MIT Probability Course Summarized

Zaid Alyafeai

March 2018

Contents

1 Models and Axioms

Sample Space: All possible outcomes of a certain experiment.

Event: Subset of the sample space.

Here are the axioms of probability

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. Given a countable set of disjoint sequences of events $\{A\}_n$ we have

$$P(\cup A_i) = \sum P(A_i)$$

2 Conditioning and Bays Rule

We define conditional probability as

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

and

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

Note that conditional probabilities are probability distributions that follow the axioms of probability.

Exercise. Let A be the event that there is a plane in the sky and B the event that we get an alarm from a certain radar that there is a plane. The probability that there is a plane and the radar alarms

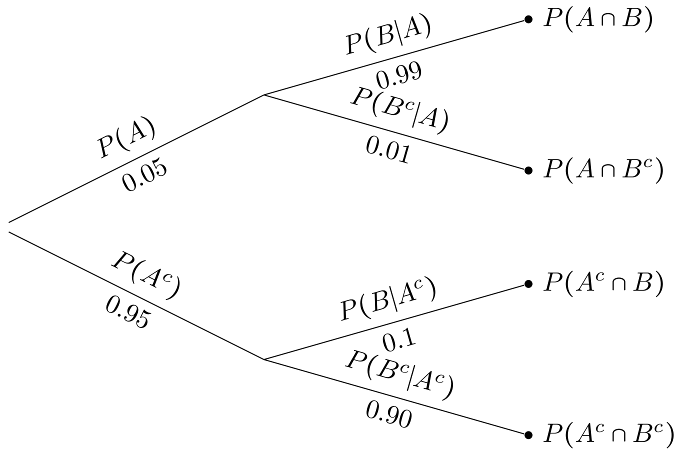


Figure 1: Tree Diagram

$$P(A \cap B) = P(A)P(B|A) = 0.05 \times 0.99 = 0.0495$$

The probability that there is a plane and the radar doesn't work

$$P(A \cap B^c) = P(A)P(B^c|A) = 0.05 \times 0.01 = 0.0005$$

We find the total probability of B as

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(A)P(B|A) + P(A^c)P(B|A^c) \end{aligned}$$

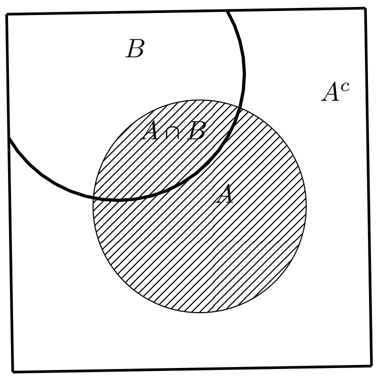


Figure 2: $P(B) = P(A \cap B) + P(A^c \cap B)$

Using that we have

$$P(B) = 0.05 \times 0.99 + 0.95 \times 0.10 = 0.1445$$

We can generalize the joint probability to three variables

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

Given disjoint events A_1, A_2, A_3 then we can deduce the total probability using divide and conquer

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \end{aligned}$$

The final version of the Bays rule for these disjoint evidences

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

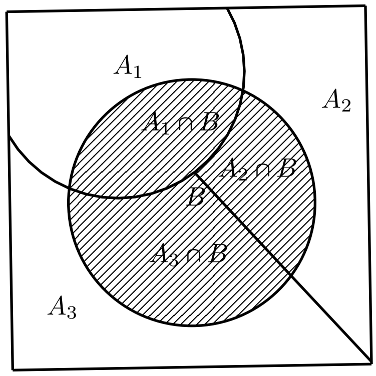


Figure 3: Total probability as a sum of three events

3 Independence

Two events A and B are independent if

$$P(B|A) = P(B)$$

Hence the event B doesn't change if A occurs. Note that independent events are different than disjoint events. Actually if A and B are disjoint then they must be dependent since $P(B|A) = 0$ but $P(B) > 0$ hence $P(B|A) \neq P(A)$. Also note that by the Bays rule we have

$$P(A \cap B) = P(A)P(B)$$

We define pairwise independence of a sequence of events $\{A\}_n$ as

$$P(\cap A_i) = \prod P(A_i)$$

which follows for any subset of the collection of $\{A\}$.

