September 26, 2017

CS 361: Probability & Statistics

Random variables

Continuous random variables

- Some quantities we would like to model don't fit into any of the probability theory we have discussed so far: things like height, weight, temperature are all quantities that we can think of as appearing in one experimental setup or another
- It's a bit tricky to formally think about sample spaces and events with these kinds of experiments and we won't do so in this course

Continuous random variables

- Instead, we will think of each continuous random variable as being associated with a function p(x) called a probability density function, which we can use to compute probabilities
- Roughly, if p(x) is a probability density function then for an infinitesimally small interval dx

$$p(x)dx = P(\{\text{event that } X \text{ takes a value in the range } [x, x + dx]\})$$

- So in order to be a density function, p(x) must be non-negative

Probability density functions

- Now if p(x) is the density function for some continuous random variable and

$$p(x)dx = P(\{\text{event that } X \text{ takes a value in the range } [x, x+dx]\})$$

- We can compute for a < b

$$P(\{X \text{ takes on a value in the range } [a,b]\}) = \int_a^b p(x)\,dx$$

- Observe, then, that

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

Density functions: summary

- Density functions are non-negative
- (Density functions can take on values greater than 1)
- Integrating density functions allows us to compute the probability that a continuous random variable takes on a value within a range
- Density functions integrate to 1 if we let x run from negative infinity to infinity

Density functions

- Non-negative functions which integrate to 1 are the density functions of some continuous random variable
- A continuous random variable X can be associated to a non-negative function which integrates to 1 that we can use to compute probabilities for X
- If we have a density for a variable, great, we can compute probabilities
- Many times, we are trying to come up with a sensible density to describe a continuous phenomenon

Example

Suppose we have a physical system that produces random numbers in the range 0 to ϵ with $\epsilon > 0$

No number outside this range can appear, but every number within this range is equally likely.

If X is a random variable that tells us which number was generated, what is its density function?

Recall our interpretation of the density function

We are going to want p(x) to be 0 when $x < 0$ or $x > \epsilon$

We will also want p(x) to have the same value, c, for every other x since we want all outcomes in our range to be equally likely

$$p(x) = \begin{cases} 0, & \text{for } x < 0 \\ 0, & \text{for } x > \epsilon \\ c, & \text{otherwise} \end{cases}$$

Example

$$p(x) = \begin{cases} 0, & \text{for } x < 0 \\ 0, & \text{for } x > \epsilon \\ c, & \text{otherwise} \end{cases}$$

In order to figure out what c should be, recall that we must have

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{0}^{\epsilon} p(x)\,dx = 1$$

But

$$\int_{0}^{\epsilon} c\,dx = c\epsilon$$

So we wind up with

$$p(x) = \begin{cases} 0, & \text{for } x < 0 \\ 0, & \text{for } x > \epsilon \\ \frac{1}{\epsilon}, & \text{otherwise} \end{cases}$$

Observe that we can have p(x) > 1 if

$\epsilon < 1$

Expected value: motivation

Think about the game we mentioned earlier. Flip a coin, heads with probability p, tails with (1-p). When it's heads you pay me q, tails I pay you r.

Should we bother to play this game?

Recall our relative frequency interpretation of probabilities

$$Npq - N(1-p)r$$

Approx payout for N games

$$pq - (1-p)r$$

Average payout per game

Expected value

Definition: 5.6 *Expected value*

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P, we define the expected value

$$\mathbb{E}[X] = \sum_{x \in \mathcal{D}} x P(X = x).$$

This is sometimes written which is $\mathbb{E}_P[X]$, to clarify which distribution one has in mind

Example

- Flip a fair coin, $P(T)=P(H)=1/2$, if it comes up heads you pay me 1, if it comes up tails, I pay you 1. Let X be the random variable corresponding to my income
- $$\begin{aligned} E[X] &= 1/2(1) + 1/2(-1) \\ &= 0 \end{aligned}$$

Example

- Flip a fair coin, heads you pay me 2, tails I pay you 1
- $E[X] = (1/2)(2) - (1/2)(1)$
  $= 1/2$
- My "expected income" from a game is 1/2, which we note is not an amount I could actually earn in one run of the game
- Playing this game is good for me and bad for you

Expectation

- If f is a function of a random variable, then f(X) is a random variable as well. And we can write the expectation of f(X) as

$$E[f] = \sum_{x \in D} f(x) P(X = x)$$

Expectation

- With continuous variables, we compute expectations with an integral

Definition: 5.8 *Expected value*

Given a continuous random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P, we define the expected value

$$\mathbb{E}[X] = \int_{x \in \mathcal{D}} x p(x) dx.$$

This is sometimes written $\mathbb{E}_p[X]$, to clarify which distribution one has in mind.

Expectation

Definition: 5.9 *Expectation*

Assume we have a function f that maps a continuous random variable X into a set of numbers $\mathcal{D}_f$. Then $f(x)$ is a continuous random variable, too, which we write F. The expected value of this random variable is

$$\mathbb{E}[f] = \int_{x \in \mathcal{D}} f(x)p(x)dx$$

which is sometimes referred to as "the expectation of f". The process of computing an expected value is sometimes referred to as "taking expectations".

