Wald's Identity

Implications for psychometric function,chronometric function, and Weber's law

M. N. Shadlen,
June 2003
revised June 2004

Introduction and road map

This tutorial is an attempt to supply the background for understanding a class of models known as diffusion – or random walk – to barrier. The model has been applied to a variety of decision problems that arise in quality control and psychology. Wald contributed the sequential probability ratio test, in which noisy evidence accumulates (or meanders) toward one of a pair of criteria. Stone and Laming were the first to apply this idea to perceptual decisions, and Roger Ratcliff is probably the most active modern champion. I was motivated to write this tutorial after reading Stephen Link's papers and book. Link championed the idea that such a model can account for a wide variety of observations in sensory psychology: speed-accuracy tradeoff, Weber's law, most of what we already know under the theory of signal detection, Fechner's and Stevens' laws concerning magnitude estimation, and so on. Link's ideas have much going for them. The object of this tutorial is not to reiterate what is in his book. Rather it is to provide a firm background in the mathematical foundations. Late in the tutorial, I spend some time deriving Weber's law and elaborating Link's argument. But the main thrust of the tutorial is on Wald's identity and especially the background in stochastic processes that is required. My real motivation in writing this tutorial is to bring my colleagues and students up to speed so that we can all work on the problem that fascinates me most right now: decisions among N>2 alternatives and the incorporation of prior probability and reward expectation in the decision process.

The organization of the tutorial is under revision. The main goal is to understand the derivation of Wald's identity and to recognize it as a martingale. To make sense of this, we must acquaint ourselves with the definition of martingales. To make sense of Wald's identity and use it, we must be comfortable with the moment generating function. Both of these topics fall outside the background of many neuroscientists. So we walk through this very slowly, building intuitions along the way. I take a long time to accomplish what Karlin & Taylor do in about two pages. I am also taking a more direct assault on Wald's identity. K and T get to Wald's identity by developing a general "trick" for making martingales using the eigenvectors (or eigenfunctions) of markov transition matrices (or functions). I only develop these concepts in the particular setting of the random walk. My approach is extremetly intuitive, but it is a bit repetitious. If you don't want the long ride, see pp. 242-243 of K and T. I mention this to preempt any concern that this seems like a hard way to go about proving Wald's identity. Not only is this simple (once you get the hang of it), but it provides some deep understanding that lets you march right through response times!

In parts 2 and 3, we use Wald's identity to derive the psychometric function (accuracy) and the chronometric function (decision time), respectively. This turns out to be a piece of cake after part 1. In the final sections, we explore some related topics. I plan to incorporate new sections on the PMF under assymetric barriers and the CMF for each type of choice.

1. Wald's Identity

Notation and preliminaries

The outline goes as follows. First we state Wald' s identity as a martingale. It turns out that proving Wald' s identity can boil down to proving that a special sequence of random numbers is a martingale. The proof comes in two parts. The first part is the more difficult and also the more general. It is to see that eigenfunctions of the transition probability functions of a Markov chain can be used to create martingales. The second part requires showing that $e^{\theta x}$ is just such an eigenfunction for the Markov process of partial sums, $x = Y_1 + \ldots + Y_n$.

I have adopted the usual convention of writing a moment generating function (mgf) as $M[\theta] = E[e^{\theta x}]$, where E is expectation, x is a random variable whose distribution posesses the mgf. Link puts a negative sign in the exponent. This leads to a nicer set of functions, but it is not the standard way to write the mgf. I will try to be clear in comparing the derived expressions to Link' s. It is usually a matter of a sign change on θ.

1.1 What is a martingale?

A martingale is a stochastic process, meaning a sequence of random numbers, that defines a fair game. Suppose you have a process that gives the random sequence,

$Y_0, Y_1, \ldots, Y_n$.

Apply some function to these values to obtain a new sequence,

$X_0, X_1, \ldots, X_n$.

In principle, you can use the whole history of $Y_0 \ldots Y_n$ to make each X_n, but for our purposes, we will only work with functions that ignore history (i.e., Markov processes):

$X_0 = f[Y_0], X_1 = f[Y_1], \ldots, X_n = f[Y_n]$

We say that X is a martingale with respect to the stochastic process that gave rise to Y if and only if

$$\text{for all } n, \quad |E[X_n]| < \infty$$

$$E[X_{n+1} \mid Y_0, Y_1, \ldots, Y_n] = X_n \tag{1}$$

In words, whatever value the sequence of X's has yielded up to the nth term, the expectation of the next term is just what has been attained to that point.

- **Example**

For example, if the Y are iid random vars with mean 0, then if the new process describing X_n is the sum of the Y_0 ... Y_n , then X is a martingale. The sum meanders about in the form of a random walk or diffusion, but whatever the sum has become on the nth addition is the expectation of the process on the next step. This would not be the case if the E[Y]≠0. But we could make our function be

$$X_n = \sum_{i=1}^{n} Y_i - n\,E[Y] \tag{2}$$

to discount the affect of adding the mean of Y at every step. We'll exploit this "undoing the effect of the incrementing operation" in a moment to construct Wald's martingale.

There is a convenient property of martingales that we will exploit. Since the expectations do not change, it turns out that if you stop the process based on some test, including an evaluation of the Y or the sum of Y_i , the expectation of X_n does not change.

1.2 What is a moment generating function?

Moment generating functions are essentially Laplace transforms of probability density functions. They have several interesting properties. Two are essential for the topic at hand. The first is well known and has to do with sums of random variables. The second is less well known. It concerns a special point where a moment generating function passes through the value 1. We need to develop intuitions about this special value, and to do that we need to have a feel for what gives a moment generating function its shape. Let's start with basics.

Consider a random variable, x, with probability density function f[x]. The moment generating function is a new function, that is an argument of a new variable -- like a Laplace or Fourier transform -- that we'll call θ. θ is not a random number. It's the argument to the function. The moment generating function transforms the random variable, x, and returns the average value. It is

$$M[\theta] == E\{e^{\theta x}\} \tag{3}$$

The expectation is over all possible values of the rv, x. Of course, we know how to calculate the expectation of $e^{\theta x}$. It's the weighted sum over all possible values that x can take. The weights are given by f(x). In other words, it's the average value of $e^{\theta x}$.

$$M[\theta] == E\{e^{\theta x}\} = \int_{-\infty}^{\infty} e^{\theta x} f[x]\, dx \tag{4}$$

Check point. This way of writing an expectation should be obvious and intuitive for the same reason that the expected value of x (i.e., the mean) is

$$\text{mean}[x] == E\{x\} = \int_{-\infty}^{\infty} x f[x]\, dx \tag{5}$$

The mean is also known as the first moment. Not surprisingly, the second moment is

$$2\,\mathrm{nd\;moment} == \mathrm{E}\{x^2\} = \int_{-\infty}^{\infty} x^2\, f[x]\, dx \tag{6}$$

That's a reminder about how to calculate the expectation. It also reveals why E{$e^{\theta\, x}$} is called a moment generating function.

