Chapter 3 Some Special Distributions

3.1 The Binomial and Related Distributions

Bernoulli Distribution

Bernoulli experiment and Bernoulli distribution

- A **Bernoulli experiment/trial** has only two possible outcomes, e.g. success/failure, heads/tails, female/male, life/death, nondefective/defective, etc.
- The outcomes are typically coded as 0 (failure) or 1 (success).
- **Definition**: **Bernoulli distribution**, $X \sim \mathrm{Bern}(p)$:

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p, \quad 0 \le p \le 1.$$

- **Properties**:
  1. The pmf is $p(x) = p^x(1-p)^{1-x}$ for $x = 0, 1$.
  2. The mean is $\mathsf{E}X = \mu = 1 \cdot p + 0 \cdot (1-p) = p$.
  3. Since $\mathsf{E}(X^2) = 1^2 \cdot p + 0^2(1-p) = p$,

$$\sigma^2 = \mathsf{Var}(X) = \mathsf{E}(X^2) - \mu^2 = p - p^2 = p(1-p).$$

Binomial Distribution

Definition of Binomial distribution

A **binomial distribution** is a common probability distribution that occurs in practice. It arises in the following situation:

(1) There are n independent trials.

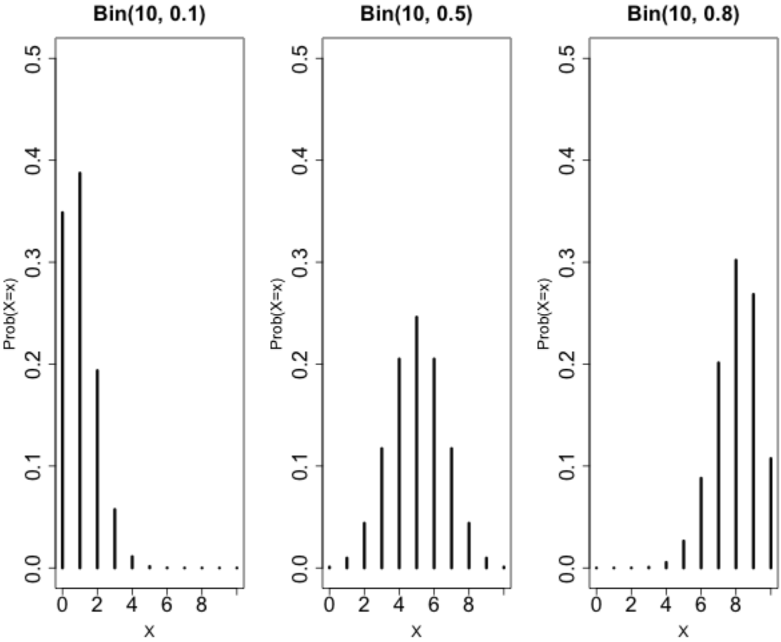
(2) Each trial results in a “success” or “failure”.

(3) The probability of success in each and every trial is equal to p.

If the random variable X counts the number of successes in the n trials, then X has a binomial distribution with parameters n and p:

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

Remark 1 The Bernoulli distribution is a special case of Binomial distribution with $n = 1$.


Bin(10, 0.1)
Bin(10, 0.5)
Bin(10, 0.8)
Prob(X=x)
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
x

Properties of Binomial distribution

If $X \sim \text{Bin}(n, p)$, then

1. The probability distribution of X is

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

   for $x = 0, 1, 2, \ldots, n$
2. $\mathsf{E}(X) = \mu = np$.
3. $\sigma^2 = np(1-p)$.

Note:

(1) $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. Recall that this is called a *combination* and is read "n choose x".

(2) $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1$

The mgf of a binomial distribution is

$$\begin{aligned}
M(t) &= \sum_x e^{tx} p(x) = \sum_x e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_x \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
&= [(1-p) + pe^t]^n, \ \forall t. \\
M'(t) &= n[(1-p) + pe^t]^{n-1}(pe^t), \\
M'(t) &= n[(1-p) + pe^t]^{n-1}(pe^t) + n(n-1)[(1-p) + pe^t]^{n-2}(pe^t)^2,
\end{aligned}$$

which gives that

$$\mu = M'(0) = np,$$

and

$$\sigma^2 = M''(0) - \mu^2 = np(1-p).$$

Theorem 3.1.1

Let X_1, X_2, ..., X_m be independent random variables such that X_i has a $\text{Bin}(n_i, p)$ distribution, for $i = 1, 2, \ldots, m$. Let

$$Y = \sum_{i=1}^{m} X_i.$$

Then, $Y \sim \text{Bin}(\sum_{i=1}^{m} n_i, p)$.

Proof.

The mgf of X_i is $M_{X_i}(t) = (1 - p + pe^t)^{n_i}$. By independence, we see

$$M_Y(t) = \prod_{i=1}^{m} \left(1 - p + pe^t\right)^{n_i} = \left(1 - p + pe^t\right)^{\sum_{i=1}^{m} n_i}.$$

□

Example

Consider the following settings. Is X a binomial random variable?

1. Let X equal the number times the ball lands in red in 10 spins of a roulette wheel (on a roulette wheel, there are 38 slots: 18 red, 18 black, and 2 green). Yes, $X \sim \text{Bin}(n = 10, p = 18/38)$
2. Let X equal the number of rainy days in the month of May. No, since trials are not independent.
3. Let X equal the number of black chips when drawing 2 chips *with* replacement from a bowl containing 2 black and 3 red chips. Yes, $X \sim \text{Bin}(n = 2, p = 2/5)$.
4. Let X equal the number of black chips when drawing 2 chips *without* replacement from a bowl containing 2 black and 3 red chips. No, since trials are not independent and the probability of success does not remain constant from trial to trial.
5. Let X equal the average weight of 20 randomly selected UI students. No, since X is not counting the number of "successes".

Suppose that 60% of adults have had their wisdom teeth removed. Suppose 10 adults are randomly selected. Assume independence.

- Find the probability that exactly 3 have had their wisdom teeth removed.

  Solution:

  This is a "binomial setting" (i.e. it satisfies the 3 requirements in the definition). So $X \sim \mathrm{Bin}(n = 10, p = 0.60)$, hence

$$\begin{aligned} P(X = 3) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{10}{3} 0.60^3 (1-0.60)^{10-3} \\ &= 120(0.60)^3(0.40)^7 \\ &= 0.04247 \end{aligned}$$

  where $\binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$.

- If 10 adults are randomly selected, how many do we expect to have had their wisdom teeth pulled, on average?
  Solution:
  $X \sim \text{Bin}(10, 0.60)$, so
  $$E(X) = np = 10(0.60) = 6.$$
- Determine σ.
  Solution:
  $X \sim \text{Bin}(10, 0.60)$, so
  $$\sigma^2 = np(1-p) = 10(0.60)(1-0.60) = 2.40$$
  and
  $$\sigma = \sqrt{2.40} = 1.549.$$

Example 3.1.4

Suppose a random experiment that has success probability p. Let X be the number of successes throughout n independent repetitions of the random experiment. Then as the number of experiments increases to infinity, the relative frequency of success, X/n, converges to p in the following sense:

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) = 0. \quad \text{for any } \varepsilon > 0\,.$$

Solution: Recall Chebyshev's inequality:

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2,$$

so we see

$$P(|X/n - p| \geq \varepsilon) \leq \text{Var}(X/n)/\varepsilon^2.$$

Interpretation: The relative frequency of success is close to the probability of p of success, for large values of n. This is the so-called **Weak Law of Large Numbers**, which will be discussed in Chapter 5.

Multinomial Distribution

- The binomial distribution can be generalized to the multinomial distribution. For simplicity, we consider the trinomial distribution.
- Consider a random experiment with three mutually exclusive and exhaustive events, C_1, C_2 and C_3. Let $p_i = P(C_i)$ for $i = 1, 2, 3$. Thus, $p_1 + p_2 + p_3 = 1$.
- Repeat the above experiment n independent times. Define the random variable X, Y, and Z to be the number of times that event C_1, C_2, C_3 occur. Then X, Y, and Z are nonnegative random variables such that $X + Y + Z = n$.
- This distribution of (X, Y) is called the **trinomial distribution**.

