Lecture 26 : Probability in a continuous context

We now take a brief look at the relationship between integration and probability. By a <u>continuous random variable</u> X, we shall mean some observable quantity that can take on a range of continuous values (real numbers as opposed to integers for example). For any such X, we may construct a <u>probability density function</u> f, where the area under the function f over the interval $[a,b]$ represents the probability that X lies within $[a,b]$:

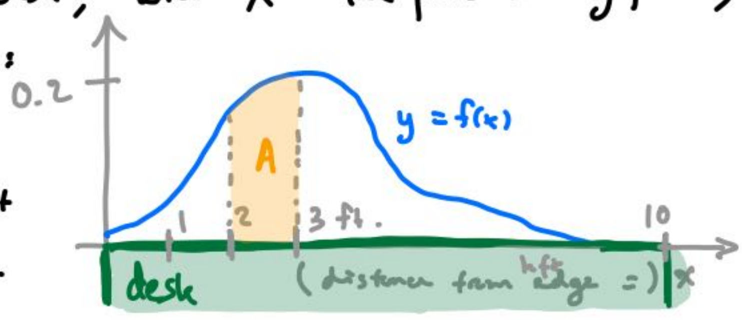
$$P(a \le X \le b) = \int_a^b f(x)\,dx.$$

x represents possible values for X

This function f must satisfy

- $f(x) \ge 0$ (negative probabilities make no sense!)
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (the probability that X takes <u>some</u> value is 1, i.e. 100%.).

So if the fellow who lives in the ceiling is about to drop a pin onto the desk, and X = the pin's landing point, maybe f looks like this: it does <u>not</u> mean that the probability the pin lands 3 ft from the edge is $0.2 = 20\%$..



It does mean that the probability that the pin lands between 2 and 3 ft from the edge of the desk is $\int_2^3 f(x)\,dx \approx 0.18 = 18\%$, i.e, the area A. (Note that it is fine for f to exceed 1 in places so long as the total area underneath is 1.)

Now imagine the ceiling guy drops $N > 1000$ pins on the desk. We expect the pin density at this point to closely match the graph of f: so dividing $[0,10]$ into n subintervals of length $\Delta x = \frac{10}{n}$, the # of pins in the ith subinterval is

$$P_i = P(x_{i-1} \le X \le x_i) \cdot N \approx f(x_i)(\Delta x)\,N.$$

We can say these are all approximately at x_i. So the average location of all the dropped pins is

$$X_{avg} = \frac{X_1 + X_2 + \cdots + X_N}{N} \approx \frac{\sum_{i=1}^n x_i P_i}{N}$$

$$\approx \sum_{i=1}^n x_i f(x_i)\,\Delta x$$

$$\xrightarrow[n\to\infty]{} \overline{X} := \int_0^{10} x f(x)\,dx.$$

More generally, define the mean μ (or expected value $E(X)$) by

same as μ

$$\mu := \int_{-\infty}^{\infty} x f(x)\,dx$$

and the variance σ^2 (or standard deviation σ) by

$$\sigma^2 := \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

(This measures how spread out the distribution is.)

Finally the <u>median</u> m is the value of X for which

$$\int_m^\infty f(x)\,dx = 0.5 .$$

Ex 1/ Is $f(x) = \frac{1}{1+x^2}$ a valid probability density function?

$$\int_{-\infty}^\infty f(x)\,dx = \arctan(x) \Big|_{-\infty}^{\infty} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi .$$

No, but $f(x) = \frac{1/\pi}{1+x^2}$ would be OK. //

Ex 2/ Let X be a cont. random variable with density function

$$f(x) = \begin{cases} (a-1)\,x^{-a}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

Find μ. (Assume $a>2$.)

