MA 578 — Bayesian Statistics

Homework 3

(Due: Tuesday, 10/8/19)

1. BDA problem 3.8.

2. BDA problem 3.9. In particular, make sure to show that

$$n(\bar{x}-\mu)^2 + \kappa_0(\mu-\mu_0)^2 = (n+\kappa_0)(\mu-\mu_1)^2 + \frac{\kappa_0 n}{\kappa_0+n}(\mu_0-\bar{x})^2$$

   for $\mu_1 := (n\bar{x} + \kappa\mu_0)/(n+\kappa_0)$.

   [*] Show that a similar relation holds for vectors $\bar{x}$, μ, and μ_0,

$$n(\bar{x}-\mu)(\bar{x}-\mu)^\top + \kappa_0(\mu-\mu_0)(\mu-\mu_0)^\top = (n+\kappa_0)(\mu-\mu_1)(\mu-\mu_1)^\top + \frac{\kappa_0 n}{\kappa_0+n}(\mu_0-\bar{x})(\mu_0-\bar{x})^\top,$$

   with $\mu_1 := n/(n+\kappa_0)\bar{x} + \kappa_0/(n+\kappa_0)\mu_0$.

3. BDA problem 3.12.

4. Consider the same setup we discussed in class for the normal case: $X_1, \ldots, X_n \,|\, \mu, \sigma^2 \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, or, in vector notation, $X = (X_1, \ldots, X_n) \,|\, \mu, \sigma^2 \sim N(\mu \mathbf{1}_n, \sigma^2 I_n)$, where $\mathbf{1}_n$ is a vector of ones of length n. Assume a conjugate prior for μ and σ^2, that is, $\mu \,|\, \sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$ and $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$.

   (a) Show that $\sigma^2 \,|\, \mu \sim \text{Inv-}\chi^2(\nu_\mu, \sigma_\mu^2)$ where $\nu_\mu := \nu_0 + 1$ and

$$\sigma_\mu^2 := \frac{\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2}{\nu_0+1}.$$

   (b) Show that by taking $\mathbb{P}(X, \sigma^2 \,|\, \mu)$ and marginalizing out σ^2 we have

$$\mathbb{P}(X \,|\, \mu) \propto \left[1 + \frac{(X-\mu\mathbf{1}_n)^\top(X-\mu\mathbf{1}_n)}{\nu_\mu\sigma_\mu^2}\right]^{-\frac{\nu_\mu+n}{2}},$$

   that is, $X \,|\, \mu \sim t_{\nu_\mu}(\mu\mathbf{1}_n, \sigma_\mu^2 I_n)$, a *multivariate* t distribution with ν_μ degrees of freedom[1].

   (c) Now take the joint $\mathbb{P}(X, \mu, \sigma^2)$ we derived in class and marginalize out σ^2 to show that

$$\mathbb{P}(\mu \,|\, X) \propto \left[1 + \frac{(\mu-\mu_n)^2}{\nu_n\sigma_n^2/\kappa_n}\right]^{-\frac{\nu_n+1}{2}},$$

   with μ_n, κ_n, ν_n, and σ_n^2 being the posterior updated parameters we discussed in class, and so $\mu \,|\, X \sim t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$, a t distribution with ν_n degrees of freedom[2].

---

[1] Check the appendix in BDA for the definition.

[2] That is a shifted and scaled t-distribution, so check BDA again for the precise definition. If you want to relate to the traditional t distribution then $\mu = \mu_n + \sigma_n/\sqrt{\kappa_n}T$ with $T \,|\, X \sim t_{\nu_n}$.

5. [*] Suppose that for $i = 1, \dots, n$, $X_i \mid \alpha, \beta \overset{\text{ind}}{\sim} \mathsf{Gamma}(\alpha, \beta)$ and you wish to define a joint prior on both parameters by first conditioning on α.

   (a) Show that, conditional on α, the conjugate prior for β is $\beta \mid \alpha \sim \mathsf{Gamma}(\kappa_0 \alpha, \lambda_0)$ and that we have Jeffreys prior when $\kappa_0 = \lambda_0 = 0$. Find the conditional posterior $\beta \mid \alpha, X$.

