Homework 6

Fred Boehm

12/2/2020

Part A.i

Compute the covariance matrix of Z

$$
\begin{aligned}
Var(Z_1) &= \frac{1}{2}Var(X_1 + X_2) = \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right) \\
Var(Z_2) &= \frac{1}{2}Var(X_1 - X_2) = \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)
\end{aligned}
$$

$$
Cov(Z_1, Z_2) = \mathbb{E}(Z_1Z_2) - \mathbb{E}Z_1\mathbb{E}Z_2 = \mathbb{E}(Z_1Z_2)
$$

and

$$
\mathbb{E}(Z_1Z_2) = \frac{1}{2}\mathbb{E}\left(X_1^2 - X_2^2\right) = \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)
$$

Thus,

$$
\Sigma = \frac{1}{2}\left(\begin{array}{cc} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{array}\right)
$$

Compute Λ

The precision matrix is:

$$
\Lambda = \frac{1}{2\sigma_1^2\sigma_2^2}\left(\begin{array}{cc} \sigma_1^2 + \sigma_2^2 & \sigma_2^2 - \sigma_1^2 \\ \sigma_2^2 - \sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{array}\right)
$$

Compute eigenvalues and eigenvectors of Σ

$$
\begin{aligned}
0 &= det\begin{pmatrix}\frac{1}{2} & a+b-\lambda & b-a \\ & b-a & a+b-\lambda\end{pmatrix} \\
&= (\frac{a+b}{2} - \lambda)^2 - (\frac{b-a}{2})^2 \\
&= ab + \lambda^2 - \lambda(a+b) = (a-\lambda)(b-\lambda)
\end{aligned}
$$

Thus, the eigenvalues of Σ are σ_1^2 and σ_2^2.

So we now have the two eigenvalues, σ_1^2 and σ_2^2

We then seek the corresponding eigenvectors.

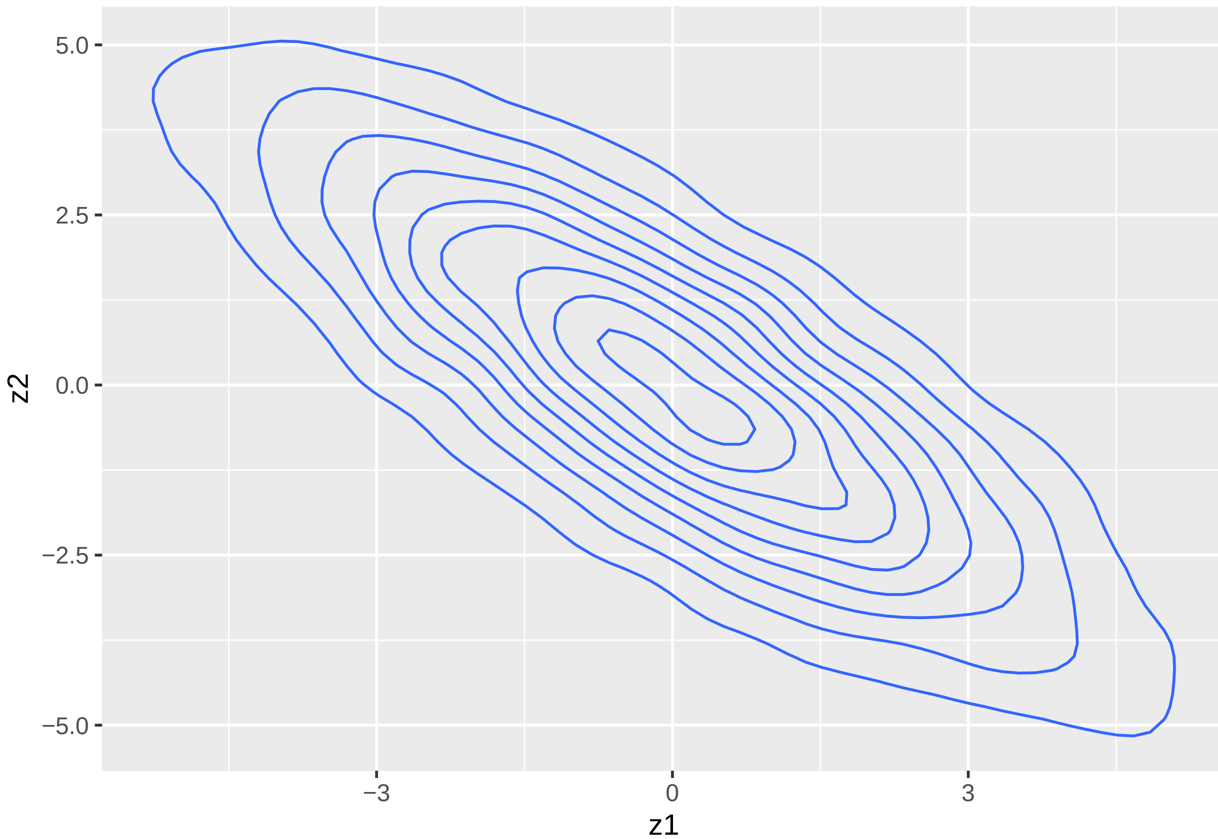
We see that these are $(1, 1)$ and $(1, -1)$.

