L}(x_i \mid \mu)\, P(\mu)}{\int \mathcal{L}(x_i \mid \mu)\, P(\mu)\, d\mu} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)}{2\sigma^2}}$$

We can observe that our posterior is a Gaussian with $\mu = x_i$ and $\sigma^2 = \sigma^2$




$$\langle \mu \rangle = \int \mu\, P(\mu \mid x_i)\, d\mu = \mu \,.$$

$$\langle \sigma_\mu^2 \rangle = \int P(\mu \mid x_i)\, (\mu - \langle \mu \rangle)^2\, d\mu = \sigma^2$$

What happens when we have multiple observations? $\{X\} = \{x_1, \dots, x_n\}$

If the observations are I.I.D., then

$$\mathcal{L}(\{X\}|\mu) = \prod_{i=1}^{N} \mathcal{L}(x_i|\mu)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)}{2\sigma^2}} = \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^N \exp\left[\sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

BUT $\sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2\sigma^2} = \sum \frac{(x_i^2 - 2x_i\mu + \mu^2)}{2\sigma^2} = -\frac{(\bar{x}-\mu)^2}{2\bar{\sigma}^2}$

with $\bar{x} = \sum_i \frac{x_i}{N}$ and $\bar{\sigma}^2 = \frac{\sigma^2}{N}$

IMPORTANT It follows that the overall position will scale as $\sqrt{N}$ in terms of precision.

# UNDERSTANDING SOME BASIC CONCEPTS

## THE CONCEPT OF MODEL SELECTION

In Bayesian statistic we can also compare models. We can define the preference among competing models $M_1$ and $M_2$ as

$$\frac{P(M_1 | X)}{P(M_2 | X)} = \frac{P(X | M_1)}{P(X | M_2)} \boxed{\frac{P(M_1)}{P(M_2)}} \rightarrow \text{ODDS RATIO}$$

$$P(X | M_1) = \int P(X | \theta, M_1) \cdot P(\theta | M_1) \, d\theta$$

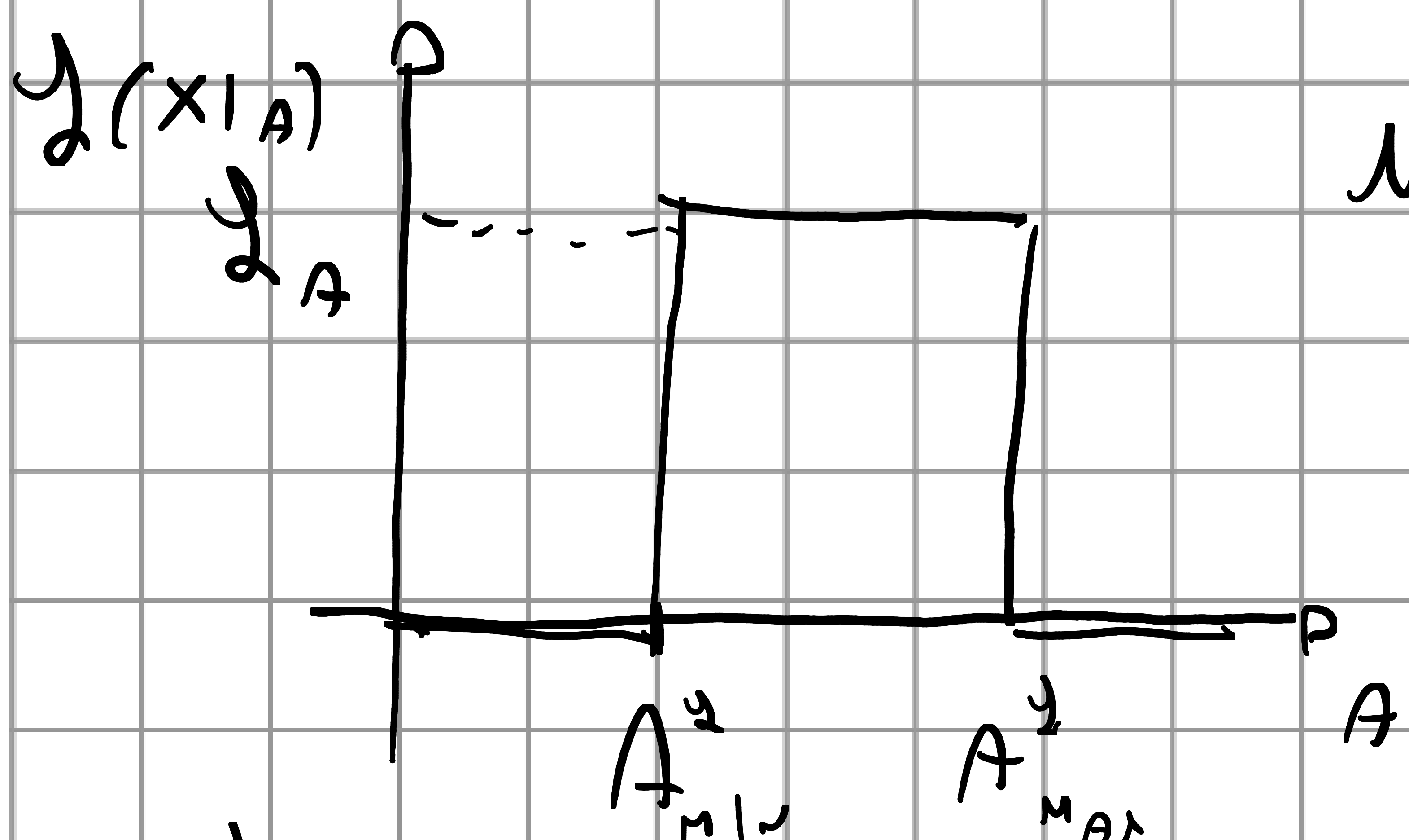
$\hookrightarrow$ MARGINAL LIKELIHOOD.

