Overview

- Statistics important for game analysis
- Probability important for statistics
- So, understand some basic probability
- Also, probability useful for game development
- What are some examples of probabilities needed for game development?

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
- Roll die (d6) and get 6 – elementary event
- Roll die (d6) and get even number – compound event, consists of elementary events 2, 4, and 6
- Pick card from standard deck and get queen of spades – elementary event
- Pick card from standard deck and get face card – compound event
- Observe players logging into MMO and see if time between two arrivals is more than 15 seconds

Outline

- Introduction *(done)*
- Probability *(next)*
- Probability Distributions

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- Roll D6: Events: 1, 2, 3, 4, 5, 6 – exhaustive, mutually exclusive
- Roll D6: Events: get even number, get number divisible by 3, get a 1 get a 5 – exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds – exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds – exhaustive, but overlap

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

• Roll die (d6) and get 6
• Roll die (d6) and get even number
• Pick card from standard deck and get queen of spades
• Pick card from standard deck and get face card
• Observe players logging into MMO and see if time between two arrivals is more than 15 seconds

Overview

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

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

Outline

- Introduction *(done)*
- Probability *(next)*
- Probability Distributions

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- Roll D6: Events: 1, 2, 3, 4, 5, 6 – exhaustive, mutually exclusive
- Roll D6: Events: get even number, get number divisible by 3, get a 1 get a 5 – exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds – exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds – exhaustive, but overlap

Overview

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

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

Outline

- Introduction *(done)*
- Probability *(next)*
- Probability Distributions

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- Roll D6: Events: 1, 2, 3, 4, 5, 6 – exhaustive, mutually exclusive
- Roll D6: Events: get even number, get number divisible by 3, get a 1 get a 5 – exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds – exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds – exhaustive, but overlap

Overview

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

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

Outline

- Introduction *(done)*
- Probability *(next)*
- Probability Distributions

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- Roll D6: Events: 1, 2, 3, 4, 5, 6 – exhaustive, mutually exclusive
- Roll D6: Events: get even number, get number divisible by 3, get a 1 get a 5 – exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds – exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds – exhaustive, but overlap

Overview

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

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

Outline

- Introduction *(done)*
- Probability *(next)*
- Probability Distributions

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- Roll D6: Events: 1, 2, 3, 4, 5, 6 – exhaustive, mutually exclusive
- Roll D6: Events: get even number, get number divisible by 3, get a 1 get a 5 – exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds – exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds – exhaustive, but overlap

Overview

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

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

Outline

- Introduction *(done)*
- Probability *(next)*
- Probability Distributions

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- Roll D6: Events: 1, 2, 3, 4, 5, 6 – exhaustive, mutually exclusive
- Roll D6: Events: get even number, get number divisible by 3, get a 1 get a 5 – exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds – exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds – exhaustive, but overlap
Probability – Definition

- Probability – likelihood of event to occur, measured by ratio of favorable cases to unfavorable 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 ¼ to 1, ½ to 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 League of Legends in lab versus some other game. Say, 30% LoL. 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.
  \[ P(A \text{ or } B) = P(A) + P(B) \]
  - 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., 2d6: A= die 1 get 5, B= die 2 gets 6. Independent, since result of one roll doesn’t affect roll of other.
  - Probability both occur \[ P(A \text{ and } B) = P(A) \times P(B) \]
  - e.g., 2d6: prob of “snake eyes” is P(1) \times P(1) = 1/6 \times 1/6 = 1/36.
- Not independent: One occurs affects probability that other occurs.
  - Probability both occur \[ P(A \text{ and } B) = P(A) \times P(B \mid A) \]
    - Where P(B \mid 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? You might think that = P(mage) \times P(defeat B) = 0.1 \times 0.8 = 0.08.
      - But likely not independent. P(defeat B \mid mage) < 80%. So, need not-independent formula P(mage) \times P(defeat B \mid mage).

