You may use the following facts without proof.

Fact 1: The probability density function (pdf) of a chi-square χ^2_k random variable, with $k > 0$ degrees of freedom, is

$$f_{\chi^2_k}(x) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ denotes the gamma function for $\alpha > 0$.

The mean and variance of a χ^2_k random variable are given by k and $2k$, respectively.

Fact 2: The pdf of a Beta(α, β) distribution, for $\alpha, \beta > 0$, is

$$f_{\text{Beta}(\alpha,\beta)}(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & x \in (0, 1), \\ 0, & \text{otherwise.} \end{cases}$$

Part I.

Let $X_1, \ldots, X_n$ be iid continuous random variables with pdf $f_X(x) > 0$, for $x \in \mathbb{R}$, and cdf $F_X(x)$. Also let $U_1, \ldots, U_n$ be iid uniform$[0, 1]$ random variables. Be sure to justify your answers, stating any standard results used.

1. Show that $F_X(X_1)$ follows the same distribution as U_1.

2. For $x \in \mathbb{R}$, let $F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i \le x\}$ be the empirical cdf of $X_1, \ldots, X_n$, where $\mathbf{1}\{\cdot\}$ is an indicator function. Show that

$$\sqrt{n}(F_n(x) - F_X(x)) \xrightarrow{d} N\left(0, F_X(x)(1 - F_X(x))\right)$$

holds as $n \to \infty$, for any given $x \in \mathbb{R}$.

3. Letting $n = 3$, parts **(a)-(d)** below regard the order statistics $U_{(1)}, U_{(2)}, U_{(3)}$ from U_1, U_2, U_3.

a) For $0 \le u_1 \le u_2 \le u_3 \le 1$, show that the joint cdf of $(U_{(1)}, U_{(2)}, U_{(3)})$ is given by

$$\begin{aligned} F_{U_{(1)},U_{(2)},U_{(3)}}(u_1, u_2, u_3) &= 6P(U_1 \le u_1, U_2 \le u_2, U_3 \le u_3, U_1 \le U_2 \le U_3) \\ &= 6u_3u_2u_1 - 3u_2^2u_1 - 3u_3u_1^2 + u_1^3. \end{aligned}$$

b) Show that the joint pdf of $(U_{(1)}, U_{(2)}, U_{(3)})$ is

$$f_{U_{(1)},U_{(2)},U_{(3)}}(u_1, u_2, u_3) = \begin{cases} 6, & 0 \le u_1 \le u_2 \le u_3 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

c) Based on **Problem 3(b)**, show that the marginal pdf of $U_{(2)}$ is

$$f_{U_{(2)}}(u_2) = \begin{cases} 6u_2(1 - u_2), & u_2 \in [0, 1], \\ 0, & \text{otherwise.} \end{cases}$$

d) Find the conditional distribution of $U_{(1)}$ given $U_{(2)}$ and identify/name the distribution.

Part II.

For **Problems 4.–5.**, let $U_1, \ldots, U_n$ again denote a random sample from the uniform$[0,1]$ distribution, where the sample size $n \geq 3$ is always odd. Let $U_{(1)}, \ldots, U_{(n)}$ denote the order statistics of $U_1, \ldots, U_n$ and let $\hat{\theta}_U \equiv U_{((n+1)/2)}$ denote the sample median.

4. Verify that the pdf of $\hat{\theta}_U$ is given by

$$f_{\hat{\theta}_U}(t) = \frac{n!}{(\frac{n-1}{2})!(\frac{n-1}{2})!} t^{(n-1)/2}(1-t)^{(n-1)/2}, \quad 0 < t < 1,$$

and identify/name the distribution of $\hat{\theta}_U$.

5. Letting $\theta_U = 1/2$ denote the population median of the uniform$[0,1]$ distribution, show that

$$\sqrt{n}(\hat{\theta}_U - \theta_U) \xrightarrow{d} N(0, 1/4)$$

as odd $n \to \infty$.

Hint: You may use the fact that if a random variable B follows the Beta(α, β) distribution, then B is equal in distribution to $Y/(Y+Z)$, where $Y \sim \chi^2_{2\alpha}$ and $Z \sim \chi^2_{2\beta}$ are two independent χ^2-random variables with degrees of freedom 2α and 2β, respectively.

Part III.

Let $X_1, \ldots, X_n$ be iid continuous random variables with pdf $f_X(x) > 0$, for $x \in \mathbb{R}$, and cdf $F_X(x)$. Further, let $\hat{\theta}_X$ denote the sample median of $X_1, \ldots, X_n$ and let $\theta_X = F_X^{-1}(1/2)$ denote the population median, i.e., $F_X(\theta_X) = 1/2$.

6. Find the asymptotic distribution of $\sqrt{n}(\hat{\theta}_X - \theta_X)$, assuming that n is odd and $n \to \infty$.

Hint: Use the conclusions from previous problems and the fact $F_X^{-1}(F_X(x)) = x$ for any $x \in \mathbb{R}$ because F_X is strictly increasing.

1. We have

$$\begin{aligned} P(F_X(X_1) \le x) &= P(X_1 \le F_X^{-1}(x)) \\ &= F_X(F_X^{-1}(x)) \\ &= x, \end{aligned}$$

where we used the fact that F_X is strictly increasing (and thus bijective) since $f_X(x) > 0$ for $x \in \mathbb{R}$.

2. Fix x. We have

$$E[F_n(x)] = \frac{1}{n}\sum_{i=1}^{n} E[\mathbf{1}\{X_i \le x\}] = \frac{1}{n} nP(X_1 \le x) = P(X_1 \le x) = F_X(x),$$

where the second equality makes use of the iid property of the sample. Also,

$$\begin{aligned} \mathrm{Var}(\mathbf{1}\{X_i \le x\}) &= E[\mathbf{1}\{X_i \le x\}^2] - (E[\mathbf{1}\{X_i \le x\}])^2 \\ &= E[\mathbf{1}\{X_i \le x\}] - (E[\mathbf{1}\{X_i \le x\}])^2 \\ &= F_X(x) - F_X(x)^2 = F_X(x)(1 - F_X(x)). \end{aligned}$$

Since $\mathbf{1}\{X_i \le x\}$ are iid, $i = 1, \ldots, n$, by the CLT we have the desired result.

