Statistics

2023 Lectures
Part 6 - Selected Families of Distributions

Institute of Economic Studies
Faculty of Social Sciences
Charles University

Bernoulli trials and Alternative (0-1) distribution

- Bernoulli trials refer to independent repetitions of some experiment in which we are interested in an event A (success) that occurs in each trial with the same probability p, while A^c denotes the failure with probability $1-p$
- $X = \begin{cases} 1 \text{ if } A \text{ occurs with } P(X=1) = p \\ 0 \text{ if } A^c \text{ occurs with } P(X=0) = 1-p = q \end{cases}$
- we write $X \sim ALT(p)$
- $E(X) = p$ and $VarX = pq$
- since $0^n = 0$ and $1^n = 1$,

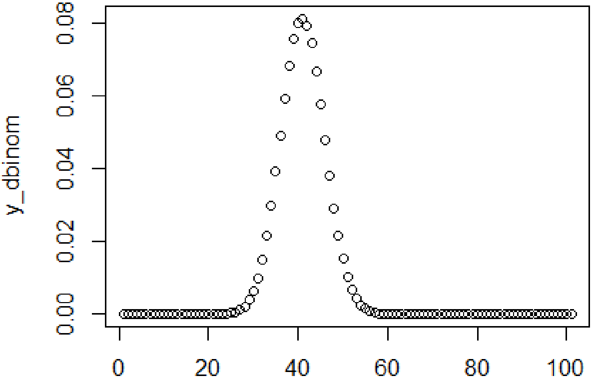
$$m_n = E(X^n) = E(X) = p, \quad \forall n \geq 1$$
$$m_X(t) = E(e^{tX}) = e^{t\cdot 0}q + e^{t\cdot 1}p = pe^t + q, \quad t \in \mathbb{R}.$$

Binomial distribution

- plays central role in probability theory and statistics and refers to a total number of successes in n Bernoulli trials
- $S_n = X_1 + \cdots + X_n$, where $X_i \sim ALT(p)$ independent and we write $S_n \sim BIN(n, p)$
- $$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, \dots, n.$$
- $E(S_n) = np$ and $VarS_n = npq$
- $$m_{S_n}(t) = (pe^t + q)^n, \ \ t \in \mathbb{R}.$$
- if $S_n \sim BIN(n, p)$ then $S_n^{'} = n - S_n \sim BIN(n, 1-p)$ useful for cases when “success” and “failure” is reversed

Binomial distribution: Example

Example 61: Consider $X \sim BIN(100, 0.4)$.
The graph of the probability mass function



Binomial distribution

Example 62: Assume that about one birth in 80 is a twin birth. What is the probability that there will be no twins births among the next 200 births in the maternity ward of a given hospital?

Let $S_{200} \sim BIN(200, 1/80)$ denotes the number of twin births among the next 200 births, we need to compute $P(S_{200} = 0)$.

$$P(S_{200} = 0) = \binom{200}{0} \left(\frac{1}{80}\right)^0 \left(\frac{79}{80}\right)^{200} \approx 0.0808$$
$$= \left(1 - \frac{200/80}{200}\right)^{200} \approx e^{-200/80} \approx 0.0821.$$

The exponential approximation is based on

$$\lim_{n \to \infty} \left(1 - \frac{c}{n}\right)^n = e^{-c}$$

which works well if the probability of success p is small and number of trials n is large.

Geometric distribution

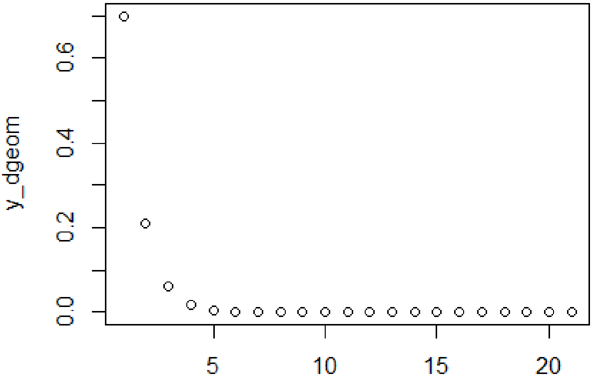
- refers to number of failures preceding the first success
- $P(X = k) = q^k p, \quad k = 0, 1, 2, \ldots$ and we write $X \sim GEO(p)$
- geometric because $P(X \geq k) = q^k p + q^{k+1} p + \cdots = q^k$
- also the distribution of rv $Y = X + 1$ (number of trials up to and including the first success) is called geometric
- $$F_X(x) = \begin{cases} 0, & x < 0; \\ \sum_{0 \leq k \leq x} q^k p, & 0 \leq x < \infty. \end{cases}$$
- $E(X) = \frac{q}{p}$ and $VarX = \frac{q}{p^2}$
- $$m_X(t) = \frac{p}{1 - qe^t} \quad \text{for } qe^t < 1 \text{ (else it is } +\infty).$$

