Probability and Statistics

Alvin Lin

Probability and Statistics: January 2017 - May 2017

Continuous Random Variables

Let X be a continuous random variable. The probability distribution or probability density function (pdf) of X is a function $f(x)$ such that for any two numbers a and b with $a \leq b$.

$$P(a \leq X \leq b) = \int_a^b f(x)\ \mathrm{d}x$$

For a function $f(x)$ to be a valid pdf, the following conditions must begin satisfied:

1. $f(x) \geq 0$ for any $x \in (-\infty, \infty)$.
2. $\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = 1$

A continuous random variable X is said to have a uniform distribution on the interval $[A, B]$ if the pdf of X is:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & , A \leq x \leq B \\ 0 & \text{, otherwise} \end{cases}$$

Example

The direction of an imperfection with respect to a reference line on a circular object such as tires, brake rotors, etc. is in general subject to uncertainty. Consider the reference line connected the valve stem on a tire rim to the center point. Let X be the angle measured clockwise to the location of the imperfection. One possible pdf for X is:

$$f(x) = \begin{cases} \frac{1}{360} & , 0 \leq x < 360 \\ 0 & , otherwise \end{cases}$$

Find the probability that there is an imperfection between the 45° and the 90° region on the rim.

$$\begin{aligned}
P(45 \leq X \leq 90) &= \int_{45}^{90} \frac{1}{360} \, \mathrm{d}y \\
&= \frac{1}{360} [y]_{y=45}^{y=90} \\
&= \frac{1}{360} (90 - 45) \\
&= \frac{45}{360} \\
&= \frac{1}{8}
\end{aligned}$$

Cumulative Distribution Function

The cumulative distribution function $F(x)$ for a continuous random variable x is defined for every number by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(s) \, \mathrm{d}s$$

Proposition

Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Then for any number a:

$$\begin{aligned}
P(X > a) &= 1 - P(X \leq a) \\
&= 1 - \int_{-\infty}^{a} f(s) \, \mathrm{d}s \\
&= 1 - F(a)
\end{aligned}$$

For any numbers a and b with $a \leq b$:

$$\begin{aligned}
P(a \leq X \leq b) &= \int_{a}^{b} f(s) \, \mathrm{d}s \\
&= \int_{-\infty}^{a} f(s) \, \mathrm{d}s + \int_{a}^{b} f(s) \, \mathrm{d}s - \int_{-\infty}^{a} f(s) \, \mathrm{d}s \\
&= \int_{-\infty}^{b} f(s) \, \mathrm{d}s - \int_{-\infty}^{a} f(s) \, \mathrm{d}s \\
&= F(b) - F(a)
\end{aligned}$$

Example

Let p be a number between 0 and 1. The $(100p)$th percentile of the distribution of a continuous random variable X, denoted by $\eta(p)$, is defined by:

$$p = P(x \leq \eta(p)) = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(x)\ \mathrm{d}x$$

Using the previous tire rim example, find $\eta(0.9)$, i.e. the 90th percentile.

$$\begin{aligned}
p &= P(x \leq \eta(p)) \\
0.9 &= P(X \leq \eta(0.9)) \\
&= F(\eta(0.9)) \\
&= \int_{-\infty}^{\eta(0.9)} f(x)\ \mathrm{d}x \\
&= \int_{0}^{\eta(0.9)} \frac{1}{360}\ \mathrm{d}x \\
&= \frac{1}{360}[x]_{x=0}^{x=\eta(0.9)} \\
&= \frac{1}{360}[\eta(0.9) - 0] \\
&= \frac{\eta(0.0)}{360} \\
\eta(0.9) &= (0.9)(360)
\end{aligned}$$

The median of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile and also has special importance.

$$\tilde{\mu} = \eta(0.5)$$

$$0.5 = F(\eta(0.5)) = F(\tilde{\mu}) = \int_{-\infty}^{\tilde{\mu}} f(x)\ \mathrm{d}x$$

Expected Values and Variance

The expected values and variance of a continouous random variable are similar to those of a discrete random variable. Instead of $\sum$, we use $\int$.
Expected value of X:

$$E(X) = \int_{-\infty}^{\infty} xf(x)\ \mathrm{d}x$$

where $f(x)$ is the pdf of X.

Expected value of $Y = h(X)$:

$$E(Y) = E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) \, \mathrm{d}x$$

Variance of X:

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) \, \mathrm{d}x$$

Standard Deviation of X:

$$\sigma_x = \sqrt{V(X)}$$

Example

Using the previous tire rim example:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} f(x) \, \mathrm{d}x \\
&= \int_{0}^{360} x \frac{1}{360} \, \mathrm{d}x \\
&= \frac{1}{360} \frac{1}{2} [x^2]_{x=0}^{x=360} \\
&= \frac{1}{2} \frac{1}{360} [360^2 - 0^2] \\
&= \frac{1}{2} \frac{1}{360} 360^2 \\
&= \frac{1}{2} 360 = 180 \\
V(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} (x - 180)^2 f(x) \, \mathrm{d}x \\
&= \int_{0}^{360} (x - 180)^2 f(x) \, \mathrm{d}x \\
&= \ldots
\end{aligned}$$

We can also use the proposition that $V(X) = E(X^2) - [E(X)]^2$:

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\ \mathrm{d}x \\
&= \int_{0}^{360} x^2 f(x)\ \mathrm{d}x \\
&= \int_{0}^{360} x^2 \frac{1}{360}\ \mathrm{d}x \\
&= \frac{1}{360} \int_{0}^{360} x^2\ \mathrm{d}x \\
&= \frac{1}{360}\frac{1}{3}[x^3]_{x=0}^{x-360} \\
&= \frac{1}{3}\frac{1}{360}[360^3 - 0^3] \\
&= \frac{1}{3}360^2 \\
V(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{360^2}{3} - 180^2 \\
&= \frac{1}{3}2^2 180^2 - 180^2 \\
&= 180^2[\frac{4}{3} - 1] \\
&= \frac{180^2}{3} \\
\sigma_x &= \sqrt{V(X)} \\
&= \frac{180}{\sqrt{3}} \\
&\approx 104
\end{aligned}
$$

Discrete vs Continuous Median

Given discrete quiz scores 6, 7, and 9, the median score is 7. For continuous values, the generation property of median $\tilde{\mu}$ is:

$$\tilde{\mu} = \eta(0.5)$$

A continuous random variable X is said to have a normal distribution with parameters μ and σ (or μ and σ_2), where:

$$-\infty < \mu < \infty$$

$$\mu \in (-\infty, \infty)$$

$$\mu \in \mathbb{R}$$

$$\mu \text{ is a real number}$$

and $\sigma > 0$, if the pdf of X is:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu)^2/(2\sigma)^2}$$

for any $x \in \mathbb{R}$.

You can find all my notes at `http://omgimanerd.tech/notes`. If you have any questions, comments, or concerns, please contact me at alvin@omgimanerd.tech