MIE 1807
Principles of Measurement

2017-01-16

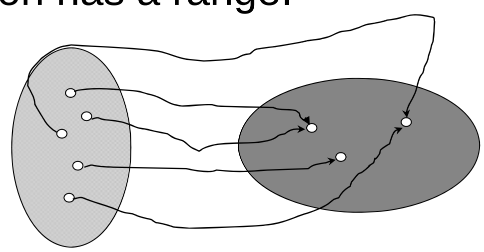
Neil Montgomery

The Basic Rules: the simple (trivial) motivating examples

- Toss a coin
- Throw a six-sided die
- Throw two six-sided dice
- Good or bad pedagogy?
  - “Boring” “Unrealistic”
  - Completely understandable
- “Probability” requires the definition of two specific kinds of functions.

Functions

- A function has a domain.
- A function has a range.



- The function is a rule
- And then you are corrupted by calculus.

Corruption by Calculus

- You spend years worrying about what the rules (i.e. functions) *look* like.
  - Call them all *f(x)*
  - Continuous?
  - Differentiable?
  - Max/min/graphs/drawing/etc.
- We don't do these things in probability to the two kinds of functions of interest.

A *probability* is a *function*

Consider the toss of a coin. The *outcomes* are:

$$S = \{\mathrm{H}, \mathrm{T}\}$$

The *sets of outcomes*, or *events* are:

$$\emptyset \to 0$$
$$\{H\} \to 0.5$$
$$\{T\} \to 0.5$$
$$\{H, T\} \to 1$$

This rule is called a *probability*

The most basic rules

$$P(\emptyset) = 0 \qquad P(S) = 1$$

For any event $A \subset S$:

$$0 \leq P(A) \leq 1$$

Another example

Consider the throw of a die. The *outcomes* are:

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

Simple example, but there are $2^6 = 64$ possible *events*. . . .

$$\emptyset, \{1\}, \{2\}, \{2, 4, 6\}, \{1, 2, 3, 4\}, \ldots$$

Some probabilities:

$$P(\{1\}) = \frac{1}{6} = P(\{2\}) \text{ etc.} \qquad P(\{1, 2\}) = \frac{2}{6}$$

Another Rule

$$A = \{2, 3, 4\} \qquad B = \{2, 4, 6\}$$

$$A \cup B = \{2, 3, 4, 6\}$$

$$A \cap B = \{2, 4\}$$

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

$$= \frac{4}{6}$$

Important Special Case

$$A = \{2, 3, 4\} \qquad C = \{1, 6\}$$

$$A \cup C = \{1, 2, 3, 4, 6\}$$

$$A \cap C = \emptyset$$

$$\begin{aligned} P(A \cup C) &= P(A) + P(C) \\ &= \frac{5}{6} \end{aligned}$$

Generalization

If $A_1, A_2, \ldots$ are *mutually disjoint* (i.e. $A_i \cap A_j = \emptyset$), then

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots$$

Summary of the first of two new kinds of function

- A *probability* is a
  - *Function*
  - whose domain is a collection of subsets of a sample space *S*
  - whose range is the unit interval [0,1]
  - that satisfies the basic rules on the previous slides

What if you have some partial information about an event?

- You are *willing* to bet money that the toss of a die will result in a 1 or a 2.
- The die is tossed.
- I tell you “an odd number appeared!”
- I give you a chance to quit the game.
- What would your *rational* choice be?
- What if instead I told you “a number larger or equal to 2 occurred”?

Analysis…

A priori $S = \{1,2,3,4,5,6\}$

$$A = \{1,2\} \qquad P(A) = \frac{1}{3}$$

"An odd number appeared" $S' = \{1,3,5\}$

$$A' = \{1\} \qquad P'(A') = \frac{1}{3}$$

"≥ 2 appeared" $S^* = \{2,3,4,5,6\}$

$$A^* = \{2\} \qquad P^*(A') = \frac{1}{5}$$

Conditional Probability

Given events A and B with $P(B) > 0$, the *conditional probability of A given B* is:

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

Die toss example 1: $A = \{1,2\}$ $B = \{1,3,5\}$

$$P(A|B) = \frac{P(\{1\})}{P(\{1,3,5\})} = \frac{1/6}{1/2} = \frac{1}{3}$$

Conditional Probability

Die toss example 2: $A = \{1, 2\}$ $B = \{2, 3, 4, 5, 6\}$

$$P(A|B) = \frac{P(\{2\})}{P(\{2, 3, 4, 5, 6\})} = \frac{1/6}{5/6} = \frac{1}{5}$$

(You might be interested in Bayes' formula, for your own education. See textbook #2 for details.)

