All of Statistics - Chapter 11 Solutions

Aug 24, 2020

1.

The posterior density is

$$f(\theta \mid X^n) \propto \mathcal{L}(\theta) f(\theta) \propto \exp(-g(\theta))$$

where

$$\begin{aligned}
2g(\theta) &= \frac{1}{\sigma^2} \sum_i (X_i - \theta)^2 + \frac{1}{b^2}(\theta - a)^2 \\
&= \frac{b^2 \left(\sum_i X_i^2 - 2X_i\theta + \theta^2\right) + \sigma^2 \left(\theta^2 - 2a\theta + a^2\right)}{\sigma^2 b^2} \\
&= \frac{\theta^2 \left(nb^2 + \sigma^2\right) - 2\theta \left(\sigma^2 a + b^2 \sum_i X_i\right)}{\sigma^2 b^2} + \text{const.} \\
&= \frac{nb^2 + \sigma^2}{\sigma^2 b^2} \left(\theta - \frac{\sigma^2 a + b^2 \sum_i X_i}{nb^2 + \sigma^2}\right)^2 + \text{const.}
\end{aligned}$$

It follows that the posterior density is normal with variance

$$\tau^2 = \frac{\sigma^2 b^2}{nb^2 + \sigma^2} = \left(\frac{n}{\sigma^2} + \frac{1}{b^2}\right)^{-1} = \left(\frac{1}{\operatorname{se}^2} + \frac{1}{b^2}\right)^{-1}$$

and mean

$$\bar{\theta} = \frac{\sigma^2 a + b^2 \sum_i X_i}{nb^2 + \sigma^2} = \frac{1}{b^2 / \operatorname{se}^2 + 1} a + \frac{1}{1 + \operatorname{se}^2 / b^2} \frac{1}{n} \sum_i X_i.$$

2.

a)

```
np.random.seed(1)
samples = np.random.randn(100) + 5.
```

b)

The posterior is

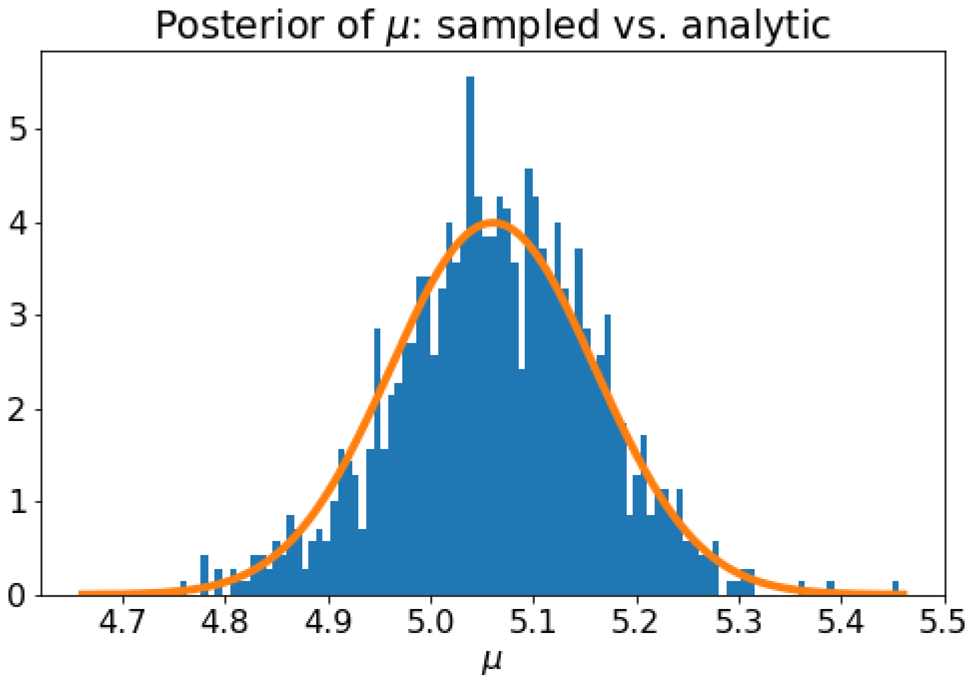
$$f(\mu \mid X^n) = \mathcal{L}(\mu) f(u) = \mathcal{L}(\mu) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_i (X_i - \mu)^2 \right).$$

