Example 1

- The average number of robberies in a day is four in a certain big city.
- What is the probability that six robberies occurring in two days?

<u>Solution</u>

- Let X be the number of robberies in two days.
- Then $X \sim P(\lambda)$ where $\lambda = 2 \times 4 = 8$.
- $\Pr(X = 6) = \frac{e^{-8}(8)^6}{6!} = 0.1222.$

Microsoft Excel Worksheet

The following properties of the Poisson distribution are useful:

✓ Let X follows the $Poisson(\lambda_1)$ distribution. Let Y follows the $Poisson(\lambda_2)$ distribution. X and Y are independent. Then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$.

This has been applied in this slide. The average number of robberies in a day is four; therefore, if we define $X_1 = \#$ of robberies in day 1; $X_2 = \#$ of robberies in day 2, then $X_1 \sim P(4), X_2 \sim P(4)$. X_1 and X_2 are independent as they are the number of occurrences of events in different days. So we conclude $X = X_1 + X_2 \sim P(4 + 4) = P(8)$.

✓ Let X be the number of occurrences of events in a period of time T; it has the $Poisson(\lambda)$ distribution. If Y is the number of occurrences of events in the period of time tT, then $Y \sim Poisson(t\lambda)$. Note that this can also be used to identify that $X \sim Poisson(8)$ given in the slide.

Example 3

A can company reports that the number of breakdowns per 8 hour shift on its machine-operated assembly line follows a Poisson distribution, with a mean of 1.5.

(a) What is the probability of exactly two breakdowns during the midnight shift?

(b) What is the probability of fewer than two breakdowns during the afternoon shift?

(c) What is the probability that no breakdowns during three consecutive 8-hour shifts?

(Assume the machine operates independently across shifts.)

In the lecture video, Prof. Chan solved the problem jointly using the Poisson distribution and binomial distribution. This problem, on the other hand, can also be solved using the properties in the last page.

- ✓ Let X_i = the number of breakdowns during the ith 8-hour shift. Then $X_i \sim Poisson(1.5)$.
- ✓ Let $Y = X_1 + X_2 + X_3$, then Y is the number of breakdowns during three consecutive 8-hour shift. Therefore $Y \sim Poisson(4.5)$.
- ✓ Part (c) is asking $Pr(Y = 0)$, which is $e^{-4.5}\frac{4.5^0}{0!} = e^{-4.5}$.

Page 4-73 &133

4.5 Poisson Approximation to the Binomial Distribution

Theorem 4.5

- Let X be a **Binomial** random variable with parameters n and p. That is

  $$\Pr(X = x) = f_X(x) = {}_nC_r\, p^x q^{n-x}, \text{ where } q = 1 - p.$$

- Suppose that $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains a constant as $n \to \infty$.
- Then X will have approximately a Poisson distribution with parameter np. That is

$$\lim_{\substack{p\to 0\\ n\to\infty}} \Pr(X = x) = \frac{e^{-np}(np)^x}{x!}$$

Theorem

Theorem

- If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$,
- then as $n \to \infty$,

$$Z = \frac{X - np}{\sqrt{npq}} \text{ is approximately } \sim N(0,1)$$

The approximation of the binomial distribution by Poisson or Normal distributions is under totally different scenarios.

✓ Poisson distribution is a good approximation to the Binomial(n, p) distribution, when n is large but np is small (so that $n(1-p)$ is large).

✓ Normal distribution is a good approximation to the Binomial(n, p) distribution, when n, np, $n(1-p)$ are all large. Practically, we require $np > 5$ and $n(1-p) > 5$.

Keep in mind that these are just approximations, they couldn't give you the exact value. Roughly speaking, how good the approximation is depends on how the corresponding conditions are satisfied.

Mean and Variance of Cont Uniform RV

Theorem 4.6

If X is uniformly distributed over $[a, b]$, then

$$E(X) = \frac{a+b}{2}, \qquad \text{and} \quad V(X) = \frac{1}{12}(b-a)^2.$$

Proof

$$E(X) = \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_{x=a}^{b}$$

$$= \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

Comparing this properties with the discrete random variables, we observe that the continuous uniform distribution over an interval has nicer formulae for evaluating the expectation and the variance.

But keep in mind that these formulae are applicable only when that the distribution is defined on a single interval.

Page 4-90

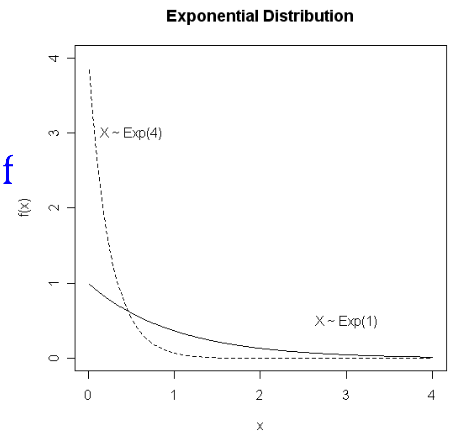
4.7 Exponential Distribution

Definition 4.6

- A continuous random variable X assuming all nonnegative values is said to have an exponential distribution with parameter $\alpha > 0$ if its probability density function is given by

  $f_X(x) = \alpha e^{-\alpha x}$, for $x > 0$.

  and 0 otherwise.
- Note : $\int_{-\infty}^{\infty} f(x)\, dx = 1$



Exponential distribution has two alternative definitions, one is given on the page above; the other is given on page 4-94 that $f_X(x) = \frac{1}{\mu}e^{-x/\mu}$ for $x > 0$, where the parameter $\mu > 0$. Both definitions can be seen frequently in the literature.

