Machine Learning 2018 – Probability Review

Kien C Nguyen

November 14, 2018

VEF ACADEMY

Permutations

Result

Number of different ordered arrangements (permutations) of n distinct objects

$$P_n = n(n-1)(n-2)\dots 3.2.1 = n!$$

Example 1: How many ways can you arrange 4 students in a row of 4 chairs?

Example 2: How many different letter arrangements can we get from the letters B, E, T, T, E, R?

Combinations

Result

Number of different groups of r objects taken from n objects $(0 \leq r \leq n)$

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Example 1: Andy attends a Machine Learning class at the VEF Academy which has a total of 28 students (including Andy). Andy wants to find 3 more students in the class to form a group of 4 (which Andy plans to name "Fantastic Four") for the term projects. How many ways can Andy pick the 3 students?
Example 2: (Binomial Theorem) Expand the polynomial $(x+y)^n$ and explain the coefficients!

Python functions for permutations and combinations

```
>>> # Permutation
>>> import math
>>> math.factorial(5)
120

>>> # Combination
>>> from scipy import special
>>> scipy.special.comb(5, 2, exact=True,\
repetition=False)
10L
```

Sample space

- The sample space of an experiment is the set of all possible outcomes of the experiment. E.g.,
  - In the experiment of flipping 1 coin (Bernoulli trial), the sample space is $S = \{H, T\}$.
  - What is the sample space in the experiment of flipping 2 coins?
  - We rank all the students in a class of 30 students based on the weighted average of their scores. Suppose that no two students have the same weighted average score. What is the sample space of the rankings?
  - A smart phone manufacturer conducts an experiment in which they measure (in hours) the lifetime of one of their smart phone models, SP-X. The sample space is $S = \{x : 0 \leq x < \infty\}$.

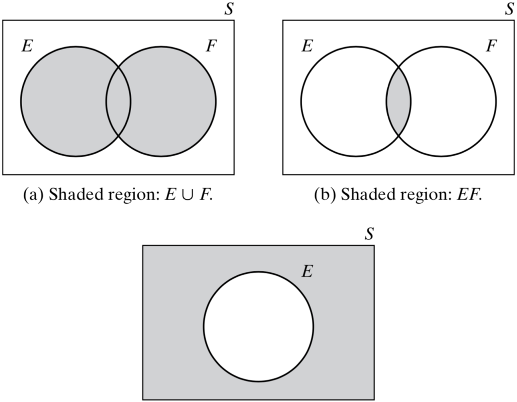
Events

- Any subset E of the sample space is an event. An event is a set of possible outcomes of the experiment. E.g.
  - In the experiment of flipping 2 coins, the event that the first coin turns Heads:
    $E = \{(H,H),(H,T)\}$
  - In the experiment of measuring the lifetime of the smart phone, the following event
    $E = \{x : 10,000 \leq x \leq 20,000\}$.
- Union: $E \bigcup F$ is the event that consists of all outcomes that are either in E or in F or in both E and F. E.g.,
  $E = \{(H,H),(H,T)\}$, $F = \{(H,T),(T,H)\}$, then
  $E \bigcup F = \{(H,H),(H,T),(T,H)\}$.
- Intersection $E \bigcap F$ is the event that consists of all outcomes that in both E and F. With E and F from above,
  $E \bigcap F = \{(H,T)\}$.

Events

- Complement: For any event E, the complement of E, E^C is defined to consist of all outcomes in the sample space S that are not in E.
- $E = \{(H, H), (H, T)\} \Rightarrow E^C = \{(T, H), (T, T)\}$.
  $E = S \Rightarrow E^C = \emptyset$.
- Figures from Ross [1]

Events



(a) Shaded region: $E \cup F$.

(b) Shaded region: EF.

(c) Shaded region: E^c.

FIGURE 2.1: Venn Diagrams

Events

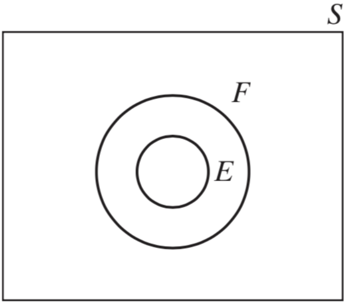


FIGURE 2.2: $E \subset F$

Axioms of Probability

Consider an experiment with the sample space S. For each event E of the sample space S, a number $P(E)$ is defined such that the following three axiom hold :

Axiom 1: $0 \leq P(E) \leq 1$

Axiom 2: $P(S) = 1$

Axiom 3: For any sequence of mutually exclusive events $E_1, E_2, \ldots$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Such a $P(E)$ is referred to as the probability of the event E. In this case, the event space is called a σ-algebra.

