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)

See notes in "Lecture 1".

2 Random Variable and Distribution

2.1 Random Variable

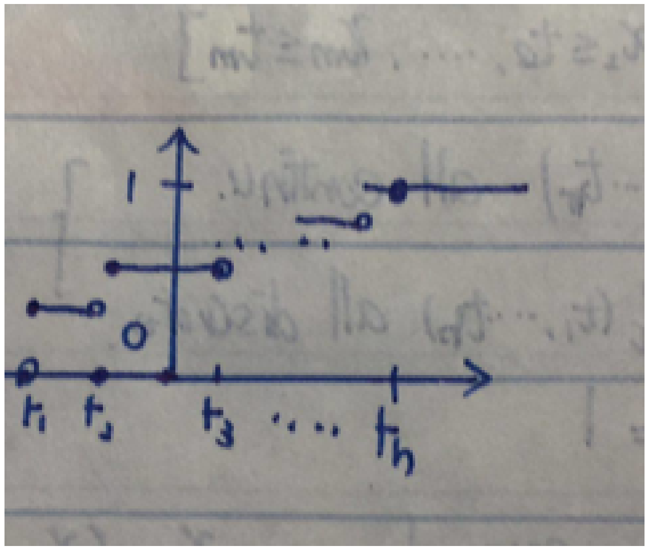
Discrete Random Variables(D.R.V) $x \to t_1, t_2, \cdots, t_n$

Continuous Random Variable(C.R.V) $x \to [a, b]$ a range of value.

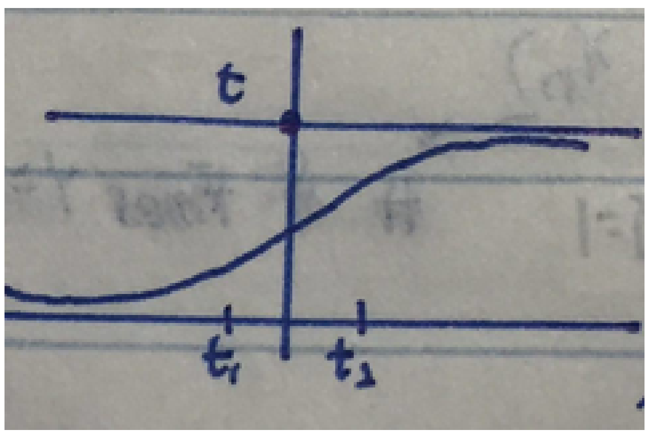
2.2 Distribution Function

CDF: Cumulative Distribution Function
$F_x(t) = P_r[x \leq t]$, probability can only increase

1. For Discrete:



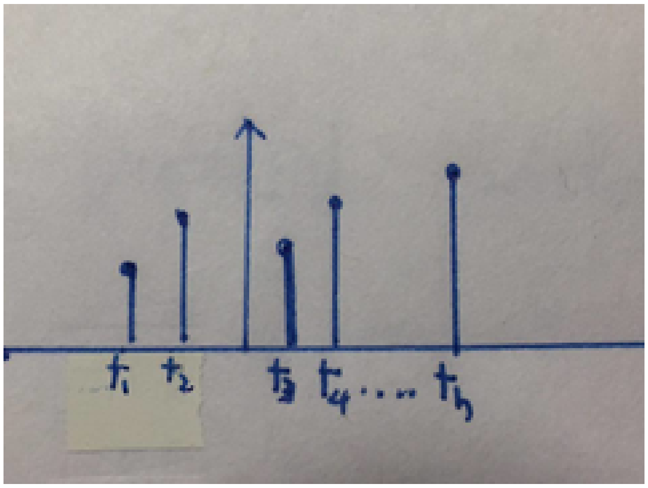
2. For Continuous:



pdf or pmf: Probability Density(Mass) Function

1. For Discrete:

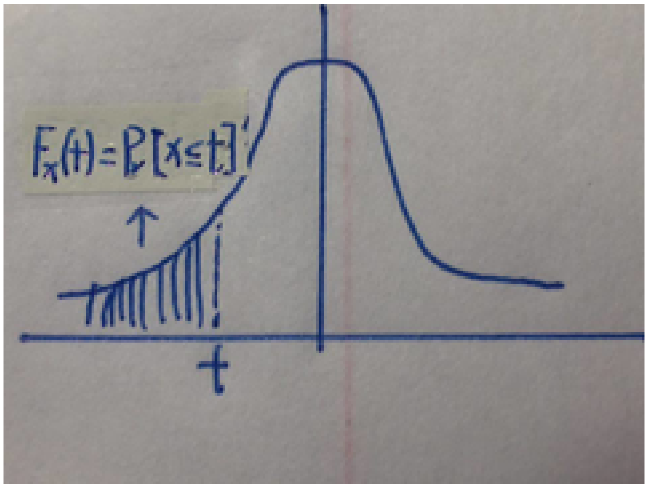
$$x : F_x(t) = P_r[x = t]$$



2. For Continuous:

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

   i $f_x(t) \geq 0$
   ii $\int_{-\infty}^{\infty} f_x(t)dt = 1$

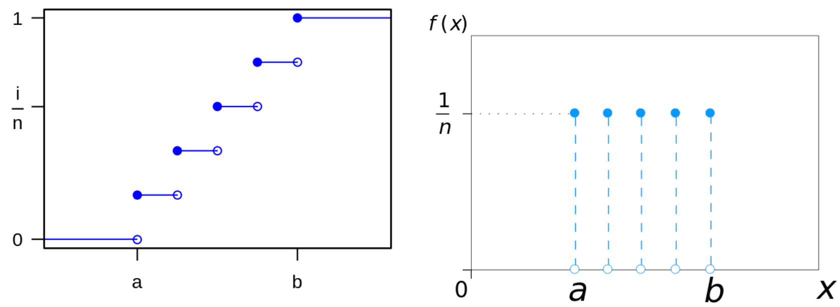


2.3 Discrete Distribution

2.3.1 Discrete Uniform Distribution

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

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



2.3.2 Bernoulli Distribution

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

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

2.4 Bernoulli Distribution

2.5 Binomial Distribution

numbers of 0's in independent Bernoulli trial with $P_r[0] = P$
P.D.F: $b(t|p,n) = \binom{n}{t} p^t (1-p)^{n-t}$
C.D.F: $B(t|p,n) = \sum_{n=0}^{t} b(t|p,n)$

2.6 Continuous Uniform

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

2.7 Normal Random

$mean = \mu$ and $std. = \sigma$
P.D.F: $\phi(t|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp(\frac{t-\mu}{\sigma})^2$
C.D.F: $\Phi(t|\mu,\sigma) = \int_{-\infty}^{t} \phi(t|\mu,\sigma)dt$

2.8 Random Vector

$X_N = (x_1, x_2, \cdots x_N)$ can be continuous or discrete
C.D.F: $F_x(t_1, t_2, \cdots t_n) = P_r[x_1 \leq t_1, x_2 \leq t_2, \cdots, x_n \leq t_n]$

P.D.F : $\begin{cases} \frac{d}{dt_1 dt_2 dt_3 \cdots dt_n} F(t_1, t_2, \cdots t_n) = f_x(t_1, \cdots, n) & \text{all continuous} \\ P_r[x_1 = t_1, x_2 = t_2, \cdots, x_n = t_n] = f_x(t_1, \cdots, n & \text{all discrete} \end{cases}$
both are joint distribution R.V. $x_1, \cdots, x_n$

3 Multivariate Distributions

3.1 Discrete Multivariate Distribution

$y \to 1, 2, \cdots, r; \quad P_r[y = r_1] = P_u; \quad \sum P_u = 1$
repeat n times, X_n= number of times y= k occurs $\underset{\sim}{x} = (x_1, \cdots, x_n)$ $x_1 =$ number of times y=1; $x_n =$ number of times y=r

3.2 Multinomial Distribution

P.D.F: $f_{\underset{\sim}{x}}(t_1, t_2, \cdots, t_r) = P_r[x_1 = t_1, x_2 = t_2, \cdots, x_n = t_n] = \binom{n}{t_1, t_2, \cdots, t_r} P_1^{t_1}, P_2^{t_2}, \cdots P_r^{t_r}$
$\binom{n}{t_1, t_2, \cdots, t_r} = \frac{n!}{t_1! t_2! \cdots t_n!}$
$\underset{\sim}{x} = (x_1, x_2)$both continuous, $\underset{\sim}{\mu} = \binom{\mu_1}{\mu_2}$
$\sigma_1^2 \to x_1, \sigma_2^2 \to x_1, \sigma_{12} \to x_1 x_2$
Covariance Matrix: $\sum = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$

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

$\phi(t_1, t_2 | \binom{\mu}{n}, \sum) = \frac{1}{\sqrt{2\pi Det(\sum)}} exp[(t - \underset{\sim}{\mu})^T \sum^- 1(t - \underset{\sim}{\mu}), t = \binom{t_1}{t_2}$

joint p.d.f. $f_{(x1,x2)}(t_1, t_2)$

$x_1 \backslash x_2$	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
$f_x(t)$	0.3	0.45	0.45	

$f_{x_1|x_2=0}$ in (1): $\frac{0.1}{0.7} = \frac{1}{7}$
$f_{x_1|x_2=0}$ in (2): $\frac{0.4}{0.7} = \frac{4}{7}$
$f_{x_1|x_2=0}$ in (3): $\frac{0.2}{0.7} = \frac{2}{7}$

3.3 Marginal Distribution

$\underset{\sim}{x} = (x_1, x_2, \cdots, x_n) \to f_x(t_1, \cdots, t_n)$
p.d.f.: $(x_1, x_2, \cdots, x_n) = \underset{\sim}{\overline{x}}$
$f_{\overline{x}} = (t_1, t_2, \cdots, t_n) = \int_{t_{k+1}, \cdots, t_n}^{\infty} f_x(t_1, \cdots, t_n) dt_{k+1}, \cdots, t_n = \sum_{t_{k+1}} \sum_{t_{k+2}} \cdots \sum_{t_n} f_x(t_1, \cdots, t_n)$

