Lecture 6

Continuous Probability Distributions

Statistics for
Civil & Environmental Engineers

Uniform Distribution(1)

- **Definition**
  - **A uniformly distributed random variate can have any value in an interval *a* to *b* with equal likelihood**
- **Uniform pdf**
  - $$f_X(x) = \frac{1}{b-a} \qquad \text{for} \quad a \le x \le b$$
    $$= 0 \qquad \text{otherwise}$$
    $$\text{where } a, b = \text{population parameters}$$
- **Population parameters**
  - **Mean=**
  - **Variance=**

Uniform Distribution(2)

- **Parameter estimators**
  - **MOM:** $\hat{a} = \bar{x} - \sqrt{3}s$, $\hat{b} = \bar{x} + \sqrt{3}s$
  - **MLE:** $\hat{a} = x_{\min}$, $\hat{b} = x_{\max}$
- **Distribution shapes**

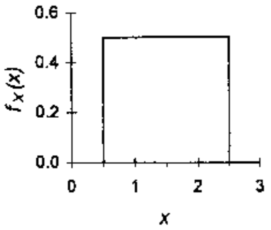


FIGURE 4.2.1
The pdf, $f_X(x)$, for a uniform (0.5, 2.5) variate.

Uniform Distribution Example

Example 4.19. Density of concrete. Consider the data of Table 1.2.1. This shows the densities and compressive strengths of concrete samples. It may be assumed (as in Chapter 3) that the marginal density of this concrete can be approximated by a uniform distribution. The pdf is shown in Fig. 3.3.6*b*, assuming a = 2400 kg/m^3 and b = 2500 kg/m^3. From the data, the lowest of the 40 values is 2411 kg/m^3 and the highest value is 2488 kg/m^3; also, the estimated mean and standard deviation are 2445 kg/m^3 and 15.99 kg/m^3 (from Table 1.2.2), respectively. If one uses the method of moments to estimate the two parameters, then by solving from the two relationships for the mean and variance just given [Eqs. (4.2.2*a*) and (4.2.2*b*)],

$$\hat{a} = x - \sqrt{3}\hat{s} = 2417 \text{ kg/m}^3 \qquad \text{and} \qquad \hat{b} = \bar{x} + \sqrt{3}\hat{s} = 2473 \text{ kg/m}^3.$$

Because some of the data are outside the range given by the values of the parameters, these estimates by the method of moments are unacceptable.

Exponential Distribution(1)

❖ **Definition**

- **The exponential distribution models the time (or length or area) between Poisson events**
- **Hann(1994). Statistical methods in hydrology**

Exponential Distribution

The probability distribution of the time, T, between occurrences of the event can be found by noting that the prob(T < t) is equal to 1 − prob(T > t). The prob(T > t) is equal to the probability of no occurrences in time t which is $f_X(0; \lambda t)$ or $e^{-\lambda t}$. Thus

$$\text{prob}(T \leq t) = P_T(t; \lambda) = 1 - e^{-\lambda t} \qquad (4.21)$$

Exponential Distribution(2)

❖ **Exponential pdf & cdf**

- **pdf:** $f_X(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0,\ \lambda > 0$

  $= 0 \quad \text{otherwise}$

  where $x =$ continuous random variable

  $\lambda =$ population parameter

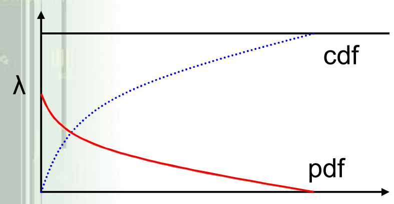
- **cdf:** $F_X(x) = 1 - e^{-\lambda x}$

Exponential Distribution(3)

- **Population parameters**
  - **Mean=**
  - **Variance=**
  - **Skewness coef.= 2**
- **Parameter estimator**
  -
- **Comments**
  - **special case of the gamma distribution (*r* =1)**

Exponential Distribution(3)

- **Distribution shapes**



Exponential Distribution Example

Example 4.20. Floods affecting construction. An engineer constructing a bridge across a river is concerned about the possible occurrence of a flood exceeding 100 m^3/s, which can seriously affect his work. If a flow of such magnitude is exceeded once in five years on average, on the basis of recorded data, what is the chance that the work, which is scheduled to last 14 months, can proceed without interruption or detrimental effects?

