STAT - 324

Probability & Statistics for Engineers

Chapter 1:

Introduction to Statistics & Data Analysis:

Statistics:
- → Collection of data
- → Displaying and organizing data
- → Descriptive statistical measures
- → Data Analysis and inferences

* Population (whole set) — Sample $\equiv$ observations (subset)

N: population size — n: sample size

$X_1, X_2, \dots, X_N$ — $x_1, x_2, \dots, x_n$ sample values

statistics obtained from the sample are used to estimate (approximate) the parameters of the population.

* Scientific data

* Statistical Inference: (1) Estimation → point est. → interval est.

(2) Hypotheses testing

* Measures of location (Central tendency):

sample mean $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

note: $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

* Measures of variability (Dispersion)

Sample variance $S^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ → degrees of freedom

Standard deviation $S = \sqrt{S^2}$

note: $S^2 = \dfrac{\sum x_i^2 - n(\bar{x})^2}{n-1}$

STAT-324

Probability: Chapter 2

* Sample space S [set of all possible outcomes of a random experiment]
* Event $A \subseteq S$

$\phi \subseteq S$ impossible event

$S \subseteq S$ sure event

$n(S)$ no. of elements in S

$n(A)$ no. of elements in A (occurrences)

→ Operations on events:

* Complement $A^c = \{x \in S : x \notin A\}$
* Intersection $A \cap B = AB = \{x \in S : x \in A \text{ and } x \in B\}$

(both A and B occur together)

- if $A \cap B = \phi \Rightarrow A$ and B are disjoint (mutually exclusive)

* Union $A \cup B = \{x \in S : x \in A \text{ or } x \in B\}$

(At least one of A, B occurs)

→ Counting sample points:

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

note: $n! = n.(n-1)!$

- no. of ways of selecting r objects from n objects without regard to order.
- no. of combinations of n distinct objects taken r at a time.
- no. of different selections of r objects from n objects
- no. of ways of dividing a set of n objects into two sets (r) & $(n-r)$

$\binom{n}{n} = 1, \; \binom{n}{0} = 1, \; \binom{n}{1} = n, \; \binom{n}{r} = \binom{n}{n-r}$

→ Probability of an event:

- $0 \leq P(A) \leq 1$, $P(S) = 1$, $P(\Phi) = 0$

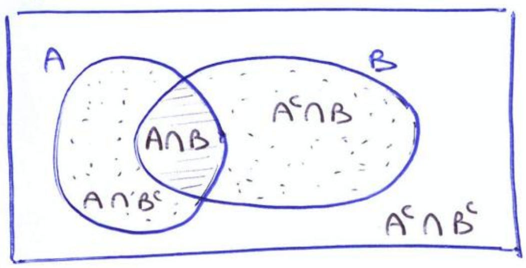
$$P(A) = \frac{n(A)}{n(S)}$$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

if A, B are disjoint $\underline{P(A \cap B) = 0}$

$$P(A \cup B) = P(A) + P(B)$$

S



- $P(A^c) = 1 - P(A)$
- $P(A \cap B^c) = P(A) - P(A \cap B)$
- $P(A^c \cap B^c) = 1 - P(A \cup B)$

<u>D' Morgans:</u> $A^c \cap B^c = (A \cup B)^c$

$A^c \cup B^c = (A \cap B)^c$

* Conditional probability:

probability of occuring an event A, given & it's known that some event B has occured - conditional prob. of A given B

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

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

$\boxed{\text{Multiplicative rule:}}$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1 A_2)$$

Independent events:

- if A_1, A_2, A_3 are independent:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2)\,P(A_3)$$

- A, B are independent events:

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

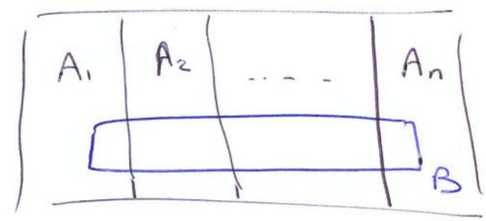
$P(B|A) = P(B)$ or

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

* Bayes' Rule

Defn.: Events $A_1, A_2, \cdots, A_n$ constitute a partition of the sample space S if:

- $\bigcup_{i=1}^{n} A_i = S$
- $A_i \cap A_j = \phi \quad \forall\ i \neq j$



Theorem: Total probability:

If $A_1, A_2, \cdots, A_n$ constitute a partition of S then for any event B:

$$\boxed{P(B) = \sum_{k=1}^{n} P(A_k)\ P(B|A_k)}$$

Bayes' rule: $$\boxed{P(A_i|B) = \frac{P(A_i)\,(P(B|A_i)}{P(B)}}$$

Discrete r.v.'s:

* Prob. distribution function

$$f(x) = P(X = x) \quad \begin{cases} f(x) \geq 0 \\ \sum_x f(x) = 1 \end{cases}$$

x	x_1	x_2	$\cdots$	x
$f(x)$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x$

* Cummulative distribution function CDF

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

$$F(x) = \begin{cases} 0 & x < x_1 \\ \boxed{\quad} & x_1 \leq x < x_2 \\ \vdots & \vdots \\ 1 & x \geq x \end{cases}$$

$$f(x_i) = F(x_i) - F(x_{i-1})$$

* $$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$
$$= F(b) - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a) + f(a)$$

$$P(a < X < b) = F(b) - F(a) - f(b)$$

Mean $\mu = E(X) = \sum_x x f(x)$

Variance $\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$

where $E(X^2) = \sum_x x^2 f(x)$

Continuous r.v.'s

$P(X = x) = 0 \ , \ \forall x$

* Prob. density function

$$f(x) = \begin{cases} \boxed{\quad} & x_0 < x < x_1 \\ 0 & \text{elsewhere} \end{cases} \quad \begin{cases} f(x) \geq 0 \\ \int_{-\infty}^{\infty} f(x)\,dx = 1 \end{cases}$$

* Cummulative distribution function CDF

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

* $$P(a < X < b) = P(a \leq X \leq b) = F(b) - F(a)$$

Mean $\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$

Variance $\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$E(aX \pm b) = a\,E(X) \pm b$$

$$Var(aX \pm b) = a^2\, Var(X)$$

---

if X, Y are independent r.v.'s

$$E(aX \pm bY) = a\,E(X) \pm b\,E(Y)$$

$$Var(aX \pm bY) = a^2\, Var(X) + b^2\, Var(Y)$$

Chebyshev's theorem:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

lower bound

$$P(-k\sigma < X - \mu < k\sigma) \geq 1 - \frac{1}{k^2}$$

note: $P(|X - \mu| \geq k\sigma) = 1 - P(|X - \mu| < k\sigma)$