Bayesian Estimators from Posterior Distr.

Posterior-EV: expectation for θ on average, given observed data

$$\hat{\theta}_{PE} = \mathbb{E}(\theta|x) \text{ with } \theta|x \overset{posterior}{\sim} \dots$$ Bayes-optimal decision for L2 Loss (#)

$Var(\theta|x) =$ **Posterior-Variance:** $\mathbb{E}[(\theta - \hat{\theta}_{PE})^2]$ & $sd(\theta|x) = \sqrt{\quad''\quad}$, via Posterior-Uncertainty

Scalar Bayesian Measure of Param. Uncertainty after data observation

- Very easy to obtain for conjugate Priors where we always get Posterior from known Standard Distribution
- Can be quite hard to calculate for non-standard Posteriors. May be a tough integral or infinite sum

(#) $r(d,p|x) = E[L_2(d(X),\theta)|x] = E[(d-\theta)^2|x]$
$= E[d^2 - 2d\theta + \theta^2|x] = d^2 - 2dE(\theta|x) + E(\theta^2|x)$
$d^* = \underset{d}{\arg\min}\, r(d,p|x): \frac{\partial}{\partial d} r(\,) = 2d - 2E(\theta|x) \overset{!}{=} 0 \Leftrightarrow d = E(\theta|x)$
Bayes-optimal Risk $r^*(p) \overset{r(d^*,p|x)}{=} E[L_2(d^*,\theta)|x] = E[(E(\theta|x)-\theta)^2|x]$
$= Var(\theta|x)$

Posterior-Mode / Max. A-posterior (MAP): MAP is the posterior value with highest density, so most probable θ given observed data

$$\hat{\theta}_{MAP} = \underset{\theta}{\arg\max}(p(\theta|x))$$ Bayes-optimal for 0-1 Loss ($\varepsilon \to 0$)

- Only explicitly known for few Standard Posteriors, so in general MAP needs to be numerically approximated / calculated
- If const. flat (Laplace-) Prior $p(\theta) \propto$ const. it holds $p(\theta|x) \propto L(\theta) \Rightarrow \hat{\theta}_{MAP} = \hat{\theta}_{MLE}$ with nice ML-properties also true for MAP

For other Priors: MAP $\hat{=}$ penalised MLE maximizing penalised Log-Likelihood $\ell(\theta) + \log(p(\theta))$

Sensitivity Analysis How robust /stable are Estimators?
e.g. Estimator Matrix for different Prior-Hyperparams
Point Estimator, HPD for varying Hyperparam-combinations

Posterior-Median: "middle / central" posterior value given observed data

$$\hat{\theta}_{med} = Median(\theta|x)$$ Bayes-optimal for L1 loss

- Only easy to obtain for know Posterior Distr. where can simply use Distr. Fct. bzw. its inverse (quantile function)
- robust vs outliers and invariant under monotone transf. of param. (both PE and MAP are not transf.-invariant)

× Robust comes in handy when we (as often in reality) have to approximate Posterior by randomly drawing from it!

Note that all three point estimators are (nearly) identical for (nearly) symmetrical unimodal posteriors

But which is the best one? That depends on how we define decision-theoretical Loss on $d = \hat{\theta}$

In Decision Theory we conduct inference directly for Decision. Decision can be value / choice of point estimator or Hypothesis after completed Test.
$\downarrow \mathcal{D} = \Theta$ $\qquad \downarrow \mathcal{D} = \{$All possible Hypotheses$\}$

Decision Function $d: \mathcal{X} \to \mathcal{D},\ x \mapsto d(x)$ maps from sample space to decision space

Win Function also valid approach
We evaluate (point est.) choice with a Loss function $L: \mathcal{D} \times \Theta \to \mathbb{R},\ (d,\theta) \mapsto L(d,\theta)$

L_p-Loss $L_p(d,\theta) = |d-\theta|^p$ — L1 Loss: $L_1(d,\theta) = |d-\theta|$ — L2 Loss: $L_2(d,\theta) = (d-\theta)^2$

0-1 loss $L_\varepsilon(d,\theta) = \begin{cases} 1 & \text{if } |d-\theta| > \varepsilon \\ 0 & \text{if } |d-\theta| \le \varepsilon \end{cases}$

- Loss fct. depends on decision d and unknown $\overset{true}{\text{param}}$ θ, not quite Bayesian actually
- For given Loss fct., before observing data x, $d(X)$ is random variable and therefore also $L(d(x),\theta)$