   (b) Marginalize out β from the likelihood to obtain

$$\mathbb{P}(X \mid \alpha) = \lambda_0^{-n} \frac{\Gamma\big((n+\kappa_0)\alpha\big)}{\Gamma(\alpha)^n \Gamma(\kappa_0 \alpha)} \left(\prod_{i=1}^{n} \frac{X_i}{\lambda_0} \right)^{\alpha - 1} \left(1 + \sum_{i=1}^{n} \frac{X_i}{\lambda_0} \right)^{-(n+\kappa_0)\alpha},$$

   that is, $X/\lambda_0 \sim \mathsf{Inv\text{-}Dirichlet}(\alpha, \dots, \alpha, \kappa_0 \alpha)$[3].

   (c) Finally, show that Jeffreys prior for α is

$$\mathbb{P}(\alpha) \propto \Big[n\psi_1(\alpha) + \kappa_0^2 \psi_1(\kappa_0 \alpha) - (n+\kappa_0)^2 \psi_1\big((n+\kappa_0)\alpha\big) \Big]^{1/2},$$

   where ψ_1 is the *trigamma*[4] function.

6. [*] You observe $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu_i, \sigma^2)$, that is, $X = (X_1, \dots, X_n) \mid \mu \sim N(\mu, \sigma^2 I_n)$, and take a conjugate prior for μ, $\mu_i \overset{\text{iid}}{\sim} N(\mu_0, \tau^2)$.

   (a) Show that marginally $X \sim N(\mu_0 \mathbf{1}_n, (\sigma^2 + \tau^2) I_n)$. Moreover, derive the posterior for μ as $\mu \mid X \sim N(\mu_1, \omega \sigma^2 I_n)$, with $\mu_1 := \omega X + (1-\omega)\mu_0 \mathbf{1}_n$. What is ω as a function of σ^2 and τ^2?

   Suppose that you want to follow a frequentist analysis and so you assume that there is a "true" value μ^* for μ. Under this setup, you evaluate an estimator $\widehat{\mu}$ using a *mean squared error* (MSE), defined as

$$\begin{aligned} \mathrm{MSE}(\mu) &= \mathbb{E}_{X \mid \mu^*}\left[(\widehat{\mu} - \mu^*)^\top (\widehat{\mu} - \mu^*) \right] \\ &= \mathrm{tr}\big(\mathrm{Var}_{X \mid \mu^*}[\widehat{\mu}]\big) + \big(\mathbb{E}_{X \mid \mu^*}[\widehat{\mu}] - \mu^*\big)\big(\mathbb{E}_{X \mid \mu^*}[\widehat{\mu}] - \mu^*\big)^\top, \end{aligned}$$

   where tr is the trace operator. Note that the MSE can be decomposed into variance and (squared) bias components.

   Now you want to compare two estimators in particular: the MLE for μ, $\widehat{\mu}_{\text{MLE}} = X$, and the posterior mean μ_1.

   (b) Show that the MLE is unbiased, but that μ_1 is only unbiased when $\mu_1^* = \dots = \mu_n^* = \mu_0$.

   (c) Show that $\mathrm{MSE}(\widehat{\mu}_{\text{MLE}}) = n\sigma^2$ and $\mathrm{MSE}(\mu_1) = \omega^2 n \sigma^2 + (1-\omega)^2 (\mu_0 \mathbf{1}_n - \mu^*)^\top (\mu_0 \mathbf{1}_n - \mu^*)$ and so $\mathrm{MSE}(\mu_1) \le \mathrm{MSE}(\widehat{\mu}_{\text{MLE}})$ whenever $(\mu_0 \mathbf{1}_n - \mu^*)^\top (\mu_0 \mathbf{1}_n - \mu^*) \le n(\sigma^2 + 2\tau^2)$.

[3] Check wikipedia for the definition of an *inverted* Dirichlet.

[4] Again, wikipedia is your friend.

(d) (Empirical Bayes) To elicit the hyper-parameters μ_0 and τ^2, take them as the maximizers of $\log \mathbb{P}(X)$. In particular, show that $\mu_0 = \bar{X}$ and that

$$\omega = 1 - \frac{\sigma^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2 / n}$$

so that $\mu_1 = \omega X + (1 - \omega) \bar{X} \mathbf{1}_n$.[5]

[5] μ_1 is then a *James-Stein* estimator, and can be shown to always beat the MLE with respect to MSE if n is not very small.