SCS 1307
Probability Distributions

Complete Guide with Solved Examples

Created by: M.Y. Hanan Mohamed

Based on Dr. Dilshani Tissera's Course Material

University of Colombo, School of Computing

October 4, 2025

Contents

1 Discrete Probability Distributions

1.1 Binomial Distribution

Binomial Distribution

A binomial situation arises when:

- There are n independent trials
- Each trial has exactly two outcomes: "success" or "failure"
- Probability of success p remains constant for all trials
- We count the number of successes X in n trials

Notation: $X \sim \text{Bin}(n, p)$
Probability Mass Function:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Properties of Binomial Distribution

If $X \sim \text{Bin}(n, p)$, then:

$$\begin{aligned} \text{Mean: } E(X) &= np \\ \text{Variance: } \text{Var}(X) &= np(1-p) = npq \quad \text{where } q = 1 - p \\ \text{Standard Deviation: } \sigma &= \sqrt{npq} \end{aligned}$$

Example 1: Coin Tosses

Find the probability of getting exactly 2 heads in 6 tosses of a fair coin.
Solution:
Let X = number of heads. Then $X \sim \text{Bin}(6, 0.5)$

$$\begin{aligned} P(X = 2) &= \binom{6}{2} (0.5)^2 (0.5)^4 \\ &= \frac{6!}{2!4!} \times (0.5)^6 \\ &= 15 \times \frac{1}{64} \\ &= \boxed{\frac{15}{64} = 0.234} \end{aligned}$$

Example 2: Fair Coin - Multiple Probabilities

Toss a fair coin 3 times. Find the probability of:

(a) 3 heads

(b) 2 tails and 1 head

(c) At least 1 head

(d) Not more than 1 tail

Solution:
Let $X \sim \text{Bin}(3, 0.5)$ be number of heads
(a) 3 heads:

$$\begin{aligned} P(X = 3) &= \binom{3}{3}(0.5)^3(0.5)^0 \\ &= 1 \times (0.5)^3 \\ &= \boxed{\frac{1}{8} = 0.125} \end{aligned}$$

(b) 2 tails and 1 head (i.e., 1 head):

$$\begin{aligned} P(X = 1) &= \binom{3}{1}(0.5)^1(0.5)^2 \\ &= 3 \times \frac{1}{8} \\ &= \boxed{\frac{3}{8} = 0.375} \end{aligned}$$

(c) At least 1 head:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \binom{3}{0}(0.5)^0(0.5)^3 \\ &= 1 - \frac{1}{8} \\ &= \boxed{\frac{7}{8} = 0.875} \end{aligned}$$

(d) Not more than 1 tail (i.e., at least 2 heads):

$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) \\ &= \binom{3}{2}(0.5)^3 + \binom{3}{3}(0.5)^3 \\ &= 3 \times \frac{1}{8} + 1 \times \frac{1}{8} \\ &= \boxed{\frac{4}{8} = 0.5} \end{aligned}$$

Example 3: Die Rolling

Find the probability that in 5 tosses of a fair die, a 3 appears:

(a) Twice

(b) At most once

(c) At least two times

Solution:
Let $X \sim \text{Bin}(5, \frac{1}{6})$ be number of 3's
(a) Twice:

$$\begin{aligned} P(X = 2) &= \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 \\ &= 10 \times \frac{1}{36} \times \frac{125}{216} \\ &= \boxed{0.161} \end{aligned}$$

(b) At most once:

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{5}{0}\left(\frac{5}{6}\right)^5 + \binom{5}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4 \\ &= \boxed{0.804} \end{aligned}$$

(c) At least two times:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - 0.804 \\ &= \boxed{0.196} \end{aligned}$$

Example 4: Family Children

A family has 4 children. Assuming $P(\text{boy}) = 0.5$, find:

(a) At least 1 boy

(b) At least 1 boy and at least 1 girl

Solution:
Let $X \sim \text{Bin}(4, 0.5)$ be number of boys

(a) At least 1 boy:

$$\begin{aligned}P(X \geq 1) &= 1 - P(X=0)\\ &= 1 - \binom{4}{0}(0.5)^4\\ &= 1 - \frac{1}{16}\\ &= \boxed{\frac{15}{16} = 0.9375}\end{aligned}$$

