Assignment 2: CS215

Satyam Sinoliya, 23B0958
Vaibhav Singh, 23B1068
Shaik Awez Mehtab, 23B1080

Solution 1

Task A

When $X \sim \text{Ber}(p)$, PGF of X is

$$G_{\text{Ber}}(z) = \mathbb{E}(z^X) \tag{1}$$
$$= \sum_{n=0}^{\infty} P[X = n]z^n \tag{2}$$

Since $P[X = 0] = (1 - p)$, $P[X = 1] = p$, $P[X = n] = 0$ when $n > 1$,

$$G_{\text{Ber}}(z) = P[X = 0]z^0 + P[X = 1]z^1 \tag{3}$$
$$= (1 - p) + pz \tag{4}$$

Task B

When $X \sim \text{Bin}(n, p)$, PMF of X is

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \text{ for } k \leq n. \tag{5}$$

and $P[X = k] = 0$ for $k > n$.

$$G_{\text{Bin}}(z) = \sum_{k=0}^{\infty} P[X = k]z^k \tag{6}$$
$$= \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} z^k \tag{7}$$
$$= \sum_{k=0}^{n} \binom{n}{k} (pz)^k (1 - p)^{n-k} \tag{8}$$
$$= (1 - p + pz)^n. \tag{9}$$

By equation 4, $G_{\text{Bin}}(z) = (1 - p + pz)^n = (G_{\text{Ber}}(z))^n$. Hence proved.

Task C

By the definition of PGF,

$$G(z) = \sum_{n=0}^{\infty} P[X_1 = n]z^n \tag{10}$$

Let $(G(z))^k = \sum_{n=0}^{\infty} a_n z^n$. Now, $G_\Sigma(z)$ is

$$G_\Sigma(z) = \sum_{n=0}^{\infty} P[X = n]z^n \tag{11}$$
$$= \sum_{n=0}^{\infty} P[X_1 + \cdots + X_k = n]z^n \tag{12}$$
$$= \sum_{n=0}^{\infty} \sum P[X_1 = i_1, \ldots, X_k = i_k]z^n \tag{13}$$

where $i_1 + \cdots + i_k = n$.

$$= \sum_{n=0}^{\infty} \sum P(i_1) \ldots P(i_n)z^n \tag{14}$$

Now, $G(z) = \sum_{n=0}^{\infty} P(n)z^n$. And since a_n is coefficient of z^n in $(G(z))^k = (\sum_{n=0}^{\infty} P(n)z^n)^k$.

$$a_n = \sum P(i_1)P(i_2)\dots P(i_k) \text{ where } i_1 + \cdots + i_k = n \tag{15}$$

By equation 14

$$G_\Sigma(z) = \sum_{n=0}^{\infty} a_n z^n = (G(z))^k \tag{16}$$

Hence Proved.

Task D

When $X \sim \text{Geo}(p)$, PMF of X,

$$P[X = k] = (1-p)^{k-1}p \tag{17}$$

for $k > 0$. $P[X = 0] = 0$. Now, PGF of X,

$$G_{\text{Geo}}(z) = \sum_{k=0}^{\infty} P[X = k]z^k \tag{18}$$

$$= \sum_{k=1}^{\infty} P[X = k]z^k \tag{19}$$

$$= \sum_{k=1}^{\infty} p(1-p)^{k-1}z^k \tag{20}$$

$$= \sum_{k=1}^{\infty} pz(z - zp)^{k-1} \tag{21}$$

$$= pz\sum_{k=0}^{\infty} (z - zp)^k \tag{22}$$

$$= \frac{pz}{1 - z + pz} \tag{23}$$

Task E

By equation 9,

$$G_{\text{Bin}}(z) = (1 - p + pz)^n = G_X^{(n,p)}(z). \tag{24}$$

For $Y \sim \text{NegBin}(n, p)$, Y represents the number of independent coin throws required to get n heads of a coin. Let X_i represents the number of throws of coin required after getting $(i-1)^{\text{th}}$ head to get the i^{th} head. Since all of the coin throws are independent, the outcome of a given throw doesn't depend on the previous coins' output. Thus, X_i is just the number of throws to get a head when a coin in thrown, where each $X_i \sim \text{Geo}(p)$ since each coin is same with probability of getting head as p.
Y can be written as $Y = X_1 + X_2 + \cdots + X_k$. Using equations 16 and 23,

$$G_Y^{(n,p)}(z) = (G_{\text{Geo}}(z))^n \tag{25}$$

$$= \left(\frac{pz}{1 - z + pz}\right)^n \tag{26}$$

$$\tag{27}$$

$$G_X^{(n,p^{-1})}(z^{-1}) = \left(1 - \frac{1}{p} + \frac{1}{pz}\right)^n \tag{28}$$

$$= \left(\frac{1 - z + pz}{pz}\right)^n \tag{29}$$

$$\left(G_X^{(n,p^{-1})}(z^{-1})\right)^{-1} = \left(\frac{pz}{1 - z + pz}\right)^n \tag{30}$$

$$= G_Y^{(n,p)}(z) \tag{31}$$

Hence Proved.

