Statistics for Data Science

Lecture 3

Transforms and Expectations

Ken Cameron

Recall

- Random variable X has a cdf Fx(x)
- Cumulative distribution function.
- We also covered
  - Mass density functions
  - Probability density functions

Content

- Expected Values
- Estimating a random variable
- Functions of a random variable
- Support Sets
- Another theorem

Expected Value

- One more basic definition…
- The expected value of a random variable is the long-run average value of repetitions
- More specifically, it is the probability-weighted average of all possible values
- What is the expected value of rolling a fair six-sided die?
- What about if the die isn’t fair?

Expected Value

- The expected value of a random variable is denoted by E.

- For a discrete random variable x, the expected value

$$\mathrm{E}[x] = \sum_{x} x\, p(x)$$

- We have already seen one example of this: the arithmetic mean, which is the expected value when every value of x is equally likely
- Fair dice roll:
  - $\mathrm{E}[x] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = 3.5$

Expected Value

- The expected value is more powerful, for example, an unfair dice roll:
  - $1\frac{1}{12} + 2\frac{1}{12} + 3\frac{1}{12} + 4\frac{1}{12} + 5\frac{1}{12} + 6\frac{7}{12} = 4.75$

- Or the amount you should expect to win if you play roulette:
  - $\mathrm{E}[\text{profit on £1}] = -£1\frac{37}{38} + £35\frac{1}{38} = -£0.0526$

Expected values

- So E g(X) is

$$\int_{-\infty}^{\infty} g(x) f_x(x) dx$$ if X is continuous

$$\sum_{x \in X} g(x) f_x(x) \quad = \quad \sum_{x \in X} g(x) P(X = x)$$ if X is discrete

If the sum or integral exists. If E|g(X)| = ∞ then E g(X) does not exist.

Minimising Distance

- Suppose we measure the distance between a random variable X and a constant b by $(X – b)^2$.

- The closer to X that b is, the smaller the squared value above.

- If we can determine the value of b that minimises $E(X\text{-}b)^2$ we will have a good predictor of X.

Without using calculus

- $E(X - b)^2 = E(X-EX + EX - b)^2$ Add +/- EX which does nothing.

  $= E((X - EX) + (EX-b))^2$ Group terms.

  $= E(X-EX)^2 + (EX - b)^2 + 2E((X-EX)(EX-b))$ Expanded sqr

Note,

$$E((X - EX)(EX - b)) = (EX - b)E(X - EX) = 0$$

- $(EX - b)$ is a constant and comes out of the expectation

$$E(X - EX) = EX - EX = 0$$

$$E(X - b)^2 = E(X - EX)^2 + (EX - b)^2$$

- Choose $b = EX$

$$\min_b E(X - b)^2 = E(X - EX)^2$$

Functions of a Random Variable

- Any function of X, call it g(X) is also a random variable.

- We can introduce Y, a new random variable to represent this:

$$Y = g(X)$$

Sample Spaces

- As Y is a function of X we can describe the probabilistic behaviour of Y in terms of X. So for any set A:

$$P(Y \in A) = P(g(X) \in A)$$

- So if y = g(x), g defines a mapping from the sample space of X $\mathcal{X}$ to a new sample space $\mathcal{Y}$ of Y.

$$g(x): \mathcal{X} \rightarrow \mathcal{Y}$$

Inverse mapping

- We can also associate an inverse mapping g^{-1}

- A mapping from subsets of Y to subsets of X, defined by

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}$$

Sets into Sets

- $g^{-1}(A)$ is the set of points in $\mathcal{X}$ that g(x) takes into the set A.

- A can be a point set {y}.

$$g^{-1}(\{y\}) = \{x \in \mathcal{X}: g(x) \in y\}$$

- We can write $g^{-1}(\{y\})$ as $g^{-1}(y)$ in this case.

- $g^{-1}(y)$ can still be a set when there are multiple x for which $g(x) = y$.

- Now for any set $A \subset \mathcal{Y}$,

$$P(Y \in A) = P(g(X) \in A) = P(\{x \in \mathcal{X}: g(x) \in A\}) = P(X \in g^{-1}(A))$$

Defines the pdf of Y. Showing the pdf satisfies the axioms left as an exercise.

If X is discrete, then $\mathcal{X}$ is countable.

- Y = g(X) is $\mathcal{Y} = \{y: y = g(x), x \in \mathcal{X}\}$ is also a countable set.

- Therefore Y is also a discrete random variable.

- The pmf $f_y(y)$ can be found by identifying $g^{-1}(y)$ for each $y \in \mathcal{Y}$ and summing the appropriate probabilities ($x \in g^{-1}(y)$)

