Overview

- Continuous Random Variables
- Cumulative Distribution Function
- How do we calculate the area under a curve ?
- Continuous Random Variables: CDF and PDF
- Expectation and Variance
- Conditional Probability Density Function
- Chain Rule for PDFs
- Bayes Rule for PDFs
- Independence

Continuous Random Variables

All RVs up to now have been discrete:

- Take on distinct values e.g. in set $\{1, 2, 3\}$
- Often represent binary values or counts

What about continuous RVs ?

- Take on real-values
- e.g. travel time to work, temperature of this room, fraction of Irish population supporting Scotland in the rugby

Cumulative Distribution Function

Suppose Y is a random variable, which may be discrete or continuous valued.

- $F_Y(y) := P(Y \leq y)$ is the **cumulative distribution function** (CDF).
- CDF exists and makes sense for both discrete and continuous valued random variables
- When Y takes discrete values $\{y_1, \cdots, y_m\}$, then $F_Y(y) = \sum_{j:y_j \leq y} P(Y = y_j)$
- $F_Y(-\infty) = 0$, $F_Y(+\infty) = 1$.
- Also,

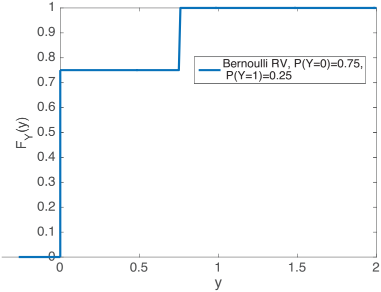
$$P(Y \leq b) = P(Y \leq a) + P(a < Y \leq b)$$
$$i.e.\ F_Y(b) = F_Y(a) + P(a < Y \leq b)$$

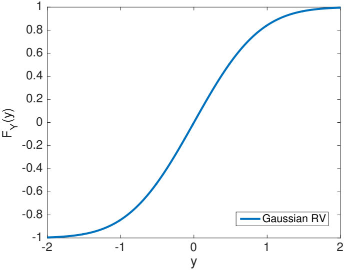
Therefore,

$$P(a < Y \leq b) = F_Y(b) - F_Y(a)$$

Cumulative Distribution Function

Examples of CDFs for discrete and continuous valued RVs:

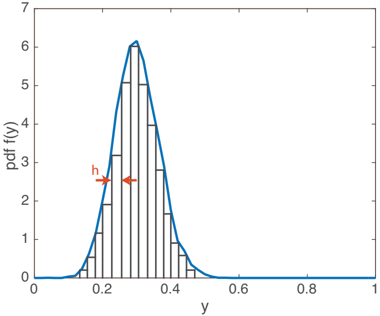




- Observe that CDF always starts at 0 and rises to 1
- CDF never decreases

How do we calculate the area under a curve ?

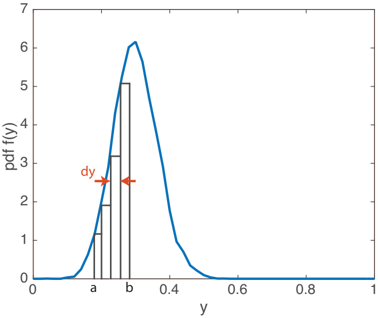
- Fit a series of rectangles under the curve, each of width h:



- We know the area under a rectangle, its the height$\times$width h
- Add up the areas of all the rectangles to get an estimate of the area under the curve
- As h gets smaller and smaller ($h \to 0$) this value becomes closer and closer to the true area[1]

[1] The maths needed to analyse this convergence is beyond this module, but if interested take a look at https://en.wikipedia.org/wiki/Riemann_integral and https://en.wikipedia.org/wiki/Lebesgue_integration.

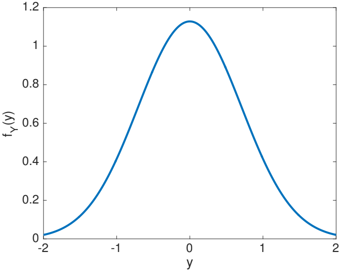
How do we calculate the area under a curve ?

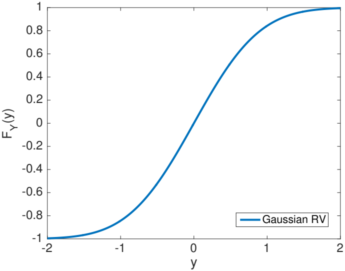


- Think of $f(y)dy$ as the area of the rectangle between y and $y + dy$ with dy infinitesimally small.
- Write the area under curve between a and b as $\int_a^b f(y)dy$
- Think of integral as the sum of areas of rectangles each of width h as $h \to 0$. Integral symbol $\int$ is supposed to be suggestive of a sum. Can think of dy as h (infinitesimally small).

How do we calculate the area under a curve ?

