Statistical Inference

B. Statistical Data Science 2nd Year   Indian Statistical Institute

Teacher: Soham Sarkar

Exercise Series 4 (Solutions)

Solution 1. (a)

$$f_p(x) = p^x(1-p)^{1-x} = e^{x\log(p)+(1-x)\log(1-p)} = e^{x\log\left(\frac{p}{1-p}\right)+\log(1-p)}, \quad x \in \{0,1\}.$$

Therefore, $f_p(x)$ is of the form $e^{a(p)\tau(x)-b(p)}c(x), \quad x \in \mathcal{S}$ with

$$a(p) = \log\left(\frac{p}{1-p}\right), \tau(x) = x, b(p) = -\log(1-p), c(x) = 1 \text{ and } \mathcal{S} = \{0,1\}.$$

So, f_p belongs to the one-parameter exponential family. A minimal sufficient statistic for p is $\sum_{i=1}^{n} X_i$.

(b)

$$f_\lambda(x) = e^{-\lambda}\frac{\lambda^x}{x!} = e^{x\log(\lambda)-\lambda}\frac{1}{x!}, \quad x \in \{0,1,2,\ldots\}.$$

Therefore, $f_\lambda(x)$ is of the form $e^{a(\lambda)\tau(x)-b(\lambda)}c(x), \quad x \in \mathcal{S}$ with

$$a(\lambda) = \log(\lambda), \tau(x) = x, b(\lambda) = \lambda, c(x) = \frac{1}{x!} \text{ and } \mathcal{S} = \{0,1,2,\ldots\}.$$

So, f_λ belongs to the one-parameter exponential family. A minimal sufficient statistic for λ is $\sum_{i=1}^{n} X_i$.

(c)

$$f_p(x) = p(1-p)^x = e^{x\log(1-p)+\log(p)}, \quad x \in \{0,1,2,\ldots\}.$$

So, $f_p(x)$ is of the form $e^{a(p)\tau(x)-b(p)}c(x), \quad x \in \mathcal{S}$ with

$$a(p) = \log\left(\frac{p}{1-p}\right), \tau(x) = x, b(p) = -\log(1-p), c(x) = 1 \text{ and } \mathcal{S} = \{0,1\}.$$

A minimal sufficient statistic for p is $\sum_{i=1}^{n} X_i$.

Note: A similar derivation can be done for the other version of $\mathsf{Geometric}$ distribution: $f_p(x) = p(1-p)^{x-1}, \quad x \in \{1,2,\ldots\}$.

(d) $f_\theta = \mathsf{Uniform}(\theta,1)$, $\theta < 1$:

$$f_\theta(x) = \frac{1}{1-\theta}, \quad x \in (\theta,1).$$

The support $\mathcal{S} = (\theta,1)$ of X depends on the unknown parameter θ. Therefore, the $\mathsf{Uniform}(\theta,1)$ distribution does not belong to the exponential family. The same is true for $\mathsf{Uniform}(\theta,\theta+1)$ for $\theta \in \mathbb{R}$.

(e) $f_\theta = \mathsf{Normal}(0, \sigma^2)$, $\theta = \sigma^2$:

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} = e^{-\frac{x^2}{2\sigma^2} - \log(\sigma)} \frac{1}{\sqrt{2\pi}}, \quad x \in \mathbb{R}.$$

The pdf is of the form $e^{a(\sigma)\tau(x) - b(\sigma)} c(x), \quad x \in \mathcal{S}$ with

$$a(\sigma) = -\frac{1}{2\sigma^2}, \tau(x) = x^2, b(\sigma) = \log(\sigma), c(x) = \frac{1}{\sqrt{2\pi}} \text{ and } \mathcal{S} = \mathbb{R}.$$

$\mathsf{Normal}(0, \sigma^2)$ belongs to the one-parameter exponential family. A sufficient statistic for σ is $\sum_{i=1}^{n} X_i^2$.

