STATISTICS AND PROBABILITY

Table des matières

Chapitre 1

Continuous distributions

1.1 General presentation

A continuous random variable (CRV) take on values at every point over a given interval. Thus CRV have no gap or unassumed values.

- Measuring the thickness of an item
- time required to complete a task
- temperature of a solution

-Measure of time, temperature, weight, height, ...

These can potentially take on any value, depending only on the ability to measure accurately.

The probability of a specific value for continuous variable is zero. The probability function for discrete variable is not applicable any more. We introduce density function for the probability function for continuous random variables.

– **The probability density function**

The probability density function $f(x)$, of random variable X has the following properties :

1. $f(x) > 0$ for all value of x
2. The area under the probability density function $f(x)$ over all values of the random variable X within its range, is equal to 1.
3. The probability that X lies between two values is the area under the density function graph between the two values.
   $P(a < X < b) = \int_a^b f(x)dx$
4. The cumulative density function $F(x_0)$ is the area under the probability density function $f(x)$ from the minimum x value up to x_0. $F(x_0) = \int_{x_m}^{x_0} f(x)dx$ where x_m is the minimum value of the random variable x.

– **Probability as an Area**

- Shaded area under the curve is the probability that X is between a and b. (Note that the probability of any individual value is zero)

- The total area under the curve $f(x)$ is 1

- The area under the curve $f(x)$ to the left of x_0 is $F(x_0)$, where x_0 is any value that the random variable can take. $F(x_0) = P(X \leq x_0)$

1.2 Uniform distribution

The uniform distribution is a probability distribution that has equal probabilities for all equal-width intervals within the range of the random variable

Solving a uniform distribution problem :

Density of probability : $f(x) = \begin{cases} \frac{1}{b-a} \text{ if } a \leq x \leq b \\ 0 \text{ if not} \end{cases}$ where : $f(x)$ = value of the density function at any x value, a = minimum value of x, and b = maximum value of x.

The mean is : $\mu = \frac{a+b}{2}$ and the standard deviation $\sigma^2 = \frac{(b-a)^2}{12}$

Examples : Problem 6.1, 6.2 page 202 and 2013.

Problems 6.1 and 6.2 page 204

1.3 Normal distribution

The normal distribution closely approximates the probability distributions of a wide range of random variables

- Total sales or production

- The patterns of stock and bond prices are often modeled using the normal distribution

- Distributions of sample means approach a normal distribution given a "large" sample size

Density of probability : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$ where :

μ is the mean of x, σ the standard deviation of x, $\pi = 3.1416$, $e = 2.7183$ $x =$ any value of the continuous variable from $-\infty$ to $+\infty$

- It is bell shaped, Mean, Median and Mode are Equal, Location is determined by the mean, μ and Spread is determined by the standard deviation σ. It is denoted $X \sim N(\mu, \sigma)$

Solving a normal problem :

We introduce cumulative distribution function to obtain probabilities for a specified normal distribution :

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

The probability for a range of values is measured by the area under the curve.

$P(a < X < b) = F(a) - F(b)$ GRAPHIC here.

To compute the probability for normal distribution with a specific mean and variance, we convert to standard normal distribution with mean 0 and variance 1

Indeed, the standardized normal distribution with mean 0 and variance 1 is denoted Z. $Z \sim N(0,1)$

The result : If $X \sim N(\mu, \sigma)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Reading the Standard Normal Distribution table :

Note When $np > 5$ and $n(1-p) > 5$ the binomial distribution can be approximated to Normal distribution with mean np and standard deviation $\sqrt{np(1-p)}$.

However, the translation of a discrete distribution to a continuous distribution is not completely straightforward. correction of +0.5 or - 0.5 or $\mp$0.5 depending on the problem , is required This correction ensures that most of the binomial problem´s information is correctly transferred to the normal curve analysis. This correction is called the correction for continuity, which is made during conversion of a discrete distribution into a continuous distribution. In the following table are some guidelines for the correction for continuity.

Values being determined	Correction
$x >$	$+0.5$
$x \geq$	-0.5
$x <$	-0.5
$x \leq$	$+0.5$
$\leq x \leq$	-0.5 and $+0.5$
$< x <$	$+0.5$ and -0.5
$x =$	-0.5 and $+0.5$

1.4 Exponential distribution

It is a continuous distribution closely related to the Poisson distribution. Whereas the Poisson distribution is discrete and describes random occurrences over some interval, the exponential distribution is continuous and describes a probability distribution of the times between random occurrences.

– Used to model the length of time between two occurrences of an event (the time between arrivals)

Examples :

- Time between trucks arriving at an unloading dock
- Time between transactions at an ATM Machine
- Time between phone calls to the main operator

– It is a Random variables with positive values ; its distribution is not symmetric. It is skewed to the right, The x values range from zero to infinity, Its apex is always at $x = 0$. The exponential random variable T $(t > 0)$ has a probability density function $f(t) = \lambda e^{-\lambda t}$ for all $t > 0$.
  Where, λ is the mean number of occurrences per unit time, t is the number of time units until the next occurrence and $e = 2.71828$
  T is said to follow an exponential probability distribution defined by a single parameter, its mean λ (lambda).
  Graph, page 224.

– The distribution has a mean of $\frac{1}{\lambda}$ and a variance of $\frac{1}{\lambda^2}$.

– Probabilities of the Right tail of the Exponential Distribution are given by :
  $P(x \geq x_0) = e^{-\lambda x_0}$
  $x_0 \geq 0$ is the fraction of the interval or the number of intervals between arrivals in the probability question and λ is the average arrival rate.

Example 1 : Arrivals at a bank are Poisson distributed with a λ of 1.2 customer every minute. What is the average time between arrivals and what is the probability that at least 2 minutes will elapse between one arrival and the next arrival ?

Example 2 : A manufacturing firm has been involved in statistical quality control for several years. As part of the production process, parts are randomly selected and tested. From the records of these tests, it has been established that a defective part occurs in a pattern that is Poisson distributed on the average of 1.38 defects every 20 minutes during production runs. Use this information to determine the probability that less than 15 minutes will elapse between any two defects.