4. **Classical Probability Distributions**

4.1 **Discrete Models**

Recall from 3.1, that for a discrete population random variable $X$, we have

**Definition:** $f(x)$ is a **probability distribution function** if, for all $x$,

$$f(x) \geq 0 \quad \text{AND} \quad \sum_{x} f(x) = 1.$$  

The resulting **cumulative distribution function** (cdf) is defined as, for all $x$,

$$F(x) = P(X \leq x) = \sum_{x \leq x} f(x_i),$$

and is piecewise constant, increasing from 0 to 1. Therefore, for any two population values $a < b$, it follows that

$$P(a \leq X \leq b) = \sum_{a}^{b} f(x) = F(b) - F(a^{-}).$$

**Definition:** The **mean** (or **expected value**) of $X$ is given by $\mu = E[X] = \sum_{x} x f(x).$

**Definition:** The **variance** of $X$ is given by either of the two equivalent forms:

$$\sigma^2 = E[(X - \mu)^2] = \sum_{x} (x - \mu)^2 f(x)$$

$$\sigma^2 = E[X^2] - \mu^2 = \sum_{x} x^2 f(x) - \mu^2$$

[Diagram showing the cumulative distribution function (CDF) and the probability distribution function (PDF) for a discrete random variable $X$. The CDF is piecewise constant, increasing from 0 to 1. The PDF consists of bars at the values $x_1, x_2, x_3, \ldots$, with heights $f(x_1), f(x_2), f(x_3), \ldots$. The total area under the PDF is equal to 1.]
Example:

**PREVENTING CHRONIC DISEASE**

PUBLIC HEALTH RESEARCH, PRACTICE, AND POLICY

VOLUME 3. NO. 1

JANUARY 2006

ORIGINAL RESEARCH

Identifying Geographic Disparities in the Early Detection of Breast Cancer Using a Geographic Information System

Jane A. McBryde, PhD, Patrick L. Rumington, MD, Ronald E. Gangnon, PhD, Lucmen Harriaran, Ludmila D. Andersen, MS

**POPULATION** = Women diagnosed with breast cancer in Dane County, 1996-2000

Among other things, this study estimated that the rate of “breast cancer in situ (BCIS),” which is diagnosed almost exclusively via mammogram, is approximately 12-13%. That is, for any individual randomly selected from this population, we have a binary variable

\[ BCIS = \begin{cases} 
1, & \text{with probability 0.12} \\
0, & \text{with probability 0.88}. 
\end{cases} \]

In a random sample of \( n = 100 \) breast cancer diagnoses, let

\[ X = \# \text{BCIS cases} \quad (0,1,2,\ldots,100). \]

Questions:

- How can we model the probability distribution of \( X \), and under what assumptions?

- Probabilities of events, such as \( P(X = 0) \), \( P(X = 20) \), \( P(X \leq 20) \), etc.?

- Mean \# BCIS cases = ?

- Standard deviation of \# BCIS cases = ?

Full article available online at this [link](link).
Binomial Distribution (Paradigm model = coin tosses)

**Binary random variable:**

\[
Y = \begin{cases} 
1, & \text{Success (Heads)} \\
0, & \text{Failure (Tails)} 
\end{cases}
\]

- Probability: \( P(\text{Success}) = \pi \)
- Probability: \( P(\text{Failure}) = 1 - \pi \)

**Experiment:** \( n = 5 \) independent coin tosses

**Sample Space** \( S = \{(H \ H \ H \ H \ H), \ldots, (T \ T \ T \ T \ T)\} \)

\( \#(S) = 2^5 = 32 \)

**Random Variable:** \( X = \text{“# Heads in } n = 5 \text{ independent tosses (0, 1, 2, 3, 4, 5)”} \)

**Events:**

- “\( X = 0 \)” = Exercise
  \[ \#(X = 0) = \binom{5}{0} = 1 \]
- “\( X = 1 \)” = Exercise
  \[ \#(X = 1) = \binom{5}{1} = 5 \]
- “\( X = 2 \)” = Exercise
  \[ \#(X = 2) = \binom{5}{2} = 10 \]
- “\( X = 3 \)” = see above
  \[ \#(X = 3) = \binom{5}{3} = 10 \]
- “\( X = 4 \)” = Exercise
  \[ \#(X = 4) = \binom{5}{4} = 5 \]
- “\( X = 5 \)” = Exercise
  \[ \#(X = 5) = \binom{5}{5} = 1 \]

**Recall:** For \( x = 0, 1, 2, \ldots, n \), the combinatorial symbol \( \binom{n}{x} \) – read “\( n \)-choose-\( x \)” – is defined as the value \( \frac{n!}{x!(n-x)!} \), and counts the number of ways of rearranging \( x \) objects among \( n \) objects. See Appendix > Basic Reviews > Perms & Combos for details.

