STAT 24400 Lecture 7
Section 3.6 Functions of 2+ Random Variables
Section 3.7 Order Statistics

Yibi Huang
Department of Statistics
University of Chicago

Functions of 2+ Random Variables

Sum of Two Discrete Random Variables

If X and Y are discrete random variables with joint PMF $p(x, y)$, the PMF for $Z = X + Y$ is

$$p_Z(z) = \mathrm{P}(X + Y = z) = \sum_{\{(x,y):x+y=z\}} p(x, y) = \sum_x p(x, z - x).$$

Example 1: Sum of Independent Binomial R.V.'s

Suppose $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$ are independent. Their joint PMF is thus

$$\begin{aligned} p(x, y) &= \binom{m}{x} p^x (1-p)^{m-x} \cdot \binom{n}{y} p^y (1-p)^{n-y} \\ &= \binom{m}{x}\binom{n}{y} p^{x+y} (1-p)^{m+n-(x+y)}, \qquad \begin{array}{c} 0 \leq x \leq m, \\ 0 \leq y \leq n. \end{array} \end{aligned}$$

The PMF for $Z = X + Y$ is thus

$$\begin{aligned} p_Z(z) = \sum_x p(x, z-x) &= \sum_{x=0}^{z} \binom{m}{x}\binom{n}{z-x} p^z (1-p)^{m+n-z} \\ &= \binom{m+n}{z} p^z (1-p)^{m+n-z}, \end{aligned}$$

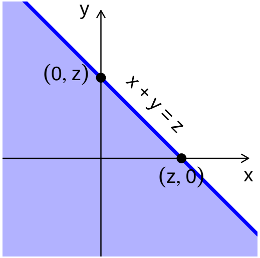
where $\binom{m+n}{z} = \sum_{x=0}^{z} \binom{m}{x}\binom{n}{z-x}$ is the Vandermonde identity.

This shows $X + Y \sim \text{Bin}(m+n, p)$.

Sum of Two Continuous Random Variables

Suppose X and Y are continuous random variables with joint PDF $f(x, y)$. The CDF for $Z = X + Y$ is integrating $f(x, y)$ over the shaded region $\{(x, y)\colon x + y \le z\}$.

$$
\begin{aligned}
F_Z(z) = \mathrm{P}(Z \le z) &= \iint_{\{(x,y):x+y\le z\}} f(x, y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x, y)\mathrm{d}y\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x, v - x)\mathrm{d}v\mathrm{d}x \quad \text{let } y = v - x \\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x, v - x)\mathrm{d}x\mathrm{d}v \left(\begin{matrix}\text{swapping order}\\ \text{of integration}\end{matrix}\right)
\end{aligned}
$$



The PDF is thus

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)\mathrm{d}x.$$

Example 2: Sum of Two Independent Gamma R.V.'s

Suppose $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$ are indep. Their joint PDF is

$$f(x,y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \cdot \frac{\lambda^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y} = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)}, \qquad \begin{array}{l} x > 0 \\ y > 0. \end{array}$$

$\Rightarrow$ $f(x, z-x)$ is defined for $x > 0$ and $z - x > 0$, i.e., $0 \leq x \leq z$.
The PDF for $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x) \mathrm{d}x = \frac{\lambda^{\alpha+\beta} e^{-\lambda z}}{\Gamma(\alpha)\Gamma(\beta)} \int_{x=0}^{z} x^{\alpha-1}(z-x)^{\beta-1} \mathrm{d}x.$$

Making a change of Variable: $u = x/z \;\Rightarrow\; x = zu$, $\mathrm{d}x = z\mathrm{d}u$, we get

$$\begin{aligned} \int_{x=0}^{z} x^{\alpha-1}(z-x)^{\beta-1} \mathrm{d}x &= \int_0^1 (uz)^{\alpha-1}(z-uz)^{\beta-1} z \mathrm{d}u \\ &= z^{\alpha+\beta-1} \int_0^1 u^\alpha (1-u)^{\beta-1} \mathrm{d}u = z^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \end{aligned}$$

Plugging $\int_{x=0}^{z} x^{\alpha-1}(z-x)^{\beta-1}\mathrm{d}x = z^{\alpha+\beta-1}\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ back to $f_Z(z)$, we get

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\mathrm{d}x = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-\lambda z}, \quad z > 0,$$

which is exactly the PDF for Gamma$(\alpha+\beta, \lambda)$.

