Prob: cond	p. 442
Prob: Indep	p. 443
Bernoulli	p. 445
Bayes'	p. 455
Gen Bayes'	p. 457

Today:

- ▷ Prob, cond prob
- ▷ Bayes'
- ▷ Bernoulli

Read 7.3 about Bayesian spam filters

Discussion: Bayes' Theorem: $p(F|E) = \dfrac{p(E|F)\cdot p(F)}{p(E|F)\cdot p(F) + p(E|\overline{F})\cdot p(\overline{F})}$

Let's say we want to determine if an incoming mail is spam based on the fact that it contains the words "Nigerian Prince"

Let's denote the following events:

E: contains "Nigerian Prince"

F: is spam

Furthermore, let's suppose that we have 100 old mails that we have classified as spam or not spam by hand.

Let's say that 50 of them are spam. So $p(F) = \frac{50}{100} = \frac{1}{2}$

and $p(\overline{F}) = \frac{1}{2}$ as well.

Furthermore, we can look at the pile of spam and count that $\frac{40}{50} = \frac{4}{5}$ of them contain "Nigerian Prince", so $p(E|F) = \frac{4}{5}$

And of the non-spam $\frac{1}{50}$ contained "Nigerian Prince", so $p(E|\overline{F}) = \frac{1}{50}$

We can now use Bayes' theorem: $\dfrac{\frac{4}{5}\cdot\frac{1}{2}}{\frac{4}{5}\cdot\frac{1}{2} + \frac{1}{50}\cdot\frac{1}{2}} = \frac{40}{41} \approx 0{,}97$

And if it is over a threshold, say 0,9, we mark as spam.

7.3.2 Two boxes with ○ (white) and ● (blue) balls:

First pick random box, then random ball from that box.

p(Ball came from Box 1, given that blue ball was picked)?

Define events:

E: Get blue ball ● , $\bar{E}$: Get white ball ○

F: Choose box 1 , $\bar{F}$: choose box 2

So the question is now $p(F \mid E)$?

We know $p(F) = \frac{1}{2}$ and $p(\bar{F}) = \frac{1}{2}$.

$p(E \mid F) = \frac{3}{5}$, $p(E \mid \bar{F}) = \frac{1}{5}$

Use Bayes': $p(F \mid E) = \dfrac{p(E \mid F) \cdot p(F)}{p(E \mid F) \cdot p(F) + p(E \mid \bar{F}) \cdot p(\bar{F})} = \dfrac{\frac{3}{5} \cdot \frac{1}{2}}{\frac{3}{5} \cdot \frac{1}{2} + \frac{1}{5} \cdot \frac{1}{2}} = \frac{3}{4}$

7.3.4 Opium use and opium test

Real: Not using
Test: using

outcomes:

	Using	Not using
Test pos	TP	FP
Test neg	FN	TN

TP: True positive
FP: False positive
FN: False negative
TN: True negative.

We know FP = 2% and FN = 5% and can infer the rest.

	Using	Not using
Test pos	0,95	0,02
Test neg	0,05	0,98

1% = 0,01 of population using

Eg. to find TP; think: Say I am using and I take test. 5% of time it will answer (incorrectly) negative. The remaining must be 95%.

Define events E, F:

E: Tested pos $\quad \bar{E}$: Tested neg

F: Actually uses $\quad \bar{F}$: Does not use

Let's collect probabilities:

$p(F) = 0,01 \qquad p(\bar{F}) = 0,99$

$p(E|F) = TP = 0,95 \qquad p(E|\bar{F}) = FP = 0,02$

$p(\bar{E}|F) = FN = 0,05 \qquad p(\bar{E}|\bar{F}) = TN = 0,98$

1) $p(\bar{F}|\bar{E})$? $\dfrac{p(\bar{E}|\bar{F}) \cdot p(\bar{F})}{p(\bar{E}|\bar{F}) \cdot p(\bar{F}) + p(\bar{E}|F)p(F)} = \dfrac{0,98 \cdot 0,99}{0,98 \cdot 0,99 + 0,05 \cdot 0,01} \approx 0,999$

