CS 661, Spring 2013
Midterm Solutions

1. (a) The likelihood function $L_n(\alpha)$ of an IID set of samples $X_1, \ldots, X_n$ is the product of the density function evaluated at the n samples, so

$$L_n(\alpha) = \prod_{i=1}^{n} f(X_i) = \alpha^n \left(\prod_{i=1}^{n} \frac{1}{X_i} \right)^{\alpha+1}.$$

Taking the natural log gives

$$\mathcal{L}_n(\alpha) = \ln(L_n(\alpha)) = n \ln \alpha - (\alpha + 1) \sum_{i=1}^{n} \ln X_i.$$

We then compute the derivative of $\mathcal{L}_n(\alpha)$ with respect to α to get

$$\frac{d}{d\alpha} \mathcal{L}_n(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} \ln X_i.$$

Then equating this to 0 and solving for α yields

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln X_i}.$$

The second derivative of $\mathcal{L}_n(\alpha)$ is

$$\frac{d^2}{d\alpha^2} \mathcal{L}_n(\alpha) = -\frac{n}{\alpha^2} < 0,$$

so $\hat{\alpha}$ is a local maximizer of $\mathcal{L}_n(\alpha)$ and $L_n(\alpha)$. Evaluating $\mathcal{L}_n(\alpha)$ at the boundary $\alpha = 0$ is $-\infty$, so $\hat{\alpha}$ is the global maximizer of $\mathcal{L}_n(\alpha)$ and $L_n(\alpha)$. Thus, $\hat{\alpha}$ is the MLE of α.

(b) Let X have a lognormal(μ, σ^2) distribution, and let $X_1, \ldots, X_n$ be i.i.d. copies of X. For $j = 1, 2$, let $\mu_j = \mathcal{E}[X^j]$ be the jth theoretical moment and $M_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$ be the jth sample moment. We know that $\mu_1 = \exp(\mu + \frac{\sigma^2}{2})$, and variance $\mathcal{V}(X) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1)$. Thus,

$$\mu_2 = \mathcal{V}(X) + \mu_1^2 = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1) + \exp(2\mu + \sigma^2) = \exp(2\mu + 2\sigma^2).$$

To derive the method-of-moments estimators of μ and σ^2, we set $\mu_1 = M_1$ and $\mu_2 = M_2$, and solve for the parameters μ and σ^2. First, $\mu_1 = M_1$ gives $\exp(\mu + \frac{\sigma^2}{2}) = M_1$, which implies

$$\mu = \ln(M_1) - \frac{\sigma^2}{2}. \tag{1}$$

Also, $\mu_2 = M_2$ gives

$$\exp(2\mu + 2\sigma^2) = M_2, \tag{2}$$

and substituting (1) into (2) yields $M_2 = \exp(2\ln(M_1) + \sigma^2) = M_1^2 e^{\sigma^2}$, so

$$\sigma^2 = \ln(M_2/M_1^2) = \ln(M_2) - 2\ln(M_1).$$

Then plugging this into (1) gives $\mu = 2\ln(M_1) - \frac{1}{2}\ln(M_2)$. Hence, the method-of-moments estimators of μ and σ^2 are

$$\begin{aligned} \hat{\mu} &= 2\ln(M_1) - \frac{1}{2}\ln(M_2), \\ \hat{\sigma}^2 &= \ln(M_2) - 2\ln(M_1). \end{aligned}$$

2. (a) We define $\hat{\gamma}_n$ as

$$\hat{\gamma}_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in A),$$

where I denotes the indicator function.

(b) Note that

$$\mathcal{E}[\hat{\gamma}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}[I(X_i \in A)] = \frac{1}{n}\sum_{i=1}^{n} \mathcal{P}(X_i \in A) = \gamma$$

since the X_i are IID. Hence, $\hat{\gamma}_n$ is an unbiased estimator of γ.

(c) Note that $I(X_i \in A)$, $i = 1, 2, \ldots, n$, are IID because X_i, $i = 1, 2, \ldots, n$, are IID. Also, $\hat{\gamma}_n$ is the sample average of the $I(X_i \in A)$, $i = 1, 2, \ldots, n$, each of which has mean $\mathcal{E}[I(X_i \in A)] = \gamma$, which we assumed satisfies $0 < \gamma < 1$, so γ is finite. Hence, $\hat{\gamma}_n \approx \gamma$ for large n by the ordinary law of large numbers.

