De Finetti’s Theorem

Chaithanya and Shravan

Priors in Bayes rule

- Bayes rule:

parameters observations

priors on parameters

$$\underbrace{P(\theta|x)}_{} = \frac{P(x|\theta)P(\theta)}{P(x)} \propto P(x|\theta)\boxed{P(\theta)}$$

- Independently discovered by Bayes and Laplace in 18th century.
- Priors are part of our assumptions and can affect our conclusions directly or indirectly.
- So how do we choose the right priors? Which assumptions are justified to support our priors?

Example: flipping a coin

- Task: determine the bias of a coin
- Experiment:
    1. Flip the coin
    2. Observe: $x_i = \{0,1\}$
    3. Repeat n times
- Observations: $\{x_1, x_2, x_3, \dots, x_n\}$
- Analysis: Use Bayes rule to estimate the bias.

Example: flipping a coin

- What is a good prior in this experiment?
  - No bias because classical mechanics is deterministic ?
  - Suppose we *can't* know all the information about the initial conditions, how does this lack of information justify a particular prior ?
  - Is uniform prior a good prior ? Uniform in which parameters ?
  - What about a prior that has correlations ? Ex: More likely to get a '1' after a '0' (coin with a memory)
  - When is a conditional IID prior a good prior ? When is it justified ?

Example: flipping a coin

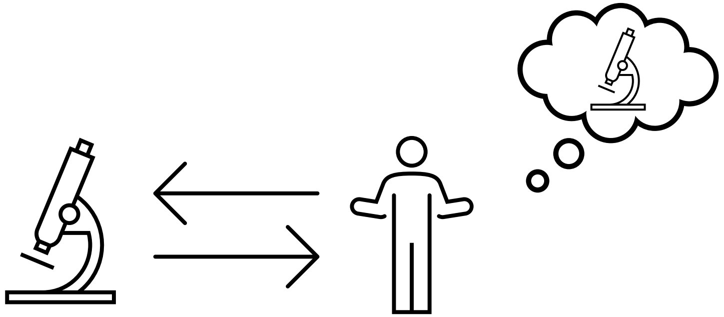
- Starting from different priors can lead to different conclusions.
- Ex: Suppose that we get observations that look like following

$$0100100010001010001000001001 01 \text{ .... (No '11's)}$$

- Someone starting with a IID prior will converge to a different distribution compared to someone starting with a prior that allows for correlations.
- Allowing for a prior with all possible correlations is intractable.
- But what is the operational justification for an IID prior ?
- Operationally, what does it even mean to say that the coin has a bias?
- These are the questions that De Finetti's theorem tries to answer!!!!

Tomography with a Bayesian friend

- All probabilities are subjective. Then what is an *unknown* probability in tomography or characterization protocols? *Unknown* to who?



- What does it mean to talk about expectation values with respect to an *unknown* probability?
- Introducing a third person only complicates the matter further.
- De Finetti's theorem also helps address this question by giving an operational interpretation to the tomography (characterization) experiments.

Tomography with a Bayesian friend

- De Finetti’s theorem also justifies the use of parameters in tomography for Bayesians.
- Like that of bias for a coin, or a quantum state for a qubit.
- Broadly, De Finetti’s theorem gives an operational understanding of many assumptions that we make in probability theory.

Some notations

- In the following slides, we will talk about joint probability of *N* trials of discrete binary random variables each of which can assume *2* different values.

$$p(x_1, x_2, \dots, x_N) \;\; x_i \in \{0,1\} \; \forall \, i \in \{1, \dots, N\}$$

- Permutation of indices (read as re-labelling the indices). As is the usual case, we denote permutation as a bijective function from a set to itself.

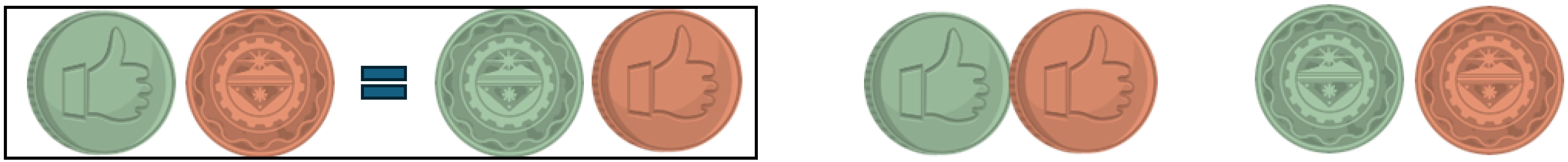
$$\pi : \{1, \dots, N\} \rightarrow \{1, \dots, N\} \quad \pi(i) \neq \pi(j) \text{ for } i \neq j$$

