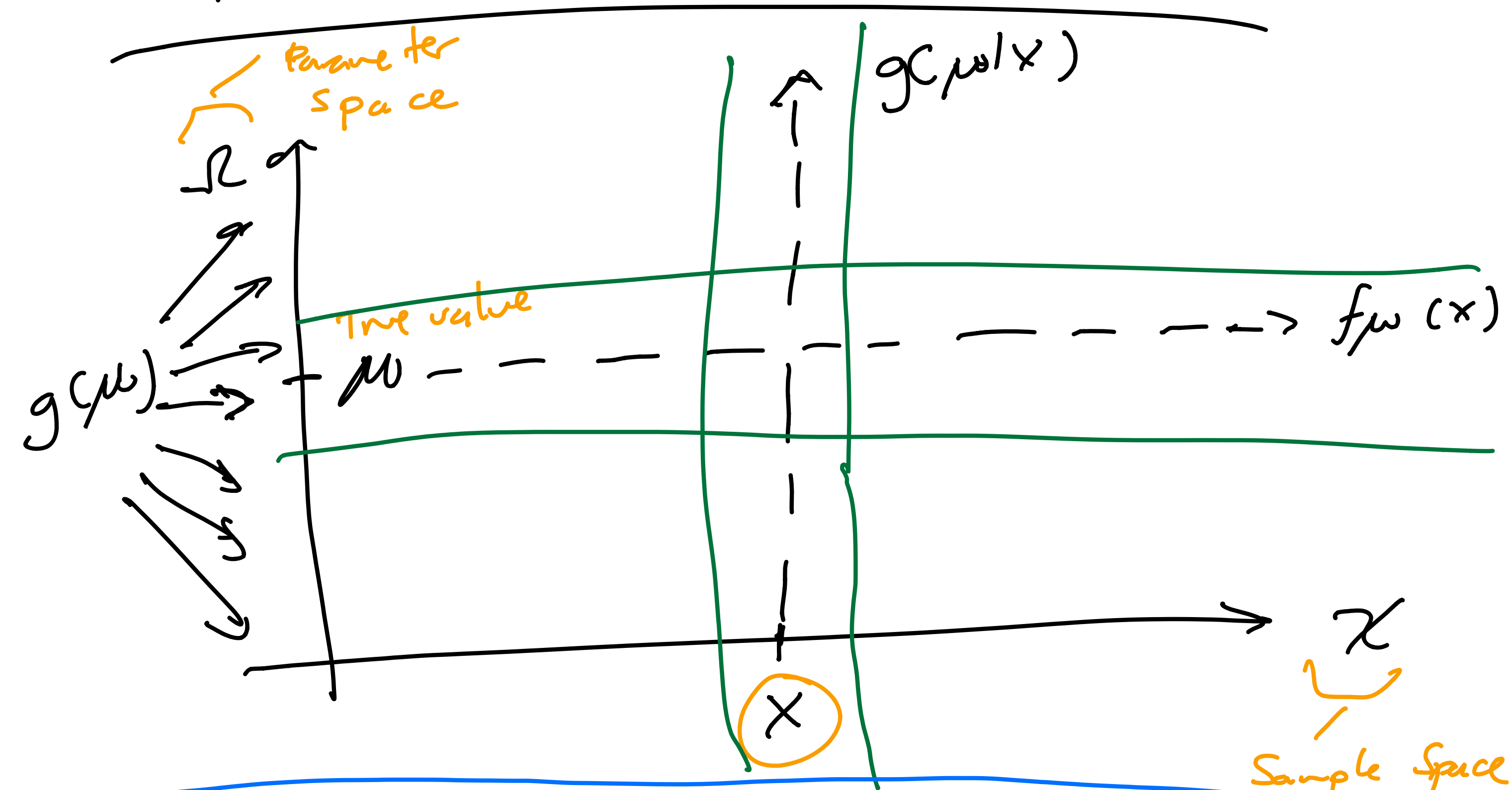
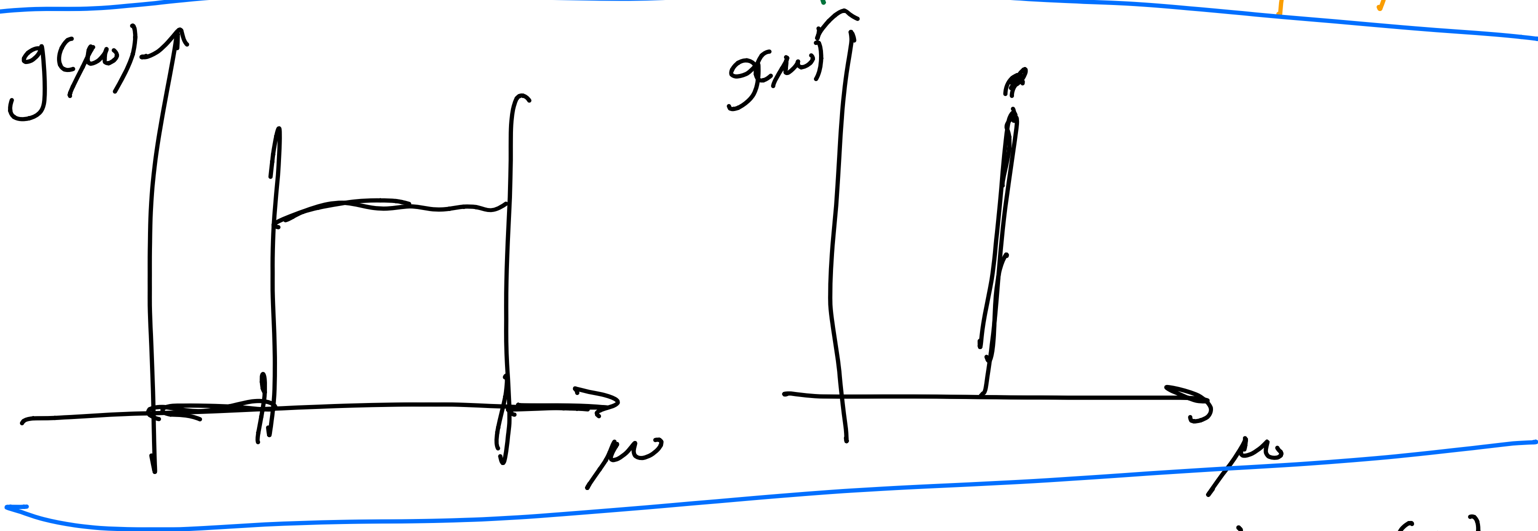
Bayesian / Frequentist Comparison List.





