STAT 206

Random Experiment:

Sept. 7, 2018

- uncertain outcomes

Sample Space: S, Ω

- set of all possible outcomes {HH, HT, TH, TT} } Discrete (ints) / Continuous (reals)
  - → Event: subset of S (A)

Probability Definitions:

1) Classical: # items in A / # items in S.
   - → All outcomes in S must be equal prob.
   - → S must be finite. Circular logic. → counting could be impractical
2) Long Run Freq: Prob: outcomes as trials → ∞
   - → hard to apply
3) Subjective Prob: Prob based on private info
   - → Bayesian stats (ex: stocks)

Counting Rules:

Sept 10, 2018

→ Addition: P(A or B) = P(A) + P(B) ← (watch mutual exclusivity)

→ Multiplication: P(A AND B) = P(A) * P(B)

Example:

1) Pick 2 nums from {1...9}. P(1 is odd?)

|S| = 81. |A| = | first odd, AND second even | OR + | first even, AND second odd |

= 20 + 20 = 40.

∴ P = 40/81

More Rules:

→ n^k rule: Repetitions "with replacement" allowed, order matters. Selecting k objects from n.

→ permutation rule: Repetitions not allowed, order matters. k objects from n.

→ # outcomes: $^nP_k = \frac{n!}{(n-k)!}$

Example:

6 letter word from $\{A..F\}$. P(ABC beside eachother?)

Treat them as 1 letter.

$\#$ ways = (# ways to group) * (# ways to place the group).

$= 6 \times 24$

→ Combination Rule: same as permutations but order doesn't matter.

→ $^nC_k = \frac{n!}{k!(n-k)!}$

Example:

$\{A,B,C,D\}$ $n=4$ $k=2$ $^4C_2 = 6$.

Random 5 cards dealt. P(3 hearts, 2 spades)

$|S| = {}^{52}C_5$ $|A| = {}^{13}C_3 \times {}^{13}C_2$

$P(3 \text{ hearts}) = {}^{13}C_3 \times {}^{39}C_2$

Properties of ^nC_k:

1) $^nC_k = {}^nC_{n-k}$

2) $^nC_k = {}^{n-1}C_k + {}^{n-1}C_{k-1}$

Ex: 1) $\{A...G\}$ $n=7$ $k=3$. Possibilities where A doesn't appear?

$= {}^6C_3$

2) Possibilities where A appears? $^6C_2$ $\therefore {}^6C_3 + {}^6C_2 = {}^7C_3$

3) Proof: $\underbrace{n \times {}^{n-1}C_{k-1}}_{\text{choose captain then remaining team}} = \underbrace{k \times {}^nC_k}_{\text{chose team then captain}}$

4) $^{m+n}C_k = \sum_{j=0}^{k} ({}^mC_j)({}^nC_{k-j})$

m boys, n girls. k people selected.

Cases:

$^mC_0 * {}^nC_k$

$^mC_1 * {}^nC_{k-1}$

$\vdots$

$^mC_k * {}^nC_0$

$\Bigg\} \sum_{j=0}^{k} ({}^mC_j)({}^nC_{k-j})$

Scanned by CamScanner

Example: - Pigeon hole principle:

1) Prove there will be at least 2 people with the same initials in Waterloo. There are $> 26^3$ ppl in waterloo.

Pigeon Hole:

n slots, k objects, $k > n$, then there must be overlap.

Example:

Birthday Problem:

→ If $n > 23$, P(at least 2 same birthday) > 0.5.

$$P = 1 - P(\text{all distinct})$$

$$= 1 - \frac{365 \times 364 \times \ldots \times 342}{365^{24}} = 1 - \frac{{}^{365}P_{24}}{365^{24}}$$

Sept 12, 2018

	order	~~order~~
Rep	n^k	${}^{n+k-1}C_k$
~~Rep~~	${}^nP_k$	${}^nC_k$

Example:

1) Rand. 5 hand card is dealt.

a) P(4 of a kind)? KKKK2

$$P = \frac{13 \times 1 \times 1 \times 1 \times 48}{{}^{52}C_5}$$

b) P(full house)? AAA22

$$\frac{13 \times {}^4C_3 \times 12 \times {}^4C_2}{{}^{52}C_5}$$

2) Rand. string of $\{0,1\}$. $L = 10$. P(4 zeros)?

