CS70: Lecture 28.

Continuous Probability

1. Conditional Probability (Recap: revisit $G(p)$)
2. Continuous Probability: Examples
3. Continuous Probability: Events
4. Continuous Random Variables

Recap: Conditional distributions

$X \mid Y$ is a RV:

$$\sum_x p_{X|Y}(x \mid y) = \sum_x \frac{p_{XY}(x,y)}{p_Y(y)} = 1$$

Multiplication or Product Rule:

$$p_{XY}(x,y) = p_X(x)p_{Y|X}(y \mid x) = p_Y(y)p_{X|Y}(x \mid y)$$

Total Probability Theorem: If $A_1, A_2, \ldots, A_N$ partition Ω, and $P[A_i] > 0 \ \forall i$, then

$$p_X(x) = \sum_{i=1}^{N} P[A_i]P[X = x \mid A_i]$$

Nothing special about just two random variables, naturally extends to more.

Let's visit the mean and variance of the geometric distribution using conditional expectation.

Revisiting mean of geometric RV $X \sim G(p)$

X is **memoryless**

$$P[X = n+m \mid X > n] = P[X = m].$$

Thus $E[X \mid X > 1] = 1 + E[X]$.

Why? (Recall $E[g(X)] = \sum_l g(l)P[X = l]$)

$$\begin{aligned} E[X \mid X > 1] &= \sum_{k=1}^{\infty} kP[X = k \mid X > 1] \\ &= \sum_{k=2}^{\infty} kP[X = k-1] \qquad (\textit{memoryless}) \\ &= \sum_{l=1}^{\infty} (l+1)P[X = l] \qquad (l = k-1) \\ &= E[X+1] = 1 + E[X] \end{aligned}$$

Revisiting mean of geometric RV $X \sim G(p)$

X is **memoryless**

$$P[X = k+m \mid X > k] = P[X = m].$$

Thus $E[X \mid X > 1] = 1 + E[X]$.

We have $E[X] = P[X = 1]E[X \mid X = 1] + P[X > 1]E[X \mid X > 1]$.

$$\begin{aligned} &\Rightarrow E[X] = p.1 + (1-p)(E[X]+1) \\ &\Rightarrow E[X] = p + 1 - p + E[X] - pE[X] \\ &\Rightarrow pE[X] = 1 \\ &\Rightarrow E[X] = \frac{1}{p} \end{aligned}$$

Derive the variance for $X \sim G(p)$ by finding $E[X^2]$ using conditioning.

Summary of Conditional distribution

For Random Variables X and Y, $P[X = x \mid Y = k]$ is the **conditional distribution** of X given $Y = k$

$$P[X = x \mid Y = k] = \frac{P[X = x, Y = k]}{P[Y = k]}$$

Numerator: Joint distribution of (X, Y).

Denominator: Marginal distribution of Y.

(Aside: surprising result using conditioning of RVs):

Theorem: If $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$ are independent, then $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

"Sum of independent Poissons is Poisson."

Sum of Independent Poissons is Poisson

Intuition based on Binomial limiting behavior

- $X_1 \sim B(n, p_1)$ where $p_1 = \frac{\lambda_1}{n}$, n is large, λ_1 is constant
- $X_2 \sim B(n, p_2)$ where $p_2 = \frac{\lambda_2}{n}$, n is large, λ_2 is constant

Question: What is (a good approximation to) $Y = X_1 + X_2$? (X_1, X_2 independent)

$$X_1 : \ T \ T \ T \ T \ H \ T \ T \ T \ \cdots \ H \ \cdots$$

H appears with probability p_1

$$X_2 : \ T \ T \ H \ T \ T \ T \ T \ T \ \cdots \ H \ \cdots$$

H appears with probability p_2

$$Y : \ T \ T \ H \ T \ H \ T \ T \ T \ \cdots \ 2H \ \cdots$$

H appears with probability $p_1 + p_2$, $2H$ appears with $p_1 p_2$

Intuition: If $p_1 = \frac{\lambda_1}{n}$ and $p_2 = \frac{\lambda_2}{n}$, then $p_1 p_2 = \frac{\lambda_1 \lambda_2}{n^2}$
$\Rightarrow 2H$ will essentially NEVER appear!

