Homework 1

Kevin Yang

September 28, 2021

Exercise 1

Let $X = (x_1, x_2, \ldots, x_n)$
$x_i \sim Pois(\lambda)$ $\qquad\qquad \lambda \sim \Gamma(\alpha, \beta)$

We have that:

$$
\begin{aligned}
P(\lambda \mid X) &= \frac{P(X \mid \lambda)P(\lambda)}{P(X)} \\
&\propto P(X \mid \lambda)P(\lambda) \\
&\propto \lambda^{n\bar{x}} e^{-n\lambda} \lambda^{\alpha-1} e^{-\beta\lambda} \\
&= \lambda^{n\bar{x}+\alpha-1} e^{-(n+\beta)\lambda}
\end{aligned}
$$

Therefore, $P(\lambda \mid X) \sim \Gamma(n\bar{x} + \alpha, \beta + n)$ and the Gamma distribution is conjugate to the Poisson distribution.

Exercise 2

Let $s = (s_1, s_2, \dots, s_n)$, $s' = (s'_1, s'_2, \dots, s'_n)$ and $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$

Detailed balance equation: $\pi(s)P(s, s') = \pi(s')P(s', s)$

w.l.g consider an update for s_1

Case 1: $s_{-1} \neq s'_{-1}$

$$\pi(s)P(s, s') = \pi(s')P(s', s) = 0$$

Case 2: $s_{-1} = s'_{-1}$

$$
\begin{aligned}
\pi(s)P(s, s') &= \pi(s)P(s'_1 \mid s'_{-1}) \\
&= \pi(s)\frac{\pi(s')}{\sum_z \pi(z, s'_{-1})} \\
&= \pi(s')\frac{\pi(s)}{\sum_z \pi(z, s_{-1})} \\
&= \pi(s')P(s_1 \mid s_{-1}) \\
&= \pi(s')P(s', s)
\end{aligned}
$$

Consider the move from s to s'. The acceptance probability for MH will be

$$\frac{\pi(s')P(s', s)}{\pi(s)P(s, s')}$$

We have that

$$
\begin{aligned}
\frac{P(s', s)}{P(s, s')} &= \frac{\frac{\pi(s)}{\sum_z \pi(z, s_{-1})}}{\frac{\pi(s')}{\sum_z \pi(z, s'_{-1})}} \\
&= \frac{\pi(s)}{\pi(s')}
\end{aligned}
$$

So

$$\frac{\pi(s')P(s', s)}{\pi(s)P(s, s')} = \frac{\pi(s')\pi(s)}{\pi(s)\pi(s')} = 1$$

Exercise 3

a)

```
##1. enumeration and conditioning:


## condition and marginalize:
## compute joint:
p = np.zeros((2, 2, 2, 2))  # c,s,r,w
for c in range(2):
    for s in range(2):
        for r in range(2):
            for w in range(2):
                p[c, s, r, w] = p_C(c) * p_S_given_C(s, c) * p_R_given_C(r, c) * p_W_given_S_R(w, s, r)


p_C_given_W = np.zeros(2)
for c in range(2):
    for s in range(2):
        for r in range(2):
            p_C_given_W[c] += p[c, s, r, 1]

p_C_given_W /= np.sum(p_C_given_W)

print('There is a {:.2f}% chance it is cloudy given the grass is wet'.format(p_C_given_W[1] * 100))
```

b)

```
##2. ancestral sampling and rejection:
# https://www.cs.ubc.ca/~fwood/CS532W-539W/lectures/mcmc.pdf

num_samples = 10000
samples = np.zeros(num_samples)
rejections = 0
i = 0
while i < num_samples:
    c = np.argmax(np.random.multinomial(1, [p_C(0), p_C(1)]))
    s = np.argmax(np.random.multinomial(1, [p_S_given_C(0, c), p_S_given_C(1, c)]))
    r = np.argmax(np.random.multinomial(1, [p_R_given_C(0, c), p_R_given_C(1, c)]))
    w = np.argmax(np.random.multinomial(1, [p_W_given_S_R(0, s, r), p_W_given_S_R(1, s, r)]))
    if w != 1:
        rejections += 1
        continue
    else:
        samples[i] = c
        i += 1

print('The chance of it being cloudy given the grass is wet is {:.2f}%'.format(samples.mean() * 100))
print('{:.2f}% of the total samples were rejected'.format(100 * rejections / (samples.shape[0] + rejections)))
```