$f_\theta = \mathsf{Normal}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$:

$$\begin{aligned} f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} &= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \log(\sigma)} \frac{1}{\sqrt{2\pi}}, \quad x \in \mathbb{R} \\ &= e^{a_1(\mu,\sigma^2)\tau_1(x) + a_2(\mu,\sigma^2)\tau_2(x) - b(\mu,\sigma^2)} c(x), \quad x \in \mathcal{S}, \end{aligned}$$

with $a_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2}, \tau_1(x) = x^2, a_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}, \tau_2(x) = x, b(\mu, \sigma^2) = \frac{\mu^2}{2\sigma^2} + \log(\sigma), c(x) = \frac{1}{\sqrt{2\pi}}$ and $\mathcal{S} = \mathbb{R}$. $\mathsf{Normal}(\mu, \sigma^2)$ belongs to the two-parameter exponential family. A minimal sufficient statistic for (μ, σ^2) is $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$.

(f) Cauchy distribution cannot be expressed in the exponential form. It does not belong to the exponential family.

(g) For $\mathsf{Laplace}(a, b)$,

$$f_{a,b}(x) = \frac{1}{2b} e^{-\frac{|x-a|}{b}}, x \in \mathbb{R}.$$

If $a = 0$, then the pdf becomes $e^{-\frac{|x|}{b} - \log(2b)}, \quad x \in \mathbb{R}$. So, the $\mathsf{Laplace}(0, \theta)$ distribution belongs to the one-parameter exponential family. A minimal sufficient statistic in this case is $\sum_{i=1}^{n} |X_i|$.

If $a \neq 0$, then the Laplace distribution does not belong to the exponential family.

(h) $\mathsf{Normal}(\theta, \theta^2)$:

$$f_\theta(x) = e^{-\frac{1}{2\theta^2}x^2 + \frac{1}{\theta}x - \log(\theta) - \frac{1}{2}} \frac{1}{\sqrt{2\pi}}, \quad x \in \mathbb{R}. \quad \text{(check (g) above)}$$

So, $\mathsf{Normal}(\theta, \theta^2)$ belongs to the two-parameter exponential family with $a_1(\theta) = -\frac{1}{2\theta^2}, \tau_1(x) = x^2, a_2(\theta) = \frac{1}{\theta}, \tau_2(x) = x, b(\theta) = \log(\theta) + \frac{1}{2}, c(x) = \frac{1}{\sqrt{2\pi}}$ and $\mathcal{S} = \mathbb{R}$. A minimal sufficient statistic for θ is $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$.

$\mathsf{Normal}(\theta, \theta)$:

$$f_\theta(x) = e^{-\frac{1}{2\theta}x^2 - \frac{\theta}{2} - \frac{1}{2}\log(\theta)} \frac{1}{\sqrt{2\pi}} e^x, \quad x \in \mathbb{R}.$$

This distribution belongs to the one-parameter exponential family with $a(\theta) = -\frac{1}{2\theta}, \tau(x) = x^2, b(\theta) = \frac{\theta}{2} - \frac{\log(\theta)}{2}, c(x) = \frac{1}{\sqrt{2\pi}} e^x$ and $\mathcal{S} = \mathbb{R}$. A minimal sufficient statistic for θ is $\sum_{i=1}^{n} X_i^2$.

(i)
$$f_{\alpha,\beta}(x) = x^{\alpha-1}(1-x)^{\beta-1} = e^{(\alpha-1)\log(x)+(\beta-1)\log(1-x)}, \quad x \in (0,1).$$

So, $\mathsf{Beta}(\alpha,\beta)$ belongs to the two-parameter exponential family with $a_1(\alpha,\beta) = \alpha - 1, \tau_1(x) = \log(x), a_2(\alpha,\beta) = \beta - 1, \tau_2(x) = \log(1-x), b(\alpha,\beta) = 0, c(x) = 1$ and $\mathcal{S} = (0,1)$. A minimal sufficient statistic for (α,β) is

$$\left(\sum_{i=1}^n \log(X_i), \sum_{i=1}^n \log(1-X_i)\right) \equiv \left(\log \prod_{i=1}^n X_i, \log \prod_{i=1}^n (1-X_i)\right) \equiv \left(\prod_{i=1}^n X_i, \prod_{i=1}^n (1-X_i)\right).$$

(j)
$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = e^{-\lambda x + (\alpha-1)\log(x) + \alpha\log(\lambda) - \log(\Gamma(\alpha))}, \quad x \in (0,\infty).$$

So, $f_{\alpha,\lambda}$ belongs to the two-parameter exponential family with $a_1(\alpha,\lambda) = -\lambda, \tau_1(x) = x, a_2(\alpha,\lambda) = \alpha - 1, \tau_2(x) = \log(x), b(\alpha,\lambda) = \log(\Gamma(\alpha)) - \alpha\log(\lambda), c(x) = 1$ and $\mathcal{S} = (0,\infty)$. A minimal sufficient statistic for (α,λ) is $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n \log(X_i)\right) \equiv \left(\sum_{i=1}^n X_i, \prod_{i=1}^n X_i\right)$.

