Elements of Probability and Statistics

Lecture 09: Continuous Probability Distributions

IAI, TCG-CREST | September 15, 19 & 22, 2023

9.1 Continuous Distributions

For discrete distributions, we observed that non-zero probabilities according to the PMF existed only for the supports. Therefore the CDF was a discontinuous function that had jumps at each points in the support.

For continuous distributions, the CDF is defined to be any continuous and differentiable function, with the exception of a finite number of points allowed where the CDF is continuous but not differentiable.

Any random variable with such a continuous CDF is defined to be a continuous random variable.

Probability Density Function (PDF): The PDF of a continuous random variable X with CDF F is defined as the derivative of the CDF $f(x) = F'(x)$. The support of the random variable X is the set of all x where $f(x) > 0$.

The PDF $f(x)$ is not a probability. To obtain a probability $P(X \leq x)$, an integral of the PDF is calculated.

Prop. From PDF to CDF: For a continuous random variable X with PDF f, the probability $P(X \leq x)$, which is the CDF, is given by,

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt \ .$$

The probability at any point is zero, i.e., $P(X = x) = 0 \ \forall x$. This also implies that when calculating probabilities, the endpoints do not necessarily need to be specified. Thus, $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$. For any such interval,

$$P(a < X < b) = F(b) - F(a) = \int_{a}^{b} f(x)dx \ .$$

In general for any arbitrary region $A \subseteq \mathbb{R}$,

$$P(X \in A) = \int_{A} f(x)dx \ .$$

Defn. The PDF of a continuous random variable f is a valid PDF if it satisfies two conditions:

1. Non-negativity: $f(x) \geq 0$.
2. Integrates to one: $\int_{-\infty}^{\infty} f(x)dx = 1$.

The above definition implies that the CDF is non-decreasing.

Ex. 1. Logistic Distribution: The CDF of a Logistic Distribution is,

$$F(x) = \frac{e^x}{1+e^x}, \ x \in \mathbb{R} \ .$$

The PDF can be obtained by differentiating the CDF,

$$f(x) = \frac{e^x}{(1+e^x)^2}, \ x \in \mathbb{R} \ .$$

Let $X \sim$ Logistic. Then the probability of an interval such as $P(-2 < X < 2)$ can be calculated in one of two ways. The first is using the CDF:

$$P(-2 < X < 2) = \int_{-2}^{2} \frac{e^x}{(1+e^x)^2} dx = F(2) - F(-2) \approx 0.76 \ .$$

The second approach is from the integration of the PDF. By substituting $u = 1 + e^x$ so that $du = e^x dx$,

$$\int_{-2}^{2} \frac{e^x}{(1+e^x)^2} dx = \int_{1+e^{-2}}^{1+e^2} \frac{1}{u^2} du = \left(-\frac{1}{u}\right)\Bigg|_{1+e^{-2}}^{1+e^2} \approx 0.76 \ .$$

Ex. 2. Rayleigh Distribution: The CDF of a Rayleigh distribution is,

$$F(x) = 1 - e^{-x^2/2}, \ x > 0 \ .$$

The PDF is,

$$f(x) = xe^{-x^2/2}, \ x > 0 \ .$$

Let $X \sim$ Rayleigh. Then the probability of the interval $P(X > 2)$ is,

$$P(X > 2) = \int_{2}^{\infty} xe^{-x^2/2} dx = 1 - F(2) \approx 0.14 \ .$$

Since the PDF is the derivative of the CDF, which is $P(a < X < b)$, the PDF can be interpreted as a function whose value in an interval is proportional to the probability of X in that interval. The value of the PDF function dictates the change in the CDF function. Over a small ϵ interval around a point a, if we assume the PDF is contant, then the probability is,

$$P(a - \epsilon/2 < X < a + \epsilon/2) = \int_{a-\epsilon/2}^{a+\epsilon/2} f(x)dx \approx f(a)\epsilon \ .$$

This is approximately the change in the CDF that occurs around point a.

