# 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)** $X \mathrel{\hat{=}}$ prob. dist. $P_X$ on set $\mathcal{A}_X$

NOTATION: $X \sim P$ for $P_X = P$ leave out subscript if clear

UNLIKE THE BOOK, I ALWAYS DISTINGUISH $X$ and $x$

$\Pr(\underset{\text{RV}}{X} = \underset{\text{outcome}}{x}) = P_X(x) = P(x)$

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

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

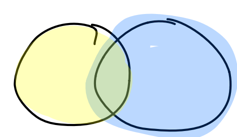
e.g. if $X$ random variable on $\{1, \ldots, 6\}$:

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

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

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

"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 "contingency table":

| $y \backslash x$ | 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.

$$Pr(X=x) = \sum_y Pr(X=x, Y=y) \quad \text{etc.}$$

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

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

How about:

independent!

| | |
|---|---|
| 15% | 60% |
| 5% | 20% |

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 $P(x|y) = \dfrac{P(x,y)}{P(y)}$

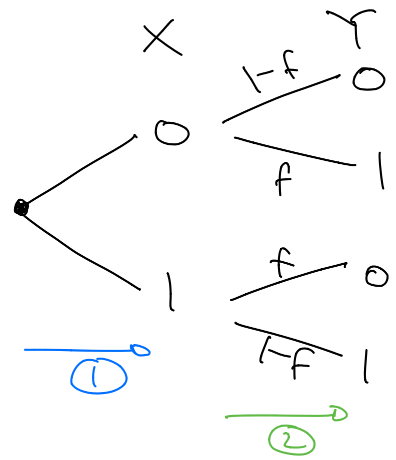
* $P(y|x)$ is prob. dist in $y$ for each fixed $x$ !

Two simple rewritings:

* $P(x,y) = P(x) P(y|x) = P(y) P(x|y)$

$\quad$ (1) $\quad$ (2)

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




FORWARD

* Bayes rule:

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

INVERSE

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

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

e.g. decoding the repetition code $R_3$: assume $S \sim$ Uniform$(\{0,1\})$ $\Big\}$ all indep.

$R_1 = S \oplus N_1, \dots, R_3 = S \oplus N_3$ where $N_1, N_2, N_3 \sim$ Bern$(f)$

Assume we received $r = r_1 r_2 r_3$. How should we estimate $s$?

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

$$\Rightarrow \frac{P(S=0|r)}{P(S=1|r)} = \frac{P(r|S=0)}{P(r|S=1)} = \frac{P(r|T=000)}{P(r|T=111)} = \prod_{k=1}^{3} \underbrace{\frac{P(r_k|T_k=0)}{P(r_k|T_k=1)}}_{\frac{1-f}{f} \text{ if } r_k=0, \text{ else } \frac{f}{1-f}}$$

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

= majority vote

## Combining independent RV's:

$\boxed{\text{Quiz:}}$ ① Let $S, K \overset{iid}{\sim}$ Uniform$(\{0,1\})$, $T = S \oplus K$.

Are $S$ and $T$ independent?

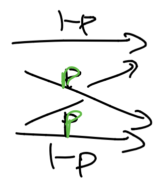
Yes: $P(S=s, T=t)$
$= P(S=s, K=s\oplus t) = \frac{1}{4}$

② How to label two dice w/ numbers from $\{0,1,\dots,6\}$ s.th. their sum $\sim$ Uniform$(\{1,2,\dots,12\})$

A: 123456
B: 000666

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

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




* $\Pr(Y=k) = \binom{n}{k} \underbrace{p^k (1-p)^{n-k}}_{\text{probability of any such string}}$ $\qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$ binomial coefficient

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

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

  ↯ $X \sim \text{Uniform}(\{\pm 1\})$, $Y = -X$ $\Longrightarrow$ $E[XY] = -1$, $E[X] = E[Y] = 0$

  NOT indep

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

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

* $Var(cX) = c^2 Var(X)$

* If $X, Y$ independent:

  $Var(X+Y) = Var(X) + Var(Y)$ (B)

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

Examples

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

$p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2$
$= p(1-p)$

Interpretation?

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

Pf: $\Pr(X \geq t) = \sum_{x \geq t} P(x) \leq \sum_{x \geq t} P(x) \frac{x}{t} \leq \frac{E[X]}{t}$ $\square$

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

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

Pf: Apply Markov to $Y = (X - EX)^2$. $\square$

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

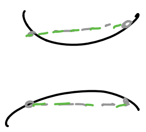
$\Longrightarrow \Pr(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}$

WHP: empirical averages $\approx$ expectation value

Pf: $E\bar{X} = \mu$ & $Var(\bar{X}) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{\sigma^2}{n}$. $\leadsto$ Chebyshev. $\square$

One last reminder:

f is called convex / concave if $p \cdot f(x) + (1-p) f(x') \geq f(px + (1-p)x')$ (convex) / $\leq$ (concave) $\quad \forall p \in [0,1]$, $x, x'$



exp, $x^2$, ...  Sufficient for convex: $f'' \geq 0$

log, $\sqrt{x}$, ...  concave: $f'' \leq 0$

Strongly convex / concave if "=" only holds for $p=0$ or $p=1$

if > or <

Jensen's inequality: If f convex / concave: $E[f(X)] \geq f(EX)$ / $\leq$

* If strongly convex/concave: "=" iff X is constant.

From next week on we will study:

Entropy of $X \sim P$:

$$H(X) = H(P) = \mathbb{E}\left[\log \frac{1}{P(X)}\right] = \sum_x P(x) \log \frac{1}{P(x)} = -\sum_x P(x) \log P(x)$$

ALWAYS BASE 2 $\qquad 0 \cdot \log 0 \equiv 0$

* $0 \leq H(X) \leq \log(\#\text{outcomes})$

= iff X constant $\qquad$ = iff X uniformly random

$\mathbb{E}\left[\log \frac{1}{P(X)}\right] \leq \log \mathbb{E}\left[\frac{1}{P(X)}\right]$ by Jensen

* $X \sim \text{Bernoulli}(p)$: binary entropy function

$$H(X) = -p \cdot \log(p) - (1-p)\log(1-p)$$