Ex. 2.11.1: Posterior inference: suppose you have a Beta(4, 4) prior distribution on the probability θ that a coin will yield a 'head' when spun in a specified manner. The coin is independently spun ten times, and 'heads' appear fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3. Calculate your exact posterior density (up to a proportionality constant) for θ and sketch it.

Answer:

We are asked

$$\begin{aligned} p(\theta|y<3) &= p(\theta|(y=0)+(y=1)+(y=2)) \\ &\propto p((y=0)+(y=1)+(y=2)|\theta)p(\theta) \\ &= (p(y=0|\theta)+p(y=1|\theta)+p(y=2|\theta))p(\theta). \end{aligned}$$

Each term separately will yield $\mathrm{Beta}(4+y,\,4+(10-y))$, so that the final probability distribution is plainly

$$\frac{1}{3}\left[\mathrm{Beta}(4,\,14)+\mathrm{Beta}(5,\,13)+\mathrm{Beta}(6,\,12)\right].$$

Those have means 4/18, 5/18, 6/18, so that the mean is 5/18. More concretely, if

$$p(x)=\sum_i w_i p_i(x),$$

then the expectation of stuff is given by

$$\begin{aligned} \mathrm{E}\cdot &= \int \mathrm{d}x\; p(x)\cdot \\ &= \sum_i w_i \int \mathrm{d}x\; p_i(x)\cdot \\ &= \sum_i w_i \mathrm{E}_i \cdot . \end{aligned}$$

It then follows that

$$\begin{aligned} \mathrm{E}\theta &= \frac{1}{3}\left[\mathrm{E}_1\theta+\mathrm{E}_2\theta+\mathrm{E}_3\theta\right] \\ &= \frac{5}{18}. \end{aligned}$$

The same proccess can yield the new variance.

The corresponding computer graph is shown in Figure **??**.

Ex. 2.11.2: Predictive distributions: consider two coins, C_1 and C_2, with the following characteristics: $\Pr(\text{heads}|C_1)=0.6$ and $\Pr(\text{heads}|C_2)=0.4$. Choose one of the coins at random and imagine spinning it repeatedly. Given that the first two spins from the chosen coin are tails, what is the expectation of the number of additional spins until a head shows up?

Answer:

Here I'll represent the data by D.

The conditional probability that we make n tosses until heads is found is

$$p(n|C_i)=(\Pr(\text{tails}|C_i))^{n-1}\Pr(\text{heads}|C_i),$$

so that

$$\begin{aligned} p(n|D) &= \sum_i p(n,\,C_i|D) \\ &= \sum_i p(n|C_i)\Pr(C_i|D). \end{aligned}$$

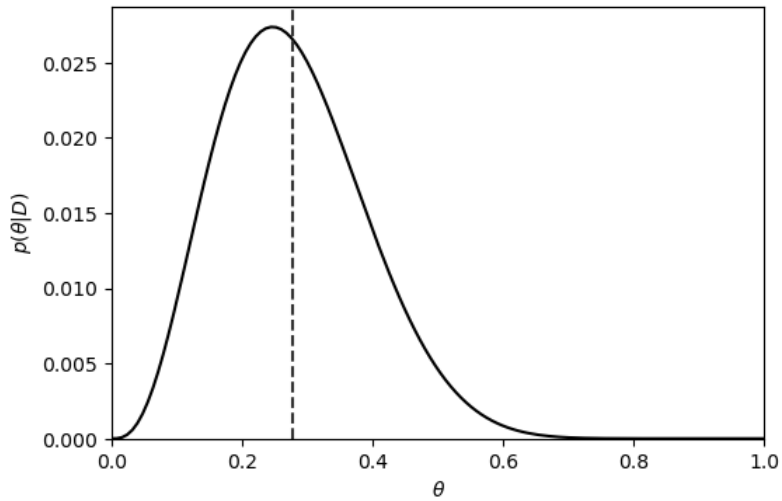

Figure 1: Probability distribution for θ given the number of heads is less than 3.

