CS 661, Spring 2009
Midterm Solutions

1. (a) Since the variance is known to be 1, then density of the normal with mean μ is $f(x) = (1/\sqrt{2\pi})e^{-(x-\mu)^2/2}$. Thus, for a sample $X_1, \ldots, X_n$, the likelihood function $L(X_1, \ldots, X_n; \mu)$ is

$$L(X_1, \ldots, X_n; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(X_i-\mu)^2/2} = (2\pi)^{-n/2} \exp\left(\sum_{i=1}^{n} \frac{-(X_i-\mu)^2}{2}\right).$$

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

$$\ln L(X_1, \ldots, X_n; \mu) = -\frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2}.$$

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

$$0 = \frac{d}{d\mu}\left[-\frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2}\right] = \sum_{i=1}^{n} X_i - n\mu,$$

which implies

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Thus, the MLE of μ is $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(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 $k = 1, 2$, let $\mu_k = \mathcal{E}[X^k]$ be the kth theoretical moment and $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ be the kth 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. Note that $\ln Z = \sum_{i=1}^n \ln X_i$. When n is large, $\sum_{i=1}^n \ln X_i$ is the sum of a large number of i.i.d. random variables, which is approximately normally distributed by the central limit theorem. Also, $Z = \exp\{\ln Z\} = \exp\{\sum_{i=1}^n \ln X_i\}$, and we know that if Y is normally distributed, then e^Y has a lognormal distribution. Thus, the distribution of Z should be approximately lognormal when n is large.

3. Let $X_1, X_2, \ldots, X_n$ be the first 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, $n = 50$ and $\bar{X}_{50} = 0.79$, so $T_{50} = -1.485$. Since $|T_{50}| < 1.96$, 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.

4. (a) We can generate $N \sim \text{Poisson}(\lambda)$ using inversion as follows. First generate $U \sim \text{unif}[0,1]$, and set $X = k$ if

$$\sum_{i=0}^{k-1} p(i) < U \leq \sum_{i=0}^{k} p(i).$$

Also, note that the probability mass function satisfies $p(0) = e^{-\lambda}$ and the following holds for each $i \geq 1$:

$$p(i) = p(i-1)\frac{\lambda}{i}.$$

Hence, our complete algorithm to generate N is as follows:

1. Generate $U \sim \text{unif}[0,1]$.
2. Set $k = 0$, $P = e^{-\lambda}$ and $F = P$.
3. Do while $F \leq U$
4. Let $P = P\lambda/k$ and $F = F + P$.
5. Enddo
6. Return $N = k$.

(b) We will use inversion to generate a sample from density f_1. First, we need to compute the corresponding CDF F_1. For $0 \leq x \leq 1$, we get

$$F_1(x) = \int_0^x 2u^2\,du = \frac{2}{3}x^3.$$

For $2 \leq x \leq 4$, we get

$$F_1(x) = \int_0^1 2u^2\,du + \int_2^x ((4-u)/6)\,du = 1 - \frac{(4-x)^2}{12}.$$

Thus, the CDF is

$$F_1(x) = \begin{cases} 0 & \text{if } x < 0, \\ 2x^3/3 & \text{if } 0 \leq x \leq 1, \\ 2/3 & \text{if } 1 < x < 2, \\ 1 - (4-x)^2/12 & \text{if } 2 \leq x \leq 4, \\ 1 & \text{if } x > 4, \end{cases}$$

which we now need to invert. For $0 \leq U \leq 2/3$, we set $2x^3/3 = U$, and solve for x, giving $x = (3U/2)^{1/3}$. For $2/3 < U \leq 1$, we set $1 - (4-x)^2/12 = U$, and solve for x, giving $x = 4 \pm \sqrt{12(1-U)}$. Since we need $2 \leq x \leq 4$ when $2/3 < U \leq 1$, we set $x = 4 - \sqrt{12(1-U)}$. Hence, our algorithm for generating a sample X from density f_1 is as follows:

1. Generate $U \sim \text{unif}[0,1]$.
2. If $0 \leq U \leq 2/3$, then return $X = (3U/2)^{1/3}$, and stop.
3. If $2/3 < U \leq 1$, then return $X = 4 - \sqrt{12(1-U)}$, and stop.

