Binomial and Normal Distributions

1. Binomial distribution
2. Poisson distribution
3. Normal distribution
4. In-class exercise

Binomial and Normal Distributions

Binomial Distribution

Jacob Bernoulli
1713

Discrete Probability Distributions

- Random variable X can only assume countably many discrete values: x_1, x_2, x_3, …
- Probability distribution function $f(x)$

$$f(x) = \begin{cases} p_j \equiv P(X = x_j) & \text{if } x = x_j \\ 0 & \text{otherwise} \end{cases}$$

- Cumulative distribution function

$$F(x) = \sum_{x_j \le x} f(x_j) = \sum_{x_j \le x} p_j$$

- Properties

$$P(a < X \le b) = F(b) - F(a) = \sum_{a < x_j \le b} p_j$$

$$\sum_j p_j = 1$$

Binomial Distribution

- The binomial distribution is a discrete probability distribution that has applications in manufacturing, medicine and other fields
- The distribution governs the probability that an event A will occur a certain number of times in n independent trials
- The probability that A occurs in any one trial is $P(\mathrm{A}) = p$
- The probability that A will not occur in any one trial is $P(\mathrm{A}^c) = q = 1\text{-}p$
- The random variable X = number of times A occurs in n trials

Binomial Distribution

- $X = \{0, 1, 2, \ldots, n\}$
- If $x = m$, then A occurred in m trials and did not occur in n-m trials
- X has the probability distribution function:

$$f(x) = \binom{n}{x} p^x q^{n-x} \qquad \binom{n}{x} \equiv \frac{n!}{x!(n-x)!}$$

- Limiting cases ($x = 0$, $x = n$):

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = 1 \qquad f(x=0) = \binom{n}{0} p^0 q^{n-0} = q^n$$

$$\binom{n}{n} = \frac{n!}{n!(n-n)!} = 1 \qquad f(x=n) = \binom{n}{n} p^n q^{n-n} = p^n$$

UMASS.

Binomial Distribution

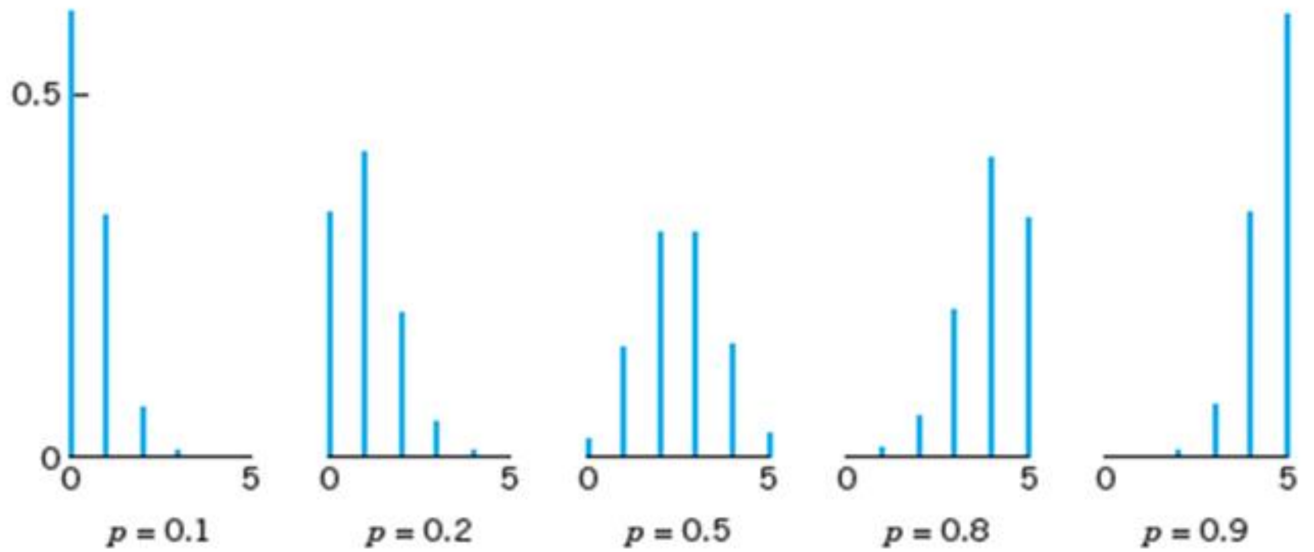


Fig. 517. Probability function (2) of the binomial distribution for $n = 5$ and various values of p