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

The expectation is said to exist only if the integral is a finite value, i.e., if it does not diverge.

When working with any function of a random variable g, the expected value of the function is $E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$.

9.2 Uniform Distribution

A uniform distribution is defined as one that has a constant PDF over an interval.

If $U \sim \text{Unif}(a, b)$, then its PDF is,

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

The CDF is,

$$F(x) = \begin{cases} 0 & \text{if } x \le a \ , \\ \frac{x-a}{b-a} & \text{if } a < x < b \ , \\ 1 & \text{if } x \ge b \ . \end{cases}$$

A special distribution from the family of Uniform Distributions is $\text{Unif}(0, 1)$. From the CDF, we see that for $X \sim \text{Unif}(0, 1)$, $P(X < x) = x$.

Thm. The expectation and variance of $U \sim \text{Unif}(a, b)$:

$$E(U) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{a+b}{2} \ .$$

Similarly, the variance can be calculated to be $Var(U) = \frac{(b-a)^2}{12}$.

Another way of calculating the expectation and variance is by using location-scale transformation.

Defn. Location-scale Transformation: For a random variable X, the random variable $Y = \sigma X + \mu$ with $\sigma > 0$ is said to be a location-scale transformation of X.

When any transformation is applied on a random variable, the support of the transformed random variable must be checked to verify whether a distribution of the same family has been reached. If $X \sim \text{Unif}(a, b)$, then $Y = cX + d$ with $c > 0$ is still a uniform distribution as $Y \sim \text{Unif}(ca + d, cb + d)$. An example of a transformation that will not be uniform in general is $Z = X^2$. Similarly, for discrete distributions like $\text{Bin}(n, p)$ a location-scale transformation will not necessarily preserve the support of the family of distribution, e.g., for $X \sim \text{Bin}(n, p)$, $Y = X + 5$ or $Z = 2X$ will not be in the family of $\text{Bin}(n, p)$.

Starting from a simpler $U \sim \text{Unif}(0, 1)$, we can find the expectation and variance, and use location-scale transformation to find it for the general uniform distribution. For $U \sim \text{Unif}(0, 1)$,

$$\begin{aligned} E(U) &= \int_0^1 x dx = \frac{1}{2}, \\ E(U^2) &= \int_0^1 x^2 dx = \frac{1}{3}, \text{ and so} \\ Var(U) &= \frac{1}{12} \ . \end{aligned}$$

Now for a general $X \sim \text{Unif}(a, b)$, as $X = a + (b - a)U$,

$$E(X) = E(a + (b - a)U) = \frac{a+b}{2}, \text{ and}$$
$$Var(X) = Var(a + (b - a)U) = \frac{(b-a)^2}{12} \ .$$

Defn. Universality of Uniform: Let F be a CDF which is strictly increasing on the support of the distribution, and hence F^{-1} exists as a function from $(0, 1)$ to $\mathbb{R}$. Then the following are true.

(i) Let $U \sim \text{Unif}(0, 1)$ and $X = F^{-1}(U)$. Then X is a random variable with CDF F.

(ii) Let X be a random variable with CDF F. Then $F(X) \sim \text{Unif}(0, 1)$.

Note that there is a difference in the meanings of the following notations: For X with CDF F we write $F(x) = P(X \leq x)$, whereas $F(X)$ is $F(X(s))\ \forall s \in S$.

Proof. (i) For all $x \in \mathbb{R}$,

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

(ii) Let $Y = F(X)$. As Y takes values in $(0, 1)$, the CDF $P(Y \leq y) = 0$ for $y \leq 0$, and $P(Y \leq y) = 1$ for $y \geq 1$. For $y \in (0, 1)$,

$$P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y \ .$$

Therefore $Y \sim \text{Unif}(0, 1)$.

