

Bayesian Updates

January 5, 2023

Chapter 1

Preliminaries

1.1 Circular Distributions

I'm anticipating that I might need to put more words into this later on, so am leaving space for them here.

1.1.1 Von Mises distribution

The Von-Mises distribution is given by:

$$f(x, \mu, \kappa) = \frac{1}{2\pi I_0 \kappa} \exp(\kappa \cos(x - \mu)), \quad -\pi \leq x \leq \pi,$$

where $I_0(\cdot)$ is the 0th modified Bessel function, where the n th modified Bessel function is given by

$$I_n(\kappa) = \frac{1}{\pi} \int_0^\pi \cos(n\theta) \exp(\kappa \cos \theta) d\theta.$$

The circular mean of the Von-Mises distribution is given by:

$$\mathbb{E} [\exp i\theta] = \frac{I_1(\kappa)}{I_0(\kappa)} e^{i\mu}.$$

In general, this can be seen via

$$\begin{aligned}
\mathbb{E}[e^{in\theta}] &= \frac{1}{2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \exp(in\theta) \exp(\kappa \cos(\theta - \mu)) d\theta \\
&= \frac{1}{2\pi I_0(\kappa)} \int_{-\pi-\mu}^{\pi-\mu} \exp(in(\psi + \mu)) \exp(\kappa \cos \psi) d\psi \\
&= \frac{e^{in\mu}}{2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \exp(in\theta) \exp(\kappa \cos \theta) d\theta \\
&= \frac{e^{in\mu}}{2\pi I_0(\kappa)} \int_{-\pi}^{\pi} (\cos(n\theta) + i \sin(n\theta)) \exp(\kappa \cos \theta) d\theta \\
&= \frac{e^{in\mu}}{2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \cos(n\theta) \exp(\kappa \cos \theta) d\theta \\
&= \frac{I_{|n|}(\kappa)}{I_0(\kappa)} e^{in\mu}.
\end{aligned}$$

Note that we remove the sin integral by using the fact that the integral of an odd function over a symmetric, periodic interval is 0.

Chapter 2

Problem Statement

2.1 Setup

Goal: Given a single measurement of a Bernoulli random variable and a Von-Mises prior distribution, calculate the posterior distribution and approximate to a Von-Mises distribution.

- t - time step
- d_t - Grover depth of quantum circuit at time t
- Y_t - random variable representing a single shot measurement y_t of the quantum circuit at time t
- $\Pi(\theta|Y_1 = y_1, \dots, Y_t = y_t) = \Pi(\theta|\mathbf{Y}_t)$ - 'true' posterior at time t (though values for $t' < t$ have been used to approximate the earlier distributions)
- $\hat{\Pi}(\theta|Y_1 = y_1, \dots, Y_t = y_t) = \hat{\Pi}(\theta|\mathbf{Y}_t)$ - approximate posterior at time t .

According to Bayes rule:

$$\Pi(\theta|Y_t = y_t, \mathbf{Y}_{t-1}) = \frac{\Pi(Y_t = y_t|\theta)\Pi(\theta|\mathbf{Y}_{t-1})}{\Pi(Y_t = y_t)},$$

so we need to compute each of the quantities on the RHS.

At time t , we make a measurement y_t of $Y_t \sim \text{Ber}(p_t)$ at a Grover depth of d_t where

$$p_t = \frac{1}{2}(1 - \cos((4d_t + 2)\hat{\mu}_{t-1})).$$

Thus,

$$\Pi(Y_t = y_t|\theta) = \frac{1}{2}(1 + (-1)^{y_t} \cos((4d_t + 2)\hat{\mu}_{t-1})).$$

For convenience, let $\lambda_t = 4d_t + 2$.

To simplify some of the computations, we're going to assert that the posterior follows a Von-Mises distribution after every update, so we calculate the new values $\hat{\mu}_t, \hat{\kappa}_t$ and generate our approximate posterior

$$\hat{\Pi}(\theta|\mathbf{Y}_t) \sim VM(\hat{\mu}_t, \hat{\kappa}_t).$$

2.2 Single shot updates

For simplicity, we're going to consider the first step of the update, which makes things a lot nicer. In this case, we want to know what the circular mean of the posterior distribution is after updating.

