Stochastic Optimization - I

Review of Basic Probability Theory

Random Inflows

Introduction

- Decisions relating to most water resources systems need to be made in the face of hydrologic uncertainty.
- The hydrologic variables such as rainfall in a command area, inflow to a reservoir, evapotranspiration of crops which influence decision making in water resources, are all random variables.
- Optimization models developed for water resources management must therefore be formulated to give optimal decisions with an indication of the associated hydrologic uncertainty.

Stochastic Optimization - Approaches

- Two classical approaches to deal with the hydrologic uncertainty in optimization models are
  - Implicit Stochastic Optimization (ISO) and
  - Explicit Stochastic Optimization (ESO).
- Implicit Stochastic Optimization (ISO)
  - Hydrologic uncertainty is implicitly incorporated
  - Optimization model itself is a deterministic model, in which the hydrologic inputs are varied with a number of equi-probable sequences and the deterministic optimization model is run once with each of the input sequences.
  - Output set is then statistically analyzed to generate a set of optimal decisions.

Stochastic Optimization - Approaches

- Explicit Stochastic Optimization (ESO)
  - Stochastic nature of the inputs is explicitly included in the optimization model through their probability distributions.
  - Optimization model is a stochastic model and a single run of the model specifies the optimal decisions.
  - Two commonly used ESO techniques
    - Chance Constrained Linear Programming (CCLP), and
    - Stochastic Dynamic Programming (SDP)
- Background of probability theory is essential for ESO

Review of Basic Probability Theory

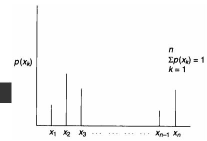
- ***Random Variable***
  - A variable whose value is not known or cannot be measured with certainty (or is nondeterministic) is called a random variable.
  - Examples of random variables of interest in water resources are rainfall, streamflow, time between hydrologic events (e.g. floods of a given magnitude), evaporation from a reservoir, groundwater levels, re-aeration and de-oxygenation rates etc.
  - Any function of a random variable is also a random variable (r.v).
  - We use an upper case letter to denote a random variable and the corresponding lower case letter to denote the value that it takes.
  - For example, daily rainfall may be denoted as X. The value it takes on a particular day is denoted as x.
  - We then associate *probabilities* with events such as $X \geq x$, $0 \leq X \leq x$, etc.

Discrete and Continuous Random Variables

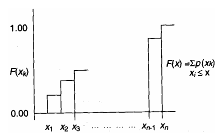
- If a r.v. X can take on only discrete values $x_1, x_2, x_3, \ldots$, then X is a discrete random variable.
- An example of a discrete random variable is the number of rainy days in a year which may take on values such as, 10, 20, 30, ...
- A discrete random variable can assume a finite number of values.
- If a r.v. X can take on *all* real values in a range, then it is a continuous random variable.
- Most variables in hydrology are continuous random variables.
- The number of values that a continuous random variable can assume is infinite.

Probability Distributions

- For a discrete random variable, there are spikes of probability associated with the values that the random variable assumes.
- For discrete random variables, the probability distribution is called a *probability mass function* and in case of continuous random variables it is called a *probability density function* (*pdf*).
- The cumulative distribution function, $F(x)$, represents the probability that X is less than or equal to x_k, and is shown in Figure for a discrete r.v. i.e. $F(x_k) = P(X \le x_k)$



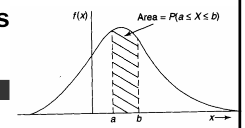
Probability mass function



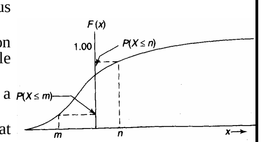
Cumulative Probability distribution

Probability Distributions

- The probability density function (pdf) of a continuous random variable is denoted by $f(x)$.
- Probability distributions of continuous random variables are smooth curves.
- The cumulative distribution function (cdf) of a continuous random variable is denoted by $F(x)$.
- It is a non-decreasing function with a maximum value of 1.
- The cdf represents the probability that X is less than or equal to x, i.e. $F(x) = P(X \le x)$.



Probability density function (pdf)



Cumulative Probability distribution (cdf)

Probability Distributions

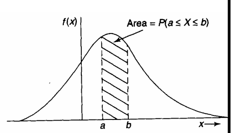
- Any function $f(x)$ defined on the real line can be a valid probability density function if and only if
  - $f(x) \ge 0$ for all x, and
  - $f(x) \le 1$ for all x
- The pdf and the cdf are related by

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

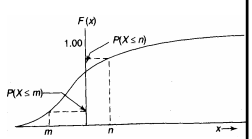
- For a continuous random variable, probability of the random variable taking a value exactly equal to a given value is zero, because

$$P(X = d) = P(d \le X \le d) = \int_d^d f(x)\,dx = 0$$

Area under the curve to the left of $x = a$ is $\text{Prob}(X \le a)$
Area under the curve to the left of $x = b$ is $\text{Prob}(X \le b)$
Area between $x = a$ and $x = b$ is $\text{Prob}[a \le X \le b]$



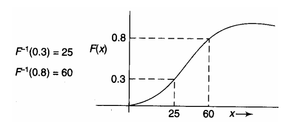
Probability density function (pdf)



Cumulative Probability distribution (cdf)

Probability Distributions

For any given α, $0 \le \alpha \le 1$, we may determine a value x from the cumulative distribution such that $F(x) = \alpha$. We then denote, $x = F^{-1}(\alpha)$.