${}^{10}C_4$ positions for 0. That's it. $P = {}^{10}C_4 / 2^{10}$

Scanned by CamScanner

Pepys-Newton Problem:

Roll 6 die, 12, or 18. P(at least 1,2,3 sixes)? Highest?

a) 6 die, 1 six.
$P = 1 - P(\text{no } 6)$
$= 1 - \frac{5^6}{6^6}$

b) 12 die, 2 six.
$P = 1 - P(\text{no } 6) - P(1\ 6)$
$= 1 - \frac{5^{12}}{6^{12}} - \frac{5^{11} \times {}^{12}C_1}{6^{12}}$

c) 18 die, 3 six
"

Example:

1) Rand. words $\{m,i,s,s,i,s,s,i,p,p,i\}$ using all letters w/o rep.

a) how many words?
$\frac{11!}{4!\,2!\,4!}$ ← total letters ← repeated
$({}^{11}C_4 \times {}^{7}C_4 \times {}^{3}C_2)$

b) P(all S together)?
$\frac{8!}{4!\,2!} \cdot \frac{4!\,2!\,4!}{11!} = \frac{8!\,4!}{11!}$

Axiomatic Definition:

Probability is a fcn mapping from events → [0, 1] satisfying:

a) $P(S) = 1$, $P(\phi) = 0$

b) If $A \cap B = \emptyset$ (no overlap), then $P(A \cup B) = P(A) + P(B)$
↳ Mutually exclusive

Theorem: If $A \subset B$, $P(A) \le P(B)$
Let $B = A \cup (B - A)$. Then $P(B) = P(A) + P(B - A)$

Theorem: $P(A^c) = 1 - P(A)$
$A \cup A^c = S$.

Inclusion/Exclusion Principle:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example: Sept 14, 2018

1) Box contains 3 red, 4 black cards. 4 picked. P(at least 2 red)?

$$P = \frac{{}^3C_2 \times {}^4C_2 + {}^3C_3 \times {}^4C_1}{{}^7C_4} \quad \text{NOT} \quad \frac{{}^3C_2 \times {}^5C_2}{{}^7C_4} \text{ (double counting)}$$

Definitions:

- A, B are mutually exclusive if both can't happen at once.
  - ↳ $A \cap B = \phi$
- A, B are exhaustive if at least one must happen.
  - ↳ $A \cup B = S$

If both, then $A = \sim B$.

Laws of Prob:

Incl/Excl for 3 events:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(B \cap C) + P(C \cap A)] + P(A \cap B \cap C)$$

Example:

1) 2 die. P(at least 1 six)?

$P = 1 - \frac{5^2}{6^2} = {}^{11}/_{36}$ OR A_1 = first six, A_2 = second six.

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = {}^1/_6 + {}^1/_6 - {}^1/_{36}$$

2) $S = \{1, \ldots, 150\}$. How many are rel. prime to 70?

$70 = 2 \times 5 \times 7$ $\quad P = 150 - |(2 \cup 5 \cup 7)|$

$= 150 - (|2| + |5| + |7| - \ldots)$

DeMontfort's Problem:

$n = 4$. Let A_i = i^{th} letter in correct env.

$$P(A_1 \cup \ldots \cup A_N) = P(A_1) + P(A_2) + P(A_3) + P(A_4) \quad {}^4C_1 \quad P = 4\left(\frac{1}{4}\right) = 1$$

$$- [P(A_1 \cap A_2) + \cdots] \quad {}^4C_2 \quad P = 6\left(\frac{1}{12}\right) = \frac{1}{2}$$

$$+ [P(A_1 \cap A_2 \cap A_3) + \ldots] \quad {}^4C_3 \quad P = 4\left(\frac{1}{24}\right) = \frac{1}{6}$$

$$- P(A_1 \cap A_2 \cap A_3 \cap A_4) \quad {}^4C_4 \quad P = \frac{1}{24}$$

$P = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots$

As $n \to \infty$, $P = 1 - \frac{1}{e}$

Conditional Probability:

$P(A \text{ given } B) = P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

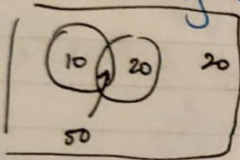
Sept 17, 2018

More Laws:

$P(A \cup B)^c = P(A^c \cap B^c)$ } DMG
$P(A \cap B)^c = P(A^c \cup B^c)$

Example:

1) 100 stdts. 60 MATH 135, 70 STAT 230, 50 both.
P(std taking ex. 1 course)?



$P = 30/100$

OR DMG:
$P(A \cap B^c) + P(A^c \cap B)$

Conditional Probability Pt. 2:

$P(A)$ = unconditional prob.
$P(A \cap B) = P(A|B) \cdot P(B)$

Example:

1) sum = 8. 2 die. A: sum = 8. B: first roll 3.
$P(A) = \frac{5}{36}$. $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{1}{36}}{\frac{1}{6}} = \dfrac{1}{6}$.

2) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)}{P(B)} = \dfrac{\frac{3}{8}}{\frac{7}{8}} = \dfrac{3}{7}$

Statistical Independence:

A, B independent if $P(A|B) = P(A)$
$\hookrightarrow \dfrac{P(A \cap B)}{P(B)} = P(A)$ $P(A \cap B) = P(A) * P(B)$

If A, B are m.e., they aren't independent.

1) If A, B ind. Are A^c and B ind? A^c, B^c, A, B^c? ✓

Law of Total Probability:

$B_1, B_2, \dots B_N$ are m.e. and exhaustive.

$P(A) = P(A|B_1)\cdot P(B_1) + \dots + P(A|B_N)P(B_N). = \sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)$

Example:

		R	B
1)	Box1	2	8
	Box2	3	7
	Box3	7	3

1 Box and card selected.

$P(\text{card is red}) = P(A) = ?$

$P(A) = P(A|B_1)\cdot P(B_1) + P(A|B_2)\cdot P(B_2) + P(A|B_3)\cdot P(B_3)$

2) A: sum 7. B = first roll 3.

$P(A) = 1/6$ $\quad P(A|B) = 1/6$ Independence...

3) C 206 : 30%.
D 237 : 40%.
E 334 : 50%.

Assume independence. $P(2 \text{ A's})$?

3 cases ([206, 237], etc).

$P(C \cap D \cap E^c) = 0.3 \times 0.4 \times 0.5$ (b/c independent)

$P(C^c \cap D \cap E) =$

$P(C \cap D^c \cap E) =$

4) $A = p, B = q$. $p + q + r = 1$ r(neither wins) Each game is ind. $P(A \text{ wins})$?

Cases: A win; Draw, Win; Draw, Draw, Win;

$p + rp + r^2p + \dots$

$p(1 + r + r^2 + \dots)$

$p\left(\frac{1}{1-r}\right) = \frac{p}{p+q}.$

Proof of Law of Total Prob:

$A = (A \cap B_1) \cup \dots \cup (A \cap B_N)$ ← because of exhaustivity

$= \sum P(A|B_i) \cdot P(B_i)$

Example:

1) Late: walk - 0.7   Drive - 0.1   Bus - 0.4 (equal chance)

a) P(late)? $= P(L|W) \cdot P(W) + P(L|D) \cdot P(D) + P(L|B) \cdot P(B)$

$= \frac{1}{3}(0.7 + 0.1 + 0.4)$

b) $P(B|L)$? ← Bayesian updating

Baye's Rule $\left\{ = \frac{P(B \cap L)}{P(L)} = \frac{P(L|B) \cdot P(B)}{P(L)} = \frac{\frac{1}{3} \cdot 0.4}{\frac{1}{3}(0.7+0.1+0.4)} \right.$

2) 100 cards · 50: born in July? 50: do drugs?

$P(yes) = P(Y|card\ 1) \cdot P(card\ 1) + \underbrace{P(Y|card\ 2)}_{\text{solve}} \cdot P(card\ 2)$

Bayes Rule: Sept 19, 2018

We know: $P(B_i)$ and $P(A|B_i)$. We want: $P(B_i|A)$ "posterior"

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} = \frac{P(A|B_i) \cdot P(B_i)}{P(A)}$$

Example:

1) 1% of humans have some disease, A. Person gets tested w/ 95% accuracy (+/-) both ways

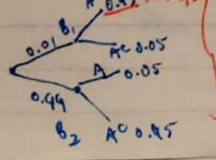
Test is pos. P(person has disease)?

B_1: have disease, B_2 = don't. A = test is positive.

1) $S = B_1 \cup B_2$ ✓   2) $B_1 \cap B_2 = \emptyset$ ✓

$P(B_1) = 0.01$ $P(B_2) = 0.99$ $P(A|B_1) = 0.95$ $P(A|B_2) = 0.05$

$$P(B_1|A) = \frac{(0.95)(0.01)}{(0.95)(0.01) + (0.05)(0.99)} = 16\%$$



can do w/ tree diagram

2) Box has 60 red, 40 blue cards.

Red: Were you born in Jan → March?

Blue: Do you like J. B.?

Pick card at random, answers question, puts card back. 100 students. 27 said yes

$P(yes) = 27/100$. $P(blue) = 0.4$. $P(red) = 0.6$.

$P(yes) = P(yes|red) \cdot P(red) + P(yes|blue) \cdot P(blue)$

$0.27 = (\frac{1}{4}) \cdot 0.6 + x \cdot 0.4$.

