Hypothesis Testing

1 Formulating hypothesis

We want to test whether new production line leads to a decrease in the average battery lifetime.

Null hypothesis H_0 : The mean battery life is unchange

$\mu = 500 \qquad \sigma = 50$

Alternate hypothesis : The mean battery lifetime has decreased

$\mu < 500$

This is a one-sided test since we are only concerned about the reduction of lifetime

2 Test statistic based on sample mean $\bar{X}_n$

$\sigma = 50 \qquad n = 20$

Since n is reasonably large $Z \sim N$

$$\therefore \; Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

3 Critical value and rejection

$\alpha = 0.05$

We need to find value of z_α such that

$P(Z < z_\alpha) = \alpha$

$= 0.5$

From standard normal tables the critical value is

$z_{0.05} = -1.645$

which means that 5% of area under standard normal curve lies to the left of -1.645

$\therefore$ Reject H_0 if $Z < -1.645$

4 $\bar{X}_n = 480$

$$Z = \frac{480 - \mu_0}{\sigma/\sqrt{n}}$$

$$= \frac{480 - 500}{50/\sqrt{20}}$$

$$= \frac{-20 \times \sqrt{20}}{50}$$

$$= -1.789$$

5 Power function

$$\pi(\mu) = P(Z < -1.645 | \mu)$$

$$= P\left(\frac{\bar{X}_n - 500}{50/\sqrt{20}} < -1.645 | \mu\right)$$

$$= P\left(\bar{X}_n < 500 - \frac{1.645 \times 50}{\sqrt{20}} | \mu\right)$$

$$= P(\bar{X}_n < 481.61 | \mu)$$

Standardizing $\Rightarrow \phi\left(\frac{481.61 - \mu}{50/\sqrt{20}}\right)$

$$= \phi\left(\frac{481.61 - 470}{11.18}\right)$$

$$= 0.8508$$

$$= 85\%$$

6 P value is the probability of a test

$z = -1.789$ under H_0

$$P(z < -1.789) = \phi(-1.789) \approx 0.0368$$

at $\alpha = 0.05$ $p = 0.0368 < 0.05 \Rightarrow$ reject

$\alpha = 0.05$ $p = 0.0368 > 0.01 \Rightarrow$ dont reject

7 ~~Ans~~ 95% confidence

$$X_n - Z_{0.05} \times \frac{\sigma}{\sqrt{n}}$$

$$= 480 \pm 1.96 . 11.18$$

$$= 480 \pm 21.91$$

$\therefore$ Confidence interval is $[458.09, 501.91]$

Bayesian Statistics

Problem A

Part 1

a Given that prior information is modelled by a Beta distribution

$$\theta \sim \beta(\alpha, \beta) \qquad B(\alpha, \beta) \Rightarrow f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$\therefore$ Prior is given by

$$\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Given $\alpha = 2 \quad \beta = 8$

$$\Rightarrow \frac{\Gamma(2)\Gamma(8)}{\Gamma(10)} = \frac{1!\,7!}{9!}$$

$$= \frac{1! \times \cancel{7!}}{9 \times 8 \times \cancel{7!}}$$

$$= 1/72$$

$$\therefore \text{Prior} = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{1/72}$$

$$= 72\theta^{2-1}(1-\theta)^{8-1}$$

$$= 72\theta(1-\theta)^7$$

b Expectatiof θ

for $\beta(\alpha, \beta)$ Expectation is given by $\frac{\alpha}{\alpha+\beta}$

$\therefore$ for $\theta \sim \beta(2,8)$

$$E[\theta] = \frac{2}{2+8}$$

$$= \underline{\underline{0.2}}$$

Part 2

a For a Bernoulli variable if θ is the probability of success then probability of failure is $(1-\theta)$ and

$$f(X=x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

$$\text{Likelihood} = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{1-\sum_{i=1}^{n} x_i}$$

b given $n=10$ and number of success $=3$

$$\therefore f_{10}(X=3|\theta) = \theta^3(1-\theta)^7$$

Part 3

a Posterior distribution = Prior × Likelihood

$$= 72\theta(1-\theta)^7\,\theta^s(1-\theta)^{n-s}$$

where s is the number of success
n is the number of trials

$$\text{Posterior} = 72\,\theta^{1+s}(1-\theta)^{7+(n-s)}$$

$$\propto \theta^{1+s}(1-\theta)^{7+(n-s)}$$

$n = 10$ $s = 3$

$$\propto \theta^{1+3}(1-\theta)^{7+7}$$

$$\propto \theta^4(1-\theta)^{14}$$

$\alpha - 1 = 4$ $\beta - 1 = 14$
$\alpha = 5$ $\beta = 15$

$\therefore$ Posterior $\sim \beta(5, 15)$

b $E[\theta/X] = E[\beta(5,15)]$

$$= \frac{5}{5+15}$$

$$= \underline{\underline{0.25}}$$

Part 4

a Predictive probability $= \int_0^1 P(X_{n+1}=1/\theta) \cdot \text{Prior}$

Prior $\propto \dfrac{\theta(1-\theta)^7}{B(2,8)}$ $\quad P(X_{n+1}=1/\theta) = \theta$