Since we have no information as to which coin we picked, we'll take $\Pr(C_i) = \frac{1}{2}$ which, together with $\Pr(D|C_i) = (\Pr(\text{tails}|C_i))^2$ yield $\Pr(C_i|D) = (\Pr(\text{tails}|C_i))^2 \,/\, \left[(\Pr(\text{tails}|C_i))^2 + (\Pr(\text{tails}|C_{\bar{i}}))^2\right]$, where $\bar{i}$ is not i. Numerically, this yields $P(C_2|D) = 0.69$, nice.

Anyways, we also have $p(n|C_i) = (\Pr(\text{tails}|C_i))^{n-1}\Pr(\text{heads}|C_i)$, so we get our final result of

$$p(n|D) = \frac{(\Pr(\text{tails}|C_1))^{n+1}\Pr(\text{heads}|C_1) + (\Pr(\text{tails}|C_2))^{n+1}\Pr(\text{heads}|C_2)}{(\Pr(\text{tails}|C_1))^2 + (\Pr(\text{tails}|C_2))^2}.$$

Our final calculation is then $\mathrm{E}[n] = \sum_{n=1} np(n)$, which can be done analitically just fine but I wont, and yields about 2.24.

Ex. 2.11.3: Predictive distributions: let y be the number of 6's in 1000 rolls of a fair die.

a Sketch the approximate distribution of y, based on the normal approximation.

b Using the normal distribution table, give approximate 5%, 25%, 50%, 75% and 95% points for the distribution of y.

Answer:

We have $y \sim \mathrm{B}(1000, 1/6)$, with mean $166.\overline{6}$ and deviation 11.8. So I should draw a bell curve which should have a width of 23.6 at 0.6 its height, centered in 166.6. As for the table thingy, the following Python interpreter run has been performed

```
In [1]: import scipy.stats as st

In [2]: [ 166.6+11.8*st.norm.ppf(x) for x in [0.05, 1/4, 1/2, 3/4, 0.95] ]
Out[2]:
[147.1907272019726,
 158.64102094768623,
 166.6,
 174.55897905231376,
 186.00927279802738]
```

Ex. 2.11.4: Predictive distributions: let y be the number of 6's in 1000 independent rolls of a particular real die, which may be unfair. Let θ be the probability that the die lands on '6'.

Suppose your prior distribution for θ is as follows:

$$\begin{aligned}\Pr(\theta = 1/12) &= 0.25\\ \Pr(\theta = 1/6) &= 0.5\\ \Pr(\theta = 1/4) &= 0.25.\end{aligned}$$

a Using the normal approximation for the conditional distributions, $p(y|\theta)$, sketch your approximate prior predictive distribution for y.

b Give approximate 5%, 25%, 50%, 75%, and 95% points for the distribution of y. (Be careful here: y does not have a normal distribution, but you can still use the normal distribution as part of your analysis.)

Answer:

Again we have $p(y) = \sum_\theta p(y|\theta)p(\theta)$. Under the normal approximation, $y|\theta \sim \mathcal{N}(n\theta,\, n\theta(1-\theta))$, so we'll have

$$y|\theta \sim 0.25 \cdot N(83.3,\, 8.74^2) + 0.5 \cdot N(166.7,\, 11.8^2) + 0.25 \cdot N(250, 13.7),$$

which is a pretty disjoint (and thus easily drawable) mixture if you ask me.

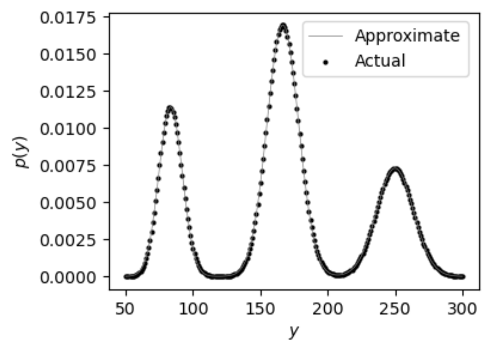

Figure 2: Actual and normally-approximated predictive prior distribution for y

As for the approximate points, we can essentially regard the distributions as disjoint, so that the 5% point is the 20% of the first one, the 25% can be approximated as, say, the middle point between the first two Gaussians, the 50% point as the mean 166.7 of the middle Gaussian, and so on. So, *a ojo*, they should be about 70, 120, 167, 210, and 270. But I shall put this handwavery to the test via numerical integration, which is performed in the script '2.11.4.py'. For the real values it yields 76, 120, 167, 207, and 261. For the normal-plus-disjointness-approximated values we get 76, 125, 167, 208, and 262.

