# Statistics and Machine Learning 1

# Lecture 7E: The Laplace Approximation

Mark Muldoon
Department of Mathematics, Alan Turing Building
University of Manchester

Week 7

# Nonconjugate priors & the Laplace approximation

When the likelihood and prior aren't conjugate, one can sometimes make progress by approximating the posterior with a Gaussian centered on the MAP estimate: machine learners call this the *Laplace approximation*.

It's closely related to the idea that, near a maximum at $x_\star$, a function $g(x)$ can be approximated via it's Taylor series:

$$\begin{aligned} g(x) &= g(x_\star) + (x - x_\star)g'(x_\star) + \frac{1}{2}(x - x_\star)^2 g''(x_\star) + \dots \\ &\approx g(x_\star) + \frac{1}{2}(x - x_\star)^2 g''(x_\star) \end{aligned}$$

where I have used the fact that, as $x_\star$ is a maximum, $g'(x_\star) = 0$.

# Mean and variance of the Laplace approximation

If we have a posterior density $f(x)$ that has a single maximum at $x_\star$, then the Laplace approximation involves saying that $f$ is approximately a normal distribution with mean $\mu_L$ and variance $\sigma_L^2$ satisfy $\mu_L = x_\star$ and

$$\frac{-1}{\sigma_L^2} = \left( \frac{d^2}{dx^2} \log(f) \right|_{x=x_\star} = \left( \frac{\frac{d^2 f}{dx^2}}{f} \right|_{x=x_\star}$$

# Example: Laplace approx. is exact for a Gaussian

The density for a Gaussian with mean $\mu$ and variance $\sigma^2$ has a global maximum at $x_\star = \mu$ and the log of the density is

$$\log(f(x)) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

so

$$\frac{d^2}{dx^2}\left(\log(f(x))\right) \;=\; \frac{d^2}{dx^2}\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \;=\; \frac{d}{dx}\left(-\frac{(x-\mu)}{\sigma^2}\right) \;=\; \frac{-1}{\sigma^2}$$

In this case the Laplace approximation would be $\mathcal{N}(\mu_L, \sigma_L^2)$ with

$$\mu_L = x_\star = \mu \;\text{ and }\; \sigma_L^2 = \frac{-1}{\left(\frac{d^2}{dx^2}\log(f(x))\right|_{x=x_\star}} = \frac{-1}{-1/\sigma^2} = \sigma^2$$

and so agrees exactly with the original distribution.

# Example: Laplace approx. to the Beta distribution

Suppose we didn't know about conjugate priors and so wanted to make a Laplace approximation to the Beta-distributed posterior in our polling example. A Beta distribution with shape parameters $\alpha > 1$ and $\beta > 1$ has its mode at $p_\star = (\alpha - 1)/(\alpha + \beta - 2)$.

The log of the density function is

$$\log\left(\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)\right) + (\alpha - 1)\log(p) + (\beta - 1)\log(1 - p)$$

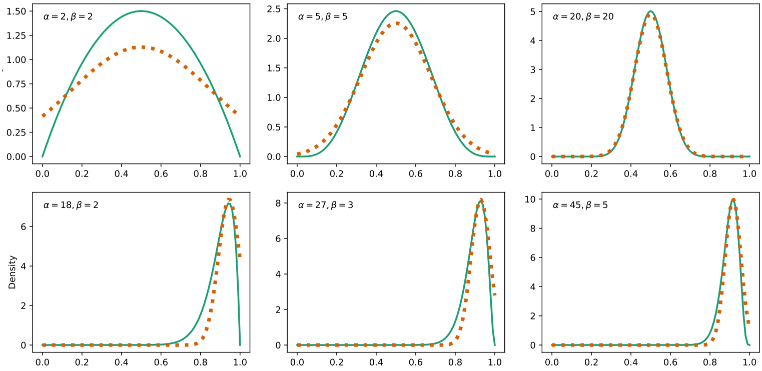
so its second derivative is

$$\frac{d^2}{dp^2}\log(f(x)) = -\frac{\alpha - 1}{p^2} - \frac{\beta - 1}{(1-p)^2}.$$

Substituting $p_\star$ into this and tidying up yields a Laplace approximation $\mathcal{N}(\mu_L, \sigma_L^2)$ with

$$\mu_L = \frac{\alpha - 1}{\alpha + \beta - 2} \quad \text{and} \quad \sigma_L^2 = \frac{(\alpha - 1)(\beta - 1)}{(\alpha + \beta - 2)^3}.$$

# Laplace approx. for Beta distributions




Above: densities for various Beta distributions (green) along with their Laplace approximations (orange, dotted). In general, the approximation is better when $\alpha + \beta$ is large and when the peak of the Beta distribution is symmetric.