Basics of continuous distributions

Li Li

Department of Mathematics and Statistics

Continuous random variables

$X = [a, b]$

A continuous random variable is a variable that can take uncountable many values.

$X = \{x \mid a < x < b\}$

- Continuous random variables almost never take an exact prescribed value c with a nonzero probability, but there is a positive probability that its value will lie in particular intervals which can be arbitrarily small.

  $P(X = c) = 0$

- For example, a random variable measuring the time taken for something to be done is continuous since there are an infinite number of possible times that can be taken.

Continuous random variables

A discrete random variable is a random variable that can take up to a countable number of values.

- Discrete random variables can take an exact prescribed value c in the sample space with nonzero probability.
- For example, the number of heads in 10 tosses of fair coins can take 5 with a probability of 0.246.

Sometimes, we use continuous random variables to approximately describe discrete random variables that has many values. For example,

- A random variable measuring the level of cholesterol in milligrams per deciliter can be considered as continuous approximately since there are many possible values it can take.

Probability of an event

- For a continuous random variable X, an event is typically written as $\{a < X < b\}$.

$P(X=c)$ ✗

- The probability of an event is written as $P(a < X < b)$.

Since there are uncountable number of possible values in the interval, a **probability density function (PDF)** is used to specify the probability of the random variable falling within a particular range of values. A probability density function, denoted by $f(x)$, a is a function such that

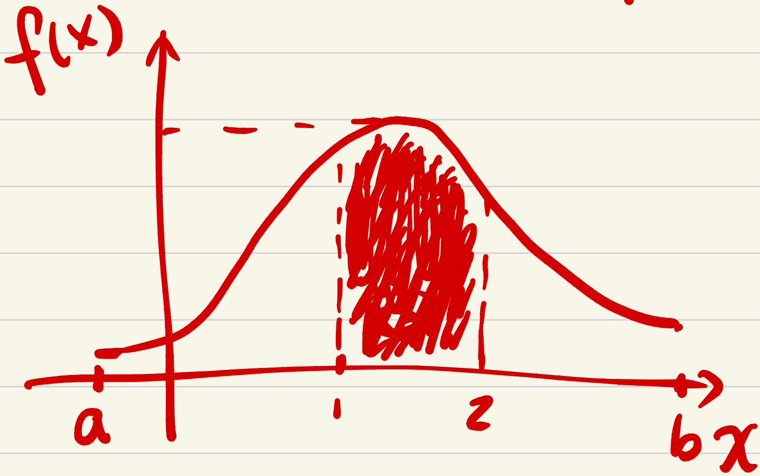
- $P(a < X < b) = \int_a^b f(x)dx$ (the area underneath the curve and between a and b).
- $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$. (the area is exactly 1 underneath it). $= \int_a^b f(x)dx = 1$

For example, if $f(x) = 0.1$ for $0 < x < 10$ and 0 elsewhere, then $P(1 < X < 3) = \int_1^3 0.1dx = 0.2$.

$= 0.1(3-1)$

$P(-5 < X < 3) = \int_0^3 0.1\,dx$
$= 0.3$

$f(x) \geq 0$



$$p(1 < X < 2)$$

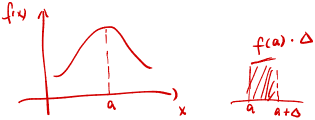
$$= \int_1^2 f(x)\,dx$$

$$= \int_1^2 f(s)\,ds$$

$\cancel{\sum f(x) = 1}$

$\int_a^b f(x)\,dx = 1$

Interpretation of the probability density function



- Over a very short interval, $P(a < X < a + \triangle) \approx f(a)\triangle$. So we can interpret the density $f(a)$ as the “rate” of probability for X taking value around a.
- A relative likelihood that the value of the random variable would equal that sample. The value of the PDF at two different samples can be used to infer how much more likely it is that the random variable would equal one sample compared to the other sample.

