Chapter 5: Solutions to Exercises

Exercise 1

To find the inverse of the c.d.f., set $u = F(x \mid c, \beta)$ and solve for x:

$$\begin{aligned} u &= \exp\Big(-e^{-(x-c)/\beta}\Big) \\ -\log u &= e^{-(x-c)/\beta} \\ -\log(\log 1/u) &= (x-c)/\beta \\ -\beta\log(\log 1/u) + c &= x. \end{aligned}$$

Thus, letting

$$G(u) = -\beta\log(\log 1/u) + c,$$

we have that $G(U) \sim \text{Gumbel}(c, \beta)$ if $U \sim \text{Uniform}(0, 1)$.

Exercise 2

See Figure 1. See Appendix for source code. Roughly speaking, by inspecting the plots, the approximations seem to be failing to converge—they start approaching 0, but then get "kicked out" again. The issue is that every once in a while, a value of X_i occurs that is so astronomically large compared to all previous values, that it significantly shifts the sample average.

Exercise 3

Part (a)

First, note that, abbreviating $x = x_{1:n}$,

$$\begin{aligned} \mathbb{E}\Big(\frac{1}{p(x|\boldsymbol{\theta})}\Big| x\Big) &= \int \frac{1}{p(x|\theta)} p(\theta|x)d\theta \\ &= \int \frac{1}{p(x|\theta)} \frac{p(x|\theta)p(\theta)}{p(x)} d\theta \\ &= \frac{1}{p(x)} \int p(\theta)d\theta = \frac{1}{p(x)} \end{aligned}$$

by Bayes' theorem, and using that $p(x_{1:n}|\theta) > 0$ for all θ to justify the cancellation.

Jeffrey W. Miller (2015). *Lecture Notes on Bayesian Statistics.* Duke University, Durham, NC.