- Mean and variance: $\mu = np \quad \sigma^2 = npq$
- Equal probabilities of success and failure ($p = q = 0.5$)

$$f(x) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} \left(\frac{1}{2}\right)^n$$

Binomial Distribution Example

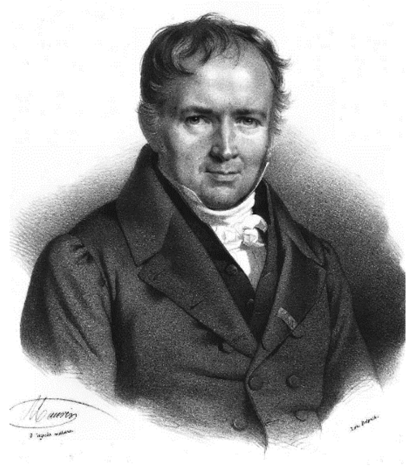
- A plant manufactures solar thin films. Assume the probability of manufacturing a single acceptable film is $P(\text{A}) = p = 0.99$.
- What is the probability that only a single film in a lot of 50 films will be defective?

$$\binom{n}{x} = \binom{50}{49} = \frac{50!}{49!(50-49)!} = 50$$

$$f(x = 49) = \binom{50}{49} 0.99^{49} 0.01^{1} = 0.306$$

Binomial and Normal Distributions

Poisson Distribution



Siméon Denis Poisson
1713

Poisson Distribution

- If $p \rightarrow 0$ and $n \rightarrow \infty$ such that $\mu = np \rightarrow$ constant, the binomial distribution converges to the Poisson distribution:

$$f(x) = \frac{\mu^x}{x!} e^{-\mu} \qquad x = 0, 1, 2, \ldots$$

- Mean and variance: $\sigma^2 = \mu$
- When applicable, the Poisson distribution is more convenient to use then the binomial distribution

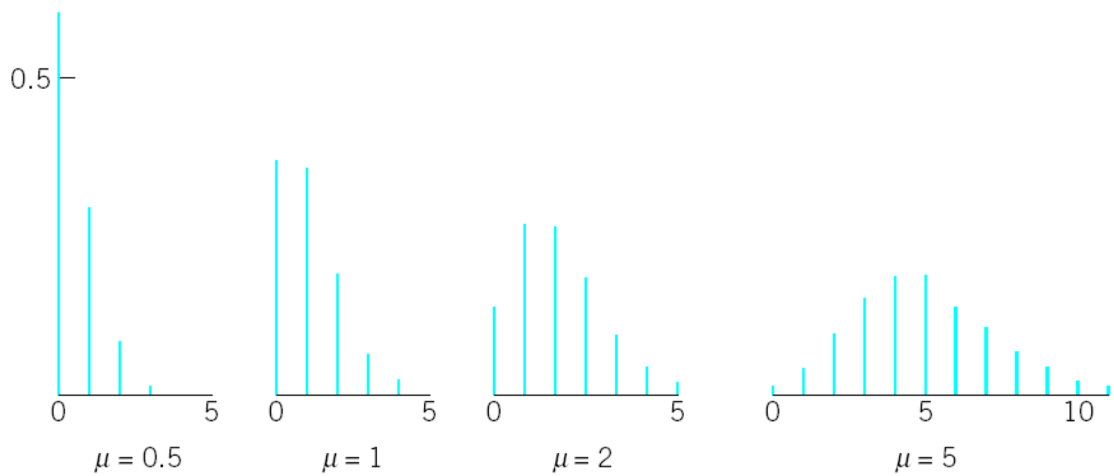


Fig. 517. Probability function (5) of the Poisson distribution for various values of μ

Poisson Distribution Example

- A plant manufactures solar thin films. Assume the probability of manufacturing a single defective film is $P(\text{A}) = p = 0.005$.
- What is the probability that a lot of 100 films will contained more than 2 defective films?

