ISYE 6832

Simulation Theory and Methods Homework 2

Jacob Aguirre *Email: aguirre@gatech.edu*
Instructor: Dr. Eunhye Song

Problem 1. Recall that the exponential distribution has the support, $x > 0$. The shifted exponential distribution is defined as the distribution of a random variable that has nonzero shift from an exponential distribution.

1. Derive the pdf of a shifted exponential distribution with shift c and rate λ.

   Proof. The cdf of an exponential distribution is just $1 - e^{-\lambda x}$. By introducing a constant shift c, the equation modifies to $1 - e^{-\lambda(x-c)}$ to accommodate the shift in the distribution's location. To obtain the pdf of this shifted exponential distribution, we differentiate this adjusted CDF with respect to x, yielding $\lambda e^{-\lambda(x-c)}$. □

2. Derive the MLEs of c and $\lambda, \hat{c}, \hat{\lambda}$, respectively estimated from the size$-n$ i.i.d. sample from the shifted exponential distribution, $X_1, \dots, X_n$.

   Proof. Recall that MLE equates to maximizing the log-likelihood function. So if we have n observations, the log likelihood is

   $$\log\left(\lambda^n e^{-\lambda \sum_{i=1}^n x_i - nc}\right) = n\log(\lambda) - \lambda \sum_{i=1}^n x_i - n\lambda c.$$

   By differentiating this expression with respect to c and λ,

   $$\frac{d}{dc}(n\log(\lambda) - \lambda \sum_{i=1}^n x_i - n\lambda c) = -n\lambda,$$

   $$\frac{d}{d\lambda}(n\log(\lambda) - \lambda \sum_{i=1}^n x_i - n\lambda c) = \frac{n}{\lambda} - \sum_{i=1}^n x_i - nc.$$

   Clearly, the optimal value of c, denoted as $\hat{c}$, is the minimum of the sample set $\{X_1, X_2, \dots, X_n\}$, given the monotonically increasing nature of the log-likelihood function with respect to c. Furthermore, by equating the derivative with respect to λ to zero, we find the estimator $\hat{\lambda}$ as $\frac{1}{\bar{X} - \min\{X_1, X_2, \dots, X_n\}}$, where $\bar{X}$ is the sample mean. □

3. Do these MLEs achieve consistency (in-probability convergence to the true c and λ)? Prove your claim.

Proof. According to the weak law of large numbers and Slutsky's theorem, both estimators demonstrate consistency. This implies that $\bar{X}$ approaches the actual mean $\frac{1}{\lambda} + c$ as the sample size increases, and the limit of the minimum value of the sample set, $\lim_{n\to\infty} \min\{X_1, \ldots, X_n\}$, converges to c. Consequently, the probability that the absolute difference between $\hat{\lambda}_n$ and λ is less than any positive number ϵ approaches 1, $P(|\hat{\lambda}_n - \lambda| < \epsilon) = 1$, and similarly, the probability that the absolute difference between $\hat{c}_n$ and c is less than ϵ also approaches 1 as n tends to infinity, $\lim_{n\to\infty} P(|\hat{c}_n - c| < \epsilon) = 1$, aligning with the expectations. Thus, we're done. □

4. In this case, we can obtain the exact sampling distribution of $\sqrt{n}(\hat{c} - c)$. Find the sampling distribution and show your derivation. What is the asymptotic distribution of $\sqrt{n}(\hat{c} - c)$?

Proof. We already know that convergence in probability occurs. Since convergence pertains to a specific point, and again recalling that convergence in probability suggests almost sure convergence, we almost surely get

$$\lim_{n\to\infty} \sqrt{n}(\hat{c} - c) = 0.$$

□

5. Derive the asymptotic distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$.

Proof. Let us first define $X_1 = \min\{X_1, \ldots, X_n\}$ and introduce a function $Y = \frac{1}{\bar{X} - X_1}$. Then we can just take $\lim_{n\to\infty}$ as

$$\begin{aligned}
\lim_{n\to\infty} \sqrt{n}(\hat{\lambda} - \lambda) &= \lim_{n\to\infty} \sqrt{n}\left(\frac{1}{\bar{X} - X_1} - \lambda\right) \\
&= \lim_{n\to\infty} \sqrt{n}(Y - \lambda) \\
&= \lim_{n\to\infty} \sqrt{n}\left(Y_0 + \frac{1}{y_0^2}\left(Y - \frac{1}{\lambda}\right) - \lambda\right) \\
&= \lim_{n\to\infty} \sqrt{n}\left(\frac{1}{y_0^2}\left(Y - \frac{1}{\lambda}\right)\right) \\
&= \lim_{n\to\infty} \sqrt{n}\frac{1}{y_0^2}\left((\bar{X} - c - \lambda) - (X_1 + c)\right)^{-1} \\
&\xrightarrow[\text{using slutsky}]{P} \lim_{n\to\infty} \sqrt{n}\frac{1}{y_0^2}\left(\bar{X} - C - \lambda\right)^{-1} \\
&\xrightarrow[\text{by c.l.t.}]{} \mathcal{N}(0, \frac{1}{\lambda^2})
\end{aligned}$$