3. **a**) Let $\mathcal{U} = \{(u_1, u_2, u_3) : 0 \le u_1 \le u_2 \le u_3 \le 1\}$. Then $(U_{(1)}, U_{(2)}, U_{(3)})$ is supported on $\mathcal{U}$. For $(u_1, u_2, u_3) \in \mathcal{U}$,

$$\begin{aligned} F_{U_{(1)},U_{(2)},U_{(3)}}(u_1, u_2, u_3) &= P(U_{(1)} \le u_1, U_{(2)} \le u_2, U_{(3)} \le u_3) \\ &= P(\{U_1 \le u_1, U_2 \le u_2, U_3 \le u_3, U_1 \le U_2 \le U_3\} \quad \text{or} \\ &\qquad \{U_1 \le u_1, U_3 \le u_2, U_2 \le u_3, U_1 \le U_3 \le U_2\} \quad \text{or} \\ &\qquad \{U_2 \le u_1, U_1 \le u_2, U_3 \le u_3, U_2 \le U_1 \le U_3\} \quad \text{or} \\ &\qquad \{U_2 \le u_1, U_3 \le u_2, U_1 \le u_3, U_2 \le U_3 \le U_1\} \quad \text{or} \\ &\qquad \{U_3 \le u_1, U_1 \le u_2, U_2 \le u_3, U_3 \le U_1 \le U_2\} \quad \text{or} \\ &\qquad \{U_3 \le u_1, U_2 \le u_2, U_1 \le u_3, U_3 \le U_2 \le U_1\}) \\ &= 6P(U_1 \le u_1, U_2 \le u_2, U_3 \le u_3, U_1 \le U_2 \le U_3) \\ &= 6\int_0^{u_1}\int_{t_1}^{u_2}\int_{t_2}^{u_3} 1 dt_3 dt_2 dt_1 \\ &= 6\int_0^{u_1}\int_{t_1}^{u_2} (u_3 - t_2) dt_2 dt_1 \\ &= 6\int_0^{u_1} [u_3(u_2 - t_1) - [u_2^2/2 - t_1^2/2]) dt_1 \\ &= 6[u_3 u_2 u_1 - u_3 u_1^2/2 - u_2^2 u_1/2 + u_1^3/6] \end{aligned}$$

The equalities follow from U_1, U_2, U_3 being iid Unif$(0, 1)$ with joint pdf 1.

b) For $(u_1, u_2, u_3) \in \mathcal{U}$,

$$f_{U_{(1)},U_{(2)},U_{(3)}}(u_1, u_2, u_3) = \frac{\partial^3}{\partial u_1 \partial u_2 \partial u_3} F_{U_{(1)},U_{(2)},U_{(3)}}(u_1, u_2, u_3) = 6.$$

c) For $0 \le u_1 \le u_2 \le 1$,

$$\begin{aligned} f_{U_{(1)},U_{(2)}}(u_1, u_2) &= \int_{u_2}^{1} f_{U_{(1)},U_{(2)},U_{(3)}}(u_1, u_2, u_3) du_3 \\ &= 6(1 - u_2). \end{aligned}$$

So

$$f_{U_{(1)}|U_{(2)}}(u_1 \mid u_2) = \frac{f_{U_{(1)},U_{(2)}}(u_1, u_2)}{f_{U_{(2)}}(u_2)} = \frac{6(1-u_2)}{6u_2(1-u_2)} = \frac{1}{u_2}.$$

Thus, given $U_{(2)} = u_2$, $U_{(1)}$ follows a uniform distribution on $[0, u_2]$.

4. Write

$$\begin{aligned} F_{\hat{\theta}_U}(t) &= P(\hat{\theta}_U \le t) = P(U_{(\frac{n+1}{2})} \le t) \\ &= P\left(\text{At least } \frac{n+1}{2} \text{ variables among } U_1, \ldots, U_n \text{ are } \le t\right) \\ &= \sum_{j=(n+1)/2}^{n} P\left(\text{Exactly } j \text{ variables among } U_1, \ldots, U_n \text{ are} \le t\right) \\ &= \sum_{j=(n+1)/2}^{n} \binom{n}{j} P\left(U_1 \le t, \ldots, U_j \le t, U_{j+1} > t, \ldots, U_n > t\right) \\ &= \sum_{j=(n+1)/2}^{n} \binom{n}{j} t^j (1-t)^{n-j}, \end{aligned}$$

due to $U_1, \ldots, U_n$ being iid Unif$(0, 1)$. So

$$\begin{aligned} f_{\hat{\theta}_U}(t) &= \frac{d}{dx} F_{\hat{\theta}_U}(t) \\ &= \sum_{j=(n+1)/2}^{n-1} \binom{n}{j} j t^{j-1}(1-t)^{n-j} - \sum_{j=(n+1)/2}^{n-1} \binom{n}{j} (n-j) t^j (1-t)^{n-j-1} + n t^{n-1} \\ &= \sum_{j=(n+1)/2}^{n-1} \binom{n}{j} j t^{j-1}(1-t)^{n-j} - \sum_{j'=(n+3)/2}^{n} \binom{n}{j'-1} (n-j'+1) t^{j'-1} (1-t)^{n-j'} + n t^{n-1} \\ &= \sum_{j=(n+1)/2}^{n-1} \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j} - \sum_{j=(n+3)/2}^{n} \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j} + n t^{n-1} \\ &= \sum_{j=(n+1)/2}^{(n+1)/2} \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j} - \sum_{j=n}^{n} \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j} + n t^{n-1} \\ &= \frac{n!}{\left(\frac{n-1}{2}\right)! \left(\frac{n-1}{2}\right)!} t^{\frac{n-1}{2}} (1-t)^{\frac{n-1}{2}} - n t^{n-1} + n t^{n-1} \\ &= \frac{n!}{\left(\frac{n-1}{2}\right)! \left(\frac{n-1}{2}\right)!} t^{\frac{n-1}{2}} (1-t)^{\frac{n-1}{2}}, \quad t \in (0, 1), \end{aligned}$$

which is our desired result. Here the third equality applies change-of-variable $j' = j + 1$ and the notation j' is replaced by j in the fourth equality and onward. The sample median follows a Beta distribution with $\alpha = \beta = (n + 1)/2$.