$P_r[x=0] = P_r[x_2=0, x_1=1] + P_r[x_2=0, x_1=2] + P_r[x_2=0, x_1=3]$
$P_r[x=1] = P_r[x_2=1, x_1=1] + P_r[x_2=1, x_1=2] + P_r[x_2=1, x_1=3]$
$\phi(t_1, t_2|\underset{\sim}{\mu}, \sum)$
$\phi(t_1) = \int_{-\infty}^{\infty} \phi(t_1, t_2|\cdots)dt_2$
Find p.d.f.$f(t_1, t_2)$
Example: $x_1 \to$ Height, $x_2 \to$ Gender $\begin{cases} 0 & \text{male} \\ 1 & \text{female} \end{cases}$
for $x_2 = 0 \to$ 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 $x_2 = 1 \to$ Height $\sim N(65, 4.2) \Leftrightarrow f_{x_1|x_2}(t_1|t_2=1)\phi(t_1|\phi=65, \sigma=4.2$
marginal distribution of height for people
$f_{x_1}(t_1) = f_{x_1|x_2}(t_1|t_2=0) * f_{x_2}(0) + f_{x_1|x_2}(t_1|t_2=1) * f_{x_2}(1) = \phi(t_1|69, 4.5) * 0.5 + \phi(t_2|65, 4.2) * 0.5 = \phi(t_1|\frac{69+65}{2}, \sqrt{\frac{4.5^2+4.2^2}{2}})$

3.4 Conditional Distribution in 2 Variables

$x = (x_1, x_2) \to$ joint $f(t_1, t_2)$
$f_{x_1|x_2}(t_1|t_2) = \frac{f_x(t_1,t_2)}{f_x(t_2)}$
$f_x(t_1, t_2) = f_{x_1|x_2}(t_1|t_2) * f_{x_2}(t_2)$
$f_{x_2|x_1}(t_2|t_1) = \frac{f_x(t_1,t_2)}{f_{x_1}(t_1)}$
$f_x(t_1, t_2) = f_{x_2|x_1}(t_2|t_1) * f_{x_1}(t_1)$

4 Bayes

4.1 Bayes Formula

$f_{x_1|x_2}(t_1|t_2) = \frac{(f_{x_2|x_1}(t_2|t_1)*f_{x_1}(t_1))}{f_{x_2}(t_2)}$
Discrete $x_1 : f_{x_2}(t_2) = \sum_{t_1} f_{x_2|x_1}(t_2|t_1) * f_{x_1}(t_1)$
Continuous $x_1 : f_{x_2}(t_2) = \int_{-\infty} \infty f_{x_2|x_1}(t_2|t_1) * f_{y_1}(t_1)dt$
Example: A person has height 6?7"
$f(x_2=0|x_1=6'7") = \frac{f_{x_1|x_2}(6'7"|x_2=0)*f_{x_2}(0)}{f_{x_1}(67)}$
$f(x_2=1|x_1=6'7") = \frac{f_{x_1|x_2}(6'7"|x_2=1)*f_{x_2}(0)}{f_{x_1}(67)}$
Example: In a box $\frac{1}{4}$ of coins are fake, $\frac{3}{4}$ of coins are real
The probability to get fake: $P_r[head] = \frac{1}{3}, P_r[tail] = \frac{2}{3}$
The probability to get real: $P_r[head] = \frac{1}{2}, P_r[tail] = \frac{2}{2}$
Take a random coin selected, n= 20 times, t= 7 heads, what is $P_r[real]$? what is $P_r[false]$?
x_1= number of heads in n= 20 trials
$x_2 = \begin{cases} 0, \text{fake} & f_{x_2}(0) = \frac{1}{4} \\ 1, \text{real} & f_{x_2}(1) = \frac{3}{4} \end{cases}$
$f(t_2=0|x_1=7, n=20) = \frac{f_{x_1|x_2}(7|fake, n=20)*f_{x_2}(0)}{f_{x_1}(7)} = \binom{20}{7}\frac{1}{3}^7\frac{2}{3}^1 3 * 0.25 = 0.45$

$f(t_2 = 1|x_1 = 7, n = 20) = \frac{f_{x_1|x_2}(7|real, n=20)*f_{x_2}(1)}{f_{x_1}(7)} \binom{20}{7} \frac{1}{2}^7 \frac{1}{2}^1 3 * 0.75 = 0.55$

4.2 Bayes Classfication

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

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

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

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

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 $P_r[y = k, \underset{\sim}{x}]$

k= category= 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

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

4.5 Modify Bayes Rule(Uneven Cost)

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

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

real:$P_r[x_2 = 1|x_1 = 7] * C_{11} + P_r[x_2 = 0|x_1 = 7] * C_{12}$
fake:$P_r[x_2 = 1|x_1 = 7] * C_{21} + P_r[x_2 = 0|x_1 = 7] * C_{22}$

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