Random Variables and Distributions
Lecture for Economics 240A

Douglas G. Steigerwald

UC Santa Barbara

September 2011
What is a fair bet in a game of chance?
You are offered to bet on the outcome of a die roll
  - payout - $1 for each pip showing

*If the die is fair, should you pay $4 to take this bet?*

How would you determine if you should take this bet?
Framework: Random Variable

- the sample space is \{S1, \ldots, S6\}
  - we must convert outcomes to dollar amounts
- \(X\) a random variable that denotes $payoff from each outcome

\[
\begin{array}{cccc}
S1 & S2 & \cdots & S6 \\
\downarrow & \downarrow & \cdots & \downarrow \\
X: & 1 & 2 & \cdots & 6 \\
\end{array}
\]

- \(X(\omega)\) is a mapping from sample space to \(\mathbb{R}\)
  - not random
  - not a variable
Distribution

*Distribution* - enumerates values the random variable takes and the associated probabilities

- **Discrete random variable** takes a countable number of values
- **Values** \( X = x_i \)
  - 6 distinct values \( x_1 = 1, \ldots, x_6 = 6 \)
- **Probabilities** \( \mathbb{P}(X = x_i) \)
  - \( \mathbb{P}(X = 1) = \mathbb{P}(S1) = 1/6 \)
  - countable number of values allows \( \mathbb{P}(X = x_i) > 0 \)
  - \( \sum_{i=1}^{6} \mathbb{P}(X = x_i) = 1 \)

- **Distribution of** \( X \)

\[
\mathbb{P}(X = x_i) = 1/6 \quad i = 1, \ldots, 6
\]

- a multinomial distribution (with equal probabilities)
Operators

*Operator - a specific mathematical procedure*

- **Summation Operator**
  - arguments: sequence of real numbers \( \{w_i\}_{i=1}^n \)
  - \( w_i \) can be a function of real numbers: \( w_i = x_i p_i \)

\[
\sum_{i=1}^{n} w_i \equiv w_1 + w_2 + \cdots + w_n
\]

- **Expectation Operator**
  - arguments: random variable outcomes and probabilities
  - defined for all \( k > 0 \) (\( k^{th} \) moment of \( X \))
  - discrete random variable
    - countable number of outcomes \( \rightarrow \) summation operator

\[
\mathbb{E} \left( X^k \right) = \sum_{i=1}^{n} x_i^k \cdot \mathbb{P} \left( X = x_i \right)
\]
Initial Question Answered

*If the die is fair, should you pay $4 to take this bet?*

- What is the expected payoff from a roll of the die?
- Answer: the expected value, or mean, of $X$

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

$$= \sum_{i=1}^{6} x_i \cdot \frac{1}{6}$$

$$= \frac{1}{6} \sum_{i=1}^{6} i = 3.5$$

- note, this value is not an outcome, so the mean is not "the outcome we would expect to occur"
- if we play many times, the average payoff is $3.50$

- You are asked to pay $4 to place a bet that returns $3.50 on average
Distribution
Discrete - Binomial

- Example - sex at birth

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

- 2 values - binomial
  - values are \{0, 1\} - Bernoulli

- Distribution
  - $\Pr(X = x_i) = p_i \quad i = 0, 1$

- Mean
  - $\mathbb{E}(X) = 0 \cdot (1 - p_1) + 1 \cdot p_1 = p_1$
Distribution
Continuous - Normal (Gaussian)

- Example - temperature deviations from 70°
  - continuous random variable - uncountable number of values

- Probabilities
  - $P(X = x_i) = 0$ for all $i$
  - density function $f(x)$ satisfies:
    - $\int_{x_1}^{x_2} f(x) \, dx = P(x_1 < X < x_2)$
    - $\int_{-\infty}^{\infty} f(x) \, dx = 1$

- Moments
  - $E(X^k) = \int_{-\infty}^{\infty} x^k f(x) \, dx$ (discrete $\sum_{i=1}^{n} x_i^k P(X = x_i)$

- Gaussian: $N(\mu, \sigma^2)$
  - $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} (x-\mu)^2}$
  - $E(X) = \mu$
Mean - Mathematical Details

Definition

- if \( y \) is discrete on the set \( \{\tau_1, \tau_2, \ldots\} \), then

\[
\mathbb{E}y = \sum_{j=1}^{\infty} \tau_j \mathbb{P}(y = \tau_j)
\]

- if \( y \) is continuous with density \( f \), then

\[
\mathbb{E}y = \int_{-\infty}^{\infty} yf(y) \, dy
\]

- unify the definitions by writing the expectation as the Lebesgue integral with respect to the distribution function \( F \)

\[
\mathbb{E}y = \int_{-\infty}^{\infty} ydF(y)
\]

