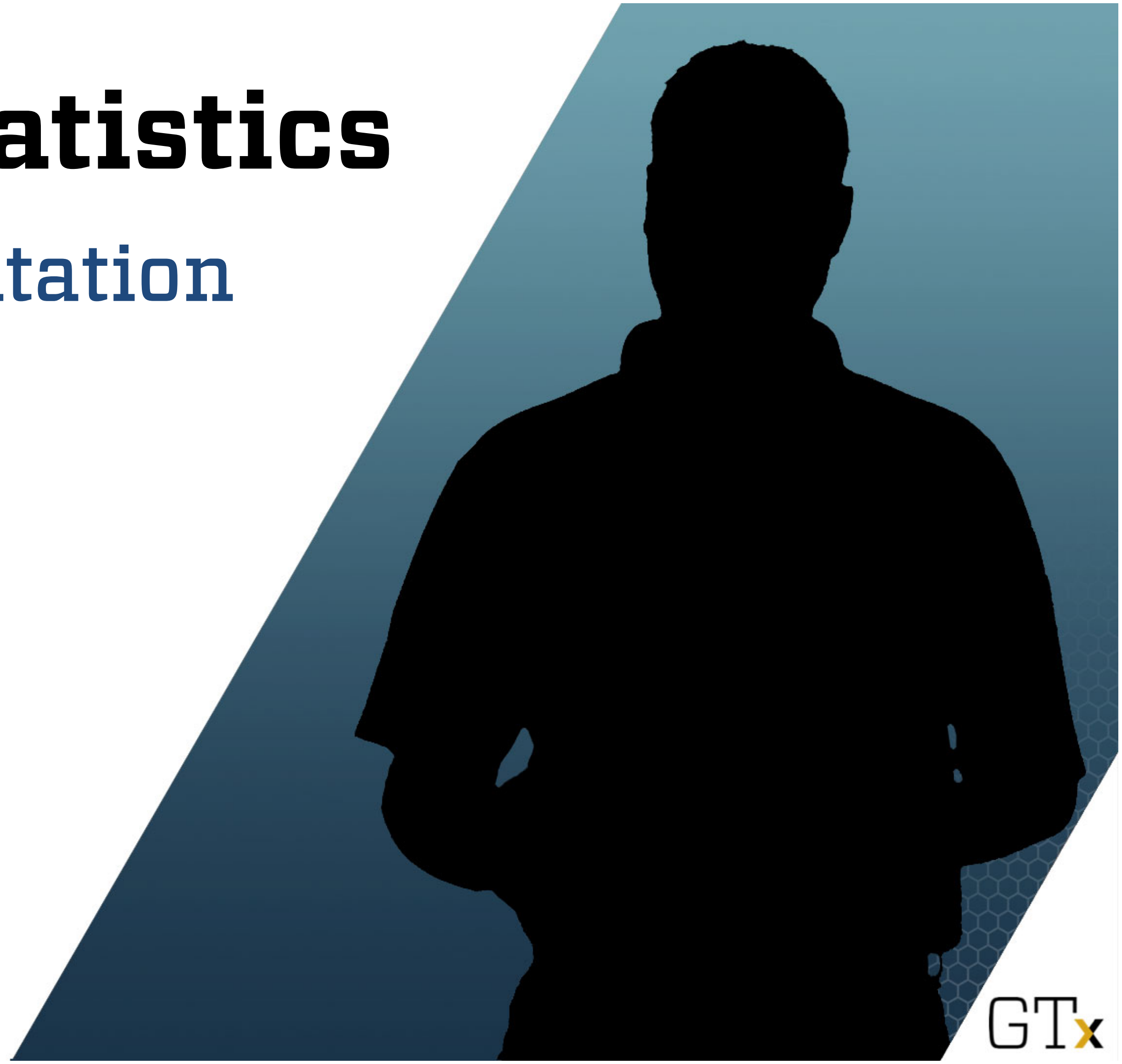
Bayesian Statistics

Bayesian Computation

Brani Vidakovic
Professor
School of Industrial and Systems Engineering



Before We Begin…

In this unit:

- Numerical Approaches
- Markov Chain Monte Carlo (MCMC)

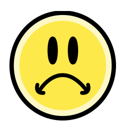
Numerical Approaches in Bayesian Computation

- Classical Statistics $\Leftarrow$ optimization
- Bayesian Statistics $\Leftarrow$ integration

Recall Bayes Theorem:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

  - Easy $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$, but $f(x|\theta)\pi(\theta)$ needs to be normalized to be a density. Normalizing constant is marginal $m(x) = \int_\Theta f(x|\theta)\pi(\theta)\, d\theta$



- Normalizing $f(x|\theta)\pi(\theta)$ was easy (no integration) for <u>conjugate cases</u>.

