Homework 2 Math 742 3.23, 6.2, 6.8 → graded for content 02/02/2026

David Cortes

Casella/Berger

3.1 Find expressions for (a) $E(X)$ and (b) $V(X)$ if X is a random variable with the general discrete uniform (N_0, N_1) distribution that puts equal probability on each of the values $N_0, N_{0+1}, \dots, N_1$. Where $N_0 \le N_1$ and N_0 and N_1 are both integers.

Identify the PMF, "Discrete Uniform on $\{N_0, \dots, N_1\}$" means each integer in the support has equal probability, so the PMF is:

$$f(x) = P(X=x) = \frac{1}{N_1 - N_0 + 1}, \quad x = N_0, N_0+1, \dots, N_1.$$

We do this because expectation/variance require the PMF.

For a discrete random variable $E[X] = \sum_x x P(X=x)$

(a) Then $E[X] = \sum_{x=N_0}^{N_1} x \cdot \frac{1}{N_1 - N_0 + 1} = \boxed{\frac{1}{N_1 - N_0 + 1} \sum_{x=N_0}^{N_1} x.}$ (i)

Now we evaluate the partial sum $\sum_{x=N_0}^{N_1} x = \sum_{x=1}^{N_1} x - \sum_{x=1}^{N_0-1} x.$

Use the standard formula $\sum_{x=1}^{m} x = \frac{m(m+1)}{2} \rightarrow \sum_{x=1}^{N_1} x = \frac{N_1(N_1+1)}{2}, \quad \sum_{x=1}^{N_0-1} x = \frac{(N_0-1)N_0}{2}$

Now combine and expand: Substract the two factor over a common denominator: factor out

$$\frac{N_1^2 + N_1}{2} - \frac{N_0^2 - N_0}{2} = \frac{N_1^2 + N_1 - N_0^2 + N_0}{2} = \frac{(N_1^2 - N_0^2) + (N_1 + N_0)}{2} = \frac{(N_1 - N_0)(N_1+N_0) + 1(N_1+N_0)}{2}$$

$$= \frac{(N_1+N_0)[(N_1 - N_0) + 1]}{2} = \frac{(N_1+N_0)(N_1 - N_0 + 1)}{2} \rightarrow \sum_{x=N_0}^{N_1} x = \frac{(N_1+N_0)(N_1-N_0+1)}{2}$$

↳ Replace/substitute in (i)

$$E[X] = \frac{1}{\cancel{(N_1 - N_0 + 1)}} \cdot \frac{(N_1+N_0)\cancel{(N_1 - N_0 + 1)}}{2} = \frac{(N_1+N_0)}{2}$$

(a) $E[X] = \frac{N_1 + N_0}{2}$

b) $\mathbb{V}[X] = \underbrace{E[X^2]}_{b_1} - \underbrace{(E[X])^2}_{b_2}$

b_1) $$E[X^2] = \sum_{x=N_0}^{N_1} x^2 \cdot \frac{1}{N_1 - N_0 + 1} = \frac{1}{N_1 - N_0 + 1} \cdot \boxed{\sum_{x=N_0}^{N_1} x^2} = \underbrace{\sum_{x=1}^{N_1} x^2}_{\boxed{b_{1.1}}} - \underbrace{\sum_{x=1}^{N_0-1} x^2}_{\boxed{b_{1.2}}}$$

Using $\boxed{\sum_{x=1}^{m} x^2 = \frac{m(m+1)(2m+1)}{6}}$ in $\boxed{b_{1.1}}$ ∧ $\boxed{b_{1.2}}$ we get

$$\boxed{b_{1.1}} \rightarrow \sum_{x=1}^{N_1} x^2 = \frac{N_1(N_1+1)(2N_1+1)}{6} \quad \text{and} \quad \boxed{b_{1.2}} \rightarrow \sum_{x=1}^{N_0-1} x^2 = \frac{(N_0-1)(N_0-1+1)(2(N_0-1)+1)}{6}$$

$$= \frac{(N_0-1)N_0(2N_0-2+1)}{6}$$

$$\boxed{b_{1.2}} \rightarrow \sum_{x=1}^{N_0-1} x^2 = \frac{(N_0-1)N_0(2N_0-1)}{6}$$

Susbtituing back $\boxed{b_{1.1}}$ ∧ $\boxed{b_{1.2}}$ in $\sum_{x=N_0}^{N_1} x^2 = \underbrace{\boxed{\sum_{x=1}^{N_1} x^2}}_{\boxed{b_{1.1}}} - \underbrace{\boxed{\sum_{x=1}^{N_0-1} x^2}}_{\boxed{b_{1.2}}}$ we get $\rightarrow$

$$\sum_{x=N_0}^{N_1} x^2 = \frac{N_1(N_1+1)(2N_1+1)}{6} - \frac{(N_0-1)N_0(2N_0-1)}{6} = \frac{(N_1(N_1+1)(2N_1+1)) - ((N_0-1)N_0(2N_0-1))}{6}$$

