02/18

- grading : on-going
- HW#3: HW#2 - Q1
  Due: 5pm 02/18 (F)

Q3: clarified

Q10: $y_i = x_i'\beta + \varepsilon_i$

- The Fisher information $I(\theta)$ in the random variable X is defined as

$$I(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial \log f(x \mid \theta)}{\partial \theta}\right)^2\right].$$

$\Rightarrow$ the expected slope of $\log f(x \mid \theta)$

$\Rightarrow$ measure the amount of information that a sample of data contains about unknown parameters.

- Under commonly satisfied conditions (true for exponential families),

$$\mathrm{E}_\theta\left[\left(\frac{\partial \log f(x \mid \theta)}{\partial \theta}\right)^2\right] = -\mathrm{E}_\theta\left[\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}\right].$$

- **Ex 3.5.4** Consider $x \sim \text{Bin}(n, p)$. Find the Fisher information, $I(p)$.

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\log f(x|p) = \log \binom{n}{x} + x \log p + (n-x) \log (1-p)$$

$$\frac{\partial \log f(x|p)}{\partial p} = \frac{x}{p} + \frac{n-x}{(1-p)} \cdot (-1)$$

$$\frac{\partial^2 \log f(x|p)}{\partial p^2} = -\frac{x}{p^2} - \frac{(n-x)}{(1-p)^2}$$

$$\rightarrow I(p) = -E_\theta\left(\frac{\partial^2 \log (f(x|p))}{\partial p^2}\right) = -E_\theta\left(-\frac{x}{p^2} - \frac{(n-x)}{(1-p)^2}\right)$$

$$= + \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} = \frac{n}{p} + \frac{n}{(1-p)} = \frac{n}{p(1-p)}$$

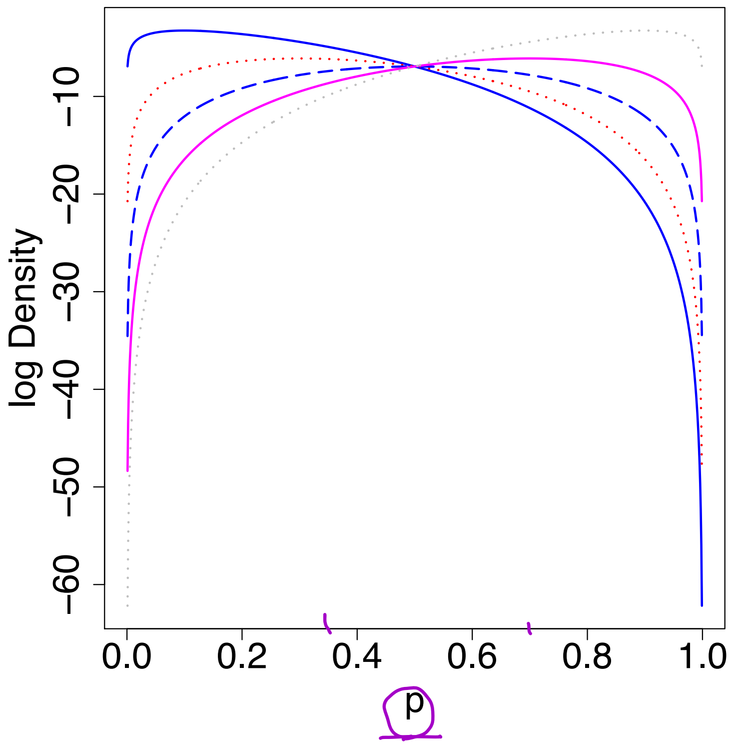
$$\pi^*(p) = Be(1/2, 1/2)$$

$$\Rightarrow \pi^*(p) \propto \sqrt{I(p)} \propto \left(\frac{1}{p(1-p)}\right)^{\frac{1}{2}} = p^{-1/2} (1-p)^{-1/2}$$

