

Calculus Review

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

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

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

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

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

$$\text{Product Rule: } (f(x)g(x))' = f'(x)g(x) + f(x)g'(x).$$

$$d(uv) = u dv + v du$$

$$\text{Integration by Parts: } \int u dv = uv - \int v du.$$

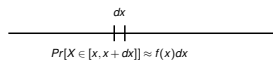
Discrete/Continuous

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

Continuous: "outcome" is real number.

Probability: Events is interval.

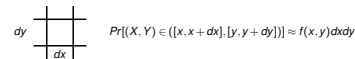
Density: $Pr[X \in [x, x+dx]] = f(x)dx$



Joint Continuous in d variables: "outcome" is $\in R^d$.

Probability: Events is block.

Density: $Pr[(X, Y) \in ([x, x+dx], [y, y+dy])] = f(x, y)dxdy$



Summary

Continuous Probability 1

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

Poll

What is true?

X has CDF $F(x)$ and PDF $f(x)$.

- (A) $Pr[X > t] = 1 - Pr[X \leq t]$.
 (B) $S(t) = Pr[X > t] = 1 - F(t)$.
 (C) $Y = 2X$, $f_Y(y) = 2f(y)$.
 (D) $Y = 2X$, $F_Y(y) = F(y/2)$.
 (E) $Y = 2X$, $f_Y(y) = \frac{1}{2}f(y/2)$.

(A), (B), (D) think events, (E) think event and density.

(C) confuses probability density of outcome with value of outcome.

Probability

Probability!

Challenges us. But really neat.

At times, continuous. At others, discrete.

Sample Space: Ω , $Pr[\omega]$.

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: $A = [a, b]$, $Pr[X \in A]$,

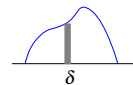
CDF: $F(x) = Pr[X \leq x]$.

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 Probability.

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

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

Total Probability Rule: $Pr[A] = Pr[A \cap B] + Pr[A \cap \bar{B}]$

$Pr[\text{"Second Heads"}] = Pr[HH] + Pr[HT]$

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.

Conditional density.

Conditional Density: $f_{X|Y}(x, y)$.

Conditional Probability: $Pr[X \in A | Y \in B] = \frac{Pr[X \in A, Y \in B]}{Pr[Y \in B]}$

$$Pr[X \in [x, x+dx] | Y \in [y, y+dy]] = \frac{f_{X,Y}(x,y)dx dy}{f_Y(y)dy}$$

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy}$$

Corollary: For independent random variables, $f_{X|Y}(x, y) = f_X(x)$.

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, $Y = U[a, a+b]$.

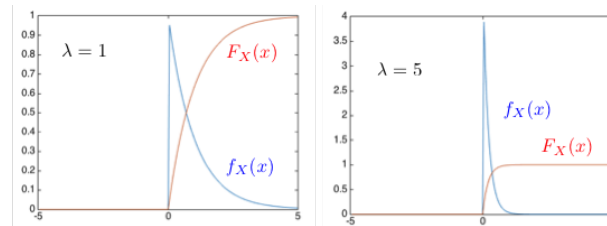
Replace b by $b-a$, use $X = U[0, 1]$, then $Y = a + (b-a)X$ is $U[a, b]$.

Expo(λ)

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 $t > 0$.

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,

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

Some Properties

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

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

'Used is as good as new.'

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

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

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

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

Expectation

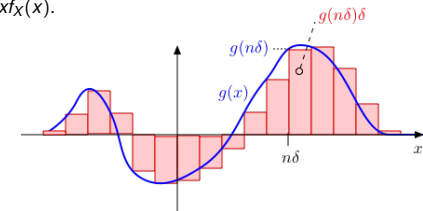
Definition: The **expectation** of a random variable X with pdf $f(x)$ is defined as

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

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,

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

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

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

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

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

Examples of Expectation

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

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

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,

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

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

Review: Law of Large Numbers.