**Note:** \( \binom{n}{r} \) is computed via the mathematical function “\( n \text{Cr} \)” on most calculators.
Probabilities:

First assume the coin is fair \((\pi = 0.5 \Rightarrow 1 - \pi = 0.5)\), i.e., equally likely elementary outcomes \(H\) and \(T\) on a single trial. In this case, the probability of any event \(A\) above can thus be easily calculated via \(P(A) = \#(A) / \#(S)\).

<table>
<thead>
<tr>
<th>(x)</th>
<th>(P(X = x) = \frac{1}{2^5} \binom{5}{x})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/32 = 0.03125</td>
</tr>
<tr>
<td>1</td>
<td>5/32 = 0.15625</td>
</tr>
<tr>
<td>2</td>
<td>10/32 = 0.31250</td>
</tr>
<tr>
<td>3</td>
<td>10/32 = 0.31250</td>
</tr>
<tr>
<td>4</td>
<td>5/32 = 0.15625</td>
</tr>
<tr>
<td>5</td>
<td>1/32 = 0.03125</td>
</tr>
</tbody>
</table>

Now consider the case where the coin is biased (e.g., \(\pi = 0.7 \Rightarrow 1 - \pi = 0.3\)). Calculating \(P(X = x)\) for \(x = 0, 1, 2, 3, 4, 5\) means summing \(P(\text{all its outcomes})\).

Example: \(P(X = 3) = \)

\[
\begin{align*}
\text{outcome} & \quad \text{via independence of } H, T \\
\underbrace{P(H H H T T)} & = (0.7)(0.7)(0.7)(0.3)(0.3) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(H H T H T)} & = (0.7)(0.7)(0.3)(0.7)(0.3) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(H H T T H)} & = (0.7)(0.7)(0.3)(0.3)(0.7) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(H T H H T)} & = (0.7)(0.3)(0.7)(0.7)(0.3) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(H T H T H)} & = (0.7)(0.3)(0.7)(0.3)(0.7) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(H T T H H)} & = (0.7)(0.3)(0.3)(0.7)(0.7) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(T H H H T)} & = (0.3)(0.7)(0.7)(0.7)(0.3) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(T H H T H)} & = (0.3)(0.7)(0.7)(0.3)(0.7) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(T H T H H)} & = (0.3)(0.7)(0.3)(0.7)(0.7) = (0.7)^3 (0.3)^2 \\
+ \underbrace{P(T T H H H)} & = (0.3)(0.3)(0.7)(0.7)(0.7) = (0.7)^3 (0.3)^2 \\
\end{align*}
\]

\[
\text{via disjoint outcomes, } \quad = \binom{5}{3} (0.7)^3 (0.3)^2
\]
Hence, we similarly have...

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\binom{5}{0} (0.7)^0 (0.3)^5 = 0.00243$</td>
</tr>
<tr>
<td>1</td>
<td>$\binom{5}{1} (0.7)^1 (0.3)^4 = 0.02835$</td>
</tr>
<tr>
<td>2</td>
<td>$\binom{5}{2} (0.7)^2 (0.3)^3 = 0.13230$</td>
</tr>
<tr>
<td>3</td>
<td>$\binom{5}{3} (0.7)^3 (0.3)^2 = 0.30870$</td>
</tr>
<tr>
<td>4</td>
<td>$\binom{5}{4} (0.7)^4 (0.3)^1 = 0.36015$</td>
</tr>
<tr>
<td>5</td>
<td>$\binom{5}{5} (0.7)^5 (0.3)^0 = 0.16807$</td>
</tr>
</tbody>
</table>

Example: Suppose that a certain medical procedure is known to have a 70% successful recovery rate (assuming independence). In a random sample of $n = 5$ patients, the probability that three or fewer patients will recover is:

Method 1: $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$$= 0.00243 + 0.02835 + 0.13230 + 0.30870 = 0.47178$$

