A gallery of useful discrete probability distributions
Bernoulli distribution

The simplest non-uniform distribution

\[ p - \text{probability of success (1)} \]
\[ 1-p - \text{probability of failure (0)} \]

\[ f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} \]

Jacob Bernoulli
(1654-1705)
Swiss mathematician (Basel)

• Law of large numbers
• Mathematical constant e=2.718...
Bernoulli distribution

\[ f(x) = P(X = x) = \begin{cases} 
  p & \text{if } x = 1 \\
  1 - p & \text{if } x = 0 
\end{cases} \]

\[ E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = 0(1 - p) + 1(p) = p \]
\[ \text{Var}(X) = E(X^2) - (EX)^2 = [0^2(1 - p) + 1^2(p)] - p^2 = p - p^2 = p(1 - p) \]
Refresher: Binomial Coefficients

\[ \binom{n}{k} = \binom{n}{k} = \frac{n!}{k!(n-k)!} , \text{ called } n \text{ choose } k \]

\[ \binom{10}{3} = \binom{10}{3} = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{3 \cdot 2 \cdot 1 \cdot 7!} = 120 , \text{ called } 10 \text{ choose } 3 \]

Number of ways to choose k objects out of n without replacement and where the order does not matter. Called binomial coefficients because of the binomial formula

\[ (p + q)^n = \sum_{x=0}^{n} C^n_x p^x q^{n-x} \]
Binomial Distribution

- **Binomially-distributed** random variable $X$ equals sum (number of successes) of $n$ independent Bernoulli trials.

- The probability mass function is:

  $$f(x) = C^n_x p^x q^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$  \hspace{1cm} (3-7)

- Based on the binomial expansion:

  $$f = (p+q)^n = \sum_{x=0}^{n} C^n_x p^x q^{n-x}$$
Binomial Mean and Variance

$X$ is a binomial random variable with parameters $p$ and $n$.

Mean:
$\mu = E(X) = np$

Variance:
$\sigma^2 = V(X) = np(1-p)$

Standard deviation:
$\sigma = \sqrt{np(1-p)}$
Matlab exercise: Binomial distribution

• Generate a sample of size 100,000 for binomially-distributed random variable X with n=100, p=0.2
• Tip: generate n Bernoulli random variables and use sum to add them up
• Plot the approximation to the Probability Mass Function based on this sample
• Calculate the mean and variance of this sample and compare it to theoretical calculations: E[X]=n*p and V[X]=n*p*(1-p)
Matlab exercise: Binomial distribution

- n=100; p=0.2;
- Stats=100000;
- r1=rand(Stats,n)<p;
- r2=sum(r1,2);
- mean(r2)
- var(r2)
- [a,b]=hist(r2, 0:n);
- p_b=a./sum(a);
- figure; stem(b,p_b);
- figure; semilogy(b,p_b,'ko-')
Poisson Distribution

• Limit of the binomial distribution when
  – $n$, the number of attempts, is very large
  – $p$, the probability of success is very small
  – $E(X)=np$ is just right

The annual numbers of deaths from horse kicks in 14 Prussian army corps between 1875 and 1894

<table>
<thead>
<tr>
<th>Number deaths</th>
<th>Observed frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>144</td>
<td>139</td>
</tr>
<tr>
<td>1</td>
<td>91</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5 and over</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>280</td>
<td>280</td>
</tr>
</tbody>
</table>

From von Bortkiewicz 1898

Siméon Denis Poisson (1781–1840)
French mathematician and physicist
Let $\lambda = np = E(x)$, so $p = \lambda/n$

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

$$= \frac{n(n-1)\ldots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \sim \frac{n^x}{x!} \left(\frac{\lambda}{n}\right)^x = \frac{\lambda^x}{x!};$$

$$\sum_x \frac{\lambda^x}{x!} = e^\lambda.$$ 

Normalization requires $\sum_x P(X = x) = 1$.

Thus $P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$
Poisson Mean & Variance

If \( X \) is a Poisson random variable, then:

- **Mean:** \( \mu = E(X) = \lambda = \mu \cdot p \)
- **Variance:** \( \sigma^2 = \text{Var}(X) = \lambda = \mu \cdot p \cdot (1-p) \approx \mu \cdot p \)
- **Standard deviation:** \( \sigma = \lambda^{1/2} \)

Note: Variance = Mean

Note: Standard deviation/Mean = \( \lambda^{-1/2} \) decreases with \( \lambda \)
Matlab exercise: Poisson distribution

• Generate a sample of size 100,000 for Poisson-distributed random variable $X$ with $\lambda = 2$

• Plot the approximation to the Probability Mass Function based on this sample

• Calculate the mean and variance of this sample and compare it to theoretical calculations:
  $E[X] = \lambda$ and $V[X] = \lambda$
Matlab exercise: Binomial distribution

- Stats=100000; lambda=2;
- r2=random('Poisson',lambda,Stats,1);
- mean(r2)
- var(r2)
- [a,b]=hist(r2, 0:max(r2));
- p_p=a./sum(a);
- figure; stem(b,p_p);
- figure; semilogy(b,p_p,'ko-')
Secretary problem or Picky bride problem

• A future bride has **known (large) number** – n – of suitors whom she dates one at a time
• She can easily **evaluate and rank the “quality” of suitors relative to each other but has no idea of its distribution**
• She has only **one chance to choose** her husband and cannot go back to the guys she dumped
• How can she **maximize the probability to choose the best groom** among n suitors?