Lecture 26: Probability in a Continuous Context

We now take a brief look at the relationship between integration and probability. By a continuous random variable $X$, 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 probability density function $f$, where the area under the function $f$ over the interval $[a,b]$ represents the probability that $X$ lies within $[a,b]$:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx.$$  

This function $f$ must satisfy

- $f(x) \geq 0$ (negative probabilities make no sense!)
- $\int_{-\infty}^{\infty} f(x) \, dx = 1$ (the probability that $X$ takes some 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 it looks like this; it does not mean that the probability the pin lands 3 ft from the edge is $0.2 = 20\%$. 

![Diagram of a probability density function with a shaded area representing a probability]
It does mean that the probability that the pin lands between 2 and 3 ft from the edge of the desk is \( \int_2^3 f(x) \, dx \approx 0.15 \), 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 \( i \)-th subinterval is

\[
\Theta_i = P(\xi_{i-1} \leq X \leq \xi_i) \cdot N = f(\xi_i)(\Delta x) \cdot N.
\]

We can say then an oil approximately at \( \xi_i \). So the average location of all the dropped pins is

\[
\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} \approx \frac{\sum_{i=1}^{n} \xi_i \cdot \Theta_i}{N}
\]

\[
\approx \frac{\sum_{i=1}^{n} \xi_i \cdot f(\xi_i) \cdot \Delta x}{N} \to \bar{X} = \int_0^{10} x f(x) \, dx.
\]

More generally, define the mean \( \mu \) (or expected value \( E(X) \)) by

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

and the variance \( \sigma^2 \) (or standard deviation \( \sigma \)) by

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

(This means how spread out the distribution is.)
Finally the median \( m \) 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?
\[
\int_{-\infty}^{\infty} f(x) \, dx = \arctan(x) \bigg|_{-\infty}^{\infty} = \frac{\pi}{2} - \frac{(-\pi)}{2} = \pi.
\]
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 \( \mu \). (Assume \( a > 2 \).)

First we check that \( f \) is a density function:
\[
\int_{-\infty}^{\infty} f(x) \, dx = \int_{1}^{\infty} (a-1)x^{-a} \, dx = \lim_{b \to \infty} \left[ \frac{b^{1-a} - 1}{1-a} \right] = 1.
\]
Now \( \mu = \int_{-\infty}^{\infty} xf(x) \, dx = \int_{1}^{\infty} (a-1)x^{-a+1} \, dx = \lim_{b \to \infty} \left[ \frac{a-1}{a-2} x^{1-a} \right] = \frac{a-1}{a-2} \).

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
\[
\lim_{x \to -\infty} F(x) = 0, \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.
\]

**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 \( S(x) = e^{-\left(\frac{x}{\theta}\right)^2} \), determine the mean and median life expectancy.

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

So \( f(x) = F'(x) = -S'(x) = \frac{x}{200} e^{-\left(\frac{x}{\theta}\right)^2} \),

for \( x \geq 0 \) and 0 for \( x < 0 \). For the mean, write
\[
\frac{1}{2} = \int_0^\infty \frac{x}{200} e^{-\left(\frac{x}{\theta}\right)^2} \, dx = -\left[ e^{-\left(\frac{x}{\theta}\right)^2} \right]_0^\infty
\]

\[
= 1 - e^{-\left(\frac{0}{\theta}\right)^2}
\]

\[
\Rightarrow e^{-\left(\frac{0}{\theta}\right)^2} = \frac{1}{2} \Rightarrow \left(\frac{m}{\theta}\right)^2 = \ln 2
\]

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

For the mean,
\[
m = \int_0^\infty x f(x) \, dx = \int_0^\infty \frac{x^2}{200} e^{-\left(\frac{x}{\theta}\right)^2} \, dx
\]

\[
= -\left[ e^{-\left(\frac{x}{\theta}\right)^2} \right]_0^\infty + \int_0^\infty e^{-\left(\frac{x}{\theta}\right)^2} \, dx
\]

\[
= 0 + \int_0^\infty e^{-\left(\frac{x}{\theta}\right)^2} \, dx
\]

\[
\Rightarrow m = 17.72 \text{ yrs.}
\]
A couple of typical density functions:

- **Uniform distribution**
  
  \[ f(x) = \begin{cases} 
  0 & , \ x < 0 \text{ or } x > M \\
  \frac{1}{2M} & , \ 0 \leq x \leq M 
  \end{cases} \]

  — 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.

- **Normal distribution**
  
  \[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  
  (has \( \mu \) and \( \sigma \) as indicated)

  — use substitution + our earlier result that
  
  \[ \int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi} \]

- **Exponential distribution**
  
  \[ f(x) = \begin{cases} 
  0 & , \ x < 0 \\
  \frac{1}{\mu} e^{-x/\mu} & , \ x \geq 0
  \end{cases} \]

  — 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 \( x \geq 0 \) (and 0 for \( x < 0 \)). Find \( P(X \leq 600) \).

**[Hint:** you may need \( e^1 = 0.368, e^2 = 0.135, e^3 = 0.050 \)]

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