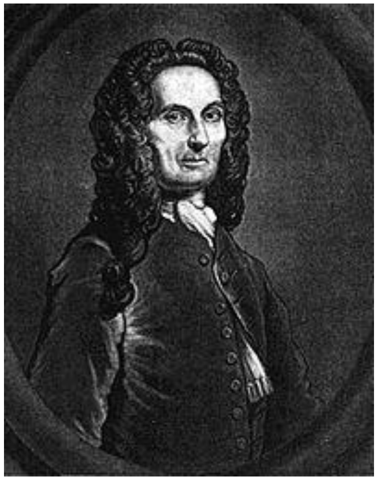
$$\mu = np = (100)(0.005) = 0.5$$

$$f(x \leq 2) = f(0) + f(1) + f(2) = e^{-0.5}\left(1 + \frac{0.5}{1} + \frac{0.5^2}{2}\right) = 0.9856$$

$$f(x > 2) = 1 - f(x \leq 2) = 0.0144$$

Binomial and Normal Distributions

Normal Distribution



Abraham de Moivre
1733

Continuous Probability Distributions

- Random variable X can assume infinitely many real values
- Cumulative distribution function:

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

- Probability distribution function:

$$f(x) = \frac{dF(x)}{dx}$$

- Properties

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(v)dv$$

$$\int_{-\infty}^{\infty} f(v)dv = 1$$

Normal Distribution

- By far the most commonly used continuous probability distribution is the normal distribution
- Also called the Gaussian distribution and the "bell shaped curve"
- Probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right] \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

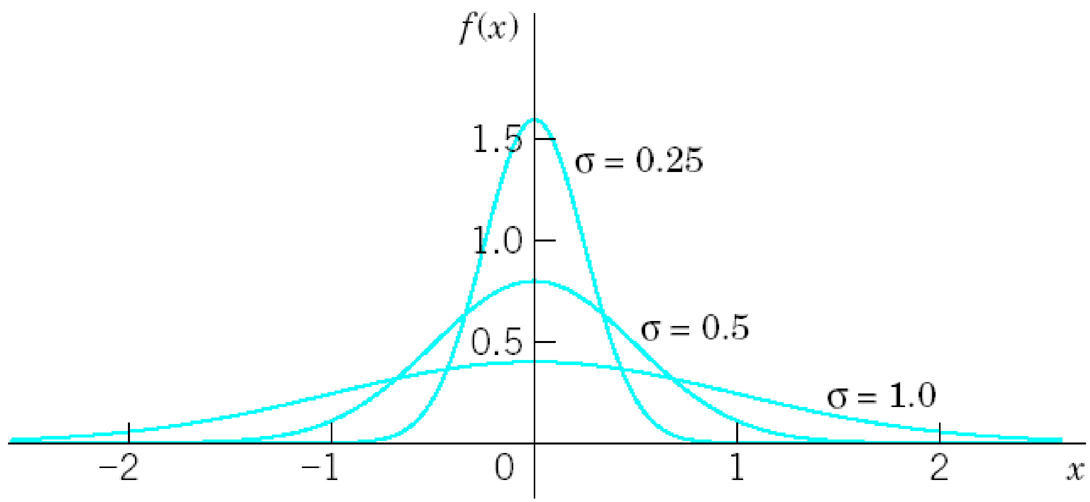


Fig. 518. Density (1) of the normal distribution with $\mu = 0$ for various values of σ

Normal Distribution

- Cumulative distribution function

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^{2}\right]dv$$

- Standardized normal distribution ($\mu = 0$, $\sigma^2 = 1$)

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-u^2/2}\,du \qquad F(x)=\Phi(z)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

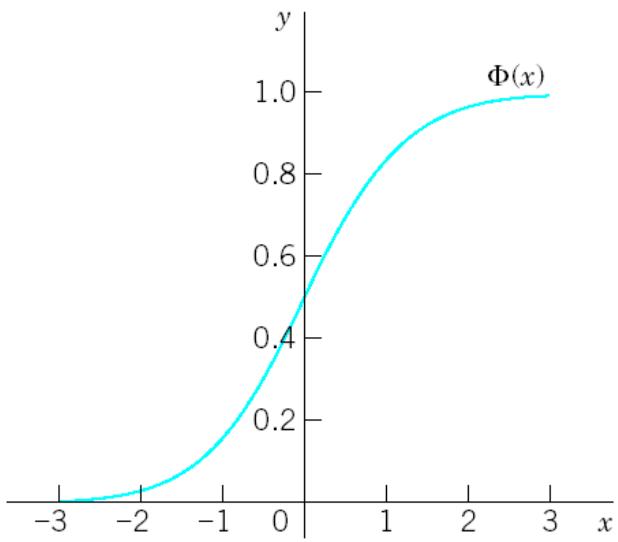


Fig. 519. Distribution function $\Phi(z)$ of the normal distribution with mean 0 and variance 1