Method 2: $P(X \leq 3) = 1 - [P(X = 4) + P(X = 5)]$

$$= 1 - [0.36015 + 0.16807] = 1 - 0.52822 = 0.47178$$

Example: The mean number of patients expected to recover is:

$$\mu = E[X] = 0 (0.00243) + 1 (0.02835) + 2 (0.13230) + 3 (0.30870) + 4 (0.36015) + 5 (0.16807)$$

$$= 3.5 \text{ patients}$$

This makes perfect sense for $n = 5$ patients with a $\pi = 0.7$ recovery probability, i.e., their product. In the probability histogram above, the “balance point” fulcrum indicates the mean value of 3.5.
**General formulation:**

### The Binomial Distribution

Let the *discrete* random variable $X$ = “# Successes in *n independent Bernoulli trials* (0, 1, 2, …, *n*),” each having constant probability $P($Success$) = \pi$, and hence $P($Failure$) = 1 - \pi$. Then the probability of obtaining any specified number of successes $x = 0, 1, 2, …, n$, is given by:

$$P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}.$$  

We say that $X$ has a Binomial Distribution, denoted $X \sim \text{Bin}(n, \pi)$. Furthermore, the mean $\mu = n \pi$, and the standard deviation $\sigma = \sqrt{n \pi (1 - \pi)}$.

**Example:** Suppose that a certain spontaneous medical condition affects 1% (i.e., $\pi = 0.01$) of the population. Let $X =$ “number of affected individuals in a random sample of $n = 300$.” Then $X \sim \text{Bin}(300, 0.01)$, i.e., the probability of obtaining any specified number $x = 0, 1, 2, …, 300$ of affected individuals is:

$$P(X = x) = \binom{300}{x} (0.01)^x (0.99)^{300-x}.$$  

The mean number of affected individuals is $\mu = n \pi = (300)(0.01) = 3$ expected cases, with a standard deviation of $\sigma = \sqrt{(300)(0.01)(0.99)} = 1.723$ cases.

### Probability Table for Binomial Dist.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\binom{n}{0} \pi^0 (1-\pi)^{n-0}$</td>
</tr>
<tr>
<td>1</td>
<td>$\binom{n}{1} \pi^1 (1-\pi)^{n-1}$</td>
</tr>
<tr>
<td>2</td>
<td>$\binom{n}{2} \pi^2 (1-\pi)^{n-2}$</td>
</tr>
<tr>
<td>etc.</td>
<td>etc.</td>
</tr>
<tr>
<td>$n$</td>
<td>$\binom{n}{n} \pi^n (1-\pi)^{n-n}$</td>
</tr>
</tbody>
</table>

**Exercise:** In order to be a valid distribution, the sum of these probabilities must = 1. Prove it.

**Hint:** First recall the Binomial Theorem: How do you expand the algebraic expression $(a+b)^n$ for any $n = 0, 1, 2, 3, …$? Then replace $a$ with $\pi$, and $b$ with $1 - \pi$. Voilà!
Comments:

- The assumption of independence of the trials is absolutely critical! If not satisfied – i.e., if the “success” probability of one trial influences that of another – then the Binomial Distribution model can fail miserably. (Example: \( X = \) “number of children in a particular school infected with the flu”) The investigator must decide whether or not independence is appropriate, which is often problematic. If violated, then the correlation structure between the trials may have to be considered in the model.

- As in the preceding example, if the sample size \( n \) is very large, then the computation of \( \binom{n}{x} \) for \( x = 0, 1, 2, \ldots, n \), can be intensive and impractical. An approximation to the Binomial Distribution exists, when \( n \) is large and \( \pi \) is small, via the Poisson Distribution (coming up…).

- Note that the standard deviation \( \sigma = \sqrt{n \pi (1 - \pi)} \) depends on the value of \( \pi \). (Later…)
How can we estimate the parameter $\pi$, using a sample-based statistic $\hat{\pi}$?

**POPULATION**

<table>
<thead>
<tr>
<th>Binary random variable</th>
</tr>
</thead>
</table>
| $Y = \begin{cases} 
1, & \text{Success with probability } \pi \\
0, & \text{Failure with probability } 1 - \pi 
\end{cases}$ |

**Experiment:** $n$ independent trials

**SAMPLE**

$0/1$ $0/1$ $0/1$ $0/1$ $0/1$ $0/1$ $\ldots$ $0/1$  
$(y_1, y_2, y_3, y_4, y_5, y_6, \ldots, y_n)$

$y_1 + y_2 + y_3 + y_4 + y_5 + \ldots + y_n$

Let $X = \# \text{Successes in } n \text{ trials} \sim \text{Bin}(n, \pi)$  
$(n - X = \# \text{Failures in } n \text{ trials})$.

