Probability Models: Models for Uncertainty (Parametric)

Based on 9.07 Lectures 2-3, 8

Discrete (Lecture 2)

- Bernoulli $B(1, p)$
- Binomial $B(n, p)$
- Poisson $P(\lambda)$

Continuous (Lecture 3)

- Exponential $E(\lambda)$
- Gamma $\Gamma(\alpha, \beta)$
- Beta $\beta eta(\alpha, \beta)$
- Gaussian $N(\mu, \sigma^2)$
- Inverse Gaussian $IG(\mu, \lambda)$

Properties To Know (Characterize Uncertainty)

1. Shape of the probability mass fct (density function)
2. Location: median, mean, mode
3. ~~Scale~~ ~~Variance~~ Spread: variance, standard deviation
4. Properties as the number of observations gets large
5. Applications
6. Domain of Definition

Question	Probability Model
Election	Binomial
Efficacy of a New Therapy	Binomial (better or not)
Noise in the fMRI scanner	Rayleigh
Background Noise in the MEG Scanner	Gaussian
Interspike Interval Distribution	Inverse Gaussian

Data Analysis Paradigm

$x_1, x_2, \ldots, x_n \qquad x_i \sim f(x|\theta) \qquad$ parametric model

If we knew θ, we could assess how well the data are described by the model

Estimation: A formal procedure that tells us how to compute a model/parameter from observed data

Estimator is a function that takes data into a value of the parameter $(x_1, \ldots, x_n) \longrightarrow \hat{\theta}$

Estimate is a specific value of the estimator

Example Gallup Poll of the Trump Approval Rating 2/12/17

$x_i \sim$ Bernoulli $\quad p \quad$ parameter (fraction that approve Trump)

1 approve $\qquad \theta = p$

0 disapprove

$Y = \sum_{i=1}^{n} x_i \sim$ Binomial $(n = 1{,}500, p)$ $\qquad$ Thinking Intuitively

= 600 approve $\qquad \hat{p} = \frac{600}{1500} = 0.4$

900 disapprove $\qquad 1 - \hat{p} = 0.6$

Example 3.2 MEG Background Noise Data

$x_1, \ldots, x_n \qquad x_i \sim N(\mu, \sigma^2) \qquad E(x_i) = \mu \qquad \theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$

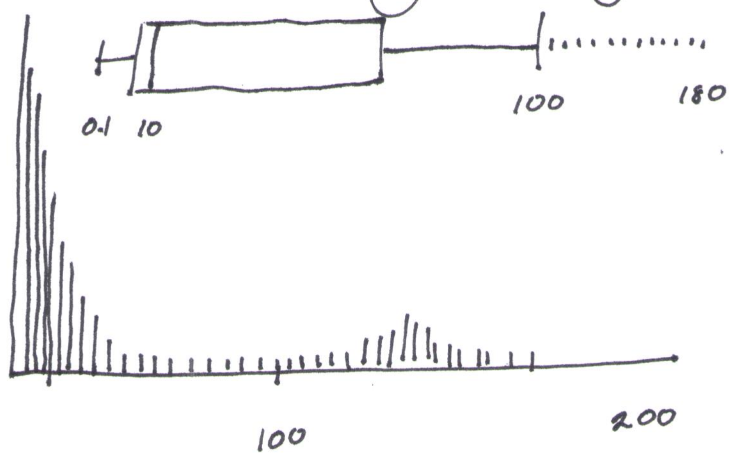
$Var(x_i) = \sigma^2$

Estimators:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \qquad \text{or } s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$f(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma^2}\right\}$$

Example 8.1 Spiking Activity of a Retinal Neuron



Box Plot

Histogram

Possible Models
Exponential
Gamma
Inverse Gaussian
Log Normal

Exponential

$f(x) = \lambda e^{-\lambda x}$ $\quad \lambda > 0 \quad x > 0 \quad \theta = \lambda$

