→ Numeric optimization

→ Bayes

- Monte Carlo
- Conjugat priors
- type of Bayesian point estimation.

Bayesian Statistics

- Probability is a belief about something.
- Idea: (1) represent of prior belief about a pamater vector Θ, as prob. dist'n.

$\Pi_\Theta(\theta) \leftarrow$ prior density.

(2) calculate corresponding "likelihood function" $(x_1, x_2, \dots, x_n) \sim f_{X_{1:n}}(x_{1:n}; \theta)$, evaluated at observed data $x_1^*, x_2^*, \dots, x_n^*$:

$$\rightarrow L(\theta) = f_{X_{1:n}}(x_{1:n}^*; \theta).$$

We have Θ is now a random vector!

$$f_{X_{1:n}|\Theta}(x_{1:n}^* \mid \Theta = \theta) = \begin{cases} f(x|\theta) \\ \pi(x|\theta) \end{cases}$$

(3) update belief using Baye's theorem: "posterior dist'n"

$$\pi_{\theta|X}(\theta \mid X = x^*) = \frac{\overset{L(\theta)}{\pi(x|\theta)}\,\pi(\theta)}{\int \pi(x|\tau)\,\pi(\tau)\,d\tau}$$

$$\propto \underline{\pi(x|\theta)\,\pi(\theta)}$$

posterior is proportional to likelihood times prior.

Conjugate prior, how do we find them?

if our prior $\pi_\theta(\theta)$ is from some family $\underline{G}$ of $\underline{\text{distributions}}$ (ex: $\underline{\text{Gamma}(\alpha, \lambda)}$)

and likelihood family $\underline{H}$,

then we say G is conjugate family to H, $\underline{\underline{\text{if}}}$ $\pi(\theta|x)$ is still in family $\underline{G}$.

$\pi_\theta(\theta) \sim \text{Beta}(\alpha, \beta)$ (prior for p)

$\pi(x|\theta) \sim \text{Bin}(n, p)$

$\pi(\theta|x) \sim \text{Beta}(\alpha^*, \beta^*)$

??? $\text{Beta}(\alpha + X, n - X + \beta)$?

$$\pi(\theta|x) \propto \left(\pi(x|\theta)\, \pi(\theta)\right)$$

"kernel of $\pi(\theta|x)$"

$$\theta^{\text{something}} (1-\theta)^{\text{something}} +$$

Beta support θ on $(0,1)$

Beta family is conjugate for many likelihood families

- Binomial
- geometric
-

Gamma (α, λ) $\rightarrow$ conjugate prior to many distribn with "rate" params.

- exp
- Pois
- ...

Strategy:

compute $\pi(\theta|x) \propto \pi(x|\theta)\pi(\theta)$, match kernel to known distribution.

Numeric Integration

$$\int \pi(x|\theta)\,\pi(\theta)\,d\theta$$

Monte-Carlo

- Idea: Expectations are integrals

$$E[X] = \int x f(x)\,dx,$$

$f(x)$ is density of X.

- SLLN: $\frac{1}{N}\sum X_i \to E[X_i]$

$$= \int x f(x)\,dx.$$

$\int g(x)\,dx$:

(1) if bounds are finite:

$\int_a^b g(x)\,dx$.

$\rightarrow$ use uniform(a,b)!

$X_i \overset{iid}{\sim} U(a,b)$, $f(x) = \frac{1}{b-a}\big|^{(a\le x\le b)}$

$$\int_a^b g(x)\,dx = (a-b)\int_a^b \frac{g(x)}{b-a}\,dx$$

$$= (a-b)\, E[g(x)],$$

where $X_i \overset{iid}{\sim} U(a,b)$.

$$\approx (a-b)\cdot \frac{1}{m}\sum_{i=1}^{m} g(X_i)$$

$\xrightarrow{a.s.}$ Integral.

(2) bounds are ifinite

$\int_0^\infty g(x)\,dx$, sample from R.V., that has support $(0,\infty)$.