Therefore, dividing by $n$…

$$\frac{X}{n} = \text{proportion of Successes in } n \text{ trials}$$

$\hat{\pi} = p \ (= \bar{y}, \text{as well})$

and hence…

$q = 1 - p = \text{proportion of Failures in } n \text{ trials}$.

**Example:** If, in a sample of $n = 50$ randomly selected individuals, $X = 36$ are female, then the statistic $\hat{\pi} = \frac{X}{n} = \frac{36}{50} = 0.72$ is an estimate of the true probability $\pi$ that a randomly selected individual *from the population* is female. The probability of selecting a male is therefore estimated by $1 - \hat{\pi} = 0.28$. 
Poisson Distribution  
(Modeling rare events)

**Discrete Random Variable:**

\[ X = \# \text{ occurrences of a (rare) event } E, \text{ in a given interval of time or space, of size } T. \quad (0, 1, 2, 3, \ldots) \]

Assume:

1. All the occurrences of \( E \) are independent in the interval.
2. The mean number \( \mu \) of expected occurrences of \( E \) in the interval is proportional to \( T \), i.e., \( \mu = \alpha T \). This constant of proportionality \( \alpha \) is called the rate of the resulting Poisson process.

Then…

The Poisson Distribution

The probability of obtaining any specified number \( x = 0, 1, 2, \ldots \) of occurrences of event \( E \) is given by:

\[
P(X = x) = \frac{e^{-\mu} \mu^x}{x!}
\]

where \( e = 2.71828 \ldots \) (“Euler’s constant”).

We say that \( X \) has a Poisson Distribution, denoted \( X \sim \text{Poisson}(\mu) \). Furthermore, the mean is \( \mu = \alpha T \), and the variance is \( \sigma^2 = \alpha T \) also.

Examples:  
# bee-sting fatalities per year, # spontaneous cancer remissions per year, # accidental needle-stick HIV cases per year, hemocytometer cell counts
Example (see above): Again suppose that a certain spontaneous medical condition $E$ affects 1% (i.e., $\alpha = 0.01$) of the population. Let $X =$ “number of affected individuals in a random sample of $T = 300$.” As before, the mean number of expected occurrences of $E$ in the sample is $\mu = \alpha T = (0.01)(300) = 3$ cases. Hence $X \sim \text{Poisson}(3)$, and the probability that any number $x = 0, 1, 2, \ldots$ of individuals are affected is given by:

$$P(X = x) = \frac{e^{-3} \cdot 3^x}{x!}$$

which is a much easier formula to work with than the previous one. This fact is sometimes referred to as the Poisson approximation to the Binomial Distribution, when $T$ (respectively, $n$) is large, and $\alpha$ (respectively, $\pi$) is small. Note that in this example, the variance is also $\sigma^2 = 3$, so that the standard deviation is $\sigma = \sqrt{3} = 1.732$, very close to the exact Binomial value.

<table>
<thead>
<tr>
<th>$x$</th>
<th>Binomial $P(X = x) = \binom{300}{x} (0.01)^x (0.99)^{300-x}$</th>
<th>Poisson $P(X = x) = \frac{e^{-3} \cdot 3^x}{x!}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.04904</td>
<td>0.04979</td>
</tr>
<tr>
<td>1</td>
<td>0.14861</td>
<td>0.14936</td>
</tr>
<tr>
<td>2</td>
<td>0.22441</td>
<td>0.22404</td>
</tr>
<tr>
<td>3</td>
<td>0.22517</td>
<td>0.22404</td>
</tr>
<tr>
<td>4</td>
<td>0.16888</td>
<td>0.16803</td>
</tr>
<tr>
<td>5</td>
<td>0.10099</td>
<td>0.10082</td>
</tr>
<tr>
<td>6</td>
<td>0.05015</td>
<td>0.05041</td>
</tr>
<tr>
<td>7</td>
<td>0.02128</td>
<td>0.02160</td>
</tr>
<tr>
<td>8</td>
<td>0.00787</td>
<td>0.00810</td>
</tr>
<tr>
<td>9</td>
<td>0.00258</td>
<td>0.00270</td>
</tr>
<tr>
<td>10</td>
<td>0.00076</td>
<td>0.00081</td>
</tr>
<tr>
<td>etc.</td>
<td>$\rightarrow 0$</td>
<td>$\rightarrow 0$</td>
</tr>
</tbody>
</table>

Area $= 1$

Area $= 1$
Why is the Poisson Distribution a good approximation to the Binomial Distribution, for large \( n \) and small \( \pi \)?

