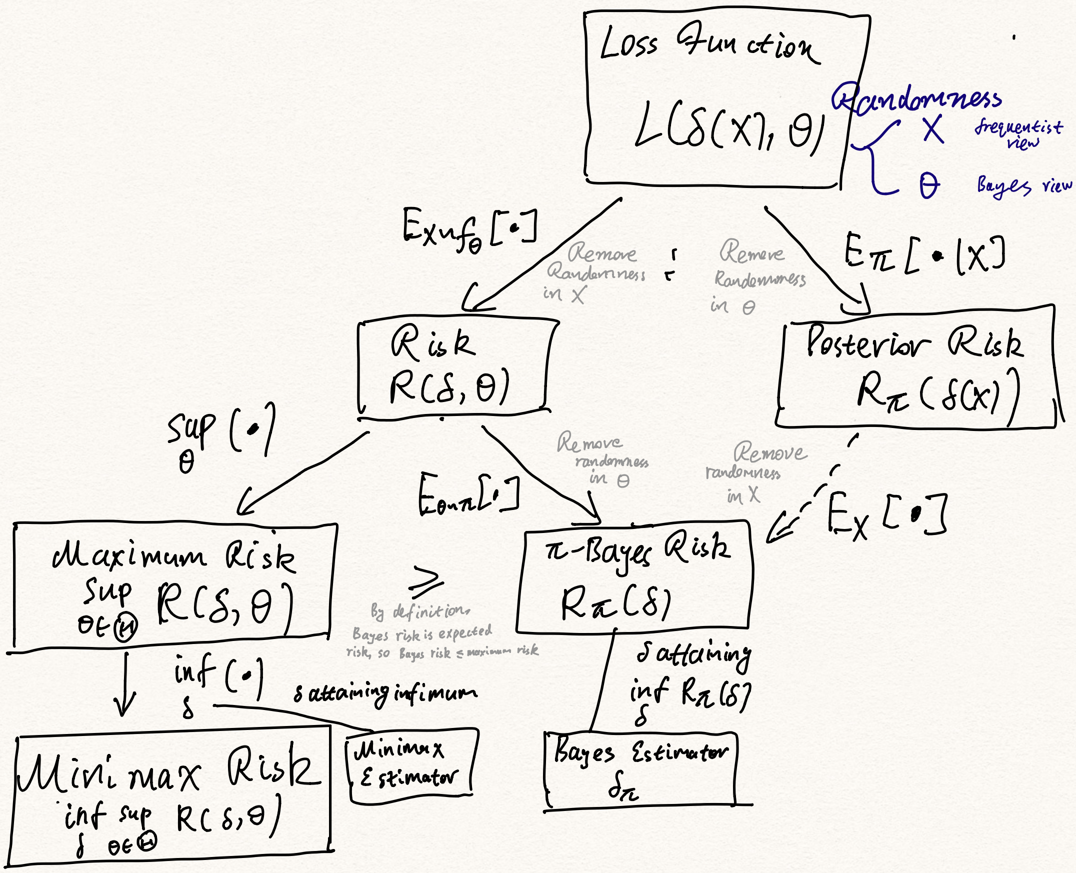
Notations

$\mathcal{X}$ - sample space
$\mathcal{A}$ - action space
π - prior/posterior
Θ - parameter space
δ - decision rule $\delta: \mathcal{X} \to \mathcal{A}$
L - Loss function $L: \mathcal{A} \times \Theta \to [0,\infty)$

Loss function is a penalty function of the decision rule for deviating from true parameter θ. Risk functions aim to discover the properties of loss function buried under its randomness in θ, x.



- Risk: Average loss across all samples

- Posterior Risk: Average loss across all θ, weighted by prior

- π-Bayes Risk: Average loss across all samples and parameters (prior)

- Minimax Risk: the risk given by decision rule δ that minimises the worst-case risk (i.e. maximum risk)

Relationships between Risks

$$R_\pi(\delta) = \int_{\Theta \times \mathcal{X}} L(\delta(x),\theta)\, p(x,\theta)\, dx\, d\theta$$

$$= \int_\Theta \underbrace{\int_{\mathcal{X}} L(\delta(x),\theta)\, p(x|\theta)\, dx}_{= R(\delta,\theta)}\, \pi(\theta)\, d\theta \quad (1)$$