$$= \int_0^1 \theta \cdot \frac{\theta(1-\theta)^7}{B(2,8)} d\theta$$

$$= \int_0^1 \frac{\theta^2 (1-\theta)^7}{B(2,8)}$$

$$= \int_0^1 \frac{B(3,8)}{B(2,8)}$$

$$B(\alpha+1, \beta) = \frac{\alpha}{\alpha+\beta} B(\alpha, \beta)$$

$$B(2+1, 8) = \frac{2}{2+8} B(2,8)$$

$$= 0.2\, B(2,8)$$

$$\text{Prob} = \frac{0.2\, B(2,8)}{B(2,8)}$$

$$= \underline{\underline{0.2}}$$

b

Part 5

Prior $\sim \beta(5, 15)$

$$\text{Prior} = \frac{\theta^4 (1-\theta)^{14}}{B(5,15)}$$

$$B(5,15) = \frac{\Gamma(5)\,\Gamma(15)}{\Gamma(20)}$$

$$= \frac{4! \times \cancel{14!}}{19 \times 17 \times 18 \times 16 \times 15 \times \cancel{14!}}$$

$$= \frac{\cancel{24}}{19 \times 17 \times \underset{3}{\cancel{18}} \times \underset{4}{\cancel{16}} \times 15}$$

$$= \frac{1}{19 \times 17 \times 12 \times 15}$$

$$\text{Prior} = 58140\ \theta^4 (1-\theta)^{14}$$
$$\propto \theta^4 (1-\theta)^{14}$$

$$\text{Likelihood} = \theta^s (1-\theta)^{n-s}$$

$$\text{Posterior} = \theta^4 (1-\theta)^{14}\, \theta^s (1-\theta)^{n-s}$$
$$= \theta^{4+s} (1-\theta)^{14+(n-s)}$$

$$\text{Posterior} \sim B(4+s+1,\ 14+(n-s)+1)$$
$$\sim B(5+s,\ 15+n-s)$$

$$\text{Predictive prob} = \int_0^1 \theta \cdot \frac{\theta^4 (1-\theta)^{14}}{B(5,15)}$$

$$= \int_0^1 \frac{\theta^5 (1-\theta)^{14}}{B(5,15)}$$

$$= \frac{B(6,15)}{B(5,15)}$$

$$= \frac{B(5+1, 15)}{B(5, 15)}$$

$$= \frac{\frac{5}{20} \cancel{B(5, 15)}}{\cancel{B(5, 15)}}$$

$$= \underline{\underline{0.25}}$$

Linear Least Squares Regression

Problem B

1 $y_i = \varphi(x_i)^T \theta_* + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \varphi(x_1)^T \\ \varphi(x_2)^T \\ \vdots \\ \varphi(x_n)^T \end{bmatrix} \theta_* + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$y = \phi \theta_* + \varepsilon$$

For least square problem

$$\hat{\theta} = (\phi^T \phi)^{-1} \phi^T y$$

$$= (\phi^T \phi)^{-1} \phi^T (\phi \theta_* + \varepsilon)$$

$$= \underbrace{(\phi^T \phi)^{-1} \phi^T \phi}_{I} \theta_* + (\phi^T \phi)^{-1} \phi^T \varepsilon$$

$$= \theta_* + (\phi^T \phi)^{-1} \phi^T \varepsilon$$

$$E[\hat{\theta}] = E[\theta_* + (\phi^T \phi)^{-1} \phi^T \varepsilon]$$

$$= E[\theta_*] + E[(\phi^T \phi)^{-1} \phi^T \varepsilon]$$

$$= \theta_* + (\phi^T \phi)^{-1} \phi^T E[\varepsilon]$$

$$= \theta_* + (\phi^T \phi)^{-1} \phi^T \times 0$$

$$= \theta_*$$

$E[\theta_*] = \theta_*$ since its a constant

$E[\varepsilon] = 0$ because $\varepsilon \sim N(0, \sigma^2)$
mean $= 0$

$\therefore$ $\hat{\theta}$ is an unbiased estimator

2 $\hat{\theta} = (\phi^T\phi)^{-1}\phi^T y$

$= (\phi^T\phi)^{-1}\phi^T(\phi\theta_* + \cancel{\phi}\varepsilon)$

$= \theta_* + (\phi^T\phi)^{-1}\phi^T\varepsilon$

$\mathrm{Cov}(\hat{\theta}) = E\left[(\hat{\theta} - E[\hat{\theta}])(\hat{\theta} - E[\hat{\theta}])^T\right]$

$\hat{\theta} - E[\hat{\theta}] = (\theta_* + (\phi^T\phi)^{-1}\phi^T\varepsilon - \theta_*)$

