ANLY601 Take Home Assignment 5

Xunge Jiang

Mar 17, 2019

1 Explain the intuition behind the hyperparameters of the Beta distribution α_1 and α_2. For example, when these values are higher, does this imply having more prior information or less? Why?

α_1 and α_2 are considered as results from previous trials where α_1 means the number of success and α_2 is the number of failures. For each time we get new results, we have to then update our α_1 and α_2 to reflect our new information. Therefore, when these values are higher, it means we are having more prior information.

2 What values for α_1 and α_2 using a Beta prior indicate an uninformative prior?

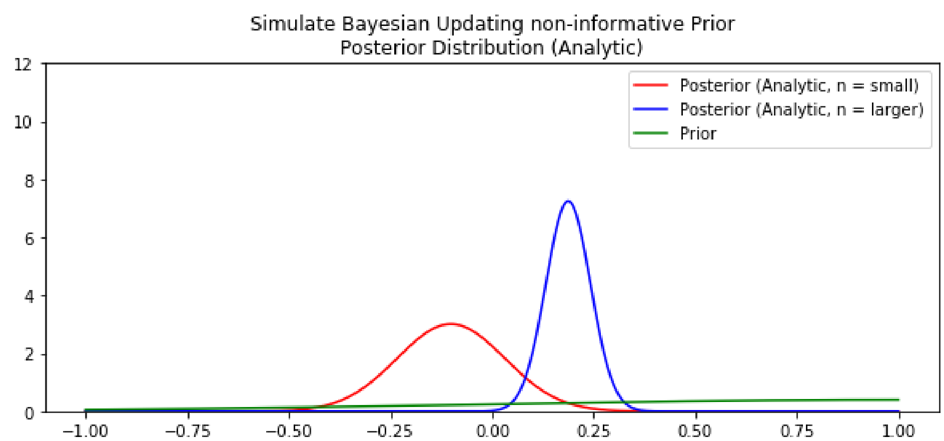
When α_1 and α_2 are equal, the Beta Prior indicates an uninformative prior since it has a uniform distribution and number of success and failures are the same (not informative).

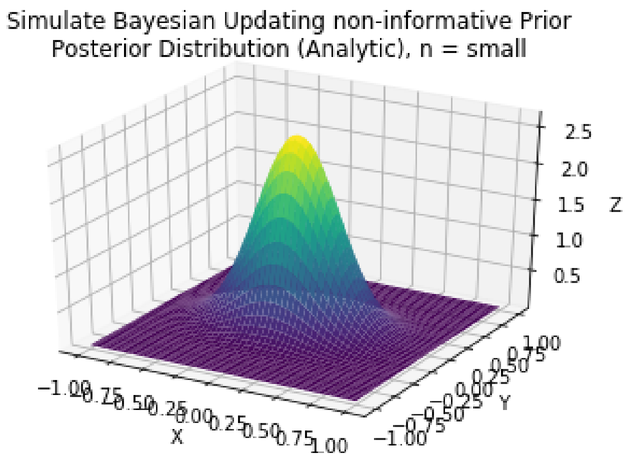
3 Independent Research. What is a Jefferies prior and how is it related to Beta?

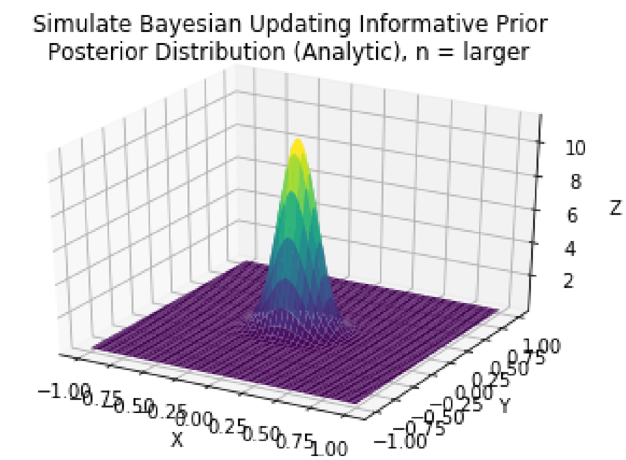
Jeffreys prior is a non-informative prior distribution and it is based on the Fisher information function of a model. Jeffreys priors are widely used in Bayesian analysis. In general, they are not conjugate priors.
Jeffrey prior is Beta Distribution only when $\alpha = \beta = \frac{1}{2}$, which has density that looks roughly like a bucket suspended above [0,1].