3) Monte Hall:

$P(D_1) = P(D_2) = P(D_3) = 1/3$. choose D_1 wlog. Monte opens a door w/o prize. S = switch doors and win. $P(S) = P(S|D_1) \cdot P(D_1) + P(S|D_2) \cdot P(D_2) + P(S|D_3) \cdot P(D_3)$, with $P(S|D_1) = 0$

$$P(S) = \overset{1}{P(S|D_2)} \cdot \overset{1/3}{P(D_2)} + \overset{1}{P(S|D_3)} \cdot \overset{1/3}{P(D_3)}$$
$$= 2/3$$

Simpson's Paradox:

Dr. A, Dr. B. Heart disease, bandaid.

		Heart	Bandaid
Dr. A:	✓	70	10
	×	20	0
Dr. B:	✓	2	81
	×	8	9

Dr A better in both, but Dr. B has higher total success rate.

More Baye's Rule: Sept 21, 2018

Box 1	7 Red	3 Blue	0.2	Prob. of box selection.
Box 2	5 Red	5 Blue	0.4	
Box 3	2 Red	8 Blue	0.4	$P(B_1 \mid \text{card is Red})$?

$$P(B_1 \mid \text{card red}) = \frac{P(Red|B_1) \cdot P(B_1)}{P(Red|B_1) \cdot P(B_1) + \ldots} = \frac{(\frac{7}{10})(0.2)}{(\frac{7}{10})(0.2) + (\frac{1}{2})(0.4) + \frac{2}{10} \cdot 0.4}$$

Scanned by CamScanner

Random Variables:

X is a random var if it takes a numerical value for every outcome of random experiment.
↳ discrete or continuous

Example:

1) Toss coin 3 times.
$X = |\#\text{ of heads} - \#\text{ of tails}|$

2) Roll die 2 times.
$X = $ sum of faces.

Distribution Function:

$f(x) = P(X = x)$ "mass fcn"
$= F(x) - F(x-1)$

Example:

1) $X =$ 1) above.

X	f
1	3/4
3	1/4

Properties:

1) $f(x_i) \geq 0$
2) $\sum f(x_i) = 1$

Cumulative Distribution Function:

$F(x) = P(X \leq x)$
$= \sum_{y \leq x} f(y)$

Properties:
1) Non decreasing fcn. $(x \leq y \Rightarrow F(x) \leq F(y))$
2) $\lim_{x \to -\infty} F(x) = 0$
3) $\lim_{x \to \infty} F(x) = 1$

Example:

x	f
0	0.3
1	0.5
2	0.2

Suppose $x < 0$: $F(x) = 0$
$x = 0$: $F(x) = 0.3$
$x = \frac{1}{2}$: $F(x) = 0.3$
$x = 1$: $F(x) = 0.8$ etc.

Expected Value:

$E(x) = \mu$ = avg value for a random var.
$= \sum_{i=1}^{n} x_i \cdot P_i$

Example: Sept 24, 2018

1) Rand 13 card hand. X = # of aces. f(x)?

x	f
0	"
1	"
2	"
3	$\frac{{}_4C_3 \times {}_{48}C_{10}}{{}_{52}C_{13}}$
4	"

2) 4 die rolled. X = max roll. f(x)?

*midterm: min: (1 - max)

x	f
1	
2	
3	
4	
5	
6	

Find CDF first.

$F(3) = P(x \leq 3)$

$= (\frac{1}{2})(\frac{1}{2})(\frac{1}{2})(\frac{1}{2}) = (\frac{3}{6})^4$

$F(2) = (\frac{2}{6})^4$

$\}\ f(3) = F(3) - F(2)$

Properties:

$E(g(x)) = \sum_{x \in X} g(x) f(x)$ LOTUS

$E(\text{const}) = \text{const}$

$E(ax+b) = aE(x) + b$ * only linear.

$E(x - E(x)) = 0.$

Variance:

Measure of how r.v. is spread across the mean

$V(x) = E[x - E(x)]^2$

↳ average of the squared deviation from the mean

St. Dev:

+ sqrt of $V(x)$

Example:

x	f	$x-\mu$	$(x-\mu)^2$
0	0.2	-1.1	1.21
1	0.5	-0.1	0.01
2	0.3	0.9	0.81

1) Find $E(x) = 1.1$
2) $E(x-\mu)^2 =$

Properties:

$V(x) = E[x-\mu]^2 = \sigma^2$
$= E(x^2) - \mu^2$ = avg of sqs − sq of avgs.

$V(\text{const}) = 0$

$V(ax+b) = a^2 Var(x)$

LOTUS: Can get $E(x^2)$ by letting $y = x^2$, and get f for y. Sept 26, 2018

→ dist of $x+y$ is NOT $f(x)+g(y)$

Distributions:

Binomial:

- n trials, 2 possible outcomes (success/failure)
- P(success) = p. same for every trial x = # of successes
- Independent trials.

$x \sim Bin(n,p)$ $E(x) = np$ $V(x) = np(1-p)$

1) n+1 possible values for x. $[0,n]$
2) $f(x) = {}^nC_x\, p^x (1-p)^{n-x}$ formula.