5. Following the hint, $\hat{\theta}_U$ is equal in distribution to $Y/(Y+Z)$ where Y and Z are iid χ^2_{n+1} random variables. Let $A_1, A_2, \ldots$ and $B_1, B_2, \ldots$ be two independent sequences of iid χ^2_1 random variables. We have that the joint distribution of Y and Z equal to that of $Y_{n+1} = \sum_{i=1}^{n+1} A_i$ and $Z_{n+1} = \sum_{i=1}^{n+1} B_i$. Since $\theta_U = 1/2$,

$$\sqrt{n+1}(\hat{\theta}_U - \theta_U) \stackrel{d}{=} \sqrt{n+1}\left(\frac{Y_{n+1} - (Y_{n+1}+Z_{n+1})/2}{Y_{n+1}+Z_{n+1}}\right) = \frac{1}{2}\frac{(n+1)^{-1/2}(Y_{n+1}-Z_{n+1})}{(n+1)^{-1}(Y_{n+1}+Z_{n+1})}, \quad (1)$$

where $\stackrel{d}{=}$ means equal in distribution.

Note that the $A_i - B_i$ and the $A_i + B_i$ are iid random variables with mean 0 and 2, respectively, and variance both equal to 4. By the CLT,

$$\frac{1}{\sqrt{n+1}}(Y_{n+1} - Z_{n+1}) = \frac{1}{\sqrt{n+1}}\sum_{i=1}^{n+1}(A_i - B_i) \xrightarrow{d} N(0,4). \quad (2)$$

Also, by the weak law of large numbers,

$$\frac{1}{n+1}(Y_{n+1} + Z_{n+1}) = \frac{1}{n+1}\sum_{i=1}^{n+1}(A_i + B_i) \xrightarrow{p} 2. \quad (3)$$

By (1)–(3) and Slutsky's theorem,

$$\sqrt{n}(\hat{\theta}_U - \theta_U) \xrightarrow{d} N(0, \frac{1}{4}).$$

6. Let $U_i = F_X(X_i)$. Then U_i, $i = 1, \ldots, n$ follow independent uniform distributions on $[0,1]$. Let $\hat{\theta}_U$ be the sample median of $U_1, \ldots, U_n$, and $\theta_U = 1/2$ be the population median of U_1. Since F_X is strictly increasing, it preserves the ordering of its arguments, and thus

$$F_X(\hat{\theta}_X) = F_X(X_{(\frac{n+1}{2})}) = \text{sample median of } \{F_X(X_i)\}_{i=1}^n = U_{(\frac{n+1}{2})} = \hat{\theta}_U.$$

Also, $F_X(\theta_X) = 1/2 = \theta_U$. Thus

$$\hat{\theta}_X = F_X^{-1}(\hat{\theta}_U), \quad \text{and } \theta_X = F_X^{-1}(\theta_U).$$

Because $\sqrt{n}(\hat{\theta}_U - \theta_U) \xrightarrow{d} N(0, \frac{1}{4})$ by Problem 5, the delta method leads to

$$\sqrt{n}(\hat{\theta}_X - \theta_X) = \sqrt{n}(F_X^{-1}(\hat{\theta}_U) - F_X^{-1}(\theta_U)) \xrightarrow{d} N(0, \frac{1}{4}\left(\frac{d}{dx}F_X^{-1}(\theta_U)\right)^2).$$

By calculus,

$$1 = \frac{d}{dx}x = \frac{d}{dx}F_X^{-1}(F_X(x)) = \left(\frac{d}{dx}F_X^{-1}\right)(F_X(x)) \cdot \frac{d}{dx}F_X(x) = \left(\frac{d}{dx}F_X^{-1}\right)(F_X(x)) \cdot f_X(x),$$

so

$$\frac{d}{dx}F_X^{-1}(\theta_U) = \frac{1}{f_X(F_X^{-1}(\theta_U))} = \frac{1}{f_X(\theta_X)}.$$

Thus,

$$\sqrt{n}(\hat{\theta}_X - \theta_X) \xrightarrow{d} N\left(0, \frac{1}{4f_X(\theta_X)^2}\right).$$

Some facts that you may use are:

Fact 1: If X is an inverse Gaussian random variable with parameters $\mu > 0, \lambda > 0$, denoted $X \sim$ IG (μ, λ), then the pdf of X is

$$f(x|\mu, \lambda) = \begin{cases} \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{ -\frac{\lambda(x-\mu)^2}{2\mu^2 x} \right\} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

For such X it follows that $E(X) = \mu$, Var$(X) = \mu^3/\lambda$ and $E(1/X) = 1/\mu + 1/\lambda$.

Fact 2: Suppose that $X_1, X_2, \ldots, X_n$ are iid random variables with $X_1 \sim$ IG (μ, λ). Let the sample mean $\bar{X} = \sum_{i=1}^n X_i/n$, then $\bar{X} \sim$ IG $(\mu, n\lambda)$ and $\lambda \sum_{i=1}^n (1/X_i - 1/\bar{X}) \sim \chi^2_{n-1}$. Here, χ^2_{n-1} denotes the χ^2 distribution with $n-1$ degrees of freedom.

Fact 3: If $Y \sim \chi^2_m$, then $E(Y) = m$, Var $(Y) = 2m$ and $E(1/Y) = 1/(m-2)$.

Fact 4: If W is a gamma random variable with parameters (α, β), that is $W \sim$ Gamma (α, β), then the pdf of W is

$$f(w|\alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} \exp\{-\beta w\} & \text{if } w > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha, \beta > 0$. Further, E$(W) = \alpha/\beta$.

Part I Suppose that $X_1, X_2, \ldots, X_n$ are iid random variables with $X_1 \sim$ IG (μ, λ). Let $\boldsymbol{\theta} = (\mu, \lambda)$, and $\omega \equiv$ Var $(X_1) = \mu^3/\lambda$.