Ex. 2.11.5: Posterior distribution as a compromise between prior information and data: let y be the number of heads in n spins of a coin, whose probability of heads is θ.

a If your prior distribution for θ is uniform on the range [0, 1], derive your prior predictive distribution for y,

$$\Pr(y = k) = \int_0^1 \Pr(y = k|\theta)\mathrm{d}\theta,$$

for each $k = 0, 1, \ldots, n$.

b Suppose you assign a Beta(α, β) prior distribution for θ, and then you observe y heads out of n spins. Show algebraically that your posterior mean of θ always lies between your prior mean, $\frac{\alpha}{\alpha+\beta}$, and the observed relative frequency of heads, $\frac{y}{n}$.

c Show that, if the prior distribution on θ is uniform, the posterior variance of θ is always less than the prior variance.

d Give an example of a Beta(α, β) prior distribution and data y, n, in which the posterior variance of θ is higher than the prior variance.

Answer:

Point a:

$$
\begin{aligned}
\Pr(y=k) &= \int_0^1 \mathrm{d}\theta \;\Pr(y=k|\theta) \\
&= \int_0^1 \mathrm{d}\theta \;\binom{n}{k}\theta^k(1-\theta)^{n-k} \\
&= \binom{n}{k}\int_0^1 \mathrm{d}\theta \;\theta^k(1-\theta)^{n-k} \\
&= \binom{n}{k}\left\{\cancelto{0}{\left[\frac{1}{k+1}\theta^{k+1}(1-\theta)^{n-k}\right]_0^1} + \frac{n-k}{k+1}\int_0^1 \mathrm{d}\theta\;\theta^{k+1}(1-\theta)^{n-k-1}\right\} \\
&= \binom{n}{k}\frac{n-k}{k+1}\frac{n-k-1}{k+2}\cdots\frac{1}{n}\int_0^1 \mathrm{d}\theta\;\theta^n \\
&= \binom{n}{k}\frac{k!(n-k)!}{(n+1)!} \\
&= \frac{1}{n+1}.
\end{aligned}
$$

Point b:

We've seen in the book that we now have $\theta|y \sim \mathrm{Beta}(y+\alpha,\, n-y+\beta)$, which has expectation value

$$
\begin{aligned}
\frac{y+\alpha}{n+\alpha+\beta} &= \frac{y}{n}\frac{n}{n+\alpha+\beta} + \frac{\alpha}{\alpha+\beta}\frac{\alpha+\beta}{n+\alpha+\beta} \\
&= w\frac{y}{n} + (1-w)\frac{\alpha}{\alpha+\beta} \quad \left(w = \frac{n}{n+\alpha+\beta} \in [0,\,1]\right).
\end{aligned}
$$

This is a linear function of w which goes from $\frac{y}{n}$ to $\frac{\alpha}{\alpha+\beta}$ as w goes from 0 to 1, thus proving the result.

Point c:

Since $\theta \sim 1$, its variance is

$$
\begin{aligned}
\int_0^1 \mathrm{d}\theta \;\left(\theta - \frac{1}{2}\right)^2 &= \frac{1}{3}\left[\left(\theta-\frac{1}{2}\right)^3\right]_0^1 \\
&= \frac{1}{12}.
\end{aligned}
$$

Now, $\theta|y \sim \mathrm{Beta}(y+1,\, n-y+1)$, which has variance

$$
\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}.
$$

As a function of y, this is proportional to the denominator $-y^2+ny+n+1$, whose maximum is at $\frac{1}{2}n$ and yields $\frac{1}{4}(2n+1)$, so that

$$
\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} \le \frac{1}{4}\frac{(2n+1)^2}{(n+2)^2\,(n+3)}.
$$

This function can be maximized analitically via derivation, but I just plotted it. It's maximum is around 3, taking the value $49/600 < 50/600 = 1/12$. Such a close call though.

As for point c, take Beta(3, 1), yielding variance 0.0375, and suppose one negative outcome is measured, taking the posterior distribution towards Beta(3, 2) with variance 0.04.