Conditional Probability

$P(E|F)$ is the conditional probability that E occurs given that F has occurred.

Definition

If $P(F) > 0$ then

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Example 1: (from Ross [1]) A fair coin is tossed twice. Assuming that the two coin tosses are independent, what is the conditional probability that both tosses land on heads, given that (a) the first flip lands on heads? (b) at least one flip lands on heads?

Conditional Probability

Example 1: (a) Let $B = \{(H, H)\}$ be the event that both tosses land on heads, $F = \{(H, H), (H, T)\}$ be the event that the first toss land on heads.

$$P(B|F) = \frac{P(BF)}{P(F)} = \frac{1/4}{1/2} = 1/2$$

Independent Events

Definition

Two events E and F are said to be independent if $P(EF) = P(E)P(F)$. Two events E and F that are not independent are said to be dependent (Ross [1]).

Example: A card is drawn at random from an ordinary deck of 52 playing cards. If E is the event that the selected card is a Queen and F is the event that it is a diamond, then E and F are independent. We have that $P(EF) = 1/52$, whereas $P(E) = 4/52$ and $P(F) = 13/52$.

Bayes's Formula

Formula

Let F_i, i=1, . . . , n, be mutually exclusive events whose union is the entire sample space. We have that

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

If the events F_i, i=1, . . . , n, are competing hypotheses, then we can use Bayes's formula to compute the conditional probabilities of these hypotheses when we have additional evidence E.

Bayes's Formula – Example (from Ross [1])

Example 1: Suppose that we have 3 cards that are identical in form, except that both sides of the first card are colored red, both sides of the second card are colored black, and one side of the third card is colored red and the other side black. The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

Let RR, BB, and RB denote, respectively, the events that the chosen card is all red, all black, or the red–black card. Also, let R be the event that the upturned side of the chosen card is red. Then the desired probability is obtained by

$$\begin{aligned} P(RB|R) &= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|RB)P(RB) + P(R|BB)P(BB)} \\ &= \frac{(1/2)(1/3)}{(1)(1/3) + (1/2)(1/3) + 0(1/3)} = \frac{1}{3} \end{aligned}$$

Random Variables

- A real-valued function defined on the outcome of a probability experiment is called a random variable (r.v.).
- $F(x) = P\{X \leq x\}$ is called the distribution function (a.k.a. cumulative distribution function, or cdf) of X.
- A random variable whose set of possible values is either finite or countably infinite is called discrete.
- If X is a discrete r.v. then the function $p(x) = P\{X = x\}$ is called the probability mass function of X.
- The expected value of X (also mean or expectation of X) is given by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Random Variables

- For a function g,

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

- The variance of X

$$var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

- The standard deviation of X: $SD(X) = \sqrt{Var(X)}$
- X is the random value whose value is equal to the value from rolling a die. Calculate $E(X)$, $var(X)$, and $SD(X)$.

Python functions for mean, var, and std dev

```
>>> import numpy as np
>>> die = [1, 2, 3, 4, 5, 6]
>>> np.mean(die)
3.5
>>> np.var(die)
2.9166666666666665
>>> np.std(die)
1.707825127659933
```

Some common types of discrete random variables

Binomial(n, p): The pmf is given by

$$\begin{aligned} p(i) &= \left(\begin{array}{c} n \\ i \end{array} \right) p^i (1-p)^{(n-i)} \\ &= \frac{n!}{i!(n-i)!} \, p^i (1-p)^{(n-i)} \\ &\quad \text{where } 0 \leq i \leq n \\ E[X] &= np \\ Var(X) &= np(1-p) \end{aligned}$$

Note that when $n = 1$, this becomes a Bernoulli random variable.

pmf of Binomial(10, 1/2)

