Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors David Karger and Nancy Lynch

Handout 46
May 2, 1998

# Wald's Theorem

## 1 Theorem Statement

Wald's Theorem concerns the expected sum of a random number of random variables. For example, suppose that I flip a coin. If I get heads, then I roll two dice. If I get tails, then I roll three dice. What is the expected sum of the dice that I roll? Wald's Theorem supplies a simple answer.

**Theorem 1 (Wald's Theorem).** *Let $Q$ be a random variable with range $\mathbb{N}$, and let $R, R_1, R_2, R_3, \ldots$ be random variables with countable ranges such that*

$$\Pr(R_i = x \mid Q \geq i) \quad = \quad \Pr(R = x)$$

*for all $i \geq 1$ and all $x$ in the range of $R$. Then*

$$\text{Ex}(R_1 + R_2 + \ldots + R_Q) \quad = \quad \text{Ex}(R) \cdot \text{Ex}(Q)$$

*provided that the expectations on the right exist and are finite.*

The first equation in the theorem statement describes a tricky condition. Informally, this equation requires every random variable $R_i$ appearing in the sum $R_1 + R_2 + \ldots + R_Q$ to have the same distribution as a "reference" random variable $R$. Note, however, that the distribution of a random variable $R_i$ can be arbitrary in the event that $R_i$ does not appear in the sum; that is, in the event that $Q < i$.

The coin-and-dice problem is solved easily using Wald's Theorem. Based on the outcome of a coin toss, I roll either two or three dice. Thus, we define $Q$ so that $\Pr(Q = 2) = \Pr(Q = 3) = \frac{1}{2}$, and we define $R, R_1, R_2, \ldots$ to take on values in the range 1 to 6 uniformly. The sum of the dice that I roll is equal to $R_1 + \ldots + R_Q$. By Wald's Theorem, this sum is equal to:

$$\begin{aligned}
\text{Ex}(R_1 + \ldots + R_Q) &= \text{Ex}(R) \cdot \text{Ex}(Q) \\
&= \frac{7}{2} \cdot \frac{5}{2} \\
&= \frac{35}{4}
\end{aligned}$$

## 2 Proof Preliminaries

The proof of Wald's Theorem requires two facts, which we state without proof. The first fact generalizes linearity of expectation to the case of an infinite sum.

**Fact 1.** *Let $R_1, R_2, R_3, \ldots$ be random variables with countable ranges. If the summation*

$$\sum_{i=1}^{\infty} |R_i|$$

*converges for all points in the sample space, then*

$$\text{Ex}\left(\sum_{i=1}^{\infty} R_i\right) = \sum_{i=1}^{\infty} \text{Ex}(R_i).$$

The second fact that we will need is a statement about *conditional expectation.* The conditional expectation of a random variable $R$ given an event $E$ is denoted by $\text{Ex}(R \mid E)$ and is defined as follows.

$$\text{Ex}(R \mid E) = \sum_{x \in \text{Range}(R)} x \cdot \Pr(R = x \mid E)$$

The fact that we need about conditional expectation is stated below.

**Fact 2.** *Let $R$ be a random variable, and let $E$ be an event. Then*

$$\text{Ex}(R) = \text{Ex}(R \mid E) \cdot \Pr(E) + \text{Ex}(R \mid \overline{E}) \cdot \Pr(\overline{E})$$

*provided that all the expectations exist and are finite.*

## 3 Proof of Wald's Theorem

We want to prove that

$$\text{Ex}(R_1 + R_2 + \ldots + R_Q) = \text{Ex}(R) \cdot \text{Ex}(Q).$$

The sum $R_1 + R_2 + \ldots + R_Q$ is awkward, because the final subscript $Q$ is a random variable. We eliminate this awkwardness by introducing some indicator random variables. In particular, let $I_i$ be an indicator for the event that $R_i$ appears in the sum $R_1 + R_2 + \ldots + R_Q$; that is, $I_i$ is the event that $Q \geq i$. With this definition, we can rewrite the sum as follows.

$$R_1 + R_2 + \ldots + R_Q \quad = \quad \sum_{i=1}^{\infty} R_i \cdot I_i$$

In effect, the indicator variables $I_i$ "mask out" random variables $R_i$ that do not appear in the sum $R_1 + R_2 + \ldots + R_Q$. The proof of Wald's Theorem begins by taking the expected value of both sides of this equation.

$$\text{Ex}(R_1 + R_2 + \ldots + R_Q) \quad = \quad \text{Ex}\left(\sum_{i=1}^{\infty} R_i \cdot I_i\right)$$

According to Fact 1, we can apply linearity of expectation to break up this infinite sum provided that the sum

$$\sum_{i=1}^{\infty} |R_i \cdot I_i|$$

always converges. At each point in the sample space, the random variable $Q$ takes on some value in $\mathbb{N}$. Thus, at each point in the sample space, only finitely many of the indicator variables $I_i$ are non-zero. Consequently, for each point in the sample space, the sum contains of only finitely many non-zero terms, and so the sum always converges. Thus, we can apply linearity of expectation and continue reasoning as follows.

$$\begin{aligned}
\text{Ex}(R_1 + R_2 + \ldots + R_Q) &= \sum_{i=1}^{\infty} \text{Ex}\left(R_i \cdot I_i\right) \\
&= \sum_{i=1}^{\infty} (\text{Ex}\left(R_i \cdot I_i \mid I_i = 1\right) \cdot \Pr(I_i = 1) + \\
&\qquad \text{Ex}\left(R_i \cdot I_i \mid I_i = 0\right) \cdot \Pr(I_i = 0)) \\
&= \sum_{i=1}^{\infty} \text{Ex}\left(R_i \mid I_i = 1\right) \cdot \Pr(I_i = 1)
\end{aligned}$$

The second equation uses Fact 2, and the third equation follows by simpifying. The next step is to simplify the conditional expectation in the last expression above. We break out and simplify this term below and then return to the main line of the argument.

$$\begin{aligned}
\text{Ex}\left(R_i \mid I_i = 1\right) &= \sum_x x \cdot \Pr(R_i = x \mid I_i = 1) \\
&= \sum_x x \cdot \Pr(R_i = x \mid Q \geq i) \\
&= \sum_x x \cdot \Pr(R = x) \\
&= \text{Ex}(R)
\end{aligned}$$

The first step uses the definition of expectation, the second step uses the definition of the indicator variable $I_i$, the third step uses the technical assumption in the theorem statement, and the final step uses the definition of expectation again. All summations are over $x$ in the range of $R_i$.

Using this result, we can wrap up the proof of Wald's Theorem.

$$\begin{aligned}
\operatorname{Ex}(R_1 + R_2 + \ldots + R_Q) &= \sum_{i=1}^{\infty} \operatorname{Ex}(R) \cdot \Pr(I_i = 1) \\
&= \operatorname{Ex}(R) \cdot \sum_{i=1}^{\infty} \Pr(I_i = 1) \\
&= \operatorname{Ex}(R) \cdot \sum_{i=1}^{\infty} \Pr(Q \geq i) \\
&= \operatorname{Ex}(R) \cdot \operatorname{Ex}(Q)
\end{aligned}$$

The constant $\operatorname{Ex}(R)$ is pulled out in the second equation. The third step uses the definition of the indicator variable $I_i$, and the final step uses an identity for the expecatation of a random variable with range $\mathbb{N}$.