## Calculus Review

$\frac{d(e^{cx})}{dx} = ce^{cx}$.

$\frac{d(x^2)}{dx} = 2x$.

$\int x dx = \frac{x^2}{2} + c$.

$\frac{d(\ln x)}{dx} = \frac{1}{x}$

Chain Rule: $\frac{d(f(g(x))}{dx} = f'(g(x))g'(x)dx$

Product Rule:
$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$.
$d(uv) = udv + vdu$

Integration by Parts: $\int udv = uv - \int vdu$.

## Summary

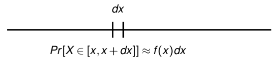
Continuous Probability 1

1. pdf: $Pr[X \in (x, x+\delta]] \approx f_X(x)\delta$.
2. CDF: $Pr[X \le x] = F_X(x) = lim_{\delta \to 0} \sum_i f_X(x_i)\delta = \int_{-\infty}^{x} f_X(y)dy$.
3. $X \sim U[a,b]$: $f_X(x) = \frac{1}{b-a} 1\{a \le x \le b\}; F_X(x) = \frac{x-a}{b-a}$ for $a \le x \le b$.
4. $X \sim Expo(\lambda)$:
   $f_X(x) = \lambda \exp\{-\lambda x\} 1\{x \ge 0\}; F_X(x) = 1 - \exp\{-\lambda x\}$ for $x \le 0$.
5. Target: $f_X(x) = 2x \cdot 1\{0 \le x \le 1\}; F_X(x) = x^2$ for $0 \le x \le 1$.
6. Joint pdf: $Pr[X \in (x, x+\delta), Y = (y, y+\delta)) = f_{X,Y}(x,y)\delta^2$.
   6.1 Conditional Distribution: $f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$.
   6.2 Independence: $f_{X|Y}(x,y) = f_X(x)$

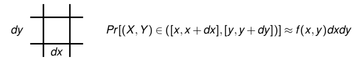
## Discrete/Continuous

Discrete: Probability of outcome $\rightarrow$ random variables, events.

Continuous: "outcome" is real number.
Probability: Events is interval.
Density: $Pr[X \in [x, x+dx]] = f(x)dx$




Joint Continuous in $d$ variables: "outcome" is $\in R^d$.
Probability: Events is block.
Density: $Pr[(X,Y) \in ([x, x+dx], [y, y+dx])] = f(x,y)dxdy$

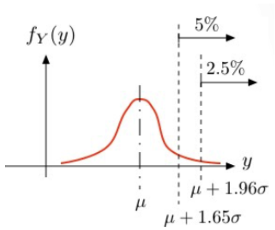



## Normal (Gaussian) Distribution.

For any $\mu$ and $\sigma$, a **normal** (aka **Gaussian**) random variable $Y$, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

**Standard normal has** $\mu = 0$ **and** $\sigma = 1$.




Note: $Pr[|Y - \mu| > 1.65\sigma] = 10\%; Pr[|Y - \mu| > 2\sigma] = 5\%$.

## Scaling and Shifting and properties

**Theorem** Let $X = \mathcal{N}(0,1)$ and $Y = \mu + \sigma X$. Then

$$Y = \mathcal{N}(\mu, \sigma^2).$$

**Theorem** If $Y = \mathcal{N}(\mu, \sigma^2)$, then

$$E[Y] = \mu \text{ and } var[Y] = \sigma^2.$$

## Review: Law of Large Numbers.

**Theorem:** Set of independent identically distributed random variables, $X_i$,
$A_n = \frac{1}{n} \sum X_i$ "tends to the mean."

Say $X_i$ have expectation $\mu = E(X_i)$ and variance $\sigma^2$.

Mean of $A_n$ is $\mu$, and variance is $\sigma^2/n$.

Used Chebyshev.

$$Pr[|A_n - \mu| > \varepsilon] \le \frac{var[A_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon} \to 0.$$

## Central Limit Theorem

**Central Limit Theorem**

Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = \mu$ and $var(X_1) = \sigma^2$. Define

$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$S_n \to \mathcal{N}(0,1), \text{as } n \to \infty.$$

That is,

$$Pr[S_n \le \alpha] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

**Proof:** See EE126.

**Note:**

$$E(S_n) = \frac{1}{\sigma/\sqrt{n}}(E(A_n) - \mu) = 0$$

$$Var(S_n) = \frac{1}{\sigma^2/n} Var(A_n) = 1.$$

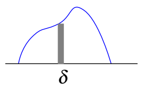
## Probability

Probability!
Challenges us. But really neat.
At times, continuous. At others, discrete.