(d) Since $I^2(X_i \in A) = I(X_i \in A)$, we have $\beta^2 = \mathcal{V}[I(X_i \in A)] = \mathcal{E}[I^2(X_i \in A)] - (\mathcal{E}[I(X_i \in A)])^2 = \mathcal{E}[I(X_i \in A)] - (\mathcal{E}[I(X_i \in A)])^2 = \gamma(1-\gamma)$, which is positive and finite since we assumed $0 < \gamma < 1$. Also, since $\hat{\gamma}_n$ is the sample average of the $I(X_i \in A)$, which are IID with mean γ, we have the following CLT:

$$\frac{\sqrt{n}}{\sqrt{\gamma(1-\gamma)}}(\hat{\gamma}_n - \gamma) \overset{\mathcal{D}}{\approx} \mathcal{N}(0,1)$$

for large n. Thus, $\beta = \sqrt{\gamma(1-\gamma)}$ and $\alpha = \gamma$.

(e) An estimator for β is $\hat{\beta}_n = \sqrt{\hat{\gamma}_n(1-\hat{\gamma}_n)}$. Another possible estimator of β is

$$\hat{\beta}_n = \left[\frac{1}{n-1}\sum_{i=1}^{n}(I(X_i \in A) - \hat{\gamma}_n)^2\right]^{1/2}.$$

Since $\hat{\gamma}_n \approx \gamma$ for large n, as we argued in part (c), the first estimator satisfies $\hat{\beta}_n = \sqrt{\hat{\gamma}_n(1-\hat{\gamma}_n)} \approx \sqrt{\gamma(1-\gamma)} = \beta$ for large n, so it satisfies a law of large numbers. The second estimator also does since it is the sample variance of the

$I(X_i \in A)$, $i = 1, 2, \ldots, n$, and we know that the sample variance of IID samples is consistent, as can be seen as follows. Note that the second estimator satisfies

$$\hat{\beta}_n = \left[\frac{n}{n-1} \frac{1}{n} \sum_{i=1}^{n} I(X_i \in A) - \frac{n}{n-1} \hat{\gamma}_n^2 \right]^{1/2}.$$

Now $n/(n-1) \approx 1$ for large n, and $\frac{1}{n} \sum_{i=1}^{n} I(X_i \in A) \approx \mathcal{E}[I(X \in A)] = \gamma$ by the law of large numbers. Thus, $\hat{\beta}_n \approx \sqrt{\gamma - \gamma^2} = \beta$.

(f) The CLT implies that for large n,

$$\begin{aligned}
0.95 &= \mathcal{P}\left\{-1.96 < \mathcal{N}(0,1) < 1.96\right\} \\
&\approx \mathcal{P}\left\{-1.96 < \frac{\sqrt{n}}{\beta}(\hat{\gamma}_n - \gamma) < 1.96\right\} \\
&\approx \mathcal{P}\left\{-1.96 < \frac{\sqrt{n}}{\hat{\beta}_n}(\hat{\gamma}_n - \gamma) < 1.96\right\} \\
&= \mathcal{P}\left\{-1.96\frac{\hat{\beta}_n}{\sqrt{n}} < \hat{\gamma}_n - \gamma < 1.96\frac{\hat{\beta}_n}{\sqrt{n}}\right\} \\
&= \mathcal{P}\left\{\hat{\gamma}_n - 1.96\frac{\hat{\beta}_n}{\sqrt{n}} < \gamma < \hat{\gamma}_n + 1.96\frac{\hat{\beta}_n}{\sqrt{n}}\right\},
\end{aligned}$$

where the third step holds since $\hat{\beta}_n \approx \beta$ for large n for $\hat{\beta}_n$ defined as either of the estimators in the previous part. Thus, an approximate 95% (two-sided) confidence interval for γ is

$$\left(\hat{\gamma}_n \pm 1.96\frac{\hat{\beta}_n}{\sqrt{n}}\right)$$

3. (a) We first compute the CDF F_1 corresponding to the density f_1. For $x < -3$, we have $F_1(x) = 0$. For $-3 \le x < -2$, we have

$$F_1(x) = \int_{-3}^{x} \frac{u+3}{2}\, du = \frac{(x+3)^2}{4}.$$

For $-2 \le x < 0$, we have $F_1(x) = 1/4$. For $0 \le x < 4$, we have

$$F_1(x) = \frac{1}{4} + \int_{0}^{x} \frac{3}{16}\, du = \frac{1}{4} + \frac{3}{16}x.$$