Sum of Independent Poissons is Poisson

Let's define events:

- A: Every Y_i has H or T for $i = 1, 2, \cdots, n$
- D: At least one Y_i has $2H$ for $i = 1, 2, \cdots, n$

We have A and D partition Ω, so

$$P[Y = k] = P[Y = k \mid A]P[A] + P[Y = k \mid D]P[D]$$

$$\begin{aligned} P[D] &= P[\cup_{i=1}^{n}(Y_i \text{ is } 2H)] \\ &\le \sum_{i=1}^{n} P[Y_i \text{ is } 2H] \\ &\le \sum_{i=1}^{n} \frac{\lambda_1\lambda_2}{n^2} = \frac{\lambda_1\lambda_2}{n} \end{aligned}$$

Sum of Independent Poissons is Poisson

Let's define events:

- A: Every Y_i has H or T for $i = 1, 2, \cdots, n$
- D: At least one Y_i is $2H$ for $i = 1, 2, \cdots, n$

We have A and D partition Ω, so

$$P[Y = k] = P[Y = k \mid A]P[A] + P[Y = k \mid D]P[D]$$

$$P[D] \le \frac{\lambda_1\lambda_2}{n}$$

$P[D] \to 0$ as n grows

$P[A] = 1 - P[D] \to 1$ as n grows

$P[Y = k \mid A] \underset{D}{=} B(n, p_1 + p_2)$

$P[Y = k] \sim B(n, p_1 + p_2)$

Limit: "$Poisson(\lambda_1) + Poisson(\lambda_2) = Poisson(\lambda_1 + \lambda_2)$"

Continuous Probability: Why do we need it?

Many settings involve uncertainty in quantities like time, distance, velocity, temperature, etc. that are **continuous-valued**.
Need to extend our discrete-probability knowledge-base to cover this.

Here are some motivating examples:

Alice and Bob decide to meet at Yali's Cafe to study for CS 70. As they have uncertain schedules, they are independently and uniformly likely to show up randomly at any time in the designated hour. They decide that whoever shows up first will wait for at most 10 minutes before leaving.
What is the probability they meet?

You break a stick at two points chosen independently uniformly at random.

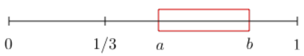
What is the probability you can make a triangle with the three pieces?

In digital video and audio, one represents a continuous value by a finite number of bits. This introduces an error perceived as noise: the quantization noise. What is the power of that noise?

Continuous Probability: Uniformly at Random in $[0, 1]$.

Choose a real number X, uniformly at random in $[0, 1]$.

What is the probability that X is exactly equal to $1/3$? Well, ..., 0.



What is the probability that X is exactly equal to 0.6? Again, 0.

In fact, for any $x \in [0, 1]$, one has $Pr[X = x] = 0$.

How should we then describe 'choosing uniformly at random in $[0, 1]$'?

Here is the way to do it:

$$Pr[X \in [a, b]] = b - a, \forall 0 \le a \le b \le 1.$$

Makes sense: $b - a$ is the fraction of $[0, 1]$ that $[a, b]$ covers.

Uniformly at Random in $[0, 1]$.

Let $[a, b]$ denote the **event** that the point X is in the interval $[a, b]$.

$$Pr[[a, b]] = \frac{\text{length of } [a, b]}{\text{length of } [0, 1]} = \frac{b - a}{1} = b - a.$$

Intervals like $[a, b] \subseteq \Omega = [0, 1]$ are **events.**
More generally, events in this space are unions of intervals.
Example: the event A - "within 0.2 of 0 or 1" is $A = [0, 0.2] \cup [0.8, 1]$.
Thus,

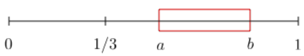
$$Pr[A] = Pr[[0, 0.2]] + Pr[[0.8, 1]] = 0.4.$$

More generally, if A_n are pairwise disjoint intervals in $[0, 1]$, then

$$Pr[\cup_n A_n] := \sum_n Pr[A_n].$$

Many subsets of $[0, 1]$ are of this form. Thus, the probability of those sets is well defined. We call such sets events.

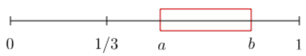
Uniformly at Random in $[0, 1]$.



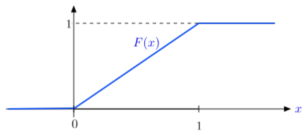
Note: A **radical** change in approach. For a finite probability space, $\Omega = \{1, 2, \ldots, N\}$, we started with $Pr[\omega] = p_\omega$. We then defined $Pr[A] = \sum_{\omega \in A} p_\omega$ for $A \subset \Omega$. We used the same approach for countable Ω.

