STAT 237
Random Variables

March 4-7, 2022

Colin Reimer Dawson

Outline

Event

An **event** is a **subset** of the sample space that does or does not contain a particular outcome/possible world.

- The coin comes up heads.
- The coin is fair.
- The last train arrived 20 minutes ago.
- The next train is operating normally.
- The next train will arrive in 10 minutes.

Three Kinds of Events

.1 Some events describe **observations** (or **data**)
  - **Examples:** the coin comes up heads, the next train will arrive in 10 minutes
  - Both Bayesian and frequentist probability use these

.2 Other events describe **hidden world states** (or **hypotheses**) (whose truth value may *never* be observable directly, or may be known to some observers but not others)
  - **Examples:** The coin is fair, the last train arrived 10 minutes ago, the next train is operating normally
  - In frequentist probability, each defines a **separate** sample space. In Bayesian probability, they are subsets of **one**

.3 Still others describe **conjunctions** or **intersections** of data and hypotheses
  - **Examples:** The coin is fair **and** the next flip will be heads, the next train is operating normally **and** it will arrive in 10 minutes

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:
  - The train will arrive in 10 minutes

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:
  - The train will arrive in 10 minutes
  - The train will arrive in 11 minutes

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:
  - The train will arrive in 10 minutes
  - The train will arrive in 11 minutes
  - The train will arrive in 11 minutes and 30 seconds

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:
  - The train will arrive in 10 minutes
  - The train will arrive in 11 minutes
  - The train will arrive in 11 minutes and 30 seconds
  - ...

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:
  - The train will arrive in 10 minutes
  - The train will arrive in 11 minutes
  - The train will arrive in 11 minutes and 30 seconds
  - ...
- We can save computation and "clutter" if we define **collections of mutually exclusive events** that differ along specific **characteristics** of interest

Organizing Events of Interest

- It would be cumbersome to have to list out every possible event of interest:
  - The train will arrive in 10 minutes
  - The train will arrive in 11 minutes
  - The train will arrive in 11 minutes and 30 seconds
  - ...
- We can save computation and "clutter" if we define **collections of mutually exclusive events** that differ along specific **characteristics** of interest
- This gives us: **Random Variables**

Outline

Random Variable

Random Variable

- A **random variable** represents some **characteristic** of a given element of the sample space.
  - Time between arrivals of the train
  - Number of black marbles we see after 3 draws
- Different outcomes can have the same or different values of a given random variable
- **Key consequence:** A random variable **partitions a sample space** into non-overlapping events, each consisting of all "worlds" that share a specific value of the variable

Example

- Let Ω = all sequences of 3 coin tosses.
- We can define a **random variable**, X, that counts **number of heads**.
- Then HHT and HTH, though different **elements** of Ω, are treated by X as equivalent:

Distribution of a Random Variable

- The expression $P(X = x)$ refers to the probability of the event consisting of **all elements of the sample space where the characteristic X refers to has the value x**

Distribution of a Random Variable

- The expression $P(X = x)$ refers to the probability of the event consisting of **all elements of the sample space where the characteristic X refers to has the value x**
- The set of possible values X can have is called its **range**

Distribution of a Random Variable

- The expression $P(X = x)$ refers to the probability of the event consisting of **all elements of the sample space where the characteristic X refers to has the value x**
- The set of possible values X can have is called its **range**
- Roughly: Random variables whose ranges are **finite** or **integer valued** are called **discrete random variables**

Distribution of a Random Variable

- The expression $P(X = x)$ refers to the probability of the event consisting of **all elements of the sample space where the characteristic X refers to has the value x**
- The set of possible values X can have is called its **range**
- Roughly: Random variables whose ranges are **finite** or **integer valued** are called **discrete random variables**
- Roughly: Random variables whose ranges are **real numbers** (or an interval of real numbers) are **continuous random variables**

Outline

Discrete Random Variables

Probability Mass Function

For a discrete random variable, we will use the shorthand $p(x)$ (or $p_X(x)$ if we need to be explicit about the random variable) to represent $P(X = x)$.

The function p (or p_X), which takes a value x and gives us the probability that X takes that value, is called the **probability mass function** (PMF) of X.

Discrete Random Variables

Probability Mass Function

For a discrete random variable, we will use the shorthand $p(x)$ (or $p_X(x)$ if we need to be explicit about the random variable) to represent $P(X = x)$.

The function p (or p_X), which takes a value x and gives us the probability that X takes that value, is called the **probability mass function** (PMF) of X.