Example 3: Sum of Two Independent Cauchy R.V.'s

Suppose X and Y are indep. Cauchy with the PDF

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty \leq x < \infty.$$

What the distribution of $T = X + Y$?

Example 3: Sum of Two Independent Cauchy R.V.'s

Suppose X and Y are indep. Cauchy with the PDF

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty \leq x < \infty.$$

What the distribution of $T = X + Y$?

Ans. One could find the PDF of $T = X + Y$ by integrating

$$f_T(t) = \int_{-\infty}^{\infty} f(x) f(t-x) \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\pi^2(1+x^2)(1+(t-x)^2)} \mathrm{d}x$$
$$= \frac{2}{\pi(4+t^2)}, \quad -\infty < t < \infty.$$

The calculation is shown in the next 4 pages.

This implies that $Z = (X+Y)/2 = T/2$ has identical distribution as X and Y.

$$f_Z(z) = 2 f_T(2z) = \frac{1}{\pi(1+z^2)}, \quad -\infty < z < \infty.$$

The first step is to find constants A, B, C, and D that satisfy

$$\frac{1}{(1+x^2)(1+(t-x)^2)} = \frac{Ax+B}{1+x^2} + \frac{Cx+D}{1+(t-x)^2},$$

where A, B, C, and D may depend on t but not on x.

Multiplying both sides by $(1+x^2)(1+(t-x)^2)$ we get

$$\begin{aligned}
1 &= (Ax+B)(1+(t-x)^2) + (Cx+D)(1+x^2) \\
&= (Ax+B)(1+t^2-2tx+x^2) + (Cx+D)(1+x^2) \\
&= (A+C)x^3 + (-2tA+B+D)x^2 + (A(1+t^2)-2tB+C)x + B(1+t^2)+D.
\end{aligned}$$

For two polynomials to be equal, their coefficients for x^3, x^2, x and 1 must match. We thus get the 4 equations

$$\begin{aligned}
0 &= A+C \\
0 &= -2tA+B+D \\
0 &= A(1+t^2)-2tB+C \\
1 &= B(1+t^2)+D
\end{aligned}$$

$$
\begin{aligned}
0 &= A + C && (1)\\
0 &= -2tA + B + D && (2)\\
0 &= A(1+t^2) - 2tB + C && (3)\\
1 &= B(1+t^2) + D && (4)
\end{aligned}
$$

From (1), we know $C = -A$.
Plugging in $C = -A$ into (3), we get

$$0 = At^2 - 2Bt = t(At - 2B) \;\Rightarrow\; At = 2B$$

Plugging in $At = 2B$ into (2), we get

$$0 = -4B + B + D = -3B + D \;\Rightarrow D = 3B.$$

Plugging in $D = 3B$ into (4), we get $1 = B(1+t^2) + 3B$, and thus

$$B = \frac{1}{4+t^2}, \quad D = 3B = \frac{3}{4+t^2}, \quad A = \frac{2B}{t} = \frac{2}{t(4+t^2)} = -C.$$

