Bayesian Inference

- As with frequentist inference, the objective is to learn properties about $\mathcal{F}$.

- Further let's emphasize that $\mathcal{F}$ is a density from a family of densities.

$$\mathcal{F} = \{ f_\mu(x) \, ; \, x \in \mathcal{X}, \mu \in \Omega \}$$

The density from which our data comes

a density for x, with parameter vector μ

Potential values x can take

x comes from a sample space.

μ from parameter space.

$\mathcal{X}$: sample space

μ (unobserved) is a point in the parameter space Ω.

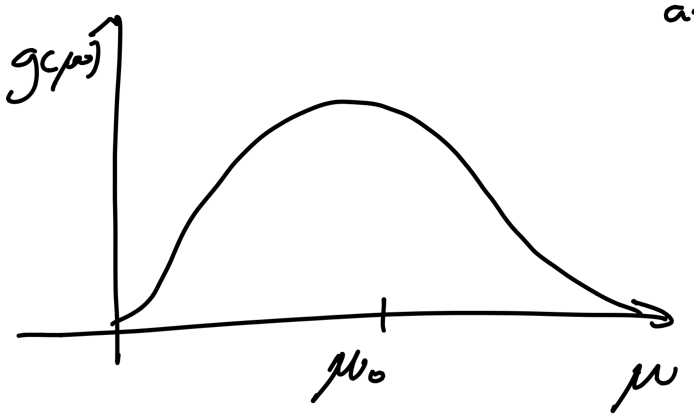
→ we observe $\underline{x}$ from $f_\mu(\underline{x})$ and we infer μ.

- Bayesian Inference adds one assumption: the knowledge of a prior density $g(\mu)$, $\mu \in \Omega$.

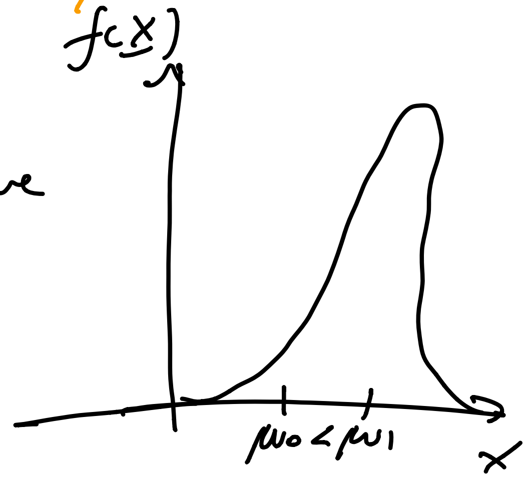
Bayes' Theorem is a rule for combining prior knowledge in $g(\mu)$ with current evidence in x.

Likelihood func.

Say, if you know



and you observe



How to consistently update our belief about μ?

Let $g(\mu | x)$ be the posterior density of μ. (after observing $\underline{x}$)

a.k.a $f(x|\mu)$

Bayes' Rule:

$$g(\mu | \underline{x}) = \frac{g(\mu) \cdot f_\mu(\underline{x})}{f(\underline{x})}$$

$$g(\mu | \underline{x}) \propto g(\mu) \cdot f_\mu(\underline{x})$$

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}}$$

Undergrad version

where $f(\underline{x})$ is the marginal density of $\underline{x}$

$$f(\underline{x}) = \int_{\Omega} f_\mu(\underline{x}) \cdot g(\mu) \, d\mu$$

Brutally differen from frequentist calculations.

Here $\underline{x}$ is fixed and our belief of μ changes with observed data!

– We can also write Bayes rule as:

$$g(\mu|x) = \underbrace{c_{\underline{x}}}_{\text{Constant that depends on } \underline{x}} \cdot f_\mu(\underline{x}) \cdot g(\mu)$$

Finally, for any two μ_1, μ_2 values on Ω, the ratio of posterior densities is given by:

$$\frac{g(\mu_1|x)}{g(\mu_2|x)} = \frac{\cancel{c_{\underline{x}}}\; g(\mu_1)\, f_{\mu_1}(\underline{x})}{\cancel{c_{\underline{x}}}\; g(\mu_2)\, f_{\mu_2}(\underline{x})}$$

"The posterior odds ratio is the prior odds ratio times the likelihood ratio".

A warmup Example

• Engineer knows she's having twins. She asks herself: what is the prob. that they'll be identical?

⇒ Doctor says: 1/3 of twin births were identical.

- Say that x is the sonogram result (either same sex or opposite sex) and same sex is observed.

- So, applying Bayes Rule.

$$\frac{g(\text{Identical} \mid \text{Same})}{g(\text{Fraternal} \mid \text{Same})} = \frac{g(\text{Identical})}{g(\text{Fraternal})} \cdot \frac{f_{\text{Identical}}(\text{Same})}{f_{\text{Fraternal}}(\text{Same})}$$

$$= \frac{1/3}{2/3} \cdot \frac{1}{1/2} = 1$$

- Fraternal and Identical are equally likely.

	Same	Diff	
Id	1/3	0	1/3
Fr	1/3	1/3	2/3

$g(\text{Identical} \mid \text{Same}) = .5$
$g(\text{Fraternal} \mid \text{Same}) = .5$

Increased from $\frac{1}{3}$ to $\frac{1}{2}$!