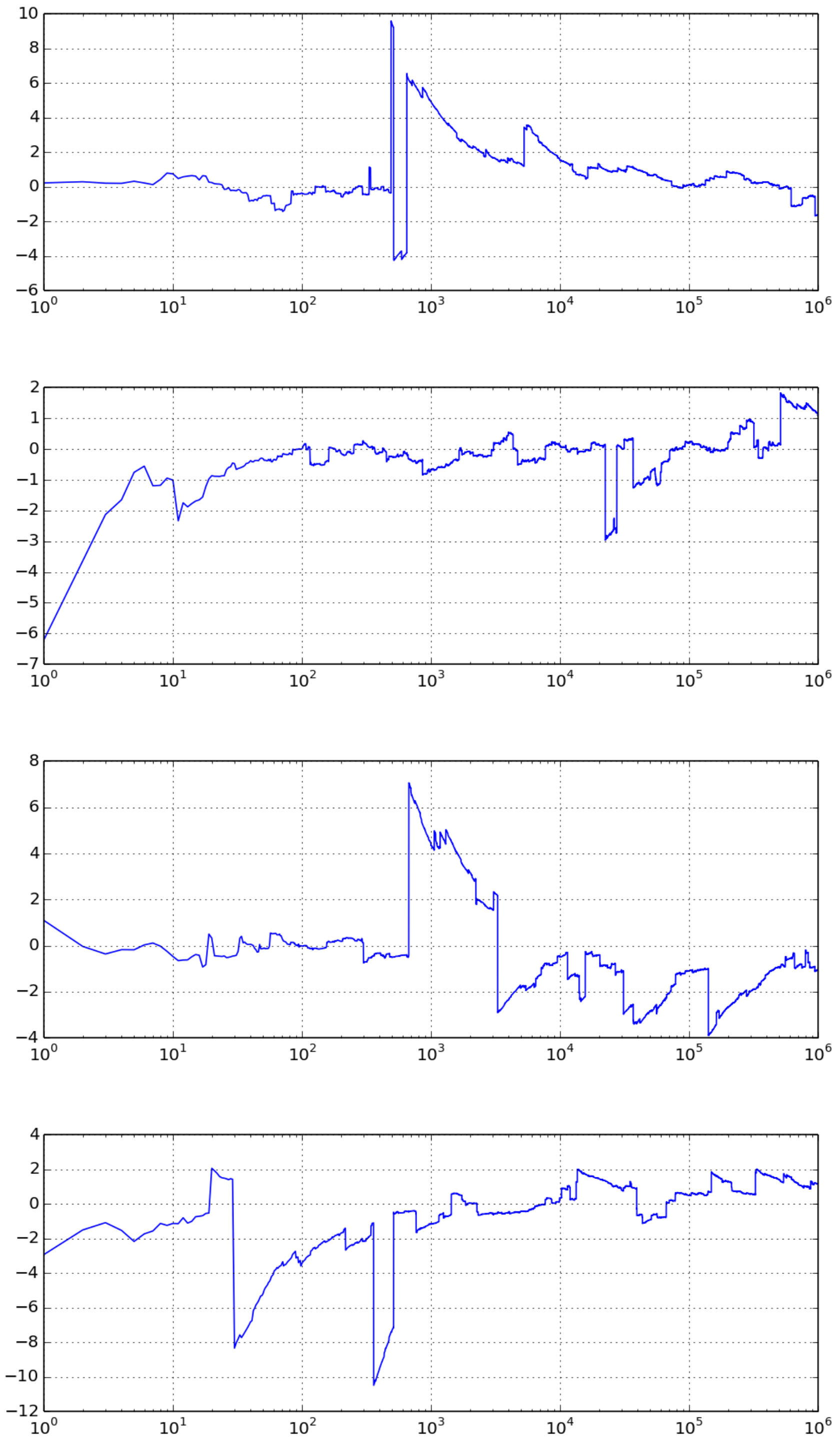

Figure 1: Lack of convergence of Monte Carlo approximations for Cauchy samples.

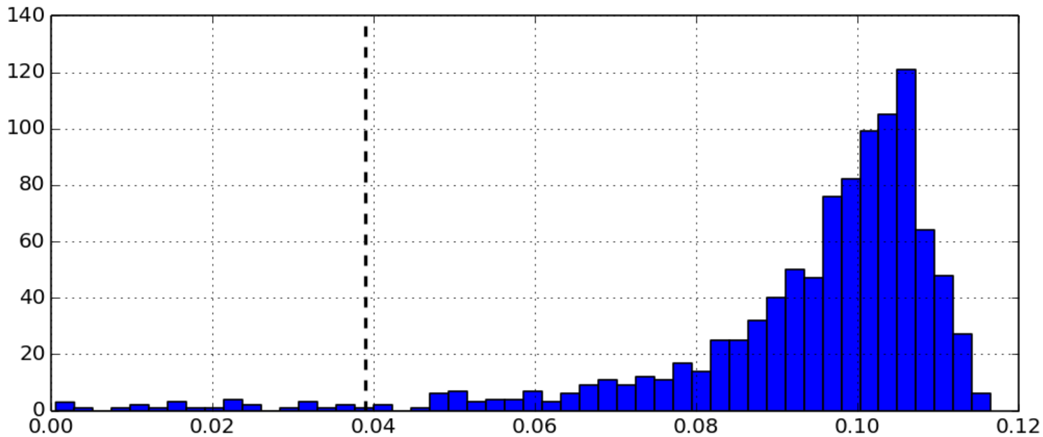

Figure 2: Histogram of harmonic mean approximations, compared with the true value (dotted line).

Since this is finite, then letting $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N \overset{\text{iid}}{\sim} p(\theta|x)$, we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{p(x|\boldsymbol{\theta}_i)} \longrightarrow \mathbb{E}\Big(\frac{1}{p(x|\boldsymbol{\theta})}\Big| x\Big) = \frac{1}{p(x)}$$

as $N \to \infty$, with probability 1, by the law of large numbers. Therefore,

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{p(x|\boldsymbol{\theta}_i)}\Big)^{-1} \longrightarrow p(x)$$

as $N \to \infty$, with probability 1.

