

Algorithmic Learning Theory

Spring 2017

Lecture 2

Instructor: Farid Alizadeh

Scribe: Chien-Ming Huang

Edit: Yuan Qu

1/25/2017

1. Review Bayes Theory(Lecture 1)
2. Random Variable and Distribution
 - (a) Random variable
 - i. DRV, discrete random variable
 - ii. CRV, continuous random variable
 - (b) Distribution function
 - i. CDF, cumulative distribution function
 - ii. pdf or pmf, probability density(Mass) function
 - (c) Discrete distribution
 - i. Discrete uniform distribution
 - ii. Beunoulli's distribution
 - iii. Binomial distribution
 - (d) Continuous distribution
 - i. Continuous uniform distribution
 - ii. Normal distribution
3. Multivariate Distributions
 - (a) Random vector
 - (b) Discrete multivariate distribution
 - (c) Binormal distribution
 - (d) Marginal distribution
 - (e) Conditional distribution
4. Bayes Classification

1 Review Bayes Theory(Lecture 1)

See notes in "Lecture 1".

2 Random Variable and Distribution

2.1 Random Variable

2.1.1 Discrete Random Variable(D.R.V)

$x \rightarrow t_1, t_2, \dots, t_n$

2.1.2 Continuous Random Variable(C.R.V)

$x \rightarrow [a, b]$ a range of value.

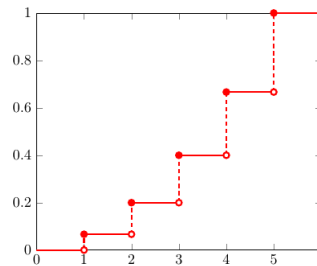
2.2 Distribution Function

2.2.1 CDF:

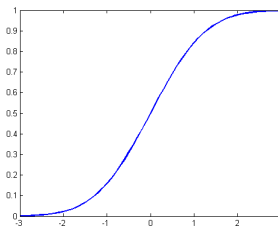
Cumulative Distribution Function

$F_x(t) = P_r[x \leq t]$, probability can only increase

1. For Discrete:



2. For Continuous:

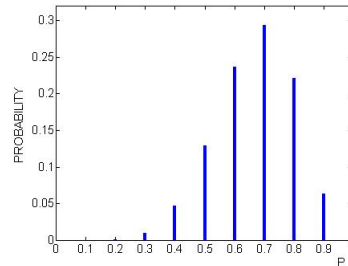


2.2.2 pdf or pmf:

Probability Density(Mass) Function

1. For Discrete:

$$x : F_x(t) = P_r[x = t]$$

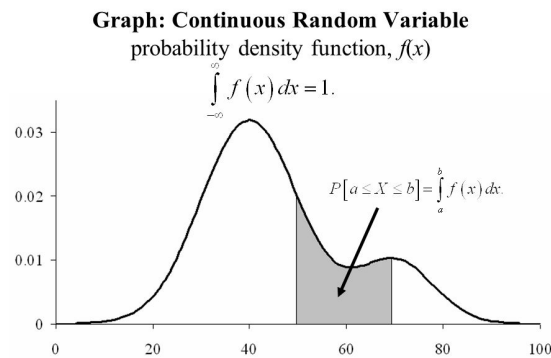


2. For Continuous:

$$f_x(t) = \frac{d}{dt} F_x(t), F_x(t) = \int_{-\infty}^t f_x(t) dt$$

$$\text{i } f_x(t) \geq 0$$

$$\text{ii } \int_{-\infty}^{\infty} f_x(t) dt = 1$$

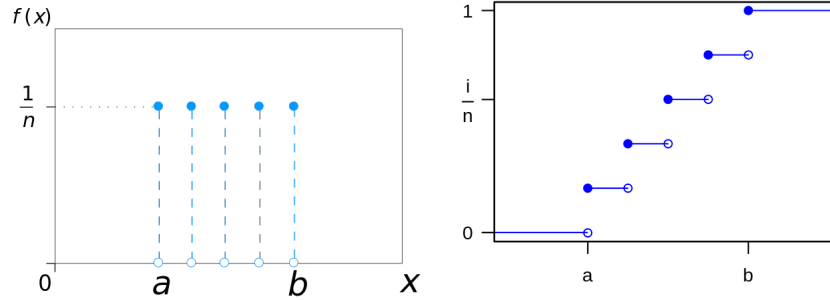


2.3 Discrete Distribution

2.3.1 Discrete Uniform Distribution

$$x : 1, 2, 3, \dots, k$$

$$\text{pdf} : u_x(t) = \begin{cases} \frac{1}{n}, & \text{if } t = 1, 2, \dots, n \\ 0, & \text{otherwise} \end{cases}$$