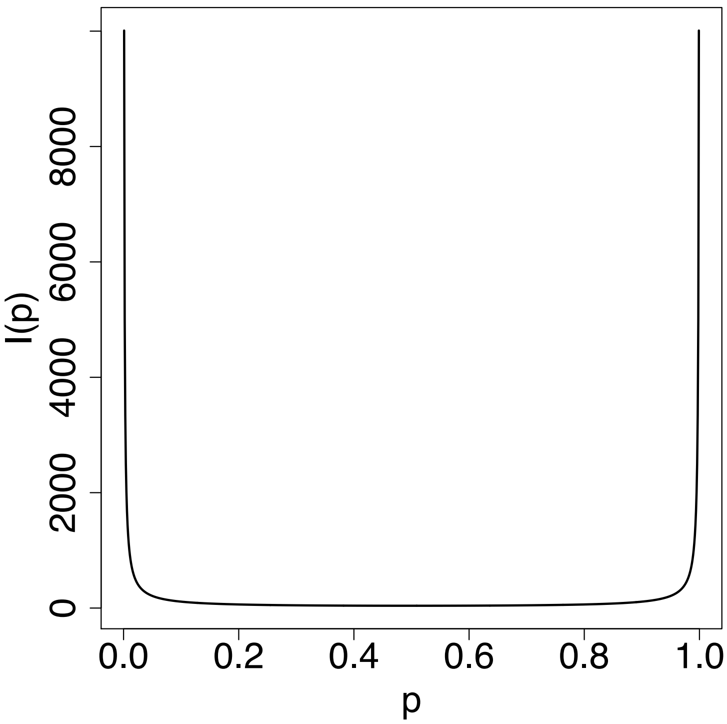
Bin(n, p) , $p \in (0, 1)$

♣ **Ex 3.5.4**: $\log(f(x \mid p))$ with $n = 10$ and $x = 1, 3, \ldots, 9$.

$\binom{n}{x} p^x (1-p)^{n-x}$



♣ **Ex 3.5.4**: $I(p)$



† The Jeffreys Prior

- Jeffreys Prior: noninformative priors in general settings based on Fisher information.

$$\pi^{\star}(\theta) \propto [I(\theta)]^{1/2}.$$

- Recall $I(\theta)$: an indicator of the amount of information brought by the sampling model (or the observation) about θ.

  i.e. large $I(\theta) \Rightarrow$ more sample info to discriminate between θ and $\theta + d\theta$.

  $\Rightarrow$ Assign more prior probability to the values that have large $I(\theta)$ so that the influence of the prior distribution is minimized.

- Jeffreys Prior:

$$\pi^{\star}(\theta) \propto [I(\theta)]^{1/2}.$$

- defined up to a normalizing constant when $\pi^{\star}(\theta)$ is proper.

- Observe for any one-to-one transform $h(\theta)$

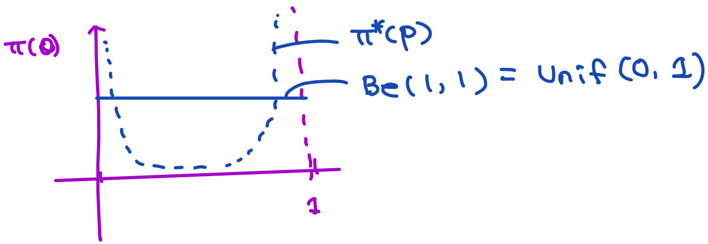
$$I(\theta) = I(h(\theta))(h'(\theta))^2$$

$\Rightarrow$ The invariant reparameterization requirement is satisfied.

- **Ex 3.5.4** Consider $x \sim \text{Bin}(n, p)$. Find the Jeffreys prior for this model.

$$\pi^*(p) \propto p^{-1/2}(1-p)^{-1/2} \Rightarrow Be\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

$p \in (0, 1)$



† The Jeffreys Prior (contd)

- For multidimensional $\boldsymbol{\theta} \in \mathbb{R}^p$,

$$\pi^*(\boldsymbol{\theta}) = \{\det(\underbrace{\boxed{\boldsymbol{I}(\boldsymbol{\theta})}}_{\substack{\text{}}}\overset{}{})\}^{1/2},$$

$p \times p$

where $\boldsymbol{I}$: $p \times p$ Fisher information matrix.