Joint pmf and joint mgf of trinomial distribution

- Joint pmf of X and Y:

$$p(x,y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y},$$

where x and y are nonnegative integers and $x + y \le n$.

- Joint mgf of X and Y:

$$M(t_1, t_2) = (p_1 e^{t_1} + p_2 e^{t_2} + p_3)^n,$$

for all $t_1, t_2 \in \mathbb{R}$.

- We see $X \sim \mathrm{Bin}(n, p_1)$ and $Y \sim \mathrm{Bin}(n, p_2)$ according to $M(t_1, 0)$ and $M(0, t_2)$.

Hypergeometric Distribution

From Binomial to Hypergeometric

An urn contains 6 red marbles and 4 blue marbles. We are going to select 3 marbles from the urn with the following two different sampling plans. Find the probability that 2 of the selected marbles will be red?

1. Sampling with replacement: each time we pick a marble out, we replace it with a marble of the same color.
2. Sampling without replacement: each time we pick a marble out, it means one less marble in the urn.

Solution:

Let X be the number of red marbles selected.

1. With replacement
   - X follows binomial distribution.
   - $n = 3, p = 6/10 = 0.6$.
   - We want the probability of $x = 2$:

$$f(2) = \binom{3}{2} 0.6^2 (1 - 0.6)^1 = 0.432.$$

2. Without replacement: X follows hypergeometric distribution...

Hypergeometric distribution

- Sampling without replacement.
- The probability of success changes from trial to trial.
- Trials are no longer independent.
- Fixed number of trials.

Definition

The pmf of the **hypergeometric random variable** X, the number of successes in a random sample of size n selected from N items of which k are labeled success and $N-k$ labeled failure, is

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \max(0, n-(N-k)) \leq x \leq \min(n, k).$$

Back to sampling plan 2

Example

An urn contains 6 red marbles and 4 blue marbles. If we select 3 marbles at random, what is the probability that 2 of them will be red?

Solution:

- Usually, if we don't mention sampling with replacement, the scenario is without replacement as in this example.
- Let X be the number of red marbles selected. Then X follows hypergeometric distribution.

Binomial and hypergeometric relationship

- If N is very large compared to n we can use the binomial distribution to approximate the hypergeometric distribution.
- Use $p = k/N$.
- Good results if $n/N \leq 0.05$.

Exp and var of hypergeometric distribution

- For hypergeometric distribution (i.e., without replacement):

$$\mathsf{E}(X) = n\frac{k}{N} \text{ and } \mathsf{Var}(X) = n\frac{k}{N}\frac{N-k}{N}\frac{N-n}{N-1}.$$

- For binomial distribution (i.e., with replacement):

$$\mathsf{E}(X) = n\frac{k}{N} \text{ and } \mathsf{Var}(X) = n\frac{k}{N}\frac{N-k}{N}.$$

- The term $\frac{N-n}{N-1}$ is also named **finite population correction factor** (FPC).

Negative Binomial Distribution

Negative Binomial Distribution

Assumption:

(1) The trials are independent of each other.

(2) Each trial results in a “success” or “failure”.

(3) The probability of success in each and every trial is equal to p.

Definition: The negative binomial random variable X is the total number of trials in the sequence to produce exactly r successes (r is a fixed positive integer):

The pmf of $X \sim \mathrm{NB}(r, p)$:

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \quad x = r, r+1, r+2, r+3, \ldots$$

Remark on different forms of negative binomial

Definition 1: The pmf of $X \sim \mathrm{NB}(r, p)$: X: total number of trials (including the rth success)

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \quad x = r, r+1, r+2, r+3, \ldots$$

A different definition of NB(r, p) is that it is the distribution of "the number of **failures** in the sequence of trials before the rth success".

Definition 2: The pmf of $X \sim \mathrm{NB}(r, p)$: X: total number of failures

$$f(x) = \binom{x+r-1}{r-1}(1-p)^{x-r}p^r, \quad x = 0, 1, 2, 3, \ldots$$

The two definitions can be distinguished by whether the support starts at $x = 0$ or at $x = r$.

We stick to Definition 1 in this class.

Geometric Distribution

Geometric distribution

- The **geometric distribution** is a special case of a negative binomial distribution with $r = 1:\ X \sim \mathrm{NB}(1, p) = \mathrm{Geo}(p)$.
- The pmf for geometric random variable is

$$f(x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots$$

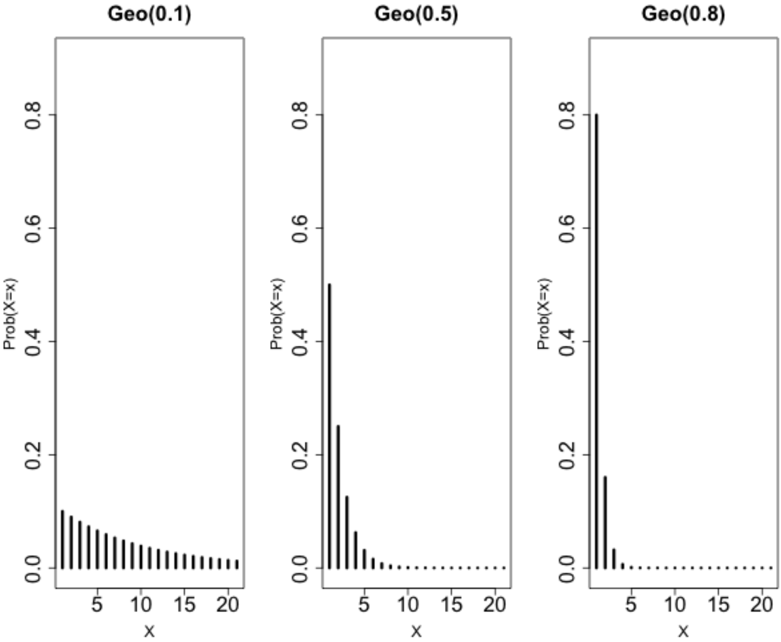
- The cdf for geometric random variable is

$$F(x) = 1 - (1-p)^{x}, \quad x = 1, 2, 3, \ldots$$

- Expectation, variance, and mgf:

$$\mathsf{E}X = \frac{1}{p} \text{ and } \mathsf{Var}(X) = \frac{1-p}{p^2}.$$

$$m(t) = \frac{pe^t}{1-(1-p)e^t},\ t < -\ln(1-p).$$


Geo(0.1)
Geo(0.5)
Geo(0.8)
Prob(X=x)
0.0
0.2
0.4
0.6
0.8
5
10
15
20
x

Memoryless property of geometric distribution

- The geometric distribution has a **memoryless property**. Suppose $X \sim \text{Geo}(p)$. Let x, y be positive integers. Then

$$P(X > x + y \,|\, X > x) = P(X > y).$$

- Tossing a coin is memoryless.
- Another example on Wikipedia: thousands of safes in a long hallway, and each safe has 500 position. An eccentric person stops at each safe once and makes a single attempt to open it.
- Based on this property, can we simply obtain the expectation of X?
- We see

$$\mathsf{E}(X) = p \cdot 1 + (1-p)\mathsf{E}(X+1),$$

which gives $\mathsf{E}(X) = 1/p$.

Chapter 3 Some Special Distributions

3.2 The Poisson Distribution

Poisson distribution

Suppose X is a discrete random variable such that X equals the number of occurrences in a given (time) interval or region where m equals the expected number of occurrences in the given interval or region. We assume that

(1) The number of occurrences in non-overlapping intervals or regions are independent.

(2) The probability of exactly 1 occurrence in a sufficiently short interval (or small region) is proportional to the length of the interval (or size of the region).

(3) The probability of 2 or more occurrences in a sufficiently short interval (or small region) is approximately equal to 0.

Then X has **a Poisson distribution with parameter m**. Using mathematical shorthand we write

$$X \sim \mathrm{Pois}(m).$$

From Binomial distribution to Poisson distribution

- Binomial pmf of $\mathrm{Bin}(n, p)$:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

- Consider the limiting case when $n \to \infty$ and $p \to 0$ while $np = m$:

$$\lim_{n\to\infty,\ p\to\frac{m}{n}} P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \frac{e^{-m} m^x}{x!},$$

which is pmf of $\mathrm{Pois}(m)$.

