Chapter 4. Continuous Random Variables and Probability Distributions

Math 3670 Summer 2024

Georgia Institute of Technology

Section 1.
Probability Density Functions

Motivation

Suppose X is the depth of a lake at a randomly chosen point on the surface.

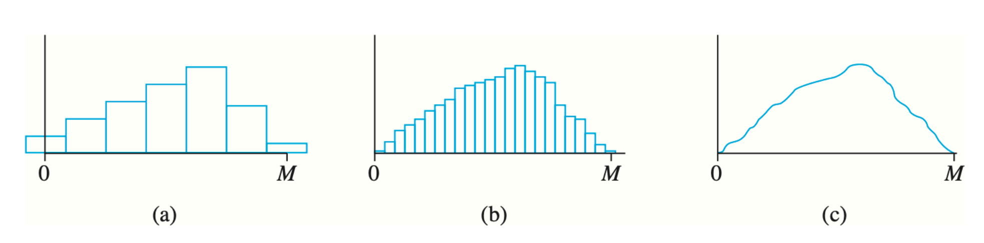
Let M be the maximum depth (in meters), so that any number in the interval $[0, M]$ is a possible value of X.

If we “discretize“ X by measuring depth to the nearest meter, then possible values are nonnegative integers less than or equal to M.

The resulting discrete distribution of depth can be pictured using a probability histogram.

Motivation

If depth is measured much more accurately and the same measurement axis, each rectangle in the resulting probability histogram is much narrower, though the total area of all rectangles is still 1.



Definition

We say a random variable X is continuous if there exists a function $f(x)$ such that

1. $f(x) \geq 0$ for all x,
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$, and
3. $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$ for all a, b.

The function $f(x)$ is called the probability density function (PDF) of X.

Example

The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is, in general, subject to uncertainty.

Consider the reference line connecting the valve stem on a tire to the center point. Let X be the angle measured clockwise to the location of an imperfection with PDF

$$f(x) = \begin{cases} \frac{1}{360}, & 0 \leq x < 360, \\ 0, & \text{otherwise.} \end{cases}$$

What is the probability that the angle is between 90° and 180°?

Definition

A continuous RV X is said to have a uniform distribution on the interval $[A, B]$ if the PDF of X is

$$f(x) =$$

We denote by $X \sim \text{Unif}(A, B)$.

Example

Let X be a continuous RV with PDF

$$f(x) = \begin{cases} 3e^{-3x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Find the probability $\mathbb{P}(X \leq 5)$.

Probability Density Functions

Properties

For a continuous RV X,

1. $\mathbb{P}(X = c) =$

2. $\mathbb{P}(a \leq X \leq b) =$

Exercise

Let X be a continuous RV with PDF

$$f(x) = \begin{cases} cx(1-x), & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

1. Find the constant $c > 0$.
2. Find the probability $\mathbb{P}(X \geq \frac{1}{3})$.

Section 2.
Cumulative Distribution Functions and Expected Values

Cumulative Distribution Functions

Definition

The cumulative distribution function $F(x)$ for a continuous RV X is defined by

$$F(x) = \mathbb{P}(X \leq x) =$$

Cumulative Distribution Functions

Example

Let $X \sim \mathsf{Unif}(A, B)$. Find the CDF.

Example

Let X be a continuous RV with PDF

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3x}{8}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

1. Find the CDF.
2. Find the probability $\mathbb{P}(1 \leq X \leq 1.5)$.

Cumulative Distribution Functions

Proposition

If X is a continuous RV with PDF $f(x)$ and CDF $F(x)$, then at every x at which the derivative $F'(x)$ exists,

$$F'(x) = f(x).$$

Percentiles of a Continuous Distribution

Definition

For $0 \leq p \leq 1$, **the $(100p)$-th percentile** of the distribution of a continuous RV X, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) =$$

In particular, the 50th percentile is called **the median** and denoted by $\tilde{\mu}$.

Example

Let X be a continuous RV with PDF

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the 25th and 50th percentiles.

Expected Values

Definition

The expected or mean value of a continuous RV X with PDF $f(x)$ is

$$\mathbb{E}[X] = \mu_X =$$

Expected Values

Example

Let X be a continuous RV with PDF

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $\mathbb{E}[X]$.

Expected Values

Proposition

If X is a continuous RV with PDF $f(x)$ and $h(X)$ is a function of X, then

$$\mathbb{E}[h(X)] =$$

Example

If you break a stick of length 1 at random into two pieces, what is the expected length of the longer piece?