- Under commonly satisfied conditions (true for exponential families),

$$\underline{I_{ij}(\boldsymbol{\theta})} = -\mathrm{E}_\theta \left[\frac{\partial^2}{\partial\ \theta_i \partial\ \theta_j} \log f(x \mid \boldsymbol{\theta}) \right].$$

- **Ex 3.5.6** Consider $x \sim \mathrm{N}(\mu, \sigma^2)$ with μ and $\boxed{\sigma}$ unknown. Find the Jeffreys prior for this model.

⋆⋆ Jeffreys (1961) was mainly emphasizing the use of the Jeffreys priors in the one-dimensional case.

$$f(x \mid \theta, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$$

$$\underbrace{I(\theta,\sigma)}_{2\times 2} = \begin{bmatrix} \bigcirc & \\ & \end{bmatrix}$$

$$\log(f(x \mid \theta,\sigma)) = -\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{(x-\theta)^2}{2\sigma^2}$$

$$\frac{\partial \log(f(x \mid \theta,\sigma))}{\partial \theta} = \frac{2(x-\theta)}{2\sigma^2} = \frac{(x-\theta)}{\sigma^2}$$

$$\frac{\partial \log(f(x \mid \theta,\sigma))}{\partial \sigma} = -\frac{1}{\sigma} + \frac{2(x-\theta)^2}{2\sigma^3} = -\frac{1}{\sigma} + \frac{(x-\theta)^2}{\sigma^3}$$

$$\frac{\partial^2 \log(f(x \mid \theta,\sigma))}{\partial \theta^2} = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2 \log(f(x \mid \theta,\sigma))}{\partial \sigma^2} = \frac{1}{\sigma^2} - \frac{3(x-\theta)^2}{\sigma^4}$$

$$\frac{\partial^2 \log(f(x \mid \theta,\sigma))}{\partial \theta\, \partial \sigma} = -\frac{2(x-\theta)}{\sigma^3}$$

$$I(\theta,\sigma) = \begin{bmatrix} -E\left(-\frac{1}{\sigma^2}\right) & -E\left(-\frac{2(x-\theta)}{\sigma^2}\right) \\ \text{''} & -E\left(\frac{1}{\sigma^2} - \frac{3(x-\theta)^2}{\sigma^4}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \underbrace{-\frac{1}{\sigma^2} + \frac{3\cancel{\sigma^2}}{\cancel{\sigma^4}\,2}} \end{bmatrix}$$

$$= -\frac{1}{\sigma^2} + \frac{3}{\sigma^2}$$

$$= \frac{2}{\sigma^2}$$

$$|I(\theta,\sigma)| = \frac{1}{\sigma^2} \cdot \frac{2}{\sigma^2}$$

$$\pi^*(\theta,\sigma) \propto |I(\theta,\sigma)|^{1/2} \propto \frac{1}{\sigma^2}$$

"joint" Jefferys prior of θ and "σ"

$\eta = \sigma^2$

$$\pi^*(\theta, \sigma^2) \propto \frac{1}{\sigma^4}$$

① θ unknown σ known

$$\pi_1^*(\theta) \propto \sqrt{\underbrace{-E\left(\frac{\partial^2 \log f(x \mid \theta,\sigma)}{\partial \theta^2}\right)}_{I(\theta)}}$$

$$\propto \underbrace{\sqrt{\frac{1}{\sigma^2}}}$$

$$\propto 1$$

② θ known σ unknown

$$\pi_2^*(\sigma) \propto \sqrt{\frac{2}{\sigma^2}}$$

$$\propto \frac{1}{\sigma}$$

$$\Rightarrow \pi_2^*(\sigma^2) \propto \frac{1}{\sigma^2}$$

$$\pi(\theta,\sigma) = \pi_1(\theta)\,\pi_2(\sigma)$$

$$\propto 1 \cdot \frac{1}{\sigma}$$

$\downarrow$

$\cdot\ \pi(\theta \mid \sigma, x) \approx N$

$\pi(\sigma^2 \mid x) = IG$

- **Any drawback?** The use of Jeffreys priors does not satisfy the Likelihood Principle!