Gamma Distribution(1)

- **Description**
  - **the probability dist of the time to the *r* th occurrence, which is the sum of *r* independent r.v. $T_1 + T_2 + \ldots + T_r$ form the exponential distribution**
- **Gamma pdf**
  - $$f_X(x;\lambda,r) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} \quad \text{for } 0 \le x \text{ with } \lambda > 0 \text{ and } r>0,$$
    $$= 0 \quad \text{otherwise}$$

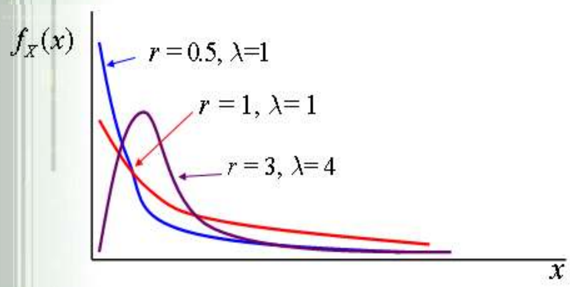
    where λ = scale parameter, x = random variable
    r = shape parameter, $\Gamma()$ = gamma function

Gamma Distribution(2)

- **Population parameters**
  - **Mean=**
  - **Variance=**
  - **Skewness coef.=**
- **Parameter estimators**
  - $\hat{\lambda} = \bar{x}/s^2,\ \hat{r} = \bar{x}^2/s^2$
- **Comments**
  - **the log-Pearson type III distribution is a 3 parameter gamma distribution**

Gamma Distribution(3)

- **Distribution shapes**



Gamma Distribution Example

Example 4.22. Pumps for a town water supply. At a remote pumping station two pumps are operated so that, when there is a breakdown, the other pump (which serves as a standby) is switched on automatically. The pumps are identical and have a mean time between breakdowns of 300 days. Determine the probability density function of the time, in days, during which the system operates until a complete breakdown.

Normal Distribution(1)

- **Exponential pdf & cdf**
  - **pdf:** $\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \quad \text{for} \quad -\infty < x < \infty$
  - **cdf:** $\Phi(x) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \quad \text{for} \quad a < x < b$
- **Comments**
  - **2 parameter distribution**
  - **skewness coef.=0 i.e. symmetrical about the mean**
  - **unbounded but if μ is greater than 3σ, the chances of *X* less than 0 are negligible in practice**
  - **reproductive properties**

Normal Distribution(2)

- **Standard Normal Distribution**
  - $Z = \dfrac{X-\mu}{\sigma}$ $\quad N \sim (0,1) \quad$ $\phi(z)=\dfrac{1}{\sqrt{2\pi}}\exp(-\dfrac{z^2}{2})$
- **Central Limit Theorem**
  - **If S_n is the sum of n independently and identically distributed random variables X_i each having a mean μ and variance σ^2 then in the limit as n approaches infinity, the distribution of S_n approaches a normal distribution with mean $n\mu$ and variance $n\sigma^2$**

Normal Distribution Example

Example 4.28. Settlement of bridge foundations. A proposed bridge across a stream is supported at the two ends and on a center pier. Although the design allows for relative settlements of the foundations, the engineer needs to keep these within limits. Settlements caused by the sum of numerous dead loads and impacts of moving vehicles may be assumed to be normally distributed, with the variability arising from the effects on the soil by the foundations. By correlating with results of tests on similar structures and soil conditions, estimates of the settlements are made as follows:

	Mean, cm	Standard deviation, cm
Left end	3.0	1.0
Center pier	5.0	1.5
Right end	3.0	1.0

It is also assumed as an initial approximation that the settlements are independent.