Putting everything together, we have

$$\begin{aligned}
\frac{1}{(1+x^2)(1+(t-x)^2)} &= \frac{Ax+B}{1+x^2} + \frac{Cx+D}{1+(t-x)^2} \\
&= \frac{2x+t}{t(4+t^2)(1+x^2)} + \frac{3t-2x}{t(4+t^2)(1+(t-x)^2)} \\
&= \frac{2x+t}{t(4+t^2)(1+x^2)} + \frac{t+2(t-x)}{t(4+t^2)(1+(t-x)^2)} \\
&= \frac{1}{(4+t^2)}\left(\frac{1}{1+x^2} + \frac{1}{1+(t-x)^2}\right) + \frac{1}{t(4+t^2)}\left(\frac{2x}{1+x^2} + \frac{2(t-x)}{1+(t-x)^2}\right).
\end{aligned}$$

The PDF for $T = X + Y$ is thus

$$f_T(t) = \int_{-\infty}^{\infty} \frac{1}{\pi^2(1+x^2)(1+(t-x)^2)} \mathrm{d}x = I + II,$$

(continued next page)

where

$$
\begin{aligned}
I &= \frac{1}{\pi^2(4+t^2)} \int_{-\infty}^{\infty} \frac{1}{1+x^2} + \frac{1}{1+(t-x)^2} \mathrm{d}x \\
&= \frac{1}{\pi^2(4+t^2)} \Big[\arctan(x) + \arctan(x-t) \Big]_{x=-\infty}^{x=\infty} \\
&= \frac{1}{\pi^2(4+t^2)} \left(\frac{\pi}{2} + \frac{\pi}{2} - (-\frac{\pi}{2}) - (-\frac{\pi}{2}) \right) = \frac{2}{\pi(4+t^2)} \\
II &= \frac{1}{\pi^2 t(4+t^2)} \int_{-\infty}^{\infty} \frac{2x}{1+x^2} + \frac{2(t-x)}{1+(t-x)^2} \mathrm{d}x \\
&= \frac{1}{\pi^2 t(4+t^2)} \left[\log(1+x^2) - \log(1+(t-x)^2) \right]_{x=-\infty}^{x=\infty} \\
&= \frac{1}{\pi^2 t(4+t^2)} \log \left[\frac{1+x^2}{1+(t-x)^2} \right]_{x=-\infty}^{x=\infty} = 0
\end{aligned}
$$

Thus

$$
f_T(t) = I + II = \frac{2}{\pi(4+t^2)} \quad \text{for } -\infty < t < \infty.
$$

Summary: Sum of Two Independent R.V.'s

Suppose all X and Y below are independent.

- If $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$, then $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$
- If $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$, then $X + Y \sim \text{Bin}(m + n, p)$
- If X and Y are both $\sim \text{Geometric}(p)$, then $X + Y \sim \text{NegBin}(2, p)$
- If $X \sim \text{NegBin}(m, p)$ and $Y \sim \text{NegBin}(n, p)$, then $X + Y \sim \text{NegBin}(m + n, p)$
- If $X \sim \text{EXP}(\lambda)$ and $Y \sim \text{EXP}(\lambda)$, then $X + Y \sim \text{Gamma}(2, \lambda)$
- If $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$, then $X + Y \sim \text{Gamma}(\alpha + \beta, \lambda)$
- If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- If X and Y are both Cauchy, then $(X + Y)/2$ is also Cauchy.

Bivariate Transformation

Suppose X and Y are continuous r.v. with joint PDF $f_{XY}(x,y)$, They are mapped onto U and V by a 1-to-1 transformation

$$\begin{cases} u = g_1(x,y) \\ v = g_2(x,y) \end{cases} \xrightarrow{\text{inverse transform}} \begin{cases} x = h_1(u,v) \\ y = h_2(u,v). \end{cases}$$

The joint PDF $f_{UV}(u,v)$ is given by

$$f_{UV}(u,v) = f_{XY}(h_1(u,v), h_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|,$$

where $\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$ is absolute value of the *Jacobian of the transformation*, defined as

$$\left| \frac{\partial(x,y)}{\partial(u,v)} \right| = \left| \begin{matrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{matrix} \right| = \left| \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \right|.$$