1. Show that $(\bar{X}, \sum_{i=1}^n [1/X_i - 1/\bar{X}])$ jointly form a complete and sufficient statistic for $\boldsymbol{\theta}$ based on $(X_1, \ldots, X_n)$.

2. Find the UMVUEs of μ and λ.

3. Prove that there is a unique maximizer $\hat{\boldsymbol{\theta}}_n$ of the likelihood function of $\boldsymbol{\theta}$ based on $(X_1, X_2, \ldots, X_n)$.

4. Show that the Fisher information matrix based on $(X_1, X_2, \ldots, X_n)$ is

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n\lambda}{\mu^3} & 0 \\ 0 & \frac{n}{2\lambda^2} \end{pmatrix}.$$

5. Find the asymptotic (bivariate) normal distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ as $n \to \infty$.

6. Show that $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$. Hint: Use **Problem 5**.

7. Find $\hat{\omega}_n$ the MLE of $\omega \equiv \text{Var}(X_1) = \frac{\mu^3}{\lambda}$.

8. Using the result of **Problem 5**, find the limiting distribution of $\sqrt{n}(\hat{\omega}_n - \omega)$ as $n \to \infty$.

9. Show that $Q((X_1, \ldots, X_n), \lambda) = \lambda \sum_{i=1}^{n}(1/X_i - 1/\bar{X})$ is a pivotal quantity .

10. Let λ_0 be a known positive value. Using the pivotal quantity from **Problem 9**, derive a size $\alpha \in (0, 1)$ test for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$.

11. Using a statistic U based on $X_1, \ldots, X_n$ and the joint pivotal quantity from **Problem 9**, construct a one sided confidence interval of the form $(0, U)$ for λ with confidence coefficient $(1 - \alpha) \in (0, 1)$.

12. Let $\boldsymbol{\theta}_0$ be a given value of $\boldsymbol{\theta}$. Derive an asymptotic size $0 < \alpha < 1$ LRT for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Hint: There is no need to expand or simplify the LRT statistic.

13. Is the Jeffreys' prior $\pi^*(\boldsymbol{\theta}) \propto (\text{det}I(\boldsymbol{\theta}))^{1/2}$ a valid density on the parameter space?

Part II For **Problems 14-16**, assume that the prior density of $\boldsymbol{\theta}$ is $\pi(\boldsymbol{\theta}) = \pi_1(\mu)\pi_2(\lambda)$, where $\pi_1(\mu)$ is the Gamma (α_μ, β_μ) density, and $\pi_2(\lambda)$ is the Gamma $(\alpha_\lambda, \beta_\lambda)$ density for known positive values of $\alpha_\lambda, \beta_\lambda, \alpha_\mu$ and β_μ. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be the observed data.

14. Derive (up to a normalizing constant) the posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ of $\boldsymbol{\theta}$.

15. Is the family of prior densities $\{\pi(\boldsymbol{\theta}) : \text{all } \alpha_\lambda, \beta_\lambda, \alpha_\mu, \beta_\mu > 0\}$ conjugate in this model? Explain.

16. Derive (up to a normalizing constant) the marginal posterior density of μ given x from the joint posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ obtained in **Problem 15**.

Part III For **Problem 17**, assume that μ is known and the prior density of λ is $\pi_2(\lambda)$, where $\pi_2(\lambda)$ is the Gamma $(\alpha_\lambda, \beta_\lambda)$ density for known values of $\alpha_\lambda > 1/2$ and $\beta_\lambda > n/\mu$.

17. Derive the Bayes estimator of λ under the loss function

$$L(\lambda, t) = \frac{3(t - \lambda)^2}{\lambda}.$$

1. The joint pdf of $(X_1, X_2, \ldots, X_n)$ is

$$(\lambda/[2\pi])^{n/2}\exp(n\lambda/\mu)\Big(\prod_{i=1}^{n} x_i\Big)^{-3/2}\exp\Big(-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n} x_i - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}\Big).$$

Since the inverse Gaussian density is in exponential family, and the parameter space contains an opening set in $\mathbb{R}^2$, by factorization theorem $(\sum_{i=1}^n X_i, \sum_{i=1}^n [1/X_i])$ and hence $(\bar{X}, \sum_{i=1}^n [1/X_i - 1/\bar{X}])$ is complete and sufficient for $\boldsymbol{\theta}$.

2. From the given facts, we know that

$$E(\bar{X}) = \mu, E\Big(\frac{n-3}{\sum_{i=1}^n [1/X_i - 1/\bar{X}]}\Big) = \lambda.$$

Since $(\bar{X}, \sum_{i=1}^n [1/X_i - 1/\bar{X}])$ is complete and sufficient for $\boldsymbol{\theta}$, $\bar{X}$ is UMVUE of μ and $(n-3)/(\sum_{i=1}^n [1/X_i - 1/\bar{X}])$ is UMVUE of λ.

3. The loglikelihood function for (μ, λ) (up to an additive constant) is

$$\log \ell(\mu,\lambda) = \frac{n}{2}\log\lambda - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n} x_i - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i} + \frac{n\lambda}{\mu}.$$

Thus,

$$\begin{aligned}\frac{\partial \log \ell(\mu,\lambda)}{\partial\lambda} &= \frac{n}{2\lambda} - \frac{\sum_{i=1}^n x_i}{2\mu^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{x_i} + \frac{n}{\mu}, \text{ and}\\ \frac{\partial \log \ell(\mu,\lambda)}{\partial\mu} &= \frac{\lambda\sum_{i=1}^n x_i}{\mu^3} - \frac{n\lambda}{\mu^2}.\end{aligned} \tag{1}$$

From (1), the solution of the likelihood equations is $\hat{\boldsymbol{\theta}}_n \equiv (\bar{X}, n/\sum_{i=1}^n [1/X_i - 1/\bar{X}])$. Let $\bar{x} = \sum_{i=1}^n x_i/n$. Note that, irrespective of the value of λ, $\frac{\partial \log \ell(\lambda,\mu)}{\partial\mu} \gtreqless 0$ if $\mu \lesseqgtr \bar{x}$. Next, $\frac{\partial \log \ell(\lambda,\bar{x})}{\partial\lambda} \gtreqless 0$ if $\lambda \lesseqgtr n/\sum_{i=1}^n [1/x_i - 1/\bar{x}]$. Thus, the likelihood function has a unique maximizer at $\hat{\boldsymbol{\theta}}_n$.