(k) The pdf of the $\mathsf{Pareto}(\mu,\alpha)$ distribution is

$$f_{\mu,\alpha}(x) = \frac{\alpha\mu^\alpha}{x^{\alpha+1}}, \quad x \geq \mu.$$

The support $\mathcal{S} = [\mu,\infty)$ of this distribution depends on the unknown parameter μ. Therefore, it is not a member of the exponential family if μ is unknown. However, if μ is known, then the only parameter is α. In that case, the pdf can be expressed as

$$f_\alpha(x) = e^{-(\alpha+1)\log(x) + \log(\alpha) + \alpha\log(\mu)}, \quad x \in [\mu,\infty),$$

which has the one-parameter exponential form with $a(\alpha) = -(\alpha+1), \tau(x) = \log(x), b(\alpha) = -\log(\alpha) - \alpha\log(\mu), c(x) = 1$ and $\mathcal{S} = [\mu,\infty)$. Keep in mind that we are treating μ to be a known constant in this case. A minimal sufficient statistic for α in this case is $\sum_{i=1}^n \log(X_i) \equiv \prod_{i=1}^n X_i$.

(l) The $\mathsf{Weibull}$ pdf $f_{\lambda,k}(x) = \frac{\lambda}{k}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$, $x \geq 0$ cannot be expressed in the exponential form if k is unknown.

If k is known, then we can write

$$f_\lambda(x) = e^{-\frac{x^k}{\lambda^k} - (k-1)\log(\lambda) + \log\left(\frac{\lambda}{k}\right)} x^{k-1}, \qquad x \geq 0,$$

which belongs to the one-parameter exponential family with $a(\lambda) = -\frac{1}{\lambda^k}, \tau(x) = x^k, b(\lambda) = (k-1)\log(\lambda) - \log\left(\frac{\lambda}{k}\right), c(x) = x^{k-1}$ and $\mathcal{S} = [0,\infty)$. In this case, a minimal sufficient statistic for λ is $\sum_{i=1}^n X_i^k$.

(m)

$$\begin{aligned} f(x_1,\ldots,x_k) &= \frac{n!}{x_1!\cdots x_k!} p_1^{x_1}\cdots p_k^{x_k} \\ &= \frac{n!}{x_1!\cdots x_k!} e^{x_1\log(p_1)+\cdots+x_k\log(p_k)}, \quad 0 \leq x_1,\ldots,x_k \leq n, x_1+\cdots+x_k = n. \end{aligned}$$

The pmf belongs to the k-parameter exponential family with

$$\begin{aligned}
&a_j(p_1,\ldots,p_k) = \log(p_j), \quad \tau_j(x_1,\ldots,x_k) = x_j, \qquad j = 1,\ldots,k,\\
&b(p_1,\ldots,p_k) = 0, \quad c(x_1,\ldots,x_k) = \frac{n!}{x_1!\cdots x_k!} \qquad \text{and}\\
&\mathcal{S} = \{x_1,\ldots,x_k : x_i = 0,1,\ldots,n \;\forall\, i = 1,\ldots,k \text{ with } x_1+\cdots+x_k = n\}.
\end{aligned}$$

A minimal sufficient statistic for $(p_1,\ldots,p_k)$ is $\left(\sum_{i=1}^n X_{1i},\ldots,\sum_{i=1}^n X_{ki}\right)$.

