Today: ASV 2.4 and 2.5

Next: ASV 3.3
Last time: Independent trials of success/failure

sequence of experiments $(\Omega_i, \mathcal{F}_i, P_i), i=1, \ldots, n$ \hspace{0.5cm} $P_i = P_j$

Common probability space: Product space

define $(\Omega, \mathcal{F}, P)$ by $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ for $\omega=(\omega_1, \omega_2, \ldots, \omega_n) \in \Omega$ $P((\omega_1, \omega_2, \ldots, \omega_n)) = P_1(\omega_1)P_2(\omega_2)\cdots P_n(\omega_n)$

The simplest situation: $\Omega_i = \{\text{success}, \text{failure}\}$ only two outcomes

$P_i(\text{success}) = p$ $P_i(\text{failure}) = 1-p$

Random variables: indicators of success

$X_i: \Omega_i \rightarrow \mathbb{R}$, $X_i(\text{success}) = 1$ $P(X_i=1) = p$ $X_i(\text{failure}) = 0$ $P(X_i=0) = 1-p$
Bernoulli random variable (parameter $p$)

Each of $X_i$'s above is called Bernoulli r.v.

**Def.** We say that r.v. $X$ has Bernoulli distribution with success rate $p$, $p \in [0,1]$, if $P(X = 1) = 1 - P(X = 0) = p$

In this case we write $X \sim \text{Ber}(p)$

Binomial random variable (parameters $n$ and $p$)

Gives the number of successful trials in the sequence of $n$ independent "Bernoulli experiments".

Let $X_1, \ldots, X_n$ independent r.v., for each $i$ $X_i \sim \text{Ber}(p)$.

Denote $S_n := X_1 + \cdots + X_n$. What are possible values of $S_n$?

$S_n \in \{0, 1, \ldots, n\}$. Compute $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$

($k$ successes out of $n$ $\rightarrow$ $n$ choose $k$, each with proba $p$)
Binomial r.v. Example

So as above is binomial r.v. with parameters \( n \) and \( p \).

**Def.** For \( n \in \mathbb{N} \) and \( 0 < p \leq 1 \), we say that r.v. \( X \) has binomial distribution with parameters \( n \) and \( p \), denoted \( X \sim B(n,p) \), if \( X \in \{0,1,\ldots,n\} \) and for any \( k \in \{0,\ldots,n\} \) \( P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \).

**Example** Choose a point uniformly a random from unit disk. You win if you are at distance \( \leq \frac{1}{2} \) to the center. You play this game 6 times. What is the probability that you win 4 or 5 times out of 6?

\( X_i = \) "win in \( i \)-th game" \( \Rightarrow S_6 = \sum_{i=1}^{6} X_i \sim B(6,\frac{1}{4}) \).

\( X_i \sim Ber(\frac{1}{4}) \) \( \Rightarrow P(S_6 \in \{4,5\}) = P(S_6 = 4) + P(S_6 = 5) = \binom{6}{4} \left(\frac{1}{4}\right)^4 \left(\frac{1}{4}\right)^2 + \binom{6}{5} \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^1 \).
**Geometric distribution (with parameter p)**

Gives the number of the first successful trial

**Def** Let \( p \in (0, 1) \). We say that r.v. \( X \) has geometric distribution with parameter \( p \) if for any \( k \in \mathbb{N} \)

\[
P(X = k) = (1-p)^{k-1} p
\]

Denote \( X \sim \text{Geom}(p) \)
Example of geometric distribution

Example You play roulette (18 red, 18 black and 0) and you bet on red on each play. What is the probability that you first win in the 5th play?

Let \( N = \) first time you win. We have sequence of independent trials, for each trial probability of success is \( P(\text{the outcome is red}) = \frac{18}{37} \)

Then \( N \sim \text{Geom}(\frac{18}{37}) \), so \( P(N = 5) = \binom{4}{19} \left( \frac{18}{37} \right) \left( \frac{19}{37} \right) = 0.033 \)
Examples of binomial/geometric distributions

Example Baseball player hits the ball 30% of times.

1) \( A = \{ \text{player bats 6 times and hits the ball exactly 4 times} \} \)

2) \( B = \{ \text{player bats consecutively and first hits the ball on 6th try} \} \)

3) \( C = \{ \text{player bats 10 times and hits the ball at least 2 times} \} \)

Let \( X \sim B(6, 0.3) \), \( Y \sim \text{Geom}(0.3) \), \( Z \sim B(10, 0.3) \)

Then

\[
P(A) = P(X = 4) = \binom{6}{4} 0.3^4 0.7^2
\]

\[
P(B) = P(Y = 6) = \ldots
\]

\[
P(C) = P(Z \geq 2) = 1 - P(Z \leq 1) = 1 - P(Z = 0) - P(Z = 1) = \ldots
\]

\[
= 1 - \binom{10}{0} 0.3^0 0.7^{10} - \binom{10}{1} 0.3^1 0.7^9
\]
**Poisson distribution**

**Def.** Let \( \lambda > 0 \) and let \( X \) be a r.v. taking values in \( \{0, 1, 2, \ldots \} \). \( X \) has Poisson distribution with parameter \( \lambda \), if

\[
P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \text{for } k \in \{0, 1, \ldots \}
\]

We write \( X \sim \text{Pois}(\lambda) \).

**Rem.** Poisson distribution describes the probability of a "rare" event happening \( k \) times after trying (repeating the experiment) "many" times. Parameter \( \lambda > 0 \) gives the "expected number" of occurrences.
Example of Poisson distribution

Observation: between 1875 and 1894 (20 years) in 14 units of Prussian army, there were 196 deaths from horse kicks, distributed in the following way:

<table>
<thead>
<tr>
<th>Number of deaths per unit per year, k</th>
<th>Number of units (in 20 years) with k deaths/year</th>
<th>( P(A_k) ) (empirical probability)</th>
<th>( P(X=k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>149</td>
<td>0.51</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>91</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>5+</td>
<td>0 (280 total)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Denote by \( A_k \) = \{unit-year chosen at random (from 280) had k deaths\), \( P(A_k) = \text{column #2} / 280 \), \( X \sim \text{Poisson} \left( \frac{196}{280} \right) \), \( P(X=0) = e^{-0.7} \approx 0.5 \)
Birthday problem: solution

What is the probability that in a room with \( k \) people there are at least 2 with the same birthday?

How big \( k \) should be so that the above probability is \( > \frac{1}{2} \)? Let \( A_k = \{ \text{at least 2 among } k \text{ share birthday} \} \)

\[
P(A_k) = 1 - P(A_k^c), \quad A_k^c = \{ \text{all } k \text{ people have bdays of different dates} \}
\]

\[
P(A_k^c) = \frac{365 \cdot 364 \cdots (365-k+1)}{365^k} \quad \text{decreasing with } k
\]

\[
P(A_k) \text{ increasing with } k, \text{ if we compute for different } k's
\]

we get \( P(A_{22}) \approx 0.148 \quad P(A_{23}) \approx 0.151 \)
The Sally Clark case

Assume independence carefully!
During 1996-1998, two infants of Sally Clark died without obvious causes, the woman was arrested on suspicion of murder.
In the court, an "expert" presented statistical data

\[ P(\text{unexplained infant death}) = \frac{1}{8500} \]

"Expert" concluded:

\[ P(\text{two infants die in one family}) = \left( \frac{1}{8500} \right)^2 \]

Clark was convicted, spent 3 years in prison. Later studies showed that 2nd infant died from natural causes.