If $X \sim \text{Pois}(m)$, then

(1) The probability distribution of X is

$$P(X = x) = \frac{e^{-m} m^x}{x!}$$

for $x = 0, 1, 2, \ldots$

(2) As usual, adding up all of the probabilities yields

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{e^{-m} m^x}{x!} = 1.$$

(3) The Poisson mgf:

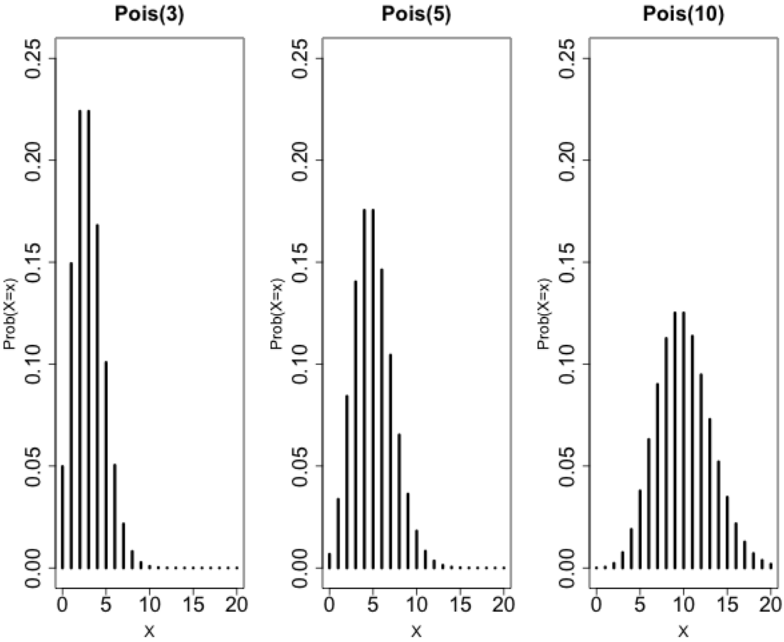
$$M(t) = e^{m(e^t - 1)} \quad \forall t,$$

$$M'(t) = e^{m(e^t - 1)}(me^t),$$

$$M''(t) = e^{m(e^t - 1)}(me^t) + e^{m(e^t - 1)}(me^t)^2.$$

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

(4) $\sigma^2 = m$.


Pois(3)
Pois(5)
Pois(10)
Prob(X=x)
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
x

Example

The average number of defects per square yard of cloth is 1.2. Assume assumptions (1)-(3) are met.

1. Find the probability of 1 defect in 1 square yard.
   Solution:
   Letting X equal the number of defects in 1 square yard, we have $X \sim \mathrm{Pois}(m = 1.2)$. Thus,
   $$P(X = 1) = \frac{e^{-m} m^x}{x!} = \frac{e^{-1.2} 1.2^1}{1!} = 0.3614.$$
2. Find the probability of 2 defects in 2 square yards.
   Solution:
   Letting Y equal the number of defects in 2 square yards, we have $Y \sim \mathrm{Pois}(m = 2 \times 1.2 = 2.4)$. Thus,
   $$P(Y = 2) = \frac{e^{-m} m^y}{y!} = \frac{e^{-2.4} 2.4^2}{2!} = 0.2613.$$

Example

The average number of defects per square yard of cloth is 1.2. Assume assumptions (1)-(3) are met.

3. Suppose we have a large pile of pieces of cloth. Each piece of cloth is exactly 1 square yard in size. If you counted the number of flaws in each piece and recorded this data, what would be the standard deviation of your dataset?
   Solution:
   If X equals the number of flaws in 1 square yard, then $X \sim \mathrm{Pois}(m = 1.2)$. Thus, the standard deviation would be

$$\sigma = \sqrt{m} = \sqrt{1.2} = 1.095.$$

Theorem 3.2.1

Suppose that $X_1, \ldots, X_n$ are independent random variables such that $X_i \sim \mathrm{Pois}(m)$ for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^n X_i$. Then $Y \sim \mathrm{Pois}(\sum_{i=1}^n m_i)$.

Proof.

We have

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= \prod_{i=1}^n \exp\{m_i(e^t - 1)\} \\ &= \exp\left\{\sum_{i=1}^n m_i(e^t - 1)\right\}. \end{aligned}$$

Thus Y has a Poisson distribution with parameter $\sum_{i=1}^n m_i$ due to the uniqueness of mgf. □

Sampling from a Poisson distribution

- Let $X \sim \text{Pois}(\lambda)$ be the number of customers arriving in 8 hours.
- We decompose $X = F + M$, where F and M are the number of female and male customers and $P(\text{female}) = p$.
- The joint pmf of F and M is

$$\begin{aligned}
P_{J,K}(j,k) &= P(F=j, M=k) \\
&= P(F=j, M=k|X=j+k) \cdot P(X=j+k) \\
&= \binom{j+k}{j} p^j (1-p)^k \cdot \frac{e^{-\lambda}\lambda^{j+k}}{(j+k)!} \\
&= \frac{e^{-\lambda p}(\lambda p)^j}{j!} \frac{e^{-\lambda(1-p)}[\lambda(1-p)]^k}{k!}.
\end{aligned}$$

- We see that $F \sim \text{Pois}(\lambda p)$ and $M \sim \text{Pois}[\lambda(1-p)]$ are independent.

Chapter 3 Some Special Distributions

3.3 The Γ, χ^2, and β Distributions

The Gamma Distribution

Gamma Function

- Gamma function:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy, \qquad \alpha > 0.$$

- Properties:

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty e^{-y} dy = 1, \\
\Gamma(\alpha) &= (\alpha - 1) \int_0^\infty y^{\alpha-2} e^{-y} dy = (\alpha - 1)\Gamma(\alpha - 1), \\
\Gamma(n) &= (n-1)!
\end{aligned}$$

Gamma Distribution

- Gamma function with $y = x/\beta$:

$$\Gamma(\alpha) = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \left(\frac{1}{\beta}\right) dx, \qquad \alpha > 0,\ \beta > 0.$$

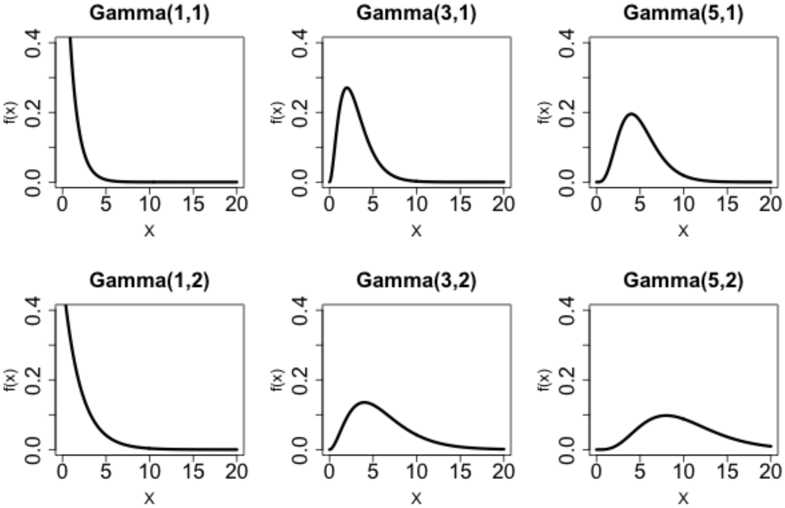
- Equivalent form:

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx.$$

- Gamma distribution $X \sim \mathrm{Gamma}(\alpha, \beta)$:

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

- Parameters: α shape, β scale.



The first row depicts the pdf of exponential distribution and the second row is for Chi-square distribution.

Properties of Gamma Distribution

- Moment generating function:

$$M_X(t) = \frac{1}{(1 - \beta t)^{\alpha}}, \; t < 1/\beta.$$

- Expectation:

$$M'(t) = (-\alpha)(1 - \beta t)^{-\alpha - 1}(-\beta),$$

$$\mu_X = \alpha\beta.$$

- Variance:

$$M''(t) = (-\alpha)(-\alpha - 1)(1 - \beta t)^{-\alpha - 2}(-\beta)^2,$$

$$\sigma_X^2 = \alpha\beta^2.$$

Suppose that a random variable X has a probability density function given by,

$$f(x) = \begin{cases} kx^3 e^{-x/2} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Find the value of k that makes $f(x)$ a density function.
- Find the expectation and variance.