Theorem 33: (memoryless property of $GEO(p)$)
Geometric distribution has the memoryless property, i.e., for $X \sim GEO(p)$ and any $n, m \in \mathbb{N} \cup \{0\}$, one has

$$P(X \geq m + n | X \geq m) = P(X \geq n).$$

Geometric distribution: Example

Example 63: Consider $X \sim GEO(0.7)$.
The graph of the probability mass function

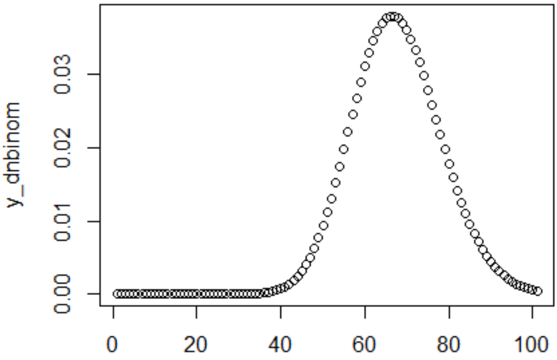


Negative binomial distribution

- rather than asking about the number of trials up to the first success, we ask about number of Bernoulli trials up to and including rth success
- the event $\{Y = n\}$ occurs if the nth trial is success and the first $n-1$ trials give exactly $r-1$ successes. Thus
$$P(Y = n) = \binom{n-1}{r-1} p^r q^{n-r}, \;\; n = r, r+1, \ldots$$
and we write $Y \sim NB(r, p)$.
- also rv $X = Y - r$, the number of failures preceding the rth success, is referred to as having a negative binomial distribution
- $E(Y) = \frac{r}{p}$ and $VarY = \frac{rq}{p^2}$
- $$m_Y(t) = \frac{p^r e^{tr}}{(1-qe^t)^r} \text{ for } qe^t < 1 \text{ (else it is } +\infty).$$

Negative distribution: Example

Example 64: Consider $X \sim NB(100, 0.6)$.
The graph of the probability mass function for Y which counts number of failures before a target number of successes is reached



Negative binomial distribution

Theorem 34: Let $S_n \sim BIN(n,p)$ and $Y_r \sim NB(r,p)$. Then for every $k = 0, 1, \ldots$

$$P(Y_r > k) = P(S_k < r).$$

Example 65: A salesman calls prospective buyer to make a sales pitch. Assume that the outcomes of consecutive calls are independent, and that on each call has 15% chance of making a sale. His daily goal is to make 3 sales, and he can make only 20 calls in a day. What is the probability that he does not achieve his daily goal?

Hypergeometric distribution

- from population of N elements with A successes sample n elements without replacement and observe the number of successes
- for max$\{0, n-(N-A)\} \leq k \leq$ min$\{n, A\}$

$$P(X=k)=\frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}$$

- to show that $\sum_k P(X=k)=1$ it suffices to compare coefficients in

$$(1+x)^N=(1+x)^A(1+x)^{N-A}$$

- $E(X)=n\frac{A}{N}$ and $VarX=n\frac{A}{N}\frac{N-A}{N}(1-\frac{n-1}{N-1})$

Properties of hypergeometric distribution

Theorem 35: Let $N \to \infty$, $A \to \infty$ and $\frac{A}{N} \to p, 0 < p < 1$. Then for every fixed n and $k = 0, 1, \ldots, n$,

