CS 661, Spring 2015
Midterm Solutions

1. (a) For outcome ω_1, we have $N(1) = 1$, $N(2) = 1$, $N(3) = 1$, $N(4) = 2$, $N(5) = 3$. For outcome ω_2, we have $N(1) = 0$, $N(2) = 0$, $N(3) = 1$, $N(4) = 1$, $N(5) = 2$. For outcome ω_3, we have $N(1) = 0$, $N(2) = 1$, $N(3) = 1$, $N(4) = 1$, $N(5) = 1$.

   (b) Let $\text{bin}(n, q)$ denote a binomial distribution with n IID trials and success probability q in each trial. Since the coin tosses are IID, the distribution of $N(5)$ is $\text{bin}(5, p)$ and the distribution of $N(16)$ is $\text{bin}(16, p)$. In general, the distribution of $N(k)$ is $\text{bin}(k, p)$, which depends on k. Also, $\mathcal{P}(N(k) = j) = \binom{k}{j} p^j (1-p)^{k-j}$ for each $j = 0, 1, 2, \ldots, k$, where $\binom{k}{j} = k!/[j!(k-j)!]$.

   (c) Note that if $N(k) = j$ for some $0 \le j \le k$, then $N(k+2)$ can only take on the values j, $j+1$ and $j+2$, with respective probabilities $(1-p)^2$, $2p(1-p)$, and p^2. The conditional probability $\mathcal{P}(N(k+2) = j \,|\, N(k) = j) = (1-p)^2$. But the unconditional probability $\mathcal{P}(N(k+2) = j) = \binom{k+2}{j} p^j (1-p)^{k+2-j}$. Since the conditional and unconditional probabilities are unequal, we have that $N(k)$ and $N(k+2)$ are dependent. Hence, the marginal (binomial) distributions of $N(k)$, $k = 1, 2, 3, \ldots$, do not completely define the probabilistic structure of the stochastic process N.

   Another way of showing the dependence is as follows. The event $\{N(k) = j, N(k+2) = j\}$ occurs if and only if there are exactly j heads in the first k tosses, and the next two tosses are tails. There are $\binom{k}{j}$ ways in which there are exactly j heads in the first k tosses. There are the same number of outcomes in which there are exactly j heads in the first k tosses and the next two tosses are tails. Each of these latter outcomes has probability $p^j(1-p)^{k+2-j}$, so

$$\mathcal{P}(N(k) = j, N(k+2) = j) = \binom{k}{j} p^j (1-p)^{k+2-j}$$

   Also,

$$\mathcal{P}(N(k) = j) = \binom{k}{j} p^j (1-p)^{k-j},$$
$$\mathcal{P}(N(k+2) = j) = \binom{k+2}{j} p^j (1-p)^{k+1-j}.$$

   Thus, since $\mathcal{P}(N(k) = j, N(k+1) = j) \neq \mathcal{P}(N(k) = j)\,\mathcal{P}(N(k+2) = j)$, we have that $N(k)$ and $N(k+2)$ are dependent.

   An even simpler way of showing $N(k)$ and $N(k+2)$ are independent is to consider the event $\{N(k) = 0, N(k+2) = 3\}$ for any $k \ge 1$. It is impossible to have $N(k) = 0$ and $N(k+2) = 3$, so $\mathcal{P}(N(k) = 0, N(k+2) = 3) = 0$. But $\mathcal{P}(N(k) = 0) = (1-p)^k$ and $\mathcal{P}(N(k+2) = 3) = \binom{k+2}{3} p^3 (1-p)^{k-1}$, so $\mathcal{P}(N(k) = 0)\mathcal{P}(N(k+2) = 3) \neq 0$. Thus, $\mathcal{P}(N(k) = 0, N(k+2) = 3) \neq \mathcal{P}(N(k) = 0)\mathcal{P}(N(k+2) = 3)$, so $N(k)$ and $N(k+2)$ are dependent.

(d) Note that $N(k) - N(j)$ is the number of heads in the $k - j$ tosses $j + 1, j + 2, j + 3, \ldots, k$. Because the tosses are IID, the distribution of $N(k) - N(j)$ is $\text{bin}(k - j, p)$. Similarly, $N(m+k) - N(m+j)$ is the number of heads in the $k - j$ tosses $m + j + 1, m + j + 2, m + j + 3, \ldots, m + k$. Because the tosses are IID, the distribution of $N(m+k) - N(m+j)$ is again $\text{bin}(k - j, p)$. Thus, the distribution of $N(m + k) - N(m + j)$ does not depend on m. Hence, the process N is said to have *stationary increments*.