- For nonconjugate cases typically we numerically compute the posterior (either by numerical integration for $m(x)$, or by sampling)

- Example 1. Assume $x|\theta \sim \mathrm{N}(\theta, 1)$ and $\theta \sim \mathrm{Cau}(0,1)$.

  The likelihood is $f(x|\theta) \propto e^{-\frac{1}{2}(x-\theta)^2}$

  The prior is $\pi(\theta) \propto \frac{1}{1+\theta^2}$

For Normal/ Cauchy pair integral

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-\theta)^2} \frac{d\theta}{1+\theta^2}$$

is not solvable in terms of elementary functions.

What to do if we need to find Bayes estimator

$$\delta_B(x) = \frac{\int_\Theta \theta f(x|\theta)\pi(\theta)\, d\theta}{\int_\Theta f(x|\theta)\pi(\theta)\, d\theta}\, ?$$

Note, here:

$$\delta_B(x) = \frac{\int_{-\infty}^{+\infty} \frac{\theta}{1+\theta^2} e^{-\frac{1}{2}(x-\theta)^2}\, d\theta}{\int_{-\infty}^{+\infty} \frac{1}{1+\theta^2} e^{-\frac{1}{2}(x-\theta)^2}\, d\theta}$$

Since $e^{-\frac{1}{2}(x-\theta)^2} \equiv e^{-\frac{1}{2}(\theta-x)^2}$

$$\delta_B(x) = \frac{\int_{-\infty}^{+\infty} \frac{\theta}{1+\theta^2} e^{-\frac{1}{2}(\theta-x)^2}\, d\theta}{\int_{-\infty}^{+\infty} \frac{1}{1+\theta^2} e^{-\frac{1}{2}(\theta-x)^2}\, d\theta}$$

→ **Sample** $\theta_1, \theta_2, \dots, \theta_N$ **from** $\mathrm{N}(x, 1)$

$$\delta_B(x) \approx \frac{\sum_{i=1}^{N} \frac{\theta_i}{1+\theta_i^2}}{\sum_{i=1}^{N} \frac{1}{1+\theta_i^2}}$$

→ **Sample** $\theta_1, \theta_2, \dots, \theta_N$ **from** $\mathrm{Ca}(0,1)$

$$\delta_B(x) \approx \frac{\sum_{i=1}^{N} \theta_i e^{-\frac{1}{2}(\theta_i - x)}}{\sum_{i=1}^{N} e^{-\frac{1}{2}(\theta_i - x)}}$$

For example, *norcau.m* calculates $\delta_B(2)$.

(Octave/ MATLAB)
norcau.m

Check with WinBUGS

$x|\theta \sim \mathrm{N}(\theta, 1)$; $\theta \sim \mathrm{Ca}(0,1)$ by numerical integration:

$x = 2$, $\delta_B(2) = 1.2821951027$ (*norcau.py*)

Laplace's Method

- Derives inspiration from the work of Laplace, 1774.
- $g(\theta) = f(x|\theta)\pi(\theta)$

approximation of un-normalized posterior by a normal distribution

- $g(\theta)$ unimodal, not too skewed
- $\hat{\theta}$ the mode of $g(\theta)$
- θ possibly multivariate
- Why not moment matching? Need mean + variance and they depend on $m(x)$.

$$\log g(\theta) \cong \log g\left(\hat{\theta}\right) - \frac{1}{2}\left(\theta - \hat{\theta}\right)' Q\left(\theta - \hat{\theta}\right)$$

$$Q_{ij} = \left[-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log g(\theta)\right]_{\theta=\hat{\theta}}$$

