# Quantile function. Normal distribution

## Probability theory

Anton Afanasev

Higher School of Economics

DSBA 221

October 14, 2023

# Seminar Overview

1. Quantile function
   - Quantiles
   - Strictly monotonic case
   - General case
2. Normal distribution
   - Introduction
   - Standard normal distribution table
   - Practice with table
   - Linear combination of normal variables

# Quantiles

## Definition

$q$-quantiles are values that partition a range of probability distribution into $q$ intervals of equal probabilities.

- Specialized $q$-quantiles:
  - 2-quantile: median;
  - 4-quantiles: quartiles $Q_1, Q_2$ and $Q_3$,
    $Q_2 \equiv$ median,
    interquartile range: $\mathrm{IQR} = Q_3 - Q_1$;
  - 100-quantiles: percentiles $P_1, P_2, \ldots, P_{99}$.
- $x_{k/q}$ is a $k^{\text{th}}$ $q$-quantile of $X$ if:

$$\begin{cases} \mathsf{P}(X < x_{k/q}) \leq k/q, \\ \mathsf{P}(X \leq x_{k/q}) \geq k/q. \end{cases}$$

- In continuous case:

$$\mathsf{P}(X \leq x_{k/q}) = F_X(x_{k/q}) = k/q.$$

# Examples of $q$-quantiles calculation

## Example

- Find the $5^{\text{th}}$ 7-quantile $x_{5/7}$ of the following sample $X$:

$$\begin{matrix} 1 & 1 & 1 & 3 & 5 & 6 & 8 \\ 8 & 10 & 10 & 11 & 12 & 12 & 15 \end{matrix}$$

- By definition:

$$\begin{cases} \mathsf{P}(X < 10) = \dfrac{8}{14} \le \dfrac{5}{7}, \\ \mathsf{P}(X \le 10) = \dfrac{10}{14} \ge \dfrac{5}{7}, \end{cases} \quad \text{and also} \quad \begin{cases} \mathsf{P}(X < 11) = \dfrac{10}{14} \le \dfrac{5}{7}, \\ \mathsf{P}(X \le 11) = \dfrac{11}{14} \ge \dfrac{5}{7}. \end{cases}$$

- Both values 10 and 11 are fit, which means that any value from the interval $[10, 11]$ is correct.
- By default, the average 10.5 is taken.

# Examples of $q$-quantiles calculation

## Example

- Find the $5^{\text{th}}$ 7-quantile $x_{5/7}$ of the following sample $X$:

$$\begin{array}{ccccccc} 1 & 1 & 1 & 3 & 5 & 6 & 8 \\ 8 & 10 & 10 & 11 & 12 & 12 & \end{array}$$

- By definition:

$$\begin{cases} \mathsf{P}(X < 10) = \dfrac{8}{13} \leq \dfrac{5}{7}, \\ \mathsf{P}(X \leq 10) = \dfrac{10}{13} \geq \dfrac{5}{7}, \end{cases} \quad \text{but} \quad \begin{cases} \mathsf{P}(X < 11) = \dfrac{10}{13} \not\leq \dfrac{5}{7}, \\ \mathsf{P}(X \leq 11) = \dfrac{11}{13} \geq \dfrac{5}{7}. \end{cases}$$

- Without any ambiguity the correct value is 10.

# Quantile function

## Strictly monotonic c.d.f.

### Definition

$p$-quantile is a value $x_p$ from range of random variable $X$, such that:

$$\forall p \in [0,1]: \quad F_X(x_p) = p.$$

- If $F_X(x)$ is strictly monotonic, it has an inverse, which is denoted as $Q(p)$ and is called quantile function:

$$Q(p) = F_X^{-1}(x_p).$$

- Domain and image are swapped:

$$F_X(x): \mathbb{R} \to [0,1] \qquad \Longrightarrow \qquad Q(p): [0,1] \to \mathbb{R}.$$

# Problem 1

Derive a quantile function for a variable $X \sim \text{Cauchy}(0, 1)$, which has the following p.d.f. on $x \in \mathbb{R}$:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

# Problem 1

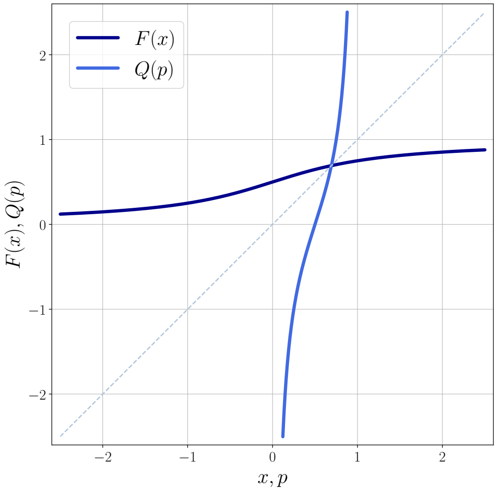


Figure: C.d.f. $F(x) = \dfrac{1}{\pi} \arctan x + \dfrac{1}{2}$ and quantile function $Q(p) = -\cot(\pi p)$.

