Given Ω is chosen, a probability law on Ω is a mapping $\mathbb{P}$ that assigns a number to every event (i.e. to every subset of Ω) such that:

1) $\mathbb{P}(A) \geq 0$ for every event A
2) $\mathbb{P}(\Omega) = 1$ (normalization)
3) Additivity Rules
   i) If $A \cap B = \emptyset$ (A, B are events) then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
   ii) If $A_1, A_2, A_3, \ldots$ is a countable sequence of mutually disjoint events (i.e. $A_i \cap A_j = \emptyset \ \forall i, j$), then $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

$\bigcup_{n=1}^{\infty} A_n$ — union of all of them

So, given an event $A \subset \Omega$, $\mathbb{P}(A)$ is a model for "the likelihood that the outcome of the uncertain experiment is in A"

"Event A occurs" means "outcome of experiment is in A"

Product Rule: If $D_1 \supset D_2 \supset D_3 \cdots \supset D_n$ is a nested decreasing sequence of positive-probability events, then

$$\mathbb{P}(D_n) = \mathbb{P}(D_1)\mathbb{P}(D_2 | D_1)\mathbb{P}(D_3 | D_2) \cdots \mathbb{P}(D_n | D_{n-1})$$

If $A_1, A_2, \ldots, A_n$ is a partition of Ω, then $C_1, C_2, \ldots, C_n$ partitions B, where

$$C_k = B \cap A_k \quad 1 \leq k \leq n$$

Idea: $B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n)$
$= (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$

By additivity,

LAW OF TOTAL PROBABILITY

$\mathbb{P}(B) = \mathbb{P}(C_1) + \cdots + \mathbb{P}(C_n)$
$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \cdots + \mathbb{P}(B \cap A_n)$

$$\boxed{\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n)\mathbb{P}(A_n)}$$

Given Ω, $\mathbb{P}$, if $A_1, \ldots, A_n$ are events that partition Ω and have nonzero probability, then for any event B, and for any $k : 1 \leq k \leq n$,

$$\mathbb{P}(A_k | B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\mathbb{P}(B|A_1)\mathbb{P}(A_1) + \cdots + \mathbb{P}(B|A_n)\mathbb{P}(A_n)}$$

"PROOF:" Numerator $= \mathbb{P}(B \cap A_k)$ by definition of conditional probability

Denominator $= \mathbb{P}(B)$ by total probability theorem

$$\frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A_k | B)!$$

Next Circle of Ideas: Independence

Given, Ω, $\mathbb{P}$. Say two events, $A, B \subset \Omega$ are independent when $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Same as $\mathbb{P}(A|B) = \mathbb{P}(A)$ when $\mathbb{P}(B) > 0$ — Knowing B occurred says nothing more about whether A occurred
$\mathbb{P}(B|A) = \mathbb{P}(B)$ when $\mathbb{P}(A) > 0$

Knowing that A occurred says nothing more about whether B occurred.

CAUTION: Independence is "about" not only the events and how they sit in Ω, but also (more crucially) about $\mathbb{P}$!!

Common error: if "$A \cap B = \emptyset$, then A and B are independent"

WRONG whenever $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$

Conditional Independence: $\Omega, \mathbb{P}$; say events A and B are conditionally independent given (event) C when

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

$\mathbb{P}(A | B \cap C) = \mathbb{P}(A|C)$; $\mathbb{P}(B | A \cap C) = \mathbb{P}(B|C)$

Knowledge of B gives no function info about probability of A on top of knowledge of C

To see this just play w/ formulas

$$\mathbb{P}(A | B \cap C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \frac{\mathbb{P}(A \cap B | C)\mathbb{P}(B|C)}{\mathbb{P}(B|C)} = \mathbb{P}(A|C)$$

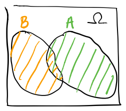
Given a discrete r.v X w/ $p_X(x)$ pmf, define the expected value (or expectation)

$$\mathbb{E}(X) = \sum_{x \in X} x \, p_X(x)$$

Given: Ω and $\mathbb{P}$ and two events $A, B \subset \Omega$, define $\mathbb{P}(A|B)$ = "Probability of A given B"

Idea: Given event B occurs, what's the likelihood that A occurs?