Part (b)

See Appendix for source code. Here are the values of the harmonic mean approximations for 5 independent sets of samples:

$$0.109 \quad 0.097 \quad 0.106 \quad 0.100 \quad 0.094$$

Based on these, we might expect the true value of the marginal likelihood to be somewhere between 0.08 and 0.12. However, the true value is actually

$$\mathcal{N}(2 \mid 0, \lambda^{-1} + \lambda_0^{-1}) = \mathcal{N}(2 \mid 0, 1 + 10^2) = 0.0389.$$

The harmonic mean approximations are quite far away from the true value.

Figure 2 shows a histogram using 1000 independent sets of samples, compared with the true value. The vast majority of the time, the harmonic mean is considerably higher than the true value.

Part (c)

Using $\lambda_0 = 1/100^2$, the harmonic mean gives us

$$0.094 \quad 0.109 \quad 0.048 \quad 0.095 \quad 0.097$$

while the true value is 0.00399. The harmonic mean is now even more misleading.

Exercise 4

See Appendix for source code. See module text for plot and discussion.

Exercise 5

Let $S \subset A$. We have $Z \in S$ if and only if one of the following events occurs:

$$\begin{aligned}
&\{X_1 \in S\} \\
&\{X_1 \notin A, X_2 \in S\} \\
&\{X_1 \notin A, X_2 \notin A, X_3 \in S\} \\
&\vdots
\end{aligned}$$

and these events are mutually exclusive. Hence,

$$\begin{aligned}
\mathbb{P}(Z \in S) &= \mathbb{P}(X_1 \in S) + \mathbb{P}(X_1 \notin A, X_2 \in S) + \mathbb{P}(X_1 \notin A, X_2 \notin A, X_3 \in S) + \cdots \\
&\stackrel{\text{(a)}}{=} \sum_{k=1}^{\infty} \Big[\prod_{i=1}^{k-1} \mathbb{P}(X_i \notin A)\Big] \mathbb{P}(X_k \in S) \\
&\stackrel{\text{(b)}}{=} \mathbb{P}(X \in S) \sum_{k=0}^{\infty} \mathbb{P}(X \notin A)^k \\
&\stackrel{\text{(c)}}{=} \frac{\mathbb{P}(X \in S)}{1 - \mathbb{P}(X \notin A)} = \frac{\mathbb{P}(X \in S)}{\mathbb{P}(X \in A)} = \mathbb{P}(X \in S \mid X \in A)
\end{aligned}$$

where (a) is by independence, (b) since the X_i's are identically distributed, and (c) by the formula for the geometric series: $\sum_{k=0}^{\infty} a^k = 1/(1-a)$ for $a \in [0, 1)$.

A Source code in Julia language

Exercise 2

```
using PyPlot, Distributions
draw_now() = (pause(0.001); get_current_fig_manager()[:window][:raise_]())

M = 10^6
for rep = 1:5
    x = rand(Cauchy(0,1),M)
```

```
    MC = cumsum(x)./[1:M]

    figure(1,figsize=(10,4)); clf(); hold(true); grid(true)
    semilogx(1:M, MC)
    draw_now()
    savefig("cauchy$rep.png",dpi=120)
end
```

Exercise 3

```
using PyPlot, Distributions
draw_now() = (pause(0.001); get_current_fig_manager()[:window][:raise_]())

lambda = 1.0
lambda_0 = 1/10^2
likelihood(x,theta) = pdf(Normal(0,1/sqrt(lambda)),x-theta)

x = 2.0
L = lambda + lambda_0
M = lambda*x / L

N = 10^6
R = 5
values = zeros(R)
for rep = 1:R
    theta_samples = rand(Normal(M,1/sqrt(L)),N)
    harmonic_mean = 1/mean(1.0./likelihood(x,theta_samples))
    values[rep] = harmonic_mean
end

truth = pdf(Normal(0,sqrt(1/lambda + 1/lambda_0)),x)

println(values[1:5])
println(truth)

figure(1,figsize=(10,4)); clf(); hold(true); grid(true)
plt.hist(values,50)
plot([truth,truth],[0,ylim()[2]],"k--",linewidth=2)
draw_now()
savefig("harmonic_mean.png",dpi=120)
```

