# Notes 3: Discrete Probability Distributions

Julio Garín
Department of Economics

Statistics for Economics

Spring 2012

# Discrete and Continuous Random Variables

Table: Examples of Discrete Random Variables

| **Experiment** | **Random Variable** ($x$) | Possible values |
|---|---|---|
| Sell an automobile | Gender of the customer | 0 if male; 1 if female |
| Coin is tossed three times | Number of heads | |

Table: Examples of Continuous Random Variables

| **Experiment** | **Random Variable** ($x$) | Possible values |
|---|---|---|
| Bid in an auction | Maximum bid | |
| Fill a soft drink can | Number of onces | |

# Some Concepts

A *random variable* is a real-valued function for which the domain is the sample space.

A random variable $X$ is said to be *discrete* if it can assume only a finite or countable infinite number of distinct values

The expression $(X = x)$ can be read, *the set of all points in $S$ assigned the value $x$ by the random variable $X$*.

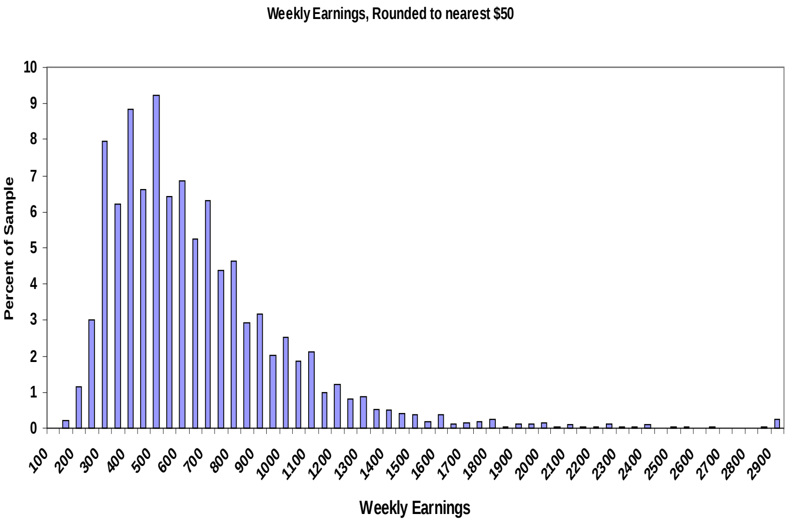
# Some Concepts

The probability that $X$ takes on the value $x$, $P(X = x)$, is defined as the *sum of the probabilities of all sample points in* $S$ that are assigned the value $x$. We will sometimes denote $P(X = x)$ by $f(x)$.

- There are $n$ possible outcomes.
- $P(X = x_i)$ is the probability that outcome $i$ will happen.
  - Called the PDF: Probability Distribution Function (P. Density F.)

The probability distribution for a discrete variable $X$ can be represented by a formula, a table, or a graph that provides $f(x) = P(X = x) \quad \forall x$.

# Histogram and PDF



# Some Concepts

- Similar to the previous section:

For any discrete probability distribution, the following must be true:

1. $0 \leq f(x) \leq 1 \quad \forall x$.
2. $\sum_x f(x) = 1$, where the summation is over all values of $x$ with nonzero probability.

- Why would I care?

Table: Probability Distribution for the Number of Automobiles Sold During a Day

| $x$ | $f(x)$ |
|---|---|
| 0 | 0.18 |
| 1 | 0.39 |
| 2 | 0.24 |
| 3 | 0.14 |
| 4 | 0.04 |
| 5 | 0.01 |
| **Total** | 1 |

# Cumulative Distribution Function

The distribution function $F(x)$, also called the cumulative distribution function (CDF) or cumulative frequency function, describes the probability that a r.v. $X$ takes on a value less than or equal to a number $x$.

- Mathematically.
- Why would I care?
  - Many times, it is useful to work with complements:
    - CDF $= P(x \leq k)$

$$P(x > k) = 1 - P(x \leq k)$$

  - Exmple: The birthday problem.

