Percentiles and Mode

20M

The official reading defines the $100p^{\text{th}}$ ***percentile*** of a random variable X as the value π_p that satisfies **both** of these inequalities:

- $\Pr(X < \pi_p) \le p$
- $\Pr(X \le \pi_p) \ge p$

The definition above applies to both discrete and continuous distributions. This definition can be simplified if X is **continuous**. Recall that for a continuous random variable X, it is always true that $\Pr(X < \pi_p) = \Pr(X \le \pi_p)$. Therefore, the two inequalities can be restated as:

$$p \le \Pr(X \le \pi_p) \le p$$

This can be further restated as:

$$F(\pi_p) = \Pr(X \le \pi_p) = p$$

In other words, the $100p^{\text{th}}$ percentile of a continuous random variable X is the value π_p that satisfies $F(\pi_p) = p$.

There are special names for certain percentiles:

- The 25^{th} percentile is the 1^{st} quartile
- The 50^{th} percentile is the ***median***, or the 2^{nd} quartile
- The 75^{th} percentile is the 3^{rd} quartile

CONTINUOUS EXAMPLE

Consider the following example:

A random variable X has the following probability distribution:

$$f(x) = 0.5x, \qquad 0 < x < 2$$

Determine the median and the 3rd quartile of X.

The median is the 50th percentile of X, which is $\pi_{0.5}$. The 3rd quartile is its 75th percentile, which is $\pi_{0.75}$.

Determine the CDF of X to compute its percentiles.

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0.25x^2, & 0 < x < 2 \\ 1, & x \geq 2 \end{cases}$$

Since X is continuous, $\pi_{0.5}$ is the value of X such that $F(\pi_{0.5}) = 0.5$. Therefore,

$$F(\pi_{0.5}) = 0.5$$

$$0.25(\pi_{0.5})^2 = 0.5$$

$$\pi_{0.5} = \boldsymbol{\sqrt{2}}$$

Similarly, the CDF evaluated at $x = \pi_{0.75}$ must produce a probability of 0.75. Therefore,

$$F(\pi_{0.75}) = 0.75$$

$$0.25(\pi_{0.75})^2 = 0.75$$

$$\pi_{0.75} = \sqrt{\mathbf{3}}$$

DISCRETE EXAMPLE

Consider another example:

A random variable X has the following probability distribution:

$$p(x) = \begin{cases} 0.40, & x = 1 \\ 0.20, & x = 2 \\ 0.15, & x = 5 \\ 0.25, & x = 8 \\ 0, & \text{otherwise} \end{cases}$$

Determine the median and the 3rd quartile of X.

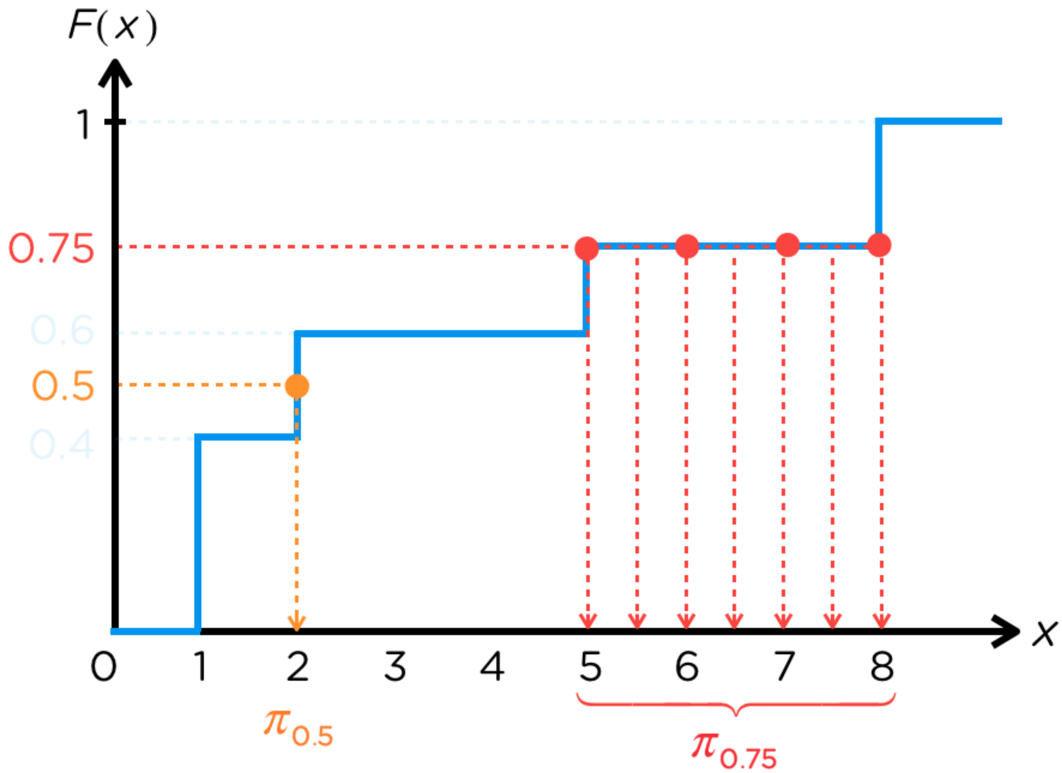
Since X is discrete in this example, the shortcut formula $F(\pi_p) = p$ cannot be used. The full definition of a percentile must be used instead, which can be portrayed visually.

