ACM 116: Lecture 2

Agenda

- Independence
- Bayes' rule

- Discrete random variables
  - Bernoulli distribution
  - Binomial distribution
- Continuous Random variables
  - The Normal distribution
- Expected value of a random variable

Bayes' rule

$B_1, \ldots, B_n$ mutually exclusive,
$\bigcup_{i=1}^{n} B_i = \Omega$, then for any event A,

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

because

- $P(A|B_j)P(B_j) = P(A \cap B_j)$
- $\sum P(A|B_i)P(B_i) = P(A)$

Example : Polygraph tests

(lie-detector tests)

- Event +/- : polygraph reading is positive/negative
- Event T : subject telling the truth
- Event L : subject is lying

Studies of polygraph reliability

$$P(+|L) = .88 \qquad \Rightarrow P(-|L) = .12$$
$$P(-|T) = .86 \qquad \Rightarrow P(+|T) = .14$$

On a particular question, a vast majority of the subjects have no reason to lie :

$$P(T) = .99 \qquad \Rightarrow P(L) = .01$$

Example : Polygraph tests

Probability that a person is telling the truth when the polygraph is +

$$P(T|+) = \frac{P(+|T)P(T)}{P(+|T)P(T) + P(+|L)P(L)}$$
$$= \frac{.14 \times .99}{.14 \times .99 + .88 \times .01} = .94$$

Conclusion : screening this population of largely innocent people, 94% of positive readings will be in error.

This examples illustrates the dangers in using screening procedures on large populations.

Independence

Two events are independent if the chance of getting B is the same whether A occurred or not.

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

i.e.

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

Definition : Two events are independent if

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

Independence

Example : Sex distribution for 3 children

$$\{b,b,b\},\{b,b,g\},\ldots,\{g,g,g\} \qquad \text{each with } p = 1/8$$

- Event A : there is at most one girl
- Event B : family has children of both sexes

$$P(A) = 4/8 \qquad P(B) = 6/8$$
$$P(A \cap B) = 3/8 = P(A)P(B)$$

because $A \cap B = \{b,b,g\},\{b,g,b\},\{g,b,b\}$

$\Rightarrow$ Independence

Independence : generalization

The collection of events $A_1, A_2, \ldots, A_n$ are said to be mutually independent if

$$P(A_{i_1} \cap \ldots \cap A_{i_n}) = P(A_{i_1}) \ldots P(A_{i_m})$$

for any subcollection $A_{i_1}, \ldots A_{i_m}$.

Remark : pairwise independence is not enough !

Example : two coin tosses

- A : first toss H
- B : second toss H
- C : exactly one H

Pairwise independence but it is clear that A, B, C are not independent.

Independence : generalization

Example : System made of several components

1. In a series : system fails if any of the component fails

$$P(\text{works}) = (1-p)^n$$

e.g. $p = .5, \quad n = 10, \quad P(\text{works} = .6)$

2. Parallel : system fails if all components fail

$$P(\text{fails}) = p^n$$

e.g. $P(\text{fails}) = (.5)^{10} \simeq 10^{-13}$

Random variables

Sometimes, you are not interested in all the details of the outcome of an experiment, but rather interested in the numerical value of some quantity determined by the outcome of an experiment.

Example : Survey, random sample of size $N = 100$

R.V. = proportion of people who favor candidate A. We are not interested in individual responses but rather in the # of people supporting candidate A.

Example : Physics

R.V. = # particles hitting a detector. We are not interested in individual particles.

Definition

A r.v. is a function from the sample space to the real numbers

Discrete Random Variables

R.v. that can take on only a discrete set of values (finite or countable)

Example : experiment that consists of sampling people from a large population until you find someone with a given disease. X is the number of people you sampled.

