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.

Potential values x can take

μ from parameter space.

$$\underbrace{\mathcal{F}}_{\text{The density from which our data comes}} = \{ \underbrace{f_\mu(x)}_{\text{a density for } x\text{, with parameter vector } \mu} ; \underbrace{x \in \mathcal{X}}_{x \text{ comes from a sample space.}} , \mu \in \Omega \}$$

$\mathcal{X}$: sample space

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

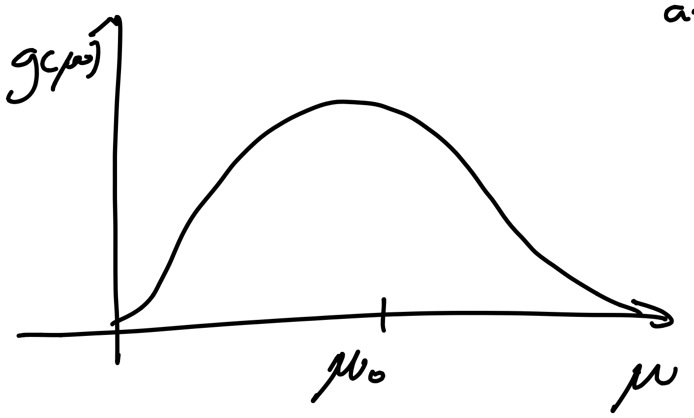
$\rightarrow$ we observe $\underset{\sim}{x}$ from $f_\mu(\underset{\sim}{x})$ and we infer μ.

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

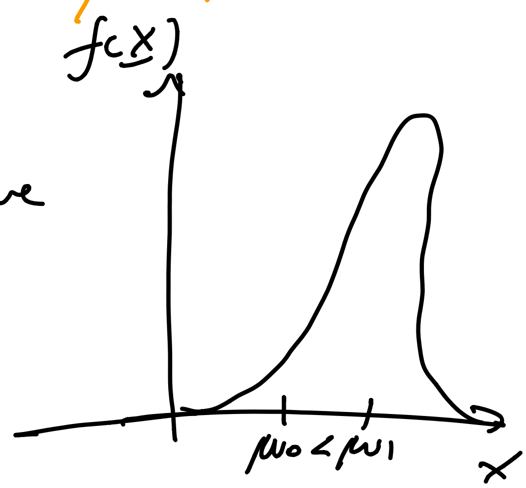
Bayes' Theorem is a rule for combining prior knowledge in $g(\mu)$ with current evidence in x.

Likelihood func.

Say, if you know



and you observe



How to consistently update our belief about μ?

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

<u>Bayes' Rule:</u>

$$g(\mu|\underline{x}) = \frac{g(\mu)\cdot \overbrace{f_\mu(\underline{x})}^{\text{a.k.a } f(x|\mu)}}{f(\underline{x})}$$

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

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

Undergrad version

where $f(\underline{x})$ is the <u>marginal density</u> 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".

$\boxed{\text{A warmup Example}}$

- 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 $\frac{1}{2}$!