Note that

$$\frac{1}{n} \sum_i (X_i - \mu)^2 = \frac{1}{n} \sum_i X_i^2 - 2\mu X_i + \mu^2 = \mu^2 - \frac{2\mu}{n} \sum_i X_i + \text{const.}$$
$$= \left(\mu - \frac{1}{n} \sum_i X_i \right)^2.$$

Therefore, the posterior is a normal distribution with mean $n^{-1} \sum_i X_i$ and variance σ^2/n (see Part (c) for a plot).

c)



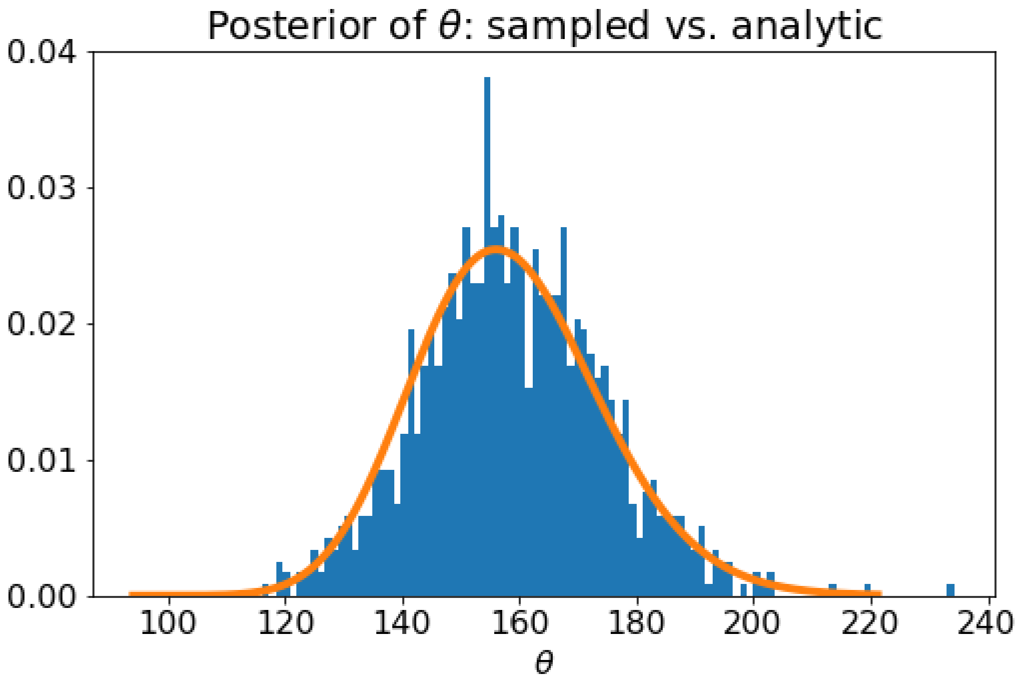
The plot is generated by the code below.

```
post_mu_mean = np.mean(samples)
post_mu_std = np.sqrt(1. / samples.size)
post_mu_grid = np.linspace(post_mu_mean - 4. * post_mu_std,
                           post_mu_mean + 4. * post_mu_std, 1000)
post_mu_pdf = scipy.stats.norm.pdf(post_mu_grid, loc=post_mu_mean,
                                   scale=post_mu_std)
np.random.seed(1)
post_mu_samples = post_mu_std * np.random.randn(1000) + post_mu_mean

plt.figure(figsize=(1.618*5., 5.))
plt.hist(post_mu_samples, bins=100, density=True)
plt.plot(post_mu_grid, post_mu_pdf, linewidth=4)
plt.title('Posterior of $\\mu$: sampled vs. analytic')
plt.xlabel('$\\mu$')
```

d)

$\theta \mid X^n$ is log-normally distributed since $\log \theta = \mu$ and $\mu \mid X^n$ is normally distributed.