$$P(X = k) \longrightarrow_{N\to\infty} \binom{n}{k} p^k (1-p)^{n-k}.$$

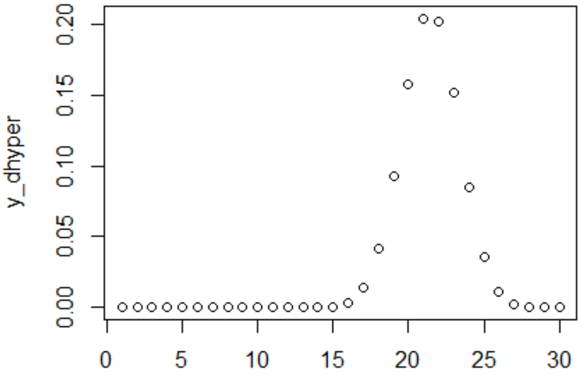
- I.e., the hypergeometric distribution converges to binomial if the size of the finite population increases

Theorem 36: Let X and Y be independent rv's with binomial distributions $X \sim BIN(m, p)$ and $Y \sim BIN(n, p)$. Then

$$P(X = k | X + Y = r) = \frac{\binom{m}{k}\binom{n}{r-k}}{\binom{m+n}{r}}.$$

Hypergeometric distribution: Example

Example 66: Consider X with hypergeometric distribution with parameters $N = 70$, $A = 50$ and $n = 30$.
The graph of the probability mass function



Hypergeometric distribution

Example 67: A class consists of 10 boys and 12 girls. The teacher selects 6 children at random. Is it more likely that she chooses 3 boys and 3 girls or that she chooses 2 boys and 4 girls?

- A generalization: Pólya urn scheme:
  Urn (population) consists of A balls (elements) of one kind and $N-A$ elements of another kind. Each time a ball is sampled, put back and c balls of the same kind is added to the urn.
- if $c=0$ then we have binomial distribution
- if $c=-1$ then we have hypergeometric distribution

Poisson distribution

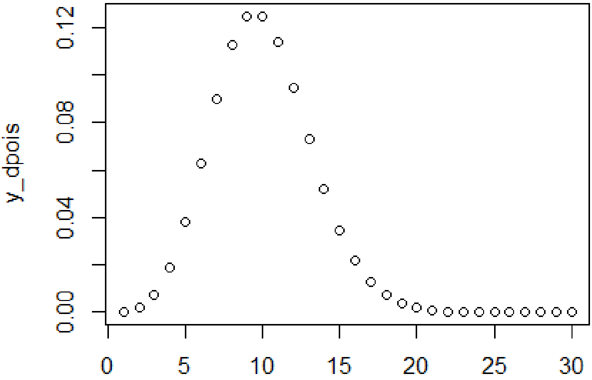
- a rv X is said to have a Poisson distribution if for some $\lambda > 0$,
$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \;\; k = 0, 1, \ldots$$
and we write $X \sim POI(\lambda)$
- a limit case of the binomial distribution

Theorem 37:
If $p \to 0, n \to \infty$ and $np \to \lambda > 0$ then
$$\lim \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

- $E(X) = \lambda$ and $VarX = \lambda$
- $m_X(t) = e^{\lambda(e^t - 1)}, \;\; t \in \mathbb{R}$

Poisson distribution: Example

Example 68: Consider $X \sim POI(10)$.
The graph of the probability mass function



Properties of Poisson distribution

- Poisson distribution is closed under addition of independent random variables.

Theorem 38: If $X \sim POI(\lambda_1)$ and $Y \sim POI(\lambda_2)$ are independent, then

$$X + Y \sim POI(\lambda_1 + \lambda_2).$$

- if X and Y are independent with Poisson distribution then conditional distribution of X given $X + Y$ is binomial

Theorem 39: If $X \sim POI(\lambda_1)$ is independent of $Y \sim POI(\lambda_2)$ then for $k = 0, \ldots, n$

$$P(X = k | X + Y = n) = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k}.$$

