this time: multivariate θ

next time: ~↓

(J) read: ch. 1–4; (5, 6) new

AMS 206
20 Feb 18

(1)

(G) ch. 1–4

THT 1 quiz 3 due 11:59 pm tonight

Quiz 4 now assigned: drop-dead due date $\boxed{\text{6 Mar 18}}$

$$\begin{array}{ccc} 2 & 3 & 4 \\ \begin{bmatrix} 16 \\ 16 \\ \vdots \end{bmatrix} & \begin{bmatrix} 16.0 \\ 16.0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 15.98 \\ 16.01 \\ \vdots \end{bmatrix} \end{array}$$

deterministic | stochastic / probabilistic

$$F \sim p(F)$$

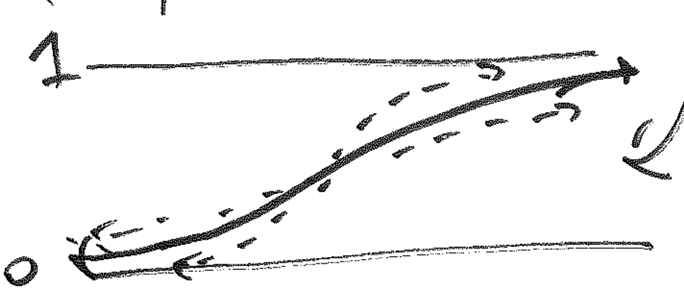
(lik) $(Y_i \mid F\, \mathcal{B}) \stackrel{IID}{\sim} F$

$(i = 1, \dots, n)$

underlying CDF

$\boxed{Y_i \text{ exch. } (\mathcal{B})}$

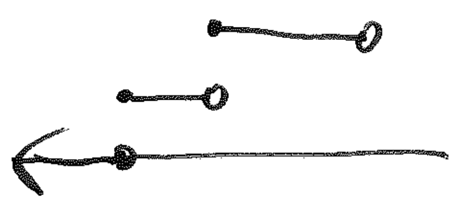
(suppose Y_i continuous on $\mathbb{R}$)



LoS (Poisson) case study:

F = CDF of Poisson(λ), $(\lambda > 0)$

F discrete



putting priors on functions: Bayesian nonparametric (BNP) methods

meta conjecture | all successful machine learning methods are BNP in disguise

2 ways to solve model uncertainty problem

1. BNP
2. Bayesian cross-validation

(k = 2)

temporary model

$$(\mu, \sigma^2) \sim p(\mu, \sigma^2) \quad G$$

$$(y_i \mid \mu, \sigma^2, B) \overset{IID}{\sim} N(\mu, \sigma^2)$$

$$(i = 1, \dots, n)$$

$\theta = (\mu, \sigma^2)$; $\underset{k \times 1}{\theta}$

1 2

$$\boxed{p(\theta \mid y)} = c \, \boxed{p(\theta)} \, p(y \mid \theta) = c \, p(\theta) \boxed{\ell(\theta \mid y)}$$

$\boxed{k = 100}$

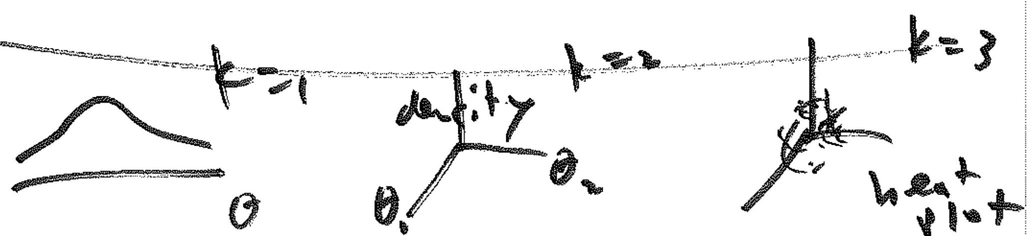
$$\int c \cdot p(\theta) \, \ell(\theta \mid y) \, d\theta = 1$$

Θ ← space of possible θ values

usually $\Theta = \mathbb{R}^k$

$$c \underbrace{\int \cdots \int}_{\leftarrow 100 \rightarrow} p(\theta) \, \ell(\theta \mid y) \, d\theta = 1$$

(k = 100)-dimensional density $p(\theta \mid y)$



$$\boxed{\underset{\sim}{\theta} = (\theta_1, \dots, \theta_k)}$$ look at marginals:

$$p(\theta_1 | y) = \underbrace{\int \cdots \int}_{(k-1)} p(\underset{\sim}{\theta} | y) \, d\theta_2 \, d\theta_3 \cdots d\theta_k$$

