Overview

• Statistics important for game analysis
• Probability important for statistics
• So, understand some basic probability
• Also, probability useful for game development

Probability – Definitions

• Exhaustive set of events – set of all possible outcomes of experiment/observation
• Mutually exclusive sets of events – elementary events in that do not overlap

Probability Introduction

• Probability – way of assigning numbers to outcomes to express likelihood of event
• Event – outcome of experiment or observation
  – Elementary – simplest type for given experiment
  – Joint/Compound – more than one elementary

Outline

• Introduction (done)
• Probability (next)
• Probability Distributions

Overview

• Probability attack will succeed
• Probability loot from enemy contains rare item
• Probability enemy spawns at particular time
• Probability action (e.g., building a castle) takes particular amount of time
• Probability players at server

4/24/2018
Probability – Definition

- **Probability** – likelihood of event to occur, measured by ratio of favorable cases to all cases
- Set of rules that probabilities must follow
  - Probabilities must be between 0 and 1 (but often written/said as percent)
  - Probabilities of set of exhaustive, mutually exclusive events must add up to 1
- e.g., d6: events 1, 2, 3, 4, 5, 6. Probability of 1/6th to each
  - legal set of probabilities
- e.g., d6: events 1, 2, 3, 4, 5, 6. Probability of ½ to roll 1, ½ to roll 2, and 0 to all the others
  - Also legal set of probabilities
- Not how honest d6’s behave in real life!

Assigning Probabilities

- **Classical** (by theory)
  - In many cases, exhaustive, mutually exclusive outcomes equally likely
    - assign each outcome probability of 1/n
  - e.g., d6: 1/6, Coin: prob heads ½, tails ½, Cards: pick Ace 1/13
- **Empirically** (by observation)
  - Obtain data through measuring/observing
  - e.g., Watch how often people play FIFA 18 in lab versus some other game. Say, 30% FIFA. Assign that as probability
- **Subjective** (by hunch)
  - Based on expert opinion or other subjective method
  - e.g., e-sports writer says probability Team SoloMid (League team) will win World Championship is 25%

Rules About Probabilities (1 of 2)

- **Complement**: A an event. Event “A does not occur” called complement of A, denoted A’
  - P(A’) = 1 - P(A)
  - e.g., d6: P(6) = 1/6, complement is P(6’) and probability of not 6 is 1 - 1/6, or 5/6
    - Note: when using p, complement is often q
- **Mutually exclusive**: Have no simple outcomes in common – can’t both occur in same experiment
  - P(A or B) = P(A) + P(B)
  - e.g., d6: P(3 or 6) = P(3) + P(6) = 1/6 + 1/6 = 2/6

Rules About Probabilities (2 of 2)

- **Independence**: One occurs doesn’t affect probability that other occurs
  - e.g., 1d6: A die 1 gets 5, B: die 2 gets 6. Independent, since result of one roll doesn’t affect roll of other
  - Probability both occur P(A and B) = P(A) x P(B)
  - e.g., 2d6: prob of “snake eyes” is P(1) x P(1) = 1/6 x 1/6 = 1/36
- **Not independent**: One occurs affects probability that other occurs
  - Probability both occur P(A and B) = P(A) x P(B | A)
  - Where P(B | A) means the prob B given A happened
  - e.g., MMO has 10% mages, 40% warriors, 80% Boss defeated. Probability Boss fights mage and is defeated?
    - Probability Boss fights mage and is defeated?
  - You might think that = P(mage) x P(defeat B | mage) = .10 * .8  = .08
  - But likely not independent. P(defeat B | mage) < 80%. So, need non-independent formula P(mage)* P(defeat B | mage)

Probability Example

- Probability drawing King?
  - Probability drawing King?)
    - P(K) = \( \frac{1}{4} \)
  - Draw, put back. Now?
• Probability drawing King?
  \( P(K) = \frac{1}{4} \)
• Draw, put back. Now?
  \( P(K) = \frac{1}{4} \)
• Probability not King?
  \( P(K') = 1 - P(K) = \frac{3}{4} \)
• Draw, put back. 2 Kings?
  \( P(K) \times P(K) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \)

• Probability drawing King?
  \( P(K) = \frac{1}{4} \)
• Draw, put back. Now?
  \( P(K) = \frac{1}{4} \)
• Probability not King?
  \( P(K') = 1 - P(K) = \frac{3}{4} \)
• Draw, put back. 2 Kings?
  \( P(K) \times P(K) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \)

• Draw. King or Queen?
  \( P(K \text{ or } Q) = P(K) + P(Q) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \)
• Draw, put back.  Not 2 Kings?
  \( P(K) = \frac{1}{4} \)

• Probability drawing King?
  \( P(K) = \frac{1}{4} \)
• Draw, put back. Now?
  \( P(K) = \frac{1}{4} \)
• Probability not King?
  \( P(K') = 1 - P(K) = \frac{3}{4} \)
• Draw, put back. 2 Kings?
  \( P(K) \times P(K) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \)
Probability Example
- Draw. King or Queen?
  \[ P(K \text{ or } Q) = P(K) + P(Q) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \]
- Draw, put back. Not 2 Kings?
  \[ P(K') \times P(K') = \frac{3}{4} \times \frac{3}{4} = \frac{9}{16} \]
- Draw, don't put back. Not 2 Kings?
  \[ P(K') \times P(K') = \frac{3}{4} \times \frac{2}{3} = \frac{6}{12} = \frac{1}{2} \]
- Draw, don't put back. King 2\textsuperscript{nd} card?