Probability distribution : X can take on values $x_1, x_2, \ldots$

$$p(x_i) = P(X = x_i)$$

$$\sum_i p(x_i) = 1$$

Cumulative distribution function (cdf)

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty$$

Examples

- The Bernoulli distribution
- The binomial distribution
- The Poisson distribution (later)

The Bernoulli distribution

Bernoulli r.v. takes on only two values : 0 or 1.

$$X = \begin{cases} 0 & \text{w.p.} \quad 1-p \\ 1 & \text{w.p.} \quad p \end{cases}$$

e.g. interested in whether some event A occurs or not :

$$I_A = \begin{cases} 0 & \text{if } A \text{ does not occur} \\ 1 & \text{if } A \text{ occurs} \end{cases}$$

$$P(I_A = 1) = P(A)$$

The Binomial distribution

Binomial distribution

- n *independent* experiments or trials are performed
- n is *fixed*
- Each experiment results in a success with a probability p, and failure with a probability $1 - p$

Then total number of successes is a binomial r.v. with parameters n and p.

X can take on values $0, 1, \ldots, n$ and

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

Why ?

The Binomial distribution

- Each configuration with k successes is equally likely $\longrightarrow p^k(1-p)^{n-k}$
- How many such configurations ? $\longrightarrow \binom{n}{k}$

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

Example : Power supply problem

- N customers use electric power intermittently (say $N = 100$)
- Assume each customer has the same probability p of requiring power at any given time t (say $p = 1/3$)

Suppose power is adjusted to k power units. How large should we choose k s.t. the probability of overload is $\leq .1\%$?

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

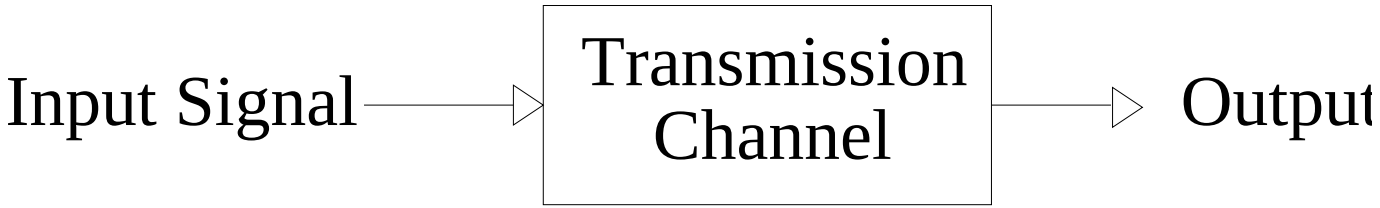
Take the minimum k s.t. $P(X > k) \leq 10^{-4}$, which gives

$$k \sim 51 \qquad .01\%$$

$$k \sim 48 \qquad .1\%$$

$$k \sim 45 \qquad 1\%$$

Example : Digital Communications



Channel is noisy : each bit has probability p of being incorrectly transmitted. To improve reliability, send bits a repeated number of times n. Say n odd $\longrightarrow$ Majority decoder

$$X = \#\text{errors} \sim \text{Binomial}(n, p)$$

$$P(\text{Bit correctly decoded}) = P(X < n/2)$$

E.g. $n = 5, \quad p = .1$

$$P(\text{Bit correctly decoded}) = P(X \leq 2) = .9914$$

$\longrightarrow$ Improvement in reliability.

Continuous random variables

A R.V, X is a mapping

$$\begin{aligned} X \quad &: \quad \Omega \to \mathbb{R} \\ &\qquad \omega \mapsto X(\omega) \end{aligned}$$

- Cumulative distribution function (CDF) :

$$F(x) = P(X \leq x)$$

- X is said to be a continuous random variable if there exists f s.t.

