Chapter 5

Joint Probability Distributions and Random Samples

5.1

Jointly Distributed Random Variables

Joint Probability Mass Function

Let X and Y be two discrete rv's defined on the sample space of an experiment. The *joint probability mass function* $p(x, y)$ is defined for each pair of numbers (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

Let A be the set consisting of pairs of (x, y) values, then

$$P\left[(X, Y) \in A\right] = \sum_{(x,y) \in A} \sum p(x, y)$$

Marginal Probability Mass Functions

The *marginal probability mass functions* of X and Y, denoted $p_X(x)$ and $p_Y(y)$ are given by

$$p_X(x) = \sum_y p(x, y) \qquad p_Y(y) = \sum_x p(x, y)$$

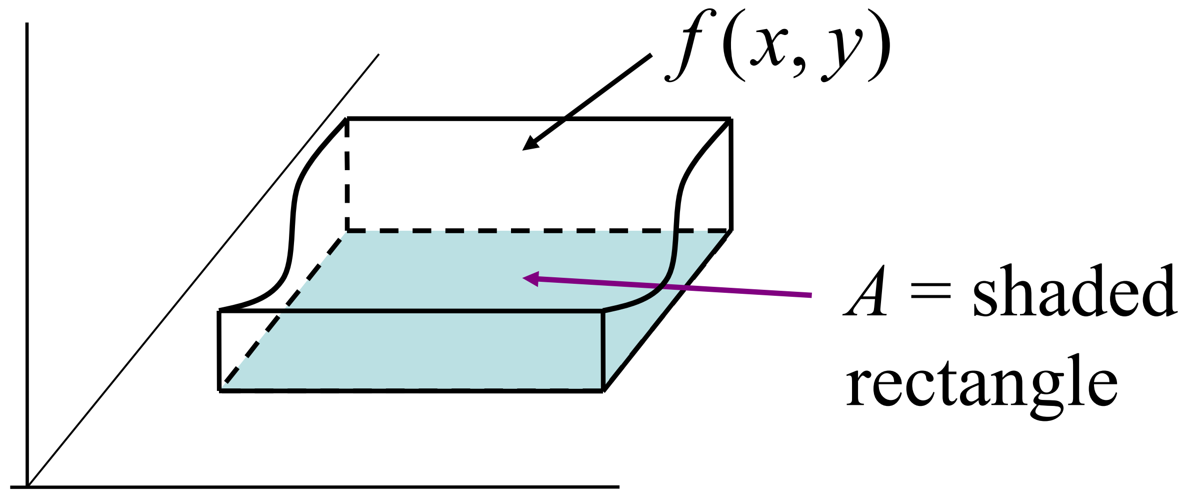
Joint Probability Density Function

Let X and Y be continuous rv's. Then $f(x, y)$ is a *joint probability density function* for X and Y if for any two-dimensional set A

$$P\left[\left(X,Y\right)\in A\right]=\iint\limits_{A} f(x,y)\,dx\,dy$$

If A is the two-dimensional rectangle $\left\{(x,y): a\le x\le b, c\le y\le d\right\}$,

$$P\left[\left(X,Y\right)\in A\right]=\int_{a}^{b}\int_{c}^{d} f(x,y)\,dy\,dx$$



$$P\left[\left(X, Y\right) \in A\right]$$

= Volume under density surface above A

Marginal Probability Density Functions

The *marginal probability density functions* of X and Y, denoted $f_X(x)$ and $f_Y(y)$, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \quad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx \quad \text{for } -\infty < y < \infty$$

Independent Random Variables

Two random variables X and Y are said to be *independent* if for every pair of x and y values

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

when X and Y are discrete or

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

when X and Y are continuous. If the conditions are not satisfied for all (x, y) then X and Y are *dependent.*

More Than Two Random Variables

If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

If the variables are continuous, the joint pdf is the function *f* such that for any *n* intervals $[a_1, b_1]$, …, $[a_n, b_n]$,

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n)$$

$$= \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n) dx_n \ldots dx_1$$

Independence – More Than Two Random Variables

The random variables $X_1, X_2, \ldots, X_n$ are *independent* if for every subset $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ of the variables, the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

Conditional Probability Function

Let X and Y be two continuous rv's with joint pdf $f(x, y)$ and marginal X pdf $f_X(x)$. Then for any X value x for which $f_X(x) > 0$, *the conditional probability density function of Y given that $X = x$* is

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \quad -\infty < y < \infty$$

If X and Y are discrete, replacing pdf's by pmf's gives the *conditional probability mass function of Y when $X = x$*.