Then, lets substitute back into b_1

$$E[X^2] = \frac{1}{N_1 - N_0 + 1} \cdot \left[\frac{N_1(N_1-1)(2N_1+1) - (N_0-1)N_0(2N_0-1)}{6}\right]$$

Now we have

$$\mathbb{V}[X] = E[X^2] - (E[X])^2$$

To reduce the algebra, set

$$a = N_0,\ b = N_1,\ d = b - a,\ \rightarrow d = N_1 - N_0$$

$$\rightarrow N_1 - N_0 + 1 = b - a + 1 = d + 1$$

$E[X] = \frac{a+b}{2}$, but $b = a+d$, so $E[X] = \frac{a+(a+d)}{2} = \frac{2a+d}{2} = a + \frac{d}{2}$

Now square it:

$$(E[x])^2 = \left(a + \frac{d}{2}\right)^2 \longrightarrow \left(a+\frac{d}{2}\right)^2 = a^2 + 2a \cdot \frac{d}{2} + \left(\frac{d}{2}\right)^2 = a^2 + ad + \frac{d^2}{4}$$

$$\rightarrow (E[X])^2 = a^2 + ad + \frac{d^2}{4}$$

for $E[X^2] = \frac{1}{b-a+1} \cdot \frac{b(b+1)(2b+1)-(a-1)a(2a-1)}{6}$, since $b-a+1 = d+1$, this is

$$E[x^2] = \frac{1}{d+1} \cdot \frac{b(b+1)(2b+1)-(a-1)a(2a-1)}{6}$$, Expand $b(b+1)(2b+1)$

$(b^2+b)(2b+1)$

$(b^2+b)(2b+1) = (b^2+b) \cdot 2b + (b^2+b) \cdot 1$

$(b^2+b) \cdot 2b = 2b^3 + 2b^2$

$(b^2+b) \cdot 1 = b^2 + b$

Add them $b(b+1)(2b+1) = 2b^3 + 2b^2 + b^2 + b = 2b^3 + 3b^2 + b$

Expand $(a-1)a(2a-1)$

$(a^2-a)(2a-1)$

$(a^2-a)(2a-1) = (a^2-a) \cdot 2a - (a^2-a) \cdot 1$

$(a^2-a) \cdot 2a = 2a^3 - 2a^2$, $(a^2-a) \cdot 1 = a^2 - a$

Substract $(a^2-a)(2a-1) = (2a^3-2a^2) - (a^2-a) = 2a^3 - 2a^2 - a^2 + a = 2a^3 - 3a^2 + a$

Now compute

$$b(b+1)(2b+1) - (a-1)a(2a-1)$$
$$= (2b^3+3b^2+b) - (2a^3-3a^2+a)$$
$$= 2b^3 + 3b^2 + b - 2a^3 + 3a^2 - a$$

So

$$E[X^2] = \frac{1}{d+1} \cdot \frac{2b^3+3b^3+b-2a^3+3a^2-a}{6}$$, Now substitute $b = a+d$ and simplify

We need b^3, and b^2

$$b^2 = (a+d)^2 = a^2 + 2ad + d^2$$

$$b^3 = (a+d)^3 = a^3 + 3a^2d + 3ad^2 + a^3$$

plug it into $2b^3+3b^2+b$

$$2b^3 = 2(a^3+3a^2d+3ad^2+d^3) = 2a^3+6a^2d+6ad^2+2d^3$$

$$3b^2 = 3(a^2+2ad+d^2) = 3a^2+6ad+3d^2$$

$$b = a+d$$

Adding them all we get:

$$2b^3+3b^2+b = (2a^3+6a^2d+6ad^2+2d^3) + (3a^2+6ad+3d^2) + (a+d)$$

Now put this into the whole numerator for $2b^3+3b^2+b-2a^3+3a^2-a$:

$$\left[(2a^3+6a^2d+6ad^2+2d^3) + (3a^2+6ad+3d^2) + (a+d)\right] - 2a^3+3a^2-a$$

Combining like terms:

$2a^3-2a^3=0$, $a-a=0$, $3a^2+3a^2=6a^2$, we get

$6a^2d+6ad^2+2d^3+6a^2+6ad+3d^2+d. \rightarrow$

$$E[x^2] = \frac{d}{d+1}\cdot\frac{6a^2d+6ad^2+2d^3+6a^2+6ad+3d^2+d}{6}$$, factor out $(d+1)$

numerator $6a^2d+6a^2 = 6a^2(d+1)$, $6ad^2+6ad = 6ad(d+1)$, $\rightarrow$

$$6a^2(d+1)+6ad(d+1)+(2d^3+3d^2+d)$$

$$E[x^2] = \frac{1}{d+1}\cdot\frac{6a^2(d+1)+6ad(d+1)+(2d^3+3d^2+d)}{6}$$, Now, split the fraction in 3

$$E[x^2] = \frac{6a^2(d+1)}{6(d+1)} + \frac{6ad(d+1)}{6(d+1)} + \frac{2d^3+3d^2+d}{6(d+1)} = a^2+ad+\frac{2d^3+3d^2+d}{6(d+1)}$$

$$2d^3+3d^2+d = d(2d^2+3d+1) = d[(2d+1)(d+1)] \rightarrow$$

$$E[x^2] = a^2+ad+\frac{d(2d+1)}{6}$$

$$Var(X) = E[X^2] - (E[X])^2 = \left(a^2 + ad + \frac{d(2d+1)}{6}\right) - \left(a^2 + ad + \frac{d^2}{4}\right)$$

$$Var(X) = \frac{d(2d+1)}{6} - \frac{d^2}{4}$$, common denominator 12:

$$= \frac{d(2d+1)}{6} = \frac{2d(2d+1)}{12} \quad , \quad \frac{d^2}{4} = \frac{3d^2}{12} \quad ,$$ so...

