HW 1.

Bayesian Inference

STA 427/527, Fall 2019, Xin Wang

Probability

- An **experiment** is the process by which an observation/outcome is made.
- An **outcome** of an experiment is any possible observation of that experiment. Outcomes are often called sample points.
- The **sample space** of an experiment is the set consisting of all possible sample points. (S)
- An **event** is a set of outcomes of an experiment, or a subset of S. That is a set of sample points.

"flip a coin" : outcome: "head" "tail"

$S = \{head, tail\}$

"toss a die" outcome: 1, 2, 3, 4, 5, 6.

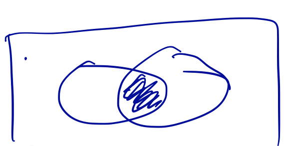
$S = \{1, 2, 3, 4, 5, 6\}$

event : even points $A = \{2, 4, 6\}$

- **Probability:** Suppose S is a sample space, A is a subset of S. A probability measure $P(\cdot)$ is a function that maps events in S to real numbers.

1. For any event A, $P(A) \geq 0$
2. $P(S) = 1$
3. For any countable collection $A_1\ A_2, \ldots$ (finite sequence of events) of pairwise mutually exclusive events.
   $(A_i \cap A_j = \phi \quad i \neq j)$
   $P(A_1 \cup A_2 \cdots) = P(A_1) + P(A_2) + \cdots$

- Consider $P(A \cap B)$



$$\boxed{P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)}$$

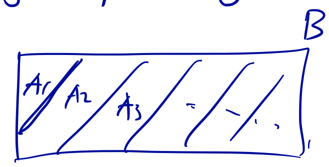
conditional probability: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ $\quad P(B|A) = \dfrac{P(A \cap B)}{P(A)}$

$P(B|A)$

- **Bayes' theorem**

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \boxed{\frac{P(B|A)\,P(A)}{P(B)}}$$

Extension : $B = A_1 \cup A_2 \cdots \cup A_k \qquad A_i \cap A_j = \phi \quad i \neq j$



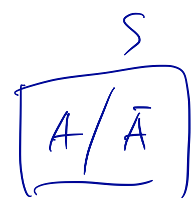
$$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{P(B)}$$

$$= \frac{P(B|A_i)\,P(A_i)}{\Sigma P(B\cap A_i)} = \frac{P(B|A_i)\,P(A_i)}{\Sigma P(B|A_i)\,P(A_i)}$$

- **Example:** The prevalence of heart disease in a certain population is 10%. A screening test for heart disease has 99% sensitivity and 90% specificity, where sensitivity measures the probability of positive that is correctly identified as such and specificity measures the probability of negative that is correctly identified as such. Suppose everyone in the population is given the screening test. What is the probability that one individual with positive test result actually has heart disease?

A: has heart disease. B: postive test result.

$P(A|B) = ?$



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

$P(A) = 0.1 \qquad P(\bar{A}) = 0.9$

$P(B|A) = 0.99 \qquad P(\bar{B}|\bar{A}) = 0.9$

$$P(B) = P(B\cap A) + P(B\cap \bar{A})$$

$$= P(B|A)\,P(A) + P(B|\bar{A})\,P(\bar{A})$$

$P(B|\bar{A}) = 1 - P(\bar{B}|\bar{A})$

$= 1 - 0.9 = 0.1$

$$P(A|B) = \frac{0.99 \times 0.1}{0.99 \times 0.1 + 0.1 \times 0.9}$$

$$= 0.524$$

Frequentist vs Bayesian

- Suppose θ is the unknown parameter, and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ are observations.
- Frequentist

① parameter θ is a fixed value

② use y to estimate θ, denoted as $\hat{\theta}$.

③ $\hat{\theta}$ is a random variable $\hat{\theta}(y)$

④ The inference is based on that data are "repeatable".

Example: $Y_i \sim N(\mu, \sigma^2), i=1,\ldots,n$, σ^2 is unknown.

observations $y_1, y_2 \ldots y_n$.

Construct a confidence interval for μ, $100(1-\alpha)\%$ $\alpha = 0.05$

$\bar{y} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$, $[\bar{y} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{y} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}]$

95% CI.

- Bayesian

① θ is a random variable (prior)

② how can make inference about θ?

observations $\underset{\sim}{y} = (y_1 \ldots y_n)^T$

we can find the distribution of θ conditional on the given data set $\underset{\sim}{y}$ (fixed) [posterior]

Bayes' Theorem

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

$\theta \quad \underset{\sim}{y}$

$$\boxed{P(\theta|\underset{\sim}{y}) = \frac{P(\underset{\sim}{y}|\theta)\,\pi(\theta)}{P(\underset{\sim}{y})}}$$

  - Steps in Bayesian
    1. Data distribution (Likelihood)

       $P(\underset{\sim}{y}|\theta)$: specify a model

    2. Prior:

       prior knowledge of θ, $\pi(\theta)$

    3. Posterior

$$P(\theta|\underset{\sim}{y}) = \frac{P(y|\theta)\,\pi(\theta)}{P(\underset{\sim}{y})}$$

$$P(\underset{\sim}{y}) = \int_{\Theta} P(\underset{\sim}{y},\theta)\,d\theta = \int_{\Theta} P(\underset{\sim}{y}|\theta)\,\pi(\theta)\,d\theta \longrightarrow P(\underset{\sim}{y},\theta) = P(\underset{\sim}{y}|\theta)\,\pi(\theta)$$

In practice: $\boxed{P(\theta|\underset{\sim}{y}) \propto P(\underset{\sim}{y}|\theta)\,\pi(\theta)}$

Likelihood. $Y_i \sim N(\mu, \sigma^2)$, σ is known.

