Bayesian estimation and hypothesis testing

Andrew Thangaraj

IIT Madras

Section 1

Bayesian estimation

Parameter estimation

$$X_1, \ldots, X_n \sim \text{iid } X, \text{parameter } \theta$$

- Two schools of thought for design of estimators

Parameter estimation

$$X_1, \ldots, X_n \sim \text{iid } X, \text{parameter } \theta$$

- Two schools of thought for design of estimators
- Frequentist: treat θ as an unknown constant
    - Method of moments
    - Maximum likelihood

Parameter estimation

$$X_1, \ldots, X_n \sim \text{iid } X, \text{parameter } \theta$$

- Two schools of thought for design of estimators
- Frequentist: treat θ as an unknown constant
  - Method of moments
  - Maximum likelihood
- Bayesian: treat θ as a random variable with a known distribution
  - Bayesian estimation

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $\underset{\text{event}}{S} = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(\overbrace{p = 0.25}^{A}|\overbrace{S}^{B}) = P(S|p = 0.25)\underbrace{P(p = 0.25)}_{0.5}/P(S) =$
      $0.25^3 \times 0.75^2 \times 0.5/P(S) = 0.25$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.5/\underline{P(S)} = 0.25$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.5/\underline{P(S)} = 0.75$

Use Bayes' rule

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.5/P(S) = 0.25$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.5/P(S) = 0.75$
    - ★ $P(S) = 0.25^3 \times 0.75^2 \times 0.5 + 0.75^3 \times 0.25^2 \times 0.5 = 0.25^2 \times 0.75^2 \times 0.5$

$S = (S \text{ AND } (p=0.25)) \text{ OR } (S \text{ AND } (p=0.75))$

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.5/P(S) = 0.25$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.5/P(S) = 0.75$
    - ★ $P(S) = 0.25^3 \times 0.75^2 \times 0.5 + 0.75^3 \times 0.25^2 \times 0.5 = 0.25^2 \times 0.75^2 \times 0.5$
  - Estimator 1: Since $P(p = 0.75|S) > P(p = 0.25|S)$, we could estimate $\hat{p} = 0.75$

Example 1: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \text{Uniform}\{0.25, 0.75\}$
  - Assume p is chosen first at random according to the above distribution
  - Once p is chosen, the samples are drawn according to Bernoulli(p)
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) =$ $0.25^3 \times 0.75^2 \times 0.5/P(S) = 0.25$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.5/P(S) = 0.75$
    - ★ $P(S) = 0.25^3 \times 0.75^2 \times 0.5 + 0.75^3 \times 0.25^2 \times 0.5 = 0.25^2 \times 0.75^2 \times 0.5$
  - Estimator 1: Since $P(p = 0.75|S) > P(p = 0.25|S)$, we could estimate $\hat{p} = 0.75$
  - Estimator 2: Posterior mean, $\hat{p} = \underbrace{0.25\, P(p = 0.25|S) + 0.75\, P(p = 0.75|S)}_{\text{Conditional expected value}} = 0.625$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.9/P(S) = 0.75$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.9/P(S) = 0.75$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.1/P(S) = 0.25$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.9/P(S) = 0.75$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.1/P(S) = 0.25$
    - ★ $P(S) = 0.25^3 \times 0.75^2 \times 0.9 + 0.75^3 \times 0.25^2 \times 0.1 = 0.25^2 \times 0.75^2 \times 0.3$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.9/P(S) = 0.75$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.1/P(S) = 0.25$
    - ★ $P(S) = 0.25^3 \times 0.75^2 \times 0.9 + 0.75^3 \times 0.25^2 \times 0.1 = 0.25^2 \times 0.75^2 \times 0.3$
  - Estimator 1: Since $P(p = 0.25|S) > P(p = 0.75|S)$, we estimate $\hat{p} = 0.25$

Example 2: Bernoulli(p)