4 Consider a Gaussian-Inverse-Wishart Prior. How many hyper parameters are there? Explain how would you indicate prior knowledge via the manipulation of the hyperparameter(s)? Show supporting evidence / figures using your code from In-Class Exercise.

There are four hyperparameters, $\mu_0,\ k_0,\ v_0,\ \psi$.
When μ is smaller, we have more prior knowledge.
When k_0 is larger, we have more prior knowledge.
When v_0 is larger, we have more prior knowledge.
When ψ is larger, we have more prior knowledge.
I started with
$\mu = [2, 2]$ and update to $\mu = [-1e - 17, -5e - 18]$
$k_0 = 2$ and update to $k_0 = 27$
$v_0 = 10$ and update to $v_0 = 35$
$\psi = [1, 1]$ and update to $\psi = [8.74, 8.74]$
and the updating graphs and code are shown below.








```
a1 = [2, 2] #mu
a2 = 2      #k0
a3 = 10     #df/v0
a4 = [1, 1] #psi

# Prior Mean
prior_mean = a1
print('Prior mean:', prior_mean)

x1 = np.linspace(-1,1, 1000)
x2 = np.linspace(-1,1, 1000)
X1, X2 = np.meshgrid(x1, x2)
Z = np.empty(X1.shape + (2,))
Z[:, :, 0] = X1
```

```
Z[:, :, 1] = X2 # Z is the data

# Plot the prior
fig = plt.figure()
ax = plt.axes(projection='3d')
ax.plot_surface(X1, X2, Y.pdf(Z), cmap='viridis')
ax.set_xlabel('X')
ax.set_ylabel('Y')
ax.set_zlabel('Z')
ax.set(title='Prior Distribution (Gaussian Inverse Wishart)')
plt.show()
###############################################
# N = small
# Find the hyperparameters of the posterior

# update hyperparameters
a1_hat = [(a2*mu[0] + nobs*X1.mean())/(a2+nobs), (a2*mu[1] + nobs*X2.mean())/(a2+nobs)]
a2_hat = a2 + nobs
a3_hat = a3 + nobs
a4_hat = a4 + Z.var() + (a2 * nobs / (a2+nobs)) * (Z.mean() - a1) * (Z.mean() - a1)
sigma_hat = stats.invwishart.rvs(a3_hat, a4_hat, 1) / a2_hat
Y = stats.multivariate_normal(a1_hat, [[sigma_hat[0,0], 0], [0,sigma_hat[1,1]]])

# Posterior Mean
post_mean = a1_hat
print('Posterior Mean (Analytic):', post_mean)

# Plot the analytic posterior after N = 4 Observations
fig = plt.figure()
ax = plt.axes(projection='3d')
ax.plot_surface(X1, X2, Y.pdf(Z), cmap='viridis')
ax.set_xlabel('X')
ax.set_ylabel('Y')
ax.set_zlabel('Z')
ax.set(title='Simulate Bayesian Updating non-informative Prior\nPosterior Distribution (Anal
plt.show()
#####################################################

########################################################
# Simulate data  N = larger

np.random.seed(123)

nobs = 25
mu = [0, 0]
k0 = 2
```

```
df=10
psi = [1, 1]
sigma = stats.invwishart.rvs(df=df, scale=psi, size=1) / k0
Y = stats.multivariate_normal(mu, [[sigma[0,0], 0], [0,sigma[1,1]]])

a1_hat = [(a2*mu[0] + nobs*X1.mean())/(a2+nobs), (a2*mu[1] + nobs*X2.mean())/(a2+nobs)]
a2_hat = a2 + nobs
a3_hat = a3 + nobs
a4_hat = a4 + Z.var() + (a2 * nobs / (a2+nobs)) * (Z.mean() - a1) * (Z.mean() - a1)
sigma_hat = stats.invwishart.rvs(a3_hat, a4_hat, 1) / a2_hat
Y = stats.multivariate_normal(a1_hat, [[sigma_hat[0,0], 0], [0,sigma_hat[1,1]]])

# Posterior Mean
post_mean = a1_hat
print('Posterior Mean (Analytic):', post_mean)

# Plot the analytic posterior after N = 4 Observations
fig = plt.figure()
ax = plt.axes(projection='3d')
ax.plot_surface(X1, X2, Y.pdf(Z), cmap='viridis')
ax.set_xlabel('X')
ax.set_ylabel('Y')
ax.set_zlabel('Z')
ax.set(title='Simulate Bayesian Updating Informative Prior\nPosterior Distribution (Analytic
plt.show()
```