Task F

For $Y \sim \text{NegBin}(n, p)$,

$$P[Y = k] = \binom{k-1}{n-1} p^n (1-p)^{k-n} \text{ for } k \geq n \tag{32}$$

Otherwise, $P[Y = k] = 0$. PGF of Y is

$$G_Y^{(n,p)}(z) = \sum_{k=0}^{\infty} P[Y = k] z^k \tag{33}$$

$$= \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} z^k \tag{34}$$

$$= \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} p^n (1-p)^k z^{n+k} \tag{35}$$

$$= (pz)^n \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} (z - pz)^k \tag{36}$$

Using equation 27,

$$\left(\frac{pz}{1 - z + pz}\right)^n = (pz)^n \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} (z - pz)^k \tag{37}$$

$$(1 - (z - pz))^{-n} = \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} (z - pz)^k \tag{38}$$

Since z, p are arbitrary, let $z - pz = x$.

$$(1 - x)^{-n} = \sum_{r=0}^{\infty} \binom{r+n-1}{n-1} x^r = \sum_{r=0}^{\infty} \binom{-n}{r} x^r \tag{39}$$

Hence proved.

Task G

To prove: *Given PGF of a random variable X is $G(z)$, expectation of X i.e $\mathbb{E}(x) = G'(1)$*
Proof:

$$G(z) = \mathbb{E}(z^X) = \sum_{k=0}^{\infty} P[X = k] z^k \tag{40}$$

$$G'(z) = \sum_{k=0}^{\infty} kP[X = k] z^{k-1} \tag{41}$$

$$G'(1) = \sum_{k=0}^{\infty} kP[X = k] \tag{42}$$

$$= \mathbb{E}[X] \tag{43}$$

Hence Proved. Now, Let's derive means of Bernoulli, Binomial, Geometric and Negative Binomial distributions using this:

1. **Bernoulli Distribution:** Let $X \sim \text{Ber}(p)$,

$$G_{\text{Ber}}(z) = (1-p) + pz \tag{44}$$
$$G'_{\text{Ber}}(z) = p \tag{45}$$
$$G'_{\text{Ber}}(1) = p = \mathbb{E}[X] \tag{46}$$

   Thus, $\mathbb{E}[X] = p$.

2. **Binomial Distribution:** Let $X \sim \text{Bin}(n, p)$,

$$G_{\text{Bin}}(z) = (1-p+pz)^n \tag{47}$$
$$G'_{\text{Bin}}(z) = np(1-p+pz)^{n-1} \tag{48}$$
$$G'_{\text{Bin}}(1) = np = \mathbb{E}[X] \tag{49}$$

   Thus, $\mathbb{E}[X] = np$.

3. **Geometric Distribution:** Let $X \sim \text{Geo}(p)$,

$$G_{\text{Geo}}(z) = \frac{pz}{1-z+pz} \tag{50}$$
$$G'_{\text{Geo}}(z) = \frac{p(1-z+pz) - pz(p-1)}{(1-z+pz)^2} \tag{51}$$
$$= \frac{p}{(1-z+pz)^2} \tag{52}$$
$$G'_{\text{Geo}}(1) = \frac{p}{p^2} = \frac{1}{p} = \mathbb{E}[X] \tag{53}$$

   Thus, $\mathbb{E}[X] = \frac{1}{p}$.

4. **Negative Binomial Distribution:** Let $X \sim \text{NegBin}(n, p)$,

$$G_{\text{NegBin}}(z) = \left(\frac{pz}{1-z+pz}\right)^n \tag{54}$$
$$G'_{\text{NegBin}}(z) = n\left(\frac{pz}{1-z+pz}\right)^{n-1}\left(\frac{p}{(1-z+pz)^2}\right) \tag{55}$$
$$G'_{\text{NegBin}}(1) = \frac{n}{p} = \mathbb{E}[X] \tag{56}$$

   Thus, $\mathbb{E}[X] = \frac{n}{p}$.