Probability Example

- Probability drawing King?
  \[ P(K) = \frac{1}{4} \]
  - Draw, put back. Now?
Probability Example

- 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?
  \[ P(K) = \frac{1}{4} \]
  \[ P(K') = 1 - P(K) = \frac{3}{4} \]
Probability Example

• Probability drawing King?
  \[ P(K) = \frac{1}{4} \]

• Draw, put back. Not 2 Kings?
  \[ P(K') \times P(K') = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \]

• 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?
  \[ P(K') \times P(K') = \frac{3}{4} \times \frac{3}{4} = \frac{9}{16} \]

• Draw, don’t put back. Not 2 Kings?
  \[ P(K') \times P(K' | K') = \frac{3}{4} \times \frac{2}{3} = \frac{6}{12} = \frac{1}{2} \]

• Draw, don’t put back. King 2nd card?
  \[ P(K' | K') = \frac{1}{3} \]

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

<table>
<thead>
<tr>
<th></th>
<th>[ a + b ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>[ a + b ]</td>
</tr>
<tr>
<td>Median</td>
<td>[ a + b ]</td>
</tr>
<tr>
<td>Mode</td>
<td>N/A</td>
</tr>
<tr>
<td>Variance</td>
<td>[ (b - a + 1)^2 - 1 ]</td>
</tr>
<tr>
<td>Std Dev</td>
<td>[ \sqrt{\text{Variance}} = 1.7 ]</td>
</tr>
</tbody>
</table>
Binomial Distribution Example (1 of 3)

How to assign probabilities?

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

\[ P(X=2) = ? \]

Binomial Distribution Example (2 of 3)

All equally likely \((p = 1/8 \text{ for each})\)
\[ P(HHT) + P(HTH) + P(THH) = 3/8 \]

Can draw histogram of number of heads

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

Excel: \texttt{binom.dist}()

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

Binomial Distribution Example (3 of 3)

How to assign probabilities?

• Could measure (empirical)
  – Q: how?
• Could use “hunch” (subjective)
  – Q: what do you think?
• Could use theory (classical)
  – Math using our probability rules (not shown)
  – Enumerate (next)
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 units
  2. Number of events in one unit independent of number of events in any other unit
  3. Events occur singly (not simultaneously). In other words, as unit gets smaller, probability of two events occurring approaches 0

Poisson Distribution Example

- Number of games student plays per day averages one per day
- Number of games played per day independent of all other days
- Can only play 1 game at a time
- What’s probability of playing two games next day?
- In this case, the value of $\lambda = 1$

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

Outline

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

So far random variable could take only discrete set of values

Q: What does that mean?
Q: What other distributions might we consider?
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 $P(X=x) \rightarrow$ 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., $P(\text{Height} = 60.1946728163 \text{ inches})$
  • Instead, ask about range of values
    – e.g., $P(59.5 < X < 60.5)$
• Uses calculus (integrate area under curve) (not shown here)

What continuous distribution is especially important?

The Normal Distribution

• “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 normal (“bell curve”)
  – Basis for inferential statistics (e.g., statistical tests)
  – “Bridge” between probability and statistics
• Aka “Gaussian” distribution

A Normal Distribution

• Many normal distributions
• However, “the” normal distribution refers to standard normal
  – Subtract mean ($\mu$)
  – Divide by standard deviation ($\sigma$)

$Z = \frac{X - \mu}{\sigma}$

Use to predict how likely an observed sample is given a population mean.
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?
  - Use some inferential statistics (parametric tests)
  - Can use Empirical Rule
- How? Several ways. One:
  - Normal probability plot – graphical technique to see if data set is approximately normally distributed

Normality Testing with a Histogram

- Use histogram shape to look for “bell curve”

Q: What distributions are these from? Any normal?

Yes

No

They are all from normal distribution! Suffer from:
- Binning (not continuous)
- Few samples (15)
Normality Testing with a Quantile-Quantile Plot

- Quantiles of one versus another
- If line \( \rightarrow \) same distribution

1. Order data
2. Compute Z scores (normal)
3. Plot data (y-axis) versus Z (x-axis)
   - Normal? \( \rightarrow \) line

Quantile-Quantile Plot Example – Order Data

- Unordered
  - 7.19
  - 6.31
  - 5.89
  - 4.50
  - 3.77
  - 4.25
  - 5.19
  - 5.79
  - 6.79

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

N = 9 data points

Quantile-Quantile Plot Example – Compute Z scores

- Divide into \( N+1 = 10 \)
  - 10% = 0.84
  - 20% = 0.52
  - 30% = 0.25
  - 40% = 0
  - 50% = 0
  - 60% = 0.25
  - 70% = 0.52
  - 80% = 0.84
  - 90% = 1.28

Z-Table

- Tells what cumulative percentage of the standard normal curve is under any point (Z-score). Or, \( P(-\infty \rightarrow Z) \)
- \( \text{NORMSINV(area)} \) – provide Z for area under standard normal curve

(Only some points shown)
Quantile-Quantile Plot Example – Plot

Linear? → Normal

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

Examples of Normality Testing with a Quantile-Quantile Plot