One Special Γ: Exponential Distribution

Exponential Distribution

- X has **exponential distribution** $X \sim \exp(\beta)$ if X is $\Gamma(\alpha = 1, \beta)$.
- pdf and cdf:

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$F(x) = 1 - e^{-x/\beta} \quad x > 0.$$

- Expectation: β.
- Variance: β^2.

Memorylessness of Exponential Distribution

The exponential distribution is "memoryless":

$$P(X > x) = e^{-x/\beta}.$$

Let $x > 0, y > 0$, then

$$P(X > x + y \,|\, X > x) = \frac{e^{-(x+y)/\beta}}{e^{-x/\beta}} = e^{-y/\beta} = P(X > y).$$

- The only memoryless continuous probability distributions are the exponential distributions.
- The only memoryless discrete probability distributions are the geometric distributions

Applications of Exponential Distribution

The exponential distribution is frequently used to model the following:

- Life time of electronic components
- The time until a radioactive particle decays.
- The time until default (on payment to company debt holders) in reduced form credit risk modeling.
- The time it takes for a bank teller to serve a customer.

From Poisson Distribution to Exponential Distribution

- Suppose the event occur in time according to a Poisson process with parameter λ, i.e., $X \sim \text{Poisson}(\lambda)$.
- Over the time interval $(0, t)$, the probability of have x occurrence is

$$P(X = x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}.$$

- Let T be the length of time until the first arrival.
- Target: distribution of T, i.e., $F(t) = P(T \leq t)$.
- The waiting time of the first event is great than t, which is equivalent to zero occurrence over time interval $(0, t)$:

$$P(T > t) = P(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}.$$

$$P(T \leq t) = 1 - e^{-\lambda t},$$

so $T \sim \exp(1/\lambda)$.

Motivation of Additivity of Gamma Distribution

What is the distribution of the waiting time of the kth occurrence?

In other words, if T_1 is the waiting time of the 1st occurrence, T_2 is the waiting time of the 2nd occurrence, ... what is the distribution of

$$T = T_1 + T_2 + \ldots + T_k.$$

Answer: suppose each $T_i \sim \exp(\beta) \sim \Gamma(1, \beta)$, then $T \sim \Gamma(k, \beta)$.

Why?

Theorem 3.3.2

If $X_1, \dots, X_n$ are independent random variables and that

$$X_i \sim \text{Gamma}(\alpha_i, \beta) \quad \text{for } i = 1, \dots, n.$$

Let $Y = \sum_{i=1}^n X_i$. Then $Y \sim \text{Gamma}\Big(\sum_{i=1}^n \alpha_i,\ \beta\Big)$.

Solution Sketch:

$$M_Y(t) = \prod_{i=1}^n (1 - \beta t)^{-\alpha_i} = (1 - \beta t)^{-\sum_{i=1}^n \alpha_i},\ t < 1/\beta.$$

Another Special Γ: χ^2 Distribution

Chi-square Distribution

- A **Chi-square distribution** with parameter r, $X \sim \chi^2(r)$, is defined to be $\mathrm{Gamma}(\alpha = r/2, \beta = 2)$.
- pdf

$$f(x) = \begin{cases} \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- mgf:

$$M_X(t) = (1-2t)^{-r/2}, \ t < 1/2.$$

- Expectation: r.
- Variance: $2r$.

CDF of Chi-squared Distribution

The cdf of a Chi-squared distribution is

$$F(x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} t^{r/2-1} e^{-t/2} dt\,,$$

which does **not** have a closed form expression.

Instead, at each x, the value of $F(x)$ can be evaluated numerically.

Example

Suppose $X \sim \chi^2(10)$. Determine the probability $P(3.25 \leq X \leq 20.5)$.

Table II
Chi-Square Distribution

The following table presents selected quantiles of chi-square distribution, i.e., the values x such that

$$P(X \leq x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2}\, dw,$$

for selected degrees of freedom r.

	$P(X \leq x)$							
r	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000

Corollary 3.3.1

If $X_1, \ldots, X_n$ are independent random variables and that

$$X_i \sim \chi^2(r_i) \quad \text{for } i = 1, \ldots, n.$$

Let $Y = \sum_{i=1}^n X_i$. Then $Y \sim \chi^2\left(\sum_{i=1}^n r_i\right)$.

Theorem 3.3.1

Suppose $X \sim \chi^2(r)$. If $k > -r/2$ then $\mathsf{E}\left(X^k\right)$ exists and it is given by

$$\mathsf{E}(X^k) = \frac{2^k \Gamma(\frac{r}{2} + k)}{\Gamma(\frac{r}{2})}, \text{ if } k > -\frac{r}{2}.$$

The Beta Distribution

Motivating example: Suppose that X_1 has a Gamma$(\alpha, 1)$ distribution, that X_2 has a Gamma$(\beta, 1)$, and that X_1, X_2 are independent, where $\alpha > 0$ and $\beta > 0$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1/(X_1 + X_2)$. Show that Y_1 and Y_2 are independent.

Solution:

One-to-one transformation: $x_1 = y_1 y_2$ and $x_2 = y_1(1 - y_2)$,
$0 < x_1, x_2 < \infty$ gives $0 < y_1 < \infty,\ 0 < y_2 < 1$.

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1,$$

Joint pdf of X_1 and X_2:

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x_1^{\alpha-1} x_2^{\beta_1} e^{-x_1-x_2},\ 0 < x_1, x_2 < \infty.$$

Joint pdf of Y_1 and Y_2:

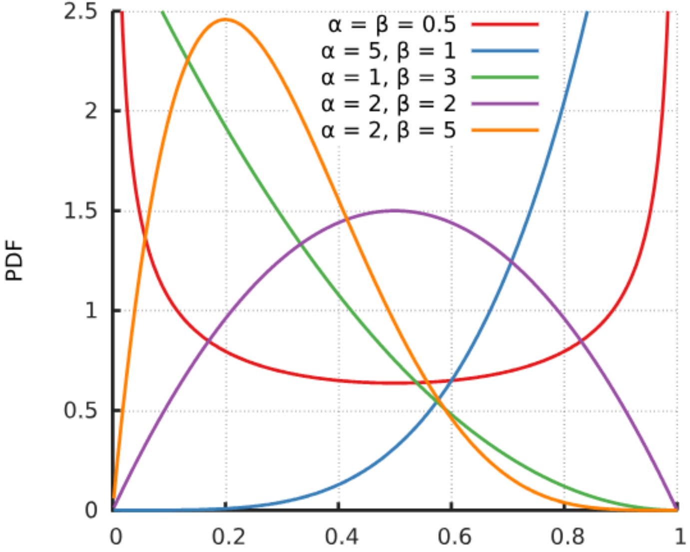
$$f_{Y_1,Y_2}(y_1, y_2) = \left(\frac{1}{\Gamma(\alpha+\beta)} y_1^{\alpha+\beta-1} e^{-y_1}\right)\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_2^{\alpha-1}(1 - y_2)^{\beta-1}\right).$$

Beta Distribution

We say Y_2 follows a **beta distribution** with parameters α and β.

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \ 0 < y < 1.$$

- Mean: $\dfrac{\alpha}{\alpha+\beta}$.
- Variance: $\dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.
- Beta distribution is **not** a special case of Gamma distribution.


α = β = 0.5
α = 5, β = 1
α = 1, β = 3
α = 2, β = 2
α = 2, β = 5
PDF
0
0.5
1
1.5
2
2.5
0
0.2
0.4
0.6
0.8
1

Chapter 3 Some Special Distributions

3.4 The Normal Distribution

- Gamma function:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy, \qquad \alpha > 0.$$

- Properties:

$$\Gamma(1) = \int_0^\infty e^{-y} dy = 1,$$
$$\Gamma(\alpha) = (\alpha - 1) \int_0^\infty y^{\alpha-2} e^{-y} dy = (\alpha-1)\Gamma(\alpha-1),$$
$$\Gamma(n) = (n-1)!$$

- What is $\Gamma(0.5)$?