Normal Distribution(3)

- **Central Limit Theorem**
  - **If S_n is the sum of *n* independently and identically distributed random variables X_i each having a mean μ and variance σ^2 then in the limit as *n* approaches infinity, the distribution of S_n approaches a normal distribution with mean $n\mu$ and variance $n\sigma^2$**

Normal Distribution Example(2)

Example 4.29. Concrete densities. The mean and standard deviation of the densities of concrete from a particular mix given in Table 1.2.2 are (very close to) 2445 and 16 N/mm^2, respectively. We do not know the vaules pertaining to the population. However, let us assume that the standard deviation of 16 N/mm^2 is the true value. From the values of skewness and kurtosis in Table 1.2.2 it seems reasonable to assume that the density of concrete represented here is from a normal population; a uniform distribution is an alternative used, for instance, in Example 3.37. Then the distribution of $\overline{X}_n = \sum_{i=1}^{n} X_i/n$ will be approximately $N(\mu, 16^2/n)$ even for small values of n. We note from Table C.1 that, for example, $\Phi(2.575) = .995$. Thus we can say, using Eq. (4.2.20),

$$\Pr[-2.575 \times 16/\sqrt{n} \le (\overline{X}_n - \mu_x) \le 2.575 \times 16/\sqrt{n}] = .99.$$

This means that if we have a sample size, say, $n = 16$,

$$\Pr[-10.3 \le (\overline{X}_n - \mu) \le 10.3] = .99.$$

The implication is that even with a small sample size of 16 we are 99 percent confident that the mean can be estimated within 10 N/mm^2 of the true value. Of course, if the variance is larger (or the sample size is smaller) the difference will be greater. There will be more about these aspects in Chapter 5.

(2-parameter) Lognormal Distribution

- **Random variables**
  - **X: lognormally distribution**
  - **Y=LN(X): normally distributed**
- **Features**
  - **Population parameters**

$$\mu_X = \exp[\mu_Y + \sigma_Y^2 / 2]$$
$$\sigma_X^2 = \mu_X^2[\exp(\sigma_Y^2) - 1]$$
$$\sigma_X = 3CV_X + CV_X^3$$

  - **Bounded**

Lognormal Distribution Example(1)

Example 4.31. Lognormal distribution of timber strengths. Without the zero value, the summary data in Table 1.2.2 representing the modulus of rupture have a coefficient of skewness of .54. Although this result is not highly significant, we may, in the first instance, fit a lognormal distribution to the data.

From Eq. (4.2.28*c*), we have

$$\sigma_{\ln(X)} = \sqrt{\ln(V_X^2 + 1)} = \sqrt{\ln(0.24^2 + 1)} = 0.237.$$

From Eq. (4.2.28*c*)

$$\mu_{\ln(X)} = \ln\left[\frac{\mu_X}{(V_X^2 + 1)^{1/2}}\right] = \ln\left[\frac{39.33}{(0.24^2 + 1)^{1/2}}\right] = 3.62.$$

Also from Eq. (4.2.28*d*), the median is

$$m_X = \frac{39.33}{(0.24^2 + 1)^{1/2}} = 37.19 \text{ N/mm}^2.$$

Lognormal Distribution Example(2)

For example, to determine the modulus of rupture that is exceeded 95 percent of the time, we solve the inverse $\Phi(x) = .05$ using the normal distributiom. Thus $z = -1.645$ for $\Phi(z) = .05$ and the corresponding y is given by

$$\frac{y - 3.64}{0.238} = -1.645.$$

Hence $x = \exp(y) = \exp(3.62 - 0.237 \times 1.645) = 25.28 \text{ N/mm}^2$.

We may also be interested to know, for example, the probability that the modulus of rupture of a randomly selected timber is not less than 20 N/mm^2. This is given by

$$\Pr[X \geq 20] = 1 - F_X(20) = 1 - F_Z\left[\frac{\ln(20) - 3.62}{0.237}\right] = 1 - F_z(-2.634)$$

$$= 1 - .996 = .004.$$