"Unity" of the PMF

The probability axioms imply that

$$\sum_{x \in \text{Range}(X)} p(x) = 1$$

Example: Bernoulli Distribution

- If X can only take two possible values, we can label one of them 1 and the other 0. Then X has a **Bernoulli distribution**

Example: Bernoulli Distribution

- If X can only take two possible values, we can label one of them 1 and the other 0. Then X has a **Bernoulli distribution**
- Since $p_X(0)$ is determined by $p_X(1)$ (because $p_X(0) + p_X(1) = 1$), the PMF of X is **determined by a single number**

Example: Bernoulli Distribution

- If X can only take two possible values, we can label one of them 1 and the other 0. Then X has a **Bernoulli distribution**
- Since $p_X(0)$ is determined by $p_X(1)$ (because $p_X(0) + p_X(1) = 1$), the PMF of X is **determined by a single number**
- Defining $\mu := p_X(1)$, the PMF is then

$$p_X(x) = \begin{cases} \mu & \text{for } \mathsf{x} = 1 \\ 1 - \mu & \text{for } \mathsf{x} = 0 \\ 0 & \text{otherwise} \end{cases}$$

Example: Bernoulli Distribution

- If X can only take two possible values, we can label one of them 1 and the other 0. Then X has a **Bernoulli distribution**
- Since $p_X(0)$ is determined by $p_X(1)$ (because $p_X(0) + p_X(1) = 1$), the PMF of X is **determined by a single number**
- Defining $\mu := p_X(1)$, the PMF is then

$$p_X(x) = \begin{cases} \mu & \text{for x = 1} \\ 1 - \mu & \text{for x = 0} \\ 0 & \text{otherwise} \end{cases}$$

- The value μ is called a **parameter** of the distribution of X

Bernoulli PMF, More Concisely

The following are equivalent ways of writing the PMF of a Bernoulli random variable

$$p_X(x) = \begin{cases} \mu & \text{for x = 1} \\ 1 - \mu & \text{for x = 0} \\ 0 & \text{otherwise} \end{cases}$$

$$p_X(x) = \begin{cases} \mu^x (1 - \mu)^{1-x} & \text{if x = 0,1} \\ 0 & \text{otherwise} \end{cases}$$

Outline

Parameters and Hypotheses

- Many random variables have distributions (PMFs for discrete R.V.s) that are defined by a small number of **parameters** (often one or two)

Parameters and Hypotheses

- Many random variables have distributions (PMFs for discrete R.V.s) that are defined by a small number of **parameters** (often one or two)
- Examples:

Parameters and Hypotheses

- Many random variables have distributions (PMFs for discrete R.V.s) that are defined by a small number of **parameters** (often one or two)
- Examples:
  - Bernoulli PMF is defined by $\mu := p(1)$

Parameters and Hypotheses

- Many random variables have distributions (PMFs for discrete R.V.s) that are defined by a small number of **parameters** (often one or two)
- Examples:
  - Bernoulli PMF is defined by $\mu := p(1)$
  - Poisson PMF is defined by λ (the mean)

Parameters and Hypotheses

- Many random variables have distributions (PMFs for discrete R.V.s) that are defined by a small number of **parameters** (often one or two)
- Examples:
  - Bernoulli PMF is defined by $\mu := p(1)$
  - Poisson PMF is defined by λ (the mean)
  - Normal distribution (continuous) is defined by μ (the mean) and σ (the standard deviation)

Parameters and Hypotheses

- Many random variables have distributions (PMFs for discrete R.V.s) that are defined by a small number of **parameters** (often one or two)
- Examples:
  - Bernoulli PMF is defined by $\mu := p(1)$
  - Poisson PMF is defined by λ (the mean)
  - Normal distribution (continuous) is defined by μ (the mean) and σ (the standard deviation)
- Often in statistics we have a reasonable model of **what family** a random variable's distribution is in (e.g., Bernoulli, Normal), but don't know the **value(s) of the parameter(s)**

Parameters and Hypotheses

- In a Bayesian framework, that means there is a **possible world**, or **hypothesis** corresponding to **each possible value of a parameter**

Parameters and Hypotheses

- In a Bayesian framework, that means there is a **possible world**, or **hypothesis** corresponding to **each possible value of a parameter**
- Then, **within each of these hypotheses** (that is, **conditioned on** the parameter value), the random variable in question has the distribution defined by that parameter value.