Example:

1) 10 coin tosses, $p(H) = 0.6$. $P(7 \text{ heads})$?
$P = f(7) = {}^{10}C_7\, (0.6)^7 (0.4)^3$

2) Random walk: $P(\text{right}) = 0.7$ $P(\text{left}) = 0.3$. P(back to start?)
right, left = 10 each.
$f(10) = {}^{20}C_{10}\, 0.7^{10}\, 0.3^{10}$

Geometric:
- Same as Bin, but x = # of failures before 1^{st} success.
$X \sim Geom(p)$
- $f(x) = (1-p)^x p$

~~Hypergeometric:~~ mistake: Negative Binomial.
- X = # of failures before k^{th} success. $x \sim$ Neg. Bin

Example: Sep 28, 2018

1) Coin 4 times. A: exactly 1 head. B = at least one head.
$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow$

$P(B) = 1 - P(0 \text{ heads})$
$= 1 - {}^4C_0 (0.5)^0 (0.5)^4$

$P(A \cap B) = P(A)$ (subset).
$P(A) = {}^4C_1 (0.5)^1 (0.5)^3$

2) Prove $E(x) = np$ for $x \sim Bin(n, p)$.
Let $x_i = 1$ if i^{th} outcome is ✓, 0 else.
$\therefore x = \sum_{i=1} x_i$

x_i	f
0	$1-p$
1	p

$E(x_i) = p$. $\therefore E(x) = \sum E(x_i) = np$

3) P that avg Jeopardy player lasts at least 4 shows?
$P = 1 - P(\text{lose}) - P(\text{win lose}) - P(\text{win win lose})$ Geometric.
$= 1 - \frac{2}{3} - (\frac{1}{3})(\frac{2}{3}) + (\frac{1}{3})(\frac{1}{3})(\frac{2}{3})$

Negative Binomial:
$P(X = x)$ = x failures before k^{th} success.
$x+k-1$ trials, $k-1$ successes. Success on k^{th} trial.
$${}^{x+k-1}C_{k-1} \, p^k (1-p)^{(x+k-1)-(k-1)}$$

Example:

1) Best of 7. P(Red Sox win a game) = 60%.

P(win in 6) = 2 failures before 4th success. $= {}^5C_3 \cdot (0.6)^3 (0.4)^2 \cdot 0.6$

$P = P(\text{w in } 4) + P(5) + P(6) + P(7)$.

Hypergeometric (For real):

N objects r successes N-r fails. pick n obj w/o replacement.

x = # of successes. x ~ Hypergeometric.

Example:

1) Rand. hand of 13 cards. X = # of kings.

N = 52. n = 13. r = 4.

$$P(X=3) = \frac{{}^4C_3 \times {}^{48}C_{10}}{{}^{52}C_{13}} = \frac{{}^rC_x \cdot {}^{N-r}C_{n-x}}{{}^NC_n}$$

2) 15 cards in a box. 6 red, 9 blue. Pick 8. Oct 1, 2018

P(w/o replacement, 3 reds?):

$$= \frac{{}^6C_3 \times {}^9C_5}{{}^{15}C_8} \quad \text{Hyper geo} \quad \cong 0.38$$

P(w replacement?):

$$= {}^8C_3 \; 0.4^3 \; 0.6^5 \quad \cong 0.278$$

3) 1500 cards: 600 R, 900 B. Pick 8 w/o replacement.

$$P = \frac{{}^{600}C_3 \times {}^{900}C_5}{{}^{1500}C_8} \cong 0.27 \cong \text{with replacement}$$

If $N \gg 1$ and $n \ll N$, hypergeometric ~ binomial.

"Binomial Approximation" → chance doesn't chance much between trials.

Binomial vs Negative Binomial:
X: # of failures before k^{th} success

Example:
1) 10% have O+ blood.
i) 100 pplane selected. $P(6\ O^+)$?
Binomial: ${}^{100}C_6 \cdot (0.10)^6 (0.90)^{94}$

ii) keep picking until 6 O^+. $P(100 \text{ trials})$?
Neg. Binomial: ${}^{99}C_5 \cdot (0.10)^5 (0.9)^{94} \times (0.10)$

Proof Σ hypergeometric $= 1$?
$$P(X=x) = \frac{{}^{r}C_x \; {}^{N-r}C_{n-x}}{{}^{N}C_n}$$ Story proof.

