BAYESian ESTIMATION

BAYES RULE !

Dr. Test for Disease

H_0: sick

H_1: not sick

Test declare sick or

Test declares Healthy

given

$P(\text{type I error}) = 0.02 = P(\text{Test Health} \mid \text{sick})$

$P(\text{type II error}) = 0.01 = P(\text{Test sick} \mid \text{healthy})$

pretty small so "trust the test" ?

Suppose you have additional information (PRIOR "to seeing your data") e.g. $\underbrace{\text{common disease}}_{P(\text{sick}) = 1/10}$ or $\underbrace{\text{rare}}_{P(\text{sick}) = 1/100{,}000}$

How to incorporate this extra info. in your decision ? BAYES RULE.

E.g. suppose your test declares you sick.

$$\star = P(\text{sick} \mid \underbrace{\text{test declared me sick}}_{\text{TDS}}) = \frac{P(\text{TDS} \mid \text{sick})\, P(\text{sick})}{P(\text{TDS})} = \frac{P(\text{TDS} \mid \text{sick})\, P(\text{sick})}{P(\text{TDS} \mid \text{sick})\, P(\text{sick}) + P(\text{TDS} \mid \text{not sick})\, P(\text{not sick})}$$

"POSTERIOR"

$\star = 0.91$ if $p(\text{sick}) = 1/10$

$\star = 0.001$ if $p(\text{sick}) = 1/100{,}000$

General Hypotheses

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1$

<u>Frequentist/classical probability</u> : prob. of event A can be observed as a relative frequency over replications of experiment ($\hat{p} \to p$ as $n \to \infty$)

<u>Subjective approach</u> : prob. as degree of belief in something.

<u>PRIOR</u> — hard data (prior data) / subjective $\xrightarrow{\text{Bayes rule update}}$ frequentist prob. / degree of belief — <u>POSTERIOR</u>

~·~

How to turn the knowledge of a posterior probability to a $\underbrace{\text{decision about the state of the world}}_{\text{inference}}$?

Our parameter θ is the realization of a r.v. having density $p(\theta)$ ← prior or $\underbrace{\text{posterior}}_{\text{given the data \& after the update}}$

How to predict/infer the value of a hidden/unobservable r.v.

Suppose you form a predictor/estimator called $\tilde{\theta}$.

Error of prediction: $\tilde{\theta} - \theta$

Define MSE : $E(\tilde{\theta}-\theta)^2$ ← gives a measure of accuracy in the $E(\ )^2$ metric

MAE : $E|\tilde{\theta}-\theta|$ ← Mean Absolute error

0-1 error $E\, 1_{\{\tilde{\theta} \neq \theta\}} = P(\tilde{\theta} \neq \theta)$ ← more intuitive if θ is discrete valued

$\ell(\tilde\theta, \theta)$ = loss associated with predictor $\tilde\theta$ & true state θ

$E\,\ell(\tilde\theta,\theta)$ = expected loss A.K.A. "<u>Risk</u>"

$\ell(\tilde\theta,\theta) = (\tilde\theta-\theta)^2$ L^2 loss

$\ell(\tilde\theta,\theta) = |\tilde\theta-\theta|$ L^1 loss

$\ell(\tilde\theta,\theta) = \begin{cases} 1 & \text{if } \tilde\theta\neq\theta \\ 0 & \text{if } \tilde\theta=\theta \end{cases}$ 0-1 loss

$$\boxed{\text{given } p(\theta) \;\&\; E\theta = \int \theta\, p(\theta)\, d\theta \quad \text{or } E\theta = \sum_\theta \theta\, p(\theta)}$$

1) <u>Find $\tilde\theta$ that minimizes L^2 loss</u>

$$E(\underset{\text{ct.}}{\tilde\theta} - \underset{\text{r.v.}}{\theta})^2 = E(\tilde\theta - E\theta + E\theta - \theta)^2 = E(\tilde\theta - E\theta)^2 + 2E\left[(\tilde\theta - E\theta)(E\theta-\theta)\right] + E(\underbrace{E\theta-\theta}_{\text{var }\theta})^2$$

it is constant

$$= E(\tilde\theta - E\theta)^2 + \text{var}\,\theta$$

$\boxed{\tilde\theta = \text{mean}}$

$\tilde\theta = E\theta$ is the minimizer $\Rightarrow$ $\tilde\theta = E\theta$ is the <u>best</u> predictor of r.v. θ

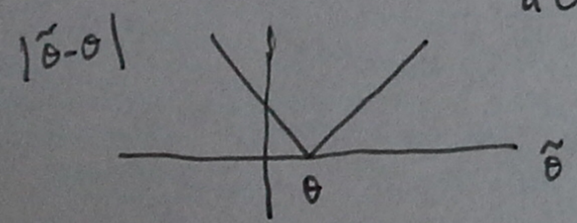
w.r.t. L^2 loss

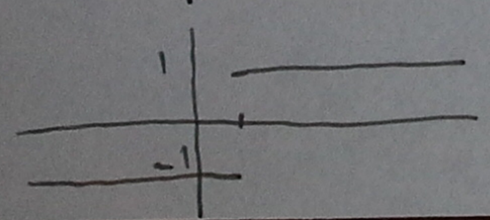
2) $E|\tilde\theta-\theta| = L_1$ loss

Assume θ is continuous r.v. with density $p(\theta)$

$$0 = \frac{d}{d\tilde\theta} E|\tilde\theta-\theta| = \frac{d}{d\tilde\theta}\int_{-\infty}^{\infty} |\tilde\theta-\theta|\, p(\theta)\, d\theta = \int_{-\infty}^{\infty} \frac{d}{d\tilde\theta}|\tilde\theta-\theta|\; p(\theta)\, d\theta = -\int_{\tilde\theta}^{\infty} p(\theta)\, d\theta + \int_{-\infty}^{\tilde\theta} p(\theta)\, d\theta = 0 \text{ when}$$