Definition

The variance of a continuous random variable X with PDF $f(x)$ is

$$\mathrm{Var}(X) =$$

The standard deviation (SD) of X is

$$\sigma_X =$$

Expected Values

Proposition

$$\begin{aligned} \mathrm{Var}(X) &= \\ \mathrm{Var}(aX + b) &= \end{aligned}$$

Exercise

Let X be a continuous RV with PDF

$$f(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2 - x, & 1 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

Find the CDF and draw the graph.

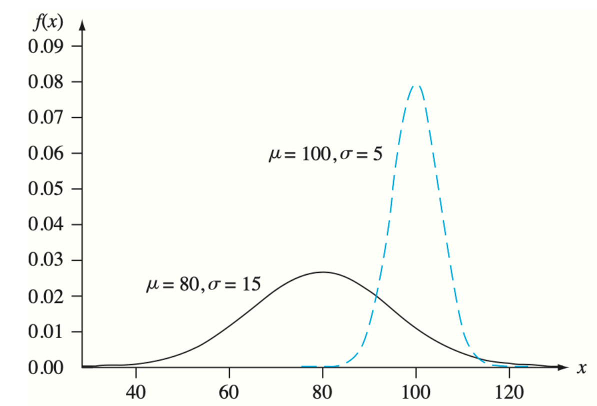
Section 3.
The Normal Distribution

Definition

A continuous RV X is said to have a normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if the PDF of X is

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

We denote by $X \sim N(\mu, \sigma^2)$. Note that $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$.


f(x)
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
μ = 100, σ = 5
μ = 80, σ = 15
40
60
80
100
120
x

Definition

The normal distribution with parameters $\mu = 0, \sigma = 1$ is called the standard normal distribution.

Usually, it is denoted by $Z \sim N(0, 1)$.

The PDF is

$$f(z; 0, 1) =$$

and the CDF is

$$\Phi(x) =$$

Example

For $Z \sim N(0, 1)$, find

1. $\mathbb{P}(Z \leq 1.25)$
2. $\mathbb{P}(Z > 1.25)$
3. $\mathbb{P}(Z \leq -1.25)$
4. $\mathbb{P}(-.38 \leq Z \leq 1.25)$

Percentiles of the Standard Normal Distribution

Example

Find the 75th percentile of the standard normal distribution.

Percentiles of the Standard Normal Distribution

Definition

z_α will denote the value on the z axis for which α of the area under the z curve lies to the right of z_α.

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1-\alpha)$th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Nonstandard Normal Distributions

Proposition

If $X \sim N(\mu, \sigma^2)$, then $aX + b$ is also normal and

$$aX + b \sim$$

In particular,

Example

The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions.

The article "Fast-Rise Brake Lamp as a Collision-Prevention Device" (Ergonomics, 1993: 391–395) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean value 1.25 sec and standard deviation of .46 sec.

What is the probability that reaction time is between 1.00 sec and 1.75 sec?

Exercise

(4.3-32) Suppose the force acting on a column that helps to support a building is a normally distributed random variable X with mean value 15.0 kips and standard deviation 1.25 kips.

Find $\mathbb{P}(X \leq 15)$ and $\mathbb{P}(14 \leq X \leq 18)$.

Example

IQ in a particular population (as measured by a standard test) is known to be approximately normally distributed with $\mu = 100$ and $\sigma = 15$.

What is the probability that a randomly selected individual has an IQ of at least 125?

The Normal Distribution and Discrete Populations

Proposition

Let $X \sim \text{Bin}(n, p)$.

If the binomial probability histogram is not too skewed, then X has approximately a normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$.

In practice, the approximation is adequate if

$$np \geq 10, \qquad n(1-p) \geq 10,$$

since there is then enough symmetry in the underlying binomial distribution.

Example

Suppose that 25% of all students at a large public university receive financial aid.

Let X be the number of students in a random sample of size 50 who receive financial aid, so that $p = .25$.

What is the probability that at most 10 students receive aid?

Exercise

(4.3-55) Suppose only 75% of all drivers in a certain state regularly wear a seat belt.

A random sample of 500 drivers is selected.

What is the probability that

1. Between 360 and 400 (inclusive) of the drivers in the sample regularly wear a seat belt?
2. Fewer than 400 of those in the sample regularly wear a seat belt?