Independence

The obviously interesting thing about the first example is that the probability did not change with extra information.

$$P(A|B) = \frac{1/6}{1/2} = \frac{1}{3} = P(A)$$

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

Independence: $P(A \cap B) = P(A)P(B)$

The second of the two new kinds of functions

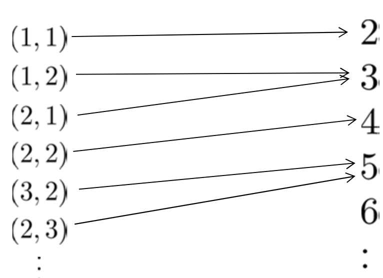
Sample spaces can be unwieldy, let alone collections of subsets thereof.

Consider tossing two dice, with 36 outcomes and $2^{36} = 6.9 \times 10^9$ events.

A more natural setting is usually to consider a numerical result relating to the random outcome, such as "the sum of the two dice".

The second new kind of function

We'll mostly study functions with sample space domains and *real valued* ranges.



The thing to notice about this kind of function

Any subset of the real line you can think of implies (via this function) some set of outcomes.

Examples:

"anything less than 4" $\Longrightarrow \{(1,1),(1,2),(2,1)\}$

"bigger than or equal to 11"

$$\Longrightarrow \{(5,6),(6,5),(6,6)\}$$

And what can be done with "sets of outcomes"?

This new kind of function needs a name

A real-valued function of a sample space is called a *random variable*.

The conventional notation:

$$X, Y, Z, W, X_1, X_2, Y_1, X_{12}, \text{ etc.}$$

Argument of the function is usually omitted!

Calculus versus probability again

In calculus the fundamantal question is: what does the function *look like*.

Tangents, areas, derivatives, minima, maxima. Notation: $f(x)$ etc.

In probability the fundamental question is: what is its *distribution*.

i.e. the range of values and their probabilities.

The fundamental distribution descriptors

First, two examples:

Coin toss where H leads to gain of \$1 and T leads to loss of \$1.

$X_1(H) = 1$ and $X_1(T) = -1$

$P(X_1 = 1) = 0.5$ and $P(X_1 = -1) = 0.5$

The fundamental distribution descriptors

Die toss where gain in X_2 dollars is $10\times$ the outcome.

$$P(X_2 = 10i) = \frac{1}{6}$$

In both cases we came up with a function of the form:

$$P(X = x_i) = p(x_i) = p_i$$

Where $p_i \geq 0$ and $\sum p_i = 1$.

Descriptor 1: *Probability mass function*

Such a function is called a pmf and is used when the random variable takes on discrete values.

Completely describes the distribution for such a random variable.

Fundamental use is to compute probabilities:

$$P(X \in A) = \sum_{x_i \in A} p(x_i)$$

Fundamental distribution descriptors

For discrete r.v., there is the p.m.f.:

$$P(X = x_i) = p(x_i) = p_i$$

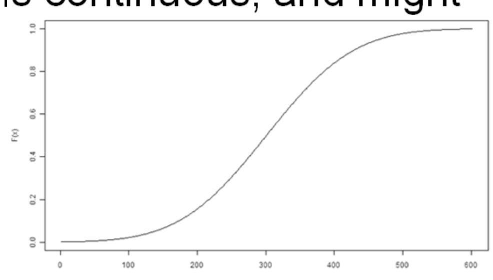
Where $p_i \geq 0$ and $\sum p_i = 1$.

For *all* random variables there is the *cumulative distribution function* (cdf):

$$F(x) = P(X \leq x)$$

"Continuous" random variables

- A random variable that takes on *real* values on an *interval* is usually called a *continuous* random variable.
- It's *cdf* is what is continuous, and might look like this:



Probability Density Function

The analogue to p.m.f. for continuous random variables is the p.d.f. (the derivative of the c.d.f.) Notation: $f(x)$. Synonym: “density”

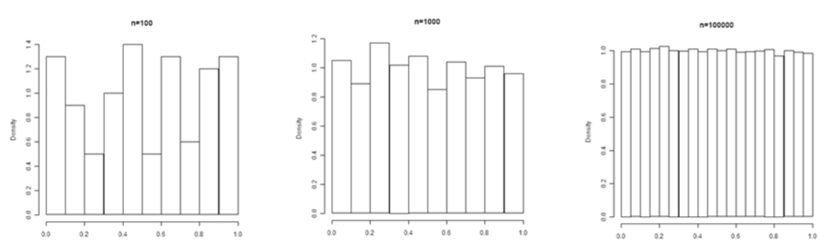
(Think of a p.d.f. as the “limiting” histogram as $n \to \infty$.)

