Recall: a <u>probability measure</u> on $\mathcal{U}$ is $p : \mathcal{P}(\mathcal{U}) \to \mathbb{R}$ such that

(i) $p(A) \geq 0 \quad \forall A \subseteq \mathcal{U}$

(ii) $p(\mathcal{U}) = 1$

(iii) $p\left(\dot{\bigcup}_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i)$, $A_i \cap A_j = \emptyset$ for $i \neq j$

<u>Remarks</u>

$p(\bar{A}) = 1 - p(A)$, since $\mathcal{U} = A \,\dot{\cup}\, \bar{A}$

$p(\emptyset) = 0$, since $\mathcal{U} = \mathcal{U} \,\dot{\cup}\, \emptyset$

$A \subseteq B \Rightarrow p(A) \leq p(B)$, since $B = (B - A) \,\dot{\cup}\, A$

$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$$p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(A \cap B) - p(A \cap C) - p(B \cap C) + p(A \cap B \cap C)$$

etc..

<u>Examples</u>

Since the probability of an outcome is the ratio of the measure of the (weighted) favourable set to the measure of the (weighted) universe of possibilities,

- Flip coin twice, then
  $$\mathcal{U} = \{(h,h), (h,t), (t,h), (t,t)\}$$
  and $p(\geq \text{head}) = 3/4$.
- Toss 3 dice, then $p(\text{all even}) = 3^3/216 = 27/216$
  and $p(\text{at least one odd}) = 1 - 27/216$
- Pick one card from a pack of 52,
  then
  $$p(\text{ace or spade}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

Useful Formulae

- \# permutations of n different objects $= |S_n| = n!$
- \# ways of choosing k things from n with replacement $= n^k$
- \# " " " " " " " without replacement

$$= \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!\,(n-k)!}$$

This is often denoted ${}^nC_k$ or $\binom{n}{k}$. Notice that numerically $\binom{n}{k} = \binom{n}{n-k}$

We could also think of this as a permutation of the n objects with the k to be chosen occupying the first k slots and the $(n-k)$ rejected being in the last slots. Hence this is the # permutations of n objects comprising 2 types (the 'chosen' and the 'others') with k of type 1 and $r = n-k$ of type 2. This could be notated

$$\binom{n}{k \;\; r} := \frac{n!}{k!\;r!} = \binom{n}{k} = \binom{n}{r}$$

- Since determining asymptotic behaviour of factorials gets messy, a useful approximation is

$$n! \sim \sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}} \quad \text{(Stirling)}$$

- \# permutations of $n = n_1 + \dots + n_t$ objects comprising t different types is $\binom{n}{n_1 \dots n_t} = \dfrac{n!}{n_1! \dots n_t!}$

e.g. # permutations of $\{a, a, a, b, b, c\}$ is $\dfrac{6!}{3!\,2!\,1!} = 60$

- $(x+y)^n = \sum_0^n \binom{n}{r} x^r y^{n-r}$
- $(x_1 + \dots + x_t)^n = \sum \binom{n}{r_1 \dots r_t} x_1^{r_1} \dots x_t^{r_t}$ with the sum taken over all $r_i \geq 0$ with $r_1 + \dots + r_t = n$

Examples

- Pond has 50 fish of which 10 are tagged, then 9 fish are netted.

$$p(\text{exactly 2 netted are tagged}) = \frac{\binom{10}{2}\binom{40}{7}}{\binom{50}{9}}$$

- Now the pond has 10 A-fish, 15 B-fish, 20 C-fish, and 5 others. Again net 9 fish.

$$p\left(\text{exactly } \begin{matrix} 2 \text{ A-fish} \\ 3 \text{ B-fish} \\ 2 \text{ C-fish} \end{matrix}\right) = \frac{\binom{10}{2}\binom{15}{3}\binom{20}{2}\binom{5}{2}}{\binom{50}{9}}$$

- Can also build a probability list of events, e.g., values of sum of 2 dice ...

2	1/36	6	5/36	10	3/36
3	2/36	7	6/36	11	2/36
4	3/36	8	5/36	12	1/36
5	4/36	9	4/36		

So the sums are not equi-probable, yet we can use this data to weight our events ...

$$p(\text{get a prime sum}) = p(\text{get } 2, 3, 5, 7, \text{ or } 11)$$
$$= \frac{1}{36} + \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{2}{36} = \frac{15}{36}$$

Conditional Probability

If A and B are independent events then $p(A \cap B) = p(A)p(B)$. Now let $p(A \mid B)$ denote the probability of A given B, then with A and B still independent we get

$$p(A \cap B) = p(A \mid B)\, p(B).$$

This inspires a definition useful in dependent situations (provided $p(B) \neq 0$) of

$$p(A \mid B) := \frac{p(A \cap B)}{p(B)}.$$

Examples

- Bucket of 10 things has 3 defectives. Examining things successively, find p(last defective is 5^{th} thing examined).