▪ 1.2.1 The MGF can be used to calculate moments

Take the derivative of M[θ] with respect to θ

$$\begin{aligned} M'[\theta] &= \partial_\theta \left(\int_{-\infty}^{\infty} e^{\theta\, x}\, f[x]\, dx\right) \\ &= \int_{-\infty}^{\infty} x e^{\theta\, x}\, f[x]\, dx \end{aligned} \tag{7}$$

Null

When $\theta = 0$, the exponential goes to 1, leaving an expression for the 1st moment, or mean

$$M'[0] = \int_{-\infty}^{\infty} x\, f[x]\, dx \tag{8}$$

If we take the 2nd derivative, and evaluate at $\theta = 0$, we get the 2nd moment

$$\begin{aligned} M''[\theta] &= \partial_\theta \left(\int_{-\infty}^{\infty} x\, e^{\theta\, x}\, f[x]\, dx\right) \\ &= \int_{-\infty}^{\infty} x^2\, e^{\theta\, x}\, f[x]\, dx \end{aligned} \tag{9}$$

$$M'[0] = \int_{-\infty}^{\infty} x^2\, f[x]\, dx$$

and so on.

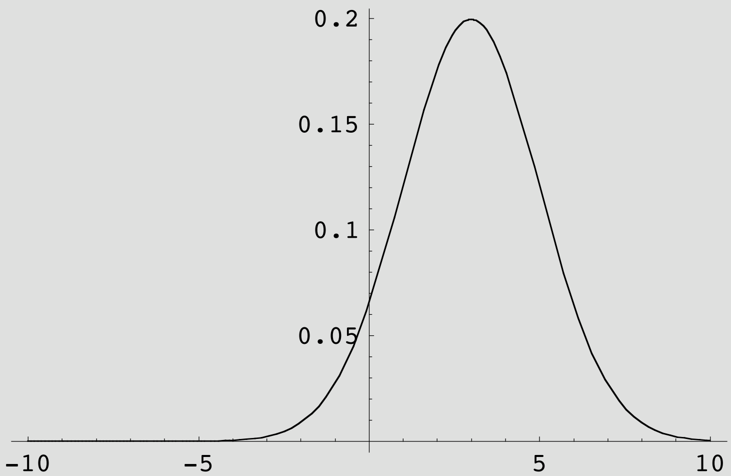
▪ 1.2.2 An example

Let's look at the normal distribution with mean = μ, variance = σ^2 along with its moment generating function. The normal PDF is

$$\mathbf{npdf[x\_,\ \mu\_,\ \sigma\_]\ :=\ \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}}$$

Choose particular values for the mean and standard deviation

```
μ = 3;
σ = 2;
Plot[npdf[x, μ, σ], {x, -10, 10}]
```




```
- Graphics -
```

Ensure that the pdf has area 1

```
Integrate[npdf[x, μ, s], {x, −Infinity, Infinity}, Assumptions → s > 0]

1
```

Moment generating function is expectation of $e^{\theta x}$

```
Integrate[Exp[θ x] npdf[x, m, s], {x, −Infinity, Infinity}, Assumptions → s > 0]
```

$$e^{m\,\theta + \frac{s^2\,\theta^2}{2}}$$

At this point, I highly recommend running through the matlab tutorial momentGeneratingFuncTutorial.m It lacks the symbolic math, but it provides essential intuitions. The main thing we want to accomplish is to get an understanding for the shape of the MGF and in particular the value of θ where it equals 1. We call this special root of the MGF θ_1 . In the next bit, we'll use matlab to derive θ_1 for the normal distribution.

■ Show θ_1 for the normal distribution

```
npdf[x_, μ_, σ_] := E^(-(x-μ)^2/(2 σ^2)) / (Sqrt[2 π] σ)
```

$$\text{npdf}[x\_, \mu\_, \sigma\_] := \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$$

Ensure that the pdf has area 1

```
Integrate[npdf[x, m, s], {x, -Infinity, Infinity}, Assumptions → s > 0]

1
```

Moment generating function is expectation of $e^{\theta x}$

```
Integrate[Exp[θ x] npdf[x, m, s], {x, -Infinity, Infinity}, Assumptions → s > 0]
```

$$e^{m\,\theta+\frac{s^2\,\theta^2}{2}}$$

Set this equal to 1 or take log and set equal to 0. The nonzero root is θ_1

$$\text{Solve}\left[m\,\theta+\frac{s^2\,\theta^2}{2}==0,\ \theta\right]$$

$$\left\{\{\theta\to 0\},\left\{\theta\to-\frac{2\,m}{s^2}\right\}\right\}$$

(Note to JP: This is the bit that think that mean/var is the invariant of interest in these models.)

TO BE CONTINUED

■ 1.3 Wald's identity and Wald's martingale

Suppose $Y_0 = 0$ and $Y_1, Y_2, \ldots$ are independent identically distributed (iid) random variables with associated PDF, g[Y], and moment generating function, $M[\theta]$

$$\begin{aligned} M[\theta] &= E[e^{\theta Y}] \\ &= \int_{-\infty}^{\infty} e^{\theta Y}\, g[Y]\, dY \end{aligned} \tag{10}$$

I dropped the subscript on the Y because the Y are identically distributed. Generate a new list of random variables as follows.

$$W_0 = 1$$
$$W_n = M[\theta]^{-n}\, e^{\theta(Y_0 + \ldots + Y_n)} \tag{11}$$

We can also write the sum of the random variables as S_n.

$$W_n = M[\theta]^{-n}\, e^{\theta S_n} \tag{12}$$

Remember that S_n is also a sequence of random numbers. It is a random walk,
and we refer to it as a markov process of partial sums. It' s a markov process because $E[S_{n+1}]$ depends only on the values of S_n and we say " partial sums " to remind ourselves that it is the sum up to n terms. In the next section, we will show that the sequence, W_0, W_1, ..., W_n is a martingale and that the expectation of W_n is 1 for all n. Both statements are expressed in the following equation.

$$E[W_{n+1} \mid S_0, S_1, \ldots, S_n] = W_n = 1 \tag{13}$$

Wald's identity is usually written in terms of Equation 5.

$$E\left[\frac{e^{\theta S_n}}{M[\theta]^n}\right] = 1 \tag{14}$$

■ 1.4 Derivation part 1

The process we wish to study is the random walk formed by accumulating ranom numbers. We will refer to the sum after each step n by S_n.

$$S_0 = 0$$
$$S_1 = Y_1$$
$$S_2 = Y_1 + Y_2 \tag{15}$$
$$S_n = Y_1 + Y_2 + \ldots + Y_n$$

Notice that the S_n form a sequence of random numbers. As mentioned above, it is a Markov process.