Examples

Suppose a species of bacteria typically lives 4 to 6 hours and the rate of dying over the two hours is constant. Denote X as the time of death for a randomly selected bacteria.

(a) What is the probability that a bacterium lives exactly 5 hours? 0

(b) What is the probability that the bacterium dies between 5 hours and 5.01 hours?

$$\frac{5.01-5}{6-4}$$

Examples

Example

Let X be a continuous random variable with a probability density function (PDF)

$$f(x) = \begin{cases} cx^2, & -1 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the value of c?

(b) Find $P(0.5 < X < 1.5)$.

a) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

$= \int_{-1}^{2} c x^2\,dx$

$= \left.\frac{cx^3}{3}\right|_{-1}^{2} = \frac{c \cdot 2^3}{3} - \left[\frac{c \cdot (-1)^3}{3}\right]$

$= \frac{c \cdot 8}{3} + \frac{c}{3} = 3c = 1$

$c = \frac{1}{3}$

b) $= \int_{0.5}^{1.5} \frac{1}{3} x^2\,dx$

$= \left.\frac{x^3}{9}\right|_{0.5}^{1.5} = 0.3612$

Example

Exercise

Let a random variable X have a probability density function (PDF)

Exponential density.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$\lambda > 0$

(a) What values of λ would make the density function valid?

(b) Find $P(X > 3)$.

$e^0 = 1 \quad e^{+\infty} = +\infty$

$e^{-\infty} = 0$

a) $\int_0^{+\infty} \lambda e^{-\lambda x} \, dx = 1$

$= \left[- e^{-\lambda x} \right] \Big|_0^{+\infty} = [-0] - [-e^{-\lambda \cdot 0}]$

$f(x) < \infty$ for all x, $= 1$ hence $\lambda > 0$

Exercise

Cumulative distribution function

The **cumulative distribution function (CDF)** of a continuous random variable X, denoted as $F(x)$, is defined as

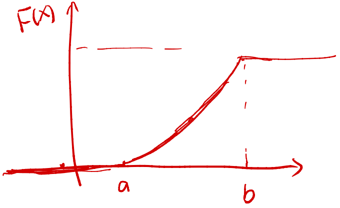
$$\boxed{F(x) = P(X \le x)} = \int_{-\infty}^{x} f(u)du$$

$f(u)du$

$f(y)dy$

$(-\infty, x]$

for $-\infty < x < \infty$.
For example, if $f(x) = 1$ for $0 < x < 1$ and zero elsewhere.
Then



$$F(x) = \begin{cases} 0 & x < a \\ \int_a^x f(u)du & a < x < b \\ 1 & x > b \end{cases}$$

Where do we use cumulative distribution function?

Median $\longleftrightarrow$ 50th percentile of the distribution

Q_1 (First quartile) $\longleftrightarrow$ 25th percentile of the distribution.

- It is used to define percentiles (theoretical counterparts of quantiles): *u*-th percentile of a distribution is defined as a value c so that $F(c) = u/100$. For example, 25-th percentile of the distribution F is c so that $F(c) = 0.25$.
- It is used in survival analysis where $F(t) = P(T < t)$ describes the probability of a subject lives up to t unit of time.

  $P(T > 2) = 1 - F(2)$
- It is used in countless statistical theories for establishing properties of statistical analysis.

Examples

Let X be a continuous random variable with PDF

$$f(x) = \begin{cases} \frac{1}{16}(x+7), & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

Find the cumulative distribution function (CDF) of X.

$$F(x) = \begin{cases} 0 & x<0 \\ \int_0^x \frac{1}{16}(u+7)\,du = \frac{1}{16}\left(\frac{x^2}{2}+7x\right) & 0 \leq x \leq 2 \\ 1 & x>2 \end{cases}$$

$$\frac{d\left(\frac{1}{16}\left(\frac{x^2}{2}+7x\right)\right)}{dx} = \frac{1}{16}\left(\frac{2x}{2}+7\right)$$

$$\int_0^x \frac{1}{16}(u+7)\,du = \frac{1}{16}\left(\frac{u^2}{2}+7u\right)\Big|_0^x = \frac{1}{16}\left(\frac{x^2}{2}+7x\right) - 0 \qquad = \frac{1}{16}(x+7) \qquad 0 \leq x \leq 2$$

