50008 - Probability and Statistics - Lecture 3

Oliver Killane

24/01/22

Continuous Random Variables

Lecture Recording

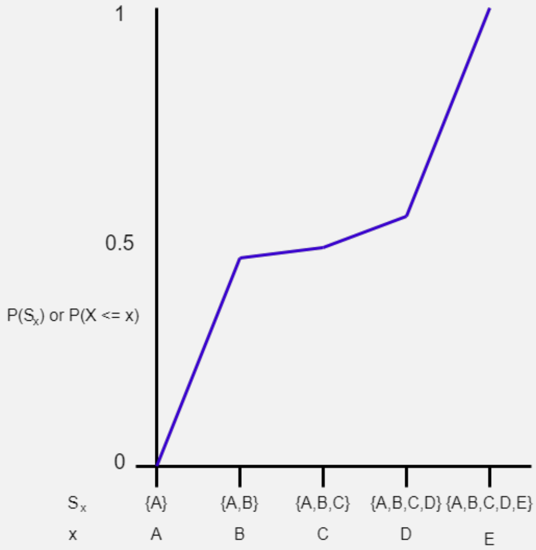
Lecture recording is available here

For continuous random variables we want to track quantities in $\mathbb{R}$ (e.g temperature, volume, other probabilities).

Induced Probability Terms

$$S_x = \{s \in S | X(s) \leq x\}$$
$$P_X(X \leq x) = P(S_x) = F_X(x)$$

S_x is the elements of the sample space up to an including x. Hence the porbability of getting S_x is the cumulative probability.



Definition: Probability Density Function

For a random variable $X : S \to \mathbb{R}$ the induced probability is defined as:

$$P_X((-\infty, x]) = P(S_X) = F_X(X)$$

A variable X is **absolutely continuous** if $\exists f_X : \mathbb{R} \to \mathbb{R}$ such that:

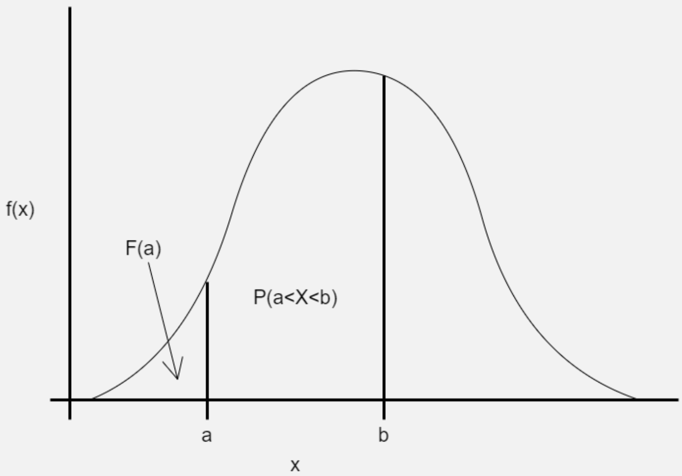
$$F_X(x) = \int_{u=-\infty}^{x} f_X(u)du$$

$$f_x(x) = F'(x) = \frac{d}{dx}F_X(x)$$

Where f_X is the **probability density function** (**pdf**).

To find probability that $X \in (a, b]$:

$$P_X(a < X \leq b) = P_X(X \leq b) - P_X(X \leq a) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$



- We can use $<$ and $\leq$ interchangeably as $P(X = x) = 0 \Leftrightarrow P(X \leq x) \equiv P(X < x)$.
- Probability of any event is zero: $P_X(X = y) = 0$, any elementary event $\{x\}$ where $x \in \mathbb{R}$ has zero probability.
- However the sum of a range of events probabilities is not zero.
- Hence the range of a continuous random variable is uncountable (i.e as $\mathbb{R}$ is also).

$$\forall x \in \mathbb{R}. f_X(x) >= 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(x)d_x = 1$$

Example: Defining a continuous random variable

Given some continuous random variable x with a probability density function given as:

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & otherwise \end{cases}$$

For some unknown constant c

To find the value of c we use the requirement that the cumulative distribution must sum to 1:

$$\int_0^3 cx^2 = 1 \rightsquigarrow [\frac{cx^3}{3}]_0^3 = 1 \rightsquigarrow (9c) - 0 = 1 \rightsquigarrow c = {}^1/_9$$

Hence:

$$f(x) = \begin{cases} \frac{x^2}{9} & 0 < x < 3 \\ 0 & otherwise \end{cases}$$

Hence we can specify the cumulative probability distribution as:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^3}{27} & 0 < x < 3 \\ 1 & x \geq 0 \end{cases}$$

We can then calculate probabilities using the cumulative distribution:

$$P(1 < X < 2) = F(2) - F(1) = \frac{2^3}{27} - \frac{1^3}{27} = \frac{7}{27} \approx 0.259$$

