Probability Distributions: Continuous

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

FEBRUARY 28, 2017

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 density at a point x is denoted $f(x)$.
- Density behaves like probability:
  - $f(x) \geq 0$, for all x
  - $\int_x f(x) = 1$
- Even though $P(X = 1.5) = 0$, density allows us to ask other questions:
  - Intervals: $P(1.4999 < X < 1.5001)$
  - Relative likelihood: is 1.5 more likely than 0.8?

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.
  - $P(X \geq a) = \int_{x=a}^{\infty} f(x)$
  - $P(X \leq a) = \int_{x=-\infty}^{a} f(x)$
  - $P(a \leq X \leq b) = \int_{x=a}^{b} f(x)$
- 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.

College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

Probability Distributions: Continuous

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

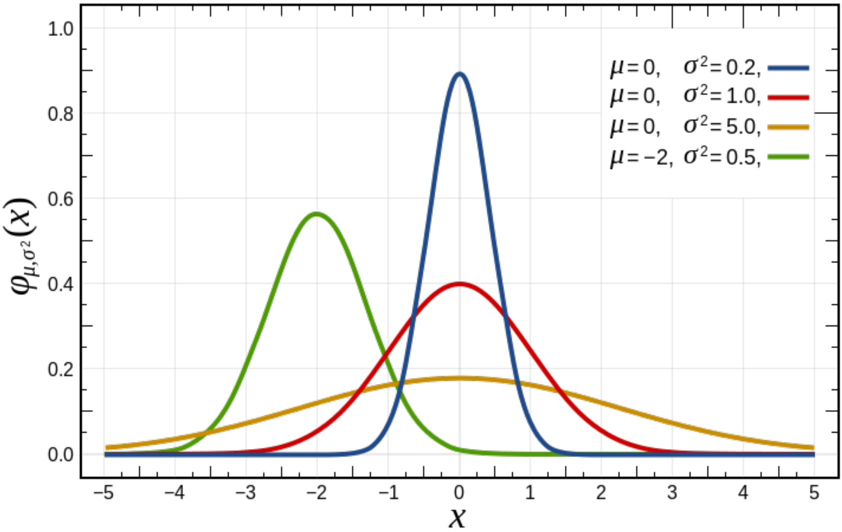
FEBRUARY 28, 2017

The normal distribution

- The most common continuous distribution is the *normal* distribution, also called the *Gaussian* distribution.
- The density is defined by two parameters:
  - μ: the *mean* of the distribution
  - σ^2: the *variance* of the distribution (σ is the *standard deviation*)
- The normal density has a "bell curve" shape and naturally occurs in many problems.

Carl Friedrich Gauss
1777 – 1855

The normal distribution



The normal distribution

- The probability density of the normal distribution is:

$$f(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\substack{\text{Does not} \\ \text{depend on } x}} \underbrace{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}_{\substack{\text{Largest when } x=\mu; \\ \text{shrinks as } x \text{ moves} \\ \text{away from } \mu}}$$

- Notation: $\exp(x) = e^x$
- If X follows a normal distribution, then $\mathbb{E}[X] = \mu$.
- The normal distribution is symmetric around μ.

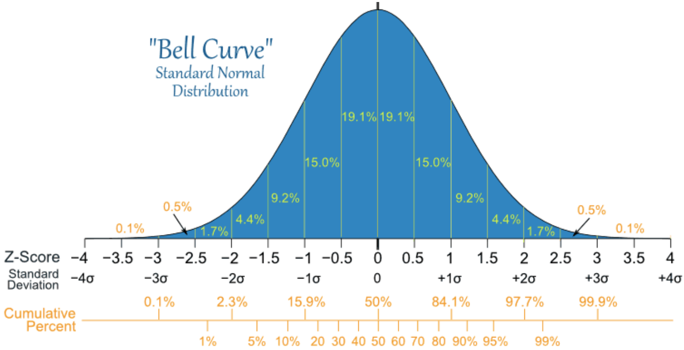
The normal distribution

- What is the probability that a value sampled from a normal distribution will be within *n* standard deviations from the mean?
- $P(\mu - n\sigma \leq X \leq \mu + n\sigma) = ?$

The normal distribution

- What is the probability that a value sampled from a normal distribution will be within *n* standard deviations from the mean?
- $P(\mu - n\sigma \leq X \leq \mu + n\sigma) = ?$
$= \int_{x=\mu-n\sigma}^{\mu+n\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x=\mu-n\sigma}^{\mu+n\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

The normal distribution



Applying the normal distribution

- Most variables in the real world don't follow an exact normal distribution, but it is a very good approximation in many cases.

