Week 4 – Continuous random variables

Juan Pablo Lewinger

Last class

- Discrete random variables:
  - Take finite or countable value
  - Completely characterized by their pdf or cdf
- Bernoulli: $X = \begin{cases} 1 & Heads \\ 0 & Tails \end{cases}$
- Binomial: number of successes in n independent Bernoulli trials with identical probability of success p.
  - $X \sim Binomial(n, p)$, $x \in \{0, 1, \ldots n\}$
- Geometric distribution: number of trials until first success in repeated Bernoulli experiments with identical probability of success p.
  - $X\ Geometric(p)$, $x \in \{1, 2, \ldots\}$

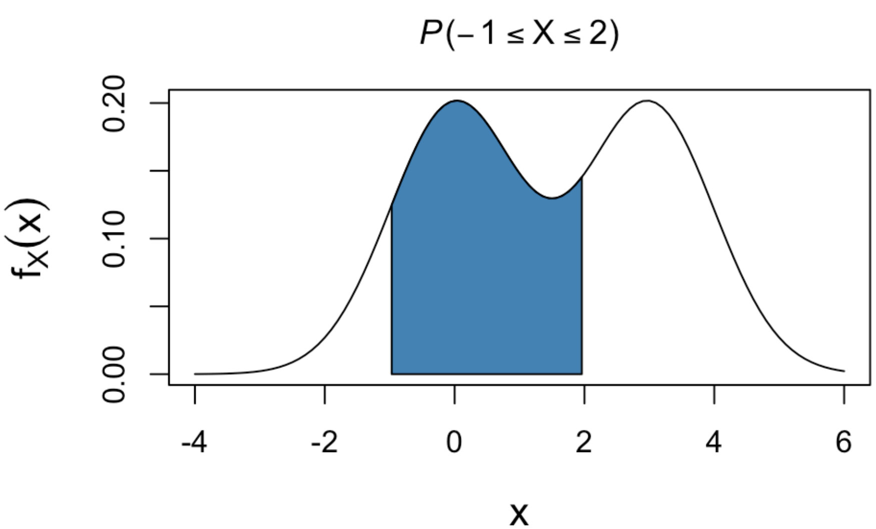
Continuous random variables

A continuous random variable is a function $X : \Omega \longrightarrow \mathbb{R}$ that takes on uncountable infinite values and such that for $a \leq b$:

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

for some function $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$ and $\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$

$f_X : \mathbb{R} \longrightarrow \mathbb{R}$ is called the probability density function (or density function) of the random variable X



Probability density function

- More generally, if $S \subset \mathbb{R} : P(X \in S) = \int_S f_X(x)\,dx$

Properties of the pdf: - $P(-\infty < X < +\infty) = \int_{-\infty}^{\infty} f_X(t)\,dt = 1$

- $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$
- $f_X(x) = \frac{d}{dx}F_X(x) = F_X'(x)$ (Fundamental theorem of calculus, see next slide)
- $P(x - \frac{\epsilon}{2} < X < x + \frac{\epsilon}{2}) = \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f_X(t)\,dt \approx \epsilon f_X(x)$
- $$f_X(x) = \lim_{\epsilon \to 0} \frac{P(x - \frac{\epsilon}{2} < X < x + \frac{\epsilon}{2})}{\epsilon}$$, i.e. represents the density of probability 'mass' at point x
- For a continuous random variable X, $P(X = x) = 0$

Fundamental theorem of calculus

Part I: Let f be a continuous real-valued function defined on a closed interval $[a, b]$. Let F be the function defined, for all x in $[a, b]$, by

$$F(x) = \int_a^x f(t)\,dt$$

Then F is uniformly continuous on $[a, b]$ and differentiable on the open interval (a, b), and

$$F'(x) = f(x)$$

for all x in (a, b)

