Implication: If a sufficient statistic T exists, we can immediately restrict attention to its distribution when deriving an LRT.

Example 8.7. Suppose $X_1, X_2, ..., X_n$ are iid exponential(θ), where $\theta > 0$. Consider testing

$$H_0 : \theta = \theta_0$$
$$\text{versus}$$
$$H_1 : \theta \neq \theta_0.$$

(a) Show that the LRT statistic based on $\mathbf{X} = \mathbf{x}$ is

$$\lambda(\mathbf{x}) = \left(\frac{e}{n\theta_0}\right)^n \left(\sum_{i=1}^n x_i\right)^n e^{-\sum_{i=1}^n x_i/\theta_0}.$$

(b) Show that the LRT statistic based on $T = T(\mathbf{X}) = \sum_{i=1}^n X_i$ is

$$\lambda^*(t) = \left(\frac{e}{n\theta_0}\right)^n t^n e^{-t/\theta_0},$$

establishing that $\lambda^*(t) = \lambda(\mathbf{x})$, as stated in Theorem 8.2.4.
(c) Show that

$$\lambda^*(t) \leq c \iff t \leq c_1 \text{ or } t \geq c_2,$$

for some c_1 and c_2 satisfying $c_1 < c_2$.

Example 8.8. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$; i.e., both parameters are unknown. Set $\boldsymbol{\theta} = (\mu, \sigma^2)$. Consider testing

$$H_0 : \mu = \mu_0$$
$$\text{versus}$$
$$H_1 : \mu \neq \mu_0.$$

The null hypothesis H_0 above looks simple, but it is not. The relevant parameter spaces are

$$\begin{aligned} \Theta_0 &= \{\boldsymbol{\theta} = (\mu, \sigma^2) : \ \mu = \mu_0, \ \sigma^2 > 0\} \\ \Theta &= \{\boldsymbol{\theta} = (\mu, \sigma^2) : -\infty < \mu < \infty, \ \sigma^2 > 0\}. \end{aligned}$$

In this problem, we call σ^2 a **nuisance parameter**, because it is not the parameter that is of interest in H_0 and H_1. The likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}. \end{aligned}$$

$$\lambda(\underset{\sim}{x}) = \frac{\sup_{\Theta_0} L(\theta|\underset{\sim}{x})}{\sup_{\Theta} L(\theta|\underset{\sim}{x})} = \frac{\sup_{\sigma^2} L(\mu_0, \sigma^2|\underset{\sim}{x})}{\sup_{\mu,\sigma^2} L(\mu, \sigma^2|\underset{\sim}{x})}$$

Unrestricted MLE: In Example 7.6 (notes, pp 33), we showed that

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \overline{X} \\ S_b^2 \end{pmatrix} = \begin{pmatrix} \overline{X} \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{pmatrix}$$

$\lambda(\underset{\sim}{x}) \le c$

maximizes $L(\boldsymbol{\theta}|\mathbf{x})$ over Θ.

$\log \lambda(\underset{\sim}{x}) \le \log c$

Restricted MLE: It is easy to show that

$\log \frac{\sup L(\ )}{\sup L(\ )} \le \log c$

$$\widehat{\boldsymbol{\theta}}_0 = \begin{pmatrix} \mu_0 \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 \end{pmatrix}$$

$\log \sup L - \log \sup L \le \log c$

maximizes $L(\boldsymbol{\theta}|\mathbf{x})$ over Θ_0.

$\sup \log L - \sup \log L \le \log c$

(a) Show that

$$\lambda(\mathbf{x}) = \frac{L(\widehat{\boldsymbol{\theta}}_0|\mathbf{x})}{L(\widehat{\boldsymbol{\theta}}|\mathbf{x})} = \left[\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \mu_0)^2}\right]^{n/2}.$$

$\sum_{i=1}^{n}(x_i - \overline{x} + \overline{x} - \mu_0)^2$

(b) Show that

$$\lambda(\mathbf{x}) \le c \iff \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| \ge c'.$$

$\sup_{\theta\in\Theta_0} \log L(\theta|\underset{\sim}{x}) - \sup_{\theta} \log L(\theta|\underset{\sim}{x}) \le c'$

This demonstrates that the "one-sample t test" is a LRT under normality.