- $\Pi(\theta) \sim VM(\mu, \kappa)$ - prior
- $\Pi(Y|\theta) \sim \text{Ber}(\frac{1}{2}(1 - \cos(\lambda\theta)))$

This gives us:

$$\begin{aligned} \Pi(Y = y) &= \int_{-\pi}^{\pi} \Pi(Y = y|\theta) \Pi(\theta) d\theta \\ &= \frac{1}{2\pi I_0(\kappa)} \int_{-\pi}^{\pi} \frac{1}{2} (1 + (-1)^y \cos(\lambda\theta)) \exp(\kappa \cos(\theta - \mu)) d\theta \\ &= \frac{1}{2\pi I_0(\kappa)} \left(\int_{-\pi}^{\pi} \frac{1}{2} \exp(\kappa \cos(\theta - \mu)) d\theta \right. \\ &\quad \left. + (-1)^y \int_{-\pi}^{\pi} \frac{1}{2} \cos(\lambda\theta) \exp(\kappa \cos(\theta - \mu)) d\theta \right) \\ &= \frac{1}{4\pi I_0(\kappa)} \left(2\pi I_0(\kappa) + (-1)^y \int_{-\pi}^{\pi} \frac{e^{i\lambda\theta} + e^{-i\lambda\theta}}{2} \exp(\kappa \cos(\theta - \mu)) d\theta \right) \\ &= \frac{1}{2} \left(1 + (-1)^y \cos(\lambda\mu) \frac{I_\lambda(\kappa)}{I_0(\kappa)} \right) \end{aligned}$$

where in the penultimate line, we use the expression for the n th circular moment. Putting this all together, and letting

$$C(y, \lambda, \mu, \kappa) = \frac{\frac{1}{2} \frac{1}{2\pi I_0(\kappa)}}{\frac{1}{2} (1 + (-1)^y \cos(\lambda\mu) \frac{I_\lambda(\kappa)}{I_0(\kappa)})} = \frac{1}{2\pi (I_0(\kappa) + (-1)^y \cos(\lambda\mu) I_\lambda(\kappa))}$$

gives

$$\begin{aligned}
\mathbb{E}[e^{i\theta}|Y=y] &= C(y, \lambda, \mu, \kappa) \int_{-\pi}^{\pi} e^{i\theta} (1 + (-1)^y \cos(\lambda\theta)) \exp(\kappa \cos(\theta - \mu)) d\theta \\
&= C(y, \lambda, \mu, \kappa) \left(\int_{-\pi}^{\pi} e^{i\theta} \exp(\kappa \cos(\theta - \mu)) d\theta \right. \\
&\quad \left. + (-1)^y \int_{-\pi}^{\pi} e^{i\theta} \cos(\lambda\theta) \exp(\kappa \cos(\theta - \mu)) d\theta \right) \\
&= C(y, \lambda, \mu, \kappa) \left(2\pi I_1(\kappa) e^{i\mu} \right. \\
&\quad \left. + (-1)^y \int_{-\pi}^{\pi} e^{i\theta} \left(\frac{e^{i\lambda\theta} + e^{-i\lambda\theta}}{2} \right) \exp(\kappa \cos(\theta - \mu)) d\theta \right) \\
&= 2\pi C(y, \lambda, \mu, \kappa) \left(I_1(\kappa) e^{i\mu} + \frac{(-1)^y}{2} \left(I_{\lambda+1}(\kappa) e^{i(\lambda+1)\mu} + I_{\lambda-1}(\kappa) e^{-i(\lambda-1)\mu} \right) \right)
\end{aligned}$$

where in the penultimate line, we use the fact that

$$\int_{-\pi}^{\pi} e^{in\theta} \exp(\kappa \cos(\theta - \mu)) d\theta = 2\pi I_0(\kappa) \mathbb{E}[e^{in\theta}] = I_{|n|}(\kappa) e^{in\mu}.$$

This gives us that

$$\mathbb{E}[e^{i\theta}|Y=y] = \frac{I_1(\kappa) e^{i\mu} + \frac{(-1)^y}{2} (I_{\lambda+1}(\kappa) e^{i(\lambda+1)\mu} + I_{\lambda-1}(\kappa) e^{-i(\lambda-1)\mu})}{I_0(\kappa) + (-1)^y \cos(\lambda\mu) I_{\lambda}(\kappa)}.$$