Exercise 4

```
module GPS

using PyPlot, Distributions
draw_now() = (pause(0.001); get_current_fig_manager()[:window][:raise_]())

# Latitude/longitude measurements from some points in Eno River State Park, NC.
# (Obtained using Google maps)
```

```
data = [36.077916 -79.009266
        36.078032 -79.009180
        36.078129 -79.009094
        36.078048 -79.008891
        36.077942 -79.008962
        36.089612 -79.035760 # outlier
        36.077789 -79.008917
        36.077563 -79.009281]

# latitude only:
x = data[:,1]
println("mean(x) = ",mean(x))
println("median(x) = ",median(x))

theta_0 = 36.07
sigma_0 = 0.02
sigma = 0.0002

# prior
P = Cauchy(theta_0,sigma_0)

log_likelihood(x,theta) = sum(logpdf(Cauchy(theta,sigma),x))
likelihood(x,theta) = exp(log_likelihood(x,theta))

# importance sampling distribution
Q = Cauchy(median(x),1e-4) # (to choose the scale, I cheated by looking at
    graph of the likelihood)

lower,upper = 36.06,36.095
t = linspace(lower,upper,1000)
ticks = lower:0.005:upper

# Plot histogram
figure(1,figsize=(10,4)); clf(); hold(true)
subplots_adjust(bottom = 0.2)
plt.hist(x,linspace(lower,upper,100))
xlim(lower,upper)
xticks(ticks,ticks)
yticks(0:4)
xlabel("Latitude (degrees)",fontsize=14)
ylim(0,4)
draw_now()
savefig("gps-histogram.png",dpi=120)

# Plot prior, posterior, and proposal densities
figure(2,figsize=(10,4)); clf(); hold(true)
subplots_adjust(bottom = 0.2)
plot(t,pdf(P,t)./maximum(pdf(P,t)),"g",label="prior",linewidth=2)
posterior = [likelihood(x,th)*pdf(P,th) for th in t]
plot(t,posterior./maximum(posterior),"r",label="posterior",linewidth=2)
```

```
plot(t,pdf(Q,t)./maximum(pdf(Q,t)),"b",label="proposal",linewidth=2)
xlim(lower,upper)
ylim(0,1.1)
yticks([])
xticks(ticks,ticks)
xlabel("\$\\theta\$ (degrees, latitude)",fontsize=16)
legend(loc="upper right",numpoints=1,labelspacing=0.1,fontsize=15)
draw_now()
savefig("gps-curves.png",dpi=120)

# Approximating the marginal likelihood with simple Monte Carlo versus
   Importance sampling
N = 10^6

theta = rand(P,N)
MC = cumsum([likelihood(x,th) for th in theta])./[1:N]

theta = rand(Q,N)
IS = cumsum([likelihood(x,th)*pdf(P,th)/pdf(Q,th) for th in theta])./[1:N]

@printf("MC = %.8e\n",MC[end])
@printf("IS = %.8e\n",IS[end])

# Plot approximations over time
figure(3,figsize=(10,4)); clf(); hold(true)
subplots_adjust(bottom = 0.2)
semilogx(1:N,MC,"g",label="Monte Carlo",linewidth=2)
semilogx(1:N,IS,"b",label="Importance sampling",linewidth=2)
xlabel("N (# of samples used in the approximation)",fontsize=16)
ylabel("approx of \$p(x_{1:n})\$",fontsize=16)
legend(loc="lower right",numpoints=1,labelspacing=0.1,fontsize=15)
draw_now()
savefig("MC-vs-IS.png",dpi=120)

end
```