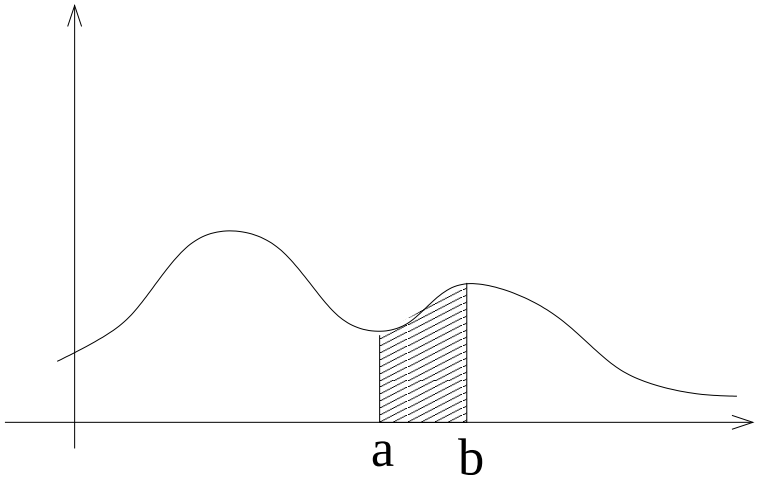
$$P(X \in A) = \int_A f(x)\, dx$$

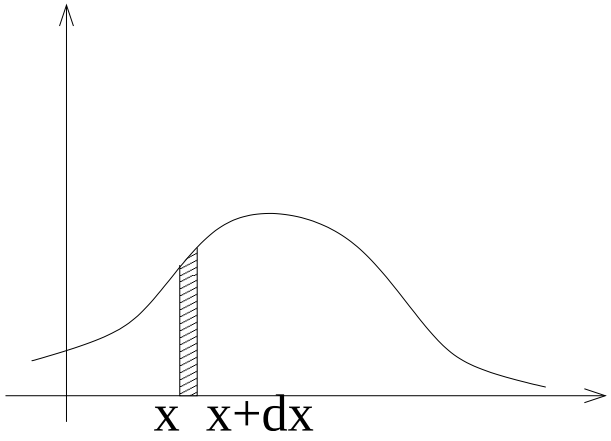
for any 'reasonable' set A

$\to$ f is called the density function

Continuous random variables

Interpretation





The normal distribution

Central role in probability and statistics

- Introduced by Gauss to model measurement errors
- Sum of independent r.v.'s is approximately normal (CLT)

Two parameters μ (mean), σ (standard deviation)

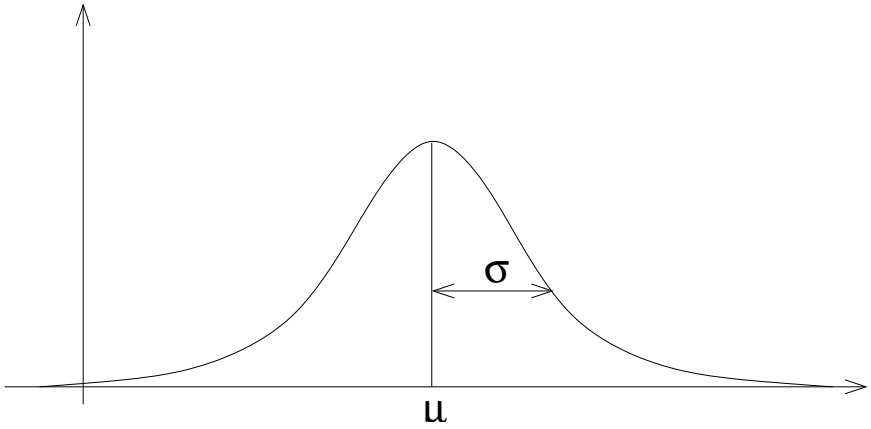
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Notation : $X \sim N(\mu, \sigma^2)$

"X follows a normal distribution with parameters μ and σ^2"

The normal distribution

This is the bell curve



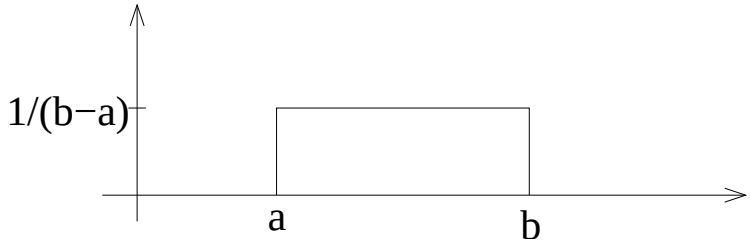
- Symmetric around μ
- Bell-shaped
- Spread is about σ

The uniform distribution

Interval $[a, b]$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Notation : $X \sim U[a, b]$