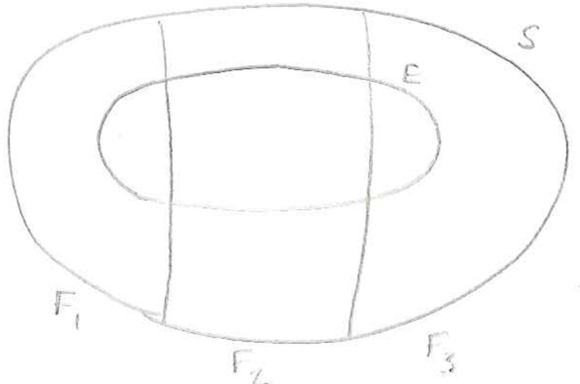
2) $p(F|E)$? $\dfrac{p(E|F) \cdot p(F)}{p(E|F)p(F) + p(E|\bar{F})p(\bar{F})} = \dfrac{0,95 \cdot 0,01}{0,95 \cdot 0,01 + 0,02 \cdot 0,99} \approx 0,324$

7.3.8 Events: E, F_1, F_2, F_3 from S

$p(E|F_1) = \frac{1}{8}$; $p(F_1) = \frac{1}{4}$

$p(E|F_2) = \frac{1}{4}$; $p(F_2) = \frac{1}{4}$ $\Bigg\}$ Sum = 1

$p(E|F_3) = \frac{1}{6}$; $p(F_3) = \frac{1}{2}$



$p(F_1|E)$?

$$p(F_1|E) = \frac{p(E|F_1)\,p(F_1)}{p(E|F_1)p(F_1) + p(E|F_2)p(F_2) + p(E|F_3)p(F_3)}$$ ← Generalized Bayes'

$$= \frac{\frac{1}{8}\cdot\frac{1}{4}}{\frac{1}{8}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{6}\cdot\frac{1}{2}} = \frac{3}{17} \approx 0{,}17$$

7.3.10

work

	Car	Bus	Bike
Late (prob)	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{1}{20}$
Choice (prob)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

He is now late; boss wants to find prob that he used car.

Define events: E, F_1, F_2, F_3:

E: late , F_1: car , F_2: Bus , F_3: Bike

Boss wants $p(F_1|E)$.

$p(E|F_1) = \frac{1}{2}$ $\quad p(E|F_2) = \frac{1}{5}$ $\quad p(E|F_3) = \frac{1}{20}$

$p(F_1) = \frac{1}{3}$ $\quad p(E|F_3) = \frac{1}{3}$ $\quad p(F_3) = \frac{1}{3}$

a) $$p(F_1|E) = \frac{p(E|F_1)\,p(F_1)}{p(E|F_1)p(F_1) + p(E|F_2)p(F_2) + p(E|F_3)p(F_3)}$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{3}\left(\frac{1}{2} + \frac{1}{5} + \frac{1}{20}\right)} = \frac{2}{3} = 0{,}66$$

b) Now boss knows choice prob: $p(F_1) = 0{,}3$, $p(F_2) = 0{,}1$, $p(F_3) = 0{,}6$

$$p(F_1|E) = \frac{\frac{1}{2}\cdot 0{,}3}{\frac{1}{2}\cdot 0{,}3 + \frac{1}{5}\cdot 0{,}1 + \frac{1}{20}\cdot 0{,}6} = \frac{3}{4} = 0{,}75$$

SE 7.10 p, q prime , $n = p \cdot q$

Random num m , $1 \le m < n$ (n diff choices of m)

p(m divides n)?

Recall exercise 6.1.44.

we counted # of m's that divide n to be $p \cdot q - (p+q-1)$

So $|E| = p \cdot q - (p+q-1)$

$|S| = n$ $\Rightarrow$ prob $= \dfrac{p \cdot q - (p+q-1)}{n}$

SE 7.12 Events $E_1, E_2, E_3, \ldots, E_n$ $p(E_i) > 0$ $\forall i$

Show $p(E_1 \cap E_2 \cap \cdots \cap E_n) = p(E_1) \cdot p(E_2 | E_1) \cdot p(E_3 | E_1 \cap E_2) \cdots p(E_n | E_1 \cap \ldots \cap E_{n-1})$ (*)

Induction:

Basis: $n = 2$

$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2 | E_1)$ ← Follows from def $p(E|F) = \dfrac{p(E \cap F)}{p(F)}$