$$Var(X) = \frac{2d(2d+1)}{12} - \frac{3d^2}{12} = \frac{2d(2d+1) - 3d^2}{12} = \frac{4d^2 + 2d - 3d^2}{12} = \frac{d^2 + 2d}{12}$$

$$Var(X) = \frac{d(d+2)}{12}$$ ← everything is expresed only in $d = N_1 - N_0$, then...

$$Var(X) = \frac{(N_1 - N_0)((N_1 - N_0) + 2)}{12}$$

3.4 A man with n keys wants to open his door and tries the keys at random. Exactly one key will open the door. Find the mean number of trials if,

a) unsuccessful keys are not eliminated from further selections.

b) unsuccesful key are eliminated.

- There are n keys
- Exactly **one** key opens the door.
- the drunk guy pick the keys "randomly" until the door get open.

And let X be the number of trials (attempts) until the first time you open the door.

S = success on trial (the door opens) F = failure on a trial (door dosen't open)

a) Unsuccessful keys are not eliminated. This means, after a failure, you put the key back, and next trial you again choose uniformly from all n keys.

Probability of success/failure on any trial. Because each trial you choose randomly among all n keys:

$P(S) = \frac{1}{n}$ (only 1 good key out of n) , $P(F) = 1 - P(S) = 1 - \frac{1}{n} = \frac{n-1}{n}$

Write the event $\{X=x\}$ as a sequence "$X=x$" means: you fail the first $x-1$ times, then succeed on the x-th time. So:

$$\{X=x\} = \underbrace{F\,F \cdots F}_{x-1 \text{ times}}\,S.$$

Now, lets compute $P(X=x)$. Because we are replacing keys, each trial has the same succees probabitity $\frac{1}{n}$ and failures are $\frac{n-1}{n}$ each time. Thus

$$p(X=x) = \left(\frac{n-1}{n}\right)^{x-1}\left(\frac{1}{n}\right), \quad x = 1,2,3,\ldots$$ This is the geometric pmf with parameter $p=\frac{1}{n}$

Now, lets use the mean of a geometric random variable, For geometric (p) defined as "# trials until first success" $E[X] = \frac{1}{p}$ here $p=\frac{1}{n}$, so $E(x) = \frac{1}{1/n} = n$

What happens when we eliminate the unsuccessful key?.

(b) Unsuccesful keys are eliminated. This meens, after you try a key and it fails, we discard it, so we never pick it again. Basically we are sampling whitout replacement from n keys until we get the good key.

The worst scenario, since I never repeat a failed key is: try all $n-1$ wrong keys first, and at the end the rigth one. So $X \in \{1,2,\ldots,n\}$

$P(x=1) \to P(x=1) = \frac{1}{n}$, $P(X=2) \to P(x=2) = \underbrace{\text{first trial wrong key}}_{\frac{n-1}{n}} \cdot \underbrace{\text{second trial good key}}_{\frac{1}{n-1}}$

$$P(x=2) = \frac{\cancel{n-1}}{n} \cdot \frac{1}{\cancel{n-1}} = \frac{1}{n}$$

$$P(X=3) = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \frac{1}{n-2} = \frac{1}{n}$$

General pattern $P(X=k) = \frac{1}{n}$ for $k \in \{1,\ldots,n\}$ the event $X=k$ means: first $k-1$ picks are wrong, and the k-th pick is correct.

trial 1 $\to$ wrong: $\frac{n-1}{n}$

trial 2 $\to$ wrong: $\frac{n-2}{n-1}$

trial 3 $\to$ wrong: $\frac{n-3}{n-2}$

trial $k-1 \to$ Wrong: $\frac{n-(k-1)}{n-(k-2)} = \frac{n-k+1}{n-k+2}$

trial $k \to$ good! , After removing $k-1$ wrong keys, $n-(k-1)$: $\frac{1}{n-k+1}$

$$P(X=k) = \left(\frac{\cancel{n-1}}{n}\right)\left(\frac{\cancel{n-2}}{\cancel{n-1}}\right)\cdots\left(\frac{\cancel{n-k+1}}{\cancel{n-k+2}}\right)\left(\frac{1}{\cancel{n-k+1}}\right) = \frac{1}{n}$$

