Probability and Statistics

Alvin Lin

Probability and Statistics: January 2017 - May 2017

Discrete Random Variables

A discrete random variable is a random variable whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on (a countably infinite set). A countable set is finite or countably infinite.

- Finite sets:

$$\left\{1\right\}$$

$$\left\{A, B, E\right\}$$

- Countably infinite sets:

$$\left\{2, 4, 6, 8, \dots\right\}$$

$$\left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\}$$

$$\left\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\right\}$$

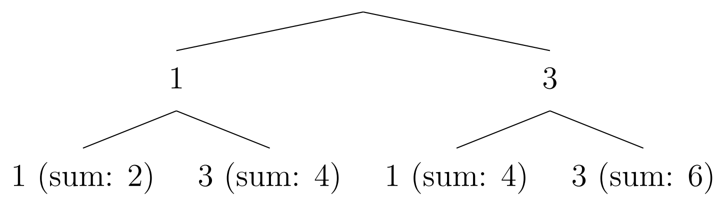
$$\left\{(m, n) \in \mathbb{Z} \times \mathbb{Z} \mid m \geq 0, n \geq 0\right\}$$

A random variable is continuous if both of the following apply:

1. Its set of possible values consists either of all members in a single interval on the number line, possibly infinite in extent (eg: $(-\infty, \infty), (0, \infty), \dots$) or all numbers in a disjoint union of such intervals.

Probability Mass Function

There are two balls marked 1 and 3, respectively. Select a ball twice with replacement.



What is the probability that the sum is 4?

$$\frac{2}{4} = \frac{1}{2}$$

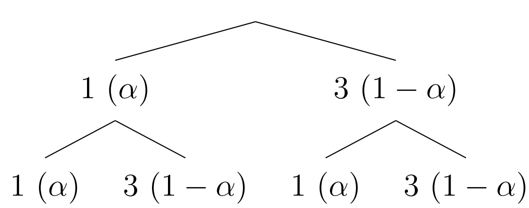
Let X be the random variable indicating the value of the sum of the two numbers.

$$\begin{aligned} p(4) &= P(X = 4) \\ &= P(\{\omega \in S \mid X(\omega) = 4\}) \\ &= P(\{(1,3),(3,1)\}) \\ &= \frac{2}{4} \end{aligned}$$

$p(X = 4)$ is the probability mass function (pmf), a function that gives the probability that a random variable is equal to some exact number.

Example

Using the same problem as before, the selections are independent and the probability that we get a 1 in a selection is α.



$$
\begin{aligned}
p(2;\alpha) &= P(X=2) \\
&= \alpha^2 \\
p(4;\alpha) &= P(X=4) \\
&= \alpha(1-\alpha) + (1\alpha)\alpha \\
&= 2\alpha(1-\alpha) \\
p(6;\alpha) &= (1-\alpha)^2
\end{aligned}
$$

α can take various values. The collection of $p(x;\alpha)$

$$(p(x;\frac{1}{2}), p(x;\frac{1}{6}), \dots)$$

is a family of probability distributions.

Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ of a discrete random variable X with pmf $p(x)$ is defined for every number x by:

$$F(x) = P(X \leq x) = \sum_{y;y\leq x} p(y)$$

Example

Using the same problem as before:

$$\begin{aligned}
F(2) &= P(X \leq 2) \\
&= \frac{1}{4} \\
&= \sum_{y \leq 2} p(y) \\
&= p(2) \\
&= \frac{1}{4} \\
F(4) &= P(X \leq 4) \\
&= \sum_{y \leq 4} p(y) \\
&= p(2) + p(4) \\
&= \frac{1}{4} + \frac{2}{4} \\
&= \frac{3}{4} \\
F(6) &= P(X \leq 6) \\
&= \sum_{y \leq 6} p(y) \\
&= p(2) + p(4) + p(6) \\
&= \frac{1}{4} + \frac{2}{4} + \frac{1}{4} \\
&= 1
\end{aligned}$$

$$
\begin{aligned}
P(4 \leq X \leq 6) &= \frac{3}{4} \\
&= F(6) - F(2) \\
&= (\frac{1}{4} + \frac{2}{4} + \frac{1}{4}) - (\frac{1}{4}) \\
&= \frac{3}{4} \\
P(2 \leq X \leq 4) &= \frac{3}{4} \\
&= F(4) = P(X \leq 4) \\
&= \frac{3}{4}
\end{aligned}
$$

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

$$P(a \leq X \leq b) = F(b) - F(a-)$$

where $a-$ represents the largest possible X value that is strictly less than a. Notation:

$$
\begin{array}{rl}
\textit{probability} & P(\textit{event} \in S) \\
\textit{pmf of a random variable} & p(\textit{a number}) \\
\textit{cdf of a random variable} & F(\textit{a number})
\end{array}
$$

Expected Values

Let X be a discrete random variable with set of possible values D and pmf $p(x)$. The expected value or mean value of X is:

$$E(X) = \mu_x = \mu = \sum_{x \in D} xp(x)$$

If the random variable X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$ is:

$$E[h(x)] = \sum_{x \in D} h(x)p(x)$$

If $h(X)$ is of very special type $h(X) = aX + b$, where a and b are constants:

$$E[h(x)] = E(aX + b) = \mu_{ax+b} = aE(x) + b$$

Example

Extending from the same problem:

$$
\begin{aligned}
E(X) &= \mu_x \\
&= \sum_{x \in D} x p(x) \\
&= 2p(2) + 4p(4) + 6p(6) \\
&= 2\frac{1}{4} + 4\frac{2}{4} + 6\frac{1}{4} \\
&= \frac{2+8+6}{4} \\
&= 4 \\
E(\mathrm{e}^X) &= \sum_{x \in D} \mathrm{e}^x p(x) \\
&= \mathrm{e}^2 p(2) + \mathrm{e}^4 p(4) + \mathrm{e}^6 p(6) \\
&= \mathrm{e}^2 \frac{1}{4} + \mathrm{e}^4 \frac{2}{4} + \mathrm{e}^6 \frac{1}{4} \\
&= \frac{1}{4}[e^2 + 2e^4 + \mathrm{e}^6] \\
E(10X - 20) &= (10 \times 2 - 20)p(2) + (10 \times 4 - 20)p(4) + (10 \times 6 - 20)p(6) \\
&= 10E(X) - 20 \\
&= 10 \times 4 - 20 \\
&= 20
\end{aligned}
$$

Variance and Standard Deviation

Let X have pmf $p(x)$ and expected value μ. The variance of X is:

$$V(X) = \sigma_x^2 = \sigma^2 = \sum_{x \in D} (x - \mu)^2 p(x) = E((x - \mu)^2)$$

The standard deviation of X is:

$$\sigma_x = \sigma = \sqrt{V(X)} = \sqrt{\sigma_x^2}$$

Example

Extending from the same problem:

$$\begin{aligned} V(X) &= (2-4)^2 p(2) + (4-4)^2 p(4) + (6-4)^2 p(6) \\ &= 2^2 \frac{1}{4} + 0^2 \frac{2}{4} + 2^2 \frac{1}{4} \\ &= 2 \\ V(10X-20) &= 10^2 V(X) \\ &= 100 \times 2 \\ &= 200 \end{aligned}$$

If you have any questions, comments, or concerns, please contact me at
alvin@omgimanerd.tech