(c) The CDF of the Weibull density f_2 is $F_2(x) = 1 - e^{-\sqrt{x}}$ for $x \geq 0$. Thus, to apply inversion, we set $F_2(x) = U$ for $U \sim \text{unif}[0,1]$, and solve for x. Thus, we get $1 - e^{-sqrtx} = U$, so $X = (-\ln(1-U))^2$. Hence, our algorithm for generating a sample X from density f_2 is as follows:

1. Generate $U \sim \text{unif}[0,1]$.
2. Return $X = (-\ln(1-U))^2$, and stop.

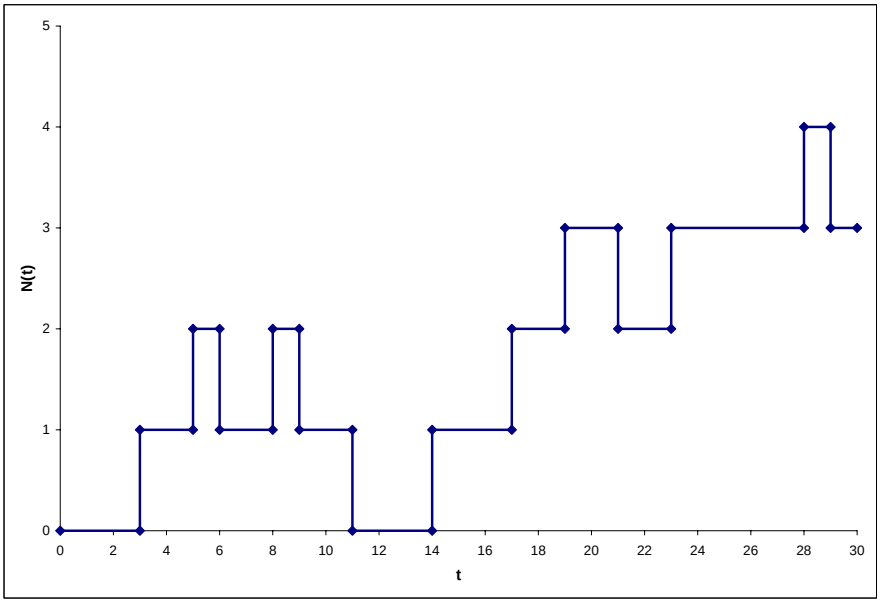
(d) We generate a sample X from density f using composition, as follows:

1. Generate independent $U_1, U_2 \sim \text{unif}[0,1]$.
2. If $U_1 \leq 3/4$, then
3. If $0 \leq U_2 \leq 2/3$, then return $X = (3U_2/2)^{1/3}$, and stop.
4. If $2/3 < U_2 \leq 1$, then return $X = 4 - \sqrt{12(1-U_2)}$, and stop.
5. If $U_1 > 3/4$, then return $X = (-\ln(1-U_2))^2$, and stop.

(e) To generate a sample of Z, we can use the following algorithm:

1. Generate $U_0 \sim \text{unif}[0,1]$.
2. Use U_0 to generate N as in part (a).
3. Set $Z = 0$.
3. For $i = 1$ to N, do
4. Generate independent $U_1, U_2 \sim \text{unif}[0,1]$, independent of any previous U's.
5. Use U_1, U_2 to generate X_i as in part (d).
6. Let $Z = Z + U_i$.
7. Enddo
8. Return Z.

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



The times t at which $N(t)$ changes are 0, 3, 5, 6, 8, 9, 11, 14, 17, 19, 21, 23, 28, 29.

(b) Calculating the area under the $N(t)$ plot from $t = 0$ to $t = 30$ and dividing by 30, we get 49/30.