Symmetric Probability Distributions (5 min)

- Symmetric Distributions: Invariant under permutations of variables

$$p(x_1, \dots, x_N) = p\left(x_{\pi(1)}, \dots, x_{\pi(N)}\right) \quad \forall \text{ permutations } \pi \text{ on } \{1, \dots, N\}$$

- Operationally, order of outcomes/trials are irrelevant. Only Frequency matters
- Consider a simple example: 2 trials of coin toss



$p(1-p)$ / 1/2 / 0 $\quad$ p^2 / 1/2 / 0 $\quad$ $(1-p)^2$ / 1/2 / 0

Anticorrelated distribution / Correlated distribution

Imposing more than SPD’s

- Suppose I have repeated the same coin toss N times. Is it enough if I choose a symmetric distribution of N variables to ensure a conditional iid like distribution ?

- What if I had instead decided to do N+M coin toss?
  - To stay consistent, I would demand that if I had chosen to perform M additional tosses, the prior for N tosses (current distribution) should just be the marginal distribution of the prior for $N + M$ tosses (what if scenario).
  - I also want the prior for $N + M$ tosses to be symmetric.

Exchangeable Probability Distributions

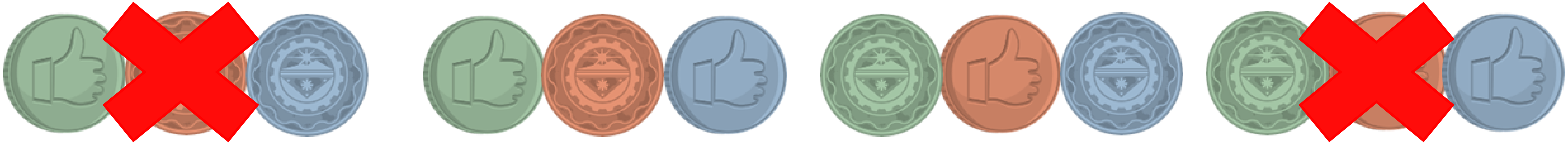
- Exchangeable Distributions: Symmetric and...

$$p(x_1, \ldots, x_N) = \sum_{x_{N+1}, \ldots, x_M} p(x_1, \ldots, x_N, x_{N+1}, \ldots, x_M) \;\forall\; M > 0$$

- Operationally, an exchangeable distribution is one in which the order of random variables is irrelevant not only for the trials performed but also for any *additional* trials that *might have been* performed

Symmetric vs Exchangeable

- By definition, all exchangeable distributions are symmetric. What about the reverse ?
- Consider the anticorrelated case plus one more toss. Since we want marginal distribution to be anticorrelated, we only have four outcomes with nonzero probability.



- Exchangeability should imply that the marginal over any two trials must be anti-correlated, which is clearly not the case.

De Finnetti's representation theorem for binary variables

- **Theorem:** Suppose we have an exchangeable probability distribution over *N* random binary variables($x_i \in \{0,1\}$), there exists a unique probability distribution *P(q)* such that,

$p(x_1, \ldots, x_N) = \int_0^1 P(q)\, q^n\, (1 - q)^{N-n}\, dq,$	$\int_0^1 P(q)\, dq = 1$

- Here, $n = \sum_i x_i$, number of times outcome is one, and q can be interpreted as the probability of obtaining one.

Implications of De finnetti's theorem

- Provides an objective criterion (exchangeability) that must be satisfied under which the conditional i.i.d assumption is valid.
- Any agent with an exchangeable prior, can now *proceed as if* the exists of an *unknown objective probability q,* associated with the coin. The ignorance of *q* is captured by the probability on the simplex, *P(q)*.
- Applying Bayes rule on prior, based on outcomes, makes $P(q)$ sharply peaked at some value. This provides an operational description for tomography.

Proof, step one

- First, since our prior is exchangeable, it is enough to keep count of the frequency of ones.
- Let $p(n, N)$ denote probability of n ones in N trials. Order is immaterial, and thus,

$$p(n, N) = \binom{N}{n} p(x_1 = 1, \ldots, x_n = 1, x_{n+1} = 0, \ldots, x_N = 0)$$