If we then take expectations over Y (i.e. multiply by $\Pi(Y=y)$ and sum) this gives us

$$\mathbb{E}[e^{i\theta}] = \frac{I_1(\kappa)}{I_0(\kappa)} e^{i\mu}.$$

So, we can infer that we do not expect the angular parameter μ to move. To infer something about κ , we need to consider $\mathbb{E}[R]$. As before, let's consider $\mathbb{E}[R|Y=y]$. From the above, we can deduce that

$$\begin{aligned}
\mathbb{E}[R|Y=y]^2 &= \frac{\left(I_1 + \frac{(-1)^y}{2} (e^{i\lambda\mu} I_{\lambda+1} + e^{-i\lambda\mu} I_{\lambda-1}) \right) \left(I_1 + \frac{(-1)^y}{2} (e^{-i\lambda\mu} I_{\lambda+1} + e^{i\lambda\mu} I_{\lambda-1}) \right)}{(I_0 + (-1)^y \cos(\lambda\mu) I_{\lambda})^2} \\
&= \frac{I_1^2 + \frac{1}{4} (I_{\lambda+1}^2 + I_{\lambda-1}^2) + \frac{1}{2} \cos(2\lambda\mu) I_{\lambda+1} I_{\lambda-1} + (-1)^y I_1 (I_{\lambda+1} + I_{\lambda-1}) \cos \lambda\mu}{(I_0 + (-1)^y \cos(\lambda\mu) I_{\lambda})^2} \\
&= \frac{N_y^2}{(I_0 + (-1)^y \cos(\lambda\mu) I_{\lambda})^2},
\end{aligned}$$

where for brevity, we have suppressed the argument κ for each of the Bessel functions I_{ν} .

Calculating $\mathbb{E}[R]$ then, is achieved by multiplying by $\Pi(Y = y)$, square-rooting, and summing. This results in the sum of the square roots of the numerators multiplied by a constant factor of $\frac{1}{2I_0(\kappa)}$, i.e.

$$\begin{aligned}\mathbb{E}[R] &= \frac{N_0 + N_1}{2I_0(\kappa)} \\ &= \frac{1}{2I_0(\kappa)} \sqrt{I_1^2 + \frac{1}{4} (I_{\lambda+1}^2 + I_{\lambda-1}^2) + \frac{1}{2} \cos(2\lambda\mu) I_{\lambda+1} I_{\lambda-1} + I_1 (I_{\lambda+1} + I_{\lambda-1}) \cos(\lambda\mu)} \\ &\quad + \frac{1}{2I_0(\kappa)} \sqrt{I_1^2 + \frac{1}{4} (I_{\lambda+1}^2 + I_{\lambda-1}^2) + \frac{1}{2} \cos(2\lambda\mu) I_{\lambda+1} I_{\lambda-1} - I_1 (I_{\lambda+1} + I_{\lambda-1}) \cos(\lambda\mu)}\end{aligned}$$

2.3 Gaussian case

2.3.1 Posterior

Now we're going to assume a different prior and posterior:

- $\Pi(\theta) \sim \mathcal{N}(\mu, \sigma^2)$ - prior
- $\Pi(Y = y|\theta) \sim \text{Ber}(\lambda) = \frac{1}{2}(1 + (-1)^y \cos(\lambda\theta))$

Bayes rule, again, states that

$$\Pi'(\theta|Y = y) = \frac{\Pi(\theta)\mathcal{L}(y, \theta)}{\Pi(Y = y)}. \quad (2.1)$$

We assume that

$$\Pi(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}, \quad (2.2)$$

we know that

$$\mathcal{L}(y, \theta) = \frac{1}{2}(1 + (-1)^y \cos(\lambda\theta)), \quad (2.3)$$

and by definition

$$\Pi(Y = y) = \int \mathcal{L}(y, \theta) \Pi(\theta) d\theta. \quad (2.4)$$

Let us now define the *bias* to be

$$\Lambda(\theta) = 2\mathcal{L}(0, \theta) - 1, \quad (2.5)$$

which gives the likelihood as

$$\mathcal{L}(y, \theta) = \frac{1}{2}(1 + (-1)^y \Lambda(\theta)). \quad (2.6)$$

Recognise that in this case $\Lambda(\theta) = \cos(\lambda\theta)$.