Application of Poisson distribution

- number of occurrences of some event in a given time interval (arrivals of customers at service stations; fire alarms in given time; number of phone calls on call centers etc.)
- recall that we say that f is $o(x)$ if $\lim_{x\to 0} \frac{f(x)}{x} = 0$.
  - if $\lim_{x\to 0} \frac{h(x)}{x} = c$ then $h(x) = cx + o(x)$.
  - if $f_1, \ldots, f_N$ are $o(x)$ then also $f_1 + \cdots + f_N$ is $o(x)$.
- Assumption 1: The number of events occurring in two non-overlapping time intervals are independent
- Assumption 2: The probability of at least one event occurring in an interval of length Δt is $\lambda \Delta t + o(\Delta t)$ for some constant $\lambda > 0$
- Assumption 3: The probability of two or more events occurring in an interval of length Δt is $o(\Delta t)$

Application of Poisson distribution

- Let $N[0,t)$ be the number of events prior time t and set $P_n(t) := P(N[0,t) = n)$, the probability of exactly n events prior time t. Then under assumptions 1, 2 and 3

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \ldots.$$

Example 69: Suppose that traffic accidents on a given intersection occur according to a Poisson process, with the rate on Saturday being twice the rate on weekdays and the rate on Sunday being double the rate on Saturdays. The total rate is about five accidents per week. What is the probability of four accidents on weekdays in a given week?

Uniform distribution

- we write $X \sim U[A, B]$, $A < B$
- density function

$$f_X(x) = \begin{cases} \frac{1}{B-A}, & \text{if } A \leq x \leq B; \\ 0, & \text{otherwise.} \end{cases}$$

- distribution function

$$F_X(t) = \begin{cases} 0, & \text{if } t < A; \\ \frac{t-A}{B-A}, & \text{if } A \leq t \leq B; \\ 1, & \text{if } t > B. \end{cases}$$

- $E(X) = \frac{A+B}{2}$, $VarX = \frac{(B-A)^2}{12}$
- moment generating function

$$m_X(t) = \begin{cases} \frac{e^{Bt}-e^{At}}{t(B-A)}, & t \neq 0; \\ 1, & t = 0. \end{cases}$$

Exponential distribution

- describes time between events in a Poisson process; used for modeling a random time of occurrence of some event or length of occurrence, e.g. life expectancy of some device, modeling of response times of computer servers, etc.
- we denote $X \sim EXP(\lambda)$, $\lambda > 0$.
- cumulative distribution function

$$F_X(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

- $E(X) = \frac{1}{\lambda}$, $VarX = \frac{1}{\lambda^2}$
- moment generating function

$$m_X(t) = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda \text{ (else it is } +\infty).$$

Theorem 40: (memoryless property of $EXP(\lambda)$)
If X is a random variable with exponential distribution then for all $x, y > 0$

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

Memory-less property

Example 70: Studies of a single-machine-tool system showed that the mean time the machine operates before breaking down is 10 hours.

i) Determine the failure rate and find the probability that the machine operates for at least 12 hours before breaking down

ii) If the machine has been operating 8 hours, what is the probability that it will last another 4 hours?

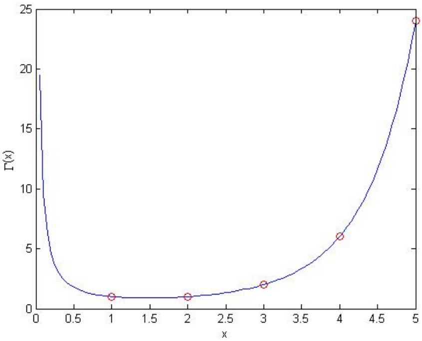
Gamma function

Definition 27: For $t \geq 0$ we define gamma function as

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx.$$

Properties:

- $\Gamma(1) = \Gamma(2) = 1$
- $\Gamma(t+1) = t\Gamma(t)$ for $t \in \mathbb{R}_+$
- $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$



Gamma function is a natural extension of factorial function to positive real numbers

Gamma distribution

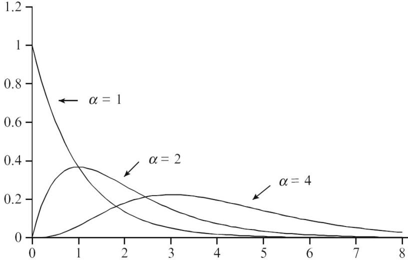
- we write $X \sim GAM(\alpha, \lambda)$
- density function

$$f(x) = \begin{cases} Cx^{\alpha-1}e^{-\lambda x}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

