STAT 5203 lecture 3

Important dates
Sept 15 : HW 1 due
Sept 24 : HW 2 due
Sept 29 : Midterm

Admin
1. Monday office hours (Before 6)
2. HW 2 due date (22nd or 24th)
3. Videos on birthday problem and matching problem

Agenda
1. Recap on conditional probability + law of total probability
2. Bayes rule
3. Independent events
4. Random variables
5. Intro to pdfs and cdfs

Recap: Conditional prob $A, B \subseteq S$

Conditional prob

$$\mathbb{P}(A) > 0 \qquad \mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} \Rightarrow \mathbb{P}(AB) = \mathbb{P}(B|A)\mathbb{P}(A)$$

Chain rule / mult. rule

$$\Rightarrow \mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$$
$$= \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)$$

law of total prob.

Today: Bayes rule "flip conditional probabilities"

Setup: Suppose we want $\mathbb{P}(B|A)$, but we only know $\mathbb{P}(A|B)$, $\mathbb{P}(B)$, ...

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(AB) + \mathbb{P}(AB^c)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} \quad \text{Bayes rule}$$

Example: A lab test is 95% correct in detecting a disease when it is present. The false positive rate is 1%. If 0.5% of the population has the disease, what is the prob that someone has the disease given that they test positive?

D = the person has the disease
P = the person tests positive

$$0.95 = \mathbb{P}(P|D)$$
$$0.01 = \mathbb{P}(P|D^c)$$
$$0.005 = \mathbb{P}(D) \qquad \mathbb{P}(D^c) = 0.995$$

$$\mathbb{P}(D|P) = \frac{\mathbb{P}(P|D)\mathbb{P}(D)}{\mathbb{P}(P|D)\mathbb{P}(D) + \mathbb{P}(P|D^c)\mathbb{P}(D^c)} = \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times (1 - 0.005)}$$

"with probability"

Example: A student is taking a multiple choice exam. Student knows the answer w.p. p and they guess w.p. $1-p$. If the student guesses, they get the question correct w.p. $1/m$ (m = # of choices).

Suppose a student gets the answer correct. What is the prob that they guessed?

A = Student knows answer
B = Student gets answer correct

General Bayes formula: If $F_1, \dots, F_n$ are disjoint and $\bigcup_{i=1}^{n} F_i = S$, then for some $E \subseteq S$,

$$\mathbb{P}(F_i \mid E) = \frac{\mathbb{P}(E \mid F_i)\mathbb{P}(F_i)}{\sum_{j=1}^{n} \mathbb{P}(E \mid F_j)\mathbb{P}(F_j)}$$

$\mathbb{P}(F_i)$ "prior probability"
$\mathbb{P}(F_i \mid E)$ "posterior probability"

Example: A plane is missing and has been assumed to have crashed in one of 3 regions (R_1, R_2, R_3) each with equal probability. Assume that if the plane crashed in R_1, the prob. of finding it is 0.8.

What is prob that the plane crashed in R_i ($i = 1, 2, 3$) if the search of R_1 is unsuccessful?

R_i = plane is in region i $\quad P(R_i) = 1/3 \quad i = 1, 2, 3$
E = region 1 search was unsuccessful

$$\mathbb{P}(R_1 \mid E) = \frac{\mathbb{P}(E \mid R_1)\,\mathbb{P}(R_1)}{\sum_{j=1}^{3} \mathbb{P}(E \mid R_j)\,\mathbb{P}(R_j)} = \frac{0.2 \times \frac{1}{3}}{0.2 \times \frac{1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{11}$$

$$\mathbb{P}(R_2 \mid E) = \mathbb{P}(R_3 \mid E) = \frac{5}{11}$$ (exercise)

	$\mathbb{P}(R_i)$	$\mathbb{P}(R_i \mid E)$
$i=1$	$1/3$	$1/11$
$i=2$	$1/3$	$5/11$
$i=3$	$1/3$	$5/11$

Independent Events

DEF: $A, B \subseteq S$, A and B are independent if $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$

* Equivalently, if $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$. Then A and B are indep iff

$\mathbb{P}(A \mid B) = \mathbb{P}(A)$
$\mathbb{P}(B \mid A) = \mathbb{P}(B)$

* (HW) If $A \cap B = \phi$, then they are not independent

Other fact: suppose A, B indep.
Then, ① A, B^c indep. want to prove that $\mathbb{P}(AB^c) = \mathbb{P}(A)\mathbb{P}(B^c)$
② A^c, B indep.
③ A^c, B^c indep.