Exercise: In Example 7.7 (notes, pp 34-35), derive the LRT statistic to test

$$H_0 : p_1 = p_2$$
versus
$$H_1 : p_1 \neq p_2.$$

Exercise: In Example 8.4 (notes, pp 67), show that the LRT statistic is

$$\lambda(\mathbf{x}) = \lambda(x_1, x_2, x_3, x_4) = \prod_{i=1}^{4}\left(\frac{2864}{4x_i}\right)^{x_i}.$$

Also, show that

$$\lambda(\mathbf{x}) \le c \iff -2\ln\lambda(\mathbf{x}) \ge c'.$$

Under $H_0 : p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$, we will learn later that $-2\ln\lambda(\mathbf{X})$ is distributed approximately as χ_3^2. This suggests a "large-sample" LRT, namely, to reject H_0 if $-2\ln\lambda(\mathbf{x})$ is "too large." We can use the χ_3^2 distribution to specify what "too large" actually means.

8.2.2 Bayesian tests

Remark: Hypothesis tests of the form

$$H_0 : \theta \in \Theta_0$$
$$\text{versus}$$
$$H_1 : \theta \in \Theta_0^c,$$

where $\Theta_0^c = \Theta \setminus \Theta_0$, can also be carried out within the Bayesian paradigm, but they are performed differently. Recall that, for a Bayesian, all inference is carried out using the posterior distribution $\pi(\theta|\mathbf{x})$.

Realization: The posterior distribution $\pi(\theta|\mathbf{x})$ is a valid probability distribution. It is the distribution that describes the behavior of the random variable θ, updated after observing the data $\mathbf{x}$. In this light, the probabilities

$$\begin{aligned} P(H_0 \text{ true}|\mathbf{x}) &= P(\theta \in \Theta_0|\mathbf{x}) = \int_{\Theta_0} \pi(\theta|\mathbf{x}) d\theta \\ P(H_1 \text{ true}|\mathbf{x}) &= P(\theta \in \Theta_0^c|\mathbf{x}) = \int_{\Theta_0^c} \pi(\theta|\mathbf{x}) d\theta \end{aligned}$$

make perfect sense and be calculated (or approximated) "exactly." Note that these probabilities make no sense to the non-Bayesian. S/he regards θ as fixed, so that $\{\theta \in \Theta_0\}$ and $\{\theta \in \Theta_0^c\}$ are not random events. We do not assign probabilities to events that are not random.

Example 8.9. Suppose that $X_1, X_2, ..., X_n$ are iid Poisson(θ), where the prior distribution for $\theta \sim \text{gamma}(a, b)$, a, b known. In Example 7.10 (notes, pp 38-39), we showed that the posterior distribution

$$\theta|\mathbf{X} = \mathbf{x} \sim \text{gamma}\left(\sum_{i=1}^{n} x_i + a,\ \frac{1}{n + \frac{1}{b}}\right).$$

As an application, consider the following data, which summarize the number of goals per game in the 2013-2014 English Premier League season:

Goals	0	1	2	3	4	5	6	7	8	9	10+
Frequency	27	73	80	72	65	39	17	4	1	2	0

There were $n = 380$ games total. I modeled the number of goals per game X as a Poisson random variable and assumed that $X_1, X_2, ..., X_{380}$ are iid Poisson(θ). Before the season started, I modeled the mean number of goals per game as $\theta \sim \text{gamma}(1.5, 2)$, which is a fairly diffuse prior distribution.

