# ISTA 116: Statistical Foundations for the Information Age

Continuous Random Variables

9 November 2011



## Outline





## Types of Random Variables

- We can classify random variables based on the types of values they take on.
- **Discrete** random variables take on discrete values (e.g., categories or integers)
- **Continuous** random variables take on continuous values (e.g., real numbers).



## Continuous Random Variables

- When random variables take on continuous values, there are necessarily infinitely many distinct possibilities.
- What's the probability of any one of them?
- Sort of like throwing a dart and hitting a target the size of an atom; but less probable.
- The strange thing about the infinity of continuity is that individual values can have probability zero, but the whole range still has a probability of 1.
  - Zeno's Arrow

## Continuous CDF

- Since individual points have probability zero, a probability mass function wouldn't be informative.
- However, the CDF still makes sense. It's defined exactly the same way as for discrete random variables:

**(The Cumulative Distribution Function)**

A random variable, $X$, can be characterized by its **cumulative distribution function**, $F_X$, which takes values and returns *cumulative* probabilities:

$$F_X(x) = P(X \leq x)$$



## Probabilities in Intervals

- Most of the time, the events we're interested in are made up of *intervals*: What's the probability of falling between two particular values.
- In these cases, the fact that points have zero probability actually simplifies our life a bit: we don't have to distinguish between strict and soft inequalities:

$$\begin{aligned} P(X \leq x) &= P(X < x) + P(X = x) \\ &= P(X < x) + 0 \\ &= P(X < x) \\ P(X \geq x) &= P(X > x) + P(X = x) \\ &= P(X > x) + 0 \\ &= P(X > x) \end{aligned}$$



## Probabilities in Intervals

**(Computing Interval Probabilities for Continuous RVs)**

1. For a continuous random variable $X$, the CDF also gives us $P(X < a)$.
2. So, $P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X \leq b)$
3. All of the above are equal to $F_X(b) - F_X(a)$.
4. Similarly, $P(X > x) = P(X \geq x) = 1 - F_X(x)$

- Probabilities of most more complex events can be obtained from the rules of probability.



## Density

- Although points have neither "mass" nor "volume", they can still have **density**.
- Density is like the "speed of accumulation" of probability at a certain point.
- For super small intervals centered at the point, we can approximate the probability by density times length.

## The Probability Density Function

- A probability mass function is useless for a continuous distribution; but a probability *density* function is informative.

**(The Probability Density Function)**

For a continuous random variable, its **probability density function** gives the "rate of change" (amount of probability accumulated per unit) in the CDF at a point:

$$f_X(x) = \frac{d}{dx} F_X(x)$$



## The PDF and Probability as Area

- If we graph the PDF of a random variable, then probabilities are areas under the curve
- Area (Probability "Mass") $\approx$ Width (Volume) $\times$ Height (Density)

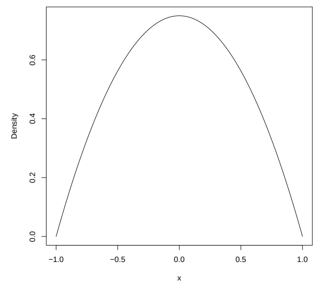


Figure: An Example PDF graph



## The PDF and Probability as Area

- In particular, the CDF is the total area under the curve to the left of a point

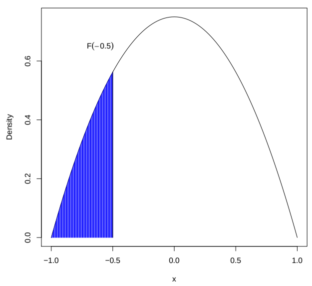


Figure: Value of the CDF is the Area of the Blue Region



## PDF and CDF Relationship

- We can invert the relationship between the PDF and CDF:

**(PDF CDF Relationship)**

For a continuous random variable, the CDF at a point is the "area under the (PDF) curve" to the left of the point:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

## Mean and Variance

- The mean and variance of a continuous random variable are the same as for a discrete distribution, but replacing sums with integrals, and the PMF with the PDF (times an "infinitesimal" differential):

**(Mean and Variance of a Continuous Random Variable)**

For a continuous random variable $X$:

$$\mu_X = E(X) = \int_{\text{Range of } X} x \cdot f(x)dx$$

$$\sigma_X^2 = E((X-\mu_X)^2) = \int_{\text{Range of } X} (x-\mu_X)^2 \cdot f(x)dx$$