4 Counting

We define the number of outcomes of a multi-stage experiment with $\{a\}_n$ outcomes as $\prod a_i$.
Here are some facts about counting

1. **Permutation** defines the number of ways of permuting n elements which equals $n!$.
2. **Subsets** the number of subsets of an n element set equals 2^n.
3. **Combination** choosing an ordered subset with k elements out of n set elements is defined as $n(n-1)\cdots(n-(k+1))$

4. **Choose** choosing an unordered subset of k elements out of an n element set is $\binom{n}{k}$

5. **Partitioning** The number of ways to partition a set of N elements into k subsets of sizes $n_1 + n_2 + \cdots + n_k = N$ is $\frac{N!}{\prod n_i}$

5 Discrete Random Variables

We define a random variable X as a function that maps from the sample space to the real numbers

$$X : \Omega \to \mathbb{R}$$

We associate with the random variable a probability mass functions defined as

$$\begin{aligned} P_X(x) &= P(X = x) \\ &= P(\{\omega \in \Omega \mid X(\omega) = x\}) \end{aligned}$$

5.1 Marginal and joint Distributions

By definition we have the joint distribution of two random variables X, Y as

$$P_{X,Y}(x, y) = P(X = x, Y = y)$$

The marginal distribution can be computed as

$$P_X(x) = \sum_y P_{X,Y}(x, y)$$

5.2 Geometric Distribution

Number of trails to get a success. Where k is the number of trails and p is the probability of success

$$P_X(k) = P(X = k) = (1 - p)^k p$$

5.3 Binomial Distribution

Total number of heads after n trials is a random variable X with probability of success p then

$$P_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

5.4 Expectation

The expectation of a random variable X is defined as

$$\mathbb{E}[X] = \sum_x xP(X = x)$$

Let $g(X)$ be a function of the random variable then

$$\mathbb{E}[g(X)] = \sum_x g(x)P(X = x)$$

There are many properties

(a) Expectation of a real number α is $\mathbb{E}[\alpha] = \alpha$

(b) Expectation is a linear function

$$\begin{aligned}
\mathbb{E}[\alpha X + \beta] &= \sum_x (\alpha x + \beta) P_X(x) \\
&= \alpha \sum_x x P_X(x) + \beta \sum_x P_X(x) \\
&= \alpha\, \mathbb{E}[X] + \beta
\end{aligned}$$

(c) Linearity of expectation for two random variables X, Y

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_x \sum_y (x + y) P_{X,Y}(x, y) \\
&= \sum_x \sum_y x P_{X,Y}(x, y) + \sum_x \sum_y y P_{X,Y}(x, y) \\
&= \sum_x x \sum_y P_{X,Y}(x, y) + \sum_y y \sum_x P_{X,Y}(x, y) \\
&= \sum_x x P_X(x) + \sum_y y P_Y(y) \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

5.5 Variance

The variance of a random variable X is defined as the squared difference of that variable to the expectation

$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

This simplifies to

$$\begin{aligned}\text{var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 - 2X\,\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\,\mathbb{E}[X\,\mathbb{E}[X]] + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

5.6 Conditional Random Variables

We define a conditional PMF as

$$P_{X|A}(x) = P(X = x|A)$$

For instance we can condition on the geometric random variable as $P_{X|X>2}(x)$ to include only the successes after 2 trails.

5.7 Law of Total Expectation

Considering the collection of disjoint events $\{A\}_n$ such that $\cup A_i \cap B = B$ the by the total probability law by defining the random variable X as taking values from B

$$P_X(x) = \sum_i P(A_i)P(X = x|A_i)$$

Then multiply by x and sum

$$\mathbb{E}[X] = \sum_i P(A_i)\,\mathbb{E}[X|A_i]$$

5.8 Expectation and Variance of Independent Random Variables

Let X, Y be independent random variables then

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x \sum_y xyP_{X,Y}(x,y) \\ &= \sum_x \sum_y (xy)P_X(x)P_Y(y) \\ &= \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

Also we have

$$\begin{aligned}
\mathrm{var}(X+Y) &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\
&= \mathbb{E}[X^2] + 2\,\mathbb{E}[XY] + \mathbb{E}[X^2] - \mathbb{E}[X]^2 - 2\,\mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[Y]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\
&= \mathrm{var}(X) + \mathrm{var}(Y)
\end{aligned}$$

5.9 Expectation of Binomial RV

Consider a binomial random variable X as a sum of n Bernoulli independent random variables $X = \sum_i X_i$ where $X_i = 1$ if it is a success experiment. Note that

$$\mathbb{E}[X] = \mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i] = \sum_i p = np$$

Hat problem.