- α is the shape parameter and λ is the scale parameter
- normalizing constant $C = \frac{\lambda^\alpha}{\Gamma(\alpha)}$
- $E(X) = \frac{\alpha}{\lambda}, VarX = \frac{\alpha}{\lambda^2}$
- moment generating function

$$m_X(t) = \frac{1}{(1 - \frac{t}{\lambda})^\alpha} \quad \text{for } t < \lambda$$

- $m_k = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\lambda^k}$



Gamma distributions, $\lambda = 1$.

Properties of Gamma distribution

Theorem 41: If $X \sim GAM(\alpha_1, \lambda)$ and $Y \sim GAM(\alpha_2, \lambda)$ are independent then $X + Y \sim GAM(\alpha_1 + \alpha_2, \lambda)$.

Theorem 42: If $X \sim GAM(\alpha, \lambda)$ then $Y = 2\lambda X \sim GAM(\alpha, \frac{1}{2})$.

Special cases:

- for $\alpha = 1$, $GAM(1, \lambda) = EXP(\lambda)$
- let $Y = Z^2$ and $Z \sim N(0, 1)$. Then

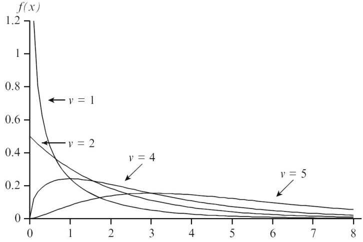
$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y \geq 0; \\ 0, & \text{otherwise}, \end{cases}$$

and hence $Y \sim GAM\left(\frac{1}{2}, \frac{1}{2}\right)$.

χ^2 distribution

Definition 28: For integer ν, the distribution $GAM\left(\frac{\nu}{2}, \frac{1}{2}\right)$ is called chi-square distribution with ν degrees of freedom. We write $X \sim \chi^2_\nu$.

- let $Y = Z^2$ and $Z \sim N(0,1)$. Then $Y \sim \chi^2_1$.



Chi-square distributions.

Theorem 43:

a) If $Z_1, \ldots, Z_n$ are iid with $N(0,1)$ then

$$X = Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n.$$

b) If $X_1, \ldots, X_n$ are independent rv's with $\chi^2_{\nu_1}, \ldots, \chi^2_{\nu_n}$, respectively, then

$$Y = X_1 + \cdots + X_n \sim \chi^2_\nu, \quad \text{with } \nu = \nu_1 + \cdots + \nu_n.$$

- χ^2 distribution is very important for testing hypotheses

Normal distribution

- the single most important distribution in the probability theory and statistics; important also in diverse areas of applications, e.g., normal distribution fits data on heights and weights of human and animal populations
- we denote $X \sim N(\mu, \sigma^2)$
- density function

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

- $E(X) = \mu, VarX = \sigma^2$
- moment generating function

$$m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, \quad t \in \mathbb{R}.$$

Theorem 44: Let X and Y be independent rv's with $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Then $U = \alpha X + \beta Y$ has distribution $N(\alpha\mu_1 + \beta\mu_2, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2)$.

Normal distribution

Example 71: Assume that height of men in a certain population is normal with mean $\mu_M = 175$ centimeters and standard deviation $\sigma_M = 5$ centimeters. The height of women is also normal, with mean $\mu_W = 170$ centimeters and $\sigma_W = 4$ centimeters.
One man and one woman are selected at random. What is the probability that the woman selected is taller than the man selected?

Example 71 cont.: Chances of $X \sim N(175, 25)$ being negative are of order $\Phi(-35) = 1 - \Phi(35) < 10^{-100}$.

Lognormal distribution

- widely applicable in modeling, for example, where there is a multiplicative product of many small independent factors, e.g., the long-term return rate on a stock investment can be considered as the product of the daily return rates

Definition 29: X has lognormal distribution if $Y = \log X$ has normal distribution.