4. From (1) we have

$$\begin{aligned}\frac{\partial^2 \log \ell(\mu,\lambda)}{\partial\lambda^2} &= -\frac{n}{2\lambda^2}\\ \frac{\partial^2 \log \ell(\mu,\lambda)}{\partial\mu\partial\lambda} &= \frac{\sum_{i=1}^n x_i}{\mu^3} - \frac{n}{\mu^2}, \text{ and}\\ \frac{\partial^2 \log \ell(\mu,\lambda)}{\partial\mu^2} &= -\frac{3\lambda\sum_{i=1}^n x_i}{\mu^4} + \frac{2n\lambda}{\mu^3}.\end{aligned}$$

Hence,

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n\lambda}{\mu^3} & 0 \\ 0 & \frac{n}{2\lambda^2}\end{pmatrix}.$$

5. By the asymptotic normality of MLE,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N_2\left(\mathbf{0}, \begin{pmatrix} \frac{\mu^3}{\lambda} & 0 \\ 0 & 2\lambda^2 \end{pmatrix}\right).$$

6. Since

$$(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}),$$

and $\frac{1}{\sqrt{n}} \to 0$, by Slutsky's theorem $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} 0$, which is equivalent to $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{P} 0$.

7. Let $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\lambda}_n)$. Since $\hat{\boldsymbol{\theta}}_n$ is the MLE of θ, by the invariance property of MLE, the MLE of $\omega = \omega(\theta)$ is $\hat{\omega}_n = \rho(\hat{\boldsymbol{\theta}}_n) = \frac{\hat{\mu}_n^3}{\hat{\lambda}_n}$.

8. Note that $\partial\omega(\theta)/\partial\theta = (3\mu^2/\lambda, -\mu^3/\lambda^2)^T$. Let $V(\boldsymbol{\theta})$ be the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$. By Delta method,

$$\sqrt{n}(\hat{\omega}_n - \omega) \xrightarrow{d} N(0, (3\mu^2/\lambda, -\mu^3/\lambda^2)V(\boldsymbol{\theta})(3\mu^2/\lambda, -\mu^3/\lambda^2)^T).$$

That is $\sqrt{n}(\hat{\omega}_n - \omega) \xrightarrow{d} N(0, 9\mu^7/\lambda^3 + 2\mu^6/\lambda^2)$.

9. Since $\lambda\sum_{i=1}^n(1/X_i - 1/\bar{X}) \sim \chi^2_{n-1}$, $Q((X_1, \ldots, X_n), \boldsymbol{\theta}) = \lambda\sum_{i=1}^n(1/X_i - 1/\bar{X})$ is a pivotal quantity.

10. Under H_0, $\lambda_0\sum_{i=1}^n(1/X_i - 1/\bar{X}) \sim \chi^2_{n-1}$. Thus, a size α test for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$ is

$$\text{Reject } H_0 \text{ if } \lambda_0\sum_{i=1}^n(1/X_i - 1/\bar{X}) > \chi^2_{n-1,1-\alpha},$$

where $\chi^2_{n,\alpha}$ is the αth quantile of χ^2_n.

11. A $(1-\alpha)$ confidence interval for λ can be found from $\{\lambda : Q((X_1, \ldots, X_n), \boldsymbol{\theta}) \leq \chi^2_{n-1,1-\alpha}\}$. Thus, $(0, \chi^2_{n-1,1-\alpha}/[\sum_{i=1}^n(1/X_i - 1/\bar{X})])$ is a $(1-\alpha)$ confidence interval for λ.

12. The LRT statistic is

$$\lambda(\mathbf{x}) = \frac{\ell(\boldsymbol{\theta}_0)}{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\ell(\boldsymbol{\theta})} = \frac{\ell(\boldsymbol{\theta}_0)}{\ell(\hat{\boldsymbol{\theta}}_n)}.$$

We know that $-2\log\lambda(\mathbf{x}) \xrightarrow{d} \chi^2_2$ as $n \to \infty$. Thus, an asymptotic size α LRT for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ is

$$\text{Reject } H_0 \text{ if } -2\log\lambda(\mathbf{x}) > \chi^2_{2,1-\alpha}.$$

13. The Jeffreys' prior density of $\boldsymbol{\theta}$ is

$$\pi^*(\boldsymbol{\theta}) \propto (\det I(\boldsymbol{\theta}))^{1/2} \propto \frac{1}{\sqrt{\lambda\mu^3}} \quad \mu > 0, \lambda > 0.$$

Since

$$\int_0^\infty \int_0^\infty \pi^*(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty,$$

it is not a valid pdf.

14. The posterior density of θ is

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto \pi(\boldsymbol{\theta}) f(\boldsymbol{x}|\boldsymbol{\theta}) \\
&\propto \pi_1(\mu)\pi_2(\lambda)\lambda^{n/2}\exp(n\lambda/\mu)\exp\Big(-\frac{\lambda}{2\mu^2}\sum_{i=1}^n x_i - \frac{\lambda}{2}\sum_{i=1}^n \frac{1}{x_i}\Big) \\
&\propto \mu^{\alpha_\mu - 1}\exp(-\beta_\mu\mu)\lambda^{n/2+\alpha_\lambda-1}\exp\Big(-\lambda\Big[\beta_\lambda + \frac{\sum_{i=1}^n x_i}{2\mu^2} + \frac{1}{2}\sum_{i=1}^n \frac{1}{x_i} - \frac{n}{\mu}\Big]\Big) \quad \mu > 0, \lambda > 0.
\end{aligned}$$

15. Since the posterior density $\pi(\theta|\boldsymbol{x})$ is not of the form of a product of two independent gamma densities, the prior family is not conjugate for the likelihood.

