Week 1

- experiment: process where we obtain data from
  - can control factors that influent the outcome
  - outcome not deterministic
  - specify all probable outcome in advance.
- sample space: collection of all probable outcomes
- event in sample space: collection of some of the outcomes in a sample space
- Simple sample space: SP with n outcomes and each outcome has equal probability
- law of large numbers: more experiments performed, observed P will trend to close to its theoretical P. $\frac{\text{\# time of observing event}}{n} \to P$ as $n \uparrow$

Week 2 Conditional Probability

- Probability in simple sample space (SSS) $P = $ # outcomes in the event / # outcomes in SS

  multiplicative counting method

  total count = product of the counts of individual steps
- combination $C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ ← choose(n, k). $P_n^k = \frac{n!}{(n-k)!}$ → prod(n, n-k+1)
- conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$ ← 分母是原来的SS

  $P(A|B)$ usually $\neq P(B|A)$
- independent events: 另一件是否发生不影响此事件 $P(A|B) = P(A)$

  $P(A|B) = \frac{P(A\&B)}{P(B)}$ if $= P(A) \Rightarrow P(A\&B) = P(A)P(B)$

Week 3 PDF, CDF

RV 经运算得到的也是RV

- Random variable (RV): function on SS. numerical value based on experiment outcome
- collection of P for all possible values of an RV: distribution of RV

  sum of P in any distribution is 1.0
- probability mass function (p.m.f) $f(x) = P(X = x)$ $\in [0, 1]$
- cumulative distribution function (c.d.f) $F(X) = P(X \le x)$ ↗ ; → 1 when $x \to \infty$
- Bernoulli distribution: experiment where outcomes can be categorized into two categories: success and failure $f(x) = \begin{cases} p, & x=1 \\ 1-p, & x=0 \\ 0, & \text{otherwise} \end{cases}$
- Binomial distribution: experiment of n independent Bernoulli trials

  RV X that counts the total number of success will follow a Binomial distribution

  $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, \cdots, n$ → dbinom(x, n, p)
- Continuous RV have uncountably many different values. ⇒ $P(Y=y) = 0$. for any y

  $P(a \le X \le b) = \int_a^b f(x)\,dx$, $f(x)$ is p.d.f of x probability density function

  $F(b) = P(X \le b) = \int_{-\infty}^b f(x)\,dx$. $F(b)$ c.d.f cumulative distribution function

  * $f(x)$ non negative. $F(x)$ non decreasing, → 1 when $x \to \infty$.

  randomly selected student ⇒ uniform distribution is used in the process

Week 4 E(X), Var(X)

- Expectation of RV is the weighted avg of all possible values by their P

  X discrete RV with p.m.f $f(x)$ $E(X) = \sum_{\text{all values of } x} x f(x)$

  continuous RV with p.d.f $f(x)$ $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$
- if X is RV, function of X is another RV. $\begin{cases} \text{discrete}: & E(g(X)) = \sum_{\text{all values of } x} g(x) f(x) \\ \text{continuous}: & E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx \end{cases}$
- if $g(X) = aX + b$, then $E(g(X)) = E(aX+b) = aE(X) + b$
- $X_1, X_2, X_3, \cdots, X_n$ be RV

  $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$

  * $E(aX_1 + bX_2 + c) = aE(X_1) + bE(X_2) + c$

  $E(X_1 X_2) = E(X_1)E(X_2)$ iff X_1 and X_2 are statistically independent
- $Var(X) = E[(X-\mu)^2]$ 写作 σ_x^2 standard deviation $\sigma_x = \sqrt{Var(X)}$

  * $Var(X) \ge 0$; $Var(aX + c) = a^2 Var(X)$

  $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$ iff X_1, X_2 independent

  ☆ $Var(aX_1 + bX_2) = a^2 Var(X_1) + b^2 Var(X_2) + 2ab\, Cov(X, Y)$

  $Var(aX_1 + bX_2) = a^2 Var(X_1) + b^2 Var(X_2) + 2ab(\rho_{X_1 X_2} \sigma_{X_1} \sigma_{X_2})$

Week 5 Cov(X,Y), ρ_{xy}

- bivariate distribution: distribution of a pair of RV

  When (X, Y) are independent RV, bivariate distribution $f(x, y) = f(x) f(y)$
- distribution of single RV X or Y are called the marginal distributions, in relation to their joint distribution for (X, Y) > sum(car$probabilities[car$X ==1])
- Covariance: $Cov(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$

  $\begin{cases} cov(X,Y) > 0 & \text{positive correlation} & \text{大 x 更有可能得大 y} \\ cov(X,Y) < 0 & \text{negative correlation} & \text{小 y} \\ cov(X,Y) = 0 & \text{uncorrelated} & \text{一个波动时另一个平稳} \end{cases}$

  X, Y independent ⇒ must be uncorrelated

$cov(X, X) = E[(X-\mu_X)^2] = Var(X) \ge 0$.

correlation is not causality!