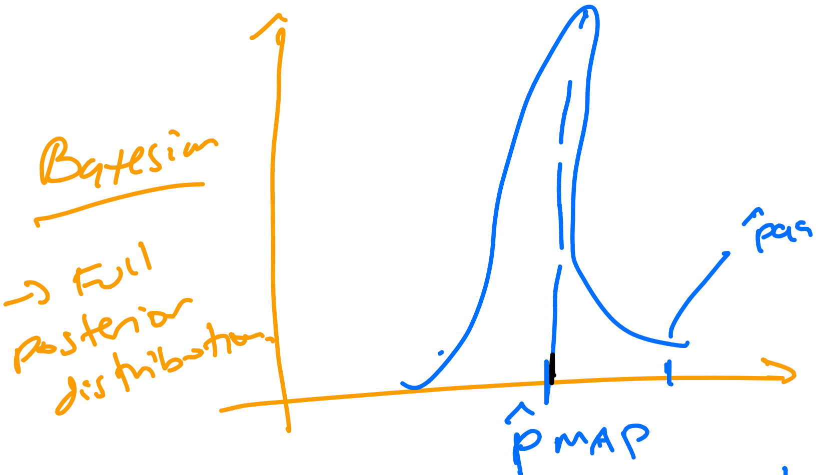
- Bayesian inference requires prior $g(\mu)$. when there is prior data / experience this very useful. If not you can pick the prior to be nearly uninformative.

- Modern data analysis problems usually have "go-to" algorithms. → Frequentist.

- Bayesian analysis answers all possible questions at once...

$$\bar{X} = \frac{K}{n}$$

Estimates $E(p)$.



(Maximum a posteriori estimate).

- Modern methods end up combining both kinds of reasoning to a degree.