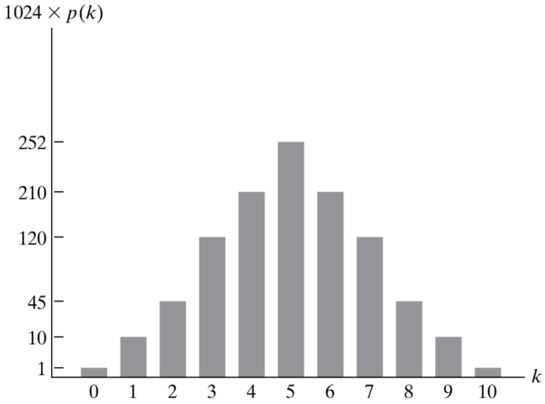


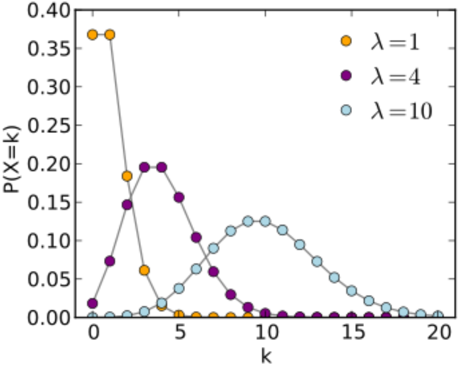
FIGURE 4.5 Graph of $p(k) = \binom{10}{k}\left(\frac{1}{2}\right)^{10}$

Some common types of discrete random variables

Poisson(λ): The pmf is given by

$$\begin{aligned} p(i) &= \frac{e^{-\lambda}\lambda^{i}}{i!} \qquad i \geq 0 \\ E[X] &= Var(X) = \lambda \end{aligned}$$

pmf of Poisson(λ)



Continuous Random Variables

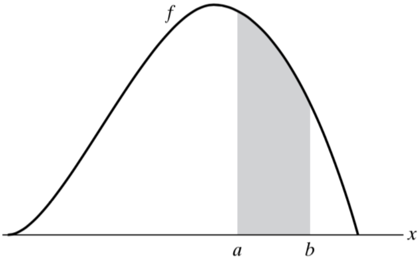
- Random variables whose set of possible values is uncountable
- X is a continuous random variable if there exists a nonnegative function f, defined for all real $x \in (-\infty, \infty)$ having the property that, for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)dx$$

$$\begin{aligned} F(x) &= P\{X \leq x\} = \int_{-\infty}^{x} f(x)dx \\ 1 &= \int_{-\infty}^{\infty} f(x)dx \\ P(a \leq x \leq b) &= \int_a^b f(x)dx \end{aligned}$$

- $f(x)$ is called the probability density function (pdf) of random variable X. $F(x)$ is the cumulative distribution function (cdf) of X.

pmf of Poisson(λ)



$P(a \le X \le b)$ = area of shaded region

FIGURE 5.1: Probability density function f.

Continuous Random Variables

Example: Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & otherwise \end{cases}$$

(a) What is the value of C?
(b) Find $P\{X > 1\}$.

Expectation and Variance

Expectation of X

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

Expectation of $g(X)$

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

The variance of X

$$var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

The standard deviation of X: $SD(X) = \sqrt{Var(X)}$

Example

Example: (from Ross [1]) Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

(a) Calculate $E(X)$.
(b) Calculate $var(X)$.

Normal random variables

A normal random variable X with mean μ and variance σ^2 is a continuous random variable whose probability density function (pdf) is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

A standard normal random variable is a normal random variable with $\mu = 0$ and $\sigma = 1$.

Normal random variables

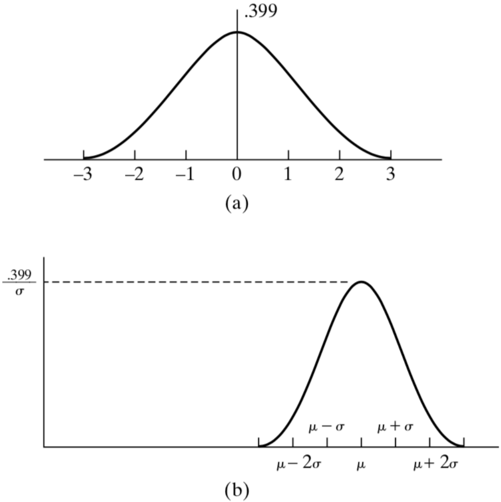


FIGURE 5.5: Normal density function: (a) $\mu = 0, \sigma = 1$; (b) arbitrary μ, σ^2.