- Target:

$$\Gamma(0.5) = \int_0^\infty y^{-1/2} e^{-y} dy.$$

- Define $y = x^2$, so $dy = 2xdx$, and

$$\Gamma(0.5) = 2\int_0^\infty e^{-x^2} dx = \int_{-\infty}^\infty e^{-x^2} dx.$$

- **Key fact:**

$$\Gamma(0.5) = \int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}.$$

Key fact:

$$\Gamma(0.5) = \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Proof:

$$\begin{aligned}
\{\Gamma(0.5)\}^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right) \\
&= \left(\int_{-\infty}^{\infty} e^{-y^2} dy\right)\left(\int_{-\infty}^{\infty} e^{-z^2} dz\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-y^2-z^2} dydz \\
&= \int_{0}^{2\pi} d\theta \int_{0}^{\infty} e^{-r^2} rdr = \frac{1}{2}\int_{0}^{2\pi} d\theta \int_{0}^{\infty} e^{-r^2} dr^2 \\
&= \pi.
\end{aligned}$$

In the second last inequality, polar coordinates are used such that $x = r\sin\theta$ and $y = r\cos\theta$.

From

$$\Gamma(0.5) = \int_{-\infty}^{\infty} \exp\left\{-x^2\right\} dx = \sqrt{\pi},$$

we see that

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = \sqrt{2\pi}.$$

Can we construct a distribution related to the integral above?

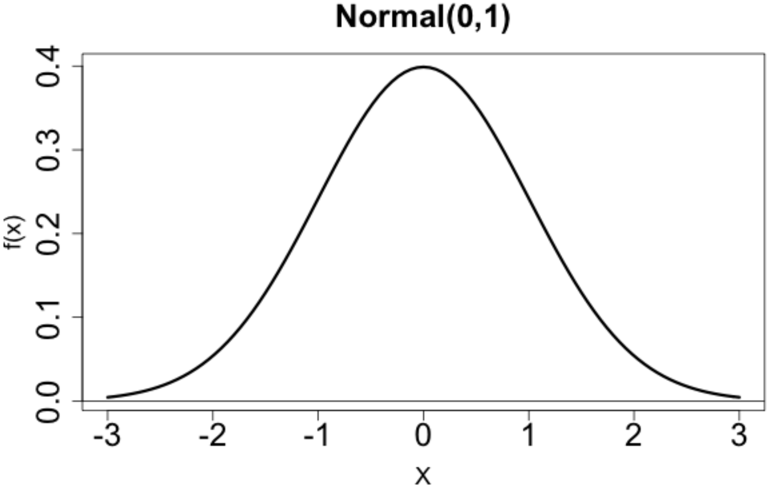
Normal Distribution, aka Gaussian Distribution

A random variable Z is said to follow a **standard normal distribution** if it has pdf

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \qquad -\infty < z < \infty$$

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

Density Function of Standard Normal Distribution



Properties of Standard Normal Distribution

Moment generating function:

$$\begin{aligned}
M_Z(t) &= \mathrm{E}\exp(tZ) \\
&= \int_{-\infty}^{\infty} \exp(tz)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)dz \\
&= \exp\left(\frac{1}{2}t^2\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(z-t)^2\right)dz \\
&= \exp\left(\frac{1}{2}t^2\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}w^2\right)dw,\ w = z - t \\
&= \exp\left(\frac{1}{2}t^2\right),\ -\infty < t < \infty.
\end{aligned}$$

Expectation: $M_Z'(t) = t\exp\left(\frac{1}{2}t^2\right) \Rightarrow \mathrm{E}(Z) = 0.$
Variance: $M_Z''(t) = \exp\left(\frac{1}{2}t^2\right) + t^2\exp\left(\frac{1}{2}t^2\right) \Rightarrow \mathrm{Var}(Z) = 1.$

Let $X = \sigma Z + \mu$, where the random variable $Z \sim \mathrm{N}(0, 1)$, and $-\infty < \mu < \infty$, $\sigma > 0$ are two parameters. What is the pdf of X?

Answer:

- The pdf of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}.$$

- $X = \sigma Z + \mu \iff Z = \frac{X-\mu}{\sigma}$.
- The Jacobian is σ^{-1}.
- Then the pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < x < \infty.$$

We say that X has a **normal distribution** if it has the pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < x < \infty,$$

where $-\infty < \mu < \infty$ and $\sigma^2 > 0$ are two parameters. Write $X \sim \mathrm{N}(\mu, \sigma^2)$.

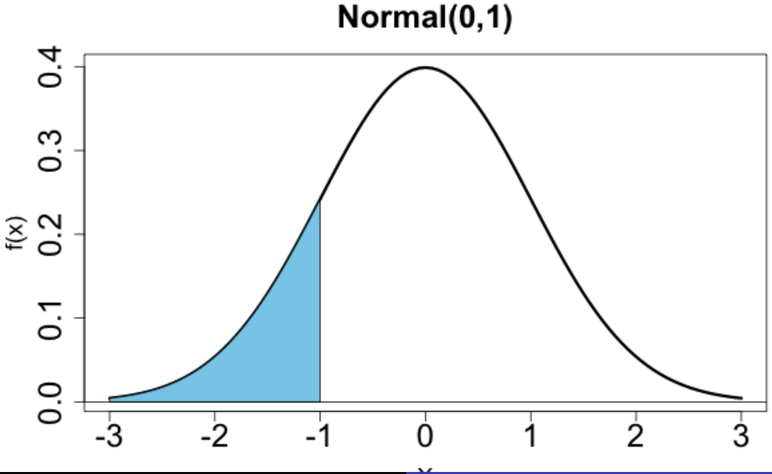
Properties:

- Expectation: μ. We call μ the location parameter
- Variance σ^2. We call σ the scale parameter.
- Moment generating function:

$$\begin{aligned} \mathrm{E}\exp(tX) &= \mathrm{E}\exp(t(\sigma Z + \mu)) = \exp(\mu t)\,\mathrm{E}\exp(t\sigma Z) \\ &= \exp(\mu t)\exp\left(\frac{1}{2}\sigma^2 t^2\right) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right). \end{aligned}$$

Denote by $\Phi(z)$ the cdf of the standard normal distribution $N(0,1)$, that is to say,

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt.$$



From its picture, show that

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

Table III in Appendix C offers an abbreviated table of probabilities for a standard normal.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

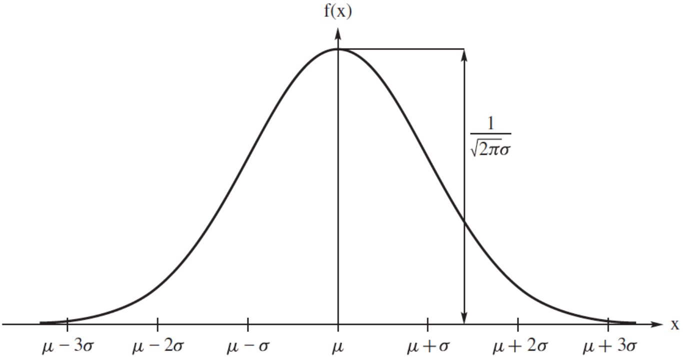
Example

Let Z denote a normal random variable with mean 0 and standard deviation 1. Find

1. $P(Z > 2)$,
2. $P(-2 \leq Z \leq 2)$.

Example. Let X be $\mathrm{N}(2, 25)$. Find $P(0 < X < 10)$.

Example. Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$. Find $P(\mu - 3\sigma < X < \mu + 3\sigma)$.



Theorem 3.4.1 From Normal to Chi-square

Suppose $Z \sim \mathrm{N}(0, 1)$, then $W = Z^2$ follows $\chi^2(1)$.

Proof.