Ex. 2.11.6: Predictive distributions: Derive the mean and variance (2.17) of the negative binomial predictive distribution for the cancer rate example, using the mean and variance formulas (1.8) and (1.9).

Answer:

Formulas (1.8) and (1.9) are

$$\begin{aligned}
\mathrm{E}(u) &= \mathrm{E}(\mathrm{E}(u|v)), \\
\mathrm{var}(u) &= \mathrm{E}(\mathrm{var}(u|v)) + \mathrm{var}(\mathrm{E}(u|v)).
\end{aligned}$$

We're thus asked the mean and variance of the distribution $p(y) = \int \mathrm{d}\theta \; p(y|\theta)p(\theta)$, where $\theta \sim \mathrm{Gamma}(\alpha,\, \beta)$ and $y|\theta \sim \mathrm{Poisson}(10n\theta)$. Direct application follows:

$$\begin{aligned}
\mathrm{E}(y) &= \mathrm{E}(\mathrm{E}(y|\theta)) \\
&= \mathrm{E}(10n\theta) \\
&= 10n_j\mathrm{E}(\theta) \\
&= 10n\frac{\alpha}{\beta}. \\
\mathrm{var}(y) &= \mathrm{E}(\mathrm{var}(y|\theta)) + \mathrm{var}(\mathrm{E}(y|\theta)) \\
&= \mathrm{E}(10n\theta) + \mathrm{var}(10n\theta) \\
&= 10n\frac{\alpha}{\beta} + (10n)^2\frac{\alpha}{\beta^2},
\end{aligned}$$

where the mean and variance of the Gamma and Poisson have been Googled and used straight away.

Ex. 2.11.7: Noninformative prior densities:

a For the binomial likelihood, $y \sim \mathrm{Bin}(n,\, \theta)$, show that $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is the uniform prior distribution for the natural parameter of the exponential family.

b Show that if $y = 0$ or n, the resulting posterior distribution is improper.

Answer:

We can express the binomial likelihood in exponential form as follows

$$\begin{aligned}
\mathrm{Bin}(y|n,\, \theta) &= \binom{n}{y}\theta^y(1-\theta)^{n-y} \\
&= \binom{n}{y}(1-\theta)^n\left(\frac{\theta}{1-\theta}\right)^y \\
&= \binom{n}{y}(1-\theta)^n\exp\left(y\log\left(\frac{\theta}{1-\theta}\right)\right).
\end{aligned}$$

It then becomes clear that the natural parameter is $\phi = \log\left(\frac{\theta}{1-\theta}\right)$. Since the transformation between ϕ and θ is one-to-one,

$$\begin{aligned}
p(\phi) &= p(\theta)\left|\frac{\mathrm{d}\phi}{\mathrm{d}\theta}\right|^{-1} \\
&= \theta^{-1}(1-\theta)^{-1}\left|\frac{\theta}{1-\theta}(1-\theta)^2\right| \\
&= 1.
\end{aligned}$$

As for the posterior distribution, we have

$$\begin{aligned}
p(\theta|y) &\propto p(y|\theta)p(\theta) \\
&\propto \theta^{y-1}(1-\theta)^{n-y-1}.
\end{aligned}$$

Whether $y = 0$ or $y = n$ we have something of the form $\eta^{-1}(1-\eta)^{n-1}$, η being either θ or $1-\theta$, so that the integral of this quantity over η is precisely the normalization required, but it yields

$$\begin{aligned}\int_0^1 \mathrm{d}\eta\, \eta^{-1}(1-\eta)^{n-1} &= \int_0^1 \mathrm{d}\eta \sum_{k=0}^{n-1} \binom{n-1}{k} \eta^{-1}\eta^k \\ &= \sum_{k=0}^{n-1} \binom{n-1}{k} \int_0^1 \mathrm{d}\eta\, \eta^{k-1} \\ &= \int_0^1 \mathrm{d}\eta\, \eta^{-1} + \sum_{k=1}^{n-1} \binom{n-1}{k} \int_0^1 \mathrm{d}\eta\, \eta^{k-1} \\ &= \infty + \text{something finite,}\end{aligned}$$

whence this integral is improper.

Ex. 2.11.8: Normal distribution with unknown mean: a random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose your prior distribution for θ is normal with mean 180 and standard deviation 40.

a Give your posterior distribution for θ. (Your answer will be a function of n.)

b A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$. (Your answer will still be a function of n.)

c For $n = 10$, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$.

d Do the same for $n = 100$.