Mean, Variance and Quantiles

Lecture Recording

Lecture recording is available here

Definition: Expected (Continuous)

The **mean** or **expected** of a continuous random variable X:

$$\mu_X = E_X(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

For a function of interest that is applied to the random variable $g : \mathbb{R} \to \mathbb{R}$:

$$E_X(g(X)) \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

- $E(aX + b) = aE(X) + b$
- $E(g(X) + h(X)) = E(g(X)) + E(h(X))$

Definition: Variance (Continuous)

The variance of a continuous random variable X:

$$\sigma_X^2 = Var_X(X) = E((X - \mu_X)^2) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx$$

We can show this as:

$$\begin{aligned} Var_X(X) &= \int_{-\infty}^{\infty} x^2 f_X(x)ds - \mu_X^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

For a linear transformation:

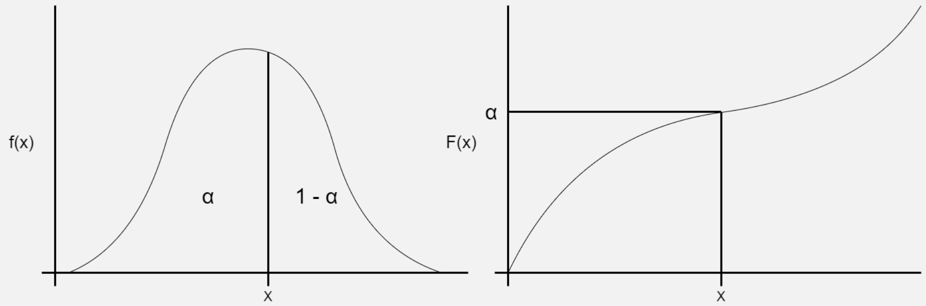
$$Var(aX + b) = a^2 Var(X)$$

Definition: Quartiles

The lower, upper quartiles and median are points

For a continuous random variable X, we define the α-**Quantile** $Q_X(\alpha)$ where $0 \leq \alpha \leq 1$ as the lowest X such that:

$$P(X \leq Q_X(\alpha)) = \alpha \quad \text{or in other words} \quad Q_X(\alpha) = F_X^{-1}(\alpha)$$



Using Q_X we can define some standard quantiles:

- **Quartiles** Lower Quartile ($\alpha = {}^1/_4$), Median ($\alpha = {}^1/_2$) and Upper Quartile ($\alpha = {}^3/_4$)
- **Percentiles** The nth percentile: $\alpha = \dfrac{n}{100}$

Example: Basic continuous random variable

Given continuous random variable X:

$$f(x) = \begin{cases} \frac{x^2}{9} & 0 < x < 3 \\ 0 & otherwise \end{cases}$$

We can calculate the expected:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{-\infty}^{0} xf(x)dx + \int_{0}^{3} xf(x)dx + \int_{3}^{\infty} xf(x)dx \\ &= \int_{-\infty}^{0} x \times 0dx + \int_{0}^{3} xf(x)dx + \int_{3}^{\infty} x \times 0dx \\ &= \int_{0}^{3} xf(x)dx = \int_{0}^{3} \frac{x^3}{9}dx = \left[\frac{x^4}{36}\right]_0^3 \\ &= \frac{9}{4} = 2.25 \end{aligned}$$

We can calculate the variance:

$$\begin{aligned} Var(X) &= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu_X^2 \\ &= \int_{-\infty}^{0} x^2 f_(x)dx + \int_{0}^{3} x^2 f(x)dx + \int_{3}^{\infty} x^2 f(x)dx - \mu_X^2 \\ &= \int_{0}^{3} x^2 f(x)dx - \mu_X^2 = \int_{0}^{3} \frac{x^5}{9}dx - \mu_X^2 \\ &= 27 - \mu_X^2 = 27 - 2.25 = 24.75 \end{aligned}$$

we can calculate the median, we ignore the range $x > 3$ as the median must be below this.

$$\begin{aligned} 0.5 &= \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{0} f(y)dy + \int_{0}^{x} f(y)dy = \int_{0}^{x} f(y)dy \\ 0.5 &= \int_{0}^{x} \frac{y^2}{9} = \left[\frac{y^3}{27}\right]_0^x = \frac{x^3}{27} \\ x &= \sqrt[3]{0.5 \times 27} \approx 2.38 \end{aligned}$$

Notable Continuous Distributions

Lecture Recording

Lecture recording is available here

Definition: Continuous Uniform Distribution

A continuous random variable with equal probability of being any value within a range:

For $X \sim U(a, b)$:

PDF	CDF	Expected	Variance
$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$	$F_X(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$	$\mu = \dfrac{a+b}{2}$	$\sigma^2 = \dfrac{(b-a)^2}{12}$

The standard uniform distribution is defined as $X \sim U(0, 1)$:

PDF	CDF	Expected	Variance
$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & otherwise \end{cases}$	$F_X(x) = \begin{cases} 0 & x \leq a \\ x & a < x < b \\ 1 & x \geq b \end{cases}$	$\mu = {}^1/_2$	$\sigma^2 = {}^1/_{12}$

Other uniform distributions can be mapped linearly to the standard uniform.