1. Since $W = Z^2$, then $Z = \sqrt{W}$ when $Z \geq 0$ and $Z = -\sqrt{W}$ if $Z < 0$.
2. The pdf of Z is $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}$.
3. The pdf of W is

$$
\begin{aligned}
& f_W(w) \\
=& \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\sqrt{w})^2}{2}\right\} \left|\frac{1}{2\sqrt{w}}\right| + \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\sqrt{w})^2}{2}\right\} \left|-\frac{1}{2\sqrt{w}}\right| \\
=& \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{w}} \exp\left\{-\frac{w}{2}\right\} \text{ recall that } \Gamma(0.5) = \sqrt{\pi} \\
=& \frac{1}{\Gamma(0.5) 2^{0.5}} w^{0.5-1} \exp\left\{-\frac{w}{2}\right\},
\end{aligned}
$$

which is the pdf of $\Gamma(\alpha = 0.5, \beta = 2)$, which is $\chi^2(1)$. □

Theorem 3.4.2

Let $X_1, \ldots, X_n$ be independent random variables such that X_i follows $\mathrm{N}(\mu_i, \sigma_i^2)$. Then, for constants $a_1, \ldots, a_n$,

$$Y = \sum_{i=1}^{n} a_i X_i \quad \text{follows} \quad \mathrm{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Proof.

$$M_Y(t) = \prod_{i=1}^{n} \exp\left\{t a_i \mu_i + \frac{1}{2} t^2 a_i^2 \sigma_i^2\right\}$$
$$= \exp\left\{t \sum_{i=1}^{n} a_i \mu_i + \frac{1}{2} t^2 \sum_{i=1}^{n} a_i^2 \sigma_i^2\right\}.$$

Recall that the mgf of $\mathrm{N}(\mu, \sigma^2)$ is $\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$ □

Corollary 3.4.1

Let $X_1, \ldots, X_n$ be i.i.d. with common distribution $\mathrm{N}(\mu, \sigma^2)$. Then,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{follows} \quad \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Let X_1, X_2, and X_3 be i.i.d. random variables with common mgf $\exp\{t + 2t^2\}$.

1. Compute the probability $P(X_1 < 3)$.
2. Derive the mgf of $Y = X_1 + 2X_2 - 2X_3$.
3. Compute the probability $P(Y > 7)$.

Chapter 3 Some Special Distributions

3.5 The Multivariate Normal Distribution

Standard Bivariate Normal Distribution

A n-dim random vector $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^\top$ is said to have a **standard bivariate normal distribution** if its pdf is

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_i^2}{2}\right\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{z}\right\}.$$

It can be shown that $\mathrm{E}\boldsymbol{Z} = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{Z}) = \boldsymbol{I}_n$.

Write $\boldsymbol{Z} \sim \mathrm{N}_n(\mathbf{0}, \boldsymbol{I}_n)$.

- Let $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$ where $\boldsymbol{A}$ is a (nonsingular) $n \times n$ matrix and $\boldsymbol{\mu}$ is an n-dim column vector. We introduce notation $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$. Then

$$\mathrm{E}\boldsymbol{X} = \boldsymbol{A}\mathrm{E}\boldsymbol{Z} + \boldsymbol{\mu}.$$

$$\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Sigma}.$$

- Transformation:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$$

gives

$$\boldsymbol{Z} = \boldsymbol{A}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}).$$

Then Jocobian is $|\boldsymbol{A}|^{-1} = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}$.

- The pdf of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-n/2}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}.$$

Definition of Multivariate Normal Distribution

A n-dim random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ is said to have a **bivariate normal distribution with mean μ and variance-covariance matrix Σ** if its pdf is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-n/2} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}.$$

Write $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

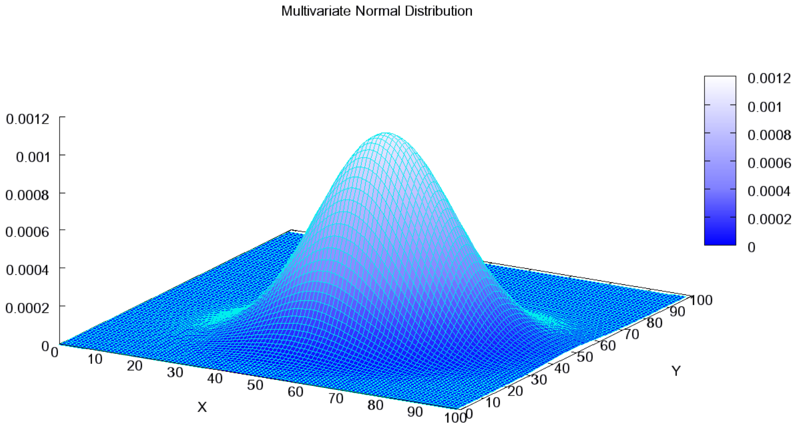
Non-matrix Expression of Bivariate Normal PDF

Suppose $\boldsymbol{X} \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then we write

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

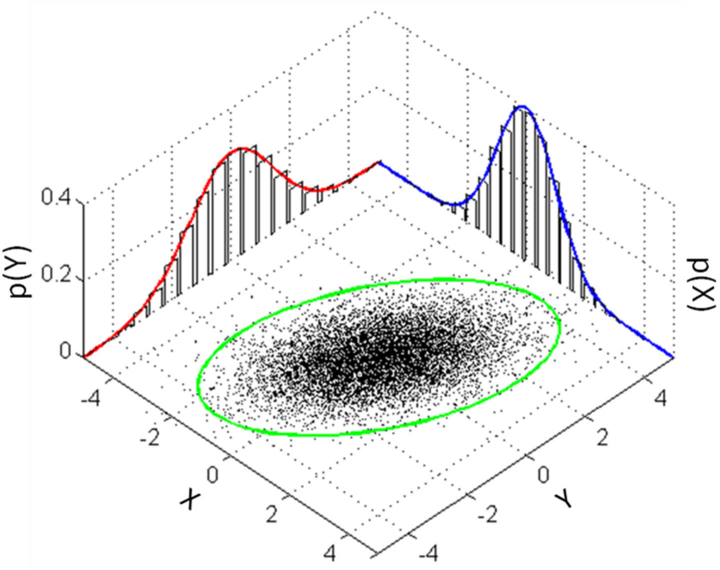
Hence the pdf of $\boldsymbol{X}$ can be expressed as

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times \exp\left\{ -\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$



The plot is from Wikipedia `https://en.wikipedia.org/wiki/Multivariate_normal_distribution`.

Marginal Distributions



The plot is from Wikipedia `https://en.wikipedia.org/wiki/Multivariate_normal_distribution`.

What is the marginal distribution of bivariate normal distribution?

Moment Generating Functions

- Multivariate standard normal distribution, $\boldsymbol{Z} \sim \mathrm{N}_n(\mathbf{0}, \boldsymbol{I})$:

$$M_{\boldsymbol{Z}}(\boldsymbol{t}) = \mathrm{E}(\exp(\boldsymbol{t}^\top \boldsymbol{Z})) = \mathrm{E}\left[\prod_{i=1}^n \exp(t_i Z_i)\right] = \prod_{i=1}^n \mathrm{E}(\exp(t_i Z_i))$$

$$= \exp\left\{\frac{1}{2}\sum_{i=1}^n t_i^2\right\} = \exp\left\{\frac{1}{2}\boldsymbol{t}^\top \boldsymbol{t}\right\}.$$

- Multivariate normal distribution, $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, recall $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$ and $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Sigma}$:

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{t}) &= \mathrm{E}(\exp(\boldsymbol{t}^\top \boldsymbol{X})) \\ &= \mathrm{E}(\exp(\boldsymbol{t}^\top \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{t}^\top \boldsymbol{\mu})) \\ &= \exp(\boldsymbol{t}^\top \boldsymbol{\mu})\mathrm{E}\left[\exp\left\{(\boldsymbol{A}^\top \boldsymbol{t})^\top \boldsymbol{Z}\right\}\right] \\ &= \exp(\boldsymbol{t}^\top \boldsymbol{\mu})\left[\frac{1}{2}\exp\left\{\boldsymbol{t}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{t}\right\}\right] \\ &= \exp(\boldsymbol{t}^\top \boldsymbol{\mu})\exp[(1/2)\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}]. \end{aligned}$$

Theorem

Suppose $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{X} + \boldsymbol{b}$ where $\boldsymbol{C}$ is a full rank $m \times n$ matrix and b is an $m \times 1$ vector. Then $\boldsymbol{Y} \sim \mathrm{N}_m(\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{b},\, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top)$.

Proof.