2. (a) The likelihood function $L(X_1, \ldots, X_n)$ is

$$L(X_1, \ldots, X_n; p) = \prod_{i=1}^{n} (1-p)^{X_i - 1} p = (1-p)^{\sum_{i=1}^{n} X_i - n} p^n.$$

To derive the MLE for p, we need to find the value of $0 \leq p \leq 1$ that maximizes $L(X_1, \ldots, X_n; p)$. Equivalently, we can find the value of p that maximizes

$$\ln L(X_1, \ldots, X_n; p) = \left(\sum_{i=1}^{n} X_i - n \right) \ln(1-p) + n \ln(p).$$

Taking the derivative of $\ln L(X_1, \ldots, X_n; p)$ with respect to p and setting it equal to zero yields

$$\frac{d}{dp} \left[\left(\sum_{i=1}^{n} X_i - n \right) \ln(1-p) + n \ln(p) \right] = - \left(\sum_{i=1}^{n} X_i - n \right) \frac{1}{1-p} + \frac{n}{p} = 0,$$

which implies

$$- \left(\sum_{i=1}^{n} X_i - n \right) p + n(1-p) = 0.$$

Thus,

$$\left(\sum_{i=1}^{n} X_i \right) p = n,$$

so

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i},$$

which satisfies $0 \leq \hat{p} \leq 1$.

We also need to check that $\hat{p}$ is a local maximizer rather than a local minimizer. To do this, we need that the second derivative at $\hat{p}$ is negative:

$$\frac{d^2}{dp^2} \left[\left(\sum_{i=1}^{n} X_i - n \right) \ln(1-p) + n \ln(p) \right] = - \left(\sum_{i=1}^{n} X_i - n \right) \frac{1}{(1-p)^2} - \frac{n}{p^2} < 0$$

for all $0 \leq p \leq 1$, so $\hat{p}$ is a local maximizer.

We finally need to check the log likelihood function at the boundaries $p = 0$ and $p = 1$. In both cases, $\ln L(X_1, \ldots, X_n; p) = -\infty$, so $\hat{p}$ is the global maximizer and thus the MLE.

(b) We have assumed that a and b are unknown in the unif$[a, b]$ distribution, so we have two unknown parameters to estimate. Thus, we equate the sample kth moment M_k to the true kth moment μ_k for $k = 1, 2$, where $M_k = \sum_{i=1}^{n} X_i^k$. The first true moment of F is $\mu_1 = (a + b)/2$, and the variance is $(b - a)^2/12$. Since variance $= \mu_2 - \mu_1^2$, the second true moment is

$$\mu_2 = \text{variance} + \mu_1^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} = \frac{a^2 + b^2 + ab}{3}.$$

Now setting $\mu_1 = M_1$ leads to $(a + b)/2 = M_1$, so

$$a = 2M_1 - b. \tag{1}$$

Then let $\mu_2 = M_2$ to get $(a^2 + b^2 + ab)/3 = M_2$, and substituting in $a = 2M_1 - b$ yields

$$\frac{(2M_1 - b)^2 + b^2 + (2M_1 - b)b}{3} = M_2.$$

Simple algebra gives us

$$b^2 - 2M_1 b + (4M_1^2 - 3M_2) = 0,$$

and the quadratic equation then results in

$$b = \frac{2M_1 \pm \sqrt{4M_1^2 - 4(4M_1^2 - 3M_2)}}{2} = M_1 \pm \sqrt{3(M_2 - M_1^2)}.$$

Since the right endpoint b of the uniform distribution must be larger than the sample mean M_1, we then have the method-of-moments estimator of b as

$$\hat{b} = M_1 + \sqrt{3(M_2 - M_1^2)}.$$

Putting this back into equation (1) leads to method-of-moments estimator of a as

$$\hat{a} = M_1 - \sqrt{3(M_2 - M_1^2)}.$$

3. (a) For each $i = 1, 2, \ldots, n$, let $W_i = I(-1 \le 3X_i^5 \ln Y_i / \sin(X_i\sqrt{Z_i} \le 5)$, where I denotes the indicator function, which equals 1 (resp., 0) when its argument is true (resp., false). We define $\hat{\theta}_n$ as

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} W_i$$

(b) Note that

$$\begin{aligned}\mathcal{E}[\hat{\theta}_n] &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}[I(-1 \le 3X_i^5 \ln Y_i / \sin(X_i\sqrt{Z_i} \le 5)] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{P}(-1 \le 3X_i^5 \ln Y_i / \sin(X_i\sqrt{Z_i} \le 5) = \theta\end{aligned}$$

since the (X_i, Y_i, Z_i) are IID triples. Hence, $\hat{\theta}_n$ is an unbiased estimator of θ.

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

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

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

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

(e) An estimator for β is $\hat{\beta}_n = \sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}$. Since $\hat{\theta}_n \approx \theta$ for large n, as we argued in part (c), the estimator satisfies $\hat{\beta}_n = \sqrt{\hat{\theta}_n(1-\hat{\theta}_n)} \approx \sqrt{\theta(1-\theta)} = \beta$ for large n, so $\hat{\beta}_n$ obeys a law of large numbers.