We assume that the Y_i are independent and that each is described by a probability density function, g[Y], and moment generating function $M_Y[\theta]$. Throughout this tutorial, we will write $M[\theta]$ without a subscript to designate the MGF of the random variable, Y. It' s important to keep track in your mind that the Y are the increments in the steps that comprise the random walk.

■ Now generate a new stochastic process based on the S_i.

$$X_n = f[S_n] = e^{\theta S_n} \tag{16}$$

where θ is any constant. We' ll think about θ as a variable later on. But for now,
we can imagine that θ is a particular value. So really, what we' re generating is not a sequence of random numbers but a family of such sequences depending on what we fill in for θ. Don' t get hung up on this. It it bothers you, just assume $\theta = 1$, but don' t drop it from the expressions. Let' s look at our new stochastic process

$$X_0 = 1 \text{ because } S_0 = 0$$

$$X_1 = e^{\theta\, Y_1} \text{ because } S_1 = Y_1$$

$$X_2 = e^{\theta(Y_1 + Y_2)} \text{ because } S_2 = Y_1 + Y_2 \tag{17}$$

$$X_n = e^{\theta(Y_1 + Y_2 + ... + Y_n)} \text{ because } S_n = Y_1 + Y_2 + ... + Y_n$$

In short, the X_i comprise another stochastic process that we have spawned using the function, f. Suppose we have observed such a process to the n^{th} step.That means that X_n is known but X_{n+1} is yet to be determined. What is the expectation of X_{n+1}, conditional on knowledge of X_n

$$E\{X_{n+1} \mid X_n\} = E\{f[S_{n+1}] \mid S_n\} \tag{18}$$

Using the exponential function, f, this is

$$E\{X_{n+1} \mid X_n\} = E\{e^{\theta S_{n+1}} \mid S_n\} \tag{19}$$

The conditional here reminds us that S_n is known. The random variable S_{n+1} can be decomposed into the part that is known and a random increment

$$= E\{e^{\theta(S_n + Y_{n+1})}\} \tag{20}$$

Take the nonrandom part out of the expectation.

$$= e^{\theta S_n}\, E\{e^{\theta\, Y}\} \tag{21}$$

I dropped the subscript on Y because the random increments are always drawn from the same kind of distribution. They are iid.Notice that the expectation in this expression is the moment generating function, $M[\theta]$, associated with the random variable, Y.

$$E\{f[S_{n+1}] \mid S_n\} = M[\theta]\, f[S_n] \tag{22}$$

Remember,for any given value of θ,the moment generating function,$M[\theta]$ is just a number.To make this more transparent, I'm going to substitute the constant λ for $M[\theta]$.This is just to make a point.It allows us to see something special about the function, f,that we are using to generate our random sequence.

$$E\{f[S_{n+1}]\} = \lambda\, f[S_n] \tag{23}$$

When we write things this way, we can understand why f is called an eigenfunction and λ (i.e., $M[\theta]$) an eigenvalue for the stochastic process (markov process of partial sums). Remember, we know S_n. The function, f, has this cool property that if you feed it the state of the stochastic process at step n, it spits out a number that, upon multiplication by λ, is the expectation of the stochastic process at the next step.It' s probably obvious where we' re heading with all this. With our eigenfunction $f[S_n] = e^{\theta S_n}$ in hand, we can create a martingale by undoing the multiplication by λ that occurs with each step.

Let' s break this down by looking at the expectations for a few steps (n). Start before we take the first step. What' s the expectation of $f[S_0]$?

$$E\{e^{\theta S_0}\} = 1 \text{ because } S_0 = 0 \quad (24)$$

That was easy. Let's look at step n=1

$$E\{e^{\theta S_1}\} = M[\theta] \text{ because } S_1 = Y_1\text{. We might want to write this as} \\ E\{e^{\theta S_1}\} = \lambda \quad (25)$$

On to step n = 2. What is $E\{e^{\theta S_2}\}$?

Recall that $E\{f[S_2] \mid S_1\} = \lambda f[S_1]$

That' s fine, but it assumes we know S_1. To get the expectation of $f[S_2]$, we need to average over all possible values for S_1.Let' s write this by taking a second expectation over all the possible S_1 .I' ll indicate this by writing a subscript on the E for expectation.

$$\begin{aligned} E\{e^{\theta S_2}\} &= \mathbf{E}_{S_1}\left\{E\{f[S_2] \mid S_1\}\right\} \\ &= \mathbf{E}_{S_1}\{\lambda f[S_1]\} \\ &= \lambda E\{f[S_1]\} \\ &= \lambda^2 \end{aligned} \quad (26)$$

That last step works because the expectation is just equation (18).Not to beat it to death, but before we get general, let's do one more step (n=3).

$$\begin{aligned} E\{e^{\theta S_3}\} &= \mathbf{E}_{S_2}\left\{E\{f[S_3] \mid S_2\}\right\} \\ &= \mathbf{E}_{S_2}\{\lambda f[S_2]\} \\ &= \lambda E\{f[S_2]\} \\ &= \lambda^3 \end{aligned} \quad (27)$$

Notice how we are exploiting the markov property. The random variable at step n depends only on th e value it attained at step n − 1. Obviously, our eigenfunction, f, is playing akey role here. Andremember, λ is really the moment generating function evaluated at some value, θ.We may as well write things out for the n^{th} step.

$$\begin{aligned} E\{e^{\theta S_n}\} &= \mathbf{E}_{S_{n-1}}\left\{E\{f[S_n] \mid S_{n-1}\}\right\} \\ &= \mathbf{E}_{S_{n-1}}\{\lambda f[S_{n-1}]\} \\ &= \lambda E\{f[S_{n-1}]\} \\ &= \lambda^n \end{aligned} \quad (28)$$

That gives us Wald's identity.

$$\lambda^{-n} E\{e^{\theta S_n}\} = 1 \quad (29)$$

although it is typically written with the constant inside the expectation

$$E\left\{\frac{e^{\theta S_n}}{M[\theta]^n}\right\} = 1 \quad (30)$$

I substituted the mgf for λ. I want to emphasize that (23) holds for any value of θ. We're going to pick a particular value in section 2.