  Let A = "5^{th} thing examined is defective"

  B = "exactly 2 defectives found in 1^{st} 4 examined"

  We want $p(A \wedge B)$.

  $$p(B) = \frac{\binom{3}{2}\binom{7}{2}}{\binom{10}{4}} \quad \text{and} \quad p(A|B) = \frac{1}{6}$$

  only 1 defective left!

  Hence $p(A \wedge B) = p(A|B)\,p(B) = \frac{1}{6}\,\frac{\binom{3}{2}\binom{7}{2}}{\binom{10}{4}}$.

- Have 2 buckets ... α has 5 blue and 4 white balls

  β " 4 " " 5 " "

  Now transfer 1 ball at random from α to β.

  Find p (now pick blue ball from β).

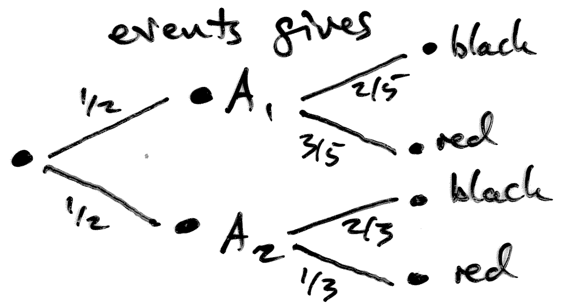
  Let αB = "pick blue from α", αW = "pick white from α",

  βB = " " " " β", βW = " " " " β".

  Then
  $$p(\beta B) = p(\beta B|\alpha B)\,p(\alpha B) + p(\beta B|\alpha W)\,p(\alpha W)$$
  $$= \tfrac{5}{10}\cdot\tfrac{5}{9} + \tfrac{4}{10}\cdot\tfrac{4}{9}$$

- Have buckets A_1, A_2 with A_1 having 2 black and 3 red balls, and A_2 having 2 black and 1 red ball. Choose bucket at random then pick ball at random, then find p(red ball chosen).

  A first intuition might be $p(\text{red}) = \frac{3+1}{8} = \frac{1}{2}$ however this would be wrong. Diagramming the events gives



  $$\Rightarrow p(\text{red}) = \tfrac{1}{2}\cdot\tfrac{3}{5} + \tfrac{1}{2}\cdot\tfrac{1}{3} = \underline{\underline{\tfrac{7}{15}}}$$

So far, we've looked at 'a priori' probabilities; reasoning about probabilities of subsequent events given data about earlier ones. However, sometimes we want to argue in reverse; reasoning 'a posteriori' about earlier events given data about later ones.

Suppose $\mathcal{U} = \bigcup_{i=1}^{r} A_i$ and B any event with $p(B) > 0$.

Then $B = \bigcup_{i=1}^{r} (B \cap A_i)$ and

$$p(B) = \sum_1^r p(B \cap A_i) = \sum_1^r p(B \mid A_i)\, p(A_i)$$

Now, for each j, $p(A_j \mid B) = \dfrac{p(A_j \cap B)}{p(B)}$, so then

$$p(A_j \mid B) = \frac{p(B \mid A_j)\, p(A_j)}{\sum_{i=1}^{r} p(B \mid A_i)\, p(A_i)}$$

This is Bayes' formula, and the study of such things is often called Baysian probability.

Examples

- Have buckets A_1 with 2 red balls and 4 white balls
  - A_2 — 1 — — — 2 — —
  - A_3 — 5 — — — 4 — — .

