Continuous Random Variables

Terminology

Informally, a random variable X is called **continuous** if its values x form a "continuum", with $P(X = x) = 0$ for each x. By contrast, a discrete random variable is one that has a finite or countably infinite set of possible values x with $P(X = x) > 0$ for each of these values. Analogous terminology is applied to the distributions associated with such random variables: a **discrete distribution** is one associated with a discrete random variable, and a **continuous distribution** is one associated with a continuous random variable.

General concepts and formulas for continuous random variables

- **Cumulative distribution function (c.d.f.):**
  - **Notation and definition:** $F(x) = P(X \leq x)$.
  - **Properties:** (1) $F(x)$ is non-decreasing, (2) $0 \leq F(x) \leq 1$, (3) $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to +\infty$.
- **Probability density function (p.d.f.):**
  - **Notation and definition:** $f(x) = F'(x)$ (There is no **direct** definition of $f(x)$ in terms of X, without the detour via the c.d.f. $F(x)$; in particular, $f(x)$ is **not** equal $P(X = x)$ as in the discrete case.)
  - **Properties:** (1) $f(x) \geq 0$, (2) $\int f(x)dx = 1$ (the integral being over the "range" of f).
  - **Relations between $f(x)$ and $F(x)$:** $f(x)$ is the derivate of $F(x)$; $F(x)$ is an anti-derivative (i.e., integral) of $f(x)$.
- **Probability computations via p.d.f.'s and c.d.f.'s:** $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$
- **Median and percentiles:** The median of a distribution is an x-value at which $F(x) = 0.5$, i.e., such that $X \leq x$ with 50% probability. The median is also called the 50-th percentile. More general "percentiles" are defined analogously: for example, the 90-th percentile is an x-value such that $F(x) = 0.9$ i.e., such that $X \leq x$ with 90% probability; it is denoted by $\pi_{0.9}$.
- **Mode of a distribution:** A point x where the density $f(x)$ is maximal.
- **Expectation, variance, and moment-generating function:** These are defined in the same way as for discrete random variables, with sums replaced by integrals: $E(X) = \int xf(x)dx$, $\mathrm{Var}(X) = E(X^2) - E(X)^2$, $M(t) = E(e^{tX}) = \int e^{tx}f(x)dx$, etc., where the integrals are extended over the range of $f(x)$.

Notes and tips

- **Note that the definition of a density function, $f(x)$, for a continuous random variable is completely different from that for a discrete random variable:** In the discrete case, $f(x)$ is defined as $f(x) = P(X = x)$ and called a probability mass function (p.m.f.). In the continuous case, $f(x)$ is called a probability density function (p.d.f.), but it is no longer given by $P(X = x)$ (the latter probability is always 0 for continuous X); it can only be obtained indirectly as the derivative of the c.d.f. $F(x)$.
- **Keep track of the range of a density:** Make it a habit to always write down the range of a density $f(x)$ along with the formula (i.e., the interval of x on which the function is defined and non-zero), and keep this range in mind when trying to determine the correct integration limits. For example, if a problem asks for $P(X \geq 2)$ and the density $f(x)$ "lives" on the interval $[1, 4]$, the integral giving $P(X \geq 2)$ should be from 2 to 4; if, instead, $f(x)$ were defined and non-zero on the infinite interval $[1, \infty)$, the integral would be from 2 to infinity.
- **For probabilities involving continuous random variables, "$\leq$" and "$<$", and similarly "$\geq$" and "$>$", are interchangeable:** For example, $P(X \leq 2)$ is the same as $P(X < 2)$, since the difference between these two probabilities, $P(X = 2)$, is 0 for a continuous random variable. By the same token, when specifying the range of a density functions, it doesn't matter whether or not one includes the boundary points in the range. For example, in the definition "$f(x) = 2x$ for $0 < x < 1$" one could replace the range "$0 < x < 1$" by "$0 \leq x \leq 1$" without affecting any probability calculations involving this density.

Important continuous distributions

The following is a list of essential formulas for continuous distributions which you should memorize. The most important such distribution is the exponential distribution. Another, very important distribution, the normal distribution, will be covered separately later.

1. **Exponential distribution:**
   - **P.d.f.:** $f(x) = \frac{1}{\theta} e^{-x/\theta}$ $(0 \leq x < \infty)$
   - **C.d.f.:** $F(x) = 1 - e^{-x/\theta}$ $(0 \leq x < \infty)$
   - **Tail formula for exponential distribution:** $P(X > x) = e^{-x/\theta}$ $(0 \leq x < \infty)$
   - **Expectation:** $\mu = \theta$
   - **Variance:** $\sigma^2 = \theta^2$
2. **Uniform distribution on interval $[a, b]$:**
   - **P.d.f.:** $f(x) = 1/(b - a)$ $(a \leq x \leq b)$.
   - **Expectation:** $\mu = (a + b)/2$ (i.e., the midpoint of the interval $[a, b]$)
   - **Variance:** $\sigma^2 = (b - a)^2/12$