1. ***Statistical decision theory.*** A decision-theoretic approach to the estimation of an unknown parameter $\theta$ introduces a loss function $\mathrm{L}(\widehat{\theta} - \theta)$ which, loosely speaking, quantifies the cost of deciding that the parameter has the value $\widehat{\theta}$ when it is in fact equal to $\theta$. Denote by $\boldsymbol{X}$ the vector of observations. An estimate

$$\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$$

may be chosen to minimize the *posterior expected loss*:

$$\rho(\widehat{\theta} \,|\, \boldsymbol{x}) \;\stackrel{\triangle}{=}\; \mathrm{E}_{\Theta \,|\, \boldsymbol{X}}[\mathrm{L}(\widehat{\theta} - \Theta) \,|\, \boldsymbol{x}] = \int \mathrm{L}(\widehat{\theta} - \theta)\, f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta$$

see handout # 4. This optimal choice of $\widehat{\theta}$ is called a *Bayes estimate* for the loss function $\mathrm{L}(\widehat{\theta} - \theta)$. Show the following:

**(a)** If

$$\mathrm{L}(\widehat{\theta} - \theta) = |\widehat{\theta} - \theta|$$

where $|\cdot|$ denotes absolute value, then any posterior median of $\theta$ is a Bayes estimate of $\theta$.

**(b)** If $k_0$ and $k_1$ are nonnegative numbers, not both zero, and

$$\mathrm{L}(\widehat{\theta} - \theta) = \begin{cases} k_0 \cdot (\theta - \widehat{\theta}), & \text{if } \theta \geq \widehat{\theta} \\ k_1 \cdot (\widehat{\theta} - \theta), & \text{if } \theta < \widehat{\theta} \end{cases}$$

then any $k_0/(k_0 + k_1)$ quantile of the posterior distribution $f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})$ is a Bayes estimate of $\theta$.

Recall that we have shown in handout # 4 that the posterior mean $\mathrm{E}_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{X} = \boldsymbol{x})$, if it exists, is the unique Bayes estimate of $\theta$ for the popular squared-error loss $\mathrm{L}(\widehat{\theta} - \theta) = (\widehat{\theta} - \theta)^2$.

*Comment:* For simplicity, assume that the posterior pdf is continuous and all its moments (that you need) exist.

2. ***Bayesian estimators are biased.*** Prove that the posterior mean, based on a proper prior distribution, cannot be an unbiased estimator, except in degenerate problems.

*Comments.* We need to show that, for proper prior distributions,

$$\widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{x}) = \mathrm{E}_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})$$

is a *biased* estimator of $\theta$. An estimator $\widehat{\theta}(\boldsymbol{X})$ is an *unbiased* estimator of $\theta$ if

$$\mathrm{E}_{\boldsymbol{X} \mid \Theta}[\widehat{\theta}(\boldsymbol{X}) \mid \theta] = \theta.$$

*Hints.* Use proof by contradiction, where you start by assuming that

$$\mathrm{E}_{\boldsymbol{X} \mid \Theta}[\widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{X}) \mid \theta] = \theta$$

holds (i.e. that $\widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{X})$ is an *unbiased* estimator of $\theta$) and ultimately reach a contradiction by evaluating

$$\mathrm{BMSE}\{\widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{X})\} = \mathrm{E}_{\boldsymbol{X},\Theta}([\widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{X}) - \Theta]^2).$$

To do that, start by computing

$$\mathrm{E}_{\boldsymbol{X},\Theta}[\Theta\, \widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{X})]$$

via iterated expectations. There are two ways to apply iterated expectations; use both.

3. Consider $X[n]$, $n = 0, 1, \ldots, N-1$ modeled as conditionally independent, identically distributed (i.i.d.) random variables given a parameter $\lambda$, following the Poisson distribution:

$$p_{X \mid \Lambda}(x[n] \mid \lambda) = \frac{\lambda^{x[n]}}{x[n]!} \exp(-\lambda), \quad x[n] = 0, 1, \ldots$$

for $n = 0, 1, \ldots, N-1$ and $\lambda > 0$. We choose the exponential prior pdf for $\lambda$:

$$\pi(\lambda) = b \, \exp(-b\,\lambda) \cdot i_{[0,+\infty)}(\lambda)$$

where $b > 0$ is a *known* constant.