The plot is generated by the code below.

```
post_theta_med = np.exp(post_mu_mean)
post_theta_std = np.sqrt((np.exp(post_mu_std**2) - 1.) \
                         * np.exp(2. * post_mu_mean + post_mu_std**2))
post_theta_grid = np.linspace(post_theta_med - 4. * post_theta_std,
                              post_theta_med + 4. * post_theta_std, 1000)
post_theta_pdf = scipy.stats.lognorm.pdf(post_theta_grid, post_mu_std,
                                         scale=post_theta_med)
post_theta_samples = np.exp(post_mu_samples)

plt.figure(figsize=(1.618*5., 5.))
plt.hist(post_theta_samples, bins=100, density=True)
plt.plot(post_theta_grid, post_theta_pdf, linewidth=4)
plt.title('Posterior of $\\theta$: sampled vs. analytic')
plt.xlabel('$\\theta$')
```

e)

Evaluating the code below yields [4.86, 5.25] as an approximate 95% confidence interval for μ.

```
index = int(post_mu_samples.size * 0.025)
sorted_post_mu_samples = np.sort(post_mu_samples)
post_mu_ci_95_lo = sorted_post_mu_samples[  index    ]
post_mu_ci_95_hi = sorted_post_mu_samples[-(index+1)]
```

f)

Since $\mathbb{P}(\mu \leq b) = \mathbb{P}(\theta \leq e^b)$ and $\mathbb{P}(\mu \geq a) = \mathbb{P}(\theta \geq e^a)$, it is sufficient to exponentiate the lower and upper bounds from Part (e) to get a 95% confidence interval for θ.

Evaluating the code below yields [129.64, 190.77] as an approximate 95% confidence interval for θ.

```
post_ci_95_theta_lo = np.exp(post_mu_ci_95_lo)
post_ci_95_theta_hi = np.exp(post_mu_ci_95_hi)
```

3.

The posterior density is proportional to

$$f(\theta \mid X^n) \propto \mathcal{L}(\theta) f(\theta) = \frac{1}{\theta} \prod_i \left[\frac{1}{\theta} I_{(X_i, \infty)}(\theta) \right] = \frac{1}{\theta^{n+1}} I_{(X_{(n)}, \infty)}(\theta).$$

In particular, $f(\theta \mid X^n)$ is a power law density with normalizing constant

$$c = \int_{X_{(n)}}^{\infty} \frac{1}{\theta^{n+1}} d\theta = \frac{1}{n X_{(n)}^n}.$$

4.

a)

By equivariance, the MLE is

$$\hat{\tau} = \hat{p}_2 - \hat{p}_1 = \frac{40}{50} - \frac{30}{50} = 0.2.$$

The standard error is (see Chapter 9 Question 7 Part (c))

$$\widehat{\text{se}}(\hat{\tau}) = \sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{50} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{50}} \approx 0.089.$$

Therefore, a 90% confidence interval for τ is

$$\hat{\tau} \pm 1.645 \cdot \widehat{\text{se}}(\hat{\tau}) \approx [0.053, 0.35].$$

The computation above is replicated in the code below.

```
n_patients = 50
placebo_success = 30
treatment_success = 40

mle_p1 = placebo_success / n_patients
mle_p2 = treatment_success / n_patients
mle_tau = mle_p2 - mle_p1
mle_tau_se = np.sqrt(  mle_p1 * (1. - mle_p1) / n_patients \
                     + mle_p2 * (1. - mle_p2) / n_patients)
mle_tau_ci_90_lo = mle_tau - 1.645 * mle_tau_se
mle_tau_ci_90_hi = mle_tau + 1.645 * mle_tau_se
```

b)

Evaluating the code below yields approximately the same standard error and confidence interval found in Part (a).

```
n_sims = 10**6

np.random.seed(1)
bstrap_p1_samples = np.random.binomial(n=n_patients, p=mle_p1,
                                       size=n_sims) / n_patients
bstrap_p2_samples = np.random.binomial(n=n_patients, p=mle_p2,
                                       size=n_sims) / n_patients
bstrap_tau_samples = bstrap_p2_samples - bstrap_p1_samples
bstrap_tau_mean = np.mean(bstrap_tau_samples)
bstrap_tau_std = np.std(bstrap_tau_samples)
bstrap_tau_ci_90_lo = bstrap_tau_mean - 1.645 * bstrap_tau_std
bstrap_tau_ci_90_hi = bstrap_tau_mean + 1.645 * bstrap_tau_std
```

c)