Expected value of a random variable

Definition The expected value of a r.v. X is denoted by $E(x)$ and defined by

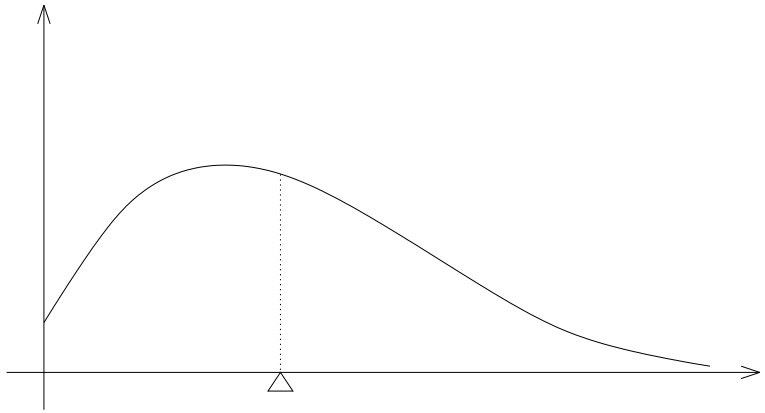
$$E(X) = \sum_i x_i p(X = x_i) = \sum_i x_i p_i$$

provided that the sum is absolutely convergent, i.e.

$$\sum_i |x_i| p(x_i) < \infty$$

Other names : Expectation, Mean

Interpretation



$E(x)$: point at which the histogram balances out

Expected value

Example $X \sim$ Binomial(n, p)

$$E(X) = np$$

i.e. np is the mean of the r.v. X.

For a continuous r.v., sum is replaced by an integral.

Definition X continuous r.v. with density $f(x)$, then

$$E(X) = \int x f(x)\, dx$$

Example : $X \sim N(\mu, \sigma^2)$

$$E(X) = \mu$$

Expectations of functions of r.v.'s

X random variable. We are not interested in $E(X)$ but in $E(g(X))$.

Example : Kinetic theory of gas : distribution of the modulus of the velocity v

$$f_V(v) = \frac{\sqrt{2/\pi}}{\sigma^3} v^2 e^{-v^2/(2\sigma^2)}$$

and we are interested in $Y = \frac{1}{2} m X^2$, the kinetic energy. What is $E(Y)$?

Theorem : Let $Y = g(X)$

- X discrete with distribution function $p(x)$

$$E(Y) = \sum_x g(x) p(x)$$

  provided sum is absolutely convergent

- X with density $f(x)$

$$E(Y) = \int g(x) f(x)\, dx$$

Expectations of functions of r.v.'s

Proof (discrete case)

$$E(Y) = \sum_i y_i P(Y = y_i)$$

Let A_i be the set of x's such that $g(x) = y_i$.

$$P(Y = y_i) = p_Y(y_i) = \sum_{x \in A_i} p(x)$$

$$E(Y) = \sum_i y_i p_Y(y_i) = \sum_i \sum_{A_i} g(x)p(x)$$

$$= \sum_x g(x)p(x)$$

Expectations of functions of r.v.'s

Example : Kinetic theory of gas

$$
\begin{aligned}
E(Y) &= \int \frac{1}{2} m x^2 f_X(x)\, dx \\
&= \frac{m}{2} \frac{\sqrt{2/\pi}}{\sigma^3} \int_0^\infty x^4 e^{-x^2/(2\sigma^2)}\, dx \\
&= \frac{m}{2} \sqrt{2/\pi}\sigma^2 \int_0^\infty u^4 e^{-u^4/2}\, du \\
&= \frac{3}{2} m\sigma^2
\end{aligned}
$$

Linearity of the expectation

- $X_1, \ldots, X_n$ n r.v.'s with expectation $E(X_i)$
- $Y = b_1 X_1 + \ldots + b_n X_n$

Then

$$E(Y) = b_1 E(X_1) + \ldots + b_n E(X_n)$$

e.g. the expected value of a sum of r.v.'s is equal to the sum of their expected values.

Example : Hat problem

- n men attend a dinner
- leave their hat
- take a random hat when they leave

N : $\#$ of men who pick their own hat. What is $E(N)$?

Linearity of the expectation

$N = I_1 + \ldots + I_n$ with

$$I_i = \begin{cases} 1 & \text{if person } i \text{ picks his hat} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(N) &= E(I_1 + \ldots I_n) \\ &= E(I_1) + \ldots + E(I_n) \\ E(I_i) &= P(I_i = 1) = \frac{1}{n} \\ \Rightarrow E(N) &= 1 \end{aligned}$$