Random Variables

- A random variable is a function whose value depends on the outcome of a chance event
- *X* is a random variable,
  x is a possible value of *X* (a realization of *X*)
- Discrete RV
  - Takes on values from a discrete set

\# of years until a certain flood stage returns
\# of times reservoir storage drops below a level

- Continuous
  - Takes on values from a continuous set

e.g., Rainfall, Streamflow, Temperature, Concentration

Independent RVs

- If the distribution of *X* is not influenced by the value taken by *Y*, and vice versa, the two random variables are said to be *independent*.
- For two independent random variables, the joint probability is the product of the separate probabilities.

$$\Pr\{a \le X \le b \text{ and } c \le Y \le d\} = \Pr\{a \le X \le b\}\Pr\{c \le Y \le d\}$$

$$F_{XY}(x,y) = F_X(x)F_Y(y) \qquad f_{XY}(x,y) = f_X(x)f_Y(y)$$

Marginal Distributions

- Marginal CDF of *X*, is CDF of *X* ignoring *Y*

$$F_X(x) = \Pr\{X \le x\} = \lim_{y\to\infty} F_{XY}(x,y)$$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy$$

Conditional Distributions

Conditional CDF is the CDF for *X* given that *Y* has taken a particular value

$$F_{X|Y}(x|y) = \Pr\{X \le x \text{ given } Y = y\}$$

$$F_{X|Y}(x|y) = \Pr\{X \le x | Y = y\} = \frac{\int_{-\infty}^{x} f_{XY}(s,y)ds}{f_Y(y)}$$

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}$$

Expectation

$$E[g(X)]=\int_{-\infty}^{+\infty} g(x)f_X(x)dx \qquad X \text{ is continuous}$$

$$E[g(X)]=\sum_{x_i} g(x_i)p_X(x_i) \qquad X \text{ is discrete}$$

Note

$$E[X]=\int_{-\infty}^{+\infty} xf_X(x)dx \qquad E[X]=\sum_{x_i} x_i p_X(x_i)$$

Example1:

- Principle - Replacement of uncertain quantities by either expected, median or worst-case values can grossly affect the evaluation of project performance when important parameters are highly variable.
- elevation of reservoir water surface varies from year to year depending on the inflow and demand for water.

المبدأ - يمكن أن يؤدي استبدال الكميات غير المؤكدة من خلال القيم المتوقعة أو المتوسطة أو أسوأ الحالات إلى تأثير كبير على تقييم أداء المشروع عندما تكون المعلومات المهمة متغيرة بدرجة كبيرة.
يتفاوت ارتفاع سطح مياه الخزان من سنة إلى أخرى حسب التدفق والطلب على المياه.

possible pool levels	probability of each level	recreation potential in visitor-days per day for reservoir with different pool levels
10	0.10	25
20	0.25	75
30	0.30	100
40	0.25	80
50	0.10	70

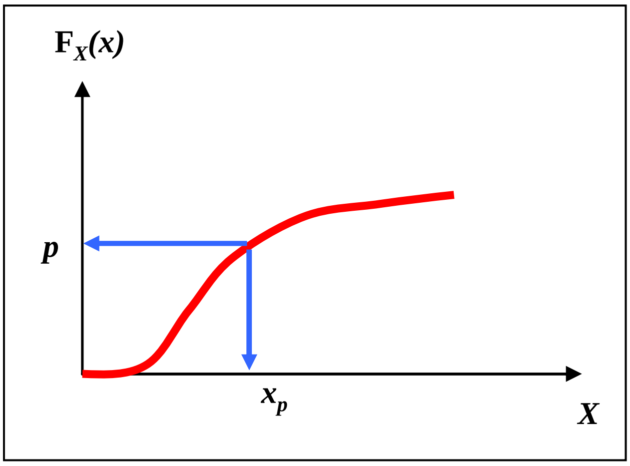
Average pool level

$$\overline{L} = 10(0.10) + 20(0.25) + 30(0.30) + 40(0.25) + 50(0.10) = 30$$

$$VD(\overline{L}) = 100$$

Average Visitation Rate

$$\overline{VD} = 0.10VD(10) + 0.25VD(20) + 0.30VD(30) + 0.25VD(40) + 0.10VD(50)$$
$$= 0.10(25) + 0.25(75) + 0.30(100) + 0.25(80) + 0.10(70)$$
$$= 78.5 \text{ visitor - days}$$