## The Continuous Uniform Distribution

- The simplest example of a continuous random variable is one with a (continuous) **uniform distribution**
- Recall that the discrete uniform distribution had the same probability at every value in a range.
- The continuous uniform distribution has the same *density* at every (continuous) value in a range.
- Its only parameters are the locations of the endpoints of the range.
  - We write

$$X \sim \mathcal{U}(a,b)$$



## The Continuous Uniform Distribution

- If $X$ ranges from $a$ to $b$, has the same density everywhere, what must that density be?
  - Version 1: You want to travel 1 mile, starting at $a$ o'clock, and ending at $b$ o'clock, traveling at a constant speed. What speed do you need to go?
  - Version 2: You have to draw a straight line on a graph from point $a$ to point $b$ which is the top of a rectangle with an area of 1. How high must the rectangle be?



## The Continuous Uniform Distribution

- If $X \sim \mathcal{U}(a,b)$, what is $F_X(x)$?
  - Version 1: If you're going at a constant rate, $\frac{1}{b-a}$ miles per hour, starting at $a$ o'clock, how far will you have gone by $x$ o'clock?
  - Version 2: If your rectangle goes from $a$ on the left to $x$ on the right, and is $\frac{1}{b-a}$ units high, what's its area?

## The Continuous Uniform Distribution

- If $X \sim \mathcal{U}(a, b)$, where should its mean be?
  - Hint: Think about the mean-as-balance-point
- The variance is $\frac{1}{12}(b-a)^2$ (we'd need calculus to go through the derivation).



## The Normal Distribution

- Probably the most important distribution in statistics is the **Normal Distribution**.
- This is the distribution whose density is a "bell curve".
- Lots of things in nature are approximately Normally distributed:
  - Heights
  - Blood Pressures
  - IQs
  - Machine-made part sizes
- In general, when there are lots of tiny, independent factors influencing a quantity, it tends toward a Normal distribution (we'll see why this is so later)



## Properties of the Normal Distribution

- The Normal Distribution is directly parameterized by its mean ($\mu$) and its standard deviation ($\sigma$).
- Its PDF (although we won't do calculations with it in this class) is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$$

- Its CDF doesn't have an algebraic form. In R we can use `pnorm()` to compute it to a good approximation.