# Example: PDF and CDF

Table: Probability Distribution for the Number of Automobiles Sold During a Day

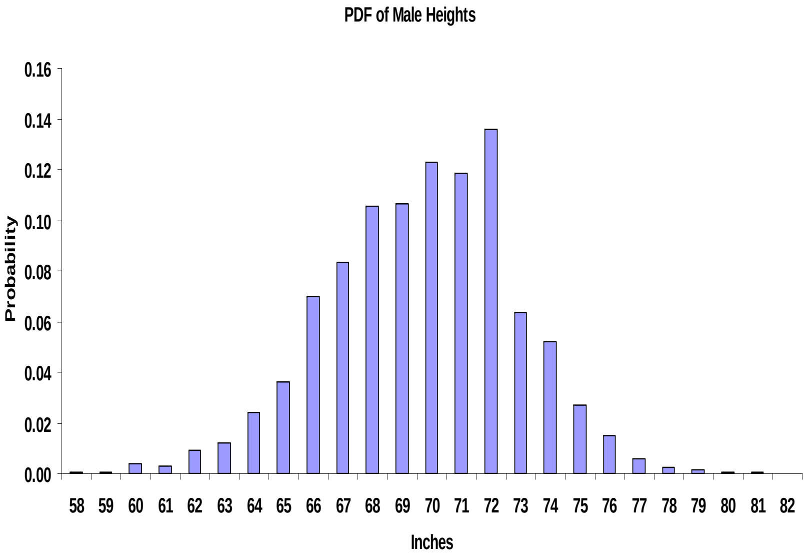
| $x$ | $f(x)$ | $F(x)$ |
|---|---|---|
| 0 | 0.18 | 0.18 |
| 1 | 0.39 | 0.57 |
| 2 | 0.24 | 0.81 |
| 3 | 0.14 | 0.95 |
| 4 | 0.04 | 0.99 |
| 5 | 0.01 | 1 |
| **Total** | 1 | |

# Example: PDF

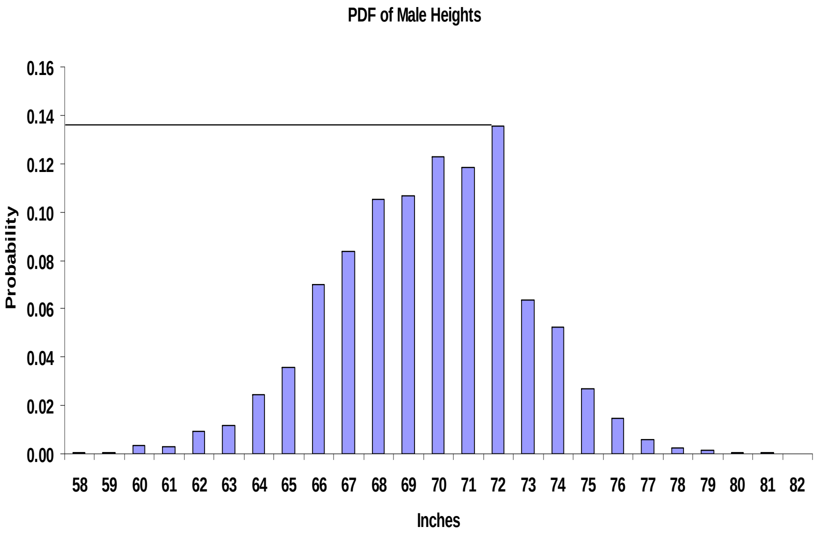
Table: PDF of Height of Males (In Inches)

| Height | Fraction | Height | Fraction |
|---|---|---|---|
| 58 | 0.0006 | 66 | 0.0700 |
| 59 | 0.0006 | 67 | 0.0834 |
| 60 | 0.0036 | 68 | 0.1054 |
| 61 | 0.0028 | 69 | 0.1067 |
| 62 | 0.0093 | 70 | 0.1229 |
| 63 | 0.0120 | 71 | 0.1183 |
| 64 | 0.0243 | 72 | 0.1357 |
| 65 | 0.0359 | 73 | 0.0635 |

# PDF


PDF of Male Heights
Probability
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
Inches