To memorize the formula, keep in mind that

$$f_{UV}(u,v)\mathrm{d}u\mathrm{d}v = f_{XY}(x,y)\mathrm{d}x\mathrm{d}y,$$

so informally

$$f_{UV}(u,v)\mathrm{d}u\mathrm{d}v = f_{XY}(x,y)\underbrace{\left|\frac{\partial(x,y)}{\partial(u,v)}\right|}_{\displaystyle\downarrow\atop\displaystyle\frac{\mathrm{d}x\mathrm{d}y}{\mathrm{d}u\mathrm{d}v}}\mathrm{d}u\mathrm{d}v$$

Example 4 — Gamma Again

Suppose $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$ are independent. Find the joint and marginal PDF's for

$$U = X + Y \quad \text{and} \quad V = \frac{X}{X + Y}.$$

Example 4 — Gamma Again

Suppose $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$ are independent. Find the joint and marginal PDF's for

$$U = X + Y \quad \text{and} \quad V = \frac{X}{X+Y}.$$

The inverse transformation is

$$\begin{aligned} X &= UV \\ Y &= U - X = U - UV = U(1-V) \end{aligned}$$

The Jacobian is

$$\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = |-uv - u(1-v)| = u$$

The joint PDF for (X, Y) (from Example 2) is

$$f_{XY}(x, y) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)}, \quad x > 0, y > 0.$$

The joint PDF for (U, V) is

$$\begin{aligned} f_{UV}(u, v) &= f_{XY}(uv, u(1-v)) \cdot u \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} (uv)^{\alpha-1} (u(1-v))^{\beta-1} e^{-\lambda u} u \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha+\beta-1} v^{\alpha-1} (1-v)^{\beta-1} e^{-\lambda u} \\ &= \underbrace{\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} u^{\alpha+\beta-1} e^{-\lambda u}}_{\text{PDF for Gamma}(\alpha+\beta,\lambda)} \cdot \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} v^{\alpha-1} (1-v)^{\beta-1}}_{\text{PDF for BETA}(\alpha,\beta)} \end{aligned}$$

The joint PDF for (X, Y) (from Example 2) is

$$f_{XY}(x,y) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)}, \quad x > 0, y > 0.$$

The joint PDF for (U, V) is

$$\begin{aligned}
f_{UV}(u,v) &= f_{XY}(uv, u(1-v)) \cdot u \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} (uv)^{\alpha-1} (u(1-v))^{\beta-1} e^{-\lambda u} u \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha+\beta-1} v^{\alpha-1} (1-v)^{\beta-1} e^{-\lambda u} \\
&= \underbrace{\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} u^{\alpha+\beta-1} e^{-\lambda u}}_{\text{PDF for Gamma}(\alpha+\beta,\lambda)} \cdot \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} v^{\alpha-1} (1-v)^{\beta-1}}_{\text{PDF for BETA}(\alpha,\beta)}
\end{aligned}$$

This shows

- $U = X + Y \sim \text{Gamma}(\alpha+\beta, \lambda)$, $V = \frac{X}{X+Y} \sim \text{BETA}(\alpha, \beta)$
- U and V are independent

Example 5 — Normal

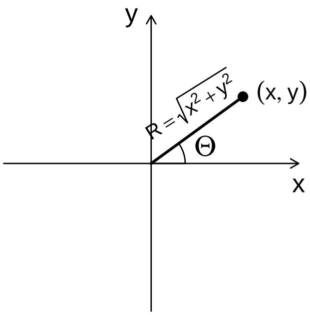
Suppose $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent.
Find the joint and marginal PDF's for

$$R = \sqrt{X^2 + Y^2} \quad \text{and}$$
$$\Theta = \tan^{-1}(Y/X)$$

so that $-\pi < \Theta \leq \pi$.