5 Show that the variance of the Bernoulli Distribution is equal to $\mu(1-\mu)$.

The expected value of Bernoulli Distribution:

$$\begin{aligned} E[x] &= \sum xp(x) && (1)\\ &= 1 \times p(1) + 0 \times p(0) && (2)\\ &= \mu + 0 * (1-\mu) && (3)\\ &= \mu && (4) \end{aligned}$$

Also,

$$\begin{aligned} E[x^2] &= \sum x^2 p(x) && (5)\\ &= 1^2 p(1) + 0^2 p(0) && (6)\\ &= \mu + 0 * (1-\mu) && (7)\\ &= \mu && (8) \end{aligned}$$

By formula, $Var[x] = E[x^2] - E[x]^2$,

$$
\begin{aligned}
Var[x] &= E[x^2] - E[x]^2 && (9)\\
&= \mu - \mu^2 && (10)\\
&= \mu(1-\mu) && (11)
\end{aligned}
$$

6 Show that the Entropy of a Bernoulli Random Variable X is equal to $H[X] = -\mu ln[\mu] - (1-\mu)ln[1-\mu]$

The probability mass functions of Bernoulli Random Variable X is

$$
Bern[x] = \begin{cases} \mu^x(1-\mu)^{1-x} & , \ x = 0,1 \\ 0 & , \ x \neq 0,1 \end{cases}
$$

Therefore,

$$
\begin{aligned}
H[x] &= -\sum_{x \in \{0,1\}} p(x)\, ln\, p(x) && (12)\\
&= -p(x=0)\, ln\, p(x=0) - p(x=1)\, ln\, p(x=1) && (13)\\
&= -(1-\mu)\, ln\, (1-\mu) - \mu\, ln\, \mu && (14)
\end{aligned}
$$

7 Given random variable X, with Gaussian likelihood p(X | μ), and Gaussian prior p(μ), the posterior will also be Gaussian. Corollary: the product of two Gaussian PDFs is Gaussian.

Let $f(x)$ and $g(x)$ be Gaussian PDFs with arbitrary means μ_f and μ_g and standard deviations σ_f and σ_g :

$$
f(x) = \frac{1}{\sqrt{2\pi}\sigma_f} e^{-\frac{(x-\mu_f)^2}{2\sigma_f^2}} \quad and \quad g(x) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(x-\mu_g)^2}{2\sigma_g^2}}
$$

The product of these two is

$$
f(x)g(x) = \frac{1}{2\pi\sigma_f\sigma_g} e^{-\left(\frac{(x-\mu_f)^2}{2\sigma_f^2} + \frac{(x-\mu_g)^2}{2\sigma_g^2}\right)}
$$

Let

$$
\beta = \frac{(x-\mu_f)^2}{2\sigma_f^2} + \frac{(x-\mu_g)^2}{2\sigma_g^2}
$$

By some transformation:

$$\beta = \frac{x^2 - 2\frac{\mu_f {\sigma_g}^2 + \mu_g {\sigma_f}^2}{{\sigma_f}^2 + {\sigma_g}^2} x + \frac{\mu_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}}{2\frac{\sigma_f^2 \sigma_g^2}{{\sigma_f}^2 + {\sigma_g}^2}}$$

