# Home reading and exercises: Conjugate priors (and non-informative priors)

Read the following:

- LN pages 56-59 on change of variables and Fisher information (until the equation (4.20))

## Exercises:

**1.**

Let $\{f(\cdot \mid \theta) : \theta \in \Theta\}$ be a parametric model and let $(x_1, ..., x_n)$ be an i.i.d. sample from $f(\cdot \mid \theta_0)$, for some $\theta_0 \in \Theta$. Assume that $t : \mathcal{X}^n \to \mathbb{R}^d$ is a sufficient statistic for the parametric model.

From lecture we already know that

$$\pi(\theta \mid x) = \pi(\theta \mid t(x)).$$

a) Let $\hat{\theta}_n : \mathcal{X}^n \to \Theta$ denote a point estimator derived from the posterior (e.g., the posterior mean or mode). Argue that there exists a function $\tilde{\theta}_n : \mathbb{R}^d \to \Theta$ such that
$$\hat{\theta}_n(x) = \tilde{\theta}_n(t(x)).$$

Bayesian point estimators share this property with other point estimators that are derived from the likelihood, such as the maximum likelihood estimator (MLE). Next assume that $t$ is complete, $E\hat{\theta}_n(X)^2 < \infty$ and that $\hat{\theta}$ is unbiased, i.e. $E\hat{\theta}_n(X) = \theta_0$.

**Theorem (Lehmann-Scheffé):** Let $\{f(\cdot \mid \theta) : \theta \in \Theta\}$ be a parametric model for data $x$ with a sufficient and complete statistic $t(x)$. Any unbiased, quadratically integrable estimator $\hat{\theta}_n$ is a function of $x$ only through $t(x)$ if and only if for any other unbiased square-integrable estimator $\hat{\theta}'_n$,

$$E(\hat{\theta}_n - \theta_0)^2 \leq E(\hat{\theta}'_n - \theta_0)^2.$$

b) Apply the Lehmann-Scheffé theorem to conclude that for any other unbiased square-integrable estimator $\hat{\theta}'_n$

$$E(\hat{\theta}_n - \theta_0)^2 \leq E(\hat{\theta}'_n - \theta_0)^2.$$

The message of this exercise is that if the sufficient statistic is complete, which it often is, then unbiased Bayesian point estimators are $L_2$-optimal in the class of unbiased estimators. They share this property with maximum likelihood estimators.

Of course, the direct usefulness of this result is limited, since the presence of the prior tends to cause bias for point estimators. And quite often we don't even care about unbiasedness (see e.g. *bias-variance tradeoff*). The frequentist performance of Bayesian estimators will be considered again in the future and reaches the conclusion that the posterior generally gives rise to optimal point estimators in an asymptotic sense within much wider class of estimators.

Nevertheless, the mentioned bias can be controlled or even eliminated if one chooses the prior by the methods of empirical Bayes.

Consider a Normal-Normal model

$$\theta \sim \mathcal{N}(\mu, \tau^2)$$
$$X_1, ..., X_n \mid \theta \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2).$$

We know that the posterior mean is given by

$$\hat{\theta}_n(x) = \frac{n\tau^2}{n\tau^2 + \sigma^2}\bar{x} + \frac{\sigma^2}{n\tau^2 + \sigma^2}\mu,$$

where $\bar{x}$ is the sample average.

c) Find the bias of $\hat{\theta}_n$ as an estimator of $\theta$ and propose a method to eliminate the bias.

Well done! You have constructed optimal unbiased point estimator in the sense of Lehmann-Scheffé theorem.

**2. Select your favorite exercise from the exercises 2-5 in the LN page 47 and solve it. Naturally, you are free to solve more than one exercise.**

**3. Solve the exercises 2-3 in the LN page 67.**