Purdue-NCKU program

Lecture 1
Review of distribution theory

Dr. Qifan Song

Random Variable (r.v.)

- Random variables are the *numerical* outcomes of random events
  - examples:
    - number of heads in a coin flipping experiment
    - quality score of a batch of products
  - Plays an important role in statistical theory
  - Source of randomness: measurement error, sampling, treatment variation...
- *Random* is not arbitrary.

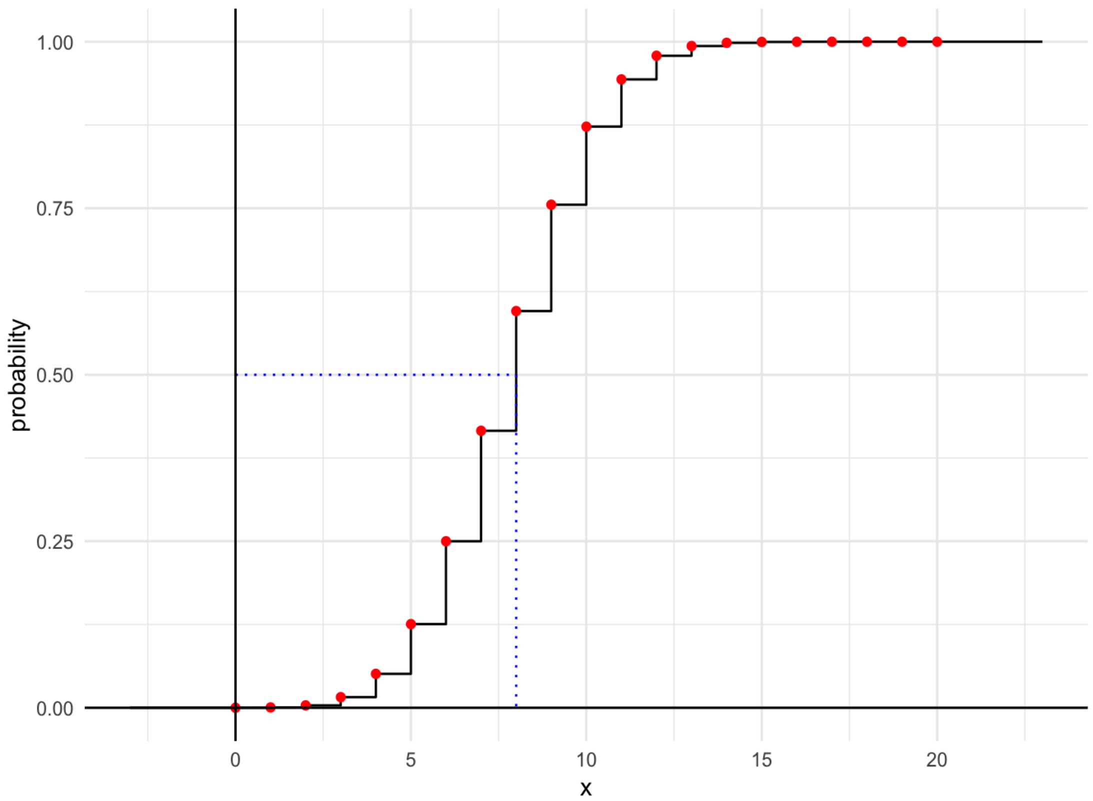
Characterize the randomness

Cumulative Distribution Function of r.v. X

- $F_X(t) = P(X \leq t)$.
- Properties
  - $\lim_{t\to\infty} F_X(t) = 1$ and $\lim_{t\to-\infty} F_X(t) = 0$
  - F_X is non-decreasing
  - F_X is right-continuous
- Immediate result
  - $P(a < X \leq b) = F_X(b) - F_X(a)$
  - $P(X = a) = F_X(a) - \lim_{t\uparrow a} F_X(t)$

