infinite sets

Bijection Rule: If $f: A \to B$ is a bijection, then $|A| = |B|$

Finite cardinality definitions:

- If A is a set and $n \in \mathbb{N}$, then $|A| = n$ means $|A| = |\{1, 2, 3 \ldots, n\}|$
- If $|A| = n$ for some n, then we say A is finite
- If $|A| = k$ and $|A| = \ell$ then $k = \ell$

Sum rule: If A and B disjoint then $|A \cup B| = |A| + |B|$

- WTS If $|A| = k$ and $|B| = \ell$ then $|A + B| = k + \ell$

product Rule: If A and B disjoint then $|A \times B| = |A \cdot B|$

- WTS If $|A| = k$ and $|B| = \ell$ then $|A \times B| = k\ell$

Quotient Rule: If R is an equivalence relation on A, and every equivalence class of R has k elements then $|A/R| = |A|/k$

i.e. $\forall a \in A$, $|[a]_R| = k$

2^n: number of subsets of set of size n

$n!$: number of permutations of A

$\binom{n}{k}$: number of k element subsets of A

$\sum_{k=0}^{n} expr$:
① choose k
② choose object that has "expr" diff ways

$k \cdot \ell$:
① do something one of k ways
② do something one of ℓ ways

Number of ways	Process
a	Choose an element of A
ab	Choose an element of A and then an element of B
$a + b$	Choose an element of A or an element of B
$\sum_i a_i$	Choose i and then an element of A_i
2^a	Choose a subset of A
$a!$	Choose a permutation of A
$a!/(a-k)!$	Choose a sequence of k different elements of A
$\binom{a}{k}$	Choose a subset of k elements of A
a/k	Choose an element of A/R where $\forall a, \lvert[a]\rvert = k$

proof: $\binom{n}{k} = \binom{n}{n-k}$

- let $|A|$ be set with n elements
- let $X = \{B \subseteq A \mid |B| = k\}$, so $|X| = \binom{n}{k}$
- let $Y = \{C \subseteq A \mid |C| = n-k\}$, so $|Y| = \binom{n}{n-k}$
- let $f: X \to Y$ given by $f(B) = A \setminus B$
    - then f is bijective
    - so LHS $= |X| = |Y| =$ RHS

proof: $2^n = \sum_{k=0}^{n} \binom{n}{k}$

- let $|A|$ be set with n elements
- let $X = 2^A$, so $|X| = \binom{n}{k}$
- let $Y = 2^A$, we construct an element of Y by first choosing $k \in [0 \dots n]$, then choosing subset $B \subseteq A$ of size k. So $|Y| = \sum_{k=0}^{n} \binom{n}{k}$
- so LHS $= |X| = |Y| =$ RHS

Exercise: $\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ ← fill remaining slots with elements of b

choose k

choose k slots in tuple

choose k elements of A

$$pr(X|Y) = \frac{pr(Y|X)\,pr(X)}{pr(Y)}$$ Baye's rule

$$pr(Y) = pr(Y|X)\,pr(X) + pr(Y|\text{not } X)\,pr(\text{not } X)$$

proof: $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$

- want to choose k objects out of $\{1, \dots, n\}$
  - $\binom{n-1}{k}$ ways of choosing k without choosing the last one
  - $\binom{n-1}{k-1}$ ways of choosing k including n i.e. choosing $k-1$ out of $\{1, \dots, n-1\}$

Conditional probability: If A and $B \subseteq S$ are events, $pr(B|A) := \frac{pr(B \cap A)}{pr(A)}$

Law of total probability: $pr(B) = \sum pr(B|A_i)\,pr(B_i)$

$$pr(B) = pr(B|A)\,pr(A)$$

Sample space: S is a set

outcome: an element $s \in S$

event: a subset $E \subseteq S$

Kolmogorov's axioms

probability measure: pr on a sample space S is a function $pr: 2^S \to \mathbb{R}$ satisfying:

1. $\forall E,\ pr(E) \geq 0$
2. $pr(S) = 1$
3. If E_1 and E_2 are disjoint, $pr(E_1 \cup E_2) = pr(E_1) + pr(E_2)$

probability space: (S, pr) where pr is a probability measure on S

proof: $\forall E \subseteq S,\ pr(E) \leq 1$

- $S = E \cup (S \setminus E)$ $\quad$ * $(S \setminus E) \cap E = \emptyset$
- $pr(S) = 1$
  $= pr(E \cup (S \setminus E))$
  $= pr(E) + pr(S \setminus E)$ $\quad$ * by axiom 3
- $pr(E) = 1 - pr(S \setminus E)$ $\quad$ * $pr(S \setminus E) \geq 0$ by axiom 1
  ≤ 1

random variable: a real-valued RV of X is a function $X: S \to \mathbb{R}$

In general, a T-valued RV of Z is a function $Z: S \to T$

- If $X: S \to \mathbb{R}$ and $Y: S \to \mathbb{R}$ then $X+Y: S \to \mathbb{R}$ is given by
$$\begin{cases} (X+Y)(s) := X(s) + Y(s) \\ (XY)(s) := (X(s))(Y(s)) \\ (\sin(X))(s) := \sin(X(s)) \\ (X/Y)(s) := (X(s))/(Y(s)) \end{cases}$$

- If $X: S \to \mathbb{R}$ and $x \in \mathbb{R}$ then $(X = x)$ is the event $\{s \mid X(s) = x\}$

"middle" of distribution

expected value: a weighted average of values of X, weighted by probabilities. $E(X) = \sum_{s \in S} X(s) \cdot pr(\{s\}) = \sum_{x \in \mathbb{R}} x \cdot pr(X = x)$

probability Mass Function of X (PMF): $PMF_X(x) := pr(X = x)$ $\quad PMF_X : \underset{\text{output}}{\mathbb{R}} \to \underset{\text{prob}}{\mathbb{R}}$