> cov( ) , cor( )

- cov 问题: value depends on unit

  correlation coefficient $\rho_{xy} = \frac{cov(X,Y)}{\sigma_x \sigma_y} \in [-1, 1]$

  $|\rho_{XY}|$ closer to 1, more pronounced linear relationship between X and Y

  $cov(X,Y) = E[XY] - E[X]E[Y]$
- Conditional expectation: $E(X|Y=y) = \sum_{\text{all values of } x} x f(X=x|Y=y)$

  $E(g(X)|Y=y) = \sum_{\text{all values of } x} g(x) f(X=x|Y=y)$

$E(X|Y=y) = \int_{-\infty}^{\infty} x f(x|Y=y)\,dx$, where $f(x|Y=y) = \frac{f(X, Y=y)}{f(Y=y)}$

Week 2 Permutation test

- Pool all m+n observations into a vector (m for treatment & n for control)

REPEAT
- Generate the first sample of size m without replacement.
- Use the remaining n observations as the second sample.
- Calculate the difference in means or another statistic between two samples.

until we reach a large sample size for the statistic

```
# Step 1: the pooled data vector
data = c(2, 1, 6, 5)
obs_difference=mean(c(6,5))-mean(c(2,1))
# Step 2
permutation=c()
for (i in 1:10^5) {
  S1=sample(data,2)
  permutation[i]=mean(S1)-(sum(data)-sum(S1))/2
}
sum(permutation>obs_difference)/10^5
hist(permutation)
```

Calculate the fraction of times the generated statistics exceed the observed statistic (multiply by 2 for a two-sided test).

assumption: avg viral load is same across both groups.

effect of direct gaze is no different from that of averted gaze

Week 3 Inverse Transform Method

The inverse transform method

Find the inverse CDF of the distribution (from (0, 1) → X)

REPEAT

We "spin the wheel" ⟺ generate $U \sim U(0,1)$ Apply U to Inverse CDF

If $U \in [0, .50]$, let $X = 1$ "strawberry"

If $U \in [.50, .60]$, let $X = 2$ "mango"

If $U \in [.60, 1.0]$, let $X = 3$ "vanilla"

$F^{-1}(q) =$ "strawberry" for $0 \le q < 0.5$
$=$ "Mango" for $0.5 \le q < 0.6$
$=$ "Vanilla" for $0.6 \le q \le 1$

- Consider a target RV X with distribution function $F(x)$
- Derive the inverse CDF $F^{-1}(q)$

REPEAT
- Generate $U \sim U(0,1)$. Denote the realized value of U as q.
- Then obtain X = $F^{-1}(q)$.

Until enough number of X has been obtained

```
# ice-cream example: # P(X=1)=0.5, P(X=2)=0.1, P(x=3)=0.4
n=1000
#create a placeholder for n obs as the output
output=c(0)
for(i in 1:n){
  spin=runif(1)
  if(spin<0.5){
    output[i]=1
  }
  else if(spin<0.6){
    output[i]=2
  }
  else {
    output[i]=3
  }
}
prop.table(table(output))
```

```
# PS: the results can be obtained by:
sample(c(1,2,3),n,replace=TRUE,prob=c(.5,.1,.4))
or
sample(c("Strawberry","Mango","Vanilla"),n,replace=TRUE,prob=c(.5,.1,.4))
```

- 多元分布没有 inverse distribution function
- Inverse CDF $F^{-1}(q)$ difficult to derive

Week 4

- Sample Moments: measure key distributional aspects of a variable

  k-th moment of an RV X: $\mu_k' = E(X^k)$

  k-th central moment: $\mu_k = E((X - EX)^k)$

- 1^{st} (central) moments $\mu_1' = E(X)$ mean $\mu_1 \equiv 0$ ← $E(X - E(X)) = E(X) - E(X) = 0$.
- 2^{nd} central moments $\mu_2 = E((X-EX)^2)$ ← variance

  compare variation of variables, use Coefficients of Variation (CV) ⟶ sample variance / standard deviation $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)$ $CV = \frac{\text{sample std}}{\text{sample mean}}$
- 3^{rd} central moments $\mu_3 = E((X-EX)^3)$

  只要f(x)有对称性 ⇒ $\mu_3 = 0$.