For a continuous space, e.g., $\Omega = [0, 1]$, we cannot start with $Pr[\omega]$, because this will typically be 0. Instead, we start with $Pr[A]$ for some events A. Here, we started with A = interval, or union of intervals.

Uniformly at Random in $[0,1]$.



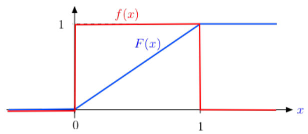
Note: $Pr[X \leq x] = x$ for $x \in [0,1]$. Also, $Pr[X \leq x] = 0$ for $x < 0$ and $Pr[X \leq x] = 1$ for $x > 1$. Let us define $F(x) = Pr[X \leq x]$.



Then we have $Pr[X \in (a,b]] = Pr[X \leq b] - Pr[X \leq a] = F(b) - F(a)$.

Thus, $F(\cdot)$ specifies the probability of all the events!

Uniformly at Random in $[0,1]$.



$$Pr[X \in (a,b]] = Pr[X \leq b] - Pr[X \leq a] = F(b) - F(a).$$

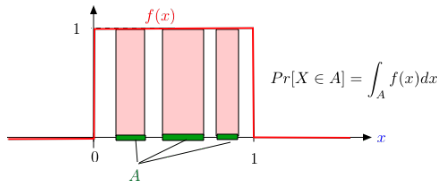
An alternative view is to define $f(x) = \frac{d}{dx}F(x) = 1\{x \in [0,1]\}$. Then

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

Thus, the probability of an event is the integral of $f(x)$ over the event:

$$Pr[X \in A] = \int_A f(x)dx.$$

Uniformly at Random in $[0,1]$.



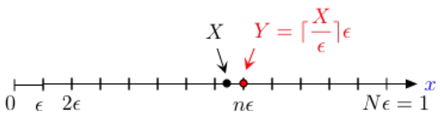
Think of $f(x)$ as describing how
one unit of probability is spread over $[0,1]$: uniformly!

Then $Pr[X \in A]$ is the probability mass over A.

Observe:

- This makes the probability automatically additive.
- We need $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Uniformly at Random in $[0,1]$.



Discrete Approximation: Fix $N \gg 1$ and let $\varepsilon = 1/N$.

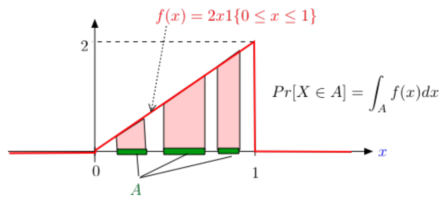
Define $Y = n\varepsilon$ if $(n-1)\varepsilon < X \leq n\varepsilon$ for $n = 1,\ldots,N$.

Then $|X - Y| \leq \varepsilon$ and Y is discrete: $Y \in \{\varepsilon, 2\varepsilon, \ldots, N\varepsilon\}$.

Also, $Pr[Y = n\varepsilon] = \frac{1}{N}$ for $n = 1,\ldots,N$.

Thus, X is 'almost discrete.'

Nonuniformly at Random in $[0,1]$.



This figure shows a different choice of $f(x) \geq 0$ with $\int_{-\infty}^{\infty} f(x)dx = 1$.

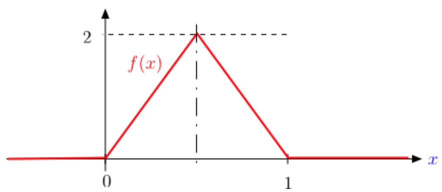
It defines another way of choosing X at random in $[0,1]$.

Note that X is more likely to be closer to 1 than to 0.

One has $Pr[X \leq x] = \int_{-\infty}^{x} f(u)du = x^2$ for $x \in [0,1]$.

Also, $Pr[X \in (x, x+\varepsilon)] = \int_{x}^{x+\varepsilon} f(u)du \approx f(x)\varepsilon$.

Another Nonuniform Choice at Random in $[0,1]$.



This figure shows yet a different choice of $f(x) \geq 0$ with $\int_{-\infty}^{\infty} f(x)dx = 1$.

It defines another way of choosing X at random in $[0,1]$.

Note that X is more likely to be closer to $1/2$ than to 0 or 1.