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Suppose that $p \sim \{\overset{0.9}{0.25}, \overset{0.1}{0.75}\}$
- Samples: $1, 0, 1, 1, 0$
  - Notation: $S = (X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0)$
  - Estimate using Bayes' rule
    - ★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.9/P(S) = 0.75$
    - ★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.1/P(S) = 0.25$
    - ★ $P(S) = 0.25^3 \times 0.75^2 \times 0.9 + 0.75^3 \times 0.25^2 \times 0.1 = 0.25^2 \times 0.75^2 \times 0.3$
  - Estimator 1: Since $P(p = 0.25|S) > P(p = 0.75|S)$, we estimate $\hat{p} = 0.25$
  - Estimator 2: Posterior mean, $\hat{p} = 0.25\, P(p = 0.25|S) + 0.75\, P(p = 0.75|S) = 0.375$

Bayesian estimation

$$X_1, \ldots, X_n \sim \text{iid } X, \text{parameter } \Theta$$

Capital theta

- Prior distribution of Θ: $\Theta \sim \underbrace{f_\Theta(\theta)}_{\text{PMF (or) PDF}}$

Bayesian estimation

$$X_1, \ldots, X_n \sim \text{iid } X, \text{parameter } \Theta$$

- Prior distribution of Θ: $\Theta \sim f_\Theta(\theta)$
- Samples: $x_1, \ldots, x_n$, Notation: $S = (X_1 = x_1, \ldots, X_n = x_n)$
- Bayes' rule: posterior $\propto$ likelihood $\times$ prior

$$\underbrace{P(\Theta = \theta | S)}_{\text{Posterior}} = \underbrace{P(S | \Theta = \theta)}_{\text{likelihood}} \underbrace{f_\Theta(\theta)}_{\text{prior}} / P(S)$$

Normalization → does not depend on θ (integrated over θ)

Discrete $P(S) = \sum_{\theta} P(S|\theta = \theta) f_\theta(\theta)$

Continuous $P(S) = \int_{\theta} P(S|\theta = \theta) f_\theta(\theta)\, d\theta$

Bayesian estimation

$$X_1, \ldots, X_n \sim \text{iid } X, \text{parameter } \Theta$$

- Prior distribution of Θ: $\Theta \sim f_\Theta(\theta)$
- Samples: $x_1, \ldots, x_n$, Notation: $S = (X_1 = x_1, \ldots, X_n = x_n)$
- Bayes' rule: posterior $\propto$ likelihood $\times$ prior

$$P(\Theta = \theta | S) = \underline{P(S|\Theta = \theta) f_\Theta(\theta)} / P(S)$$

$\arg\max_\theta P(\Theta = \theta | S)$

- Estimation using "posterior" probability
  - Posterior mode: $\hat{\theta} = \arg\max_\theta P(S|\Theta = \theta) f_\Theta(\theta)$
  - Posterior mean: $\hat{\theta} = E[\Theta|S]$, mean of posterior distribution
    - ★ $(\Theta|S)$ may be a known distribution, and its mean might become a simple formula in some cases

Discrete: $\sum_\theta \theta P(\Theta = \theta | S)$

Continuous: $\int_\theta \theta \left(\text{posterior density}\right) d\theta$

Meaning of prior distribution

- Prior distribution
  - Captures what we might know about the parameter
  - This could be using some scientific model or expert opinion
- Posterior $\propto$ Likelihood of samples $\times$ Prior
  - Intuitively understood as incorporating "data" into prior
  - Useful in modeling
- What if we do not know anything?
  - You can choose a flat prior, uniform over the entire range
- Lots of debates between frequentists vs Bayesians
  - Search "frequestist vs Bayesian"

Section 2

Choice of prior and examples

How to pick prior?

- Flat, uninformative
  - Nearly flat over the interval in which the parameter takes value
  - This usually reduces to something close to maximum likelihood
- Conjugate priors
  - Pick a prior so that the posterior is in the same class as prior
  - Examples
    - ★ Prior: Normal and Posterior: Normal
    - ★ Prior: Beta and Posterior: Beta
- Informative priors
  - This needs some justification from the domain of the problem
  - Parameterize the prior so that its flatness can be controlled

Bernoulli(p) samples with uniform prior

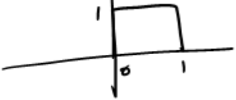
$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

bold face

"random variable"

- Prior $\mathbf{p} \sim \text{Uniform}[0, 1]$, continuous distribution

pdf



Bernoulli(p) samples with uniform prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