Knowledge that B occurs is gonna effect the likelihood of A's occurrence.



Intuitively: $A|B$ is "fraction of B's $\mathbb{P}$-mass that also lies in A"

Motivates the definition

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Undefined for $\mathbb{P}(B) = 0$

Observation: Given $B \subset \Omega$ w/ $\mathbb{P}(B) > 0$, as A runs over all events, $\mathbb{P}(A|B)$ defines a new probability law on Ω

- $\mathbb{P}(A|B) \geq 0 \quad \forall A \subset \Omega$
- $\mathbb{P}(\Omega) = 1$ (by 2nd reality check above)
- If $A_1 \cap A_2 = \emptyset$, then

disjoint

$(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B)$

divide both sides by $\mathbb{P}(B)$! $\Rightarrow \mathbb{P}((A_1 \cup A_2) \cap B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B)$

$$\mathbb{P}((A_1 \cup A_2) | B) = \mathbb{P}(A_1 | B) + \mathbb{P}(A_2 | B)$$

Counting Principle:

Re-count # of subsets of $\Omega = \{s_1, \ldots, s_n\}$.

We can view each subset as arising from a multi stage building process.

Stage 1: Choose whether to put s_1 in the subset - either yes or no $(n_1 = 2)$

Stage 2: Choose whether to put s_2 in the subset - either yes or no $(n_2 = 2)$

⋮

Stage n: Choose whether to put s_n in the subset - either yes or no $(n_n = 2)$

$$\#(\text{subsets}) = n_1 n_2 \cdots n_n = 2^n$$

"n choose k"

$$\binom{n}{k} := \frac{n(n-1)\cdots(n-k+1)}{k!}$$

Comment: Amazing that this is an integer!
Can also write

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This comes up in situations involving independent trials

Idea: Perform a random experiment repeatedly + independently

If experiment has 2 outcomes, call them Bernoulli Trials. Generally visualize coin flips in this case - H, T possible outcomes at each stage - if $\mathbb{P}(\{H\}) = p$, $\mathbb{P}(\{T\}) = 1 - p$ and you perform experiment n times, for each k, $0 \leq k \leq n$,

$$\mathbb{P}(\text{sequence you get}) = p^k (1-p)^{n-k}, \text{ where } k = \#(H) \text{ you flip}$$

This is true REGARDLESS of the order each H, T appears in the sequence of flips.

$\mathbb{P}(\text{you get } k \text{ H's}) = \#(\text{ways of picking } k \text{ spots in } n \text{ available spots})$

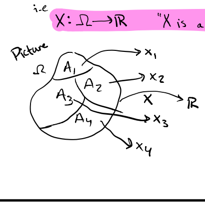
$$= \binom{n}{k} p^k (1-p)^{n-k} \quad 0 \leq k \leq n$$

Next BIG TOPIC: DISCRETE RANDOM VARIABLES

Start with Ω and $\mathbb{P}$: a discrete random variable (r.v.) is a real valued function with domain Ω that takes on only finite or countably infinite number of different values.

i.e. $X : \Omega \to \mathbb{R}$ "X is a mapping from Ω to real's"

Picture



i.e. $X(s) = x_1 \ \forall s \in A_1$, etc

Variance of X

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

$$\sigma_X = +\sqrt{\mathrm{Var}(X)} \quad \leftarrow \text{Standard deviation}$$

p_X is defined as follows: for every possible value of $x \in X$,

$$p_X(x) = \mathbb{P}(A_x) \text{ where } A_x = \{s \in \Omega : X(s) = x\}$$

i.e. $p_X(x)$ = the probability that the r.v. X takes on the specific value x

Book uses $\mathbb{P}(\{X = x\})$ or $\mathbb{P}(X = x)$ ← misuse of notation compared to above