Under the prior $f(p_1, p_2) = 1$, the posterior is proportional to

$$f(p_1, p_2 \mid X_1, X_2) \propto p_1^{X_1}(1 - p_1)^{n - X_1} p_2^{X_2}(1 - p_2)^{n - X_2}.$$

By Theorem 2.33, the posterior is a product of independent distributions with densities

$$g_i(p_i) \propto p_i^{X_i}(1 - p_i)^{n - X_i}.$$

It follows that each is a Beta distribution with parameters $\alpha_i = X_i + 1$ and $\beta_i = n - X_i + 1$.

Evaluating the code below yields a posterior mean of approximately 0.19 and posterior 90% confidence interval of [0.047, 0.34].

```
np.random.seed(1)
post_p1_samples = np.random.beta(a=placebo_success + 1,
                                 b=n_patients - placebo_success + 1,
                                 size=n_sims)
post_p2_samples = np.random.beta(a=treatment_success + 1,
                                 b=n_patients - treatment_success + 1,
                                 size=n_sims)
post_tau_samples = post_p2_samples - post_p1_samples
post_tau_mean = np.mean(post_tau_samples)
index = int(n_sims * 0.05)
sorted_post_tau_samples = np.sort(post_tau_samples)
post_tau_ci_90_lo = sorted_post_tau_samples[  index   ]
post_tau_ci_90_hi = sorted_post_tau_samples[-(index+1)]
```

d)

Let

$$g(p_1, p_2) = \log\left(\left(\frac{p_1}{1 - p_1}\right) \div \left(\frac{p_2}{1 - p_2}\right)\right).$$

By equivariance, the MLE is $\hat{\psi} = g(\hat{p}_1, \hat{p}_2) \approx -0.98$. The Fisher information matrix is (see Chapter 9 Question 7 Part (b))

$$I(p_1, p_2) = \operatorname{diag}\left(\frac{n}{p_1\,(1-p_1)}, \frac{n}{p_2\,(1-p_2)}\right).$$

Moreover,

$$\nabla g(p_1, p_2)^\intercal = \left(\frac{1}{p_1\,(1-p_1)}, \frac{1}{p_2\,(1-p_2)}\right).$$

Therefore, by the delta method,

$$\widehat{\text{se}}(\hat{\psi}) = \sqrt{\hat{\nabla} g^\intercal I(\hat{p}_1, \hat{p}_2)^{-1} \hat{\nabla} g} = \sqrt{\frac{1}{\hat{p}_1\,(1-\hat{p}_1)\,n} + \frac{1}{\hat{p}_2\,(1-\hat{p}_2)\,n}} \approx 0.46.$$

A 90% confidence interval for ψ is

$$\hat{\psi} \pm 1.645 \widehat{\text{se}}(\hat{\psi}) \approx [-1.73, 0.23].$$

The computation above is replicated in the code below.

```
mle_p1_ratio = mle_p1 / (1. - mle_p1)
mle_p2_ratio = mle_p2 / (1. - mle_p2)
mle_psi = np.log(mle_p1_ratio / mle_p2_ratio)
mle_psi_se = np.sqrt(  1./ (mle_p1 * (1. - mle_p1) * n_patients) \
                     + 1./ (mle_p2 * (1. - mle_p2) * n_patients))
mle_psi_ci_90_lo = mle_psi - 1.645 * mle_psi_se
mle_psi_ci_90_hi = mle_psi + 1.645 * mle_psi_se
```

e)

Evaluating the code below yields a posterior mean of approximately -0.95 and posterior 90% confidence interval of [-1.70, -0.22].

