LECTURE 10
QUANTILES (PERCENTILES), SYMMETRIC DISTRIBUTIONS, LOGISTIC DISTRIBUTION

1 Definition and properties

Let X be a continuously distributed random variable with a CDF F_X. Suppose that the CDF function is monotone increasing everywhere. In this case, the inverse function of F_X exists. If the support of the distribution is $\mathbb{R} = (-\infty, \infty)$,

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow (0,1), \text{ and} \\ F_X^{-1} : (0,1) &\rightarrow \mathbb{R}. \end{aligned}$$

The values $F_X^{-1}(\tau)$ for $\tau \in (0,1)$ are known as the *quantiles* of X (or of the distribution of X).

Definition 1. Suppose that F_X is a continuous and monotone increasing CDF, and let F_X^{-1} be its inverse function. For $\tau \in (0,1)$, the τ-th quantile is defined as

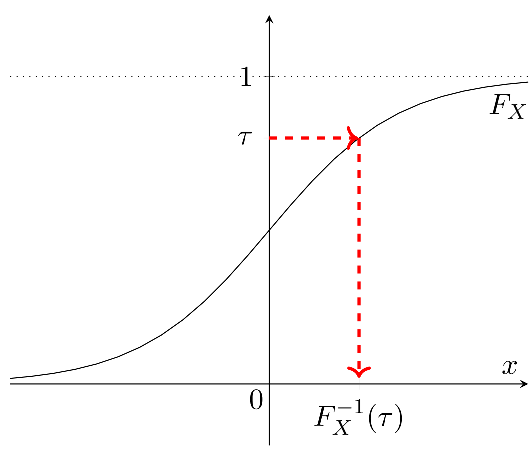
$$q_\tau = F_X^{-1}(\tau).$$

Remark. 1. Quantiles are also often referred to as *percentiles*. However, in the latter case they are indexed by *percentages* $100 \times \tau$ rather than fractions $\tau \in (0,1)$.

2. Similarly to means, variances, and other moments, quantiles are the properties of distributions. Their interpretation will be discussed below.

Construction of quantiles for strictly increasing continuous CDFs is shown graphically in Figure 1.

Figure 1: Construction of the τ-th quantile $q_\tau = F_X^{-1}(\tau)$



Recall that an inverse function must satisfy the following property:

$$F_X(F_X^{-1}(\tau)) = \tau. \tag{1}$$

Hence, the τ-th quantile must satisfy

$$\begin{aligned} P(X \leq q_\tau) &= F_X(q_\tau) \\ &= F_X(F_X^{-1}(\tau)) \\ &= \tau, \end{aligned}$$

where the equality in the first line holds by the definition of the CDF, the equality in the second line holds by the definition of the quantile q_τ, and the last equality holds by the result in (1) for monotone continuous functions. This result gives an important interpretation of quantiles: the probability of drawing a value below (to the left of) the τ-th quantile is equal to τ. We will state this result as a theorem:

Theorem 2. *Suppose that X is distributed with F_X, where F_X is a continuous and monotone increasing CDF. Let $q_\tau = F_X^{-1}(\tau)$ be the τ-th quantile of F_X. Then,*

$$F_X(q_\tau) = P(X \leq q_\tau) = \tau.$$

Several commonly used quantiles/percentiles are given their own names:

Median $\tau = 0.5$. According to Theorem 2, the probability to be below the median is 50%.

Deciles $\tau = 0.1, 0.2, \ldots, 0.9$. The value $\tau = 0.1$ corresponds to the first decile, $\tau = 0.2$ corresponds to the second decile, and etc. The probability to be below the first decile is 10%. The fifth decile coincides with the median.

Quartiles $\tau = 0.25, 0.50, 0.75$. The value $\tau = 0.25$ corresponds to the first quartile, $\tau = 0.50$ corresponds to the second quartile (or the median), and $\tau = 0.75$ corresponds to the third quartile. The *interquartile range* (IQR) is defined as the difference between the third and first quartiles:

$$IQR = q_{0.75} - q_{0.25}.$$

Along with variances (or standard deviations), IQR is often used as measures of dispersion or spread of a distribution. Note that the interval $[q_{0.25}, q_{0.75}]$ contains 50% of the probability mass:

$$\begin{aligned} P(q_{0.25} < X < q_{0.75}) &= P(X < q_{0.75}) - P(X < q_{0.25}) \\ &= 0.75 - 0.25 \\ &= 0.5. \end{aligned}$$

Thus, a longer IQR corresponds to a more disperse distribution.