Bin (n, p) $\Rightarrow$ $\pi^*(p) = Be(1/2, 1/2)$

- **Ex 3.5.7** Consider $n \mid p \sim \mathrm{NB}(x, p)$ where n: total # of trials, x: # of successes (pre-determined), p: prob. of success. Find the Jeffreys prior for this model.

$$p(n \mid p) = \binom{n-1}{x-1} p^x (1-p)^{n-x}$$

$$\Rightarrow \quad \tilde{\pi}^* \propto p^{-1} (1-p)^{-1/2}$$

$$\neq p^{-1/2} (1-p)^{-1/2}$$

† Reference Priors: Read CR §3.5.4.

- For multidimensional $\theta = (\theta_1, \theta_2)$, distinguish between θ_1 (parameters of primary interest) and θ_2 (nuisance parameters)
  - ⋆⋆ Conditional on θ_1, define prior $\pi(\theta_2 \mid \theta_1)$ as the Jeffreys prior associated with $f(x \mid \theta)$ when θ_1 fixed.
  - ⋆⋆ Integrate out θ_2 according to $\pi(\theta_2 \mid \theta_1)$,

$$p(x \mid \theta_1) = \int_{\Theta_2} p(x \mid \theta_1, \theta_2)\pi(\theta_2 \mid \theta_1) d\theta_2.$$

  - ⋆⋆ Find the Jeffreys prior $\pi(\theta_1)$ based on $p(x \mid \theta_1)$.
  - ⋆⋆ Set $\pi(\theta_1, \theta_2) = \pi(\theta_1)\pi(\theta_2 \mid \theta_1)$.

  Note: Choosing different nuisance parameters generates different priors.

- For one dimensional θ, the reference prior is identical to the Jeffreys prior

$$\begin{cases} x \mid \theta \sim N(\theta, \sigma^2) \\ \theta \sim N(\mu, \tau^2) \end{cases}$$

$$\begin{cases} \theta \sim N(\underline{\mu}, \underline{10}) \leftarrow \\ \theta \sim N(\mu, \underline{10^5}) \end{cases}$$

$\dagger$ Objectivity (JB §3.7)

① $f(x \mid \theta)$

② $L(\theta, d)$

→ ③ $\pi(\theta)$

- To most non-Bayesians, classical statistics is "objective", while Bayesian statistics is "subjective".

- No method of analysis is fundamentally "objective": e.g. the choice of a loss function is almost always subjective.

  ⋆⋆ The use of a prior distribution introduces another subjective feature into the analysis.

- When there is an overwhelming amount of data, virtually *any* analysis would yield the same conclusion.

  ⋆⋆ In such cases, choices of feature as $f(x \mid \theta)$ will have a serious bearing on the conclusion.

† Posterior Validation and Robustness (CR 3.6, JB 4.7.1 and 4.7.2)

- Do slight changes in the prior distribution cause significant changes in the decision and, if so, what should be done?

- JB Example 2, p111 We observe $X \sim \mathrm{N}(\theta, 1)$ and subjectively specify a prior median of 0 and prior quantiles of ± 1.

  ⋆⋆ Either the $C(0, 1)$ (π_C) or $\mathrm{N}(0, 2.19)(\pi_N)$ densities are thought to be reasonable matches to prior beliefs.

  ⋆⋆ The difference between the two priors is mainly in the functional form (which is difficult to determine).

$x \approx 0$

$|x| \gg 0$

  ⋆⋆ Suppose θ is estimated under the squared-error loss, so that the posterior means will be used. Does it matter whether we use π_C or π_N?