16. The marginal posterior density

$$\begin{aligned}
\pi(\mu|\boldsymbol{x}) &= \int_0^\infty \pi(\mu, \lambda|\boldsymbol{x}) d\lambda \\
&\propto \mu^{\alpha_\mu - 1}\exp(-\beta_\mu\mu)\int_0^\infty \lambda^{n/2+\alpha_\lambda-1}\exp\Big(-\lambda\Big[\beta_\lambda + \frac{\sum_{i=1}^n x_i}{2\mu^2} + \frac{1}{2}\sum_{i=1}^n \frac{1}{x_i} - \frac{n}{\mu}\Big]\Big) d\lambda \\
&\propto \mu^{\alpha_\mu - 1}\exp(-\beta_\mu\mu)\Big[\beta_\lambda + \frac{\sum_{i=1}^n x_i}{2\mu^2} + \frac{1}{2}\sum_{i=1}^n \frac{1}{x_i} - \frac{n}{\mu}\Big]^{-(n/2+\alpha_\lambda)} \quad \mu > 0.
\end{aligned}$$

17. Note that $\pi(\lambda|\boldsymbol{x})$ is the density of Gamma $(\alpha_\lambda + n/2, \beta_\lambda + \sum_{i=1}^n x_i/[2\mu^2] + \sum_{i=1}^n [1/\{2x_i\}] - n/\mu)$. Let $\alpha'_\lambda = \alpha_\lambda + n/2$ and $\beta'_\lambda = \beta_\lambda + \sum_{i=1}^n x_i/[2\mu^2] + \sum_{i=1}^n [1/\{2x_i\}] - n/\mu$. Note that

$$E[L(\lambda, t)|\boldsymbol{x}] = \int_0^\infty \frac{3(t-\lambda)^2}{\lambda}\pi(\lambda|\boldsymbol{x}) d\lambda.$$

Now,

$$\int_0^\infty \frac{3(t-\lambda)^2}{\lambda}\pi(\lambda|\boldsymbol{x}) d\lambda \propto \int_0^\infty (t-\lambda)^2 \lambda^{(\alpha'_\lambda - 1)-1}\exp(-\beta'_\lambda\lambda) d\lambda,$$

which is minimized at the mean of Gamma $(\alpha'_\lambda - 1, \beta'_\lambda)$. Thus the Bayes estimator of λ under the loss function $L(\lambda, t)$ is

$$\frac{\alpha'_\lambda - 1}{\beta'_\lambda} = \frac{\alpha_\lambda + [n-2]/2}{\beta_\lambda + \sum_{i=1}^n x_i/[2\mu^2] + \sum_{i=1}^n [1/\{2x_i\}] - n/\mu}.$$

Part I

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X, Y, X_n, $n \geq 1$, denote (Borel measurable) random variables (r.v.'s). Denote the set of real numbers as $\mathbb{R}$ and denote the Borel σ-algebra on $\mathbb{R}$ as $\mathcal{B}(\mathbb{R})$.

1. Give the following definitions:

- **a**. State the defining properties of $\mathcal{F}$ with respect to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- **b**. State the defining properties of $\mathbb{P}$ with respect to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- **c**. State the defining properties of X to be a (Borel measurable) random variable.
- **d**. Define the meaning of almost sure convergence of X_n to X as $n \to \infty$ (written as $X_n \overset{a.s}{\to} X$).
- **e**. Define the meaning of convergence of X_n to X in probability as $n \to \infty$ (written as $X_n \overset{p}{\to} X$).
- **f**. For a given $p > 0$, define the meaning of L_p-convergence of X_n to X as $n \to \infty$ (written as $X_n \overset{L_p}{\to} X$).
- **g**. Define the meaning of convergence of X_n to X in distribution as $n \to \infty$ (written as $X_n \overset{d}{\to} X$).
- **h**. For a random variable X, define the (induced) probability measure P_X and the distribution function F_X.
- **i**. Define what it means for X to be an absolutely continuous random variable.
- **j**. Let $X \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ (i.e. X is integrable) and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Define the meaning of the conditional expectation of X given $\mathcal{G}$ (written as $\mathbb{E}(X|\mathcal{G})$).
- **k**. Explain the meaning of the notation $\mathbb{E}(X|Y)$.

Part II

Suppose X, X_n, $n \geq 1$, denote independent r.v.'s, where $X \equiv 0$ and, for each $n \geq 1$, X_n has a probability measure P_{X_n} on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with a density function (with respect to the Lebesgue measure m) given by

$$\left(\frac{dP_{X_n}}{dm}\right)(u) \equiv f_{X_n}(u) = \begin{cases} 2n(1-nu) & 0 < u < \frac{1}{n}, \\ 0 & \text{otherwise.} \end{cases}$$

2. Argue carefully (stating any standard results) that there exists one probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which r.v.'s X, X_n for $n \geq 1$ can all be defined with the marginal distributions above.

Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ be **some** probability space with r.v.'s Y, Y_n, $n \geq 1$. Suppose that $Y \equiv 0$ and that, for each $n \geq 1$, Y_n has the same marginal distribution as X_n above ($X_n \stackrel{d}{=} Y_n$). The point here is that $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and the r.v.'s Y_n may not be the same as $(\Omega, \mathcal{F}, \mathbb{P})$ and the r.v.'s X_n in **Question 2**. We do not suppose that the r.v.'s Y_n, $n \geq 1$, are independent.

3. Show $Y_n \xrightarrow{p} Y$ as $n \to \infty$.

4. For what values of $p > 0$ does $Y_n \xrightarrow{L_p} Y$ as $n \to \infty$?

5. Does $Y_n \xrightarrow{d} Y$ as $n \to \infty$?

6. Show that $Y_n \xrightarrow{a.s} Y$ as $n \to \infty$ (Hint: Verify $\sum_{n=1}^{\infty} \tilde{\mathbb{P}}(|Y_n| > \epsilon) < \infty$ for any $\epsilon > 0$.)

Part III

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space; let X, Y, X_1, X_2 denote integrable r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$; and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra.

7. If $\mathbb{E}(X|Y) = Y$ and $\mathbb{E}(Y|X) = X$, prove $X = Y$ (a.s.).

8. If X, Y are independent, show that $\mathbb{E}(X|Y) = \mathbb{E}(X)$ (a.s.).

9. If $X \sim Uniform(-1, 1)$ and $Y = X^2$, show that $\mathbb{E}(X|Y) = \mathbb{E}(X)$ (a.s.). In this example, are X, Y independent?