(b) At least 1 boy and at least 1 girl:

$$\begin{aligned}P(1 \leq X \leq 3) &= 1 - P(X=0) - P(X=4)\\ &= 1 - (0.5)^4 - (0.5)^4\\ &= 1 - \frac{1}{16} - \frac{1}{16}\\ &= \boxed{\frac{14}{16} = 0.875}\end{aligned}$$

Example 5: Expected Value

In 100 tosses of a fair coin, what is the expected number of heads?
Solution:
$X \sim \text{Bin}(100, 0.5)$

$$\begin{aligned}E(X) &= np\\ &= 100 \times 0.5\\ &= \boxed{50}\end{aligned}$$

Example 6: Defective Bolts

If the probability of a defective bolt is 0.1, find the mean and standard deviation for the number of defective bolts in 400 bolts.
Solution:
$X \sim \text{Bin}(400, 0.1)$
Mean:

$$E(X) = np = 400 \times 0.1 = \boxed{40}$$

Standard Deviation:

$$\begin{aligned}\sigma &= \sqrt{npq}\\ &= \sqrt{400 \times 0.1 \times 0.9}\\ &= \sqrt{36}\\ &= \boxed{6}\end{aligned}$$

Example 7: Finding Parameters

Random variable X has binomial distribution with mean 5.76 and standard deviation 1.92. Find $P(X = 6)$.

Solution:

Given: $E(X) = 5.76$ and $\sigma = 1.92$

We know: $E(X) = np$ and $\mathrm{Var}(X) = npq$

$$\begin{aligned} np &= 5.76 \\ npq &= (1.92)^2 = 3.6864 \end{aligned}$$

Dividing:

$$\begin{aligned} q &= \frac{3.6864}{5.76} = 0.64 \\ p &= 1 - 0.64 = 0.36 \\ n &= \frac{5.76}{0.36} = 16 \end{aligned}$$

Therefore, $X \sim \mathrm{Bin}(16, 0.36)$

$$\begin{aligned} P(X = 6) &= \binom{16}{6}(0.36)^6(0.64)^{10} \\ &= \boxed{0.198} \end{aligned}$$

1.2 Poisson Distribution

Poisson Distribution

The Poisson distribution models the number of events occurring in a fixed interval of time or space when events occur:

- Randomly and independently
- At a constant average rate λ

Notation: $X \sim \text{Po}(\lambda)$
Probability Mass Function:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, 3, \ldots$$

where $\lambda > 0$ is the average number of occurrences in the interval.

Properties of Poisson Distribution

If $X \sim \text{Po}(\lambda)$, then:

$$\text{Mean: } E(X) = \lambda$$
$$\text{Variance: } \text{Var}(X) = \lambda$$
$$\text{Standard Deviation: } \sigma = \sqrt{\lambda}$$

Note: Mean equals variance!

Applications of Poisson Distribution

The Poisson distribution is used for counting rare events:

- Number of car accidents on a road in one day
- Number of accidents in a factory in one week
- Number of phone calls to a switchboard per minute
- Number of insurance claims per month
- Number of bacteria in a liquid sample
- Number of radioactive disintegrations per second

Example 1: Basic Poisson Calculation

If $X \sim \text{Po}(2)$, find:

(a) $P(X = 4)$

(b) $P(X \geq 3)$

Solution:

(a) $P(X = 4)$:

$$\begin{aligned} P(X = 4) &= \frac{e^{-2} \cdot 2^4}{4!} \\ &= \frac{e^{-2} \cdot 16}{24} \\ &= \boxed{0.0902} \end{aligned}$$

(b) $P(X \geq 3)$:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[\frac{e^{-2} \cdot 2^0}{0!} + \frac{e^{-2} \cdot 2^1}{1!} + \frac{e^{-2} \cdot 2^2}{2!}\right] \\ &= 1 - e^{-2}[1 + 2 + 2] \\ &= 1 - 5e^{-2} \\ &= \boxed{0.323} \end{aligned}$$