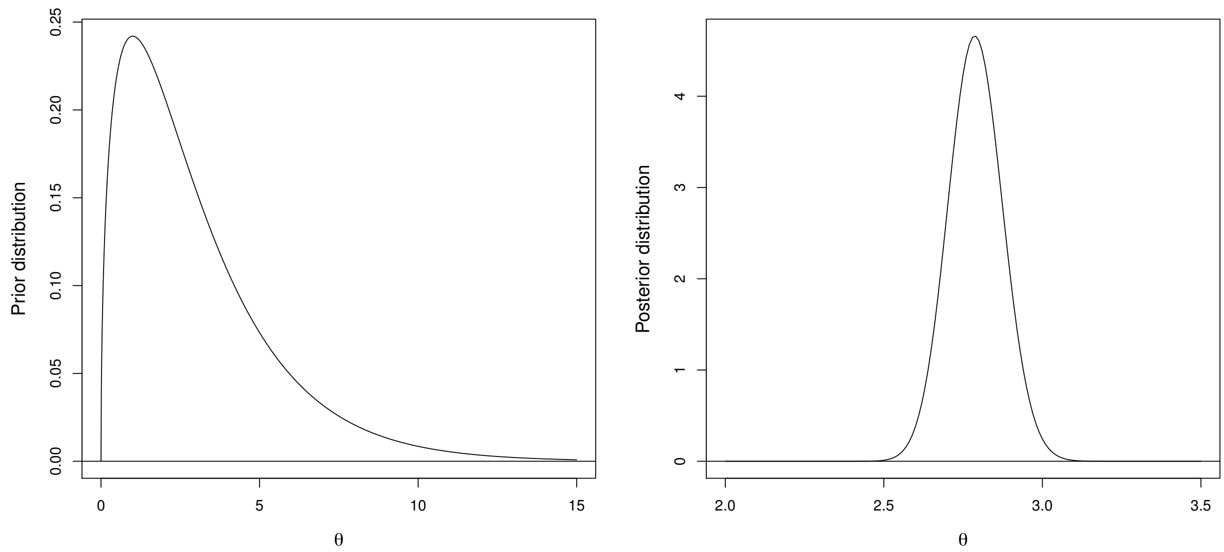

Figure 8.1: 2013-2014 English Premier League data. Prior distribution (left) and posterior distribution (right) for θ, the mean number of goals scored per game. Note that the horizontal axes are different in the two figures.

Based on the observed data, I used R to calculate

```
> sum(goals)
[1] 1060
```

The posterior distribution is therefore

$$\theta|\mathbf{X} = \mathbf{x} \sim \text{gamma}\left(1060 + 1.5,\ \frac{1}{380 + \frac{1}{2}}\right) \stackrel{d}{=} \text{gamma}(1061.5, 0.002628).$$

I have depicted the prior distribution $\pi(\theta)$ and the posterior distribution $\pi(\theta|\mathbf{x})$ in Figure 8.1. Suppose that I wanted to test $H_0 : \theta \geq 3$ versus $H_1 : \theta < 3$ on the basis of the assumed Bayesian model and the observed data $\mathbf{x}$. The probability that H_0 is true is

$$P(\theta \geq 3|\mathbf{x}) = \int_3^\infty \pi(\theta|\mathbf{x})d\theta \approx 0.008,$$

which I calculated in R using

```
> 1-pgamma(3,1061.5,1/0.002628)
[1] 0.008019202
```

Therefore, it is far more likely that H_1 is true, in fact, with probability over 0.99.

8.3 Methods of Evaluating Tests

Setting: Suppose $\mathbf{X} = (X_1, X_2, ..., X_n) \sim f_{\mathbf{X}}(\mathbf{x}|\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$ and consider testing

$$H_0 : \theta \in \Theta_0$$
versus
$$H_1 : \theta \in \Theta_0^c,$$

where $\Theta_0^c = \Theta \setminus \Theta_0$. I will henceforth assume that θ is a scalar parameter (for simplicity only).

8.3.1 Error probabilities and the power function

Definition: For a test (with test function)

$$\phi(\mathbf{x}) = I(\mathbf{x} \in R), \quad \text{RR}$$

we can make one of two mistakes:

1. Type I Error: Rejecting H_0 when H_0 is true
2. Type II Error: Not rejecting H_0 when H_1 is true.

Therefore, for any test that we perform, there are four possible scenarios, described in the following table:

		Decision	
		Reject H_0	Do not reject H_0
Truth	H_0	Type I Error	☺
	H_1	☺	Type II Error

Calculations:

$P(\text{reject } H_0 \mid \theta \in \Theta_0)$
$= P(\underset{\sim}{X} \in RR \mid \theta \in \Theta_0) = E_\theta[\phi(\mathbf{X})]$
where $\theta \in \Theta_0$

1. Suppose $H_0 : \theta \in \Theta_0$ is true. For $\theta \in \Theta_0$,
$$P(\text{Type I Error}|\theta) = P_\theta(\mathbf{X} \in R) = E_\theta[I(\mathbf{X} \in R)] = E_\theta[\phi(\mathbf{X})].$$