Then canceling in order $(n-1), (n-2) \cdots (n-k+1)$

This means X is uniform on $\{1, 2, \dots, n\}$

Now, lets compute $E[X]$. By definition $E[X] = \sum_{k=1}^{n} k\, P(X=k)$ since $P(X=k) = \frac{1}{n}$

$$E[X] = \sum_{k=1}^{n} k\left(\frac{1}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} k \quad \text{using} \quad \sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

$$E[X] = \frac{1}{n}\cdot\frac{n(n+1)}{2} = \frac{n+1}{2}$$

3.23 The Pareto Distribution, with parameters α and β, has pdf

$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}$; where $\alpha < x < \infty$, $\alpha > 0$, $\beta > 0$

(a) Verify that $f(x)$ is a PDF

(b) Derive the mean and variance of this distribution

(c) Prove that the variance does not exists if $\beta \leq 2$

(a) A function is a PDF if:

(1) It is nonnegative

(2) It integrates to 1

why? More...

For (1) $x \geq \alpha$, $\beta > 0$, $\alpha^\beta > 0$, and $x^{\beta+1} > 0$, so $\frac{\beta\alpha^\beta}{x^{\beta+1}} \geq 0$

for $x < \alpha$, $f_X(x) = 0$

So $f_X(x) \geq 0$ for all x.

(2) Total integral equals 1.

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} 0\,dx + \int_{-\infty}^{\infty} \frac{\beta\alpha^\beta}{x^{\beta+1}}\,dx = \int_{-\infty}^{\infty} \beta\alpha^\beta x^{-(\beta+1)}\,dx$$

Now Integrate $x^{-(\beta+1)}$: Recall the power rule: for $k \neq -1 \longrightarrow \int x^k\,dx = \frac{x^{k+1}}{k+1}$

Here $k = -(\beta+1)$, so $k+1 = -\beta \neq 0$ since $\beta > 0 \longrightarrow \int x^{-(\beta+1)}\,dx = \frac{x^{-\beta}}{-\beta}$

So $$\int_{\alpha}^{\infty} \beta\alpha^\beta x^{-(\beta+1)}\,dx = \beta\alpha^\beta\left[\frac{x^{-\beta}}{-\beta}\right]_{\alpha}^{\infty} = \beta\alpha^\beta\left(0 - \frac{\alpha^{-\beta}}{-\beta}\right)$$

Why is the "∞" term 0? Because $x^{-\beta} = \frac{1}{x^{\beta}} \longrightarrow 0$ as $x \longrightarrow \infty$ when $\beta > 0$

$= \beta \alpha^{\beta} \left(\frac{\alpha^{-\beta}}{\beta} \right) = \alpha^{\beta} \alpha^{-\beta} = 1$ so the PDF integrates to 1. ✓

ⓑ Derive mean and variance

By definition (continuos case)

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{\alpha}^{\infty} x \cdot \frac{\beta \alpha^{\beta}}{x^{\beta+1}} dx \longrightarrow E[X^2] = \int_{\alpha}^{\infty} x^2 \cdot \frac{\beta \alpha^{\beta}}{x^{\beta+1}} dx.$$

Compute $E[X]$

$$E[X] = \int_{\alpha}^{\infty} x \cdot \frac{\beta \alpha^{\beta}}{x^{\beta+1}} dx = \int_{\alpha}^{\infty} \beta \alpha^{\beta} x^{1-(\beta+1)} dx = \int_{\alpha}^{\infty} \beta \alpha^{\beta} x^{-\beta} dx$$

Integrate $x^{-\beta}$. there are two cases:

- If $\beta \neq 1$, $\int x^{-\beta} dx = \frac{x^{1-\beta}}{1-\beta} \longrightarrow E[X] = \beta \alpha^{\beta} \left[\frac{x^{1-\beta}}{1-\beta} \right]_{\alpha}^{\infty}$ = ?

Now lets check the convergence at ∞:

If $\beta > 1 \longrightarrow 1-\beta < 0$, so $x^{1-\beta} \longrightarrow 0$ as $x \longrightarrow \infty$

If $\beta \leq 1 \longrightarrow 1-\beta \geq 0$, so $x^{1-\beta}$ does NOT go to 0; the integral diverges. ✓

Assuming $\beta > 1$ (so the mean exists), we get

$$E[X] = \beta \alpha^{\beta} \left(0 - \frac{\alpha^{1-\beta}}{1-\beta} \right) = \beta \alpha^{\beta} \cdot \frac{\alpha^{1-\beta}}{\beta - 1} = \frac{\beta \alpha}{\beta - 1}$$ OK

Compute $E[X^2] = \int_{\alpha}^{\infty} x^2 \cdot \frac{\beta \alpha^{\beta}}{x^{\beta+1}} dx = \int_{\alpha}^{\infty} \beta \alpha^{\beta} x^{2-(\beta+1)} dx = \int_{\alpha}^{\infty} \beta \alpha^{\beta} x^{1-\beta} dx$. Again, two cases

Integrate $x^{1-\beta}$. If $1 - \beta \neq -1 \rightarrow \beta = -2 \rightarrow \int x^{1-\beta} dx = \frac{x^{2-\beta}}{2-\beta}$

