4.1: Probability Density Functions and Cumulative Distribution Functions

Taylor

University of Virginia

Motivation

In chapter 3, our discrete rvs had a finite or countably infinite number of possible values it could take on. Now we'll talk about continuous rvs. They can take on a whole interval/range of possible values.

Definition

Let X be a continuous rv.

Definition

The **probability density function** pdf of X is the function $f(x)$ such that for any $a \leq b$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

...it also has these properties

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

Definition

Definition

The **cumulative distribution function (cdf)** for a cts rv X is this function

$$F(x) = \int_{-\infty}^{x} f(s) ds$$

Another fact

If X is continuous with pdf and cdf $f(x)$ and $F(x)$, respectively, then

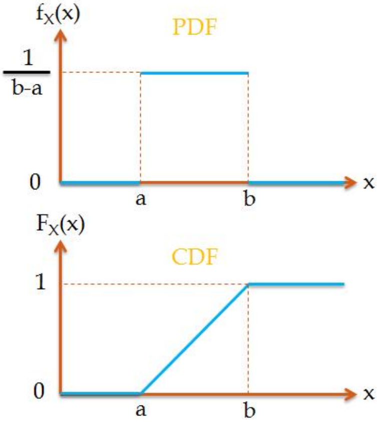
$$f(x) = \frac{d}{dx}F(x)$$

(because of the fundamental theorem of calculus)

Definition

Example of continous rv No. 1: the **uniform distribution** on the interval $[a, b]$:

$$f(x) = f(x; a, b) = (b - a)^{-1}, \qquad a \leq x \leq b$$



Derivation

assuming $a \leq x \leq b$...

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(s)ds \\
&= \int_{a}^{x} \frac{1}{b-a} ds \\
&= \frac{1}{b-a}(s)\Big|_{s=a}^{s=x} \\
&= \frac{x-a}{b-a}
\end{aligned}
$$

Note

Here is an important thing to remember about continuous rvs:

$$P(X = c) = \int_c^c f(x)dx = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} f(x)dx = 0$$

so for any a, b

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b)$$

(we don't have to be careful about our inequalities like we do with discrete rvs...)

Another fact

Let X be a cts rv with pdf $f(x)$ and cdf $F(x)$. Then for any a

$$P(X > a) = 1 - F_X(a)$$

$$P(a \leq X \leq b) = P(a < X \leq b) = F_X(b) - F_X(a)$$

Definition

Definition

A **percentile** $\eta(p)$ associated with some percent p of a cts rv X is some number such that

$$p = F[\eta(p)] = \int_{-\infty}^{\eta(p)} f(s)ds$$

i.e. it's some number such that p % of the data is behind it

Note: **quantiles** are just special cases of percentiles (e.g. quartiles = 25th,50th,75th percentiles) (e.g. deciles = 10th,20th,30th,40th,50th,60th,70th,80th,90th percentiles)

Note: the **median** is the 50th percentile

Definition

Definition

A cts rv X is **symmetric** if there is some point c such that

$$f(c-s) = f(c+s)$$

for all s

Example

This is example 4.9 from page 167. We have the density $f(x) = \frac{3}{2}(1 - x^2)$ as long as $X \in [0, 1]$. Find the 25% percentile for this distribution.

Example

This is example 4.9 from page 167. We have the density $f(x) = \frac{3}{2}(1 - x^2)$ as long as $X \in [0, 1]$. Find the 25% percentile for this distribution.

$$\int_0^{\eta(.25)} \frac{3}{2}(1 - x^2)dx = .25 \iff$$

$$\int_0^{\eta(.25)} \frac{3}{2}dx - \int_0^{\eta(.25)} \frac{3}{2}x^2 dx = .25 \iff$$

$$\frac{3}{2}\eta - \frac{1}{2}\eta^3 = .25$$

Example

X follows a **normal** or **gaussian distribution** if

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

show that it is symmetric

Example

X follows a **normal** or **gaussian distribution** if

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

show that it is symmetric

$$\begin{aligned} f(\mu - s) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(-s)^2}{2\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(s)^2}{2\sigma^2}\right] \\ &= f(\mu + s) \end{aligned}$$

and this holds for arbitrary s