- **A stochastic process resembling X_n is a martingale**

I want to make a minor adjustment to our stochastic process from above. Recall that the stochastic process , X_i, was generated by taking the random sums (i.e., our random walk) and exponentiating. All I want to do now is divide by the eigenvalues we' ve been referring to as λ for convenience but which we know are really M[θ]. To be cute, I' ll call this new stochastic process W_0, W_1 , ..., W_n

$$W_n = \lambda^{-n} f[S_n] = \frac{e^{\theta S_n}}{M[\theta]^n} \tag{31}$$

I wrote this two ways. The first expression makes it obvious that this stochastic process is generated using the eigenfunction and eigenvalue associated with the markov process of partial sums. The second version makes the connection to Wald' s identity. In fact, we already know that the E {W_n} = 1 for all n. But just for the heck of it, let' s prove that the stochastic process, W_0, W_1 , ..., W_n is a martingale.

To prove this, we need to show that

$$E\{W_{n+1} \mid Y_0, Y_1, \ldots, Y_n\} = W_n \tag{32}$$

Our W depends on the sums of the Y_i. In other words, the S_i we' ve been talking about. Moreover, because the process is markov, we don' t need to write down the whole history. The most recent state covers everything we need to know about the process. So we need only convince ourselves that

$$E\{W_{n+1} \mid S_n\} = W_n \tag{33}$$

From the definition of W, we have

$$E\{W_{n+1} \mid S_n\} = E\left\{\lambda^{-(n+1)}\, e^{\theta S_{n+1}} \;\middle|\; S_n\right\} \tag{34}$$

Using the same stragey as before, we will break S_{n+1} into the part we know (i.e., S_n) and the new random bit that got added on in the n + 1 step.

$$\begin{aligned} &= E\left\{\lambda^{-(n+1)}\, e^{\theta(S_n + Y_{n+1})}\right\} \\ &= \left(\lambda^{-n}\, e^{\theta S_n}\right) \lambda^{-1}\, E\{e^{\theta Y_{n+1}}\} \\ &= \left(\lambda^{-n}\, e^{\theta S_n}\right) \lambda^{-1}\, \lambda \\ &= W_n \end{aligned} \tag{35}$$

Let me walk you through those four lines. The first is just expressing S_{n+1} as the sum through the n^{th} step plus the increment. The next step pulls all the non − random stuff out of the expectation. Remember, we' re assuming we know S_n. I also grouped together in the parenthetical a term that you should recognize as equivalent to W_n. That left a λ^{-1} straggling. To get to third line we just simplified the expectation. Like before, we recognize that there' s nothing special about the $(n+1)^{th}$ increment. So we forget subscript from Y. In fact, that expectation is just M[θ]. We' re using λ to represent M[θ]. So to be consistent, we write it that way. Now we can see that the λ' s outside the parentheses annihilate each another. That leaves W_n, which completes the proof. It is worth pointing out that this anihiliation of λ' s obscures the important fact that the constant, θ, in the exponential is the same constant that we use in argument to the moment generating funcion, M[θ].

■ **Summary**

The cumulative sum of iid random variables, $S_n = Y_1 + Y_2 + ... + Y_n$,
is a stochastic process that describes a ranom walk in discrete time steps,
n. We constructed a new stochastic process by transforming the S_n . The transformation involves exponentiating S_n (times a constant) and dividing by another constant raised to the n^{th} power. The latter constant is the MGF associated with the Y variable that increments S_n at each step of the random walk. Actually it is the MGF evaluated at some value, θ. This θ is the constant that multiplies S_n in the term $e^{\theta S_n}$. The new stocastic process, W_i, is a martingale whose expectation is 1. Because $\{W_i\}$ is a martingale,
we can stop it using any rule we want to apply without affecting the expectaion. In other words,
Wald' s identity will hold for the sum even when we stop it using decision criteria.

2. From Wald's identity to the logistic response function

Consider the same random walk, $S_n = Y_1 + Y_2 + ... + Y_n$ as above. But add the following decision rule. If S_n is between than some upper barrier value, +A, and some lower barrier –B, take another sample. Otherwise stop the process. Note that A and B are positive numbers. The idea is that if the path stops at the +A barrier, the process makes a decision in favor of proposition 'A'. If it stops by hitting the lower barrier, it makes a decision in favro of proposition 'B'.

Let's take it as given that the process will eventually stop. Perhaps I'll fill in the proof one day. Here, we need to recognize that the random walk is stochastic. Any particular instantiation of the stochastic process (random walk) is a path that could end at +A or –B. In fact, we assume that when the process stops, the sum does not overshoot +A or –B, but hits these values exactly. That's an important (and often questionable) assumption, but it certainly holds when the the values for the Y are tiny compared to the barrier heights. The stopping time or number of steps, n, is a random value, which we will explore in the next section. In this section, let's focus on which barrier the path hits first, that is, the probability of making an 'A' versus a 'B' decision. Of course this depends on the random variables used for the increments and also on the values we choose for A and B.

