Probability Theory Refresher

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 \( \mathbf{A} \) (finite set):
\[
P : \mathbf{A} \rightarrow \mathbb{R}_{\geq 0}, \quad \sum_{a \in \mathbf{A}} P(a) = 1
\]
e.g. Bernoulli (C):
\[\mathbf{A} = \{0, 1\}, \quad P(1) = p, \quad P(0) = 1 - p\]
Uniform (A):
\[P(a) = \frac{1}{\# \mathbf{A}} \quad \forall a \in \mathbf{A}
\]

Random variable (RV) \( X \) is prob. dist. \( P_X \) on set \( \mathbf{A} \)

NOTATION:
\[X \sim P \quad \text{for} \quad P_X = P\]

\[
\Pr(X = x) = P_X(x) \equiv P(x)
\]

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

\[
\Pr(\text{condition on } X) = \sum_{x \text{ condition holds}} P(x) = \Pr(X \in \{x \text{ s.t. condition holds}\})
\]

e.g. if \( X \) random variable on \( \{1, 2, 3, 4, 5, 6\} \):
\[
\Pr(X \text{ even } | X \neq 2) = \Pr(X \in \{4, 6\}) = P(4) + P(6)
\]

\[
\begin{align*}
\Pr(A \text{ or } B) &= \Pr(A) + \Pr(B) - \Pr(A \text{ and } B) \\
&\leq \Pr(A) + \Pr(B)
\end{align*}
\]

"union bound"

\[
\begin{align*}
X \text{ RV, } f \text{ function } \rightarrow Y = f(X) \text{ RV} \\
\Pr(Y = y) &= \sum_{x : f(x) = y} \Pr(X = x) \quad \text{or simply} \quad P(y) = \sum_{f(x) = y} P(x)
\end{align*}
\]

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 \( \mathcal{A}_X \times \mathcal{A}_Y \). Similar for tuples.
Concise by "contingency table":

- Marginal distributions of \( X \) & \( Y \):
  \[
P(X) = \sum_Y P(X, Y) \quad \text{and} \quad P(Y) = \sum_X P(X, Y)
  \]

  i.e.
  \[
  P_r(X = x) = \sum_Y P_r(X = x, Y = y) \quad \text{etc.}
  \]

- \( X, Y \) are called independent if \( P(X, Y) = P(X) \cdot P(Y) \)

Conditional prob. dist. of \( Y \) given \( X \):

\[
P_r(Y = y | X = x) = \frac{P_r(X = x, Y = y)}{P_r(X = x)}
\]

**NOTATION:** \( P_r(Y | X = x) \), \( P_r(X | Y = y) \), \( P(Y | X) \), ...

i.e. \( P(Y | X) = \frac{P(X, Y)}{P(X)} \) and \( P(X | Y) = \frac{P(X, Y)}{P(Y)} \)

- \( P(Y | X) \) is prob. dist in \( Y \) for each fixed \( X \) ?

Two simple rewritings:

- \( P(X, Y) = \frac{P(X) \cdot P(Y | X)}{\text{eq. } \frac{P(X) \cdot P(Y | X)}{P(Y)}} = P(Y) \cdot P(X | Y) \)

  eg. \( X \) channel input, \( Y \) channel output

Bayes rule:

\[
P(X | Y) = \frac{P(X | Y) \cdot P(Y)}{P(X)} = \frac{P(X | Y) \cdot P(Y)}{\sum_{Y'} P(X | Y') \cdot P(Y')}
\]

eg. \( P(\text{pos} | \text{sick}) = P(\text{neg} | \text{healthy}) = 90\% \), \( P(\text{sick}) = 1\% \)

\[
\Rightarrow P(\text{sick} | \text{pos}) = \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.1 \times 0.99} = \frac{1}{12} < 10\% \quad \text{?}
\]
E.g. decoding the repetition code $R_3$: assume $S \sim \text{Uniform}\{0,1\}$. All shunts

$R_1 = S \oplus N_1 \ldots$, $R_2 = S \oplus N_2$ were $N_1, N_2, N_3 \sim \text{Bern}(f)$

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

\[
P(S|r) = \frac{P(r|S)P(S)}{P(r)} = \frac{1}{2}
\]

\[
\frac{P(S=0|r)}{P(S=1|r)} = \frac{P(r|S=0)}{P(r|S=1)} = \frac{P(r|T=000)}{P(r|T=111)} = \frac{1}{f} \text{ if } f \neq 0, \text{ else } \frac{1}{1-f}
\]

