Probability Theory Refresher (§2)

Will be slightly informal (but in a way that can be made completely rigorous)

Axiomatic approach → text book / after class. When in doubt: ASK!

Probability distribution on $\mathcal{A}$ (finite set): $P: \mathcal{A} \to \mathbb{R}_{\geq 0}$, $\sum_{a \in \mathcal{A}} P(a) = 1$

e.g. Bernoulli(f): $\mathcal{A} = \{0,1\}$, $P(1) = f$, $P(0) = 1-f$

Uniform($\mathcal{A}$): $P(a) = \frac{1}{\#\mathcal{A}}$ $\forall a \in \mathcal{A}$

Random variable (RV) $X \;\hat{=}\;$ prob. dist. P_X on set $\mathcal{A}_X$

NOTATION: $X \sim P$ for $P_X = P$

UNLIKE THE BOOK, I ALWAYS DISTINGUISH X and x

$\Pr(\underset{\text{RV}}{X} = \underset{\text{outcome}}{x}) = P_X(x) = P(x)$ we leave out subscript if clear!

$\Pr(X \in S) = \sum_{x \in S} P(x)$

$\Pr(\text{condition on } X) = \sum_{x \text{ cond. holds}} P(x) = \Pr(X \in \{x \text{ s.th. condition holds}\})$

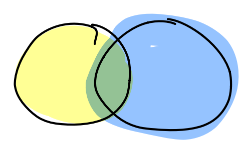
e.g. if X random variable on $\{1, \ldots, 6\}$:

$\Pr(X \text{ even and } X \neq 2) = \Pr(X \in \{4, 6\}) = P(4) + P(6)$

* $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \underbrace{\Pr(A \text{ and } B)}_{=0 \text{ if mutually exclusive}}$

$\leq \Pr(A) + \Pr(B)$

"union bound"



* X RV, f function $\Rightarrow$ $Y = f(X)$ RV

$\Pr(Y = y) = \sum_{x : f(x) = y} \Pr(X = x)$ or simply $P(y) = \sum_{f(x) = y} P(x)$

More than one random variable

How to describe "pair of RVs" (X, Y)? "Joint prob. dist":

$\Pr(X = x, Y = y) = P_{(X,Y)}(x, y) = P_{XY}(x, y) = P(x, y)$

i.e. (X, Y) is RV on $\mathcal{A}_{XY} = \mathcal{A}_X \times \mathcal{A}_Y$. Similar for tuples.

* Can visualize by "probability table" or "contingency table":

Y \ X	SUMMER	WINTER	
SUN	30%	10%	40%
RAIN	20%	40%	60%
	50%	50%	

* marginal distributions of X & Y:

$$P(x) = \sum_y P(x,y) \quad \& \quad P(y) = \sum_x P(x,y)$$

i.e. $Pr(X=x) = \sum_y Pr(X=x, Y=y)$ etc.

* X, Y are called independent if $P(x,y) = P(x) \cdot P(y)$

NOT independent! $p(\text{SUN}, \text{SUMMER}) \neq p(\text{SUN}) \cdot p(\text{SUMMER})$

How about:

15%	60%
5%	20%

independent!

Conditional prob. dist. of Y given X:

$$Pr(Y=y \mid X=x) := \frac{Pr(X=x, Y=y)}{Pr(X=x)}$$

NOTATION: $P_{Y|X=x}(y)$, $P_{Y|X}(y|x)$, $p(y|x)$, ...

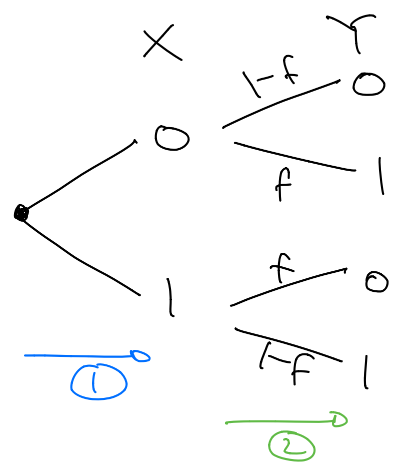
i.e. $P(y|x) = \dfrac{P(x,y)}{P(x)}$ and $P(x|y) = \dfrac{P(x,y)}{P(y)}$

* $P(y|x)$ is prob. dist in y for each fixed x ?

