A (very) brief introduction to probability

Purpose: Build models of experiments with random outcomes & analyze these models.

By random outcomes we mean anything that we cannot predict with certainty.
E.g. roll dice, toss coins, throw darts, etc.

Ingredients of a probability model

"Defn" of probability space

- Sample space – Ω
  Set of all possible outcomes
  Elements of Ω are called sample points
- The set of events $\mathcal{F}$
  An event is a subset of Ω.
  E.g. roll a die $\Omega = \{1,2,3,4,5,6\}$
  The subset $A = \{2,4,6\}$, $A \subseteq \Omega$ is the event the roll is even.
  The event "not a 6" is $B = \{1,2,3,4,5\} \subseteq \Omega$
  $\mathcal{F} = \{A, B, \dots\}$
- Probability measure (probability distribution)
  P a function $: \mathcal{F} \longrightarrow \mathbb{R}$
  that associates with every event A a real number $P(A)$ called its probability.
  P must satisfy the following properties
  i) $0 \le P(A) \le 1 \quad \forall A \in \mathcal{F}$
  ii) $P(\Omega) = 1$ & $P(\emptyset) = 0$
  iii) For any sequence of pairwise disjoint events $A_1, A_2, \dots$ we have $P(\bigcup_i A_i) = \sum_i P(A_i)$.

If A, B not disjoint, can write $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Generalizes to $A_1, \dots, A_n$ as well.

$$P(A_1 \cup \dots \cup A_n) = \sum P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \dots + (-1)^n P(A_1 \cap \dots \cap A_n)$$

Ex: If modeling a fair die

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$\mathcal{F}$ = set of subsets of Ω $= \{\emptyset, \{1\}, \{2\}, \dots, \{6\}, \{1,2\}, \dots\dots, \Omega\}$

$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = \frac{1}{6}$

$P(\{2,3\}) = \frac{2}{6}$, etc.

If have a biased die, say the side with 6 is heavy, so lands on 6 more often, could have

$P(\{1\}) = \dots = P(\{5\}) = \frac{1}{10}$, $P(\{6\}) = \frac{1}{2}$

$P(\{1,6\}) = \frac{1}{10} + \frac{1}{2}$, etc.

Rmk: If Ω is finite, can specify P by giving $P(\omega)$ $\forall \omega \in \Omega$. These determine P uniquely.
E.g. if we know the prob a die is 1, 2, 3, 4, 5, 6 then know P(die is odd).

Q: What if Ω is infinite?
Ex: Flip a fair coin until 1st head. Count the number of flips.

$\Omega = \{\infty, 1, 2, 3, \dots\}$

$P(k) = P(\text{takes } k \text{ flips to get 1st } H) = \frac{1}{2^k}$

$P(\infty) = 1 - \sum_{k \in \mathbb{N}} P(k) = 1 - \sum_{k \geq 1}^{\infty} \frac{1}{2^k} = 0.$

As in the finite case, P is determined by specifying $P(\omega)$ $\forall \omega \in \Omega$.

Random Variables

Sometimes interested not in the outcome itself, but a number associated with an outcome.

Defn: A random variable (RV) is a function from Ω to $\mathbb{R}$.

Ex: Roll 2 fair dice. $\Omega = \{(a,b) : a,b \in \{1,\dots,6\}\}$.

random variables

X_1 = outcome of 1st die.

X_2 = — " — 2nd die.

X_3 = Sum of the outcomes

$$P(X_3 = 3) = P(\{(1,2),(2,1)\}) = \frac{2}{36}.$$

$$P(X_1 = 2, X_3 = 8) = P(\{2,6\}) = \frac{1}{36}.$$

Defn: Let X be a RV. The probability distribution of the RV X is the collection of probabilities $P(X \in B)$ for "reasonable" sets B of real numbers.

Ex 1) We say the RV X has the Bernoulli distribution with parameter p if

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

(Think of a biased coin with prob. p of heads)

2) We say the RV X has the Binomial distribution with parameters n, p if

$$P(X = m) = \begin{cases} \binom{n}{m} p^m (1-p)^{n-m} & \text{if } 0 \le m \le n \\ 0 & \text{otherwise.} \end{cases}$$

(Ex: Toss a Bernoulli(p) coin n times & let X be the number of heads)

3) We say the RV X has the Poisson distribution with parameter λ if

$$P(X=k)=\begin{cases} e^{-\lambda}\frac{\lambda^k}{k!} & \text{if } k\in\mathbb{Z}_{\geq 0} \\ 0 & \text{otherwise.}\end{cases}$$

Ex: Pick a number uniformly at random from $[0,2]$.

X is equally likely to lie anywhere in $[0,2]$.

$P(x\in[a,b])=?$ Should be the proportion of $[0,2]$ covered by $[a,b]$ so if $0\leq a<b\leq 2$ should have $P(x\in[a,b])=\frac{b-a}{2}$.

We cannot specify such a distribution by specifying $P(X=t)$ $\forall t\in\mathbb{R}$ since $P(X=t)=0$ $\forall t$.

Defn: Let X be a RV. If a function f satisfies

$$P(X\leq b)=\int_{-\infty}^{b} f(x)\,dx \quad \forall\, b\in\mathbb{R},$$

then f is called the prob. density fcn (pdf) of X.