First we check that f is a density function:

$$\int_{-\infty}^\infty f(x)\,dx = \int_1^\infty (a-1)x^{-a}\,dx = \lim_{b\to\infty} \left(-x^{-a+1}\right)\Big|_1^b$$

$$= \lim_{b\to\infty} \left(1 - \frac{1}{b^{a-1}}\right) = 1 . \quad ✓$$

Now $\mu = \int_{-\infty}^\infty x f(x)\,dx = \int_1^\infty (a-1)x^{-a+1} = \lim_{b\to\infty} \frac{a-1}{-a+2}\, x^{-a+2} \Big|_1^b$

$$= \frac{a-1}{a-2} .$$ //

We can also talk about the cumulative distribution function

$$F(x) = \int_{-\infty}^x f(x)\,dx = P(X \leq x) .$$

By the FTC, $F'(x) = f(x)$. Moreover, we have $\lim\limits_{x \to -\infty} F(x) = 0$, and $\lim\limits_{x \to \infty} F(x) = 1$.

Ex 3 / If the survival function $S(x)$ of some species yields the probability that an individual will still be alive at age x, and $S(x) = e^{-(x/20)^2}$, determine the mean and median life expectancy.

First, let X be the lifespan (in years) and $f(x)$ its density function; then $S(x) = P(X \geq x) \left(= 1 - F(x)\right)$.

So $f(x) = F'(x) = -S'(x) = \frac{x}{200} \cdot e^{-(x/20)^2} = \frac{x}{200} e^{-\left(\frac{x}{20}\right)^2}$

for $x \geq 0$ and 0 for $x < 0$. For the median, write

$$\frac{1}{2} = \int_0^m \frac{x}{200} e^{-\left(\frac{x}{20}\right)^2} dx = -e^{-(x/20)^2} \Big|_0^m$$

$$= 1 - e^{-(m/20)^2}$$

$$\Rightarrow e^{-(m/20)^2} = 1/2 \underset{\ln}{\Rightarrow} \left(\frac{m}{20}\right)^2 = \ln 2$$

$$\Rightarrow m = 20\sqrt{\ln 2} \approx 16.65 \text{ yrs.}$$

For the mean,

Use fact that $\int_0^\infty e^{-x^2} dx = \frac{1}{2}\sqrt{\pi}$ + substitution

$$\mu = \int_0^\infty x f(x)\, dx = \int_0^\infty \frac{x^2}{200} e^{-\left(\frac{x}{20}\right)^2} dx$$

$$= \underbrace{-x e^{-\left(\frac{x}{20}\right)^2} \Big|_0^{\text{“}\infty\text{”}}}_{=0} + \int_0^\infty e^{-(x/20)^2} dx = 10\sqrt{\pi}$$

$u = x \quad dv = \frac{x}{200} e^{-(x/20)^2} dx$

$du = dx \quad v = -e^{-(x/20)^2}$

≈ 17.72 yrs. //

A couple of typical density functions:

- <u>Uniform distribution</u>

$$f(x) = \begin{cases} 0, & x<0 \text{ or } x>M \\ \frac{1}{M}, & 0 \le x \le M \end{cases}$$

— e.g. if you arrive at a red light which is red for M seconds in its cycle, and X = # of seconds to wait for green.

- <u>Normal distribution</u>

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

(has μ and σ as indicated)

— use substitution + our earlier result that

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

- <u>Exponential distribution</u>

$$f(x) = \begin{cases} 0, & x<0 \\ \frac{1}{\mu} e^{-x/\mu}, & x \ge 0 \end{cases}$$

— arises in context of waiting times (for service, for a predator to find prey, etc.)

TRY: The probability density function of X is given by $f(x) = \frac{1}{300} e^{-x/300}$ for $x \ge 0$ (and 0 for $x<0$). Find $P(X \le 600)$. [Hint: you may need $e^{-1} = 0.368$, $e^{-2} = 0.135$, $e^{-3} = 0.050$]

[Ans: $\frac{1}{300}\int_0^{600} e^{-x/300}\,dx = -e^{-x/300}\Big|_0^{600} = 1 - e^{-2} \approx 0.865 = 86.5\%$]