- Prior $\mathbf{p} \sim$ Uniform$[0, 1]$, continuous distribution
- Samples: $x_1, \ldots, x_n$
- Posterior: $\mathbf{p}|(X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto \overbrace{P(X_1 = x_1, \ldots, X_n = x_n|\mathbf{p} = p)}^{\text{likelihood}}\overbrace{f_{\mathbf{p}}(p)}^{\text{prior density}}$
    $f_{\mathbf{p}}(p) = 1,\ 0 \le p \le 1$
  - Posterior density $\propto p^w(1-p)^{n-w}$, $0 \le p \le 1$
    - ★ $w = x_1 + \cdots + x_n$: number of 1s in samples

$$\text{pdf} \propto \underbrace{p^w(1-p)^{n-w},\ 0 < p < 1}_{\text{Beta distribution}}$$

Bernoulli(p) samples with uniform prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

Beta(α, β)

pdf $\propto x^{\alpha-1}(1-x)^{\beta-1}$

$0 < x < 1$

mean $= \frac{\alpha}{\alpha+\beta}$

- Prior $\mathbf{p} \sim \text{Uniform}[0,1]$, continuous distribution
- Samples: $x_1, \ldots, x_n$
- Posterior: $\mathbf{p} | (X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \mathbf{p} = p) f_{\mathbf{p}}(p)$
  - Posterior density $\propto p^w (1-p)^{n-w}$, $0 \leq p \leq 1$
    - ★ $w = x_1 + \cdots + x_n$: number of 1s in samples
- Posterior density: Beta$(\overbrace{w+1}^{\alpha}, \overbrace{n-w+1}^{\beta})$
  - Posterior mean $= \frac{w+1}{\underbrace{w+1}_{\alpha}+\underbrace{n-w+1}_{\beta}} = \frac{w+1}{n+2} = \frac{x_1+\cdots+x_n+1}{n+2}$

$\hat{p}$

Bernoulli(p) samples with uniform prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

- Prior $\mathbf{p} \sim \text{Uniform}[0, 1]$, continuous distribution
- Samples: $x_1, \ldots, x_n$
- Posterior: $\mathbf{p} | (X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \mathbf{p} = p) f_{\mathbf{p}}(p)$
  - Posterior density $\propto p^w (1-p)^{n-w}$, $0 \leq p \leq 1$
    - $w = x_1 + \cdots + x_n$: number of 1s in samples
- Posterior density: $\text{Beta}(w+1, n-w+1)$
  - Posterior mean $= \frac{w+1}{w+1+n-w+1} = \frac{w+1}{n+2} = \frac{x_1+\cdots+x_n+1}{n+2}$

$$\hat{p} = \frac{X_1 + \cdots + X_n + 1}{n+2}$$

$$\hat{p}_{ML} = \frac{X_1 + \cdots + X_n}{n}$$

Bernoulli(p) samples with beta prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

- Prior $\mathbf{p} \sim \text{Beta}(\alpha, \beta)$, continuous distribution
  - $f_{\mathbf{p}}(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$, $0 \leq p \leq 1$

Bernoulli(p) samples with beta prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

- Prior $\mathbf{p} \sim \text{Beta}(\alpha, \beta)$, continuous distribution
  - $f_{\mathbf{p}}(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$, $0 \leq p \leq 1$
- Samples: $x_1, \ldots, x_n$
- Posterior: $\mathbf{p}|(X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \mathbf{p} = p) f_{\mathbf{p}}(p)$
  - Posterior density $\propto p^{w+\alpha-1}(1-p)^{n-w+\beta-1}$, $0 \leq p \leq 1$
    - ★ $w = x_1 + \cdots + x_n$: number of 1s in samples

Bernoulli(p) samples with beta prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