Step: Assume (*), show $p(\underbrace{E_1 \cap \ldots \cap E_n} \cap E_{n+1}) = p(E_1) \cdot p(E_2|E_1) \cdots p(E_{n+1} | E_1 \cap \ldots E_n)$

let $F = E_1 \cap \cdots \cap E_n$

so

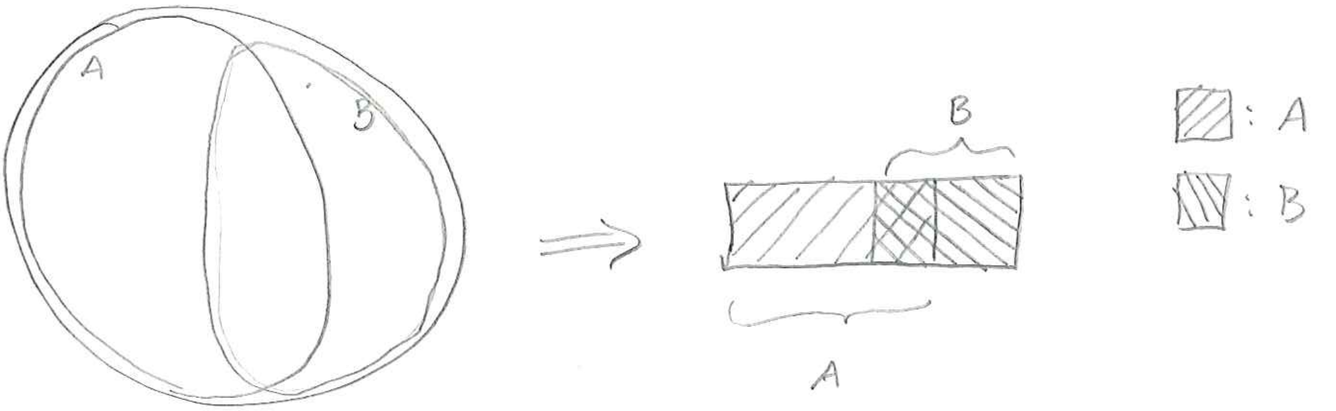
$p(F \cap E_{n+1}) \overset{\text{By basis}}{=} p(F) \cdot p(E_{n+1} | F)$

$= \underline{p(E_1) \cdot p(E_2 | E_1) \cdots p(E_n | E_1 \cap \cdots \cap E_{n-1})} \cdot p(E_{n+1} | F)$

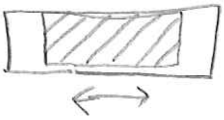
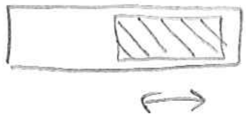
expand F and apply induction hyp

$= \text{—}\,||\,\text{—} \cdot \underbrace{p(E_{n+1} | E_1 \cap \cdots \cap E_n)}_{\text{Expand F}}$

SE 7.18 Events A, B; $p(A) = \frac{2}{3}$, $p(B) = \frac{1}{2}$

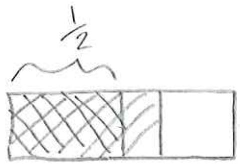


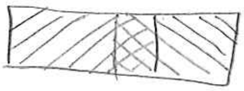
Venn diagram (2D)

We are allowed to slide  and 

Such that they overlap more or less

a) $p(\underbrace{A \cap B})$, Largest vs. smallest?

Largest: Move B inside A: 

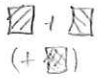
smallest: Move A to one side and B to another 

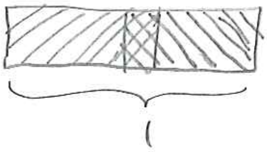
or

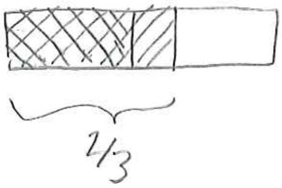
$\frac{2}{3} - \frac{1}{2} = \frac{4}{6} - \frac{3}{6} = \frac{1}{6}$

b) $p(\underbrace{A \cup B})$, Largest vs. smallest?

Combined



Largest: 

Smallest: More B inside A 

Jan 15. 2.

a) Choose 5 nums from $\{1, 2, 3, \ldots, 8\}$ at random

Show $\exists$ pair x, y of picked nums s.t. $x+y=9$

Consider $1+8=9$, $2+7=9$, $3+6=9$, $4+5=9$.

