**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

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

\[ \Rightarrow E[X] = p \cdot 1 + (1-p) \cdot (E[X] + 1) \]

\[ \Rightarrow E[X] = p + 1 - p + E[X] - pE[X] \]

\[ \Rightarrow pE[X] = 1 \]

\[ \Rightarrow E[X] = \frac{1}{p} \]

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

---

**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 \( E[g(X)] = \sum g(i)P[X = i] \))

\[ E[X \mid X > 1] = \sum_{k=1}^{\infty} kP[X = k \mid X > 1] \]

\[ = \sum_{k=1}^{\infty} kP[X = k - 1] \quad \text{(memoryless)} \]

\[ = \sum_{i=1}^{\infty} (i + 1)P[X = i] \quad (i = k - 1) \]

\[ = E[X + 1] = 1 + E[X] \]

---

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

\( X \) is memoryless

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

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

**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 \( p_1 = \frac{1}{2} \), \( n \) is large, \( \lambda_1 \) is constant
- \( X_2 \sim B(n, p_2) \) where \( p_2 = \frac{1}{2} \), \( n \) is large, \( \lambda_2 \) is constant

**Question:** What is (a good approximation to) \( Y = X_1 + X_2 \)?

\( (X_1, X_2) \) independent

\[ X_1 : \ T \ T \ T \ T \ H \ T \ T \ \cdots \ H \ \cdots \]

\( H \) 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 \( p_1p_2 \)

\( 2H \) appears with probability \( p_1p_2 \)

**Intuition:** If \( p_1 = \frac{1}{2} \) and \( p_2 = \frac{1}{2} \), then \( p_1p_2 = \frac{1}{4} \)

\[ \Rightarrow 2H \text{ will essentially NEVER appear!} \]
We call such sets events.

We then defined

Instead, we start with

we cannot start with

length of

we started with

\[ \sum_{i=1}^{n} P[Y_i \text{ is } 2H] \]

\[ \sum_{i=1}^{n} \frac{\lambda_i \lambda_2}{n} \]

\[ P[D] \leq \frac{\lambda_1 \lambda_2}{n} \]

\[ P[Y = k | A] = P(Y = k | A)[P[A] + P[Y = k | D]P[D] \]


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 \(P[X = x] = 0\).

How should we then describe ‘choosing uniformly at random in \([0,1]\)?

Here is the way to do it:

\[ P[X \in [a,b]] = \frac{b-a}{1} = b - a. \]

Intervals like \([a,b] \subset \Omega = [0,1]\) 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,

\[ P[A] = P[0.0.2] + P[0.8,1] = 0.4. \]

More generally, if \(A_n\) are pairwise disjoint intervals in \([0,1]\), then

\[ P[\cup_n A_n] = \sum_n P[A_n]. \]

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

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 independently and uniformly decide that whoever shows up first will wait for at most 10 minutes before leaving.

What is the probability they meet?

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 \([0,1]\).

Let \([a,b]\) denote the event that the point \(X\) is in the interval \([a,b]\).

\[ P([a,b]) = \frac{b-a}{1} = b - a. \]

Intervals like \([a,b] \subset \Omega = [0,1]\) 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,

\[ P[A] = P[0.0.2] + P[0.8,1] = 0.4. \]

More generally, if \(A_n\) are pairwise disjoint intervals in \([0,1]\), then

\[ P[\cup_n A_n] = \sum_n P[A_n]. \]

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

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

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

![Uniformly at Random in [0, 1].](image)

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.’

![Uniformly at Random in [0, 1].](image)

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

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

This makes the probability automatically additive.

We need $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

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

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

![Uniformly at Random in [0, 1].](image)

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

![Nonuniformly at Random in [0, 1].](image)

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, $\Pr[X \in [0, 1/3]] = \int_0^{1/3} 4x\,dx = 2\left[\frac{x^2}{2}\right]_{1/3} = \frac{2}{9}$

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

![Another Nonuniform Choice at Random in [0, 1].](image)
**General Random Choice in \( \mathbb{R} \)**

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 \( a_1 < a_2 < a_3 < \cdots < a_n \),

\[
Pr[X \in (a_i, b_i)] = Pr[X \in (a_1, b_1)] + \cdots + Pr[X \in (a_n, b_n)] = F(b_1) - F(a_1) + \cdots + F(b_n) - F(a_n).
\]

Let \( f(x) = \frac{1}{\Delta} F(x) \). Then,

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

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 \( f_X(x) \).

**Example: CDF**

Example: hitting random location on gas tank.

Random location on circle.

Random Variable: \( Y \) distance from center.

Probability within \( y \) of center:

\[
Pr[Y \leq y] = \frac{\text{area of small circle}}{\text{area of dartboard}} = \frac{\pi y^2}{\pi} = y^2.
\]

Hence,

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

**Calculation of event with dartboard..**

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

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

\[
Pr[0.5 < Y \leq 0.6] = Pr[Y \leq 0.6] - Pr[Y \leq 0.5] = F_Y(0.6) - F_Y(0.5) = .36 - .25 = .11
\]

**PDF.**

Example: “Dart” board.
Recall that

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

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

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

**Discrete Approximation**

Fix \( \epsilon \ll 1 \) and let \( Y = n \epsilon \) if \( X \in (n \epsilon, (n+1) \epsilon) \).
Thus, \( Pr[Y = n \epsilon] = F_X(n \epsilon) - F_X((n+1) \epsilon) \).

Note that \( |X - Y| \leq \epsilon \) and \( Y \) is a discrete random variable.
Also, if \( f_X(x) = \frac{1}{\Delta} F_X(x) \), then \( F_X(x + \epsilon) - F_X(x) \approx f_X(x) \epsilon \).

Hence, \( Pr[Y = n \epsilon] \approx f_X(n \epsilon) \epsilon \).
Thus, we can think of \( X \) of being almost discrete with \( Pr[X = n \epsilon] \approx f_X(n \epsilon) \epsilon \).
**Random Variables**

Continuous random variable $X$, specified by

1. $F_X(x) = Pr[X \leq x]$ for all $x$.

Cumulative Distribution Function (cdf).

1. $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$.
2. $F_X(x) \leq F_X(y)$ if $x \leq y$.

Or $f_X(x)$, where $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$ or $f_X(x) = \frac{dF_X(x)}{dx}$.

Probability Density Function (pdf).

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

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

**Expo($\lambda$)**

The exponential distribution with parameter $\lambda > 0$ is defined by
\[ f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\} \]
\[ F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases} \]

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

**Summary**

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

**Target**

A discrete random variable $X$, specified by

1. $f_X(n\delta)\delta$ for $n = \ldots, -2, -1, 0, 1, 2, \ldots$ with $Pr[X = n\delta] = f_X(n\delta)\delta$.

- PDF $f_X(n\delta)\delta$.
- CDF $F_X(x)$.

**A Picture**

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

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

\[ Pr[x < X < x + \delta] \approx f_X(x)\delta \]
\[ Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(u)\,du \]