Solution 2

Task A

To prove:
Let X be a continuous real-valued random variable with CDF : $\mathbb{R} \to [0,1]$. Assume that F_X is invertible. Then the random variable $Y := F_X(X) \in [0,1]$ is uniformly distributed in $[0,1]$
Proof:
F_X by definition can also be written as

$$F_X(x) = P(X \leq x) \tag{57}$$

Define a new random variable Y,

$$Y = F_X(X) \tag{58}$$

Y is the result of applying CDF F_X to the random variable X. To prove the theorem, assume $y \in [0, 1]$. So, the probablity that $Y \leq y$ is:

$$P(Y \leq y) = P(F_X(X) \leq y) \tag{59}$$

It is assumed that $F_X(x)$ is invertible, so,

$$P(Y \leq y) = P(X \leq F_X^{-1}(y)) \tag{60}$$

which is basically, probablity that X is less that or equal to $F_X^{-1}(y)$. This can be written in the CDF form, which is $F_X(F_X^{-1}(y))$. So,

$$P(Y \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y \tag{61}$$

So,

$$P(Y \leq y) = y \tag{62}$$

where $y \in [0, 1]$, which is the CDF of uniform distributon in $[0, 1]$. So, Y is a uniform distributon in $[0, 1]$ regardless of X.

Task B

According to the theorem proved above, CDF of any random variable X mapped with itself gives a uniform random variable Y in $[0, 1]$. So, let $Y \sim \text{Uniform}(0, 1)$. Then for any random variable X,

$$F_X(X) = Y \tag{63}$$

$$X = F_X^{-1}(Y) \tag{64}$$

Algorithm $\mathcal{A}$:

1. Input: A sample y from the uniform distributon on $[0, 1]$.
2. Transformation:
   - Apply the inverse CDF to y to compute a sample u.
   - Define $\mathcal{A}(u) = u = F_X^{-1}(y)$
3. Output: The random variable $U = F_X^{-1}(Y)$

This gives us the correct required random variables as, CDF of U is $F_U(u)$,

$$P(U \leq u) = P(F_X(Y) \leq u) \tag{65}$$

$$F_U(u) = P(F_X(F_X^{-1}(X) \leq u)) \tag{66}$$

$$F_U(u) = P(X \leq u) \tag{67}$$

$$F_U(u) = F_X(u) \tag{68}$$

$$\tag{69}$$

U and X have the same CDF, which was initially required.
Proof:
F_X by definition can also be written as

$$F_X(x) = P(X \leq x) \tag{70}$$

Define a new random variable Y,

$$Y = F_X(X) \tag{71}$$

Y is the result of applying CDF F_X to the random variable X. To prove the theorem, assume $y \in [0,1]$. So, the probability that $Y \leq y$ is:

$$P(Y \leq y) = P(F_X(X) \leq y) \tag{72}$$

It is assumed that $F_X(x)$ is invertible, so,

$$P(Y \leq y) = P(X \leq F_X^{-1}(y)) \tag{73}$$

which is basically, probability that X is less that or equal to $F_X^{-1}(y)$. This can be written in the CDF form, which is $F_X(F_X^{-1}(y))$. So,

$$P(Y \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y \tag{74}$$

So,

$$P(Y \leq y) = y \tag{75}$$

where $y \in [0,1]$, which is the CDF of uniform distributon in $[0,1]$. So, Y is a uniform distributon in $[0,1]$ regardless of X.

Task B

According to the theorem proved above, CDF of any random variable X mapped with itself gives a uniform random variable Y in $[0,1]$. So, let $Y \sim \text{Uniform}(0,1)$. Then for any random variable X,

$$F_X(X) = Y \tag{76}$$

$$X = F_X^{-1}(Y) \tag{77}$$

Algorithm $\mathcal{A}$:

1. Input: A sample y from the uniform distributon on $[0,1]$.
2. Transformation:
   - Apply the inverse CDF to y to compute a sample u.
   - Define $\mathcal{A}(u) = u = F_X^{-1}(y)$
3. Output: The random variable $U = F_X^{-1}(Y)$

This gives us the correct required random variables as, CDF of U is $F_U(u)$,

$$P(U \leq u) = P(F_X(Y) \leq u) \tag{78}$$

$$F_U(u) = P(F_X(F_X^{-1}(X) \leq u)) \tag{79}$$

$$F_U(u) = P(X \leq u) \tag{80}$$

$$F_U(u) = F_X(u) \tag{81}$$

$$\tag{82}$$

U and X have the same CDF, which was initially required.