Section 4.
The Exponential and Gamma Distributions

The Exponential Distribution

Definition

A random variable X is said to have **an exponential distribution** with parameter $\lambda > 0$ if the PDF of X is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

We denote by $X \sim \mathrm{Exp}(\lambda)$.

The Exponential Distribution

Proposition

For $X \sim \mathrm{Exp}(\lambda)$,

$$\begin{aligned} \mathbb{E}[X] &= \\ \mathrm{Var}(X) &= \\ F(X) &= \end{aligned}$$

Example

The article "Probabilistic Fatigue Evaluation of Riveted Railway Bridges" (J. of Bridge Engr., 2008: 237–244) suggested the exponential distribution with mean value 6 MPa as a model for the distribution of stress range in certain bridge connections.

Let's assume that this is in fact the true model.

Find the probability that stress range is at most 10 MPa.

Proposition

Suppose that the number of events occurring in any time interval of length t has a Poisson distribution with parameter αt.

Further assume that numbers of occurrences in nonoverlapping intervals are independent of one another.

Then the distribution of elapsed time between the occurrence of two successive events is exponential with parameter $\lambda = \alpha$.

Example

Suppose that calls are received at a 24-hour "suicide hotline" according to a Poisson process with rate a $\alpha = 5$ call per day.

Let X be the number of days X between successive calls.

What is the probability that more than 2 days elapse between calls?

The Exponential Distribution

Memoryless Property

For $X \sim \mathrm{Exp}(\lambda)$,

$$\mathbb{P}(X \geq s + t | X \geq s) =$$

The Gamma Distribution

Definition

For $\alpha > 0$, the Gamma function is defined by

$$\Gamma(\alpha) =$$

For example,

1. $\Gamma(1) =$
2. $\Gamma(2) =$
3. In general, $\Gamma(n) =$
4. $\Gamma(1/2) =$

The Gamma Distribution

Definition

A random variable X is is said to have **a Gamma distribution** with parameters $\alpha, \beta > 0$ if the PDF of X is

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

We denote by $X \sim \mathrm{Gamma}(\alpha, \beta)$.

If $\beta = 1$, X is called a standard Gamma random variable.

α is called the shape parameter.

β is called the scale parameter.

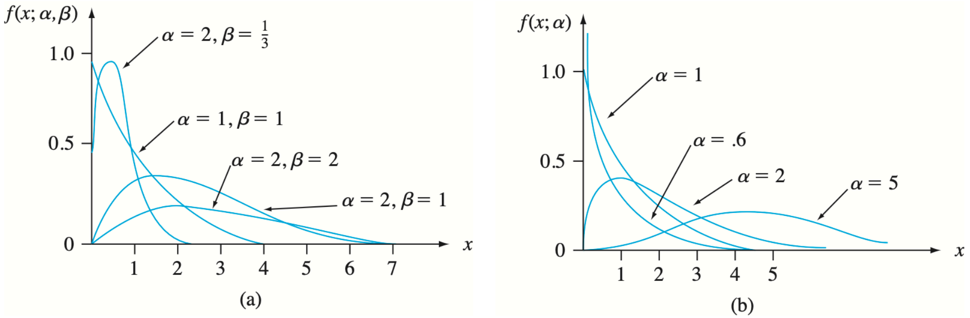


Figure 4.27 (a) Gamma density curves; (b) standard gamma density curves

Proposition

For $X \sim \mathrm{Gamma}(\alpha, \beta)$,

$$\mathbb{E}[X] = \alpha\beta$$
$$\mathrm{Var}(X) = \alpha\beta^2.$$

Definition

A random variable X is said to have **a chi-squared distribution** with parameter ν if $X \sim \mathrm{Gamma}(\nu/2, 2)$.

The parameter ν is called the number of degrees of freedom of X.

We denote by $X \sim \chi^2(\nu)$.

Exercise

(4.4-70) If $X \sim \mathrm{Exp}(\lambda)$, find the $100p$-th percentile and the median for $0 < p < 1$.

Section 6.
Probability Plots

Motivation

An investigator will often have obtained a numerical sample $x_1, x_2, \cdots, x_n$ and wish to know whether it is plausible that it came from a population distribution of some particular type (e.g., from a normal distribution).

For one thing, many formal procedures from statistical inference are based on the assumption that the population distribution is of a specified type.