The two models can also have different # parameters.

EXAMPLE Imagine that we want to compare

We prefer among two models. $M_A$, $M_B$

$M_A$: Has uniform likelihood to explain the observation $X$, with a parameter $A$ and




has a prior $\pi(A) = \frac{\Theta(|A_{Max}-A_{m}|)}{A_{Max}-A_{Min}}$

$$\mathcal{L}_A = \int_{A_{MIN}}^{A_{MAX}} \mathcal{L}(X|A) \frac{1}{A_{Max}-A_{MIN}} \, dA$$

$= \mathcal{L}_A \frac{A^1_M - A^1_m}{A_{MAX}-A_{MIN}}$. We see that the evidence is the maximum likelihood times posterior volume over prior volume.

## PP-PLOT

UNIFORM DISTRIBUTION

$$p(a|b) = \frac{p(b|a)\,p(b)}{p(a)} \qquad \boxed{C = \int_{-\infty}^{a} p(a|b_T)\, da}$$

$$\frac{\partial c}{\partial a} = p(a|b_T) \Rightarrow p_c = p_a \cdot \left|\frac{\partial c}{\partial a}\right|^{-1} = 1$$

# GW POPULATION INFERENCE

We want to estimate a set of population parameters $\lambda$ in common to an entire population.

$p_{pop}(\theta | \lambda) \Rightarrow$ process that generates observations.

The probability of having $N_{Dete}$ with parameters $\theta_i$ is

$$p(\{\theta\} | \lambda) = \prod_{i=1}^{N} \frac{p_{pop}(\theta_i | \lambda)}{\int d\theta \, p_{pop}(\theta | \lambda)}$$

Unfortunately our detections have a selection bias, i.e. we are not able to see all the $\theta$.

$$\Rightarrow \quad p_D(\{\theta\} | \lambda) = \prod_{i=1}^{N} \frac{p_{pop}(\theta_i | \lambda) \overbrace{p(DET | \theta_i)}^{\sim 1}}{\int p_{pop}(\theta_i | \lambda) \, p(DET | \theta)}$$

$\sim 1$ as we know we detect event

NOTE The definition of $p(DET|\theta)$ come from the likelihood:

$$p(DET|\theta) = \int_{X_{DET}} \mathcal{L}(X|\theta)\, dX$$

Moreover we need to remember that we are not able to measure perfectly $\theta_i$, i.e. there exist a noise process for which exist $\mathcal{L}(X|\theta_i)$

$$\Rightarrow p(X_i|\lambda) = \int d\theta\, \mathcal{L}(X_i|\theta)\, p(\theta|\lambda) =$$

$$= \frac{\int d\theta\, \mathcal{L}(X|\theta)\, p_{pop}(\theta|\lambda)}{\int d\theta\, p(DET|\theta)\, p_{pop}(\theta|\lambda)\, d\theta}$$

The overall pop posterior is

$$p(\lambda|\{X\}) = p(\lambda) \prod_{i=1}^{N} \frac{\int d\theta\, \mathcal{L}(X|\theta)\, p_{pop}(\theta|\lambda)}{\int d\theta\, p_{DET}(\theta)\, p_{pop}(\theta|\lambda)}$$

# BRIGHT SIREN CASE

Let us write the likelihood

$$\mathcal{L}(x|H_0) = \frac{\int dz\, \mathcal{L}^{EM+GW}(x|z, H_0)\, P_{pop}(z)\, dz}{\int P^{EM+GW}_{Det}(z, H_0)\, P_{pop}(z)\, dz}$$

In the local universe we can approximate $d_L \approx \frac{cz}{H_0}$ and $P(z) \propto z^2$.