Task E(B)

We have a random variable X which can take values from $\{-h, -h+2, \ldots, h-2, h\}$. Each ball makes h random binary decisions(left or right) as it descends. If we let Y be the number of times the ball moves right, the final position of the ball will be given by,

$$X = -h + 2Y \tag{83}$$

where Y is a **binomial variable** because in simple terms it is the summation of h bernoulli decisions each with probability $\frac{1}{2}$.

$$Y \sim \text{Bin}(h, \frac{1}{2}) \tag{84}$$

For a particular pocket $X = 2i$, the corresponding value of Y is:

$$Y = \frac{h+2i}{2} \tag{85}$$

Thus, the probability that the ball lands in the pocket $X = 2i$ is the probability that $Y = \frac{h+2i}{2}$. Using Binmial distribution, this is:

$$P_h[X = 2i] = P_h\left[Y = \frac{h+2i}{2}\right] = \binom{h}{\frac{h+2i}{2}}\left(\frac{1}{2}\right)^h \tag{86}$$

This is the **closed form expression for** $P_h[X = 2i]$
Now, we need to show $P_h[X = 2i]$ approximates to normal distribution for very large h.
Using **stiriling's approximation** for large n, which states that:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{2}\right)^n \tag{87}$$

We can convert the factorials in the binomial coefficient:

$$\binom{h}{r} \approx \frac{h!}{r!(h-r)!} \tag{88}$$

Using stirlings approximation, we have,

$$\binom{h}{y} = \frac{\sqrt{2\pi h}\left(\frac{h}{e}\right)^h}{\sqrt{2\pi y}\,(\frac{y}{e})^y \cdot \sqrt{2\pi(h-y)}\left(\frac{h-y}{e}\right)^{h-y}} \tag{89}$$

where $y = \frac{h+2i}{2}$ For large h, we can simplify this assuming small i (relative to h). In particular, $\frac{h+2i}{2}$ can be written as $\frac{h}{2}$, leading to:

$$\binom{h}{\frac{h+2i}{2}} \approx \frac{2^h}{\sqrt{\pi h}}e^{-\frac{2i^2}{h}} \tag{90}$$

Substituting it back in P_h gives:

$$P_h[X = 2i] \approx \frac{2^h}{\sqrt{\pi h}}e^{-\frac{2i^2}{h}}\left(\frac{1}{2}\right)^h \tag{91}$$

Simplifying the powers of 2 gives:

$$P_h[X = 2i] \approx \frac{1}{\sqrt{\pi h}}e^{-\frac{2i^2}{h}} \tag{92}$$

which is basically normal distribution with $\mu = 0$ and $\sigma^2 = h/2$

Solution 3

Task D

Given, PDF of Gamma-distribution Gamma(k, θ) is $f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$. First moment of it is

$$\mu_1^{\text{Gamma}} = \mathbb{E}[X] \tag{93}$$
$$= \int_{-\infty}^{\infty} x \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} dx \tag{94}$$
$$= \frac{1}{\theta^k \Gamma(k)} \int_{-\infty}^{\infty} x^k e^{-\frac{x}{\theta}} dx \tag{95}$$

Let $u = \frac{x}{\theta}$, then $\theta du = dx$

$$= \frac{\theta}{\Gamma(k)} \int_{-\infty}^{\infty} u^k e^{-u} du \tag{96}$$
$$= \frac{\theta \Gamma(k+1)}{\Gamma(k)} \tag{97}$$

Since $\Gamma(k+1) = k\Gamma(k)$

$$= k\theta \tag{98}$$

Second moment of it is

$$\mu_2^{\text{Gamma}} = \mathbb{E}[X^2] \tag{99}$$
$$= \int_{-\infty}^{\infty} x^2 \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} dx \tag{100}$$
$$= \frac{1}{\theta^k \Gamma(k)} \int_{-\infty}^{\infty} x^{k+1} e^{-\frac{x}{\theta}} dx \tag{101}$$

Let $u = \frac{x}{\theta}$, then $\theta du = dx$

$$= \frac{\theta^2}{\Gamma(k)} \int_{-\infty}^{\infty} u^{k+1} e^{-u} du \tag{102}$$
$$= \frac{\theta^2 \Gamma(k+2)}{\Gamma(k)} \tag{103}$$
$$= (k+1)k\theta^2 \tag{104}$$

Thus, $\mu_1^{\text{Gamma}} = k\theta$, $\mu_2^{\text{Gamma}} = (k+1)k\theta^2$.
Estimate the best gamma distribution approximation to the given data by equating first and second moments. For $X \sim \text{Gamma}(n, p)$, density function is

$$f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \tag{105}$$

$\mu_1 = k\theta$ and $\mu_2 = k(k+1)\theta^2$.

Task E

- Log likelihood for binomimal distribution ≈ -2.157
- Log likelihood for gamma distribution $= -$`inf`

Thus binomial distribution is a better fit for the data.