Theorem: Set of independent identically distributed random variables, X_i ,

$$A_n = \frac{1}{n} \sum X_i \text{ "tends to the mean."}$$

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

Mean of A_n is μ , and variance is σ^2/n .

Used Chebyshev.

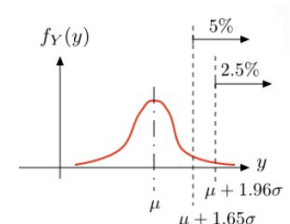
$$\Pr[|A_n - \mu| > \varepsilon] \leq \frac{\text{var}[A_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon} \rightarrow 0.$$

Normal (Gaussian) Distribution.

For any μ and σ , 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: $\Pr[|Y - \mu| > 1.65\sigma] = 10\%$; $\Pr[|Y - \mu| > 2\sigma] = 5\%$.

Central Limit Theorem

Central Limit Theorem

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

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

Then,

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

That is,

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

Proof: See EE126.

Note:

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

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

Confidence Intervals.

Recall: $A_n = \frac{1}{n} \sum X_i$, for X_i identical and independent.

For $\mu = E(X_i)$ and variance σ^2 . Mean of A_n is μ , and variance is σ^2/n .

Recall Chebyshev: $Pr[|A_n - \mu| > \varepsilon] \leq \frac{\text{var}[A_n]}{\varepsilon^2}$

Implies to get confidence $1 - \delta$ we need

$$\text{var} A_n \varepsilon^2 = \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \leq \delta \text{ or } n \geq \frac{\sigma^2}{\varepsilon^2} \frac{1}{\delta}$$

Central Limit Theorem:

$$Pr[|A_n - \mu| > \varepsilon] \leq C \int_{x \geq \varepsilon} e^{-\frac{x^2}{2\text{var}A}} \leq C e^{-\frac{\varepsilon^2}{2\text{var}A}}$$

for $\varepsilon > \sqrt{\text{var}A}$ (C is roughly $2/\sqrt{2\pi}$)

Implies to get confidence $1 - C\delta$ we need

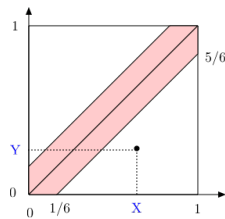
$$e^{-\frac{\varepsilon^2}{2\text{var}A}} \leq \delta \implies -\frac{n\varepsilon^2}{2\sigma^2} \leq \log \delta \implies n \geq \frac{2\sigma^2}{\varepsilon^2} \log \frac{1}{\delta}.$$

Examples: 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?



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

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

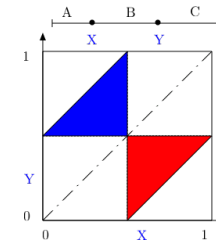
The shaded area are the pairs where $|X - Y| < 1/6$, i.e., such that they meet.

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

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 $X = \text{Expo}(\lambda)$ and $Y = \text{Expo}(\mu)$ 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, $\int_0^\infty x \lambda e^{-\lambda x} dx = \lambda \left[-\frac{x e^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2} \right]_0^\infty = \frac{1}{\lambda}$.

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

Minimum of n i.i.d. Exponentials.

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

What is true?

(A) Z is exponential.

(B) Parameter is n .

(C) $\lim_{N \rightarrow \infty} (1 - n/N)^N \rightarrow e^{-n}$

(D) $E[Z] = 1/n$.

(C) is an argument for (A), (B) and (D).

Maximum of n i.i.d. Exponentials

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

Calculate $E[Z]$.

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

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

From memoryless property of the exponential.

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

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

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

Hence,

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

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

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

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

Summary

Continuous Probability

- ▶ Continuous RVs are similar to discrete RVs
- ▶ Think that $X \in [x, x + \varepsilon]$ with probability $f_X(x)\varepsilon$
- ▶ Sums become integrals,
- ▶ The exponential distribution is magical: memoryless.