Probability as a measure

Outline

1) What is a probability?
2) Measures and integrals
3) Densities
4) Probability spaces

Probability as a measure

Outline

1) What is a probability?
2) Measures and integrals
3) Densities
4) Probability spaces

What is probability?

Two common answers:

1) <u>Frequentist answer</u>: Relative frequency over many repetitions of a given experiment.

2) <u>Bayesian answer</u>: Degree of belief that something is true, or will happen.

Many disagreements, e.g. about what kinds of things we can meaningfully assign prob. to.

- Prob. (Die roll lands on 4)
- Prob. (Surgery is successful)
- Prob. (Harris wins upcoming election)
- Prob. (Subatomic particle has predicted mass)
- Prob. ($P = NP$)
- Prob. (20^{th} digit of $\sqrt{2}$ is 5)

Bayesians get mileage by putting probs on everything.

Frequentists try to avoid this.

Many controversies about this!

Mathematical Probability

Fortunately, these disagreements don't extend to the mathematical construct of probability

Mathematical answer: A function P mapping (some) subsets of a sample space $\mathcal{X}$ to $[0,1]$, which is additive over disjoint sets:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{if } A_i \cap A_j = \emptyset, \text{ all } i \neq j$$

and has $P(\mathcal{X}) = 1$

Originally invented to analyze games of chance
Classical theory for discrete events defined by Laplace

Problems remained for continuous variables, e.g.

- treatment of pathological sets
- conditioning on probability zero events

Kolmogorov (1933) recognized:

- probability is a special case of a measure,
- expectation is an integral against a prob. measure

Measure theory basics

<u>Measure theory</u> is a rigorous grounding for probability theory [subject of 205A]

Simplifies notation & clarifies concepts, especially around integration & conditioning

Given a set $\mathcal{X}$, a measure μ maps subsets $A \subseteq \mathcal{X}$ to non-negative numbers $\mu(A) \in [0,\infty]$

<u>Example</u> $\mathcal{X}$ countable (e.g. $\mathcal{X} = \mathbb{Z}$)

<u>Counting measure</u> $\#(A) = \#$ points in A

<u>Example</u> $\mathcal{X} = \mathbb{R}^n$

<u>Lebesgue measure</u>
$$\lambda(A) = \int \cdots \int_A dx_1 \cdots dx_n = \text{Volume}(A)$$

<u>Standard Gaussian distribution</u>:

$$P_Z(A) = \mathbb{P}(Z \in A) \quad \text{where } Z \sim N(0,1)$$
$$= \int_A \phi(x)\,dx \qquad \phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

<u>NB</u> Because of pathological sets, $\lambda(A)$ can only be defined for certain subsets $A \subseteq \mathbb{R}^n$

In general, the domain of a measure μ is a collection of subsets $\mathcal{F} \subseteq 2^{\mathcal{X}}$

(power set)

$\mathcal{F}$ should satisfy certain closure properties (technical term: <u>σ-field</u>) (not important for us)

① $\mathcal{X} \in \mathcal{F}$

② If $A \in \mathcal{F}$ then $\mathcal{X} \setminus A \in \mathcal{F}$

③ If $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

<u>Ex</u>: $\mathcal{X}$ countable, $\mathcal{F} = 2^{\mathcal{X}}$

<u>Ex</u>: $\mathcal{X} = \mathbb{R}^n$, $\mathcal{F}$ = Borel σ-field $\mathcal{B}$

$\mathcal{B}$ = smallest σ-field including all open rectangles

$$(a_1, b_1) \times \cdots \times (a_n, b_n) \qquad a_i < b_i \quad \forall i$$

Given a <u>measurable space</u> $(\mathcal{X}, \mathcal{F})$ a <u>measure</u> is a map $\mu : \mathcal{F} \to [0, \infty]$ with

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad \text{for disjoint } A_1, A_2, \ldots \in \mathcal{F}$$

$$\mu(\emptyset) = 0$$

μ <u>probability measure</u> if $\mu(\mathcal{X}) = 1$

Push-forward

Suppose we have a measure μ on $\mathcal{X}$,
and a (nice) function $f : \mathcal{X} \to \mathcal{Y}$