(n)

$$\begin{aligned}
f(x,y) &= \frac{1}{2\pi\,\sigma_x\,\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{1}{\sigma_x^2}x^2 + \frac{1}{\sigma_y^2}y^2 - \frac{2\mu_x}{\sigma_x^2}x - \frac{2\mu_y}{\sigma_y^2}y\right.\right.\\
&\qquad\qquad\left.\left. - \frac{2\rho}{\sigma_x\,\sigma_y}xy + \frac{2\rho\,\mu_x}{\sigma_x\,\sigma_y}y + \frac{2\rho\,\mu_y}{\sigma_x\,\sigma_y}x\right\}\right]\\
&= \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{1}{\sigma_x^2}x^2 + \frac{1}{\sigma_y^2}y^2 - \frac{2\mu_x}{\sigma_x^2}x - \frac{2\mu_y}{\sigma_y^2}y - \frac{2\rho}{\sigma_x\,\sigma_y}xy + \frac{2\rho\,\mu_x}{\sigma_x\,\sigma_y}y + \frac{2\rho\,\mu_y}{\sigma_x\,\sigma_y}x\right.\right.\\
&\qquad\qquad\left.\left. - \log(\sigma_x) - \log(\sigma_y) - \frac{1}{2}\log(1-\rho^2)\right\}\right]\frac{1}{2\pi}.
\end{aligned}$$

Clearly, this belongs to the multi-parameter exponential family. A minimal sufficient statistic for $(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$ is $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n Y_i^2, \sum_{i=1}^n X_iY_i\right)$. Another minimal sufficient statistic for this model is $\left(\overline{X},\overline{Y},S_x^2,S_y^2,R_{xy}\right)$, where

$$\begin{aligned}
\overline{X} &= \frac{1}{n}\sum_{i=1}^n X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i, \quad S_x^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2, \quad S_y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i-\overline{Y})^2,\\
R_{xy} &= \frac{\sum_{i=1}^n (X_i-\overline{X})\,(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^n (X_i-\overline{X})^2\;\sum_{i=1}^n (Y_i-\overline{Y})^2}}
\end{aligned}$$

are the sample mean, sample variance and sample correlation coefficient.

Solution 2. (a) There are several ways of deriving the expectations.

- Using the result derived in the class, the pdf of $X_{(1)}$ is

$$\begin{aligned}
f_{(1)}(y) &= \frac{n!}{(1-1)!\,(n-1)!}\left(\frac{y}{\theta}\right)^{1-1}\left(1-\frac{y}{\theta}\right)^{n-1}\frac{1}{\theta}, \qquad 0<y<\theta\\
&= \frac{n}{\theta}\left(1-\frac{y}{\theta}\right)^{n-1}, \qquad 0<y<\theta.
\end{aligned}$$

  Similarly, the pdf of $X_{(n)}$ is

$$\begin{aligned}
f_{(n)}(y) &= \frac{n!}{(n-1)!\,(n-n)!}\left(\frac{y}{\theta}\right)^{n-1}\left(1-\frac{y}{\theta}\right)^{n-n}\frac{1}{\theta}, \qquad 0<y<\theta\\
&= \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1}, \qquad 0<y<\theta.
\end{aligned}$$

  Now, $\mathbb{E}(X_{(1)})$ can be calculated as

$$\mathbb{E}(X_{(1)}) = \int y f_{(1)}(y)\,\mathrm{d}y = \int_0^\theta y\frac{n}{\theta}\left(1-\frac{y}{\theta}\right)^{n-1}\mathrm{d}y$$

$$
\begin{aligned}
&= \int_0^1 n\theta(1-z)z^{n-1}\,\mathrm{d}z \qquad \left(\text{substitute } 1-\frac{y}{\theta}=z\right)\\
&= n\theta\left(\frac{1}{n}-\frac{1}{n+1}\right) = \frac{\theta}{n+1}.
\end{aligned}
$$

Similarly,

$$
f_{(n)}(y) = \frac{n!}{(n-1)!\,(n-n)!}\left(\frac{y}{\theta}\right)^{n-1}\left(1-\frac{y}{\theta}\right)^{n-n}\frac{1}{\theta} = \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1}, \qquad 0<y<\theta.
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}(X_{(n)}) = \int y f_{(n)}(y)\,\mathrm{d}y &= \int_0^\theta y\frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1}\,\mathrm{d}y\\
&= \int_0^1 n\theta z^n\,\mathrm{d}z \qquad \left(\text{substitute } \frac{y}{\theta}=z\right)\\
&= n\theta\frac{1}{n+1} = \frac{n\theta}{n+1}.
\end{aligned}
$$

- Another important result, which helps in calculating expectations in some cases is the following:

  If X is a non-negative continuous random variable, i.e., $\mathbb{P}(X\geq 0)=1$, then

$$
\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X>t)\,\mathrm{d}t = \int_0^\infty \{1-\mathbb{P}(X\leq t)\}\,\mathrm{d}t.
$$

  To use this result, it is enough to know the cdf of X. Even knowing the pdf is not necessary. Now, recall from the class that

$$
\mathbb{P}(X_{(1)}>t) = \begin{cases} 1 & \text{if } t\leq 0,\\ \left(1-\dfrac{t}{\theta}\right)^n & \text{if } 0<t<\theta,\\ 0 & \text{if } t\geq\theta. \end{cases}
$$

  Therefore, using the result above, we get

$$
\mathbb{E}(X_{(1)}) = \int_0^\theta \left(1-\frac{t}{\theta}\right)^n\,\mathrm{d}t = \theta\int_0^1 z^n\,\mathrm{d}z = \frac{\theta}{n+1}.
$$

  Again, for $X_{(n)}$, we get

$$
\mathbb{P}(X_{(n)}>t) = 1-\mathbb{P}(X_{(n)}\leq t) = \begin{cases} 1 & \text{if } t\leq 0,\\ 1-\left(\frac{t}{\theta}\right)^n & \text{if } 0<t<\theta,\\ 0 & \text{if } t\geq\theta. \end{cases}
$$

  So, with the above result, we have

$$
\mathbb{E}(X_{(n)}) = \int_0^\theta \left\{1-\left(\frac{t}{\theta}\right)^n\right\}\,\mathrm{d}t = \theta - \int_0^\theta \frac{t^n}{\theta^n}\,\mathrm{d}t = \theta - \frac{\theta}{n+1} = \frac{n}{n+1}\theta.
$$

  Note: Be mindful that the formula for the expectation given above only applies to non-negative random variables. In some cases, it is possible to make a random variable positive by using "translation". Suppose $X\sim\mathsf{Uniform}(a,b)$, where $a<0$. If we define $Y=X+c$, where $c>|a|$, then $Y\sim\mathsf{Uniform}(a+c,b+c)$ and $a+c>0$. Since Y is non-negative, we can use the formula to compute $\mathbb{E}(Y)$. Finally, $\mathbb{E}(X)$ would simply be $\mathbb{E}(Y)-c$.

Note: Notice the stark difference between $\mathbb{E}(X_{(1)})$ and $\mathbb{E}(X_{(n)})$. While $\mathbb{E}(X_{(1)}) = \dfrac{\theta}{n+1}$, $\mathbb{E}(X_{(n)}) = \dfrac{n}{n+1}\theta$. So, $\mathbb{E}(X_{(n)}) = n\mathbb{E}(X_{(1)})$. Compared to these, we have $\mathbb{E}(X) = \dfrac{\theta}{2}$. Therefore, the extreme order statistics are pretty far from the actual random variable.

(b) $\mathbb{E}(R_n) = \mathbb{E}(X_{(n)}) - \mathbb{E}(X_{(1)}) = \dfrac{n}{n+1}\theta - \dfrac{1}{n+1}\theta = \dfrac{n-1}{n+1}\theta.$

(c) $\mathbb{E}(R_n) \to \theta$ as $n \to \infty$. Therefore, for large sample sizes, the sample range $R_n = X_{(n)} - X_{(1)}$ gets close to the population range $\theta - 0 = \theta$.

Solution 3. Using the result derived in the class, the pdf of the sample median is

$$\begin{aligned}
f_{\left(\frac{n+1}{2}\right)}(y) &= \frac{n!}{\left(\frac{n+1}{2}-1\right)\left(n-\frac{n+1}{2}\right)} \{F_X(y)\}^{\frac{n+1}{2}-1} \{1-F_X(y)\}^{n-\frac{n+1}{2}} f_X(y) \\
&= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} [F_X(y)\{1-F_X(y)\}]^{\frac{n-1}{2}} f_X(y).
\end{aligned}$$

- We say that a pdf f_X is symmetric about θ if $f_X(y-\theta) = f_X(\theta-y)\ \forall\, y$. It can be verified that if f_X is symmetric about θ and F_X is the corresponding cdf, then F_X satisfies $F_X(y-\theta) = 1-F_X(\theta-y)$.

