FoDA L5 : Bayesian Inference

Bayes' Rule

$$\Pr(M \mid D) \propto \Pr(D \mid M) \cdot \Pr(M)$$

prior

proportional to

$$f(x) \propto g(x)$$
$$\Rightarrow f(x) = c \cdot g(x)$$

fixed, unknown constant

pdfs: needed for continuous R.Vs

$$p(M \mid D) \propto f(D \mid M) \cdot \pi(M)$$

posterior　　likelihood　　prior

if $\pi(M) = c$

$\hookrightarrow$ flat prior

model = H = height of university students.

data $D = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}$

prior $\pi(H)$ mean 66 inches
stdev 6 inches

$$\pi(H) = N_{66,6}(H) = \frac{1}{\sqrt{\pi\, 72}} \exp\left(\frac{-(\mu_H - 66)^2}{(2 \cdot 6^2)}\right)$$

Lets say $\underline{\text{MLE}}$ $\frac{1}{|D|} \sum_{x \in D} x = 5.5$ feet

Maximum likelihood estimate

$\times 12$

66 inches

$x_i \sim N_{\mu_H, 2}$ inches, independent

$$\underbrace{f(D \mid H)}_{\text{likelihood}} = \prod_{x \in D} g(x) = \underbrace{\prod_{x \in D}}_{\text{\prod}} \left(\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{1}{\underbrace{8}_{2 \cdot (\sigma)^2}} (\mu_H - x)^2 \right) \right)$$

$D = \{x_1, x_2, \dots, x_n\}$

$$\underbrace{p(H \mid D)}_{\text{posterior}} \propto f(D \mid H) \cdot \pi(H) \; N_{66,6}$$

$$f(D \mid H) \cdot \frac{1}{\sqrt{\pi 72}} \exp\left(-(\mu_H - 66)^2 / 72 \right)$$

log-posterior

$$\ln(p(H \mid D)) \propto \ln(f(D \mid H)) + \ln(\pi(H)) + C$$

$$\propto \left(\sum_{x \in D} \left(-\frac{1}{8} (\mu_H - x)^2 \right) \right) - \frac{1}{72} (\mu_H - 66)^2 + C$$

$$\propto - \left(\sum_{x \in D} 9 (\mu_H - x)^2 \right) - (\mu_H - 66)^2 + C$$

product

som

$$\ln\left(\prod_{x \in D} g(x)\right) = \sum_{x \in D}\left(\ln\left(g(x)\right)\right)$$

Weight of Prior

Data $D: x \sim N_{\mu_H, 2}$ $\rightarrow$ Variance 4

Prior $\pi(\mu_H) = N_{66, \cancel{6}}$ $\rightarrow$ 36 $\qquad \frac{4}{36} = \frac{1}{9}$

$0.1 \rightarrow 0.01 \qquad \frac{4}{0.01} = 400$

$\hookrightarrow$ prior worth ~~2~~400 data points

$$\sum_{x \in D} (x - \mu_H)^2 + 400\,(66 - x)^2$$

$|D| = n = 100 ? \Rightarrow$ trust prior more

if $|D| = n = 10{,}000 \Rightarrow$ trust data more

Weighted Average

prior

66

$$\boxed{\begin{array}{ll} \text{set} & \text{points } x_1, x_2, \ldots x_n \\ & \text{weights } w_1, w_2, \ldots w_n \end{array}}$$

X_{n+1}

w_{n+1}

900

1 1 1

$$w\text{-ave}(X, w) = \frac{72 \sum_{i=1}^{n} w_i x_i}{72 \sum_{i=1}^{n} w_i}$$

uniform ave

$$w_i = \frac{1}{n} \implies \frac{\sum_i \frac{1}{n} x_i}{\sum_{i=1}^{n} \frac{1}{n}} = \frac{\frac{1}{n}\sum_i x_i}{1}$$

independent

Data + Gaussian noise

↳ MLE $\min_{\mu} \sum_{x_i \in D} (x_i - \mu)^2$

sum of squared errors

↳ prior

↳ Gaussian prior

↳ weighted SSE

↳ Regularizer $\sum_{x_i \in D} (x_i - \mu)^2 + R(\mu)$

Power of Posterior

$$M^* = \max_M p(M \mid D)$$ ← MAP estimate

- compare posteriors of models M_1, M_2

$$\frac{p(M_1 \mid D)}{p(M_2 \mid D)} = 1.3 \text{ or } 100$$

- marginalize over models M_1, M_2, M_3

new data x'

$M(x') \to v$

$$\hat{v} = \frac{p(M_1 \mid D) \cdot M_1(x') + p(M_2 \mid D) \cdot M_2(x') + p(M_3 \mid D) \cdot M_3(x')}{\sum_i p(M_i \mid D)}$$

- confidence intervals on models

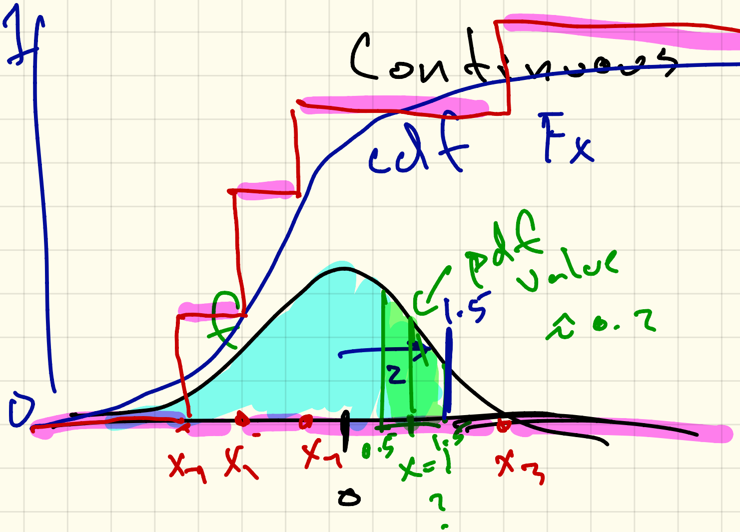
$M \in \mathbb{R}$

w/ 95% confidence

$M \in [61, 68 \text{ inches}]$

Discrete R.V. $X \in [A, B, C]$

$Pr[X = A] = 0.4$



$X \in \mathbb{R}$

$X \sim N(0, 2)$

$Pr[X = 1] = 0$

$Pr[x \in [0.5, 1.5]] = \int_{x \in 0.5}^{1.5} f_X(x)$

cdf $F_X(t) = Pr[X \le t] = \int_{x=-\infty}^{t} f_X(x)$