Let us define *expected bias* b and the *chi function* χ as

$$b = \int \Pi(\theta) \Lambda(\theta) d\theta \quad (2.7)$$

$$\chi = \frac{1}{\sigma^2} \int (\theta - \mu) \Pi(\theta) \Lambda(\theta) d\theta \quad (2.8)$$

Now, putting that in to equation (2.4) gives

$$\Pi(Y = y) = \frac{1}{2} \left[\int \Pi(\theta) d\theta + (-1)^y \int \Pi(\theta) \Lambda(\theta) d\theta \right]. \quad (2.9)$$

The first part equals 1 by normalisation. The second part is the expected bias by definition, i.e.,

$$\Pi(Y = y) = \frac{1}{2} \left[1 + (-1)^y b \right]. \quad (2.10)$$

The expected bias b is given by

$$b = \int \Pi(\theta) \Lambda(\theta) d\theta \quad (2.11)$$

$$= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} \cos(\lambda\theta) d\theta, \quad (2.12)$$

which apparently ([1], equation 153 on page 25) is an 'identity' for $\sigma > 0$ and $\mu, \lambda \in \mathbb{R}$:

$$b(\mu, \sigma) = e^{-\frac{1}{2}\lambda^2\sigma^2} \cos(\lambda\mu). \quad (2.13)$$

Putting this all together gives for the posterior:

$$\Pi'(\theta|Y = y) = \frac{\Pi(\theta)\mathcal{L}(y, \theta)}{\Pi(Y = y)} \quad (2.14)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{e^{-\frac{1}{2}(\frac{\theta-\mu}{\sigma})^2} (1 + (-1)^y \cos(\lambda\theta))}{1 + (-1)^y e^{-\frac{1}{2}\lambda^2\sigma^2} \cos(\lambda\mu)}. \quad (2.15)$$

2.3.2 Expected values

Now, we're interested in the following quantities:

- $\mathbb{E}_y(\text{Var}_\theta(\theta|Y))$ - The expected posterior variance
- \mathcal{V} - The variance reduction factor
- $\mathbb{E}(\theta|Y = 0), \text{Var}(\theta|Y = 0)$ - Posterior mean and variance when measure $Y = 0$
- $\mathbb{E}(\theta|Y = 1), \text{Var}(\theta|Y = 1)$ - Posterior mean and variance when measure $Y = 1$

Theorem 12 (together with equation (113)) of [1] states that the *expected posterior variance* is given by

$$\mathbb{E}_y(\text{Var}_\theta(\theta|Y)) = \sigma^2(1 - \sigma^2\mathcal{V}), \quad (2.16)$$

with

$$\mathcal{V} = \frac{1}{4} \left[\sum_{y \in \{0,1\}} \frac{I_1(y)^2}{I_0(y)} - \mu^2 \right], \quad (2.17)$$

and with

$$I_k(y) = \int \theta^k \mathcal{L}(y, \theta) \Pi(\theta) d\theta \quad (2.18)$$

the k -th moment of the function $\mathcal{L}(y, \cdot) \Pi(\cdot)$.

Now by writing the expected bias b and chi function χ , equations (2.7) and (2.8) respectively, in terms of the moments, you can show (equations (132-135) from [1]) that for a two-outcome likelihood function, the variance reduction factor can be written as

$$\mathcal{V} = \begin{cases} \frac{\chi^2}{1-b^2}, & |b| < 1 \\ 0, & |b| = 1. \end{cases} \quad (2.19)$$

The Gaussian prior has a nice property: differentiating the expected bias w.r.t. the prior mean gives the chi function, i.e.

$$\chi(\mu, \sigma) = \frac{\partial}{\partial \mu} b(\mu, \sigma), \quad (2.20)$$

resulting in

$$\chi(\mu, \sigma) = -\lambda e^{-\frac{1}{2}\lambda^2\sigma^2} \sin(\lambda\mu) \quad (2.21)$$

and now the variance reduction factor can be written as

$$\mathcal{V} = \mathcal{V}(\mu, \sigma) = \frac{\partial_\mu b(\mu, \sigma)^2}{1 - b(\mu, \sigma)^2} \mathbb{1}_{\Lambda \notin \{\pm 1\}}, \quad (2.22)$$

where $\mathbb{1}_{\Lambda \notin \{\pm 1\}}$ denotes the indicator function which is equal to 1 when $\Lambda \notin \{\pm 1\}$ and 0 otherwise.