This integral may not be defined or may not be finite. The cases we encounter will have well-behaved expectations, however

Linearity of expectation

- For random variables X and Y, and constant *k* we have

$$E[X+Y]=E[X]+E[Y]$$

$$E[kX]=kE[X]$$

Mean and variance

- The expected value of a random variable is also called the **mean**
- The **variance** of a random variable is defined as

$$\mathrm{var}[X] = E[(X - E[X])^2]$$

Properties of variance

Useful Facts: 5.3 *Variance*

1. For any constant k, $\mathsf{var}[k] = 0$
2. $\mathsf{var}[X] \geq 0$
3. $\mathsf{var}[kX] = k^2\mathsf{var}[X]$
4. if X and Y are independent, then $\mathsf{var}[X+Y] = \mathsf{var}[X] + \mathsf{var}[Y]$

You will prove some of these

Standard deviation

Definition: 5.13 *Standard deviation*

The **standard deviation** of a random variable X is defined as

$$\mathsf{std}\left(\{X\}\right) = \sqrt{\mathsf{var}[X]}$$

Covariance

- The **covariance** of X and Y is defined as

$$\mathrm{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- Note that cov(X,X) = var[X]

Another expression for variance

By definition we have

$$var[X] = E[(X - E[X])^2]$$

Which we rewrite to reduce confusion

$$var[X] = E[(X - \mu_X)^2]$$

Expanding

$$var[X] = E[X^2 - 2X\mu_X + \mu_x^2]$$

Using linearity of expectation

$$var[X] = E[X^2] - 2\mu_X E[X] + \mu_x^2$$

Or

$$var[X] = E[X^2] - 2E[X]E[X] + E[X]^2$$

Giving our final expression

$$\boxed{var[X] = E[X^2] - (E[X])^2}$$

Another expression for covariance

$$\begin{aligned}
\mathrm{cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\
&= \mathbb{E}[(XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y])] \\
&= \mathbb{E}[XY] - 2\mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \boxed{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].}
\end{aligned}$$

A couple of results

Useful Facts: 5.6 *Variance and Covariance*

1. if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
2. if X and Y are independent, then $\mathsf{cov}(X, Y) = 0$.

Proofs are straightforward, in the book

Example: coin flip

- We flip a biased coin that has P(H)=p and P(T) = 1-p and let X be a random variable with value 1 for heads and 0 for tails
- What is E[X]?

$$1p + 0(1-p) = p$$

- What is Var[X]?

$$\mathrm{var}[X] = E[X^2] - E[X]^2$$

$$\mathrm{var}[X] = p - p^2$$

Overloaded terms?

- You may have noticed we are using some terms like mean, variance, standard deviation, that we have used before
- Earlier we developed concepts for datasets and now we have concepts for random variables
- They aren't very different concepts
- If we suppose that each data item in a dataset has probability 1/N and consider the random variable given by reporting the value of the data item. These concepts will be equal

$$\mathbb{E}[x] = \sum_i x_i p(x_i) = \frac{1}{N} \sum_i x_i = \text{mean}\left(\{x\}\right).$$

Markov's inequality

This theorem says that for any random variable X and any value a, we have

$$P(\{|X| \geq a\}) \leq \frac{E[|X|]}{a}$$

A random variable is unlikely to have an absolute value much larger than the mean of its absolute value

If, for instance, we took a = 10 E[| X |], we'd get

$$P(\{|X| \geq 10E[|X|]\}) \leq .10$$

Indicator functions

- To make Markov's inequality easy to prove we will introduce the useful notion of an indicator function

Definition: 5.16 *Indicator functions*

An indicator function for an event is a function that takes the value zero for values of X where the event does not occur, and one where the event occurs. For the event $\mathcal{E}$, we write

$$\mathbb{I}_{[\mathcal{E}]}(X)$$

for the relevant indicator function.

- We have following fairly immediately, since I is 0 everywhere but on $\mathcal{E}$

$$\mathbb{E}[\mathbb{I}_{[\mathcal{E}]}] = P(\mathcal{E})$$

Proving Markov

Proving

$$P(\{|X| \geq a\}) \leq \frac{E[|X|]}{a}$$

First note that for all a we have

$$a\mathbb{I}_{[\{|X|\geq a\}]}(X) \leq |X|$$

Since I(X) will be 0 for any X < a and 1 for any value >= a

Taking expectations, we get

$$E[a\mathbb{I}_{[\{|X|\geq a\}]}(X)] \leq E[|X|]$$

We can pull out the a due to the linearity of expectation

$$E[\mathbb{I}_{[\{|X|\geq a\}]}(X)] \leq \frac{E[|X|]}{a}$$

And finally using

$$\mathbb{E}\left[\mathbb{I}_{[\mathcal{E}]}\right] = P(\mathcal{E})$$

Our left hand side is

$$P(|X| \geq a)$$

Which gives the desired result