Examples

Let the random variable X have PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the cumulative distribution function (CDF) of X.

$$F(x) = \begin{cases} 0 & x \le 0 \\ \int_0^x \lambda e^{-\lambda u}\, du & x > 0 \end{cases}$$

$$= \underline{1 - e^{-\lambda x}}$$

Examples

Let the number of hours for a randomly selected LED light bulb have PDF

$$f(x) = \begin{cases} \frac{1}{40{,}000} e^{-\frac{x}{40{,}000}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the 25-th percentile and 99-th percentile of the distribution.

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\frac{1}{40{,}000}x} & x > 0 \end{cases}$$

c: 25th percentile

$F(c) = 0.25$

$$1 - e^{-\frac{1}{40,000}c} = 0.25$$

$$0.75 = e^{-\frac{1}{40,000}c}$$

$$\log(0.75) = -\frac{1}{40,000}c$$

$$-0.287 = -\frac{1}{40,000}c$$

$$c = 11,507$$

d: 99th percentile

$F(d) = 0.99$

$$1 - e^{-\frac{1}{40,000}d} = 0.99$$

$$\log(0.01) = -\frac{1}{40,000}d$$

$$-4.605 = -\frac{1}{40,000}d$$

$$d = 184,207$$

Using CDF to obtain PDF

A logistic distribution is a continuous distribution that appears in logistic regression and neural networks. A logistic distribution has CDF as follows

$$F(x) = \frac{1}{1 + e^{-x}}, -\infty < x < +\infty.$$

Find its density function.

$$f(x) = \frac{dF(x)}{dx} = \frac{-(e^{-x})'}{(1 + e^{-x})^2} = \frac{e^{-x}}{(1 + e^{-x})^2}, -\infty < x < +\infty.$$

Interesting facts about CDF

- CDF of a continuous random variable is always continuous while CDF of a discrete random variable is only right-continuous.
- Transforming a continuous random variable X using its own CDF function $F(x)$ follows a continuous uniform distribution, i.e. $Y = F(X)$ has density $f(y) = 1$ for $0 < y < 1$ and $f(y) = 0$ elsewhere.

e.g $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$

$Y = F(X) = 1 - e^{-\lambda X}$

Density of Y: $f(y) = 1$, $0 < y < 1$

Mean

The **mean (or the expected value of** X**)**, denoted as μ or $E(X)$ for a continuous distributed is defined as

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

For discrete distributions

$$E(X) = \sum_{x_i} x_i f(x_i)$$

For example, if

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$E(X) = \int_0^1 x * 1dx = 0.5.$

Example

If a PDF has the following form

$$f(x) = \begin{cases} x^2/3, & -1 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

Find its mean.

$$= \left. \frac{x^4}{12} \right|_{x=-1}^{x=2}$$

$$\mu = \int_{-1}^{2} x \cdot \frac{x^2}{3} \, dx = \int_{-1}^{2} \frac{x^3}{3} \, dx = \left. \frac{x^4}{12} \right|_{-1}^{2}$$

$$= \frac{16}{12} - \frac{1}{12} = \frac{15}{12} = \frac{5}{4}$$

$$= \boxed{}$$

1.25

Mean vs sample mean

- Mean alone refers to the mean of a distribution or a population.
- When only a sample from a population is available, denoted by $x_1, x_2, \ldots x_n$, we calculate sample mean using

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

  Sample mean is a good way to estimate the distribution mean/population mean (see Chapter 7).
- An interval estimate of mean is available (Chapter 8).