2. Suppose $H_1 : \theta \in \Theta_0^c$ is true. For $\theta \in \Theta_0^c$,
$$P(\text{Type II Error}|\theta) = P_\theta(\mathbf{X} \in R^c) = 1 - P_\theta(\mathbf{X} \in R) = 1 - E_\theta[\phi(\mathbf{X})] = E_\theta[1 - \phi(\mathbf{X})].$$
$P(\text{do not reject } H_0 \mid \theta \in \Theta_1) = 1 - P(\underset{\sim}{X} \in RR \mid \theta \in \Theta_1) = 1 - E_\theta[\phi(\underset{\sim}{X})]$ where $\theta \in \Theta_1$

It is very important to note that both of these probabilities depend on θ. This is why we emphasize this in the notation.

Definition: The **power function** of a test $\phi(\mathbf{x})$ is the function of θ given by

$$\beta(\theta) = P_\theta(\mathbf{X} \in R) = E_\theta[\phi(\mathbf{X})]. \quad = \begin{cases} \text{Prob of Type I error if } \theta \in \Theta_0 \\ 1 - \text{Prob of Type II error if } \theta \in \Theta_1 \end{cases}$$

In other words, the power function gives the probability of rejecting H_0 **for all** $\theta \in \Theta$. Note that if H_1 is true, so that $\theta \in \Theta_0^c$,

$$\beta(\theta) = P_\theta(\mathbf{X} \in R) = 1 - P_\theta(\mathbf{X} \in R^c) = 1 - P(\text{Type II Error}|\theta).$$

Example 8.10. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(\mu, \sigma_0^2)$, where $-\infty < \mu < \infty$ and σ_0^2 is known. Consider testing

$$\begin{gathered} H_0 : \mu \leq \mu_0 \\ \text{versus} \\ H_1 : \mu > \mu_0. \end{gathered}$$

The LRT of H_0 versus H_1 uses the test function

$$\phi(\mathbf{x}) = \begin{cases} 1, & \dfrac{\overline{x} - \mu_0}{\sigma_0/\sqrt{n}} \geq c \\ 0, & \text{otherwise.} \end{cases}$$

The power function for this test is given by

$$\begin{aligned} \beta(\mu) = P_\mu(\mathbf{X} \in R) &= P_\mu\left(\frac{\overline{X} - \mu_0}{\sigma_0/\sqrt{n}} \geq c\right) \\ &= P_\mu\left(\overline{X} \geq \frac{c\sigma_0}{\sqrt{n}} + \mu_0\right) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{\sigma_0/\sqrt{n}} \geq \frac{\frac{c\sigma_0}{\sqrt{n}} + \mu_0 - \mu}{\sigma_0/\sqrt{n}}\right) = 1 - F_Z\left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}}\right), \end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$ and $F_Z(\cdot)$ is the standard normal cdf.

Exercise: Determine n and c such that

$$\begin{aligned} \sup_{\mu \leq \mu_0} \beta(\mu) &= 0.10 \\ \inf_{\mu \geq \mu_0 + \sigma_0} \beta(\mu) &= 0.80. \end{aligned}$$

- The first requirement implies that $P(\text{Type I Error}|\mu)$ will not exceed 0.10 for all $\mu \leq \mu_0$ (H_0 true).
- The second requirement implies that $P(\text{Type II Error}|\mu)$ will not exceed 0.20 for all $\mu \geq \mu_0 + \sigma_0$ (these are values of μ that make H_1 true).