Types of CDF

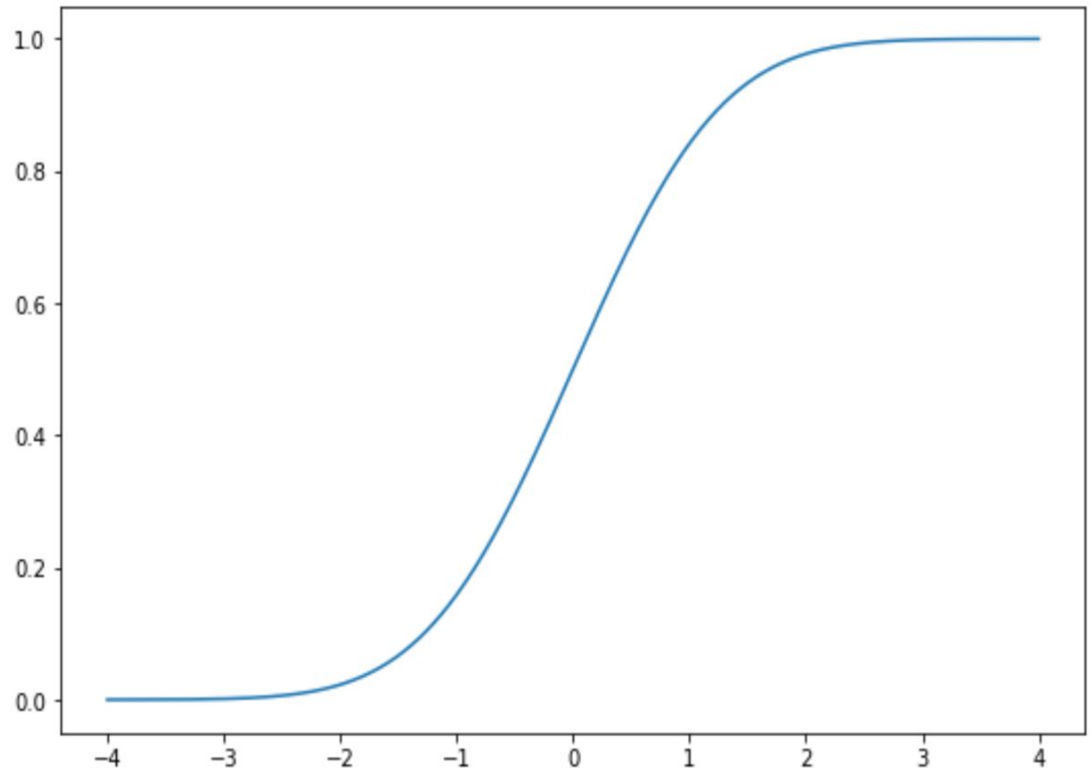
Step function – discrete random variable



Piece-wise constant function, with countable-many jump points

Types of CDF

Continuously differentiable function − continuous random variable



Smoothly increasing function

Discrete Random Variable

- Jump points: $x_1, \ldots, x_n, \ldots$
- Jump heights: $p_1, \ldots, p_n, \ldots$, with $\sum p_i = 1$
- Probability Mass function (PMF): $Pr(X = x_i) = p_i$

Expected value of h(X):

$$Eh(X) = \sum p_i h(x_i),$$

given $\sum p_i |h(x_i)| < \infty$.

Mean value of X, i.e., $E(X)$ measures the average outcomes. It reflects the center of the distribution *given the dispersion of the distribution is under control*.

For example: X with PMF $Pr(X = n) = 3/[n^2\pi^2]$ for all non-zero integer n doesn't have a mean.

Unifrom Discrete Distribution

$$P(X = x_i) = \text{constant}$$

- Equal-chance outcomes
- Counting problem: $P(X \in A) = |A|/|\mathcal{X}|$

Bernoulli-related Distributions

$$P(X = 1) = p \text{ and } P(X = 0) = 1 - p$$

related distribution

- repeated Bernoulli experiment with the same p
- Binomial/ Negative Binomial/ Geometric distributions
- Closeness under addition

Poisson Distribution

$$P(X = n) = \lambda^n e^{-\lambda}/n! \text{ for all non-negative integer } n$$

- Closeness under addition
- Used to model rare events: how many 100-year floods can occur during 100 years?
- Used to model point processes: how many customers within a time interval

Continuous Random Variable

- $Pr(X = x) = 0$ for any x
- $Pr(X \in x \pm \epsilon) > 0$ if $F'_X(x) \neq 0$
- $Pr(X \in [a,b]) = F_X(b) - F_X(a) = \int_a^b F'_X(t)dt$
- probability density function (pdf) $f_X(t) = F'_X(t)$.