Consider the problem where we have n number of peoples attending a party. Every person leaves his hat before entering the party. When the party ends every person gets a random hat. What is the expected number of people to get their own hats? what is the variance ?

Let X be the number of people to get their hat back. Then X can be written as a sum of Bernoulli variables $X = \sum_i X_i$

$$\mathbb{E}[X] = \mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i] = \sum_i 1/n = n/n = 1$$

The variance is more complicated to evaluate

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

First note that

$$\mathbb{E}[X]^2 = 1^2 = 1$$

For the the first expectation

$$\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{E}\left[\left(\sum_i X_i\right)^2\right] \\
&= \mathbb{E}[\sum_i X_i^2] + \mathbb{E}[\sum_{i\neq j} X_i X_j] \\
&= \sum_i \mathbb{E}[X_i^2] + \sum_{i\neq j} \mathbb{E}[X_i X_j] \\
&= n/n + \sum_{i\neq j} P(X_i X_j = 1) \\
&= 1 + \sum_{i\neq j} P(X_i)P(X_j = 1|X_i = 1) \\
&= 1 + \sum_{i\neq j} \frac{1}{n(n-1)} = 1 + \frac{n(n-1)}{n(n-1)} = 2
\end{aligned}$$

Hence

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2 - 1 = 1$$

6 Continuous Random Variables

We define the following

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Where $f_X(x)$ is defined is the probability density function.

Consider the case where $\delta \to 0$ then the area under the curve can be approximated as

$$P(x \leq X \leq x + \delta) \approx f_X(x)\delta$$

6.1 Cumulative Distribution

The cumulative distribution is defined as

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt$$

6.2 Expectation

It is defined like in the discrete case

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

6.3 Uniform Distribution

Defined as

$$T(n) = \begin{cases} \frac{1}{b-a} & a \le X \le b \\ 0 & \text{otherwise} \end{cases}$$

The expectation can be evaluated directly as the center of gravity

$$\mathbb{E}[X] = \frac{b+a}{2}$$

7 Multiple Continuous Random Variables

We define the joint distribution

$$P((X,Y) \in S) = \int \int_S f_{X,Y}(x,y) dx dy$$

The marginal distribution

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$$

The conditional distribution

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Buffon's Needle

Given a needle with length $l < d$ find the probability of the needle intersecting one of the lines

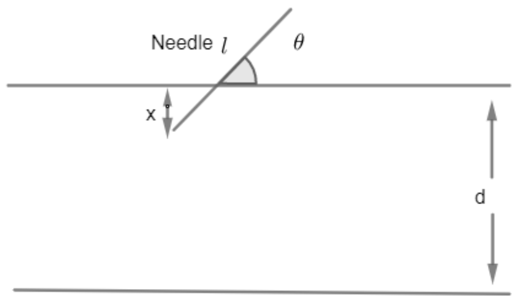


Figure 4: Needle problem

Note first that X, θ are uniform independent random variables with $0 \le x \le d/2$; $0 \le \theta \le \pi/2$

$$f(x,\theta) = f_X(x) f_\theta(\theta) = \frac{2}{d} \times \frac{2}{\pi} = \frac{4}{d\pi}$$

Note that the needle intersects one of the lines if $x \le \frac{l}{2}\sin(\theta)$ hence

$$P\left(x \le \frac{l}{2}\sin(\theta)\right) = \int_0^{\pi/2} \int_0^{l/2\sin(\theta))} \frac{4}{d\pi} dx d\theta = \frac{4}{d\pi} \int_0^{\pi/2} \int_0^{l/2\sin(\theta))} dx d\theta = \frac{2l}{\pi d}$$

8 Derived Distributions

Given the cumulative distribution we can find the marginal distribution

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

Exercise. Given $Y = X^3$ where X is uniform in the interval $[0,3]$, find the distribution of Y?

Consider the cumulative distribution then

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^3 \leq y) \\
&= P(X \leq y^{1/3}) \\
&= \frac{1}{2} y^{1/3}
\end{aligned}$$

Then we have

$$\frac{d}{dy} F_Y(y) = \frac{1}{6y^{2/3}}$$

Exercise. Given that $Y = aX + b$ with $a > 0$ then find $f_Y(y)$ using $f_X(x)$.