Start by sketching a graph of the CDF. This is represented by the blue line in the figure below. To determine the $100p^{\text{th}}$ percentile, draw a horizontal line $F(x) = p$.

- In most cases, this line will intersect the CDF at one point. The x-coordinate of the point of intersection is the $100p^{\text{th}}$ percentile.

- In the case of multiple points of intersection, the x-coordinates of all points of intersection will count as the percentile, i.e. there will be multiple points corresponding to the same percentile.

The following graph illustrates $\pi_{0.5}$ and $\pi_{0.75}$.



The line $F(x) = 0.5$ intersects the CDF at $x = 2$. Therefore, $\pi_{0.5} = \mathbf{2}$.

We can verify that our answer is correct by checking if the percentile satisfies the two inequalities.

- $\Pr(X < 2) = \Pr(X \le 1) = 0.4 \le 0.5$
- $\Pr(X \le 2) = 0.6 \ge 0.5$

This confirms the median of X is 2.

Notice that for any p where $0.4 \le p \le 0.6$, the line $F(x) = p$ will intersect the CDF at $x = 2$. Therefore, 2 is not only the 50th percentile; it is any percentile of X between the 40th and the 60th.

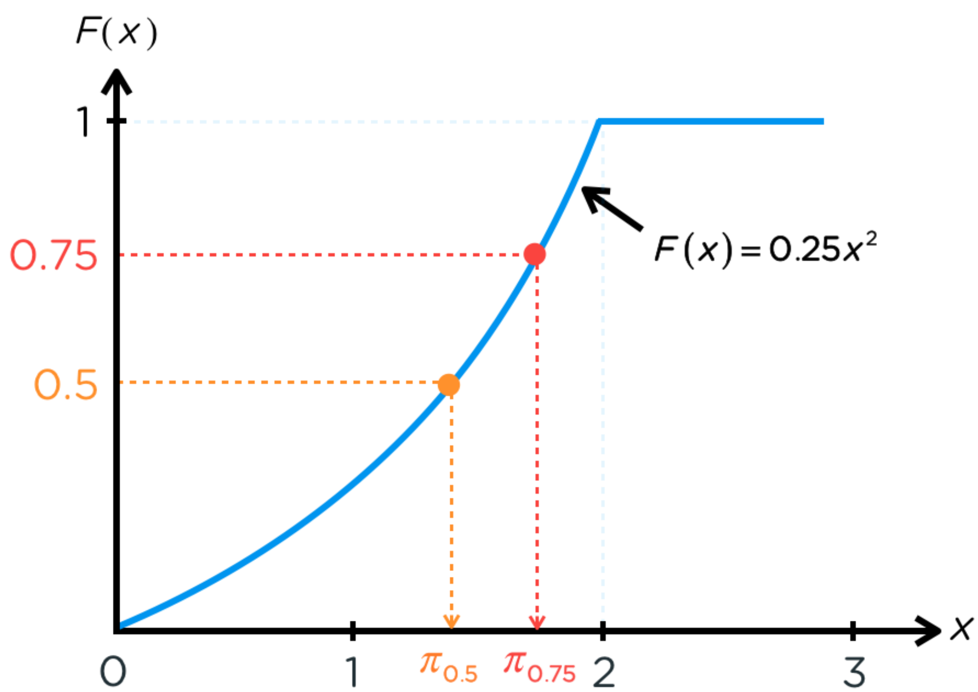
The line $F(x) = 0.75$ does not intersect the CDF at just one point. It intersects the CDF at every point from $x = 5$ to $x = 8$. That means every value in the interval from 5 to 8 (including 5 and 8) is the 75th percentile. This can be confirmed using the percentile definition, as both inequalities will be satisfied for **any** $\pi_{0.75}$ where $5 \leq \pi_{0.75} \leq 8$.