For instance, $Pr[X \in [0,1/3]] = \int_0^{1/3} 4xdx = 2\left[x^2\right]_0^{1/3} = \frac{2}{9}$.

Thus, $Pr[X \in [0,1/3]] = Pr[X \in [2/3,1]] = \frac{2}{9}$ and $Pr[X \in [1/3,2/3]] = \frac{5}{9}$.

General Random Choice in $\Re$

Let $F(x)$ be a nondecreasing function with $F(-\infty) = 0$ and $F(+\infty) = 1$.

Define X by $Pr[X \in (a,b]] = F(b) - F(a)$ for $a < b$. Also, for $a_1 < b_1 < a_2 < b_2 < \cdots < b_n$,

$$\begin{aligned} & Pr[X \in (a_1,b_1] \cup (a_2,b_2] \cup (a_n,b_n]] \\ & = Pr[X \in (a_1,b_1]] + \cdots + Pr[X \in (a_n,b_n]] \\ & = F(b_1) - F(a_1) + \cdots + F(b_n) - F(a_n). \end{aligned}$$

Let $f(x) = \frac{d}{dx}F(x)$. Then,

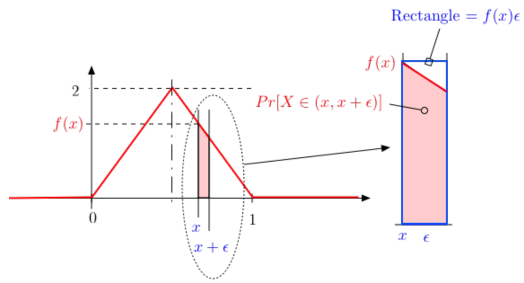
$$Pr[X \in (x, x+\varepsilon]] = F(x+\varepsilon) - F(x) \approx f(x)\varepsilon.$$

Here, $F(x)$ is called the cumulative distribution function (cdf) of X and $f(x)$ is the probability density function (pdf) of X.

To indicate that F and f correspond to the RV X, we will write them $F_X(x)$ and $f_X(x)$.

$Pr[X \in (x, x+\varepsilon)]$

An illustration of $Pr[X \in (x, x+\varepsilon)] \approx f_X(x)\varepsilon$:



Thus, the pdf is the 'local probability by unit length.'

It is the 'probability density.'

Discrete Approximation

Fix $\varepsilon \ll 1$ and let $Y = n\varepsilon$ if $X \in (n\varepsilon, (n+1)\varepsilon]$.

Thus, $Pr[Y = n\varepsilon] = F_X((n+1)\varepsilon) - F_X(n\varepsilon)$.

Note that $|X - Y| \le \varepsilon$ and Y is a discrete random variable.

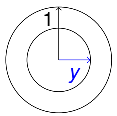
Also, if $f_X(x) = \frac{d}{dx}F_X(x)$, then $F_X(x+\varepsilon) - F_X(x) \approx f_X(x)\varepsilon$.

Hence, $Pr[Y = n\varepsilon] \approx f_X(n\varepsilon)\varepsilon$.

Thus, we can think of X of being almost discrete with $Pr[X = n\varepsilon] \approx f_X(n\varepsilon)\varepsilon$.

Example: CDF

Example: hitting random location on gas tank.
Random location on circle.



Random Variable: Y distance from center.
Probability within y of center:

$$\begin{aligned} Pr[Y \le y] &= \frac{\text{area of small circle}}{\text{area of dartboard}} \\ &= \frac{\pi y^2}{\pi} = y^2. \end{aligned}$$

Hence,

$$F_Y(y) = Pr[Y \le y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \le y \le 1 \\ 1 & \text{for } y > 1 \end{cases}$$

Calculation of event with dartboard..

Probability between .5 and .6 of center?
Recall CDF.

$$F_Y(y) = Pr[Y \le y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \le y \le 1 \\ 1 & \text{for } y > 1 \end{cases}$$

$$\begin{aligned} Pr[0.5 < Y \le 0.6] &= Pr[Y \le 0.6] - Pr[Y \le 0.5] \\ &= F_Y(0.6) - F_Y(0.5) \\ &= .36 - .25 \\ &= .11 \end{aligned}$$

PDF.