For such an f, $P(X\in B)=\int_B f(x)\,dx$.

Ex: If $X\sim \text{Uniform}[0,2]$, then $f(x)=\frac{1}{2}\mathbb{1}_{[0,2]}$ is a pdf for X.

Rmk: If X has a pdf, then $P(X=b)=0$ $\forall\, b\in\mathbb{R}$.

Q: Which f's can be pdf's?

A: Any $f\geq 0$ s.t. $\int_{-\infty}^{\infty} f(x)\,dx=1$ (since that is $P(X\in(-\infty,\infty))=P(\Omega)$)

A useful object to describe the distribution of a RV whether it is discrete, has density, or neither, is the cumulative distribution function (c.d.f.) defined by

$$F_X(t) := P(X \leq t).$$

If X has density $f_X(t)$ then $f_X(t) = F_X'(t)$.

<u>Ex:</u> A RV Z has the normal distr with parameters μ, σ^2, $Z \sim N(\mu, \sigma^2)$ if Z has density

$$f_Z(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The case $\mu = 0, \sigma = 1$ is called the standard normal & the density denoted $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

<u>Q:</u> Is $f_Z(t)$ a density? Clearly $f_Z(t) \geq 0$, so is $\int_{\mathbb{R}} f_Z(t)\,dt = 1$?

By a change of variable it is enough to do the case of $N(0,1)$.

$$\left(\int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right)^2 = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \iint e^{-\frac{x^2+y^2}{2}} dx\,dy$$

$$= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} r\,dr\,d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}} r\,dr = \frac{1}{2\pi} \cdot 2\pi \int_0^{\infty} e^{-s} ds = 1.$$

<u>Ex:</u> A RV X has the exponential distribution with parameter λ if it has density

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0. \end{cases}$$

Independence

<u>Defn:</u> Events A, B are called independent if $P(A \cap B) = P(A)\,P(B)$.

Defn: Events $A_1, \dots, A_n$ are independent if
$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$
for all $1 \le i_1 < i_2 < \dots < i_k \le n$.

Rmk: $A_1, \dots, A_n$ indep. is not the same as A_i, A_j indep $\forall i,j$.

Ex: Toss a fair coin 3 times.
A_1 - event 1st 2 match
A_2 - event last 2 match
A_3 - event 1st & last match.
A_1, A_2, A_3 - pairwise indep. but A_1, A_2, A_3 not indep.

Defn: Random variables $X_1, X_2, \dots, X_n$ are indep if $\forall B_1, \dots, B_n \subseteq \mathbb{R}$, the events
$X_1 \in B_1, \dots, X_n \in B_n$ indep.
I.e. $\forall 1 \le i_1 < \dots < i_k \le n$,
$$P(X_{i_1} \in B_{i_1}, \dots, X_{i_k} \in B_{i_k}) = \prod_{l=1}^{k} P(X_{i_l} \in B_{i_l}).$$

Ex: If $X_1, X_2, \dots, X_n$ indep Bernoulli(p), then
$S_n = X_1 + \dots + X_n \sim$ Binomial(n, p)

Expectation

Defn: Let X be a discrete RV.
The expected value of X, EX is defined to be
$$EX := \sum_t t P(X = t)$$
where the sum ranges over the possible values of X.
If X has density f, then EX is defined as
$$EX = \int_{\mathbb{R}} t f(t)\, dt.$$

Think of EX has follows: If you "measured" X a lot of times, it would be EX on average.
The precise statement behind that is

<u>Thm:</u> (The weak law of large numbers)
Let $X_1, X_2, \ldots$ be independent, identically distributed (i.i.d.) random variables with finite expectation.
Let $S_n = X_1 + \cdots + X_n$ & $\mu = EX_1$. Then

$$\frac{S_n}{n} \xrightarrow[\text{in probability}]{P} \mu$$

i.e. $\forall \varepsilon > 0$, $P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$.

Given a s'ce of RVs $X_1, X_2, X_3, \ldots$, what does it mean for $X_1, X_2, X_3, \ldots$ to converge to a RV X?

<u>Def:</u> A sequence of RVs $X_1, X_2, \ldots$ cv to a RV X in probability if $\forall \varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

We write this as $X_n \xrightarrow{P} X$.

<u>Def:</u> A sequence of RVs $X_1, X_2, \ldots$ cv to a RV X (all defined on the same space Ω) almost surely if

$$P(\lim_{n\to\infty} X_n = X) = 1$$

$$\parallel$$

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}).$$

Write as $X_n \xrightarrow{a.s.} X$

Defi: A sequence of RVs $X_1, X_2, \ldots$ cv to a RV X in distribution if

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t) \quad \forall t \text{ where } F_X(t) \text{ is cts.}$$

where $F_X(t)$ stands for the cdf of a RV X

i.e. $F_X(t) := P(X \le t)$.

Write as $X_n \xrightarrow{d} X$.

Defi: A sequence of RVs $X_1, X_2, \ldots$ cv to X (all defined on the same space) if

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0.$$

Thm: Suppose $X_1, X_2, \ldots, X$ are RVs defined on the same space (Ω, F, P). Then

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$$

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{P} X$$

Exercise: For all other implications come up with examples that show they fail.