A quick review:

We are considering taking possible actions $a \in \mathcal{A}$, and unknown quantities affecting our decision are represented by $\theta \in \Theta$.

In the estimation context, the action is just an estimate of θ, $\hat{\theta}(x)$.

The loss function describes the consequences of taking action a when the true state of nature is θ. We write it $L(\theta, a)$ or $L(\theta, \hat{\theta}(x))$.

Ultimately, we want to *choose* an action a or an estimate $\hat{\theta}(x)$. Our choice is driven by looking at a particular *risk function.*

So far we have seen just one strategy, the Bayes rule, which chooses $\hat{\theta}(x)$ to minimize the Bayes risk.

Recall the different risk functions:

1. Posterior risk (depends on x and the form of $\hat{\theta}$)

$$r(\hat{\theta}|x) = \int L(\theta, \hat{\theta}(x)) f(\theta|x) d\theta$$

2. Frequentist risk (depends on θ and the form of $\hat{\theta}$)

$$R(\theta, \hat{\theta}) = \int L(\theta, \hat{\theta}(x)) f(x|\theta) dx$$

3. Bayes risk (depends on the form of $\hat{\theta}$)

$$r(f, \hat{\theta}) = \int \int L(\theta, \hat{\theta}(x)) f(x, \theta) dx\, d\theta$$

Completely equivalently, we could write

1. Posterior risk: $r(\hat{\theta}|x) = E_{\theta|X}[L(\theta, \hat{\theta}(x))]$

2. Frequentist risk: $R(\theta, \hat{\theta}) = E_{X|\theta}[L(\theta, \hat{\theta}(X))]$

3. Bayes risk: $r(f, \hat{\theta}) = E_{\theta,X}[L(\theta, \hat{\theta}(X))]$

By iterated expectation, we also have that

$$r(f, \hat{\theta}) = E_{\theta}[E_{X|\theta}[L(\theta, \hat{\theta}(X))]] = E_{\theta}[R(\theta, \hat{\theta})]$$

and

$$r(f, \hat{\theta}) = E_X[E_{\theta|X}[L(\theta, \hat{\theta}(X))]] = E_X[r(\hat{\theta}|X)]$$

Example: Suppose $X_1, \ldots, X_n | \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$, where θ is known. Let the prior distribution for σ^2 be inverse gamma with parameters a and b. The prior PDF is

$$f(\sigma^2; a, b) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp\{-b/\sigma^2\}$$

- Find the posterior distribution for σ^2.
- What is the Bayes estimator under squared error loss?
- What is the Bayes estimator under absolute error loss?
- What is the Bayes estimator under zero-one loss?

You may use the fact the mean of an $InverseGamma(a, b)$ distribution is $b/(a-1)$ when $a > 1$, the mode is $b/(a+1)$, and the median is not available in closed form.

One final note about Bayes rules: under weak conditions, they are admissible. The intuition for this is that if there existed a rule that had lower risk, it would also have lower Bayes risk.

Here is one set of conditions:

Suppose that $\Theta \subseteq \mathbb{R}$ and that $R(\theta, \hat{\theta})$ is a continuous function of θ for every $\hat{\theta}$. Let f be a prior density that assigns positive probability to any open subset of Θ. Let $\hat{\theta}^f$ be a Bayes rule, with finite Bayes risk. Then $\hat{\theta}^f$ is admissible.

We'll now consider a different strategy for choosing an action, called a minimax rule. To motivate this, consider the following example.

An investor is deciding whether or not to purchase \$1000 of risky ZZZ bonds. If the investor buys the bonds, they can be redeemed at maturity for a net gain of \$500. There could, however, be a default on the bonds, in which case the original \$1000 investment would be lost. If the investor doesn't buy the bonds, she will put her money in a "safe" investment, for which she will be guaranteed a net gain of \$300 over the same time period. She estimates the probability of a default to be 0.1.

- Describe the parameter space Θ and the space of possible actions $\mathcal{A}$.
- What is the prior distribution?
- For each possible $\theta \in \Theta$ and $a \in \mathcal{A}$, compute the loss.
- Is any action inadmissible?

In the previous example, the Bayes rule is for the investor to buy the bonds, since this minimizes her expected loss (maximizes her expected gain) relative to the prior distribution for a default occurring.

However, suppose the investor is very conservative, and wants to choose a strategy to minimize the "worst case scenario." This is known as the minimax strategy – it minimizes the maximum loss that could occur.

Writing the frequentist risk of action a as $R(\theta, a)$, the maximum risk

$$\bar{R}(a) = \sup_{\theta} R(\theta, a)$$

Which action in the example minimizes $\bar{R}(a)$?

In the estimation context, our possible actions are estimators $\hat{\theta}$. Then the maximum risk is

$$\bar{R}(\hat{\theta}) = \sup_{\theta} R(\theta, \hat{\theta})$$

A decision rule that minimizes the maximum frequentist risk is called a minimax rule. The estimator $\hat{\theta}$ is a minimax rule (under a particular loss function) if

$$\sup_{\theta} R(\theta, \hat{\theta}) = \inf_{\tilde{\theta}} \sup_{\theta} R(\theta, \tilde{\theta})$$

Example (continued): Suppose $X \sim N(\theta, 1)$ and we are estimating θ under squared error loss. Consider $\hat{\theta}_c(x) = cx$.

- Calculate $\sup_{\theta} R(\theta, \hat{\theta}_c)$.

- Use this to determine the minimax estimator of θ.

In general it can be difficult to find minimax rules. One connection to Bayes rules you should be aware of is the following:

Suppose that $\hat{\theta}$ is the Bayes rule with respect to some prior f. Suppose further that $\hat{\theta}$ has constant risk: $R(\theta, \hat{\theta}) = c$ for some c. Then $\hat{\theta}$ is minimax.

Example: Suppose $X|p \sim Bin(n, p)$ and the loss is squared error.

- Show $\hat{p} = X/n$ is not minimax. *Hint: Consider the randomized estimator*

$$\tilde{p} = \left\{ \begin{array}{ll} X/n & \text{with probability } 1 - \frac{1}{n+1} \\ 1/2 & \text{with probability } \frac{1}{n+1} \end{array} \right\}$$

- Consider the Bayes estimator when $p \sim Beta(a, b)$. Find a and b so that the Bayes estimator has constant frequentist risk. This estimator is then minimax.