Exercise Sheet #3

due on Monday, 15 May 2023, 10am sharp

1. **Theory Question:** Gamma functions, distributions, in one and more dimensions.

   (a) In the lecture, we encountered the **Gamma** distribution, the exponential family with pdf

   $$p(x \mid a, b) = \mathcal{G}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \qquad \text{with} \qquad \Gamma(z) := \int_0^\infty t^{z-1} e^{-t}\, \mathrm{d}t.$$

   We already saw that the Gamma function is a generalization of the factorial function: $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$. Like the Beta integral, it is arguably an intractable object, although extremely good numerical approximations exist. So, as in the Beta case, we can construct a more tractable approximation by constructing the *Laplace approximation*. To do so, consider the unnormalized density $\tilde{p}(x \mid a, b) = x^{a-1} e^{-bx}$, and compute the first two derivatives of its logarithm. Use them to find the mode, and the curvature at the mode, and construct a Taylor approximation. Interpret it as the logarithm of a Gaussian distribution,[1] and find its parameters, as in the lecture. Write out an explicit expression for mean and covariance of this Gaussian approximation (in terms of a, b) for full marks.

   Then consider the special case[2] $b = 1$. Show how it can be used to construct an analytic approximation for the Gamma function, known as *Stirling's approximation* (though it can be traced back to Abraham de Moivre's *Doctrine of Chances*, 1733).

   (b) We now move to the multivariate form of the Gamma distribution. Assume we are given n observations $\boldsymbol{x}_i \in \mathbb{R}^d$ drawn iid. from the multivariate Gaussian likelihood

   $$p(\boldsymbol{x}_i \mid \mu, \Sigma) = \mathcal{N}(\boldsymbol{x}_i \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu})^\intercal \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right).$$

   Assume that we know the mean $\boldsymbol{\mu} \in \mathbb{R}^d$, but not the (symmetric positive definite) precision matrix Σ^{-1}. Show that the conjugate prior for Σ^{-1} under this likelihood is given by the *Wishart distribution*

   $$p(\Sigma^{-1} \mid W, \nu) = \mathcal{W}(\Sigma^{-1} \mid W, \nu) = \frac{|\Sigma^{-1}|^{(\nu-d-1)/2}}{2^{d\nu/2} |W|^{\nu/2} \Gamma_d(\nu/2)} \exp\left(-\frac{1}{2} \operatorname{tr}(W \Sigma^{-1}) \right).$$

   Where tr is the trace. You do not need to prove the form of the normalization constant. It is known as the *multivariate Gamma function* and is given by

   $$\Gamma_d(z) = \pi^{d(d-1)/4} \prod_{i=1}^{d} \Gamma\left(z + \frac{1-i}{2} \right).$$

   What is the posterior distribution for Σ^{-1}?

2. **Practical Question:** can be found in `Ex03.ipynb`

---

[1] As in Laplace's original work, we make a small error here by changing the support of the distributions from $(0, \infty)$ for the Gamma to $(-\infty, \infty)$ for the Gaussian.

[2] The $b = 1$ case is historically called the *Erlang distribution*. The Danish engineer Agner K. Erlang (1878–1929) worked for Copenhagen's Telephone company. In a seminal paper in 1909 he argued that the number of telephone calls in a given interval follows a Poisson distribution. As shown in the lecture, the Gamma distribution is the conjugate prior to the Poisson. In the paper, Erlang effectively uses conjugate prior marginalization to predict how many phone calls Copenhagen's switch boards would need to be able to handle.