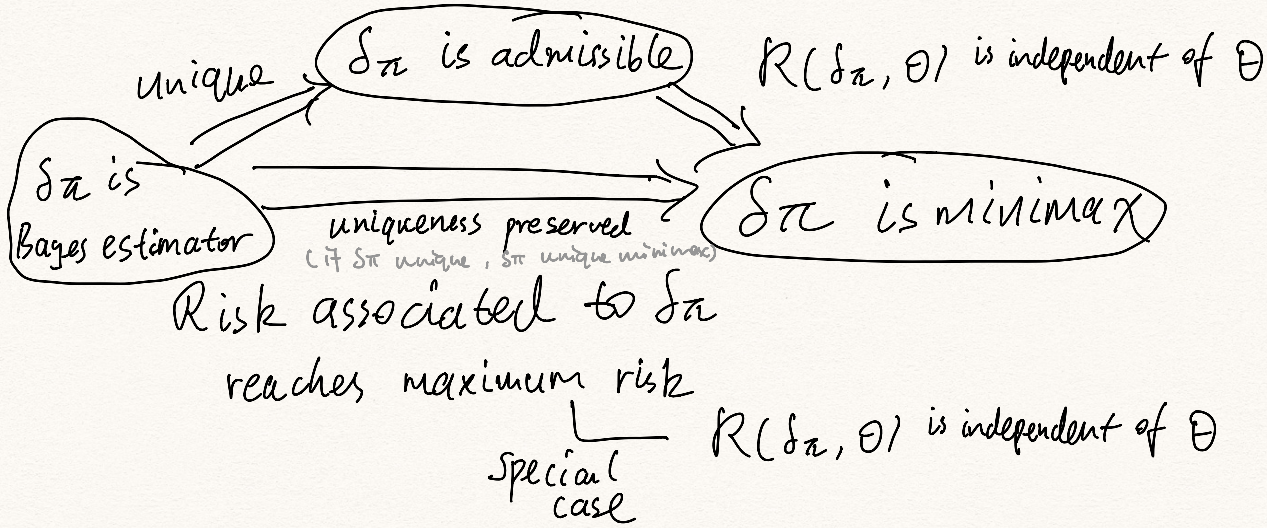
$p(x)$ is unknown so in practices (1) is used to find $R_\pi(\delta)$

$$= \int_{\mathcal{X}} \underbrace{\int_\Theta L(\delta(x),\theta)\, \pi(\theta|x)\, d\theta}_{R_\pi(\delta(x))}\, p(x)\, dx \quad (2)$$

note $p(x) = \int f_\theta(x)\pi(\theta)d\theta$ is called prior predictive function

(2) Explains trivially why π that minimises $\delta \mapsto R_\pi(\delta(x))$ $\forall x \in \mathcal{X}$ must also minimise $\delta \mapsto R_\pi(\delta)$

Relationship between estimators



Lehmann - Scheffe Theorem

This is a powerful guidance on finding UMRVE, it is an extension of the Rao-Blackwell's theorem.

Lemma If $\exists$ complete statistics $T=T(X)$, $g_1(T)$, $g_2(T)$ are unbiased estimators of $g(\theta)$, then $g_1 = g_2$ a.e. (i.e.

Proof. $E(g_1(T)-g_2(T)) = g(\theta)-g(\theta)=0$

Define $g(T) := g_1(T)-g_2(T)$, $E(g(T))=0$

by completeness of T, $P(g(T)=0) = P(g_1(T)=g_2(T)) = 1$

$\square$

Combining the lemma with Rao-Blackwell theorem yields Lehmann-Scheffe theorem:

Restrict on square loss function and unbiased estimators suppose sufficient, complete statistics T exists,

(1) if $\hat{g}(X)$ can be written as a function of T, and $\hat{g}(X)$ is unbiased estimator of $g(\theta)$, then $\hat{g}$ is unique UMVUE of $g(\theta)$

(2) If $\check{g}(X)$ is unbiased estimator of $g(\theta)$, then $\hat{g}(X) := E_\theta(\check{g}(X) \mid T)$ is unique UMVUE of $g(\theta)$

(3) If UMVUE of $g(\theta)$ exists, it must be a function of T

Proof. (2) By tower rule,

$$E_\theta(\hat{g}(X)) = E_\theta(E_\theta(\check{g}(X)|T)) = E_\theta(\check{g}(X)) = g(\theta)$$

so $\hat{g}$ is unbiased. Rao-blackwell yields

$$Var_\theta(\hat{g}(X)) \le Var_\theta(\check{g}(X)),$$

By lemma, $\hat{g}$ is unique(a.e) nomatter what $\check{g}(X)$ is

so $\hat{g}$ is UMVUE

(1) trivial by (2) and the lemma

(3) Suppose $\hat{g}(X)$ is UMVUE of $g(\theta)$,

Take $g^*(X) = E_\theta(\hat{g}(X) | T = T(X))$, by Rao-blackwell,

$$Var_\theta(\hat{g}(X)) \ge Var_\theta(g^*(X))$$

but by UMVUE property,

$$Var_\theta(\hat{g}(X)) \le Var_\theta(g^*(X))$$

so $Var_\theta(\hat{g}(X)) = Var_\theta(g^*(X))$

$\Rightarrow \hat{g} = g^*$ a.e. Note g^* is a function of $T = T(X)$

so $\hat{g}$ is also a function of T $\square$