CS 661, Spring 2012
Midterm Solutions

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

$$L_n(\theta) = \prod_{i=1}^{n} p(X_i) = \theta^N (1-\theta)^{n-N},$$

where $N = \sum_{i=1}^{n} X_i$ is the number of the n samples that equal 1. Taking the log gives

$$\mathcal{L}_n(\theta) = \ln(L_n(\theta)) = N \ln\theta + (n-N)\ln(1-\theta).$$

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

$$\frac{d}{d\theta}\mathcal{L}_n(\theta) = \frac{N}{\theta} - \frac{n-N}{1-\theta}.$$

Then equating this to 0 and solving for θ yields

$$\hat{\theta} = \frac{N}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Computing the second derivative of $\mathcal{L}_n(\theta)$ gives

$$\frac{d^2}{d\theta^2}\mathcal{L}_n(\theta) = -\frac{N}{\theta^2} - \frac{n-N}{(1-\theta)^2} < 0,$$

so $\hat{\theta}$ is a local maximizer of $\mathcal{L}_n(\theta)$ and $L_n(\theta)$. Now we consider the boundary values for θ, which are 0 and 1. Note that $\mathcal{L}_n(0) = \mathcal{L}_n(1) = -\infty$, so $\hat{\theta}$ is the global maximizer of $\mathcal{L}_n(\theta)$ and $L_n(\theta)$, which makes it the MLE for θ.

(b) For a gamma(α, β) distribution, the population mean is $\mu_1 = \alpha\beta$ and the population variance is $\alpha\beta^2$, so the population second moment is $\mu_2 = \alpha\beta^2 + \alpha^2\beta^2$. Let $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ be the kth sample moment, $k = 1, 2$. Then set $\mu_1 = M_1$ and $\mu_2 = M_2$, giving us the equations $\alpha\beta = M_1$ and $\alpha\beta^2 + \alpha^2\beta^2 = M_2$, and we want to solve for α and β in these two equations. The first equation yields $\alpha = M_1/\beta$, and substituting this into the second equation gives $M_1\beta + M_1^2 = M_2$, so $\beta = (M_2 - M_1^2)/M_1$. Then substituting this into the first equation gives $\alpha = M_1^2/(M_2 - M_1^2)$. Thus, the method-of-moments estimators of α and β are $\hat{\alpha} = M_1^2/(M_2 - M_1^2)$ and $\hat{\beta} = (M_2 - M_1^2)/M_1$.

2. (a) For each i,

$$\mathcal{E}[I\{X_i \le x\}] = 1 \times \mathcal{P}\{X_i \le x\} + 0 \times \mathcal{P}\{X_i > x\} = \mathcal{P}\{X_i \le x\} = F(x).$$

Thus,

$$\mathcal{E}[\hat{F}_n(x)] = \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^{n} I\{X_i \le x\}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}\left[I\{X_i \le x\}\right] = \frac{1}{n}\sum_{i=1}^{n} F(x) = F(x),$$

so $\hat{F}_n(x)$ is an unbiased estimator of $F(x)$ for each x.

Let $Y_i = I\{X_i \leq x\}$. Since the X_i are IID, the Y_i are also IID. The mean of Y_i is $\mathcal{E}[Y_i] = (1 \times \mathcal{P}\{X_i \leq x\}) + (0 \times \mathcal{P}\{X_i > x\}) = \mathcal{P}\{X_i \leq x\} = F(x)$. Since the mean $F(x)$ satisfies $0 < F(x) < 1$, the law of large numbers implies that $\bar{Y}_n \approx F(x)$ for large n, where $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$.

(b) Since the $Y_i = I\{X_i \leq x\}$ are IID with mean $E[Y_i] = F(x) \leq 1 < \infty$, we have that $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n Y_i \approx \mathcal{E}[Y_i] = F(x)$ for large n by the LLN.