An Example

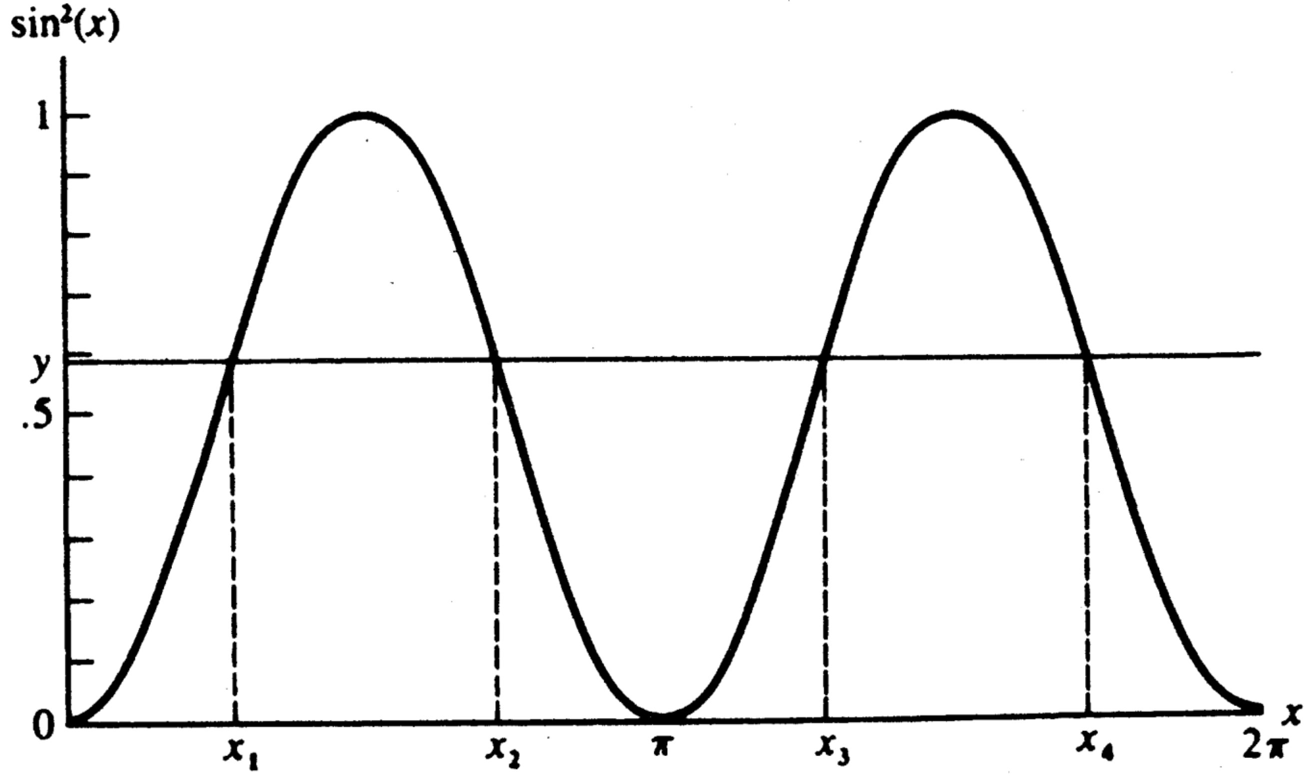
- Random Variable X with uniform distribution on interval $(0, 2\pi)$

$$f_x(x) = \begin{cases} 1/(2\Omega) & 0 < x < 2\Omega \\ 0 & Otherwise \end{cases}$$

- Now consider

$$Y = \sin^2(X)$$

Example



An Example

$$P(Y <= y) = P(X <= x_1) + P(x_2 <= X <= x_3) + P(X >= x_4)$$

- From symmetry of $\sin^2(x)$ and that X has uniform distribution

$$P(X <= x_1) = P(X >= x_4) \text{ and } P(x_2 <= X <= x_3) = 2P(x_2 <= X <= \pi)$$

- So

$$P(Y <= y) = 2P(X < = x_1) + 2P(x_2 <= X <= \pi)$$

An Example

- Where x_1 and x_2 are the two solutions to:

$$\sin^2(x) = y \quad 0 < y < \pi$$

- The resulting cdf expression was not simple, even though it looked like it should be when we started.

- It's important to keep track of the sample spaces of the random variable to avoid confusion.

Support Sets

- When the transform is from X to Y = g(X) it is convenient to use:

- $\mathcal{X} = \{x: f_x(x) > 0\}$ and $\mathcal{Y} = \{y: y = g(x) \text{ for some } x \in \mathcal{X}\}$ [1]

- The pdf of X is positive only on the set $\mathcal{X}$ and zero elsewhere.

- This known as the support set of a distribution.

Monotone functions

- It's easier if the g(x) are monotone.

- Functions are monotone if:

$$u > v \Rightarrow g(u) > g(v)$$

Or,

$$u < v \Rightarrow g(u) > g(v)$$

- If x –> g(x) is monotone then it is one-to-one and onto from $\mathcal{X}$ -> $\mathcal{Y}$
  - Each x goes to only one y.
  - And each y comes from at most one x
- And for $\mathcal{Y}$ defined as in [1] for each y $\in \mathcal{Y}$ there is an x $\in \mathcal{X}$ such that g(x) = y.

- If g is monotone then g^{-1} is single-valued.

$$g^{-1}(y) = x \quad \text{if and only if } y = g(x)$$

- If g is increasing then:

$$\{x \in X: g(x) <= y\} = \{x \in X : g^{-1}(g(x)) <= g^{-1}(y)\} = \{x \in X : x <= g^{-1}(y)\}$$

- If g is decreasing then:

$$\{x \in X: g(x) <= y\} = \{x \in X : g^{-1}(g(x)) >= g^{-1}(y)\} = \{x \in X : x > g^{-1}(y)\}$$

- If g(x) is increasing then:

$$F_y(y) = \int_{-\infty}^{g^{-1}(y)} f(x)dx = F_x(g^{-1}(y))$$

- If g(x) is decreasing then:

$$F_y(y) = \int_{g^{-1}(x)}^{-\infty} f(x)dx = F_x(g^{-1}(y))$$

Theorem

- These results allow us to state the following theorem:

- Let X have cdf $F_x(x)$, let $Y = g(X)$ and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as in [1]

  a. If g is an increasing function on $\mathcal{X}$, $F_y(y) = F_x(g^{-1}(y))$ for $y \in \mathcal{Y}$.

  b. If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_y(y) = 1 - F_x(g^{-1}(y))$ for $y \in \mathcal{Y}$.