Note that

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(aX + b \leq y) \\
&= P(X \leq (y-b)/a) \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}$$

By taking the derivative with respect to y

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{a} F_X\left(\frac{y-b}{a}\right)$$

8.1 Convolution formula

Given $W = X + Y$ where X, Y are independent random variables then

$$\begin{aligned}
P_W(w) &= P(X + Y = w) \\
&= \int_{-\infty}^{\infty} f_{X,Y}(x, w-x) dy \\
&= \int_{-\infty}^{\infty} f_X(x) f_Y(w-x) dx
\end{aligned}$$

9 The Law of Iterated Expectations

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

Consider $g(Y) = \mathbb{E}[X|Y]$ as a random variable

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_y \mathbb{E}[X|Y=y]P_Y(y)\\ &= \sum_y \sum_x xP_{X|Y}(X=x|Y=y)P_Y(y)\\ &= \sum_x \sum_y xP_{X,Y}(X=x,Y=y)\\ &= \sum_x xP_X(x)\\ &= \mathbb{E}[X]\end{aligned}$$

Student's Score Problem

Let X be the student score of a random student and Y the section number taking values in $\{0,1\}$. Given that $\mathbb{E}[X|Y=1] = 90$ and $\mathbb{E}[X|Y=2] = 60$, then calculate the expected quiz score of a random student.

Note that by the law of iterated expectations

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \sum_y \mathbb{E}[X|Y=y]P_{X|Y}(X|Y=y) = \frac{1}{3}\times 90 + \frac{2}{3}\times 60 = 70$$

Store Problem

Define N as the number of stores and X_i as the money spent on a store i then calculate the expected money spent in total.

Note that we need to evaluate $Y = \sum_{i=1}^{N} X_i$. Also note that by the linearity of expectations

$$\mathbb{E}[Y|N=n] = n\,\mathbb{E}[X]$$

Hence we have for a general N

$$\mathbb{E}[X|N] = N\,\mathbb{E}[X]$$

Then by the law of iterated expectations

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|N]] = \mathbb{E}[N\,\mathbb{E}[Y|N]] = \mathbb{E}[N]\,\mathbb{E}[X]$$

Where the last equality follows by independence of the store number and the money spent.

10 Bernoulli Process

A sequence of independent Bernoulli trials with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. We also assume p is constant during the process.

Exercise. Find the probability of getting $X_i = 1$ for all processes.

Assume that there are N trials, then

$$P\left(X_1, X_2, \cdots, X_N = 1\right) = \cup_{i=1}^{N} P(X_i = 1) = p^N$$

Then taking $N \to \infty$

$$P(X_1 = 1, X_2 = 1, \cdots) = \lim_{N\to\infty} p^N = 0$$

10.1 Time until kth Arrival

Consider $Y_k = \sum_{i=1}^{k} T_i$ where T_i is a geometric random variable , then we have

$$\begin{aligned} P(Y_k = t) &= P(k-1 \text{ arrivals in } [1, \cdots, t-1] \text{ , arrival at } t) \\ &= \binom{t-1}{k-1} p^{k-1}(1-p)^{t-k} p \end{aligned}$$

11 Poisson Process

Define $p(k, \tau)$ as the probability of k arrivals in duration τ this is a probability distribution where $\sum_k p(k, \tau) = 1$. Disjoint intervals have independent number of arrivals with τ small then

$$p(k, \tau) = \begin{cases} 1 - \lambda\tau & k = 0 \\ \lambda\tau & k = 1 \\ 0 & k > 1 \end{cases}$$

Then we can define

$$\mathbb{E}[\text{Number of arrivals in } [0, \tau]] = \lambda\tau$$

Then we can define the expected number of arrivals per unit length

$$\lambda = \lim_{\tau\to 0} \frac{P(1, \tau)}{\tau}$$

We can use the Bernoulli processes by assuming that τ is very small, then

$$P(k, \tau) = \binom{n}{k}\left(\frac{\tau\lambda}{n}\right)^k \left(1 - \frac{\lambda\tau}{n}\right)^{n-k}$$

By taking $\tau \to 0$

$$p(k, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$$

12 Markov Chains

It models customer services with the following distributions

1. Customer arrival using a Bernoulli distribution with parameter p
2. Customer departure using a geometric distribution with parameter q
3. State X_n which measures the number of customers at step n