Moreover we will assume that the EM EM counterpart measures the redshift precisely while the GW the luminosity distance.

We have $\mathcal{L}^{EM+GW}(x|z, H_0) \propto \mathcal{L}^{EM}(\hat{z}|z)\, \mathcal{L}^{GW}(x|d_L)$

$$\propto \delta(\hat{z}-z)\, \mathcal{L}^{GW}(x|d_L)$$

$$\mathcal{L}(X|H_0) \propto \frac{\mathcal{L}_{GW}(X|d_L(\hat{z},H_0))\,\hat{z}^2}{\int p^{EM+GW}_{DET}(z,H_0)\,z^2\,dz}$$

Since we do not know the GW likelihood but we know the posterior, we can write

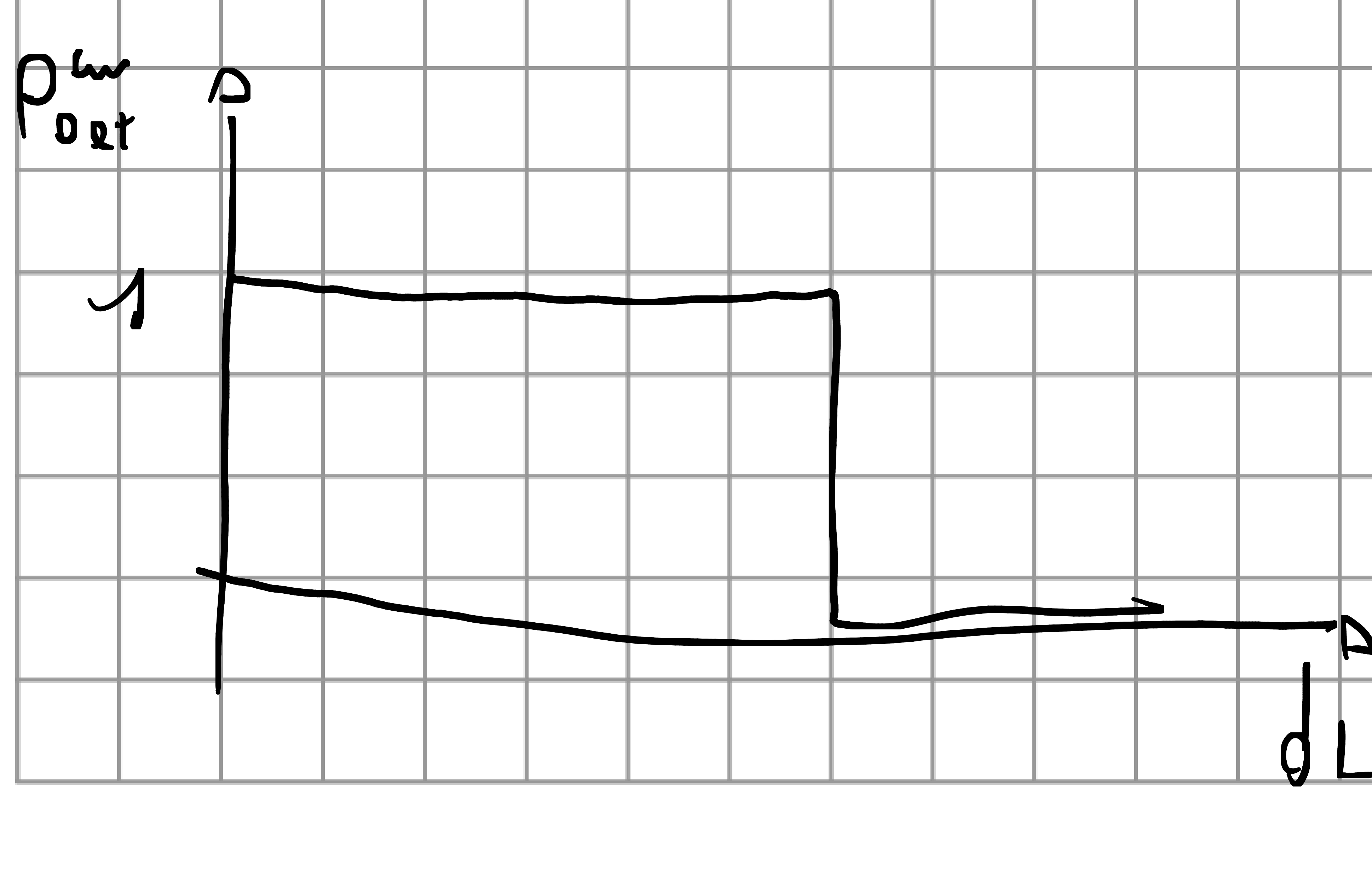
$$\mathcal{L}_{GW}(X|d_L) \propto \frac{p(d_L|X)}{\pi(d_L)}$$