Poisson Distribution:
$X \sim$ Poisson
$x = 0, 1, 2, \ldots \qquad f(x) = \frac{e^{-\mu}\mu^x}{x!}$

$x = \#$ of successes in a time period. (ex: website hits in 1hr)
↳ No upper limit

1) $\Sigma = 1$?
$$\frac{e^{-\mu}\mu^0}{0!} + \frac{e^{-\mu}\mu^1}{1!} + \ldots$$
$$e^{-\mu}\underbrace{\left(1 + \frac{\mu}{1!} + \frac{\mu^2}{2!} + \ldots\right)}_{\text{maclauren series } e^x} = e^{-\mu} \cdot e^{\mu} = 1$$

2) $E(x)$? (μ)
$$= \sum_{x_i=0}^{\infty} x_i f(x_i) = 0 \cdot e^{-\mu}\frac{\mu^0}{0!} + 1 \times e^{-\mu}\frac{\mu^1}{1!} + \ldots$$
$$= \mu e^{-\mu}\left[1 + \frac{\mu}{1!} + \frac{\mu^2}{2!} + \ldots\right] = \mu$$

Try: variance $V(x)$ also $= \mu$.
If n is large and p is small, then $bin(n,p) \sim Poisson(\mu = np)$
↳ if p large, do $1-p$ (failure)

Example:

1) 200 ppl at party. P(2 born on Jan 1)?

Bin(n,p): $X = 2$ success = born on Jan 1. $p = 1/365$ $n = 200$.

$P = {}^{200}C_2 \cdot \left(\frac{1}{365}\right)^2 \left(\frac{364}{365}\right)^{198} \cong 0.086$

~Poisson: $P = \frac{e^{-\mu}\mu^x}{x!}$ where $\mu = np$, $x = 2$ — n large, p small

$= \frac{e^{-np}(np)^2}{2!} \doteq 0.08$

2) 122 tickets, 120 seats on flight. P(person no-shows) = 0.03

P(more people than seats?)

3) Select ppl w/ replacement until 5 left handed. Oct 3, 2018

20% lefties. P(at least 20?)

Method 1: X = # of trials. $P(X \geq 20)$?

$P(X=20) = {}^{19}C_4 (0.20)^5 (0.80)^{15}$. Similarly, $P(X=21)$, etc...

Method 2: Neg. bin. distr. Y = # of failures before 5^{th} success.

$\underbrace{\text{\# of failures}}_{x-k}$

P(Y at least 15)?

Importance of Poisson:

n = # of trials $\to \infty$

p = prob of success $\to 0$

$np = \mu$.

} $Bin(n,p) \longrightarrow Poi(np = \mu)$

Airline example:

P(1 no show) + P(0 no shows)

${}^{122}C_1 (0.03)^1 (0.97)^{121} + {}^{122}C_0 (0.03)^0 (0.97)^{122}$ } can ~ Poi

$\mu = 0.03 \times 122 = 3.66$

$\approx P(X=0) + P(X=1)$

$\approx \frac{e^{-3.66} 3.66^0}{0!} + \frac{e^{-3.66} 3.66^1}{1!}$

Proof:

$$f(x) = {}^nC_x\, p^x (1-p)^{n-x} \qquad n\to\infty \quad p\to 0 \quad np=\mu$$

$$= \frac{n!}{x!(n-x)!}\left(\frac{\mu}{n}\right)^x\left(1-\left(\frac{\mu}{n}\right)\right)^{n-x}$$

$$= \frac{n!}{x!(n-x)!}\,\frac{\mu^x}{n^x}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$

$$= \frac{\mu^x}{x!}\left[\frac{n!}{(n-x)!\,n^x}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}\right]$$

$$= \frac{\mu^x}{x!}\left[\frac{n\cdot(n-1)(n-2)\dots(n-x+1)}{n^x}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}\right]$$

$$= \frac{\mu^x}{x!}\left[\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-x+1}{n^x}\right)\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}\right]$$

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

$$= \frac{\mu^x}{x!}(1\cdot 1\cdot 1\cdots 1)(1^{-x})(e^{-\mu})$$

= poisson

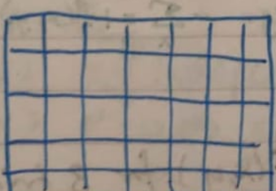
Poisson Process:

1) No upper bound?
2) # of failures make no sense/irrelevant?

} then, apply poisson

Example:

1) Rain falls at rate λ. x: # of drops in any square.



Assumptions:
(i) • Independence
(ii) • Individuality
• Proportionality (iii)

(i) # of successes independent if intervals disjoint
(ii) if areas $\to 0$, $P(2+)$ successes in a square $\to 0$.
(iii) $2\times$ area $= 2\times$ prob of raindrops.

} If these are satisfied, and X = # of successes in an interval t, $X \overset{Poi}{\sim} (\lambda t)$

2) # of eqmakes is a poisson process. $\lambda = 6$/year.