Example: Mapping to Standard Uniform

Given $X \sim U(2, 5)$ find the expected, variance and median.

Take $Y \sim U(0, 1)$, $X = 3 \times Y + 2$.

Distribution	**Expected**	**Variance**	**Median**
Y	0.5	${}^1/_{12}$	0.5
X	3.5	${}^3/_4$	3.5

Definition: Exponential Distribution

Given a rate of events λ, what is the probability of waiting X time for the event to occur.

For $X \sim Exponential(\lambda)$ or $X \sim Exp(\lambda)$ where $\lambda > 0$:

PDF	CDF	Expected	Variance
$f_X(x) = \lambda e^{-\lambda x}$ where $x \geq 0$	$F_X(x) = 1 - e^{-\lambda x}$ where $x \geq 0$	$\mu_X = \frac{1}{\lambda}$	$\sigma^2 = \frac{1}{\lambda^2}$

The distribution has the **Lack of memory property**, namely the time waited already does not affect the next part of the distribution (same shape).

$$P(X > x+t|X > t) = \frac{P(X > x + t \cap X > t)}{P(X > t)} = \frac{P(X > x + t)}{P(X > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = P(X > x)$$

$$P(X > x + t|X > t) = P(X > x)$$

This distribution can be combined with Poisson. Given $X \sim Poisson(\lambda)$ (events occurring in a given time frame), the time between events is modelled by $X \sim Exponential(\lambda)$ (interval time for one event).

There is a variant with θ as the parameter for the distribution where $\theta = \frac{1}{\lambda}$.

Definition: Normal Distribution

Given a mean value (μ) and a variance (σ^2) from the mean the symmetrical distribution is a **Normal Distribution**.

For $X \sim Normal(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma^2)$ where $\sigma > 0$:

PDF	CDF
$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt$

The **Standard/Unit Normal Distribution** is $X \sim N(0, 1)$:

PDF	CDF
$\phi(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{1}{2}x^2\right\}$	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$

We can apply linear functions:

$$X \sim N(\mu, \sigma^2) \to \text{ and } aX + b \sim N(a\mu + b, a^2\sigma^2)$$

Hence we can use the **Standard Normal Distribution**:

$$X \sim N\mu, \sigma^2 \Rightarrow \frac{X - \mu}{\sigma} \sim N(0, 1) \text{ and hence } P(X \leq x) = \Phi(\frac{x - \mu}{\sigma})$$

Definition: Lognormal Distribution

Given $X \sim N(\mu, \sigma^2)$ and $Y = e^X$ we can compute the **PDF** of Y:

$$f_Y(y) = \frac{1}{\sigma y \sqrt{2\pi}} exp\left[-\frac{(\log y - \mu)^2}{2\sigma^2}\right]$$

Central Limit Theorem

Lecture Recording

Lecture recording is available here

Definition: Moment Generating Function

The moment generating function M_X for a continuous random variable X is:

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

Assuming the calculus within the $E(\dots)$ is valid, the nth moment is given by:

$$E[X^n] = \left.\frac{d^n M_x(t)}{dt^n}\right|_{t=0}$$

If the integral does not exist, the **characteristic function** $\phi_X(t) = M_X(\iota t)$ can be used (ι is imaginary unit).

Example: Expected and Variance

$$\begin{aligned}
E[X] &= \left.\frac{dM_x(t)}{dt}\right|_{t=0} \\
&= \left.\frac{dE[e^{tX}]}{dt}\right|_{t=0} \\
&= \left.\frac{d\int_{-\infty}^{\infty} e^{tx} f_X(x)dx}{dt}\right|_{t=0} \\
&= \left.\int_{-\infty}^{\infty} xe^{tx} f_X(x)dx\right|_{t=0} \\
&= \int_{-\infty}^{\infty} xe^{0x} f_X(x)dx \\
&= \int_{-\infty}^{\infty} xf_X(x)dx
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= \left.\frac{d^2M_x(t)}{dt^2}\right|_{t=0} \\
&= \left.\frac{d^2E(e^{tX})}{dt^2}\right|_{t=0} \\
&= \left.\frac{d^2\int_{-\infty}^{\infty} e^{tx} f_X(x)dx}{dt^2}\right|_{t=0} \\
&= \left.\frac{d\int_{-\infty}^{\infty} xe^{tx} f_X(x)dx}{dt}\right|_{t=0} \\
&= \left.\int_{-\infty}^{\infty} x^2e^{tx} f_X(x)dx\right|_{t=0} \\
&= \int_{-\infty}^{\infty} x^2e^{0x} f_X(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f_X(x)dx
\end{aligned}$$

$$Var[X] = E[X^2] - (E[X])^2$$

Product of Random Variables

Given independent random variables $Z_1, Z_2, \ldots, Z_n$:

$$E[\prod_{i=1}^{n} Z_i] = \prod_{i=1}^{n} E[Z_i]$$

The sum of the random variables is the products of their **Moment Generating Functions**.

