Exercise Sheet #7

Solution

December 4, 2018

1 The original Laplace approximation

We consider the un-normalized Beta Function

$$\tilde{p}(r) = r^{(a-1)}(1-r)^{(b-1)} \tag{1}$$

and its logarithm

$$\log \tilde{p}(r) = (a-1) \cdot \log(r) + (b-1) \cdot \log(1-r) \tag{2}$$

(a)

The first derivative of (2) is given by

$$\begin{aligned} g(r) := \frac{\partial \log \tilde{p}(r)}{\partial r} &\overset{(2)}{=} \frac{\partial}{\partial r}(a-1) \cdot \log(r) + (b-1) \cdot \log(1-r) \\ &= \frac{(a-1)}{r} - \frac{(b-1)}{(1-r)} \end{aligned}$$

and the second derivative by

$$\begin{aligned} \psi(r) := \frac{\partial^2 \log \tilde{p}(r)}{(\partial r)^2} = \frac{\partial}{\partial r} g(r) = \frac{\partial}{\partial r} \frac{(a-1)}{r} &- \frac{(b-1)}{(1-r)} \\ = -\frac{(a-1)}{r^2} &- \frac{(b-1)}{(1-r)^2} \end{aligned}$$

(b)

Using exercise 1 (a), we find $\hat{r}$, such that $g(\hat{r}) = 0$

$$\begin{aligned} g(\hat{r}) = \frac{(a-1)}{\hat{r}} - \frac{(b-1)}{(1-\hat{r})} &\overset{!}{=} 0 \\ \iff \hat{r} &= \frac{a-1}{a+b-2} \end{aligned} \tag{3}$$

and we immediatly see that $a + b \neq 2$ must be fulfilled.

We also compute $\psi(\hat{r})$:

$$\begin{aligned}\psi(\hat{r}) = -\frac{(a-1)}{(\frac{a-1}{a+b-2})^2} - \frac{(b-1)}{(1-\frac{a-1}{a+b-2})^2} &= -\frac{(a+b-2)^2}{(a-1)} - \frac{(a+b-2)^2}{(b-1)} \\ &= -\frac{(a+b-2)^3}{(a-1)(b-1)}\end{aligned} \tag{4}$$

We have two constraints that restrict the values of a and b.

$$\hat{r} \in (0,1) \tag{5}$$

$$-\psi(\hat{r}))^{-1} \geq 0 \text{ the variance must be non-negative.} \tag{6}$$

If we rewrite Eq. (3) as

$$\hat{r} = \frac{1}{1 + \underbrace{\frac{b-1}{a-1}}_{>0}}$$

we see that condition (5) requires $(a-1)$ and $(b-1)$ to have the same sign i.e. $0 < a, b < 1$ or $a, b > 1$. By looking at condition (6) and Eq. (4) we see $a, b > 1$ is required for a positive variance.

(c)

We construct a Taylor approximation of $\log(\tilde{p}(r))$ around $\hat{r}$:

$$\begin{aligned}\log \tilde{p}(r) &\approx \log \tilde{p}(\hat{r}) + g(\hat{r})(r-\hat{r}) + \frac{1}{2}\psi(\hat{r})(r-\hat{r})^2 \\ &\overset{g(\hat{r})=0}{=} \log \tilde{p}(\hat{r}) - \frac{(r-\hat{r})^2}{-2\psi^{-1}(\hat{r})}\end{aligned}$$

which gives us an approximation for $\tilde{p}(r)$:

$$\tilde{p}(r) \approx \tilde{p}(\hat{r}) \cdot e^{-\frac{(r-\hat{r})^2}{-2\psi^{-1}(\hat{r})}}$$

So now we can construct an approximation of the Beta Function:

$$\begin{aligned}B(a,b) = \int_0^1 \tilde{p}(r)\,dr \approx \int_{-\infty}^{\infty} \tilde{p}(r)\,dr \approx \int_{-\infty}^{\infty} \tilde{p}(\hat{r}) \cdot e^{-\frac{(r-\hat{r})^2}{-2\psi^{-1}(\hat{r})}}\,dr& \\ = \hat{r}^{(a-1)}(1-\hat{r})^{(b-1)} \int_{-\infty}^{\infty} e^{-\frac{(r-\hat{r})^2}{-2\psi^{-1}(\hat{r})}}\,dr& \\ = \hat{r}^{(a-1)}(1-\hat{r})^{(b-1)} \sqrt{2\pi(-\psi^{-1}(\hat{r}))}& \\ = \left(\frac{a-1}{a+b-2}\right)^{(a-1)} \left(1 - \frac{a-1}{a+b-2}\right)^{(b-1)} \sqrt{2\pi\frac{(a-1)(b-1)}{(a+b-2)^3}}&\end{aligned}$$