# $Pr(X = 72)$

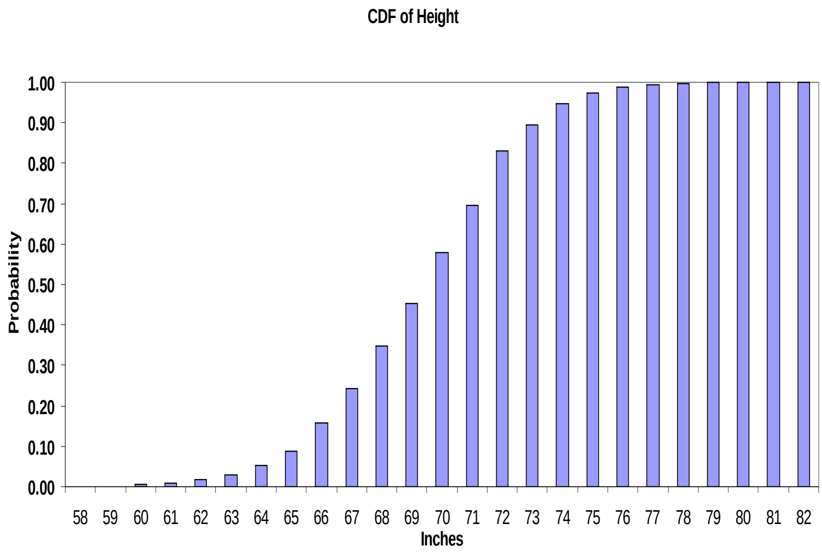

PDF of Male Heights
Probability
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
Inches

# Example: CDF

Table: CDF of Height of Males (In Inches)

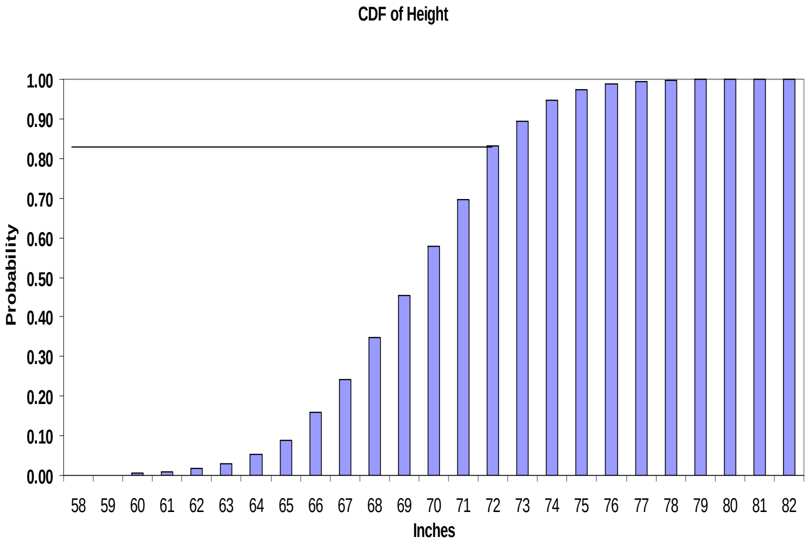
| Height | Fraction | Height | Fraction |
|---|---|---|---|
| 58 | 0.0006 | 66 | 0.1590 |
| 59 | 0.0012 | 67 | 0.2425 |
| 60 | 0.0049 | 68 | 0.3479 |
| 61 | 0.0077 | 69 | 0.4547 |
| 62 | 0.0170 | 70 | 0.5776 |
| 63 | 0.0289 | 71 | 0.6959 |
| 64 | 0.0530 | 72 | 0.8317 |
| 65 | 0.0892 | 73 | 0.8952 |

# PDF



CDF of Height
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
Inches

# $F(72)$



CDF of Height
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Probability
58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
Inches

# Example

Table: Probability Distribution for a r.v.

| $x$ | $f(x)$ |
|---|---|
| 20 | 0.20 |
| 25 | 0.15 |
| 30 | 0.25 |
| 35 | 0.40 |

1. Is this probability distribution valid?
2. What is the probability that $x = 30$.
3. What is the probability that $x$ is less than or equal to 25?
4. What is the probability that $x$ is greater than 30?

# Uniform Distribution

- Not only tables and grahps can give us probabilities.

The *discrete uniform distribution* can be represented by,

$$f(x) = \frac{1}{n}$$

where $n$ the number of values the variable may assume.

- Example I: Rolling a die.
- Example II: Pclretchfpecqh.
- Tables vs. graphs vs. formulas.

Moments

# Expected Value

Let $X$ be a discrete r.v. with the probability function $f(x)$. Then the expected value of $X$, $\mathbb{E}(X)$, is defined to be

$$\mathbb{E}(X) = \mu = \sum_x x f(x)$$

if the sum is absolutely convergent.