- JB Ex, p 111 & p 195(contd)

Table 4.7. Posterior Means.

x	0	1	2	4.5	10
$\delta^C(x)$	0	0.52	1.27	4.09	9.80
$\delta^N(x)$	0	0.69	1.37	3.09	6.87

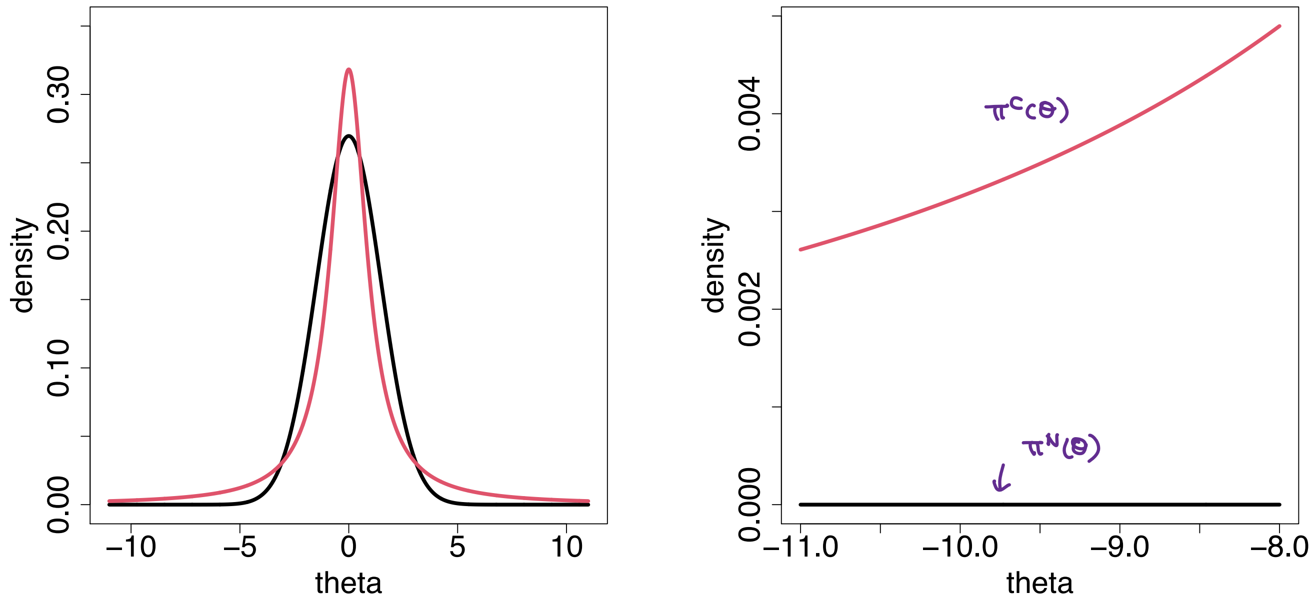
⋆⋆ For small x, $\delta^C(x)$ and $\delta^N(x)$ are quite close $\Rightarrow$ indicate some degree of robustness with respect to choice of the prior

⋆⋆ For moderate or large x, substantial difference between $\delta^C(x)$ and $\delta^N(x)$ $\Rightarrow$ not robust to reasonable variation in the prior

• JB Ex, p 111 & p 195(contd): Compare priors via $m(x)$ and eliminate from considerations priors which seem to be ruled out by data.

x	0	4.5	6.0	10
$m(x\mid\pi_N)$	0.22	0.0093	0.00079	3.5×10^{-8}
$m(x\mid\pi_C)$	0.21	0.018	0.0094	0.0032

• JB Ex, p 111 & p 195(contd): One of the main sources of nonrobustness in estimation (in specific for moderate or large x), will be seen to be the **degree of flatness of the prior tail**.



- Sensitivity of Bayesian analysis to possible misspecification of the prior distribution (assuming the likelihood is known).

  ⋆⋆ Try different reasonable priors and evaluate how a change in the prior changes the inference about the parameter of interest.

- How to robustify our priors?

  ⋆⋆ Robust prior distributions: Parameterized distributions as insensitive as possible to small variations in the prior information.

  e.g. $t-$distributions are preferable to normal priors in the normal case.

- Robustify the conjugate priors by *hierarchical modeling*. e.g.;

$$\begin{aligned} \lambda &\sim \pi_2(\lambda), \\ \theta \mid \lambda &\sim \pi_1(\theta \mid \lambda), \\ x \mid \theta &\sim f(x \mid \theta), \\ \Rightarrow \quad \pi(\theta) &= \int \pi_1(\theta \mid \lambda)\pi_2(\lambda)d\lambda. \end{aligned}$$

An additional level in the prior modeling increases the robustness of the prior distributions.