Example: CDF $F_Y(y)$ in right-hand plot is area under curve in left-hand plot between $-\infty$ and y i.e. $F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt$



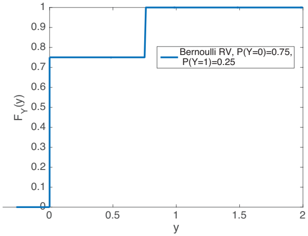


Continuous Random Variables: CDF and PDF

- For a continuous-valued random variable Y there exists a function $f_Y(y) \geq 0$ such that:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt$$

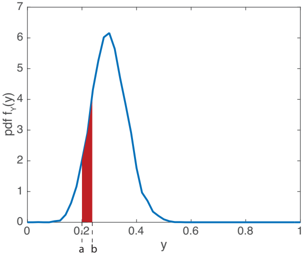
- cf $F_Y(y) = \sum_{j:y_j \leq y} P(Y = y_j)$ in discrete-valued case
- f_Y is called the **probability density function** or **PDF** of Y.
- $\int_{-\infty}^{\infty} f(y)dy = 1$ (since $\int_{-\infty}^{\infty} f(y)dy = F_Y(\infty) = P(Y \leq \infty) = 1$)
- Note that tricky to define PDF f_Y for a discrete random variable since its CDF has "jumps" in it.



Continuous Random Variables: CDF and PDF

- It follows that

$$
\begin{aligned}
&P(a < Y \leq b) \\
&\qquad = F_Y(b) - F_Y(a) \\
&\qquad = \int_{-\infty}^{b} f_Y(t)dt - \int_{-\infty}^{a} f_Y(t)dt \\
&\qquad = \int_{a}^{b} f_Y(t)dt
\end{aligned}
$$



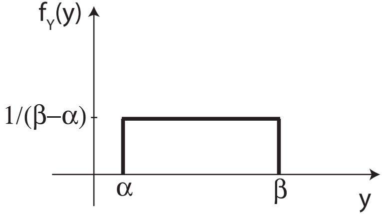
- The probability density function $f(y)$ for random variable Y is not a probability e.g. it can take values greater than 1.
- Its the area under the PDF between points a and b that is the probability $P(a < Y \leq b)$

Example: Uniform Random Variables

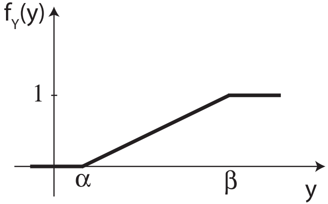
Y is a **uniform random variable** when it has PDF:

$$f_Y(y) = \begin{cases} \frac{1}{\beta-\alpha} & \textit{when } \alpha \leq y \leq \beta \\ 0 & \textit{otherwise} \end{cases}$$

PDF

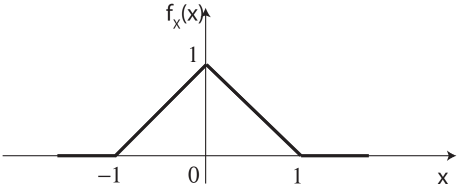


CDF



- For $\alpha \leq a \leq b \leq \beta$: $P(a \leq Y \leq b) = \frac{b-a}{\beta-\alpha}$
- rand() function in Matlab.
- A bus arrives at a stop every 10 minutes. You turn up at the stop at a time selected uniformly at random during the day and wait for 5 minutes. What is the probability that the bus turns up ?

Example



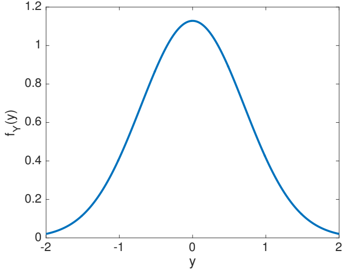
- Check the area under the PDF is 1. Area of left-hand triangle is $1/2$, area of right-hand triangle same. Total is 1.
- What is $P(0 \leq X \leq 1)$? Its the area under the PDF between points 0 and 1 i.e. the area of the right-hand triangle. So $P(0 \leq X \leq 1) = 0.5$.
- What is $P(0 \leq X \leq \infty)$? $f_X(x) = 0$ for $x > 1$, so $P(0 \leq X \leq \infty) = P(0 \leq X \leq 1) = 0.5$

The Normal Distribution

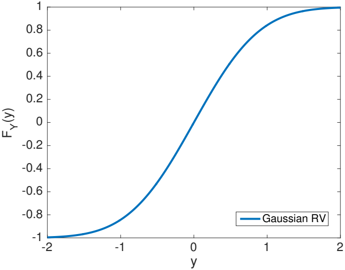
Y is a **Normal random variable** $Y \sim N(\mu, \sigma^2)$ when it has PDF:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

PDF



CDF



$\mu = 0$, $\sigma = 1$

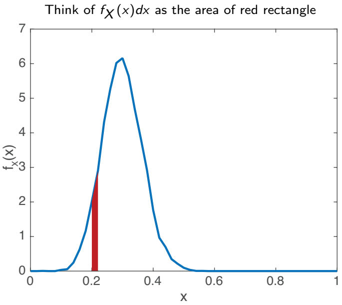
- $E[Y] = \mu$, $Var(Y) = \sigma^2$
- Symmetric about μ and defined for all real-valued x
- A Normal RV is also often called a **Gaussian random variable** and the Normal distribution referred to as the Gaussian distribution.