So the first equation is a Gaussian function. Compare the terms in the second equation to a usual Gaussian form

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x^2 - 2\mu x + \mu^2)}{2\sigma^2}}$$

Since a term ϵ that is independent of x can be added to complete the square in β, this is sufficent to complete the proof in cases where the normalisation can be ignored. The product of two Gaussian PDFs is proportional to a Gaussian PDF with a mean that is half the coefficient of x in second equation and a standard deviation that is the square root of half of the denominator i.e.

$$\sigma_{fg} = \sqrt{\frac{\sigma_f^2 \sigma_g^2}{\sigma_f^2 + \sigma_g^2}} \quad and \quad \mu_{fg} = \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}$$

i.e. the variance σ_{fg}^2 is half the harmonic mean of the individual variances σ_f^2 and $sigma_g^2$, and the mean μ_{fg} is the sum of the individual means μ_f and μ_g weighted by their variances. In general, the product is not itself a PDF as, due to the presence of the scaling factor, it will not have the correct normalisation. The product $f(x)g(x)$ can now be written in the usual Gaussian form directly, with an unknown scaling constant (this may be sufficient in cases where renormalisation can be applied). Alternatively, proceeding from the second equation, suppose that ϵ is the term required to complete the square in β i.e.

$$\epsilon = \frac{\left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}\right)^2 - \left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}\right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{(\sigma_f^2 + \sigma_g^2)}} = 0$$

Adding this term to β gives

$$\beta = \frac{x^2 - 2x\frac{\mu_f {\sigma_g}^2 + \mu_g {\sigma_f}^2}{{\sigma_f}^2 + {\sigma_g}^2} + \frac{\mu_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}}{2\frac{\sigma_f^2 \sigma_g^2}{{\sigma_f}^2 + {\sigma_g}^2}} + \frac{\left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}\right)^2 - \left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}\right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{(\sigma_f^2 + \sigma_g^2)}}$$

$$= \frac{x^2 - 2x\frac{\mu_f {\sigma_g}^2 + \mu_g {\sigma_f}^2}{{\sigma_f}^2 + {\sigma_g}^2} + \left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}\right)^2}{2\frac{\sigma_f^2 \sigma_g^2}{{\sigma_f}^2 + {\sigma_g}^2}} + \frac{\frac{\mu_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2} - \left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}\right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{(\sigma_f^2 + \sigma_g^2)}}$$

By some manipulation and calculation, we get

$$\beta = \frac{\left(x - \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_f^2 + \sigma_g^2}\right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{(\sigma_f^2 + \sigma_g^2)}} + \frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)} = \frac{\left(x - \mu_{fg}\right)^2}{2\sigma_{fg}^2} + \frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)}$$

Therefore,

$$f(x)g(x) = \frac{1}{2\pi\sigma_f\sigma_g} exp\Big[-\frac{(x - \mu_{fg})^2}{2\sigma_{fg}^2}\Big] exp\Big[-\frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)}\Big]$$

Therefore, the product of two Gaussians PDFs $f(x)$ and $g(x)$ is a Gaussian PDF:

$$f(x)g(x) = \frac{S_{fg}}{\sqrt{2\pi}\sigma_{fg}} exp\Big[-\frac{(x - \mu_{fg})^2}{2\sigma_{fg}^2}\Big]$$

where

$$\sigma_{fg} = \sqrt{\frac{\sigma_f^2 \sigma_g^2}{\sigma_f^2 + \sigma_g^2}} \quad and \quad \mu_{fg} = \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{{\sigma_f}^2 + {\sigma_g}^2}$$

$$S_{fg} = \frac{1}{\sqrt{2\pi(\sigma_f^2 + \sigma_g^2)}} exp\Big[-\frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)}\Big]$$

Conclusion: The product of two Gaussian PDFs is Gaussian.