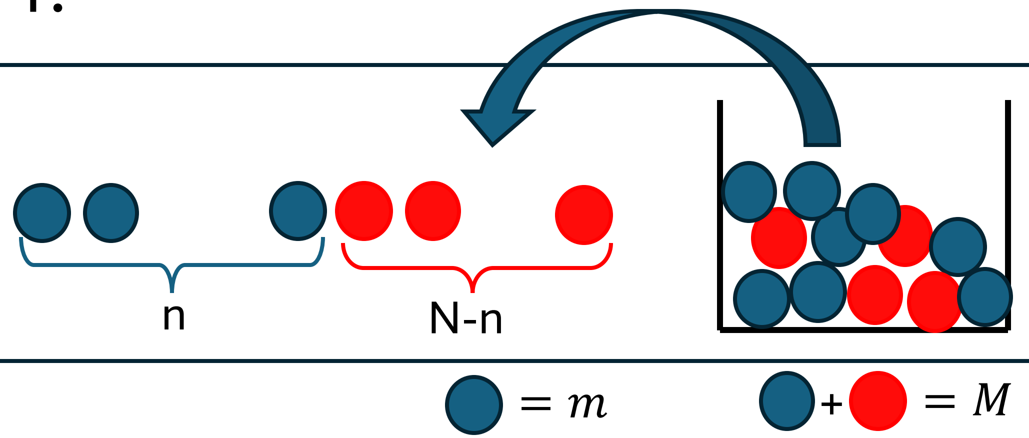
- Using law of total probability, for any $M > N$,

$$p(n, N) = \binom{N}{n} \sum_m p(x_1 = 1, \ldots, \ldots, x_N = 0 \,|m, M)\, p(m, M)$$

Proof, step two

- Consider, $p(\,1,\ldots,1,0,\ldots,0\,|m,M)$. It is independent of ordering and only depends on the frequency of 1.

$p(1,\ldots,1,0\ldots,0\,|m,M) \quad \equiv$



- Taking $M \to \infty$,

$$p(1,\ldots,1,0\ldots,0\,|m,M) \sim \frac{m^n(M-m)^{N-n}}{M^N} = z^n(1-z)^{N-n} \quad ; \quad z = \frac{\mathrm{m}}{\mathrm{M}}$$

- Combining everything,

$$p(n,N) = \binom{N}{n}\int_0^1 P(z)\, z^n\,(1-z)^{N-n}\,dz$$

Di Finetti's general representation theorem

- We can generalize the representation theorem beyond binary variables.
- **Theorem:** Suppose we have an exchangeable probability distribution over *N* random variables, there exists a unique probability distribution *P(**p**)* such that,

$$p^e(x_1, \ldots, x_N) = \int_{\mathcal{S}_k} P(\mathbf{p}) \prod_{i=1}^{N} p_{x_i} d\mathbf{p} = \int_{\mathcal{S}_k} P(\mathbf{p}) p_1^{n_1} \ldots p_k^{n_k} d\mathbf{p}$$

- Here,

$$\mathbf{p} = (p_1, \ldots, p_k) \quad n_i = \{\#j | x_j = i\} \quad \mathcal{S}_k = \{\mathbf{p} | p_j \geq 0, \sum_{i=1}^{k} p_i = 1\}$$

Quantum Scenario ?

- Quantize the coin. Goal is to estimate the density matrix, instead of the bias.
- The usual assumption in Quantum State Tomography is just like the i.i.d for probabilities. You have access to N copies of the same state

$$\rho^{(N)} = \rho \otimes \rho \otimes \cdots \otimes \rho$$

- When is this assumption really valid ? What is it's operational meaning ?
- Can we use the techniques that we have just seen ?

Learning from a Bayesian perspective

Density Matrix
Outcomes from measurements
prior on space of density matrices

- Quantum Bayes rule

$$P_{\text{post}}(\rho|x) = \frac{P(x|\rho)P_{\text{prior}}(\rho)}{\int P(x|\rho)P_{\text{prior}}(\rho)d\rho} \propto P(x|\rho)\boxed{P_{\text{prior}}(\rho)}$$

- Prior distribution over the space of density matrices. Update the prior to get the posterior using the bayes rule.

We update the probability distribution over the density matrices, not the density matrices themselves!

- Define the prior and posterior density matrices as

$$\rho_{\text{prior}} = \int d\rho \;\; P_{\text{prior}}(\rho)\,\rho \;\; ; \rho_{\text{posterior}} = \int d\rho \;\; P_{\text{post}}(\rho|x)\,\rho$$

Symmetric Quantum Priors

- State must be invariant under any permutation of indices

$$\boxed{\rho^{(N)} = \hat{\pi}_N\, \rho^{(N)} \hat{\pi}_N{}^{\dagger} \quad \forall\, \pi_N \;\; \text{permutations over N indeces}}$$

- Consider *any* orthonormal basis $\{\, |\mathrm{i}_1\rangle|\mathrm{i}_2\rangle \ldots |\mathrm{i}_\mathrm{n}\rangle\}$, we can write

$$\rho^{(N)} = \sum R_{i_1 \ldots i_n; j_1 \ldots j_n}\; |i_1 \ldots i_n\rangle\langle j_1 \ldots j_n|$$

