MAT 2377 3X (Spring 2011)

§4.1-4.4 **Continuous Random Variables**

Definition : A random variable X is said to be continuous if its cumulative distribution function F_X is a continuous function.

Definition : Let X be a continuous random variable with the c.d.f. F_X. The **probability density function** of X is

$$f(x) = \begin{cases} F'(x), & \text{if } F'(x) \text{ exists,} \\ 0, & \text{otherwise,} \end{cases}$$

where F' is the derivative of F.

Note : We sometimes denote the probability density function (p.d.f.) of X as f_X.

Example 1 : Consider a Poisson process with a rate λ. Let X be the length of the interval required to observe a change in the Poisson process. Show that the c.d.f. of X is

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

Note : The function F is a continuous function.

Properties of f :

1. $f(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. **[Computational Property]** Let $A \subseteq \mathbb{R}$, then

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

   In particular,

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

Note : if a function $f : \mathbb{R} \to \mathbb{R}$ satisfies the above properties (1) and (2), then we say that f is a probability density function.

Other properties of f :

1. A single value does not have a mass, i.e. for all $x \in \mathbb{R}$,

$$P(X = x) = 0.$$

2. Let $a, b \in \mathbb{R}$, such that $a < b$, then

$$\begin{aligned} & P\left(a < X < b\right) = P\left(a \leq X < b\right) \\ = \;& P\left(a \leq X \leq b\right) = P\left(a < X \leq b\right) \\ = \;& \int_a^b f(x)\, dx = F(b) - F(a) \end{aligned}$$

Definition : Let X be a continuous random variable with the probability density function f. Let $h : \mathbb{R} \to \mathbb{R}$ be a function from the set of real number to the set of real numbers. The **expected value** of $h(X)$ is

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

Definition : Let X be a continuous random variable with the p.d.f. f. Its *mean* (also called its *expected value*) is defined as

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

Its *variance* is defined as

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx$$

Its *standard deviation* is defined as

$$\sigma = \sqrt{V(X)}.$$

Alternative formula for the variance : The following computational formula can be used to compute the variance.

$$\sigma^2 = V(X) = E[X^2] - \mu^2 = \left(\int_{-\infty}^{\infty} x^2\, f(x)\, dx \right) - \mu^2.$$

Example 2 : The time to complete a task (in hours) can be represented as a continuous random variable X with the following probability density function :

$$f(x) = \begin{cases} 0.2, & 0 < x < 4 \\ 0.04, & 4 < x < 9 \\ 0, & \text{otherwise} \end{cases}$$

(a) Determine the cumulative distribution function of X.

(b) Determine the probability that the time to complete the task will be more than 2 hours, but less than 7 hours.

Note : Use two different methods to obtain the answer.

(c) Determine a time x such that 90% of the tasks are completed in at most x hours.

(d) Determine the mean and the variance of X.

(e) The cost of a task is \$1500 per hour. What is the expected cost of a task ?

(f) What is the variance of the cost of a task ?

§4.8 **Exponential Distribution**

Definition : Let X be the length of an interval required to observe a change in the Poisson process with rate λ. We say that X has an exponential distribution.

From Example 1, we know that its c.d.f. is

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

We differentiate to obtain its p.d.f.

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

Its mean and variance are, respectively

$$\mu = E[X] = \frac{1}{\lambda} \qquad \text{and} \qquad \sigma^2 = V(X) = \frac{1}{\lambda^2}$$

Lack of memory property of the exponential :
Let X be an exponential random variable and let s and t be two positive real numbers. Then,

$$P(X > s + t \mid X > s) = P(X > t).$$

Example 3 : The errors on a data tape occur at a rate 1 error per 35,000 centimeters. We assume that the errors can be modeled as a Poisson process.

(a) Determine the length of tape that is required to be 90% certain that there are no errors.

(b) We verify a tape for errors. There are no errors on the first 20,000 centimeters. What is the probability that there are no errors on the next $30,000$ centimeters ?

§4.9 **Erlang Distribution**

Definition : Let X be the distance required for r changes in a Poisson process of rate λ. We say that X has an *Erlang* distribution with parameters λ and r.

It c.d.f. is (for $x > 0$) :

$$\begin{aligned} F(x) &= P(X \leq x) = 1 - P(X > x) \\ &= 1 - P(\text{“at most } r-1 \text{ changes} \\ &\qquad\qquad \text{in the interval } [0,x]\text{”}) \\ &= 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}\,(\lambda\, x)^k}{k!} \end{aligned}$$

Its mean and variance are, respectively

$$\mu = E[X] = \frac{r}{\lambda} \qquad \text{and} \qquad \sigma^2 = V(X) = \frac{r}{\lambda^2}$$

Example 4 : Suppose that counts recorded by a Geiger counter follow a Poisson process with an average of three counts per minute.
(a) What is the probability that the waiting time for 2 counts will be at most 30 seconds ?
(b) What is the mean time between counts ?
(c) Compute the expected waiting time to observe 4 particles.
(d) Give the standard deviation for the waiting time to observe 4 particles.

