Topic 4: Continuous random variables

Rohini Somanathan

Course 003, 2018

Continuous random variables

Definition (Continuous random variable): *An r.v.* X *has a continuous distribution if there exists a non-negative function* f *defined on the real line such that for any interval* A*,*

$$P(X \in A) = \int_A f(x)\,dx$$

The function f *is called the probability density function (PDF) of* X*. Every PDF must satisfy:*

1. *Nonnegative:* $f(x) \geq 0$
2. *Integrates to 1:* $\int_{-\infty}^{\infty} f(x)\,dx = 1$

The CDF of X *is given by:*

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

A continuous random variable is a random variable with a continuous distribution.

Continuous r.v.s: caveats and remarks

1. The density is not a probability and it is possible to have $f(x) > 1$.
2. $P(X = x) = 0$ for all x. We can compute a probability for X being very close to x by integrating f over an ϵ interval around x.
3. Any function satisfying the properties of the PDF, represents the density of some r.v.
4. For defining probabilities, it doesn't matter whether we include or exclude endpoints:

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Expectation of a continuous r.v.

Definition (Expectation of a continuous random variable): *The expected value or mean of a continuous r.v.* X *with PDF* f *is*

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

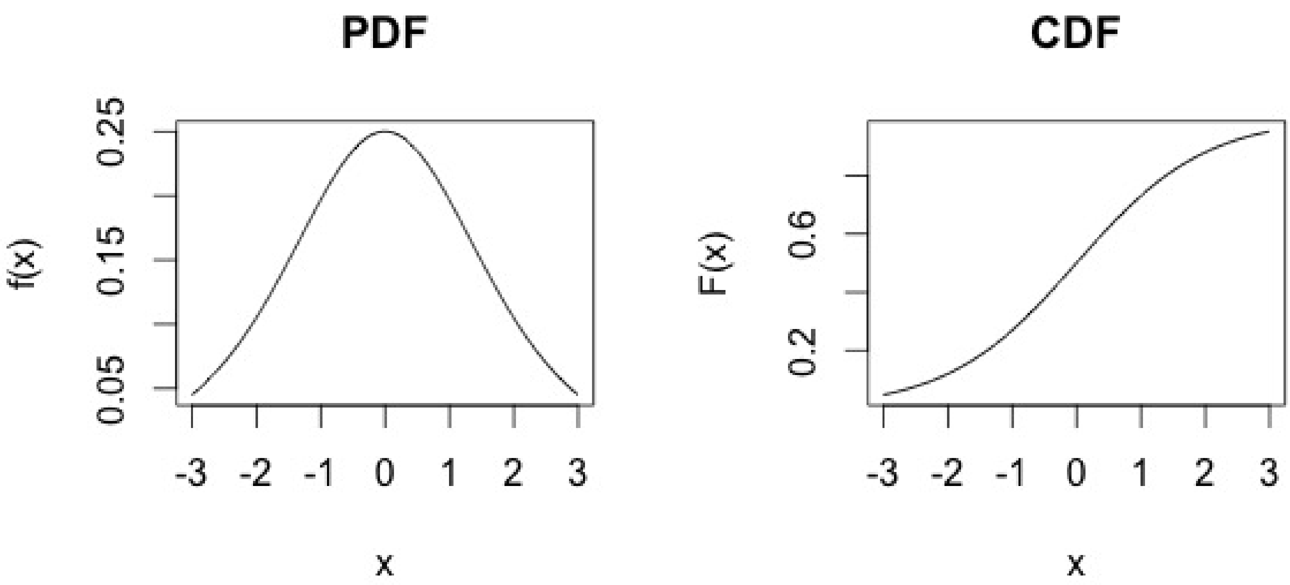
Result (Expectation of a function of X): *If* X *is a continuous r.v.* X *with PDF* f *and* g *is a function from* $\mathbb{R}$ *to* $\mathbb{R}$*, then*

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

The Logistic distribution

The CDF of a logistic distribution is $F(x) = \frac{e^x}{1+e^x}, \quad x \in \mathbb{R}$.

We differentiate this to get the PDF, $f(x) = \frac{e^x}{(1+e^x)^2}, \quad x \in \mathbb{R}$



This is a similar to the normal distribution but has a closed-form CDF and is computationally easier.

For example: $P(-2 < X < 2) = F(2) - F(-2) = \frac{e^2-1}{1+e^2} = .76$

The Uniform distribution

Definition (Uniform distribution): *A continuous r.v. U has a Uniform distribution on the interval (a, b) if its PDF is*

$$f(x; a, b) = \frac{1}{b-a} I_{(a,b)}(x)$$

We denote this by $U \sim Unif(a, b)$

We often write a density in this manner- using the indicator variable to describe its support. This is a valid PDF since the area under the curve is the area of a rectangle with length $(b-a)$ and height $\frac{1}{(b-a)}$. The CDF is the accumulated area under the PDF:

$$F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x \leq b \\ 1 & \text{if } x \geq b \end{cases}$$

Graph the PDF and CDF of $Unif(0, 1)$.