The stopped random walk, $\hat{S}_n$, which I will denote with a hat, is either + A or –B. It is a random number,
but it' s an easy one to work with. What is the mean (expectation) of $\hat{S}_n$? Let P_A be the probability that the process stops by hitting the upper barrier first.It' s just a weighted sum of + A and –B :

$$E\{\hat{S}_n\} = P_A\, A + (1 - P_A)\,(-B) \tag{36}$$

This somewhat trivial random variable, $\hat{S}_n$, also has a moment generating function. It' s just a weighted sum also :

$$M_{\hat{S}_n}[\theta] = E\left[e^{\theta \hat{S}_n}\right] = P_A\, e^{\theta A} + (1 - P_A)\, e^{\theta(-B)} \tag{37}$$

Let's relate this expression to Wald's identity.

$$E\left[\frac{e^{\theta S_n}}{M[\theta]^n}\right] = 1 \tag{38}$$

Because the term inside the expectation is a martingale, we know that this expectation is the same even if the process is stopped according to some rule, even a rule that is based on the observations of the stochastic process itself. This is the main virtue of martingales. We are exploiting the optional stopping theorem to add a hat to Wald's identity.

$$E\left[\frac{e^{\theta \hat{S}_n}}{M[\theta]^n}\right] = 1 \tag{39}$$

The expession holds for any particular choice of θ.
But it would be nice to lose the denominator. So let's exploit the intuition we developed about the moment generating function. For a large class of distributions, `Yi`, there exists a value for θ such that
$M[\theta_1]=1$, where $\theta_1 \neq 0$.(Link uses $\ominus^*$ to refer to this special value, θ_1. The notation is problematic for mathematica; so I'm sticking with the subscript.) Note that when $\theta = \theta_1$, the denominator in Wald's identity disappears, leaving behind an expectation that looks something like the mgf for the stopped sum, $\hat{S}_n$. In fact, it is not the moment generating function, but moment generating function evaluated at a particular value of its argument $\theta = \theta_1$.

$$M_{\hat{S}_n}[\theta_1] = E\left[e^{\theta_1 \hat{S}_n}\right] = 1 \tag{40}$$

We can now relate the two expressions for moment generating functions

$$P_A\, e^{\theta_1 A} + (1 - P_A)\, e^{\theta_1 (-B)} = E\left[e^{\theta_1 \hat{S}_n}\right] = 1 \tag{41}$$

This leads to a simple expression of th eprobability of hitting the A barrier first,
i.e., the probability of a decision in favor of 'A'.

- **Use mathematica to simplify the algebra**

$$\textbf{Solve}\left[P_A\, e^{\theta_1 A} + (1 - P_A)\, e^{\theta_1 (-B)} == 1,\ P_A\right]$$

$$\left\{\left\{P_A \to \frac{-1 + e^{B\,\theta_1}}{-1 + e^{A\,\theta_1 + B\,\theta_1}}\right\}\right\}$$

Notice that when there is no drift, the probability of absorption is determined solely by geometry, as seen by taking the limit

$$\textbf{Limit}\left[\frac{1 - e^{-B\,\theta_1}}{e^{A\,\theta_1} - e^{-B\,\theta_1}},\ \theta_1 \to 0\right]$$

$$\frac{B}{A + B}$$

When A=B, we have symmetric diffusion

Solve[$P_A\, e^{\theta_1 A} + (1 - P_A)\, e^{\theta_1 (-A)}$ == 1, P_A]

$$\left\{\left\{P_A \to \frac{1}{1 + e^{A \theta_1}}\right\}\right\}$$

which is recognized as a logistic function of the argument, $-A\,\theta_1$

FullSimplify[$\frac{1 - e^{-A \theta_1}}{-e^{-A \theta_1} + e^{A \theta_1}}$]

$$\frac{1}{1 + e^{A \theta_1}}$$

The algebra is transparent once you see that the denominator can be factored into

Factor[$-e^{-A \theta_1} + e^{A \theta_1}$]

$$e^{-A \theta_1} (-1 + e^{A \theta_1}) (1 + e^{A \theta_1})$$

We recognize that this is the same equation that arises in Brownian motion to symmetric barrier at A when the drift rate is μ and variance is 1, because $\theta_1 = -2\,\mu / \sigma^2$ for the normal distribution. See RewardEquations_fromRoss (and below). Remember also that Link defines the mgf as E[$e^{-\theta x}$]. That leads to a more standard looking logistic because of the negative exponent.

3. From Wald's identity to the response time function

To get the response time function, we need to solve for the number of steps for the accumulation, $S_n = Y_1 + \ldots + Y_n$, to reach ±A. I know of three ways to show that the stopped process has an expectation,

$$E[S_n] = E[n]\, E[Y]. \tag{42}$$

One that does not use martingales can be found on p. 105 of Ross (2nd Edition). Another uses martingale based on the sum:

$$X_n = S_n - n\, E[Y], \quad E[X_n] = E[X_1] = 0$$

which gives the desired result in 1 step. I will use Wald's martingale because it is a little less obvious and because it's nice to see that the moment generating functions can really serve the purpose of generating moments. The key to the use of the martingales is that one can exploit the "optional stopping theorem," which states essentially, that if $\{X_n\}$ is a martingale, then stopping the process at n=T does not change E[X_n]. This seems pretty intuitive because E[X_n] = E[X_{n-1}]=X_0. So start with Wald's martingale

$$E[M[\theta]^{-n}\, e^{\theta S_n}] = 1$$

and differentiate with respect to θ.

- **Here's a little reminder about derivatives of products and the use of the chain rule:**

- **Returning to the derivation, the derivative with respect to θ is**

$$E[e^{S\theta} S M[\theta]^{-n} - e^{S\theta} n M[\theta]^{-1-n} M'[\theta]] = 0$$

Suppose we set $\theta = 0$. We know that $M[\theta] = 1$ (because $M[\theta]$ is the expectation of $e^{\theta Y}$) and most importantly, we know that the derivative of the mgf evaluated at 0 is the 1st moment, that is E[Y].

$$\begin{aligned} &E[S - n E[Y]] = 0 \\ &E[S] - E[n] E[Y] = 0 \\ &E[S] = E[n] E[Y] \end{aligned} \tag{43}$$

which is what we obtained using the other methods. Rearranging, we see that expectation of the number of steps, n, for the sum of Y to reach ±A is

$$E[n] = \frac{E[S]}{E[Y]} \tag{44}$$

The numerator, E[S], is just the average stopping position. Since this is either A or -A,

$$\begin{aligned} E[S] &= P_A A + (1 - P_A)(-A) \\ &= (2 P_A - 1) A \end{aligned}$$

$$E[n] = \frac{(2 P_A - 1) A}{\mu} \tag{45}$$

Substitute the logistic function for $P_A = \frac{1}{1+e^{\theta_1 A}}$.