Quantiles

- X is RV
- p-th quantile is xp

$$F_X(x_p) = \Pr\{X < x_p\} = p$$

Median: $x_{0.50}$ equally likely to be above as below that value

Interquartile Range : $[x_{0.25}, x_{0.75}]$ range of values that the random variable might assume.

p^{th} quantile is also the 100-p percentile
Floodplain management - the 100-year flood $x0.99$
Water quality management - minimum 7-day-average low flow expected once in 10 years: 10 percentile of the distribution of the annual minima of the 7-day average flows

- Observed values, sample of size n

$$\{x_1, x_2, ..., x_n\}$$

- Order statistics (observations ordered by magnitude

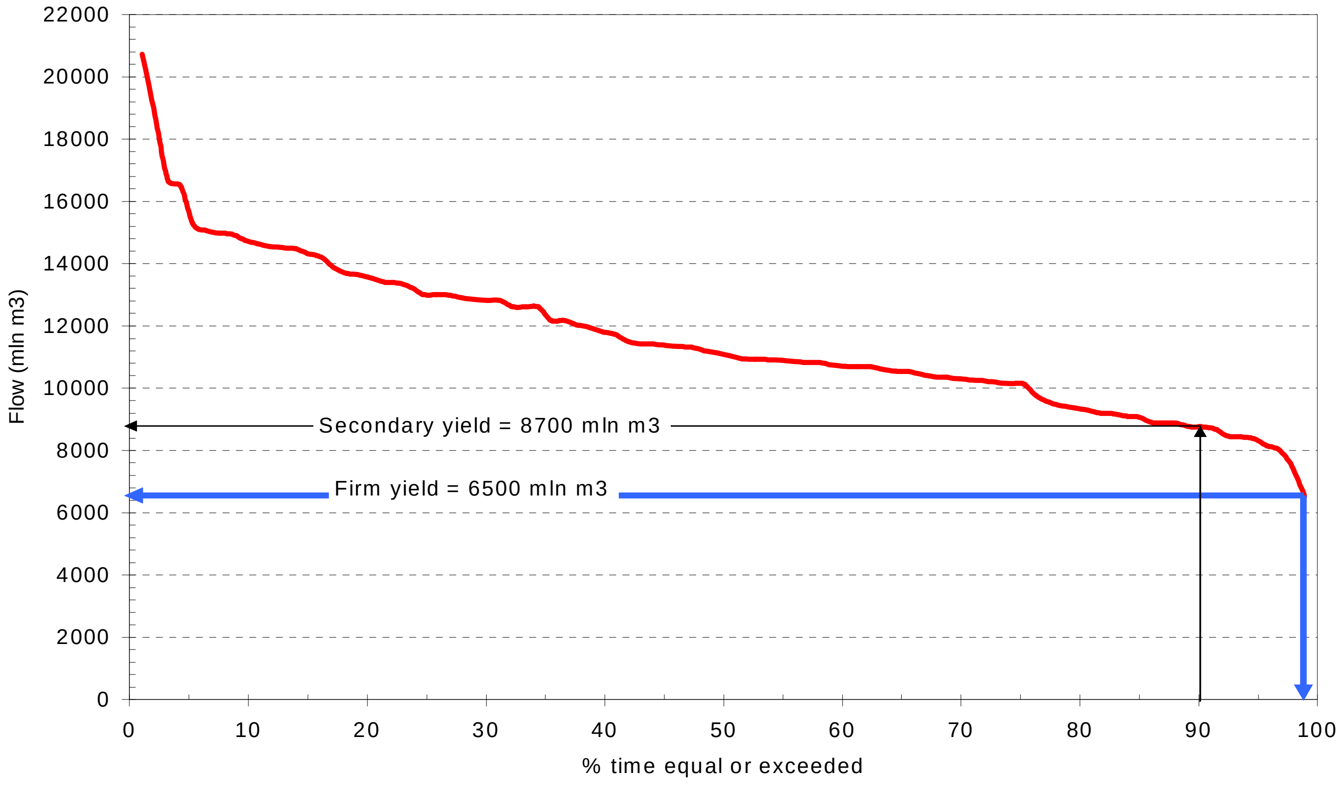
$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$$