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= \mathrm{E}(\exp(\boldsymbol{t}^\top \boldsymbol{Y})) \\
&= \mathrm{E}(\exp(\boldsymbol{t}^\top \boldsymbol{C}\boldsymbol{X} + \boldsymbol{t}^\top \boldsymbol{b})) \\
&= \exp(\boldsymbol{t}^\top \boldsymbol{b})\mathrm{E}\left[\exp\left\{(\boldsymbol{C}^\top \boldsymbol{t})^\top \boldsymbol{X}\right\}\right] \\
&= \exp(\boldsymbol{t}^\top \boldsymbol{b})\left[\exp\left\{(\boldsymbol{C}^\top \boldsymbol{t})^\top \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top \boldsymbol{t}\right\}\right] \\
&= \exp(\boldsymbol{t}^\top (\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{b}))\exp[(1/2)\boldsymbol{t}^\top \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top \boldsymbol{t}].
\end{aligned}
$$

□

Linear combinations of normal random vector is still normally distributed.

Corollary

Suppose $\boldsymbol{X} = (X_1, X_2)^\top \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $X_1 \sim N(\mu_1, \sigma_1^2)$.

Proof.

Take $\boldsymbol{C} = (1, 0)$. □

Corollary 3.5.1

Extension: Suppose $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $\boldsymbol{X}_1 \sim N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}^2)$, where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$$

and

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Marginal distributions of multivariate normal vector are still normal distributions.

Independence

1. Recall that if X_1 and X_2 are independent, then $\mathrm{Cov}(X_1, X_2) = 0$.
2. However, if $\mathrm{Cov}(X_1, X_2) = 0$, then X_1 and X_2 is not necessarily independent.
3. However, if $\mathrm{Cov}(X_1, X_2) = 0$, and both X_1 and X_2 are normally distributed, then X_1 and X_2 must be independent.

Theorem 3.5.2

Suppose $\boldsymbol{X}$ has a $\mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Conditional Distributions

Theorem 3.5.3

Suppose $\boldsymbol{X}$ has a $\mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Assume that $\boldsymbol{\Sigma}$ is positive definite. Then the conditional distribution of $\boldsymbol{X}_1 | \boldsymbol{X}_2$ is

$$\mathrm{N}_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

The conditional distributions of a multivariate normal vector are also normal.

Example 3.5.2

Consider a bivariate normal random variable (X_1, X_2), with pdf

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\times \exp\left\{ -\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

The conditional distribution of X_1 given $X_2 = x_2$ is

$$\mathrm{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x-\mu_2)\ ,\ \sigma_1^2(1-\rho^2) \right)$$

Multivariate Normal and χ^2

Theorem 3.5.4

Suppose $\boldsymbol{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}$ is positive definite. Then the random variable $W = (\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ has a $\chi^2(n)$ distribution.

Chapter 3 Some Special Distributions

3.6 t- and F-Distributions

t-Distribution

Motivation

- If $X_1, \ldots, X_n$ is a random sample from $\mathrm{N}(\mu, \sigma^2)$, then

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

follows the standard normal distribution.

- Thus

$$P(-z_{0.025} < Z < z_{0.025}) = 95\%,$$
$$P\left(\bar{X} - z_{0.025}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right) = 95\%,$$

which is actually a 95% **confidence interval** for the location parameter μ. The confidence interval will be discussed in Chapter 4.2.

- In practice, σ is unknown, we may use

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

- But $P(-z_{0.025} < t < z_{0.025}) = 95\%$ does **not** hold anymore.
- What is the distribution of t?
- W.S.Gosset (who published the result with the name of Student) derived the distribution of t, which is named **Student's t-distribution.**
- The resulting confidence interval $P(-t_{0.025} < t < t_{0.025}) = 95\%$ is wider than the one based on normal distribution, because we have less information on σ.

Definition of t- Distribution

Suppose $W \sim \mathrm{N}(0,1)$, $V \sim \chi^2(r)$ and that W and V are independent. The pdf of

$$T = \frac{W}{\sqrt{V/r}}.$$

is

$$f(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\sqrt{\pi r}}\left(1+\frac{t^2}{r}\right)^{-\frac{r+1}{2}}, \ -\infty < t < \infty.$$

This distribution is called **t-distribution** with r degrees of freedom.

1. The joint pdf of W and V is

$$f_{W,V}(w,v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) \frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} v^{\frac{r}{2}-1} e^{-\frac{v}{2}},\ -\infty < w < \infty,\ v > 0$$

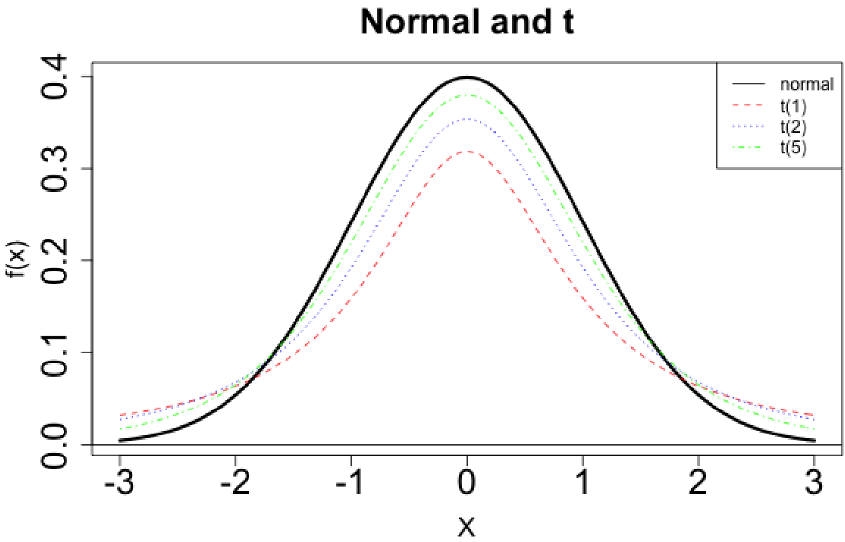
2. Transformation $(w,v) \to (t,v)$:

$$t = \frac{w}{\sqrt{\frac{v}{r}}},\ u = v \Rightarrow w = \frac{t\sqrt{u}}{\sqrt{r}},\ v = u.\ |J| = \frac{\sqrt{u}}{\sqrt{r}}.$$

3. Joint pdf of T and U is

$$\begin{aligned} f_{T,U}(t,u) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{ut^2}{2r}\right) \frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} \frac{\sqrt{u}}{\sqrt{r}} \\ &= \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}} \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}} \times \text{ pdf of } \Gamma\left(\frac{r+1}{2}, 2\left(1 + \frac{t^2}{r}\right)^{-1}\right) \end{aligned}$$

which gives the pdf of T.


Normal and t
normal
t(1)
t(2)
t(5)
f(x)
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x

Mean and Variance of t-Distribution

By the definition,

$$T = W(V/r)^{-1/2}, \ W \text{ and } V \text{ are independent.}$$

Suppose $r/2 - k/2 > 0$, we have

$$\mathrm{E}(T^k) = \mathrm{E}\left[W^k \left(\frac{V}{r}\right)^{-k/2}\right] = \mathrm{E}(W^k)\mathrm{E}\left[\left(\frac{V}{r}\right)^{-k/2}\right]$$

$$= \mathrm{E}(W^k)\frac{2^{-k/2}\Gamma(r/2 - k/2)}{\Gamma(r/2)r^{-k/2}}, \text{ if } k < r.$$

Recall for $X \sim \chi^2(r)$. If $k > -r/2$, then

$$\mathrm{E}(X^k) = \frac{2^k\Gamma(r/2 + k)}{\Gamma(r/2)}.$$

Therefore $\mathrm{E}(T) = E(W) = 0$ if $r > 1$,
and $\mathrm{Var}(T) = \mathrm{E}(T^2) = \frac{r}{r-2}$ if $r > 2$.

Properties of t-Distribution

- The density function of t-distribution is symmetric, bell-shaped, and centered at 0.
- The variance of t-distribution is larger than the standard normal distribution.
- The tail of t-distribution is heavier (larger kurtosis).
- As the degree of freedom increases, the density function of t-distribution converges to the density of the standard normal distribution. This is called **convergence in distribution**, as will be discussed in Example 5.2.3.

Student's Theorem

Suppose $X_1, \cdots, X_n$ are iid $\mathrm{N}(\mu, \sigma^2)$ random variables. Define the random variables,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Then

1. $\overline{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$;
2. $\overline{X}$ and S^2 are independent;
3. $(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$;
4. The random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t-distribution with $n-1$ degrees of freedom.