$$
\begin{aligned}
E[n] &= \frac{\left(\frac{2}{1+e^{\theta_1 A}} - 1\right) A}{\mu} \\
&= \frac{1 - e^{\theta_1 A}}{1 + e^{\theta_1 A}} \left(\frac{A}{\mu}\right) \\
&= \left(\frac{e^{\frac{-\theta_1 A}{2}}}{e^{\frac{-\theta_0 A}{2}}}\right) \frac{1 - e^{\theta_1 A}}{1 + e^{\theta_1 A}} \left(\frac{A}{\mu}\right) \\
&= \frac{e^{\frac{-\theta_1 A}{2}} - e^{\frac{-\theta_1 A}{2}}\ e^{\theta_1 A}}{e^{\frac{-\theta_1 A}{2}} + e^{\frac{-\theta_1 A}{2}}\ e^{\theta_1 A}} \left(\frac{A}{\mu}\right) \\
&= \frac{e^{\frac{-\theta_1 A}{2}} - e^{\frac{\theta_1 A}{2}}}{e^{\frac{-\theta_1 A}{2}} + e^{\frac{\theta_1 A}{2}}} \left(\frac{A}{\mu}\right) \\
&= -\mathrm{Tanh}\left[\frac{\theta_1 A}{2}\right]\left(\frac{A}{\mu}\right)
\end{aligned}
\tag{46}
$$

Note that for Brownian motion, we assume the Y are distributed as normal with mean, μ, and variance σ^2 . In that case, $\Theta_1 = \frac{-2\,\mu}{\sigma^2}$, using the standard definition of the mgf (i.e., the one we've been using here, not Link's). Therefore

$$E[n] = \mathrm{Tanh}\left[\frac{\mu A}{\sigma^2}\right]\left(\frac{A}{\mu}\right)$$

■ **Some mathematica expressions to check the algebra above...**

4. Weber's law from a Poisson difference

Link's ideas on this are mainly in Link Ch. 12, pp. 192ff. I think it would be helpful to step more carefully through the logic and the math. The logic behind the theory is similar to the approach one would take using the theory of signal detection. It may be helpful to expose the common set of ideas. That way we can see where Link's theory departs from SDT and appreciate its advantages. In both theories, we compare a base stimulus and a test stimulus. We assign internal representation of base and test using random variables and we consider the discriminability of the two signals or the detectability of the difference.

■ The formulation from Signal Detect Theory

■ Overview

Suppose the base stimulus, $S_{1\,L}$, gives rise to a (neural) signal whose mean intensity is $\mu_{1\,L} = L_1\,\alpha$, where α is some kind of unit intensity value which scaled by L_1. To this base, we compare a second stimulus, $S_{1\,M}$. It gives rise to a signal whose mean intensity $\mu_{1\,M} = M_1\,\alpha$. We can think of the first stimulus as a pedestal value and the second a pedestal plus an increment. The stimulus difference associated with the increment, ΔS_1, gives rise to a difference in signals, $\Delta\mu_1 = \alpha\,(M_1 - L_1)$. To theorize about Weber's law, we consider the size of this increment that is required to produce some criterion level of discriminability. We then do the same thing for a scaled version of the base stimulus. In that case we are comparing signals, $S_{2\,L}$ and $S_{2\,M}$, with mean intensties, $\mu_{2\,L} = L_2\,\alpha$ and $\mu_{2\,M} = M_2\,\alpha$. Weber's law states that the criterion performance is achieved for increments in stimulus that scale with the baseline stimulus intensity. That is,

$$\frac{\alpha\,(M_1 - L_1)}{L_1\,\alpha} = \frac{\alpha\,(M_2 - L_2)}{L_2\,\alpha} \qquad (1)$$

which implies that discriminability depends only on the ratio of the signal intensities

$$\frac{M_1}{L_1} = \frac{M_2}{L_2}. \qquad (2)$$

Our job is to explain why this would be true. In order to make statements about the discriminability of two signals, we have to specify (assume) something about the variability of the random signals.

■ Constant variance assumption

If we assume that all the signals are Normally distributed with means specified by the μ_{ij} and identical variance, we would require the same difference in mean intensity to produce equal discriminability, we cannot explain Weber's law, t

$$\Delta\mu_1 = \Delta\mu_2$$

Under our scaling assumption, this implies

$$(M_1 - L_1) = (M_2 - L_2)$$

which cannot be true if (2) holds.

Clearly, the problem is the scaling assumption that the stimulus gives rise to an internal representation that is proportiona its intensity. One solution is to assume that the (neural) signals scale by the log of stimulus intensity. If $\mu_{ij} = \log S_{ij}$, the

$$\log(k\;S_{ij}) = \log(k) + \log(S_{ij}).$$

So, one way SDT can give us Weber's law is to assume a log transformation of the stimulus intensity into its neural signal and constant noise. What is worth noting is that SDT is not really achieving Weber's law in this case, except by fiat. It gives us Weber's law only by enforcing a log transformation at the level of stimulus encoding. In general, we think that the variance of neural signals is not constant but instead varies with mean intenseity, and it is often the case that larger stimluli lead to proportionately larger neural responses. In that case, SDT simply fails to exp

■ Poisson assumption

A more realistic formulation is that the variance of internal signals scales in some way with the mean value. Clearly, if that is the case, then the log transformation would not achieve Weber's law. In fact, we would need a larger increment than
the one predicted by Weber's law because there would be more noise associated with the larger signals, hence the need for a larger separation of μ_{2M} from μ_{2L}. Let's see how this plays out.

{I'm copying some of what I wrote above. I might move it here.}

Suppose the base stimulus, S_{1L}, gives rise to a (neural) signal whose mean intensity is $\mu_{1L} = L_1\,\alpha$, where α is some kind of unit intensity value which scaled by L_1. To this base, we compare a second stimulus, S_{1M}. It gives rise to a signal whose mean intensity $\mu_{1M} = M_1\,\alpha$. We can think of the first stimulus as a pedestal value and the second a pedestal plus an increment. The stimulus difference associated with the increment, ΔS_1, gives rise to a difference in signals, $\Delta\mu_1 = \alpha\,(M_1 - L_1)$. To theorize about Weber's law, we consider the size of this increment that is required to produce some criterion level of discriminability. We then do the same thing for a scaled version of the base stimulus. In that case we are comparing signals, S_{2L} and S_{2M}, with mean intensties, $\mu_{2L} = L_2\,\alpha$ and $\mu_{2M} = M_2\,\alpha$.
If the μ_x are Poisson means,
then SDT tells us that the discriminability of S_{1M} from S_{1L} is governed by the difference of the mean neural signals, $\alpha\,(M_1 - L_1)$, in relation to the mean uncertainty, $\sqrt{\alpha\,(M_1 + L_1)}$.
If $M_1 = w\,L_1$, the probability that a random draw x_M is larger than x_L is given by the function pYgtX, below. Notice that this function does not give us Weber's law. We can also see this by deriving D' as a function of the two Poisson rates. Clearly, D' is not the same when the rates scale.

```
<< Statistics`DiscreteDistributions`
```

Begin by defining a joint Poisson surface

```
p1[x_, y_, r1_, r2_] :=
  (PDF[PoissonDistribution[r1], x]) ( PDF[PoissonDistribution[r2], y]
   )
```

The probability that the 2nd rv is greater than the 1st is given by

```
pYgtX[a_, b_, r1_, r2_] := Sum[ p1[a, b, r1, r2], {a, 0, r1 * 10}, {b, a + 1, r2 * 10}] +
    0.5 (Sum[ p1[a, b, r1, r2], {a, 0, r1 * 10}, {b, a, a}])
```

where ties are equally likely to be classified as a<b or b<a.