The use of such a procedure is inappropriate if the actual underlying probability distribution differs greatly from the assumed type.

Motivation

Understanding the underlying distribution can sometimes give insight into the physical mechanisms involved in generating the data.

An effective way to check a distributional assumption is to construct what is called a probability plot.

The essence of such a plot is that if the distribution on which the plot is based is correct, the points in the plot should fall close to a straight line.

If the actual distribution is quite different from the one used to construct the plot, the points will likely depart substantially from a linear pattern.

Sample Percentiles

Definition

Order the n sample observations from smallest to largest.

Then the i-th smallest observation in the list is taken to be the $[100(i - 0.5)/n]$-th sample percentile.

Sample Percentiles

Example

The sample consisting of $n = 20$ observations on dielectric breakdown voltage of a piece of epoxy resin is

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.5	28.87	29.11	29.13	29.5	30.88

Find the sample percentiles.

Normal Probability Plot

A plot of the n pairs

$$([100(i - .5)/n]\text{-th z percentile}, i\text{-th smallest observation})$$

on a two-dimensional coordinate system is called a normal probability plot.

If the sample observations are in fact drawn from a normal distribution with mean value μ and standard deviation σ, the points should fall close to a straight line with slope σ and intercept μ.

Thus a plot for which the points fall close to some straight line suggests that the assumption of a normal population distribution is plausible.

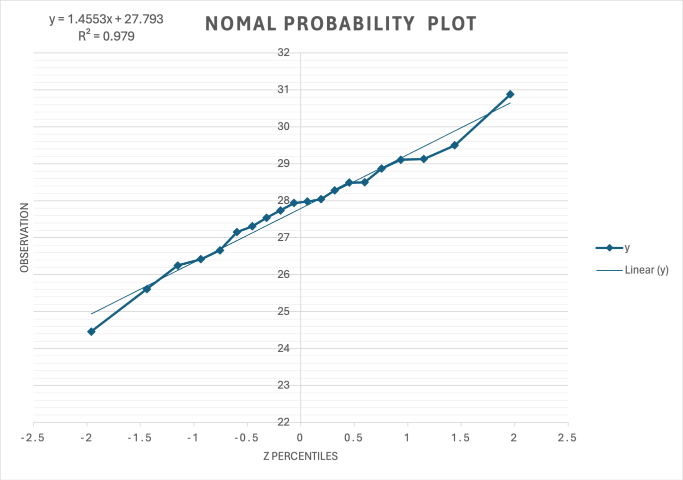
Example

The sample consisting of $n = 20$ observations on dielectric breakdown voltage of a piece of epoxy resin is

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.5	28.87	29.11	29.13	29.5	30.88

Observation	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
z percentile	−1.96	−1.44	−1.15	−.93	−.76	−.60	−.45	−.32	−.19	−.06
Observation	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
z percentile	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96

Example


y = 1.4553x + 27.793
R² = 0.979
NOMAL PROBABILITY PLOT
OBSERVATION
Z PERCENTILES
32
31
30
29
28
27
26
25
24
23
22
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
y
Linear (y)

Example

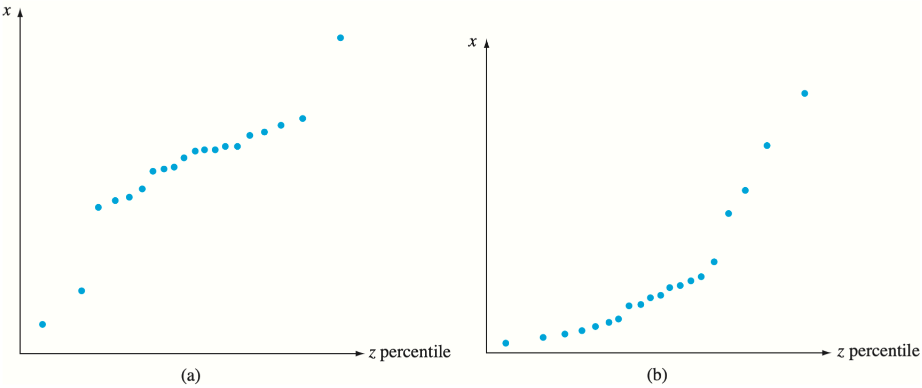


Figure 4.37 Probability plots that suggest a nonnormal distribution: (a) a plot consistent with a heavy-tailed distribution; (b) a plot consistent with a positively skewed distribution

Example