# Cauchy distribution

- Cauchy distribution of $X$ with location of the peak $x_0$ and scale parameter $\gamma = \frac{1}{2}\text{IQR} \iff X \sim \text{Cauchy}(x_0, \gamma)$.
- P.d.f.:
$$f_X(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]}.$$
- C.d.f.:
$$F_X(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}.$$
- Mean:
$$\mathsf{E}(X) \text{ does not exist.}$$
- Variance:
$$\mathsf{V}(X) = \infty.$$

# Quantile function

## General case

- If $F_X(x)$ is NOT strictly monotonic, inverse function does not exist.
- Generally, the quantile is a set of $x$, satisfying:

$$Q(p) = [\sup\{x : F_X(x) < p\},\ \sup\{x : F_X(x) \leq p\}]\,.$$

- To eliminate ambiguity, the smallest possible value is used, which can be found as

$$Q(p) = \inf\{x \in \mathbb{R} : p \leq F_X(x)\},$$

capitalizing on the right-continuity of $F_X(x)$.
- $Q(p)$ turns out to be left-continuous.

# Problem 2

Derive and sketch a quantile function for $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

# Problem 2

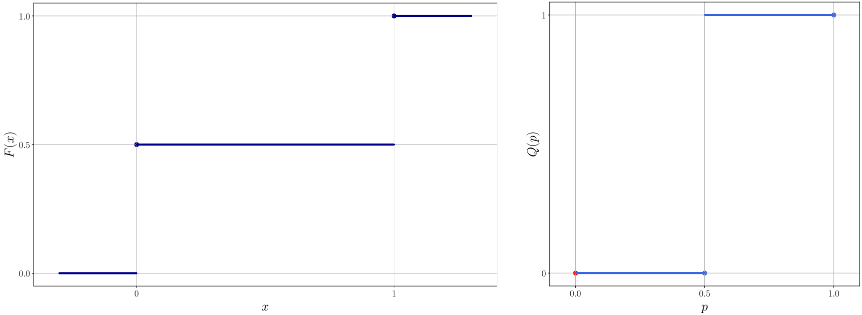


Figure: C.d.f. $F(x)$ and quantile function $Q(p)$ of $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.

# Normal (Gaussian) distribution

- Normal distribution of $X$ with mean $\mu$ and variance $\sigma^2$ $\iff$

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right).$$

- P.d.f.:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- C.d.f.:

$$F_X(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right].$$

- Mean:

$$\mathsf{E}(X) = \mu.$$

- Variance:

$$\mathsf{V}(X) = \sigma^2.$$

# Standard normal distribution

- Standard normal distribution is usually denoted as $Z$:

$$Z \sim \mathcal{N}(0,1).$$

- P.d.f.:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

- Its c.d.f. is denoted as $\Phi(z)$.
- Symmetrical property:

$$\Phi(-z) = 1 - \Phi(z).$$

# $\Phi(z)$ table

$0 \leq z < 2$

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |

# $\Phi(z)$ table

$2 \leq z < 4$

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |
| 3.5 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 |
| 3.6 | 0.9998 | 0.9998 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |
| 3.7 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |
| 3.8 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |
| 3.9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

# Problem 3

Consider normally distributed random variable $Z \sim \mathcal{N}(0, 1)$.

1. What is $\mathsf{P}(Z > 1.2)$?
2. What is $\mathsf{P}(-1.24 \leq Z \leq 1.86)$?

# Problem 3

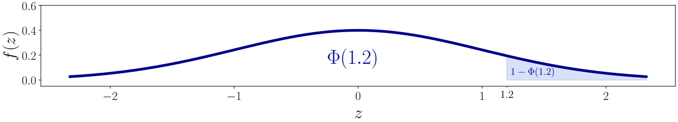


Figure: P.d.f. $f(z)$ with highlighted $\mathsf{P}(Z > 1.2)$.

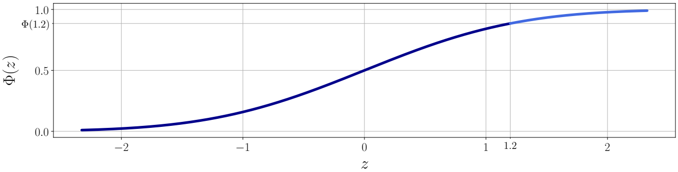


Figure: C.d.f. $\Phi(z)$ with highlighted $\mathsf{P}(Z > 1.2)$.

# Problem 3

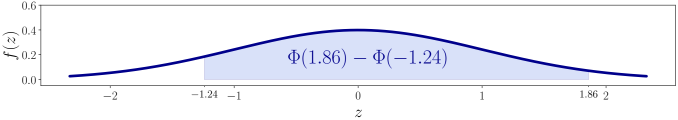


Figure: P.d.f. $f(z)$ with highlighted $\mathsf{P}(-1.24 \leq Z \leq 1.86)$.

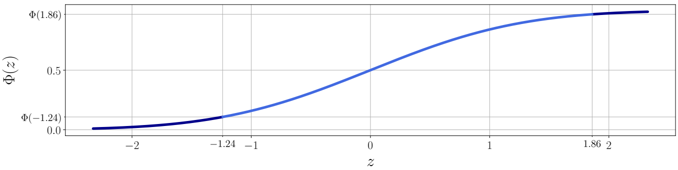


Figure: C.d.f. $\Phi(z)$ with highlighted $\mathsf{P}(-1.24 \leq Z \leq 1.86)$.