P(2 year period, 7 e.q's)?

X = # of eq in 2 years. $X \sim Poi(6 \times 2) = Poi(12)$

$$P(X=7) = \frac{e^{-12} 12^7}{7!}$$

Example

1) # accidents on 401 in 1hr $\lambda = 6$/hour.

a) $P(X>3)$ in one hour? $t = 1$ hour

$X \sim Poi(6 \times 1) = Poi(6)$

$P(X>3) = 1 - P(0) - P(1) - P(2) - P(3)$

b) Ex. 1 accident in 15 min? $t = 0.25$

$X \sim Poi(6 \times 0.25)$

c) Look at 8 dif hours. P(2 of these hours have >3 accidents)

# of trials = 8. p(success) = part a) = 0.84

$P(Y=2) = {}^8C_2 (0.84)^2 (0.16)^6$

d) X = # of hours have to survey before at least 4 hours w/ more than 3 accidents.

$P(X=7)$: First 6 trials, 3 successes + success on 7th.

${}^6C_3 (0.84)^3 (0.16)^3 \times 0.84$

Multivariate Distributions:

Joint Probability Dist. Fcn:

$f(x,y) = P(X=x, Y=y)$

- Non neg. - $\sum_x \sum_y f(x,y) = 1$

Ex: X = # heads, Y = |# heads − # tails|. Find $f(x,y)$ for 3 trials.

$X = 0, 1, 2, 3$

$Y = 3, 1$

X \ Y	1	3
0	0	1/8
1	3/8	0
2	3/8	0
3	0	1/8

Scanned by CamScanner

Marginal Distributions:

Ind. distr from joint distr? Yes.

$f_X(x) =$

x	f_X
0	1/8
1	3/8
2	3/8
3	1/8

y	f_Y
1	3/4
3	1/4

Just add rows/cols.

Can't go from marginal → joint, usually.

Independent Rand. Variables:

$f(x,y) = f_x(x) \cdot f_y(y) \quad \forall x,y$

↳ $P(A \cap B) = P(A)P(B)$, basically

} if ind, can go from marg → joint (just multiply)

Covariance:

$$Cov(x,y) = E\{(x-\underset{E(x)}{\mu_x})(y-\underset{E(y)}{\mu_y})\}$$

x = beers, y = gpa. If $x > \mu_x$, then usually $y < \mu_x$. prod = negative.

-Sign of covariance tells us direction of slope x,y.

Correlation Coefficient:

$$\rho_{x,y} = \frac{Cov(x,y)}{s.d(x)\, s.d(y)}$$

} strength of relationship

i) $|\rho_{xy}| \le 1$

ii) $\rho_{xy} = 1 \Rightarrow$ exact relationship.

iii) x,y independent $\Rightarrow \rho = 0$.

iv) $\rho = 0 \not\Rightarrow x,y$ independent. (measures linear relationship)

↳ x: -1 : $1/3$; 0 : $1/3$; 1 : $1/3$ $\quad y = x^2$, but $\rho = 0$.

v) High $\rho \not\Rightarrow$ causation.

vi) If $Y = ab + x$, $|\rho_{xy}| = 1$: 1 if $b>0$, -1 if $b<0$

Scanned by CamScanner

Example:

1) Prove $E(ax+b) = aE(x)+b$

$Y = ax+b$

$$E(ax+b) = \sum_{i=1}^{n} (ax_i + b) f(x_i) \quad g(x_i)$$

$$= \sum_{i=1}^{n} a x \cdot f(x_i) + \sum_{i=1}^{n} b \cdot f(x_i) \; = aE(x)+b$$

2) 2^k \$ if head appears first time on k^{th} trial, for $k: 1 \to 5$.
for $k \geq 6$, lose \$256. $E(x)$?

x	
2	1/2
4	1/4
8	1/8
16	1/16
32	1/32
-256	$1 - \sum$ above $= 1/32$

$E(x) = \sum x_i f(x_i) = -3$

3) Blood test. Test n together. If fails, test each indv.
Avg. tests you must do? n people, p(disease) = p. independent.

x	f
1	$(1-p)^n$
1+n	$1 - (1-p)^n$

$E(x) =$.

4) Prove $\sigma^2 = E(x-\mu)^2 = E(x^2) - \mu^2$

$= E(x^2 - 2\mu x + \mu^2)$

$= E(x^2) - E(2\mu x) + E(\mu^2)$

$= E(x^2) - 2\mu E(x) + \mu^2$

$= E(x^2) - \mu^2$

5) Prove $V(ax+b) = a^2 Var(x)$

$V(ax+b) = E(ax+b - E(ax+b))^2$

$= E(ax+b - aE(x) - b)$

$= E(a(x-E(x)))^2$

$= a^2 E(x-E(x))^2 = a^2 V(x)$

6) $p = 0.2$, $n = 100$ $\quad x \sim Bin(100, 0.2)$ $\quad$ cost of accident (x^2+x+3)

Avg cost?