Expected Value of X, E(X)

Expected value of X, $E(X)$

$$E(X) = \sum_k x_k\, p(x_k) \quad \text{for discrete r.v.s}$$

where $p(x_k)$ is $\text{Prob}(X = x_k)$;

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{for continuous r.v.s}$$

where $f(x)$ is the pdf of the r.v. X.
The mean of an r.v. X, denoted as μ, is equal to the expected value, i.e.

$$\mu = E(X)$$

Variance of X, Var(X)

$$\text{Var}(X) = \sigma^2 = E(X - \mu)^2$$

$$= \sum_k (x_k - \mu)^2\, p(x_k) \quad \dots X \text{ discrete}$$

$$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx \quad \dots X \text{ continuous}$$

Standard deviation, $\sigma = +\sqrt{\sigma^2}$ (+ve square root of variance).
Coefficient of variation, C_v

$$C_v = \frac{\sigma}{\mu}$$

Example

Probability density function (pdf) of a random variable X is

$f(x) = 3x^2 \qquad 0 \le x \le 1$

$\quad = 0 \qquad$ else where

Determine 1. cumulative distribution function (cdf)

2. Expected value, E(X); 3. Variance, Var (X); 4. $P[X \ge 0.6]$; 5. $P[0.4 \le X \le 0.7]$

1. Cumulative distribution function,

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^x 3x^2\,dx = x^3 \qquad 0 \le x \le 1$$

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

$$E(X) = \int_0^1 x\, 3x^2\,dx = 3/4$$

Example – contd.

3. $\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$$\mathrm{Var}(X) = \int_0^1 (x-(3/4))^2\, 3x^2\,dx$$

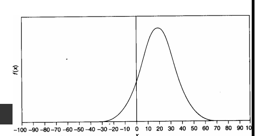
$$= \left[\frac{3x^5}{5} + \frac{27x^3}{48} - \frac{18x^4}{16}\right]_0^1 = \frac{3}{80} = 0.0375$$

4. $P[X \ge 0.6] = 1 - P[X \le 0.6]$

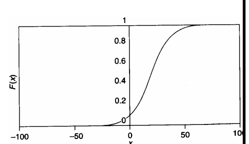
$= 1 - F(0.6)$

$= 1 - (0.6)^3 = 0.784$

Normal Distribution



Normal probability density function (pdf)



Cumulative Probability distribution (cdf)

- Three commonly used distributions in water resources are: Normal, Lognormal and Exponential distributions.
- The *pdf* of the normal distribution is given by $f(x)$.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty \le x \le +\infty$$

- The cdf of Normal distribution, i.e. $F(x)$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad -\infty \le x \le \infty$$

Normal Distribution

- Mean and standard deviation are two important parameters of Normal distribution.
- Standardization, $Z=(X-\mu)/\sigma$
- Z will follow normal distribution with N(0,1).
- PDF of Z is symmetrical about zero.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \qquad -\infty \le z \le +\infty$$

- Values of $\phi(z)$ obtained by numerical integration are used in the computations for normal distributions.

Normal Distribution Example

The monthly streamflow at a reservoir site is represented by a random variable *X* which follows normal distribution with a mean of 30 units and a standard deviation of 15 units.

Find 1. P[X>45]; 2. *P*[X < 20] and

3. The flow value which will be exceeded with a probability of 0.9.

1. $P[X \ge 45] = P[((X-\mu)/\sigma) \ge ((45-30)/15)]$

$= P[Z \ge 1]$

$= 1 - P[Z \le 1]$

$= 1 - 0.8413$ from Table 6.1

$= 0.1587$

2. $P[X \le 20] = P[Z \le ((20-30)/15)]$

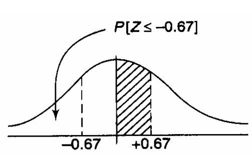
$= P[Z \le (-10/15)]$

$= P[Z \le -0.67]$

$= 0.5 -$ (Area under the std. normal curve between 0 and $+ 0.67$)

$= 0.5 - 0.2486$ (from Table 6.1)

$= 0.2514$



Normal Distribution Example - Contd.

3. The problem is to find x such that $P[X \ge x] = 0.9$

$P[X \ge x] = 0.9$

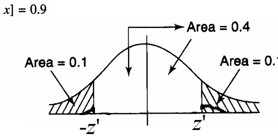
$P[Z \ge ((x-\mu)/\sigma)] = 0.9$

i .e. $P[Z \ge ((x-30)/15)] = 0.9$

i.e. $1 - P[Z \le z] = 0.9$

where $z = (x-30)/15$

i.e. $P[Z \le z] = 0.1$



$(x-30)/15 = -1.28$

$x = 10.8$ units.

Thank You