c)

```
##gibbs sampling
num_samples = 10000
samples = np.zeros(num_samples)
state = np.zeros(4, dtype='int')
# c,s,r,w, set w = True

c, s, r, w = 0, 1, 2, 3
i = 0
state[w] = 1
while i < num_samples:
    state[c] = np.argmax(np.random.multinomial(1, p_C_given_S_R[:, state[s], state[r]]))
    state[s] = np.argmax(np.random.multinomial(1, p_S_given_C_R_W[state[c], :, state[r], state[w]]))
    state[r] = np.argmax(np.random.multinomial(1, p_R_given_C_S_W[state[c], state[s], :, state[w]]))

    samples[i] = state[c]
    i += 1


print('The chance of it being cloudy given the grass is wet is {:.2f}%'.format(samples.mean() * 100))
```

Results:

```
There is a 57.58% chance it is cloudy given the grass is wet
The chance of it being cloudy given the grass is wet is 58.05%
34.73% of the total samples were rejected
The chance of it being cloudy given the grass is wet is 58.91%

Process finished with exit code 0
```

Exercise 4

MH within Gibbs on blocks w and $\hat{t}$:

<u>w block</u>

Let $q(w, w')$ be the proposal distribution. We also have that,

$$\pi(w) \propto p(t, x, \sigma^2, w, \alpha)$$

$$= \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t_n - w^T x_n}{\sigma}\right)^2} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{1}{2}\left(\frac{w_d}{\sqrt{\alpha}}\right)^2}$$

So the update probability is,

$$r = \frac{\pi(w')q(w', w)}{\pi(w)q(w, w')}$$

$$= \frac{\prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t_n - w'^T x_n}{\sigma}\right)^2} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{1}{2}\left(\frac{w'_d}{\sqrt{\alpha}}\right)^2} q(w', w)}{\prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t_n - w^T x_n}{\sigma}\right)^2} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{1}{2}\left(\frac{w_d}{\sqrt{\alpha}}\right)^2} q(w, w')}$$

$$\propto e^{-\frac{1}{2}\sum_{n=1}^{N}\left(\frac{t_n - w'^T x_n}{\sigma}\right)^2 - \left(\frac{t_n - w^T x_n}{\sigma}\right)^2} e^{-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{w'_d}{\sqrt{\alpha}}\right)^2 - \left(\frac{w_d}{\sqrt{\alpha}}\right)^2} \frac{q(w', w)}{q(w, w')}$$

<u>$\hat{t}$ block</u>

Let $q(w, w')$ be the proposal distribution.

$$\pi(\hat{t}) \propto N(w^T \hat{x}, \sigma^2)$$

The update probability is,

$$r = \frac{\pi(\hat{t}')q(\hat{t}', \hat{t})}{\pi(\hat{t})q(\hat{t}, \hat{t}')}$$

$$\propto \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{t}' - w^T \hat{x}}{\sigma}\right)^2} q(\hat{t}', \hat{t})}{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{t} - w^T \hat{x}}{\sigma}\right)^2} q(\hat{t}, \hat{t}')}$$

$$= e^{-\frac{1}{2}\left(\frac{\hat{t}'^2 - \hat{t}^2 + 2\hat{t}w^T\hat{x} - 2\hat{t}'w^T\hat{x}}{\sigma^2}\right)} \frac{q(\hat{t}', \hat{t})}{q(\hat{t}, \hat{t}')}$$