Example. (Value-at-Risk) In Finance, Value-at-Risk (VaR) is a commonly used measure of risk. VaR is typically defined on the distribution of losses. Let L denote a random loss on some investment portfolio. Given $\tau \in (0,1)$ (typically τ is a very small number such as 0.05 or 0.01), VaR_τ is a number such that the probability that the loss exceeds VaR_τ is τ:

$$P(L > VaR_\tau) = \tau.$$

For example, suppose that for $\tau = 0.05$ the VaR equals $100,000. This means that the probability of losing over $100,000 is 5%, i.e. with probability 95% the loss is not going exceed $100,000. Thus, the definition can be alternatively stated as

$$P(L \leq VaR_\tau) = 1 - \tau,$$

from which it follows that VaR_τ is the $(1-\tau)$-th quantile of the distribution of losses.

Example. Consider $X \sim Uniform(0,1)$. In this case, for $x \in (0,1)$, the CDF of X is given by $F_X(x) = x$. Hence, given $\tau \in (0,1)$,

$$q_\tau = \tau.$$

More generally, suppose that $X \sim Uniform(a,b)$. In this case, for $x \in (a,b)$,

$$F_X(x) = \frac{x-a}{b-a}.$$

Using Theorem 2 to find the τ-th quantile, we can set an equation

$$\tau = \frac{q_\tau - a}{b-a}.$$

Solving the equation for q_τ, we find

$$q_\tau = \tau(b-a) + a.$$

For example, the median of the uniform distribution is given by

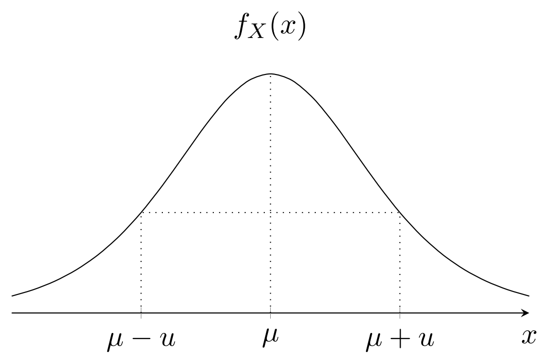
$$\begin{aligned} q_{0.5} &= 0.5(b-a)+a \\ &= \frac{a+b}{2}. \end{aligned}$$

In this example, the median coincides with the mean (the center of the interval (a, b)), and therefore, 50% of the probability mass is below the mean. Such a situation occurs, for example, when the PDF is symmetric around the mean.

2 Symmetric distributions

Many important distributions in probability and statistics have PDFs that are symmetric around some point in their support, say μ. A graph of such a PDF is shown in Figure 2.

Figure 2: Symmetric around μ PDF f_X



Definition 3. A distribution with PDF f_X is symmetric around μ if

$$f_X(\mu+u) = f_X(\mu-u)$$

for any $u \in \mathbb{R}$.

Example. (Logistic distribution) We say that X has a logistic distribution if its CDF is given by

$$F_X(x) = \frac{1}{1+e^{-x}}. \tag{2}$$

(Check that this function is indeed a CDF!) The logistic CDF corresponds to the following

PDF:

$$\begin{aligned} f_X(x) &= \frac{d}{dx}(1+e^{-x})^{-1} \\ &= \frac{e^{-x}}{(1+e^{-x})^2} \\ &= \frac{e^{-x}}{1+2e^{-x}+e^{-2x}} \\ &= \frac{e^{-x}}{1+2e^{-x}+e^{-2x}} \times \frac{e^{2x}}{e^{2x}} \\ &= \frac{e^{x}}{e^{2x}+2e^{x}+1} \\ &= \frac{e^{x}}{(1+e^{x})^2}. \end{aligned}$$