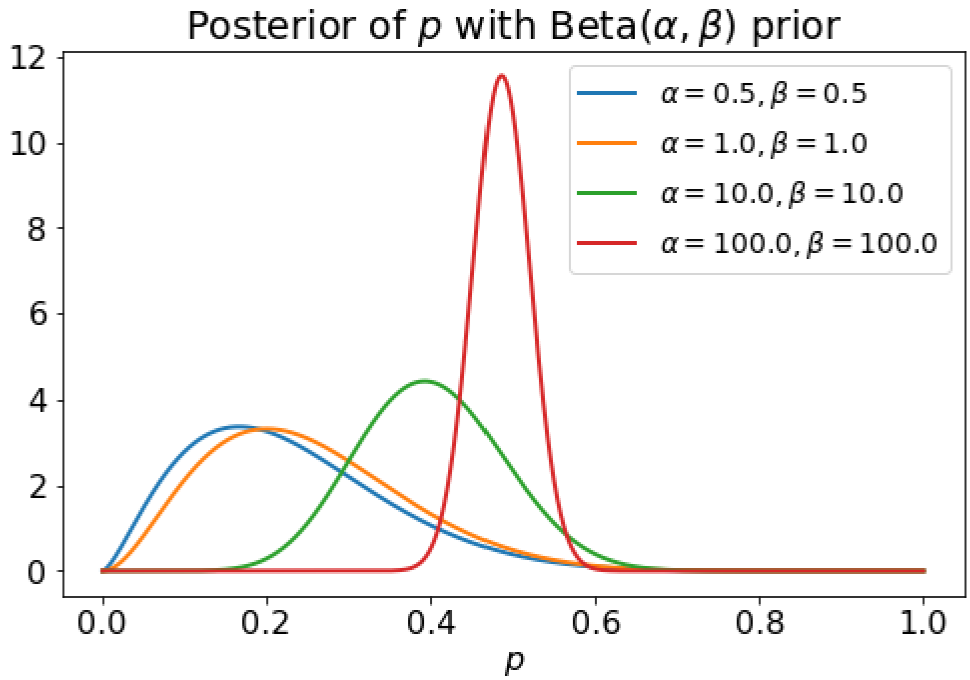
```
post_p1_samples_ratio = post_p1_samples / (1. - post_p1_samples)
post_p2_samples_ratio = post_p2_samples / (1. - post_p2_samples)
post_psi_samples = np.log(post_p1_samples_ratio / post_p2_samples_ratio)
post_psi_mean = np.mean(post_psi_samples)
sorted_post_psi_samples = np.sort(post_psi_samples)
post_psi_ci_90_lo = sorted_post_psi_samples[  index   ]
post_psi_ci_90_hi = sorted_post_psi_samples[-(index+1)]
```

5.

Let n be the number of trials k be the number of successes (in this case, 10 and 2). Then,

$$f(p \mid X^n) = \mathcal{L}(p) f(p) \propto p^k (1-p)^{n-k} p^{\alpha-1} (1-p)^{\beta-1}$$

and hence the prior is a conjugate prior. In particular, the posterior is a Beta distribution with $\bar{\alpha} = k + \alpha$ and $\bar{\beta} = n - k + \beta$.



6.

a)

The likelihood is

$$\mathcal{L}(\lambda) = \prod_i \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \propto \lambda^{\sum_i X_i} e^{-n\lambda}.$$

The prior is a Gamma distribution:

$$f(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}.$$

It follows that the posterior is also a Gamma distribution with parameters $\bar{\alpha} = \alpha + \sum_i X_i$ and $\bar{\beta} = \beta + n$. The posterior mean is $\bar{\alpha}/\bar{\beta}$.

b)

The Jeffreys prior is

$$f(\lambda) = I(\lambda)^{1/2} = \lambda^{-1/2}.$$

Combining this with the likelihood computed in Part (a), the posterior is a Gamma distribution with parameters $\bar{\alpha} = 1/2 + \sum_i X_i$ and $\bar{\beta} = n$.

7.