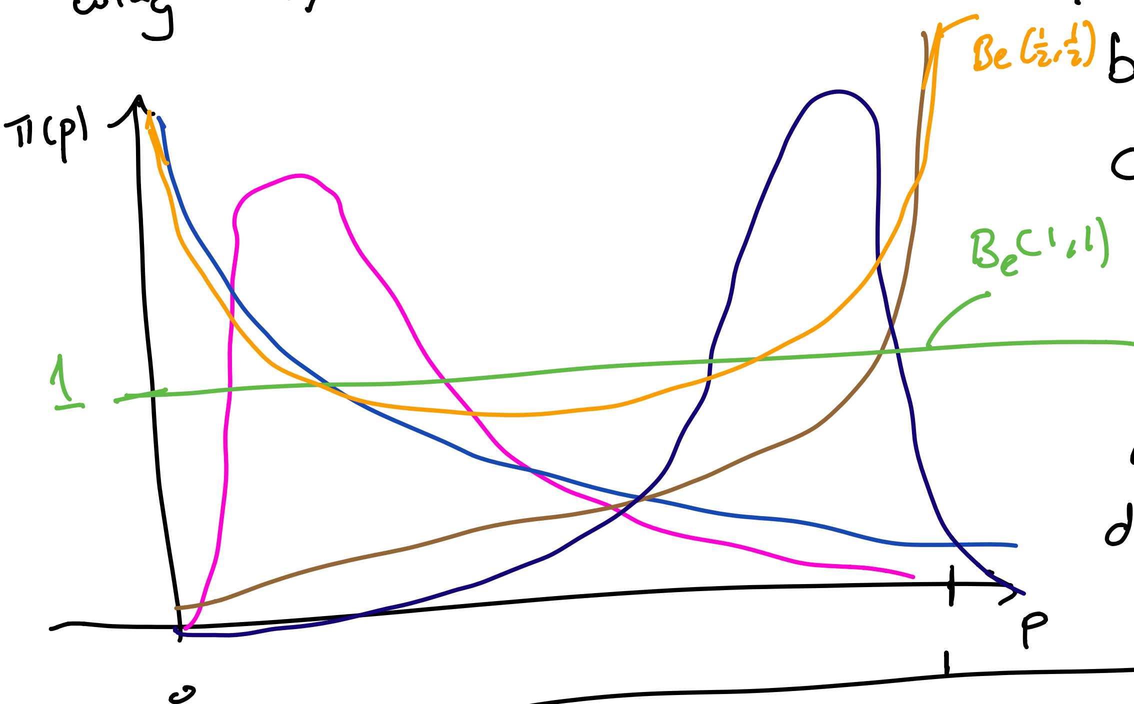
A more elaborate Bayesian Inference Example.

$X \sim Bin(n, p)$

$$P(X = k \mid p, n) = \underbrace{\binom{n}{k}}_{= \frac{n!}{k! \, (n-k)!}} \cdot p^k (1-p)^{n-k}$$

Prob. that # of successes is exactly k

Furthermore, let's assume that before observing our data our belief of p looks either of the following ways:



All of these beliefs about p can be modeled using the same distribution w/ different parameters

$$\boxed{\text{The Beta Distribution!}}$$

$p \sim \text{Beta}(\alpha, \beta)$ By definition of Beta density

$$g(p \mid \alpha, \beta) = \frac{p^{\alpha-1} \cdot (1-p)^{\beta-1}}{\boxed{B(\alpha, \beta)}}$$

to normalize the pdf to 1.

Beta function

$$B(\alpha, \beta) = \int_0^1 p^{\alpha-1} (1-p)^{\beta-1} \, dp$$

- The Beta function can be written in terms of the Gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-z} \, dx \quad \text{for } z > 0$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- Further, the Gamma function can be thought of as a continuous generalization of the factorial function.

$$\Gamma(n+1) = n!$$

Beta Distr. Properties

$$E(p \mid \alpha, \beta) = \frac{\alpha}{\alpha+\beta}$$

For our following analysis, it is convenient to reparameterize our Beta distribution such that the mean can be represented using one $\underline{1}$ parameter.

$$\underline{\mu = \frac{\alpha}{\alpha+\beta}} \; ; \; \underline{M = \alpha+\beta}$$

$$\rightarrow g(p \mid \mu, M) = \text{Beta}(M \cdot \mu, \; M \cdot (1-\mu))$$

By the way, under this parametrization

$$E(p \mid \mu, M) = \mu \; ; \; V(p \mid \mu, M) = \frac{\mu(1-\mu)}{M+1}$$

- Using all these definitions, we can finally compute posterior distr:

$$\underbrace{g(p \mid k)}_{\text{our belief about } p \text{ after observing } k \text{ successes in } n \text{ trials}} \; \underbrace{\propto}_{\text{Proportional to}} \; f_p(k) \cdot \overbrace{g(p \mid \underbrace{\mu, M}_{\text{Prior hyperparameters}})}^{\text{Prior}}$$

$$g(p \mid k) \propto p^{\underline{k}} (1-p)^{n-k} \cdot p^{\underline{M\mu - 1}} \cdot (1-p)^{M(1-\mu)-1}$$

$$[n-k+M(1-\mu)]-1$$

$$\propto p^{(k+M\mu)-1} \cdot (1-p)^{[\ldots]}$$

$$\propto \text{Beta}\left[k+\overbrace{M\mu}^{\text{Equivalent to "successes"}},\; n-k+\overbrace{M(1-\mu)}^{\text{Failures}}\right]$$

So, if the <u>prior</u> for p is <u>Beta</u> and the likelihood is <u>Binomial</u>, then, the <u>posterior</u> of p is also <u>Beta</u>.

This kind of prior is called a <u>conjugate</u> prior. i.e. the <u>posterior</u> is of the same family of the <u>prior</u>.