$$M_{Z_1+Z_2}(t) = E[e^{t(Z_1+Z_2)}] = E[e^{tZ_1}e^{tZ_2}] = E[e^{tZ_1}]E[e^{tZ_2}] = M_{Z_1}(t)M_{Z_2}(t)$$

$$S_n = \sum_{i=1}^{n} Z_i \Rightarrow M_{S_n}(t) = \prod_{j=1}^{n} MX_j(t)$$

Central Limit Theorem

Definition: Central Limit Theorem

Given $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables from any distribution with mean μ and finite variance σ^2.

$$S_n = \sum_{i=1}^{n} X_i$$

Hence we have a distribution with a known expected and variance, so can form a **Normal Distribution**.

$$\begin{array}{ccc} Y = S_n & E(Y) = n\mu & Var(Y) = n\sigma^2 \\ Y = S_n - n\mu & E(Y) = 0 & Var(Y) = n\sigma^2 \\ Y = \dfrac{S_n - n\mu}{\sqrt{n}\sigma} & E(X) = 0 & Var(X) = 1 \end{array}$$

Y can now be used to approximate a **Standard Normal Distribution**.

$$\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

This implies that for large (but finite n):

$$\overline{X} \approx N(\mu, \frac{\sigma^2}{n}) \quad \text{and} \quad \sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2)$$

Where $\overline{X}$ is the average value of the random variables $\frac{\sum_{i=0}^{n} X_i}{n}$.

The approximation holds for all distributions (including discrete), and is exact when the random variables are from the same **normal distribution**.

An attempt at CLT proof

Given the random variables $X_1, X_2, \ldots, X_n$ we can standardize and get their sum:

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}\sigma} = \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}\sigma} \quad \text{where} \quad Y_i = X_i - \mu$$

The moment generating function of Z_n is the product of the **moment generating functions** of the Y (all identically distributed, so identical **MGFs**).

$$M_{Z_n}(t) = \left(M_Y \left(\frac{t}{\sqrt{n}\sigma} \right) \right) \quad \text{where } M_y \text{ is the moment generating function for all } Y_i$$

We can then expand the M_Y around 0 using Taylor's Theorem:

$$M_Y(t) = M_Y(0) + M'_Y(0)t + {}^1\!/\!{}_2 M''_Y(0)t^2 + O(t^3)$$

$O(t^3)$ is the error term of our approximation, as this is for higher powers, it has a small effect so can be ignored

The derivatives of the **MFG** are:

$M'_Y(0) = E(Y_i) = 0$ due to shift performed earlier and $M''_Y(0) = E(Y_i^2) = \sigma^2 + E(Y_i)^2 == \sigma^2 + 0 = \sigma^2$

Hence we can derive:

$$M_Y(t) = 1 + \frac{\sigma^2 t^2}{2} + O(t^3)$$

Hence we can scale t, and ignore the error term for simplicity:

$$M_Y \left(\frac{t}{\sqrt{n}\sigma} \right) = 1 + \frac{t^2}{2n}$$

As the error term gets very small, we can use limits to get an approximation for $M_{Z_n}(t)$.

$$lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} (1 + \frac{t^2}{2n} + O(n^{-3/2}))^n = e^{t^2/2}$$

Note that $\lim_{m\to\infty}(1 + \frac{x}{m})^m = e^x$.

Example: Coin Tossing

Consider a set of count tosses, each are Bernoulli discrete random variables (take values 0 or 1).

$$X_1, X_2, X_3, \ldots, X_n \text{ where } \mu = p \text{ and } \sigma^2 = p(1-p)$$

The total score of toin tosses can be modelled as a binomail distribution:

$$\sum_{i=1}^{n} X_i \text{ is } X \sim Binomial(n,p) \text{ with } E(X) = np \text{ and } Var(X) = np(1-p)$$

For large n can also model it as a normal distribution:

$$\sum_{i=1}^{n} X_i \text{ is } X \sim N(n\mu, n\sigma^2) \equiv N(np, np(1-p))$$

As the number of events (coin tosses) tends to infinity, the distributions tend to look identical.