Data3401 Week 1

January 16, 2020

Probability: experiments, outcomes, sample spaces, events

- Experiments (or trials) generate outcomes.
- Outcomes are random.
- Probability is a way to describe randomness.
- S or Ω ("big omega") $\equiv$ the sample space $\equiv$ the set of all possible outcomes.
- Outcomes are elements of Ω. Call them $\omega_1, \omega_2, \omega_3, \ldots$
- Events are subsets of Ω. Call them A, B, C etc.
- ω membership $\equiv$ event occurrence: suppose an experiment generates outcome ω and $\omega \in A, \omega \in B, \omega \notin C$
- We say events A and B occurred, event C did not occur, $A \cap B$ occurred; $A \cap C$ did not; $B \cup C$ _____

Handmade example

What are the $\omega' s$?

What are some events?

What type of diagram does this remind you of?

Expressing probabilities

- Lots of ways to write the probabilty of an event
- prob of event A = $P(A) = P(\{\omega \in \Omega : \omega \in A\}) = P(\omega \in A)$ all mean the same thing
- prob events A and B both occur = $P(\omega \in A \text{ and } \omega \in B) = P(\omega \in A \cap B) = P(A \cap B) \equiv P(AB)$
- prob event A or B occurs = $P(\omega \in A \text{ or } \omega \in B) = P(\omega \in A \cup B) = P(A \cup B)$

Example cont'd

Use your eyeball to calculate probabilities of events from previous slide.

Probability is math...

- ... but mathematicians claimed it only recently. In the 1930's, A. N. Kolmogorov formulated probability as a branch of measure theory.
- Probability is a measure defined on sets (events).
- Guess which guy is Kolmogorov ...

The three K axioms, some first principles

1. For any event A, $P(A) \geq 0$
2. $P(\Omega) = 1$
3. If $A_1, A_2, A_3, \ldots$ are disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

4. $P(\emptyset) = 0$
5. For any event A, $0 \leq P(A) \leq 1$
6. $P(A^C) =$ ____________________ = the rule of complements or the one minus trick.
7. $P(A \cup B) =$ ____________________

Math notation for sets

- $\emptyset \equiv \{\} \equiv$ the empty set $\equiv$ the set with no elements.
- Elements versus subsets: $\omega_1 \in \Omega, \omega_1 \in A, A \subset \Omega$.
- Every set is a subset of itself; $\emptyset$ is a subset of every set.
- $2^A \equiv$ the power set of $A \equiv$ the set of all subsets of A. It's a set of sets, so: $A \in 2^A, \emptyset \in 2^A, \{\emptyset, A\} \subset 2^A$.
- $A^C \equiv$ the complement of $A \equiv$ "A complement" $\equiv$ "not A" $\equiv$ all the elements of Ω which are not in A
- Intersection: $A \cap B \equiv$ the set of all ω's in __________
- Union: $A \cup B \equiv$ the set of all ω's in __________
- A and B are disjoint if they do not intersect $\equiv \quad A \cap B = \emptyset$. Remember _______________ for representing sets?
- Disjoint unions: sometimes we write $A \uplus B$ instead of $A \cup B$ to emphasize that are A and B are disjoint.
- BIG set operators work like this: $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n$.

Conditional probability adjusts $P(A)$ conditional on information about event B.

'given information' is usually $\omega \in B$
'conditional' $\equiv$ 'given'

Definition

The conditional probability of event A given event B is

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

- It's a good exercise to rewrite this with the ω's.
- Consider fish example again ...

(Unconditional) probability vs. conditional probability

Rearranging the definition of $P(A|B)$

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A|B)P(B) &= \frac{P(A \cap B)}{P(B)}P(B) \\
P(A|B)P(B) &= P(A \cap B)
\end{aligned}
$$

- Same thing for B given A:

$$P(B|A)P(A) = P(B \cap A)$$

- Right hand sides (RHS) are equal, so:

$$P(A|B)P(B) = P(B|A)P(A)$$

Conditional probability: Bayes' rule

- Continuing from previous slide:

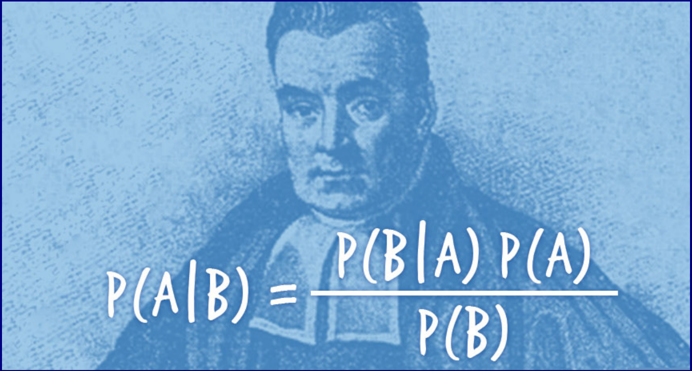
$$
\begin{aligned}
P(A|B)P(B) &= P(B|A)P(A) \\
\frac{P(A|B)P(B)}{P(B)} &= \frac{P(B|A)P(A)}{P(B)} \\
P(A|B) &= P(B|A)\frac{P(A)}{P(B)}
\end{aligned}
$$

- This formula shows how to 'reverse' the conditioning.
- We can make it go the other way:

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

- This result is Bayes' rule.

