"The square root of Bayesian inference" sketch
(rev 2020–01-06)

J. Köplinger

jenskoeplinger@gmail.com

31 August 2019

Outline

Bayesian inference - quick summary

What Bayesian inference does and doesn't do

Bayesian inference:

- What is does: Give credible intervals for variable model parameters θ from past data.
- What is doesn't: Give confidence intervals for future data x from fixed model parameters.

Likelihood and prior distribution

$$\mathrm{P}(\theta \mid x) = \frac{f(x \mid \theta)\mathrm{P}(\theta)}{\int_\theta f(x \mid \theta')\mathrm{P}(\theta')\, d\theta'}$$

Schematically:

- $f(x \mid \theta)$ is likelihood function for data x given some model parameter(s) θ.
- $\mathrm{P}(\theta)$ is the prior distribution that encodes prior data, as well as model assumption on how the data distributes.
- $\mathrm{P}(\theta \mid x)$ is the posterior distribution of model parameters θ given observations and prior.
- The maximum of $\mathrm{P}(\theta \mid x)$ gives the most likely θ. Credible intervals for θ correspond to areas under the curve.

Conjugate priors

$$\mathrm{P}(\theta \mid x) = \frac{f(x \mid \theta)\mathrm{P}(\theta)}{\int_\theta f(x \mid \theta')\mathrm{P}(\theta')\, d\theta'}$$

- If $\mathrm{P}(\theta \mid x)$ is in the same family of functions as $\mathrm{P}(\theta)$, then $\mathrm{P}(\theta)$ is called conjugate prior to $\mathrm{P}(\theta \mid x)$.
- Posterior distributions $\mathrm{P}(\theta \mid x)$ can then become new priors $\mathrm{P}(\theta)$ once more data comes in.

Radioactive decay

For example, in radioactive decay we expect event times x to distribute exponentially, given a decay rate θ. The Bayesian ansatz is:

$$f(x \mid \theta) \propto \theta e^{-\theta x} \quad (x \geq 0),$$
$$\mathrm{P}(\theta) = \Gamma(\alpha, \beta) \propto \theta^{\alpha-1} e^{-\beta\theta},$$
$$\mathrm{P}(\theta \mid x) = \frac{f(x \mid \theta)\mathrm{P}(\theta)}{\int_\theta f(x \mid \theta')\mathrm{P}(\theta')\, d\theta'}.$$

- $f(x \mid \theta)$ is the likelihood for decay time x given a decay rate θ.
- In the prior, $\mathrm{P}(\theta)$, the α and β are hyperparameters that encode previous data.
- With these choices, $\mathrm{P}(\theta)$ is conjugate prior to the posterior $\mathrm{P}(\theta \mid x)$, because both are gamma distributions in general.

Radioactive decay: Step by step example

To give a step-by-step example:

- Pick a noninformative prior, e.g. $\alpha = 1$, $\beta = 0$, $\mathrm{P}(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta} = \mathrm{const.}$
- Perform a single measurement for decay time, x_1.
- The likelihood of θ from this is $f(\theta \mid x_1) \propto \theta e^{-\theta x_1}$.
- Identify this as likelihood of x_1, i.e. $f(x_1 \mid \theta) \equiv f(\theta \mid x_1)$.
- Update the prior with this data, to get the posterior:

$$\mathrm{P}(\theta \mid x_1) \propto f(x_1 \mid \theta)\mathrm{P}(\theta) \propto \theta e^{-\theta x_1}.$$

Radioactive decay: Criterion for "making good choices"

The posterior after taking one measurement x_1,

$$\mathrm{P}(\theta \mid x_1) \propto \theta e^{-\theta x_1},$$

can then be identified as a new prior

$$\mathrm{P}_1(\theta) = \Gamma(\alpha_1, \beta_1) \propto \theta^{\alpha_1 - 1} e^{-\beta_1 \theta},$$

with hyperparameters $\alpha_1 = 2$, $\beta_1 = x_1$.

- More data changes the shape of $\mathrm{P}(\theta) = \Gamma(\alpha, \beta)$.
- A sharper peak in $\mathrm{P}(\theta)$ means smaller uncertainty in model parameter θ. This is the quality criterion that the (subjective) choices for prior and likelihood were indeed "good choices".