- density function

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \quad x > 0.$$

- $E(X) = e^{\mu + \frac{\sigma^2}{2}}, VarX = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$

Example 72: If $Y \sim N(0, 1)$ and $Y = \log X$ then X has lognormal distribution with $E(X) = e^{\frac{1}{2}}$ and not $e^{E(Y)} = 1$!

Expectation of random vectors

- For a random vector X we can extend the definition of expected value to a vector of expected values and variance to a variance matrix of a random vector.

Definition 30: Let $X = (X_1, \ldots, X_n)^\top$ be a random vector. Then

a) the expected value of X is a vector

$$E(X) = (E(X_1), \ldots, E(X_n))^\top$$

b) if $E(X_i^2) < \infty, i = 1, \ldots, n$, then $VarX$, the variance matrix of X, is defined as

$$\begin{aligned} VarX = & \begin{pmatrix} VarX_1 & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_1, X_2) & VarX_2 & \cdots & Cov(X_2, X_n) \\ \vdots & & \ddots & \vdots \\ Cov(X_1, X_n) & Cov(X_2, X_n) & \cdots & VarX_n \end{pmatrix} \\ = & E((X - EX)(X - EX)^\top). \end{aligned}$$

Variance matrix and correlation matrix

c) if $E(X_i^2) < \infty, i = 1, \ldots, n$, then $CorrX$, the correlation matrix of X, is defined as

$$CorrX = \begin{pmatrix} 1 & Corr(X_1, X_2) & \cdots & Corr(X_1, X_n) \\ Corr(X_1, X_2) & 1 & \cdots & Corr(X_2, X_n) \\ \vdots & & \ddots & \vdots \\ Corr(X_1, X_n) & Corr(X_2, X_n) & \cdots & 1 \end{pmatrix}.$$

Theorem 45: (without proof) (Hőlder inequality)
Let $E|X|^p < \infty$ and $E|Y|^q < \infty, p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$. Then

$$E(XY) \leq (E|X|^p)^{\frac{1}{p}} (E|Y|^q)^{\frac{1}{q}}.$$

Special case: $E(XY) \leq \sqrt{E(X^2)E(Y^2)}$.

Properties of variance matrix

Theorem 46: (without proof)

Let $VarX_i < \infty, i = 1, \ldots, n$. Then

i) the variance matrix of X is symmetric and positive semidefinite;

ii) for any $a = (a_1, \ldots, a_n)^\top$ and $B = (b_{ij})_{m\times n}$

$$Var(a + BX) = B\,VarX\,B^\top;$$

iii) if $X_1, \ldots, X_n$ are mutually independent, or at least uncorrelated, then $VarX$ is a diagonal matrix;

iv) if $VarX$ is singular then there exist nonzero real constants $a_1, \ldots, a_{n+1}$ and $Q \subset S$ with $P(Q) = 1$ such that

$$\sum_{i=1}^{n} a_i X_i(s) = a_{n+1}$$

for all $s \in Q$.

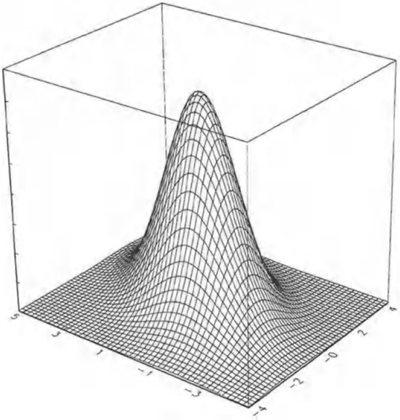
Bivariate normal distribution

- $X \sim N_2\left(\left(\begin{array}{c}\mu_1\\ \mu_2\end{array}\right), \left(\begin{array}{cc}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{array}\right)\right)$, where $\mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 > 0, \rho \in [-1, 1]$.
- joint density function

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}$$

- both X_1 and X_2 have marginal normal distribution $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively
- $E(X_1) = \mu_1, E(X_2) = \mu_2$
- $Var(X_1) = \sigma_1^2, Var(X_2) = \sigma_2^2, Corr(X_1, X_2) = \rho$

Bivariate normal distribution



Example 71 cont.: Assume the heights of siblings jointly normally distributed and additionally correlated with $\rho = 0.6$. If we sample a brother and a sister, what is the probability that the sister is taller than her brother?