§4.5 **Uniform Distribution**

Definition : Let X be a random variable with the following probability density function

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

We say that X follows a uniform distribution on the interval $[a, b]$.

The cumulative distribution function for X is

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

The means of X is $\mu = E[X] = \dfrac{a+b}{2}$.

The variance of X is $\sigma^2 = V(X) = \dfrac{(b-a)^2}{12}$.

§4.6 Normal Distribution :

Definition : Let μ be a real number and σ be a positive real number. A random variable X with the probability density function (p.d.f.)

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

is called a normal random variable with the parameters μ and σ.

Note : The mean and the variance of X are

$$E[X] = \mu \quad \text{and} \quad V[X] = \sigma^2.$$

Notation : $X \sim N(\mu, \sigma^2)$ means that X has a normal distribution with mean μ and variance σ^2.

Definition : Let Z be a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$, then we say that X has a standard normal distribution. Its p.d.f. is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

It c.d.f. is

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)\, dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Remarks :

1. Some values for $\Phi(z)$ are found in Table III in the Appendix A.
2. ϕ is symmetric about $z = 0$.
   (a) $\Phi(0) = P(Z \leq 0) = 0.5$
   (b) $P(Z \leq -z) = P(Z \geq z)$

Example 5 :
Suppose that Z has a standard normal distribution, i.e. $Z \sim N(0,1)$. Determine

1. $P(0.53 < Z < 2.06)$
2. $P(-2.63 \leq Z \leq 0.53)$
3. $P(Z > 14.5)$
4. c such that $P(Z > c) = 0.10$
5. c such that $P(-c < Z < c) = 0.95$

Note : The following Theorem informs us that we can always transform a normal random variable in order to obtain a standard normal random variable.

Standardization Theorem : Let $X \sim N(\mu, \sigma^2)$. If

$$Z = \frac{X - E[X]}{\sqrt{V(X)}} = \frac{X - \mu}{\sigma}$$

then Z has a $N(0,1)$ distribution.

Consequence : Let $X \sim N(\mu, \sigma^2)$.

$$\begin{aligned} F_X(x) &= P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

Moreover,

$$\begin{aligned} P(a \leq X \leq b) &= F_X(b) - F_X(a) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Example 6 : Suppose that the width (in cm) of some items produced in aluminium is normally distributed with a mean of 2 and a standard deviation of 0.007.
(a) Tolerance levels are 2.0 ± 0.012. What percentage of the items do not conform to the tolerance levels ?
(b) The width (in cm) of 90% of the items is below what value ?

§4.7 **Normal Approximation of the Binomial Distribution and of the Poisson Distribution**

Theorem [DeMoivre-Laplace] : Let X be binomial random variable with parameters n and p and let

$$Z = \frac{X - E[X]}{\sqrt{V[X]}} = \frac{X - np}{\sqrt{np(1-p)}}.$$

Then,

$$\lim_{n\to\infty} F_Z(z) = \Phi(z),$$

where Φ is the cumulative distribution function for the standard normal distribution.

Remarks :

1. Is n is large, then $F_Z(z) \approx \Phi(z)$.
2. The approximation is good when

$$n\,p > 5 \quad \text{and} \quad n\,(1-p) > 5$$

Theorem : Let X be a Poisson random variable with mean λ and let

$$Z = \frac{X - E[X]}{\sqrt{V[X]}} = \frac{X - \lambda}{\sqrt{\lambda}}.$$

Then,

$$\lim_{\lambda\to\infty} F_Z(z) = \Phi(z),$$

where Φ is the c.d.f. for the standard normal.

Remarks :

1. If λ is large, then $F_Z(z) \approx \Phi(z)$.
2. The approximation is good when

$$\lambda > 5$$

Steps to follow for the normal approximation :

- **[Step 1] :** Verify that the rule of thumb for a good approximation is satisfied.
- **[Step 2] :** Express the event in terms of inclusive inequalities, for example $P(a \leq X \leq b)$.
- **[Step 3 - continuity correction] :** Add 0.5 to the upper limit and subtract 0.5 from the lower limit.

$$P(a \leq X \leq b) = P(a - 0.5 < X < b + 0.5).$$

- **[Step 4 - normal approximation] :** Standardize and approximate the probability of the event with the c.d.f. for the standard normal.

$$\begin{aligned} & P(a \leq X \leq b) \\ = \; & P(a - 0.5 < X < b + 0.5) \\ \approx \; & \Phi\left(\frac{b + 0.5 - \mu_X}{\sigma_X}\right) - \Phi\left(\frac{a - 0.5 - \mu_X}{\sigma_X}\right) \end{aligned}$$

Example 7 : Suppose that X has a Poisson distribution with mean $\lambda = 10$. Approximate the following probabilities :

(a) $P(X = 11)$. **(b)** $P(9 \leq X \leq 11)$.

Example 8 : Suppose that X has a binomial distribution with parameters $n = 100$ et $p = 0.06$.

Approximate $P(5 \leq X < 9)$.