$(\hat{\theta} - E[\hat{\theta}])(\hat{\theta} - E[\hat{\theta}])^T$

$= (\cancel{\theta_*} + (\phi^T\phi)^{-1}\phi^T\varepsilon - \cancel{\theta_*})(\cancel{\theta_*} + (\phi^T\phi)^{-1}\phi^T\varepsilon - \cancel{\theta_*})^T$

$= (\phi^T\phi)^{-1}\phi^T\varepsilon\left[(\phi^T\phi)^{-1}\phi^T\varepsilon\right]^T$

$= (\phi^T\phi)^{-1}\phi^T\varepsilon\varepsilon^T\phi(\phi^T\phi)^{-1}$

$\mathrm{Cov}(\hat{\theta}) = E\left[(\phi^T\phi)^{-1}\phi^T\varepsilon\varepsilon^T\phi(\phi^T\phi)^{-1}\right]$

$= (\phi^T\phi)^{-1}\phi^T E[\varepsilon\varepsilon^T]\phi(\phi^T\phi)^{-1}$

$E[\varepsilon\varepsilon^T] = \mathrm{cov}(\varepsilon)$

$= \sigma^2$

$\therefore \mathrm{cov}(\hat{\theta}) = (\phi^T\phi)^{-1}\phi^T\sigma^2\phi(\phi^T\phi)^{-1}$

$= \sigma^2(\phi^T\phi)^{-1}\underbrace{\phi^T\phi(\phi^T\phi)^{-1}}_{I}$

$= \sigma^2(\phi^T\phi)^{-1}$

2 For computing OLS

$$\hat{\theta} = (\phi^T\phi)^{-1}\phi^T y$$

The term $(\phi^T\phi)^{-1}$ is critical for this calculation. Numerical instability arises when $(\phi^T\phi)^{-1}$ is ill conditioned.

The condition number $k(\phi^T\phi)$ measures how sensitive matrix inversion is to small changes.

$$k(\phi^T\phi) = \frac{\lambda_{max}}{\lambda_{min}}$$

Low k value : well conditioned, stable inversions

High k value : ill conditioned, small errors in ϕ amplify into large errors in $\hat{\theta}$

given $n = 50$ $d = 45$

theoretically $\lambda_{max} \approx (\sqrt{n} + \sqrt{d})^2$ for $\phi^T\phi$

$\lambda_{min} = (\sqrt{n} - \sqrt{d})^2$

$\lambda_{max} = 189.86$ $\lambda_{min} = 0.13$

$$k = \frac{189.86}{0.13}$$

$$\approx 1460$$

$k > 10^3 \Rightarrow$ ill conditioned

Inorder to tackle these problems of numerical instability we introduce regularization ~~p~~ methods

Eg: Ridge regression which adds a penalty to loss function biasing $\hat{\theta}$ towards zero

Principal component Analysis which projects ϕ to a k dimensional space thereby reducing ~~da~~ dimensionality

Part c

1 $$E[R[\hat{\theta}]] - \sigma^2 = \underbrace{\|E[\hat{\theta}] - \theta^*\|_{\hat{\Sigma}}^2}_{\text{Bias}} + \underbrace{E[\|\hat{\theta} - E[\hat{\theta}]\|_{\hat{\Sigma}}^2]}_{\text{variance}}$$

Bias measures how far the average estimate $E[\hat{\theta}]$ is from the true θ^*. High bias oversimplifies the model ie underfit.

Variance measures how much $\hat{\theta}$ fluctuates around its mean high variance occurs when the complex model overfit to the noise.

Trade off in model complexity

Low complexity model
- Low variance
- High bias : cause errors, misclassifications, etc

High complexity model
- High variance : sensitive to noise data
- Low bias

Optimal complexity
- Balanced bias and variance
- minimizes excess risk

Problem D

1

$$\Pi = \Phi(\Phi^T\Phi)^{-1}\Phi^T$$

$$\Pi^2 = \Phi(\Phi^T\Phi)^{-1}\underbrace{\Phi^T\Phi(\Phi^T\Phi)^{-1}}_{I}\Phi^T$$

$$= \Phi(\Phi^T\Phi)^{-1} I \Phi^T$$

$$= \Phi(\Phi^T\Phi)^{-1}\Phi^T$$

$$= \underline{\underline{\Pi}}$$

$$\Pi^T = \left[\Phi(\Phi^T\Phi)^{-1}\Phi^T\right]^T$$

$$= (\Phi^T)^T\left[(\Phi^T\Phi)^{-1}\right]^T\Phi^T$$

$$= \Phi\left[(\Phi^T\Phi)^T\right]^{-1}\Phi^T$$

$$= \Phi(\Phi^T\Phi)^{-1}\Phi^T$$

$$= \underline{\underline{\Pi}}$$

2

$$\hat{y} = \Phi\hat{\theta}$$

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T y \qquad \Pi = \Phi(\Phi^T\Phi)^{-1}\Phi^T$$

$$\therefore \hat{y} = \Phi(\Phi^T\Phi)^{-1}\Phi^T y$$

$$= \Pi y$$