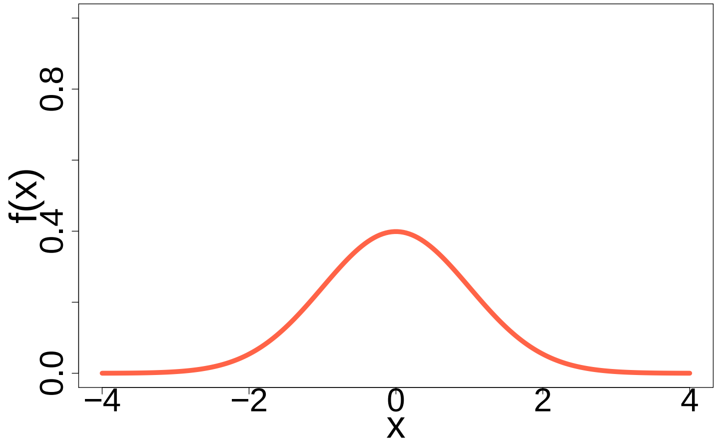


Figure: Density of the $\mathcal{N}(0, 1)$ distribution

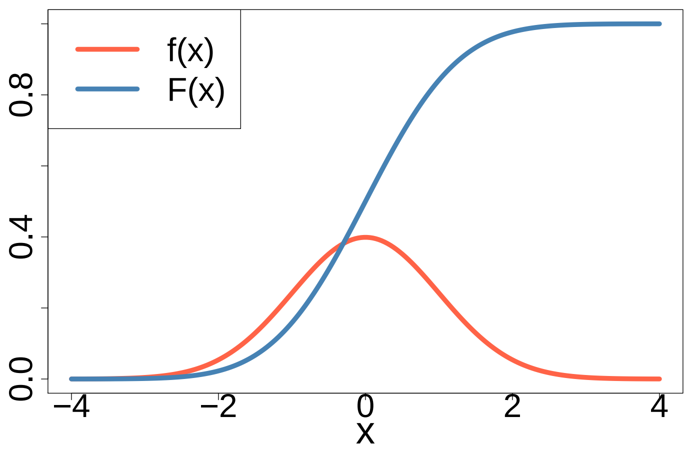


Figure: Density and CDF of the $\mathcal{N}(0,1)$ distribution



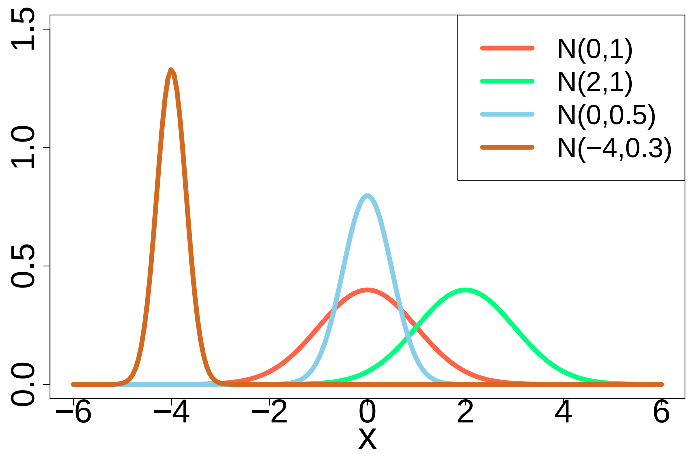


Figure: Densities of Normal Distributions with Various $\mu$ and $\sigma$



## The Standard Normal Distribution

- What do you think happens if we take a random variable $X \sim \mathcal{N}(\mu, \sigma)$, and convert all its values into $z$-scores?
- What distribution will the $z$-scores have?
- What happens to the mean of a data set if we subtract the mean from every value?
- What happens to the standard deviation of a data set if we divide every value by the standard deviation?
- What happens to the *shape* of the distribution?



## The Standard Normal Distribution

- The $\mathcal{N}(0,1)$ is obtained by converting *any* Normal distribution to $z$-scores.
- This distribution is given a special name: the **Standard Normal** distribution.
- Sometimes use $Z$ to mean a random variable that has the Standard Normal distribution.
- If $X \sim \mathcal{N}(\mu, \sigma)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.
- You will sometimes see $\phi(z) = f_Z(z)$ and $\Phi(z) = F_Z(z)$ to represent the Standard Normal PDF and CDF.

## The Standard Normal Distribution

- What's the point?
- There's no formula for the CDF of the Normal; it has to be approximated. *But*, any probability for any Normal can be computed by converting to $z$-scores, and computing probabilities for the Standard Normal.
- We can compute Standard Normal CDF values once, and store the results to be used for any later computations we might need.
- Before modern computers, Standard Normal CDF tables had to be used.



## Normal Probabilities

- For example, if we have $X \sim \mathcal{N}(\mu, \sigma)$, then

$$\begin{aligned}
P(a < X < b) &= P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}) \\
&= P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}) \\
&= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}) \\
P(X < b) &= \Phi(\frac{b-\mu}{\sigma}) \\
P(X > a) &= 1 - \Phi(\frac{a-\mu}{\sigma})
\end{aligned}$$



## Some Rules of Thumb

- You can do quick, "back of the envelope" approximations of Normal probabilities by remembering a few common values. If $X \sim \mathcal{N}(\mu, \sigma)$:

$$\begin{aligned}
P(\mu - \sigma \leq X \leq \mu + \sigma) &= P(-1 < Z < 1) \approx 0.680 \\
P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= P(-2 < Z < 2) \approx 0.950 \\
P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= P(-3 < Z < 3) \approx 0.997
\end{aligned}$$

- In other words, only 32%, 5% and 0.3% of Normally distributed data lies more than 1, 2 and 3 standard deviations from the mean (respectively).
- This is known as the $68 - 95 - 99.7$ rule of thumb



## Some Rules of Thumb

- Often we are interested in the reverse: what *values* capture a certain *probability*? (I.e., what's a "typical" range?)
- For example, we might want to solve for $q$ to capture 95% of cases:

$$P(q \leq X \leq q) = P(\frac{-q-\mu}{\sigma} < Z < \frac{q-\mu}{\sigma}) \approx 0.95$$

- Or, in terms of the complement:

$$P(|X| > q) = P(|Z| < \frac{q-\mu}{\sigma}) \approx 1 - 0.95$$