  Suppose $p(A_1) = 1/3$, $p(A_2) = 1/6$, and $p(A_3) = 1/2$.
  Event: Pick a box then pick a ball.

  $$p(\text{red}) = p(\text{red} \mid A_1)\, p(A_1) + p(\text{red} \mid A_2)\, p(A_2) + p(\text{red} \mid A_3)\, p(A_3)$$
  $$= 2/6 \cdot 1/3 + 1/3 \cdot 1/6 + 5/9 \cdot 1/2 = 4/9 .$$

  Now suppose the outcome is picking a red ball. Given that, what is $p(A_1)$?

  $$p(A_1 \mid \text{red}) = \frac{p(A_1 \cap \text{red})}{p(\text{red})} = \frac{p(\text{red} \mid A_1)\, p(A_1)}{p(\text{red})} = \frac{2/6 \cdot 1/3}{4/9} = \frac{1}{4}.$$

- Have machines A_1, A_2, A_3 producing products with 2%, 1%, 3% defectives respectively. Given that A_1, A_2, A_3 supply 35%, 25%, 40% respectively of the total output,

$$p(\text{defective selected came from } A_3) = p(A_3 \mid \text{defective})$$
$$= \frac{p(\text{defective} \mid A_3)\, p(A_3)}{p(\text{defective})} = \frac{\frac{3}{100}\cdot\frac{40}{100}}{\frac{215}{10000}} = \frac{120}{215}.$$

<u>Definitions</u>

The set of all possible outcomes is the <u>samplespace</u>. Sometimes we use "sample space" to denote (by abuse of notation) the set of events <u>caused</u> by the possible outcomes.

A <u>random variable</u> is a function defined on a sample space.

<u>Examples</u>

- Roll a pair of dice.
  The set $S = \{(1,1), (1,2), \ldots, (6,6)\}$ of all 36 pairs of numbers is the sample space, and a <u>multivariate</u> random variable X could be defined on S by
  $$X(a,b) := (a,b).$$
  Given the same S, we could define a <u>single-variate</u> random variable Y on S by
  $$Y(a,b) := a+b.$$
  We might even choose to call $S' = \{2, 3, \ldots, 12\}$ of all 11 values of Y a sample space (by common abuse of notation).
- Consider a bucket of n things, where n_1 are red and $n_2 = n - n_1$ are white.
  Event: select (without replacement) r things.
  Here $S = \{(r,0), (r-1,1), \ldots, (0,r)\}$ where (a,b) means that a selected things are red and b are white.

Now define X on S by $X(a,b) := a$, so that we count only the number of red things selected.
Then
$$p(X = x) = \frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n}{r}}.$$

This gives rise to the <u>hypergeometric distribution</u>.

<u>Definition</u>

For a random variable X (discrete), the function
$$f(x) := p(X = x)$$
is the <u>probability density function</u> (or p.d.f.) of X.

<u>Examples</u>

- Toss a coin until the 1st head appears, then $S = \{1, 2, 3, \ldots\}$, and let X be defined on S by $X(a) := a$. Then
$$p(X=1) = \tfrac{1}{2}, \quad p(X=2) = \left(\tfrac{1}{2}\right)^2, \ldots$$
so
$$f(x) = \left(\tfrac{1}{2}\right)^x$$
is the p.d.f. for X. Now let
$$F(x) := p(X \le a) = \sum_{x \le a} f(x)$$
$$= \tfrac{1}{2} + \left(\tfrac{1}{2}\right)^2 + \cdots + \left(\tfrac{1}{2}\right)^x = 1 - \left(\tfrac{1}{2}\right)^x.$$

This 'cumulative' function is called the <u>probability distribution function</u> for X, and in this example is a single-variate distribution.
Notice that
$$p(a \neq X \le b) = \sum_{x \ngeq a}^{x \le b} f(x) = F(b) - F(a).$$

- A die has 3 faces with 1, 2 with 2 and 1 with 3.
Event: Roll die 5 times recording

$$x_1 = \#1\text{'s}, \quad x_2 = \#2\text{'s}, \quad x_3 = \#3\text{'s}.$$

Then $S = \{(5,0,0), (4,1,0), (4,0,1), \ldots, (0,0,5)\}$.
Now let X be defined on S by $X(a,b,c) := (a,b,c)$, then $f(a,b,c) = p(X = (a,b,c))$ is a p.d.f. for X and

$$f(a,b,c) = \binom{5}{a\ b\ c} \left(\tfrac{3}{6}\right)^a \left(\tfrac{2}{6}\right)^b \left(\tfrac{1}{6}\right)^c.$$

Actually, c is determined by a and b, $c = 5-b-a$, so

$$f(a,b) = \frac{5!}{a!\,b!\,(5-a-b)!} \left(\tfrac{1}{2}\right)^a \left(\tfrac{1}{3}\right)^b \left(\tfrac{1}{6}\right)^{5-a-b}$$

and

$$F(a,b) = \sum_{x \le a} \sum_{y \le b} f(x,y)$$

is the corresponding multivariate distribution function.