Two simple rewritings:

* $P(x,y) = \underbrace{P(x)}_{①} \underbrace{P(y|x)}_{②} = P(y) P(x|y)$

e.g. X channel input, Y channel output, $P(y|x)$ channel



FORWARD

* Bayes rule:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_{x'} P(y|x') P(x')}$$

INVERSE

e.g. $P(\underset{y}{\text{pos}} \mid \underset{x}{\text{sick}}) = P(\text{neg} \mid \text{healthy}) = 90\%$, $P(\text{sick}) = 1\%$

$$\Rightarrow P(\underset{x}{\text{sick}} \mid \underset{y}{\text{pos}}) = \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.1 \times 0.99} = \frac{1}{12} < 10\% \; ?$$

e.g. decoding the repetition code R_3: assume $S \sim$ Uniform($\{0,1\}$)
all independent RV's

$Y_1 = S \oplus N_1, \dots, Y_3 = S \oplus N_3$ where $N_1, N_2, N_3 \sim$ Bernoulli(f)
Sum modulo two (XOR)

Assume we received $y = y_1 y_2 y_3$. How should we estimate s?

$$P(s|y) = \frac{P(y|s)\,P(s)}{P(y)} \qquad P(s) = \tfrac{1}{2}, \quad P(y) \text{ fixed}$$

$$\Rightarrow \frac{P(S=0|y)}{P(S=1|y)} = \frac{P(y|S=0)}{P(y|S=1)} = \frac{P(y|X=000)}{P(y|X=111)} = \prod_{k=1}^{3} \underbrace{\frac{P(y_k|X_k=0)}{P(y_k|X_k=1)}}_{\frac{1-f}{f} \text{ if } y_k=0, \text{ else } \frac{f}{1-f}}$$

$$= \underbrace{\left(\frac{1-f}{f}\right)}_{>1 \text{ if } f<50\%}^{\#0\text{'s} - \#1\text{'s}} = \begin{cases} >1 & \text{if } \#0\text{'s} > \#1\text{'s} \\ <1 & \text{if } \#1\text{'s} > \#0\text{'s} \end{cases}$$

= majority vote

Combining independent RV's: independent and identical distribution

QUIZ: ① Let $X, N \overset{\text{IID}}{\sim}$ Uniform($\{0,1\}$), $Y = X \oplus \underbrace{N}_{\text{uniform!}}$.
Are X and Y independent?

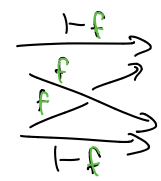
YES! $\Pr(X=x, Y=y) = \Pr(X=x, N=x\oplus y) = \frac{1}{4} = \Pr(X=x)\Pr(Y=y)$

② How to label two dice w/ numbers from $0, \dots, 6$ such that their sum is $\sim$ Uniform($\{1, 2, \dots, 12\}$)?

A: 123456
B: 000666

Binomial(n, f): Distribution of $Y = X_1 + \dots + X_n$ where $X_i \overset{\text{IID}}{\sim}$ Bernoulli(f)

* e.g. number of bit flips when we send n bits through



* $\Pr(Y=k) = \binom{n}{k} \underbrace{f^k (1-f)^{n-k}}_{\text{probability of any such string}}$

$\binom{n}{k}$: # bitstrings with k ones, $n-k$ zeros

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{binomial coefficient}$$

"Numerical" random variables

If $X \sim P$ is RV with values in $\mathcal{A} \subseteq \mathbb{R}$:

Expectation value (mean): $EX = E[X] = \sum_x P(x) \cdot x$

* $E[f(X)] = \sum_x P(x) \cdot f(x)$ "law of the unconscious statistician"

* $E[cX] = c \cdot E[X]$ & $E[X+Y] = E[X] + E[Y]$ (A)

* If X, Y independent: $E[XY] = E[X] \cdot E[Y]$ $\sum_{x,y} p(x) p(y)\, xy$

  ↯ $X \sim \text{Uniform}(\{\pm 1\})$, $Y = -X$ $\Rightarrow E[XY] = -1$, $E[X] = E[Y] = 0$
  NOT indep

Variance: $\text{Var}(X) = E\left[(X - EX)^2\right]$

$= \sum_x P(x)(x - EX)^2 = E[X^2] - E[X]^2$

* $\text{Var}(cX) = c^2 \,\text{Var}(X)$

* If X, Y independent:

  $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ (B)

  use that $E[XY] = E[X] \cdot E[Y]$

Examples

P	Bernoulli(f)	Binomial(n,f)	
E	f	$n \cdot f$	(A)
Var	$f(1-f)$	$n \cdot f \cdot (1-f)$	(B)

$E\left[(X - EX)^2\right] = E\left[(X - f)^2\right]$
$= f(1-f)^2 + (1-f)(0-f)^2 = f(1-f)$

Interpretation?

Markov inequality: If $X \geq 0$: $\boxed{\Pr(X \geq t) \leq \frac{E[X]}{t} \quad (\forall t > 0)}$

Chebyshev inequality: $\boxed{\Pr\left(|X - EX| \geq \varepsilon\right) \leq \frac{\text{Var}(X)}{\varepsilon^2}}$

with high probability
WHP deviation from mean is of order $\sqrt{\text{Var}(X)}$

Law of large numbers: $X_1, \ldots, X_n \overset{\text{IID}}{\sim} P$ w/ mean μ, variance σ^2,
$\bar{X} := \frac{1}{n}(X_1 + \ldots + X_n)$.

$\Rightarrow \Pr(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}$

WHP: empirical averages ≈ expectation value

We will look into these results more carefully next week and then discuss "entropy" and compression!