Table: Probability Distribution of Cars Sold During a Day

| $x$ | $f(x)$ | $xf(x)$ |
|---|---|---|
| 0 | 0.18 | |
| 1 | 0.39 | |
| 2 | 0.24 | |
| 3 | 0.14 | |
| 4 | 0.04 | |
| 5 | 0.01 | |
| **Total** | 1 | $\sum_x xf(x)$ |

# Variance

If $X$ is a r.v. with mean $\mathbb{E}(X) = \mu$, the *variance* of a r.v. $X$ is defined to be the expected value of $(x - \mu)^2$. That is,

$$\text{Var}(X) = \mathbb{E}\left[(X - \mu)^2\right]$$

The *standard deviation* of $X$ is the positive square root of Var$(X)$

Table: Probability Distribution of Cars Sold During a Day

| $x$ | $f(x)$ | $(x-\mu)^2$ | $f(x)$ | $(x-\mu)^2 f(x)$ |
|---|---|---|---|---|
| 0 | 0-1.5 | 2.25 | 0.18 | |
| 1 | 1-1.5 | 0.25 | 0.39 | |
| 2 | 2-1.5 | 0.25 | 0.24 | |
| 3 | 3-1.5 | 2.25 | 0.14 | |
| 4 | 4-1.5 | 6.25 | 0.04 | |
| 5 | 5-1.5 | 12.25 | 0.01 | |
| **Total** | | | | $\sum_x (x-\mu)^2 f(x)$ |

# Example

Table: Probability Distribution for $x$

| $x$ | $f(x)$ |
|---|---|
| 0 | 1/8 |
| 1 | 1/4 |
| 2 | 3/8 |
| 3 | 1/4 |

1. Calculate:
   - Mean.
   - Variance.
   - Standard Deviation.
2. Draw the PDF.

# Some Theorems (and Proofs) Related to Moments

Let $X$ be a discrete random variable with probability function $f(x)$ and $c$ be a constant. Then

$$\mathbb{E}(c) = c$$

Let $X$ be a discrete random variable with probability function $f(x)$, $g(X)$ be a function of $X$, and $C$ be a constant. Then

$$\mathbb{E}\left[cg(X)\right] = c\mathbb{E}\left[g(X)\right]$$

Let $X$ be a discrete random variable with probability function $f(x)$, $g_1(X), g_2(X), \cdots, g_k(X)$, be $k$ functions of $X$. Then

$$\mathbb{E}\left[g_1(X) + g_2(X) + \cdots + g_k(X)\right] = \mathbb{E}\left[g_1(X)\right] + \mathbb{E}\left[g_2(X)\right] + \cdots + \mathbb{E}\left[g_k(X)\right]$$

Let $X$ be a discrete random variable with probability function $f(x)$ and mean $\mathbb{E}(X) = \mu$. Then

$$\text{Var}(X) = \sigma^2 = \mathbb{E}\left[(X - \mu)^2\right] = \mathbb{E}(X^2) - \mu^2$$

## Binomial Probability Distribution

# Binomial Distribution

- Motivation.

A *binomial experiment* possesses the following properties:

1. The experiment consists of a fixed number, $n$, of identical trials.
2. Each trial results in one of two outcomes: success, S, or failure, F.
3. The probability of success on a single trial is equal to some value $p$ and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
4. The trials are independent.
5. The random variable of interest is $X$, the number of successes observed during the $n$ trials.

- Example: Tossing a coin $n$ times.

# Determining whether a particular experiment is a Binomial

An early-warning detection system for aircraft consist of four identical radar units operating independently of one another. Suppose that each has a probability of 0.95 of detecting an intruding aircraft. When an intruding aircraft enters the scene, the random variable of interest is $X$, the number of radar units that do not detect the plane. Is this a binomial experiment?

1. The experiment consists of a fixed number, $n$, of identical trials.
2. Each trial results in one of two outcomes: success, S, or failure, F.
3. The probability of success on a single trial is equal to some value $p$ and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
4. The trials are independent.
5. The random variable of interest is $X$, the number of successes observed during the $n$ trials.

# Example: Martin Clothing Store

Consider the purchase decision of the next three customers who enter the Martin Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is 0.30.

1. What is the probability that two of the next three customers will make a purchase?

- Number of experimental outcomes providing exactly 2 successes in the $n = 3$ trials.
- Probability of purchases.
- Table with trial outcomes.
- What is the probability of a particular sequence of trial outcomes with $x$ successes in $n$ trials?