Independence of multiple events: $A_1, A_2, \dots, A_n$

DEF: $A_1, A_2, \dots, A_n$ are pairwise indep. if $\forall i, j \in \{1, \dots, n\}$ $i \neq j$
$\mathbb{P}(A_i A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$

DEF: $A_1, A_2, \dots, A_n$ are mutually indep if $\forall$ subsets of events $A_{i_1}, A_{i_2}, \dots, A_{i_r}$ $2 \le r \le n$, we have

$$\mathbb{P}(A_{i_1} A_{i_2} \dots A_{i_r}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_r})$$

e.g., A_1, A_2, A_3, A_4, A_5. Then $\mathbb{P}(A_1 A_2 A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$
$\mathbb{P}(A_1 A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$
$\mathbb{P}(A_1 A_2 A_3 A_5) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\mathbb{P}(A_5)$
$\vdots$

Mutual independence $\Rightarrow$ pairwise indep.
$\not\Leftarrow$

e.g., tossing 2 coins → $S = \{HH, HT, TH, TT\}$
rolling 2 dice

$\mathbb{P}(HH) = \frac{1}{4} = \frac{\# HH}{\text{total}}$

$\mathbb{P}(HH) = \mathbb{P}(H) \times \mathbb{P}(H) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Repeated independent trials

e.g., toss $\underbrace{\text{two coins}}_{\text{2 trials}}$ $\mathbb{P}(H) = \frac{1}{2}$

$\boxed{\text{e.g.}}$ Roll a dice 5 times. What is the prob. of rolling a 5 exactly 2 times?

E_i = roll a 5 on the i^{TH} trial $\mathbb{P}(E_i) = \frac{1}{6}$ $\mathbb{P}(E_i^c) = \frac{5}{6}$

$$\mathbb{P}(E_1 E_2 E_3^c E_4^c E_5^c) = \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3^c)\mathbb{P}(E_4^c)\mathbb{P}(E_5^c) = \left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right)^3$$

5 5 ≠5 ≠5 ≠5
5 ≠5 ≠5 5 ≠5
$\vdots$

$$\mathbb{P}(2\ 5\text{'s}) = \underbrace{\#\text{ of ways to get 2 5s in 5 rolls}}_{\binom{5}{2}} \times \left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right)^3$$

In general, if we have N independent trials and in each trial $\mathbb{P}(E_i) = p$ ← success in trial i

$\mathbb{P}(K \text{ success}) = \binom{N}{K} p^K (1-p)^{N-K}$ ← Binomial distribution

Useful: If $A_1, \dots, A_n$ are indep. and $\mathbb{P}(A_i) = p_i$

$\mathbb{P}$(at least on A_i occurs) $= 1 - \mathbb{P}$(none of $A_1, \dots, A_n$ occur)
$= 1 - (1-p_1) \times (1-p_2) \times \dots \times (1-p_n)$ } Birthday problem

e.g., $\mathbb{P}$(m failures before n success) $= \binom{m+n-1}{m} p^n (1-p)^m$ Negative Binomial

Ch 3 DeGroot: Random variables, distributions, pmf, pdf, cdfs

Random variables outcomes $\to$ functions of outcomes

e.g., toss 2 coins $S = \{HH, HT, TH, TT\}$

# of heads for	
HH	2
HT, TH	1
TT	0

$\mathbb{P}(1 \text{ head}) = \frac{1}{2}$

DEF: A random variable, X, is a function from the sample space to the real numbers $X: S \to \mathbb{R}$

X = # of heads
$X: S \to \{0, 1, 2\}$

$S = \{HH, HT, TH, TT\}$

$HH \to 2 \in \mathbb{R}$, $HT \to 1 \in \mathbb{R}$, $TH \to 1 \in \mathbb{R}$, $TT \to 0 \in \mathbb{R}$

$\mathbb{P}(X = 0) = \frac{1}{4}$
$\mathbb{P}(X = 1) = \frac{1}{2}$
$\mathbb{P}(X = 2) = \frac{1}{4}$

Random Variab

Connection between r.v. and probabilities

More generally, $E \subset \mathbb{R}$ Then $\mathbb{P}(X \in E) = \mathbb{P}(\{s \in S : X(s) \in E\})$

e.g. $\mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

e.g., toss two coins. Start with \$1. If H then double our fortune, T then lose everything