Lebesgue Integration
Mean - Mathematical Details

Existence

- if $\mathbb{E}y = \int_{-\infty}^{\infty} y \, dF(y)$ is not finite, then separately evaluate the integrals

\[
l_1 = \int_{0}^{\infty} y \, dF(y)
\]
\[
l_2 = -\int_{-\infty}^{0} y \, dF(y)
\]

- if $l_1 = \infty$ and $l_2 < \infty$, it is typical to define $\mathbb{E}y = \infty$
- if $l_1 < \infty$ and $l_2 = \infty$, it is typical to define $\mathbb{E}y = -\infty$
- if both $l_1 = \infty$ and $l_2 = \infty$, then $\mathbb{E}y$ is undefined

- if $\mathbb{E} |y| = \int_{-\infty}^{\infty} |y| \, dF(y) = l_1 + l_2 < \infty$
  - then $\mathbb{E}y$ exists and is finite
  - common to say, the mean $\mathbb{E}y$ is "well defined"
Mean - Alternative Definition

- for any non-negative random variable \( y \)

\[
\mathbb{E} y = \int_0^\infty \mathbb{P} (y > u) \, du
\]

- Proof: Let \( F^* (x) = \mathbb{P} (y > x) = 1 - F(x) \)
  - \( F(x) \) is the distribution function for \( y \)

- integration by parts

\[
\begin{align*}
\mathbb{E} y &= \int_0^\infty ydF (y) = - \int_0^\infty ydF^* (y) \\
&= - [yF^*(y)]_0^\infty + \int_0^\infty F^*(y) \, dy \\
&= \int_0^\infty \mathbb{P} (y > u) \, du.
\end{align*}
\]
Cumulative Probability Distribution

- definition: \( P(X \leq x) \)
  - implication: \( P(X > x) = 1 - P(X \leq x) \)

- Discrete random variable
  - fraction of outcomes \( \leq x \)

- Continuous random variable
  - \( F(x) = \int_{-\infty}^{x} f(t) \, dt \)
    - \( 0 \leq F(x) \leq 1 \)
    - Cumulative Distribution Function (CDF)

- Yields percentiles
  - 95th percentile: value of \( x \) such that \( F(x) = .95 \)
What did I get on the SAT?

- **how to interpret scores from the SAT**
  - $X$ - score on mathematics component
  - $X \in (200, 800)$

- **Percentiles and scores**
  - $X = 800 \rightarrow 99.93$ percentile
    - 7 in 10,000 get a perfect score
  - 88th percentile $\rightarrow X = 640$
    - 88 percent of scores are $\leq 640$
Review

Where to place bets

- **Roulette:** 38 slots - \{00, 0, 1, 2, \ldots, 36\}

  \[
  \begin{array}{cccccccc}
  S00 & S0 & S1 & S2 & \cdots & S36 \\
  \downarrow & \downarrow & \downarrow & \downarrow & \cdots & \downarrow \\
  X : & 00 & 0 & 1 & 2 & \cdots & 36 \\
  W : & 0 & 0 & 0 & 1 & \cdots & 1 \\
  \end{array}
  \]

- **Bet:** $1 on ball in even numbered slot, payoff - $2
  
  - \(P(\text{win}) \equiv (W = 1) = \frac{18}{38}\)
  - \(\mathbb{E}(\text{winnings}) = 2 \cdot \frac{18}{38} - 1 = -\frac{1}{19} \approx -0.05263157894736842\

- **Lotteries:** Payoffs set by state law
  
  - \(\mathbb{E}(\text{winnings}) = -0.50\)

- **Stock Market:** (long term holding of index, annual return)
  
  - \(\mathbb{E}(\text{winnings}) = +0.10\)
Lebesgue Integration - Intuition

- Integral of non-negative function between limits $a$ and $b$ can be interpreted as "area under the curve"
  - Continuous function on closed, bounded interval - area could be defined as an integral
  - Computed via approximation of the region by polygons
    - Approximation introduced by Bernhard Riemann in 1854

- Lebesgue integration - introduced by Henri Lebesgue in 1904:
  - Extends the integral to more irregular functions
  - Extends integration onto spaces more general than the real line
Lebesgue Integration - Graph Definition

- insight - one should be able to rearrange the values of the function while preserving the integral
  - Riemann integration - divide the base of the figure (the $x$-axis) into equal segments, sum the areas of the vertical rectangles
  - Lebesgue integration - divide the height of the figure (the $y$-axis) into equal segments, sum the areas of the horizontal rectangles

- Lebesgue integral of $f : \mathbb{R} \to \mathbb{R}^+$ is the sum of the thin horizontal strips between $y = t$ and $y = t + dt$

\[
\int f = \int_0^\infty f^*(t) \, dt
\]

- $f^*(t) = \mu(\{x | f(x) > t\})$
- $\mu(\cdot)$ a measure
  - assigns to each set $A$ of real numbers a nonnegative number $\mu(A)$
  - "size" of $A$ (should agree with length of an interval)