(c) Note that $\mathcal{E}[Y_i^2] = (1^2 \times \mathcal{P}\{X_i \leq x\}) + (0^2 \times \mathcal{P}\{X_i > x\}) = \mathcal{P}\{X_i \leq x\} = F(x)$, so $\sigma^2 \equiv \mathcal{V}[Y_i] = F(x) - F^2(x) = F(x)(1 - F(x))$. Since $0 < F(x) < 1$, we have that $0 < \sigma^2 < 1 < \infty$. Thus, the central limit theorem implies

$$\frac{\sqrt{n}}{\sigma}(\bar{Y}_n - F(x)) \approx \mathcal{N}(0,1)$$

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

(d) Since $\beta = F(x)(1 - F(x))$, one estimator of β is $\hat{\beta}_n = \hat{F}_n(1 - \hat{F}_n(x))$. Also, since $\beta = \sigma$, another estimator of β is $S_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (Y_i - \hat{F}_n(x))^2}$.

3. (a) From the CLT from the previous question, we have that

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

and also that

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

where $S_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (Y_i - \hat{F}_n(x))^2}$. Therefore, since $\Phi^{-1}(0.975) = 1.96$, where Φ denotes the CDF of a standard normal, we get that

$$\begin{aligned}
0.95 &= \mathcal{P}\{-1.96 \leq \mathcal{N}(0,1) \leq 1.96\} \\
&\approx \mathcal{P}\left\{-1.96 \leq \frac{\sqrt{n}}{S_Y}(\bar{Y}_n - \gamma) \leq 1.96\right\} \\
&= \mathcal{P}\left\{-\frac{1.96 S_Y}{\sqrt{n}} \leq (\bar{Y}_n - \gamma) \leq \frac{1.96 S_Y}{\sqrt{n}}\right\} \\
&= \mathcal{P}\left\{\bar{Y}_n - \frac{1.96 S_Y}{\sqrt{n}} \leq \gamma \leq \bar{Y}_n + \leq \frac{1.96 S_Y}{\sqrt{n}}\right\}.
\end{aligned}$$

Hence, an approximate 95% two-sided confidence interval for γ is $[\bar{Y}_n \pm 1.96 S_Y/\sqrt{n}]$. If the lower (resp., upper) endpoint is below 0 (resp., above 1), then it can be replaced with 0 (resp., 1).

(b) By slightly modifying the previous part, we can obtain an approximate 95% lower one-sided confidence interval for p to be $[\bar{Y}_n - 1.65 S_Y/\sqrt{n}, 1]$. Then if $\bar{Y}_n - 1.65 S_Y/\sqrt{n} > 0.8$, then reject the null hypothesis H_0. If $\bar{Y}_n - 1.65 S_Y/\sqrt{n} \leq 0.8$, then don't reject H_0.

4. (a) We have that f_1 is the density function of a Weibull distribution with shape parameter $\alpha = 3$ and scale parameter $\beta = 4$, so its CDF is $F_1(x) = [1 - \exp\{-(x/4)^3\}]I_{[0,\infty)}(x)$. To generate a sample $Y \sim F_1$, we can use inversion. Let $U \sim \text{unif}[0,1]$. Setting $F_1(Y) = U$ yields $1 - \exp\{-(Y/4)^3\} = U$. Solving for Y in terms of U then gives $Y = 4[-\ln(1-U)]^{1/3}$. Thus, an algorithm for generating $Y \sim F_1$ is as follows:

> 1. Generate IID $U \sim \text{unif}[0,1]$.
> 2. Return $Y = 4[-\ln(1-U)]^{1/3}$.

(b) The derivative of $f_2(x)$ is $f_2'(x) = \frac{1}{1-e^{-1}}[-e^{-x}+2]I_{[0,1]}(x)$, which is strictly positive for $x \in [0,1]$ since $0 < e^{-x} \leq 1$ for all $0 \leq x \leq 1$. Thus, on its support $[0,1]$, the density f_2 is strictly increasing, so its maximum occurs at the right endpoint $x = 1$. Let $m = f_2(1) = (1+e^{-1})/(1-e^{-1})$ be the maximum of $f_2(x)$. Now we can use acceptance-rejection with a uniform majorizing function $g(x) = I_{[0,1]}(x)$. The algorithm to generate $Z \sim f_2$ is then as follows.

> 1. Generate $U_1, U_2 \sim \text{unif}[0,1]$ IID.
> 2. Let $Z_1 = U_1$ and $Z_2 = mU_2$.
> 3. If $Z_2 \leq f_2(Z_1)$, then set $Z = Z_1$ and stop;
>   else, discard (Z_1, Z_2) and return to step 1.