$\rightsquigarrow$ new measure ν on $\mathcal{Y}$, defined by
where μ-mass in $\mathcal{X}$ gets mapped to

Push-forward measure: $\nu(B) = \mu(\{x : f(x) \in B\})$

More compactly, $\nu = \mu \circ f^{-1}$

not inverse function

for pre-image $f^{-1}(B) = \{x : f(x) \in B\} \subseteq \mathcal{X}$

Important because it gives us the dist. of a function of a random variable

Ex Define $f(z) = \max\{0, z\}$,

let $Y = f(Z)$ where $Z \sim N(0,1)$

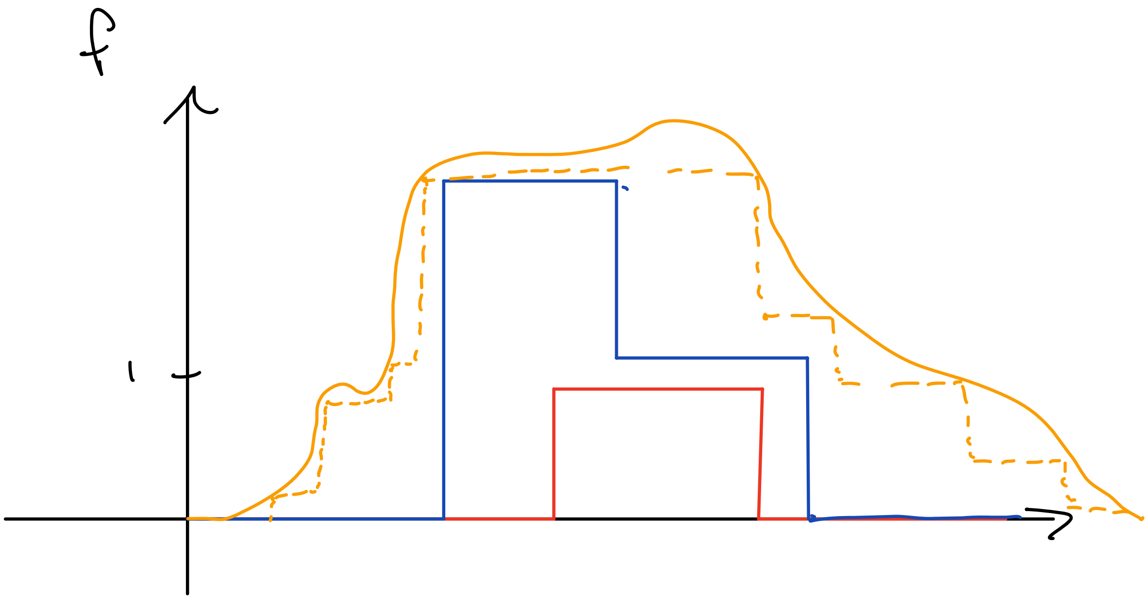
Define
$$\begin{aligned} P_Y(B) &= \mathbb{P}(Y \in B) \\ &= \mathbb{P}(f(Z) \in B) \\ &= P_Z(f^{-1}(B)) \end{aligned}$$

$\Rightarrow$ P_Y is push-forward of P_Z, through f

Integrals

Measures let us define <u>integrals</u> that put weight $\mu(A)$ on $A \subseteq \mathcal{X}$

Define $\int 1\{x \in A\} \, d\mu(x) = \mu(A)$, extend to other functions by linearity & limits:



Indicator $\int 1\{x \in A\} \, d\mu(x) = \mu(A)$

Simple Function $\int \left(\sum c_i 1\{x \in A_i\} \right) d\mu(x) = \sum c_i \mu(A_i)$

"Nice enough" (measurable) + function $\int f(x) \, d\mu(x)$ approximated by simple functions

Examples:

Lebesgue integral

Counting: $\int f \, d\# = \sum_{x \in \mathcal{X}} f(x)$

Lebesgue: $\int f \, d\lambda = \int \cdots \int f(x) \, dx_1 \cdots dx_n$

Gaussian: Note $\int 1_A(x) \, dP_Z(x) = P_Z(A) = \int_{-\infty}^{\infty} 1_A \phi \, dx$

By extension,

$$\int f \, dP_Z = \int f(x) \phi(x) \, dx = \mathbb{E}[f(Z)]$$

To evaluate $\int f \, dP_Z$ rewrite as $\int f \phi \, dx$.

density [can't always do this]

e.g. Binom

It is nice to turn integrals we care about into Lebesgue integrals. When can we do this?

Densities

λ and P above are closely related. Want to make this precise.

Given $(\mathcal{X}, \mathcal{F})$, two measures P, μ

We say P is <u>absolutely continuous</u> wrt μ if $P(A) = 0$ whenever $\mu(A) = 0$

Notation: <u>$P \ll \mu$</u> or we say μ <u>dominates</u> P

If $P \ll \mu$ then (under mild conditions) we can always define a <u>density function</u>

$\rho : \mathcal{X} \to [0, \infty)$ with

$$P(A) = \int_A \rho(x) \, d\mu(x)$$

$$\int f(x) \, dP(x) = \int f(x) \rho(x) \, d\mu(x)$$

Sometimes written $\rho(x) = \frac{dP}{d\mu}(x)$, called <u>Radon-Nikodym derivative</u>

Densities are very useful:

Turn $\int f(x)dP(x)$ into something we know how to evaluate, such as

1) $\int_{\mathcal{X}} f(x)p(x)dx$ (X continuous, $\mathcal{X} \subseteq \mathbb{R}^n$)

$p(x)$ called <u>probability density function</u> (pdf)

2) $\sum_{x \in \mathcal{X}} f(x)p(x)$ (X discrete, $\mathcal{X}$ countable)

$p(x)$ called <u>probability mass function</u> (pmf)

Often define distributions by giving their density wrt some known measure, e.g.

<u>Ex</u>: Binom(n, θ) pmf : $p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ ...

(density p wrt counting measure on $\mathcal{X} = \{0, \ldots, n\}$

Note this dist. has <u>no</u> density wrt Lebesgue:

$$\int_{\{0,\ldots,n\}} p(x)\,dx = 0 \quad \text{for any function } p$$

Probability space, random variables

Problem setup may have many random outcomes with complex relationships to one another

Convenient to start with abstract outcome $\omega \in \Omega$

- represents "everything that happens"
- quantities of interest are functions of ω

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space

$\omega \in \Omega$ called outcome
$A \in \mathcal{F}$ called event
$\mathbb{P}(A)$ called probability of A

A random variable is a function $X: \Omega \to \mathcal{X}$

We say X has distribution Q $(X \sim Q)$ if

$$\mathbb{P}(X \in B) = \mathbb{P}(\{\omega : X(\omega) \in B\}) = Q(B)$$

$Q(B)$ is the push-forward of $\mathbb{P}$: $Q(B) = \mathbb{P} \circ X^{-1}(B)$

Idea applies more generally: if μ measure on $\mathcal{X}$, $f: \mathcal{X} \to \mathcal{Y}$ induces new measure $\nu(B) = \mu(f^{-1}(B))$

Can write events involving many R.V.s:

$$\mathbb{P}(X > Y > Z \geq 0) = \mathbb{P}(\{\omega : \cdots\})$$

The <u>expectation</u> is an integral w.r.t. $\mathbb{P}$

$$\mathbb{E}[f(X,Y)] = \int_\Omega f(X(\omega), Y(\omega))\, d\mathbb{P}(\omega)$$

To do real calculations we must eventually boil $\mathbb{P}$ or $\mathbb{E}$ down to concrete integrals/sums/etc.

If $\mathbb{P}(A) = 1$ we say A occurs <u>almost surely</u>

More in Keener ch. 1, much more in Stat 205A