So $E[X^2] = \beta \alpha^{\beta} \left[\frac{x^{2-\beta}}{2-\beta} \right]_{\alpha}^{\infty}$

Check convergence at ∞:

If $\beta > 2 \longrightarrow 2-\beta < 0 \longrightarrow x^{2-\beta} \longrightarrow 0$ as $x \to \infty$

If $\beta \leq 2 \longrightarrow$ It diverges (lets formalize this in part ⓒ)

Assuming $\beta > 2$:

$$E[X^2] = \beta\alpha^\beta \left(0 - \frac{\alpha^{2-\beta}}{2-\beta}\right) = \beta\alpha^\beta \cdot \frac{\alpha^{2-\beta}}{\beta-2} = \frac{\beta\alpha^2}{\beta-2}$$

Compute $Var(X)$, By definition $Var(X) = E[X^2] - (E[X])^2$

$$Var(X) = \frac{\beta\alpha^2}{\beta-2} - \left(\frac{\beta\alpha}{\beta-1}\right)^2 \longrightarrow \text{factor out } \alpha^2:$$

$$Var(X) = \alpha^2\left(\frac{\beta}{\beta-2} - \frac{\beta^2}{(\beta-1)^2}\right) \longrightarrow$$ put the two fractions over a common denominator $(\beta-2)(\beta-1)^2$:

$$\frac{\beta}{\beta-2} = \frac{\beta(\beta-1)^2}{(\beta-2)(\beta-1)^2}, \quad \frac{\beta^2}{(\beta-1)^2} = \frac{\beta^2(\beta-2)}{(\beta-1)^2(\beta-2)} \longrightarrow$$

$$Var(X) = \alpha^2 \cdot \frac{\overset{\beta}{\cancel{\beta(\beta-1)^2 - \beta^2(\beta-2)}}}{(\beta-2)(\beta-1)^2}$$? This matches the Standard Pareto variance formula ?

ⓒ Prove the variance does not exist if $\beta \leq 2$

Two show $Var(X)$ does not exist, it's important to show $E[X^2] = \infty$

We already got $E[X^2] = \int_\alpha^\infty \beta\alpha^\beta x^{1-\beta}\,dx$

Lets analyze the integral

$\int_\alpha^\infty x^{1-\beta}\,dx$

$\longrightarrow$ If $\beta < 2 \longrightarrow 1-\beta > -1$ and $\int_\alpha^\infty x^{1-\beta}dx$ diverges because the power is too large (the tail does not decay fast enough).

$\longrightarrow$ If $\beta = 2 \longrightarrow 1-\beta = -1$ and $\int_\alpha^\infty x^{-1}dx = \int_\alpha^\infty \frac{1}{x}\,dx = \operatorname{Log} x \Big|_\alpha^\infty = \infty$ Also diverges

Therefore $E[X^2] = \infty$ for $\beta \leq 2$, so $Var(X)$ does not exist for $\beta \leq 2$ ✓

6.2 Let $X_1, X_2, \dots, X_n$ be independent random variables with densities

$$f_{X_i}(x|\theta) = \begin{cases} e^{i\theta - x} & : x \geq i\theta \\ 0 & : x < i\theta \end{cases}$$

Prove that $T = \min(X_i/i)$ is a sufficient statistic for θ.

Road map:

① Redefine the PDF by using an indicator function:

$$f_{X_i}(x|\theta) = e^{i\theta - x} \mathbf{1}\{x \geq i\theta\}$$

② Find an expression for the joint PDF: $f_X(x|\theta)$

③ Rewrite the joint PDF using the statistic $T = \min(X_i/i)$

④ Use the factorization theorem to show that T is sufficient for θ.

① The density is only nonzero when $x \geq i\theta$. The clean way to write "only if $x \geq i\theta$" is with an indicator function.

$$1_{[i\theta,\infty)}(x) = \begin{cases} 1, & x \geq i\theta \\ 0, & x < i\theta \end{cases} \longrightarrow \text{Rewrite} \longrightarrow f_{X_i}(x|\theta) = e^{i\theta - x} 1_{[i\theta,\infty)}(x)$$

② Since $X_1, \dots, X_n$ are independent variables, the joint probability density function is the product of the individual marginal PDFs

$$f_X(x|\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta) \xrightarrow{\text{substitute the expression from ①}} f_X(x|\theta) = \prod_{i=1}^{n} \Big(\underbrace{e^{i\theta - x_i}}_{\text{Part A}} \cdot \underbrace{1_{[x_i \geq i\theta]}}_{\text{part B}} \Big)$$

Part A (Exponentials):

$$\prod_{i=1}^{n} e^{i\theta - x_i} = e^{\sum_{i=1}^{n}(i\theta - x_i)} = e^{\theta\sum_{i=1}^{n} i - \sum_{i=1}^{n} x_i}$$

Recall that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \longrightarrow e^{\frac{n(n+1)}{2}\theta - \sum_{i=1}^{n} x_i} = e^{\frac{n(n+1)}{2}\theta} \cdot e^{-\sum_{i=1}^{n} x_i}$

Part B (Indicators):

$$\prod_{i=1}^{n} 1_{\{x_i \geq i\theta\}} = 1_{\{x_1 \geq \theta\}} \cdot 1_{\{x_2 \geq 2\theta\}} \cdot \dots \cdot 1_{\{x_n \geq n\theta\}}$$

For the product of these indicators to be 1, all conditions must be true simultaneosly

$$x_1 \geq \theta \wedge \frac{x_2}{2} \geq \theta \wedge \dots \wedge \frac{x_n}{n} \geq \theta$$

?