Example 5 — Normal

Suppose $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent.
Find the joint and marginal PDF's for



$$R = \sqrt{X^2 + Y^2} \quad \text{and}$$
$$\Theta = \tan^{-1}(Y/X)$$

so that $-\pi < \Theta \leq \pi$.
The inverse transformation is

$$X = R\cos\Theta, \quad Y = R\sin\Theta$$

The Jacobian is

$$\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

The joint PDF for (X, Y) is

$$f_{XY}(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \quad -\infty < x, y < \infty.$$

The joint PDF for (R, Θ) is

$$\begin{aligned} f_{R\Theta}(r, \theta) &= f_{XY}(r\cos\theta, r\sin\theta) \cdot r \\ &= \underbrace{\frac{1}{2\pi}}_{\text{PDF of } \Theta} \cdot \underbrace{re^{-r^2/2}}_{\text{PDF of } R}, \qquad \begin{matrix} -\pi < \Theta \leq \pi \\ 0 \leq r < \infty \end{matrix} \end{aligned}$$

This shows

- Θ is Uniform on $(-\pi, \pi)$
- $R = \sqrt{X^2 + Y^2}$ has the PDF $f_R(r) = re^{-r^2/2}$ for $r \geq 0$
- Θ and R are independent

Example 6 — Quotient of Two Standard Normal

Suppose $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent.

a. Find the joint PDF for $U = X/Y$ and $V = Y$.
b. Find the marginal PDF for $U = X/Y$.

Example 6 — Quotient of Two Standard Normal

Suppose $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent.

a. Find the joint PDF for $U = X/Y$ and $V = Y$.
b. Find the marginal PDF for $U = X/Y$.

The inverse transformation is

$$X = UV, \quad Y = V$$

The Jacobian is

$$\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = |v|.$$

As the joint PDF for (X,Y) is $f_{XY}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$ for $-\infty < x, y < \infty$, the joint PDF for (U,V) is

$$f_{UV}(u,v) = f_{XY}(uv, v) \cdot |v| = \frac{1}{2\pi}|v| e^{-v^2(1+u^2)/2}, \; -\infty < u, v < \infty.$$

We can obtain the marginal PDF of U by integrating the joint PDF over v.

$$\begin{aligned}
f_U(u) &= \int_{-\infty}^{\infty} f_{UV}(u,v)\mathrm{d}v \\
&= \int_{-\infty}^{0} f_{UV}(u,v)\mathrm{d}v + \int_{0}^{\infty} f_{UV}(u,v)\mathrm{d}v \\
&= 2\int_{0}^{\infty} f_{UV}(u,v)\mathrm{d}v \quad (\text{since } f_{UV}(u,v) = f_{UV}(u,-v)) \\
&= \frac{1}{\pi}\int_{0}^{\infty} ve^{-v^2(1+u^2)/2}\mathrm{d}v \\
&= \frac{1}{\pi(1+u^2)}\int_{0}^{\infty} ze^{-z^2/2}\mathrm{d}z \quad \left(\text{letting } v=\frac{z}{\sqrt{1+u^2}} \Rightarrow \mathrm{d}v=\frac{\mathrm{d}z}{\sqrt{1+u^2}}\right) \\
&= \frac{1}{\pi(1+u^2)}, \quad -\infty < u < \infty.
\end{aligned}$$

Observe that $U = X/Y$ has the **Cauchy distribution** in L04.

Order Statistics

i.i.d. Random Sample

Suppose $X_1, \ldots, X_n$ are *independent and identically distributed ("i.i.d.")*, from a distribution with CDF F

- Independence $\Rightarrow F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^n F(x_i)$
- If X_i's are discrete, then the joint PMF is a product of the individual PMF

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n).$$

- If X_i's are continuous, then the joint PDF is a product of the PDF $f()$ for an individual X_i:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n).$$

Order statistics

The order statistics of a random sample $X_1, \dots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, \dots, X_{(n)}$ and they satisfy

$$X_{(1)} \leq \dots, \leq X_{(n)}.$$

In other words,

$$\begin{aligned} X_{(1)} &= \min_{1 \leq i \leq n} X_i, \\ X_{(2)} &= \text{second smallest } X_i, \\ &\vdots \\ X_{(k)} &= \text{kth smallest } X_i, \\ &\vdots \\ X_{(n)} &= \max_{1 \leq i \leq n} X_i \end{aligned}$$

Note: if there are ties, the same value appears multiple times.
e.g., if $(X_1, X_2, X_3) = (3, 5, 3)$, then $X_{(1)} = X_{(2)} = 3$ and $X_{(3)} = 5$.