Suppose we have observations from $Unif(0, 1)$. We can transform these into a sample from $\tilde{U} \sim Unif(a, b)$:

$$\tilde{U} = a + (b-a)U$$

The Probability Integral Transformation

Result: *Let* $\mathbf{X}$ *be a continuous random variable with the distribution function* $\mathbf{F}$ *and let* $\mathbf{Y} = \mathbf{F}(\mathbf{X})$. *Then* $\mathbf{Y}$ *must be uniformly distributed on* $[\mathbf{0}, \mathbf{1}]$. *The transformation from* $\mathbf{X}$ *to* $\mathbf{Y}$ *is called the probability integral transformation.*

We know that the distribution function must take values between 0 and 1. If we pick any of these values, y, the y^{th} quantile of the distribution of X will be given by some number x and

$$Pr(Y \leq y) = Pr(X \leq x) = F(x) = y$$

which is the distribution function of a uniform random variable.

This result helps us generate random numbers from various distributions, because it allows us to transform a sample from a uniform distribution into a sample from some other distribution provided we can find F^{-1}.

Example: Given $U \sim \text{Unif}(0,1)$, $\log\left(\frac{U}{1-U}\right) \sim$ Logistic.

($F(x) = \frac{e^x}{1+e^x}$, set this equal to u and solve for x)

The Normal distribution

This symmetric bell-shaped density is widely used because:

1. **It captures many types of natural variation quite well: heights-humans, animals and plants, weights, strength of physical materials, the distance from the centre of a target.**

2. **It has nice mathematical properties: many functions of a set normally distributed random variables have distributions that take simple forms.**

3. **Central limit theorems are fundamental to statistic inference. The sample mean of a large random sample from *any* distribution with finite variance is approximately normal.**

The Standard Normal distribution

Definition (Standard Normal distribution): *A continuous r.v. Z is said to have a standard Normal distribution if its PDF φ is given by:*

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

The CDF is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)\,dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$E(Z) = 0, Var(Z) = 1$ *and* $X = \mu + \sigma Z$ *is has a Normal distribution with mean* μ *and variance* σ^2.

($E(\mu + \sigma Z) = E(\mu) + \sigma E(Z) = \mu$ and $Var(\mu + \sigma Z) = \sigma^2 Var(Z) = \sigma^2$)

Approximations of Normal probabilities

If $\mathbf{X} \sim \mathbf{N}(\mu, \sigma^2)$ *then*

- $\mathbf{P}(|\mathbf{X} - \mu| < \sigma) \approx 0.68$
- $\mathbf{P}(|\mathbf{X} - \mu| < 2\sigma) \approx 0.95$
- $\mathbf{P}(|\mathbf{X} - \mu| < 3\sigma) \approx 0.997$

Or stated in terms of z*:*

- $\mathbf{P}(|\mathbf{Z}| < 1) \approx 0.68$
- $\mathbf{P}(|\mathbf{Z}| < 2) \approx 0.95$
- $\mathbf{P}(|\mathbf{Z}| < 3) \approx 0.997$

The Normal distribution therefore has very thin tails relative to other commonly used distributions that are symmetric about their mean (logistic, Student's-t, Cauchy).

The Gamma distribution

The gamma function of α is defined as

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \tag{1}$$

Definition (Gamma distribution): *It is said that a random variable X has a Gamma distribution with parameters α and β ($\alpha > 0$ and $\beta > 0$) if X has a continuous distribution for which the PDF is:*

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{(0,\infty)}(x)$$

To see that this is a valid PDF: Substitute for y in the expression $\Gamma(\alpha)$ by $y = \beta x$ (so $dy = \beta dx$)

