Probability Distributions: Continuous

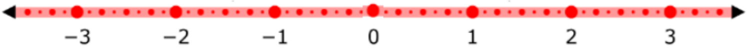
Introduction to Data Science Algorithms

Dirk Grunwald

- Continious random variables
- Uniform
- Exponential
- Gamma
- Pareto
- Normal
- Multvariate normal - used in machine learning
- Dirichlet - used in machine learning

Continuous random variables

- Today we will look at *continuous* random variables:
  - Real numbers: $\mathbb{R}$; $(-\infty, \infty)$
  - Positive real numbers: $\mathbb{R}^+$; $(0, \infty)$
  - Real numbers between -1 and 1 (inclusive): $[-1, 1]$
- The *sample space* of continuous random variables is uncountably infinite.



Continuous distributions

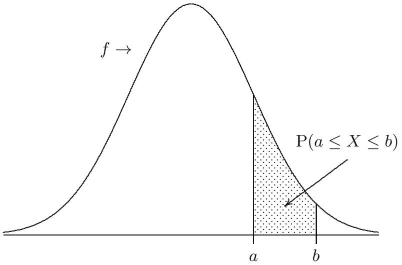
- Last time: a discrete distribution assigns a probability to every possible outcome in the sample space
- How do we define a continuous distribution?
- Suppose our sample space is all real numbers, $\mathbb{R}$.
  - What is the probability of $P(X = 20.1626338)$?
  - What is the probability of $P(X = -1.5)$?

Continuous distributions

- Last time: a discrete distribution assigns a probability to every possible outcome in the sample space
- How do we define a continuous distribution?
- Suppose our sample space is all real numbers, $\mathbb{R}$.
  - What is the probability of $P(X = 20.1626338)$?
  - What is the probability of $P(X = -1.5)$?
- **The probability of any continuous event is always 0.**
  - Huh?
  - There are infinitely many possible values a continuous variable could take. There is zero chance of picking any one exact value.
  - We need a slightly different definition of probability for continuous variables.

Probability density

- A *probability density function* (PDF, or simply *density*) is the continuous version of probability mass functions for discrete distributions.
- The probability at any point is zero, but the probably within a range is defined. For example, $\int_a^b f(x)$ is $P(a \leq X \leq b)$.



- Easier to express using the Cummulative Distribution Function (CDF): $P(a \leq X \leq b) = F(b) - F(a)$.

Probability of intervals

- While the probability for a specific value is 0 under a continuous distribution, we can still measure the probability that a value falls within an interval.
  - $P(X \geq a) = \int_{x=a}^{\infty} f(x) = 1 - F(a)$
  - $P(X \leq a) = \int_{x=-\infty}^{a} f(x) = F(a)$
  - $P(a \leq X \leq b) = \int_{x=a}^{b} f(x) = F(b) - F(a)$
- This is analogous to the disjunction rule for discrete distributions.
  - For example if X is a die roll, then $P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$
  - An integral is similar to a sum

Likelihood

- The *likelihood function* refers to the PMF (discrete) or PDF (continuous).
- For discrete distributions, the likelihood of x is $P(X = x)$.
- For continuous distributions, the likelihood of x is the density $f(x)$.
- We will often refer to likelihood rather than probability/mass/density so that the term applies to either scenario.
- The Cummualative Distribution Function $F_X(x) = P(X \leq x)$ is valid for either discrete or continious distributions.

Uniform Distribution

- $X \sim U(a,b)$: X is distributed as the Uniform distribution over a and b, for $b > a$.
- PDF is $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.
- CDF $F(x) = \frac{x-a}{b-a}$ for $a \leq x \leq b$.

- *Random number generators* typically return samples $U(0,1)$.
- These values are used to generate other random variates

Exponential distribution

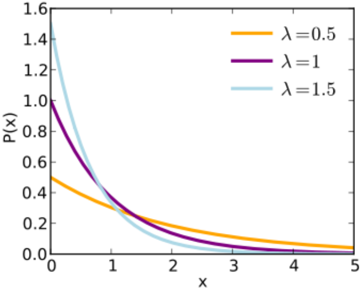
- The exponential distribution is over positive real numbers (including zero), with the highest density at zero and decaying as x increases
- Sample space: $[0, \infty)$
- The probability density function is:

$$
\begin{aligned}
f(x) &= \lambda e^{-\lambda x} \\
F(X) &= 1 - e^{-\lambda x}
\end{aligned}
$$

  - Exponential is *memoryless*: $P[T > s + t \,|\, T > s] = P[T > t]$
  - If interarrival time is $\sim E(\lambda)$ then number of arrivals is Poission distribution with parameter $1/\lambda$. Fundemental to queueing theory, less so for data science.

Exponential distribution

- A good model for:
  - The length of a phone call
  - Time between packets in the internet
  - The time between shooting stars during a meteor shower
  - The distance between cracks in a pipeline
- The parameter $\lambda > 0$ controls how quickly the density decays.



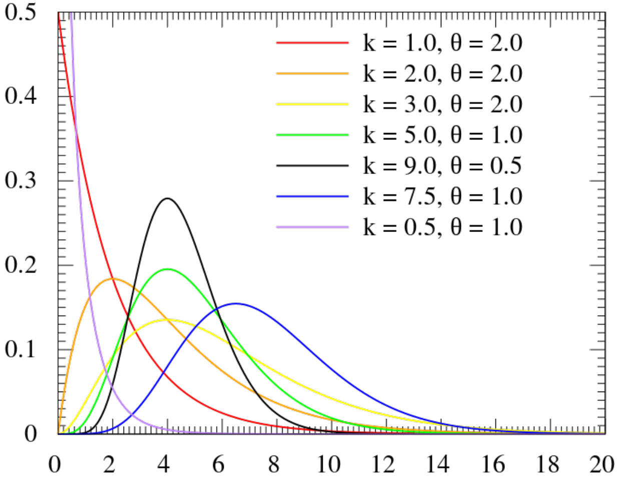
Gamma distribution

- The gamma distribution is a generalization of the exponential distribution (and others)
- Two parameters: shape $k > 0$, scale $\theta > 0$
- PDF:

$$f(x) = \frac{x^{k-1} \exp(-\frac{x}{\theta})}{\theta^k \Gamma(k)}$$

- Equivalent to exponential distribution when $k = 1$, $\theta = \frac{1}{\lambda}$

Gamma distribution



Pareto Distribution

- Pareto distribution: $X \sim f(\alpha)$ where $f(x) = \alpha / x^{\alpha+1}$
- Vilfredo Pareto noticed that income was distributed $\sim C/x^{\alpha}$ for some constants X and $\alpha > 0$.

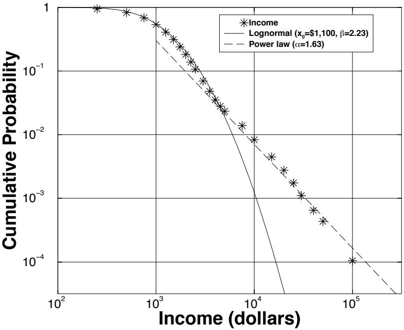


Fig. 1. The power law and lognormal fits to the 1935-36 U.S. income data. The solid line represents the lognormal fit with $x_0 = \$1,100$ and $\beta = 2.23$. The straight dashed line represents the power law fit with $\alpha = 1.63$ (Badger 1980, with permission from Taylor & Francis Ltd.).