Part II: Let f be a real-valued function on a closed interval $[a, b]$ and F antiderivative of f in (a, b):

$$F'(x) = f(x)$$

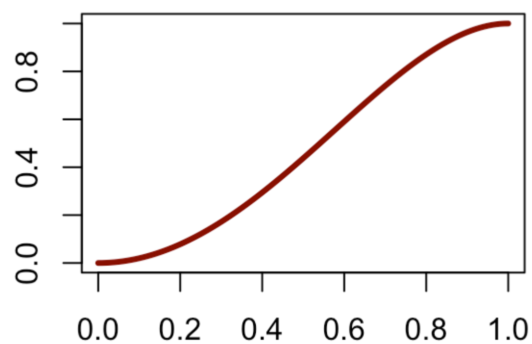
If f is (Riemann) integrable on $[a, b]$ then

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

Example

Example: Let a continuous random variable X be given that takes values in $[0, 1]$, and whose distribution function is given by:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x^2 - x^4 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

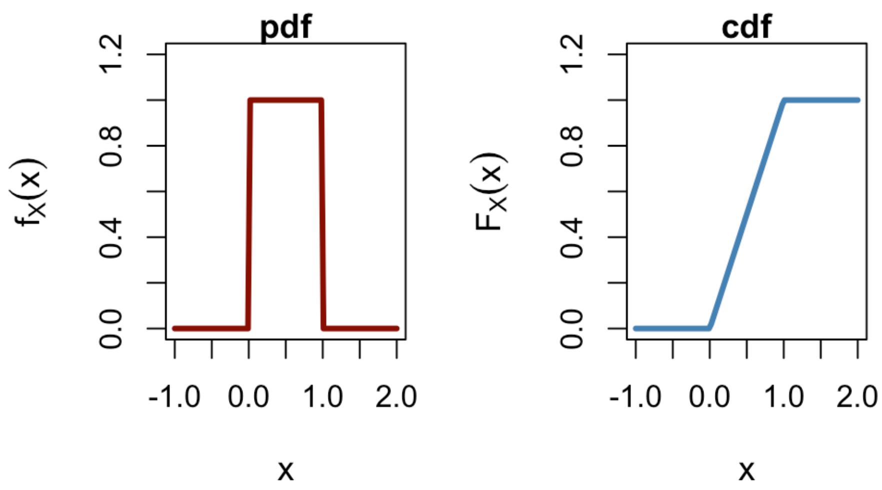


1. Compute $P(\frac{1}{4} \le X \le \frac{3}{4})$
2. What is the probability density function of X?
3. Compute $P(\frac{1}{4} \le X \le \frac{1}{2} \text{ or } X > \frac{3}{4})$

Uniform distribution, $X \sim U[a,b]$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

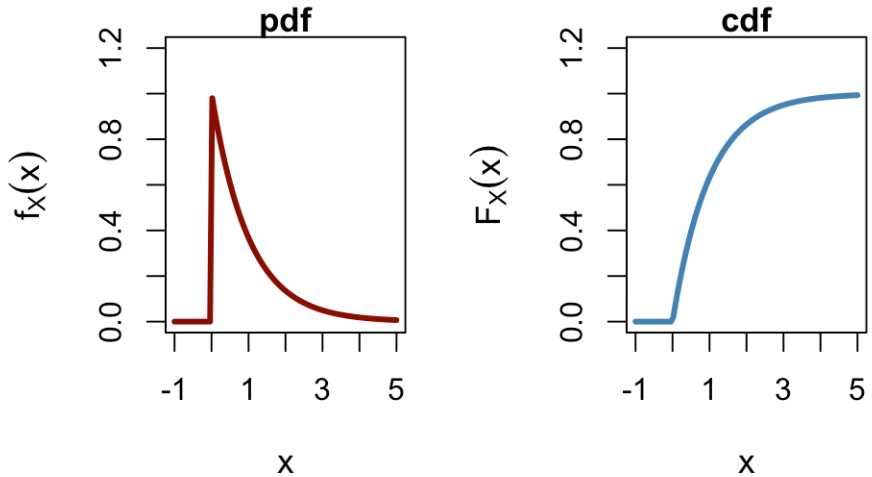
$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x > b \end{cases}$$