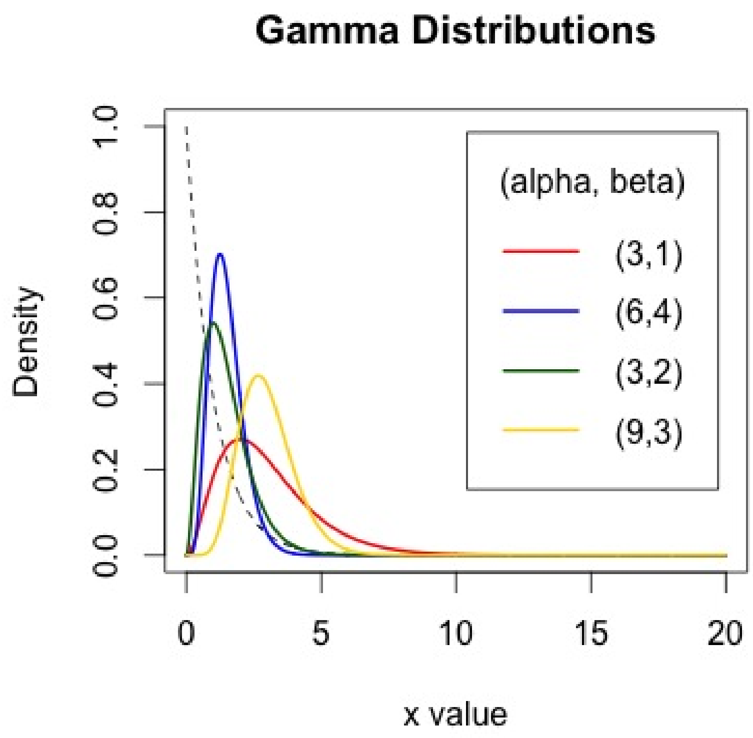
$$\Gamma(\alpha) = \int_0^\infty (x\beta)^{\alpha-1} e^{-x\beta} \beta dx$$

or

$$1 = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx$$

$$E(X) = \frac{\alpha}{\beta} \text{ and } Var(X) = \frac{\alpha}{\beta^2}$$

Features of the Gamma density



The density is decreasing for $\alpha \leq 1$, and hump-shaped and skewed to the right otherwise. It attains a maximum at $x = \frac{\alpha-1}{\beta}$

Higher values of α make it more bell-shaped and higher β increases the concentration at lower x values.

The Exponential distribution

We have a special name for a Gamma distribution with $\alpha = 1$:

Definition (Exponential distribution): *A continuous random variable* X *has an exponential distribution with parameter* β *(* $\beta > 0$*) if its PDF is:*

$$f(x;\beta) = \beta e^{-\beta x} I_{(0,\infty)}(x)$$

We denote this by $X \sim Expo(\beta)$*. We can integrate this to verify that for* $x > 0$*, the corresponding CDF is*

$$F(x) = 1 - e^{-\beta x}$$

The exponential distribution is memoryless:

$$P(X \geq s+t | X \geq s) = \frac{X \geq s+t}{X \geq s} = \frac{e^{-\beta(s+t)}}{e^{-\beta s}} = P(X \geq t)$$

The χ^2 distribution

Definition (χ^2 distribution): *The* χ^2_ν *distribution is the* **Gamma**$(\frac{\nu}{2}, \frac{1}{2})$ *distribution. Its PDF is therefore*

$$f(x;\nu) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} I_{(0,\infty)}(x)$$

Notice that for $\nu = 2$, the χ^2 density is equivalent to the exponential density with $\beta = \frac{1}{2}$. It is therefore decreasing for this value of ν and hump-shaped for other higher values.

The χ^2 is especially useful in problems of statistical inference because if we have ν independent random variables, $X_i \sim N(0,1)$, the sum $\sum\limits_{i=1}^{\nu} X_i^2 \sim \chi^2_\nu$ Many of the estimators we use in our models fit this case (i.e. they can be expressed as the sum of independent normal variables)

Gamma applications

Survival analysis:

The risk of failure at any point t is given by the **hazard rate**,

$$h(t) = \frac{f(t)}{S(t)}$$

where $S(t)$ is the **survival function**, $1 - F(t)$. The exponential is memoryless and so, if failure hasn't occurred, the object is as good as new and the hazard rate is constant at β. For **wear-out** effects, $\alpha > 1$ and for **work-hardening** effects, $\alpha < 1$

Relationship to the Poisson distribution:

If Y, the number of events in a given time period t has a poisson density with parameter λ, the rate of success is $\gamma = \frac{\lambda}{t}$ and the time taken to the first breakdown $X \sim \text{Gamma}(1, \gamma)$.

Example: A bottling plant breaks down, on average, twice every four weeks, so $\lambda = 2$ and the breakdown rate $\gamma = \frac{1}{2}$ per week. The probability of less than three breakdowns in 4 weeks is $P(X \leq 3) = \sum\limits_{i=0}^{3} e^{-2}\frac{2^i}{i!} = .135 + .271 + .271 + .18 = .857$ Suppose we wanted the probability of no breakdown? The time taken until the first break-down, x must therefore be more than four weeks. This $P(X \geq 4) = \int\limits_4^\infty \frac{1}{2}e^{-\frac{x}{2}}dx = e^{-\frac{x}{2}}\Big|_4^\infty = e^{-2} = .135$

Income distributions that are uni-modal