2.3.2 Bernoulli Distribution

$$\text{pdf: } f_x(t) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{CDF: } F_x(t) = \begin{cases} 0, & x \leq 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

2.3.3 Binomial Distribution

numbers of 0's in independent Bernoulli trial with $P[0] = p$

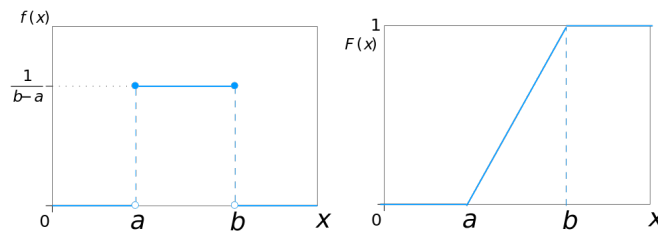
$$\text{pdf: } b(t|p, n) = \binom{n}{t} p^t (1-p)^{n-t}, \quad \binom{n}{t} = \frac{n!}{t!(n-t)!}$$

$$\text{CDF: } B(t|p, n) = \sum_{n=0}^t b(t|p, n)$$

2.4 Continuous distribution

2.4.1 Continuous uniform distribution

$$u(t|a, b) = \begin{cases} \frac{1}{b-a}, & a \leq t \leq b \\ 0, & \text{otherwise} \end{cases}$$



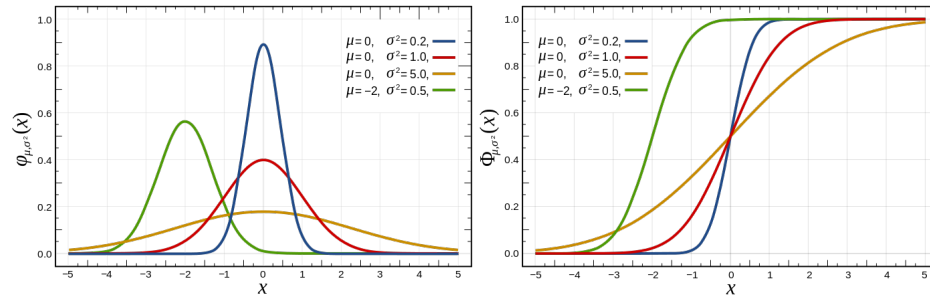
2.4.2 Normal distribution

mean = μ and std. = σ

$$\text{pdf: } f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{CDF: } \frac{1}{2} [1 + \text{erf}(\frac{x-\mu}{\sigma\sqrt{2}})]$$

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$$



3 Multivariate Distributions

3.1 Random Vector

$X_N = (x_1, x_2, \dots, x_N)$ can be continuous or discrete.

$$\text{CDF: } F_x(t_1, t_2, \dots, t_n) = P_r[x_1 \leq t_1, x_2 \leq t_2, \dots, x_n \leq t_n]$$

$$\text{pdf: } \begin{cases} \frac{\partial}{\partial t_1 \partial t_2 \partial t_3 \dots \partial t_n} F(t_1, t_2, \dots, t_n) = f_x(t_1, \dots, t_n), & \text{all continuous} \\ P_r[x_1 = t_1, x_2 = t_2, \dots, x_n = t_n] = f_x(t_1, \dots, t_n), & \text{all discrete} \end{cases}$$

both are joint distribution R.V. x_1, \dots, x_n

3.2 Discrete Multivariate Distribution

$$Y \rightarrow 1, 2, \dots, r; \quad P[Y = r_1] = P_u; \quad \sum P_u = 1$$

repeat n times, x_k = number of times $Y = k$ occurs

$$\underline{x} = (x_1, \dots, x_n)$$

$$\text{pdf: } f_x(x_1, x_2, \dots, x_n) = P[x_1 = t_1, x_2 = t_2, \dots, x_n = t_n] \binom{n}{t_1, t_2, \dots, t_n} p_1^{t_1} p_2^{t_2} \dots p_n^{t_n}$$