to refer to $\mathbb{P}(A_x)$, where $A_x = \{s \in \Omega : X(s) = x\}$

Things to note about p_X:
- $p_X(x) \geq 0$ for all possible values of X (why? cause for any x, $p_X(x)$ is $\mathbb{P}(\text{an event}) \geq 0$!)
- If V is any finite or countably infinite set of possible values of X, then if we set
  B = the event "$X \in V$"
  i.e. $B = \{s \in \Omega : X(s) \in V\}$

then

$$\mathbb{P}(B) = \sum_{x \in V} p_X(x)$$

21:44

Binomial RV

Given positive integer n for some $p \in [0, 1]$, the Binomial (n, p) pmf defined as follows:

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & 0 \leq k \leq n \\ 0, & \text{all other } k \end{cases}$$

This could arise from
Ω = {all H-T sequences of length n}
$\mathbb{P}(\text{any sequence}) = p^{\#H}(1-p)^{\#T}$

like the outcome of n-independent flips of a possibly unfair coin.

$X(\text{any sequence}) = \#(\text{Heads})$ in sequence

Geometric Random Variable

Geometric(p):

$$p_X(k) = p(1-p)^{k-1}, \quad 1 \leq k < \infty$$

A possible $\Omega, \mathbb{P}$? You have a coin where $\mathbb{P}(H) = p$; Ω = set of all semi-infinite $(1 \leq n < \infty)$ of H, T; $X(s)$ = index of first head for all $s \in \Omega$

Reality check: recall $\sum_{x \in X} p_X(x) = 1$ for any discrete r.v X.

Verify this for geometric r.v.'s.

$$\sum_{x \in X} p_X(k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{m=0}^{\infty} (1-p)^m = p \frac{1}{1-(1-p)} = 1 \checkmark$$

Next, multiple discrete rvs and joint pmfs, etc
Given $\Omega, \mathbb{P}$ and two discrete rvs X, Y defined in Ω, define the joint pmf of X and Y via

$$p_{X,Y}(x, y) = \mathbb{P}(\underbrace{\{X = x\}}_{A_x} \cap \underbrace{\{Y = y\}}_{B_y})$$

$\forall x \in X, y \in Y$.

Note: for any set V of possible value pairs for X, Y, we have

$$\sum_{x,y \in V} p_{X,Y}(x, y) = \mathbb{P}(\text{event that } (X, Y) \in V)$$

Since X, Y are discrete rvs, they have their own pmfs p_X, p_Y. These are determined as follows from the joint pmf $p_{X,Y}(x, y)$:

① $p_X(x) = \sum_y p_{X,Y}(x, y) \quad \forall x$

② $p_Y(y) = \sum_x p_{X,Y}(x, y) \quad \forall y$

Why are these true?

$$p_X(x) = \mathbb{P}(\{X = x\}) = \mathbb{P}(A_x) = \underbrace{\sum_y \mathbb{P}(A_x \cap B_y)}_{\text{by Total Probability Thm}} = p_{X,Y}(x, y) \ \forall x$$

These definitions generalize in an obvious way to > 2 rv's defined on same $\Omega, \mathbb{P}$.

KEY THING:
- joint pmf determines the marginals
- marginals do **NOT** determine the joint

Recall the expected value rule: Given X, $p_X(x)$, $Y = g(X)$, have

$$\mathbb{E}(Y) = \sum_x g(x) p_X(x)$$

← good bc saves us from computing $p_Y(y)$ first before $\mathbb{E}(Y)$

Similarly, given X, Y w/ joint pmf $p_{X,Y}(x, y)$ and some real valued function $Z = g(X, Y)$, we have

$$\mathbb{E}(Z) = \sum_{x \in X} \sum_{y \in Y} g(x, y) p_{X,Y}(x, y)$$

← don't need $p_Z(z)$ to get $\mathbb{E}(Z)$

Next, **Conditional Stuff**

Given $\Omega, \mathbb{P}$ and a discrete rv X defined on Ω and an event $A \subset \Omega$ w/ $\mathbb{P}(A) > 0$, and a possible value x for X, the conditional pmf of X given A is defined as

$$p_{X|A}(x) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)} = \text{"}\mathbb{P}(B|A) \text{ where } B \text{ is the event } \{X = x\}\text{"}$$