Example 2: Bacteria Count

The mean number of bacteria per milliliter of liquid is 4. Find the probability that in 1ml there will be:

(a) No bacteria

(b) 4 bacteria

(c) Less than 3 bacteria

Solution:
$X \sim \text{Po}(4)$
(a) No bacteria:

$$\begin{aligned} P(X = 0) &= \frac{e^{-4} \cdot 4^0}{0!} \\ &= e^{-4} \\ &= \boxed{0.0183} \end{aligned}$$

(b) 4 bacteria:

$$\begin{aligned} P(X = 4) &= \frac{e^{-4} \cdot 4^4}{4!} \\ &= \frac{256e^{-4}}{24} \\ &= \boxed{0.195} \end{aligned}$$

(c) Less than 3 bacteria:

$$\begin{aligned} P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-4}\left[1 + 4 + \frac{16}{2}\right] \\ &= e^{-4} \times 13 \\ &= \boxed{0.238} \end{aligned}$$

Example 3: Scaling Poisson Parameter

Using data from previous example where mean is 4 per ml, find the probability that in 3ml of liquid there will be less than 2 bacteria.
Solution:
For 3ml: $Y \sim \text{Po}(12)$ (since λ scales with volume)

$$\begin{aligned} P(Y < 2) &= P(Y = 0) + P(Y = 1) \\ &= \frac{e^{-12} \cdot 12^0}{0!} + \frac{e^{-12} \cdot 12^1}{1!} \\ &= e^{-12}(1 + 12) \\ &= 13e^{-12} \\ &= \boxed{7.99 \times 10^{-5}} \end{aligned}$$

Example 4: Call Center

A call center receives calls at a mean rate of 3 per minute. Find the probability that during a randomly selected 3 minutes there will be no calls.
Solution:
For 3 minutes: $X \sim \text{Po}(9)$

$$\begin{aligned} P(X = 0) &= \frac{e^{-9} \cdot 9^0}{0!} \\ &= e^{-9} \\ &= \boxed{0.000123} \end{aligned}$$

1.3 Other Discrete Distributions

1.3.1 Uniform Distribution

Discrete Uniform Distribution

All outcomes are equally likely.
Sample Space: $S = \{1, 2, 3, \ldots, k\}$
PMF: $P(X = x) = \frac{1}{k}$ for $x = 1, 2, \ldots, k$
Mean: $\mu = \frac{k+1}{2}$
Variance: $\sigma^2 = \frac{k^2-1}{12}$

1.3.2 Geometric Distribution

Geometric Distribution

Models the number of trials until the first success in a sequence of independent Bernoulli trials.
Notation: $X \sim \mathrm{Geo}(p)$
PMF: $P(X = x) = p(1-p)^{x-1}$ for $x = 1, 2, 3, \ldots$
Mean: $\mu = \frac{1}{p}$
Variance: $\sigma^2 = \frac{1-p}{p^2}$

Example: Geometric Distribution

The probability of a male or female child is equal. Find the probability that a family's fourth child is their first son.
Solution:
$X \sim \mathrm{Geo}(0.5)$ where $X =$ trial number of first boy

$$\begin{aligned}
P(X = 4) &= (0.5)(1-0.5)^{4-1} \\
&= 0.5 \times (0.5)^3 \\
&= 0.5 \times 0.125 \\
&= \boxed{0.0625}
\end{aligned}$$

1.3.3 Negative Binomial Distribution

Negative Binomial Distribution

Models the number of trials until the k-th success.
Notation: $X \sim \mathrm{NB}(k, p)$
PMF: $P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$ for $x = k, k+1, \ldots$
Mean: $\mu = \frac{k}{p}$
Variance: $\sigma^2 = \frac{k(1-p)}{p^2}$
Note: Geometric distribution is special case when $k = 1$.

