Probability & Functions

Matt Mettler

University of Illinois Urbana Champaign

mettler3@illinois.edu

Math Camp

August 8, 2025

Set Theory

Definition

Sets are a collection of elements

Sets are noted by a capitol letter and elements of a set are indicated by lower case letters

Example

- let X be a set of all real numbers x where $0 \leq x \leq 1$, then we know $\frac{1}{2}$ is an element of X and can write that as $\frac{1}{2} \in X$
- Let Z be a set of whole numbers $Z = \{0, 1, 2, 3, 4, ...n\}$
- Let C be a set of Star Trek Captains C={Picard, Janeway, Sisko, Kirk, ..., Pike}

Set Theory 2

- Sets are countable if C is finite, or corresponds to a one to one function that maps onto the set of natural numbers i.e. has as many elements as there are positive integers, for example $X = \{1, 2, 3...100, ...N\}$
- Sets are countably infinite if the set is not finite.
- Suppose we have a set C of natural numbers between 5 and 12. $C = \{5, 6, 7, 8, 9, 10, 11, 12\}$ – the cardinality (number of elements is 8) – Countable set
- Suppose we have a set C of all natural numbers. $C = \{1, 2, 3, ...\mathbb{N}\}$ –infinitely countable set

Set Theory 3

Experiments and outcomes as sets

- Suppose we have an experiment where the outcomes are known beforehand and the conditions are repeatable. We call this a random experiment and all possible outcomes are the sample space.
- The sample space of outcomes is a set. (Sets are noted as capital letters and sample space is often noted as Ω I stick to capital letters in these slides).
- For instance, think of a coin toss where our outcomes are H (heads) and T (tails). Thus our set outcomes C are $C = \{H, T\}$.
- Or a roll of a six sided die $C = \{1, 2, 3, 4, 5, 6\}$.

- We can also think of more complex sample spaces, such as ordered pairs of outcomes. Suppose we roll a red die and a white die. We can, presumably, repeatedly roll these die under the same conditions. The sample space is the ordered outcomes of the two die. $X = \{(1,1), ..., (1,6), (2,1), ..., (6,6)\}$

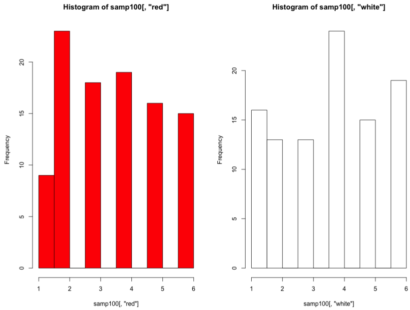


Figure: R-output

- Can we code this in R?

Set Theory 5.1

Rolling the dice (R Answer)

```
<<chunkfig, include=T, echo=F, fig=T>>=
# we can create a random experiment of die in r
roll<-function(n){
x<-1:6
samp_space<-matrix(ncol=2,nrow=n,dimnames = list(c(),c("red","white")))
for(i in 1:n){

red<-sample(x,1,replace=TRUE)
white<-sample(x,1,replace=TRUE)

samp_space[i,1]<-red
samp_space[i,2]<-white
}
return(samp_space)
}

samp100<-roll(100)
par(mfrow = c(1,2))
hist(samp100[,"red"],col="red")
hist(samp100[,"white"],col="white")
```

- Empty sets are denoted by $\emptyset$
- The complement of an even A is the set of all elements in C which are not in A. complements are denoted as A^c also, $A^c = \{x \in C : x \notin A\}$
- If each element of set A is also an element of set B, then A is a subset of B, $A \subset B$, if $B \subset A$ then we can say that $A = B$
- The union of A and B is the set of all elements that are in A or in B or in both, $A \cup B$
- The intersection of A and B is the set of all elements that are in both A and B, $A \cap B$
- Two events (subsets) are disjointed if they do not have elements in common $A \cap B = \emptyset$ also called a disjointed set.

- Let's say we have a 20 sided die, and the experiment is to roll the die and record the number facing up after the die stops. The sample space is $C = \{1, 2, 3, 4, 5, ...20\}$
- Lets say we have event Q as the subset where the roll is $W = \{19\}$, $X = \{1, 5, 10\}$, $Y = \{5, 12, 14, 20\}$ and $Z = \{8, 12, 20\}$.

Example

- $X^c = \{2, 3, 4, 6, 7, 8, 9, 11, ..., 20\}$
- $X \cup Y = \{1, 5, 10, 12, 14, 20\}$
- $X \cap Y = \{5\}$
- $W \cap Z = \emptyset$
- $(X \cup Y) \cap (X \cup Z) = \{1, 5, 10, 12, 14, 20\} \cap \{1, 5, 8, 10, 12, 20\} = \{1, 5, 10, 12, 20\}$

Function

A function is a relationship among sets where the input (domain) has exactly one output in the co-domain and can be traced back to its input.