Combining independent RV's:

**Quiz:**

1. Let $S, K \sim \text{Uniform}\{0,1\}$, $T = S \oplus K$. Are $S$ and $T$ independent?

2. How to label two dice with numbers from $\{0,1,\ldots,5\}$ s.t. their sum $\sim \text{Uniform}\{11,12,\ldots,12\}$

Binomial$(n,p)$: Distribution of $Y = X_1 + \ldots + X_n$ were $X_i \sim \text{Bernoulli}(p)$

\[
\begin{align*}
\Pr(Y = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \frac{p^k (1-p)^{n-k}}{(1-p)^n}
\end{align*}
\]

\text{Expected number of bit flips when we send } n \text{ bits through}

"Numerical" random variables

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

**Expectation value (mean):** $EX = E(X) = \sum_{x} P(x) \cdot x$
**E[f(X)] = \sum_x P(x) \cdot f(x) \quad \text{"law of the unconscious statistician"}

\[ E[cX] = c \cdot E[X] \quad \text{&} \quad E[X+Y] = E[X] + E[Y] \]

* If \( X, Y \) independent: \( E[XY] = E[X] \cdot E[Y] \)

\( X \sim \text{Uniform(\{-1, 1\}), } Y = -X \Rightarrow E[XY] = -1 \), \( E[X^2] = E[Y] = 0 \)

**Variance:** \( \text{Var}(X) = E[(X - E[X])^2] \)

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

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

* If \( X, Y \) independent:

\[ \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \]

\( \therefore \) use that \( E[XY] = E[X] \cdot E[Y] \)

**Examples**

<table>
<thead>
<tr>
<th>( P )</th>
<th>Bernoulli (p)</th>
<th>Binomial (n \cdot p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E )</td>
<td>p</td>
<td>np</td>
</tr>
<tr>
<td>( \text{Var} )</td>
<td>p(1-p)</td>
<td>n \cdot p \cdot (1-p)</td>
</tr>
</tbody>
</table>

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

**Interpretation?**

**Markov inequality:** If \( X \geq 0 \):

\[ \Pr(X \geq t) \leq \frac{E[X]}{t} \quad (\forall t > 0) \]

**Pf:**

\[ \Pr(X \geq t) = \sum_{x \geq t} P(x) \leq \sum_{x \geq t} P(x) \cdot \frac{x}{t} \leq \frac{E[X]}{t} \]

**Chebyshev inequality:**

\[ \Pr(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} \]

**Pf:** Apply Markov to \( Y = (X - E[X])^2 \).

**Law of large numbers:** \( X_1, \ldots, X_n \sim P \) \( \text{with mean } \mu \), variance \( \sigma^2 \)

\[ \bar{X} := \frac{1}{n}(X_1 + \ldots + X_n) \]

\[ \Pr\left( |\bar{X} - \mu| \geq \varepsilon \right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \]

**Pf:** \( E\bar{X} = \mu \) & \( \text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}(X_1 + \ldots + X_n) = \frac{\sigma^2}{n} \). \( \Rightarrow \) Chebyshev.
One last reminder:

- \( f \) is called convex if \( p \cdot f(x) + (1-p) \cdot f(x') \geq f(px + (1-p)x') \) for all \( p \in [0,1] \).
- \( f \) is called concave if \( p \cdot f(x) + (1-p) \cdot f(x') \leq f(px + (1-p)x') \) for all \( p \in [0,1] \).

- \( \exp(x) \cdot x^2 \) is sufficient for convex: \( f'' \geq 0 \)
- \( \log(x) \) is sufficient for concave: \( f'' \leq 0 \)

- Strongly convex/concave if "=" only holds for \( p=0 \) or \( p=1 \).

Jensen's inequality: If \( f \) convex, then \( E[f(X)] \geq f(E[X]) \) is concave.

* If strongly convex/concave, then "=" if \( X \) is constant.

From next week on we will study:

**Entropy of \( X \sim P \):**

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

\[ \text{Always base 2} \]

* \( 0 \leq H(X) \leq \log (\text{#outcomes}) \)
  - if \( X \) constant
  - if \( X \) uniformly random

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

\[
H(X) = -p \cdot \log(p) - (1-p) \log(1-p)
\]

\[ H(\text{Bin}(1,p)) \]

\[ \text{Graph: } H(p) = -p \cdot \log(p) - (1-p) \cdot \log(1-p) \]