Example: Oil Drilling

An exploratory oil well has 20% chance of striking oil. What is the probability that the third strike comes on the seventh well drilled?
Solution:
$X = 7$, $k = 3$, $p = 0.2$

$$\begin{aligned} P(X = 7) &= \binom{7-1}{3-1}(0.2)^3(0.8)^{7-3} \\ &= \binom{6}{2}(0.2)^3(0.8)^4 \\ &= 15 \times 0.008 \times 0.4096 \\ &= \boxed{0.049} \end{aligned}$$

1.3.4 Hypergeometric Distribution

Hypergeometric Distribution

Models sampling **without replacement** from a finite population.
Setup: Population of N objects with k successes and $N - k$ failures. Sample n objects.
Notation: $X \sim \text{Hypergeometric}(N, k, n)$
PMF: $P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$
Mean: $\mu = \frac{nk}{N}$
Variance: $\sigma^2 = \frac{nk(N-k)(N-n)}{N^2(N-1)}$

Example: Card Drawing

A deck has 20 cards: 6 red and 14 black. Draw 5 cards randomly without replacement. What is the probability of exactly 4 red cards?
Solution:
$N = 20$, $k = 6$, $n = 5$, $x = 4$

$$\begin{aligned} P(X = 4) &= \frac{\binom{6}{4}\binom{14}{1}}{\binom{20}{5}} \\ &= \frac{15 \times 14}{15504} \\ &= \frac{210}{15504} \\ &= \boxed{0.0135} \end{aligned}$$

2 Continuous Probability Distributions

2.1 Continuous Random Variables

Continuous Random Variable

A random variable that can take any value in an interval (uncountably infinite values).
Examples: Time, weight, height, temperature, distance

Probability Density Function (PDF)

For continuous random variable X, the PDF $f(x)$ has properties:

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $P(a < X \leq b) = \int_a^b f(x)\,dx$

Important: For continuous RV, $P(X = a) = 0$ for any specific value a.

Expectation and Variance for Continuous RV

Expectation (Mean):

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

Variance:

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = E(X^2) - \mu^2$$

where

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

Example 1: Finding Constant k

A continuous random variable has PDF $f(x) = kx$ for $0 \leq x \leq 4$.

(a) Find k

(b) Find $P(1 \leq X \leq 2.5)$

Solution:

(a) Using $\int_0^4 f(x)\,dx = 1$:

$$\begin{aligned}
\int_0^4 kx\,dx &= 1 \\
k\left[\frac{x^2}{2}\right]_0^4 &= 1 \\
k \times 8 &= 1 \\
k &= \boxed{\frac{1}{8}}
\end{aligned}$$

(b) $f(x) = \frac{x}{8}$ for $0 \le x \le 4$:

$$\begin{aligned}
P(1 \le X \le 2.5) &= \int_1^{2.5} \frac{x}{8}\,dx \\
&= \frac{1}{8}\left[\frac{x^2}{2}\right]_1^{2.5} \\
&= \frac{1}{16}[(2.5)^2 - 1^2] \\
&= \frac{1}{16}[6.25 - 1] \\
&= \boxed{\frac{5.25}{16} = 0.328}
\end{aligned}$$

Example 2: Expectation and Variance

If X has PDF $f(x) = \frac{3x^2}{64}$ for $0 \le x \le 4$, find $E(X)$ and $\text{Var}(X)$.
Solution:
Find $E(X)$:

$$\begin{aligned}
E(X) &= \int_0^4 x \cdot \frac{3x^2}{64}\,dx \\
&= \frac{3}{64}\int_0^4 x^3\,dx \\
&= \frac{3}{64}\left[\frac{x^4}{4}\right]_0^4 \\
&= \frac{3}{64} \times \frac{256}{4} \\
&= \frac{3 \times 64}{64} \\
&= \boxed{3}
\end{aligned}$$

Find $E(X^2)$:

$$\begin{aligned}
E(X^2) &= \int_0^4 x^2 \cdot \frac{3x^2}{64}\,dx \\
&= \frac{3}{64}\int_0^4 x^4\,dx \\
&= \frac{3}{64}\left[\frac{x^5}{5}\right]_0^4 \\
&= \frac{3}{64} \times \frac{1024}{5} \\
&= \boxed{9.6}
\end{aligned}$$

Find Var(X):

$$\begin{aligned}
\text{Var}(X) &= E(X^2) - [E(X)]^2 \\
&= 9.6 - 3^2 \\
&= 9.6 - 9 \\
&= \boxed{0.6}
\end{aligned}$$