Pure Gibbs on blocks w and $\hat{t}$:

<u>w block</u>

$$\log p(\boldsymbol{w} \mid \boldsymbol{t}, \boldsymbol{x}, \sigma^2 \alpha) \propto \log p(\boldsymbol{w}, \boldsymbol{t}, \boldsymbol{x}, \sigma^2, \alpha)$$

$$= \log \prod_{n=1}^{N} p(t_n \mid \boldsymbol{w}, \boldsymbol{x}_n, \sigma^2) p(\boldsymbol{w} \mid 0, \alpha \boldsymbol{I})$$

$$\propto \frac{-\sum_{n=1}^{N}(t_n - \boldsymbol{w}^T \boldsymbol{x}_n)^2}{2\sigma^2} - \frac{1}{2} \boldsymbol{w}^T (\alpha \boldsymbol{I})^{-1} \boldsymbol{w}$$

$$\propto -\frac{1}{2\sigma^2}(\|\boldsymbol{t} - \boldsymbol{x}^T \boldsymbol{w}\|^2) - \frac{1}{2} \boldsymbol{w}^T (\alpha \boldsymbol{I})^{-1} \boldsymbol{w}$$

$$= -\frac{1}{2\sigma^2}(\boldsymbol{t}^T \boldsymbol{t} - 2\boldsymbol{t}^T \boldsymbol{x}^T \boldsymbol{w} + \boldsymbol{w}^T \boldsymbol{x}\boldsymbol{x}^T \boldsymbol{w}) - \frac{1}{2} \boldsymbol{w}^T (\alpha \boldsymbol{I})^{-1} \boldsymbol{w}$$

$$= -\frac{1}{2}(\boldsymbol{w} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{w} - \boldsymbol{\mu}) \sim N(\boldsymbol{\mu}, \Sigma)$$

$$\boldsymbol{\mu} = \sigma^{-2}\Sigma \boldsymbol{x}^T \boldsymbol{t} \qquad \qquad \Sigma^{-1} = \sigma^{-2}\boldsymbol{x}\boldsymbol{x}^T + (\alpha \boldsymbol{I})^{-1}$$

$\hat{\boldsymbol{t}}$ block

$$p(\hat{\boldsymbol{t}} \mid \boldsymbol{t}, \boldsymbol{x}, \sigma^2, \boldsymbol{w}, \alpha) \propto e^{-\frac{1}{2}(\frac{\hat{t} - \boldsymbol{w}^T \hat{x}}{\sigma})^2} \sim Normal(\boldsymbol{w}^T \hat{\boldsymbol{x}}, \sigma^2)$$

Exercise 5

M = number of documents, K = number of topics, V = vocabulary size
$N_{w,i}$ = number of times word w is assigned to topic i
N_i = number of words assigned to topic i
$N_{j,i}$ = number of words in document j assigned to topic i
x_{lj} = l^{th} word in document j (observed)
z_{lj} = topic assignment for the lth word in document j

Joint log likelihood:

$$\log p(z, w \mid \alpha, \beta) \propto \sum_{j=1}^{M}\left(\sum_{i=1}^{K} \log \Gamma(N_{ji} + \alpha_i)\right) - \log\Gamma\left(\sum_{i=1}^{K} N_{ji} + \alpha_i\right)$$

$$+ \sum_{i=1}^{K}\left(\sum_{w=1}^{V} \log\Gamma(N_{wi} + \beta_w)\right) - \log\Gamma\left(\sum_{w=1}^{V} N_{wi} + \beta_w\right)$$

Conditional:

$$p(z_{l,j} = k \mid z^{-lj}, x, \alpha, \beta) = \frac{1}{Z} a_{ji} b_{wi}$$

$$a_{ji} = N_{ji}^{-lj} + \alpha \qquad b_{wi} = \frac{N_{wi}^{-lj} + \beta}{N_i^{-lj} + V\beta} \qquad Z = \sum_{i}^{K} a_{ji} b_{wi}$$