This is equivalent to saying that the minimum of these values must be greater than or equal to θ.

$$1\{\min(\frac{x_1}{1}, \frac{x_2}{2}, \dots, \frac{x_n}{n}) \geq \theta\}$$

Now combining Part A and Part B

$$f_X(x|\theta) = e^{\frac{n(n+1)}{2}\theta} \cdot e^{-\sum_{i=1}^{n} x_i} \cdot 1_{\{\min(x_i/i) \geq \theta\}}$$

③ Rewrite the joint PDF using the statistic $T = \min(X_i/i)$

We define our statistic $T = \min(X_i/i)$. Substituting T into the indicator function from step ②, we get the joint PDF purely in terms of the data, the statistic T, and the parameter θ:

$$f_X(x|\theta) = \underbrace{e^{\frac{n(n+1)}{2}\theta} \cdot 1_{\{T \geq \theta\}}}_{\text{Terms involving } \theta \text{ and } T} \cdot \underbrace{e^{-\sum_{i=1}^{n} x_i}}_{\text{Terms involving only } x}$$

④ Use the factorization theorem to show that T is sufficient for θ

The factorization theorem states that $T(X)$ is a sufficient statistic for $\theta \Leftrightarrow$ the joint PDF can be factored into two non-negative functions:

$$f_X(x|\theta) = g(T(x)|\theta) \cdot h(x) \longrightarrow \text{from our result in step ③ we can}$$

identify these functions:

1. $g(T|\theta)$: This depends on the data only through the statistic T:

$$g(T|\theta) = e^{\frac{n(n+1)}{2}\theta} \cdot 1_{\{T \geq \theta\}}$$

2. $h(x)$: This depends only on the data x and not on θ:

$$h(x) = e^{-\sum_{i=1}^{n} x_i}$$

Since the joint PDF $f_X(x|\theta)$ factors into a part that depends on θ only through T and a part that does NOT depend on θ, $T = \min(X_i/i)$ is a sufficient statistic for θ

6.8 Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with location PDF $f(x-\theta)$. Show that the order statistics, $T(X_1, X_2, \ldots, X_n) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ are a sufficient statistic for θ and no further reductions is possible. More specifically, a population has a location PDF if there exists a fixed PDF f_0 such that for some unknown location parameter θ,

$$f_X(x|\theta) = f_0(x-\theta), \text{ where } -\infty < \theta < \infty$$

Equivalently, if $Y \sim f_0$, then $X = \theta + Y$

Location PDF means there exists a fixed PDF f_0 (no params) such that

$f_X(x|\theta) = f_0(x-\theta)$, where θ is the unknown location parameter.

Equivalently, if $Y \sim f_0$, then $X = Y + \theta$, So, the density of one observation X_i is

$$f_{X_i}(x_i|\theta) = f_0(x_i-\theta), \quad -\infty < x_i < \infty$$

Joint PDF of the sample

Because $X_1, \ldots, X_n$ are i.i.d from that location family,

$$f_X(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_0(x_i-\theta)$$

This expression already has no extra indicators (Support is $\mathbb{R}$ with full density) So we don't need $1\{\cdot\}$ functions here.

for a specific sample $x = (x_1, \ldots, x_n)$ define the order statistics:

The order statistics $X_1 \le X_2 \le \cdots \le X_n$ represent the sorted values of the sample $x_1, \ldots, x_n$. A crucial property of the product function is that the order of the terms does not matter (multiplication is commutative).

Now lets Rewrite the joint PDF in terms of the order statistics using the ordered sample

$$f_X(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_0(x_i-\theta)$$

Because $x_1, \ldots, x_n$ is just a permutation of $x_{(1)}, \ldots, x_{(n)}$

Note: The product $\prod_{K=1}^{n} f_0(X_{(K)}-\theta)$ does NOT depend on which permutation π was used: The same multiset of values appears in the product no matter in which order you multiply them. Then:

The θ-dependent part is entirely in $\prod_{K=1}^{n} f_0(X_{(K)}-\theta)$, also any information about the original order $x_1, \dots, x_n$ vs the sorted order is independent of θ. We can therefore see the joint PDF as: $f_X(x_1, \dots, x_n | \theta) = g(X_{(1)}, \dots, X_{(n)}, \theta)\, h(x_1, \dots, x_n)$

with $g(X_{(1)}, \dots, X_{(n)}, \theta) = \prod_{K=1}^{n} f_0(X_{(K)}-\theta)$ and a h that is "1" or some combinatoria constant depending only on the permutation/multiplicity structure of the sample but not on θ. Formally, if all X_i are distinct, there are $n!$ permutations that give the same ordered vector; but all those permutations have identical value of $\prod_{K=1}^{n} f_0(X_{(K)}-\theta)$, so that multiplicity can be absorved into h, which does not depend on θ.