STAT 206B
Topic: Exchangeability & de Fenetti's Theorem

Winter 2022

† Exchangeability

- PH Section 2.7 (CR 3.8.2 or JB 3.5.7)
- So far we have been taking the notion of the model $y_i \overset{iid}{\sim} p(y \mid \theta)$ for observations $y_1, \ldots, y_n$.
- For this topic, let's focus on the actual distribution of the outcomes $(y_1, \ldots, y_n)$.
- Conditional Independence and Exchangeability.
- de Fenetti's Theorem.

PH Example p27 Participants in the 1998 General Social Survey were asked whether or not they were generally happy. Let Y_i be the random variable associated with this questions, so that

$$Y_i = \begin{cases} 1 & \text{if participant } i \text{ says that they are generally happy,} \\ 0 & \textit{otherwise}. \end{cases}$$

- Suppose someone told you the numerical value of θ, the rate of happiness among the 1272 respondents to the question.

- We assume 10 participants are sampled randomly ($10 \ll 1271$) and build a probability model for $p(y_1, \ldots, y_n)$ with $n = 10$.

PH Example p27 (contd) We may build a model as follow;

- There is some common numerical value of θ, the rate of happiness among the respondents to the question.

- It is reasonable to assume the Y_i's as conditionally independent and identically distributed given θ (or at least approximately so due to the finite population).

$$
\begin{aligned}
\Pr(Y_i = y_i \mid \theta, Y_j = y_j, i \neq j) &= \theta^{y_i}(1-\theta)^{1-y_i} \\
\Pr(Y_1 = y_1, \ldots, Y_{10} = y_{10} \mid \theta) &= \prod_{i=1}^{10} \theta^{y_i}(1-\theta)^{1-y_i} \\
&= \theta^{\sum_{i=1}^{10} y_i}(1-\theta)^{\sum_{i=1}^{10}(1-y_i)}
\end{aligned}
$$

- Let's think about a distribution $p(y_1, \ldots, y_n)$ describing the actual observables.

† PH Def 3 (or Def 3.8.2): Finite Exchangeability

Let $p(y_1, \ldots, y_n)$ be the joint density of $Y_1, \ldots, Y_n$. If

$$p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n})$$

for all permutations π of $\{1, \ldots, n\}$, then $Y_1, \ldots, Y_n$ are exchangeable.

n= 3

$p(Y_1, Y_2, Y_3)$

$= p(Y_2, Y_1, Y_3)$

$\{1, 0$

$1, 0, 1$

- Exchangeable if reordering of $(Y_1, \ldots, Y_n)$ does not change the joint distribution of $(Y_1, \ldots, Y_n)$.

  $\Leftrightarrow$ Roughly speaking, $Y_1, \ldots, Y_n$ are exchangeable if the subscripts (the "labels" identifying the individual random quantities) are **uninformative**

- Exchangeable random variables are not necessarily independent.

† Conditional Independence (continue PH Example)

- If θ is uncertain, we describe our belief about it with $\pi(\theta)$. The marginal joint distribution of $Y_1, \ldots, Y_n$ is then

$$
\begin{aligned}
p(y_1, \ldots, y_{10}) &= \int_0^1 p(y_1, \ldots, y_{10} \mid \theta)\pi(\theta) d\theta \\
&= \int_0^1 \theta^{\sum_{i=1}^{10} y_i}(1-\theta)^{\sum_{i=1}^{10}(1-y_i)}\pi(\theta) d\theta
\end{aligned}
$$

- For example, we have

\# of 1s = 6 , n=10 ,

$$
\begin{aligned}
p(1,0,0,1,0,1,1,0,1,1) &= \int_0^1 \theta^6(1-\theta)^4\pi(\theta) d\theta \\
p(1,1,0,0,1,1,0,0,1,1) &= \int_0^1 \theta^6(1-\theta)^4\pi(\theta) d\theta
\end{aligned}
$$

$\Leftrightarrow$ $Y_1, \ldots, Y_n$ **are exchangeable under this model of beliefs!**

† Summary!