Task F

- Log likelihood for two-component Gaussian mixture is -2.183

Thus, binomial distribution is slightly better fit than the given two-component gaussian mixture (whose variance is assumed to be 1).

Solution 4

Solution 5

Task A

Given, PDF of GMM variable X is $f_X = \sum_{i=1}^{K} p_i P[X_i = x]$. Let it's CDF be F_X. Then $F_X(x)$ is given by

$$
\begin{aligned}
F_X(x) &= P[X \leq x] && (106)\\
&= \int_{-\infty}^{x} f_X(t)dt && (107)\\
&= \int_{-\infty}^{x} \sum_{i=1}^{K} p_i P[X_i = t]dt && (108)\\
&= \sum_{i=1}^{K} p_i \int_{-\infty}^{x} P[X_i = t]dt && (109)\\
&= \sum_{i=1}^{K} p_i P[X_i \leq x] && (110)\\
&= \sum_{i=1}^{K} p_i F_{X_i}(x) && (111)
\end{aligned}
$$

Where $F_{X_i}(x) = P[X_i \leq x]$ is CDF of X_i.
Now, let CDF of output of the given algorithm be $F_{\mathcal{A}}(x) = P[\mathcal{A} \leq x]$. Since the events that we choose $\mathcal{A}$ to be from the distribution i (say E_i) are disjoint for $i = 1, \ldots, k$.

$$
\begin{aligned}
F_{\mathcal{A}}(x) &= P[\mathcal{A} \leq x] && (112)\\
&= \sum_{i=1}^{K} P[E_i] \cdot P[\mathcal{A} \leq x | E_i] && (113)\\
&= \sum_{i=1}^{K} p_i F_{X_i}(x) && (114)\\
&= F_X(x) && (115)
\end{aligned}
$$

We know that PDF of a random variable X with CDF $F_X(x)$ is $\frac{\partial F_X}{\partial x}$. Thus,

$$
\begin{aligned}
f_{\mathcal{A}}(x) &= \frac{\partial F_{\mathcal{A}}}{\partial x} && (116)\\
&= \frac{\partial F_X}{\partial x} && (117)\\
&= f_X && (118)
\end{aligned}
$$

Since x was arbitrary, for every $u \in \mathbb{R}$, $f_{\mathcal{A}}(u) = f_X(u)$. i.e the algorithm indeed samples from the GMM variable's distribution.

Task B

Since

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} t \cdot P[X = t]dt && (119)\\
&= \int_{-\infty}^{\infty} t \cdot \sum_{i=1}^{K} p_i P[X_i = t]dt && (120)\\
&= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} P[X_i = t]dt && (121)\\
&= \sum_{i=1}^{K} p_i \mathbb{E}[X_i] && (122)\\
&= \sum_{i=1}^{K} p_i \mu_i && (123)
\end{aligned}$$

Let $\mu = \mathbb{E}[X]$.

$$\begin{aligned}
\text{Var}[X] &= \int_{-\infty}^{\infty} (t-\mu)^2 P[X = t]dt && (124)\\
&= \int_{-\infty}^{\infty} (t-\mu)^2 \sum_{i=1}^{K} p_i P[X_i = t]dt && (125)\\
&= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} (t-\mu)^2 P[X_i = t]dt && (126)\\
&= \sum_{i=1}^{K} p_i \text{Var}[X_i] && (127)\\
&= \sum_{i=1}^{K} p_i \sigma_i^2 && (128)
\end{aligned}$$

Let $\sigma^2 = \text{Var}[X]$.

$$\begin{aligned}
\text{MGF}_X(t) &= \int_{-\infty}^{\infty} e^{tX} P[X = x]dx && (129)\\
&= \int_{-\infty}^{\infty} e^{tX} \sum_{i=1}^{K} p_i P[X_i = x]dx && (130)\\
&= \sum_{i=1}^{K} p_i \int_{-\infty}^{\infty} e^{tX} P[X_i = x]dx && (131)\\
&= \sum_{i=1}^{K} p_i \text{MGF}_{X_i}(t) && (132)\\
&= \sum_{i=1}^{K} p_i e^{t\mu_i + \frac{1}{2}t^2\sigma_i^2} && (133)
\end{aligned}$$

Task C

Given $Z = \sum_{i=1}^{K} p_i X_i$, where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=1}^{K} p_i X_i\right] \tag{134}$$

$$= \sum_{i=1}^{K} p_i \mathbb{E}[X_i] \tag{135}$$

$$= \sum_{i=1}^{K} p_i \mu_i \tag{136}$$