Continuous Random Variables

Prof. Edmund Lam

Department of Electrical and Electronic Engineering
The University of Hong Kong

ELEC2844: Probabilistic Systems Analysis
(Second Semester, 2023–24)

https://www.eee.hku.hk/~elec2844

Random variables

Continuous random variable X:

- has a nonnegative function

$$\boxed{f_X(x) \geqslant 0} \tag{1}$$

  for all x, called probability density function (PDF)
- Integrating ("continuous summing") this PDF gives us the probability of an event:

$$\boxed{\mathbb{P}(X \in B) = \int_B f_X(x)\,dx} \tag{2}$$

  for every subset B of the entire space

Random variables

Continuous random variable X:

- When B includes "everything" (Ω), then

$$\boxed{\mathbb{P}(X \in \Omega) = \int_\Omega f_X(x)\,dx = 1} \tag{3}$$

- For B to refer to a range $[a, b]$,

$$\boxed{\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(x)\,dx} \tag{4}$$

- Integral is zero when the range contains only a single number, so

$$\mathbb{P}(a \leqslant X \leqslant B) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leqslant X < b) = \mathbb{P}(a < X \leqslant b)$$

Random variables

Continuous random variable X:

- Normalization:

$$\boxed{\mathbb{P}(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(x)\,dx = 1} \tag{5}$$

- For an interval $[x, x + \delta]$:

$$\mathbb{P}([x, x+\delta]) = \int_x^{x+\delta} f_X(t)\,dt \approx f_X(x) \cdot \delta$$

  So $f_X(x)$ is the "probability mass per unit length" near x
- Since $f_X(x)$ is not the probability of any event, it does not have to be $\leqslant 1$.

Example

Is $f_X(x)$ a PDF?

$$f_X(x) = \begin{cases} \frac{1}{2\sqrt{x}} & 0 < x \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

ANS: Even though $f_X(x)$ becomes infinitely large as $x \to 0$, we still have $f_X(x) \geqslant 0$ for all x, and

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 \frac{1}{2\sqrt{x}}dx = \sqrt{x}\Big|_0^1 = 1$$

So this is a PDF.

Example

Let $f_X(x) = \begin{cases} x^2 & |x| < a \\ 0 & \text{otherwise.} \end{cases}$ be a PDF. Find a.

ANS: We integrate over the entire space, and set it to one:

$$\int_{-a}^{a} x^2 dx = \frac{x^3}{3}\Big|_{-a}^{a} = \frac{2a^3}{3} = 1$$

Therefore, $a = \sqrt[3]{\frac{3}{2}}$.

1 Expectation, mean, variance
2 Common continuous random variables
3 Cumulative density function
4 Derived distributions
5 Joint PDF, conditioning, independence
6 Continuous Bayes' rule

Expectation

The expectation or expected value or mean of a continuous r.v. X is

$$\boxed{\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx} \tag{6}$$

Similar to the discrete case, it is a very useful number as a "representative" value of X, and is also viewed as the "center of gravity."

Note: There can be problematic cases with integrals: even if $f_X(x)$ is symmetric about 0, the expected value is in fact not defined, e.g. $f_X(x) = 1/\pi(1+x^2)$

Expectation and variance

The formulas for continuous r.v. are similar to those for discrete:

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx \tag{7}$$

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \tag{8}$$

$$\text{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f_X(x)\, dx \tag{9}$$

$$\text{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 \geqslant 0 \tag{10}$$

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b \tag{11}$$

$$\text{var}(Y) = a^2 \text{var}(X) \tag{12}$$

where $Y = aX + b$

1. Expectation, mean, variance
2. Common continuous random variables
3. Cumulative density function
4. Derived distributions
5. Joint PDF, conditioning, independence
6. Continuous Bayes' rule

Continuous random variable

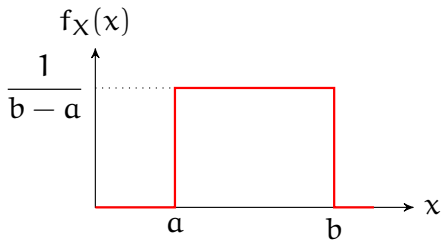
It is useful to know the mean and variance of some common distributions. A summary:

	mean: $\mathbb{E}(X)$	variance: $\text{var}(X)$
uniform: $X \sim U(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
exponential: $X \sim \text{Exp}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
normal: $X \sim \mathcal{N}(\mu, \sigma^2)$	μ	σ^2

(1) Uniform random variable

Continuous uniform random variable over a range $[a, b]$ has PDF

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$



(1) Uniform random variable

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \cdot \tfrac{1}{b-a}\,dx = \tfrac{1}{b-a} \cdot \tfrac{1}{2}x^2\Big|_a^b = \tfrac{a+b}{2}$$

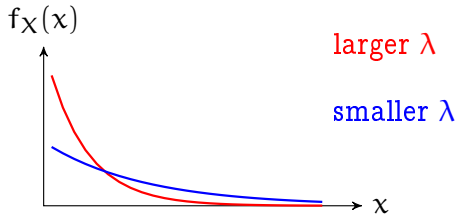
$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_a^b x^2 \cdot \tfrac{1}{b-a}\,dx = \tfrac{1}{b-a} \cdot \tfrac{1}{3}x^3\Big|_a^b = \tfrac{a^2+ab+b^2}{3}$$

$$\mathrm{var}(X) = \tfrac{a^2+ab+b^2}{3} - \left(\tfrac{a+b}{2}\right)^2 = \tfrac{(b-a)^2}{12}$$

(2) Exponential random variable

Exponential random variable with parameter $\lambda > 0$ has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$



(2) Exponential random variable

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbb{E}(X) = \int_0^{\infty} x\lambda e^{-\lambda x}\,dx = (-xe^{-\lambda x})\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx = 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty} = \frac{1}{\lambda}$$

$$\mathbb{E}(X^2) = \int_0^{\infty} x^2\lambda e^{-\lambda x}\,dx = (-x^2e^{-\lambda x})\Big|_0^{\infty} + \int_0^{\infty} 2xe^{-\lambda x}\,dx = \frac{2}{\lambda}\mathbb{E}(X) = \frac{2}{\lambda^2}$$

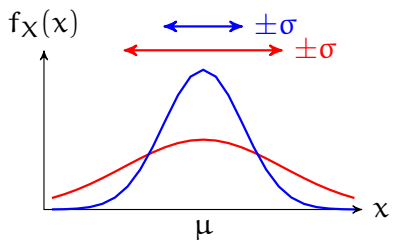
$$\mathrm{var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

(3) Normal random variable

Gaussian or normal r.v. with parameters $(\mu, \sigma > 0)$ has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \tag{15}$$

This is probably the most common distribution of all (with good scientific reason)!



