Bayesian Estimation

Dr. Test for Disease

H₀: sick
H₁: not sick

Test declare sick or Test declares Healthy

\[ P \left( \text{type I error} \right) = 0.02 = P \left( \text{Test Healthy | sick} \right) \]

\[ \text{given} \quad P \left( \text{type II error} \right) = 0.01 = P \left( \text{Test sick | healthy} \right) \]

Pretty small so "trust the test"?

"Prior" (to seeing your data)

Suppose you have additional information e.g. common disease or rare

\[ P(\text{sick}) = \frac{1}{10} \quad P(\text{sick}) = \frac{1}{100,000} \]

How to incorporate this extra info. in your decision? Bayesian Rule.

E.g. suppose your test declares you sick.

\[ \star = P \left( \text{sick | test declared me sick} \right) = \frac{P(\text{TDS | sick}) P(\text{sick})}{P(\text{TDS})} = \frac{P(\text{TDS | sick}) P(\text{sick})}{P(\text{TDS | sick}) P(\text{sick}) + P(\text{TDS | not sick}) P(\text{not sick})} \]

\[ \star = 0.91 \quad \text{if} \quad P(\text{sick}) = \frac{1}{10} \]

\[ \star = 0.001 \quad \text{if} \quad P(\text{sick}) = \frac{1}{100,000} \]

General Hypothesis
H₀: θ = θ₀
H₁: θ ≠ θ₀
Frequentist/classical probability: prob. of event $A$ can be observed as a relative frequency over replications of experiment ($\hat{p} \to p$ as $n \to \infty$)

Subjective approach: prob. as degree of belief in something.

Prior \quad hard data (prior data) \quad \text{Bayes rule update} \quad \text{frequentist prob. posterior}

\quad \text{degree of belief}

How to turn the knowledge of a posterior probability to a decision about the state of the world?

Our parameter $\theta$ is the realization of a r.v. having density $p(\theta)$

How to predict/infer the value of a hidden/unobservable r.v. $^\text{prior or posterior}$ given the data & after the update

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

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

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

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

0-1 error $E 1{\tilde{\theta} = \theta} = p(\tilde{\theta} = \theta)$ \quad - more intuitive if $\theta$ is discrete valued
\[ E(L(\tilde{\theta}, \theta)) = \text{loss associated with predictor } \tilde{\theta} \text{ & true state } \theta \]

\[ E(R(\tilde{\theta}, \theta)) = \text{expected loss} \quad \text{A.K.A. } \textit{Risk} \]

\[ \ell(\tilde{\theta}, \theta) = (\tilde{\theta} - \theta)^2 \quad \text{L}^2 \text{ loss} \]
\[ \ell(\tilde{\theta}, \theta) = |\tilde{\theta} - \theta| \quad \text{L}^1 \text{ loss} \]
\[ \ell(\tilde{\theta}, \theta) = \begin{cases} 1 & \text{if } \tilde{\theta} \neq \theta \\ 0 & \text{if } \tilde{\theta} = \theta \end{cases} \quad \text{0-1 loss} \]

\( \text{given } p(\theta) \text{ & } \mathbb{E}\theta = \int \theta p(\theta) d\theta \)

or \( \mathbb{E}\theta = \sum \theta p(\theta) \)

1) Find \( \tilde{\theta} \) that minimizes \( L^2 \) loss

\[
E(\tilde{\theta} - \theta)^2 = E(\tilde{\theta} - \theta + \theta - \theta)^2 = E(\tilde{\theta} - \theta)^2 + 2E[(\tilde{\theta} - \theta)(\theta - \theta)] + E(\theta - \theta)^2
\]

\[
\text{if } \theta \text{ is constant, } \text{var } \theta = E(\tilde{\theta} - \theta)^2 + \text{var } \theta
\]

\[
\tilde{\theta} = \text{mean} \]

\[
\tilde{\theta} = \mathbb{E}\theta \text{ is the minimizer } \Rightarrow \tilde{\theta} = \mathbb{E}\theta \text{ is the best predictor of r.v. } \theta
\]

2) \( E|\tilde{\theta} - \theta| = L_1 \) loss

\[
0 = \frac{d}{d\mathbb{E}\theta} E|\tilde{\theta} - \theta| = \frac{d}{d\mathbb{E}\theta} \int |\tilde{\theta} - \theta| p(\theta) d\theta = \int \frac{d}{d\mathbb{E}\theta} |\tilde{\theta} - \theta| p(\theta) d\theta = -\int p(\theta) d\theta + \int p(\theta) d\theta = 0 \text{ when } \mathbb{E}\theta = \text{median}
\]