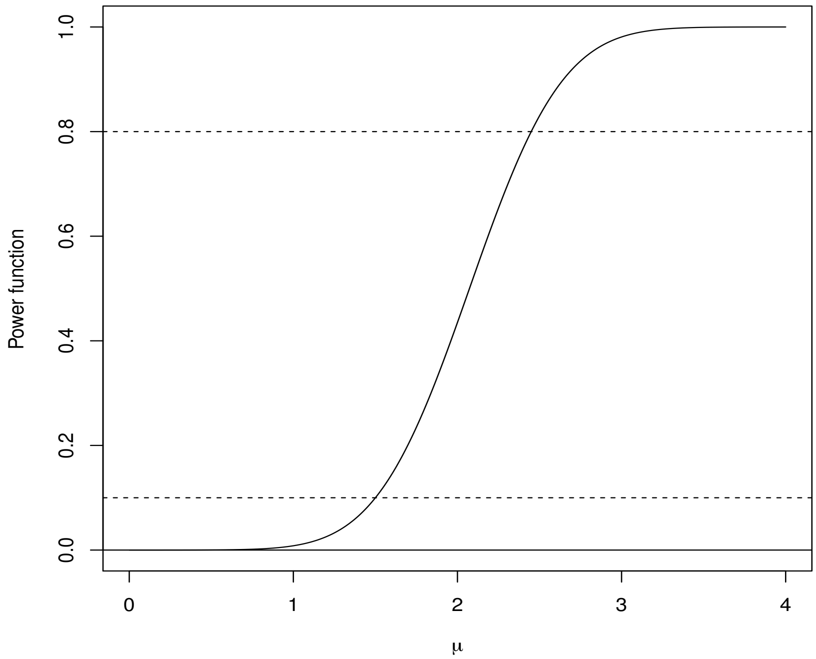

Figure 8.2: Power function $\beta(\mu)$ in Example 8.10 with $c = 1.28$, $n = 5$, $\mu_0 = 1.5$ and $\sigma_0 = 1$. Horizontal lines at 0.10 and 0.80 have been added.

Solution. Note that

$$\begin{aligned}\frac{\partial}{\partial\mu}\beta(\mu) &= \frac{\partial}{\partial\mu}\left[1 - F_Z\left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}}\right)\right] \\ &= \frac{\sqrt{n}}{\sigma_0} f_Z\left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}}\right) > 0;\end{aligned}$$

i.e., $\beta(\mu)$ is an **increasing** function of μ. Therefore,

$$\sup_{\mu \leq \mu_0} \beta(\mu) = \beta(\mu_0) = 1 - F_Z(c) \stackrel{\text{set}}{=} 0.10 \implies c = 1.28,$$

the 0.90 quantile of the $\mathcal{N}(0,1)$ distribution. Also, because $\beta(\mu)$ is increasing,

$$\begin{aligned}\inf_{\mu \geq \mu_0 + \sigma_0} \beta(\mu) = \beta(\mu_0 + \sigma_0) &= 1 - F_Z(1.28 - \sqrt{n}) \stackrel{\text{set}}{=} 0.80 \\ &\implies 1.28 - \sqrt{n} = -0.84 \\ &\implies n = 4.49,\end{aligned}$$

which would be rounded up to $n = 5$. The resulting power function with $c = 1.28$, $n = 5$, $\mu_0 = 1.5$ and $\sigma_0 = 1$ is shown in Figure 8.2.

Definition: A test $\phi(\mathbf{x})$ with power function $\beta(\theta)$ is a **size** α test if

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

The test $\phi(\mathbf{x})$ is a **level** α test if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

Note that if $\phi(\mathbf{x})$ is a size α test, then it is also level α. The converse is not true. In other words,

$$\{\text{class of size } \alpha \text{ tests}\} \subset \{\text{class of level } \alpha \text{ tests}\}.$$

Remark: Often, it is unnecessary to differentiate between the two classes of tests. However, in testing problems involving discrete distributions (e.g., binomial, Poisson, etc.), it is generally not possible to construct a size α test for a specified value of α; e.g., $\alpha = 0.05$. Thus (unless one randomizes), we may have to settle for a level α test.

Important: As the definition above indicates, the **size** of any test $\phi(\mathbf{x})$ is calculated by maximizing the power function over the null parameter space Θ_0 identified in H_0.

Example 8.11. Suppose X_1, X_2 are iid Poisson(θ), where $\theta > 0$, and consider testing

$$H_0 : \theta \geq 3$$
versus
$$H_1 : \theta < 3.$$

We consider the two tests

$$\begin{aligned} \phi_1 = \phi_1(x_1, x_2) &= I(x_1 = 0) \\ \phi_2 = \phi_2(x_1, x_2) &= I(x_1 + x_2 \leq 1). \end{aligned}$$

The power function for the first test is

$$\beta_1(\theta) = E_\theta[I(X_1 = 0)] = P_\theta(X_1 = 0) = e^{-\theta}.$$

Recall that $T = T(X_1, X_2) = X_1 + X_2 \sim \text{Poisson}(2\theta)$. The power function for the second test is

$$\beta_2(\theta) = E_\theta[I(X_1 + X_2 \leq 1)] = P_\theta(X_1 + X_2 \leq 1) = e^{-2\theta} + 2\theta e^{-2\theta}.$$

I have plotted both power functions in Figure 8.3 (next page).