✓ These two definitions are equivalent since if we do the reparameterization $\alpha = 1/\mu$.

✓ If we use the definition given on page 4-90, $E(X) = 1/\alpha$ and $V(X) = 1/\alpha^2$; see page 4-91. The cdf is given by $F_X(x) = 1 - e^{-\alpha x}$; see page 4-99.

✓ If we use the definition given on page 4-94, $E(X) = \mu$ and $V(X) = \mu^2$; see page 4-94. The cdf is given by $F_X(x) = 1 - e^{-x/\mu}$.

Exponential distribution is popularly used to model the survival (recovery) time of a patient in the medical research, where $P(X > t) = 1 - F_X(t)$ is called the survival function. It is the probability that the survival (recovery) time of a patient is greater than t.

4.8 Normal Distribution

Definition 4.7

- The random variable X assuming all real values, $-\infty < x < \infty$, has a **normal** (or **Gaussian**) distribution if its probability density function is given by
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty,$$
where $-\infty < \mu < \infty$ and $\sigma > 0$.
- It is denoted by $N(\mu, \sigma^2)$.
- μ and σ are called parameters of the normal distribution.

Normal distribution is the most important and popularly used distribution in statistics. Pay extra attention to the properties given on pages 4-106 to 4-110.

✓ The density is symmetric about μ, which is the expectation and the median of the distribution; one direct consequence is $Pr(X \le \mu) = Pr(X \ge \mu) = 0.5$. μ is also called the location parameter, which determines the location of the center of the distribution.

✓ $\sigma^2 = V(X)$ is the shape parameter (also called the dispersion parameter in the literature), which determines the shape of the density function.

✓ No matter what are the values for μ and σ^2, the density is positive for $x \in \mathbb{R}$. It gets closer and closer to (but never touch) 0, when x approaches ∞ or approaches $-\infty$.

✓ The standardization $Z = \frac{X-\mu}{\sigma}$ is very important. The density becomes symmetric about 0. That is for any $z \in \mathbb{R}$, $Pr(Z \le -z) = Pr(Z \ge z)$. $E(Z) = 0$ and $V(Z) = 1$. With this standardization, for $x_1 < x_2$, $Pr(x_1 < X < x_2)$ (with μ and σ^2 being any given values) can always be obtained from the table for Z. See page 4-111 of the lecture slides for more details.

Properties of the normal distribution (Continued)

- The importance of the standardized normal distribution is the fact that it is tabulated.
- Whenever X has distribution $N(\mu, \sigma^2)$, we can always simplify the process of evaluating the values of $\Pr(x_1 < X < x_2)$ by using the transformation $Z = (X - \mu)/\sigma$. Hence $x_1 < X < x_2$ is equivalent to
$$(x_1 - \mu)/\sigma < Z < (x_2 - \mu)/\sigma.$$
- Let $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$. Then
$$\Pr(x_1 < X < x_2) = \Pr(z_1 < Z < z_2).$$

ST2334 Probability and Statistics CYM Special Probability Distributions 4-111

Read the examples thereafter to understand how all the above properties are jointly applied to solve various problems.

Properties of the normal distribution (Continued)

2. The maximum point occurs at $x = \mu$ and its value is
$$\frac{1}{\sqrt{2\pi}\sigma}$$
3. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
4. The total area under the curve and above the horizontal axis is equal to 1.
$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1.$$

Deriving $\int_{-\infty}^{\infty} f_X(x)dx = 1$ is beyond the scope of this module. But if you are interested in it, read the link:

https://en.wikipedia.org/wiki/Gaussian_integral

Example 4

- The breakdown voltage of a randomly chosen diode of a particular type is known to be normally distributed.
- What is the probability that a diode's breakdown voltage is within 1 s.d. of its mean value?

Solution

- This question can be answered without knowing either μ or σ^2, as long as the distribution is known to be normal.
- That is, the answer is the same for **any** normal distribution.

For any normal random variable X, the probability that X is within c s.d. of its mean value is always deterministic, where $c > 0$ is a known constant. In particular, assume $X \sim N(\mu, \sigma^2)$,

$$Pr(\mu - c\sigma < X < \mu + c\sigma) = Pr\left(-c < \frac{X-\mu}{\sigma} < c)\right) = Pr(|Z| < c),$$

which does not depend on μ and σ.

Continuity Correction

Note: In the above calculations, we have made the continuity correction to improve the approximation. In general, we have:

(a) $\Pr(X = k) \approx \Pr(k - ½ < X < k + ½)$.

(b) $\Pr(a \le X \le b) \approx \Pr(a - ½ < X < b + ½)$.

$\Pr(a < X \le b) \approx \Pr(a + ½ < X < b + ½)$.

$\Pr(a \le X < b) \approx \Pr(a - ½ < X < b - ½)$.

$\Pr(a < X < b) \approx \Pr(a + ½ < X < b - ½)$.

(c) $\Pr(X \le c) = \Pr(0 \le X \le c) \approx \Pr(-½ < X < c + ½)$.

(d) $\Pr(X > c) = \Pr(c < X \le n) \approx \Pr(c + ½ < X < n + ½\,)$.

Be aware of the correction rules given in this page, when you use normal distribution to approximate the probabilities based on the binomial random variables.