Notes: Continuous Random Variables

CS 3130/ECE 3530: Probability and Statistics for Engineers

September 25, 2014

Review: A random variable on a sample space Ω is just a function $X : \Omega \to \mathbb{R}$.

So far, our sample spaces have all been discrete sets, and thus the output of our random variables have been restricted to discrete values. What if the sample space is continuous, such as $\Omega = \mathbb{R}$? This means that the output of a random variable $X : \Omega \to \mathbb{R}$ could possibly take on a continuum of values.

Example: Let's say we record the time elapsed from the start of class to when the last person arrives. This is a continuous random variable T that takes values from 0 to 80 minutes. What is the probability that $T = 5$? Well, if the precision of my watch only goes up to minutes, then I might find the last person arrives during the fifth minute. But if my precision is in seconds, it is less likely that the last person arrives exactly 5 minutes late. Then if I have a stopwatch that goes to hundredths of a second, it seems almost impossible that the last person will come in at 5 minutes on the dot. As the precision of our measurements get better and better, the probability goes down. If we were able to measure at infinite precision, the probability $P(T = 5)$ would be zero! However, the probability that the last person arrives *between* 5 and 6 minutes late is nonzero. In other words, $P(5 \leq T \leq 6)$ is not zero.

Probability density functions:
A **probability density function (pdf)** for a continuous random variable X is a function f that describes the probability of events $\{a \leq X \leq b\}$ using integration:

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

Due to the rules of probability, a pdf must satisfy $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

Continuous random variables also have cdfs. The cdf is given by

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\, dx$$

Important! A pdf is not a probability! Only when you integrate it does it give you a probability. In fact, you can have pdf's where the value $f(x)$ is greater than one for some x. (We'll see some examples below).

Exercise: Are the following functions pdf's? If so, what are the cdf's?

$$f(x) = \begin{cases} \frac{1}{2} - 2x^2 & \text{if } x \in [-1, 1] \\ 0 & \text{elsewhere} \end{cases}$$

Answer: No, $f(1) = -\frac{3}{2}$, and a pdf must always be non-negative.

$$f(x) = \begin{cases} \sin(x) & \text{if } x \in [\frac{\pi}{2}, \pi] \\ 0 & \text{elsewhere} \end{cases}$$

Answer: Yes, here we have $f(x) \geq 0$, and we can verify that it integrates to one over the range it is non-zero:

$$\begin{aligned} \int_{\pi/2}^{\pi} f(x)dx &= \int_{\pi/2}^{\pi} \sin(x)dx \\ &= -\cos(x)\Big|_{x=\pi/2}^{x=\pi} \\ &= -\cos(\pi) + \cos(\pi/2) = 1 \end{aligned}$$

Uniform Distribution:
The uniform distribution is the continuous equivalent of "equally likely" outcomes that we had in the discrete case. The pdf for a uniform random variable on the interval $[\alpha, \beta]$ is

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

If X is a such a random variable, we write $X \sim U(\alpha, \beta)$. Notice that if $\beta - \alpha < 1$, then $f(x)$ will be *greater* than one in $[\alpha, \beta]$.

Exercise: What is the cdf for $U(\alpha, \beta)$? Verify that the density for $U(\alpha, \beta)$ integrates to one.

Answer: The cdf for $U(\alpha, \beta)$ for $\alpha \leq a \leq \beta$ is given by

$$\begin{aligned} F(a) &= \int_{\alpha}^{a} \frac{1}{\beta - \alpha} dx \\ &= \frac{x}{\beta - \alpha}\Big|_{x=\alpha}^{x=a} \\ &= \frac{a - \alpha}{\beta - \alpha} \end{aligned}$$

We can verify this integrates to one by plugging in $a = \beta$, giving $F(1) = 1$. For $a < \alpha$, $F(a) = 0$, and for $a > \beta$, $F(b) = 1$.