Sample Space: $\Omega$, $Pr[\omega]$.
Event: $Pr[A] = \sum_{\omega \in A} Pr[\omega]$
$\sum_\omega Pr[\omega] = 1$.
Random variables: $X(\omega)$.
Distribution: $Pr[X = x]$
$\sum_x Pr[X = x] = 1$.

Random Variable: $X$
Event: $A = [a,b]$, $Pr[X \in A]$,
CDF: $F(x) = Pr[X \le x]$.
PDF: $f(x) = \frac{dF(x)}{dx}$.
$\int_{-\infty}^{\infty} f(x) = 1$.

Continuous as Discrete.
$Pr[X \in [x, x+\delta]] \approx f(x)\delta$




## Probability Rules are all good.

Conditional Probabity.

Events: $A, B$

Discrete: "Heads", "Tails", $X = 1$, $Y = 5$.

Continuous: $X$ in $[.2, .3]$. $X \in [.2, .3]$ or $X \in [.4, .6]$.

Conditional Probability: $Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$

$Pr$["Second Heads"|"First Heads"],
$Pr[X \in [.2, .3] | X \in [.2, .3]$ or $X \in [.5, .6]]$.

Total Probability Rule: $Pr[A] = Pr[A \cap B] + Pr[A \cap \overline{B}]$
$Pr$["Second Heads"] $= Pr[HH] + Pr[HT]$
$B$ is First coin heads.
$Pr[X \in [.45, .55]] = Pr[X \in [.45, .50]] + Pr[X \in (.50, .55]]$
$B$ is $X \in [0, .5]$

Product Rule: $Pr[A \cap B] = Pr[A|B] Pr[B]$.
Bayes Rule: $Pr[A|B] = Pr[B|A] Pr[A] / Pr[B]$.

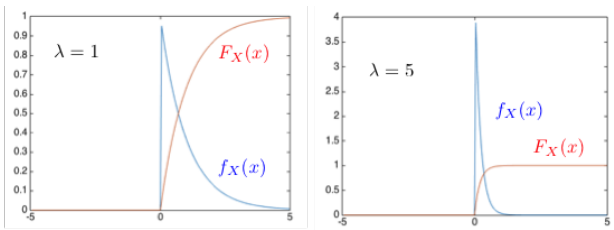
All work for continuous with intervals as events.

## $Expo(\lambda)$

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 $t > 0$.

## Some Properties

1. ***Expo* is memoryless.** Let $X = Expo(\lambda)$. Then, for $s, t > 0$,

$$\begin{aligned} Pr[X > t + s \mid X > s] &= \frac{Pr[X > t + s]}{Pr[X > s]} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= Pr[X > t]. \end{aligned}$$

'Used is a good as new.'

2. **Scaling *Expo*.** Let $X = Expo(\lambda)$ and $Y = aX$ for some $a > 0$. Then

$$\begin{aligned} Pr[Y > t] &= Pr[aX > t] = Pr[X > t/a] \\ &= e^{-\lambda(t/a)} = e^{-(\lambda/a)t} = Pr[Z > t] \text{ for } Z = Expo(\lambda/a). \end{aligned}$$

Thus, $a \times Expo(\lambda) = Expo(\lambda/a)$.

Also, $Expo(\lambda) = \frac{1}{\lambda} Expo(1)$.

## More Properties

3. **Scaling Uniform.** Let $X = U[0,1]$ and $Y = a + bX$ where $b > 0$. Then,

$$\begin{aligned} Pr[Y \in (y, y+\delta)] &= Pr[a + bX \in (y, y+\delta)] = Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b})] \\ &= Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b})] = \frac{1}{b}\delta, \text{ for } 0 < \frac{y-a}{b} < 1 \\ &= \frac{1}{b}\delta, \text{ for } a < y < a + b. \end{aligned}$$

Thus, $f_Y(y) = \frac{1}{b}$ for $a < y < a + b$. Hence, $Y = U[a, a+b]$.

Replace $b$ by $b - a$, use $X = U[0,1]$, then $Y = a + (b-a)X$ is $U[a,b]$.