  偏度

  * skewness coefficient $= E(\frac{X-\mu}{\sigma})^3$: measure how unsymmetrical (or skewed) a distribution is

  - *A positive value implies there is a longer right tail (right skewed)*
  - *A negative value there is a longer left tail (left skewed)*

  General cut-off values:

  Significantly left skewed -1 *Moderately left skewed* -0.5 *Not skewed* 0.5 *Moderately right skewed* 1 *Significantly right skewed*

  0

  range $[-\infty, +\infty]$

  library(moments)
  skewness(…)
  kurtosis(…)
- 4^{th} central moment $\mu_4 = E(X-EX)^4$

  kurtosis $= E(\frac{X-\mu}{\sigma})^4$ 峰度

  3为临界 $\begin{cases} >3 & \text{tail heavy} \\ <3 & \text{tail thin} \end{cases}$

Week 5

"D" is a binary random variable (cause), indicating whether an individual is exposed to a treatment.

"Y" is the outcome (random) variable for an individual.

-Treatment group: Those who are exposed to the treatment. $D=1$
-Control group: Those who are NOT exposed to the treatment. $D=0$

Causal effect (or treatment effect) = the change in the average of Y as a result of D

$Y = Y^0$ if $D=0$
$\quad = Y^1$ if $D=1$. $\quad Y_0$ and Y_1 potential outcomes.

Only two potential outcomes (Y^0 or Y^1) is observable; unobservable outcome: counterfactual outcome

average treatment effect of population (ATE) : $E(\delta) = E(Y^1 - Y^0)$ 针对all individuals

average treatment effect of treated (ATT) : $E(\delta \mid D=1)$

average treatment effect of controlled (ATC) : $E(\delta \mid D=0)$

$$\underbrace{E(\delta)}_{ATE} = P(D=1) \times \underbrace{E(\delta \mid D=1)}_{ATT} + P(D=0) \times \underbrace{E(\delta \mid D=0)}_{ATC}$$

	"Treated world"	"Untreated world"	
Ppl in the treated group ($D=1$)	(a) $E(Y^1 \mid D=1)$	(c) $E(Y^0 \mid D=1)$	$E(\delta \mid D=1) =$ (a)-(c)
Ppl in the control group ($D=0$)	(b) $E(Y^1 \mid D=0)$	(d) $E(Y^0 \mid D=0)$	$E(\delta \mid D=0)$ = (b) - (d)

Naive estimate for the treatment effect $= E(Y^1 \mid D=1) - E(Y^0 \mid D=0)$

↳ NE doesn't work in observation studies because studies in treatment is not randomly assigned.

Differential Treatment Effect Bias (DTEB) when $E(\delta \mid D=1) \neq E(\delta \mid D=0)$? 恒大于0吗 Bias可小于0.

当 baseline bias = 0 时 DTEB = NE - ATE.

- Suppose the ATT is +\$3000 and the ATC is +\$1500
- Suppose the Y^0 for both group is \$1000
- 10% of the population are treated (Pr(D=1)=0.1).
- What is the ATE and the naïve estimate (based on an infinitely large sample)?

The ATE= $E(\delta)$ = 0.1*3000+0.9*1500=\$1650
The naïve estimate = $E(Y^1 \mid D=1) - E(Y^0 \mid D=0)$ = (\$1000+\$3000) – (\$1000) = \$3000 (=ATE) ?

$E(Y^1 \mid D=1) \quad E(Y^0 \mid D=0)$

Bias = \$ 3000- 1650 = \$1,350

MSBA accepted rate 30%

	$E(Y^1 \mid D)$	$E(Y^0 \mid D)$
Treatment group (D=1)	\$10k	\$6k
Control group (D=0)	\$8k	\$5k

ATT = 4k $\quad$ ATC = 3k

ATE = 30% · 4k + 70% · 3k = 3.3k

NE = 5k $\quad$ Baseline Bias = 1k

Total Bias = 5 - 3.3 = 1.7k $\quad \therefore$ DTEB = TB - BB = 0.7k

- Randomized Design $\quad$ DTEB = 0, Baseline Bias = 0. $\Rightarrow$ NE = ATE.

