NCTU Machine Learning Hw2

Liang Yu Pan 0486016

November 27, 2015

1 Information Theory

(a)

Overview We use the differential entropy theorem with **Lagrange multipliers** to prove that the probability distribution which maximize the differential entropy is **Gaussian distribution**.

Useful formula

$$H(x) = -\int p(x) \ln p(x) dx \tag{1}$$

$$\int_{-\infty}^{\infty} p(x) dx = 1 \tag{2}$$

$$\int_{-\infty}^{\infty} x p(x) dx = \mu \tag{3}$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = \sigma^2 \tag{4}$$

Solve p(x) By **Lagrange multipliers** and formula(1),(2),(3),(4), we get the formula

$$L = -\int_{-\infty}^{\infty} p(x) \ln p(x) dx + \lambda_1 (\int_{-\infty}^{\infty} p(x) dx - 1) + \lambda_2 (\int_{-\infty}^{\infty} x p(x) dx - \mu) + \lambda_3 (\int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx - \sigma^2) \tag{5}$$

Then solve $\frac{\delta L}{\delta p(x)} = 0$, $\frac{\delta L}{\delta p(x)} = -\ln p(x) - 1 + \lambda_1 + \lambda_2 x + \lambda_3 (x-\mu)^2$
We get

$$p(x) = e^{-1+\lambda_1+\lambda_2 x+\lambda_3 (x-\mu)^2} \tag{6}$$

Substitute (6) into (2), (3), (4), we can get

$$\lambda_1 = 1 - \frac{1}{2}\ln 2\pi\sigma^2 \tag{7}$$

$$\lambda_2 = 0 \tag{8}$$

$$\lambda_3 = \frac{1}{2\sigma^2} \tag{9}$$

Then substitute (7),(8),(9) into (5), we can get

$$p(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{10}$$

which is **Gaussian distribution**.

(b)

derive entropy Substitute (10) into (1), we get $\frac{1}{2}(\ln 2\pi\sigma^2 + 1)$

2 Bayesian Inference for the Gaussian

(a)

Overview Because that $Posterior \propto Likelihood \times Prior$, then Λ_{MAP} can solved by $argmax_\Lambda(Posterior)$.

Solve Posterior

$$Prior : W(\Lambda|W_0, V_0) = B|\Lambda|^{\frac{V_0-D-1}{2}} e^{-\frac{1}{2}tr(W_0^{-1}\Lambda)} \tag{11}$$

$$B(W_0, V_0) = |W_0|^{-\frac{V_0}{2}}[2^{\frac{V_0 D}{2}}\pi^{\frac{D(D-1)}{4}}\prod_{i=1}^{D}\Gamma(\frac{V_0+1-i}{2})]^{-1}, D = 2 \tag{12}$$

$$Likelihood : \prod_{n=1}^{N} N(X_n|\mu, \Lambda^{-1}) \propto |\Lambda|^{\frac{N}{2}} e^{-\frac{1}{2}tr(\Lambda S)} \tag{13}$$

$$S = \Sigma_n (X_n - \mu)(X_n - \mu)^T \tag{14}$$

By (11), (13), we can get $Posterior \propto |\Lambda|^{\frac{V_0-D-1+N}{2}} e^{-\frac{1}{2}tr((W_0^{-1}+S)\Lambda)}$

Solve Λ_{MAP} To $argmax_\Lambda(Posterior)$, we solve $\frac{\partial\Lambda_{MAP}}{\partial\Lambda} = 0$ We get $\Lambda_{MAP} = (V_0 + N - D - 1)(W_0^{-1} + S)^{-1}$

(b)