- Now let A be an event for an experiment with

$$p(A) = p, \quad \text{so } p(\bar{A}) = q = 1-p.$$

Repeating the experiment n times (independently) and recording $x = \#A$'s gives

$$S = \{0, 1, \ldots, n\}, \quad \text{and we let } X(a) := a.$$

Now if α is any particular sequence of A's and $\bar{A}$'s then $p(\alpha) = p^x q^{n-x}$ if α has x A's. Hence

$$p(\text{exactly } x \text{ } A\text{'s in } n \text{ repetitions}) = \binom{n}{x} p^x q^{n-x}.$$

Notice that

$$\sum_0^n f(x) =: F(n)$$
$$= q^n + npq^{n-1} + \binom{n}{2} p^2 q^{n-2} + \ldots + p^n$$
$$= (q+p)^n = ((1-p)+p)^n = 1$$

hence $F(x)$ is called the <u>binomial distribution</u> and such an experiment is a <u>Bernoulli Experiment</u>.

If $n=1$ this is called a Bernoulli distribution.

(Investigating what happens to the ratio x/n as $n \to \infty$ leads to "the law of large numbers", a standard statistical technique.)

E.g. In 'bridge', a 'hand' is the receipt of 13 cards, so

$$p(\text{no aces in hand}) = \frac{\binom{48}{13}}{\binom{52}{13}} \cong 0.304$$

Assuming perfect shuffling each time, repeat 5 times, then

$$p\left(\begin{matrix}\text{no aces in } x \\ \text{of 5 'hands'}\end{matrix}\right) = \binom{5}{x}(0.304)^x(0.696)^{5-x}.$$

If n is large, using Stirling's formula to approximate $r!$ can help the computation.

This can all be generalised to a multinomial distribution in an obvious way ...

Assume a Bernoulli experiment with $A_1, \dots, A_r$ as all possible events (exhaustive and exclusive) with probabilities $p_1, \dots, p_r$.

Now repeat the experiment n times, then

$$S = \{(x_1, \dots, x_r) \mid \Sigma x_i = n\}$$

is the sample space, and we define X on S by $X(a_1, \dots, a_r) := (a_1, \dots, a_r)$. Then

$$p(X = (x_1, \dots, x_r)) = \binom{n}{x_1 \ \dots \ x_r} p_1^{x_1} \cdots p_r^{x_r}.$$

Again we have

$$F(n) = \Sigma f(x_1, \dots, x_r)$$
$$= (p_1 + \dots + p_r)^n = 1$$

hence the moniker "multinomial" (again there are really only $(r-1)$ variables involved).

- With the Bernoulli experiment used for the binomial distribution, this time let

$$X = \# \bar{A}\text{'s before the } 1^{st} A$$

i.e., the number of failures before the first success. Then

$$p(X = x) = (1-p)^x p = f(x).$$

Notice that

$$F(a) = \sum_{x \le a} f(x) = \sum_{x \le a} (1-p)^x p = p \sum_{x \le a} (1-p)^x,$$

hence this is called the <u>geometric distribution</u>.

Notice also that

$$F(\infty) = p \sum_0^{\infty} (1-p)^x = p \frac{1}{1-(1-p)} = 1,$$

so

$$p(X > a) = \sum_{a+1}^{\infty} (1-p)^x p = p(1-p)^{a+1} \sum_0^{\infty} (1-p)^x$$

$$= (1-p)^{a+1}$$

and

$$p(X \le a) = F(a) = 1 - (1-p)^{a+1}$$

Eg. Roll a regular die and let A = "4 displayed".

$$p(1^{st}\ 4 \text{ seen on } 6^{th} \text{ trial}) = p(X=5)$$

$$= (5/6)^x (1/6) \cong 0.067$$

$$p(X \geq 5) = p(X > 4) = (1 - 1/6)^5 \cong 0.402$$

$$p(X \le 4) = p(X < 5) \cong 1 - 0.402 = 0.598$$

- This time, repeat the Bernoulli trials until have exactly r A's (successes), then if $X = \# \bar{A}$'s (failures) then $x + r$ = # trials to get precisely r successes. So

$$p\begin{pmatrix}\text{exactly } r\ A\text{'s} \\ \text{takes } r+x \text{ tries}\end{pmatrix} = \underbrace{\binom{r+x-1}{r-1} p^{r-1} (1-p)^x}_{\text{get } r-1\ A\text{'s and } x\ \bar{A}\text{'s in } (r+x-1) \text{ trials}} \underbrace{p}_{\text{succeed in last trial}}$$

$$= \binom{r+x-1}{r-1} p^r q^x = f(x), \text{ note } r \text{ constant!}$$