3) \( 0-1 \) loss \( \Rightarrow \) minimize \( E1{[\tilde{\theta} \neq \theta]} = P(\tilde{\theta} \neq \theta) \) \( \Rightarrow \) maximization \( P(\tilde{\theta} = \theta) \)

Assume \( \theta \) is discrete, taking values \( \theta_1, \theta_2, \ldots, \theta_k \)

Choose \( \tilde{\theta} = \text{mode} \)
BAYESIAN APPROACH

1) \( \theta \) is considered a realization of a r.v. coming from prior density \( p(\theta) \),

2) using data \( X_1, \ldots, X_n \) iid from \( f_\theta(\cdot) \) to extract info about \( \theta \).
   Use Bayes Rule to get \( p(\theta | X_1, \ldots, X_n) \),
   posterior density

3) Use a measure of location of the posterior in order to predict/estimate \( \theta \).
   mean, median
   or mode

\[ \sim \sim \]

ex A bag of coins, pick a coin at random & start tossing it \( \Rightarrow X_1, \ldots, X_n \sim \text{Bernoulli}(\theta) \)

\[ p(\theta) \]
How to get posterior (Bayes Rule)

\[
p(\theta | x_1, \ldots, x_n) = \frac{P(x_1, \ldots, x_n | \theta) P(\theta)}{P(x_1, \ldots, x_n)} = \frac{P(x_1, \ldots, x_n | \theta) P(\theta)}{\int P(x_1, \ldots, x_n | \theta = a) P(\theta = a) \, da} = \ast
\]

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

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

Then \( \ast = \frac{f_{T \mid \theta}(T) \cdot h(x_1, \ldots, x_n) P(\theta)}{\int f_{T \mid \theta=a}(T) \cdot h(x_1, \ldots, x_n) P(\theta = a) \, da} = \frac{f_{T \mid \theta}(T) P(\theta)}{\int f_{T \mid \theta=a}(T) P(\theta = a) \, da} \)
Example \[ x_1, \ldots, x_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2) \text{ where } \theta \sim \mathcal{N}(\mu, \tau^2) \]

Assume \( \sigma^2, \mu, \tau^2 \) known.

Sufficient statistic is \( T = \bar{X} = \frac{1}{n} \sum x_i \sim \mathcal{N}(\theta, \sigma^2/n) \).

Denote \( \sigma_n^2 = \sigma^2/n \)

\[
f_{T|\theta}(t) = \frac{1}{\sqrt{2\pi} \sigma_n^2} e^{-\frac{(t-\theta)^2}{2\sigma_n^2}}
\]

Posterior \[
p(\theta | T = t) = \frac{f_{T|\theta}(t) \ p(\theta)}{\int f_{T|\theta}(t) \ p(\theta = a) \ da}
\]

\[
f_{T|\theta}(t) \ p(\theta) = \frac{1}{\sqrt{2\pi} \sigma_n^2} e^{-\frac{(t-\theta)^2}{2\sigma_n^2}} \cdot \frac{1}{\sqrt{2\pi} \tau^2} e^{-\frac{(\theta-\mu)^2}{2\tau^2}}
\]

\[
= \text{constant} \cdot e^{-\frac{(t-\theta)^2}{2\sigma_n^2} - \frac{(\theta-\mu)^2}{2\tau^2}}
\]

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

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

\[
= \frac{1}{\sigma_n^2 \tau^2 / (\tau^2 + \sigma_n^2)}
\]
Bayes estimator $\tilde{\theta} = \frac{\bar{X} + \frac{1}{n} \mu}{\tau^2 + \frac{1}{n}}$

\[\begin{align*}
\tilde{\theta} &\approx \mu \text{ for small } n \text{ (large } \sigma^2) \\
\text{but} \\
\tilde{\theta} &\approx \bar{X} \text{ for large } n \\
\uparrow \\
\frac{\sigma^2}{n} &= \frac{\sigma^2}{n} \to 0 \\
\text{not depending} &\text{ on the prior}
\end{align*}\]

Posterior $= \text{constant} \cdot \exp\left\{ - \frac{(\tilde{\theta} - \nu)^2}{2 \frac{\sigma^2 \tau^2}{6n^2 \tau^2/\left(\tau^2 + 6\sigma^2\right)}} \right\}$ where $\nu = \frac{\tau^2 \bar{X} + 6\sigma^2 \mu}{\tau^2 + 6\sigma^2}$

Posterior is $N(\nu, \frac{\sigma^2 \tau^2}{\tau^2 + 6\sigma^2})$