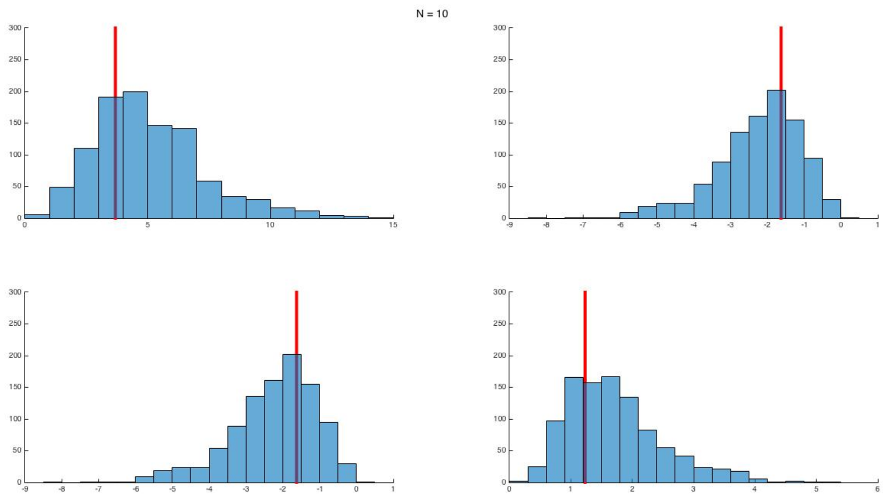

Figure 1: 1000 samples wishart distribution for N=10

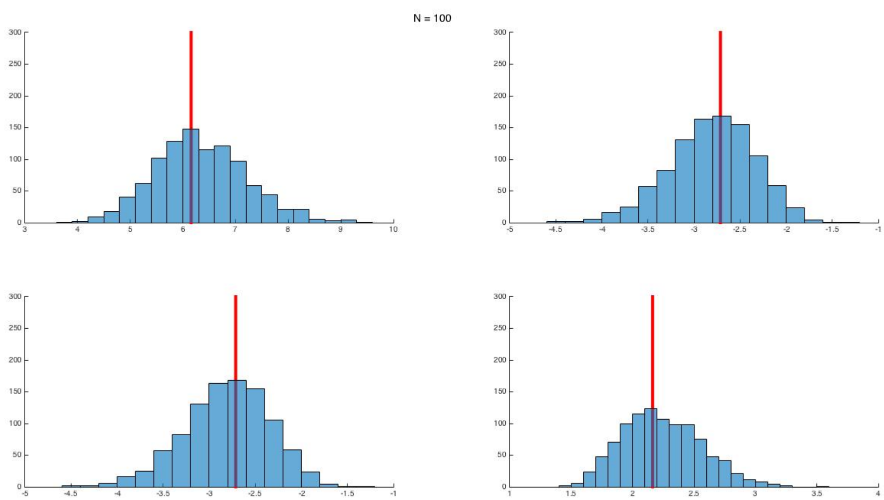

Figure 2: 1000 samples wishart distribution for N=100

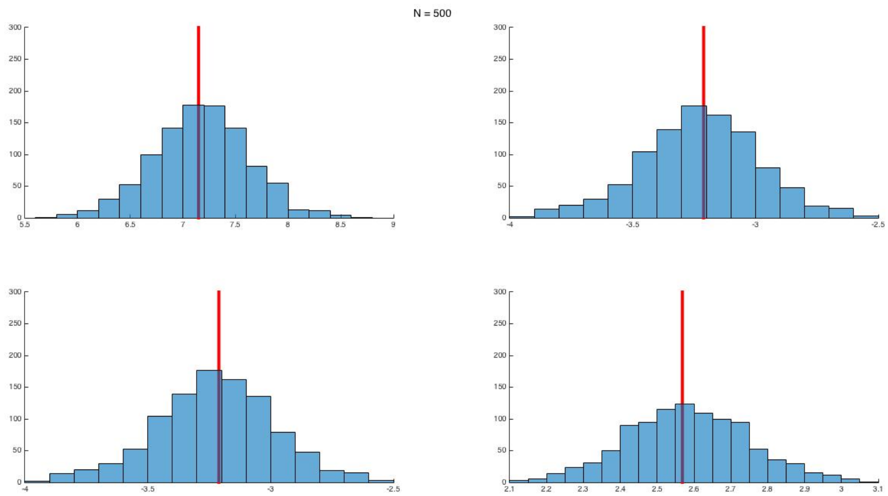

Figure 3: 1000 samples wishart distribution for N=500

3 Bayesian Inference for the Binomial

(a)

Overview Because that $Posterior \propto Likelihood \times Prior$, then μ_{MAP} can solved by $argmax_{\mu}(Posterior)$.

Solve Posterior

$$Prior : Beta(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \tag{15}$$

$$Likelihood : Bin(m_1|N,\mu) = \binom{N}{m_1}\mu^{m_1}(1-\mu)^{N-m_1} \tag{16}$$

By (15), (16), we can get $Posterior \propto \mu^{m_1+a-1}(1-\mu)^{m_2+b-1}, m_2 = N - m_1$

Solve μ_{MAP} To $argmax_{\mu}(Posterior)$, we first log Posterior, get $(m_1 + a - 1)\log(\mu) + (m_2 + b - 1)\log(1-\mu)$. Then, $\frac{\partial \mu_{MAP}}{\partial \mu} = 0$, we get $\mu_{MAP} = \frac{m_1+a-1}{m_1+m_2+a+b-2}$

(b)

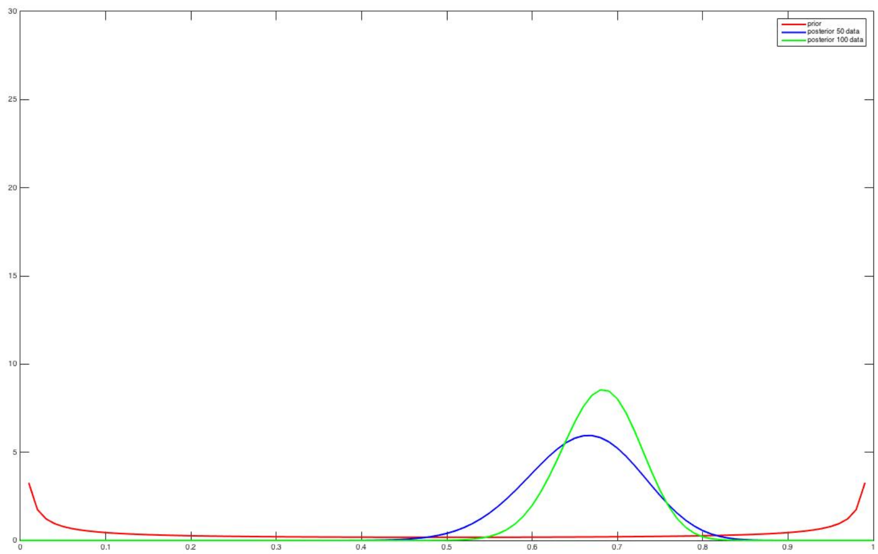

Figure 4: prior distribution and posterior distribution for 50 data and all data respectively