Observe that for any A w/ $\mathbb{P}(A) > 0$, $p_{X|A}(x)$ as x ranges over X's value space defines a pmf - i.e. $p_{X|A}(x) \geq 0 \ \forall x$ and $\sum_{x \in X} p_{X|A}(x) = 1$

Here's a fact that's similar to (and follows directly from) the Total Probability Thm: If events $A_1, A_2, \ldots, A_n$ partition Ω, and $\mathbb{P}(A_k) > 0$ for $1 \leq k \leq n$, then for any discrete rv X on Ω,

$$p_X(x) = \sum_{k=1}^{n} p_{X|A_k} \mathbb{P}(A_k)$$

More often, encounter conditional pmf of X given some other rv Y (defined on same $\Omega, \mathbb{P}$) - given X, Y defined on $\Omega, \mathbb{P}$, conditional pmf of X givend Y is defined for all x and for all y with $\mathbb{P}(\{Y = y\}) = p_Y(y) > 0$ as

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

← Same as $p_{X|A}(x)$ where $A = \{Y = y\}$

Standard Notation

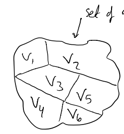
Note that for any y w/ $p_Y(y) > 0$, $p_{X|Y}(x|y)$ as x ranges over X values defines a pmf

i.e. $p_{X|Y}(x|y) \geq 0$ and $\sum_{x \in X} p_{X|Y}(x|y) = 1$

Given X, p_X, and $Y = g(X)$,

$$\mathbb{E}(Y) = \sum_{x \in X} g(x) p_X(x)$$

Why?



$y_1, \ldots, y_6$ are possible Y value

V_k = set of x-values for which $g(x) = y_k$

$$p_Y(y_k) = \sum_{x \in V_k} p_X(x)$$

Thus,

$$\mathbb{E}(Y) = \sum_{y_k} y_k p_Y(y_k)$$
$$= \sum_{y_k} y_k \sum_{x \in V_k} p_X(x)$$

← for all $x \in V_k$, $g(x) = y_k$ (for $1 \leq k \leq 6$)

$$= \sum_{y_k} \sum_{x \in V_k} g(x) p_X(x)$$
$$= \sum_{x \in X} g(x) p_X(x)$$

Total probability rule: If $A_1, \ldots, A_n$ partition Ω, then

$$p_X(x) = \sum_{k=1}^{n} p_{X|A_k}(x) \mathbb{P}(A_k)$$

Given two rvs on same $\Omega, \mathbb{P}$ - say X and Y - define

$$p_{X|Y}(x|y) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(\{Y = y\})} = \frac{p_{X,Y}(x, y)}{p_Y(y)}; \quad \forall x, y \text{ w/ } p_Y(y) > 0$$

For fixed y, $p_{X|Y}(x|y)$ defines a pmf "over x-values" - i.e.

$$p_{X|Y}(x|y) \geq 0 \ \forall x \text{ and } \sum_{x \in X} p_{X|Y}(x|y) = 1$$

"$p_{X|Y}(x|y)$ = conditional pmf of X given $Y = y$"

Like the conditional "event-centered story", have a product rule of sorts

$$p_{X,Y}(x, y) = p_Y(y) p_{X|Y}(x|y) \quad \forall x \in X, y \in Y$$

OR

$$p_{X,Y}(x, y) = p_X(x) p_{Y|X}(y|x) \quad \forall x \in X, y \in Y$$

This expresses joint in terms of marginals + conditional(s).
Also have a total-probability rule of sorts:

$$p_X(x) = \sum_{y \in Y} p_Y(y) p_{X|Y}(x|y)$$

$$p_Y(y) = \sum_{x \in X} p_X(x) p_{Y|X}(y|x)$$

We know who we are yet we know not what we may be