12.1 Markov Property

The probability of an event depends solely on the event before it

$$P(X_{n+1} = j|X_n = i) = P(X_{n+1} = j|X_n = i, X_{n-1}, \cdots, X_0)$$

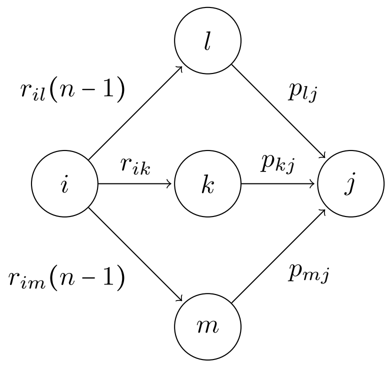
We define the probability of reaching a state j from state i after n steps as

$$r_{ij}(n) = P(X_n = j|X_0 = i)$$

Note that we have $r_{ij}(0) = 1$ if $i = j$ and $r_{ij}(0) = 0$ for $i \neq j$. Similarily we have $r_{ij}(1) = p_{ij}$

12.2 Recursive Formula

Suppose that we want to reach a state j after n steps then we calculate it for $n-1$ steps using $r_{ik}(n-1)$ where k is an intermediary state then we calculate the last step



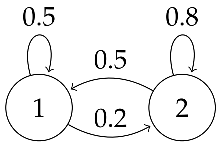
From the diagram

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj}$$

Hence we can evaluate the total probability as

$$P(X_n = j) = \sum_{i=1}^{m} P(X_0 = i)r_{ij}(n)$$

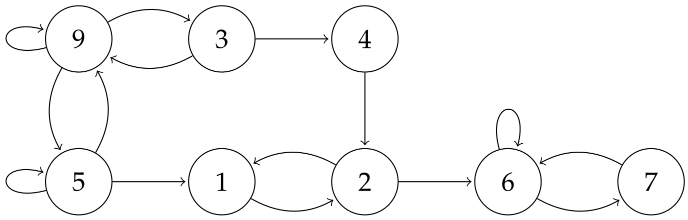
Exercise. Evaluate the following probabilities of the diagram $r_{11}(0)$, $r_{11}(11), r_{11}(n)$



Note that $r_{11}(0) = 1$, $r_{11}(1) = p_{11} = 0.5$ and

$$r_{11}(n) = 0.2 \times r_{21}(n-1) + 0.5 \times r_{11}(n-1)$$

Exercise. According to the diagram



We calculate the probabilities

$$P(X_1 = 2, X_2 = 6, X_3 = 7|X_0 = 1) = p_{12}p_{26}p_{67}$$

$$P(X_4 = 7|X_0 = 2) = p_{26}p_{67}p_{76}p_{67} + p_{26}p_{66}p_{66}p_{67} + p_{21}p_{12}p_{26}p_{67}$$

12.3 Steady State

After long time the probability of a certain state become constant regardless of the initial state i.e $r_{ij}(n) = \Pi_{ij}$ meaning that it for large n the state probability becomes constant. We define **Recurrent state** as a state that can be returned to from the initial state and **Transient state** if not recurrent. A **period** happens when there exists $d > 2$ groups such that if n is in the first group then $n+1$ will be in the next group.

Theorem. The initial state doesn't matter if the recurrent states are in a single class and the single recurrent class is not periodic.

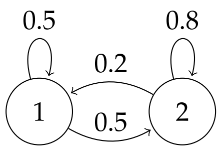
12.4 Balance Equations

Suppose there exits a steady state then

$$\Pi_j = \sum_{k=1}^{m} \Pi_k p_{kj}$$

There exits a unique solution if $\sum_j \Pi_j = 1$.

Exercise. Find the steady state probabilities of the following diagram



Then

$$\Pi_1 = 0.5 \times \Pi_1 + 0.2 \times \Pi_2$$

$$\Pi_2 = 0.5 \times \Pi_1 + 0.8 \times \Pi_2$$

We deduce then

$$0.5 \times \Pi_1 = 0.2 \times \Pi_2$$

Also we use that $\Pi_1 + \Pi_2 = 1$ to find that $\Pi_1 = 2/7$ and $\Pi_2 = 5/7$.