Combining the above, gives, together with equation (2.13) for the Gaussian prior:

$$\mathcal{V} = \frac{\lambda^2 e^{-\lambda^2\sigma^2} \sin^2(\lambda\mu)}{1 - e^{-\lambda^2\sigma^2} \cos^2(\lambda\mu)} \mathbb{1}_{\Lambda \notin \{\pm 1\}}, \quad (2.23)$$

and thereby the expected posterior variance is

$$\mathbb{E}_y(\text{Var}_\theta(\theta|Y)) = \sigma^2(1 - \sigma^2 \frac{\lambda^2 e^{-\lambda^2\sigma^2} \sin^2(\lambda\mu)}{1 - e^{-\lambda^2\sigma^2} \cos^2(\lambda\mu)} \mathbb{1}_{\Lambda \notin \{\pm 1\}}). \quad (2.24)$$

The next quantities of interest are $\mathbb{E}(\theta|Y = y)$ and $\text{Var}(\theta|Y = y)$ for $y \in \{0, 1\}$. By definition:

$$\mathbb{E}(\theta|Y = y) = \int \theta \Pi'(\theta|Y = y) d\theta, \quad (2.25)$$

and

$$\text{Var}(\theta|Y = y) = \mathbb{E}(\theta^2|Y = y) - (\mathbb{E}(\theta|Y = y))^2, \quad (2.26)$$

with

$$\mathbb{E}(\theta^2|Y = y) = \int \theta^2 \Pi'(\theta|Y = y) d\theta. \quad (2.27)$$

Starting with $\mathbb{E}(\theta|Y = y)$, we write

$$\mathbb{E}(\theta|Y = y) = \frac{1}{\Pi(Y = y)} \int \theta \Pi(\theta) \mathcal{L}(y, \theta) d\theta \quad (2.28)$$

$$= \frac{1}{\Pi(Y = y)} \int \theta \Pi(\theta) \left(\frac{1}{2} (1 + (-1)^y \cos(\lambda \theta)) \right) d\theta \quad (2.29)$$

$$= \frac{1/2}{\Pi(Y = y)} \left(\int \theta \Pi(\theta) d\theta + (-1)^y \int \theta \Pi(\theta) \Lambda(\theta) d\theta \right) \quad (2.30)$$

$$= \frac{1/2}{\Pi(Y = y)} \left(\mu + (-1)^y \int \theta \Pi(\theta) \Lambda(\theta) d\theta \right). \quad (2.31)$$

Now by equation (2.8),

$$\chi = \frac{1}{\sigma^2} \left(\int_{-\infty}^{\infty} \theta \Pi(\theta) \Lambda(\theta) d\theta - \mu \int_{-\infty}^{\infty} \Pi(\theta) \Lambda(\theta) d\theta \right), \quad (2.32)$$

and by using the definition for b (equation (2.7)):

$$\int_{-\infty}^{\infty} \theta \Pi(\theta) \Lambda(\theta) d\theta = \sigma^2 \chi + \mu b, \quad (2.33)$$

which gives:

$$\mathbb{E}(\theta|Y = y) = \frac{1/2(\mu + (-1)^y(\sigma^2 \chi + \mu b))}{\Pi(Y = y)} \quad (2.34)$$

$$= \frac{\mu + (-1)^y(\sigma^2 \chi + \mu b)}{1 + (-1)^y b} \quad (2.35)$$

$$(2.36)$$

and with equations (2.13) and (2.21) gives:

$$\mathbb{E}(\theta|Y = y) = \frac{\mu + (-1)^y e^{-\frac{1}{2}\lambda^2 \sigma^2} (\mu \cos(\lambda \mu) - \sigma^2 \lambda \sin(\lambda \mu))}{1 + (-1)^y e^{-\frac{1}{2}\lambda^2 \sigma^2} \cos(\lambda \mu)}. \quad (2.37)$$

Appendix A

Integrals

A.1 Normal Distribution

First, let us note that

$$\frac{d}{dx}(e^{-x^2}) = -2xe^{-x^2}.$$

The integrals we are interested in computing, are either of the form

$$I_{2n}(k) = \int_{-\infty}^{\infty} \theta^{2n} \cos(k\theta) e^{-\theta^2} d\theta \text{ or } I_{2n+1} = \int_{-\infty}^{\infty} \theta^{2n+1} \sin(k\theta) e^{-\theta^2} d\theta$$

Bibliography

- [1] D. E. Koh, G. Wang, P. D. Johnson, and Y. Cao. Foundations for bayesian inference with engineered likelihood functions for robust amplitude estimation. *Journal of Mathematical Physics*, 6 2022.