Reverend Thomas Bayes (1702-1761). Everyone uses Bayes's rule; his name is also associated with *Bayesian statistics* which relies on an alternative interpretation of probability. This was the subject of some disagreement in the past; not so much now.



Random variables

- Random variables are functions of ω. They extract the (quantitative) information which is carried by each ω.
- Let the sample space $\Omega =$ all the fish in the sea.
- Then each ω is one of the fish.
- Random variables are the (quantitative) questions you can ask a fish:
  - How long are you? How many siblings do you have?
  - Are you carnivorous? What percentage of your diet is vegetable matter?
  - How long was your commute today?
  - What's your favorite color?
- Before sampling, rv's live in math land. Sampling a fish and asking the question(s) generates data $\equiv$ a realization of the rv.
- In this setting, events look like $\{X < 12\}$ or $\{X < 12 \,\&\, X \geq 9\} = \{9 \leq X < 12\}$
- RV's can be **discrete** or **continuous** (chunky or smooth)

Continuous rv's are grouped into families (ch4)

- uniform
- normal (Gaussian)
- gamma, exponential, χ^2
- beta

Families defined by their probability density function (pdf)
pdf usually called $f_X(x)$ or $f(x)$
pdf depends on parameters
A family is a set of pdf's; this set indexed by the parameters
Parameter space $\equiv$ set of possible parameter values
Support $\equiv$ set of possible x-values
R functions for pdf's:

- `dunif(x,...)`
- `dnorm(x, ...)`
- type help(distributions) to see them all
- parameter values replace ...

Let X be any continuous rv with pdf $f()$

- Expected value (mean) of X:

$$E[X] \equiv \mu_X \equiv \mu \equiv \int_{-\infty}^{\infty} x\ f(x)dx$$

- Variance of X is also an expected value:

$$Var[X] \equiv E[(X-\mu)^2] \equiv \sigma_X^2 \equiv \sigma^2 \equiv \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)dx$$

- Cumulative distribution function (cdf):

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} f(t)dt$$

- The inverse of the cdf is the quantile function of X:

$$F_X^{-1}(p) \ \ or \ \ just \ \ F^{-1}(p)$$

$F^{-1}(p)$ returns the x value which has prob = p to its left.

Properties of expected value $E[\cdot]$

- $E[c] = c$
- $E[cX] = cE[X]$
- $E[X + c] = E[X] + E[c] = E[X] + c$
- $E[X + Y] = E[X] + E[Y]$
- $E[aX + bY + c] = aE[X] + bE[Y] + c$
- $Var[X + Y] = Var[X] + Var[Y]$ if X, Y are independent.
- $Var[aX] = a^2 Var[X]$

Some discrete rv's

- D_1 and D_2
- Binary trials: heads or tails? Houston or Dallas? I need an umbrella or I don't.
- Definition: $X \sim \mathit{Bernoulli}(p)$ means X has the distribution of a binary trial with $\Omega = \{0, 1\}$ and parameter $p = P(X = 1)$. By the rule of complements $P(X = 0) = 1 - p \equiv q$.
- Example from basketball: Let the rv *JH* = the outcome when James Harden shoots a free throw.

More discrete rv's

- Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \textit{Bernoulli}(p)$.
- The $\sim$ means 'is distributed as'. The *iid* means 'independent, identically distributed'. So the X_i's are independent Bernoulli trials all with the same success prob.
- Let $Y = \sum_{i=1}^{n} X_i$. So Y counts the successes.
- Definition: $Y \sim \textit{binomial}(n, p)$ means Y has the distribution of the sum of n independent $\textit{Bernoulli}(p)$ trials.
- Example with $n = 4$ and $p = 1/3$
- The pmf of Y is

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

- $\binom{n}{k}$ = 'n choose k' = number of size k sets you can select from n objects. Also called a binomial coefficient.
- Mean of Y: $\mu_Y = np$.
- Variance of Y: $\sigma_Y^2 = np(1-p) = npq$

General facts about discrete rv's

- Let X be any discrete rv (think of D_1 or a binomial rv)
- Let R be the set of possible values for X.
- Probability mass function is $p(k) = P(X = k)$ for $k \in R$
- Mean of X:
$$\mu \equiv E[X] = \sum_{k \in R} kp(k)$$
- Variance of X:
$$\sigma^2 \equiv Var[X] = \sum_{k \in R} (k - \mu)^2 p(k)$$
- Cumulative distribution function (cdf):
$$F(k) = P(X \leq k) = \sum_{j \leq k} p(j)$$
- Quantile function gets messier, have to be careful when we talk about
$$F_X^{-1}(p)$$
for discrete rv's.

Some math to recall

- $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$
- product rule: $(uv)' = u'v + uv'$
- integration by parts: $\int uv' = uv - \int u'v$
- foil
- binomial theorem: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$
- alternative formula for variance:
  $Var[X] \equiv E[(X-\mu)^2] = E[X^2] - \mu^2 = E[X^2] - E[X]^2$
- the one-minus trick
- start with Wikipedia