(3) Normal random variable

Show that the normal distribution given by Eq. (15) has unit area.

Proof: Let $I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$. The trick is to go two-dimensional by computing I^2.

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy$$

Then, we convert to polar coordinates:

$$\begin{aligned} I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r dr d\theta &= 2\pi \int_0^{\infty} e^{-r^2/2} r dr \\ &= 2\pi \int_0^{\infty} e^{-u} du = 2\pi(-e^{-u})\Big|_0^{\infty} = 2\pi \\ I &= \sqrt{2\pi} \end{aligned}$$

(3) Normal random variable

Thus, we have shown that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds = 1$$

Then, we make the change of variable $s = \dfrac{x-\mu}{\sigma}$:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx = 1$$

since $ds = dx/\sigma$.

(3) Normal random variable

Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ has mean μ and variance σ^2.

Proof: For the mean, we will appeal to the fact that it is symmetric around μ and that the integral is defined. For variance, making use of a change of variable $y = (x-\mu)/\sigma$,

$$\begin{aligned} \text{var}(X) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\left[\left(-ye^{-y^2/2}\right)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy\right] \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\left(0 + \sqrt{2\pi}\right) = \sigma^2 \end{aligned}$$

(3) Normal random variable

If $X \sim \mathcal{N}(\mu, \sigma^2)$, and a linear transformation $Y = aX + b$ where $a \neq 0$ and b are scalars,

$$\boxed{\begin{cases} X \sim \mathcal{N}(\mu, \sigma^2) \\ Y = aX + b \end{cases} \implies Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)} \tag{16}$$

Corollary:

If $Y = \dfrac{X-\mu}{\sigma} = \dfrac{1}{\sigma}X - \dfrac{\mu}{\sigma}$, then $Y \sim \mathcal{N}(0, 1)$, called standard normal.

To see why, we'll look at cumulative density function and its use in deriving transforms of random variables first.

Example

Compute $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$ for the following cases:

- $X = A\cos(\omega t + \theta)$, where $A \sim \mathcal{N}(\mu, \sigma^2)$
- $X = a\cos(\omega t + \Theta)$, where $\Theta \sim U(0, 2\pi)$
- $X = a\cos(\omega T + \theta)$, where $T \sim U(-\frac{\pi}{\omega}, \frac{\pi}{\omega})$

ANS: For $X = A\cos(\omega t + \theta)$:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(A\cos(\omega t + \theta)) = \mu\cos(\omega t + \theta)\\ \mathrm{var}(X) &= \mathrm{var}(A\cos(\omega t + \theta)) = \sigma^2\cos^2(\omega t + \theta)\end{aligned}$$

Because $\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$, we have

$$\mathbb{E}(X^2) = (\mu^2 + \sigma^2)\cos^2(\omega t + \theta)$$

Example

For $X = a\cos(\omega t + \Theta)$:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(a\cos(\omega t + \Theta))\\ &= a\int_0^{2\pi}\cos(\omega t + \theta)\cdot\frac{1}{2\pi}d\theta\\ &= 0\end{aligned}$$

This is because we are integrating a full period of a cosine function, so the net area is zero.

Example

(cont.) For $X = a\cos(\omega t + \Theta)$:

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}\left(a^2\cos^2(\omega t + \Theta)\right)\\ &= a^2\int_0^{2\pi}\cos^2(\omega t + \theta)\cdot\frac{1}{2\pi}d\theta\\ &= a^2\cdot\pi\cdot\frac{1}{2\pi}\\ &= \frac{a^2}{2}\end{aligned}$$

To compute the integral, you may want to use the relationship $\cos^2\theta = \dfrac{\cos 2\theta + 1}{2}$.

Example

For $X = a\cos(\omega T + \theta)$:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(a\cos(\omega T + \theta))\\ &= a\int_{-\pi/\omega}^{\pi/\omega}\cos(\omega t + \theta)\cdot\frac{\omega}{2\pi}dt = 0\\ \mathbb{E}(X^2) &= \mathbb{E}\left(a^2\cos^2(\omega T + \theta)\right)\\ &= a^2\int_{-\pi/\omega}^{\pi/\omega}\cos^2(\omega t + \theta)\cdot\frac{\omega}{2\pi}dt\\ &= a^2\cdot\frac{1}{2}\left(\frac{\pi}{\omega} - \frac{-\pi}{\omega}\right)\cdot\frac{\omega}{2\pi} = \frac{a^2}{2}\end{aligned}$$

Again, we are integrating a full period of cosine (or cosine squared) functions.

1. Expectation, mean, variance
2. Common continuous random variables
3. Cumulative density function
4. Derived distributions
5. Joint PDF, conditioning, independence
6. Continuous Bayes' rule

Definition

The cumulative density function (CDF) of a random variable X is denoted $F_X(x)$. This can be defined for both discrete and continuous r.v.:

$$F_X(x) = \mathbb{P}(X \leqslant x) = \begin{cases} \sum_{k \leqslant x} p_X(k) & X \text{ is discrete} \\ \int_{-\infty}^{x} f_X(t)\,dt & X \text{ is continuous} \end{cases} \tag{17}$$

It is more common to use it for the continuous case.

Interpretation: CDF is the accumulated probability up to x.

Definition

Properties of CDF:

- CDF is nondecreasing.
- The maximum of the CDF is when $x = \infty$: $F_X(\infty) = 1$.
- The minimum of the CDF is when $x = -\infty$: $F_X(-\infty) = 0$.