Ex. 1. The Logistic distribution has CDF $F(x) = \frac{e^x}{1+e^x}$, the inverse of which is $F^{-1}(u) = \log\left(\frac{u}{1-u}\right)$. The CDF of $F^{-1}(U)$ is,

$$P(\log\left(\frac{U}{1-U}\right) \leq x) = P(\frac{U}{1-U} \leq e^x) = P(U \leq \frac{e^x}{1+e^x}) = \frac{e^x}{1+e^x} \ .$$

Hence $F^{-1}(U) \sim$ Logistic. Similarly one can show for Rayleigh distribution as well.

Point (i) of the Universality also holds for discrete distributions. For a discrete random variable X with PMFs $p_0, p_1, ...$ at $j = 0, 1, ...$, the interval $(0, 1)$ is divided into countable sections of length $p_0, p_1, ...$ Then the probability of $X = j$ becomes equal to the interval length p_j.

Point (ii) of the Universality however does not hold for discrete distributions, as $F(X)$ will always be a discrete distribution.

Ex. The concept of the universality of uniform distributions is applied in the calculation of percentiles. As an example, we can consider that the marks (between 0 and 100) of several students are collected, and an assumption is made that the distribution of marks is continuous. Let the CDF of the marks distribution be F. Then $F^{-1}(U)$ provides a way to obtain quantiles or percentiles. The median marks, for which $P(\text{marks} < \text{median})$ covers half the area of the distribution can be found as $F^{-1}(0.5)$. Similarly the 0.90 percentile marks 'y' can be calculated from $F^{-1}(0.9)$, for which $P(\text{marks} < y)$ covers 90% of the area of the distribution.

Defn. For a random variable X with CDF F, the function $G(x) = 1 - F(x) = P(X > x)$

is called the survival function.

Thm. The expectation of a non-negative random variable X can be found by integrating its survival function.

$$E(X) = \int_0^\infty P(X > x)dx \ .$$

Proof. For any $x \geq 0$,

$$x = \int_0^x dt = \int_0^\infty I(x > t)dt,$$

where $I(x > t)$ is 1 if $x \geq t$, and 0 otherwise. As $X(s) = \int_0^\infty I(X(s) > t)dt \ \forall s \in S$, i.e., $X = \int_0^\infty I(X > t)$, we can consider the expectation on both sides to get,

$$E(X) = E(\int_0^\infty I(X > t)dt) = \int_0^\infty E(I(X > t))dt = \int_0^\infty P(X > t)dt \ .$$

This theorem is true for discrete random variables as well.

9.3 Normal Distributions

Defn. Standard Normal Distribution: A continuous random variable $Z \sim \mathcal{N}(0, 1)$ has the following PDF:

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

The standard normal CDF does not have a closed-form expression:

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

Some properties of the standard normal distribution:

(i) Symmetry of the PDF: $\varphi(z) = \varphi(-z)$, as $\varphi(.)$ is an even function.

(ii) Symmetry of the tail areas: $\Phi(z) = 1 - \Phi(-z)$. Proof:

$$\Phi(-z) = \int_{-\infty}^{-z} \varphi(t)dt = \int_{z}^{\infty} \varphi(u)du = 1 - \int_{-\infty}^{z} \varphi(u)du = 1 - \Phi(z).$$

The above proof uses substitution $u = -t$, and the fact that PDFs integrate to 1.

(iii) Symmetry of Z and $-Z$: If $Z \sim \mathcal{N}(0, 1)$, then $-Z \sim \mathcal{N}(0, 1)$.

This is due to: $P(-Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z) = \Phi(z)$.

