MATH 321: Mathematical Statistics

Lecture 17: Several Random Variables

Chapter 5: Distributions of Functions of Random Variables (5.3 and 5.4)

Multivariate distributions

Introduction

- Now we are extending bivariate distributions to multivariate distributions.

  The good news is that the jump from 2 random variables to 3 or 4 or n random variables is much easier than the jump from 1 to 2.

- The concepts such as marginal and conditional distributions generalize from the bivariate to the multivariate setting.

  We will start by giving these generalizations, then demonstrating via examples.

- A note on notation: Boldface letters are used to denote multiple variates. Write $\mathbf{X}$ to denote $X_1, \ldots, X_n$ and $\mathbf{x}$ to denote the sample $x_1, \ldots, x_n$.

Definitions and theorems

- Joint distributions and probabilities.
  - The random vector $\mathbf{X} = (X_1, \ldots, X_n)$ has a range that is a subset of $\mathbb{R}^n$ (n dimensions).
  - If $\mathbf{X} = (X_1, \ldots, X_n)$ a discrete random vector (the range is countable), then the **joint pmf** of $\mathbf{X}$ is the function defined by

    $$f(\mathbf{x}) = f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n) \text{ for each } (x_1, \ldots, x_n) \in \mathbb{R}^n$$

  - Finding probabilities: Then, for any $A \subset \mathbb{R}^n$,

    $$P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x})$$

  - If $\mathbf{X} = (X_1, \ldots, X_n)$ a continuous random vector, then the **joint pdf** of $\mathbf{X}$ is the function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ that satisfies

    $$P(\mathbf{X} \in A) = \int \cdots \int_A f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int \cdots \int_A f(x_1, \ldots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

- Expected values.
  - Let $g(\mathbf{x})$ be a real-valued function defined on the range of $\mathbf{X}$. The **expected value** of $g(\mathbf{X})$ is

    Discrete

    $$E[g(\mathbf{X})] = \sum_{\mathbf{x}\in\mathbb{R}^n} g(\mathbf{x})f(\mathbf{x})$$

    Continuous

    $$E[g(\mathbf{X})] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\mathbf{x})f(\mathbf{x})\mathrm{d}x_1\cdots\mathrm{d}x_n$$

- These and other definitions are analogous to the bivariate definitions, except now the sums or integrals are over the appropriate subset of $\mathbb{R}^n$ rather than $\mathbb{R}^2$.
- Marginal distributions.
  - The **marginal pdf or pmf** of any subset of the coordinates of $(X_1,\ldots,X_n)$ can be computed by integrating or summing the joint pdf or pmf over all possible values of the other coordinates.
  - Thus for example, the marginal distribution of $(X_1,\ldots,X_k)$ the first k coordinates of $(X_1,\ldots,X_n)$ is given by the pdf or pmf:

    Simple case: $n = 5$, $k = 2$
  - Even though these marginal distributions can themselves be multivariate, they are still called marginal because they have less variables than the joint distribution.
- Conditional distributions.
  - The **conditional pmf or pdf** of a subset of the coordinates of $(X_1,\ldots,X_n)$ given the value of the remaining coordinates is obtained by dividing the joint pdf or pmf by the marginal pdf or pmf of the remaining coordinates.

    $$f(x_{k+1},\ldots,x_n \mid x_1,\ldots,x_k) =$$

Example

1. Let $n = 4$ and

$$f(x_1,x_2,x_3,x_4) = \frac{3}{4}(x_1^2+x_2^2+x_3^2+x_4^2), \qquad 0 < x_i < 1, \quad i = 1,2,3,4$$

   (a) Verify $f(x_1,x_2,x_3,x_4)$ is a valid pdf.

   (b) NOTE: Probabilities ALWAYS need ALL integrals.

   Find $P(X_1 < 1/2, X_2 < 3/4, X_4 > 1/2)$.

(c) Find the marginal pdf of (X_2, X_3).

Now any probability or expected value that involves only X_1 and X_2 can be computed using this marginal pdf.

(d) Find $E(X_2 X_3)$.

(e) Find the conditional pdf $f(x_1, x_4 \mid x_2, x_3)$.

(f) Find $P(X_1 > 3/4, X_4 < 1/2 \mid X_2 = 1/3, X_3 = 2/3)$.

Independence

- Generally, we will be working with independent random variables. This is a very common assumption in probability and statistics that each observation from a random experiment is independent.

  Lets see how this impacts the definitions and theorems we just presented.

- Joint distributions:
  - Definition: Let random variables $X_1, \ldots, X_n$ have joint pdf (or pmf) $f(x_1, \ldots, x_n)$ and let $f_{X_i}(x_i)$ be the marginal pdf (or pmf) of X_i. Then $X_1, \ldots, X_n$ are **mutually independent random variables** if, for every $(x_1, \ldots, x_n)$, the joint pdf (or pmf) can be written as

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

  - Notes
    - We keep the subscripts on X_i because the marginal distributions can be different.
    - Mutual independence (strongest form) $\Longrightarrow$ Pairwise independence AND all possible subsets are independent
  - Example: Let X_1, X_2, X_3 be (mutually) independent exponential random variables with parameters $\lambda_1 = 3, \lambda_2 = 5, \lambda_3 = 1$, respectively.