Variance

The **variance of X, denoted as** $Var(X)$ **or** σ^2 , is

Discrete distr.

$$\boxed{Var(X) = E[(X-\mu)^2]} = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \qquad \sum_{x_i} (x_i - \mu)^2 f(x_i)$$

where μ is the mean of X.
For example, if

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$\sigma^2 = \int_0^1 (x-0.5)^2 * 1dx = \int_0^1 x^2 - 0.5x + 0.25dx = 0.083.$

Variance

Distr. $Var(x) = \underbrace{\sum_{x_i} x_i^2 f(x)}$ $- \mu^2$

The variance can be also be computed as

$$Var(X) = E(X^2) - (E(X))^2 = \underline{\left(\int_{-\infty}^{+\infty} x^2 f(x) dx\right) - \mu^2}$$

For example, if

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$\sigma^2 = \int_0^1 x^2 * 1dx - 0.5^2 = 0.083.$

Example

If a PDF has the following form

$$f(x) = \begin{cases} x^2/3, & -1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Find the variance of the distribution.

$$E(X^2) = \int_{-1}^{2} x^2 \cdot \frac{x^2}{3}\, dx$$

$$= \left. \frac{x^5}{15} \right|_{-1}^{2} = \left[\frac{32}{15}\right] - \left[\frac{-1}{15}\right]$$

$$= \frac{33}{15} = \frac{11}{5}$$

$$\sigma^2 = \frac{11}{5} - 1.25^2 = 2.2 - 1.5625 = \boxed{} > 0$$

Variance vs sample variance

- Variance alone refers to the variance of a distribution or a population.
- When only a sample from a population is available, denoted by $x_1, x_2, \ldots x_n$, we calculate sample variance using
$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$
Sample variance is a good way to estimate the distribution variance/population variance (see Chapter 7).
- An interval estimate of variance is available when the distribution is Normal (Chapter 4, Chapter 8).

Standard deviation

The **standard deviation** of X is the square root of its variance, denoted by σ or $SD(X) = \sqrt{Var(X)}$.

- Standard deviation alone refers to standard deviation of a distribution or a population.
- Standard deviation, unlike the variance, is expressed in the same units as the data or the random variable.
- In science, many researchers report the standard deviation of experimental data, and only effects that fall much farther than two standard deviations away from what would have been expected are considered statistically significant.

Moments

The kth moment a continuous random variable X is denoted by $E(X^k)$ (same as the notation for moments when X is a discrete variable).

$$E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx.$$

Discrete.

$$E(X^k) = \sum_{x_i} x_i^k f(x_i)$$

For example, if

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

- 1-st moment: $E(X) = \int_0^1 x * 1 dx = 0.5.$
- 2-nd moment: $E(X^2) = \int_0^1 x^2 * 1 dx = 0.333.$
- k-th moment: $E(X^k) = \int_0^1 x^k * 1 dx = \frac{1}{k+1}.$

Moment generating function

Discrete. $M(t) = \sum_{x_i} e^{tx_i} f(x_i)$

Moment generating function of X:

$$M(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx$$

as long as the integration is finite for some interval of t around 0. For example, if

Discrete

$E(g(x))$

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Continuous

$E(g(x))$

$$= \sum_{x_i} g(x_i) f(x_i) \qquad M(t) = \int_0^1 e^{tx} dx = \frac{e^t - 1}{t}. \qquad = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Use of the moment generating function

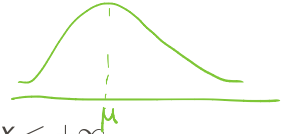
If a moment-generating function exists for a random variable X, then:

(a) The mean of X can be found by evaluating the first derivative of the moment-generating function at $t = 0$. That is: $\mu = M'(0)$.

(b) The variance of X can be found by evaluating the first and second derivatives of the moment-generating function at $t = 0$. That is: $\sigma^2 = M''(0) - (M'(0))^2$.