Hence,

$$f_X(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{x}}{(1+e^{x})^2} = f_X(-x)$$

for any $x \in \mathbb{R}$, and therefore this distribution is symmetric around $\mu = 0$. To find the τ-th quantile of this distribution, we will use again Theorem 2 and the definition of the CDF in (2):

$$\begin{aligned} \tau &= F_X(q_\tau) \\ &= \frac{1}{1+e^{-q_\tau}}. \end{aligned}$$

Solving the last equation for τ, we find:

$$q_\tau = -\log\left(\frac{1}{\tau} - 1\right).$$

To find the median of this distribution, substitute $\tau = 0.5$ to obtain

$$\begin{aligned} q_{0.5} &= -\log\left(\frac{1}{0.5} - 1\right) \\ &= -\log(1) \\ &= 0. \end{aligned}$$

Also, we have

$$
\begin{aligned}
q_{0.25} &= -\log\left(\frac{1}{0.25} - 1\right) \\
&= -\log(3) \\
&\approx -1.099, \\
q_{0.75} &= -\log\left(\frac{1}{0.75} - 1\right) \\
&= -\log\left(\frac{1}{3}\right) \\
&= \log(3) \\
&= -q_{0.25}.
\end{aligned}
$$

Not surprisingly, the quartiles are symmetric around zero. This property is not unique to quartiles. More generally, the quantiles of this distribution satisfy

$$
q_{1-\tau} = -q_\tau \text{ for any } \tau \in (0,1). \tag{3}
$$

To show that (3) (in the case of the logistic distribution), write

$$
\begin{aligned}
q_{1-\tau} &= -\log\left(\frac{1}{1-\tau} - 1\right) \\
&= -\log\left(\frac{\tau}{1-\tau}\right) \\
&= \log\left(\frac{1-\tau}{\tau}\right) \\
&= \log\left(\frac{1}{\tau} - 1\right) \\
&= -q_\tau.
\end{aligned}
$$

Note that for $\tau = 0.5$, equation (3) implies that $q_{0.5}$ must be zero:

$$
q_{0.5} = -q_{0.5},
$$

and therefore $q_{0.5} = 0$.

We will show below that, if a distribution is symmetric around μ, then its mean and median coincide with μ, and its quantiles are symmetric around μ.

Theorem 4. *Let X be distributed with a PDF f_X, which is symmetric around μ:*

$$
f_X(\mu + u) = f_X(\mu - u)
$$

for any $u \in \mathbb{R}$*. Then the mean* (EX)*, median* $(q_{0.5})$ *and quantiles* (q_τ) *of this distribution satisfy the following properties.*

(a) $EX = \mu$.

(b) $q_{0.5} = \mu$.

(c) $q_{1-\tau} - \mu = \mu - q_\tau$.

Proof. To show part (a), write

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_{-\infty}^{\infty} (x-\mu+\mu) f_X(x)dx \\ &= \int_{-\infty}^{\infty} (x-\mu) f_X(x)dx + \mu \int_{-\infty}^{\infty} f_X(x)dx \\ &= \int_{-\infty}^{\infty} (x-\mu) f_X(x)dx + \mu, \end{aligned} \tag{4}$$

where the last equality follows from the fact that PDF integrates to one. Thus, it suffices to show that

$$\int_{-\infty}^{\infty} (x-\mu) f_X(x)dx = 0. \tag{5}$$

For that purpose, write

$$\int_{-\infty}^{\infty} (x-\mu) f_X(x)dx = \int_{-\infty}^{\mu} (x-\mu) f_X(x)dx + \int_{\mu}^{\infty} (x-\mu) f_X(x)dx,$$

and consider the following change of variable:

$$\begin{aligned} u &= x - \mu, \text{ or} \\ x &= \mu + u. \end{aligned}$$

We have:

$$
\begin{aligned}
& \int_{-\infty}^{\mu} (x-\mu) f_X(x) du + \int_{\mu}^{\infty} (x-\mu) f_X(x) du \\
= & \int_{-\infty}^{0} u f_X(\mu+u) du + \int_{0}^{\infty} u f_X(\mu+u) du \\
= & \int_{0}^{\infty} (-u) f_X(\mu-u) dx + \int_{0}^{\infty} u f_X(\mu+u) du \\
= & \int_{0}^{\infty} u \left(f_X(\mu+u) - f_X(\mu-u) \right) du \\
= & \int_{0}^{\infty} (u \cdot 0) du \\
= & \ 0,
\end{aligned}
$$

where the equality in the fourth line is due to the symmetry of the PDF: $f_X(\mu+u)-f_X(\mu-u) = 0$. Hence, equation (5) holds and the result follows from (4).

Let F_X denote the CDF of X. To prove part (b), we need to show that

$$F_X(\mu) = 0.5. \tag{6}$$

Since the median satisfies

$$F_X(q_{0.5}) = 0.5,$$

the two equations will imply that

$$q_{0.5} = \mu.$$

To show (6), note that

$$
\begin{aligned}
F_X(\mu) &= \int_{-\infty}^{\mu} f_X(x) dx \\
&= 1 - \int_{\mu}^{\infty} f_X(x) dx,
\end{aligned} \tag{7}
$$

where the second line follows from the fact that the PDF integrates to one: $\int f_X(x)dx = 1$. Suppose that

$$\int_{-\infty}^{\mu} f_X(x) dx = \int_{\mu}^{\infty} f_X(x) dx. \tag{8}$$

Denoting

$$M = \int_{-\infty}^{\mu} f_X(x) dx,$$

from (7) and (8) we will have the following equation:

$$M = 1 - M,$$

which has a unique solution

$$M = 0.5.$$

This in turn implies

$$M = \int_{-\infty}^{\mu} f_X(x)dx = F_X(\mu) = 0.5.$$

Lastly, to show (8), we will use the same argument as in the prove of part (a). Consider again the same change of variable $u = x - \mu$ or $x = \mu + u$:

$$\begin{aligned}
& \int_{-\infty}^{\mu} f_X(x)dx - \int_{\mu}^{\infty} f_X(x)dx \\
= & \int_{-\infty}^{0} f_X(\mu + u)du - \int_{0}^{\infty} f_X(\mu + u)du \\
= & \int_{0}^{\infty} f_X(\mu - u)du - \int_{0}^{\infty} f_X(\mu + u)du \\
= & \int_{0}^{\infty} \left(f_X(\mu - u) - f_X(\mu + u)\right) du \\
= & \int_{0}^{\infty} 0 \cdot du \\
= & \; 0,
\end{aligned}$$

where the equality in the line before the last follows by the symmetry of the PDF.

To establish part (c), first note that by the definition of the τ-th quantile,

$$\begin{aligned}
\tau &= F_X(q_\tau) \\
&= \int_{-\infty}^{q_\tau} f_X(x)dx \\
&= \int_{-\infty}^{q_\tau} f_X(x - \mu + \mu)dx \\
&= \int_{-\infty}^{q_\tau - \mu} f_X(\mu + u)du,
\end{aligned}$$

where the last equality follows by the same change of variable that we used before in parts (a) and (b). Next, for the $(1-\tau)$-th quantile, we have

$$\begin{aligned}
1 - \tau &= F_X(q_{1-\tau}) \\
&= P(X \leq q_{1-\tau}),
\end{aligned}$$

or

$$\begin{aligned}\tau &= P(X > q_{1-\tau}) \\ &= \int_{q_{1-\tau}}^{\infty} f_X(x)dx \\ &= \int_{q_{1-\tau}}^{\infty} f_X(x-\mu+\mu)dx \\ &= \int_{q_{1-\tau}-\mu}^{\infty} f_X(\mu+u)du. \end{aligned} \tag{9}$$

Next, due to the symmetry $f_X(\mu+u) = f_X(\mu-u)$, we can write

$$\begin{aligned}\tau &= \int_{q_{1-\tau}-\mu}^{\infty} f_X(\mu+u)du \\ &= \int_{q_{1-\tau}-\mu}^{\infty} f_X(\mu-u)du \\ &= \int_{-\infty}^{-(q_{1-\tau}-\mu)} f_X(\mu+u)du \\ &= \int_{-\infty}^{\mu-q_{1-\tau}} f_X(\mu+u)du. \end{aligned} \tag{10}$$