Example

- $y = f(x)$ function $f(x)$ that maps x to y
- $f : X \to Y$

Probability

is a set function P that assigns to each event A in event space C a number $P(A)$ called the probability of the event A. More formally, $P: C \to [0,1]$

Axioms of probability set functions

- let C be the sample space and B the set of events
  1. $P(A) =\geq 0$, for all $A \in B$
  2. $P(C) = 1$
  3. if $\{A_n\}$ is a sequence of events in B and $A_m \cap A_n = \emptyset$ for all $m \neq n$, then:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Probability 2

Relative frequency

- We can work through the intuitions of relative frequency to see the first two axioms.

```
<<chunkfig , include=T, echo=F, fig=T>>=

samp1k<-as.data.frame(roll(1000))

A<-samp1k[which(samp1k$red==2),"red"]# lets look at event of red die comping up with a
2

#relative frequency is count of occurrences of event A in N repetitions

length(A)/1000 # the relative frequency

# lets redo this a bunch and see what it looks like
ntrials<-1:100
for(i in 1:length(ntrials)){
samp1k<-as.data.frame(roll(1000))

ntrials[i]<-length(samp1k[which(samp1k$red==2),"red"])/1000
}

hist(ntrials)
abline(v=1/6,col="red")
```

Chunk 2: chunkfig R Sweave

Additional properties of probability

- For each event $Z \in B$, $P(Z) = 1 - P(Z^c)$
- Probability of the empty set is zero, $P(\emptyset) = 0$
- Probability of $P(\emptyset)^c = 1$
- If X and Y are events in B, then:
  $P(X \cup Y) = P(A) + P(B) - P(A \cap B)$ This works if the events are disjointed or not.

Rolling a die

- Our sample space of our die is $C = \{1, 2, 3, 4, 5, 6\}$
- Our event space is $B = \{(\emptyset), (1), (2), (3), (4), (5), (6)\}$
- The intersections of the sets are empty so we say these are pairwise disjoint, meaning they are a mutually exclusive collection. This is also said to be exhaustive as the union of events is equal to the sample space.
- let X and Y be sets such $X = 2$ and $Y = 4$
- The probability that the roll is either X or Y is $P(X \cup Y) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - 0 = \frac{1}{3}$
- The

- What if we knew the roll was even W? what is the probability of X now? Y? union of X and Y?

Conditional probability

- The conditional probability of event A, given event B, is: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Returning to our die toss

- Using conditional probability, what is the probability of X given W?

Independence

- If the occurrence or non-occurrence of event X has no effect on the occurrence or non-occurance Y they Y is independent of X.
- $P(X|Y) = P(X)$
- $P(Y|X) = P(Y)$
- Let X and Y be two events. X and Y are independent if $P(X \cap Y) = P(X)P(Y)$
- For instance, let $X = \{1, 2, 3\}$ and $X = \{3, 4\}$ for our die toss.
- The probability of $X = \frac{1}{2}$ and $Y = \frac{1}{3}$. The intersection of X and Y is 3 which has a probability of $\frac{1}{6}$ and $P(X)P(Y) = \frac{1}{6}$

Random Variable

- A function X, that maps outcomes of the sample space B onto a measurable space E. $X : B \to E$
- Generally, and almost always for our purposes, the function X maps B onto the set of real numbers $\mathbb{R}$. Thus, $X : B \to \mathbb{R}$

- Returning to our die, we are assigning a real number to the six faces of the die.
- a coin toss can be assigned (-1,1) for heads and tails or (1,0) for heads and tails.

Discrete Random Variable

- Takes on a countable number of distinct values, event space is either finite or countable (e.g. head vs tails).
- Mass function is the function that that maps the probability of a discrete random variable being equal to a certain value. Again these are non-negative and sum to 1, given as:

$$p_X(x) = P[X = x], x \in D$$

- lets roll a fair die with number 1-6 on the faces.
- X is the face that is up for each roll of the die. The event space for X is $\{$1,2,3,4,5,6 $\}$ and
- The pmf is $p_X(i) = \frac{1}{6}$, for for i = 1,2,3,4,5,6. A outcome of 0 has $F_X = 0$ and an outcome of $1 \leq x \leq 2$, then $F_X = \frac{1}{3}$.
- An outcome of 0 has probability 0 and is outside of what we call the support.

Probability 10

Random variables

- $$p_X(x) = P(X = x) = \begin{cases} \frac{1}{6} & x = 1 \\ \frac{1}{6} & x = 2 \\ \frac{1}{6} & x = 3 \\ \frac{1}{6} & x = 4 \\ \frac{1}{6} & x = 5 \\ \frac{1}{6} & x = 6 \\ 0 & otherwise \end{cases}$$
- The support is the points in the space X which have positive probability.
- Let's toss a coin again with head=1 and tails =0. What is the pmf? What is the support?