## Some More Properties

**4. Scaling pdf.** Let $f_X(x)$ be the pdf of $X$ and $Y = a + bX$ where $b > 0$. Then

$$\begin{aligned} Pr[Y \in (y, y+\delta)] &= Pr[a + bX \in (y, y+\delta)] = Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b})] \\ &= Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b})] = f_X(\frac{y-a}{b})\frac{\delta}{b}. \end{aligned}$$

Now, the left-hand side is $f_Y(y)\delta$. Hence,

$$f_Y(y) = \frac{1}{b} f_X(\frac{y-a}{b}).$$

## Expectation

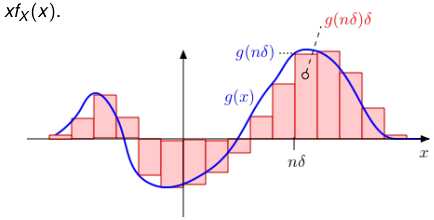
**Definition:** The **expectation** of a random variable $X$ with pdf $f(x)$ is *defined* as

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

**Justification:** Say $X = n\delta$ w.p. $f_X(n\delta)\delta$ for $n \in \mathbb{Z}$. Then,

$$E[X] = \sum_n (n\delta) Pr[X = n\delta] = \sum_n (n\delta) f_X(n\delta)\delta = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Indeed, for any $g$, one has $\int g(x) dx \approx \sum_n g(n\delta)\delta$. Choose $g(x) = x f_X(x)$.




## Examples of Expectation

1. $X = U[0,1]$. Then, $f_X(x) = 1\{0 \leq x \leq 1\}$. Thus,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x.1 dx = [\frac{x^2}{2}]_0^1 = \frac{1}{2}.$$

2. $X$ = distance to 0 of dart shot uniformly in unit circle. Then $f_X(x) = 2x1\{0 \leq x \leq 1\}$. Thus,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 2x dx = [\frac{2x^3}{3}]_0^1 = \frac{2}{3}.$$

## Examples of Expectation

3. $X = Expo(\lambda)$. Then, $f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\}$. Thus,

$$E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx = -\int_0^{\infty} x de^{-\lambda x}.$$

Recall the integration by parts formula:

$$\begin{aligned} \int_a^b u(x) dv(x) &= [u(x)v(x)]_a^b - \int_a^b v(x) du(x) \\ &= u(b)v(b) - u(a)v(a) - \int_a^b v(x) du(x). \end{aligned}$$

Thus,

$$\begin{aligned} \int_0^{\infty} x de^{-\lambda x} &= [xe^{-\lambda x}]_0^{\infty} - \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 - 0 + \frac{1}{\lambda} \int_0^{\infty} de^{-\lambda x} = -\frac{1}{\lambda}. \end{aligned}$$

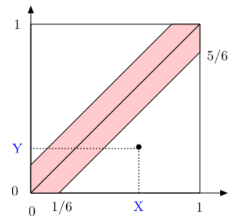
Hence, $E[X] = \frac{1}{\lambda}$.

## Meeting at a Restaurant

Two friends go to a restaurant independently uniformly at random between noon and 1pm.

They agree they will wait for 10 minutes.

What is the probability they meet?




Here, $(X, Y)$ are the times when the friends reach the restaurant.

The shaded area are the pairs where $|X - Y| < 1/6$, i.e., such that they meet.

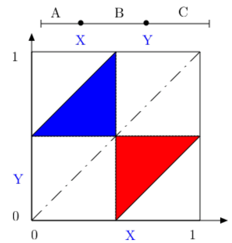
The complement is the sum of two rectangles. When you put them together, they form a square with sides $5/6$.

Thus, $Pr[\text{meet}] = 1 - (\frac{5}{6})^2 = \frac{11}{36}$.

## Breaking a Stick

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?




Let $X, Y$ be the two break points along the $[0,1]$ stick.

A triangle if $A < B + C, B < A + C$, and $C < A + B$.

If $X < Y$, this means $X < 0.5$, $Y < X + .5$, $Y > 0.5$. This is the blue triangle.

If $X > Y$, get red triangle, by symmetry.

Thus, $Pr[\text{make triangle}] = 1/4$.

## Maximum of Two Exponentials

Let $X = Expo(\lambda)$ and $Y = Expo(\mu)$ be independent.
Define $Z = \max\{X, Y\}$.

Calculate $E[Z]$.

We compute $f_Z$, then integrate.