Parameters and Hypotheses

- In a Bayesian framework, that means there is a **possible world**, or **hypothesis** corresponding to **each possible value of a parameter**
- Then, **within each of these hypotheses** (that is, **conditioned on** the parameter value), the random variable in question has the distribution defined by that parameter value.
- Examples:

Parameters and Hypotheses

- In a Bayesian framework, that means there is a **possible world**, or **hypothesis** corresponding to **each possible value of a parameter**
- Then, **within each of these hypotheses** (that is, **conditioned on** the parameter value), the random variable in question has the distribution defined by that parameter value.
- Examples:
  - If a coin is fair, then random variable X representing the outcome of a single flip has a Bernoulli distribution with $\mu = 0.5$ (where $X = 1$ means we got heads and $X = 0$ means we got tails)

Parameters and Hypotheses

- In a Bayesian framework, that means there is a **possible world**, or **hypothesis** corresponding to **each possible value of a parameter**
- Then, **within each of these hypotheses** (that is, **conditioned on** the parameter value), the random variable in question has the distribution defined by that parameter value.
- Examples:
  - If a coin is fair, then random variable X representing the outcome of a single flip has a Bernoulli distribution with $\mu = 0.5$ (where $X = 1$ means we got heads and $X = 0$ means we got tails)
  - If a coin has a tendency to favor heads slightly, perhaps $\mu = 0.51$

Parameters and Hypotheses

- In a Bayesian framework, that means there is a **possible world**, or **hypothesis** corresponding to **each possible value of a parameter**
- Then, **within each of these hypotheses** (that is, **conditioned on** the parameter value), the random variable in question has the distribution defined by that parameter value.
- Examples:
  - If a coin is fair, then random variable X representing the outcome of a single flip has a Bernoulli distribution with $\mu = 0.5$ (where $X = 1$ means we got heads and $X = 0$ means we got tails)
  - If a coin has a tendency to favor heads slightly, perhaps $\mu = 0.51$
  - We could say that **conditioned on** μ, X has a Bernoulli(μ) distribution

Conditional Distribution / PMF

- If a random variable X has a specific PMF when we restrict our attention to the specific hypothesis corresponding to the value of a parameter, we say it has that **conditional PMF**

Conditional Distribution / PMF

- If a random variable X has a specific PMF when we restrict our attention to the specific hypothesis corresponding to the value of a parameter, we say it has that **conditional PMF**
- Example: if conditioned on each value of μ, X has a Bernoulli(μ) distribution, we'd write

$$p_{X|\mu}(x \mid \mu) = \begin{cases} \mu^x(1-\mu)^{1-x} & \text{if x = 0,1} \\ 0 & \text{otherwise} \end{cases}$$

where $p_{X|\mu}$ represents the collection of possible **conditional PMFs** of X (one for each value of μ).

Conditional Distribution / PMF

- If a random variable X has a specific PMF when we restrict our attention to the specific hypothesis corresponding to the value of a parameter, we say it has that **conditional PMF**
- Example: if conditioned on each value of μ, X has a Bernoulli(μ) distribution, we'd write

$$p_{X|\mu}(x \mid \mu) = \begin{cases} \mu^x(1-\mu)^{1-x} & \text{if x = 0,1} \\ 0 & \text{otherwise} \end{cases}$$

  where $p_{X|\mu}$ represents the collection of possible **conditional PMFs** of X (one for each value of μ).
- More concisely, we can write

$$X \mid \mu \sim \text{Bernoulli}(\mu)$$

  which reads as "conditioned on μ, X **is distributed as** a Bernoulli distribution with parameter μ"

Outline

Expected Value

- The “average” value of a random variable is its **expected value**: a weighted average of the possible values it can take, where weights come from the PMF/PDF.

Expected Value

- The "average" value of a random variable is its **expected value**: a weighted average of the possible values it can take, where weights come from the PMF/PDF.

Expected Value (Discrete Case)

For a discrete random variable, X, with PMF $p(x)$, the **expected value** of X is written $\mathbb{E}[X]$ and defined as

$$\mathbb{E}[X] := \sum_{x \in \text{Range(X)}} x p(x) \tag{1}$$

Expected Value

- The "average" value of a random variable is its **expected value**: a weighted average of the possible values it can take, where weights come from the PMF/PDF.

Expected Value (Discrete Case)

For a discrete random variable, X, with PMF $p(x)$, the **expected value** of X is written $\mathbb{E}[X]$ and defined as

$$\mathbb{E}[X] := \sum_{x \in \text{Range(X)}} x p(x) \tag{1}$$

This just yields a **number**, which is also called the **mean** of X.