- Prior $\mathbf{p} \sim \text{Beta}(\alpha, \beta)$, continuous distribution
  - $f_{\mathbf{p}}(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$, $0 \le p \le 1$
- Samples: $x_1, \ldots, x_n$
- Posterior: $\mathbf{p}|(X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \mathbf{p} = p) f_{\mathbf{p}}(p)$
  - Posterior density $\propto p^{w+\alpha-1}(1-p)^{n-w+\beta-1}$, $0 \le p \le 1$
    - ★ $w = x_1 + \cdots + x_n$: number of 1s in samples
- Posterior density: $\text{Beta}(w+\alpha, n-w+\beta)$
  - Posterior mean $= \frac{w+\alpha}{w+\alpha+n-w+\beta} = \frac{w+\alpha}{n+\alpha+\beta} = \frac{x_1+\cdots+x_n+\alpha}{n+\alpha+\beta}$

Bernoulli(p) samples with beta prior

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(\mathbf{p})$$

- Prior $\mathbf{p} \sim \text{Beta}(\alpha, \beta)$, continuous distribution
  - $f_{\mathbf{p}}(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$, $0 \leq p \leq 1$
- Samples: $x_1, \ldots, x_n$
- Posterior: $\mathbf{p} | (X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \mathbf{p} = p) f_{\mathbf{p}}(p)$
  - Posterior density $\propto p^{w+\alpha-1}(1-p)^{n-w+\beta-1}$, $0 \leq p \leq 1$
    - ★ $w = x_1 + \cdots + x_n$: number of 1s in samples
- Posterior density: $\text{Beta}(w+\alpha, n-w+\beta)$
  - Posterior mean $= \frac{w+\alpha}{w+\alpha+n-w+\beta} = \frac{w+\alpha}{n+\alpha+\beta} = \frac{x_1+\cdots+x_n+\alpha}{n+\alpha+\beta}$

$$\hat{p} = \frac{X_1 + \cdots + X_n + \alpha}{n + \beta}$$

Observations for Beta prior

- Prior: Beta(α, β)
  - $\alpha, \beta \geq 0$
  - PDF $\propto p^{\alpha-1}(1-p)^{\beta-1}$, $0 < p < 1$
  - How to pick α, ~~beta~~ β?
- $\alpha = \beta = 1$: Uniform$[0, 1]$
  - Flat prior
  - Estimate close to, but not equal to, Maximum-Likelihood
- $\alpha = \beta = 0$
  - Estimate coincides with Maximum-Likelihood
- $\alpha = \beta$ (about $p = 1/2$)
  - Symmetric prior
- α, β may depend on n, the number of samples
  - $\alpha = \beta = \sqrt{n}/2$ is an *interesting* choice

Normal samples with unknown mean and known variance

$$X_1, \ldots, X_n \sim \text{iid Normal}(\underbrace{M}_{\text{r.v.}}, \sigma^2)$$

- Prior $M \sim \text{Normal}(\mu_0, \sigma_0^2)$, continuous distribution
  - $f_M(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$

Normal samples with unknown mean and known variance

continuous

$$X_1, \ldots, X_n \sim \text{iid Normal}(M, \sigma^2)$$

- Prior $M \sim \text{Normal}(\mu_0, \sigma_0^2)$, continuous distribution
  - $f_M(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$
- Samples: $x_1, \ldots, x_n$, Sample mean: $\overline{x} = (x_1 + \cdots + x_n)/n$
- Posterior: $M|\underbrace{(X_1 = x_1, \ldots, X_n = x_n)}_{S}$ is continuous
  - Posterior density $\propto f(X_1 = x_1, \ldots, X_n = x_n|M = \mu)f_M(\mu)$ likelihood prior
  - Posterior density $\propto \exp(-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}) \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$

(M|S) of $M=\mu|S$

$\mu_0, \sigma_0, x_1, \ldots, x_n$: constants

$$f(X_1=x_1, \ldots, X_n=x_n | M=\mu) \propto \underbrace{e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \cdots e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}}$$

independent

$f(X_1=x_1|M=\mu) \cdot f(X_2=x_2|M=\mu) \ldots$

Normal samples with unknown mean and known variance

$$X_1, \ldots, X_n \sim \text{iid Normal}(M, \sigma^2)$$

- Prior $M \sim \text{Normal}(\mu_0, \sigma_0^2)$, continuous distribution
  - $f_M(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$ Conjugate
- Samples: $x_1, \ldots, x_n$, Sample mean: $\overline{x} = (x_1 + \cdots + x_n)/n$
- Posterior: $M|(X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto f(X_1 = x_1, \ldots, X_n = x_n | M = \mu) f_M(\mu)$
  - Posterior density $\propto \exp(-\frac{(x_1-\mu)^2 + \cdots + (x_n-\mu)^2}{2\sigma^2}) \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}) \propto e^{-\frac{(\mu-\bar{\mu})^2}{2\bar{\sigma}^2}}$