To show that the standard normal PDF $\varphi(z)$ is a valid PDF, we can observe that it is non-negative for all z, and in order to show that it sums to 1, we need to show that $\int_{-infty}^{\infty} e^{-z^2/2}dz = \sqrt{2\pi}$. This can be shown as:

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-z^2/2}dz\right)\left(\int_{-\infty}^{\infty} e^{-z^2/2}dz\right) &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}dy\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} rdrd\theta\end{aligned}$$

The last step involved conversion to polar coordinates. Using substitutions of (i) $u =$

$r^2/2$, $du = rdr$, we continue the above to get,

$$\int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta = \int_0^{2\pi}\left(\int_0^{\infty} e^{-u} du\right) d\theta$$
$$= \int_0^{2\pi} 1 d\theta = 2\pi.$$

And therefore,

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}.$$

Defn. The expectation of $Z \sim \mathcal{N}(0,1)$ is,

$$E(Z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0.$$

As the function inside the integral is an odd function. The same argument lets us state that for all odd positive integers n, $E(Z^n) = 0$.

Defn. Calculating the variance of $Z \sim \mathcal{N}(0,1)$:

$$Var(Z) = E(Z^2) - [E(Z)]^2 = E(Z^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} z^2 e^{-z^2/2} dz$$

the last step is due to the function in the integral being an even function. Now using integration by parts with $u = z$, $du = dz$ and $v = -e^{-z^2/2}$, $dv = ze^{-z^2/2}dz$,

$$Var(Z) = \frac{2}{\sqrt{2\pi}}\left(-ze^{-z^2/2}\Big|_0^{\infty} + \int_0^{\infty} e^{-z^2/2} dz\right) = \frac{2}{\sqrt{2\pi}}\left(0 + \frac{\sqrt{2\pi}}{2}\right) = 1.$$

Defn. Normal Distribution: If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma Z$ is said to follow the Normal distribution with mean μ and variance σ^2, denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$.

The expected value and variance can be verified:

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

Defn. Standardization: From any $X \sim \mathcal{N}(\mu, \sigma^2)$, we can obtain the standard normal distribution as,

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)\ .$$

Defn. The Normal distribution CDF:

$$F(x) = P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)\ .$$

The Normal Distribution PDF:

$$f(x) = \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\ .$$

Defn. The 68-95-99.7% Rule for Normal distributions: Quick approximations of Normal probabilities can be found using this rule, which says,

$$\begin{aligned} P(|X-\mu|<\sigma) &= P(|Z|<1) \approx 0.68 \ , \\ P(|X-\mu|<2\sigma) &= P(|Z|<2) \approx 0.95 \ , \\ P(|X-\mu|<3\sigma) &= P(|Z|<3) \approx 0.997 \ . \end{aligned}$$

Q. Let $X \sim \mathcal{N}(-1, 4)$, approximate the value of $P(|X| < 3)$.

9.4 Exponential Distribution

A random variable $X \sim \text{Expo}(\lambda)$, with $\lambda > 0$, has PDF:

$$f(x) = \lambda e^{-\lambda x}, x > 0.$$

The CDF is:

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

If λ is interpreted as a rate of success per unit time, so that the average number of successes in a time interval t is λt. Then an interpretation of the Exponential distribution is that it is the distribution of the time spent until the arrival of the first success.

This interpretation is related to that of the Geometric distribution, and it can actually be shown that the Geometric distribution tends to the Exponential distribution under certain limiting conditions.

Defn. If $X \sim \text{Expo}(1)$, then $Y = \frac{X}{\lambda} \sim \text{Expo}(\lambda)$. As,

$$P(Y \leq y) = P\left(\frac{X}{\lambda} \leq y\right) = P(X \leq \lambda y) = 1 - e^{-\lambda y}.$$

The converse is true, i.e., if $Y \sim \text{Expo}(\lambda)$, then $\lambda Y \sim \text{Expo}(1)$.