$$\binom{n}{t_1, t_2, \dots, t_r} = \frac{n!}{t_1! t_2! \dots t_n!}$$

3.3 Binormal distribution

$\underline{x} = (x_1, x_2)$, both continuous

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \sigma_1^2 \rightarrow x_1, \sigma_2^2 \rightarrow x_2, \sigma_{21} \rightarrow x_1, x_2$$

$$\text{Covariance Matrix: } \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

$$\phi(t_1, t_2 | \underline{\mu}, \Sigma) = \frac{1}{\sqrt{2\pi} \cdot \text{Det}(\Sigma)} \exp[-(\underline{t} - \underline{\mu})^T \Sigma^{-1} (\underline{t} - \underline{\mu})], \underline{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

$$\phi(t_1) = \int_{-\infty}^{\infty} \phi(t_1, t_2 | \dots) dt_2$$

Joint pdf: $f_{(x_1, x_2)}(t_1, t_2)$

In a similar way, for multi-normal distribution,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \cdots & \cdots & \sigma_n^2 \end{bmatrix}$$

e.g.

$\begin{matrix} & x_1 \\ x_2 & \end{matrix}$	1	2	3	$f_{x_2}(t)$
0	0.1	0.4	0.2	0.7
1	0.2	0.05	0.05	0.3
$f_{x_1}(t)$	0.3	0.45	0.45	

$$P[x_2 = 0] = P[x_2 = 0 | x_1 = 1] + P[x_2 = 0 | x_1 = 2] + P[x_2 = 0 | x_1 = 3] = 0.7$$

$$P[x_2 = 1] = 1 - P[x_2 = 0] = 0.3$$

3.4 Marginal Distribution

$$\underline{x} = (x_1, x_2, \dots, x_n) \rightarrow f_{\underline{x}_\sim}(t_1, \dots, t_n)$$

To calculate pdf,

$$f_{\underline{x}_\sim}(t_1, t_2, \dots, t_n) = \int_{t_{k+1}, \dots, t_n}^{\infty} f_{\underline{x}_\sim}(t_1, \dots, t_n) dt_{k+1}, \dots, dt_n = \sum_{t_{k+1}} \sum_{t_{k+2}} \cdots \sum_{t_n} f_{\underline{x}_\sim}(t_1, \dots, t_n)$$

Joint pdf: $f(t_1, t_2)$

e.g. Supposed that: $x_1 \rightarrow \text{Height}$, $x_2 \rightarrow \text{Gender}$, $\begin{cases} 0 & \text{male} \\ 1 & \text{female} \end{cases}$
 For $x_2 = 0 \rightarrow \text{Height} \sim N(69, 4.5) \Leftrightarrow f_{x_1|x_2}(t_1|t_2=0)\phi(t_1|\phi=69, \sigma=4.5)$
 For $x_2 = 1 \rightarrow \text{Height} \sim N(65, 4.2) \Leftrightarrow f_{x_1|x_2}(t_1|t_2=1)\phi(t_1|\phi=65, \sigma=4.2)$

3.5 Conditional Distribution in 2 Variables

$$\begin{aligned} x &= (x_1, x_2) \rightarrow \text{joint } f(t_1, t_2) \\ f_{x_1|x_2}(t_1|t_2) &= \frac{f_x(t_1, t_2)}{f_{x_2}(t_2)} \\ f_x(t_1, t_2) &= f_{x_1|x_2}(t_1|t_2) * f_{x_2}(t_2) \\ f_{x_2|x_1}(t_2|t_1) &= \frac{f_x(t_1, t_2)}{f_{x_1}(t_1)} \\ f_x(t_1, t_2) &= f_{x_2|x_1}(t_2|t_1) * f_{x_1}(t_1) \end{aligned}$$