Screening test revisit

Let Z be a standard normal random variable and $z > 0$.

(a) Show that

$$P\{Z > z\} = P\{Z < -z\}$$

(b) Show that

$$P\{|Z| < z\} = 2P\{Z < z\} - 1$$

(c) Calculate $P\{Z < 3\}$.

(d) Calculate $P\{-2 \le Z \le 2\}$.

Solution.

(a) We note that the pdf of a standard normal random variable (r.v.) is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

where we plugged in the values $\mu = 0$ and $\sigma = 1$.

We see that $f(x) = f(-x)$ and thus $f(x)$ is symmetric about 0.

Thus it can be seen

$$P\{Z > z\} = P\{Z < -z\}$$

More rigorously,

$$P\{Z > z\} = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Using a change of variable $x = -u$, we have that

$$\begin{aligned}
P\{Z > z\} &= \int_{-z}^{-\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} x'(u) du \\
&= -\int_{-z}^{-\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \text{ where we use } \ x'(u) = -1 \\
&= \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\
&= P\{Z < -z\} \text{ Q.E.D.}
\end{aligned}$$

(b)

$$\begin{aligned} P\{|Z| < z\} &= 2P\{0 < Z < z\} \\ &= 2\left(P\{Z < z\} - 0.5\right) \\ &= 2P\{Z < z\} - 1 \text{ Q.E.D.} \end{aligned}$$

(c) $P\{Z < 3\} \approx 1.00$
Sample Python code

```
from scipy.stats import norm
norm.cdf(3)
```

(d) $P\{-2 \leq Z \leq 2\} \approx 0.95$
Sample Python code

```
from scipy.stats import norm
norm.cdf(2) - norm.cdf(-2)
```

Jointly Distributed Random Variables

The joint cumulative probability distribution function of the pair of random variables X and Y is defined by

$$F(x,y) = P\{X \leq x, Y \leq y\} \qquad -\infty < x,y < \infty$$

If X and Y are both discrete random variables, then their joint probability mass function is defined by

$$p(i,j) = P\{X = i, Y = j\}$$

The individual mass functions are

$$P\{X = i\} = \sum_j p(i,j) \qquad P\{Y = j\} = \sum_i p(i,j)$$

Jointly Distributed Random Variables

The random variables X and Y are said to be jointly continuous if there is a function $f(x, y)$, called the joint probability density function, such that for any two-dimensional set C,

$$F(x, y) = P\{(X, Y) \in C\} = \iint_C f(x, y)dxdy$$

If X and Y are jointly continuous, then they are individually continuous with density functions

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Jointly Distributed Random Variables

The random variables X and Y are independent if, for all sets A and B,

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

If X and Y are discrete random variables, then the conditional probability mass function of X given that $Y = y$ is defined by

$$P\{X = x|Y = y\} = \frac{p(x, y)}{P_Y(y)}$$

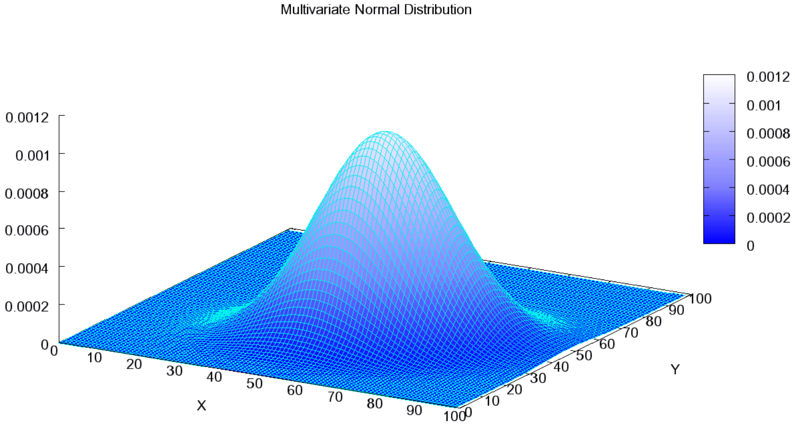
Also, if X and Y are jointly continuous with joint density function f , then the conditional probability density function of X given that $Y = y$ is given by

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

Bivariate Normal Distribution

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x - \mu_x}{\sigma_x}\right)^2 + \left(\frac{y - \mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

Bivariate Normal Distribution


Multivariate Normal Distribution
X
Y