Expectation and Variance

For dx infinitesimally small,

$$P(x \leq X \leq x + dx) = F_X(x + dx) - F_X(x)$$
$$\approx f_X(x)dx$$

so we can think of $f_X(x)dx$ as the probability that X takes a value between x and $x + dx$.



Definitions:

For discrete RV X | For continuous RV X

$$E[X] = \sum_x xP(X = x) \qquad E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$$
$$E[X^n] = \sum_x x^n P(X = x) \qquad E[X] = \int_{-\infty}^{\infty} x^n f_X(x)dx$$

As before $Var(X) = E[(X - E[X])^2) = E[X^2] - E[X]^2$.

Expectation and Variance

For both discrete and continuous random variables:

$$\begin{aligned} E[aX + b] &= aE[X] + b \\ Var(X) &= E[(X - \mu)^2] = E[X^2] - (E[X])^2 \\ Var(aX + b) &= a^2 Var(X) \end{aligned}$$

(just replace sum with integral in previous proofs)

Joint Cumulative Distribution Function

Suppose X and Y are two random variables.

- $F_{XY}(x, y) = P(X \leq x \text{ and } Y \leq y)$ is the cumulative distribution function for X and Y
- When X and Y are independent then:

$$\begin{aligned} & P(X \leq x \text{ and } Y \leq y) = P(X \leq x)P(Y \leq y) \\ i.e. \quad & F_{XY}(x, y) = F_X(x)F_Y(y) \end{aligned}$$

Joint Cumulative Distribution Function

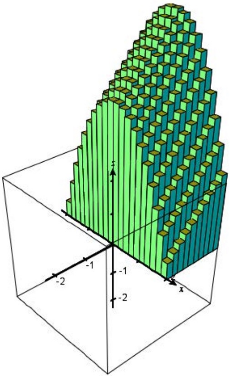
When X and Y are discrete random variables taking values $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_m\}$:

- $F_{XY}(x, y) = \sum_{i:x_i \le x} \sum_{j:y_j \le y} P(X = x_i \text{ and } Y = y_j)$

When X and Y are jointly continuous-valued random variables there exists a probability density function (PDF) $f_{XY}(x, y) \ge 0$ such that:

- $F_{XY}(x, y) = \int_{\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) du\ dv$

Can think of $P(u \le X \le u + du \text{ and } v \le Y \le v + dv) \approx f_{XY}(u, v) du\ dv$ when du, dv are infinitesimally small.



Conditional Probability Density Function

Suppose X and Y are two continuous random variables with joint PDF $f_{XY}(x, y)$. Define conditional PDF:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Compare with conditional probability for discrete RVs:

$$P(X = x|Y = y) = \frac{P(X = x \text{ and} Y = y)}{P(Y = y)}$$

Chain Rule for PDFs

Since

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

the chain rule also holds for PDFs:

$$f_{XY}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$$

Also,

- $\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = \frac{\int_{-\infty}^{\infty} f_{XY}(x,y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$
- We can marginalise PDFs:

$$\begin{aligned}\int_{-\infty}^{\infty} f_{XY}(x,y)dy &= \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dy \\ &= f_X(x)\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy = f_X(x)\end{aligned}$$

Bayes Rule for PDFs

Since

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

then

$$f_{X|Y}(x|y)f_Y(y) = f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x)$$

and so we have Bayes Rule for PDFs:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$$

Independence

Suppose X and Y are two continuous random variables with joint PDF $f_{XY}(x, y)$. Then X are Y are independent when:

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Why ?

$$\begin{aligned} P(X \leq x \text{ and } Y \leq y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) du dv \\ &= \int_{-\infty}^{x} f_X(u) du \int_{-\infty}^{y} f_Y(v) dv \\ &= P(X \leq x) P(Y \leq y) \end{aligned}$$

Example

Suppose random variable $Y = X + M$, where $M \sim N(0,1)$. Conditioned on $X = x$, what is the PDF of Y ?

- $f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y-x)^2}{2})$

Suppose that $X \sim N(0, \sigma)$. What is $f_{X|Y}(x|Y)$?

- Use Bayes Rule:

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} \\
&= \frac{\frac{1}{\sqrt{2\pi}} \exp(-\frac{(y-x)^2}{2}) \times \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2})}{f_Y(y)}
\end{aligned}
$$

- $f_Y(y)$ is just a normalising constant (so that the area under $f_{X|Y}(x|y)$ is 1).