Proof of (2): Key steps:

1. Write S^2 as a function of $(X_2 - \bar{X}, \ldots, X_n - \bar{X})$.
2. Prove $\bar{X}$ is independent of $(X_2 - \bar{X}, \ldots, X_n - \bar{X})$.
3. Thus $\bar{X}$ is independent of S^2.

1. We observe that

$$
\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) \\
&= \frac{1}{n-1}\left(\left[\sum_{i=2}^{n}(X_i - \bar{X})\right]^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right),
\end{aligned}
$$

where

$$
X_1 - \bar{X} = -\sum_{i=2}^{n}(X_i - \bar{X})
$$

since $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$.

2. We want to prove $\bar{X}$ is independent of $(X_2 - \bar{X}, \ldots, X_n - \bar{X})$. Make the transformation:

$$y_1 = \bar{x},\ y_2 = x_2 - \bar{x}, \ldots, y_n = x_n - \bar{x}.$$

Inverse functions:

$$x_1 = y_1 - \sum_{i=2}^{n} y_i,\ x_2 = y_2 + y_1, \ldots, x_n = y_n + y_1.$$

Jacobian:

$$J = \begin{vmatrix} 1 & -1 & -1 & \ldots & -1 & -1 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & 1 & 1 \end{vmatrix} = n.$$

Joint pdf of $X_1, \ldots, X_n$:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\}.$$

Joint pdf of $Y_1, \ldots, Y_n$:

$$\begin{aligned}
&f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) \\
=&\frac{n}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\left(y_1 - \sum_{i=2}^n y_i\right)^2 - \frac{1}{2}\sum_{i=2}^n (y_i + y_1)^2\right\} \\
=&\frac{n}{(\sqrt{2\pi})^n} \exp\left\{-\frac{n}{2}y_1^2 - \frac{1}{2}\left[\sum_{i=2}^n y_i^2 + \left(\sum_{i=2}^n y_i\right)^2\right]\right\}, \quad -\infty < y_i < \infty.
\end{aligned}$$

Thus Y_1 is independent of $(Y_2, \ldots, Y_n)$. Equivalently, $\bar{X}$ is independent of $(X_2 - \bar{X}, \ldots, X_n - \bar{X})$, saying $\bar{X}$ is independent of S^2.

This proof could be simplified using matrix notations.

Proof of (3): It is known that

$$V = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

follows a $\chi^2(n)$ distribution. We observe that

$$\begin{aligned}
V &= \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma} \right)^2 \\
&= \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \\
&= \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \\
&\equiv V_A + V_B.
\end{aligned}$$

We see V_B follows a $\chi^2(1)$ distribution and the distribution of V_A is our interest. We have shown that V_A and V_B are independent.

As $V \sim \chi^2(n)$, $V_B \sim \chi^2(1)$, and V_A and V_B are independent, we take mgfs on both sides of $V = V_A + V_B$. We then have

$$\begin{aligned} M_V(t) &= M_{V_A}(t) M_{V_B}(t), \\ (1-2t)^{-n/2} &= M_{V_A}(t)(1-2t)^{-1/2}, \\ M_{V_A}(t) &= (1-2t)^{-(n-1)/2}. \end{aligned}$$

We thus know that

$$V_A \sim \chi^2(n-1).$$

Proof of (4): We write

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}},$$

where the numerator follows a standard normal distribution, the denominator is $\sqrt{\chi^2(n-1)/n-1}$, and the numerator and the denominator are independent. Thus T follows $t(n-1)$.

r	$P(T \leq t)$ 0.900	0.950	0.975	0.990	0.995	0.999
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
∞	1.282	1.645	1.960	2.326	2.576	3.090

Examples

1. Assume that T has a student t-distribution with 5 degrees of freedom. Find $P(|T| > 2.571)$.
2. Suppose that the five random variables $X_1, X_2, \ldots, X_5$ are i.i.d. and each has a standard normal distribution. Determine a constant c such that the random variable

$$\frac{c(X_1 + X_2)}{(X_3^2 + X_4^2 + X_5^2)^{1/2}}.$$

   will have a t-distribution.
3. Let $X_1, \ldots, X_6$ be iid random variables each having a normal distribution with mean μ and variance σ^2. Find

$$P\left(\bar{X} - 2.571\frac{S}{\sqrt{6}} < \mu < \bar{X} + 2.571\frac{S}{\sqrt{6}}\right).$$

F-Distribution

Motivation

- Suppose $X_1, \ldots, X_n$ is a random sample from $\mathrm{N}(\mu_1, \sigma_1^2)$, and $Y_1, \ldots, Y_m$ is independently drawn from $\mathrm{N}(\mu_2, \sigma_2^2)$.
- If our interest is σ_1^2/σ_2^2, then a nature choice is

$$S_X^2/S_Y^2.$$

- The F-distribution gives us

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(n-1, m-1).$$

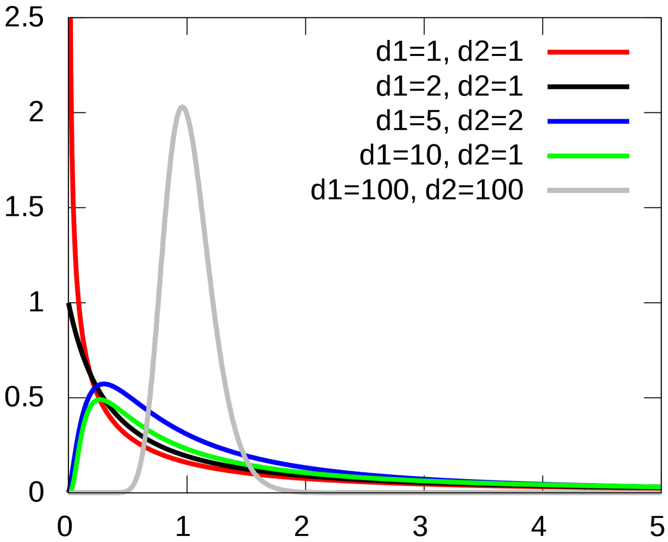
- This problem is frequently encountered in regression and analysis of variance (ANOVA).

Let U and V are two independent χ^2 random variables with degrees of freedom r_1 and r_2, respectively. The pdf of

$$W = \frac{U/r_1}{V/r_2}$$

is

$$f(w) = \frac{\Gamma(\frac{r_1+r_2}{2})(\frac{r_1}{r_2})^{r_1/2}}{\Gamma\frac{r_1}{2}\Gamma\frac{r_2}{2}} \frac{(w)^{\frac{r_1}{2}-1}}{(1+w\frac{r_1}{r_2})^{\frac{r_1+r_2}{2}}}, \ 0 < w < \infty.$$



This plot is from Wikipedia: `https://en.wikipedia.org/wiki/F-distribution`

Moments of F-Distributions

- Let F have an F-distribution with r_1 and r_2 degrees of freedom. We write
$$F = \frac{r_2}{r_1}\frac{U}{V},$$
where $U \sim \chi^2(r_1)$, $V \sim \chi^2(r_2)$, and U and V are independent.
- Thus
$$\mathrm{E}(F^k) = \left(\frac{r_2}{r_1}\right)^k \mathrm{E}(U^k)\mathrm{E}(V^{-k}).$$
- Recall, again, for $X \sim \chi^2(r)$. If $k > -r/2$, then
$$\mathrm{E}(X^k) = \frac{2^k\Gamma(r/2+k)}{\Gamma(r/2)}.$$
- We have

$\mathrm{E}(F) = \frac{r_2}{r_1} r_1 \frac{2^{-1}\Gamma(\frac{r_2}{2}-1)}{\Gamma\frac{r_2}{2}} = \frac{r_2}{r_2-2}$, being large when r_2 is large.

Facts on F-Distributions

1. If $X \sim F_{r_1,r_2}$, then $1/X \sim F_{r_2,r_1}$.
2. If $X \sim t_n$, then $X^2 \sim F_{1,n}$.
3. If $X \sim F_{r_1,r_2}$, then

$$\frac{\frac{r_1}{r_2}X}{1+\frac{r_1}{r_2}X} \sim \mathrm{Beta}\left(\frac{r_1}{2}, \frac{r_2}{2}\right).$$