Outline
- Intro (done)
- Probability (done)
- Probability Distributions (next)

Probability Distributions
- Probability distribution – values and likelihood of those values that random variable can take
  - Why? If can model mathematically, can use to predict occurrences
  - e.g., probability slot machine pays out on given day
  - e.g., probability game server hosts player today
  - e.g., probability certain game mode is chosen by player
  - Also, some statistical techniques for some distributions only

Uniform Distribution
- “So what?”
- Can use known formulas
Uniform Distribution

- “So what?”
- Can use known formulas

<table>
<thead>
<tr>
<th>Mean</th>
<th>( \frac{a + b}{2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>( \frac{a + b}{2} )</td>
</tr>
<tr>
<td>Mode</td>
<td>N/A</td>
</tr>
<tr>
<td>Variance</td>
<td>((b - a + 1)^2 - 1) / 12</td>
</tr>
<tr>
<td>Std Dev</td>
<td>( \sqrt{\text{Variance}} ) = 1.7</td>
</tr>
</tbody>
</table>

Mean = \( \frac{1 + 6}{2} = 3.5 \)
Variance = \((6 - 1 + 1)^2 - 1)/12 = 2.9
Std Dev = \( \sqrt{\text{Variance}} \) = 1.7

Binomial Distribution Example (1 of 3)

- Suppose toss 3 coins
- Random variable \( X = \) number of heads
- Want to know probability of exactly 2 heads
  \( P(X=2) = ? \)

How to assign probabilities?

- Could measure \( \text{(empirical)} \)
  - \( Q: \) how?
- Could use "hunch" \( \text{(subjective)} \)
  - \( Q: \) what do you think?
- Could use theory \( \text{(classical)} \)
  - Math using our probability rules (not shown)
  - Enumerate (next)

Binomial Distribution Example (2 of 3)

All equally likely (\( p = 1/8 \) for each)
\( P(HHT) + P(HTH) + P(THH) = 3/8 \)

Can draw histogram of number of heads

Binomial Distribution Example (3 of 3)

These are all binomial distributions

Binomial Distribution (1 of 2)

- In general, any number of trials \( n \) & any probability of successful outcome \( p \)
  (e.g., heads)

- Characteristics of experiment that gives random number with binomial distribution:
  - Experiment consists of \( n \) identical trials.
  - Each trial results in only two possible outcomes, \( S \) or \( F \)
  - Probability of \( S \) each trial is same, denoted \( p \)
  - Trials are independent
  - Random variable of interest \( X \) is number of \( S \)'s in \( n \) trials

Binomial Distribution (2 of 2)


http://www.mathnstuff.com/math/spoken/here/2class/90/binom2.gif

http://www.vassarstats.net/textbook/f0603.gif
Binomial Distribution (2 of 2)

- "So what?"
- Can use known formulas

\[
\begin{align*}
\text{Mean: } & \mu = np \\
\text{Variance: } & \sigma^2 = npq \\
\text{SD: } & \sigma = \sqrt{npq}
\end{align*}
\]

Excel: binom.dist()
- 2 heads, 3 flips
  \[= \text{binom.dist}(2, 3, 0.5, \text{FALSE}) = 0.375 \text{ (i.e., 3/8)}\]

Poisson Distribution

- Distribution of probability of events occurring in certain interval (broken into units)
  - Interval can be time, area, volume, distance
  - e.g., number of players arriving at server lobby in 5-minute period between noon-1pm
- Requires
  1. Probability of event same for all time units
  2. Number of events in one time unit independent of number of events in any other time unit
  3. Events occur singly (not simultaneously). In other words, as time unit gets smaller, probability of two events occurring approaches 0

Poisson Distributions?