5.2

Expected Values, Covariance, and Correlation

Expected Value

Let X and Y be jointly distributed rv's with pmf p(x, y) or pdf f (x, y) according to whether the variables are discrete or continuous. Then the *expected value* of a function $h(X, Y)$, denoted $E[h(X, Y)]$ or $\mu_{h(X,Y)}$

is

$$= \begin{cases} \sum_{x} \sum_{y} h(x,y) \times p(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \times f(x,y)\,dx\,dy & \text{continuous} \end{cases}$$

Covariance

The *covariance* between two rv's X and Y is

$$\text{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy & \text{continuous} \end{cases}$$

Short-cut Formula for Covariance

$$\boxed{\operatorname{Cov}(X,Y) = E(XY) - \mu_X \times \mu_Y}$$

Correlation

The *correlation coefficient* of X and Y, denoted by Corr(X, Y), $\rho_{X,Y}$, or just ρ, is defined by

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

Correlation Proposition

1. If a and c are either both positive or both negative, $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$

2. For any two rv's X and Y,

$$-1 \le \text{Corr}(X, Y) \le 1.$$

Correlation Proposition

1. If X and Y are independent, then $\rho = 0$, but $\rho = 0$ does not imply independence.

2. $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.

5.3

Estatísticas e suas Distribuições

Motivações:

As observações de diversas amostras de tamanho n oriundas da mesma distribuição da população serão geralmente diferentes. Portanto, haverão variações dos valores de qualquer função das observações.

Estatística

Uma *estatística* é qualquer quantidade cujo valor pode ser calculado dos dados amostrais. Antes de obter os dados, há incerteza sobre o valor que resultará de qualquer estatística específica. Assim, uma estatística é uma v.a. A representaremos por uma letra maiúscula; uma letra minúscula será usada para representar o valor calculado ou observado da estatística.

Amostras aleatórias

As v.a´s $X_1, \ldots, X_n$ são ditos formar uma *amostra aleatória (simples)* de tamanho n se

1. Os X_i's são v.a´s independentes.

2. Cada X_i tem a mesma distribuição de probabilidade.

A.A´s

As condições 1 e 2 são parafraseadas, dizendo que os X_i's são v.a´s iid. (independentes e idênticamente distribuidos). As condições são satisfeitas:

- No caso de amostragem com reposição ou de uma população infinita.
- Aprox. Quando n << N, na amostragem sem reposição.
- na prática, se n/N $\leq$0,05 (no máx. 5% da população coletada).

Experimentos de Simulação

As seguintes características devem ser especificadas:

1. A estatística de interesse.
2. A distribuição da população.
3. O tamanho da amostra n.
4. O número de réplicas k.

Distribuição da Média Amostral

Usando a média amostral

Seja $X_1, \ldots, X_n$ uma a.a. de uma distribuição com média μ e desvio padrão σ. Então:

$$1.\ E\left(\overline{X}\right) = \mu_{\overline{X}} = \mu$$

$$2.\ V\left(\overline{X}\right) = \sigma^2_{\overline{X}} = \sigma^2 / n$$

Além disso, com $T_o = X_1 + \ldots + X_n$,
$E\left(T_o\right) = n\mu$, $V\left(T_o\right) = n\sigma^2$, and $\sigma_{T_o} = \sqrt{n}\sigma$.