Exponential distribution, $X \sim Exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

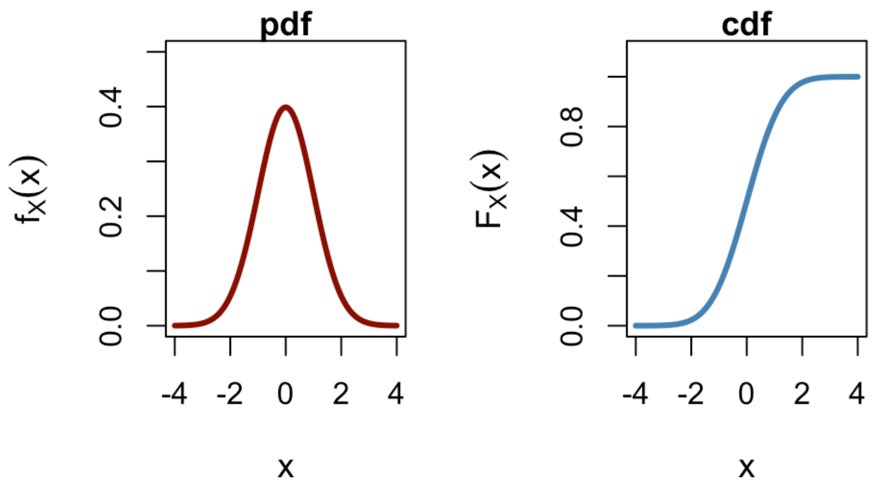
$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$



Normal distribution, $X \sim N(\mu, \sigma^2)$

$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

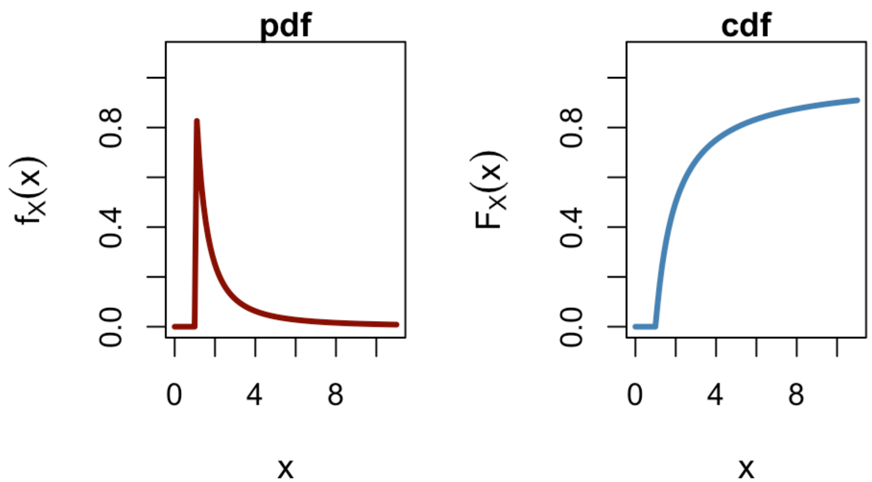
$F(X) = \Phi(x)$ There is no analytical formula but it can be numerically computed



Pareto distribution, $X \sim Pareto(x_m, \alpha)$

$$f(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{if } x \geq x_m \\ 0 & \text{if } x < x_m \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x < x_m \\ 1 - \left(\frac{x_m}{x}\right)^\alpha & \text{if } x \geq x_m \end{cases}$$