    → Exercise to show that posterior is normal
- Posterior density: Normal
  - $\underbrace{\text{Posterior mean}}_{\bar{\mu}} = \overline{x} \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \mu_0 \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}$

    $\overline{x}$: sample mean; μ_0: prior mean

Normal samples with unknown mean and known variance

$$X_1, \ldots, X_n \sim \text{iid Normal}(M, \sigma^2)$$

- Prior $M \sim \text{Normal}(\mu_0, \sigma_0^2)$, continuous distribution
  - $f_M(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$
- Samples: $x_1, \ldots, x_n$, Sample mean: $\overline{x} = (x_1 + \cdots + x_n)/n$
- Posterior: $M|(X_1 = x_1, \ldots, X_n = x_n)$ is continuous
  - Posterior density $\propto f(X_1 = x_1, \ldots, X_n = x_n | M = \mu) f_M(\mu)$
  - Posterior density $\propto \exp(-\frac{(x_1-\mu)^2 + \cdots + (x_n-\mu)^2}{2\sigma^2}) \exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$
- Posterior density: Normal
  - Posterior mean $= \overline{x} \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2} + \mu_0 \frac{\sigma^2}{n\sigma_0^2+\sigma^2}$

$$\hat{\mu}_{ML} = \frac{X_1 + \cdots + X_n}{n}$$

$$\hat{\mu} = \frac{X_1 + \cdots + X_n}{n} \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \mu_0 \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$\rightarrow 1$ as $n \to \infty$ (or) $\sigma_0 \to \infty$

$\rightarrow 0$ as $n \to \infty$ (or) $\sigma_0 \to \infty$

Oberservations for Normal prior

- Prior: Normal(μ_0, σ_0^2)
  - How to pick μ_0 and σ_0?
- Estimate is combination of data and prior
  - Prior is "updated" using data to get posterior
- If *n* is very large, $\hat{\mu} \to$ sample mean
  - Data dominates the estimate
  - Prior plays no significant role
- If *n* is small, prior contributes significantly to the estimate
  - Prior needs to have some justification when *n* is small
- If variance of prior is large compared to variance of samples, prior tends to be flat or uninformative
  - Choice of variance of prior is important

Section 3

Problems: Finding estimators

Problem 1

Suppose X is a discrete random variable taking values $\{0, 1, 2, 3\}$ with respective probabilities $\{2\theta/3, \theta/3, 2(1-\theta)/3, (1-\theta)/3\}$, where $0 \le \theta \le 1$ is a parameter. Consider the estimation of θ from samples
$2, 2, 0, 3, 1, 3, 2, 1, 2, 3$.
$\bar{x} = 1.9$

- Find the method of moments and maximum likelihood estimates.
- Using a Uniform$[0, 1]$ prior, find the posterior distribution and mean.

Method of moments

$$E[X] = \frac{2\theta}{3}\times 0 + \frac{\theta}{3}\times 1 + \frac{2}{3}(1-\theta)\times 2 + \frac{1}{3}(1-\theta)3$$
$$= \frac{\theta}{3} + \frac{4}{3} - \frac{4\theta}{3} + 1 - \theta$$
$$= \frac{7}{3} - 2\theta$$

Samples: $x_1, \ldots, x_n$

MM: Sample mean, $\bar{x} = \dfrac{x_1 + \cdots + x_n}{n}$

$$\bar{x} = \frac{7}{3} - 2\hat{\theta}_{MM}$$
$$\hat{\theta}_{MM} = \frac{1}{2}\left(\frac{7}{3} - \bar{x}\right) = \frac{1}{2}\left(\frac{7}{3} - 1.9\right)$$
$$= \frac{1.3}{6} = \frac{13}{60}$$