  For example, you can verify that $\mathsf{Normal}(0,1)$ is symmetric around 0 and the corresponding cdf satisfies $\Phi(-y) = 1-\Phi(y)$.

- If a pdf f_X is symmetric about θ, then the corresponding (population) median is θ. This can be seen by noting that the above implies $F_X(\theta) = 1/2$.

- For $X \sim f_X$, if f_X is symmetric about θ and the expectation of X exists, then $\mathbb{E}(X) = \theta$.

Now, if $f_X(\cdot)$ is symmetric around θ, then

$$\begin{aligned}
f_{\left(\frac{n+1}{2}\right)}(y-\theta) &= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} [F_X(y-\theta)\{1-F_X(y-\theta)\}]^{\frac{n-1}{2}} f_X(y-\theta) \\
&= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} [\{1-F_X(\theta-y)\}F_X(\theta-y)]^{\frac{n-1}{2}} f_X(\theta-y) \\
&= f_{\left(\frac{n+1}{2}\right)}(\theta-y).
\end{aligned}$$

So, the pdf of the sample median $X_{\left(\frac{n+1}{2}\right)}$ is also symmetric about θ. Therefore, if the expectation exists, then it must equal θ. So, the sample median would be an unbiased estimator of the population median if (a) the distribution is symmetric and (b) the expectation of the sample median exists.

(a) $\mathsf{Uniform}(0,\theta)$:

$$f_X(y) = \begin{cases} \dfrac{1}{\theta} & \text{if } 0 < y < \theta, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(y) = \begin{cases} 0 & \text{if } y < 0, \\ \dfrac{y}{\theta} & \text{if } 0 \le y \le 1, \\ 1 & \text{if } y > 1. \end{cases}$$

Therefore,

$$f_{\left(\frac{n+1}{2}\right)}(y) = \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} \left(\frac{y}{\theta}\right)^{\frac{n-1}{2}} \left(1-\frac{y}{\theta}\right)^{\frac{n-1}{2}} \frac{1}{\theta}, \qquad 0 < y < \theta.$$

You can calculate $\mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}\right)$, $\mathbb{E}\left(X^2_{\left(\frac{n+1}{2}\right)}\right)$, and hence $\mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)}\right)$ with the above pdf. Also, note that the form of the pdf implies that $X_{\left(\frac{n+1}{2}\right)}/\theta \sim \mathsf{Beta}\left(\frac{n+1}{2}, \frac{n+1}{2}\right)$. Therefore,

$$\mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}/\theta\right) = \frac{\frac{n+1}{2}}{\frac{n+1}{2}+\frac{n+1}{2}} = \frac{1}{2} \quad \Rightarrow \quad \mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}\right) = \frac{\theta}{2}$$

$$\mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)}/\theta\right) = \frac{\frac{n+1}{2}\frac{n+1}{2}}{\left(\frac{n+1}{2}+\frac{n+1}{2}\right)^2\left(\frac{n+1}{2}+\frac{n+1}{2}+1\right)} = \frac{1}{2(n+2)} \quad \Rightarrow \quad \mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)}\right) = \frac{\theta^2}{2(n+2)}.$$

The population median is $\frac{\theta}{2}$. So, the sample median is unbiased for the population median.

Note: The distribution of X is symmetric about $\frac{\theta}{2}$.

(b) $\mathsf{Uniform}(\theta, \theta+1)$:

$$f_X(y) = \begin{cases} 1 & \text{if } \theta < y < \theta+1, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(y) = \begin{cases} 0 & \text{if } y < \theta, \\ y-\theta & \text{if } \theta \le y \le \theta+1, \\ 1 & \text{if } y > \theta+1. \end{cases}$$

Therefore,

$$f_{\left(\frac{n+1}{2}\right)}(y) = \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}(y-\theta)^{\frac{n-1}{2}}(1-y+\theta)^{\frac{n-1}{2}}, \qquad \theta < y < \theta+1.$$

Expectation and variance of $X_{\left(\frac{n+1}{2}\right)}$ can be calculated directly. Otherwise, notice that the above pdf implies $X_{\left(\frac{n+1}{2}\right)} \sim \mathsf{Beta}\left(\frac{n+1}{2}, \frac{n+1}{2}\right)$. So,

$$\mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)} - \theta\right) = \frac{1}{2} \quad \Rightarrow \quad \mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}\right) = \theta + \frac{1}{2}$$

$$\mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)} - \theta\right) = \frac{1}{2(n+2)} \quad \Rightarrow \quad \mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)}\right) = \frac{1}{2(n+2)}.$$

The population median is $\theta + \frac{1}{2}$. The sample median is unbiased for the population median.