Answer:

By hypothesis, $y|\theta \sim \mathrm{N}(\theta,\, 20^2)$ and $\theta \sim \mathrm{N}(180,\, 40^2)$, so that $\bar{y}|\theta \sim \mathrm{N}(\theta,\, 20^2/n)$, and thus $\theta|\bar{y}$ is gonna be given by a normal distribution with

$$\begin{aligned}\mathrm{E}(\theta|\bar{y}) &= \frac{40^{-2}180 + n20^{-2}150}{40^{-2} + n20^{-2}}, \\ \mathrm{var}(\theta|\bar{y}) &= \frac{1}{40^{-2} + n20^{-2}}.\end{aligned}$$

I could put some numbers there, but it'd be no use. I instead ploted probability desntiy for $\theta|\bar{y}$ (vertical axis) as a function of sample size n (horizontal axis) for n between 0 and 100, as shown in Figure **??**.

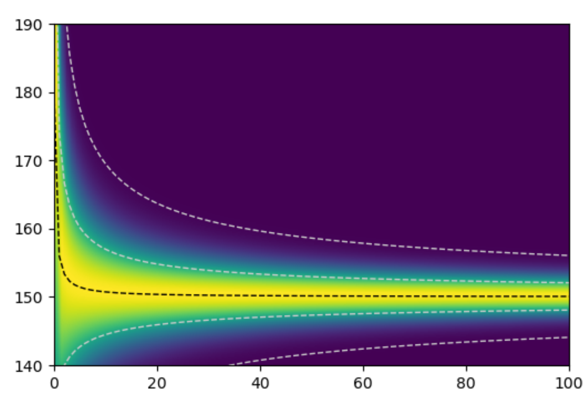

Figure 3: Non-normalized probability distribution for $\theta|\bar{y}$ as a function of θ and number of samples n. Black contour is the mean, and grey contours are 1 and 3 standard deviations from it.

Since the actual precision of the data is so small, the mean converges quickly towards the observed mean.

The posterior predictive distribution is the same, but incorporates an extra 20 in the standard deviation. In particular, for $n = 10$, the posterior probabilities for θ and $\tilde{y}$ have mean 150.73 and respectively a standard deviation of 6.24 and 26.24, so that the $\sim$ 95% intervals are 150.73 ± 12.48 and 150.73 ± 42.48. For $n = 100$, the mean becomes 150.07 and the standard deviations 1.99 and 20.199, so that the intervals are 150.07 ± 4 and 150.07 ± 48.

Ex. 2.11.9: Setting parameters for a beta prior distribution: suppose your prior distribution for θ, the proportion of Californians who support the death penalty is beta with mean 0.6 and standard deviation 0.3.

a Determine the parameters α and β of your prior distribution. Sketch the prior density function.

b A random sample of 1000 Californians is taken, and 65% support the death penalty. What are your posterior mean and variance for θ? Draw the posterior density function.

c Examine the sensitivity of the posterior distribution to different prior means and widths including a non-informative prior.

Answer:

We equate the mean and the variance

$$\frac{\alpha}{\alpha+\beta} = \mu,$$
$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \sigma^2.$$

Replacing $\alpha/(\alpha+\beta) = \mu$ and $\beta/(\alpha+\beta) = 1-\mu$ in the second equation yields

$$\alpha + \beta = \frac{\mu\,(1-\mu)}{\sigma^2} - 1,$$

which into the first equation implies

$$\alpha = \mu\left(\frac{\mu\,(1-\mu)}{\sigma^2} - 1\right),$$
$$\beta = (1-\mu)\left(\frac{\mu\,(1-\mu)}{\sigma^2} - 1\right).$$

Upon replacing the particular values $\mu = 0.6$ and $\sigma = 0.3$, we get

$$\alpha = 1, \quad \beta = \frac{2}{3}.$$

Now, after the sensus we get 650 positives and 350 negatives, whence we'll have $\theta|\mathrm{D} \sim \mathrm{Beta}(650+1, 350+\frac{2}{3})$, with mean 65.0% and standard deviation 1.4%.