Probability 11

Random variables

- $$p_X(x) = P(X = x) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2} & x = 1 \\ 0 & otherwise \end{cases}$$
- Support is $R_X = \{0, 1\}$

- Let's work out the pmf of ten tosses of a fair coin to find the probability that 8 of ten tosses are heads.
- Let's consider we have k success, n independent trials, with probability p.
- what is our k, our n, and our p?

- Before we find the pmf, let's discuss combinations and permutations.
- Permutations are k items taken from a set with n elements where order matters.
- $$^nP_k = \frac{n!}{(n-k)!}$$
- Combinations are the subsets of k items taken from a set with n elements.
- $$\binom{n}{k} = {}^nC_k = \frac{n!}{k!(n-k)!}$$

Probability 14

Random variables

- $$p_X(x) = f(k, n, p) = P(X = x) = \binom{n}{k} p^k (1-p)^n - k$$

- Returning to our problem:

$$p_X(x) = f(8, 10, .5) = P(X = x) = \binom{10}{8} .5^8 (1 - .5)^{10-8}$$

$$= \frac{10!}{8!(10-8)!} .5^8 (1 - .5)^{n-k}$$

$$= 0.043$$

- Let's try that in Rstudio

```
n<-10
k<-8
p<-.5
n_f<-factorial(n)
k_f<-factorial(k)
n_k_f<-factorial(n-k)

(n_f/(k_f*n_k_f)) *(p^k)* (1-p)^(n-k)
```

Probability 18

Random variables

- Even easier

```
> dbinom(8,10,.5)
[1] 0.04394531
> |
```

Probability 19

Random variables

- Even easier

```
> dbinom(8,10,.5)
[1] 0.04394531
>
```

- What if we wanted to know the probability of 8 or fewer heads out of ten fair flips?
- For that we would need to find the cumulative density function (cdf)
- $$p_X(x) = f(k, n, p) = P(X \leq x) = \sum_{i=o}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

```
> pbinom(8,10,.5)
[1] 0.9892578
> |
```

- **Continuous random variable** - if its cumulative distribution function is $F_X x$ is a continuous function for all $x \in R$, i.e. there are not points of discrete mass.
- Unlike a discrete random variable, the probability that X is a certain value is 0.
- Since there are uncountably many ”points” in D for a continuous random variable i.e. there are not discrete points of mass.
- Thus, $P(X = x)$ is not going to be usable for a continuous random variable.
- Instead, we must find the probability that X falls into an interval (a,b), meaning $P(a \leq X \leq b) = F_X(b) - F_X(a)$. Thus for any two numbers a and b such that $a \leq b$,we have
- $P(a < X \leq b) = \int_a^b f(x)dx$. The function $f_X(x)$ is the probability density function (pdf) of X.

Example 1.7.2. Let the random variable be the time in seconds between incoming telephone calls at a busy switchboard. Suppose that a reasonable probability model for X is given by the pdf

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-x/4} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

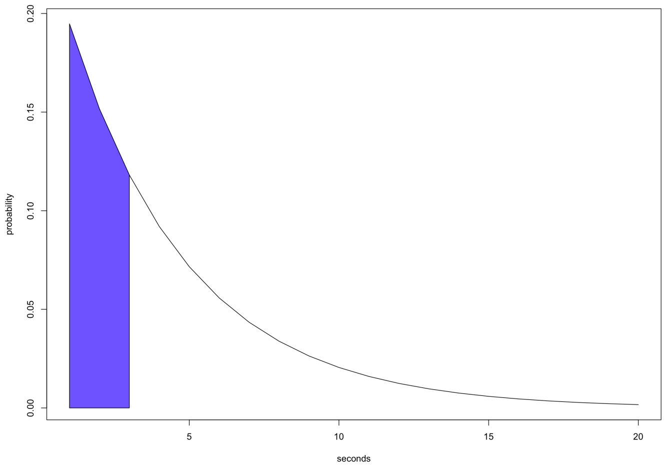
Note that f_X satisfies the two properties of a pdf, namely, (i) $f(x) \geq 0$ and (ii)

$$\int_0^\infty \frac{1}{4}e^{-x/4}\,dx = -e^{-x/4}\Big|_0^\infty = 1.$$

For illustration, the probability that the time between successive phone calls exceeds 4 seconds is given by

$$P(X > 4) = \int_4^\infty \frac{1}{4}e^{-x/4}\,dx = e^{-1} = 0.3679.$$

- What about 0-4 seconds?



- The most recognized distribution from a continuous random variable is the normal distribution with pdf:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- where u is the mean and σ is the standard deviation
- when we take a random sample from a population that is $N(\mu, \sigma)$, we will approximate the distribution of the population.