Size calculations: The size of each test is calculated as follows. For the first test,

$$\alpha = \sup_{\theta \geq 3} \beta_1(\theta) = \beta_1(3) = e^{-3} \approx 0.049787.$$

For the second test,

$$\alpha = \sup_{\theta \geq 3} \beta_2(\theta) = \beta_2(3) = e^{-6} + 6e^{-6} \approx 0.017351.$$

Both ϕ_1 and ϕ_2 are level $\alpha = 0.05$ tests.

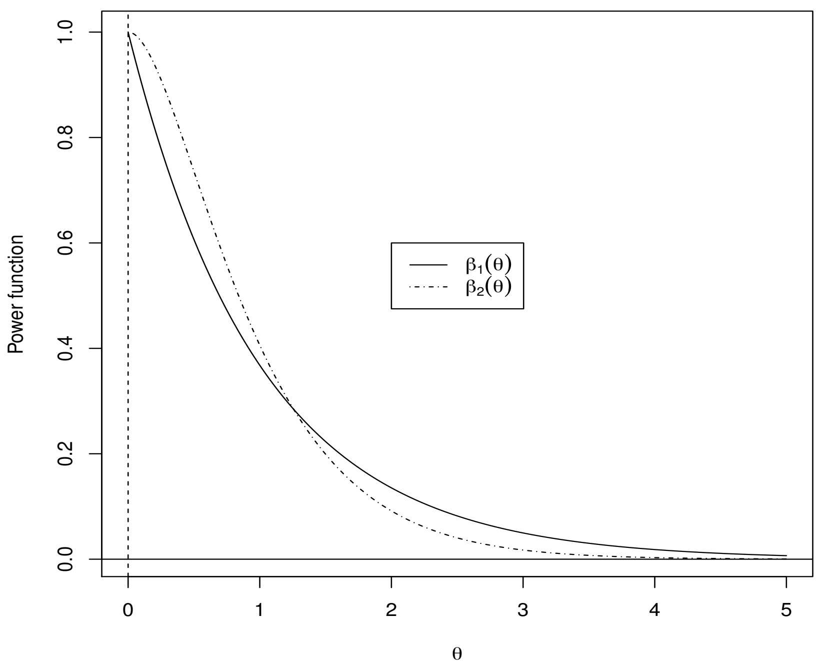

Figure 8.3: Power functions $\beta_1(\theta)$ and $\beta_2(\theta)$ in Example 8.11.

Example 8.12. Suppose $X_1, X_2, ..., X_n$ are iid from $f_X(x|\theta) = e^{-(x-\theta)}I(x \geq \theta)$, where $-\infty < \theta < \infty$. In Example 8.6 (notes, pp 72-74), we considered testing

$$\begin{gathered} H_0 : \theta \leq \theta_0 \\ \text{versus} \\ H_1 : \theta > \theta_0 \end{gathered}$$

and derived the LRT to take the form $\phi(\mathbf{x}) = I(x_{(1)} \geq c')$. Find the value of c' that makes $\phi(\mathbf{x})$ a size α test.

Solution. The pdf of $X_{(1)}$ is $f_{X_{(1)}}(x|\theta) = ne^{-n(x-\theta)}I(x \geq \theta)$. We set

$$\begin{aligned} \alpha = \sup_{\theta \leq \theta_0} E_\theta[\phi(\mathbf{X})] &= \sup_{\theta \leq \theta_0} P_\theta(X_{(1)} \geq c') \\ &= \sup_{\theta \leq \theta_0} \int_{c'}^{\infty} ne^{-n(x-\theta)}dx \\ &= \sup_{\theta \leq \theta_0} e^{-n(c'-\theta)} = e^{-n(c'-\theta_0)}. \end{aligned}$$

Therefore, $c' = \theta_0 - n^{-1}\ln\alpha$. A size α LRT uses $\phi(\mathbf{x}) = I(x_{(1)} \geq \theta_0 - n^{-1}\ln\alpha)$.