- Independence: $P(A \cap B) = P(A)P(B) \Rightarrow P(A \mid B) = P(A)$, $P(B \mid A) = P(B)$
- Total probability: $P(B) = P(B \cap S) = P(B \cap (A_1 \cup A_2 \cup \cdots \cup A_n))$
$$= P(B \cap A_1) + P(B \cap A_2) \cdots + P(B \cap A_n)$$
$$= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \cdots + P(B \mid A_n)P(A_n)$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \bar{B})P(\bar{B})}$$

A, B 独立 ⇒ A, $\bar{A}$, B, $\bar{B}$ 独立.

$$Var(X) = E(X^2) - [E(X)]^2$$

Name	Density		Moment generating function	Mean	Variance
Geometric	$(1-p)^{x-1}p$	$x = 1, 2, 3, \ldots$; $0 < p < 1$	$\frac{pe^t}{1-qe^t}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Uniform	$\frac{1}{n}$	$x = x_1, x_2, \ldots, x_n$; n a positive integer	$\frac{\sum_{i=1}^n e^{tx_i}}{n}$	$\frac{\sum_{i=1}^n x_i}{n}$	$\frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$x = 0, 1, 2, \ldots, n$; $0 < p < 1$; n a positive integer	$(q + pe^t)^n$	np	$np(1-p)$
Bernoulli or point binomial	$p^x(1-p)^{1-x}$	$x = 0, 1$; $0 < p < 1$	$q + pe^t$	p	$p(1-p)$
Hypergeometric	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\max[0, n-(N-r)] \le x \le \min(n, r)$		$n\frac{r}{N}$	$n\frac{r}{N}\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$
~~Negative binomial~~	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$x = r, r+1, r+2, \ldots$; $0 < p < 1$	$\frac{(pe^t)^r}{(1-qe^t)^r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Poisson	$\frac{e^{-k}k^x}{x!}$	$x = 0, 1, 2, \ldots$; $k > 0$	$e^{k(e^t-1)}$	k	k

Joint Distribution

- Discrete joint density: Let X, Y be discrete random variables. The order pair (X, Y) is called a two-dimentional discrete random variable.

$f_{XY}(x,y) = P[X=x \text{ and } Y=y]$ joint density for (X, Y)

- Marginal Distribution : Discrete

Discrete marginal densities: Let (X, Y) be a 2-D discrete R.V. with joint density f_{XY}. The marginal density for $\begin{cases} X \text{ is } f_X(x) = \sum_{\text{all } y} f_{XY}(x,y). \\ Y \text{ is } f_Y(y) = \sum_{\text{all } x} f_{XY}(x,y). \end{cases}$

- Joint and marginal distributions: Continuous.

$P[a \le X \le b \text{ and } c \le Y \le d] = \int_a^b \int_c^d f_{XY}(a,b)\, dy\, dx.$

$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy. \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx.$

- dnorm(x, mean = 0, sd = 1) ⇒ the pdf f(x)
- pnorm(q, mean = 0, sd = 1) ⇒ the cdf F(q)
- qnorm(p, mean = 0, sd = 1) ⇒ returns the value of x such that F(x)=p
- rnorm(n, mean = 0, sd = 1) ⇒ draw a random sample (size=n)

- $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^4 x(x/8) dx$

```
## Define the function to be integrated
>fx <- function(x) {x*x/8}
## integrate thefunction from 0 to 4
>integrate(fx, lower = 0, upper = 4)
2.666667 with absolute error < 1.1e-14
```

R做微积分

- punif(x, min=a, max=b), cdf of unif dist
  runif(n, min=0, max=1)
- dbinom(x, size, prob) pbinom(q, size, prob) rbinom(n, size, prob)

Example: automobile theft (contd.)

Suppose the insurance company charges the Brand-2 car owner \$200 premium, under the condition that the company will pay \$5000 if the car is stolen

What is the expected profit of this insurance deal?

	Brand (Y)				
	1	2	3	4	5
No (X=0)	0.928	0.968	0.994	0.993	0.991
Yes (X=1)	0.072	0.032	0.006	0.007	0.009

\$40?

$\$200 - E(g(X) \mid Y=2)$ $\quad$ *g: the payout of the company* $\quad$ $P(X=0)$
$= \$200 - (\$5000 * \Pr(X=1 \mid Y=2) + \$0 * \Pr(X=0 \mid Y=2))$
$= \$200 - \$5000 * 0.032 = \$37.66$

```
Q1                   20 obs. of 3 variables
    $ x            : int  0 0 0 0 0 1 1 1 1 1 ...
    $ y            : int  0 1 2 3 4 0 1 2 3 4 ...
    $ probabilities: num  0.08 0.07 0.06 0.01 0.01 0.06 0.11 0…
```

```
Q1=read.csv('/Users/apple/Desktop/NTU_Tri1/AN6005/PPT\'s/Week\ 5-\ l,header=T
ans1=sum(Q1$probabilities[Q1$x==2])
ans2=sum(Q1$probabilities[Q1$y>=3])
ans3=sum(Q1$probabilities[(Q1$x<=2) & (Q1$y<=2)])
ans4=sum(Q1$probabilities[Q1$x==Q1$y])
ans5=sum(Q1$probabilities[Q1$x>Q1$y])
```