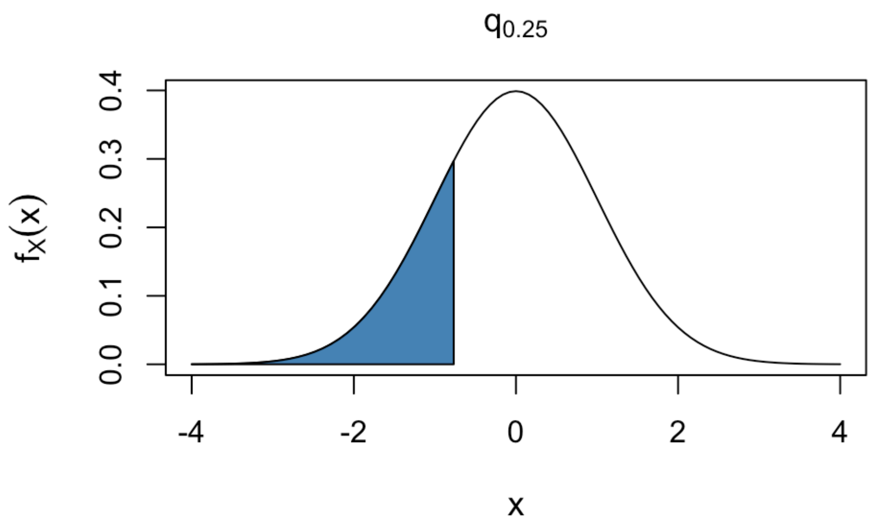


Quantiles

Let X be a continuous random variable and let p be a number between 0 and 1. The p^{th} quantile or $100p^{th}$ percentile of the distribution of X is the smallest number q_p such that:

$$F(q_p) = P(X \leq q_p) = p$$

The median of a distribution is its 50^{th} percentile.



Quantiles

Example 1: median of exponential, $X \sim Exp(\lambda)$

$$\frac{1}{2} = F(q_{0.5}) = P(X \leq q_{0.5}) = 1 - e^{-\lambda q_{0.5}}$$

$$q_{0.5} = -\frac{1}{\lambda}\log(\frac{1}{2})$$

Example 2: median of X uniform in $[1,2] \cup [3,4]$

$$q_{0.5} = 2$$

Uniform distribution in R

pmf, cdf, random generation, and quantile of uniform random variable

```
dunif(3, min=1, max=5)
```

```
## [1] 0.25
```

```
punif(3, min=1, max=5)
```

```
## [1] 0.5
```

```
runif(n=3)  #default is uniform[0,1]
```

```
## [1] 0.007446826 0.944775617 0.292623820
```

```
qunif(0.5, min=1, max=3)
```

```
## [1] 2
```

Exponential distribution in R

pmf, cdf, random generation, and quantile of an exponential random variable

```
dexp(3, rate = 0.5)
```

```
## [1] 0.1115651
```

```
pexp(3, rate = 2)
```

```
## [1] 0.9975212
```

```
rexp(n=5)  #default rate is 1
```

```
## [1] 4.60679804 0.32611096 0.01889765 1.49367378 0.70605987
```

```
c(qexp(p=0.5), -log(1/2))
```

```
## [1] 0.6931472 0.6931472
```

Normal distribution in R

pmf, cdf, random generation, and quantile of a normal random variable

```
dnorm(x=3, mean = 1, sd=2)
```

```
## [1] 0.1209854
```

```
dnorm(x=-1, mean = -2, sd=0.5)
```

```
## [1] 0.1079819
```

```
rnorm(n=5)  #default is mean=0 and sd=1, i.e. a standard normal
```

```
## [1] -0.4843401 -1.0643711  1.5258292  2.2244114  0.7464652
```

```
qnorm(0.5)
```

```
## [1] 0
```

Pareto distribution in R

pmf, cdf, random generation, and quantile of a Pareto random variable with location x_m and shape α

```
library(EnvStats)
```

```
dpareto(x=3, location = 1, shape=2)
```

```
## [1] 0.07407407
```

```
ppareto(q=3, location = 1, shape=2)
```

```
## [1] 0.8888889
```

```
rpareto(n=5, location = 1, shape=2)
```

```
## [1] 1.471863 1.356510 1.347556 1.011351 2.216670
```

```
qpareto(p=0.5, location = 1, shape=2)
```

```
## [1] 1.414214
```