- Distribution: Describes how the values of a random variable are spread.
- **Central Limit Theorem:** tells us that, when we take many random samples from any population, the sampling distribution of the sample mean will approximate a normal distribution, regardless of the populations original distribution.

For example, a fair coin toss follows a...

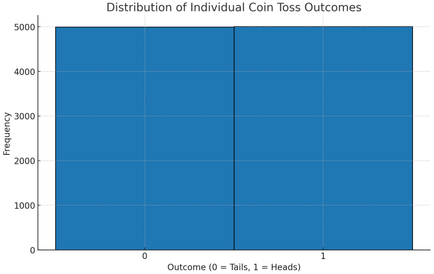


Figure: Distribution of individual coin toss outcomes.

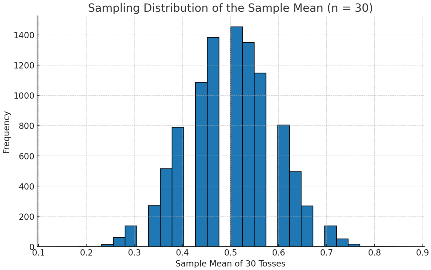


Figure: Sampling distribution of the sample mean (n=30).

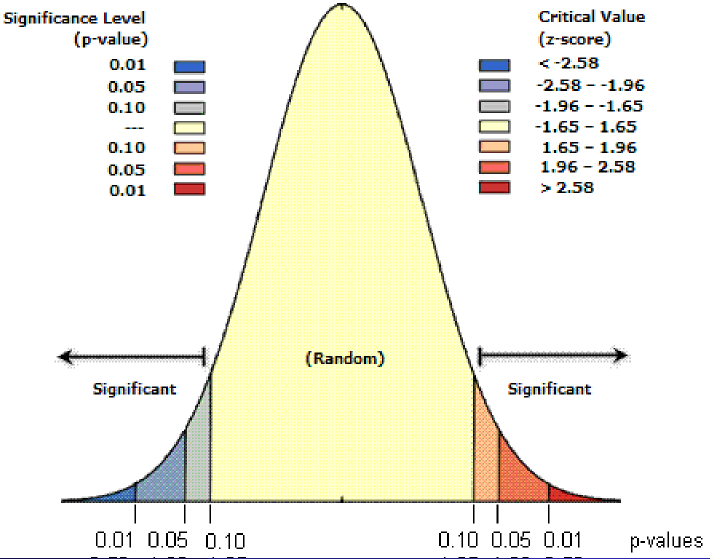
- For example, In my younger days I worked at a factory that manufactured dice. We were world famous for our six-sided die.
- however one day the machine made a weird clunking noise and the die that came out felt weird.

- The die are biased and the underlying pmf is:
- $$p_X(x) = P(X = x) = \begin{cases} \frac{1}{5} & x = 1 \\ \frac{1}{5} & x = 2 \\ \frac{1}{4} & x = 3 \\ \frac{1}{6} & x = 4 \\ \frac{1}{8} & x = 5 \\ \frac{1}{7} & x = 6 \\ 0 & otherwise \end{cases}$$
- but I didn't know this, so what can I do to see if the die are biased?

Probability 30

Distributions and the Central Limit Theorem

- We know the population mean $u = 3.5$ and sd
- $sd = \sigma = \sqrt{VAR} = E[X^2] - (E[X])^2 = \sqrt{2.9167} = 1.7078$
- we can compute the sample mean $\bar{x}$ and sd s
- we can compute standard error of our samples $SE = \frac{\sigma}{\sqrt{(n)}}$
- and Z-scores $Z = \frac{\bar{x}-u}{\frac{\sigma}{\sqrt{n}}}$


Significance Level
(p-value)
0.01
0.05
0.10

0.10
0.05
0.01
Critical Value
(z-score)
< -2.58
-2.58 – -1.96
-1.96 – -1.65
-1.65 – 1.65
1.65 – 1.96
1.96 – 2.58
> 2.58
(Random)
Significant
Significant
0.01 0.05 0.10
0.10 0.05 0.01
p-values

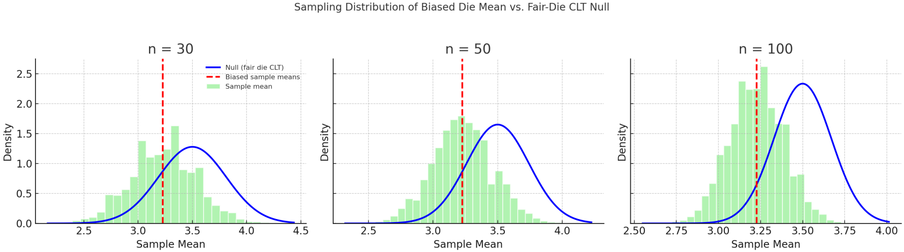


Figure: checking biased die at different sample sizes