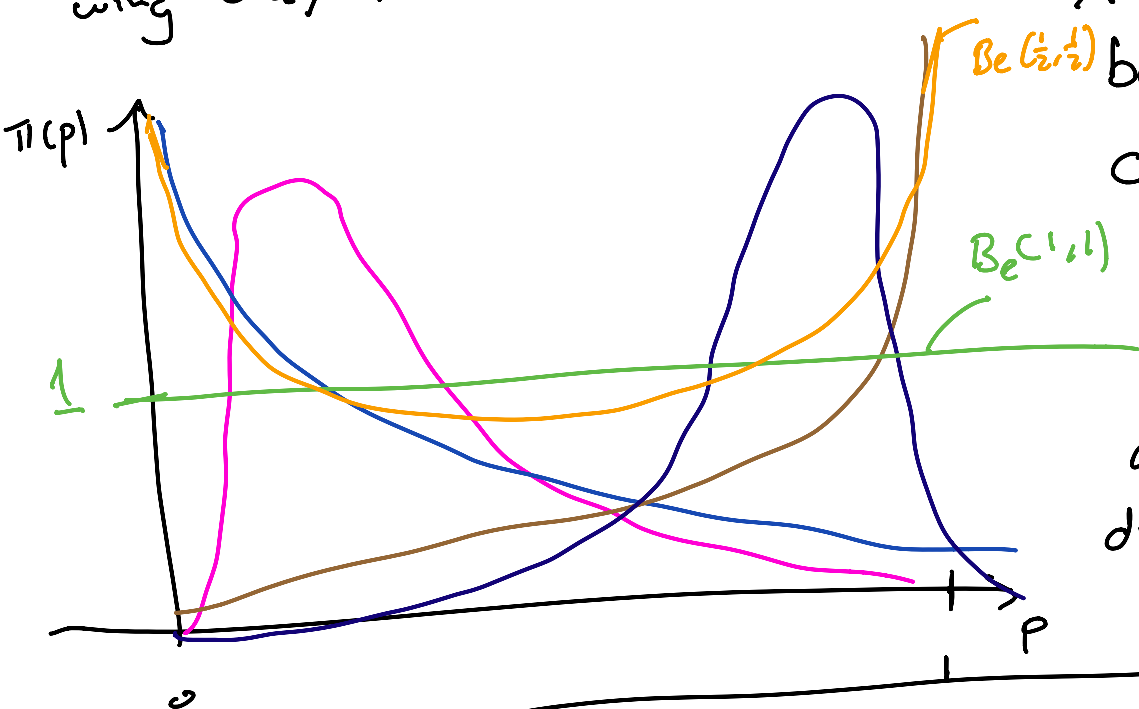
A more elaborate Bayesian Inference Example.

$X \sim Bin(n, p)$

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

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 <u>1</u> parameter.

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

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

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:

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

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

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

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

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

So, if the prior for p is Beta and the likelihood is Binomial, then, the posterior of p is also Beta ☺.

This kind of prior is called a conjugate prior. i.e. the posterior is of the same family of the prior.

→ Beta-Binomial is far from being the only conjugate pair of distributions. There are many more we can work with.

Finally, what is our prediction on the number of successes given our posterior for p?

. Using conditional probability, we can write:

$$\left[f(k \mid \mu, m) = \int_0^1 f(k \mid p) \cdot g(p \mid \mu, m)\, dp\right.$$

Marginal because it does not depend on p.

Any expressions playing the role of μ and M.

Prior

$\mathrm{Beta}(M\mu, M(1-\mu))$

$\alpha \quad \beta$

$$f(k \mid \mu, M) = \int_0^1 \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot \frac{p^{M\mu - 1} (1-p)^{M(1-\mu)-1}}{B(\mu, M)}$$

$$= \frac{\binom{n}{k}}{B(\mu, M)} \cdot \boxed{\int_0^1 p^{k+M\mu-1} (1-p)^{n-k+M(1-\mu)-1} \, dp}$$

$B(\cdot, \cdot)$

$$\frac{n!}{k!\,(n-k)!} \;\to\; \frac{\Gamma(n+1)}{\Gamma(k+1)\,\Gamma(n-k+1)}$$

$$= \frac{\Gamma(n+1)}{\Gamma(k+1)\,\Gamma(n-k+1)} \cdot \frac{\Gamma(M)}{\Gamma(M\mu)\,\Gamma(M(1-\mu))} \cdot \frac{\Gamma(k+M\mu) \cdot \Gamma(n-k+M(1-\mu))}{\Gamma(n+M)}$$

Done :)

$$B(\underset{M\mu}{\alpha}, \underset{M(1-\mu)}{\beta}) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)} \Rightarrow \frac{1}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}$$

$$= \frac{\Gamma(M)}{\Gamma(M\mu)\,\Gamma(M(1-\mu))}$$