Verify that pYgtX is 0.5 when the Poisson rates are identical

```
pYgtX[x, y, 5, 5]

0.5
```

When the 2nd rate is larger, the probability reflect this

```
pYgtX[x, y, 3, 5]

0.758373
```

But, when the rates are doubled, the probability changes, contrary to Weber's law.

```
pYgtX[x, y, 6, 10]

0.841312
```

Examine Dprime for a pair of Poisson distributed RVs

diffMean [r1_, r2_] := Sum[(x − y) p1[x, y, r1, r2], {x, 0, 10 r1}, {y, 0, 10 r2}]
diffVar[r1_, r2_] :=
  Sum[$(x - y)^2$ p1[x, y, r1, r2], {x, 0, 10 r1}, {y, 0, 10 r2}] − $(\text{diffMean[r1, r2]})^2$

diffDprime[r1_, r2_] := $\dfrac{\text{diffMean[r1, r2]}}{\sqrt{\text{diffVar[r1, r2]}}}$

```
N[diffDprime[4, 3] ]

0.377964
```

```
N[diffDprime[8, 6] ]

0.534522
```

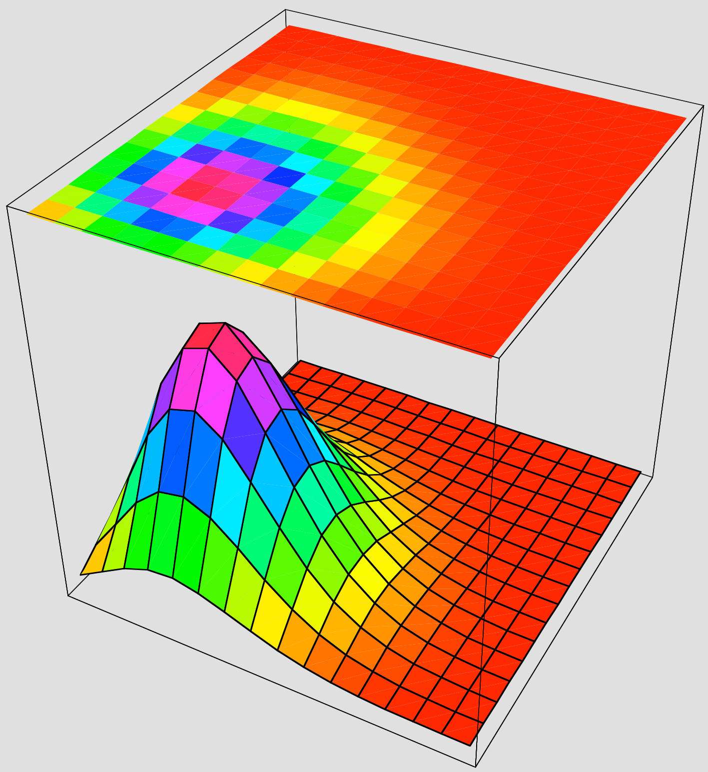
```
N[diffDprime[2, 3 / 2] ]
```

0.267261

■ **Some graphics to look at the joint probability surfaces**

```
<< Graphics`Graphics3D`
```

```
ShadowPlot3D[p1[x, y, 3, 3], {x, 0, 10},
  {y, 0, 10}, ShadowPosition → 1,  ShadowMesh →  False]
```



- Graphics3D -

■ The formulation from Wave Difference theory

■ **Short digression: the moment generating function for the Poisson Distribution**

- **Here is the development of Weber's law using Link's wave difference theory.**

Suppose that signal 1 is the sum of L Poisson intensities, each parameterized by α. We know that this is just a Poisson rv with rate parameter Lα

$$\mathbf{M_{1,L}[\theta\_] := e^{(e^{\theta}-1)\,L\,\alpha}}$$

Suppose that signal 2 is the sum of M Poisson intensities, each parameterized by α. We know that this is just a Poisson rv with rate parameter Mα

$$\mathbf{M_{1,M}[\theta\_] := e^{(e^{\theta}-1)\,M\,\alpha}}$$

What is the mfg for the sum? It is the product

$$\mathbf{M_{1,M}[\theta]\; M_{1,L}[\theta]}$$

$$e^{(-1+e^{\theta})\,L\,\alpha+(-1+e^{\theta})\,M\,\alpha}$$

We recognize that the mgf of the sum is the mgf of a Poisson distribution with parameter (L+M)α

$$\mathbf{Re[L] > 0;\; Re[M] > 0;\; \alpha > 0;}$$
$$\mathbf{Simplify\left[\frac{(-1+e^{-\theta})\,L\,\alpha+(-1+e^{-\theta})\,M\,\alpha}{(-1+e^{-\theta})\,(L+M)\,\alpha}\right]}$$

$$1$$

What is the mfg for the difference variable? It is the product

$$\mathbf{mgfDiff[\theta\_] := M_{1,M}[\theta]\; M_{1,L}[-\theta]}$$
$$\mathbf{FullSimplify[mgfDiff[\theta]]}$$

$$e^{(-1+e^{-\theta})\,L\,\alpha+(-1+e^{\theta})\,M\,\alpha}$$

```
Off[Solve::"ifun"]; Off[InverseFunction::"ifun"];
```

$$\mathbf{Solve[(-1+e^{\theta})\,L\,\alpha+(-1+e^{-\theta})\,M\,\alpha == 0,\, \theta]}$$

$$\left\{\{\theta \to 0\},\, \left\{\theta \to \mathrm{Log}\left[\frac{M}{L}\right]\right\}\right\}$$

Of course, the interesting solution here is the second one, the all important Θ_1

$$\theta_1 = \mathrm{Log}\left[\frac{M}{L}\right];$$

The probability of a particular choice is governed by the logistic function with parameter, $A\,\theta_1$.

$$p = \frac{1}{1 + e^{A\,\theta_1}}$$

$$\frac{1}{1 + \left(\frac{M}{L}\right)^{A}}$$

And this demonstrates that we will get the same proportion of choices for the same ratio, $\frac{M}{L}$, independent of α.