BAYES THEOREM

PRIOR USED FOR PEs

NOTE For $p_{DET}$ we will assume a simplified model. We will assume that EM counterparts can be always detected $p^{EM}_{DET}(z)=1$ $\forall z$, and that

$$p^{GW}_{DET}(z,H_0) = p^{GW}_{DET}(d_L) = \begin{cases} 1 & \text{if } 0 \le d_L \le d_L^{TH} \\ 0 & \text{OTHERWISE} \end{cases}$$



Then we can calculate analytically the integral at denominator.

$$\int_0^{z=\frac{d_L^{THR} \cdot H_0}{c}} z^2 \, dz = H_0^3 \cdot \left[\frac{d_L^{THR}}{3c}\right]^3 \propto H_0^3$$

Finally we have that $\mathcal{L}(x|H_0) \propto \dfrac{\dfrac{P(d_L(\hat{z}, H_0)|x)}{\pi(d_L(\hat{z}, H_0))}}{H_0^3}$

NOTE

- The selection bias proportional to $H_0^3$ is telling you that we do not favour high $H_0$. This is because we detect GWs in $d_L$ and higher $H_0$ values correspond to larger volumes.
- The posterior on $d_L$ can be calculated from posterior samples.

# BASIC GALAXY CATALOG METHOD

This time we do not measure any EM counterpart but we have a galaxy catalog.
Now we also need to use information on the sky position $\Omega$ of the GW event.

$$\mathcal{L}(x|H_0) \propto \frac{\int dz\, d\Omega\, \mathcal{L}^{GW}(x|z,\Omega,H_0)\, p_{pop}(z,\Omega)}{\int p^{GW}_{Det}(z,\Omega,H_0)\, p_{pop}(z,\Omega)\, dz\, d\Omega}$$

Note This time, we will construct $p_{pop}(z,\Omega)$ from the galaxy catalog. Namely we will have $p_{pop}(z,\Omega|CAT)$.

Let us start from the denominator. The denominator is an integral over all the sky. When averaged over all the sky, $p(z,\Omega|CAT) \simeq \frac{z^2}{4\pi}$

If the GW detection probability does not depend on a particular sky-patch,

$$\int p_{DET}(z, \Omega, H_0) \cdot \frac{z^2}{4\pi} \, dz \, d\Omega \propto H_0^3$$ as in

the previous case.

NOTE We can not make this approximation at the numerator, because the GW likelihood often can measure very well the sky position.

$$\mathcal{L}(X \mid z, H_0, \Omega) \propto \frac{p(d_L \mid X)}{\pi(d_L)} \frac{p(\Omega \mid X)}{\pi(\Omega)} \propto \frac{p(d_L \mid X)}{\pi(d_L)} \frac{\delta(\Omega - \hat{\Omega})}{1/4\pi}$$

NOTE We have made the additional assumption that estimation of $d_L$ not correlated with $\Omega$.

$$\Rightarrow \mathcal{L}(X \mid H_0) \propto \frac{1}{H_0^3} \int dz \frac{p(d_L(z, H_0) \mid X)}{\pi(d_L(z, H_0))} \cdot p_{pop}(z \mid \hat{\Omega})$$

DISTRIBUTION OF GALAXIES IN GW SKY PIXEL.

$$p_{pop}(z|\hat{\Omega}) = \frac{1}{N_{GAL}} \sum_{i=0}^{N_{GAL}} \delta(\hat{z}_i - z)$$

↳ MEASURED REDSHIFT FROM GALAXY CATALOG.

We can now make the integral in redshift.

$$\mathcal{L}(x|H_0) \propto \frac{1}{H_0^3} \frac{1}{N_{GAL}} \sum_{i=0}^{N_{GAL}} \frac{p(d_L(\hat{z}_i, H_0)|x)}{\pi(d_L(\hat{z}_i, H_0))}$$

NOTE: We have obtained that the dark siren analysis is basically a bright siren analysis with many possible hosts. The final posterior is the superposition of all the posterior.