The following table, calculated in the corresponding script to this exercise, evaluates a few alternatives for σ, ranging from *whoa* prior certainty to complete prior uncertainty.

Prior deviation	0.1%	0.5%	1.0%	5.0%	10.0%	50.0%	100%	Noninformative
Posterior mean	60.02%	60.47%	61.47%	64.57%	64.89%	65.00%	65.00%	64.97%
Posterior deviation	0.10%	0.47%	0.83%	1.44%	1.49%	1.51%	1.51%	1.51%

The noninformative alternative diminishes the posterior mean because it's centered around 50% instead of 60%, so that it's "pulling effect" is more evident.

Ex. 2.11.10: Discrete sample spaces: suppose there are N cable cars in San Francisco, numbered sequentially from 1 to N. You see a cable car at random, it is numbered 203. You wish to estimate N.

a Assume your prior distribution on N is geometric with mean 100; that is,

$$p(N) = (1/100)(99/100)^{N-1}, \quad \text{for } N = 1, 2, ...$$

What is your posterior distribution for N?

b What are the posterior mean and standard deviation of N?

c Choose a reasonable 'noninformative' prior distribution for N and give the resulting posterior distribution, mean, and standard deviation for N.

Answer:

Let $q = 99/100$. My first instinct was proposing an homogeneous sampling distribution for the number y of the observed cable car, $p(y|N) = 1/N$. But this conveys no information on the relation between y and N, and thus serves no good.

Ex. 2.11.11: Computing with a nonconjugate single parameter model: suppose $y_1, \ldots, y_5$ are independent samples from a Cauchy distribution with unknown center θ and known scale 1: $p(y_i|\theta) \propto 1/(1 + (y_i - \theta)^2)$. Assume for simplicity that the prior distribution for θ is uniform on [0, 100]. Given the observations $(y_1, \ldots, y_5) =$ (43, 44, 45, 46.5, 47.5):

a Compute the unnormalized posterior density function, $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 100$, for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta|y)$, as a function of θ.

b Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.

c Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the predictive draws.

Answer:

The following is an excerpt from the script "2.11.11.py" that does the math before plotting

```python
import numpy as np

def normalized(arr):
    return arr/arr.sum()

m = 2**12 # This is unnecessarily big I guess

y = np.array([43, 44, 45, 46.5, 47.5])
theta = np.linspace(0, 100, m+1)
b_y, b_theta = np.meshgrid(y, theta) # "Big y, big theta"

theta_updf = 1
y_gvn_theta_updf = 1/np.prod(1+(b_y-b_theta)**2, axis=1)
# item a:
theta_gvn_y_pdf = normalized(theta_updf*y_gvn_theta_updf)

np.random.seed(42)

# item b:

def draw_samples(vals, pdf, size=1):
    cdf = np.cumsum(pdf)
    udraws = np.random.uniform(size=size)
    b_cdf, b_udraws = np.meshgrid(cdf, udraws)
    ids = np.argmax(b_cdf > b_udraws, axis=1)
    return vals[ids]

theta_gvn_y_samples = draw_samples(theta, theta_gvn_y_pdf, size=1000)

# item c:
y_pred_samples = np.random.standard_cauchy(size=1000)+theta_gvn_y_samples
```

From these data, Figure **??** is constructed.

