Probability: Continuous Random Variables

... last time ...: a *continuous random variable* X
(taking values in an interval (a, b), possibly $(-\infty, \infty)$)

- has a **probability density function** f:

  $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$

- and a **(cumulative) distribution function**

  $F(x) = P(X < x) = \int_a^x f(t)dt,$ so $f(x) = F'(x)$.

- the **mean** (or **expected value**) of X is

  $\mu = \bar{x} = \int_a^b x\, f(x)\, dx$

- some famous examples: **uniform**: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$;
  exponential; **normal** (coming soon)

The Median of a Probability Density Function

The **median** of a continuous random variable X is the value x_{med} such that

$$\boxed{P(X < x_{med}) = P(X > x_{med}) = \tfrac{1}{2}}$$

In terms of the probability density function f of X,

$$\boxed{\int_a^{x_{med}} f(x)dx = \int_{x_{med}}^b f(x)dx = \tfrac{1}{2}}$$

and in terms of the distribution function F of X,

$$\boxed{F(x_{med}) = \tfrac{1}{2}}.$$

Can you interpret the mean and median geometrically/physically in terms of the graph of the density function f?

Example: the Exponential Distribution

The Exponential Distribution: probability density function is

$$f(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{\mu} e^{-\frac{t}{\mu}} & t \geq 0 \end{cases}$$

waiting time for service, lightbulb failure time, radioactive decay, ...

Find its cumulative distribution function, its mean, and its median.

- for $t \geq 0$, $F(t) = \int_0^t \frac{1}{\mu} e^{-\frac{t}{\mu}} dt = -e^{-\frac{t}{\mu}} |_0^t = 1 - e^{-\frac{t}{\mu}}$
- $\bar{t} = \int_0^\infty t \frac{1}{\mu} e^{-\frac{t}{\mu}} dt = -te^{-\frac{t}{\mu}} |_0^\infty + \int_0^\infty e^{-\frac{t}{\mu}} dt = -\mu e^{-\frac{t}{\mu}} |_0^\infty = \mu$
- $\frac{1}{2} = F(t_{med}) = 1 - e^{-\frac{t_{med}}{\mu}} \implies t_{med} = \mu \ln(2)$ (half-life!)

Example: the time T it takes for a certain radioactive atom to undergo decay is a continuous random variable with an exponential density function. If the mean decay time is 5000 years, what is the probability that the atom takes longer than 1000 years to decay?

- the probability density function for $t > 0$ is $f(t) = Ce^{-kt}$
- the distribution function is
  $F(t) = \int_0^t Ce^{-ky} dy = -\frac{C}{k} e^{-ky}|_0^t = \frac{C}{k}(1 - e^{-kt})$
- normalization: $1 = F(\infty) = \frac{C}{k} \quad \Longrightarrow \quad C = k$
- mean: $\mu = \int_0^\infty t\ ke^{-kt} dt = -te^{-kt}|_0^\infty + \int_0^\infty e^{-kt} dt = \frac{1}{k}$
- so $k = \frac{1}{5000}$
- $P(T > 1000) = 1 - F(1000) = 1 - (1 - e^{-\frac{1}{5}}) = e^{-\frac{1}{5}} \approx 0.82$

Variance and Standard Deviation

The **variance** of a continuous random variable X, taking values in (a, b), with probability density function $f(x)$, and with mean μ, is

$$\boxed{\sigma^2 = var[X] = \int_a^b (x-\mu)^2 \; f(x) \; dx}$$

and its **standard deviation** is

$$\boxed{\sigma = \sqrt{var[X]}}.$$

Example: find the mean, variance, and standard deviation of a number chosen randomly and uniformly from an interval $[0 \; , \; N]$. How likely is the number to be within one SD of the mean?

- density function is constant $f(x) = \frac{1}{N}$ for $0 \leq x \leq N$
- by symmetry, mean is $\frac{N}{2}$ (or compute $\mu = \int_0^N \frac{1}{N} x dx = \frac{N}{2}$)
- $\sigma^2 = \int_0^N \frac{1}{N}(x - \frac{N}{2})^2 = \int_0^N (\frac{x^2}{N} - x + \frac{N}{4}) = N^2(\frac{1}{3} - \frac{1}{2} + \frac{1}{4}) = \frac{N^2}{12}$
- $\sigma = \frac{N}{\sqrt{12}}$
- $P(\frac{N}{2} - \frac{N}{\sqrt{12}} < X < \frac{N}{2} + \frac{N}{\sqrt{12}}) = \int_{\frac{N}{2} - \frac{N}{\sqrt{12}}}^{\frac{N}{2} + \frac{N}{\sqrt{12}}} \frac{1}{N} dx = \frac{2}{\sqrt{12}} \approx 0.58$

Example: The Normal Distribution

The probability density function $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

describes a continuous random variable, taking values in $(-\infty, \infty)$, with a **normal (or Gaussian) distribution**, denoted $\boxed{N(\mu, \sigma^2)}$.

annual rainfall, math midterm test scores, heights,...

Exercise: given that it is normalized, $\int_{-\infty}^{\infty} f(x)dx = 1$, verify that its mean (and median) is μ, and its variance is σ^2.

Special case $N(0, 1)$ is the **standard normal distribution**:

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Its distribution function is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \quad \left(= \frac{1}{2} + \frac{1}{2} erf\left(\frac{x}{\sqrt{2}}\right) \right)$$

Exercise: what is the probability a (standard) normal random variable lies within 1 SD of its mean? $P(-1 < X < 1)$

$$= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-\frac{x^2}{2}} dx = F(1) - F(-1) = erf\left(\frac{1}{\sqrt{2}}\right) \approx 0.68$$