Using integration by parts, the expectation and variance for $X \sim \text{Expo}(1)$ can be calculated,

$$\begin{aligned} E(X) &= \int_0^\infty x e^{-x} dx = 1, \\ E(X^2) &= \int_0^\infty x^2 e^{-x} dx = 2, \\ Var(X) &= 1. \end{aligned}$$

By scale transformation $Y = \frac{X}{\lambda}$, the expectation and variance for $Y \sim \text{Expo}(\lambda)$ becomes,

$$\begin{aligned} E(Y) &= \frac{1}{\lambda} E(X) = \frac{1}{\lambda}, \\ Var(Y) &= \frac{1}{\lambda^2} Var(X) = \frac{1}{\lambda^2} \ . \end{aligned}$$

Defn. Memoryless Property: A continuous distribution is said to have a memoryless property if a random variable X from that distribution satisfies,

$$P(X \geq s+t | X \geq s) = P(X \geq t) \quad \forall s, t \geq 0 \ .$$

Thm. If X is a positive continuous random variable with the memoryless property, then X has an Exponential distribution.

Proof: We assume X is a positive continuous random variable with the memoryless property, whose CDF is F, and whose survival function is $G = 1 - F$. Now due to the memoryless property,

$$G(s+t) = G(s)G(t), \ \forall s, t \geq 0.$$

With $s = t$, we get, $G(2t) = G(t)^2$. Similarly we get $G(3t) = G(t)^3$, $G(4t) = G(t)^4$, and so on. Therefore, $G(mt) = G(t)^m$ for any positive integer m.

In a similar way, substituting t with $t/2$, $G(t/2) = G(t)^{1/2}$. For any positive integer n, we get $G(t/n) = G(t)^{1/n}$.

Combining, we get,

$$G\left(\frac{m}{n}t\right) = G\left(\frac{t}{n}\right)^m = G(t)^{m/n} \ .$$

As this holds for positive integers m, n, for any rational x the following is true,

$$G(xt) = G(t)^x \ .$$

Considering any positive real number to be a limiting value of a positive rational number, the above holds for any real x. Now considering $t = 1$, we get, $G(x) = G(1)^x$. The function that satisfies this is $G(x) = e^{-\lambda x}$, with $\lambda = -\log(G(1)) > 0$.

Thus X has an Exponential distribution.

Among discrete distributions, the Geometric distribution is the only memoryless distribution.

Defn. Stochastic Process: A stochastic process is a collection of random variables, indexed by what is usually interpreted as time.

Discrete time stochastic process are described as $X = \{X_t, t = 0, 1, 2, ...\}$ where there are countable number of random variables, indexed by non-negative integers.

Continuous time stochastic process are described as $X = \{X_t, 0 \leq t \leq \infty\}$, where there are uncountable number of random variables indexed by non-negative reals.

Defn. Poisson Process: A process of 'arrivals' in continuous time is called a Poisson process with rate λ, if the following two conditions hold:

(i) The number of arrivals that occur in an interval of length t is distributed as Pois(λt).

(ii) The number of arrivals that occur in disjoint intervals are independent of each other.

For a Poisson process, we can define the following two random variables: (i) T_1: the time until the first arrival, and (ii) N_t: the number of arrivals at or before time t. Then the event $T_1 > t$ is the same as the event $N_t = 0$. This *count-time duality* establishes a connection between a continuous random variable and a discrete random variable. In general, the event $T_n > t$ is the same as the event $N_t < n$.

For the event $T_1 > t$ or $N_t = 0$, the probability is,

$$P(T_1 > t) = P(N_t = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Thus $P(T_1 \leq t) = 1 - e^{-\lambda t}$, and therefore $T_1 \sim \text{Expo}(\lambda)$.

***Q.** Relation between Geometric and Exponential distributions: Assume that Bernoulli trials are performed in continuous time at regularly spaced times: 0, Δt, $2\Delta t$,... , where Δt is a small positive number. Let the probability of success of each trial be $\lambda\Delta t$, where λ is a positive constant. Let G be the number of failures before the first success (in discrete time), and T be the time of the first success (in continuous time).

As $\Delta t \to 0$, the CDF of T converges to the Expo(λ) CDF.

Avisek Gupta
Postdoctoral Fellow
IAI, TCG CREST.