2.2 Normal Distribution

Normal Distribution

The most important continuous distribution in statistics. Many natural phenomena follow normal distribution.
Examples: Heights, masses, ages, exam results, measurement errors
Notation: $X \sim N(\mu, \sigma^2)$
Probability Density Function:

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

Parameters:

- μ = mean (center of distribution)
- σ^2 = variance (spread of distribution)
- σ = standard deviation

Shape: Bell-shaped, symmetric about $x = \mu$

Properties of Normal Distribution

If $X \sim N(\mu, \sigma^2)$:

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

Total area under curve = 1

Empirical Rule (68-95-99.7 Rule)

For $X \sim N(\mu, \sigma^2)$:

- Approximately **68%** of data lies within $\mu \pm \sigma$
- Approximately **95%** of data lies within $\mu \pm 2\sigma$
- Approximately **99.7%** of data lies within $\mu \pm 3\sigma$

2.2.1 Standard Normal Distribution

Standard Normal Distribution

A special case with $\mu = 0$ and $\sigma^2 = 1$.
Notation: $Z \sim N(0, 1)$
Standardization Formula:

$$Z = \frac{X - \mu}{\sigma}$$

If $X \sim N(\mu, \sigma^2)$, then $Z \sim N(0, 1)$

PDF: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
CDF: $\Phi(z) = P(Z \leq z)$ (obtained from standard normal tables)

2.2.2 Operations with Normal Variables

Sum and Difference of Independent Normal Variables

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are **independent**, then:

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

Note: Variances always ADD, even for subtraction!

Multiples of Normal Variables

If $X \sim N(\mu, \sigma^2)$ and a is a constant:

$$aX \sim N(a\mu, a^2\sigma^2)$$

Example 1: Sum of Normal Variables

If $X \sim N(60, 16)$ and $Y \sim N(70, 9)$, find:

(a) $P(X + Y < 140)$

(b) $P(120 < X + Y < 135)$

(c) $P(Y - X > 7)$

Solution:
(a) Let $R = X + Y$:

$$R \sim N(60 + 70, 16 + 9) = N(130, 25)$$

$$\begin{aligned} P(R < 140) &= P\left(Z < \frac{140 - 130}{5}\right) \\ &= P(Z < 2) \\ &= \boxed{0.9772} \end{aligned}$$

(b) Using $R \sim N(130, 25)$:

$$\begin{aligned} P(120 < R < 135) &= P\left(\frac{120 - 130}{5} < Z < \frac{135 - 130}{5}\right) \\ &= P(-2 < Z < 1) \\ &= \Phi(1) - \Phi(-2) \\ &= 0.8413 - 0.0228 \\ &= \boxed{0.8185} \end{aligned}$$

(c) Let $T = Y - X$:

$$T \sim N(70 - 60, 9 + 16) = N(10, 25)$$
$$P(T > 7) = P\left(Z > \frac{7 - 10}{5}\right)$$
$$= P(Z > -0.6)$$
$$= 1 - \Phi(-0.6)$$
$$= 1 - 0.2743$$
$$= \boxed{0.7257}$$

Example 2: Multiple of Normal Variable

If $X \sim N(50, 25)$, find $P(3X > 160)$.
Solution:

$$3X \sim N(3 \times 50, 9 \times 25) = N(150, 225)$$
$$P(3X > 160) = P\left(Z > \frac{160 - 150}{15}\right)$$
$$= P(Z > 0.667)$$
$$= 1 - \Phi(0.667)$$
$$= 1 - 0.7477$$
$$= \boxed{0.2523}$$

Example 3: Real-World Application

Mr. Jones walks to the library daily. Travel time: $X \sim N(15, 4)$ minutes. Library time: $Y \sim N(25, 12)$ minutes. Find:

(i) $P(\text{away} > 45 \text{ minutes})$

(ii) $P(\text{travel time} > \text{library time})$

Solution:
(i) Total time away $= X + Y$:

$$X + Y \sim N(15 + 25, 4 + 12) = N(40, 16)$$
$$P(X + Y > 45) = P\left(Z > \frac{45 - 40}{4}\right)$$
$$= P(Z > 1.25)$$
$$= \boxed{0.1056}$$