Expected value of $h(X)$

$$Eh(X) = \int h(t) f_X(t) dt,$$

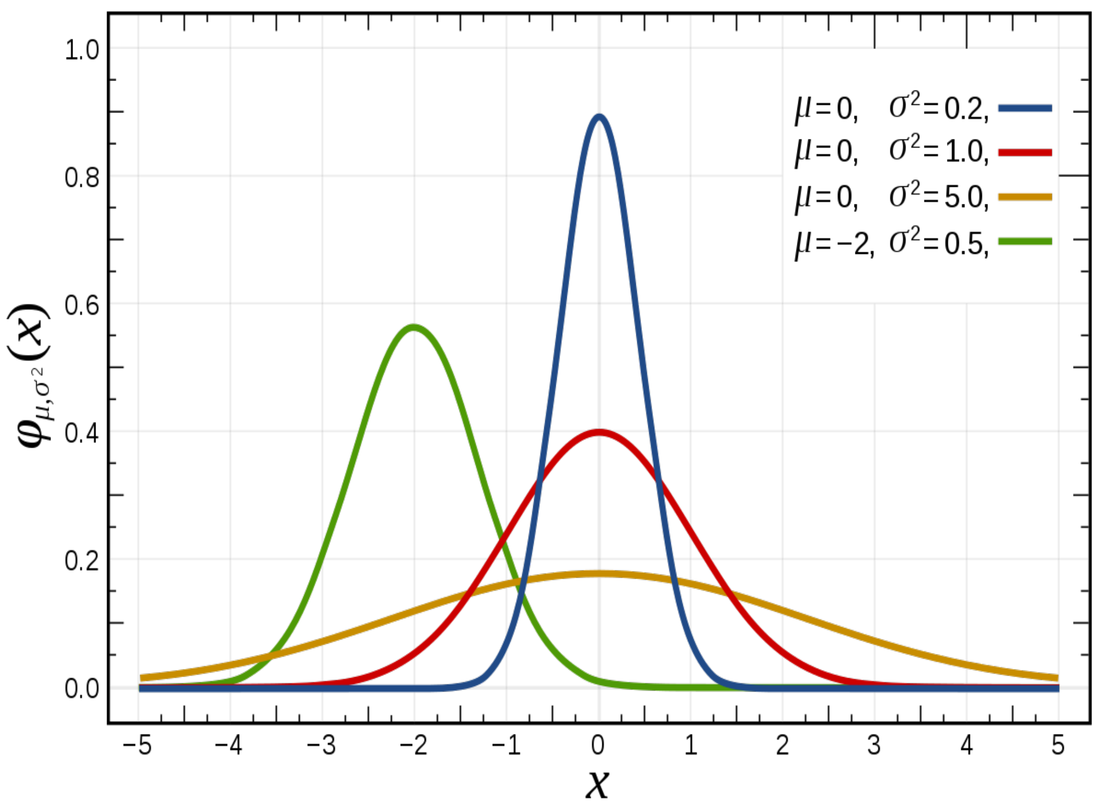
given $\int |h(t)| f_X(t) dt < \infty$.

Example: Cauchy r.v. with pdf $f_X(x) = 1/[\pi(1+x^2)]$ for all $x \in R$ doesn't have a mean.

Normal Distribution

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

- Mean is μ and variance is σ^2
- Bell shaped distribution density curve
- One of the most commonly used distribution to model random noise
- Closeness to addition (even without independence)