Mixtures of distributions

Example 1: To get to your destination you take a taxi if there is one waiting (probability 1/3) at the stand when you arrive or walk if there is no taxi waiting. A taxi takes you exactly 5 minutes. What is the cdf of the time to your destination T? Walking to your destination takes you exactly 35 minutes.

$$P(T = t) = P(T = t \mid \text{Taxi})P(\text{Taxi}) + P(T = t \mid \text{No taxi})P(\text{No taxi}) = \begin{cases} \frac{1}{3} & \text{if } t = 5 \\ \frac{2}{3} & \text{if } t = 35 \end{cases}$$

$$F_T(t) = \begin{cases} 0 & \text{if } x < 5 \\ \frac{1}{3} & \text{if } 5 \leq x < 35 \\ 1 & \text{if } x \geq 35 \end{cases}$$

Mixtures of distributions

Example 2: To get to your destination you take a taxi if one is waiting (probability 1/3) at the stand when you arrive or walk if there is no taxi waiting. Walking to your destination takes you an amount of time distributed as $Exp(\lambda_1)$ with $\lambda_1 = 1/35$. A taxi takes you an amount of time distributed as $Exp(\lambda_2)$ with $\lambda_2 = 1/5$. What is the cdf of the time to get to your destination, T?

$$F_T(t) = P(T \leq t) = P(T \leq t \mid \text{Taxi})P(\text{Taxi}) + P(T \leq t \mid \text{No taxi})P(\text{No taxi}) =$$

$$= \tfrac{1}{3}(1 - e^{-t/5}) + \tfrac{2}{3}(1 - e^{-t/35})$$

$$f_t(t) = \tfrac{d}{dt}F_T(t) = F_T'(t) = \tfrac{1}{3}(\tfrac{1}{5}e^{-t/5}) + \tfrac{2}{3}(\tfrac{1}{35}e^{-t/35})$$

Mixtures of distributions

Example 3: To get to your destination you take a taxi if one is waiting (probability 1/3) when you arrive or walk if there is no taxi. Walking to your destination takes you exactly 35 minutes. A taxi takes an amount of time distributed as $Exp(\lambda_2)$ with $\lambda_2 = 1/5$. What is the cdf of the time to your destination T?

$$F_T(t) = P(T \leq t) = P(T \leq t \mid \text{Taxi})P(\text{Taxi}) + P(T \leq t \mid \text{No taxi})P(\text{No taxi})$$

$$P(T \leq t \mid \text{No taxi}) = \begin{cases} 0 & \text{if } t < 35 \\ 1 & \text{if } t \geq 35 \end{cases}$$

$$P(T \leq t \mid \text{Taxi}) = \begin{cases} 0 & \text{if } t < 5 \\ 1 - e^{-t/5} & \text{if } t \geq 5 \end{cases}$$

$$F_T(t) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{2}{3}(1 - e^{-t/5}) & \text{if } 0 \leq x < 35 \\ \frac{1}{3} + \frac{2}{3}(1 - e^{-t/5}) & \text{if } x \geq 35 \end{cases}$$

- How about the probability density function?? *A: there isn't one!*
- This is an example of a MIXED random variable
- MIXED random variables are mixtures of discrete and continuous random variables

Discrete, continuous, and mixed random variables

- Discrete random variables have probability mass function but do not have probability density function
- Continuous random variables have probability density function but do not have probability mass function
- Mixed random variables have **neither probability mass function nor probability density function**
- **All types of random variables (discrete, continuous and mixed distributions) have cumulative distribution function!!**

Next week

- Read *4.1.2-4.1.3* from ‘Introduction to Probability, Statistics, and Random Processes’ (do the exercises!)
- Homework 2 due next Wed