(ii) $P(X > Y) = P(X - Y > 0)$:

$$\begin{aligned} X - Y &\sim N(15 - 25, 4 + 12) = N(-10, 16) \\ P(X - Y > 0) &= P\left(Z > \frac{0 - (-10)}{4}\right) \\ &= P(Z > 2.5) \\ &= \boxed{0.0062} \end{aligned}$$

2.2.3 Distribution of Sample Mean

Sampling Distribution of the Mean

If $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ has distribution:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Standard error: $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

Example: Sample Mean

A random sample of size 15 is taken from $N(60, 16)$. Find $P(\bar{X} < 58)$.
Solution:

$$\begin{aligned} \bar{X} &\sim N\left(60, \frac{16}{15}\right) \\ P(\bar{X} < 58) &= P\left(Z < \frac{58 - 60}{\sqrt{16/15}}\right) \\ &= P\left(Z < \frac{-2}{1.0328}\right) \\ &= P(Z < -1.936) \\ &= \boxed{0.0264} \end{aligned}$$

Central Limit Theorem (CLT)

If $X_1, X_2, \ldots, X_n$ is a random sample from **any distribution** with mean μ and variance σ^2, then for **large** n:

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Rule of thumb: $n \geq 30$ is usually sufficient
Key point: Even if the population is not normal, the sample mean becomes approximately normal for large samples!

Example: CLT Application

A random sample of size 30 is taken from Binomial(9, 0.5). Find $P(\bar{X} > 5)$.
Solution:
For $X \sim \text{Bin}(9, 0.5)$:

$$\begin{aligned} E(X) &= np = 9 \times 0.5 = 4.5 \\ \text{Var}(X) &= npq = 9 \times 0.5 \times 0.5 = 2.25 \end{aligned}$$

By CLT:

$$\begin{aligned} \bar{X} &\sim N\left(4.5, \frac{2.25}{30}\right) = N(4.5, 0.075) \\ P(\bar{X} > 5) &= P\left(Z > \frac{5 - 4.5}{\sqrt{0.075}}\right) \\ &= P(Z > 1.826) \\ &= \boxed{0.0340} \end{aligned}$$

2.3 Normal Approximation to Binomial

Normal Approximation to Binomial

If $X \sim \text{Bin}(n, p)$ with large n:

$$X \approx N(np, npq)$$

Guidelines for use:

- $n > 10$ and p close to $\frac{1}{2}$, OR
- $n > 30$ and p moving away from $\frac{1}{2}$
- General rule: $np > 5$ and $nq > 5$

Continuity Correction: Since binomial is discrete and normal is continuous:

$$\begin{aligned}
P(X = a) &\approx P(a - 0.5 < X < a + 0.5) \\
P(X \leq a) &\approx P(X < a + 0.5) \\
P(X < a) &\approx P(X < a - 0.5) \\
P(X \geq a) &\approx P(X > a - 0.5) \\
P(X > a) &\approx P(X > a + 0.5)
\end{aligned}$$

Example 1: Coin Tosses

Find probability of 4 to 7 heads (inclusive) in 12 tosses of a fair coin:

(a) Using binomial

(b) Using normal approximation

Solution:
(a) Using binomial: $X \sim \text{Bin}(12, 0.5)$

$$\begin{aligned}
P(4 \leq X \leq 7) &= P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) \\
&= 0.121 + 0.193 + 0.226 + 0.193 \\
&= \boxed{0.733}
\end{aligned}$$

(b) Using normal approximation:

$$\begin{aligned}
\mu &= np = 12 \times 0.5 = 6 \\
\sigma^2 &= npq = 12 \times 0.5 \times 0.5 = 3 \\
X &\approx N(6, 3)
\end{aligned}$$

With continuity correction:

$$\begin{aligned}P(4 \leq X \leq 7) &\approx P(3.5 < X < 7.5)\\&= P\left(\frac{3.5-6}{\sqrt{3}} < Z < \frac{7.5-6}{\sqrt{3}}\right)\\&= P(-1.443 < Z < 0.866)\\&= \Phi(0.866) - \Phi(-1.443)\\&= \boxed{0.732}\end{aligned}$$

Very close to exact answer!