- **Claim** If $\theta \sim \pi(\theta)$ and $Y_1, \ldots, Y_n$ are conditionally i.i.d given θ, then marginally (unconditionally on θ), $Y_1, \ldots, Y_n$ are exchangeable.

  Proof Suppose $Y_1, \ldots, Y_n$ are conditionally iid given some unknown parameter θ. Then for any permutation π of $\{1, \ldots, n\}$ and any set of values $(y_1, \ldots, y_n) \in \mathcal{Y}^n$,

$$
\begin{aligned}
p(y_1, \ldots, y_n) &= \int p(y_1, \ldots, y_n \mid \theta)\pi(\theta)d\theta \\
&= \int \prod_{i=1}^{n} p(y_i \mid \theta)\pi(\theta)d\theta \\
&= \int \prod_{i=1}^{n} p(y_{\pi_i} \mid \theta)\pi(\theta)d\theta \\
&= p(y_{\pi_1}, \ldots, y_{\pi_n})
\end{aligned}
$$

† So far we have

$$\left.\begin{array}{l} Y_1, \ldots, Y_n \mid \theta \quad \text{i.i.d} \\ \theta \sim \pi(\theta) \end{array}\right\} \underset{\overset{?!}{\Leftarrow}}{\Rightarrow} Y_1, \ldots, Y_n \quad \text{are exchangeable!} \tag{1}$$

Then how about the opposite direction?

† **Infinite Exchangeability** The infinite sequence of random quantities $y_1, y_2, \ldots$ is said to be judged (infinitely) exchangeable if every finite subsequence is judged exchangeable in the sense of Def 1 Finite exchangeability.

† PH Thorem 1: de Finetti's Theorem

Let $Y_i \in \mathcal{Y}$ for all $i \in \{1, 2, \ldots\}$. Suppose that for any n, our belief model for $Y_1, \ldots, Y_n$ is exchangeable:

$$p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n})$$

for all permutations π of $\{1, \ldots, n\}$. Then our model can be written as

$$p(y_1, \ldots, y_n) = \int p(y_1, \ldots, y_n \mid \theta)\pi(\theta) d\theta$$

for some parameter θ, some prior distribution on θ and some sampling model $p(y \mid \theta)$. The prior and sampling model depend on the form of the belief model $p(y_1, \ldots, y_n)$.

† **Continue PH example** de Finetti's Theorem says that

- There exists a probability distribution $\pi(\theta)$ such that, for every n, the joint distribution of $(Y_1, \ldots, Y_n)$ writes down as

$$p(y_1, \ldots, y_n) = \int_0^1 \underbrace{\prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{(1-y_i)}}\pi(\theta)d\theta,$$

  that is, conditional on θ, the y_i's iid Beroulli Ber(θ) random variables where $\pi(\theta)$ represents our beliefs about $\lim_{n\to\infty}\sum Y_i/n$.

† **For more general cases** where the y_i's are real valued and infinitely exchangeable, **de Finetti's Theorem** says that there exists an interesting representation, under the form

$$p(y_1, \ldots, y_n) = \int p(y_1, \ldots, y_n \mid \theta)\pi(\theta)d\theta.$$

This implies

- The assumption of exchangeability for every n

$$\Updownarrow$$

- There exists a likelihood $p(y_1, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta)$ for $y_1, \ldots, y_n$ (that is, conditionally independent).
- There exist a prior distribution on θ, $\pi(\theta)$ that represents our beliefs about the outcomes of $\{Y_1, Y_2, \ldots\}$ induced by our belief model $p(y_1, y_2, \ldots)$.

- Is infinite exchangeability reasonable?

  Possibly good approximation in many cases, but sometimes too restrictive. To be less restrictive, may consider partial exchangeability...

- Exchangeability will hold if the labels convey no information.

- When is the condition "$Y_1, \ldots, Y_n$ are exchangeable for all n" reasonable?

  For this condition to hold, we must have exchangeability and repeatability.

STAT 206B
Chapter 6: Bayesian Calculations

Winter 2022

† Bayesian Inference

- So far we have discussed the selection of adequate loss function and prior.