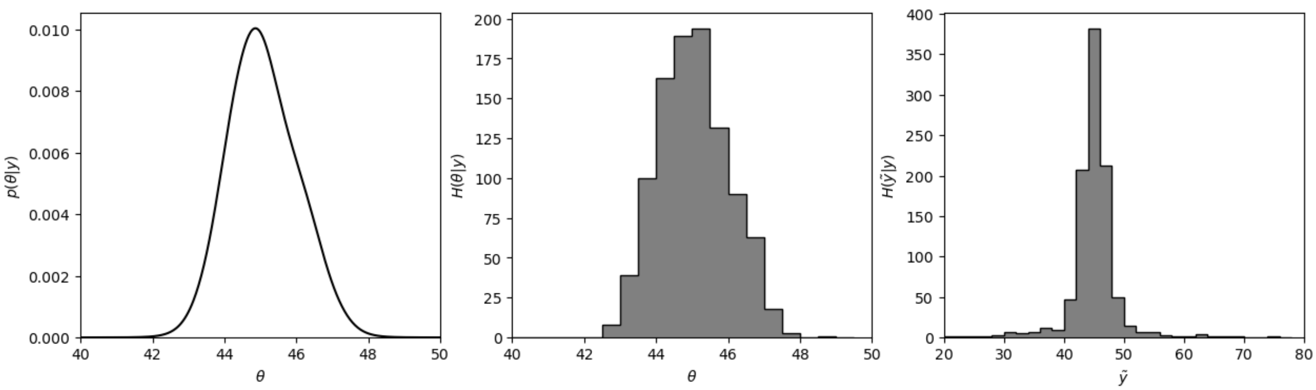

Figure 4: Binned approximation to the posterior distribution for θ (a), histogram from such a distribution (b), and histogram for the posterior predictive distribution for $y_6 = \tilde{y}$ (c).

Ex. 2.11.12: Jeffreys' prior distributions: suppose $y|\theta \sim \text{Poisson}(\theta)$. Find Jeffreys' prior density for θ, and then find α and β for which the Gamma(α, β) density is a close match to Jeffreys' density.

Answer:

Jeffreys prescribes

$$p(\theta) \propto [J(\theta)]^{\frac{1}{2}},$$

where $J(\theta)$ is the Fisher Information for θ:

$$J(\theta) = \mathrm{E}\left((\partial_\theta \log p(y|\theta))^2 \middle| \theta \right) = -\mathrm{E}\left((\partial_\theta)^2 \log p(y|\theta) \middle| \theta \right).$$

Direct calculation follows:

$$\begin{aligned} p(y|\theta) &= \text{Poisson}(\theta) \\ &= \frac{\theta^y \mathrm{e}^{-y}}{y!}, \\ \log p(y|\theta) &= y \log \theta + f(y), \\ \partial_\theta \log p(y|\theta) &= \frac{y}{\theta}, \\ J(\theta) &= \frac{1}{\theta^2} \mathrm{E}\left(y^2\right). \end{aligned}$$

It thus follows that Jeffreys' prior is the improper distribution $p(\theta) \propto 1/\theta$.

Since

$$\text{Gamma}(\theta|\alpha,\, \beta) \propto \theta^{\alpha-1} \mathrm{e}^{-\beta\theta},$$

it follows that $\alpha \to 0^+$ and $\beta \to 0^+$ should approach this distribution.

Ex. 2.11.13: Discrete data: Table **??** gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period. We use these data as a numerical example for fitting discrete data models.

a Assume that the numbers of fatal accidents in each year are independent with a Poisson(θ) distribution. Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 through 1985. under this model, give a 95% predictive interval for the number of fatal accidents in 1986. You can use the normal approximation to the gamma and Poisson or compute using simulation.

b Assume that the numbers of fatal accidents in each year are independent with a Poisson(θ) distributions with a constant rate and an exposure in each year proportional to the number of passenger miles flown.

Year	Fatal accidents	Passenger deaths	Death rate
1976	24	734	0.19
1977	25	516	0.12
1978	31	754	0.15
1979	31	877	0.16
1980	22	814	0.14
1981	21	362	0.06
1982	26	764	0.15
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

Table 1: *Worldwide airline fatalities, 1976–1985. Death rate is passenger deaths per 100 million passenger miles.* *Source:* Statistical Abstract of the United States.

Set a prior distribution for θ and determine the posterior distribution based on the data for $1976 - -1985$. (Estimate the number of passenger miles flown in each year by dividing the appropriate columns of Table **??** and ignoring round-off errors.) Give a 95% predictive interval for the number of fatal accidents in 1986 under the assumption that 8×10^{11} passenger miles are flown that year.

c Repeat (a) above, replacing 'fatal accidents' with 'passenger deaths.'

d Repeat (b) above, replacing 'fatal accidents' with 'passenger deaths.'

e In which of the cases (a)–(d) above does the Poisson model seem more or less reasonable? Why? Discuss based on general principles, without sepecific reference to the numbers in Table **??**

Incidentally, in 1986, there were 22 fatal accidents, 546 passenger deaths, and a death rate of 0.06 per 100 million miles flown. We return to this example in Exercises 3.12, 6.2, 6.3, and 8.14.