$y_1, \cdots y_n$.

$$L(\mu|\underset{\sim}{y}) = \prod_{i=1}^{n} P(y_i|\mu) = P(\underset{\sim}{y}|\mu)$$

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)}$$

- **Example:** A novelty coin company produces coins with varying levels of bias. For an individual coin, the probability of spinning "heads" is θ, where θ is drawn from a Uniform(0, 1) distribution. Consider the following thought experiment. Suppose it were possible to spin each coin in the population 100 times. Let Y represent the number of "heads" resulting from a coin's 100 spins.

1. Find the distribution of Y. That is, what is the distribution of the numbers of "heads" observed in 100 spins?

Flip a coin for 100 times.

Y is the # of heads among 100 times

$p(y)$? $p(y) = \int_0^1 p(y|\theta)\,\pi(\theta)\,d\theta \longrightarrow p(y,\theta)$

[Data distribution] $p(y|\theta) = \binom{100}{y}\theta^y(1-\theta)^{100-y}$ $Y|\theta \sim \text{Binomial}(100, \theta)$

[prior] $\pi(\theta) = 1$ $0<\theta<1$

$$p(y) = \int_0^1 \binom{100}{y}\theta^y(1-\theta)^{100-y}\cdot 1\,d\theta$$

$$= \binom{100}{y}\int_0^1 \theta^y(1-\theta)^{100-y}\,d\theta$$

$$= \binom{100}{y}\int_0^1 \theta^{(y+1)-1}(1-\theta)^{(101-y)-1}\,d\theta$$

$\alpha = y+1$ $\beta = 101-y$

$$= \binom{100}{y}\cdot\frac{\Gamma(y+1)\,\Gamma(101-y)}{\Gamma(102)}$$

$$= \frac{100!}{\cancel{y!(100-y)!}}\;\frac{\cancel{y!\,(100-y)!}}{101!}$$

$$= \frac{1}{101}, \quad y = 0, 1, 2, \ldots, 100$$

$X \sim \text{Beta}(\alpha, \beta)$ $0<x<1$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

$$\int_0^1 f(x)\,dx = 1$$

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,dx = 1$$

$$\Rightarrow \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = 1$$

$$\Rightarrow \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$\Gamma(x) = (x-1)!$

If x is a positive integer.

$\Gamma(0) = 1$.

2. Now think about the case that exactly 50 "heads" are observed in the coins we observed. Among these coins, determine the distribution of θ the actual probabilities of "heads".

$$p(\theta|y) = \frac{P(y|\theta)\,\pi(\theta)}{P(y)}$$

$$= \frac{\binom{100}{y}\theta^{y}(1-\theta)^{100-y}\cdot 1}{\frac{1}{101}}$$

$$= 101\cdot\frac{100!}{y!\,(100-y)!}\,\theta^{y}(1-\theta)^{100-y}$$

$$= \frac{101!}{y!\,(100-y)!}\,\theta^{y}(1-\theta)^{100-y}$$

$\Gamma(102) = 101!$

$\Gamma(y+1) = y!$

$\Gamma(101-y) = (100-y)!$

$$= \frac{\Gamma(102)}{\Gamma(y+1)\,\Gamma(101-y)}\,\theta^{\boxed{y+1}-1}(1-\theta)^{\boxed{101-y}-1}$$

pdf of $x \sim Beta(\alpha,\beta)$ $\quad E(x) = \frac{\alpha}{\alpha+\beta}$

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot x^{\alpha-1}(1-x)^{\beta-1} \quad 0<x<1$$

$$\theta|y \sim Beta(y+1,\ 101-y)$$

$y=50 \quad \theta|y=50 \sim Beta(51,\ 51)$

In practice: $\boxed{p(\theta|y) \propto p(y|\theta)\,\pi(\theta)}$

$$P(\theta|y) \propto \binom{100}{y}\theta^{y}(1-\theta)^{100-y}\cdot 1$$

$$\propto \boxed{\theta^{y}(1-\theta)^{100-y}}$$

$$= \theta^{\underline{y+1}-1}(1-\theta)^{\underline{101-y}-1}$$

kernel of a beta distribution. with two shape parameters $(y+1,\ 101-y)$

$$\Rightarrow \theta|y \sim Beta(y+1,\ 101-y)$$

posterior mean. $E(\theta|y) = \frac{y+1}{y+1+101-y} = \frac{y+1}{102}$

Data distribution: $P(y|\theta) = \binom{100}{y}\cdot\theta^{y}(1-\theta)^{100-y}$. $\quad$ mle (point estimator).

$$\ell(\theta|y) = \log\binom{100}{y} + y\log\theta + (100-y)\log(1-\theta) \Rightarrow \hat{\theta}_{MLE} = \frac{y}{100}$$