- e.g. Consider the problem of estimating $g(\theta)$ under the squared error loss function,

$= \pi(\theta \mid x)$

$$E^{\pi}(g(\theta) \mid x) = \int_{\Theta} g(\theta)\pi(\theta \mid x)d\theta = \frac{\int_{\Theta} \overset{\theta}{g(\theta)} f(x \mid \theta)\pi(\theta)d\theta}{\int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta}$$

  ⋆⋆ Evaluating the integrals can be quite difficult especially when Θ is high dimensional.

  ⋆⋆ Adapting a different loss function usually makes calculation even more difficult.

- CR Chapter 6 and JB Sections 4.9 & 4.10

- **Example 6.1.1** Consider $x_1, \ldots, x_n$ a i.i.d. sample from $C(\theta, 1)$, a Cauchy distribution with location parameter θ, and $\theta \sim \mathrm{N}(\mu, \sigma^2)$ with known hyperparameters μ and σ^2. The posterior distribution of θ is then

$$\pi(\theta \mid \boldsymbol{x}) \propto \underbrace{\overbrace{e^{-(\theta-\mu)^2/(2\sigma^2)}}^{\pi(\theta)} \overbrace{\prod_{i=1}^{n} \{1 + (x_i - \theta)^2\}^{-1}}^{f(x|\theta)}}_{= A(\theta)}.$$

$$= \frac{1}{C}\left(\qquad \right)$$

$\star\star$ How to make an inference about $g(\theta)$, e.g., point estimates, testing or interval estimates?

$$C = \int_{\Theta} A(\theta)\, d\theta$$

$g(\theta) = \theta$

- **Example 6.1.1**(contd) Under the squared error loss function, the Bayes estimator δ^π of θ is the posterior mean,

$$\delta^\pi(\boldsymbol{x}) = \frac{\int_{-\infty}^{\infty} \cancel{\tfrac{1}{c}}\, \theta e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^n \{1 + (x_i - \theta)^2\}^{-1} d\theta}{\int_{-\infty}^{\infty} \cancel{\tfrac{1}{c}}\, e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^n \{1 + (x_i - \theta)^2\}^{-1} d\theta}.$$

⋆⋆ Observe that closed form integrals are not available.

⋆⋆ Observe the calculation requires two numerical integrations.

⋆⋆ If we want to compute the posterior variance, it requires an additional integration, $\int_{-\infty}^{\infty} \theta^2 e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^n \{1 + (x_i - \theta)^2\}^{-1} d\theta$.

$E^\pi(\theta^2 | x) =$

† We will focus on approximations to $\pi(\theta \mid x)$ and integrals involving $\pi(\theta \mid x)$.

- Classical approximation methods (CR 6.2)

  Monte Carlo integration, Laplace analytic approximation.

- Markov chain Monte Carlo methods (CR 6.3)

  Metropolis-Hastings algorithms, the Gibbs sampler.

Note 1: Similar techniques can be used to predictive distributions, $f(y \mid x)$.

Note 2: Read Robert and Casella (1999) "Monte Carlo Statistical Methods" for more.

† Classical Approximation Methods (CR 6.2)

Recall we consider the posterior inference problem,

$$E^{\pi}(g(\theta) \mid x) = \int_{\Theta} g(\theta)\pi(\theta \mid x)d\theta = \frac{\int_{\Theta} g(\theta)f(x \mid \theta)\pi(\theta)d\theta}{\int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta}$$

- Laplace Approximation: attempt to analytically approximate the integral (does not require simulation)
- Monte Carlo methods
- Also see PH Chapter 4 (many examples with R code!) and/or Robert and Casella Chapter 3.

† Normal Approx. to Posterior - (1)

- General Idea: find a Gaussian approximation to $\pi(\theta \mid \boldsymbol{x})$.

- Consider a univariate case;

$$\pi(\theta \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{m(\boldsymbol{x})} \propto q(\theta)$$

⋆⋆ We find θ_0 and A such that $\pi(\theta \mid \boldsymbol{x}) \approx \mathrm{N}(\theta_0, A^{-1})$.