← 99 +

$(k = 100)$ $(k-1)$ (99) dimensional $\int$

Laplace approximations (1785) can be used to approx. these integrals

$(k = 10)$

$\mu \sim p(\mu)$

kmp. temp. model

$$\left\{ (Y_i | \mu, \sigma) \overset{IID}{\sim} N(\mu, \sigma_0^2) \right\} \text{ known}$$

$(i = 1, \dots, n)$

$$\sigma_*^2 = \frac{1}{\frac{1}{\sigma_\mu^2} + \frac{n}{\sigma_0^2}}$$

posterior variance

$\therefore$

$$\frac{1}{\sigma_*^2} = \frac{1}{\sigma_\mu^2} + \frac{n}{\sigma_0^2}$$

$$\begin{pmatrix}\text{post.}\\\text{precision}\end{pmatrix} = \begin{pmatrix}\text{prior}\\\text{precision}\end{pmatrix} + \begin{pmatrix}\text{lik.}\\\text{prec.}\end{pmatrix}^{(\text{data})}$$

$$\boxed{p(\mu, \sigma^2)} = p(\mu) \cdot p(\sigma^2 | \mu)$$

$$= \boxed{\underline{p(\sigma^2)} \cdot p(\underline{\mu | \sigma^2})}$$

Gaussian model

$y = (y_1, \ldots, y_n)$

$(Y_i \mid \mu\, \sigma^2\, \mathcal{G}\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma^2)$

$(i = 1, \ldots, n)$

important fact about maximum likelihood

(5)

(I) suppose you parameterize in terms of

$\theta = \sigma$

$$\ell_1(\mu\, \theta \mid y\, \mathcal{G}\, \mathcal{B}) = \prod_{i=1}^{n} p(y_i \mid \mu\, \theta\, \mathcal{G}\, \mathcal{B})$$

$$= \prod_{i=1}^{n} \frac{1}{\theta\sqrt{2\pi}} \exp\left[-\frac{1}{2\theta^2}(y_i - \mu)^2\right]$$

$$= c\, \theta^{-n} \exp\left[-\frac{1}{2\theta^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

so the log likelihood function is (ignoring irrelevant constant)

$$\ell\ell_1(\mu\, \theta \mid y\, \mathcal{G}\, \mathcal{B}) = -n \log \theta - \frac{1}{2\theta^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\frac{\partial}{\partial \mu}\ell\ell_1(\mu\, \theta \mid y\, \mathcal{G}\, \mathcal{B}) = -\frac{1}{2\theta^2}\sum_{i=1}^{n} 2(y_i - \mu)(-1)$$

$= 0$ iff $\mu = \hat{\mu}_{MLE} = \bar{y}$ (check), $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$\frac{\partial}{\partial \theta}\ell\ell_1(\mu\, \theta \mid y\, \mathcal{G}\, \mathcal{B}) = -\frac{n}{\theta} - \frac{(-2)}{2\theta^3}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$= 0 \text{ iff } \theta^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2 \to \hat{\theta}_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Now instead parameterize in terms of $\theta = \sigma^2$

$$\ell_2(\mu\,\theta \mid y\, \mathcal{G}\, \mathcal{B}) = c\, \theta^{-\frac{n}{2}} \exp\left[-\frac{1}{2\theta}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$\ell\ell_2(\mu\,\theta \mid y\, \mathcal{G}\, \mathcal{B}) = -\frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\frac{\partial}{\partial\mu}\ell\ell_2(\mu\,\theta \mid y\, \mathcal{G}\, \mathcal{B}) = -\frac{1}{2\theta}\sum_{i=1}^{n} 2(y_i - \mu)(-1)$$

$$= 0 \text{ iff } \mu = \hat{\mu}_{MLE} = \bar{y} \quad (\text{check})$$

$$\frac{\partial}{\partial\theta}\ell\ell_2(\mu\,\theta \mid y\, \mathcal{G}\, \mathcal{B}) = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$= 0 \text{ iff } \theta = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2 \to \hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Thus

$$\hat{\sigma}_{MLE} = \sqrt{(\hat{\sigma}^2)_{MLE}}$$

functional invariance of MLE

$\hat{\theta}_{MLE}$ of θ, g invertible $\to$

$$\widehat{[g(\theta)]}_{MLE} = g(\hat{\theta}_{MLE})$$