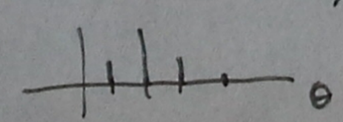
$\boxed{\tilde\theta = \text{median}}$





3) 0-1 loss $\rightsquigarrow$ minimize $E\mathbb{1}\{\tilde\theta\neq\theta\} = P(\tilde\theta\neq\theta) \Leftrightarrow$ maximizing $P(\tilde\theta=\theta)$

Assume θ is discrete, taking values $\theta_1, \theta_2, \ldots, \theta_k$



} choose $\boxed{\tilde\theta = \text{mode}}$

BAYESIAN APPROACH

1) θ is considered a unobserved realization of a r.v. coming from <u>prior</u> density $p(\theta)$

2) using data $X_1, \dots, X_n$ iid from $f_\theta(\cdot)$ to extract info about θ
   Use Bayes Rule to get $\underbrace{p(\theta \mid X_1, \dots, X_n)}_{\text{posterior density}}$

3) Use a $\underbrace{\text{measure of location}}_{\text{mean, median or mode}}$ of the posterior in order to predict/estimate θ.

~o~

<u>ex</u> $\underbrace{\text{A bag of coins}}_{p(\theta)}$, pick a coin at random & start tossing it $\rightarrow X_1, \dots, X_n \sim \text{Bernoulli}(\theta)$

How to get posterior (BAYES RULE)

$$p(\theta \mid x_1, \dots, x_n) = \frac{\overbrace{P(x_1, \dots, x_n \mid \theta)}^{\text{Likelihood}} P(\theta)}{P(x_1, \dots, x_n)} = \frac{P(x_1, \dots, x_n \mid \theta) P(\theta)}{\int P(x_1, \dots, x_n \mid \theta = a) P(\theta = a)\, da} = \circledast$$

sum if θ is discrete

If there is a sufficient statistic $T = T(x_1, \dots, x_n)$

then $P(x_1, \dots, x_n \mid \theta) = f_{T \mid \theta}(T) \cdot h(x_1, \dots, x_n)$

$$\text{Then } \circledast = \frac{f_{T \mid \theta}(T) \cdot \cancel{h(x_1, \dots, x_n)}\, P(\theta)}{\int f_{T \mid \theta = a}(T)\, \cancel{h(x_1, \dots, x_n)}\, P(\theta = a)\, da} = \frac{f_{T \mid \theta}(T)\, P(\theta)}{\int f_{T \mid \theta = a}(T)\, P(\theta = a)\, da}$$

<u>Example</u> $X_1, \dots, X_n \sim \text{iid } N(\theta, \sigma^2)$ where $\theta \sim N(\mu, \tau^2)$ ← $\boxed{\text{PRIOR } p(\theta)}$

Assume σ^2, μ, τ^2 <u>known</u>.

Sufficient statistic is $T = \bar{X} = \frac{1}{n}\sum_1^n X_i \sim N(\theta, \frac{\sigma^2}{n})$. Denote $\sigma_n^2 = \sigma^2/n$

$$f_{T|\theta}(t) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{(t-\theta)^2}{2\sigma_n^2}\right\}$$

<u>Posterior</u> $$p(\theta \mid T=t) = \frac{f_{T|\theta}(t)\, p(\theta)}{\int f_{T|\theta}(t)\, p(\theta=a)\, da}$$

$$f_{T|\theta}(t)\, p(\theta) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{(t-\theta)^2}{2\sigma_n^2}\right\} \cdot \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$= \text{constant} \cdot \exp\left\{-\frac{(t-\theta)^2}{2\sigma_n^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$\frac{(t-\theta)^2}{\sigma_n^2} + \frac{(\theta-\mu)^2}{\tau^2} = \frac{\tau^2(t^2 - 2t\theta + \theta^2) + \sigma_n^2(\theta^2 - 2\mu\theta + \mu^2)}{\sigma_n^2\tau^2} = \frac{(\tau^2+\sigma_n^2)\theta^2 - 2(\tau^2 t + \sigma_n^2\mu)\theta + \tau^2 t^2 + \sigma_n^2\mu^2}{\sigma_n^2\tau^2}$$

$$= \frac{\theta^2 - \frac{2(\tau^2 t + \sigma_n^2\mu)}{\tau^2+\sigma_n^2} + \frac{\tau^2 t^2 + \sigma_n^2\mu^2}{\tau^2+\sigma_n^2}}{\sigma_n^2\tau^2/(\tau^2+\sigma_n^2)} = \dots$$

Bayes estimator $\tilde{\theta} = \overbrace{\dfrac{\tau^2 \bar{X} + \sigma_n^2 \mu}{\tau^2 + \sigma_n^2}}^{\text{mean / median / mode}}$

$\tilde{\theta} \approx \mu$ for small n (& large σ^2)

but

$\tilde{\theta} \approx \bar{X}$ for large n

$\sigma_n^2 = \dfrac{\sigma^2}{n} \to 0$

not depending on the prior

<u>Posterior</u> $= \text{constant} \cdot \exp\left\{ - \dfrac{(\theta - \nu)^2}{2\, \sigma_n^2 \tau^2 / (\tau^2 + \sigma_n^2)} \right\}$ where $\nu = \dfrac{\tau^2 t + \sigma_n^2 \mu}{\tau^2 + \sigma_n^2}$

$$\boxed{\text{Posterior is } N\left(\nu, \frac{\sigma_n^2 \tau^2}{\tau^2 + \sigma_n^2}\right)}$$

"Conjugate prior"
same <u>shape</u> prior & posterior