Can be seen as boxes

$\underbrace{\sqcup}_{1,8}$ $\underbrace{\sqcup}_{2,7}$ $\underbrace{\sqcup}_{3,6}$ $\underbrace{\sqcup}_{4,5}$ if 2 in same box $\Rightarrow$ we have pair

4 boxes, but 5 num $\Rightarrow$ PHP $\Rightarrow$ $\exists$ pair

b) Choose $K+1$ nums from $\{1, 2, \ldots, 2K-1, 2K\}$, Sum $= 2K+1$

Same idea: $1+2K = 2K+1$, $2+2K-1 = 2K+1$, ...

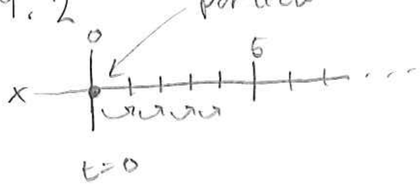
K boxes: $\underbrace{\sqcup}_{1,2K}$ $\underbrace{\sqcup}_{2,2K-1}$ ...

K boxes, but $K+1$ nums $\Rightarrow$ PHP $\Rightarrow$ $\exists$ pair

---

Jan 09. 2

a)



Every second: stay ($x_{new} = x_{old}$) — Prob q

or move right ($x_{new} = x_{old} + 1$) — Prob p

$q + p = 1$

$P_n(r)$: Prob that $x = r$ at $t = n$

Show $P_n(r) = \binom{n}{r} p^r q^{n-r}$ ← Take this as hint: Bernoulli

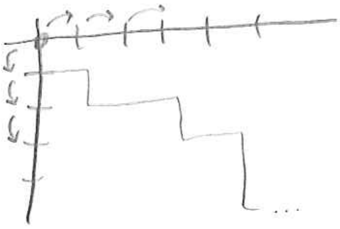
Each sec: a new (indep) trial; success: move right
failure: stay

Find prob of r success of prob p in n trials $\Rightarrow$ Exactly Bernoulli

Continued

prob

b) Introduce y-axis



Every sec: $\rightarrow$ or $\downarrow$ $\quad p + q = 1$

$Q_n(r,s)$: $x = r$ and $y = s$ at $t = n$

Observe: $r + s = n \Rightarrow s = n - r$

Specifically, if we know $x = r$, we uniquely know $y = n - r$

So it is simply the same as $P_n(r)$ (which gives prob of x-axis)

$$Q_n(r,s) = P_n(r) = \binom{n}{r} p^r q^{n-r} \quad \Big(= \underbrace{\binom{r+s}{r} p^r q^s}_{\text{Formula without } n} \Big)$$

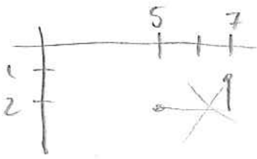
c) Given: $p = \frac{1}{3}$, $q = \frac{2}{3}$

i) Event A: pass through $(5,2)$?

$P(A)$? $\quad Q_7(5,2) = \binom{7}{5}\left(\frac{1}{3}\right)^5 \cdot \left(\frac{2}{3}\right)^2 = \frac{28}{729} \approx 0{,}038$

ii) Event B: pass through $(7,1)$?

$P(A \cap B)$? 0 as we cannot go up again



iii) Event C: pass through $(6,3)$? $\leftarrow$ $(1,1)$ compared to A

$P(A \cap C)$? Indep, so $P(A \cap C) = P(A) \cdot P(C)$

$P(C) = Q_2(1,1) \leftarrow$ from $(5,2)$ go $\rightarrow \times 1$ and $\downarrow \times 1$

$$= \binom{2}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^1 = \frac{4}{9}$$

$$P(A \cap C) = \frac{28}{729} \cdot \frac{4}{9} = \frac{112}{6561} \approx 0{,}017$$

Jan 10.2 wedding, outside, desert

rain: $5/365$ days

Bill: "It will rain that day!"

Events E, F

E: Predict ~~it~~ rains

F: Actually rains

we know about Bill

$p(E|F) = 0{,}9 = TP$

$p(E|\bar{F}) = 0{,}1 = FP$

Collect prob:

$p(F) = \frac{5}{365}$, $p(\bar{F}) = 1 - \frac{5}{365} = \frac{360}{365}$

$p(F|E)$?

$$p(F|E) = \frac{p(E|F) \cdot p(F)}{p(E|F)p(F) + p(E|\bar{F})p(\bar{F})} = \frac{0{,}9 \cdot \frac{5}{365}}{0{,}9 \cdot \frac{5}{365} + 0{,}1 \cdot \frac{360}{365}}$$

$= 0{,}11..$