Definition

We can also obtain PMF or PDF from CDF:

$$\begin{cases} p_X(k) = F_X(k) - F_X(k-1) & \text{discrete} \\ f_X(x) = \dfrac{d}{dx} F_X(x) & \text{continuous} \end{cases} \tag{18}$$

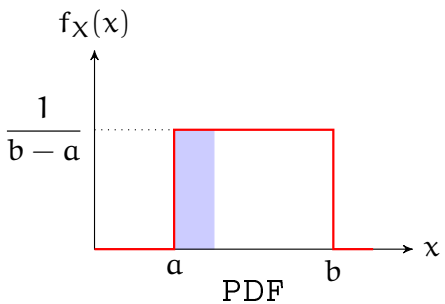
This is particularly useful when we define new r.v. through CDF, and then differentiate (for continuous r.v.) to obtain PDF.

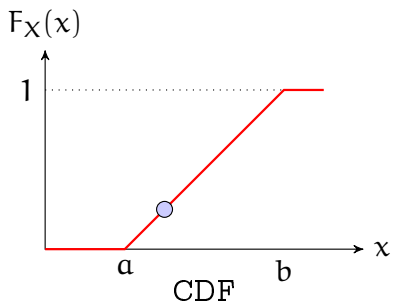
Example

For a uniform distribution $X \sim U(a, b)$, what is the CDF?

ANS: The CDF is given by

$$F_X(x) = \begin{cases} 0 & x \leqslant a \\ \int_{-\infty}^{x} f_X(x')\,dx' = \int_a^x \frac{1}{b-a}\,dx' = \frac{x-a}{b-a} & a < x \leqslant b \\ 1 & b < x \end{cases}$$





Example

Assume you take a quiz and the score is (discrete) uniform between 1 and 10. You are allowed to take the test three times and count only their maximum. What is the final PMF?

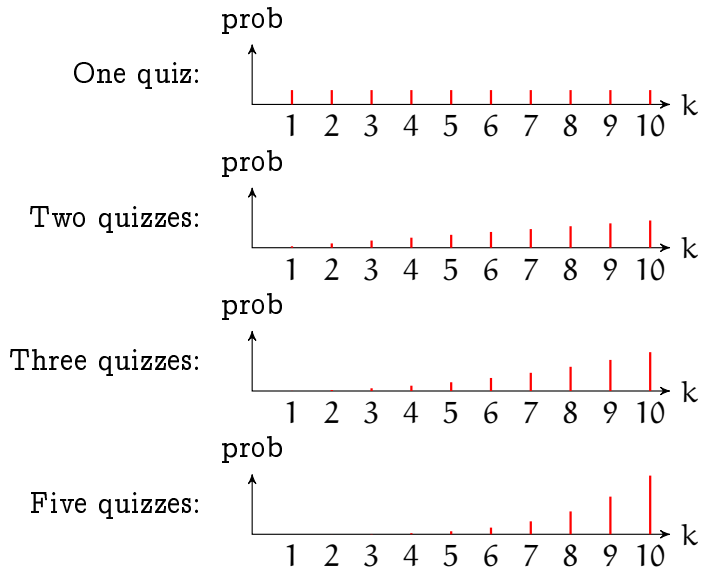
ANS: Let the quiz scores be X_1, X_2, X_3, so that $X = \max\{X_1, X_2, X_3\}$.

$$\begin{aligned} F_X(k) &= \mathbb{P}(X \leqslant k) = \mathbb{P}(X_1 \leqslant k, X_2 \leqslant k, X_3 \leqslant k) \\ &= \mathbb{P}(X_1 \leqslant k)\,\mathbb{P}(X_2 \leqslant k)\,\mathbb{P}(X_3 \leqslant k) = \left(\frac{k}{10}\right)^3 \end{aligned}$$

since the quiz score for each test is independent of each other.

$$\begin{aligned} p_X(k) &= F_X(k) - F_X(k-1) \\ &= \left(\frac{k}{10}\right)^3 - \left(\frac{k-1}{10}\right)^3 \qquad k = 1, \ldots, 10 \end{aligned}$$

Example



Geometric CDF

Remember $X \sim \text{Geo}(p)$ has PMF

$$p_X(k) = (1-p)^{k-1}p \qquad k = 1, 2, \ldots$$

Its CDF is

$$\boxed{F_X(n) = \sum_{k=1}^{n} (1-p)^{k-1}p = \frac{1-(1-p)^n}{1-(1-p)}p = 1-(1-p)^n} \tag{19}$$

for $n = 1, 2, \ldots$

Exponential CDF

$X \sim \text{Exp}(\lambda)$ has PDF

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geqslant 0$$

Its CDF is

$$\boxed{F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}} \tag{20}$$

for $x > 0$.

Geometric and exponential CDF

Geometric and exponential CDFs are closely related. By defining $\delta = -\ln(1-p)/\lambda$, we have

$$e^{-\lambda\delta} = 1 - p$$

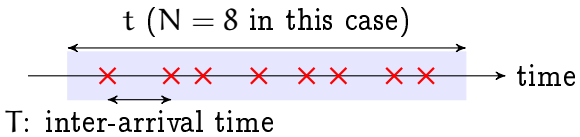
and whenever $x = n\delta$,

$$\boxed{F_{exp}(n\delta) = 1 - e^{-\lambda n\delta} = 1 - (1-p)^n = F_{geo}(n)} \tag{21}$$

Interpretation: Suppose we toss very quickly (every δ seconds, where $\delta \ll 1$) a biased coin with a very small probability of heads (equal to $p = 1 - e^{-\lambda\delta}$). The first time to obtain head, which is a geometric r.v. with parameter p, is a close approximation to an exponential r.v. with parameter λ.

Connecting exponential and Poisson random variables

Exponential random variable often has to do with *time*. It is also intimately related to the Poisson process. They are the foundations of a subject area known as queuing theory.



We start with Poisson, where people arrive at a rate of λ per unit time. Thus, within a time t, the number of arrival is $N \sim \text{Poisson}(\lambda t)$, i.e.,

$$\mathbb{P}(N = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

We are interested to find the *inter-arrival time* between consecutive people. Let this be T, which is a random variable.

