Example: Quadratic loss, \( L(θ, d) = (θ - d)^2 \)

\[
p(\theta, d) = E \left[ L(θ, d) | θ \sim π(θ) \right] = \int_0^1 L(θ, d) π(θ) dθ
\]

\[
= E_{π(θ)} [(θ - d)^2]
\]

\[
= E_{π(θ)} (θ^2) - 2dE_{π(θ)} (θ) + d^2.
\]

To find the Bayes rule,

\[
\frac{d}{dd} p(\theta, d) = -2E_{π(θ)} (θ) + 2d = d^* E_{π(θ)} (θ)
\]

Bayes rule

The corresponding Bayes risk,

\[
p^*(θ, d) = p\left(θ, d^*\right) = E_{π(θ)} (θ^2) - 2E_{π(θ)} (θ)E_{π(θ)} (θ) + E^2_{π(θ)} (θ)
\]

\[
= E_{π(θ)} (θ^2) - E^2_{π(θ)} (θ)
\]

\[
= V_{π(θ)} (θ).
\]

The Bayes rule of \( \{ θ, p = θ, π(θ), L(θ, d) = (θ - d)^2 \} \) is \( E_{π(θ)} (θ) \)

and the Bayes risk is \( V_{π(θ)} (θ) \).

- Bayes rule and risk of immediate decision
  \[ θ = \hat{θ}, f(θ) = E(θ), p^* (f(θ)) = V_{π(θ)} (θ) \]

- Bayes rule and risk after observing \( x = (x_1, ..., x_n) \)
  \[ \theta = (θ), f(θ) = E(θ|X), p^* (f(θ|x)) = V_{π(θ)} (θ|X) \]

Often we may be interested in the risk of the sampling procedure before observing the sample to decide whether or not to sample.
WE NEED TO SPECIFY FOR EACH POSSIBLE SAMPLE WHICH DECISION TO MAKE. WE HAVE A DECISION FUNCTION

\[ \delta : X \rightarrow \mathcal{D} \]

LET \( \mathcal{D} \) BE THE COLLECTION OF ALL DECISION FUNCTIONS, SO \( \delta \in \mathcal{D} \) IMPLIES \( \delta(x) \) IS IN THE DECISION SET \( \mathcal{D} \) FOR \( x \in X \).

THE RISK OF DECISION FUNCTION \( \delta \) IS:

\[ \rho(f(\theta), \delta) = \int_{x} \int_{\theta} L(\theta, \delta(x)) f(\theta, x) \, d\theta \, dx \]

Both \( \theta \) and \( x \) are random

\[ = \int_{x} \left\{ \int_{\theta} L(\theta, \delta(x)) f(\theta, x) \, d\theta \right\} f(x) \, dx \]

\[ = \int_{x} \mathbb{E}[L(\theta, \delta(x)) \mid X] f(x) \, dx \]

WE WANT TO FIND THE BAYES DECISION FUNCTION \( \delta^* \) FOR WHICH

\[ \rho(f(\theta), \delta^*) = \inf_{\delta \in \mathcal{D}} \rho(f(\theta), \delta) \]

MINIMISING \( \rho(f(\theta), \delta) \) IS EQUIVALENT TO CHOOSING THE \( \delta^*(x) \) WHICH MINIMISES THE POSTERIOR RISK \( \mathbb{E}[L(\theta, \delta^*(x)) \mid X] \)

[as \( f(x) = 0 \)]
The corresponding risk of the sampling procedure is

\[ \rho^* = \mathbb{E} \left[ \mathbb{E}(L(\theta, \hat{\theta}^*(\mathbf{x})) \mid \mathbf{x}) \right] \]

Typically considered as the risk \( \rho(f(\theta \mid \mathbf{x})) \) viewed as a random function of \( \mathbf{x} \).

Example: Quadratic loss.

\[ \hat{\theta}^*(\mathbf{x}) = \mathbb{E}(\theta \mid \mathbf{x}) \text{ for random } \mathbf{x} \]

\[ \rho^* = \mathbb{E} \left[ \mathbb{V} \rho(\theta \mid \mathbf{x}) \right] \]

Example:

Consider finding a point estimate \( \hat{\theta} \) for \( \theta \) using the loss function

\[ L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 \]

Where \( \mathbf{x} \sim \mathcal{N}(\theta, \sigma^2) \) and \( \theta \sim \mathcal{N}(\alpha, \beta) \).

The loss function is an example of generalized quadratic loss,

\[ L(\theta, \hat{\theta}) = g(\theta) (\theta - \hat{\theta})^2 \text{ where } g(\theta) = \sigma^2. \]

Note:

If \( \lambda_1 = \theta - \alpha \) and \( \lambda_2 = \theta - \alpha \) then:

\[ L(\theta, \hat{\theta}) = g(\theta) (\theta - (\theta + \epsilon))^2 = \epsilon^2 = L(\theta, \hat{\theta}) \]

The loss is symmetric (about \( \theta \)).

If \( \theta > 1 \) loss is increased compared to quadratic loss.

If \( 0 < \theta < 1 \) loss is decreased compared to quadratic loss.