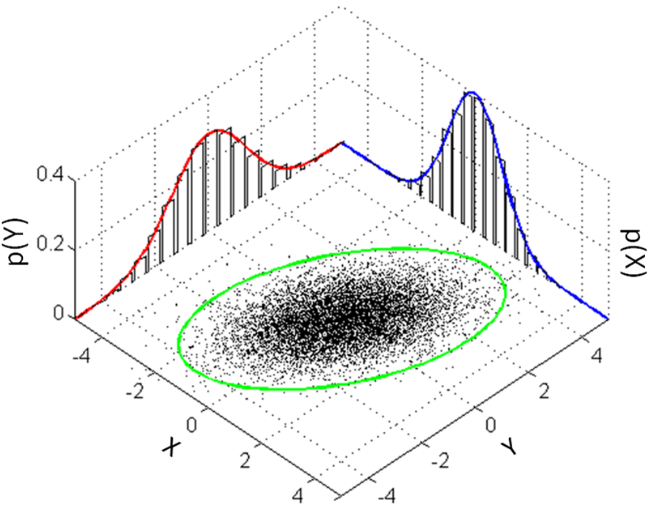
- Measurement error (e.g., from experiments) is often assumed to follow a normal distribution.
- Biological characteristics (e.g., heights of people, blood pressure measurements) tend to be normal distributed.
- Test scores

- Special case: sums of multiple random variables
  - The *central limit theorem* proves that if you take the sum of multiple randomly generated values, the sums will follow a normal distribution. (Even if the randomly generated values do not!)

Multivariate normal distribution

- What is the *joint* distribution over multiple normal variables?
- If the normal random variables are independent, the joint distribution is just the product of each individual PDF.
- But they don't have to be independent.
- We can model the joint distribution over multiple variables with the *multivariate* normal distribution.

Multivariate normal distribution



Multivariate normal distribution

- The multivariate normal distribution is a distribution over a *vector* of values $\mathbf{x}$. The mean μ is also a vector.
- In addition to the variance of each variable, each pair of variables has a *covariance*.
  - The covariance matrix for all pairs is denoted Σ.
  - The covariance indicates an association between variables. If it is positive, it means if one value increases (or decreases), the other value is also likely to increase (or decrease). If the covariance is negative, it means that if one value increases, the other is likely to decrease, and vice versa.
- $f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right)$

Computing the Mean

- If you have observations $x_1 \dots x_N$ that come from a normal distribution, what is the mean μ?
- Formula

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

Computing the Variance

- If you have observations $x_1 \dots x_N$ that come from a normal distribution, what is the variance σ^2?
- Formula

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} \tag{2}$$

Probability Distributions: Continuous

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

FEBRUARY 28, 2017

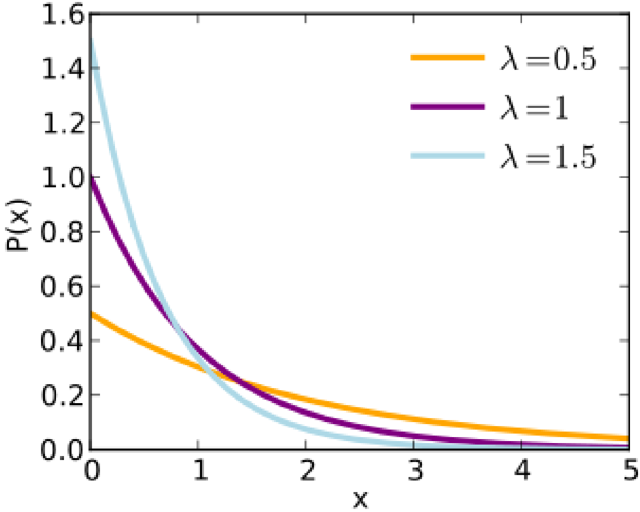
Exponential distribution

- 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:

$$f(x) = \lambda \exp(-\lambda x)$$

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

Exponential distribution



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 $k = 1$, $\theta = \frac{1}{\lambda}$

Exponential distribution