4 Bayes

4.1 Bayes Formula

$$\begin{aligned} f_{x_1|x_2}(t_1|t_2) &= \frac{(f_{x_2|x_1}(t_2|t_1) * f_{x_1}(t_1))}{f_{x_2}(t_2)} \\ \text{Discrete } x_1 : f_{x_2}(t_2) &= \sum_{t_1} f_{x_2|x_1}(t_2|t_1) * f_{x_1}(t_1) \\ \text{Continuous } x_1 : f_{x_2}(t_2) &= \int_{-\infty}^{\infty} f_{x_2|x_1}(t_2|t_1) * f_{x_1}(t_1) dt \\ \text{Example: A person has height 6'7"} & \\ f(x_2 = 0|x_1 = 6'7") &= \frac{f_{x_1|x_2}(6'7"|x_2=0) * f_{x_2}(0)}{f_{x_1}(67)} \\ f(x_2 = 1|x_1 = 6'7") &= \frac{f_{x_1|x_2}(6'7"|x_2=1) * f_{x_2}(0)}{f_{x_1}(67)} \end{aligned}$$

Example: In a box $\frac{1}{4}$ of coins are fake, $\frac{3}{4}$ of coins are real
 The probability to get fake: $P_r[\text{head}] = \frac{1}{3}, P_r[\text{tail}] = \frac{2}{3}$
 The probability to get real: $P_r[\text{head}] = \frac{1}{2}, P_r[\text{tail}] = \frac{2}{2}$
 Take a random coin selected, $n = 20$ times, $t = 7$ heads, what is $P_r[\text{real}]$? what is $P_r[\text{false}]$?

x_1 = number of heads in $n = 20$ trials

$$x_2 = \begin{cases} 0, \text{fake} & f_{x_2}(0) = \frac{1}{4} \\ 1, \text{real} & f_{x_2}(1) = \frac{3}{4} \end{cases}$$

$$\begin{aligned} f(t_2 = 0|x_1 = 7, n = 20) &= \frac{f_{x_1|x_2}(7|\text{fake}, n=20) * f_{x_2}(0)}{f_{x_1}(7)} = \binom{20}{7} \frac{1}{3}^7 \frac{2}{3}^13 * 0.25 = 0.45 \\ f(t_2 = 1|x_1 = 7, n = 20) &= \frac{f_{x_1|x_2}(7|\text{real}, n=20) * f_{x_2}(1)}{f_{x_1}(7)} = \binom{20}{7} \frac{1}{2}^7 \frac{1}{2}^13 * 0.75 = 0.55 \end{aligned}$$

4.2 Bayes Classification

$$\text{Loss}(\hat{f}, x|f), y = f(x), y = \hat{f}(x)$$

$$\text{Minimum } E_x : \text{Loss}(\hat{f}|f)$$

$$\text{Misclassification Rate: Loss}(\hat{f}, x|f) : \begin{cases} 0 & f_x = \hat{f}(x) \\ 1 & f_x \neq \hat{f}(x) \end{cases}$$

$$\text{Risk} = E_x \text{ Loss}(\hat{f}, x|f)$$

$$\text{Probability of Misclassification: } E_x = \sum_{t_i} t_i f_x(t_i) = \int_0^1 t \int_x(t) dt$$

$$x = \begin{cases} p & 0 \\ 1-p & 1 \end{cases}$$

4.3 Bayes Classification Rule

Binary choose k $P_r[y = k, x]$

$$k = \text{category} = \text{avemax} \frac{P_0[x|k]P_r[k]}{P_r[x]} \propto P_0[x|k]P_r[k]$$

4.4 Classification Cost Matrix

C_{ij} = cost of classification a + number of classification of i + number of classification of j

$$f(x) = j, E_x(\text{loss}(j = \hat{j}, x|i) = \sum_{i=1}^k P_r[i|x]C_{ij} = \sum_{i=1}^k \frac{f(x|i)P_r[i]C_{ij}}{f_x} \propto \sum_{i=1}^k f(x|i)P_r[i]C_{ij}$$

4.5 Modify Bayes Rule(Uneven Cost)

$$c = \text{avgmin} \sum_{i=1}^k f(x|i)P_r[i]C_{ij}$$

$$\text{Example: Real and Fake: } C = \begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix}$$

$$\text{real: } P_r[x_2 = 1|x_1 = 7] * C_{11} + P_r[x_2 = 0|x_1 = 7] * C_{12}$$

$$\text{fake: } P_r[x_2 = 1|x_1 = 7] * C_{21} + P_r[x_2 = 0|x_1 = 7] * C_{22}$$

$$\min : \begin{cases} 0.55 * 0 + 0.45 * 1 = 0.45 \\ 0.55 * 4 + 0.45 * 0 = 1.8 \end{cases}$$