Note that

$$\mathbb{E}\hat{\psi} = \frac{1}{n}\sum_i \mathbb{E}\left[\frac{R_i Y_i}{\xi_{X_i}}\right] = \mathbb{E}\left[\frac{R_1 Y_1}{\xi_{X_1}}\right]$$

and

$$\mathbb{E}\left[\frac{R_1 Y_1}{\xi_{X_1}}\right] = \sum_j \frac{1}{\xi_j}\mathbb{E}\left[R_1|X_1=j\right]\mathbb{E}\left[Y_1|X_1=j\right]\mathbb{P}(X_1=j)$$
$$= \sum_j \mathbb{E}\left[Y_1|X_1=j\right]\mathbb{P}(X_1=j) = \mathbb{E}Y_1.$$

Therefore, $\mathbb{E}\hat{\psi} = \mathbb{E}Y_1$. Similarly,

$$\mathbb{V}(\hat{\psi}) = \frac{1}{n}\mathbb{V}\left(\frac{R_1 Y_1}{\xi_{X_1}}\right) = \frac{1}{n}\left(\mathbb{E}\left[\left(\frac{R_1 Y_1}{\xi_{X_1}}\right)^2\right] - \mathbb{E}[Y_1]^2\right)$$

and

$$\mathbb{E}\left[\left(\frac{R_1 Y_1}{\xi_{X_1}}\right)^2\right] = \sum_j \frac{1}{\xi_j^2}\mathbb{E}\left[R_1^2 \mid X_1=j\right]\mathbb{E}\left[Y_1^2 \mid X_1=j\right]\mathbb{P}(X_1=j)$$
$$= \sum_j \frac{1}{\xi_j}\mathbb{E}\left[Y_1 \mid X_1=j\right]\mathbb{P}(X_1=j) \le \frac{1}{\delta}\mathbb{E}\left[Y_1\right].$$

Therefore,

$$\mathbb{V}(\hat{\psi}) \le \frac{1}{n}\left(\frac{1}{\delta}\mathbb{E}\left[Y_1\right] - \mathbb{E}[Y_1]^2\right) \le \frac{1}{n\delta}.$$

8.

Let $X_1, \ldots, X_n \sim N(\mu, 1)$. The MLE is $\hat{\mu} = n^{-1}\sum_i X_i$ with standard error $\text{se}(\hat{\mu}) = 1/\sqrt{n}$. Therefore, the Wald statistic is $W = \sqrt{n}\hat{\mu}$, with p-value $2\Phi(-\sqrt{n}|\hat{\mu}|)$. Clearly, if $\mu \neq 0$, then $\sqrt{n}\hat{\mu}$ diverges and hence the p-value converges to zero, correctly rejecting the null.

Next, note that

$$\mathcal{L}(\mu) f_{H_1}(\mu) = \frac{1}{b}\left(\frac{1}{\sqrt{2\pi}}\right)^{n+1}\exp\left\{-\frac{1}{2}\left[\frac{1}{b^2}\mu^2 + \sum_i (\mu - X_i)^2\right]\right\}.$$

Let

$$\sigma^2 = \frac{b^2}{1+nb^2}$$

Then,

$$\begin{aligned}
\frac{1}{b^2}\mu^2 + \sum_i (\mu - X_i)^2 &= \left(\frac{1}{b^2} + n\right)\mu^2 - 2\mu \sum_i X_i + \sum_i X_i^2 \\
&= \frac{1}{\sigma^2}\mu^2 - 2\mu \sum_i X_i + \sum_i X_i^2 \\
&= \frac{1}{\sigma^2}\left(\mu^2 - 2\sigma^2 \mu \sum_i X_i\right) + \sum_i X_i^2 \\
&= \frac{1}{\sigma^2}\left(\mu - \sigma^2 \sum_i X_i\right)^2 + \sum_i X_i^2 - \left(\sigma \sum_i X_i\right)^2.
\end{aligned}$$

Therefore,

$$\mathcal{L}(\mu) f_{H_1}(\mu) = C \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(\mu - \sigma^2 \sum_i X_i\right)^2\right\}.$$

where

$$c = \frac{\sigma}{b}\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{\frac{1}{2}\left(\sigma \sum_i X_i\right)^2 - \frac{1}{2}\sum_i X_i^2\right\}.$$

In particular, $\int \mathcal{L}(\mu) f_{H_1}(\mu) d\mu = C$. By the derivation in Section 11.8,

$$\mathbb{P}(H_0 \mid X^n = x^n) = \frac{\mathcal{L}(0)}{\mathcal{L}(0) + c} = \frac{1}{1 + c/\mathcal{L}(0)}.$$