Exercise: Is the following function a valid cdf, and if so, what is its associated pdf?

$$F(a) = \begin{cases} \exp(a) & \text{if } a \leq 0, \\ 1 & \text{otherwise.} \end{cases}$$

Answer: Yes, we can verify that $\lim_{a \to -\infty} F(a) = 0$ and that $F(0) = 1$. The pdf for this random variable comes by taking the derivative:

$$f(x) = \frac{dF(a)}{da}\Big|_{a=x} = \begin{cases} \exp(x) & \text{if } x \leq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Exponential Distribution:
pdf: $f(x) = \lambda e^{-\lambda x}$
cdf: $F(a) = 1 - e^{-\lambda a}$
Notation: $X \sim Exp(\lambda)$
What it's good for: Exponential distributions model events that occur randomly in time at some specified rate (the rate is the λ parameter). For example, we might want to model the arrival times of people coming to class, or radioactive decay (release of an ionizing particle from a radioactive material), or the time you will receive your next email. These are all "random processes" that can be modeled as exponential distributions with some rate λ.

The exponential distribution is a continuous limit of the geometric distribution (see book example also). If students are arriving to class at a rate of λ, let T be the random variable representing the time the next student arrives. What is $P(T > t)$, i.e., the probability the next student comes in beyond time t? If we break the interval $[0, t]$ into n discrete chunks of size t/n, then next arrival can be modeled as a geometric distribution with probability of arriving in a particular chunk of $p = \frac{\lambda t}{n}$. So, using the geometric distribution, $P(T > t)$ is approximately $(1 - p)^n$. We get a better approximation by increasing n, i.e., by breaking the interval into smaller chunks. In the limit we have:

$$P(T > t) = \lim_{n \to \infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}$$

So, $P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t}$, which is the cdf for the exponential distribution.

Exercise: For the exponential distribution, $Exp(\lambda)$, what is a value for λ and x that makes $f(x) > 1$?

Gaussian or Normal Distribution:
pdf: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
cdf: No explicit equation! $F(a) = \int_{-\infty}^{a} f(x)\, dx$
Notation: $X \sim N(\mu, \sigma^2)$, μ is the **mean**, and σ^2 is the **variance**
What it's good for: The Gaussian distribution is the probably the most important probability distribution in all of science. We will learn later in class that it explains randomness that comes from the addition of lots of small random disturbances. Therefore, the Gaussian distribution is widely used to model complex phenomena. The noise in many types of physical measurements is modeled as Gaussian distributed. For example, the Gaussian distribution can be argued to be a good model for "thermal noise", where random particle motions affect measurements from a sensors.

Pareto Distribution:
pdf: $f(x) = \frac{\alpha}{x^{\alpha+1}}$ for $x \geq 1$ and $f(x) = 0$ for $x < 1$.
cdf: $F(a) = 1 - \frac{1}{a^\alpha}$ for $a \geq 1$ and $F(a) = 0$ for $a < 1$.
Notation: $X \sim Par(\alpha)$
What it's good for: The Pareto distribution was originally used to model wealth distribution in economics. It models how a large portion of a society's wealth is owned by a small percentage of the population. In general, it is good for anything where probabilities are greater for some outcomes and then fall off for others (see Wikipedia for many other uses of Pareto). Computer science applications include: file sizes in internet traffic (lots of small files, fewer large ones), hard drive error lengths (lots of small errors, fewer large errors).

Quantiles:

The p**th quantile** of a random variable X is the smallest number q_p such that $F(q_p) = P(X \leq q_p) = p$. Another way to say this is $q_p = F^{-1}(p)$. The 50% quantile is called the **median**.

Notice that we can compute probabilities in an interval using the quantile function:

$$P(a \leq X \leq b) = P(\{X \leq b\} - \{X < a\}) = P(X \leq b) - P(X < a) = F^{-1}(b) - F^{-1}(a)$$

Exercise: What is the median of the Gaussian distribution $N(\mu, \sigma^2)$?

Example on board: Quantiles of the normal distribution.