Maximum Likelihood

n_0: no. of 0's in samples
n_1: " " 1's " "
n_2, n_3 ....

$$L = P(X_1 = x_1) P(X_2 = x_2) \cdots P(X_n = x_n)$$

$$L = \left(\frac{2\theta}{3}\right)^{n_0} \left(\frac{\theta}{3}\right)^{n_1} \left(\frac{2}{3}(1-\theta)\right)^{n_2} \left(\frac{1}{3}(1-\theta)\right)^{n_3}$$
$$= \frac{2^{n_0+n_2}}{3^n}\, \theta^{n_0+n_1} (1-\theta)^{n_2+n_3}$$

$$\log L = \log\left(\frac{2^{n_0+n_2}}{3^n}\right) + (n_0+n_1)\log\theta + (n_2+n_3)\log(1-\theta)$$

Differentiate w.r.t. θ & equate to 0:

$$(n_0+n_1)\frac{1}{\theta} + (n_2+n_3)\frac{1}{1-\theta}(-1) = 0$$

Solution: $\hat{\theta}_{ML} = \dfrac{n_0+n_1}{n} = \dfrac{3}{10}$

Working

Bayesian

Prior: $\Theta \sim \text{Uniform}[0,1]$, $f_\Theta(\theta) = 1,\ 0 \le \theta \le 1$

$$\text{Posterior} \propto \theta^{n_0+n_1}(1-\theta)^{n_2+n_3} \cdot \underbrace{f_\Theta(\theta)}_{1} = \theta^{n_0+n_1}(1-\theta)^{n_2+n_3}$$

$$\sim \text{Beta}(n_0+n_1+1,\ n_2+n_3+1)$$

$$\hat{\theta}_{\text{Bayes}} = \text{Posterior mean} = \frac{n_0+n_1+1}{\underbrace{n_0+n_1+1+n_2+n_3+1}_{n}} = \frac{n_0+n_1+1}{n+2}$$

$$= \frac{3+1}{10+2} = \frac{4}{12}$$

Problem 2

Consider n iid samples from a Geometric(p) distribution.

- Find the method of moments estimate.
- Find the MLE.
- Using a Uniform$[0,1]$ prior, find the posterior distribution and mean.

$X \sim \text{Geometric}(p)$

Samples: $x_1, x_2, \ldots, x_n$

$P(X_1 = x_1) = (1-p)^{x_1-1} p$

Method of moments

$E[X] = 1/p$

$\bar{x} = 1/p$

$p = \frac{1}{\bar{x}} = \frac{n}{x_1+\cdots+x_n}$

$\hat{p}_{mm} = \frac{n}{X_1+\cdots+X_n}$

Maximum Likelihood

$L = (1-p)^{x_1-1} p \cdot (1-p)^{x_2-1} p \cdots (1-p)^{x_n-1} p$

$= p^n (1-p)^{x_1+\cdots+x_n-n}$

$\log L = n \log p + (x_1+\cdots+x_n-n) \log(1-p)$

$\frac{n}{p} + (x_1+\cdots+x_n-n) \frac{1}{1-p}(-1) = 0$

$p = \frac{n}{x_1+\cdots+x_n}$

$\hat{p}_{ML} = \frac{n}{X_1+\cdots+X_n}$

Working

Bayesian

Prior: $p \sim \text{Uniform}[0,1]$

$$\text{Posterior} \propto \underbrace{p^n (1-p)^{x_1+\cdots+x_n-n}}_{\text{likelihood}} \cdot \underbrace{1}_{\text{prior}}$$

$$\sim \text{Beta}(n+1, x_1+\cdots+x_n-n+1)$$

Beta(α, β): Expected value $= \dfrac{\alpha}{\alpha+\beta}$

$$\hat{p}_{\text{Bayes}} = \text{Posterior mean} = \frac{n+1}{\cancel{n}+1+x_1+\cdots+x_n-\cancel{n}+1} = \frac{n+1}{x_1+\cdots+x_n+2}$$

Problem 3

Consider n iid samples from a Poisson(λ) distribution.