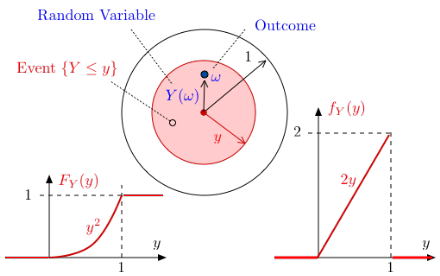
Example: "Dart" board.
Recall that

$$F_Y(y) = Pr[Y \le y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \le y \le 1 \\ 1 & \text{for } y > 1 \end{cases}$$

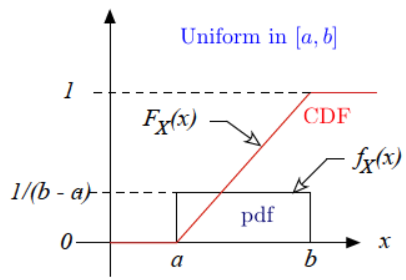
$$f_Y(y) = F_Y'(y) = \begin{cases} 0 & \text{for } y < 0 \\ 2y & \text{for } 0 \le y \le 1 \\ 0 & \text{for } y > 1 \end{cases}$$

The cumulative distribution function (cdf) and probability distribution function (pdf) give full information.
Use whichever is convenient.

Target



$U[a,b]$

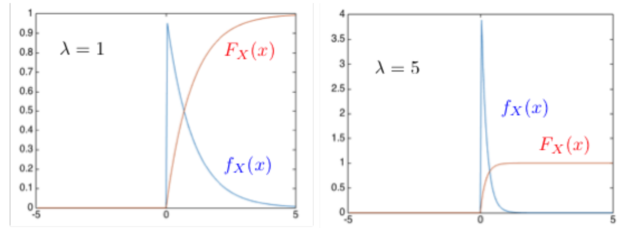


$Expo(\lambda)$

The exponential distribution with parameter $\lambda > 0$ is defined by

$$f_X(x) = \lambda e^{-\lambda x} 1\{x \ge 0\}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \ge 0. \end{cases}$$



Note that $Pr[X > t] = e^{-\lambda t}$ for $t > 0$.

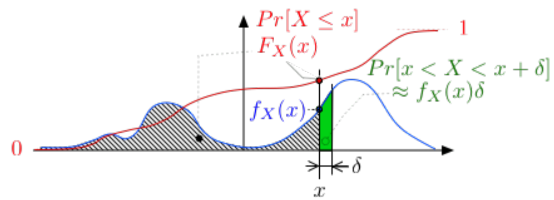
Random Variables

Continuous random variable X, specified by

1. $F_X(x) = Pr[X \le x]$ for all x.
   Cumulative Distribution Function (cdf).
   $Pr[a < X \le b] = F_X(b) - F_X(a)$
   1.1 $0 \le F_X(x) \le 1$ for all $x \in \Re$.
   1.2 $F_X(x) \le F_X(y)$ if $x \le y$.
2. Or $f_X(x)$, where $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ or $f_X(x) = \frac{d(F_X(x))}{dx}$.
   Probability Density Function (pdf).
   $Pr[a < X \le b] = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$
   2.1 $f_X(x) \ge 0$ for all $x \in \Re$.
   2.2 $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Recall that $Pr[X \in (x, x+\delta)] \approx f_X(x)\delta$. Think of X taking discrete values $n\delta$ for $n = \dots, -2, -1, 0, 1, 2, \dots$ with $Pr[X = n\delta] = f_X(n\delta)\delta$.

A Picture



The pdf $f_X(x)$ is a nonnegative function that integrates to 1.

The cdf $F_X(x)$ is the integral of f_X.

$$Pr[x < X < x + \delta] \approx f_X(x)\delta$$

$$Pr[X \le x] = F_x(x) = \int_{-\infty}^{x} f_X(u)du$$

Summary

Continuous Probability

1. pdf: $Pr[X \in (x, x+\delta]] = f_X(x)\delta$.
2. CDF: $Pr[X \le x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
3. $U[a,b]$: $f_X(x) = \frac{1}{b-a} 1\{a \le x \le b\}$; $F_X(x) = \frac{x-a}{b-a}$ for $a \le x \le b$.
4. $Expo(\lambda)$:
   $f_X(x) = \lambda \exp\{-\lambda x\} 1\{x \ge 0\}$; $F_X(x) = 1 - \exp\{-\lambda x\}$ for $x \le 0$.
5. Target: $f_X(x) = 2x1\{0 \le x \le 1\}$; $F_X(x) = x^2$ for $0 \le x \le 1$.