Thus we have convergence to a normal distribution with mean 0 and variance $\frac{1}{\lambda^2}$ through Slutsky's theorem and the central limit theorem .

□

Problem 2. Recall that in class we showed that the interpolated ecdf is pointwise (e.g., once $x \in \mathbb{R}$ is fixed) strongly consistent with the true cdf of the i.i.d. observations. Show that the interpolated ecdf is uniformly strongly consistent.

Proof. Suppose we have an interval with a certain $\epsilon > 0$. In creating an ordered set of points, let's call it Ω in doing so we should select a step size $k > \frac{1}{\epsilon}$ so that our ordered set of points satisfies

$$-\infty = \Omega_0 < \Omega_1 \leq \cdots \leq \Omega_k = \infty.$$

The intuition for the above is just that we're effectively segmenting the support of F, which here spans $(-\infty, \infty)$, into disjoint intervals. So within each interval ℓ, the function F's left limit and its actual value at Ω_ℓ are bound by the step $\frac{\ell}{k}$

$$F(\Omega_\ell^-) \leq \frac{\ell}{k} \leq F(\Omega_\ell).$$

Hence, if $\Omega_\ell > \Omega_{\ell-1}$, it follows that

$$F(\Omega_\ell^-) - F(\Omega_{\ell-1}) < \epsilon.$$

Since we have point-wise convergence,

$$F_n(\Omega_\ell) \underset{a.s.}{\longrightarrow} F(\Omega_\ell) \quad \text{and} \quad F_n(\Omega_\ell^-) \underset{a.s.}{\longrightarrow} F(\Omega_\ell^-).$$

In other words, the empirical distribution function at these points will converge to the true distribution function with a probability of 1 as we collect more and more data. Obviously the difference between these convergences narrows to zero, almost surely and so,

$$|F_n(\Omega_\ell) - F(\Omega_\ell)| \underset{a.s.}{\longrightarrow} 0 \quad \text{and} \quad |F_n(\Omega_\ell^-) - F(\Omega_\ell^-)| \underset{a.s.}{\longrightarrow} 0.$$

Therefore, for all ℓ as $n \to \infty$,

$$\Delta_n = \max_{\ell \in \mathcal{L}} \{|F_n(\Omega_\ell) - F(\Omega_\ell)|, |F_n(\Omega_\ell^-) - F(\Omega_\ell^-)|\} \underset{a.s.}{\longrightarrow} 0.$$

For any x within F's support, we need to locate Ω_ℓ such that x falls within $[\Omega_{\ell-1}, \Omega_\ell]$. Again, recall that for any point x, $F_n(x)$ is just the proportion of observations that are less than or equal to x. It follows that

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(\Omega_\ell^-) - F(\Omega_{\ell-1}) \leq F_n(\Omega_\ell^-) - F(\Omega_\ell^-) + \epsilon, \\ F_n(x) - F(x) &\geq F_n(\Omega_{\ell-1}) - F(\Omega_\ell^-) \geq F_n(\Omega_{\ell-1}) - F(\Omega_{\ell-1}) - \epsilon. \end{aligned}$$

The first equation above gives an upper bound on the difference between the empirical and true distribution functions at x. It says that this difference is at most as large as the difference between the empirical and true distribution functions at the right endpoint of the interval (Ω_ℓ^-) plus an error term ϵ. The second inequality provides a lower bound on the difference between the empirical and true distribution functions at x. Continuing, we get

$$F_n(\Omega_{\ell-1}) - F(\Omega_{\ell-1}) - \epsilon \leq F_n(x) - F(x) \leq F_n(\Omega_\ell^-) - F(\Omega_\ell^-) + \epsilon.$$

The lefthand side of the above sets a lower limit anf the right hand side sets an upper limit (notice the difference for the ϵ terms. So for any x, we assert that as $n \to \infty$,

$$|F_n(x) - F(x)| \leq \Delta_n + \epsilon \underset{a.s.}{\longrightarrow} \epsilon \qquad \Longrightarrow \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \underset{a.s.}{\longrightarrow} \epsilon.$$