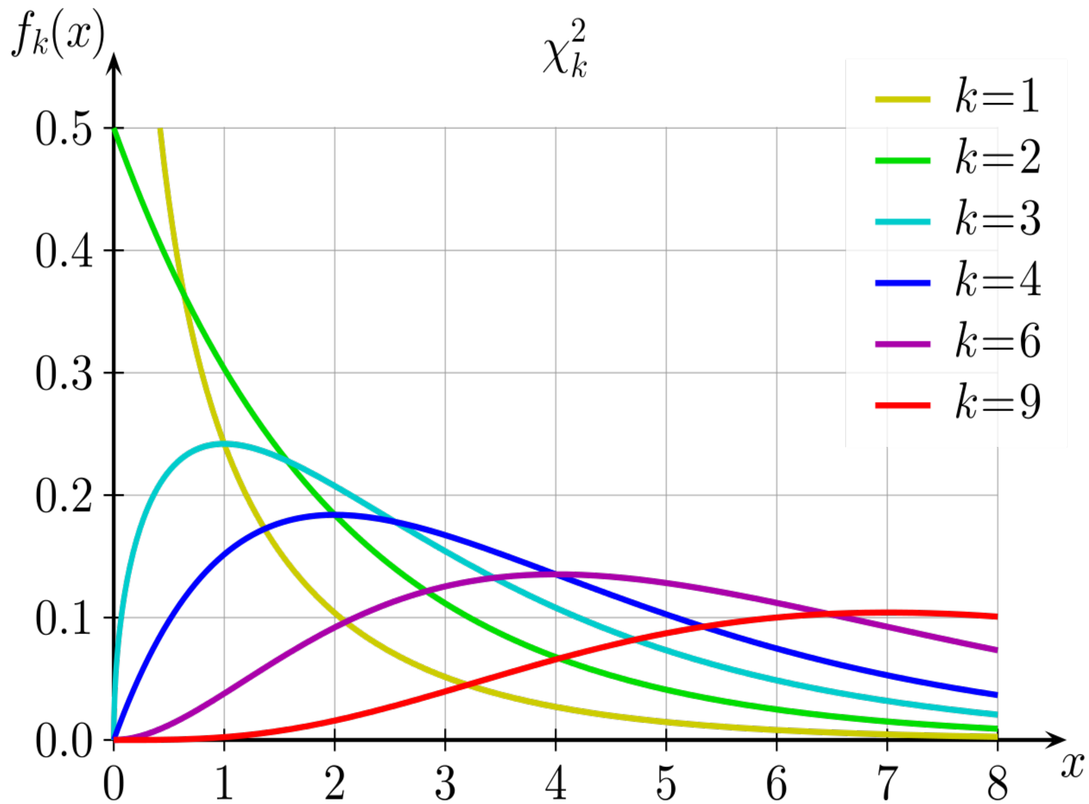


$\mu=0,\quad \sigma^2=0.2,$
$\mu=0,\quad \sigma^2=1.0,$
$\mu=0,\quad \sigma^2=5.0,$
$\mu=-2,\quad \sigma^2=0.5,$
$\varphi_{\mu,\sigma^2}(x)$
x
1.0
0.8
0.6
0.4
0.2
0.0
−5 −4 −3 −2 −1 0 1 2 3 4 5

χ_k^2 distribution

Definition: the sum of k independent squared standard normal random variables

- distribution of non-negative values
- the definition can generalized to non-integer k
- Mean is k and variance is $2k$
- often relates to the inferences about variance parameters
- reduces to exponential distribution if $k = 2$



$f_k(x)$
χ_k^2
$k=1$
$k=2$
$k=3$
$k=4$
$k=6$
$k=9$
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
7
8
x

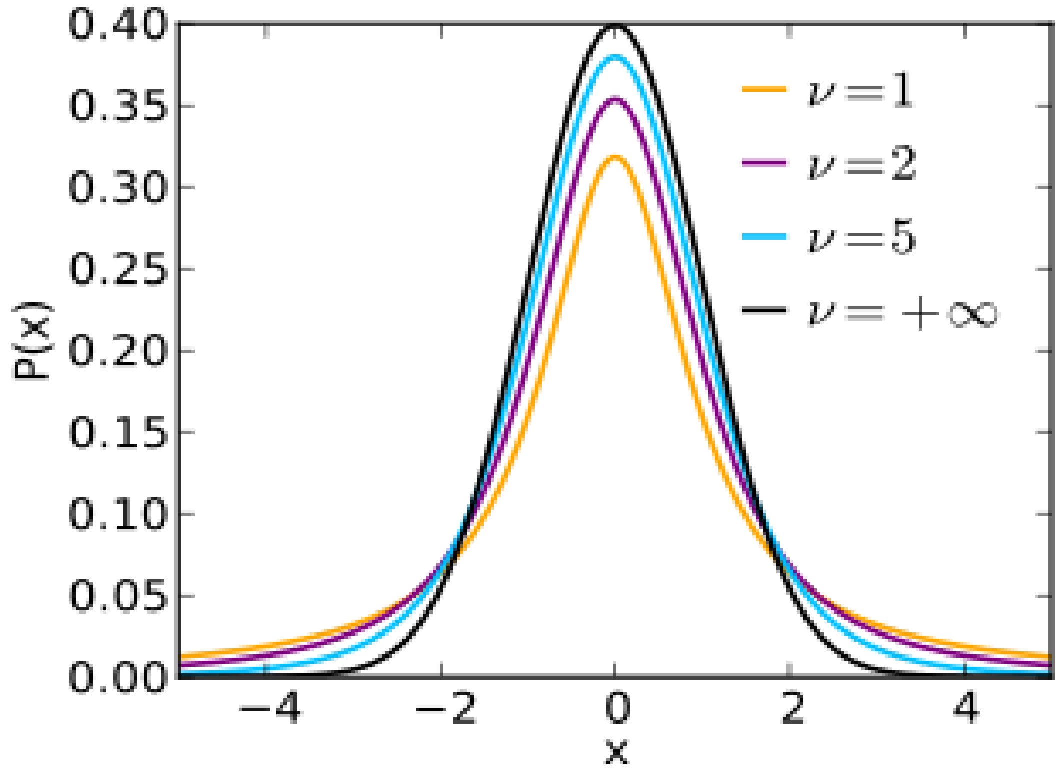
t distribution

Definition:

$$t_k = \frac{N(0, 1^2)}{\sqrt{\chi_k^2 / k}}$$

for indep normal and chi-square variables

- reduce to Cauchy distribution if $k = 1$
- reduce to Normal distribution if $k = \infty$
- Similar bell shaped distribution as Normal, but totally different tail behavior
- Commonly used for making inference about mean parameters under unknown variance.



0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu = 1$
$\nu = 2$
$\nu = 5$
$\nu = +\infty$

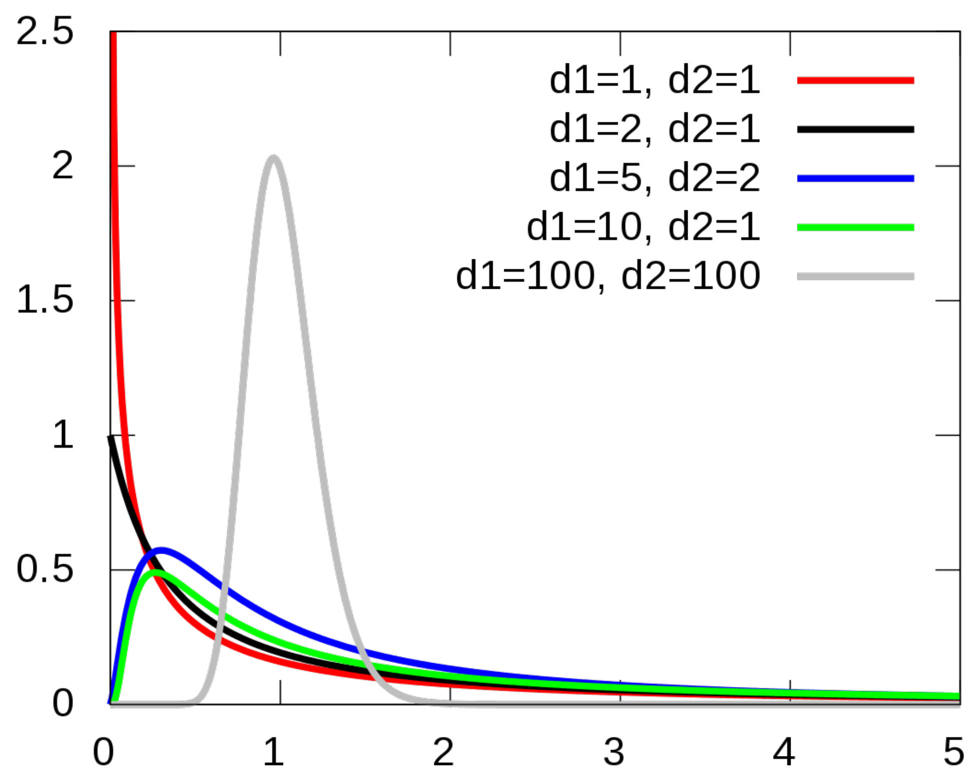
F distribution

Definition:

$$F_{m,n} = \frac{\chi^2_m/m}{\chi^2_n/n}$$

for two indep chi-square variables

- Relationship to t distribution $F_{1,k} = t_k^2$
- Commonly used for making inference about fitting or variance comparison


d1=1, d2=1
d1=2, d2=1
d1=5, d2=2
d1=10, d2=1
d1=100, d2=100
2.5
2
1.5
1
0.5
0
0
1
2
3
4
5

Multiple Variables

Given d random r.v.s $X_1, ..., X_d$

- Joint CDF $F(t_1, \ldots, t_d) = Pr(X_i \leq t_i$ for all $i)$
- Joint pdf $f(t_1, \ldots, t_d) = \partial^d F / \partial t_1 \ldots \partial t_d$
- Joint pmf $p(t_1, \ldots, t_d) = Pr(X_i = t_i$ for all $i)$
- Expected value

$$Eh(X_1, ..., X_d) = \int h(t_1, \ldots, t_d) f(t_1, \ldots, t_d) dt_1 \ldots dt_d$$