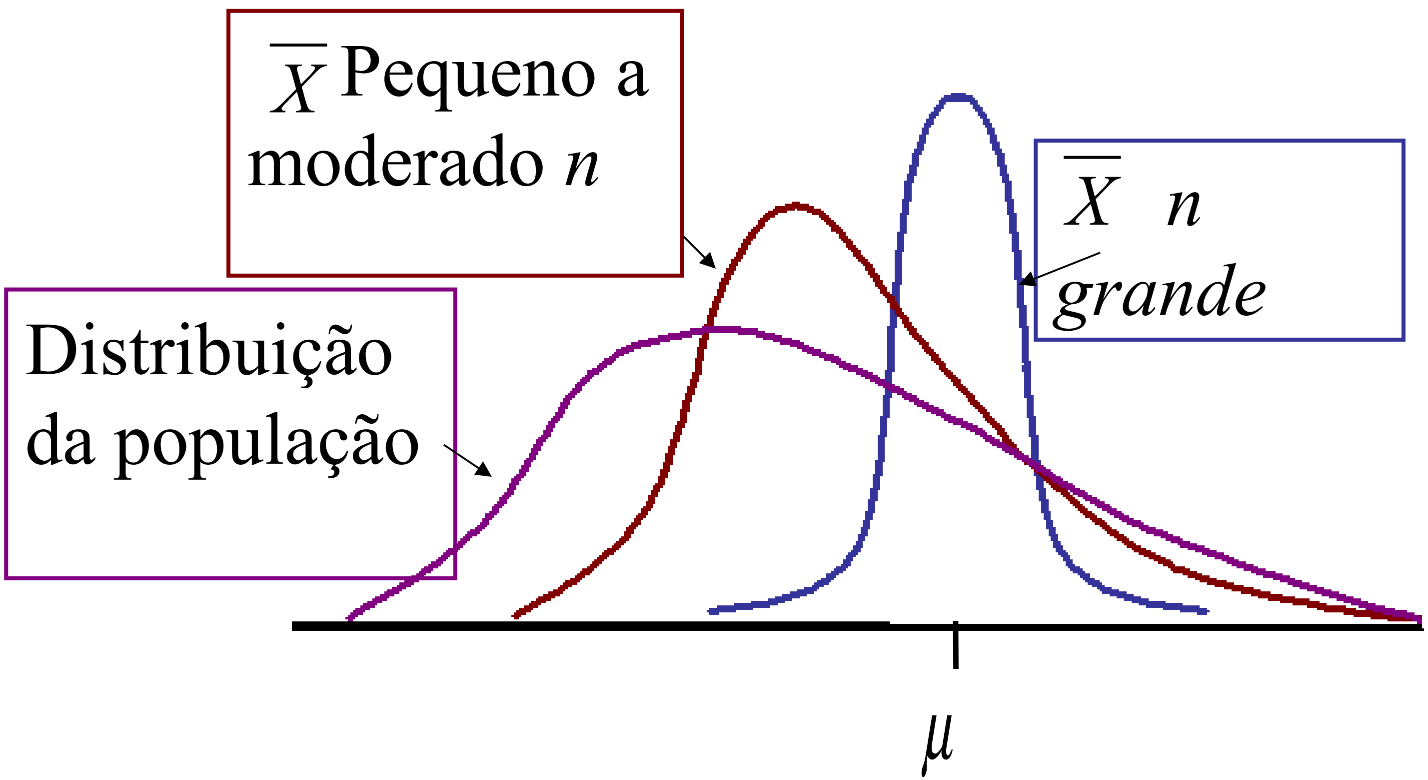
Distribuição Normal Populacional

Seja $X_1,\ldots, X_n$ uma a.a. De uma distribuição normal com média μ e desvio padrão σ . Então para qualquer n, T_o e $\overline{X}$ também são normalmente distribuidos.

Teorema do Limite Central

Seja $X_1, \ldots, X_n$ uma a.a. de uma distribuição com média μ e variância σ^2, então, para n suficientemente grande, $\overline{X}$ tem distribuição aproximadamente normal com média $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \sigma^2 / n$, e T_o também tem distribuição aproximada normal com $\mu_{T_o} = n\mu$, $\sigma_{T_o} = n\sigma^2$. Quanto maior for o valor de n, melhor será a aproximação.

O Teorema do limite central



Regra de Thumb

Se $n > 30$, o Teorema do Limite Central pode ser usado.

Aproximação à distribuição lognormal

Seja $X_1,\ldots, X_n$ uma a.a. de uma distribuição para a qual são possíveis apenas valores positivos $[P(Xi > 0) = 1]$. Então, se n é suficientemente grande, o produto $Y = X_1X_2\ldots X_n$ tem aproximadamente uma distribuição lognormal.

Distribuição de uma Combinação Linear

Combinação Linear

Dada uma coleção de n v.a $X_1, \ldots, X_n$ e n constantes numéricas $a_1, \ldots, a_n$, a v.a

$$Y = a_1 X_1 + \ldots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

É chamada uma *combinação linear* dos X_i's.

Valor esperado de uma Combinação Linear

Sejam $X_1, \ldots, X_n$ com médias $\mu_1, \mu_2, \ldots, \mu_n$ e variâncias de $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectivamente

Se os X_i's são ou não independentes,

$$E(a_1 X_1 + \ldots + a_n X_n) = a_1 E(X_1) + \ldots + a_n E(X_n)$$

$$= a_1 \mu_1 + \ldots + a_n \mu_n$$

Variancia de uma Combinação Linear

Se $X_1, \ldots, X_n$ são independentes,

$$V(a_1 X_1 + \ldots + a_n X_n) = a_1^2 V(X_1) + \ldots + a_n^2 V(X_n)$$

$$= a_1^2 \sigma_1^2 + \ldots + a_n^2 \sigma_n^2$$

e

$$\sigma_{a_1 X_1 + \ldots + a_n X_n} = \sqrt{a_1^2 \sigma_1^2 + \ldots + a_n^2 \sigma_n^2}$$

Variância de uma Combinação Linear

Para quaisquer $X_1, \ldots, X_n$,

$$V\left(a_1X_1 + \ldots + a_nX_n\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{Cov}\left(X_i, X_j\right)$$

Diferença entre duas V.A´s

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

e, se X_1 e X_2 são independentes,

$$V(X_1 - X_2) = V(X_1) + V(X_2)$$

Diferença entre v.a´s Normais

Se $X_1, X_2, \ldots X_n$ são v.a iid, então qualquer combinação linear dos X_i's também tem distribuição normal. A diferença $X_1 - X_2$ entre duas v.a´s iid normalmente distribuidas, tem também distribuição normal.