5. Background appendices

Transition functions for Markov Processes

Using an eigenfunction of a Markov process to create a martingale?

- **Start with an eigenvector of a transition matrix**

If P is a transition matrix with states in the rows, then $\vec{v}$ is a right eigenvector of P if

$\lambda\,\vec{v} = P\,\vec{v}$

where λ is a constant (eigenvalue).

Karlin writes this so that we can recognize f as a function

$$\lambda\, f[i] = \sum_{j} P_{ij}\, f[j] \tag{47}$$

Notice that the right side is just $P.\vec{f}$

The claim is that if $Y_0, Y_1, \ldots, Y_n$ is a Markov chain with transition probability matrix, P, then

$X_n = \lambda^{-n} f[Y_n]$

is a martingale. The proof is pretty simple :

$$\begin{aligned}&E[X_{n+1} \mid Y_0, \ldots, Y_n]\\ &\quad = E[\lambda^{-(n+1)}\, f[Y_{n+1}] \mid Y_0, \ldots, Y_n]\\ &\quad = \lambda^{-n}\,\lambda^{-1}\, E[f[Y_{n+1}] \mid Y_n]\end{aligned}$$

simply by substituting the definition of X_n and applying the Markov property. The next step contains funny terminology (the $P_{Y_{n,j}}$ term),

$$= \lambda^{-n} \lambda^{-1} \sum_j P_{Y_{n,j}} f[j]$$

and is worthy of a little thought. The conditional, ... | Y_n,
tells us that we are in the Y_n state. That tells us which row to look at in P. The values in this row add up to 1. They are a probability function. So, to find the expectation of f after the transition, we simply weigh the elements of f by the probabilities of going to each of the states. This gives us $E[f[Y_{n+1}]]$.

With this expression in hand, we make the obvious substitution using the fact the definition of f :

$$= \lambda^{-n} f[Y_n] = X_n$$

which completes the proof.

Moment generating functions that possess θ_1 (Link's θ^*)

I plan on adding to this section. As of July 4, 2003, I show that the mfg for a Poisson RV does not possess θ_1, but the difference of two Poisson RVs does. See matlab tutorial momentGeneratingFuncTutorial.m

The mgf for a Poisson RV. The argument α is the rate function or mean, which we see from the derivative at 0

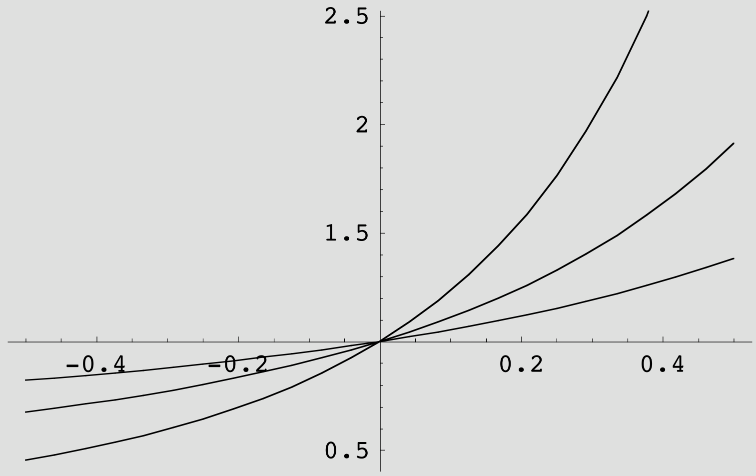
```
Unset[α]
g[x_, α_] := Exp[α (Exp[x] − 1)]
g[θ]
g'[0]
```

$$e^{(-1+e^{\theta})\alpha}$$

$$\alpha$$

Notice that it does not possess a nonzero root at g=1

```
Plot[{g[x, .5] , g[x, 1] , g[x, 2]}, {x, -.5, .5}]
```



- Graphics -

The mgf for a difference of 2 Poisson rvs. The rate is α-β. You can play around to appreciate that the slope at 0 reflects the difference. Notice that the θ_1 term is more negative, the larger the expected difference. For Link, the opposite sign convention leads to more positive Θ_1 and a logistic psychometric function with 1/(1+Exp[-Θ_1 A]).

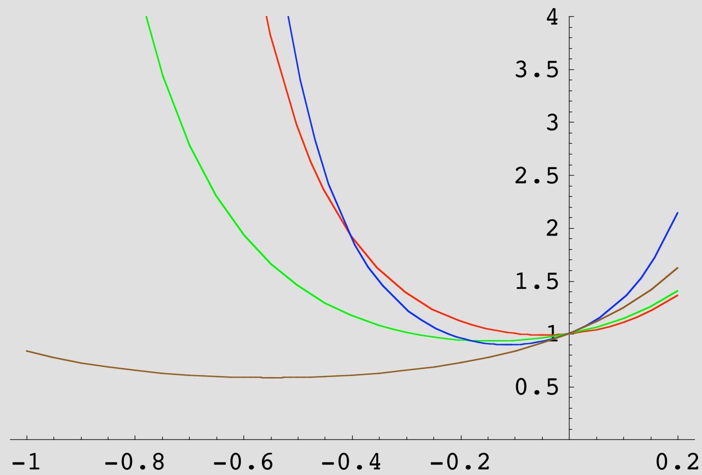
```
Unset[h]; Unset[α]; Unset[β]; Unset[a]; Unset[b];
h[x_, α_, β_] := Exp[α (Exp[x] - 1) + β (Exp[-x] - 1)]
h[θ, a, b]
h'[0, 2, 1]
```

$$e^{b\,(-1+e^{-\theta})+a\,(-1+e^{\theta})}$$

```
h′[0, 2, 1]
```

$$e^{b\,(-1+e^{-\theta})+a\,(-1+e^{\theta})}$$

```
Plot[{h[x, 5.5, 5] , h[x, 4, 3], h[x, 10, 8], h[x, 3, 1]} ,
  {x, -1, .2},  PlotStyle → {RGBColor[1, 0, 0], RGBColor[0, 1, 0],
     RGBColor[0, 0, 1], RGBColor[.5, .3, .1]},  PlotRange → {0, 4}]
```



- Graphics -