- Coherence between joint and marginal distribution definitions

- Conditional distribution = joint distribution / marginal distribution

$$f(x_1, \ldots, x_p | x_{p+1}, \ldots, x_d) = f(x_1, \ldots, x_d) / f(x_{p+1}, \ldots, x_d)$$

Independence and Covariance

- Independence: $f(t_1, t_2) = f(t_1) \times f(t_2)$ or $f(t_1, t_2) \propto g(t_1) \times h(t_2)$

- Covariance: $Cov(X_1, X_2) = E\{[X_1 - E(X_1)]\}\{[X_2 - E(X_2)]\}$

- $Var(X_1) = Cov(X_1, X_1)$

- Correlation: $Cor(X_1, X_2) = Cov(X_1, X_2)/\sqrt{Var(X_1)Var(X_2)}$

- $Cor(X_1, X_2) = \pm 1$ means $X_1 = aX_2 + b$ for non-zero a

- Independence implies 0 covariance

- 0 covariance doesn't imply independence. E.g. $f(x_1, x_2) \propto \exp\{-\sqrt{x_1^2 + x_2^2}\}$

Sum of Variables

- $X_1 - X_2 = X_1 + (-X_2)$
- $E(\sum X_i) = \sum E(X_i)$
- $Var(\sum X_i) = \sum Var(X_i) + 2\sum_{i \neq j} Cov(X_i, X_j)$

- Sample mean of i.i.d. r.v.'s X_i: $\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i$
- $E(\bar{X}) = E(X_i)$
- $Var(\bar{X}) = var(X_i)/n$

Large sample properties

Sample mean of i.i.d. r.v.'s X_i: $\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i$

- Law of Large Numbers:
  If $E(X_i) = \mu$, then as n increases, $\bar{X}$ converges (stochastically) towards μ.

- Sample mean is a good guess for population mean, when n is large; t distribution become normal distribution as k increases

- Central Limit Theory
  If $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, then as n increases, the distribution of $\sqrt{n}(\bar{X} - \mu)/\sigma$ converges to the distribution of $N(0, 1^2)$.

- probability calculation about $\bar{X}$ can be approximated by CLT:
  $\bar{X} \approx N(\mu, \sigma^2/n)$

Vector/Matrix Form

Random Vector $X = (X_1, \ldots, X_d)^\top$

- $E(X) = (E(X_1), \ldots, E(X_d))'$
- Covariance Matrix $Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]^\top\} = [Cov(X_i, Y_j)]_{i,j}$
- $Var(X) = Cov(X, X)$

Vector/Matrix form handles linear combination well

- $E(AX) = AE(X)$ where A is constant matrix or row
- $Cov(AX, BY) = ACov(X, Y)B^\top$ where A and B are constant matrix or row
- Recovers results in previous slides

Multivariate Normal Distribution $N(\mu, \Sigma)$

$X \in R^d$ with pdf

$$f(x) = (2\pi)^{-d/2}[\det(\Sigma)]^{-1/2} \exp\{-(x-\mu)^\top \Sigma^{-1}(x-\mu)/2\}$$

- $E(X) = \mu$
- $Cov(X) = \Sigma$
- Any marginal or conditional distributions are still normal distributions
- For Normal distributions, no correlation = independence (since 0 covariance leads to a proper factorization of the pdf)
- Connect to χ^2 distribution: if $X \sim N(0, \sigma^2 I_d)$, then $X^\top X/\sigma^2 = \chi^2_d$

Multivariate Normal Distribution $N(\mu, \Sigma)$

- Any linear mapping of an multivariate normal r.v. is still normally distributed
- $AX + b \sim N(A\mu + b, A\Sigma A^\top)$
- What is the distribution of sample mean of i.i.d. univariate normal r.v.'s?
- AX and BX are independent iff $A\Sigma B^\top = 0$

Distribution of sample mean/variance of iid normal

Let X_1, ..., X_n be iid normal r.v.s following $N(\mu, \sigma^2)$ distribution. Then what is the joint distribution of sample mean $\bar{X}$ and sample variance $S^2 = \sum(X_i - \bar{X})^2/(n-1)$?

Ans: (i)$\bar{X} \sim N(\mu, \sigma^2/n)$, (ii)$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ and (iii) $\bar{X}$ and S^2 are independent

Higher level understanding

- Linear projection from data to the parameter space
- Orthogonal (hence independent) decomposition of data vector

- Similar results appear in more complicated settings (e.g., regression models)

Chapter Review

- Characterization of randomness
- Different type of distributions
- Mean, variance, expected value
- Normal and Normal related distribution
- Multivariate Normal distributions