10. If $X_1 \geq X_2$ (a.s.), then prove $\mathbb{E}(X_1|Y) \geq \mathbb{E}(X_2|Y)$ (a.s.).

11. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and denote the power set of Ω as $\mathcal{F} = \mathcal{P}(\Omega)$. Let $\mathbb{P}$ be a probability measure on Ω such that $\mathbb{P}(\{\omega_i\}) = p_i$ for $i = 1, \ldots, 4$, where $0 < p_1 < 1$ and $\sum_{i=1}^{4} p_i = 1$. For given real-values x_1, x_2, x_3, x_4, define a random variable X as: $X(\omega_i) = x_i$ for $i = 1, \ldots, 4$. Let $\mathcal{G} = \{\emptyset, \{\omega_1\}, \{\omega_1\}^c, \Omega\}$ denote a sub-σ-algebra. Give an explicit form of $\mathbb{E}(X|\mathcal{G}) \equiv \mathbb{E}(X|\mathcal{G})(\omega)$, $\omega \in \mathcal{F}$, carefully justifying that your version is a conditional expectation in this set-up.

Part 1

1. Definitions:

a. $\Omega \neq \emptyset$, $\mathcal{F}$ satisfies: $\Omega \in \mathcal{F}; A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}; \{A_n\} \in \mathcal{F} \Rightarrow \cup_{n\geq 1} A_n \in \mathcal{F}$ and $\mathbb{P}$ satisfies: $\mathbb{P} : \mathcal{F} \to [0, \infty]$ and $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1, \{A_n\}_{n\geq 1} \subset \mathcal{F}$ are disjoint $\Rightarrow \mathbb{P}(\cup_{n\geq 1} A_n) = \sum_{n\geq 1} \mathbb{P}(A_n)$.

b. $X : \Omega \to \mathbb{R}$ and $\forall\ B \in \mathcal{B}(\mathbb{R}),\ X^{-1}(B) \in \mathcal{F}$.

c. $\exists A \in \mathcal{F}$ s.t $\mathbb{P}(A) = 1$, $\forall\ \omega \in A, X_n(\omega) \to X(\omega)$ as $n \to \infty$, i.e. $\mathbb{P}(X_n \to X) = 1$.

d. $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$, $\forall\ \epsilon > 0$.

e. $\mathbb{E}(|X_n - X|^p) \to 0$ as $n \to \infty$.

f. If the distribution functions of X_n, X are F_{X_n}, F_X, for $n \geq 1$ respectively and $D =$ set of discontinuity points of F_X. Then we say $X_n \xrightarrow{d} X$ if $\forall x \notin D$, $F_{X_n}(x) \to F_X(x)$, as $n \to \infty$.

g. P_X is a probability on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by $P_X(A) = \mathbb{P}(X^{-1}(A)), \forall A \in \mathcal{B}(\mathbb{R})$. $F_X : \mathbb{R} \to [0, 1]$ defined by $F_X(x) = P_X((-\infty, x]) = \mathbb{P}(X \leq x), \forall x \in \mathbb{R}$.

h. X is an absolutely continuous r.v if F_X is an absolutely continuous function, i.e. if for every $\varepsilon > 0$, there is a $\delta > 0$ such that whenever a finite sequence of pairwise disjoint sub-intervals (x_k, y_k) with $x_k < y_k \in \mathbb{R}$ satisfying $\sum_k (y_k - x_k) < \delta$, we have $\sum_k |F_X(y_k) - F_X(x_k)| < \varepsilon$. In that case, there exists nonnegative f_X s.t.

$$F_X(x) = \int_{-\infty}^{x} f_X(u)dm(u) = \int_{-\infty}^{x} f_X(u)du$$

and $\frac{dF_X}{dx} = f_X = \frac{dP_X}{dm}$ a.e.(m).

i. Conditional expectation $\mathbb{E}(X|\mathcal{G}) = X_0$ is any random variable such that $X_0 \in L_2(\Omega, \mathcal{G}, \mathbb{P})$ satisfying the following condition with probability 1 $(\mathbb{P})$:

$$\mathbb{E}\left(X\mathbf{I}_A\right) = \mathbb{E}\left(X_0\mathbf{I}_A\right),\ \forall A \in \mathcal{G}.$$

If $\mathcal{G} = \sigma\langle Y\rangle$, then $\mathbb{E}(X|\mathcal{G})$ is denoted as $\mathbb{E}(X|Y)$.

Part 2

2. Define $\mu^{(1)} = \mu_1 \equiv \mu_{F_{X_1}}$, the distribution of X_1 and $\mu_n \equiv \mu_{F_{X_n}}$, the distribution of X_n. Set the product measures $\mu^{(n+1)} = \mu^{(n)} \times \mu_{n+1}$, for $n \geq 1$. For each $n \geq 1$, $\mu^{(n)}$ is the valid measure (product measure) on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and $\mu^{(n+1)}(A \times \mathbb{R}) = \mu^{(n)}(A)$ for $A \in \mathcal{B}(\mathbb{R}^n)$. Hence, from Kolmogorov's Consistency theorem we can construct a sequence of random variables $\{X_n : n \geq 1\}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that joint distribution of $(X_1, \ldots, X_n)$ is given by $\mu^{(n)}$. Hence $\{X_n : n \geq 1\}$ is a sequence of independent r.v. with $P_{X_n} = \mu_{F_{X_n}}$, for $n \geq 1$. On this probability space, define X such that $X(\omega) = 0, \forall\omega \in \Omega$. This, being a constant function is measurable and hence a random variable on the same space.

3. Note that here $\mathbb{P}(X_n \in A) = P_{X_n}(A) = \int_A f_{X_n}(u)dm(u)$, for any $A \in \mathcal{B}(\mathbb{R})$. For any $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(X_n > \epsilon) = 0$ for $n \geq \frac{1}{\epsilon}$. Hence the convergence holds.