Finally, suppose we denote A to be the set of all (x, ϵ) where the convergence is confirmed. Thus, A can be described as an intersection and a limiting bound,

$$A := \bigcap_{\epsilon > 0} A_\epsilon := \lim_{\epsilon \to 0} A_\epsilon.$$

Given the convergence is evident for all x given any ϵ, then $1 = \lim_{\epsilon \to 0} P(A_\epsilon) = P(A)$, thereby validating our proposition. □

Problem 3. Recall that you learned the polar method to generate two independent $\mathcal{N}(0,1)$ random variables. Prove that the algorithm indeed is correct. That is, the two-dimensional vector it returns has the correct distribution.

Proof. When selecting a random point (X, Y) from a uniform distribution over the square $(-1,1) \times (-1,1)$, we consider the initial radius $U = X^2 + Y^2$. Analyzing U's distribution, we observe:

$$\begin{aligned} P(U \leq r) &= P(X^2 + Y^2 \leq r) \\ &= \frac{\text{area of circle with radius } \sqrt{r}}{\text{area of square}} \\ &= \frac{\pi \cdot r}{4\pi} \\ &= \frac{r}{4}, \end{aligned}$$

indicating a correction in the calculation to properly reflect the square's area and the proportion of the circle's area within it. This suggests U scales uniformly within $[0,1]$ given the correct normalization. For the transformation $R^2 = -2\log(U)$, if U is uniformly distributed, then R^2 follows an exponential distribution with parameter 2.

The angle θ, created by X and Y, is a uniform distribution from 0 to 2π due to the symmetry and uniformity of points within the square. This can be seen by considering the geometry of the situation: any direction from the origin within the square is equally likely, leading to

$$\begin{aligned} P(\theta \leq \phi) &= \frac{\text{arc length corresponding to } \phi}{\text{total circumference}} \\ &= \frac{\phi}{2\pi}, \end{aligned}$$

which confirms $\theta \sim U(0, 2\pi)$. Notice that the independence of radius and angle in this setup is exactly the Box-Muller transformation, where the inherent independence allows for us to generate the normally distributed variables. By Box-Muller method, where R^2 and θ are used to produce Gaussian variables, e.g., $X, Y \sim \mathcal{N}(0,1)$. □

Problem 4. Recall that in class, we studied random number generation and Box-Müeller method.

1. Implement a Python code that does the following:
   - Linear congruential generators that generate uniform random numbers z_0, a, m, and c.
   - Standard normal random variate generator that returns two i.i.d. $\mathcal{N}(0,1)$ variables using the LCG implemented above.
2. For two settings of LCGs, $(m, a, c) = (64, 21, 1)$ and $(64, 37, 1)$, respectively, produce standard normal RVs from the algorithm you implemented. Use the full period (both are full period generators. Which of the LCGs has better performances? Support your answer with your experiment results.

Proof. I just use python for coding this.

```
def generate_lcg(initial_seed, multiplier, increment, modulus) -> list:
    current_value = (multiplier * (initial_seed + increment)) % modulus
    random_sequence = [current_value / modulus]
    while True:
        current_value = (multiplier * (current_value + increment)) %mod
        new_random = current_value / modulus
        if new_random in random_sequence:
            break
        random_sequence.append(new_random)
    return random_sequence
```

Similarly for Box-Müeller method, we have

```
def generate_gaussian_pair(v1, v2):
    radius = (-2 * math.log(v1)) ** 0.5
    angle = 2 * math.pi * v2
    gaussian_x = radius * math.cos(angle)
    gaussian_y = radius * math.sin(angle)
    return gaussian_x, gaussian_y
```

To answer the second part of this question, we get a figure that looks like an Amoeba (if you're familiar with tropical geometry) that has 3 "tentacles" that spiral out from a central "locus" in the case of the non-prime generator. The other points (which stem from using the full period prime generator) don't look like they have any pattern. □

Problem 5 (6.23). We claim the following algorithm implements the inverse cdf method to generate Poisson random variables with mean λ; that is, $Pr\{X = x\} = e^{-\lambda}\lambda^x/x!$ for $x = 0, 1, 2, \ldots$. The algorithm is as follows,

1. Set $c = d = e^{-\lambda}$ and $X = 0$

```
2. Generate $U \sim U(0,1)$
3. Until $U \leq c$ do
     • $X = X + 1$
     • $d = d\lambda/X$
     • $c = c + d$
     Loop
4. Return $X$
```

First, prove that this algorithm does as claimed. Then derive the expected number of times that the loop is executed for each random variate generated. Notice that it is an increasing function of λ, so the algorithm becomes less and less efficient as λ increases. *Hint:* $Pr\{X = x\} = \lambda Pr\{X = x - 1\}/x$ for $x \geq 1$.