2 The Gamma Function

We know that the Beta Function can be written as

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{7}$$

and that the Beta function satisfies the recursive property

$$\Gamma(x+1) = x \cdot \Gamma(x) \tag{8}$$

Mean

$$\begin{aligned}
m(a,b) = \mathbb{E}_{\mathcal{B}(r;a,b)}[r] = \int_0^1 r \cdot \mathcal{B}(r;a,b)\,dr &= \int_0^1 \frac{1}{B(a,b)} r \cdot r^{a-1}(1-r)^{b-1}\,dr \\
&= \frac{\int_0^1 r^a(1-r)^{b-1}}{B(a,b)} = \frac{B(a+1,b)}{B(a,b)} \\
&\overset{(7)}{=} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&\overset{(8)}{=} \frac{a \cdot \Gamma(a)\Gamma(b)}{(a+b) \cdot \Gamma(a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \frac{a}{a+b}
\end{aligned}$$

Variance

$$\begin{aligned}
v(a,b) = \mathbb{E}_{\mathcal{B}(r;a,b)}[r^2] - m^2(a,b) &= \int_0^1 r^2 \cdot \mathcal{B}(r;a,b)\,dr - m^2(a,b) \\
&= \int_0^1 \frac{1}{B(a,b)} r^2 \cdot r^{a-1}(1-r)^{b-1}\,dr - \frac{a^2}{(a+b)^2} \\
&= \frac{B(a+2,b)}{B(a,b)} - \frac{a^2}{(a+b)^2} \\
&\overset{(7)}{=} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} - \frac{a^2}{(a+b)^2} \\
&\overset{(8)}{=} \frac{(a+1) \cdot a \cdot \Gamma(a)\Gamma(b)}{(a+b+1) \cdot (a+b) \cdot \Gamma(a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} - \frac{a^2}{(a+b)^2} \\
&= \frac{(a^2+a)(a+b) - a^2(a+b+1)}{(a+b+1)(a+b)^2} \\
&= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2b - a^2}{(a+b+1)(a+b)^2} \\
&= \frac{ab}{(a+b+1)(a+b)^2}
\end{aligned}$$

3 Stirling's approximation

We consider the unnormalized density of the Gamma distribution:

$$\tilde{p}(t|a,b) = b^a t^{a-1} e^{-bt}. \tag{9}$$

(a)

For the first derivative of $\log \tilde{p}(t|a,b)$, we get

$$\begin{aligned}\frac{\partial}{\partial t} \log \tilde{p}(t|a,b) \overset{(9)}{=} \frac{\partial}{\partial t} \log(b^a t^{a-1} e^{-bt}) &= \frac{\partial}{\partial t} \left(a \log(b) + (a-1)\log(t) - bt\right) \\ &= \frac{a-1}{t} - b\,,\end{aligned}$$

and for the second derivative

$$\frac{\partial^2}{\partial t^2} \log \tilde{p}(t|a,b) = \frac{\partial}{\partial t} \left(\frac{a-1}{t} - b\right) = -\frac{a-1}{t^2}\,.$$

(b)

We can find the mode $\hat{t} := \arg\max_{t>0} \log \tilde{p}(t|a,b)$ be finding $\hat{t}$ where the derivative is zero.

$$\frac{\partial}{\partial t} \log \tilde{p}(t|a,b) \overset{(a)}{=} \frac{a-1}{t} - b \overset{!}{=} 0$$

which is true at $\hat{t} = \frac{a-1}{b}$.
We can evaluate the Hessian at this point

$$\frac{\partial^2}{\partial t^2} \log \tilde{p}(t|a,b)\Big|_{t=\hat{t}} \overset{(a)}{=} -\frac{a-1}{\hat{t}^2} = -\frac{(a-1)b^2}{(a-1)^2} = -\frac{b^2}{a-1} = \psi(\hat{t})\,.$$

(c)

We can construct a Taylor approximation of $\log \tilde{p}(t|a,b)$

$$\log \tilde{p}(t|a,b) \approx \log \tilde{p}(\hat{t}|a,b) - \frac{1}{2}\frac{(t-\hat{t})^2}{-\psi^{-1}(\hat{t})}$$

and thus similar to 1(a) we can construct an approximation for the normalization constant $\Gamma(a)$. Note that in the following $b = 1$.

$$\begin{aligned}\Gamma(a, b=1) = \Gamma(a) = \int_0^1 \tilde{p}(t|a, b=1)\,dt \approx \int_{-\infty}^{\infty} \tilde{p}(t|a,1)\,dt \approx \int_{-\infty}^{\infty} \tilde{p}(\hat{t}|a,1) e^{-\frac{1}{2}\frac{(t-\hat{t})^2}{-\psi^{-1}(\hat{t})}}\,dt& \\ = \tilde{p}(\hat{t}|a,1) \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(t-\hat{t})^2}{-\psi^{-1}(\hat{t})}}\,dt& \\ = 1^a \hat{t}^{a-1} e^{-\hat{t}} \sqrt{2\pi(-\psi^{-1}(\hat{t}))}& \\ = (a-1)^{a-1} e^{-(a-1)} \sqrt{2\pi(a-1)}&\end{aligned} \tag{10}$$

Since we know that $\Gamma(n+1) = n!$, we can use the result we just obtained, to get Stirling's approximation

$$a! = \Gamma(a+1) \overset{(10)}{\approx} a^a e^{-a} \sqrt{2\pi a}$$