Example

Mean of a Bernoulli

A Bernoulli R.V. with parameter μ has (conditional) PMF

$$p(x \mid \mu) = \mu^x (1 - \mu)^{1-x} \text{ for } x = 0, 1$$

Example

Mean of a Bernoulli

A Bernoulli R.V. with parameter μ has (conditional) PMF

$$p(x \mid \mu) = \mu^x (1 - \mu)^{1-x} \text{ for } x = 0, 1$$

The (conditional) mean is therefore

$$\mathbb{E}[X \mid \mu] = 0 \cdot p(0 \mid \mu) + 1 \cdot p(1 \mid \mu) = 1 \cdot \mu^1 (1 - \mu)^0 = \mu$$

Expected Value of a Function

- We can compute the expected value of an **arbitrary function** of the random variable:

Expected Value of a Function

- We can compute the expected value of an **arbitrary function** of the random variable:

Expected Value of a Function of a R.V.

If f is a function mapping a real number, x, to another real number, $f(x)$, then we define

$$\mathbb{E}\left[f(X)\right] := \sum_{x \in \text{Range(X)}} f(x)p(x)$$

Variance

- A function of particular interest is $f(x) = (x - \mu)^2$, where μ is the mean (a constant), i.e., $\mu = \mathbb{E}[X]$.
- In this case, $\mathbb{E}[f(X)]$ is called the **variance**.

Variance

- A function of particular interest is $f(x) = (x - \mu)^2$, where μ is the mean (a constant), i.e., $\mu = \mathbb{E}[X]$.
- In this case, $\mathbb{E}[f(X)]$ is called the **variance**.

Variance of a R.V.

The **variance** of X is $\mathbb{E}[(X-\mu)^2]$ and measures the **average squared distance that the variable is from its mean**.

$$\mathbb{V}\mathrm{ar}[X] = \mathbb{E}\left[(X-\mu)^2\right] = \sum_{x \in \text{Range(X)}} (x-\mu)^2 p(x)$$

Example

Variance of a Bernoulli

A Bernoulli R.V. with parameter μ has (conditional) PMF

$$p(x \mid \mu) = \mu^x (1 - \mu)^{1-x} \text{ for } x = 0, 1$$

Example

Variance of a Bernoulli

A Bernoulli R.V. with parameter μ has (conditional) PMF

$$p(x \mid \mu) = \mu^x (1 - \mu)^{1-x} \text{ for } x = 0, 1$$

The (conditiona) variance is, therefore:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[X \mid \mu\right] &= (0 - \mu)^2 p(0 \mid \mu) + (1 - \mu)^2 p(1 \mid \mu) \\
&= \mu^2 (1 - \mu) + (1 - \mu)^2 \mu \\
&= \mu(1 - \mu)(\mu + 1 - \mu) \\
&= \mu(1 - \mu)
\end{aligned}$$

Outline

Cumulative Distribution Function

- Every random variable has a **cumulative distribution function**, or CDF.
- The CDF of X is a function F_X (or just F if the r.v. is clear from context), defined for each real number x as

$$F_X(x) := P(X \leq x) \tag{2}$$

- Note that such a function will always be **nondecreasing**

Continuous Random Variables

Most R.V.s that describe characteristics with a real number are **continuous**.

Continuous Random Variable

- A **continuous random variable** is one with a **continuous CDF**

Continuous Random Variables

Most R.V.s that describe characteristics with a real number are **continuous**.

Continuous Random Variable

- A **continuous random variable** is one with a **continuous CDF**
- In other words, as we increase x, $F(x) := P(X \leq x)$ increases continuously (no jumps).