Not Poisson
- Number of people arriving at restaurant during dinner hour
  - People frequently arrive in groups
- Number of students register for course in BannerWeb per hour on first day of registration
  - Prob not equal — most register in first few hours
  - Not independent — if too many register early, system crashes

Could Be Poisson
- Number of logins to MMO during prime time
- Number of groups arriving at restaurant during dinner hour
- Number of defects (bugs) per 100 lines of code
- People arriving at cash register (if they shop individually)

Phrase people use is random arrivals

Poisson Distribution Example

1. Number of games student plays per day averages 1 per day
2. Number of games played per day independent of all other days
3. Can only play one game at a time
   - What’s probability of playing 2 games tomorrow?
   - In this case, the value of \( \lambda = 1 \), want \( P(X=2) \)

\[
P(X=2) = e^{- \lambda} \frac{\lambda^2}{2!} = 0.1839
\]

Excel: poisson.dist()
- 1 game per day, chance for 2
  \[= \text{poisson.dist}(2, 1, \text{FALSE}) = 0.18394\]
Continuous Distributions

- Many random variables are continuous:
  - e.g., recording time (time to perform service) or measuring something (height, weight, strength)
- For continuous, doesn’t make sense to talk about \( P(X=x) \) → continuum of possible values for \( X \)
  - Mathematically, if all non-zero, total probability infinite (this violates our rule)
- So, continuous distributions have probability density, \( f(x) \)
  - How to use to calculate probabilities?
  - Don’t care about specific values
    - e.g., \( P(Height = 60.1946728163 \text{ inches}) \)
  - Instead, ask about range of values
    - e.g., \( P(59.5" < X < 60.5") \)
  - Uses calculus (integrate area under curve) (not shown here)

Normal Distribution (1 of 2)

- "Bell-shaped" or "Bell-curve"
  - Distribution from \(-\infty \) to \( +\infty \)
- Symmetric
- Mean, median, mode all same
  - Mean determines location, standard deviation determines "width"
- Super important!
  - Lots of distributions follow a normal curve
  - Basis for inferential statistics (e.g., statistical tests)
  - "Bridge" between probability and statistics
  - Aka "Gaussian" distribution

Normal Distribution (2 of 2)

- Many normal distributions (see right)
- However, "the" normal distribution refers to standard normal
  - Subtract mean (\( \mu \))
  - Divide by standard deviation (\( \sigma \))
  - \( X \) ~ \( \text{norm.dist(\mu, \sigma^2, 0.5, 0.1)} \)

Outline

- Intro (done)
- Probability (done)
- Probability Distributions
  - Discrete (done)
  - Continuous (next)

So far random variable could take only discrete set of values
Q: What does that mean?
Q: What other distributions might we consider?
Standard Normal Distribution

• Standardize
  – Subtract mean
  – Divide by standard deviation

• Mean \( \mu = 0 \)
• Standard Deviation \( \sigma = 1 \)
• Total area under curve = 1
  – Sounds like probability!

Using the Standard Normal

• Suppose League of Legends Champion released once every 24 days on average, standard deviation of 3 days
• What is the probability Champion released 30+ days?
  • \( x = 30, \mu = 24, \sigma = 3 \)

\[
Z = \frac{x - \mu}{\sigma} = \frac{30 - 24}{3} = 2
\]

• Want to know \( P(Z > 2) \)
  Use table (Z-table). Or Empirical Rule?

Test for Normality

• Why?
  – Can use Empirical Rule
  – Use some inferential statistics (parametric tests)
• How? Several ways. One:
  – Normal probability plot – graphical technique to see if data set is approximately normally distributed (next)

Normality Testing with a Histogram

• Use histogram shape to look for “bell curve”

Q: What distributions are these from? Any normal?
Normality Testing with a Histogram

They are all from normal distribution! Suffer from:
- Binning (not continuous)
- Few samples (15)

Quantile-Quantile Plot Example

• Do the following values come from a normal distribution?

7.19, 6.31, 5.89, 4.5, 3.77, 4.25, 5.19, 5.79, 6.79

1. Order data
2. Compute Z scores
3. Plot data versus Z

Quantile-Quantile Plot Example – Order Data

Unordered | Ordered (low to high)
---|---
7.19 | 3.77
6.31 | 4.25
5.89 | 4.50
4.50 | 5.19
3.77 | 5.89
4.25 | 5.79
5.19 | 6.31
5.79 | 6.79
6.79 | 7.19

Quantile-Quantile Plot Example – Compute Z scores

Divide into N+1 = 10

10% = ?
20% = ?
30% = ?
40% = ?
50% = 0
60% = ?
70% = ?
80% = ?
90% = ?

Want Z-score for that segment

Lookup in Z-table

Z-Table

• Tells what cumulative percentage of the standard normal curve is under any point (Z-score). Or, P(-∞ to Z)

(Note: Above for positive Z-scores – also negative tables, or diff from 50%)

\[ \text{NORMSINV(area)} \] – provide Z for area under standard normal curve
\[ \text{NORMSINV(.80)} = 0.841621 \]
Quantile-Quantile Plots in Excel

- Mostly, a manual process. Do as per above.
- Example of step by step process (with spreadsheet): [http://facweb.cs.depaul.edu/cmiller/it223/normQuant.html](http://facweb.cs.depaul.edu/cmiller/it223/normQuant.html)

Examples of Normality Testing with a Quantile-Quantile Plot