$E(x) = \lambda^{-1} \quad Var(x) = \lambda^{-2} \quad$ median $\cancel{\text{mode}} = \frac{\log 2}{\lambda} \quad$ mode $= 0$

Gamma

$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x} \quad \alpha > 0 \quad \beta > 0 \quad x > 0 \quad \theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

$E(x) = \frac{\alpha}{\beta} \quad Var(x) = \frac{\alpha}{\beta^2} \quad$ mode $= \frac{\alpha - 1}{\beta}$

Inverse Gaussian

$f(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \lambda \frac{(x-\mu)^2}{\mu^2 x}\right\} \quad \theta = \begin{bmatrix} \mu \\ \lambda \end{bmatrix}$

$E(x) = \mu \quad Var(x) = \frac{\mu^3}{\lambda} \quad$ mode $= \mu \left[\left(1 + \frac{9}{4}\left(\frac{\mu}{\lambda}\right)^2\right)^{\frac{1}{2}} - \frac{3}{2}\frac{\mu}{\lambda}\right]$

$= \mu \left[\left(1 + \frac{9}{4}\phi^2\right)^{\frac{1}{2}} - \frac{3}{2}\phi\right]$

$\phi = \frac{\lambda}{\mu}$

Log Normal

$f(x) = \left(\frac{1}{2\pi\sigma^2 x}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \frac{(\log x - \mu)^2}{\sigma^2}\right\} \quad x > 0 \quad \sigma > 0 \quad \theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$

$-\infty < \mu < \infty$

$E(x) = e^{\mu + \sigma^2/2} \quad E(x^2) = e^{2\mu + 2\sigma^2}$

$Var(x) = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2} \quad$ Mode $= e^{\mu - \sigma^2}$

median $= e^{\mu}$

Theoretical Moments

ith moment (non-central moments) about 0

$$\mu_i = \sum_{j=-\infty}^{\infty} x_j^i \, p(x_j) \qquad \mu_i = \int_{-\infty}^{\infty} x^i f(x)\,dx$$

ith central moments

$$\tilde{\mu}_i = \sum_{j=-\infty}^{\infty} (x_j-\mu_1)^i \, p(x_j) \qquad \tilde{\mu}_i = \int_{-\infty}^{\infty} (x-\mu_1)^i f(x)\,dx$$

We have

$$\mu_1 = \mu$$

$$\sigma^2 = \mu_2 - \mu_1^2$$

Cauchy distribution has no moments

$$f(x) = \frac{1}{2\pi(1+x^2)}$$

$$E(X) = \infty \quad \Rightarrow \quad E(X^r) = \infty \quad r > 1$$

ith Sample Moments

$$\hat{\mu}_i = n^{-1} \sum_{j=1}^{n} x_j^i \qquad i = 1, 2, 3, \ldots,$$

The sample moments are always finite

Def'n Method of Moments Estimation

$x_1, x_2, \ldots, x_n$ a sample from $f(x_i \mid \theta)$ a probability model with a d-dimensional unknown parameter θ

Assume that the 1st d moments $\mu_j(\theta)$ are finite for $j = 1, \ldots, d$

Find ~~the~~ $\hat{\theta}_{MM}$ the method-of-moments estimate of θ as the solution to

$$\hat{\mu}_j = \mu_j(\theta)\Big|_{\hat{\theta}_{MM}}$$

Example 8.1

$\bar{X}$ and $\hat{\sigma}^2$ are the 1st sample moment and the second sample cetral moment respectively

Inverse Gaussian Model

$\theta = \begin{bmatrix} \mu \\ \lambda \end{bmatrix} \qquad d = 2$

$E(x) = \mu \qquad Var(X) = \dfrac{\mu^3}{\lambda}$

$$\begin{aligned} \bar{X} &= \mu \\ \hat{\sigma}^2 &= \frac{\mu^3}{\lambda} \end{aligned} \quad \Rightarrow \quad \begin{aligned} \hat{\mu}_{MM} &= \bar{X} \\ \hat{\lambda}_{MM} &= \frac{\bar{X}^3}{\hat{\sigma}^2} \end{aligned}$$