13 Bayesian Statistical Inference

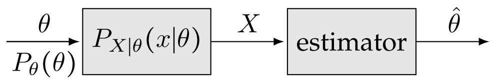


Figure 5: Bayesian Inference

This captured by the Bayes rule

$$P_{\theta|X}(\theta|X) = \frac{P_\theta(\theta)P_{X|\theta}(x|\theta)}{P_X(x)}$$

13.1 Maximum a Posteriori Probability

This defines the optimal value that optimizes the posterior probability

$$P_{\theta|X}(\theta^*|x) = \max P_{\theta|X}(\theta|x)$$

13.2 Least Minimum Squared Error

Suppose we want to estimate the best value for a random variable θ then we evaluate $\mathbb{E}[(\theta - c)^2]$

Exercise. Let $f_\theta(\theta) = 1/6$ be a uniform distribution in the interval $\theta \in [4, 10]$ then evaluate the optimal value for θ.

We evaluate the expectation of the difference squared

$$\mathbb{E}[(\theta - c)^2] = \mathbb{E}[\theta^2] - 2c\,\mathbb{E}[\theta] + c^2$$

By evaluating the derivative with respect to c

$$\frac{d\,\mathbb{E}[(\theta - c)^2]}{dc} = 0$$

We have

$$c = \mathbb{E}[\theta] = \int_4^{10} \frac{\theta}{6} d\theta = 7$$

14 Classical Inference

We want to estimate a parameter θ of the distribution $P_X(x;\theta)$

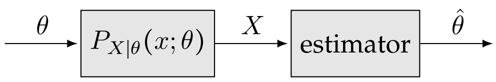


Figure 6: Bayesian Inference

14.1 Maximum likelihood estimation

$$\hat{\theta}_{\text{ML}} = \text{argmax}_\theta P_X(x;\theta)$$

Exercise. Find the maximum likelihood estimation of the exponential distribution

Assuming that we have $X_1, X_2, \cdots, X_n$ independent random variables then

$$\hat{\theta}_{\text{ML}} = \max_\theta \prod_{i=1}^{n} \theta e^{-\theta x}$$

It will be much easier to take the log since that doesn't change the optimization problem

$$\max_\theta \left(n \log \theta - \theta \sum_{i=1}^{n} x_i \right)$$

By taking the derivative and take it equal to zero

$$\theta = \frac{n}{\sum_{i=1}^{n} x_i}$$

Definition. We say an estimator to be unbiased if $\mathbb{E}[\hat{\theta}] = \theta$

Exercise. Find an unbiased estimator for the parameter p of a binomial distribution with n trials.

Let $X = X_1 + \cdots + X_n$ as a sequence of Bernoulli random variables. We estimate the parameter p as the average

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

Then this estimator is unbiased since

$$\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n} \sum_i \mathbb{E}[X_i] = p$$

14.2 Linear Regression

Given n data points (x_i, y_i) and we want to model it as a line

$$y \approx \theta_0 + \theta_1 x$$

Then we can use the squared error to evaluate the parameters

$$\min_{\theta_0,\theta_1} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$$

14.3 Probabilistic Model

We can approximate the error as a normal Gaussian distribution

$$y_i = \theta_0 + \theta_1 x_i + \epsilon_i$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Then we need to optimize

$$\max_{\theta_0,\theta_1} \exp\left\{-\frac{1}{2\sigma_2} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2\right\}$$

15 Hypothesis Testing

We define the null hypothesis

$$X \sim P_X(x : H_0)$$

and the alternative hypothesis

$$X \sim P_X(x; H_1)$$

We need a way to reject or accept H_0.

15.1 Likelihood Ratio Test

We reject H_0 if

$$\frac{P_X(x; H_1)}{P_X(x; H_0)} > \xi$$

where ξ is a critical value.

Exercise. Given n data points that are **i.i.d** (independent and identically distributed) we define

$$H_0 : X_i \sim \mathcal{N}(0, 1)$$

$$H_1 : X_i \sim \mathcal{N}(1, 1)$$

Then the ratio

$$\frac{\exp\left\{-\frac{1}{2}\sum(x_i-1)^2\right\}}{\exp\left\{-\frac{1}{2}\sum x_i^2\right\}} > \xi$$

After some manipulations we could reject H_0 if

$$\sum_{i=1}^{n} x_i > \xi'$$

where ξ' is a function of ξ.