The expected number of uniforms needed to generate Z is $2m = 2(1+e^{-1})/(1-e^{-1})$.

The previous acceptance-rejection algorithm uses a uniform majorizing function, but we can also use a non-uniform majorizing function, as follows. Specifically, since $e^{-x} \leq 1$ for all $0 \leq x \leq 1$, we have that $f_2(x) \leq [2x/(1-e^{-1})]I_{[0,1]}(x)$ for all x. Thus, if we define $g(x) = 2xI_{[0,1]}(x)$ and $c = 1/(1-e^{-1})$, then $f_2(x) \leq cg(x)$ for all x. The density g has CDF $G(x) = x^2$ for $x \in [0,1]$, $G(x) = 0$ for $x < 0$, and $G(x) = 1$ for $x > 1$. We can generate a sample W from density g using inversion; i.e., if $U \sim \text{unif}[0,1]$, then $W = \sqrt{U} \sim G$. This then leads to the following alternative acceptance-rejection algorithm.

> 1. Generate $U_1, U_2 \sim \text{unif}[0,1]$ IID.
> 2. Let $W = \sqrt{U_1}$.
> 3. If $[2W/(1-e^{-1})]U_2 \leq f_2(W)$, then set $Z = W$ and stop;
>   else, discard (U_1, U_2) and return to step 1.

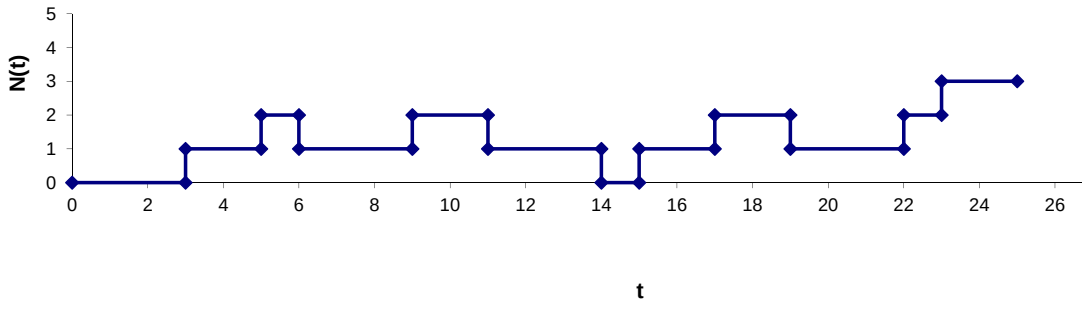
The expected number of uniforms needed to generate Z for the alternative algorithm is $2c = 2/(1-e^{-1})$. This is smaller than $2m = 2(1+e^{-1})/(1-e^{-1})$ required for the first A-R algorithm, but the second algorithm requires computing a square root, which the first algorithm doesn't require.

(c) To generate $X \sim f$, we use composition with the previous two parts. The algorithm is as follows:

> 1. Generate $U_0 \sim \text{unif}[0,1]$.
> 2. If $U_0 < 3/5$, then generate $U_1 \sim \text{unif}[0,1]$ independent of U_0, and return $X = 4[-\ln(1-U_1)]^{1/3}$ and stop.
> 3. If $3/5 \le U_0 \le 1$, then do the following:
> 4. Generate $U_1, U_2 \sim \text{unif}[0,1]$ IID and independent of U_0.
> 5. Set $Z_1 = U_1$ and $Z_2 = mU_2$, where $m = (1+e^{-1})/(1-e^{-1})$.
> 6. If $Z_2 \le f_2(Z_1)$, then return $X = Z_1$ and stop;
> else, reject (Z_1, Z_2) and return to step 4.

The expected total number of uniforms needed to generate X is $1 + [(3/5) \times 1] + [(2/5) \times 2 \times m] = (8+4m)/5$.

5. (a) Below is a plot of $N(t)$ vs. t:



The times t at which $N(t)$ changes are 0, 3, 5, 6, 9, 11, 14, 15, 17, 19, 22, 23.

(b) Calculating the area under the $N(t)$ plot from $t = 0$ to $t = 25$ and dividing by 25, we get 31/25.