Finding conjugate priors

The challenge in Bayesian inference:

- Find priors that correctly encode prior data (if any), as well as distribution assumptions of the model.
- Noninformative priors are possible, which encode only some distribution assumption but no data.
- Conjugate priors are very helpful and mathematically elegant.
- One versatile prior is the gamma distribution $\Gamma(\alpha, \beta)$.

The gamma distribution as conjugate prior

$$X \sim \Gamma(\alpha, \beta),$$
$$\mathrm{P}_\Gamma(\theta; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}.$$

The gamma distribution is conjugate prior to ...

- ... itself,
- ... the exponential distribution,
- ... the Poisson, pareto, and inverse gamma distributions,
- ... precision $1/\sigma$ of the normal distribution (known mean).

"Take the square root" of Bayesian inference

Proposition: Spinor-valued prior and posterior

Try to "take the square root" of Bayesian inference:

$$\Psi(\theta \mid x) = \frac{\phi(x \mid \theta)\,\Psi(\theta)}{\int_\theta \phi(x \mid \theta')\,\Psi(\theta')\,d\theta'}$$

- Allow $\Psi(\theta \mid x)$ and $\Psi(\theta)$ to be spinor-valued (by proposition).
- Likelihood $\phi(x \mid \theta)$ generally not positive real anymore, requires clarification.
- For now, leave interpretation or meaning open.

Relate to quantum mechanics

What is the aim?

- Find spinor-valued priors and posteriors that resemble canonical quantum mechanics
- Build a toy model to rationalize the general approach
- Frame the work to be done, to reconstruct QM exactly

Superficially, this resembles how Paul A. M. Dirac motivated the Dirac equation by linearizing $E^2 = m^2 + |\vec{p}|^2$.

Gamma distribution with complex spinors

As a concrete example, let the gamma distribution be complex-valued:

$$\Psi_\Gamma(\theta; q, m) \quad \propto \quad |\theta|^{iq}\, e^{im\theta} \qquad (\theta, q, m \in \mathbb{R})\,.$$

Terms $|\theta|^{iq} \equiv e^{iq \ln|\theta|}$ and $e^{im\theta}$ represent complex spinors.

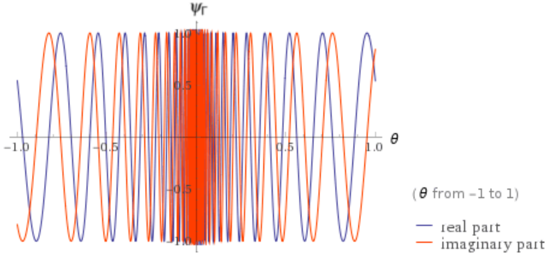

Figure: Plot of $\Psi_\Gamma(\theta; 20, 1)$ (credit: WolframAlpha, 2019)

Gamma distribution with complex spinors (ctd)

A bit more general, define $Q := \{q_j, \theta_{0,j}\}$ and

$$\Psi_\Gamma(\theta; Q, m) \quad \propto \quad \left(\prod_{j=1}^{n} |\theta - \theta_{0,j}|^{iq_j} \right) e^{im\theta} \qquad (\theta, q_j, \theta_{0,j}, m \in \mathbb{R}).$$

Why? Because then an operator $\hat{D}$ exists where $\left(\hat{D} - m\right) \Psi_\Gamma = 0$ looks like some 1D Dirac equation with $1/\theta$ fields [2]:

$$\hat{D} := -i \frac{\partial}{\partial \theta} - \sum_{j=1}^{n} \frac{q_j}{\theta - \theta_{0,j}}$$

[2] J. Köplinger, V. Dzhunushaliev, Nonassociative quantum theory, emergent probability, and coquasigroup symmetry, *arXiv*:0910.3347 (2011).