(Rule of Thumb: \( n \geq 20 \) and \( \pi \leq 0.05 \); excellent if \( n \geq 100 \) and \( \pi \leq 0.1 \)).

Let \( f_{\text{Bin}}(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \) and \( f_{\text{Poisson}}(x) = \frac{e^{-\lambda} \lambda^x}{x!} \), where \( \lambda = n\pi \).

We wish to show formally that, for fixed \( \lambda \), and \( x = 0, 1, 2, \ldots \), we have:

\[
\lim_{n \to \infty} f_{\text{Bin}}(x) = f_{\text{Poisson}}(x).
\]

Proof: By elementary algebra, it follows that…

\[
f_{\text{Bin}}(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} = \frac{n!}{x! (n-x)!} \pi^x (1 - \pi)^{n} (1 - \pi)^{-x}
\]

\[
= \frac{1}{x!} n (n-1) (n-2) \ldots (n-x+1) \pi^x \left(1 - \frac{\lambda}{n}\right)^n (1 - \pi)^{-x}
\]

\[
= \frac{1}{x!} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \ldots \frac{n-x+1}{n} (n\pi)^x \left(1 - \frac{\lambda}{n}\right)^n (1 - \pi)^{-x}
\]

\[
= \frac{1}{x!} 1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{x-1}{n}\right) \lambda^x \left(1 - \frac{\lambda}{n}\right)^n (1 - \pi)^{-x}
\]

As \( n \to \infty \), \( \pi \to 0 \),

\[
\frac{1}{x!} 1(1)(1) \ldots (1) = 1 \quad \lambda^x \quad e^{-\lambda} \quad 1^{-x} = 1
\]

\[
= \frac{e^{-\lambda} \lambda^x}{x!} = f_{\text{Poisson}}(x). \quad \text{QED}
\]
Classical Discrete Probability Distributions

**Binomial** (probability of finding \( x \) “successes” and \( n - x \) “failures” in \( n \) independent trials)

\[
X = \# \text{ successes (each with probability } \pi) \text{ in } n \text{ independent Bernoulli trials, } n = 1, 2, 3, \ldots
\]

\[
f(x) = P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n
\]

**Negative Binomial** (probability of needing \( x \) independent trials to find \( k \) successes)

\[
X = \# \text{ independent Bernoulli trials for } k \text{ successes (each with probability } \pi), \quad k = 1, 2, 3, \ldots
\]

\[
f(x) = P(X = x) = \binom{x - 1}{k - 1} \pi^k (1 - \pi)^{x-k}, \quad x = k, k + 1, k + 2, \ldots
\]

**Geometric:** \( X = \# \text{ independent Bernoulli trials for } k = 1 \text{ success} \)

\[
f(x) = P(X = x) = \pi (1 - \pi)^{x-1}, \quad x = 1, 2, 3, \ldots
\]

**Hypergeometric** (modification of Binomial to sampling without replacement from “small” finite populations, relative to \( n \))

\[
X = \# \text{ successes in } n \text{ random trials taken from a population of size } N \text{ containing } d \text{ successes, } n > \frac{N}{10}
\]

\[
f(x) = P(X = x) = \frac{\binom{d}{x} \binom{N-d}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, d
\]

**Multinomial** (generalization of Binomial to \( k \) categories, rather than just two)

For \( i = 1, 2, 3, \ldots, k \),

\[
X_i = \# \text{ outcomes in category } i \text{ (each with probability } \pi_i) \text{ in } n \text{ independent Bernoulli trials, } n = 1, 2, 3, \ldots
\]

\[
\pi_1 + \pi_2 + \pi_3 + \ldots + \pi_k = 1
\]

\[
f(x_1, x_2, \ldots, x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! \ldots x_k!} \pi_1^{x_1} \pi_2^{x_2} \ldots \pi_k^{x_k}, \quad x_i = 0, 1, 2, \ldots, n \text{ with } x_1 + x_2 + \ldots + x_k = n
\]

**Poisson** (“limiting case” of Binomial, with \( n \to \infty \) and \( \pi \to 0 \), such that \( n\pi = \lambda \), fixed)

\[
X = \# \text{ occurrences of an event having mean number of occurrences } \lambda > 0
\]

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