**(a)** Find the posterior pdf

$$f_{\Lambda \,|\, \boldsymbol{x}}(\lambda \,|\, \boldsymbol{x})$$

where $\boldsymbol{x} = [x[0], x[1], \ldots, x[N-1]]^T$. Follow the approach that we demonstrated in class, i.e. focus on the terms in the posterior pdf that depend on $\lambda$, look up the table of distributions and see what it "looks like."

**(b)** Compute the posterior mean and posterior mode, which are the Bayesian MMSE and MAP estimates of $\lambda$, respectively.

*Hint:* Your distribution table (posted on WEBCT) gives expressions for the means and modes of the listed distributions.

**(c)** Verify the obtained MAP expression by computing

$$\widehat{\lambda}_{\text{MAP}} = \arg\max_{\lambda}[\ln p_{\boldsymbol{x} \,|\, \Lambda}(\boldsymbol{x} \,|\, \lambda) + \ln \pi(\lambda)]$$

see (30) in handout # 4.

**(d)** The number of packets per unit time arriving at a node in a communication network is a Poisson random variable $X$ with rate

$$\Lambda \sim \text{Expon}(b).$$

Find the MMSE estimate of the rate $\lambda$ given the observation $X = x$. Your answer should be in terms only of $x$ and the constant $\lambda$.

4. ***Sequential-Bayesian estimation.*** Derive (14) and (15) in handout # 4 by starting with a $\mathcal{N}(\mu_0, \tau_0^2)$ prior distribution and adding data points one at a time, using the posterior pdf at each step as the prior pdf for the next.

5. ***Bayesian vs. classical DC-level estimation.*** Suppose that a measurement $x$ is recorded and modeled using the standard DC-level in Gaussian noise model:

$$f_{X\,|\,\Theta}(x\,|\,\theta) = \mathcal{N}(x\,|\,\theta, \sigma^2)$$

where the variance $\sigma^2$ is assumed to be known exactly and the unknown DC-level parameter $\theta$ is known to lie in the interval $[0, 1]$, i.e. $[0, 1]$ defines the parameter space. Consider two DC-level estimates:

(1) the ML estimate:

$$\widehat{\theta} = \widehat{\theta}(x) = \arg \max_{\theta \in [0,1]} f_{X\,|\,\Theta}(x\,|\,\theta)$$

and

(2) the posterior mean (MMSE estimate) based on the uniform $\mathrm{U}(0, 1)$ prior pdf for $\theta$.

Show that, if $\sigma^2$ is large enough, the estimate (1) has a higher (classical) mean-square error (MSE)

$$\mathrm{E}_{X\,|\,\Theta}[(\widehat{\theta} - \theta)^2\,|\,\theta]$$

than the estimate (2) for any value of $\theta \in [0, 1]$.

6. ***Quantization and MMSE estimation.*** Consider a real-valued random variable $U$ with pdf $f_U(u) = 1$ for $u \in (0, 1)$, i.e.

$$f_U(u) = \mathrm{U}(u\,|\,0, 1).$$

**(a)** Describe a simple function $q : (0, 1) \to \{0, 1, \ldots, K - 1\}$ so that the random variable $X = q(U)$ is discrete with probability mass function (pmf)

$$f_X(k) = \frac{1}{K}, \quad \text{for } k = 0, 1, \ldots, K - 1.$$

You have produced a uniform discrete random variable from a uniform continuous random variable.

**(b)** What is the minimum mean-square error (MMSE) estimator of $U$ given $X = k$? Call this estimator $\widehat{U}(k)$. Write an expression for the resulting average MSE:

$$\mathrm{E}_U\{[U - \widehat{U}(q(U))]^2\}.$$

Argue that the estimator $\widehat{U}(q(U))$ minimizes the average MSE $\mathrm{E}_U\{[U-\widehat{U}(q(U))]^2\}$ between the original input and the final output (for a fixed $q$), which is one of the key properties of a Lloyd-Max quantizer.

**(c)** Find the pmf for the random variable $\widehat{U} = \widehat{U}(q(U))$. Find $\mathrm{E}_X[\widehat{U}]$ and $\mathrm{var}_X[\widehat{U}]$. How do the mean and variance of $\widehat{U}$ compare with those of $U$? (equal, bigger, smaller?)