Note: The distribution of X is symmetric about $\theta + \frac{1}{2}$.

(c) $\mathsf{Uniform}(-\theta, \theta)$:

$$f_X(y) = \begin{cases} \frac{1}{2\theta} & \text{if } -\theta < y < \theta, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(y) = \begin{cases} 0 & \text{if } y < -\theta, \\ \frac{y+\theta}{2\theta} & \text{if } -\theta \le y \le \theta, \\ 1 & \text{if } y > \theta. \end{cases}$$

Therefore,

$$f_{\left(\frac{n+1}{2}\right)}(y) = \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\left(\frac{y+\theta}{2\theta}\right)^{\frac{n-1}{2}}\left(1-\frac{y+\theta}{2\theta}\right)^{\frac{n-1}{2}}\frac{1}{2\theta}, \qquad -\theta < y < \theta.$$

Notice that $\left(X_{\left(\frac{n+1}{2}\right)} + \theta\right)/(2\theta) \sim \mathsf{Beta}\left(\frac{n+1}{2}, \frac{n+1}{2}\right)$. So,

$$\mathbb{E}\left(\left(X_{\left(\frac{n+1}{2}\right)} + \theta\right)/(2\theta)\right) = \frac{1}{2} \quad \Rightarrow \quad \mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}\right) = 0$$

$$\mathrm{Var}\left(\left(X_{\left(\frac{n+1}{2}\right)}+\theta\right)/(2\theta)\right)=\frac{1}{2(n+2)}\quad\Rightarrow\quad\mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)}\right)=\frac{2\theta^2}{n+2}.$$

The population median is 0. The sample median is unbiased for the population median.

Note: The distribution of X is symmetric about 0.

(d) $\mathsf{Normal}(\mu,\sigma^2)$:

$$\begin{aligned} f_X(y) &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y\in\mathbb{R}, & F_X(y) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{y} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t, \quad y\in\mathbb{R},\\ &= \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right), & &= \Phi\left(\frac{y-\mu}{\sigma}\right), \end{aligned}$$

where $\phi(\cdot)$ is the pdf and $\Phi(\cdot)$ is the cdf of $\mathsf{Normal}(0,1)$. Therefore,

$$f_{\left(\frac{n+1}{2}\right)}(y)=\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\frac{n-1}{2}}\left\{1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\frac{n-1}{2}}\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right),\qquad y\in\mathbb{R}.$$

Now,

$$\begin{aligned} \mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}\right) &= \int_{-\infty}^{\infty} y\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\frac{n-1}{2}}\left\{1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\frac{n-1}{2}}\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\mathrm{d}y\\ &= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\int_{-\infty}^{\infty}(\mu+\sigma t)\left\{\Phi(t)\right\}^{\frac{n-1}{2}}\left\{1-\Phi(t)\right\}^{\frac{n-1}{2}}\phi(t)\,\mathrm{d}t\left(\text{substitute } t=\frac{y-\mu}{\sigma}\right)\\ &= \mu\int_0^1\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}u^{\frac{n-1}{2}}(1-u)^{\frac{n-1}{2}}\,\mathrm{d}u\quad(\text{substituting } u=\Phi(t))\\ &\qquad +\sigma\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\int_{-\infty}^{\infty}t\left\{\Phi(t)\right\}^{\frac{n-1}{2}}\left\{1-\Phi(t)\right\}^{\frac{n-1}{2}}\phi(t)\,\mathrm{d}t\\ &= \mu+0=\mu. \end{aligned}$$

The last line is obtained by noting that the first integrand is the pdf of $\mathsf{Beta}\left(\frac{n+1}{2},\frac{n+1}{2}\right)$ and the second integrand is an odd function of t. Such closed-form expression for the variance does not exist. But, we can make some simplifications as shown below.