Now, lets Apply the factorization theorem: Sufficiency of order statistics.

The factorization theorem says:

A Statistic $T(X)$ is sufficient for θ if the joint PDF can be written as
$f(x|\theta) = g(T(x), \theta)\, h(x)$ where: g depends on X only through $T(x)$ and on θ,
h depends on x but not on θ.

In our case:

Define the statistic $T(X) = (X_{(1)}, \dots, X_{(n)})$. Then we have $f_X(x|\theta) = \left[\prod_{K=1}^{n} f_0(X_{(K)}-\theta)\right] \cdot h$

where: $g(T(x), \theta) = \prod_{K=1}^{n} f_0(X_{(K)}-\theta)$

$h(x)$ is whatever is left and does not depend on θ (it encodes only the permutati information).

Thus the joint PDF factorized in the required way, so the order statistics are sufficient for θ.

Now, about "No further reduction is possible" part we need to show. that we can not compress $(X_{(1)}, \dots, X_{(n)})$ into a lower-dimensional statistic and still retain sufficiency. Reading Casella/Berger I found that a very standard way to argue minimal sufficiency in this setting (location family) is:

1. Using the theorem 6.2.13 "Likelihood ratio criterion"
2. For a location family $f_0(x-\theta)$, compute the Likelihood ratio $\frac{f(x|\theta)}{f(y|\theta)}$ and see when it is constant in θ.

Likelihood ratio for two samples x and y

Given two samples $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$, their joint densities are:

$$f(x|\theta)=\prod_{i=1}^{n} f_0(x_i-\theta), \quad f(y|\theta)=\prod_{i=1}^{n} f_0(y_i-\theta)$$

The Likelihood ratio is

$$R(\theta;x,y)=\frac{f(x|\theta)}{f(y|\theta)}=\prod_{i=1}^{n}\frac{f_0(x_i-\theta)}{f_0(y_i-\theta)}$$

for which pairs x,y is this ratio independent of θ?

Because f_0 is fixed and not degenerate, the only way this product can be independent of θ (for all real θ) is if, up to reordering, the sets $\{x_1-\theta,\dots,x_n-\theta\}$ and $\{y_1-\theta,\dots,y_n-\theta\}$ coincide for all θ. But subtracting the same θ from two lists preserves their "patterns", and that happens only if the ordered lists $(x_{(1)},\dots,x_{(n)})$ and $(y_{(1)},\dots,y_{(n)})$ are exactly the same.

Therefore, by theorem 6.2.13 the equivalence classes "$R(\theta;x,y)$ constant in θ" are exactly the classes of all samples having the same order statistics. So the function that labels those equivalence classes is precisely the full vector of order statistics.

Thus, no non-trivial function of the data can be sufficient and strictly simpler than the full vector of order statistic. In other words, the order statistics are minimal sufficient.

Then, By the likelihood-ratio argument (Theorem 6.2.13), any two samples give a θ-independent ratio if and only if their ordered versions are equal, so the order statistics are minimal sufficient, then no further reduction is possible.

What if y is not a permutation of x?

6.9b For the following distributions, let $X_1, \dots, X_n$ be a random sample. Find a minimal sufficient statistic for θ

$$f(x|\theta) = e^{-(x-\theta)}, \quad \text{where} \quad \theta < x < \infty, \quad -\infty < \theta < \infty$$

more specifically,

$$f(x,y) = \begin{cases} e^{-(x-\theta)} & : x > 0 \\ 0 & : x \leq 0 \end{cases}$$

Location exponential! $\rightarrow$ One observation has density:

$$f(x|\theta) = e^{-(x-\theta)} \mathbf{1}\{x \geq \theta\}$$

Equivalently $X = \theta + Y$ where $Y \sim \text{Exp}(1)$

Let $X_1, \dots, X_n$ be an i.i.d. (Independent and Identically distributed) random sample from this distribution.

By independence, the joint density is the product:

$$f(x_1, \dots, x_n | \theta) = \prod_{i=1}^{n} \underbrace{e^{-(x_i-\theta)}}\mathbf{1}\{x_i \geq \theta\}$$

$$\prod_{i=1}^{n} e^{-(x_i-\theta)} = e^{-\sum_{i=1}^{n}(x_i-\theta)} = e^{-\sum_{i=1}^{n} x_i} e^{n\theta}$$

and

$$\prod_{i=1}^{n} \mathbf{1}\{x_i \geq \theta\} = \mathbf{1}\boxed{\{\text{all } x_i \geq \theta\}} \rightarrow \text{it's equivalent to } \min_i x_i \geq \theta \rightarrow$$

$$\prod_{i=1}^{n} \mathbf{1}\{x_i \geq \theta\} = \mathbf{1}\{\theta \leq \min_i x_i\}$$ Thus the joint PDF becomes

$$f(x_1, \dots, x_n | \theta) = e^{n\theta} \cdot e^{-\sum_{i=1}^{n} x_i} \cdot \mathbf{1}\{\theta \leq \min_i x_i\}$$

Let's use Factorization theorem to get sufficiency of $\min X_i$

$$T(x) = \min_{1 \le i \le n} X_i$$

Rewrite the joint PDF as

$$f(x|\theta) = \underbrace{\left(e^{n\theta} \mathbb{1}\{\theta \le T(x)\}\right)}_{g(T(x),\theta)} \underbrace{\left(e^{-\sum_{i=1}^{n} x_i}\right)}_{h(x)}$$

Here g depends on the sample only through $T(X) = \min x_i$ and $h(x)$ does not involve θ. Therefore, by the factorization theorem, $T(x) = \min X_i$ is sufficient for θ.