In short, determining the percentiles of a discrete distribution can be tricky.

- Multiple percentiles can have the same value.
- Multiple values can correspond to the same percentile.

Based on the examples explained above, several conclusions can be made:

1. If X is **continuous**, it is easy to compute its percentiles. This is because the types of continuous random variables that we care about in this exam will have a CDF that is strictly increasing (except when it equals 0 or 1). This can be seen in the CDF plot below, based on the continuous example above.



For $0 < F(x) < 1$, the CDF is a one-to-one function of x. This implies it has an inverse function which produces unique percentiles of X.

$$\pi_p = F^{-1}(p), \text{ for } 0 < p < 1$$

2. If X is **discrete**, its CDF plot will be a step graph, as shown in the discrete example above. Since the CDF is no longer a one-to-one function, the percentiles of X have to be calculated using the full definition of percentile stated at the beginning of this section.
3. For both discrete and continuous distributions, use the **CDF** to calculate percentiles.

Example S2.1.3.1

Suppose X is a random variable that has the following density function:

$$f(x) = \frac{1}{5}\, e^{-x/5}, \qquad x > 0$$

Determine the 60th percentile of X.

Solution

Since X is continuous, solve for $\pi_{0.6}$ such that $F(\pi_{0.6}) = 0.6$.

$$\begin{aligned} F(\pi_{0.6}) &= \int_0^{\pi_{0.6}} \frac{1}{5}\, e^{-x/5}\, \mathrm{d}x \\ &= \left[-e^{-x/5}\right]_0^{\pi_{0.6}} \\ &= 1 - e^{-\pi_{0.6}/5} \end{aligned}$$

$$0.6 = 1 - e^{-\pi_{0.6}/5}$$

$$e^{-\pi_{0.6}/5} = 0.4$$

$$\begin{aligned} \pi_{0.6} &= -5\ln(0.4) \\ &= \mathbf{4.5815} \end{aligned}$$

Mode

The *mode* is the value of the random variable that maximizes the PMF or the PDF. In other words, the mode is the critical point of the **global maximum** of the PMF/PDF.

For **continuous** random variables, the mode is easy to calculate. Maximizing the PDF means calculating the first derivative of $f(x)$ and equating it to 0. Then, solve for x, which is the mode.

$$f'(x) = 0 \quad \Rightarrow \quad x = \text{mode}$$

Coach's Remarks

Technically, the solutions of $f'(x) = 0$ are the critical points of local maxima/minima of the probability function. The second derivative test can be used to verify a critical point corresponds to a local maximum; this is shown below in Example S2.1.3.2.

While not always the case, most distributions have one local maximum which coincides with the global maximum. Thus, for most exam questions, the simple approach given above will be sufficient for continuous distributions.

However, it is possible for the global maximum to exceed the local maxima found by this approach, such as when the mode is an endpoint of the domain. If you wish to be thorough, you may evaluate the density function at the endpoints and compare them to the local maxima.

While most distributions only have one mode, it is possible for a random variable to have multiple modes. This occurs when more than one value of the random variable produces the maximum PMF/PDF.

In summary, for both discrete and continuous distributions, calculate the mode(s) by maximizing the **PMF**/**PDF**.

Example S2.1.3.2

A continuous random variable X has the following PDF:

$$f(x) = 0.75\left[1 - (x-1)^2\right], \qquad 0 < x < 2$$

Determine the mode of X.

Solution

Determine the derivative of $f(x)$.

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(0.75\left[1 - (x-1)^2\right]\right) \\ &= -1.5(x-1) \end{aligned}$$

Equate the derivative to 0, and solve for x.

$$-1.5(x-1) = 0$$

$$x = \mathbf{1}$$

Thus, the mode of X is 1.

Coach's Remarks

The following step is usually not needed for the exam, but if you would like to verify that $x = 1$ corresponds to a local maximum (and not a local minimum), you can apply the second derivative test.

$$\begin{aligned} f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}[-1.5(x-1)] \\ &= -1.5 < 0 \end{aligned}$$

A negative second derivative indicates the critical point corresponds to a local maximum. This confirms the mode of X is 1.