Finally, for $x \ge 4$, we have $F_1(x) = 1$. Thus,

$$F_1(x) = \begin{cases} 0 & \text{if } x < -3, \\ \frac{(x+3)^2}{4} & \text{if } -3 \le x < -2, \\ \frac{1}{4} & \text{if } -2 \le x < 0, \\ \frac{1}{4} + \frac{3}{16}x & \text{if } 0 \le x < 4, \\ 1 & \text{if } x \ge 4. \end{cases}$$

To generate a sample Y having density f_1 using inversion, we invert its CDF F_1 by setting $F_1(Y) = U$ for $U \sim \text{unif}[0,1]$ and solving for Y in terms of U. For $0 \le U \le 1/4$, we get the equation $(Y+3)^2/4 = U$, which yields $Y = 2\sqrt{U} - 3$. For $1/4 < U \le 1$, we have the equation $1/4 + 3Y/16 = U$, which implies $Y = (16U - 4)/3$. Hence, our inversion algorithm for generating a sample Y with density f_1 is as follows:

1. Generate $U \sim \text{unif}[0,1]$.
2. If $0 \le U \le 1/4$, then return $Y = 2\sqrt{U} - 3$.
3. If $1/4 < U \le 1$, then return $Y = (16U - 4)/3$.

(b) To generate a sample W from density f_3, note that we can write

$$f_3(x) = \frac{1}{2} g_1(x) + \frac{1}{2} g_2(x),$$

where $g_1(x) = e^x I_{(-\infty,0)}(x)$ and $g_2(x) = e^{-x} I_{[0,\infty)}(x)$. Note that g_1 and g_2 are both density functions. The CDF G_1 corresponding to density g_1 is $G_1(x) = \int_{-\infty}^{x} e^u \, du$, so $G_1(x) = e^x$ for $x < 0$ and $G_1(x) = 1$ for $x \ge 0$. Density g_2 is for an exponential with rate 1, so its CDF $G_2(x) = 1 - e^{-x}$ for $x \ge 0$, and 0 for $x < 0$.

We can generate a sample $W \sim G_1$ using inversion by setting $G_1(W) = U$ for $U \sim \text{unif}[0,1]$ and solving for W in terms of U. Doing this leads to the equation $e^W = U$, so $W = \ln(U)$.

We can generate a sample $W \sim G_2$ using inversion by setting $G_2(W) = U$ for $U \sim \text{unif}[0,1]$ and solving for W in terms of U. Doing this leads to the equation $1 - e^{-W} = U$, so $W = -\ln(1-U)$.

Hence, our composition algorithm for generating a sample W having density f_3 is as follows:

1. Generate independent $U_1, U_2 \sim \text{unif}[0,1]$.
2. If $0 \le U_1 \le 1/2$, then return $W = \ln(U_2)$.
3. If $1/2 < U \le 1$, then return $W = -\ln(1 - U_2)$.

(c) To generate a sample Z with density f_2 using acceptance-rejection with f_3 as the majorizing density, we first need to determine a constant c such that $f_2(x) \le c f_3(x)$ for all x.

To determine the smallest constant c, first consider $x \ge 0$. We want to find $x \ge 0$ that maximizes the ratio

$$r(x) \equiv \frac{f_2(x)}{f_3(x)} = \frac{(1/\sqrt{2\pi}) \exp\{-x^2/2\}}{(1/2)e^{-x}} = \sqrt{\frac{2}{\pi}} \exp\left\{ \frac{-x^2}{2} + x \right\}.$$

It is then sufficient to find $x \geq 0$ that maximizes $h(x) \equiv \ln r(x) = \ln(\sqrt{2/\pi}) + x - x^2/2$. Taking the derivative of $h(x)$ gives $1-x$, which we equate to 0, yielding $x = 1$. The second derivative of $h(x)$ is -1, so $x = 1$ is the local maximizer of $h(x)$ and $r(x)$, and $r(1) = \sqrt{2/\pi}e^{1/2} = \sqrt{2e/\pi}$. Also, at the boundary $x = 0$, the ratio $r(0) = \sqrt{2/\pi} < r(1)$, so $x = 1$ is the global maximizer of $r(x)$ over $x \geq 0$. Using the similar argument for $x < 0$ shows that $x = -1$ maximizes the ratio $f_2(x)/f_3(x)$ over all $x < 0$, and $f_2(-1)/f_3(-1) = \sqrt{2e/\pi}$. Thus, we get that $c = \sqrt{2e/\pi}$ is the smallest constant such that $f_2(x) \leq cf_3(x)$ for all $x \in \Re$. We then arrive at the following acceptance-rejection algorithm to generate a sample of Z having density f_2:

1. Generate a sample of W with density f_3 using the algorithm in the previous part.
2. Generate $U_3 \sim \text{unif}[0, 1]$ independent of W.
3. If $cf_3(W)U_3 \leq f_2(W)$, then return $Z = W$; else, go to step 1.

In each iteration of the above algorithm, we require 3 uniforms (2 in step 1, and another in step 3). The expected number of iterations in the above algorithm is c, so the expected total number of uniforms needed to generate Z with the above algorithm is $3c$.

(d) Note that f_4 is the convolution of f_1 and f_2. Thus, we can express $X = Y + Z$ for Y and Z independent. To generate a sample of X, generate independent samples of Y and Z using the algorithms from the previous parts, and then set $X = Y + Z$. Generating Y takes exactly 1 uniform, and the expected number of uniforms to generate Z is $3c$, so the expected total number of uniforms to generate X is $1 + 3c$.

4. (a) The method used in the case is first to construct 95% confidence intervals for the mean interarrival and service times. Then the endpoints of the 95% confidence intervals were used as the endpoints of uniform distributions. For example, the 95% confidence interval for the mean interarrival time was (2.99, 7.17), and the interarrival time distribution was modeled as uniform on (2.99, 7.17). The service time distribution was modeled as uniform on the interval (3.10, 6.46).

This approach is not appropriate to model a distribution. Confidence intervals are used to make inferences about the (unknown) mean μ of a distribution, not about the entire distribution itself. Specifically, the interpretation of a $100(1-\alpha)\%$ confidence interval is that there is roughly a $100(1-\alpha)\%$ chance that μ lies within the confidence interval. This does not say that $100(1-\alpha)\%$ of the data will lie within the confidence interval. Nor does it say that with probability $1-\alpha$, a random observation will fall in the interval.

Moreover, the approach used in the case assumed a uniform distribution with support being the confidence interval. This also is inappropriate since it completely

ignores the shape of the histogram of the data and instead assumes observations are equally likely to fall anywhere within the support.

(b) For the interarrival data, the Input Analyzer automatically chose the sample mean 5.08 as its estimate of the (theoretical) mean for the exponential distribution. However, the results from the χ^2 test do not make any sense since the number of degrees of freedom is 0. As we discussed in class, the Input Analyzer of Arena has a quirk that it sometimes chooses the number of bins (cells, intervals) in a χ^2 test so that the number of degrees of freedom is 0, which makes no sense. However, the results for the Kolmogorov-Smirnov test are fine, with a p-value greater than 0.15, so we do not reject the null hypothesis at level $\alpha = 0.05$ that the data came from an exponential distribution with mean 5.08.

For the service-time data, Arena's Input Analyzer carries out a χ^2 test to evaluate the fit of the service-time data to an exponential distribution with mean 4.78, we get a p-value of 0.206, and so we do not reject the null hypothesis at level $\alpha = 0.05$ that the data came from an exponential distribution with mean 4.78. Moreover, the Kolmogorov-Smirnov test gives a p-value greater than 0.15, so we again do not reject the null hypothesis at level $\alpha = 0.05$.

There are additional reasons supporting the choice of exponential distributions for the service time and interarrival times. The histograms for the interarrival-time and service-time data look like exponential distributions. Also, we computed the sample means, variances, and coefficients of variation (CV) for the data, and these are given below.

Statistic	Interarrival-Time Data	Service-Time Data
Sample Mean	5.08	4.78
Sample Standard Deviation	5.96	4.76
Sample CV	1.17	1.00

The sample CVs are both reasonably close to 1, which is the theoretical CV of an exponential distribution, thus suggesting again that exponential distributions are reasonable for modeling the data.

Finally, for the interarrival-time data, we learned in class that a Poisson process is often appropriate to model arrival processes, with theoretical justification provided by the Palm-Khintchine theorem. Since a Poisson process has exponentially distributed inter-event times, this provides further support for using an exponential distribution for the interarrival times.