Example 2: Ryegrass Seeds

In a sack, 35% are ryegrass. In a sample of 400 seeds, find probability of:

(a) Less than 120 ryegrass

(b) Between 120 and 150 (inclusive)

(c) More than 160

Solution:
$X \sim \text{Bin}(400, 0.35)$
Parameters:

$$\begin{aligned}\mu &= np = 400 \times 0.35 = 140\\\sigma^2 &= npq = 400 \times 0.35 \times 0.65 = 91\\\sigma &= \sqrt{91} = 9.539\end{aligned}$$

Approximate: $X \approx N(140, 91)$
(a) Less than 120:

$$\begin{aligned}P(X < 120) &\approx P(X < 119.5) \quad \text{(continuity correction)}\\&= P\left(Z < \frac{119.5-140}{9.539}\right)\\&= P(Z < -2.149)\\&= \boxed{0.0158}\end{aligned}$$

(b) Between 120 and 150 (inclusive):

$$\begin{aligned}P(120 \leq X \leq 150) &\approx P(119.5 < X < 150.5)\\&= P(-2.149 < Z < 1.101)\\&= \Phi(1.101) - \Phi(-2.149)\\&= 0.8645 - 0.0158\\&= \boxed{0.8487}\end{aligned}$$

(c) More than 160:

$$\begin{aligned} P(X > 160) &\approx P(X > 160.5) \\ &= P(Z > 2.149) \\ &= 1 - \Phi(2.149) \\ &= \boxed{0.0158} \end{aligned}$$

Note: Symmetric about mean, so (a) and (c) are equal.

2.4 Normal Approximation to Poisson

Normal Approximation to Poisson

If $X \sim \text{Po}(\lambda)$ with **large** λ:

$$X \approx N(\lambda, \lambda)$$

Guidelines: Generally use when $\lambda > 10$ or $\lambda > 15$
Continuity correction applies (same as binomial approximation)

Example: Radioactive Counts

Radioactive disintegrations follow Poisson with mean 25 per second. Find probability that count is between 23 and 27 (inclusive):

(a) Using Poisson

(b) Using normal approximation

Solution:
(a) Using Poisson: $X \sim \text{Po}(25)$

$$\begin{aligned} P(23 \leq X \leq 27) &= \sum_{x=23}^{27} P(X = x) \\ &= \boxed{0.076342} \end{aligned}$$

(b) Using normal approximation:

$$X \approx N(25, 25)$$

With continuity correction:

$$\begin{aligned} P(23 \leq X \leq 27) &\approx P(22.5 < X < 27.5) \\ &= P\left(\frac{22.5 - 25}{5} < Z < \frac{27.5 - 25}{5}\right) \\ &= P(-0.5 < Z < 0.5) \\ &= \Phi(0.5) - \Phi(-0.5) \\ &= 0.6915 - 0.3085 \\ &= \boxed{0.383} \end{aligned}$$

Note: Normal approximation gives wider probability range for this case.

3 Practice Problems with Solutions

Problem 1: Defective Items

1% of items made by a machine are defective. A batch of 10 is inspected. What is the probability that more than one is defective?
Solution:
$X \sim \text{Bin}(10, 0.01)$

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) \\ &= 1 - [P(X = 0) + P(X = 1)] \\ &= 1 - \left[\binom{10}{0}(0.01)^0(0.99)^{10} + \binom{10}{1}(0.01)^1(0.99)^9\right] \\ &= 1 - [0.9044 + 0.0914] \\ &= 1 - 0.9958 \\ &= \boxed{0.0042} \end{aligned}$$

Problem 2: Candy Bar Sales

Pat has 40% chance of selling a candy bar at each house. He needs to sell 5 bars total. What is the probability he sells his last bar at the 11th house?
Solution:
This is negative binomial: need 5th success on 11th trial.
$X = 11$, $k = 5$, $p = 0.4$

$$\begin{aligned} P(X = 11) &= \binom{11-1}{5-1}(0.4)^5(0.6)^{11-5} \\ &= \binom{10}{4}(0.4)^5(0.6)^6 \\ &= 210 \times 0.01024 \times 0.046656 \\ &= \boxed{0.1003} \end{aligned}$$