Connecting exponential and Poisson random variables

We aim to find the CDF of T. We first note that

$$\begin{aligned}\mathbb{P}(T > t) &= \mathbb{P}(\text{inter-arrival time} > t)\\ &= \mathbb{P}(\text{no arrival within } t)\\ &= \mathbb{P}(N = 0 \text{ within } t)\end{aligned}$$

The last probability is Poisson, with $\mathbb{P}(N = 0) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$.

The CDF $F_T(t)$ computes the total probability that T is smaller than t, so

$$F_T(t) = 1 - \mathbb{P}(T > t) = 1 - e^{-\lambda t}$$

Therefore, $T \sim \text{Exp}(t)$.

Normal CDF

For $X \sim \mathcal{N}(0,1)$, the CDF is

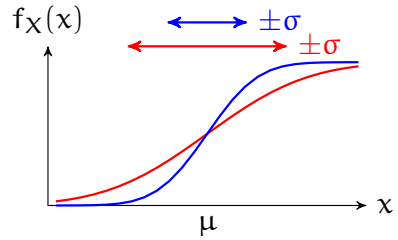
$$\boxed{F_X(x) = \Phi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt} \tag{22}$$

We cannot analytically integrate the function, but we can compute it as a function $\Phi(x)$ beforehand.

Then, for $X \sim \mathcal{N}(\mu, \sigma^2)$, we can "standardize" by defining $Y = \frac{X-\mu}{\sigma}$ such that $\mathbb{E}(Y) = 0$ and $\text{var}(Y) = 1$, and

$$\boxed{\mathbb{P}(X \leqslant x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leqslant \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(Y \leqslant \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)} \tag{23}$$

(3) Normal CDF



Example

A binary signal $\{+1, -1\}$ is transmitted through a channel corrupted with additive Gaussian noise $N \sim \mathcal{N}(0, \sigma^2)$. The receiver decides the signal is $+1$ if the received signal is $\geqslant 0$, and -1 if the received signal is < 0. What is the probability of error?

ANS: When the signal is -1, error occurs if $N \geqslant 1$.

$$\mathbb{P}(N \geqslant 1) = 1 - \mathbb{P}(N < 1) = 1 - \Phi\left(\frac{1}{\sigma}\right)$$

By symmetry, we have the same probability of error when signal is $+1$, and error occurs if $N < -1$.

As example, when $\sigma = 1$, $\Phi(1) \approx 0.84$, and probability of error is ≈ 0.16.

Median, Mode, Mean

The CDF also gives us a direct way to find some useful statistics on the distribution.

Median: The value c that separates the "higher half" and "lower half" of the distribution:

$$\boxed{\int_{-\infty}^{c} f_X(x)\, dx = \int_{c}^{\infty} f_X(x)\, dx} \tag{24}$$

Using CDF, that means:

$$\boxed{F_X(c) = \frac{1}{2}} \tag{25}$$

Example

What is the median of (i) $X \sim U(a, b)$; (ii) $X \sim \text{Exp}(\lambda)$?

ANS: (i) For a uniform distribution, $F_X(x) = \frac{x-a}{b-a}$ for $a < x \leqslant b$. Therefore,

$$F_X(c) = \frac{c-a}{b-a} = \frac{1}{2} \qquad \Rightarrow \qquad c = \frac{a+b}{2}$$

(ii) For an exponential distribution, $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$. Therefore,

$$F_X(c) = 1 - e^{-\lambda c} = \frac{1}{2} \qquad \Rightarrow \qquad c = \frac{1}{\lambda} \log 2$$

Median, Mode, Mean

Mode: The value c associated with the peak of the PDF:

$$\boxed{c = \arg\max_x f_X(x)} \tag{26}$$

(This notation means we want to find the maximum of $f_X(x)$, but we are mostly interested in the argument x that leads to this maximum, and assign c to this value.) Using CDF, that means:

$$\boxed{c = \arg\max_x \frac{d}{dx} F_X(x)} \tag{27}$$

Example

Find the mode of $F_X(x) = \begin{cases} 0 & x \leqslant 0 \\ 3x^2 - 2x^3 & 0 < x \leqslant 1 \\ 1 & x > 1 \end{cases}$.

ANS: For $0 < x \leqslant 1$, we have $f_X(x) = \frac{d}{dx} F_X(x) = 6x - 6x^2$. To find its maximum, we take the derivative and set to zero,

$$\frac{d}{dx}(6x - 6x^2) = 6 - 12x = 0 \qquad \Rightarrow \qquad x = \frac{1}{2}$$

To confirm that it is a maximum, we can take the second derivative and note that it is negative:

$$\frac{d^2}{dx^2}(6x - 6x^2) = -12$$

So, $c = \frac{1}{2}$ is the mode.

Median, Mode, Mean

Mean: We can also find the mean directly from the CDF:

$$\boxed{\mathbb{E}(X) = \int_0^\infty \big(1 - F_X(t)\big)\,dt - \int_{-\infty}^0 F_X(t)\,dt} \tag{28}$$

The proof is a little bit more involved and we do it in three steps. The discussion (and pictures) follow Section 4.4 of the reference text.

Median, Mode, Mean

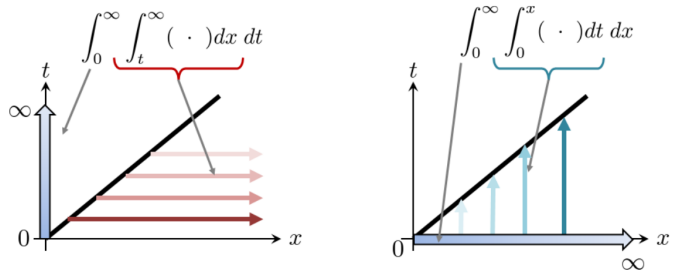
(I) Let $X \geqslant 0$. We want to show $\mathbb{E}(X) = \int_0^\infty \big(1 - F_X(t)\big)\,dt$.

We start with the right hand side:

$$\int_0^\infty \big(1 - F_X(t)\big)\,dt = \int_0^\infty \big(1 - \mathbb{P}(X \leqslant t)\big)\,dt = \int_0^\infty \big(\mathbb{P}(X > t)\big)\,dt$$
$$= \int_0^\infty \left(\int_t^\infty f_X(x)\,dx\right) dt$$

The trick is to change the integration order.