    (a) Find the joint pdf of X_1, X_2, X_3.

    (b) Find $P(1 < X_1 < 4, X_2 > 3, X_3 \leq 2)$

- If $X_1, \ldots, X_n$ are mutually independent, then knowledge about the values of some coordinates gives us no information about the values of other coordinates. Mutually independent random variables have many nice properties.
- Conditional distributions.
  - If $X_1, \ldots, X_n$ are mutually independent, we can show that the conditional distribution of any subset of the coordinates, given the values of the rest of the coordinates, is the same as the marginal distribution of the subset.
  - Example: Let $X_1, \ldots, X_4$ be mutually independent random variables. Show $f(x_3, x_4 \mid x_1, x_2) = f(x_3) f(x_4)$.

- Expected value.
  - Let $X_1, \ldots, X_n$ be mutually independent random variables. Let $g_1, \ldots, g_n$ be real-valued functions such that $g_i(x)$ is a function only of x_i, $i = 1, \ldots, n$. Then

    $$E[g_1(X_1) \cdots g_n(X_n)] =$$

  - Example: Let X_1, X_2, X_3 be independent exponential random variables with parameters $\lambda_1 = 3, \lambda_2 = 5, \lambda_3 = 1$, respectively.

    Find $E[(2X_1)(X_2 + 1)(X_3)]$.

Linear functions of random variables

Introduction and definition

- Definition: A **linear function (combination) of random variables** consists of n random variables $X_1, \ldots, X_n$ and n coefficient $a_1, \ldots, a_n$

  $$a_1X_1 + a_2X_2 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

- Why is the linear function of random variables important?

  Most of estimators of parameters are linear functions of random variables.

  1. The estimator of the population mean $\mu = E(X)$ is

  2. The estimator of the population variance $\sigma^2 = V(X)$ is

- In order to study the properties of estimators, it is necessary to know how to compute the expected value and variance of linear functions of random variables.

  We will learn how to find their distributions soon (start of MATH 321 topics).

Expected value and variance of linear functions of random variables

- We will start by demonstrating these in the simplest cases (small n and no coefficients, i.e. all $a_i = 1$), then generalize.
- Recall for when $n = 2$.

  $E(X + Y) =$

  $V(X + Y) =$

  - Now generalizing with constants a and b.

    $E(aX + bY) =$

    $V(aX + bY) =$

- Now for $n = 3$.

  $E(X + Y + Z) =$

  $V(X + Y + Z) =$

- The general pattern should be easy to see. Now we can extend this to n (still with no coefficients).

  Theorem: **Mean and variance of $\boldsymbol{X_1 + \cdots + X_n}$**

  $E\left(\sum_{i=1}^{n} X_i\right) =$

  $V\left(\sum_{i=1}^{n} X_i\right) =$

- Finally, in general we have the following theorem:
  (i) **Expected value of a linear function of random variables**

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

  (ii) **Variance of a linear function of random variables**

$$V[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

  If $X_1, \ldots, X_n$ are mutually independent (or uncorrelated),

$$V[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = \sum_{i=1}^{n} a_i^2 V(X_i)$$

- Easy way to understand and remember the variance of linear functions of random variables.
  - Lets look at the result when we square simple linear functions and expand:

$$(X_1 + X_2)^2 \quad = \quad (X_1 + X_2)(X_1 + X_2) \quad =$$

$$(X_1 + X_2 + X_3)^2 \quad =$$

$$(a_1X_1 + a_2X_2 - a_3X_3)^2 \quad =$$

  - Why is this useful? Replace all quadratic (squared) terms with variances and cross (interaction) terms with covariances.

$$V(X_1 + X_2 + X_3) \quad =$$

$$V(a_1X_1 + a_2X_2 - a_3X_3) \quad =$$

– If we have more than 3 random variables, this still works!

$$(a_1X_1 + a_2X_2 + \cdots + a_nX_n)^2 = \sum_{i=1}^{n} a_i^2 X_i^2 + 2\sum_{i<j} a_i a_j X_i X_j$$

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

– Example: Let X_1, X_2 and X_3 be random variables, where $V(X_1) = 1$, $V(X_2) = 3$, $V(X_3) = 5$, $\mathrm{Cov}(X_1, X_2) = -0.4$, $\mathrm{Cov}(X_1, X_3) = 0.5$, $\mathrm{Cov}(X_2, X_3) = 2$.

Find $V(3X_1 - X_2 + 2X_3)$.

Mgf of sums of independent random variables

Introduction

- In some applications, it is sufficient to know the mean and variance of a linear combination of random variables, say, Y. This is what we learned last section (5.3).

  However, it is often helpful to know exactly how Y is distributed (pmf / pdf / mgf). The easiest way to do this is via moment generating functions.

