Lecture 3: Continuous Distributions

IB Paper 7: Probability and Statistics

Carl Edward Rasmussen

Department of Engineering, University of Cambridge

February 1st, 2019

Continuous Random Variables

For a *continuous* random variable X, we define the *probability density function*, $p(x)$ as:

$$p(x)dx \;=\; \text{the probability that } X \text{ takes a value between } x \text{ and } x + dx.$$

The total probability must normalize to one, so:

$$\sum_x p(x \leqslant X \leqslant x + dx) \;=\; \sum_x p(x)dx \;=\; 1.$$

In the limit of small dx's the sum becomes an integral:

$$\int_x p(x)dx \;=\; 1.$$

Note: we use the same notation for probabilities (discrete), and probability densities (continuous). Sometimes probability densities are also referred to as probabilities. Which is meant, should be clear from the context.

Expectation and Cumulative probability function

The *expected value* of a continuous random variable is defined analogously to the discrete case:

$$\mathbb{E}[X] = \int_x x p(x) dx,$$

and the *variance*:

$$\mathbb{V}[X] = \int_x \big(x - \mathbb{E}[X]\big)^2 p(x) dx = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The *cumulative* distribution function is $F(X) = p(X \leqslant x)$, so

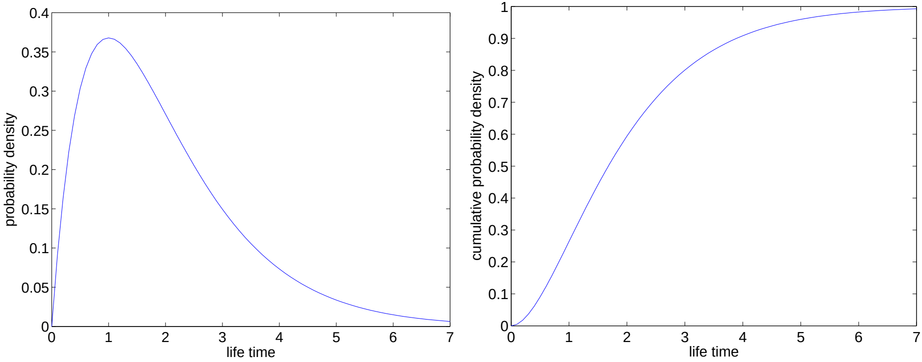
$$F(x) = \int^x p(x') dx' \qquad \text{and} \qquad \frac{dF}{dx} = p(x).$$

Depending on the context, it may be easier to work with either $p(x)$ or $F(x)$.

Example

A mechanical component has a life time X years, with probability $p(x) = x\exp(-x), x \geqslant 0$. The cumulative probability function is:

$$\begin{aligned} F(x) &= p(X \leqslant x) = \int_0^x te^{-t}dt = \big[-te^{-t}\big]_0^x + \int_0^x \exp(-t)dt \\ &= -xe^{-x} + \big[-e^{-t}\big]_0^x = 1 - e^{-x} - xe^{-x}. \end{aligned}$$



Example continued

What is the probability that failure happens within a year?
$F(1) = 1 - 2e^{-1} = 0.26$.

What is the proportion of units which will fail within a year?

What is the probability that a unit will last between 1 and 2 years?
$F(2) - F(1) = 0.33$.

What is the *most common* life time? The $\text{argmax}_x p(x) = 1$ year.

What is the *average* life time?

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x p(x) dx = \int_0^\infty x^2 e^{-x} dx = \big[-x^2 e^{-x}\big]_0^\infty + \int_0^\infty 2x e^{-x} dx \\
&= 0 + 2\big[-x e^{-x}\big]_0^\infty + 2\int_0^\infty e^{-x} dx = 0 + 2\big[-e^{-x}\big]_0^\infty = 2 \text{ years.}
\end{aligned}
$$

Time Between Poisson Events

Example: Arrivals occur randomly according to a Poisson[1] distribution $Po(\lambda)$. What is the distribution of the *time intervals* between arrivals?

Let T be the random variable representing the time between arrivals:

$$p(T = t)dt \;=\; p(t \leqslant T \leqslant t + dt),$$

ie, the probability that T lies between t and $t + dt$. This will happen if **both**

- *no* arrivals happen in the interval 0 to t. The number of arrivals in a time window t is Poisson $Po(\lambda_a = \lambda t)$. The probability of getting zero arrivals is

$$p(0 \text{ in } t) \;=\; \exp(-\lambda t).$$

- *exactly one* event happens in the interval t to $t + dt$. The number of arrivals in a time window dt is Poisson $Po(\lambda_b = \lambda dt)$. The probability of one arrival:

$$p(1 \text{ in } dt) \;=\; \lambda dt \exp(-\lambda dt).$$

Both events must happen: $p(T = t)dt = p(0 \text{ in } t) \times p(1 \text{ in } dt) = \lambda \exp(-\lambda t)dt$.