# Standardization of normal variables

- Distribution of generic normal variable $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ does not have a specific table of values $F_X(x)$.
- So it's impossible to manually calculate any generic probability $\mathsf{P}(a < X \leq b)$.
- Standardization is introduced:

$$Z = \frac{X - \mu}{\sigma}.$$

- Applying to all parts of inequality to preserve probability:

$$\mathsf{P}(a < X \leq b) = \mathsf{P}\left(\frac{a-\mu}{\sigma} < Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

- Both values can be found in $\Phi(z)$ table.

# Problem 4

Consider normally distributed random variable $X \sim \mathcal{N}(5, 4)$.

1. What is $\mathsf{P}(X > 6.4)$?
2. What is $\mathsf{P}(5.8 \leq X < 7.0)$?

# Problem 4

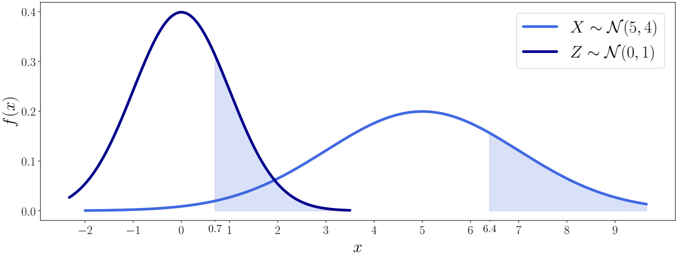


Figure: Equivalence of $\mathsf{P}(X > 6.4)$ and $\mathsf{P}(Z > 0.7)$.

# Problem 4

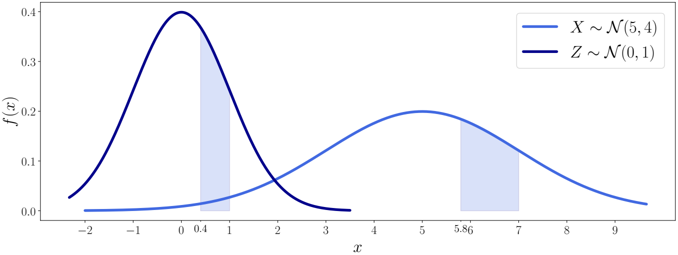


Figure: Equivalence of $\mathsf{P}(5.8 \leq X < 7.0)$ and $\mathsf{P}(0.4 \leq Z < 1.0)$.

# Problem 5

Consider normally distributed random variable $X \sim \mathcal{N}(6, 25)$.

1. Find $\mathsf{P}(6 < X < 12)$.
2. Find $\mathsf{P}(0 \leq X < 8)$.
3. Find $\mathsf{P}(-2 < X \leq 0)$.

# Problem 5

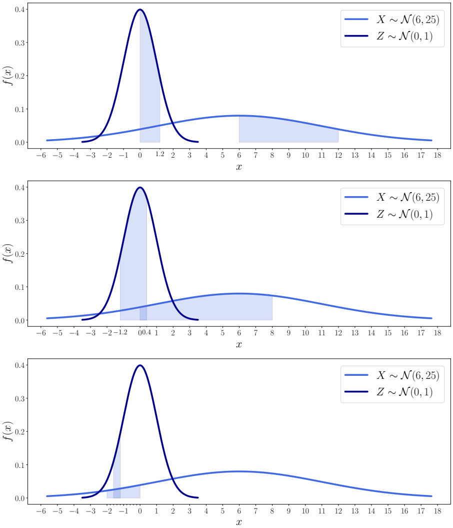


Figure: Equivalences of probabilities.

# Problem 6

The manufacturer of a brand new lithium battery claims that the mean life of a battery is 3800 hours with a standard deviation of 250 hours.

1. What percentage of batteries will last for more than 3500 hours?
2. What percentage of batteries will last for more than 4000 hours?
3. Batteries, which will last for more than $c$ hours, constitute more than $1/5$ of the population. Find the maximum possible $c$.

# Linear combination of normal variables

- There are $n$ normal random variables $X_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$.
- Let $X = \sum\limits_{i=1}^{n} a_i X_i + b$, where $a_i$ and $b$ are constant.
- Resulting $X$ has following properties:
  1. $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$,
  2. $\mu = \sum\limits_{i=1}^{n} a_i \mu_i + b$,
  3. $\sigma^2 = \sum\limits_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum\limits_{i<j}\sum a_i a_j \mathsf{Cov}(X_i, X_j)$
- If $X_i$ are pairwise independent, then

$$\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

# Problem 7

Two manufacturers of lithium batteries produce them with different specifications: the mean life of a battery from the first manufacturer is 3800 hours with a standard deviation of 250 hours, the second one's mean life is 3600 hours with standard deviation of 280 hours.

Suppose we take one battery of each kind at random and independently from each other. What is the probability that the lifetime of the first battery will be at most 50 hours greater than the second's one?

That's all Folks