Why Study Order Statistics?

- Extreme observations can be rare but catastrophic.
  Good to know their behaviors
- Sample **median** is less sensitive to outliers than the sample mean
  - If $n = 2m + 1$, then $X_{(m+1)}$ is the median
- **Quartiles** and **Percentiles** are also order statistics

Distribution of $X_{(1)}$ = Minimum

Suppose $X_1, \dots, X_n$ are i.i.d. observations from a distribution with CDF F.

What is the distribution of $X_{(1)}$?

Distribution of $X_{(1)}$ = Minimum

Suppose $X_1, \dots, X_n$ are i.i.d. observations from a distribution with CDF F.

What is the distribution of $X_{(1)}$?

$$\begin{aligned} F_{X_{(1)}}(x) &= \mathrm{P}(X_{(1)} \le x) = 1 - \mathrm{P}(X_{(1)} > x) \\ &= 1 - \mathrm{P}(X_i > x \text{ for all } i = 1, \dots, n) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

Distribution of $X_{(1)}$ = Minimum

Suppose $X_1, \dots, X_n$ are i.i.d. observations from a distribution with CDF F.

What is the distribution of $X_{(1)}$?

$$\begin{aligned} F_{X_{(1)}}(x) &= \mathrm{P}(X_{(1)} \le x) = 1 - \mathrm{P}(X_{(1)} > x) \\ &= 1 - \mathrm{P}(X_i > x \text{ for all } i = 1, \dots, n) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

If the original distribution is continuous with density $f = F'$:

$$\begin{aligned} f_{X_{(1)}}(x) = \frac{\mathsf{d}}{\mathrm{d}x} F_{X_{(1)}}(x) &= n(1 - F(x))^{n-1} \cdot \frac{\mathsf{d}}{\mathrm{d}x} F(x) \\ &= n(1 - F(x))^{n-1} \cdot f(x). \end{aligned}$$

Distribution of $X_{(m)}$ = Maximum

Suppose $X_1, \ldots, X_n$ are i.i.d. observations from a distribution with CDF F.

The CDF for $X_{(n)}$ is

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathrm{P}(X_{(n)} \leq x) \\ &= \mathrm{P}(X_i \leq x \text{ for all } i = 1, \ldots, n) \\ &= (F(x))^n \end{aligned}$$

Distribution of $X_{(m)}$ = Maximum

Suppose $X_1, \dots, X_n$ are i.i.d. observations from a distribution with CDF F.

The CDF for $X_{(n)}$ is

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathrm{P}(X_{(n)} \leq x) \\ &= \mathrm{P}(X_i \leq x \text{ for all } i = 1, \dots, n) \\ &= (F(x))^n \end{aligned}$$

If the original distribution is continuous with density $f = F'$:

$$f_{X_{(n)}}(x) = \frac{\mathsf{d}}{\mathrm{d}x} F_{X_{(n)}}(x) = n(F(x))^{n-1} \cdot \frac{\mathsf{d}}{\mathrm{d}x} F(x) = n(F(x))^{n-1} \cdot f(x).$$

Example — Order Statistics for Exponential

Suppose $X_1, \dots, X_n$ are i.i.d. Exponential(λ).

The PDF for $X_{(n)}$ is

$$f_{X_{(n)}}(x) = nF(x)^{n-1} \cdot f(x) = n(1 - e^{-\lambda x})^{n-1} \cdot \lambda e^{-\lambda x}, \quad 0 \le x \le \infty.$$

Example — Order Statistics for Exponential

Suppose $X_1, \ldots, X_n$ are i.i.d. Exponential(λ).