A one dimensional toy model of quantum mechanics

A 1D toy model of QM

- Wave functions are posteriors that are built from spinor-valued priors and generalized likelihoods over a real parameter θ.
- The (noninformative) prior for a point particle of mass m is

$$\Psi(\theta; m) := e^{im\theta}.$$

- Fields $q/(\theta - \theta_0)$ generate likelihoods

$$\phi(\theta; q, \theta_0) := |\theta - \theta_0|^{iq}.$$

- $\Psi(\theta; m)$ and $\phi(\theta; q, \theta_0)$ quantify ignorance of spinor phase.

Quantum likelihoods and posteriors

Put together, wave functions $\Psi_\Gamma(\theta; Q, m)$ are posteriors built from generalized likelihoods $\phi(Q \mid \theta)$ and spinor-valued priors $\Psi(\theta; m)$ as:

$$\Psi_\Gamma(\theta; Q, m) = \frac{\phi(Q \mid \theta)\Psi(\theta; m)}{\int_\theta \phi(Q \mid \theta')\Psi(\theta'; m)\, d\theta'}$$

- $\Psi_\Gamma(\theta; Q, m)$ give us credible intervals for θ.
- Need to clarify meaning and existence of complex-valued likelihood, measurement.
- Priors $\Psi(\theta; m)$ and likelihoods $\phi(Q \mid \theta)$ are given by definition. Can we rationalize these choices somehow?

Eigenvalue / eigenfunction rule for "making good choices"

By construction, we arrived at spinor-valued wave functions

$$\Psi_\Gamma(\theta; Q, m) \propto \left(\prod_{j=1}^{n} |\theta - \theta_{0,j}|^{iq_j} \right) e^{im\theta}.$$

These are eigenfunctions with real eigenvalue m to operator

$$\hat{D} := -i\frac{\partial}{\partial\theta} - \sum_{j=1}^{n} \frac{q_j}{\theta - \theta_{0,j}}.$$

Requiring the wave functions to be eigenfunctions to $\hat{D}$ with real eigenvalue m is therefore consistent with "making good choices" - in the Bayesian sense - for priors and likelihoods.

Measurement

- "Measurement" is setting up an experiment $\phi_{\mathrm{E}}(Q,m \mid \theta)$ that interacts with the quantum system $\Psi_{\mathrm{Q}}(\theta) \equiv \Psi_{\Gamma}(\theta; Q, m)$ in a way that exactly yields a real-valued posterior $\Psi_{\mathrm{M}}(\theta \mid Q, m)$:

$$\Psi_{\mathrm{M}}(\theta \mid Q,m) = \frac{\phi_{\mathrm{E}}(Q,m \mid \theta)\,\Psi_{\mathrm{Q}}(\theta)}{\int_{\theta} \phi_{\mathrm{E}}(Q,m \mid \theta')\,\Psi_{\mathrm{Q}}(\theta')\,d\theta'}$$

- The posterior models the measurement probability distribution; yet it is just another wave function that so happens to be real-valued.
- This is now interpreted as some projection from a quantum reality in spinor space into an anthropocentric real space.

Relation to “Quantum Bayesianism”

Shared concepts with “Quantum Bayesianism”:

- Wave functions don't “collapse” (as in Copenhagen QM). Instead, $\Psi_{\mathrm{M}}(\theta \mid Q, m)$ encodes both the observed system $\Psi_{\mathrm{Q}}(\theta)$ and the experiment setup $\phi_{\mathrm{E}}(Q, m \mid \theta)$.
- The Born rule in QM can tentatively be mapped here as mathematical method to find good choices for priors and likelihoods, as well as experiment setups $\phi_{\mathrm{E}}(Q, m \mid \theta)$ that allow for human observation.

However, the current toy model is too limited to further evaluate the Born rule mapping, since the eigenspace of $\hat{D}$ above is exactly one wave function.

Next steps

Next steps:

- Research "Quantum Bayesianism" to determine the exact context (overlap, discoveries already made, issues).
- The step $|\theta|^{iq} \equiv e^{iq \ln|\theta|}$ silently assumed a real-valued $\ln|\theta|$. This is an inconsistent limitation. Explore all $\ln|\theta| \pm i\,2\pi N$.
- Expand the 1D toy model, to using a wider class of spinors, as well as using more dimensions.
- Re-evaluate the tentative Born rule mapping in these expanded models, and the overall approach in general.

With many thanks to John Huerta for helpful conversations and advice.