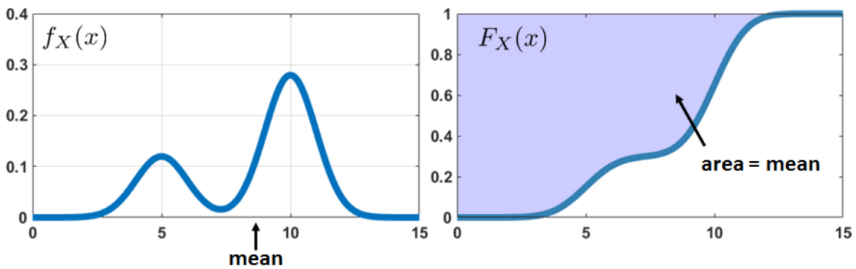
Median, Mode, Mean



$$\int_0^\infty \int_t^\infty f_X(x)\,dx dt = \int_0^\infty \int_0^x f_X(x)\,dt dx$$
$$= \int_0^\infty f_X(x) \left(\int_0^x dt\right) dx$$
$$= \int_0^\infty x f_X(x)\,dx$$
$$= \mathbb{E}(X)$$

Median, Mode, Mean

Interpretation of $\mathbb{E}(X) = \int_0^\infty \big(1 - F_X(t)\big)\,dt$:



Median, Mode, Mean

(II) Let $X < 0$. We want to show $\mathbb{E}(X) = -\int_{-\infty}^0 F_X(t)\,dt$.

We again start with the right hand side, and change the integration order:

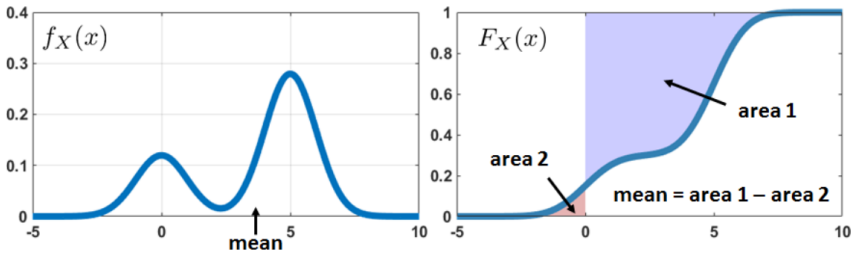
$$\int_{-\infty}^0 F_X(t)\,dt = \int_{-\infty}^0 \int_{-\infty}^t f_X(x)\,dx dt$$
$$= \int_{-\infty}^0 \int_x^0 f_X(x)\,dt dx$$
$$= \int_{-\infty}^0 -x f_X(x)\,dx$$
$$= -\mathbb{E}(X)$$

Median, Mode, Mean

(III) For a general X, we let $X = X^+ + X^-$, where X^+ is the positive part and X^- is the negative part. Then,

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X^+) + \mathbb{E}(X^-) \\ &= \int_0^\infty (1 - F_X(t))\,dt - \int_{-\infty}^0 F_X(t)\,dt\end{aligned}$$

Interpretation:



1. Expectation, mean, variance
2. Common continuous random variables
3. Cumulative density function
4. Derived distributions
5. Joint PDF, conditioning, independence
6. Continuous Bayes' rule

Procedure

Given: PDF of X, and $Y = g(X)$, find the PDF of Y.

A two-step procedure:

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(g(X) \leqslant y) = \int_{\{x \mid g(x) \leqslant y\}} f_X(x)\,dx \tag{29}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) \tag{30}$$

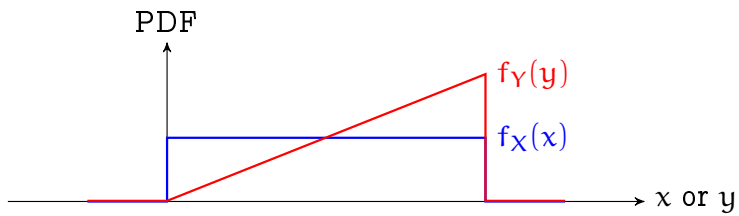
Example

$X \sim U(0,1)$ and $Y = \sqrt{X}$. Find PDF of Y.

ANS: Follow the procedure to find $F_Y(y)$ and $f_Y(y)$.

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}\left(\sqrt{X} \leqslant y\right) = \mathbb{P}(X \leqslant y^2) = y^2$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(y^2) = 2y \qquad 0 \leqslant y \leqslant 1$$



Example

$X \sim U(5, 10)$ and $Y = \dfrac{1}{X}$. Find PDF of Y.

ANS: Procedure is similar, but we need $F_X(x)$ first.

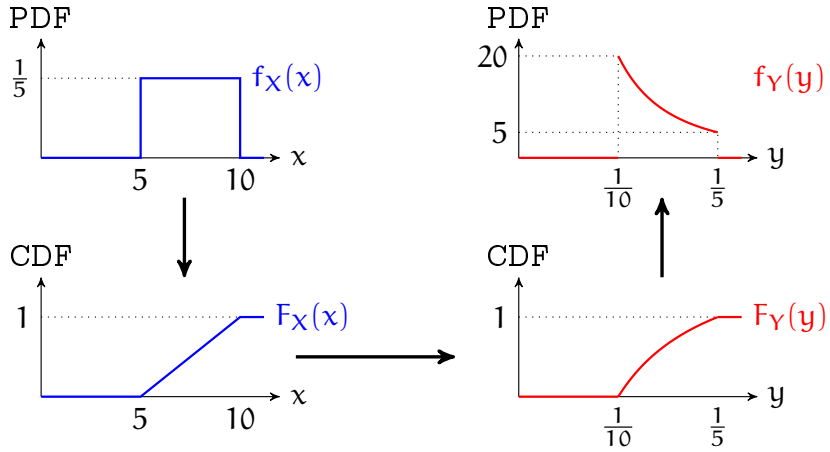
$$F_X(x) = \begin{cases} 0 & x \leqslant 5 \\ \frac{x-5}{5} & 5 < x \leqslant 10 \\ 1 & 10 < x \end{cases}$$

Then, $F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}\left(\dfrac{1}{X} \leqslant y\right) = \mathbb{P}\left(\dfrac{1}{y} \leqslant X\right) = 1 - F_X\left(\dfrac{1}{y}\right)$.

$$F_Y(y) = \begin{cases} 0 & y \leqslant \frac{1}{10} \\ 1 - (\frac{1}{y} - 5)/5 = 2 - \frac{1}{5y} & \frac{1}{10} < y \leqslant \frac{1}{5} \\ 1 & \frac{1}{5} < y \end{cases}$$

Example

Differentiating, $f_Y(y) = \begin{cases} 0 & y \leqslant \frac{1}{10} \quad \text{or} \quad \frac{1}{5} < y \\ \frac{1}{5y^2} & \frac{1}{10} < y \leqslant \frac{1}{5} \end{cases}$



Example

X has known PDF, and $Y = X^2$. Find PDF of Y.

