STAT 315: Continuous Random Variables

Luc Rey-Bellet

University of Massachusetts Amherst

luc@math.umass.edu

March 8, 2025

Continuous Random Variables

- Discrete random variables can take a discrete set of possible values like $1, 2, 3, \cdots, 20$ or all integers from $-\infty$ to ∞ and so on....
- Continuous random variables takes a continuous set of possible values like the interval $[0, 1]$ or all positive numbers $[0, \infty)$, and so on...

The pdf of a continuous RV

The probability distribution function of a continuous random variable Y is a function $f(y)$ defined for $y \in (-\infty, \infty)$ such that

- $f(y) \geq 0$
- $\int_{-\infty}^{\infty} f(y)dy = 1$

We compute probabilities by the rule

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

Example

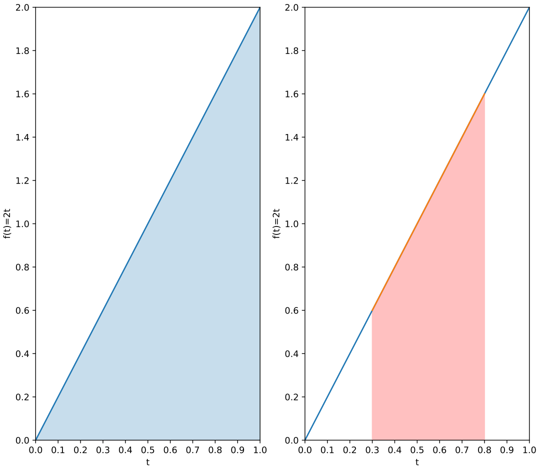


Figure: **Left**: the PDF of Y, area in blue is equal to 1. **Right:** Area in red is $P(.3 \leq Y \leq .8)$.

PDF: $f(t) = 2t$ for $0 \leq t \leq 1$

Normalization $\int_0^1 2t\, dt = 1$

$$
\begin{aligned}
P(.3 \leq Y \leq .8) &= \int_{.3}^{.8} 2tdt \\
&= t^2\big|_{.3}^{.8} \\
&= .64 - .09 \\
&= .55
\end{aligned}
$$

PDF and CDFs

The cumulative distribution function

If Y is a random variable the cumulative distribution function of a random variable Y is given by

$$F(x) = P(Y \leq x)$$

- Continuous random variables: $F(x) = \int_{-\infty}^{x} f(y)dy$
- Discrete random variables $F(x) = \sum_{y:y \leq x} p(y)$

Computing probabilities with the CDF

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

Example

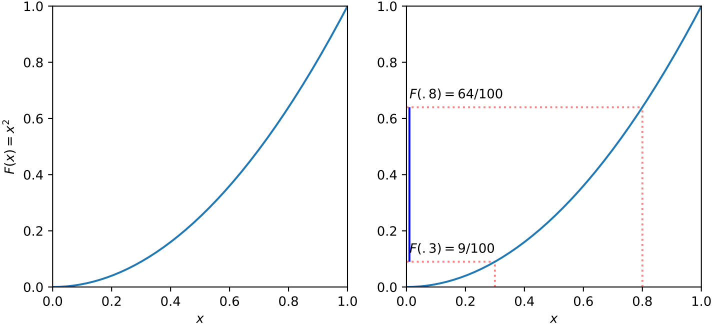


Figure: The CDF of a RV Y with density $f(y) = 2y$

$$F(x) = \begin{cases} 0 & x \le 0 \\ x^2 & 0 \le x \le 1 \\ 1 & x \ge 1 \end{cases} \qquad P(.3 \le Y \le .8) = F(.8) - F(.3) = .64 - .09$$

Examples

- Suppose Y is discrete with $P(0) = .2$, $P(1) = .4$ and $P(2) = .3$, $P(4) = .1$. What is the CDF of Y? Draw a graph.
- Suppose $f(x) = kx(1 - x)$ $0 \leq xl \leq 1$ is the PDF of a continuous random variable.
  - What is the value of k?
  - Find $P(.2 \leq Y \leq .4)$
  - Find $P(.2 < Y < .4)$
  - Find $P(Y \leq .4 | Y \geq .2)$
- Suppose $f(t) = 2e^{-2t}$ for $t \geq 0$ and $f(t) = 0$ otherwise.
  - Check that $f(t)$ is a PDF from some random variable Y.
  - Compute $P(1 \leq Y \leq 2)$.
  - Compute the CDF $F(y)$ for Y.
  - Compute $P(Y > 5 | Y > 3)$

Properties of the CDF

Properties of $F(y)$

The CDF $F(y)$ has the following property

- $F(y) \geq 0$
- $F(y)$ is increasing
- $\lim_{y \to -\infty} F(y) = 0$ and $\lim_{y \to \infty} F(y) = 1$

