Lec 05

Friday, January 13, 2023 11:47 AM

— No class on Mon 1/16 and we meet on Wed 1/18

Recap

— Sample vs variable in the random sense

— $P(y_1, y_2) = P(y_1)\, P(y_2 | y_1)$ ↳ conditionals

$= P(y_1)\, P(y_2)$

— likelihood given a sample $= [y_1, y_2, \ldots, y_n]$

$$\mathcal{L}(\mu, \sigma) = \prod_{i=1}^{n} P(y_i)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right)$$

— Maximum likelihood Estimate (MLE)

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$\hat{\sigma} = \frac{\sum_{i=1}^{n} (y_i - \hat{\mu})^2}{n}$$

→ $y_i = [y_0^i, y_1^i, \ldots, y_{K-1}^i]$ ← size K

$y_i \sim mvn(\mu, \Sigma) \sim \mathcal{N}(\mu, \Sigma)$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\mu})^T (y_i - \hat{\mu})$$

MAP estimates

prior — your knowledge about some parameters

posterior — after you encode your prior knowledge

$$\mathcal{L}(\mu, \sigma) = \prod_{i=1}^{n} P(y_i)$$

$$= P(\mu)\, P(y | \mu)$$

$p(\hat{\mu}) = N(\mu_0, \sigma_0)$

$$= \mathcal{N}(\mu_0, \sigma_0) \prod_{i=1}^{n} P(y_i | \mu)$$

$$\propto \mathcal{N}(\mu_0, \sigma_0) \exp\left(\frac{-1}{2}\left(\frac{y_i - \mu}{\sigma}\right)^2\right)$$

Maximize the likelihood to obtain values for μ, σ

$$\propto \exp\left(\frac{-1}{2} \sum \frac{(y_i - \mu)^2}{\sigma^2} + \frac{-1}{2\sigma_0^2} (\mu - \mu_0)^2\right)$$

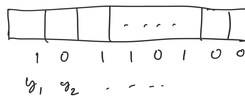
$$\log \mathcal{L}(\mu, \sigma) \propto \frac{-1}{2} \sum \frac{(y_i - \mu)^2}{\sigma^2} - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2$$

↑ Maximum of $\log \mathcal{L}(\mu, \sigma)$ = MAP estimate (maximum a-posteriori)

Randomization and design

— Silica based nanoparticles for reducing effects of fertilizers on the crop yield

— 20 seeds 1) with silica NP 2) without silica NP

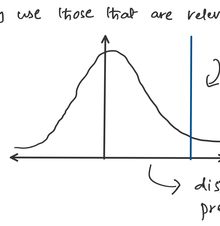


gain $= \bar{y}_1 - \bar{y}_0$ ← does not tell us the full story interms of variance (or) effects of experimental errors

Known reference distribution:

— conditions of experiment

— only use those that are relevant



10 batches of w/o modification seed data

some measure of statistical significance but we have access to only samples

significant / not ⇑ Student's t-distribution $\begin{cases} t = \text{gain apparent in your expt} \\ y_i = \text{differences between batches} \end{cases}$