GTx

For univariate parameter θ

$$Q = \left[\frac{\partial^2}{\partial \theta^2} \log g(\theta)\right]_{\theta=\hat{\theta}}$$

- $\theta | x \sim \mathrm{N}(\hat{\theta}, Q^{-1})$
- $m(x) \propto \int_\Theta g(\theta) d\theta \cong \frac{g(\hat{\theta})}{\sqrt{\det(Q/2\pi)}}$

➢ Exponentiating and integrating

$$\log g(\theta) \cong \log g(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})' Q (\theta - \hat{\theta})$$

one gets

$$\int_\Theta g(\theta) d\theta = g(\hat{\theta}) \int_\Theta e^{-\frac{1}{2}(\theta - \hat{\theta})' Q (\theta - \hat{\theta})} d\theta$$

$$= g(\hat{\theta}) \sqrt{2\pi \, \mathrm{det Q}^{-1}} = \frac{g(\hat{\theta})}{\sqrt{\det(Q/2\pi)}}$$

Example 2. $x|\theta \sim \mathrm{Ga}(r, \theta)$
$\theta \sim \mathrm{Ga}(\alpha, \beta)$.

Find Laplace's approximation to the posterior, and compare it with exact posterior (the model is conjugate).

- Exact posterior is proportional to

$$f(x|\theta)\pi(\theta) = \frac{x^{r-1}\theta^{r}}{\Gamma(r)} \times e^{-\theta x} \times \frac{\theta^{\alpha-1}\beta^{\alpha}}{\Gamma(\alpha)} \times e^{-\beta\theta}$$
$$\propto \theta^{r+\alpha-1} e^{-(\beta+x)\theta}$$

- which is the kernel of $\mathrm{Ga}(\alpha + r, \beta + x)$ distribution
- $g(\theta) = \theta^{r+\alpha-1} e^{-(\beta+x)\theta}$

$$\log g(\theta) = (r + \alpha - 1)\log\theta - (\beta + x)\theta$$

$$\frac{\partial}{\partial\theta}\log g(\theta) = \frac{r+\alpha-1}{\theta} - (\beta + x)$$

- $\frac{\partial}{\partial\theta}(\log g(\theta)) = 0,\ \boxed{\hat{\theta} = \frac{r+\alpha-1}{\beta+x}}$

  $\frac{\partial^2}{\partial\theta^2}(\log g(\theta)) = -\frac{r+\alpha-1}{(\beta+x)^2} < 0$, $\hat{\theta}$ **maximum**

- $Q = \left(-\frac{\partial^2}{\partial\theta^2}(\log g(\theta))\right)_{\theta=\hat{\theta}} = \frac{r+\alpha-1}{\hat{\theta}^2} = \frac{(\beta+x)^2}{r+\alpha-1}$

- $Q^{-1} = \frac{r+\alpha-1}{(\beta+x)^2} \Rightarrow \theta|x \overset{\text{approx}}{\sim} \mathrm{N}\left(\frac{r+\alpha-1}{\beta+x}, \frac{r+\alpha-1}{(\beta+x)^2}\right)$

$$\int_{\Theta} g(\theta)\, d\theta = \frac{g(\hat{\theta})}{\sqrt{\det(Q/2\pi)}}$$

- $\int_0^{\infty} \theta^{r+\alpha-1} e^{-(\beta+x)\theta}\, d\theta \approx \sqrt{2\pi}\,\frac{(r+\alpha-1)^{r+\alpha-\frac{1}{2}}}{(\beta+x)^{r+\alpha}}\, e^{-(r+\alpha-1)}$ **check!**

GTx

- Consult the m-file *laplace.m*

$r = 20, \quad \alpha = 5, \quad \beta = 1, \quad \text{and} \quad x = 2$

exact 95% cs , approximate 95% cs

$[5.3925, 11.9034]$ $\underbrace{[4.7994, 11.2006]}_{\text{credibility } 94.04\% \text{ (exact)}}$

$$\int g(\theta)\, d\theta = \begin{cases} 7.3228\times 10^{11}, & \text{exact} \\ 7.2974\times 10^{11}, & \text{approximate} \end{cases}$$

GTx

Summary