The PDF for $X_{(n)}$ is

$$f_{X_{(n)}}(x) = nF(x)^{n-1} \cdot f(x) = n(1 - e^{-\lambda x})^{n-1} \cdot \lambda e^{-\lambda x}, \quad 0 \leq x \leq \infty.$$

The PDF for $X_{(1)}$ is

$$\begin{aligned} f_{X_{(1)}}(x) &= n(1 - F(x))^{n-1} \cdot f(x) \\ &= n\big(1 - (1 - e^{-\lambda x})\big)^{n-1} \cdot \lambda e^{-\lambda x} \\ &= (n\lambda) e^{-(n\lambda)x}, \quad 0 \leq x \leq \infty. \end{aligned}$$

Observe that $X_{(1)} \sim$ Exponential($n\lambda$)

Joint Distribution of $X_{(1)}$ and $X_{(n)}$

Suppose $X_1, \ldots, X_n$ are i.i.d. observations from a distribution with CDF F

The joint CDF of $X_{(1)}$ and $X_{(n)}$ is

$$\begin{aligned} F_{X_{(1)},X_{(n)}}(x,y) &= \mathrm{P}(X_{(1)} \le x, X_{(n)} \le y) \\ &= \mathrm{P}(X_{(n)} \le y) - \mathrm{P}(X_{(1)} > x, X_{(n)} \le y) \\ &= \mathrm{P}(X_{(n)} \le y) - \mathrm{P}(x < X_i \le y \text{ for all } i = 1, \ldots, n) \\ &= F(y)^n - (F(y) - F(x))^n \end{aligned}$$

If continuous, we can differentiate the joint CDF to obtain the joint PDF.

$$\begin{aligned} f_{X_{(1)},X_{(n)}}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{X_{(1)},X_{(n)}}(x,y) \\ &= n(n-1) f(x) f(y) (F(y) - F(x))^{n-2}, \quad x < y. \end{aligned}$$

Example: Order Statistics for Uniform(0,1)

If $X_1, \ldots, X_n$ are i.i.d. Uniform(0,1),

$$f(x) = 1, \quad F(x) = x, \quad 0 \le x \le 1.$$

The joint PDF for $(X_{(1)}, X_{(n)})$ is

$$f_{X_{(1)},X_{(n)}}(x, y) = n(n-1)(y-x)^{n-2}, \quad 0 \le x \le y \le 1.$$

PDF for $X_{(k)}$

Suppose $X_1, \ldots, X_n$ are i.i.d. observations from a continuous distribution with CDF F and PDF f. The density of $X_{(k)}$, the kth-order statistic, is

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}[1 - F(x)]^{n-k} f(x).$$

Heuristic Proof. $\mathrm{P}(x \le X_{(k)} \le x + \mathrm{d}x)$ is the probability that

- $k-1$ observations are $\le x$, each occurs w/ prob. $F(x)$
- 1 observation is in $[x, x + \mathrm{d}x]$, which occurs w/ prob. $f(x)\mathrm{d}x$
- $n-k$ observations are $\ge x + \mathrm{d}x$, each occurs w/ prob. $1 - F(x + \mathrm{d}x) \approx 1 - F(x)$

There are $\frac{n!}{(k-1)!1!(n-k)!}$ such arrangements,
each occur with prob. $(F(x))^{k-1}[1 - F(x)]^{n-k} f(x)\mathrm{d}x$.

Example: Order Statistics for Uniform(0,1)

If $X_1, \dots, X_n$ are i.i.d. Uniform(0,1),

$$f(x) = 1, \quad F(x) = x, \quad 0 \le x \le 1.$$

The PDF for $X_{(k)}$ is

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 \le x \le 1,$$

which is the PDF for BETA$(\alpha = k, \beta = n - k + 1)$.