$E(x^2+x+3) = E(x^2) + E(x) + E(3)$

$= E(x^2) + np + 3$

$= \underline{416} + \underline{5} + 3$

$V(x) + [E(x)]^2 \quad E(x)$ for Bin

Properties of Covariance:

$Cov(x,x) = Var(x)$

$Cov(x,k) = 0$

$Cov(x,y) = E(xy) - E(x)E(Y)$

If x, y ind $\Rightarrow Cov(x,y) = 0$. BUT not $\Leftarrow$

$Cov(ax+b, y) = a\,Cov(x,y)$

$Cov(ax+b, cY+d) = ac\,Cov(x,y)$

Example:

1) Prove $E(xy) = E(x)E(y)$ for ind. x, y.

$E(xy) = \sum_{x_i=1}^{n}\sum_{Y_i=1}^{n} xy \cdot f_{xy}(x,y)$

$= \sum_{x_i}^{n}\sum_{Y_i}^{n} xy\, f_x(x) \cdot f_Y(y)$

$= \sum_{x_i} x_i f(x_i) \sum_{Y_i} y_i f_y(y_i) = E(x)(E y)$ $\quad \therefore Cov(x,y) = 0$

Relation between 2 R.V.'s: Oct 15, 2018

If $Cov(x,y) > 0 \Rightarrow x, y$ have positive linear relationship.

Proof:

$E\{(x-\mu_x)(y-\mu_y)\}$

$= E(xy - \mu_x y - \mu_y x + \mu_x \mu_y)$

$= E(xy) - \mu_x \mu_y$

Example:

1) $Cov(a+bx, x) = ?$

$= Cov(a,x) + Cov(bx, x)$

$= 0 + b \cdot Cov(x,x)$

$= b\, Var(x)$

Proof:

$$\rho = \frac{cov(x,y)}{sd(x)\,sd(y)} = \frac{Cov(x, a+bx)}{sd(x)\,sd(y)} = \frac{b\, Var(x)}{sd(x) \cdot |b| \, sd(x)} = sign(b) \cdot 1$$

Continuous Random Variables:

Any val in an interval $[a,b]$

	Discrete	Cont.
CDF	$F(x) = P(X \le x)$	$F(x) = P(X \le x)$
MF	$f(x) = P(X = x)$	$f(x) = P(X = x) = 0 \ \forall x$
Density Fcn	–	$P(a \le x \le b)$ area under f between $[a \to b]$
CDF → MF	$\sum_{y \le x} f(y)$	$F(x) = \int_{-\infty}^{x} f(y) \cdot dy$
MF → CDF	$F(x) - F(x-1)$	$\frac{d}{dx} F(x) = f(x)$
E	$E(x) = \sum x f(x)$	$\int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$
V	$E(x^2) - (E(x))^2$	$\int_{-\infty}^{\infty} x^2 f(x) \cdot dx - \mu^2$
	$\sum f(x) = 1$	$\int_{-\infty}^{\infty} f(x) dx = 1$

Uniform Distribution:

$X \sim Uni[a,b]$ if

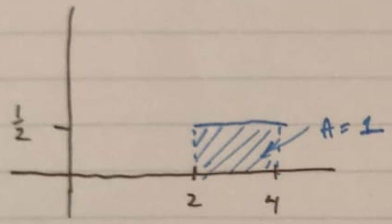
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{else} \end{cases}$$

$$E(x) = \frac{a+b}{2} \overset{\text{(proof)}}{=} \int_a^b x \cdot \frac{1}{b-a}\, dx = \left(\frac{x^2}{2}\Big|_a^b\right)\left(\frac{1}{b-a}\right)$$

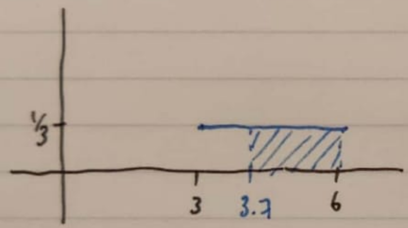
try: $V(x)$

Example:

1) $x \sim Uni[2,4]$



2) $x \sim Uni[3,6]$. $P(x \geq 3.7)$?



$$P = \frac{1}{3} \times (6-3.7) = \int_{3.7}^{6} \frac{1}{3} \cdot dy$$

Universality of the Uniform:

X is r.v. with CDF $F(x)$.

→ Generate n random #s from this dist.

If $u \sim Uni[0,1]$ then $F^{-1}(u) = x$.