PDF vs CDF for continuous random variable

By using the fundamental theorem of calculus for continuous RV we have

$$F(x) = \int_{-\infty}^{x} f(y)dy \quad \Longleftrightarrow \quad F'(x) = f(x)$$

Median and percentiles for continuous random variables

Median and p^{th} quantile

The median of Y is the value m such that

$$F(m) = P(Y \leq m) = 1/2$$

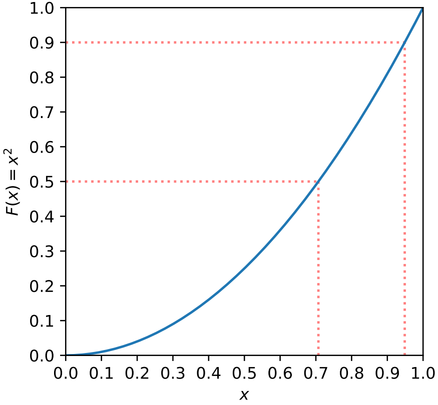
The p^{th} quantile of Y is value ϕ_p such that

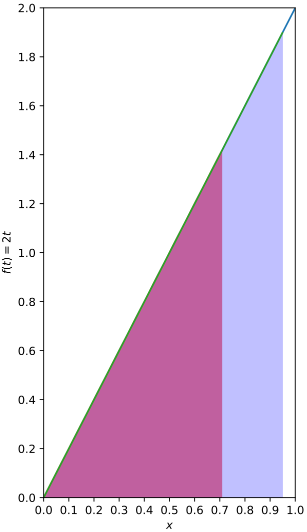
$$F(\phi_p) = P(Y \leq \phi_p) = p$$

That is we can compute quantiles by inverting the CDF and computing the inverse function F^{-1}

Example: The median and 90 percentile for PDF and CDF

$m = F^{-1}(.5) = \sqrt{.5} = .0707..$
$\phi_{.9} = F^{-1}(.9) = \sqrt{.9486..}$





Example: Pareto (power-law) distributions

The Pareto principle (the 80-20 rule) tells us that "20% of the population controls 80% of the total wealth. A reasonable model for this is the following PDF/CDF

$$F(t) = \begin{cases} 1 - \frac{1}{t^\alpha} & t \geq 1 \\ 0 & else \end{cases} \qquad f(t) = F'(t) = \begin{cases} \frac{\alpha}{t^{\alpha+1}} & t \geq 1 \\ 0 & else \end{cases}$$

For example take $\alpha = 1.2 = \frac{6}{5}$ then we have

$$\begin{array}{lll} \text{top}10\% & P(Y > t) = \frac{1}{t^{\frac{6}{5}}} = \frac{1}{10} & t_{.9} = 10^{\frac{5}{6}} = 6.... \\ \text{top}1\% & P(Y > t) = \frac{1}{t^{\frac{6}{5}}} = \frac{1}{100} & t_{.99} = 100^{\frac{5}{6}} = 46.... \\ \text{top}0.1\% & P(Y > t) = \frac{1}{t^{\frac{6}{5}}} = \frac{1}{1000} & t_{.999} = 1000^{\frac{5}{6}} = 316.... \\ \text{top}0.01\% & P(Y > t) = \frac{1}{t^{\frac{6}{5}}} = \frac{1}{10000} & t_{.9999} = 10000^{\frac{5}{6}} = 2154.... \end{array}$$

Expected values of continuous RV

Expected value of continuous RV

For a continuous RV Y with pdf $f(y)$ the expected value of Y is

$$E[Y] = \int_{-\infty}^{\infty} y f(y)\, dy$$

For a function $g(Y)$ of the RV Y we have

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y) f(y)\, dy$$

Expected values of continuous RV, cont'd

Properties of expected value

Properties

$$E[cg(Y)] = cE[g(Y)]$$
$$E[g_1(Y) + g_2(Y)] = E[g_1(Y)] + E[g_2(Y)]$$

The variance

The variance of a continuous random variable Y with pdf $f(y)$ is given by

$$\begin{aligned} V[Y] &= E\left[(Y - E[Y])^2\right] \\ &= E[Y^2] - E[Y]^2 \\ &= \int_{-\infty}^{\infty} y^2 f(y)dy - \left(\int_{-\infty}^{\infty} yf(y)dy\right)^2 \end{aligned}$$

Example

- Suppose Y has the CDF

$$F(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1}{2}y^3 + \frac{1}{2}y^2 & 0 \leq y \leq 1 \\ 1 & y \geq 1 \end{cases}$$

  Compute the mean and the variance of Y.

- Suppose the random variable Y has pdf

$$f(y) = \begin{cases} y & 0 \leq y \leq 1 \\ 2 - y & 1 \leq y \leq 2 \\ 0 & \text{else} \end{cases}$$

  - Find $F(y)$.
  - Compute mean and variance $E[Y]$ and $V(Y)$.