Combining (9) and (10), we obtain:

$$\int_{-\infty}^{q_\tau-\mu} f_X(\mu+u)du = \tau = \int_{-\infty}^{\mu-q_{1-\tau}} f_X(\mu+u)du,$$

from which the desired result follows:

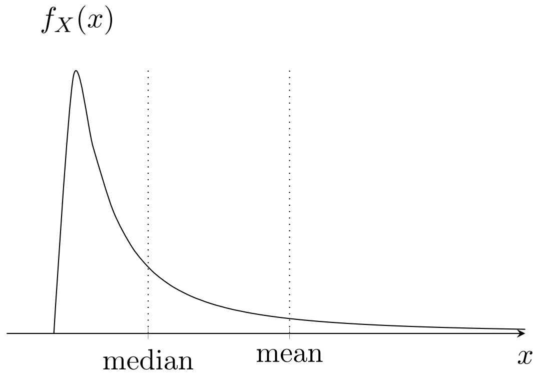
$$q_\tau - \mu = \mu - q_{1-\tau}.$$

□

3 Asymmetric distributions

While symmetric distributions are adequate for some variables, other important in economics and other disciplines variables have asymmetric distribution. Figure 3 shows the PDF of a distribution skewed to the right (with positive skewness).

Figure 3: PDF of a right-skewed distribution



In economics, for example, right-skewness characterizes the distributions of wages and incomes, house values, profits. Right-skewness means that the distribution has a very long right tail: while most of the probability mass is concentrated in the left and middle portion of the support, with a very small probability one can still draw very large values in the right tail. In the case of income distributions, right-skewness implies that a very small proportion of the population has very large incomes.

With a few exceptions, in a positive skewed distribution the mean is typically is further out in the right tail than the median. The reason for that is that very large values in the mean expression $\int x f_x(x)dx$ pull it further to the right. Since such large values occur only with a small probability, their impact on the median can be negligible. In the case of income distributions, this would imply that the majority of population have a below average income.

4 Quantiles and the uniform distribution

There is an interesting connection between quantiles and the uniform (0,1) distribution. Let X be a continuously distributed random variable with a strictly increasing CDF F_X. Consider a new random variable

$$Y = F_X(X).$$

Note that here we are applying the function F_X to a random variable. For example, if X has the logistic distribution, then $Y = (1 + e^{-X})^{-1}$. Note further that, since a CDF is bounded between zero and one, the support of Y is $(0, 1)$. To find the distribution of Y, we will use the distribution function technique from Lecture 9. For $y \in (0,1)$,

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)). \end{aligned}$$

Since $y \in (0,1)$, $F_X^{-1}(y)$ is the y-th quantile of the distribution of X:

$$F_X^{-1}(y) = q_y.$$

Therefore,

$$P(Y \leq y) = P(X \leq q_y) = y.$$

Thus, the CDF of Y is

$$F_Y(y) = P(Y \leq y) = y.$$

However, this is the CDF of the uniform $(0,1)$ distribution. We conclude that

$$F_X(X) \sim Uniform(0,1).$$

The process also works in the reverse direction. Let F_X be a CDF, and let $U \sim Uniform(0,1)$. Then, $X = F_X^{-1}(U)$ is distributed according to the CDF F_X:

$$\begin{aligned} P(X \leq x) &= P(F_X^{-1}(U) \leq x) \\ &= P(F_X(F_X^{-1}(U)) \leq F_X(x)) \\ &= P(U \leq F_X(x)) \\ &= F_X(x), \end{aligned}$$

where the last equality follows since $P(U \leq u) = u$ for any $u \in (0,1)$. The result can be used for simulating random variables with a CDF F:

1. Draw a uniformly distributed on zero-one interval random variable U.
2. Compute $X = F^{-1}(U)$. The new random variable X will be distributed according to the CDF F.