Q3: f(x) = dnorm(x,0,1) if $x > 0$,
f(x)=0.5 if $-1 \le x \le 0$,
f(x)=0 elsewhere.

```
n=1000
#create a placeholder for n obs as the output
output=c()

for(i in 1:n){
  spin=runif(1)
  # for F(x)>0.5, we know the X>0 and should be taken from a std normal dist.
  if(spin>0.5){
  output[i]=qnorm(spin)
  }
  else {
   output[i]=qunif(spin, -1,1)
  }
}
```

```
Q2=read.csv('/Users/apple/Desktop/NTU_Tri1/AN6005/PPT\'s/Week\ 5-\ Bivar
View(Q2)
Q2$probabilities <- Q2$frequencies / 200
ans21=Q2$frequencies[(Q2$x==4)&(Q2$y==0)]/sum(Q2$frequencies)
ans22=sum(Q2$frequencies[Q2$x==4])/sum(Q2$frequencies)
#conditional:
P04=(Q2$probabilities[Q2$y==0 & Q2$x==4])/sum(Q2$probabilities[Q2$x==4])
P14=(Q2$probabilities[Q2$y==1 & Q2$x==4])/sum(Q2$probabilities[Q2$x==4])
P24=(Q2$probabilities[Q2$y==2 & Q2$x==4])/sum(Q2$probabilities[Q2$x==4])
P34=(Q2$probabilities[Q2$y==3 & Q2$x==4])/sum(Q2$probabilities[Q2$x==4])
#conditional expectation
ce=sum(c(0,1,2,3)*c(P04,P14,P24,P34))
```

Inverse Transform Method

```
n=1000
#create a placeholder for n obs as the output
output=c(0)
for(i in 1:n){
  spin=runif(1)
  output[i] = qnorm(spin)
}
prop.table(table(output))
hist(output)
```

Randomized Test $\quad$ ~N(2,2) 指 μ=2, var=2

```
set.seed(100)
# Q1
e10=rnorm(650)
e11=rnorm(650)

e21=rnorm(1000)
e22=rnorm(1000)

d1=rnorm(650,2,sqrt(2))
d2=rnorm(1000,5,sqrt(2))

# control group
y0_c= 50 + e10
y1_c= 50+d1 + e11

# treatment group
y0_d= 51+ e21
y1_d= 51+d2 + e22
```

sd

```
# Q2
naive= mean(y1_d)-mean(y0_c)

ATT = mean(y1_d - y0_d)
ATC = mean(y1_c - y0_c)

ATE = 1000/(1000 + 650) * ATT+ 650/(1000 + 650) *ATC
# or
mean(c(y1_d-y0_d, y1_c - y0_c))

# Q4
random_sample=sample(1:1650,400)

y0=c(y0_d, y0_c)
y1=c(y1_d,y1_c)

naive_randomize=mean(y1[random_sample[1:200]]) - mean(y0[random_sample[201:400]])
naive_randomize

# bias comparison. notice that under randomized design, the bias becomes much small
(naive-ATE)/ATE
(naive_randomize-ATE)/ATE
```

mean(x, trim=0.05) $\quad$ 5%~95%的mean

plot(density(data\$Delay) | data\$Carrier) ← 可叠

互斥与独立: 互斥→dependence ; independence: one occurs do not give any additional info about occurrance
↓ eg

3. {Male} and {Female} are two independent events because they cannot happen at the same time.

Answers:
a. A probability distribution of a random variable only tells us the possible values of the random variable
b. Different outcomes can be assigned to the same numerical value. X=1 when dice={2,4,6}
c. It assigns one numerical value to each outcome in the sample space. 对

Response Feedback: a) is not correct, because a distribution (whether in graphic or mathematical forms) shows both the possible values of the RV and the probabilities associated with these values.

* R中 var(), sd() 都是针对sample的, 除以的是n-1.
题中要求明确说明用 sd= sqrt(mean((data-μ)^2))

R语言 Dataframe 筛选
people[, c('name','sex')]
people[people\$age>30, c('name','sex')]

$E[X] = \sum_{\text{all } x} \sum_{\text{all } y} x \cdot f_{XY}(x,y)$
$E[Y] = \sum_{\text{all } x} \sum_{\text{all } y} y \cdot f_{XY}(x,y).$ $\quad$ discrete.
$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot f_{XY}(x,y)\, dx\, dy$
$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \cdot f_{XY}(x,y)\, dx\, dy$ $\quad$ continuous.