Fundamental use:

$$P(a < X \leq b) = \int_a^b f(x)\,dx$$

Example 1

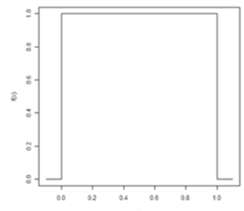
Computing problem: let X be generated randomly within the interval $(0, 1)$.



Example 1

Probability of being in (a,b) for $0 \leq a \leq b \leq 1$ should be: $b - a$

Density:



$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

"Uniform distribution on (0,1)"

Consequence of continuous r.v. "model"

Randomly generate between 0 and 1. What is the probability of getting, say, 0.2?

$$P(X = 0.2) = P(0.2 \leq X \leq 0.2) = \int_{0.2}^{0.2} 1\,dx = 0$$

$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b)$ etc.— **for continuous r.v. only!**

Independent Random Variables

Events A and B are *independent* if knowledge that A occurred would not alter your knowledge of the probability that B occurred.

Random variables X_1 and X_2 are *independent* if knowing the outcome of X_1 would not alter your assessment of the distribution of X_2.

Expected Value

Recall sample mean....

Theoretical analogue for random variables is *expected value* $E(X)$

$$E(X) = \begin{cases} \sum_x xp(x) & \text{for } X \text{ discrete} \\ \int_{\infty}^{\infty} xf(x)\,dx & \text{for } X \text{ continuous} \end{cases}$$

as long as everything converges blah blah blah...

$E(X)$ is often shortened to μ, but be careful.

Examples

Die toss where gain in X dollars is $10\times$ the outcome.

$p(x) = P(X = 10i) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$

$$E(X) = \sum_{i=1}^{6} 10i\frac{1}{6} = 35$$

Examples

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

"Uniform distribution on (0,1)"

$$E(X) = \int_{\infty}^{\infty} x f(x)\,dx = \int_0^1 x\,dx$$

$$= \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}$$

Moments

The r^{th} moment of X is:

$$E(X^r) = \begin{cases} \sum_x x^r p(x) & \text{for } X \text{ discrete} \\ \int_{\infty}^{\infty} x^r f(x)\,dx & \text{for } X \text{ continuous} \end{cases}$$

as long as everything converges blah blah blah...

In this course the first two moments will be of interest.

Variance and Standard Deviation

Recall sample variance....

Theoretical analogue for random variables is *variance* $\mathrm{Var}(X)$

$$\mathrm{Var}(X) = \begin{cases} \sum_x (x - E(X))^2 p(x) & \text{for } X \text{ discrete} \\ \int_{\infty}^{\infty} (x - E(X))^2 f(x)\,dx & \text{for } X \text{ continuous} \end{cases}$$

as long as everything converges blah blah blah...

$\mathrm{Var}(X)$ is often shortened to σ^2, but be careful.

Standard deviation: $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)} = \sigma$

Example

Die toss where gain in X dollars is $10\times$ the outcome.

$p(x) = P(X = 10i) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$

$$\mathrm{Var}(X) = \sum_{i=1}^{6} (10i - 35)^2 \frac{1}{6} = 291.67$$

$$\mathrm{sd}(X) = 17.1$$

Computing Variances

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{\infty}^{\infty} (x - E(X))^2 f(x)\,dx \\
&= \int_{\infty}^{\infty} (x^2 - 2xE(X) + E(X)^2) f(x)\,dx \\
&= \int_{\infty}^{\infty} x^2 f(x)\,dx - \int_{\infty}^{\infty} 2xE(X) f(x)\,dx + \int_{\infty}^{\infty} E(X)^2) f(x)\,dx \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

Example

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

"Uniform distribution on (0,1)"

$$\mathrm{Var}(X) = \int_{\infty}^{\infty} (x - E(X))^2 f(x)\,dx$$

$$= E(X^2) - E(X)^2$$

$$= \int_0^1 x^2 f(x)\,dx - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Some Rules

$$E(aX + b) = aE(X) + b$$

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

If X_1 and X_2 are independent:

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

but not vice-versa.

Standardization

The rules allow us to consider *standardized* versions of random variables, with mean 0 and variance 1.

$$Y = \frac{X - \mu}{\sigma}$$

$$E(Y) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}\sigma^2 = 1$$

But Y and X have the same *shape*.