[1] Recall, the Poisson: $p(r) = Po(\lambda) = \lambda^r \exp(-\lambda)/r!$.

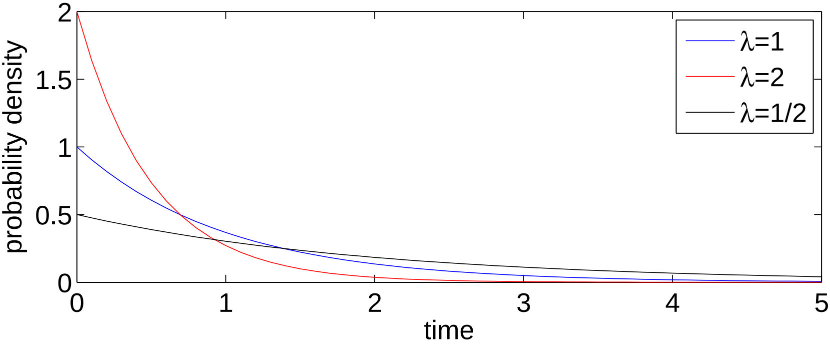
The Exponential Distribution

If

- events are independent and
- the number of events per interval are Poisson, with intensity λ

then the *time intervals* between events are exponentially distributed:

$$T \sim \mathrm{Ex}(\lambda), \qquad p(t) = \lambda \exp(-\lambda t), \text{ where } t, \lambda \geqslant 0.$$



Mean and Variance for the Exponential

The mean of the exponential

$$\mathbb{E}[T] \;=\; \int_0^\infty t p(t) dt \;=\; \int_0^\infty \lambda t \exp(-\lambda t) dt \;=\; \big[-t\exp(-\lambda t)\big]_0^\infty + \int_0^\infty \exp(-\lambda t) dt$$
$$= \big[-\tfrac{1}{\lambda}\exp(-\lambda t)\big]_0^\infty \;=\; 1/\lambda.$$

Similarly, it can be shown that the variance is

$$\mathbb{V}[T] \;=\; 1/\lambda^2.$$

Example: There are two simple ways to check whether arrival times obey a Poisson:

- mean and variance of the *number of arrivals* per unit time should be related as $\mathbb{E}[X] \simeq \mathbb{V}[X]$.
- mean and variance for *intervals between* arrivals should related as $\mathbb{E}[T]^2 \simeq \mathbb{V}[T]$.

The Gaussian or Normal Distribution

The Gaussian or Normal distribution arises in many contexts where large numbers of small random influences add up.

The Gaussian is used to approximate distributions that are not exactly Gaussian, because it is easy to compute with.

$$X \sim N(\mu, \sigma^2), \qquad p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big).$$

where μ is the mean and σ^2 is the variance.

Example: Noise is often assumed to follow a Gaussian distribution.

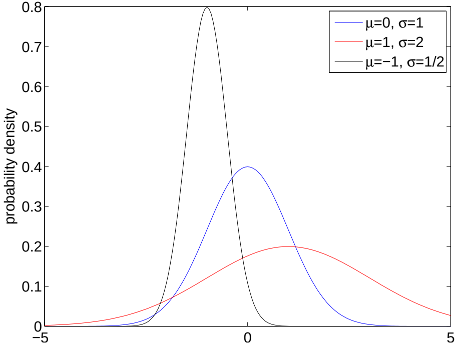
The $N(\mu = 0, \sigma^2 = 1)$ is called the *standard Normal*. If $X \sim N(\mu, \sigma^2)$, then $u = (x - \mu)/\sigma$ follows $N(0, 1)$.

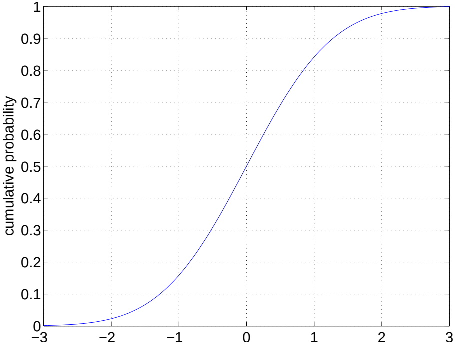
The Cumulative Gaussian

The Cumulative of the standard Gaussian is:

$$p(X \leqslant x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{(x'-\mu)^2}{2}\big)\,dx' \;=\; \Phi(x),$$

is not available in closed form.