CH 4. Fisherian Inference and Maximum Likelihood

Sir Ronald Fisher

- For a family of prob. densities $f_\mu(\underline{x})$ the log-likelihood function is defined as:

$$\underbrace{l_x}_{\text{Fixed}}(\underbrace{\mu}_{\text{Varying}}) = \log\{f_\mu(\underline{x})\}$$

The MLE is the value of $\mu \in \Omega$ that maximizes $l_x(\mu)$.

$$\hat{\mu}_{MLE} = \underset{\mu \in \Omega}{\arg\max} \{ l_x(\mu) \}$$

- Can also provide MLE's for a function of μ (i.e. $\theta = T(\mu)$) using $\hat{\theta} = T(\hat{\mu})$

Good properties about MLE

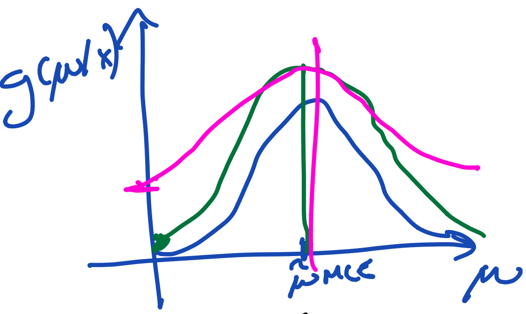
① Automatic

② Excellent frequentist properties (Bias & Variance)

③ Reasonable Bayesian Justification

$$g(\mu \mid \underline{x}) = c_x \cdot e^{l_x(\mu)}$$

Prior · normalization



- $\hat{\mu}^{MLE}$ is a maximizer of $g(\mu \mid \underline{x})$ if prior is constant.

- MLE only depends on $\mathcal{F}$ through the likelihood function. This means it also avoids the flaws of freq. inference we covered last time. (always same algorithm),

Example : — We'll use the g fn.

• We'll consider 2 potential families.

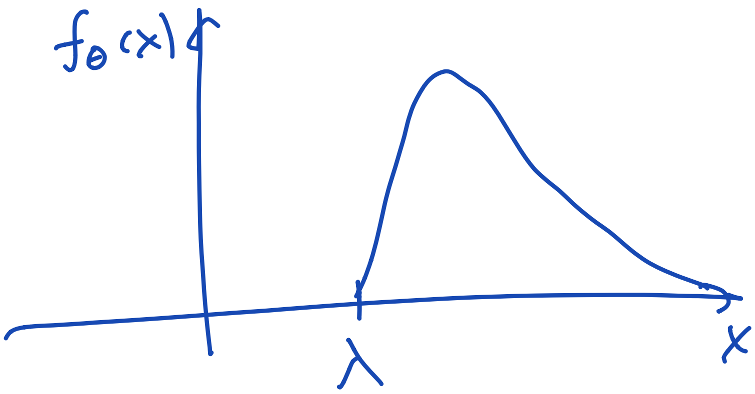
• Normal. Let $\underbrace{\theta}_{\text{Parameter vector}} = (\mu, \sigma)$

Then, $f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

• (shifted) Gamma: Let $\theta = (\lambda, \sigma, r)$

Then, $f_\theta(x) = \frac{(x-\lambda)^{r-1}}{\sigma^r \, \Gamma(r)} \cdot e^{-\frac{(x-\lambda)}{\sigma}}$ for $x \geq \lambda$, 0 otherwise.

$$f_\theta(x) = \frac{(x-\lambda)^{r-1}}{\sigma^r \, \Gamma(r)} \cdot e^{-\frac{(x-\lambda)}{\sigma}} \cdot I(x \geq \lambda)$$



Since $f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i)$ (Likelihood Function)

as we are considering iid observations ...

$$\underbrace{\ell_{\underline{x}}(\theta)}_{\text{Log-likelihood function.}} = \sum_{i=1}^{n} \log\{f_\theta(x_i)\} = \sum_{i=1}^{n} \ell_{x_i}(\theta)$$

Without proof...

For Normal densities

$$\hat{\mu}^{MLE} = \bar{X} \quad (\text{really? } ¬¬)$$

$$\hat{\sigma}^{MLE} = \left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}\right]^{1/2} \quad (\text{really? } ¬¬)$$

There does not exist $\longrightarrow$

For Shifted Gamma $\nexists$ closed form solution, so we'll numerically minimize in the computer,

where log = ln

$$f_\theta(x) = \frac{(x - \lambda)^{\gamma - 1}}{\sigma^\gamma \, \Gamma(\gamma)} \cdot e^{-\frac{(x-\lambda)}{\sigma}}$$

$$\log(f_\theta(x)) = (\gamma - 1) \cdot \log(x - \lambda) - \left[\gamma \log(\sigma) + \log(\Gamma(\gamma))\right] - \frac{(x-\lambda)}{\sigma}$$

- MLE can cause overfitting, identification and other problems in high dimensions.

However, regularized versions (such as using it with Lasso) mostly get around this issue.