☆ <u>Importance Sampling</u>.

$\int_0^\infty g(x) \cdot \frac{f(x)}{f(x)}\,dx$, $f(x)$ is some density

$= \int_0^\infty \frac{g(x)}{f(x)} \cdot f(x)\,dx$

$$= E\left[\frac{g(X)}{f(X)}\right]$$

$$\approx \frac{1}{m}\sum_{i=1}^{m}\left(\frac{g(X_i)}{f(X_i)}\right), \quad X_i \sim f(x)$$

"importance weights": w_i

$g(x) \propto f(x)$

(3) $\int_a^b e^{-x}\,dx$ Sample $U(a,b)$

$\propto \underline{\exp(1)}$, $(0,\infty)$

$(\underline{a>0})$:

$$\int_a^b e^{-x}\,dx = \int_0^\infty e^{-x}\cdot \mathbb{1}[a\le x\le b]\,dx$$

~~Sam~~

$$= \int_0^\infty h(x)\, f(x)\, dx$$

$$= E[h(X)]$$

$$\approx \frac{1}{m} \sum h(X_i)$$

$$\approx \frac{1}{m} \sum 1(a \le X_i \le b)$$

---

best approach:

Sample "heavier" from "more important" regions.

Scaled Beta.

$\int g(x)\,dx$,

We want: $f(x) \propto g(x)$.

- $\pi(\theta|X)$ ← posterior.

What about "point estimate"

- Posterior Mean:

$$\hat{\theta}_{PM} = E[\theta|X = x^*]$$

- Posterior mode

Maximum a Posterior (MAP)

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \pi(\theta|X).$$

- **Posterior Median:**

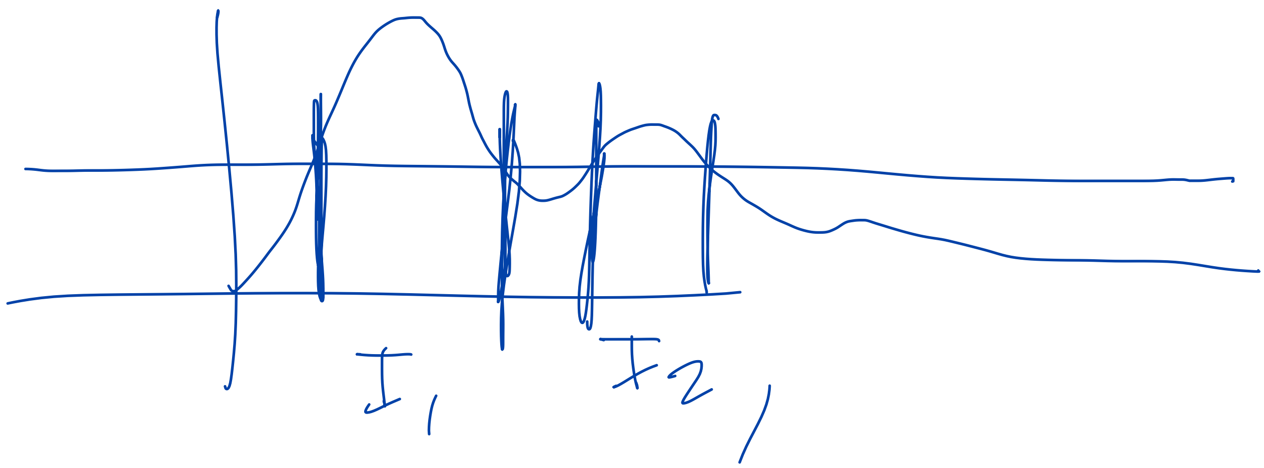
$$\hat{\theta}_{med} = \text{Med}(\theta \mid X = x^*)$$

Uncertainty.

$$CRI = P(\theta \in I_{95}) = 0.95$$

$\leftarrow \pi(\theta \mid x)$

High-density-region:



HDR: $I_1 \cup I_2$.