Proof. We seek to provide a certificate, per say, that the algorithm above does as claimed. So, let's prove this by looking at the base case of

$$P(X = 1) = \lambda e^{-\lambda}.$$

We then see that

$$d_k = \frac{\lambda}{k} d_{k-1} = \frac{\lambda^k e^{-\lambda}}{k!} = P(X = k),$$

which aligns with the PMF for the Poisson distribution. Furthermore,

$$P(X < k) = \sum_{i=1}^{k} P(X = i) = \sum_{i=1}^{k} d_k = c_k.$$

Thus we've confirmed the algorithm's functionality. For the analysis of the average iterations required, simply calculate

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot P(X = i) \\ &= \sum_{i=0}^{\infty} i \cdot \frac{\lambda^i e^{-\lambda}}{i!} \\ &= \lambda, \end{aligned}$$

underscoring that the expected count of iterations equates to λ, and so we're done. □

Problem 6 (6.30). Let X be a discrete-valued random variable that takes values $x_0 < x_1 < \ldots < x_n$ with probabilities $p_0, p_1, \ldots, p_n$, respectively. Derive, and prove the correctness of a rejection algorithm that first generates an index $I \in \{0, 1, \ldots, n\}$ using the discrete uniform distribution. What is the expected number of trials for your algorithm to generate one X? *Hint:* The following simple algorithm works but is very inefficient: (i) Generate $W \sim U(0,1)$ and set $I = \lceil (n+1)W \rceil - 1$. (ii) Generate $U \sim U(0,1)$. (iii) If $U \leq p_I$ then return $X = x_I$; otherwise go to step (i). First prove that this algorithm works, and that is has expected number of trials $n+1$. Then, derive a better algorithm by applying the concept of a majorizing function as used in the continuous case.

Proof. Observe that $P[X = x_i] = c \cdot p_i$, implying

$$\begin{aligned} P[X = x_i] &= P[U \leq p_i] \cdot P[I = i] \\ &= p_i \cdot P[I = i] \\ &= p_i \cdot P[\lceil (n+1) \cdot W \rceil - 1 = i] \\ &= p_i \cdot P[(n+1) \cdot W \leq i + 1 \mid (n+1) \cdot W \geq i] \cdot P[(n+1) \cdot W \geq i] \\ &= p_i \cdot \frac{1}{n+1}. \end{aligned}$$

So we've shown the algorithm's validity by accurately mirroring the pdf of the distribution of X on a reduced scale and providing a valid certificate. If we let Y represent the iterations required until termination, then clearly

$$\begin{aligned} \text{taking expectation } \mathbb{E}[Y] &= \sum_{j=1}^{n+1} P[Y = j] \\ &= \sum_{j=1}^{n+1} \sum_{i=1}^{n} P[X = x_i] \implies = \sum_{j=1}^{n+1} 1 \\ \text{and so } &= n + 1, \end{aligned}$$

allowing us to conclude that the expected number of draws before halting is $n + 1$.

Now, let's discuss a potential "better" algorithm, that I *attempt* to justify but am uncertain of. Suppose we take $c = \max p_0, p_1, \ldots, p_n$. We will then repeat the following,

- Generate I uniformly from $0, 1, \ldots, n$.
- Generate $U \sim U(0,1)$.
- If $U \leq \frac{p_I}{c}$, return $X = x_I$.

The choice of I is uniformly random, ensuring each x_I has an initial equal chance of selection. We also scale the acceptance probability by $\frac{1}{c}$, which ensures that the overall acceptance rate is proportional to the true distribution of X, correcting for the uniform selection of I. With c as defined, the success probability in each trial is $\sum_{i=0}^{n} \frac{1}{n+1} \cdot \frac{p_i}{c} = \frac{1}{c(n+1)}$. The expected number of trials is therefore $c(n+1)$. While the simple algorithm has an expected number of trials of $n + 1$, the improved algorithm adjusts the expected number based on the distribution's maximum probability. For distributions where c is close to $\max p_0, p_1, \ldots, p_n$, this method significantly reduces the expected number of trials, especially when $c(n + 1) < n + 1$, which is generally the case unless all p_i are equal. $\square$