Risk $R(d,\theta) \overset{\text{Loss expectation on average}}{=} \mathbb{E}_X[L(d(X),\theta)] = \int_X L(d(x),\theta) \underbrace{\boxed{f(x|\theta)}}_{\text{Likelihood / Data density}} dx$

Rarely is there a $d(\,)$ with minimal risk across all possible θ, maybe though in some restr. param-interval
Admissible Decision Rule: Dec. Fct admissable if there is no $d(\,)$ with smaller risk for all θ

Bayes Risk Posterior expected Loss $r(d,p|x) = \mathbb{E}_\theta(\mathbb{E}_X[L(d(X),\theta)]) = \int_\theta \left(\int_X L(d(x),\theta) \cdot f(x|\theta)\,dx\right) \underbrace{p(\theta)\,d\theta}_{dP(\theta)}$

evaluates uncertainty of posterior $= \mathbb{E}[L(d(X),\theta)|x] = \int_\theta L(d(X),\theta)\,p(\theta|x)\,d\theta = \int_\theta L(d(x),\theta) f(x|\theta) p(\theta)\,d\theta / f(x)$

1. average over data X (for fixed θ)
2. Then average across Prior of θ for all θ

Bayes-optimal-decision d^* minimizes Bayes Risk for given Loss function L and prior distr. p; d^* is always an admissible decision
associated Bayes-optimal risk $r^*(p) = r(d^*, p|x)$ always $\geq$ Minmax risk that results from minmax dec. $d_{minmax} = \underset{d}{\arg\min}\left(\max_\theta R(d,\theta)\right)$ conservate, protects vs. worst case

---

Interval estimators $P(\theta \in I|x) \overset{!}{=} 1-\alpha$ $\qquad \int_I p(\theta|x)\,d\theta = 1-\alpha$

Note: Scalar $Var(\theta|x)$ often much easier to obtain
[Interval] Bayesian Measure for Uncertainty of Parameter after taking observed data into account

Credibility Interval: Much more intuitive than frequentist CI as it denotes actual probability that "true" parameter lies in the interval

There are infinite intervals that satisfy above equation
HPD unimodal: $p(\theta_L|x) = p(\theta_U|x)$ no max at edges

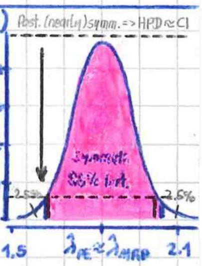
↳ **Highest Posterior Density Interval [HPD]:** $p(\theta|x) \geq p(\bar{\theta}|x)\ \forall \theta \in I, \bar{\theta} \notin I$

- Not so commonly used b.c. requires complete Posterior Distr. and numerical optimization, also not transf. invariant
- HPD always shortest (in terms of θ-axis) credibility intervall, => useful for localization procedure

(and invariant under strictly monotone transf.)
Easier is symmetrical α-cred. interv.
$I = [L,U]$ with $P(\theta < L) = P(\theta > U) = 1-\alpha/2$
Just cut upper / lower 2.5% for $\alpha =$, in R: qgamma(c(0.025, 0.975), α, β)

$p(\lambda|x)$ Post. (nearly) symm. => HPD ≈ CI

95% HPD
Interval, where upper / lower bound have same density.
For multimodal Post. can be split in several I



Future unknown x_z, $X_z|\theta \overset{\text{same } X}{\sim} \dots$ $\{X_z|X=x\}$

Bayesian Prediction Predictive Posterior Distr.: $f(x_z|\theta) = \frac{f(x_z,\theta|x)}{p(\theta|x)} \Leftrightarrow f(x_z,\theta|x) = f(x_z|\theta) \overset{posterior}{p(\theta|x)}$ joint density

=> integrate over joint dens. w.r.t. θ — predictive density $f(x_z|x) = \int f(x_z|\theta) p(\theta|x)\,d\theta$; Post. Predictive EV] easy shortcut $E(X_z|x) = \dots\dots\ E[X_z|\theta = E(\theta|x)]$ input param post. EV into Likelihood

Based on pred. distr. we can infer Var, $sd(X_z)$ and prediction intervals

For: Poisson(λ), Bin(n,π), $N(\,)$ simply = posterior EV $E_{\theta|x}(\theta)$ as $E(X_z|\theta) = c \cdot \theta$, linear in θ