One has

$$\begin{aligned} Pr[Z < z] &= Pr[X < z, Y < z] = Pr[X < z]Pr[Y < z] \\ &= (1 - e^{-\lambda z})(1 - e^{-\mu z}) = 1 - e^{-\lambda z} - e^{-\mu z} + e^{-(\lambda+\mu)z} \end{aligned}$$

Thus,

$$f_Z(z) = \lambda e^{-\lambda z} + \mu e^{-\mu z} - (\lambda+\mu)e^{-(\lambda+\mu)z}, \forall z > 0.$$

Since, $\int_0^\infty x\lambda e^{-\lambda x}dx = \lambda[-\frac{xe^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2}]_0^\infty = \frac{1}{\lambda}$.

$$E[Z] = \int_0^\infty z f_Z(z)dz = \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu}.$$

## Maximum of $n$ i.i.d. Exponentials

Let $X_1, \ldots, X_n$ be i.i.d. $Expo(1)$. Define $Z = \max\{X_1, X_2, \ldots, X_n\}$.

Calculate $E[Z]$.

We use a recursion. The key idea is as follows:

$$Z = \min\{X_1, \ldots, X_n\} + \max Y_1, \ldots, Y_{n-1}.$$

From memoryless property of the exponential.

Let then $A_n = E[Z]$. We see that

$$\begin{aligned} A_n &= E[\min\{X_1, \ldots, X_n\}] + A_{n-1} \\ &= \frac{1}{n} + A_{n-1} \end{aligned}$$

because the minimum of *Expo* is *Expo* with the sum of the rates.

Hence,

$$E[Z] = A_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} = H(n).$$

## Quantization Noise

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?

**Model:** $X = U[0,1]$ is the continuous value. $Y$ is the closest multiple of $2^{-n}$ to $X$. Thus, we can represent $Y$ with $n$ bits. The error is $Z := X - Y$.

The power of the noise is $E[Z^2]$.

**Analysis:** We see that $Z$ is uniform in $[0, a = 2^{-(n+1)}]$.

Thus,

$$E[Z^2] = \frac{a^2}{3} = \frac{1}{3}2^{-2(n+1)}.$$

The power of the signal $X$ is $E[X^2] = \frac{1}{3}$.

## Quantization Noise

We saw that $E[Z^2] = \frac{1}{3}2^{-2(n+1)}$ and $E[X^2] = \frac{1}{3}$.

The **signal to noise ratio** (SNR) is the power of the signal divided by the power of the noise.

Thus,

$$SNR = 2^{2(n+1)}.$$

Expressed in decibels, one has

$$SNR(dB) = 10\log_{10}(SNR) = 20(n+1)\log_{10}(2) \approx 6(n+1).$$

For instance, if $n = 16$, then $SNR(dB) \approx 112dB$.

## Expected Squared Distance

**Problem 1:** Pick two points $X$ and $Y$ independently and uniformly at random in $[0,1]$.

What is $E[(X-Y)^2]$?

**Analysis:** One has

$$\begin{aligned} E[(X-Y)^2] &= E[X^2 + Y^2 - 2XY] \\ &= \frac{1}{3} + \frac{1}{3} - 2\frac{1}{2}\frac{1}{2} \\ &= \frac{2}{3} - \frac{1}{2} = \frac{1}{6}. \end{aligned}$$

**Problem 2:** What about in a unit square?

**Analysis:** One has

$$\begin{aligned} E[||\mathbf{X} - \mathbf{Y}||^2] &= E[(X_1 - Y_1)^2] + E[(X_2 - Y_2)^2] \\ &= 2 \times \frac{1}{6}. \end{aligned}$$

**Problem 3:** What about in $n$ dimensions? $\frac{n}{6}$.

## Geometric and Exponential

The geometric and exponential distributions are similar. They are both memoryless.

Consider flipping a coin every $1/N$ second with $Pr[H] = p/N$, where $N \gg 1$.

Let $X$ be the time until the first $H$.

**Fact:** $X \approx Expo(p)$.

**Analysis:** Note that

$$\begin{aligned} Pr[X > t] &\approx Pr[\text{first } Nt \text{ flips are tails}] \\ &= (1 - \frac{p}{N})^{Nt} \approx \exp\{-pt\}. \end{aligned}$$

Indeed, $(1 - \frac{a}{N})^N \approx \exp\{-a\}$.

## Summary

$\boxed{\text{Continuous Probability}}$

- Continuous RVs are essentially the same as discrete RVs
- Think that $X \approx x$ with probability $f_X(x)\varepsilon$
- Sums become integrals, ....
- The exponential distribution is magical: memoryless.