Jointed PMF: $PMF_{X,Y}(x,y) := pr((X=x) \cap (Y=y))$ $\quad PMF_{X,Y} : \underset{\text{outputs}}{\mathbb{R} \times \mathbb{R}} \to \underset{\text{prob}}{\mathbb{R}}$

s	$X(s)$	$Y(s)$	$pr(\{s\})$
a	1	1	10%
b	2	3	20%
c	2	2	30%
d	3	1	40%

x	$PMF_X(x)$
1	10%
2	50%
3	40%

y	$PMF_Y(y)$
1	50%
2	30%
3	20%

linearity of expectation : $E(X+Y) = E(X) + E(Y)$ and if $c \in \mathbb{R}$, then $E(cX) = c\,E(X)$

proof : $E(X+Y) = E(X) + E(Y)$

$$= \sum_{s \in S} (X+Y)(s) \cdot Pr(\{s\})$$

$$= \sum_{s \in S} \left(X(s) + Y(s)\right) \cdot Pr(\{s\})$$

$$= \sum_{s \in S} X(s) \cdot Pr(\{s\}) + \sum_{s \in S} Y(s) \cdot Pr(\{s\})$$

$$= E(X) + E(Y)$$

proof : $E(cX) = \sum_{s \in S} c(s) \cdot X(s) \cdot Pr(\{s\})$

$$= \sum_{s \in S} c\,X(s)\,Pr(\{s\})$$

$$= c \sum_{s \in S} X(s)\,Pr(\{s\})$$

$$= c\,E(X)$$

Exercise: what is the expected sum of rolling 4 six-sided dice?

$$\begin{cases} N_1(d_1, d_2, d_3, d_4) = d_1 \\ N_2(d_1, d_2, d_3, d_4) = d_2 \\ N_3(d_1, d_2, d_3, d_4) = d_3 \\ N_4(d_1, d_2, d_3, d_4) = d_4 \end{cases}$$

$$\Sigma = N_1 + N_2 + N_3 + N_4$$

$$E(\Sigma) = E(N_1) + E(N_2) + E(N_3) + E(N_4)$$

$$E(N_1) = 1 \cdot pr(N_1 = 1) + 2 \cdot pr(N_1 = 2) + pr(N_1 = 3) + \cdots$$

$$= \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

$$= (3.5) + (3.5) + (3.5) + (3.5)$$

$$= 14$$

Exercise: what is the expected number of 3s?

$$I_i(d_1, d_2, d_3, d_4) = \begin{cases} 1 & \text{if } d_i = 3 \\ 0 & \text{if } d_i \neq 3 \end{cases}$$

N = # of 3's rolled

$$= I_1 + I_2 + I_3 + I_4$$

$$E(N) = E(I_1) + E(I_2) + E(I_3) + E(I_4)$$

$$E(I_1) = 1 \cdot pr(I_i = 1) + 0 \cdot pr(I_i = 0)$$

$$= \frac{1}{6}$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

$$= \frac{2}{3}$$

Indicator variable: of an event A is 1 for $s \in A$
0 for $s \notin A$

E_1 does not change E_2

Independent: events E_1 and E_2 are independent if $pr(E_2 | E_1) = pr(E_2)$

i.e. $pr(E_1 \cap E_2) = pr(E_1)\, pr(E_2)$

independent random variable: RV X and Y are independent if

- $E(XY) = E(X)E(Y)$
- $\forall x, y \in \mathbb{R}$, the events $(X=x)$ and $(Y=y)$ are independent
- $pr(X|Y) = pr(X)$

proof: **If X and Y are independent, then $E(XY) = E(X)E(Y)$**

- LHS : $E(XY) = \sum_{z \in \mathbb{R}} z \cdot pr(XY = z)$
- RHS : $E(X)\,E(Y) = \left(\sum_{x \in \mathbb{R}} x \cdot pr(X = x)\right)\left(\sum_{y \in \mathbb{R}} y \cdot pr(Y = y)\right)$

$$= \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} x \cdot y \cdot Pr(X = x \cap Y = y)$$

- since X and Y are independent, $Pr(X = x \cap Y = y) = Pr(X = x)\, pr(Y = y)$,

  LHS = RHS

how "spread out" distribution is

Variance: $var(X) := E((X - E(X))^2)$

proof: **alternate definition of variance is $E(X^2) - (E(X))^2$**

- $E((X - E(X))^2) = E(X^2 - 2X\,E(X) + (E(X))^2)$

$$= E(X^2) - 2E(X \overset{\text{constant}}{E(X)}) + E(\overset{\text{constant}}{E(X)^2})$$

$$= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 \qquad \text{by linearity}$$

$$= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2$$

$$= E(X^2) - 2(E(X))^2 + (E(X))^2$$

$$= E(X^2) - E(X)^2$$

proof : If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$

- $$\begin{aligned} Var(cX) &= E((cX - E(cX))^2) \\ &= E((cX - cE(X))^2) \quad \text{by linearity} \\ &= E(c^2(X - E(X))^2) \\ &= c^2 E((X - E(X))^2) \\ &= c^2 Var(X) \end{aligned}$$

Markov's inequality : If $x \geq 0$, then $Pr(X > a) \leq \frac{E(X)}{a}$

Chebychev's inequality : $\forall X$, $Pr(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$

Weak law of large numbers : If $X_1, X_2, \dots X_n$ are independent RV satisfying $E(X_i) = \mu$ and $Var(X_i)$

$$Pr\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq a\right) \leq \frac{\sigma^2}{na^2}$$