Example

For example, for the Normal distribution



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty.$$

Find the mean and variance using its moment generating function:

$$M(t) = e^{\frac{\sigma^2 t^2 + 2\mu t}{2}}$$

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$M'(t) = e^{\frac{\sigma^2 t^2 + 2\mu t}{2}} \cdot \left(\frac{2t\sigma^2 + 2\mu}{2}\right) = e^{\frac{\sigma^2 t^2 + 2\mu t}{2}} \cdot (t\sigma^2 + \mu)$$

$$M'(0) = \mu$$

$$M''(0) = \mu^2 + \sigma^2$$

$\rightarrow$ Variance $= \sigma^2$

$$y = \frac{x-\mu}{\sigma} \Rightarrow \boxed{x = y\sigma + \mu} \Rightarrow x-\mu = y\sigma$$

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \int_{-\infty}^{+\infty} e^{t(y\sigma+\mu)} \cdot \frac{1}{\sqrt{2\pi\cancel{\sigma^2}}} e^{-\frac{y^2\sigma^2}{2\sigma^2}} dy \cdot \cancel{\sigma}$$

$$= \int_{-\infty}^{+\infty} e^{ty\sigma} \cdot e^{t\mu} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

$$= e^{t\mu} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}+ty\sigma} dy$$

$$e^{-\frac{y^2}{2}+ty\sigma} = e^{-\frac{1}{2}(y^2-2ty\sigma)} = e^{-\frac{1}{2}(y-t\sigma)^2+\frac{t^2\sigma^2}{2}}$$

$$= e^{t\mu} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(y-t\sigma)^2} \cdot e^{\frac{t^2\sigma^2}{2}} dy$$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t\sigma)^2} dy = 1 \quad (\sigma = 1, \ \mu = t\sigma)$$

$$= e^{t\mu} \cdot e^{\frac{t^2\sigma^2}{2}}$$

$$= e^{\frac{t^2\sigma^2}{2}+t\mu}$$

- Let $g(X)$ be a function of a continuous random variable X where X has density function $f(x)$, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

- $E(g_1(X) + g_2(X) + \cdots + g_k(X)) = E(g_1(X)) + E(g_2(X)) + \cdots + E(g_k(X))$ where k is an integer. For example,

$$E(X-\mu)^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - \mu^2.$$

$$= E(X^2) + E[-2\mu X] + E(\mu^2)$$

$$= E(X^2) - 2\mu \cdot \underset{\mu}{\underline{E(X)}} + \mu^2 = E(X^2) - \mu^2$$

Variable transformation

- In data analysis variable transformation is the replacement of a variable by a function of that variable. It is common to perform variable transformation to recenter and rescale the variable (standardization), make highly skewed distributions less skewed (log or square root transformation).
- The theoretical counterpart of data transformation is random variable transformation. Given the distribution of a continuous random variable X, it is of interest to ask
  - what is the distribution of $\frac{X-a}{b}$ (standardization)?
  - what is the distribution of $\sqrt{X}$ (square root transformation)?
  - what is the distribution of $log(X)$ (log transformation)?

Example

If $\left(\frac{0-a}{b}, \frac{1-a}{b}\right)$

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Let b be a positive number. What are the distributions of $(X-a)/b$, $\log(X)$, and $\sqrt{X}$? The distributions are determined if you can find the PDF of both variables.

$Y = \frac{x-a}{b}$

$W = \log(x)$

$Z = \sqrt{x}$

CDF approach

$F(y) = P(Y \leq y)$

$= P\left(\frac{x-a}{b} \leq y\right)$

$= P(X \leq by + a)$

Example

$$= \frac{by+a-0}{1-0} = by+a \quad (0<by+a<1)$$

$$f(y) = F'(y) = b; \quad -\frac{a}{b} < y < \frac{1-a}{b}$$

$$-\infty < W < 0$$

$$F(w) = P(W \le w)$$

$$= P(\log(x) \le w)$$

$$= P(x \le e^{w}) = \frac{e^{w}-0}{1-0} = e^{w}$$

$$f(w) = e^{w}; \quad -\infty < w < 0$$