Now we use the minimal sufficiency criterion

Take two sample points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ Consider $\dfrac{f(x|\theta)}{f(y|\theta)}$

Use the joint PDF:

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{e^{n\theta} \cdot e^{-\sum_{i=1}^{n} x_i} \cdot \mathbb{1}\{\theta \le \min x_i\}}{e^{n\theta} \cdot e^{-\sum_{i=1}^{n} y_i} \cdot \mathbb{1}\{\theta \le \min y_i\}} = e^{-\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i} \cdot \frac{\mathbb{1}\{\theta \le \min x_i\}}{\mathbb{1}\{\theta \le \min y_i\}}$$

This factor does not depend on θ in consequence, the only way the whole ratio is constant in θ is if the indicator ratio does not depend on θ.

But, the ratio is dependent of θ if and only if:

$$\min x_i = \min y_i$$

Because if $\min X_i \ne \min y_i$ there is a range of θ values where one indicator is 1 and the other is 0, making the ratio change with θ.

Therefore, by Theorem 6.2.13 $T(x) = \min X_i$ is minimal sufficient.

Extra Problem:

Suppose that $\overbrace{X_1 \text{ and } X_2}^{\text{i.i.d}} \sim$ Bernoulli (p). Let $T(X) = (X_1 + X_2)$

Show that:

- T is sufficient for p.
- T is also minimal sufficient.

X_1 and $X_2 \sim$ Bernoulli $(p) \longrightarrow P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$

Then the PMF can be written as

$$f_{X_i}(x_i \mid p) = p^{x_i}(1-p)^{1-x_i}, \quad x_i \in \{0,1\}$$

$\rightarrow \quad T(x) = X_1 + X_2$

The Joint PMF $f(x_1, x_2 \mid p)$

By Independence $\quad f(x_1, x_2 \mid p) = f_{X_1}(x_1 \mid p)\, f_{X_2}(x_2 \mid p)$

Substitute the Bernoulli PMF:

$$f(x_1, x_2 \mid p) = \left(p^{x_1}(1-p)^{1-x_1}\right)\left(p^{x_2}(1-p)^{1-x_2}\right)$$

$$f(x_1, x_2 \mid p) = p^{x_1 + x_2}(1-p)^{(1-x_1)+(1-x_2)} = p^{x_1+x_2}(1-p)^{2-(x_1+x_2)}$$

Now recognize $x_1 + x_2 = T(x)$, so $f(x_1, x_2 \mid p) = p^{T(x)}(1-p)^{2-T(x)}$, $(x_1, x_2) \in \{0,1\}$

Factorization Theorem $\Rightarrow T$ is sufficient

$$f(x_1, x_2 \mid p) = \underbrace{p^{T(x)}(1-p)^{2-T(x)}}_{g(T(x),\, p)} \cdot \underbrace{1}_{h(x_1, x_2)}$$

Here h does not depend on p, and g depends on the data only through $T(x) = x_1 + x_2$

Therefore, by the factorization theorem, $T = X_1 + X_2$ is a sufficient statistic for p.

Now lets try to show that T is minimal sufficient

$x = (x_1, x_2)$ and $y = (y_1, y_2)$

$$\frac{f(x|p)}{f(y|p)} = \frac{p^{x_1+x_2}(1-p)^{2-(x_1+x_2)}}{p^{y_1+y_2}(1-p)^{2-(y_1+y_2)}} = p^{(x_1+x_2)-(y_1+y_2)}(1-p)^{[(2-(x_1+x_2)-(2-(y_1+y_2))]} =$$

$$\frac{f(x|p)}{f(y|p)} = \left(\frac{p}{1-p}\right)^{(x_1+x_2)-(y_1+y_2)}$$

Now, lets check when the ratio is constant in p:

- If $(x_1+x_2) = (y_1+y_2)$ $\rightarrow$ the exponent $= 0$, and the ratio $= 1$ for all p (constant)
- If $(x_1+x_2) \neq (y_1+y_2)$ $\longrightarrow$ then the ratio is $\left(\frac{p}{1-p}\right)^k$ with $k \neq 0$, wich clearly depends on p (not constant)

So the ratio $\frac{f(x|p)}{f(y|p)}$ is constant in p if and only if $T(x) = T(y)$

By theorem 6.2.13 $\rightarrow$ $T = X_1 + X_2$ is minimal sufficient.