$$\begin{aligned} \mathbb{E}\left(X^2_{\left(\frac{n+1}{2}\right)}\right) &= \int_{-\infty}^{\infty} y^2\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\frac{n-1}{2}}\left\{1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right\}^{\frac{n-1}{2}}\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\mathrm{d}y\\ &= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\int_{-\infty}^{\infty}(\mu+\sigma t)^2\left\{\Phi(t)\right\}^{\frac{n-1}{2}}\left\{1-\Phi(t)\right\}^{\frac{n-1}{2}}\phi(t)\,\mathrm{d}t\left(\text{substitute } t=\frac{y-\mu}{\sigma}\right)\\ &= \mu^2\times 1+2\mu\sigma\times 0+\sigma^2\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\int_{-\infty}^{\infty}t^2\left\{\Phi(t)\right\}^{\frac{n-1}{2}}\left\{1-\Phi(t)\right\}^{\frac{n-1}{2}}\phi(t)\,\mathrm{d}t\\ &= \mu^2+\sigma^2\frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2}\int_0^1\left\{\Phi^{-1}(u)\right\}^2u^{\frac{n-1}{2}}(1-u)^{\frac{n-1}{2}}\,\mathrm{d}u\,(\text{substituting } u=\Phi(t))\\ &= \mu^2+\sigma^2\mathbb{E}\left[\left\{\Phi^{-1}(U)\right\}^2\right],\qquad U\sim\mathsf{Beta}\left(\frac{n+1}{2},\frac{n+1}{2}\right). \end{aligned}$$

Therefore, $\mathrm{Var}\left(X_{\left(\frac{n+1}{2}\right)}\right)=\sigma^2\mathbb{E}\left[\left\{\Phi^{-1}(U)\right\}^2\right]$.

The population median is μ. So, the sample median is unbiased for the population median.

Note: The distribution of X is symmetric about μ.

(e) Exponential(λ):

$$f_X(y) = \begin{cases} \dfrac{1}{\lambda}\, e^{-\frac{y}{\lambda}} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 - e^{-\frac{y}{\lambda}} & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} f_{\left(\frac{n+1}{2}\right)}(y) &= \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} \left(1 - e^{-\frac{y}{\lambda}}\right)^{\frac{n-1}{2}} \left(e^{-\frac{y}{\lambda}}\right)^{\frac{n-1}{2}} \frac{1}{\lambda} e^{-\frac{y}{\lambda}} \\ &= \frac{1}{\lambda} \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} \left(1 - e^{-\frac{y}{\lambda}}\right)^{\frac{n-1}{2}} \left(e^{-\frac{y}{\lambda}}\right)^{\frac{n+1}{2}}, \qquad y \geq 0. \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left(X_{\left(\frac{n+1}{2}\right)}\right) &= \frac{1}{\lambda} \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} \int_0^\infty y \left(1 - e^{-\frac{y}{\lambda}}\right)^{\frac{n-1}{2}} \left(e^{-\frac{y}{\lambda}}\right)^{\frac{n+1}{2}} \mathrm{d}y \\ &= -\lambda \frac{n!}{\left\{\left(\frac{n-1}{2}\right)!\right\}^2} \int_0^1 \log(u)\, u^{\frac{n-1}{2}} (1-u)^{\frac{n-1}{2}} \mathrm{d}u \qquad \left(\text{substituting } u = e^{-\frac{y}{\lambda}}\right) \\ &= \lambda \mathbb{E}[-\log(U)], \qquad U \sim \text{Beta}\left(\frac{n+1}{2}, \frac{n+1}{2}\right). \end{aligned}$$

Similarly,

$$\mathbb{E}\left(X^2_{\left(\frac{n+1}{2}\right)}\right) = \lambda^2 \mathbb{E}[\{\log(U)\}^2], \qquad U \sim \text{Beta}\left(\frac{n+1}{2}, \frac{n+1}{2}\right).$$

The median of X is $\lambda \log(2)$. So, the sample median is not unbiased for the population median.

(f) Cauchy(μ, σ):

$$f_X(y) = \frac{\sigma}{\pi} \frac{1}{\sigma^2 + (y-\mu)^2}, \quad y \in \mathbb{R}, \qquad F_X(y) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{y-\mu}{\sigma}\right), \quad y \in \mathbb{R}.$$

This can be used to derive the pdf of the median. But, the expectation in this case does not exist. However, the distribution of X is symmetric around μ. Therefore, the distribution of the sample median is also symmetric about μ.