The Cumulative Gaussian cont.

Although not available in closed form, the cumulative Gaussian is important, as it quantifies the probability mass in the tails of a Gaussian.

The function $\Phi(x)$ is commonly tabulated, see eg. in the Engineering Mathematics Data Book.

Some useful rough rules of thumb for a Gaussian:

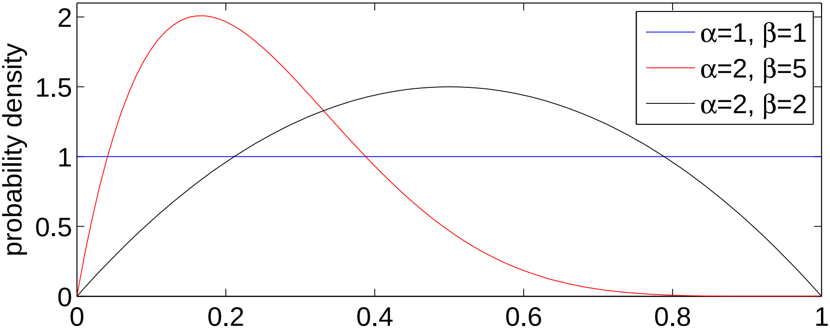
- $p(\mu - \sigma \leqslant X \leqslant \mu + \sigma) \ \simeq \ 2/3$.
- $p(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \ \simeq \ 0.95$ (95% confidence interval).

The Beta Distribution

The Beta distribution is a probability distribution *over probabilities*:

$$X \sim \mathrm{Beta}(\alpha, \beta), \quad p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \leqslant x \leqslant 1, \quad \alpha, \beta > 0,$$

where $\Gamma(\alpha) = (\alpha - 1)!$ for integer arguments.



Example: The Beta and the Binomial

Example: We make experiments with a bent coin. What is the probability π that we get Heads?

Initially, we don't know π and assume that all values are equally probable $p(\pi) = \text{Beta}(1, 1)$.

We then make 5 experiments, 4 come out Tails and 1 Head, thus Binomial: $p(x|\pi) = B(1|n = 5, \pi) = {}_5C_1\, \pi^1 (1 - \pi)^4$.

What do the experiments tell us about π?

$$p(\pi|x) \;=\; \frac{p(x|\pi)p(\pi)}{p(x)} \;=\; \text{Beta}(2, 5).$$

The above rule is an instance of Bayes rule, and is a simple consequence of the definition of conditional probability.

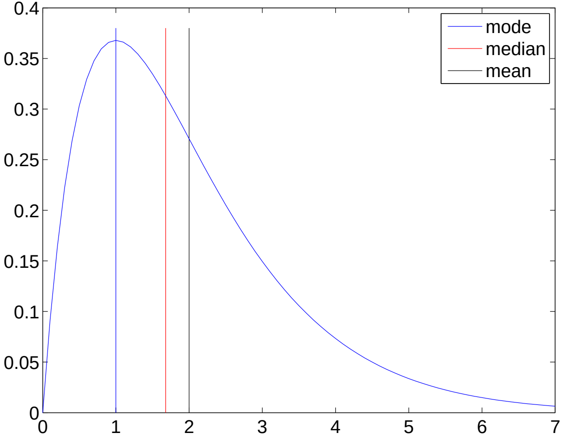
Characterizing distributions

There are many ways of characterizing discrete and continuous probability distributions. Some common ones are:

- mean or expectation, $\mathbb{E}[X] = \mu$
- variance, $\mathbb{V}[X] = \sigma^2$
- standard deviation, σ (ie the square root of the variance)
- mode, the most probable value
- median, the middle value
- quartiles, the x values such that $F(x) = 1/4$, $F(x) = 1/2$, and $F(x) = 3/4$.
- interquartile range: third minus first quartile
- skewness, definition $\mathbb{E}[(X-\mu)^3]/\sigma^3$. If the skewness is positive, the distribution is *skewed to the right*. Informally the 'tail' of the distribution is longer to the right.

Example

Here is an illustration of mode, median and mean:



In this example, the *mean* is larger than the *median*, since the distribution is *skewed to the right*.

Distribution Summaries

In different contexts, different kinds of summaries are most useful. Some summaries may even be a bit misleading…

Example: distribution of income in Britain is skewed to the right. The average income is very different from the median, since a few people have very large incomes. The logarithm of the income is much less skewed.

Example: sometimes the *variance* of the distribution can be heavily influenced by very few observations. The *interquartile range* also quantifies the *spread* of a distribution, but it is said to be *more robust toward outliers*.