Example of a Continuous CDF

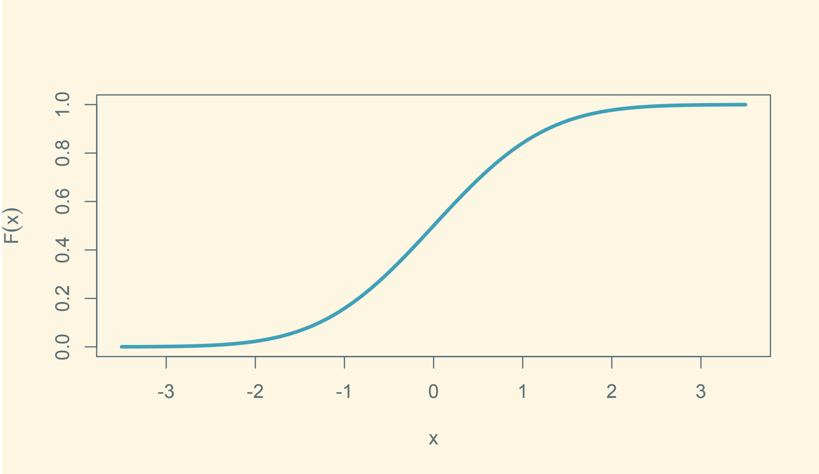

Figure: CDF of a Standard Normal distribution

Consequence of Continuity

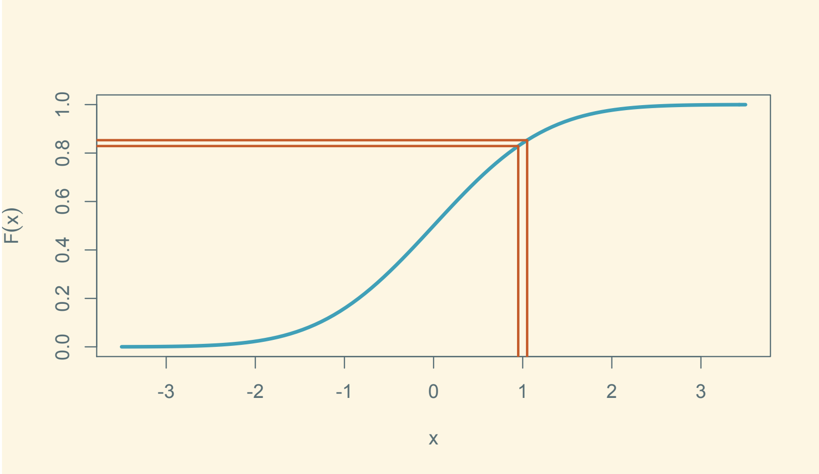

Figure: CDF of a Standard Normal distribution at 0.95 and 1.05

Nothing is Possible?

- If, for some x, $P(X = x) > 0$, then when the CDF "passed" this x, it would instantaneously "jump" up by $P(X = x)$.

Nothing is Possible?

- If, for some x, $P(X = x) > 0$, then when the CDF "passed" this x, it would instantaneously "jump" up by $P(X = x)$.
- Since continuous Random Variables have no jumps in their CDF, this implies that $P(X = x) = 0$ for **every** x

Nothing is Possible?

- If, for some x, $P(X = x) > 0$, then when the CDF "passed" this x, it would instantaneously "jump" up by $P(X = x)$.
- Since continuous Random Variables have no jumps in their CDF, this implies that $P(X = x) = 0$ for **every** x
- So... nothing can happen?

Nothing is Possible?

- If, for some x, $P(X = x) > 0$, then when the CDF "passed" this x, it would instantaneously "jump" up by $P(X = x)$.
- Since continuous Random Variables have no jumps in their CDF, this implies that $P(X = x) = 0$ for **every** x
- So... nothing can happen?
- **Zeno's Arrow Paradox:** Since an arrow's path is continuous, at any instant it travels no distance. And yet, the arrow travels.

Nothing is Possible?

- If, for some x, $P(X = x) > 0$, then when the CDF "passed" this x, it would instantaneously "jump" up by $P(X = x)$.
- Since continuous Random Variables have no jumps in their CDF, this implies that $P(X = x) = 0$ for **every** x
- So... nothing can happen?
- **Zeno's Arrow Paradox:** Since an arrow's path is continuous, at any instant it travels no distance. And yet, the arrow travels.
- Just as in motion we need a concept of **instantaneous velocity**, in probability we need a concept of **instantaneous rate of accumulation of probability**

Outline

Probability Density Function

Probability Density Function

Many continuous r.v.s can be characterized by a **probability density function**, or PDF, which is the **derivative of the CDF**; i.e., the **instantaneous rate of accumulation of probability**

The PDF has a similar role to the PMF's role for discrete variables, so we write $p_X(x)$ or just $p(x)$.