ANS: For $y \geqslant 0$,

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(X^2 \leqslant y) &= \mathbb{P}(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Differentiating,

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \qquad y \geqslant 0$$

Special case: Y is a linear function of X

Given: PDF of X, and $Y = aX + b$, find the PDF of Y.

Case 1: $a > 0$

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(aX + b \leqslant y) = \mathbb{P}\left(X \leqslant \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Case 2: $a < 0$

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(aX + b \leqslant y) = \mathbb{P}\left(X \geqslant \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Special case: Y is a linear function of X

Given: PDF of X, and $Y = aX + b$,

$$\boxed{f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)} \tag{31}$$

Example: $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|}\frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{y-b}{a}-\mu)^2/2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}} \end{aligned}$$

i.e., $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

1. Expectation, mean, variance
2. Common continuous random variables
3. Cumulative density function
4. Derived distributions
5. Joint PDF, conditioning, independence
6. Continuous Bayes' rule

Definition

Consider two continuous r.v. X and Y, with joint PDF

$$\boxed{\mathbb{P}((X,Y) \in B) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy} \tag{32}$$

for every subset B of the 2D plane.
When $B = \{(x,y) \mid a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$,

$$\boxed{\mathbb{P}(a \leqslant x \leqslant b, c \leqslant y \leqslant d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy} \tag{33}$$

Definition

We can calculate the marginal PDFs of X and Y using

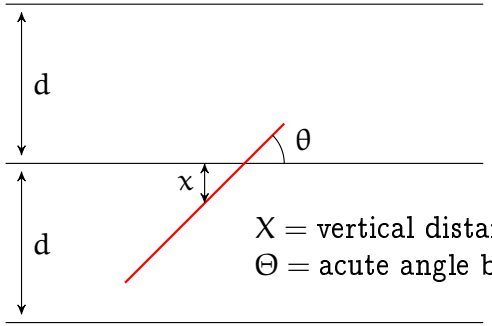
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \tag{34}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \tag{35}$$

and, for normalization, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) = 1$

Example

Buffon's Needle. We have many parallel lines at distances d apart. Assume we throw a needle of length ℓ, where $\ell < d$, on the surface at random. What is the probability that the needle will intersect one of the lines?



$X =$ vertical distance from mid needle to a line
$\Theta =$ acute angle between needle and line

$\vdots$

Example

At random: *assume a uniform joint PDF over the rectangular set*

$$S = \left\{ (x, \theta) \,\middle|\, 0 \leqslant x \leqslant \frac{d}{2}, 0 \leqslant \theta \leqslant \frac{\pi}{2} \right\}$$

This rectangle has an area $\pi d/4$, and X and Θ are independent, so

$$f_{X,\Theta}(x, \theta) = \begin{cases} \frac{4}{\pi d} & (x, \theta) \in S \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $\frac{\ell}{2} \sin \Theta$ is the *vertical* distance between the middle of the needle to the end points of the needle; the needle will intersect one of the lines if and only if

$$X \leqslant \frac{\ell}{2} \sin \Theta$$

Example

$$\begin{aligned} \mathbb{P}\left(X \leqslant \frac{\ell}{2} \sin \Theta\right) &= \int_0^{\pi/2} \int_0^{(\ell/2) \sin \theta} \frac{4}{\pi d} dx d\theta \\ &= \frac{4}{\pi d} \int_0^{\pi/2} \frac{\ell}{2} \sin \theta d\theta \\ &= \frac{2\ell}{\pi d} (-\cos \theta) \Big|_0^{\pi/2} \\ &= \frac{2\ell}{\pi d} \end{aligned}$$

This can be used as a Monte Carlo method to find π: if we drop n needles and h of them cross lines, then

$$\pi \approx \frac{2\ell}{d} \cdot \frac{n}{h}$$

Conditioning on an event

We can have conditional PDF of a continuous r.v. X given event A

$$\boxed{\mathbb{P}(X \in B \,|\, A) = \int_B f_{X|A}(x)\, dx} \tag{36}$$

with $\int_{-\infty}^{\infty} f_{X|A}(x)\, dx = 1$. Specifically, if the event is of the form $\{X \in A\}$ with $\mathbb{P}(X \in A) > 0$,

$$\mathbb{P}(X \in B \,|\, X \in A) = \frac{\mathbb{P}(X \in B, X \in A)}{\mathbb{P}(X \in A)} = \frac{\int_{A \cap B} f_X(x)\, dx}{\mathbb{P}(X \in A)}$$

$$\boxed{f_{X|\{X \in A\}}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(X \in A)} & x \in A \\ 0 & \text{otherwise.} \end{cases}} \tag{37}$$

Interpretation: Integrate all cases of $X = x$ AND belonging to A, then normalize with $\mathbb{P}(X \in A)$.

Example

A certain device has a lifetime following an exponential r.v. with parameter λ. Suppose now we know it has remained in operation for time t; and let X be the *additional* time until it fails. What is the conditional PDF of X, given it has already lasted for time t?

ANS: Define event $A = \{T > t\}$. Let us look at $\mathbb{P}(X > x \mid A)$.

$$\begin{aligned}\mathbb{P}(X > x \mid A) = \mathbb{P}(T > t + x \mid T > t) &= \frac{\mathbb{P}(T > t + x \text{ and } T > t)}{\mathbb{P}(T > t)} \\ &= \frac{\mathbb{P}(T > t + x)}{\mathbb{P}(T > t)} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}\end{aligned}$$

So $\mathbb{P}(X < x \mid A) = 1 - e^{-\lambda x}$, which means $X \sim \text{Exp}(\lambda)$.