X = fortune after 2 flips

$S = \{HH, HT, TH, TT\}$

$HH \to 4$, $HT \to 0$, $TH \to 0$, $TT \to 0$

$\mathbb{P}(X = 0) = \frac{3}{4}$
$\mathbb{P}(X = 4) = \frac{1}{4}$
$\mathbb{P}(X = a) = 0$ for $a \neq 0, 4$

DEF: The distribution of X is the collection of all probs. of the form $\mathbb{P}(X \in C)$ $\forall$ $C \subseteq \mathbb{R}$ such that $\{X \in C\}$ is an event

Discrete random variables: "There are a countable number of possible values of X"

DEF: (Probability mass function or pmf). For a discrete random variable, the pmf $p: \mathbb{R} \to [0,1]$ of X is given by

$$p(x) = \mathbb{P}(X = x) = \mathbb{P}(\{s \in S : X(s) = x\})$$

realization of X

Random variable

pmf properties: X can have values $x_1, x_2, x_3, \ldots$ The pmf satisfies

(1) $p(x_i) \geq 0$ $\quad \forall\, i = 1, 2, 3, \ldots$

(2) $p(x) = 0$ $\quad$ for $x \notin \{x_1, x_2, \ldots\}$

(3) $\sum_{i=1}^{\infty} p(x_i) = 1$

common pmfs

(1) Binomial distribution: n indep. each with success prob. p

$$\underbrace{\mathbb{P}(X = k)}_{\mathbb{P}(\underbrace{\text{Number of success}}_{X} = k)} = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n$$

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Note: define pmf for all possible values

(1a) Bernoulli distribution: Binomial dist. with $n = 1$

$$p(k) = \begin{cases} 1-p & k = 0 \\ p & \text{for } k = 1 \\ 0 & \text{o/w} \end{cases}$$

(2) Geometric distribution: indep. each with success prob. p

$X = \#$ of trials required for first success

$$\mathbb{P}(X = k) = (1-p)^{k-1} p$$

$$p(k) = \begin{cases} (1-p)^{k-1} p & k = 1, 2, \ldots \\ 0 & \text{o/w} \end{cases}$$

(3) Negative binomial

Example : The pmf of a rv X is given by

$$p(i) = \begin{cases} c\dfrac{L^i}{i!} & \text{for } i = 0, 1, 2, \ldots \\ 0 & \text{o/w} \end{cases}$$

c is a constant
$L > 0$

(a) Find the value of c

$$\sum_{i=0}^{\infty} \frac{cL^i}{i!} = 1 \Rightarrow c\underbrace{\sum_{i=0}^{\infty}\frac{L^i}{i!}}_{e^L} = 1 \rightarrow ce^L = 1 \Rightarrow c = e^{-L}$$

(Solve for c)

(b) $\mathbb{P}(X=0) = c = e^{-L}$ $\qquad \dfrac{e^{-L}L^0}{0!} = e^{-L}$

(c) $\mathbb{P}(X > 2) = \displaystyle\sum_{i=3}^{\infty} \frac{e^{-L}L^i}{i!}$

(easier) $= 1 - \mathbb{P}(X=1) - \mathbb{P}(X=0)$

$= 1 - \dfrac{e^{-L}L^1}{1!} - e^{-L}$

Remark : The above pmf is the pmf of the Poisson distr. w/ param L

DEF : (Cumulative distribution function or cdf) The cdf $F: \mathbb{R} \to [0,1]$ of an r.v X is defined as follows $F(b) = \mathbb{P}(X \leq b)$ for $b \in \mathbb{R}$

Note : If X is discrete, $F(b) = \displaystyle\sum_{x \leq b} p(x)$

e.g., flip 2 coins. X = # of heads

$$P(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \\ 0 & \text{o/w} \end{cases} \qquad F(x) = \begin{cases} 0 & (-\infty, 0) & \\ 1/4 & [0, 1) & p(0) \\ 3/4 & [1, 2) & p(0) + p(1) \\ 1 & [2, \infty) & p(0) + p(1) + p(2) \end{cases}$$

$F(1) = \mathbb{P}(X \leq 1) = \mathbb{P}(X=0) + \mathbb{P}(X=1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

$F(1.5) = \mathbb{P}(X \leq 1.5) = \mathbb{P}(X=0) + \mathbb{P}(X=1) = \frac{3}{4}$

$F(3) = \mathbb{P}(X \leq 3) = p(0) + p(1) + p(2) = 1$