Sampl $x_{(1)} = \text{smallest}$ $\qquad x_{(n)} = \text{largest}$ ng

$$x_{(i)} \approx x_p \qquad p = \frac{i}{n+1} \qquad p = \Pr\{X \le x_p\}$$

Year	Flow x	Rank i	$P(X>x)= 1-i/(n+1)= 1-p$	Ranked Flow $x_{(i)}$
1911	10817	1	0.99	6525
1912	11126	2	0.98	7478
1913	11503	3	0.97	8014
1914	11428	4	0.96	8161
1915	10233	5	0.95	8378
...				
1997	10343	87	0.06	15062
1998	14511	88	0.05	15242
1999	14557	89	0.04	16504
2000	12614	90	0.03	16675
2001	12615	91	0.02	18754
2002	16675	92=n	0.01	20725

- **Flow duration curve - Discharge vs % of time flow is equaled or exceeded.**
- **Firm yield is flow that is equaled or exceeded 100% of the time**



Example2: The following table displays the joint probabilities of different weather conditions and of different recreation benefit levels obtained from use of a reservoir in a state park:

	Possible Recreation Benefit Level		
Weather	RB_1	RB_2	RB_3
Wet	0.10	0.20	0.10
Dry	0.10	0.30	0.20

(a) Compute the probabilities of recreation levels *RB1*, *RB2*, and *RB3*, and of *dry* and *wet* weather.

$\Pr\{RB1\} = 0.1 + 0.1 = 0.2$
$\Pr\{RB2\} = 0.2 + 0.3 = 0.5$
$\Pr\{RB3\} = 0.1 + 0.2 = 0.3$

$\Pr\{\text{Dry}\} = 0.1 + 0.3 + 0.2 = 0.6$
$\Pr\{\text{Wet}\} = 0.1 + 0.2 + 0.1 = 0.4$

(b) Compute the conditional probabilities:

P(wet| *RB1*), P(*RB3*| Dry), P(*RB2*|wet)

P(wet| *RB1*) = 0.1/(0.1 + 0.1) = 0.5
P(*RB3*| Dry) = 0.2/0.6 = 1/3
P(*RB2*|wet) = 0.2/0.4 = 0.5

Example3: In flood protection planning, the 100-year flood, which is an estimate of the quantile $x_{0.99}$, is often used as a design flow. Assuming that the floods in different years are independently distributed:

(a) Show that the probability of at least one 100-year flood in a 5-year period is 0.049.
Pr{at least one 100 year flood in 5 years}
= 1 – Pr{no 100 year flood in 5 years}
$= 1 - (0.99)^{5}$
= 0.049

(b) What is the probability of at least one 100-year flood in a 100-year period?
Pr{at least one 100-year flood in a 100-year period}
= 1 – Pr{no 100 year flood in 100 years}
$= 1 - (0.99)^{100}$
= 0.634

(c) If floods at 1000 different sites occur independently, what is the probability of at least one 100-year flood at some site in any single year?

Pr{at least one 100-year flood at some site in one year} = 1 – Pr{no 100 year flood at any site in 1 year}
$= 1 - (0.99)^{1000}$
= 0.999957