SB2.1 Foundations of Statistical Inference
Problem Sheet 4

1. The risks for five decision rules $\delta_1, \ldots, \delta_5$ depend on the value of a nonnegative parameter θ and are given in the table below.

	δ_1	δ_2	δ_3	δ_4	δ_5
$0 \leq \theta < 1$	10	10	7	6	8
$1 \leq \theta < 2$	8	11	8	5	10
$2 \leq \theta$	15	11	12	14	14

   (a) Which decision rules strictly dominate δ_1?

   (b) Which decision rules are admissible?

   (c) Which decision rule is minimax?

   (d) Consider a prior π on θ which is a uniform distribution on $[0, 5]$. Which rule is Bayes with respect to π and what is its Bayes risk?

2. Let $\{X_i\}_{i=1}^n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2 I_p)$ with $\mu = (\mu_1, \ldots, \mu_p)^\top$ and $X_i = (X_{i1}, \ldots, X_{ip})^\top$. Consider the quadratic loss function $L(\mu, \hat{\mu}) = \|\mu - \hat{\mu}\|^2$. Assume that σ^2 is known.

   (a) Find the risk of the MLE $\hat{\mu} = \bar{X}$.

   (b) Find values of a for which

$$\tilde{\mu} = \left(1 - \frac{a}{\|\bar{X}\|^2}\right) \bar{X}$$

   strictly dominates the maximum likelihood estimator.

   (c) For the value of a for which the risk $R(\mu, \tilde{\mu})$ is minimized, show that

$$R(\mu, \tilde{\mu}) \leq \frac{p\sigma^2}{n} - \frac{(p-2)^2\sigma^4/n^2}{\|\mu\|^2 + p\sigma^2/n}.$$

   [*Hint: Jensen's inequality* $\mathbb{E}\left[1/V\right] \geq 1/\mathbb{E}V$ *for non-negative valued random variable* V].

   (d) Assume that only a small number $k \ll p$ of μ_i is non-zero and that they are bounded, i.e. $|\mu_i| \leq M$. Show that the risk $R(\mu, \tilde{\mu})$ remains bounded as $p \to \infty$ (while keeping n, k, M and σ^2 fixed).

3. Consider the linear regression model

$$Y_i = \beta^\top x_i + \epsilon_i, \quad \beta \in \mathbb{R}^p, \quad \epsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, 1)$$

   with $x_i = [x_{i,1}, \cdots, x_{i,p}]^\top$, $i = 1, \ldots, n$.

   (a) Write the model as $Y = X\beta + \epsilon$ where $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$ and $\epsilon \in \mathbb{R}^n$. Find the maximum likelihood estimator $\hat{\beta}$ of β and compute its bias and variance.

   (b) Assume that $X^\top X = nI_p$ where I_p is the p-dimensional identity matrix. Show that $\hat{\beta}$ is not admissible as soon as $p > 2$ and find an estimator which strictly dominates $\hat{\beta}$.

   [*Hint: Re-express likelihood in terms of* $Z = X^T Y$ *and use Question* 2].

4. Consider a hierarchical model

$$\{X_{ij}\}_{j=1}^{n_i} |\theta_i \sim \text{Bern}(\theta_i), \ \{\theta_i\}_{i=1}^{k} |\alpha, \beta \sim \text{Beta}(\alpha, \beta).$$

Write $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ and denote the observed sample mean and sample variance of $\left\{\bar{X}_i\right\}_{i=1}^{k}$ by m and v, respectively.

(a) Find $\hat{\alpha}$ and $\hat{\beta}$ using empirical Bayes, i.e. by matching the mean and variance of Beta(α, β) to m and v.

(b) Show that the posterior mean $\mathbb{E}\left[\theta_i | \{X_{ij}\}, \hat{\alpha}, \hat{\beta}\right]$ is a weighted average of m and $\bar{X}_i$.

5. Let θ be a real-valued parameter and $f(x \mid \theta)$ be the probability density function of an observation x, given θ. Let H_0 be the hypothesis that $\theta = \theta_0$ and H_1 be the hypothesis that $\theta \neq \theta_0$ and consider a prior on H_1 given by $\theta \sim g(\theta)$. The prior probability for H_0 is β.

(a) Write down expressions for the (joint) distribution $P(H, \theta)$ (observe that H can take only two values), the marginal distribution $P(x)$ and $P(H_0, \theta|x)$ and $P(H_1, \theta|x)$.

(b) Derive an expression for $\pi(H_1 \mid x)$, the posterior probability of H_1.

Now, suppose that $x_1, \ldots, x_n$ is a sample from a normal distribution with mean θ and variance v. Let $\beta = 1/2$ and let

$$g(\theta) = (2\pi w^2)^{-1/2} \exp\left\{-\theta^2/(2w^2)\right\}$$

for $-\infty < \theta < \infty$.

Show that, if $\theta_0 = 0$ and the sample mean is observed to be $10(v/n)^{1/2}$ then

(c) The likelihood ratio frequentist test of size $\alpha = 0.05$ will reject H_0 for any value of n;

(d) The posterior probability of H_0 converges to 1, as $n \to \infty$.

(e) Comment on the apparent contradiction between (c) and (d).