Another possible estimator of β is

$$\hat{\beta}_n = \left[\frac{1}{n-1}\sum_{i=1}^{n}(W_i - \hat{\theta}_n)^2\right]^{1/2}.$$

This second estimator also satisfies a law of large numbers 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

$$\hat{\beta}_n = \left[\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n} W_i - \frac{n}{n-1}\hat{\theta}_n^2\right]^{1/2}.$$

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

(f) The CLT implies that for large n,

$$\begin{aligned}
0.95 &= \mathcal{P}\left\{\mathcal{N}(0,1) \geq -1.65\right\} \\
&\approx \mathcal{P}\left\{\frac{\sqrt{n}}{\beta}(\hat{\theta}_n - \theta) \geq -1.65\right\} \\
&\approx \mathcal{P}\left\{\frac{\sqrt{n}}{\hat{\beta}_n}(\hat{\theta}_n - \theta) \geq -1.65\right\} \\
&= \mathcal{P}\left\{\hat{\theta}_n - \theta \geq -1.65\frac{\hat{\beta}_n}{\sqrt{n}}\right\} \\
&= \mathcal{P}\left\{\hat{\theta}_n + 1.65\frac{\hat{\beta}_n}{\sqrt{n}} \geq \theta\right\},
\end{aligned}$$

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% upper-one-sided confidence interval for θ is

$$\left(-\infty,\, \hat{\theta}_n + 1.65\frac{\hat{\beta}_n}{\sqrt{n}}\right]$$

Since we θ is a probability, we know that $\theta \geq 0$, so we can replace the lower confidence limit by 0 to get

$$\left[0,\, \hat{\theta}_n - 1.65\frac{\hat{\beta}_n}{\sqrt{n}}1\right]$$

as our approximate 95% upper-one-sided confidence interval for θ.

4. (a) See HW 3, problem 8 solutions.

   (b) See HW 3, problem 8 solutions.

   (c) See HW 3, problem 8 solutions.

   (d) See HW 3, problem 8 solutions.

   (e) When α is an integer, then the gamma density g is an Erlang with α stages, so we can express $Y = Z_1 + Z_2 + \cdots + Z_\alpha$, where $Z_1, Z_2, \ldots, Z_\alpha$ are IID exponential random variables with mean $-1/\theta$, which is positive since $\theta < 0$. We can generate each Z_i using inversion, i.e., $Z_i = (1/\theta)\ln(U_i)$, where $U_1, U_2, \ldots, U_\alpha \sim \text{unif}[0,1]$ IID. Thus, an algorithm to generate Y with density g is as follows:

   > 1. Generate $U_1, U_2, \ldots, U_\alpha \sim \text{unif}[0,1]$ IID.
   > 2. Return $Y = \sum_{i=1}^{\alpha}(1/\theta)\ln(U_i) = (1/\theta)\ln(\prod_{i=1}^{\alpha} U_i)$.

   (f) We first calculate

   $$\frac{h_\theta(x)}{c\, g(x)} = \exp\{-(x/\beta)^\alpha\}.$$

   Thus, we have the following acceptance-rejection algorithm to generate X having density h_θ when α is an integer and $\theta < 0$:

   > 1. Generate $U_1, U_2, \ldots, U_\alpha \sim \text{unif}[0,1]$ IID, and set $Y = (1/\theta)\ln(\prod_{i=1}^{\alpha} U_i)$.
   > 2. Generate $U \sim \text{unif}[0,1]$, independent of Y.
   > 3. If $U < \exp\{-(Y/\beta)^\alpha\}$, then return $X \leftarrow Y$, and stop. Otherwise, go to Step 1.

(g) In each iteration, generating Y in step 1 requires α uniforms, and another uniform is required in step 2. So the total number of uniforms in each iteration is $\alpha+1$. The expected number of iterations is c, so the expected number of uniforms required to generate X is $c(\alpha+1)$.

5. Let $X_1, X_2, \ldots, X_n$ be n values output by the program, and define their sample average $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. We carry out a hypothesis test with

$$\begin{aligned} H_0 &: \text{program output is i.i.d. exponential with mean 1,} \\ H_a &: \text{program output is not i.i.d. exponential with mean 1.} \end{aligned}$$

Recall that the standard deviation of an exponential mean 1 is also 1. To carry out the hypothesis test, we first define a test statistic

$$T_n = \frac{\sqrt{n}}{\sqrt{1}}\left(\bar{X}_n - 1\right).$$

Under the null hypothesis H_0, the central limit theorem implies that

$$T_n \overset{\mathcal{D}}{\approx} N(0,1)$$

for large sample sizes n, where $\overset{\mathcal{D}}{\approx}$ means "approximately distributed as." Thus, to carry out a hypothesis test at level α, we first look up the constant $z = z_\alpha$ such that $\mathcal{P}(N(0,1) > z) = \alpha/2$; e.g., if $\alpha = 0.05$, then $z = 1.96$. Then our hypothesis test is as follows: if $|T_n| > z$, then reject H_0; if $|T_n| \leq z$, then do not reject H_0.

For the data given in the problem, $n = 100$ and $\bar{X}_{100} = 1.14$, so $T_{100} = 1.4$. Since $|T_{100}| < 1.96$, we do not reject the null hypothesis at level $\alpha = 0.05$. Thus, we should not be worried (at least based on this one test) about the program.