Gamma Model

$\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \qquad d = 2$

$$\begin{aligned} \bar{X} &= \frac{\alpha}{\beta} \\ \hat{\sigma}^2 &= \frac{\alpha}{\beta^2} \end{aligned} \quad \Rightarrow \quad \begin{aligned} \hat{\alpha}_{MM} &= \frac{\bar{X}^2}{\hat{\sigma}^2} \\ \hat{\beta}_{MM} &= \frac{\bar{X}}{\hat{\sigma}^2} \end{aligned}$$

Exponential Model

$\theta = \lambda \qquad d = 1$

$\bar{X} = \lambda^{-1} \Rightarrow \hat{\lambda}_{MM} = \bar{X}^{-1}$

Or

$\hat{\sigma}^2 = \lambda^{-2} \Rightarrow \hat{\lambda}'_{MM} = \dfrac{1}{\hat{\sigma}}$

Remark 8.1

M-of-M estimates are not unique, although they are easy to compute

Uncertainty in $\bar{X}$: Gaussian Case

$X_1, \ldots, X_n \qquad X_i \sim N(\mu, \sigma^2)$

$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \dfrac{\sigma^2}{n}$

$$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{or} \quad \frac{n^{\frac{1}{2}}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

95% Confidence Interval for μ

Case i) σ^2 known $\qquad \bar{X} \pm \dfrac{1.96\,\sigma}{n}$

Case ii) σ^2 unknown and estimated by s^2 (by convention)

$$\frac{n^{\frac{1}{2}}(\bar{X} - \mu)}{s} \sim t_{n-1} \qquad \bar{X} \pm \frac{t_{n-1}(0.975)\, s}{n}$$

where $t_{n-1}(0.975)$ is the 0.975^{th} quantile of the t-distribution with $n-1$ degrees of freedom

Uncertainty in $\hat{\sigma}^2$: Gaussian Case

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-1)} \equiv \Gamma\left(\alpha = \frac{n-1}{2}, \beta = 2\right)$$

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{r=1}^{n}(X_r - \bar{X})^2}{\sigma^2} \Rightarrow \quad 95\% \text{ CI for } \sigma^2$$

$$Pr\left(\chi^2_{(n-1)}(0.025) \leq \frac{n\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{(n-1)}(0.975)\right)$$

$$= Pr\left(\chi^2_{(n-1)}(0.975)^{-1} \leq \frac{\sigma^2}{n\hat{\sigma}^2} \leq \chi^2_{(n-1)}(0.025)^{-1}\right)$$

$$= Pr\left(\chi^2_{(n-1)}(0.975)^{-1} n\hat{\sigma}^2 \leq \sigma^2 \leq \chi^2_{(n-1)}(0.025)^{-1} n\hat{\sigma}^2\right)$$

Uncertainty in $g(\bar{x})$: Delta Method

Assume g is a continuously differentiable function

then

$$g(\bar{x}) \approx N\left(g(\mu), \frac{\sigma^2 g'(\mu)^2}{n}\right)$$

where $x_1, \ldots, x_n$ $\quad x_i \sim N(\mu, \sigma^2)$ $\quad \sigma^2$ is assumed known

By the mean-value theorem and the consistency of $\bar{x}$

$$g(\bar{x}) = g(\mu) + g'(\tilde{\mu})(\bar{x} - \mu)$$

where $\bar{x} \leq \tilde{\mu} \leq \mu$

$$E[g(\bar{x})] = g(\mu) \qquad Var[g(\bar{x}) - g(\mu)] = E[g'(\tilde{\mu})^2 (\bar{x}-\mu)^2]$$

$$= g'(\tilde{\mu})^2 \frac{\sigma^2}{n}$$

$\Rightarrow$ 95% CI is

$$g(\bar{x}) \pm \frac{g'(\bar{x})\, \sigma\, 1.96}{n}$$