4. for any $p > 0$,

$$\begin{aligned} \mathbb{E}(|X_n - X|^p) &= \mathbb{E}(X^p) = \int_\Omega X^p d\mathbb{P} = \int_{\mathbb{R}} x^p dP_{X_n}(x) = \int_0^{\frac{1}{n}} x^p 2n(1-nx)dx && (1) \\ &= \frac{2}{(p+1)(p+2)} \frac{1}{n^p} \to 0, \text{ as } n \to \infty. && (2) \end{aligned}$$

5. For these random variables, $F_X = \mathbf{I}_{[0,\infty)}$ with set of discontinuity $D = \{0\}$ and

$$F_{X_n}(u) = \int_{-\infty}^u f_{X_n}(x)dx = \begin{cases} 0 & \text{, if } u < 0 \\ nu(2-nu) & \text{, if } 0 < u < \frac{1}{n} \\ 1 & \text{, if } u \geq \frac{1}{n}. \end{cases}$$

Clearly $\forall x \notin D$, $F_{X_n}(x) \to F_X(x)$, as $n \to \infty$.

6. From calculation in Question 4, we have $\mathbb{E}(X^2) = \frac{1}{6n^2}$ and hence my Markov's inequality, for each $\epsilon > 0$, $\sum \mathbb{P}(|X_n| > \epsilon) \leq \sum \frac{\mathbb{E}(X^2)}{\epsilon^2} = \frac{1}{6\epsilon^2} \sum \frac{1}{n^2} < \infty$. This, using Borel-Cantelli lemma (Proposition 7.2.3 in the text), we have $\mathbb{P}(X_n \to X) = 1$.

Part 3

7. It is easy to check $\mathbb{E}(XY) = \mathbb{E}[\mathbb{E}(XY|Y)] = \mathbb{E}[Y\mathbb{E}(X|Y)] = \mathbb{E}(Y^2)$. Similarly, $\mathbb{E}(XY) = \mathbb{E}(X^2)$. Hence $\mathbb{E}((X-Y)^2) = \mathbb{E}(X^2) + \mathbb{E}(Y^2) - 2\mathbb{E}(XY) = 0$. Since it is a non-negative r.v. with expectation 0, we have $X = Y$ a.s. ($\mathbb{P}$)

8. Note that $\mu_X = \mathbb{E}(X)$ is constant, and hence $\sigma\langle Y\rangle-$measurable. Take any $A \in \sigma\langle Y\rangle$, we have from independence, $\mathbb{E}(X\mathbf{I}_A) = E(X)E(\mathbf{I}_A) = \mu_X E(\mathbf{I}_A) = \mathbb{E}(\mu_X \mathbf{I}_A)$. So $\mathbb{E}(X|Y) = \mu_X = E(X)$ a.s. ($\mathbb{P}$).

9. In the example, clearly X, Y are not independent, but $P(X = x|Y = y) = \frac{1}{2}$ if $x = +y, -y$ and 0 otherwise. So $\mathbb{E}(X|Y) = 0$. Also, $\mathbb{E}(X) = 0$.

10. Let $\mathbb{E}(X_i|Y) = X_i^0$ be $\sigma\langle Y\rangle-$measurable functions that satisfies $\mathbb{E}(X_i\mathbf{I}_{A_i}) = \mathbb{E}(X_i^0\mathbf{I}_{A_i})$, for $A_i \in \sigma\langle Y\rangle$ for $i = 1, 2$. Since $X_1 \geq X_2$ almost surely ($\mathbb{P}$), we have $X_1\mathbf{I}_A \geq X_2\mathbf{I}_A$ almost surely ($\mathbb{P}$), and hence, by monotonicity of integrals, we have for all $A \in \sigma\langle Y\rangle$

$$\mathbb{E}(X_1^0\mathbf{I}_A) \geq \mathbb{E}(X_2^0\mathbf{I}_A), \text{ or } \int_A \left(X_1^0 - X_2^0\right) d\mathbb{P} \geq 0.$$

Since this holds for all $A \in \sigma\langle Y\rangle$, we have $X_1^0 - X_2^0 \geq 0$ or $\mathbb{E}(X_1|Y) = X_1^0 \geq X_2^0 = \mathbb{E}(X_2|Y)$ almost surely ($\mathbb{P}$).

11. Note that $\mathbb{E}(X|\mathcal{G}) = X^0$ is a $\mathcal{G}-$measurable function that satisfies

$$\mathbb{E}(X\mathbf{I}_A) = \mathbb{E}(X^0\mathbf{I}_A)$$

for all $A \in \mathcal{G}$. For this $\mathcal{G}$, the only measurable functions are indicators and constant functions. So $X^0 = a\mathbf{I}_{\{\omega_1\}} + b$ for some $a, b \in \mathbb{R}$. Let $\mu = \mathbb{E}(X) = \sum x_i p_i$. Choosing $A = \Omega$ and $A = \{\omega_1\}$ in the property above gives:

$$a\mathbb{P}(\{\omega_1\}) + b = \mu, \text{ and } X(\omega_1) = (a+b).$$

Solving the above two equations and writing $\mathbb{P}(\{\omega_1\}) = p_1, X(\omega_1) = x_1$, we get $a = \frac{x_1-\mu}{1-p_1}$ and $b = \frac{\mu - x_1 p_1}{1-p_1}$. So,

$$\mathbb{E}(X|\mathcal{G}) = \frac{1}{1-p_1}\Big((x_1 - \mu)\mathbf{I}_{\{\omega_1\}} + \mu - x_1 p_1\Big)$$

Or,

$$\begin{aligned}
\mathbb{E}(X|\mathcal{G})(\omega) &= x_1\mathbf{I}_{\{\omega_1\}}(\omega) + \left(\frac{\mu - x_1 p_1}{1-p_1}\right)\mathbf{I}_{\{\omega_2,\omega_3,\omega_4\}}(\omega) \\
&= x_1\mathbf{I}_{\{\omega_1\}}(\omega) + \left(\frac{x_2p_2 + x_3p_3 + x_4p_4}{1-p_1}\right)\mathbf{I}_{\{\omega_2,\omega_3,\omega_4\}}(\omega)
\end{aligned}$$