$$p(x) := \frac{d}{dx}F(x) \tag{3}$$

PDF as Rate of Accumulation

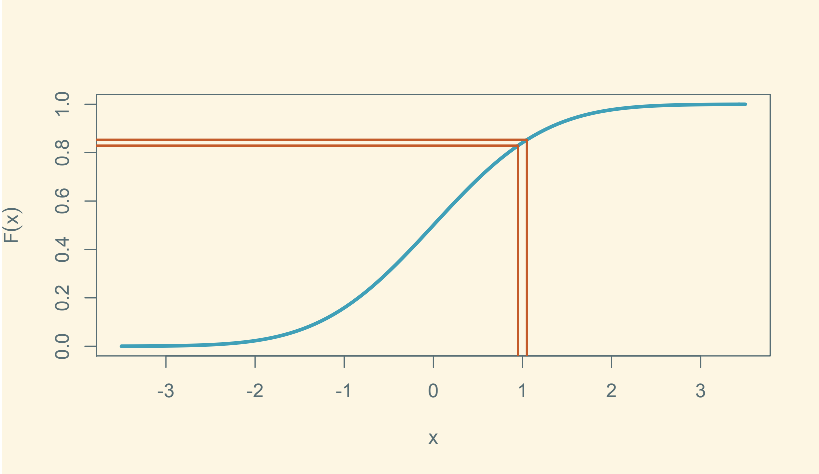

Figure: CDF of a Standard Normal distribution at 0.95 and 1.05

PDF as Rate of Accumulation

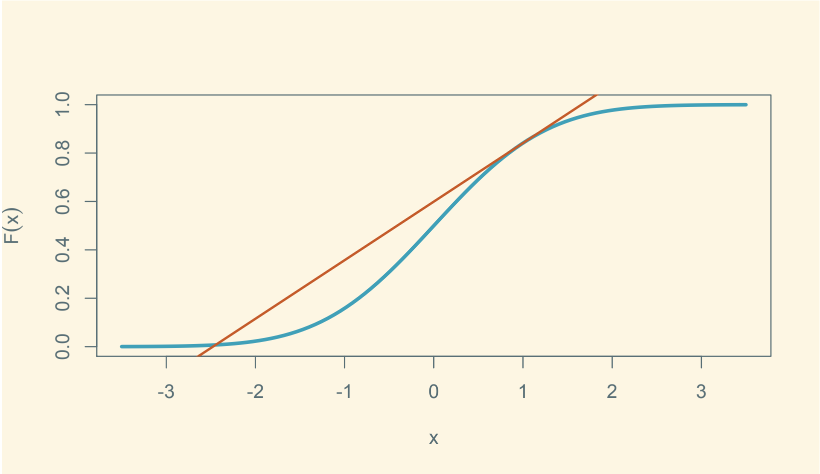

Figure: CDF of a Standard Normal distribution with tangent line at $x = 1$. The PDF at $x = 1$ is the slope of this line.

Outline

Continuous Uniform Distribution

A random variable with a **continuous uniform distribution** ranges over all reals in an interval, (a, b), with a **constant** density.

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Normal Distribution

A random variable with a **Normal Distribution** ranges over **all real numbers**. Given a **mean parameter** μ and a **variance parameter** σ^2 if its PDF is given by

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \tag{5}$$

Probability in an Interval

- By the Fundamental Theorem of Calculus, we can write

$$F(b) - F(a) = \int_a^b \left(\frac{d}{dx} F(x) \right) \, dx \tag{6}$$

$$= \int_a^b p(x) \, dx \tag{7}$$

Probability in an Interval

- By the Fundamental Theorem of Calculus, we can write

$$F(b) - F(a) = \int_a^b \left(\frac{d}{dx} F(x)\right) \, dx \tag{6}$$

$$= \int_a^b p(x) \, dx \tag{7}$$

- This gives the probability that X falls **between** a and b (i.e., how much probability does the CDF "accumulate" between a and b)

Unity of PDF

- In the limit, for **every** PDF:

$$\int_{-\infty}^{\infty} p(x)\ dx = P(-\infty < X < \infty) = 1$$

Unity of PDF

- In the limit, for **every** PDF:

$$\int_{-\infty}^{\infty} p(x)\ dx = P(-\infty < X < \infty) = 1$$

- **Note:** This means that **if we know a PDF is** $p(x) = k \cdot g(x)$**, then** k **is uniquely determined by** $g(x)$

Expected Value

- In the continuous case we simply replace the PMF by the PDF and the sum by an integral

Expected Value (Continuous Case)

For a continuous random variable, X, with PDF $p(x)$, the **expected value** of $f(X)$ (including $f(X) = X$) is

$$\mathbb{E}\left[f(X)\right] := \int_{x \in \text{Range}(X)} f(x)p(x)\ dx \tag{8}$$