Computing Probabilities

- Interval probabilities

$$P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- Sigma limits

$$P(\mu - \sigma < X \leq \mu + \sigma) \approx 68\%$$

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) \approx 95.5\%$$

$$P(\mu - 3\sigma < X \leq \mu + 3\sigma) \approx 99.7\%$$

- Samples outside the 3 sigma limit are termed outliers

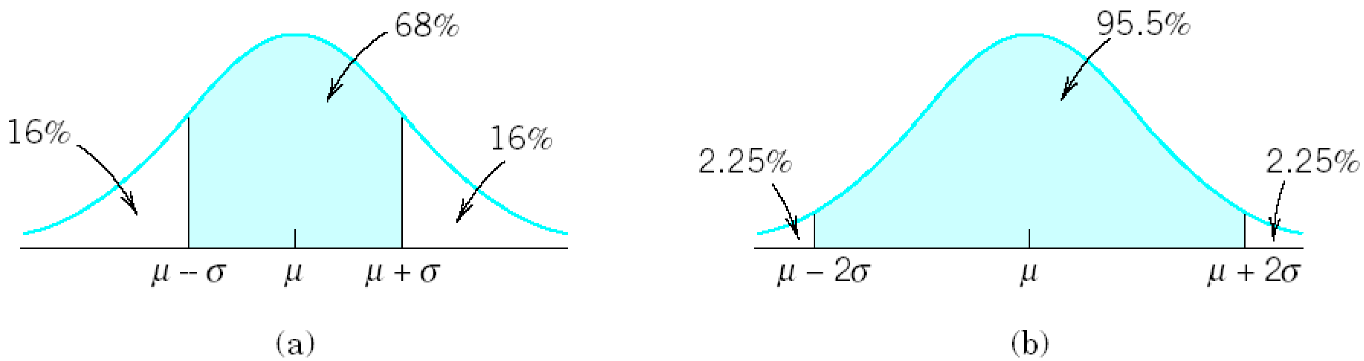


Fig. 520. Illustration of formula (6)

Normal Distribution Tables

- Values of the standardized normal distribution are tabulated in Tables A7 and A8 in the text
- Tables A7 provides probabilities for given *x* values (cumulative distribution)
- Table A8 provides *x* values for given probabilities (inverse cumulative distribution)
- To use the table note that:

$$z = \frac{x-\mu}{\sigma} \qquad \Phi(-z) = 1 - \Phi(z)$$

UMASS.

Normal Distribution Example 1

- The temperature of a bioreactor follows a normal distribution with an average temperature of 30^{o}C and a standard deviation of 1^{o}C. What percentage of the time will the temperature be within +/-0.5^{o}C of the average?
- Calculate probabilities at 29.5^{o}C and 30.5^{o}C, then calculate the difference:

$$P(X \leq 29.5) = F(29.5) = \Phi\left(\frac{29.5 - 30}{1}\right) = \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085$$

$$P(X \leq 30.5) = F(30.5) = \Phi\left(\frac{30.5 - 30}{1}\right) = \Phi(0.5) = 0.6915$$

$$P(29.5 \leq X \leq 30.5) = P(X \leq 30.5) - P(X \leq 29.5) = 0.6915 - 0.3085 = 0.3830$$

UMASS.

Normal Distribution Example 2

- The molecular weight of 10 polymer samples has been measured (all x10^{-5}):

$$x = \{1.2 \quad 0.9 \quad 1.1 \quad 1.5 \quad 1.4 \quad 0.7 \quad 1.1 \quad 0.8 \quad 1.3 \quad 1.2\}$$

- Compute the probability that a given sample will have molecular weight less than or equal to 1x10^{-5} if the true mean and variance are unknown

$$\mu \cong \bar{x} = \tfrac{1}{n}\sum_{j=1}^{n} x_j = 1.12 \qquad \sigma^2 \cong s^2 = \tfrac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2 = 0.0662$$

$$P(X \le 1) = F(1) = \Phi\left(\frac{1-1.12}{\sqrt{0.0662}}\right) = \Phi(-0.47) = 1 - \Phi(0.47) = 1 - 0.6808 \cong 0.32$$

Binomial and Normal Distributions

In-class Exercise