Today: Random variables

Next: ASV 3.2

Week 3:

- homework 1 graded, regrades open until October 16
- homework 3 (due Friday, October 21)
- midterm 1: Wednesday, October 19
Random variables

$(\Omega, \mathcal{F}, P)$ - probability space

Def A (measurable) function $X: \Omega \to \mathbb{R}$ is called a random variable.

Def Let $X$ be a random variable. The probability distribution of $X$ is the collection of probabilities $P(X \in B)$ for all $B \subset \mathbb{R}$.
If \((\Omega, \mathcal{F}, P)\) is a probability space, and \(X: \Omega \to \mathbb{R}\) is a random variable, we can define a probability measure \(\mu_x\) on \(\mathbb{R}\) given, for any \(A \subset \mathbb{R}\), by

\[
\mu_x(A) = P(X \in A) = P(\{\omega : X(\omega) \in A\})
\]

We call \(\mu_x\) the probability distribution (or law) of \(X\).

5) Toss a fair coin 4 times. \(X = \) number of tails

\[
\Omega = \{(X_1, X_2, X_3, X_4) \in \{H, T\}^4\}
\]

\(P = \) uniform on \(\Omega\)

\[
P((X_1, X_2, X_3, X_4)) = \frac{1}{2^4} = \frac{1}{16}
\]

\[
P_x(k) = \mu_x(\{k\}) = \frac{\binom{4}{k}}{16}, \quad 0 \leq k \leq 4
\]

If \(A \subset \mathbb{R}\) does not contain one of these numbers, then

\[
\mu_x(A) = 0
\]

Enough to know \(\mu_x(\{k\})\)
Probability distribution

Toss a fair coin 4 times. Let \( X = \) number of tails.

\( X \in \{0, 1, 2, 3, 4\} \)

\[
P_X(k) = P(X = k) = \frac{\left(\frac{4}{2}\right)^k}{2^4}
\]

\[0 \leq k \leq 4\]

\[
\frac{1}{4} = P_X(1)
\]

<table>
<thead>
<tr>
<th>( k )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_X(k) )</td>
<td>( \frac{1}{16} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{16} )</td>
</tr>
</tbody>
</table>

More generally, if \( X = \# \) tails in \( n \) tosses,

\[
P_X(k) = P(X = k) = \frac{\left(\frac{n}{2}\right)^k}{2^n}, \quad 0 \leq k \leq n
\]

We call this distribution Binomial with parameters \( n, \frac{1}{2} \)
4) Choose point $w$ from unit disk uniformly at random

$$X(w) = \text{dist}(0, w)$$

$$P(X \leq r) = \begin{cases} 
0, & r < 0 \\
\frac{r^2}{\pi}, & 0 \leq r \leq 1 \\
1, & r > 1
\end{cases}$$

$$P(X \in (-\infty, r]) = \mu_x((0, r]) = P(X = 0.4)?$$

What can we say about $P(X \in A)$ for other sets $A \subseteq \mathbb{R}$?

Take $A = (0.3, 0.4)$, $(-\infty, 0.4] = (-\infty, 0.3] \cup (0.3, 0.4]$.

$$\mu_x((-\infty, 0.4]) = \mu_x((-\infty, 0.3]) + \mu_x((0.3, 0.4])$$

$$= \frac{(0.3)^2}{\pi} + \frac{(0.4)^2 - (0.3)^2}{\pi} = \frac{0.07}{\pi}$$

$$P(X \in (0.4 - \varepsilon, 0.4]) = \frac{(0.4)^2 - (0.4 - \varepsilon)^2}{\pi} = \frac{2 \cdot 0.4 \varepsilon - \varepsilon^2}{\pi} \to 0 \quad \text{as} \ \varepsilon \to 0$$

For any $\varepsilon > 0$, $0 \leq P(X = 0.4) \leq P(X \in (0.4 - \varepsilon, 0.4]) = 2 \cdot 0.4 \varepsilon - \varepsilon^2 \to 0$ as $\varepsilon \to 0$.
Discrete and continuous random variables

**Discrete**: There are finitely (or countably) many possible outcomes \( \{k_1, k_2, k_3, \ldots \} \) for \( X \).

\( \mu_x \) is described by the probability mass function:

\[
p_x(k) = P(X = k) \quad \text{for} \quad k \in \{k_1, k_2, k_3, \ldots \}
\]

In this case, by the laws of probability:

\[
p_x(k) \geq 0 \quad \text{for each} \quad k, \quad \sum_{j=1}^{\infty} p_x(k_j) = 1
\]

**Continuous**: For any real number \( t \in \mathbb{R} \), \( P(X = t) = 0 \).

\( \mu_x \) is captured by understanding \( P(X \leq r) \) as a function of \( r \).

For example, \( P(X \in [a, b]) = P(X = a) + P(X \in (a, b]) \) or

\[
P(X \leq b) - P(X \leq a)
\]
Cumulative Distribution Function (CDF)

For any random variable $X$, define

CDF: $F_X(r) = P(X \leq r)$

Example: $X \sim \text{Bin}(3, \frac{1}{2})$

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_X(k)$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{3}{8}$</td>
<td>$\frac{1}{8}$</td>
</tr>
</tbody>
</table>