Moreover,

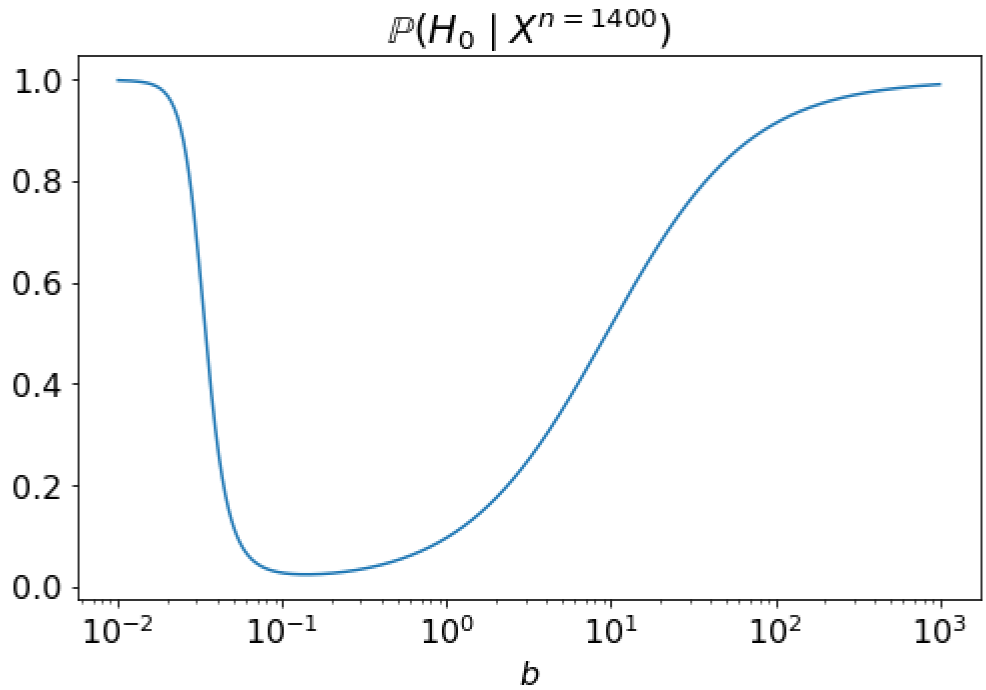
$$\mathcal{L}(0) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\sum_i X_i^2\right)$$

and hence

$$\frac{c}{\mathcal{L}(0)} = \frac{\sigma}{b}\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \exp\left\{\frac{1}{2}\left(\sigma \sum_i X_i\right)^2\right\}.$$

If $\mu \neq 0$, then the $c/\mathcal{L}(0)$ diverges and hence $\mathbb{P}(H_0 \mid X^n = x^n)$ converges to zero, correctly rejecting the null.

However, for finite n, the two tests can disagree. For example, using $n = 1400$ samples and a true mean of $\mu = 0.1$, an extremely small Wald test p-value of approximately 1.05×10^{-7} is observed. On the other hand, $\mathbb{P}(H_0 \mid X^n)$ can be close to one for particular choices of b.



Code to compute the p-value and generate the above plot is given below.

```
mu = 0.1
n_sims = 1000

np.random.seed(1)
samples = np.random.randn(n_sims) + mu

# Wald test
mle_mu = np.mean(samples)
wald = np.sqrt(n_sims) * mle_mu
p_value = 2. * scipy.stats.norm.cdf(-np.abs(wald))

# Bayesian test
b = np.linspace(1e-2, 1000., 10**6)
sigma2 = b**2 / (1. + n_sims * b**2)
a = np.sqrt(sigma2) / b * (2 * np.pi)**(-n/2. + 0.5) * np.exp(
    0.5 * sigma2 * np.sum(samples)**2)
post_prob_null = 1. / (1. + a)
plt.figure(figsize=(1.618 * 5., 5.))
plt.semilogx(b, post_prob_null)
plt.title('$\\mathbb(H_0 \\mid X)$'.format(n_sims))
plt.xlabel('$b$')
```