- Find the method of moments estimate.
- Find the MLE.
- Using a Gamma$[\alpha, \beta]$ prior, find the posterior distribution and mean.

MM

$$E[X] = \lambda = \bar{x}$$

$$\hat{\lambda}_{ML} = \frac{X_1 + \cdots + X_n}{n}$$

ML Samples: $x_1, \ldots, x_n$ $\quad P(X_i = x_i) = e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!}$

$$L = e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} \propto e^{-n\lambda} \lambda^{x_1 + \cdots + x_n}$$

$$\log L = -n\lambda + (x_1 + \cdots + x_n) \log \lambda$$

$$-n + (x_1 + \cdots + x_n) \cdot \frac{1}{\lambda} = 0$$

$$\hat{\lambda}_{ML} = \frac{X_1 + \cdots + X_n}{n}$$

$X \sim \text{Gamma}[\alpha, \beta]$: $f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x}$ $\quad \Gamma(t)$: Gamma function

$E[X] = \alpha/\beta \quad E[X^2] = \frac{\alpha^2}{\beta^2} + \alpha/\beta^2$

Working

Bayes: Prior: $\lambda \sim \text{Gamma}[\alpha, \beta]$, pdf $\propto \lambda^{\alpha-1} e^{-\beta\lambda}$

$$\text{Posterior} \propto \underbrace{e^{-n\lambda} \lambda^{x_1+\cdots+x_n}}_{\text{likelihood}} \cdot \underbrace{\lambda^{\alpha-1} e^{-\beta\lambda}}_{\text{prior}}$$

Gamma: Conjugate prior for Poisson mean

$$= \lambda^{x_1+\cdots+x_n+\alpha-1} \cdot e^{-(\beta+n)\lambda}$$

$$\sim \text{Gamma}[\underbrace{x_1+\cdots+x_n+\alpha}, \underbrace{\beta+n}]$$

$$\hat{\lambda}_{\text{Bayes}} = \underset{\text{mean}}{\text{Posterior}} = \frac{x_1+\cdots+x_n+\alpha}{n+\beta}$$

Section 4

Problems: Fitting distributions

Problem 1

Fit a Poisson distribution to the following frequency data on number of vehicles (n) making a right turn at an intersection in a 3-minute interval. Find an approximate 95% confidence interval for the sample mean using a normal approximation for the sampling distribution.

n	Frequency	n	Frequency
0	14	7	14
1	30	8	10
2	36	9	6
3	68	10	4
4	43	11	1
5	43	12	1
6	30	13+	0

$\hat{\lambda} = 3.893$

X = Number of Vehicles making a right turn in a 3-minute interval

$X \sim \text{Poisson}(\lambda)$

Samples: (after re-ordering) $\overbrace{0, 0, \ldots, 0}^{14}, \overbrace{1, 1, \ldots, 1}^{30}, \overbrace{2, 2, \ldots, 2}^{36}, \ldots\ldots, \overbrace{12}^{1}$

$\sim$ iid X

$\hat{\lambda}$ = Sample mean $\rightarrow$ how many samples? $14+30+36+\ldots$

$\rightarrow$ Sum of samples: $14\times 0 + 30\times 1 + 36\times 2 + \ldots$

Problem 2

Fit a Geometric distribution to the following frequency data on number of hops (n) between flights of birds. Find an approximate 95% confidence interval.

n	Frequency	n	Frequency
1	48	7	4
2	31	8	2
3	20	9	1
4	9	10	1
5	6	11	2
6	5	12	1

Working

$$\hat{p} = \frac{n}{X_1+\cdots+X_n}$$

$$\text{Error} = \hat{p} - p \sim N\left(0.002, \frac{(0.025)^2}{130}\right)$$

$$E[\text{Error}] = E[\hat{p}] - p$$

$$= \underbrace{E\left[\frac{n}{X_1+\cdots+X_n}\right]}_{??} - p$$

Bootstrap

$\hat{p} = 0.35$

$X_1, \ldots, X_n \sim$ iid Geometric(0.35)

$n = 130$

$\rightarrow$ N samplings:

$N = 1000$

1^{st} sampling: $x_{11}, x_{12}, \ldots, x_{1n} \rightarrow \hat{p}_1$

2^{nd} sampling: $x_{21}, x_{22}, \ldots, x_{2n} \rightarrow \hat{p}_2$

$\vdots$

$\hat{p}_{1000}$

$\sim$ Samples of $\hat{p}$

$$E[\hat{p}] = \frac{\hat{p}_1 + \cdots + \hat{p}_{1000}}{1000}$$

Problem 3

Data from a genetic experiment and expected distribution in terms of an unknown parameter θ are given in the following table.