Interpretation: Exponential r.v. is memoryless.

Conditioning on a r.v.

The conditional PDF of a r.v. X given knowledge of another r.v. Y

$$\boxed{f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}} \tag{38}$$

similarly, $\int_{-\infty}^{\infty} f_{X \mid Y}(x \mid y)\, dx = 1$.

Example

A dart lands uniformly random on a circular target of radius r. Let x and y be the Cartesian coordinates. Find the marginal PDF $f_Y(y)$ and conditional PDF $f_{X \mid Y}(x \mid y)$.

ANS: The joint PDF is $f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi r^2} & x^2 + y^2 \leqslant r^2 \\ 0 & \text{otherwise.} \end{cases}$

The marginal PDF is

$$\begin{aligned} f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx &= \frac{1}{\pi r^2} \int_{x^2+y^2 \leqslant r^2} dx \\ &= \frac{1}{\pi r^2} \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx \\ &= \frac{2}{\pi r^2} \sqrt{r^2 - y^2} \qquad \text{if } |y| \leqslant r \end{aligned}$$

This is NOT uniform!

Example

The conditional PDF is

$$\begin{aligned} f_{X \mid Y}(x \mid y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \\ &= \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2}\sqrt{r^2 - y^2}} \\ &= \frac{1}{2\sqrt{r^2 - y^2}} \qquad \text{if } x^2 + y^2 \leqslant r^2 \end{aligned}$$

Given a fixed y, this is uniform!

Conditional expectation

Let X and Y be two continuous r.v., and A an event. We can compute the conditional expectations:

$$\mathbb{E}(X \mid A) = \int_{-\infty}^{\infty} x f_{X \mid A}(x)\, dx \tag{39}$$

$$\mathbb{E}(g(X) \mid A) = \int_{-\infty}^{\infty} g(x) f_{X \mid A}(x)\, dx \tag{40}$$

$$\mathbb{E}(X \mid Y) = \int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y)\, dx \tag{41}$$

$$\mathbb{E}(g(X) \mid Y) = \int_{-\infty}^{\infty} g(x) f_{X \mid Y}(x \mid y)\, dx \tag{42}$$

Conditional expectation

Assume $A_1, \ldots, A_n$ are disjoint events partitioning the sample space, with $\mathbb{P}(A_i) > 0$ for all i.

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i)\, \mathbb{E}(X \mid A_i) \tag{43}$$

unconditional average = averaging the conditional averages

Also when $A_i = \{Y = y\}$,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \mathbb{E}(X \mid Y = y)\, f_Y(y)\, dy \tag{44}$$

Conditional expectation

Extensions to functions of several r.v.:

$$\mathbb{E}(g(X,Y) \mid Y) = \int_{-\infty}^{\infty} g(x,y) f_{X \mid Y}(x \mid y)\, dx \tag{45}$$

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty} \mathbb{E}(g(X,Y) \mid Y)\, f_Y(y)\, dy \tag{46}$$

Independence

Two r.v. X and Y are independent of each other,

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \tag{47}$$

Since $f_{X,Y}(x,y) = f_{X \mid Y}(x \mid y)\, f_Y(y)$, for all y with $f_Y(y) > 0$,

$$f_{X \mid Y}(x \mid y) = f_X(x) \tag{48}$$

Can also interchange X and Y. Furthermore,

$$\mathbb{E}(XY) = \mathbb{E}(X)\, \mathbb{E}(Y) \tag{49}$$

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\, \mathbb{E}(h(Y)) \tag{50}$$

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) \tag{51}$$

1. Expectation, mean, variance
2. Common continuous random variables
3. Cumulative density function
4. Derived distributions
5. Joint PDF, conditioning, independence
6. Continuous Bayes' rule

Definition

Unobserved phenomenon: r.v. X with $f_X(x)$
Measurement: Y, related to X through $f_{Y|X}(y|x)$
From the observation $Y = y$, what do we know about X?
$\Rightarrow$ we want to know $f_{X|Y}(x|y)$

$$\boxed{f_{X|Y}(x|y) = \frac{f_X(x)\, f_{Y|X}(y|x)}{f_Y(y)}} \tag{52}$$

This is the continuous form of Bayes' rule. Alternatively,

$$\boxed{f_{X|Y}(x|y) = \frac{f_X(x)\, f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(t)\, f_{Y|X}(y|t)\, dt}} \tag{53}$$

Example

Suppose we have electronic devices with exponential lifetime $Y \sim \text{Exp}(\lambda)$, but λ *itself is a random variable*, modeled by uniform distribution $\Lambda \sim U(1,3)$. Given we observe the lifetime of a particular device to be $Y = y$, what do we know about λ?

ANS: We want to find $f_{\Lambda|Y}(\lambda|y)$ through Bayes' rule.

$$f_\Lambda(\lambda) = \tfrac{1}{3-1} = \tfrac{1}{2} \qquad 1 \leqslant \lambda \leqslant 3$$
$$f_{Y|\Lambda}(y|\lambda) = \lambda e^{-\lambda y}$$
$$\int_{-\infty}^{\infty} f_\Lambda(t)\, f_{Y|\Lambda}(y|t)\, dt = \int_1^3 \tfrac{1}{2} t e^{-ty}\, dt$$

Example

Integration by parts:

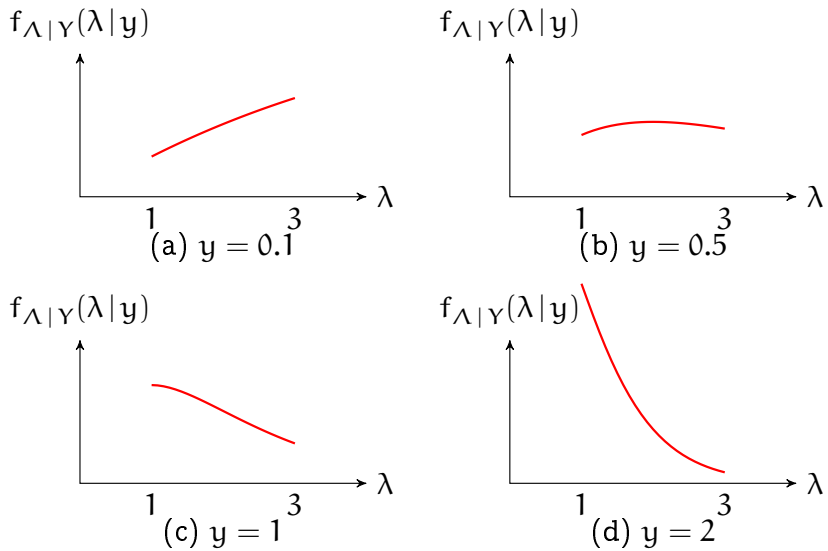
$$\int_a^b v \frac{du}{dt} dt = uv\Big|_a^b - \int_a^b u \frac{dv}{dt} dt$$

Let $v = t$ and $u = -\frac{1}{y} e^{-yt}$, $a = 1$ and $b = 3$,

$$\begin{aligned}
\tfrac{1}{2}\int_1^3 t e^{-yt} dt &= -\tfrac{t}{2y} e^{-yt}\Big|_1^3 - \int_1^3 -\tfrac{1}{2y} e^{-yt} dt \\
&= \tfrac{1}{2}\left(-\tfrac{3}{y} e^{-3y} + \tfrac{1}{y} e^{-y}\right) - \tfrac{1}{2y^2} e^{-yt}\Big|_1^3 \\
&= \tfrac{1}{2y}\left(e^{-y} - 3e^{-3y}\right) + \tfrac{1}{2y^2}\left(e^{-y} - e^{-3y}\right)
\end{aligned}$$

$$f_{\Lambda|Y}(\lambda|y) = \frac{\lambda e^{-\lambda y}}{\frac{1}{y}\left(e^{-y} - 3e^{-3y}\right) + \frac{1}{y^2}\left(e^{-y} - e^{-3y}\right)} \qquad 1 \leqslant \lambda \leqslant 3$$

Example



Special case 1: Unobserved phenomenon is discrete

Unobserved phenomenon: event A with $\mathbb{P}(A)$
Measurement: Y, related to A through $f_{Y|A}(y)$
From the observation $Y = y$, what do we know about A?
$\Rightarrow$ we want to know $\mathbb{P}(A \,|\, Y = y)$

$$\boxed{\mathbb{P}(A \,|\, Y = y) = \frac{\mathbb{P}(A)\, f_{Y|A}(y)}{f_Y(y)}} \tag{54}$$

Special case 1: Unobserved phenomenon is discrete

Case 1a: A is "binary", so $f_Y(y) = \mathbb{P}(A)\, f_{Y|A}(y) + \mathbb{P}(A^c)\, f_{Y|A^c}(y)$

$$\boxed{\mathbb{P}(A \,|\, Y = y) = \frac{\mathbb{P}(A)\, f_{Y|A}(y)}{\mathbb{P}(A)\, f_{Y|A}(y) + \mathbb{P}(A^c)\, f_{Y|A^c}(y)}} \tag{55}$$

Case 1b: A is of the form $\{N = n\}$, where $n = 1, \ldots, K$, with $p_N(n)$

$$\boxed{\mathbb{P}(N = n \,|\, Y = y) = \frac{p_N(n)\, f_{Y|N}(y|n)}{\sum_{i=1}^{K} p_N(i)\, f_{Y|N}(y|i)}} \tag{56}$$

Example

A binary signal S with $\mathbb{P}(S = 1) = p$ and $\mathbb{P}(S = -1) = 1 - p$ is transmitted, which is corrupted with independent Gaussian noise $N \sim \mathcal{N}(0, 1)$, and the received signal is $Y = N + S$. What is the probability of $S = 1$ given an observation $Y = y$?

ANS: Let $A = \{S = 1\}$.

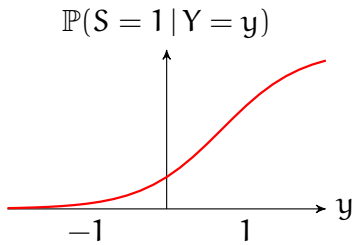
$$f_{Y|A}(y) \sim \mathcal{N}(1, 1)$$
$$f_{Y|A^c}(y) \sim \mathcal{N}(-1, 1)$$

$$\mathbb{P}(A \,|\, Y = y) = \frac{p \cdot \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}}{p \cdot \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2} + (1-p) \cdot \frac{1}{\sqrt{2\pi}} e^{-(y+1)^2/2}}$$
$$= \frac{p e^{y}}{p e^{y} + (1-p) e^{-y}}$$

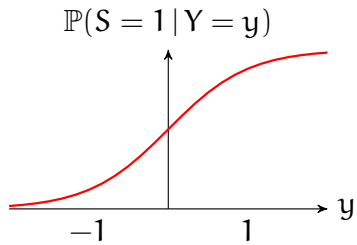
Example

$$\mathbb{P}(S = 1 \mid Y = y) = \frac{pe^{y}}{pe^{y} + (1-p)e^{-y}}$$

Let us consider specific values of p:



(a) $p = 0.2$

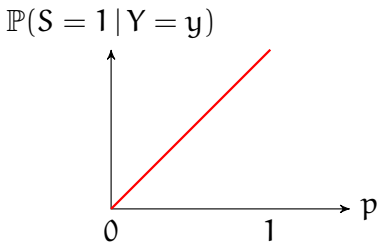


(b) $p = 0.5$

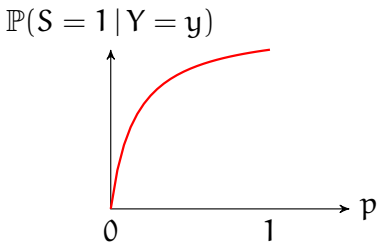
Example

$$\mathbb{P}(S = 1 \mid Y = y) = \frac{pe^{y}}{pe^{y} + (1-p)e^{-y}}$$

Let us consider specific values of y:



(a) $y = 0$



(b) $y = 1$

Conclusions

We developed the basic toolset for continuous random variables, and looked at some common distributions.