Level curves of the pdf

```r
s1 <- 1
s2 <- 10
```

```r
x1 <- rnorm(n = 10000, mean = 0, sd = sqrt(s1))
x2 <- rnorm(n = 10000, mean = 0, sd = sqrt(s2))
z1 <- (x1 + x2) / sqrt(2)
z2 <- (x1 - x2) / sqrt(2)
```

```r
library(ggplot2)
library(magrittr)
tibble::tibble(z1 = z1, z2 = z2) %>%
  ggplot() + geom_density_2d(aes(x = z1, y = z2))
```



Part A.ii

Find the minimizer

Level curves

Part A.iii

Find the minimizer

Level curves

Part B: A summary of the two articles

Bayesian inference for complex, multilevel statistical models historically involved sampling-based Markov chain Monte Carlo (MCMC) methods. In this approach, one constructs a Markov chain for which the stationary distribution is the posterior distribution for the unknowns in the statistical model. Empirically summaries of the posterior distribution are computed from thousands of samples.

Over the last 20 years, researchers in machine learning and statistics have developed variational inference methods for study of posterior distributions in complicated Bayesian models. In variational inference, one specifies a simpler family of distributions that approximate the posterior. One then chooses a distribution from the from the approximation family. The chosen distribution minimizes (over the approximation family members) the KL divergence with the posterior.

A major advantage of variational inference over sampling-based methods is the lesser computing time for variational inference. Despite this advantage, variational inference was plagued by the need for model-specific derivations. Ranganath, Gerrish, and Blei (2014) recognized this shortcoming and developed black box variational inference to address this issue. Their work enables rapid exploration of diverse collections of models.

Black box variational inference uses stochastic optimization by calculating noisy gradients of the evidence lower bound (ELBO). Ranganath, Gerrish, and Blei (2014) do this computation with Monte Carlo samples from the score function (Equation 3 in Ranganath, Gerrish, and Blei (2014)). They incorporate Rao-Blackwellization and control variates to reduce the noisiness of the gradients.

Ranganath, Gerrish, and Blei (2014) achieve impressive results in evaluating their black box variational inference methods. Figure 1 of their article illustrates the acceleration in computing time. Figure 2 shows the remarkable gradient variance reduction achieved with the Rao=-Blackwellization and control variates.

Kucukelbir et al. (2015) go a step further in developing automatic differentation variational inference (ADVI) for differentiable probability models. ADVI, which they implement in the Stan programming language, first determines an appropriate variational family. It then optimizes the variational objective function, as in standard variational inference. With this implementation, ADVI can be used for inference with any model that can be expressed in Stan.

References

Kucukelbir, Alp, Rajesh Ranganath, Andrew Gelman, and David Blei. 2015. “Automatic Variational Inference in Stan.” In *Advances in Neural Information Processing Systems*, 568–76.

Ranganath, Rajesh, Sean Gerrish, and David Blei. 2014. “Black Box Variational Inference.” In *Artificial Intelligence and Statistics*, 814–22. PMLR.