- Recall the definition: The moment generating function (mgf) of random variable X (or the distribution of X), denoted $M_X(t)$, was defined by the following in the univariate case

$$M_X(t) = \underset{\text{In general}}{E(\mathrm{e}^{tx})} \quad \rightarrow \quad \underset{\text{Discrete}}{\sum_x \mathrm{e}^{tx} f(x)} \quad \underset{\text{Continuous}}{\int_{-\infty}^{\infty} \mathrm{e}^{tx}\, f(x)\, \mathrm{d}x}$$

  Additionally, the mgf of a random variable uniquely determines its distribution (i.e. no two random variables with "different" distributions share the same pdf).

Mgf of sums of independent random variables

- Theorem: Let X and Y be independent random variables with mgfs $M_X(t)$ and $M_Y(t)$. Then the mgf of the random variable $S = X + Y$ is given by

$$M_S(t) = M_X(t) \cdot M_Y(t)$$

- Proof:

- Example 1: $X \sim \text{Normal}(\mu_1, \sigma_1^2)$ and $Y \sim \text{Normal}(\mu_2, \sigma_2^2)$ and $X \perp\!\!\!\perp Y$. Find the distribution of $S = X + Y$.

- Note: Whenever finding the distribution of a sum random variables (e.g. $X + Y$), always start with mgfs. It is usually to use the mgf rather than doing transformations using the pmf / pdf.

- Now we can extend the previous theorem to a sum of n random variables:

  Theorem: Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $Y = X_1 + \cdots + X_n$.

  $M_Y(t) =$

  In particular, if $X_1, \ldots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Y(t) =$$

- More examples:

  2. Suppose X_1 and X_2 are independent Poisson random variables with means λ_1 and λ_2, respectively. Find the distribution of $X_1 + X_2$.

     Recall that the mgf of a Poisson(λ) distribution is $M_X(t) = \mathrm{e}^{\lambda(\mathrm{e}^t - 1)}$.

3. Suppose X_1 and X_2 are *iid* Bernoulli random variables ($M_X(t) = (1-p) + pe^t$). Find the distribution of $X_1 + X_2$.

4. The same logic can be used for *iid* geometric distributions ($M_X(t) = \frac{pe^t}{1-qe^t}$) and *iid* exponential distributions ($M_X(t) = \frac{\beta}{\beta - t}$).

5. Suppose $X_1, \ldots, X_n$ are mutually independent random variables, and $X_i \sim \text{Gamma}\,(\alpha_i, \beta)$. Find the distribution of $Y = X_1 + \cdots + X_n$.

   Recall that the mgf of a Gamma(α, β) distribution is $M_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$.

- In general, we can say extend the previous examples and state the following results, which all match our previous explanations / interpretations of the relationships between these distributions:
  - Poisson:

    If $X_1, \ldots, X_n \overset{\perp\!\!\!\perp}{\sim} \text{Poisson}\,(\lambda_i)$, then $Y = X_1 + \cdots + X_n \sim \text{Poisson}\,(\lambda_1 + \cdots + \lambda_n)$.
  - Bernoulli:

    $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}\,(p)$, then $Y = X_1 + \cdots + X_n \sim \text{Binomial}\,(n, p)$.
  - Geometric:

    If $X_1, \ldots, X_r \overset{iid}{\sim} \text{Geometric}\,(p)$, then $Y = X_1 + \cdots + X_r \sim \text{Negative Binomial}\,(r, p)$.
  - Exponential:

    $X_1, \ldots, X_\alpha \overset{iid}{\sim} \text{Exponential}\,(\lambda)$, then $Y = X_1 + \cdots + X_\alpha \sim \text{Gamma}\,(\alpha, \beta)$.
  - Gamma:

    $X_1, \ldots, X_n \overset{\perp\!\!\!\perp}{\sim} \text{Gamma}\,(\alpha_i, \beta)$, then $Y = X_1 + \cdots + X_n \sim \text{Gamma}\,(\alpha_1 + \cdots + \alpha_n, \beta)$.

- Extension of previous theorem to sums of linear combinations of random variables:

  Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Let $Y = (a_1X_1+b_1)+\cdots+(a_nX_n+b_n)$. Then the mgf of Y is

  $$M_Y(t) = \left(\mathrm{e}^{t\sum b_i}\right) M_{X_1}(a_1t)\cdots M_{X_n}(a_nt)$$

  Proof:

- Example: Let $X_1 \sim \text{Normal}\,(\mu = 5, \sigma^2 = 4)$, $X_2 \sim \text{Normal}\,(\mu = 3, \sigma^2 = 8)$, and $X_1 \perp\!\!\!\perp X_2$. Find the distribution of $Y = 3X_1 + 2X_2 - 1$.

  Recall $M_X(t) = \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]$.

- Important result from this:

  Theorem: Let $X_1, \ldots, X_n$ be mutually independent random variables with $X_i \sim \text{Normal}\,(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Then,

  $$Y = \sum_{i=1}^{n}(a_iX_i + b_i) \sim \text{Normal}\left(\mu = \sum_{i=1}^{n}(a_i\mu_i + b_i),\ \sigma^2 = \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

  $\Longrightarrow$ Sum of normal random variables is __________ normal.