- The basis dependent condition for symmetric density matrices

$$R_{i_1 \ldots i_n; j_1 \ldots j_n} = R_{\pi(i_1) \ldots \pi(i_n); \pi(j_1) \ldots \pi(j_n)}$$

Exchangeable Quantum Priors

- State must be symmetric $\rho^{(N)}$
- For any $M > 0$, we should have a symmetric state $\rho^{(N+M)}$ such that $\rho^{(N)}$ is the marginal, i.e,

$$\boxed{\rho^{(N)} = \mathrm{Tr}_M\left(\rho^{(N+M)}\right)}$$

- You can give an equivalent basis dependent condition, which I am omitting for brevity.

Quantum De finnetti theorem

- Any exchangeable state of N systems can be uniquely written as

$$\boxed{\rho^{(N)} = \int d\rho\, P(\rho)\rho^{\otimes N} \quad \Big| \quad \int d\rho\, P(\rho) = 1}$$

- Objective criterion for assuming product state in QST.
- Assuming exchangeability, priors are only reflected on $P(\rho)$.
- Learning is making $P(\rho)$ sharp.

Informationally complete POVMs

- Pick a POVM $\mathcal{E} = \{E_\alpha \mid E_\alpha \geq 0, \sum_\alpha E_\alpha = 1\}$, such that the effects E_α spans the space of operators. For any operator A, we can always find scalars λ_α,
$$\sum_\alpha \lambda_\alpha E_\alpha = 1$$
- Such a POVM is called *informationally complete*. If $\mathcal{E}$ forms a basis, then scalars λ_α are unique for any operator A. Such a POVM is called *minimal informationally complete*. For any *mic* POVM, we have 1-1 correspondence between states and probability distribution
$$\rho \to \big(\mathrm{Tr}\,(\rho\, E_1), \mathrm{Tr}\,(\rho\, E_2), \dots, \mathrm{Tr}\,(\rho\, E_n)\big) \equiv (p_1, p_2, \dots, p_n)$$

Proof procedure

- Choose a *minimal informationally complete* POVM and collapse the state to a probability distribution.
- Apply the classical De finetti's theorem. Using the 1-1 correspondence, elevate the probability distribution to operators

$$\rho^{(N)} = \int P(\boldsymbol{p}) A_{\boldsymbol{p}}^{\otimes N}$$

- In general, $A_{\mathbf{p}}$, need not be psd. They are Hermitian and trace one. Show that $P(\boldsymbol{p})$ corresponding to non psd $A_{\mathbf{p}}$ should go to zero.

Importance of Exchangeability

- Exchangeability is the key operational idea that justifies using conditional IID prior in Bayes rule.
- Priors which do not assume exchangeability can come to very different conclusions.
- Fundamentally, it justifies the **use of parameters in probability theory**.
- In the example of flipping a coin, exchangeability is what is needed to assume that the property of bias for a coin.
- Same in the quantum case for assuming the property of an unknown ‘quantum state’

*Christopher A. Fuchs, Rüdiger Schack; Priors in Quantum Bayesian Inference. AIP Conf. Proc. 10 March 2009; 1101 (1): 255–259. https://doi.org/10.1063/1.3109948

Extensions to Finite Exchangeability

- What if we can don’t have time to repeat our experiment infinite number of times?
- Finite Exchangeability:

$$\{x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_m\}$$

- Implies that distribution is close to IID in variation distance*.

$$\left\| P(x_1, x_2, \dots, x_n) - \int \mu(p)\, p^k\, (1\ - p)^{n\,-k}\, \mathrm{d}p \right\| \leq 4n/m$$

* Diaconis, Persi, and David Freedman. "Finite exchangeable sequences." *The Annals of Probability* (1980): 745-764.

Extensions to Finite Exchangeability

- Quantum version**: $|\psi\rangle \in \mathrm{Sym}^{\mathrm{m}}(\mathbb{C}^{\mathrm{d}}), \quad \xi^n \coloneqq \mathrm{tr}_{\mathrm{m-n}} |\psi\rangle\langle\psi|$

$$\left\| \xi^n - \int \mu(\phi)\, |\phi\rangle\langle\phi|^n \, d\phi \right\| \leq 2\frac{dn}{m}$$

- Quantum versions of finite exchangeable De Finetti's theorems provide a natural way to quantify monogamy of entanglement in symmetric systems.

** Christandl, Matthias, et al. "One-and-a-half quantum de Finetti theorems." *Communications in mathematical physics* 273.2 (2007): 473-498.