Problem 3: First Defective

Products have 3% defective rate. What is the probability the first defective is the 5th item inspected?
Solution:
Geometric distribution: $X \sim \text{Geo}(0.03)$

$$\begin{aligned} P(X = 5) &= (0.03)(1 - 0.03)^{5-1} \\ &= 0.03 \times (0.97)^4 \\ &= 0.03 \times 0.88529 \\ &= \boxed{0.0266} \end{aligned}$$

Problem 4: Die Rolling

A die is thrown 7 times. Find the probability of exactly 3 sixes.

Solution:

$X \sim \text{Bin}(7, \frac{1}{6})$

$$\begin{aligned} P(X = 3) &= \binom{7}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4 \\ &= 35 \times \frac{1}{216} \times \frac{625}{1296} \\ &= \boxed{0.0781} \end{aligned}$$

4 Distribution Comparison Table

Distribution	PMF/PDF	When to Use	Mean	Variance
Binomial Bin(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	Fixed trials, 2 outcomes, with replacement	np	npq
Poisson Po(λ)	$\frac{e^{-\lambda}\lambda^x}{x!}$	Rare events in time/space	λ	λ
Geometric Geo(p)	$p(1-p)^{x-1}$	Trials until first success	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Neg. Binomial NB(k,p)	$\binom{x-1}{k-1}p^k(1-p)^{x-k}$	Trials until kth success	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Hypergeom. HG(N,k,n)	$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$	Sampling without replacement	$\frac{nk}{N}$	$\frac{nk(N-k)(N-n)}{N^2(N-1)}$
Uniform Unif(k)	$\frac{1}{k}$	All outcomes equally likely	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$
Normal $N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Continuous, symmetric, bell-shaped	μ	σ^2

Table 1: Summary of Probability Distributions

5 Key Formulas Quick Reference

Binomial Distribution

$X \sim \text{Bin}(n,p)$
$P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$
$E(X) = np, \quad \text{Var}(X) = npq$

Poisson Distribution

$X \sim \text{Po}(\lambda)$
$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$
$E(X) = \lambda, \quad \text{Var}(X) = \lambda$

Normal Distribution

$X \sim N(\mu,\sigma^2)$
Standardization: $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
Sum: $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$
Difference: $X-Y \sim N(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2)$
Sample Mean: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

6 Important Notes and Tips

When to Use Each Distribution

Binomial:

- Fixed number of trials
- Each trial has 2 outcomes
- Trials are independent
- Probability stays constant

Poisson:

- Events occur randomly in time/space
- Events are independent
- Average rate is known
- Counting rare occurrences

Geometric:

- Waiting for first success
- Trials are independent
- Constant probability

Normal:

- Continuous data
- Symmetric distribution
- Natural measurements
- Large sample means (CLT)

Approximations

Binomial to Normal:

- Use when $np > 5$ and $nq > 5$
- Apply continuity correction
- $X \approx N(np, npq)$

Poisson to Normal:

- Use when $\lambda > 10$

- Apply continuity correction
- $X \approx N(\lambda, \lambda)$

Common Mistakes to Avoid

1. **Forgetting continuity correction** when approximating discrete with continuous
2. **Adding variances for subtraction** - remember: $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$
3. **Confusing binomial and hypergeometric** - use hypergeometric for sampling WITHOUT replacement
4. **Wrong Poisson parameter** when scaling - multiply λ by time/space factor
5. **Forgetting to standardize** before using Z-tables
6. **Misreading "at least" and "at most"** - be careful with inequalities
7. **Using wrong variance formula** for sample mean - it's $\frac{\sigma^2}{n}$, not σ^2

Master These Distributions for Exam Success!

"The only way to learn mathematics is to do mathematics."
- Paul Halmos

Prepared by: M.Y. Hanan Mohamed
University of Colombo School of Computing
SCS 1307 - Probability & Statistics
Complete Probability Distributions Guide