COMP 680
Statistics for Computing and Data Science
Week 2: Probability and Random Variable II

Su Chen, Assistant Teaching Professor,
Rice D2K Lab

Outline

Expectation (mean)

- The expectation of a random variable X, denoted $\mathbb{E}[X]$
  - $\mathbb{E}[X] = \sum_i x_i \cdot \mathbb{P}(X = x_i)$ if X is discrete
  - $\mathbb{E}[X] = \int_{X \in \mathbb{R}} x \cdot f_X(x) dx$ if X is continuous

Expectation (mean)

- The expectation of a random variable X, denoted $\mathbb{E}[X]$
  - $\mathbb{E}[X] = \sum_i x_i \cdot \mathbb{P}(X = x_i)$ if X is discrete
  - $\mathbb{E}[X] = \int_{X\in\mathbb{R}} x \cdot f_X(x)dx$ if X is continuous
- For any function $g : \mathbb{R} \to \mathbb{R}$, the expectation of $g(X)$ is:
  - $\mathbb{E}[g(X)] = \sum_i g(x_i) \cdot \mathbb{P}(X = x_i)$ if X is discrete
  - $\mathbb{E}[g(X)] = \int_{X\in\mathbb{R}} g(x) \cdot f_X(x)dx$ if X is continuous

Expectation (mean)

- The expectation of a random variable X, denoted $\mathbb{E}[X]$
  - $\mathbb{E}[X] = \sum_i x_i \cdot \mathbb{P}(X = x_i)$ if X is discrete
  - $\mathbb{E}[X] = \int_{X \in \mathbb{R}} x \cdot f_X(x)dx$ if X is continuous
- For any function $g : \mathbb{R} \to \mathbb{R}$, the expectation of $g(X)$ is:
  - $\mathbb{E}[g(X)] = \sum_i g(x_i) \cdot \mathbb{P}(X = x_i)$ if X is discrete
  - $\mathbb{E}[g(X)] = \int_{X \in \mathbb{R}} g(x) \cdot f_X(x)dx$ if X is continuous
- Examples:

Properties of Expectation

- Linearity: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ for constants a and b and random variables X and Y.

Properties of Expectation

- Linearity: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ for constants a and b and random variables X and Y.
- Independence: If two random variables X and Y are independent, then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.
    - can you show this?

Properties of Expectation

- Linearity: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ for constants a and b and random variables X and Y.
- Independence: If two random variables X and Y are independent, then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.
    - can you show this?
- Intuition: the "typical" value of the random variable.

Variance and Standard Deviation

- The variance of a random variable X is defined to be the mean square deviation from the mean:

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

Variance and Standard Deviation

- The variance of a random variable X is defined to be the mean square deviation from the mean:

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

- The standard deviation σ_X of X is:

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

Variance and Standard Deviation

- The variance of a random variable X is defined to be the mean square deviation from the mean:

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

- The standard deviation σ_X of X is:

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

- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ for any constants a and b

Covariance

- The covariance of two random variables X and Y is defined to be:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Covariance

- The covariance of two random variables X and Y is defined to be:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- One can show: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$

Covariance

- The covariance of two random variables X and Y is defined to be:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- One can show: $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$
- If X and Y are independent:
  - $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$
  - $\mathrm{Cov}(X,Y) = 0$
  - $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Correlation Coefficient

- Pearson's correlation:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

  - $-1 \leq \rho_{X,Y} \leq 1$: a standardized measure

Correlation Coefficient

- Pearson's correlation:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

  - $-1 \leq \rho_{X,Y} \leq 1$: a standardized measure
- Describe the linear association between X and Y:
  - $\rho_{X,Y} = 1$: $Y = aX + b$ where $a = \sigma_Y / \sigma_X$
  - $\rho_{X,Y} = -1$: $Y = aX + b$ where $a = -\sigma_Y / \sigma_X$
  - $\rho_{X,Y} = 0$: no linear association or "uncorrelated"

Outline

Multinomial Distributions

- Multinomial distributions with parameters n and $\sum_{i=1}^{k} p_i = 1$

$$\mathbb{P}(X_1 = x_1, \cdots X_k = x_k) = \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

where x_i's are integers such that $\sum_{i=1}^{k} x_i = n$

Multinomial Distributions

- Multinomial distributions with parameters n and $\sum_{i=1}^{k} p_i = 1$

$$\mathbb{P}(X_1 = x_1, \cdots X_k = x_k) = \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

where x_i's are integers such that $\sum_{i=1}^{k} x_i = n$

- Binomial distribution is multinomial with $k = 2$

Gaussian Distributions

- Gaussian distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

  - mean vector $\boldsymbol{\mu}$: a vector of length k
  - covariance matrix $\boldsymbol{\Sigma}$: a $k \times k$ symmetric and positive definite matrix

Bivariate Normal Distribution

- Example of 2-dimensional Gaussian $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

  - mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$
  - covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

Bivariate Normal Distribution

- Example of 2-dimensional Gaussian $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

  - mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$
  - covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$
- 2-d random vector $\boldsymbol{X} = (X_1, X_2) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - $\mathbb{E}[X_1] = \mu_1$, $\mathbb{E}[X_2] = \mu_2$
  - $\text{Var}[X_1] = \sigma_1^2$, $\text{Var}[X_2] = \sigma_2^2$
  - $\text{Cov}[X_1, X_2] = \rho\sigma_1\sigma_2$

Properties of Gaussian Distribution

- Let $\boldsymbol{X}$ be a multivariate Gaussian random vector, then the following also follow Gaussian distribution:
  - The marginal distribution of all subsets of the components
  - The conditional distribution of the components
  - Linear combination of the components

Properties of Gaussian Distribution

- Let $\boldsymbol{X}$ be a multivariate Gaussian random vector, then the following also follow Gaussian distribution:
  - The marginal distribution of all subsets of the components
  - The conditional distribution of the components
  - Linear combination of the components
- Example $\boldsymbol{X} = (X_1, X_2) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:
  - $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim N_2(\boldsymbol{0}, \boldsymbol{I})$
  - $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$
  - $X_1 | X_2 = a \sim N(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(a - \mu_1), (1 - \rho^2)\sigma_1^2)$
  - For a constant vector a, $a^T \boldsymbol{X} \sim N(a^T \boldsymbol{\mu}, a^T \boldsymbol{\Sigma} a)$

Outline

1. Expectations, Variance and Covariance
2. Multivariate Distributions
3. Probability Inequalities
4. Transformation of Random Variables

Markov's Inequality

- Let X be a non-negative random variable and suppose that $\mathbb{E}[X]$ exists. For any $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

- Proof:

Chebyshev's Inequality

- Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$, then

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

  in particular taking $t = k\sigma$, then

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Proof:

Jensen's Inequality

- If g is a convex function then $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$; and if g is a concave function then $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$

Jensen's Inequality

- If g is a convex function then $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$; and if g is a concave function then $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$
- If a function g is a convex then it lies above any line that touches g at any point; g is concave if $-g$ is convex.

Jensen's Inequality

- If g is a convex function then $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$; and if g is a concave function then $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$
- If a function g is a convex then it lies above any line that touches g at any point; g is concave if $-g$ is convex.
- Examples:
  - $g(x) = x^2$ is convex, $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$
  - $g(x) = 1/x$ is convex for $x > 0$, $\mathbb{E}[1/X] \geq 1/\mathbb{E}[X]$
  - $g(x) = \log x$ is concave, $\mathbb{E}[\log X] \leq \log \mathbb{E}[X]$

Examples

- You flip a fair coin 100 time, give an upper bound of probability of getting 70 heads or more.

Examples

- You flip a fair coin 100 time, give an upper bound of probability of getting 70 heads or more.
- How accurate is this estimate?
    - $X \sim \text{Binomial}(n = 100, p = 0.5)$
    - estimate $\mathbb{P}(X \geq 70)$
    - apply Markov's Inequality:

Examples

- You would like to know the percentage p in general population that support certain legislation. You start a poll online to randomly survey 100 people for a Yes/No question and you estimate the proportion of Yes in the survey response.

Examples

- You would like to know the percentage p in general population that support certain legislation. You start a poll online to randomly survey 100 people for a Yes/No question and you estimate the proportion of Yes in the survey response.
- How accurate is this estimate?
  - $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
  - estimate p by $\bar{X}_n = \sum_{i=1}^{n} X_i/n$
  - apply Chebyshev's Inequality:

Outline

Functions of Random Variables

- Let X be a random variable and $Y = g(X)$ be another random variable defined by a function $g : \mathbb{R} \to \mathbb{R}$.
- The cdf of Y is given by:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_{\{x|g(x)\leq y\}} f_X(x)dx$$

- If g is a strictly increasing function, then

$$F_Y(y) = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = F_X(g^{-1}(y))$$

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)$$

General Steps of Transformation

- Let $\boldsymbol{X} = (X_1, X_2, \cdots X_n)$ be a random vector and $Y = g(X_1, X_2, \cdots X_n)$ be another random variable defined by a function $g : \mathbb{R}^n \to \mathbb{R}$.
- To find the pdf of Y :

General Steps of Transformation

- Let $\boldsymbol{X} = (X_1, X_2, \cdots X_n)$ be a random vector and $Y = g(X_1, X_2, \cdots X_n)$ be another random variable defined by a function $g : \mathbb{R}^n \to \mathbb{R}$.
- To find the pdf of Y :
  - For each value of y, find the set $A_y = \{x_1, x_2, \cdots x_n : g(x_1, x_2, \cdots x_n) \leq y\}$
  - Find the cdf

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(x_1, x_2, \cdots x_n) \leq y) = \int_{A_y} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$$

  - The pdf is $f_Y(y) = F'_Y(y)$

Examples

- Let $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1]$, find the probability density function for the following random variables:
  - $U = e^{X_1}$
  - $X^{(n)} = \max(X_1, X_2, \cdots X_n)$
  - $Y = X_1 + X_2$