Type	Frequency	Theory
1	1997	$0.25(2+\theta)$
2	906	$0.25(1-\theta)$
3	904	$0.25(1-\theta)$
4	32	0.25θ

PMF for X

$0<\theta<1$

Samples: $\overbrace{1,1,\ldots,1}^{1997},\overbrace{2,\ldots,2}^{906},\ldots$

Find the ML estimate for θ.

$$L = \left(\frac{1}{4}(2+\theta)\right)^{1997}\left(\frac{1}{4}(1-\theta)\right)^{906}\left(\frac{1}{4}(1-\theta)\right)^{904}\left(\frac{1}{4}\theta\right)^{32}$$

$$\log L = (\text{const}) + 1997\log(2+\theta) + 1810\log(1-\theta) + 32\log\theta$$

$$\frac{1997}{2+\theta} - \frac{1810}{1-\theta} + \frac{32}{\theta} = 0 \longrightarrow \text{quadratic equation}$$

Working

Section 5

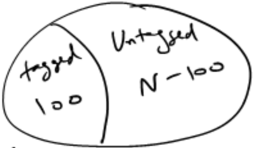
Problems: Model and estimation

Problem 1

To find the size of an animal population, 100 animals are captured and tagged. Some time later, another 50 animals are captured, and 20 of them were found to be tagged. How will you estimate the population size? What are your assumptions?

Capture-recapture

N animals



Recapture → randomly pick 50

M = # of tagged animals during recapture ~ Hypergeometric

$$P(M = 20) = \frac{\binom{100}{20}\binom{N-100}{30}}{\binom{N}{50}}$$

$$E[M] = \frac{50 \times 100}{N}$$

→ One sample from the Hypergeometric distribution

Sample mean = 20

Method of moments

$$20 = \frac{50 \times 100}{N}$$

$$N = \frac{50 \times 100}{20} = 250$$

Working

Problem 2

In a new machine, suppose that, out of 10 produced items, no item was found to be defective. How will you estimate the fraction of defective items produced by the new machine? From data collected from other similar machines, the average of the fraction of defective items was found to be 10%, and the actual fraction was between 5% and 15% in 95% of the cases.

$X_1, X_2, \ldots, X_n \sim$ iid Bernoulli(p) Goal: find p

$X_i = \begin{cases} 1 & \text{if Item } i \text{ is defective} \\ 0 & \text{otherwise} \end{cases}$ $P(X_i = 1) = p$

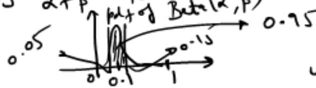
$\underline{\underline{n=10}}$ $\quad 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$: $\hat{p}_{MM} = \hat{p}_{ML} = 0$

Bayesian Prior: Beta (α, β) $\quad \hat{p}_{Bayes} = \dfrac{X_1 + \cdots + X_n + \alpha}{n + \beta} = \dfrac{\alpha}{10+\beta}$

How to pick α, β? $\longrightarrow$ Non-informative: $\alpha = \beta = 1$ $\quad \hat{p} = \dfrac{1}{11} \approx 9\%$

flat

$E[\text{Prior}] = \dfrac{\alpha}{\alpha+\beta} = \dfrac{1}{10} \Rightarrow \beta = 9\alpha$



Find α & β that satisfies this condition.

$\alpha = 13.3, \beta = 13.3 \times 9$

$\hat{p} = \dfrac{13.3}{10 + 13.3 \times 9} = 0.1025 \to 10.25\%$

Working

Problem 3

Colab sheet

$X \sim \text{Gamma}[\alpha, \beta]$: $\quad f_X(x) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$

$E[X] = \dfrac{\alpha}{\beta}$