# Binomial Probability Function

Number of experimental outcomes providing exactly $s$ successes in $n$ trials:

$$_nC_s \equiv \binom{n}{s} \equiv C(n, s) = \frac{n!}{s!(n-s)!}$$

A r.v. $X$ is said to have a *binomial probability distribution* based on $n$ trials with success probability $p$ iff

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

where,

$$x = 0, 1, 2, 3 \cdots , n \quad \text{and} \quad 0 \le p \le 1$$

# Moments and CDF of a Binomial Probability Function

If a r.v. $X$ can be described by a binomial probability distribution, then

$$\mathbb{E}(X) = \mu = np$$

and

$$\mathrm{Var}(X) = \sigma^2 = np(1-p)$$

The CDF of a binomial distribution is

$$P(x \leq a) = \sum_{x=0}^{a} \left[ {}_nC_x p^x q^{n-x} \right]$$

# Example: Martin Clothing Store

Table: Probability Distribution for the Number of Customers Making a Purchase

| $x$ | $f(x)$ |
|---|---|
| 0 | |
| 1 | |
| 2 | |
| 3 | |

- With three customers, what is the expected number of customers who will make a purchase?
- Variance and standard deviation for the number of customers who will make a purchase?
- For the next 1,000 customers entering the store, what is the variance for the number of customers who will make a purchase?

# Example 2: Births

- $P(\text{Girl on any birth}) = 0.486$
  1. Given 5 births, what is the chance that two of them will be girls?
  2. Given 3 births, what is the chance you will have 2 or more girls?

# Example 3: Call in Sick

- Average US workers have a 2% chance of losing a work day due to illness or injury.
  1. Assume 50 weeks and 5 days per week, what is the expected number of lost work days in a year?
  2. What is the standard deviation of lost days?

## Poisson Probability Distribution

# Introduction

- Motivation.
- Keep building up: Poisson and Binomial.
- Poisson experiments have two properties:
  1. The probability of an occurrence is the same for any two intervals of equal length.
  2. The occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval.
- Examples:
  - Mortatility statistics.
  - Number of tpingy errors on a text.
  - Specially in legal cases, it is used to find the probability of rare diseases (like Leukemia).
  - Machine/product failures.
  - The number of deaths by horse kicking in the Prussian army during a 20-year period.
  - Probability of criminal veredicts.
  - Number of hairs found in McDonald's hamburgers.

# Departure: Binomial

- From Binomial to Poisson: shrinking the time.
  - Example number of automobile accidents at an intersection during a time period.

A r.v. $X$ is said to have a *Poisson probability distribution* iff

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

1. Show that the probability assigned by the Poisson probability distribution satisfy the requirements for a proper probability distribution.

# Moments and Two Examples

If $X$ is a r.v. possessing a Poisson distribution with parameter $\lambda$, then

$$\mu = \mathbb{E}(X) = \lambda$$

and

$$\sigma^2 = \text{Var}(X) = \lambda$$

1. Time interval:
   - Number of arrivals at the drive-up teller window of a bank during a 15-minute period on weekday mornings. Data shows that the number of cars arriving in a 15-minute period of time is 10. What is the probability of exactly five arrivals in 15 minutes?
2. Distance interval:
   - Occurence of major defects in a highway one month after resurfacing. Assume major defects one month after resurfacing occur at the average rate of two per mile. What is the probability of no major defects in a particular 3-mile section of the highway.

# Example of a Poisson Application

Table: Number of Men Kicked to Death by Horses in Ten Prussian Army Corps (Bortkiewicz (1898))

| **Men killed per year per corp** | **Observation (Number of deaths)** | **Poisson** |
|---|---|---|
| 0 | 109 (0) | 108.7 (0.0) |
| 1 | 65 (65) | 66.3 (66.3) |
| 2 | 22 (44) | 20.2 (40.4) |
| 3 | 3 (9) | 4.1 (12.3) |
| 4 | 1 (4) | 0.6 (2.4) |
| 5+ | 0 (0) | 0.1 (0.5) |
| **Corp-years** | 200 | 200 |
| **Total deaths** | 122 | 121.9 |
| **Mean** | 0.610 | 0.610 |
| **Mean Variance** | 0.611 | 0.610 |

# Example I: Police Patrol

Suppose that a random system of police patrol is devised so that a patrol officer may visit a given beat location $X = 0, 1, 2, 3, \cdots$ times per half-hour period, with each location being visited in average once per time period. Assume that $X$ posesses, aproximately, a Poisson probability distribution.

1. What is the probability that the patrol officer will miss a given location during a half-hour period.
2. What is the probability that it will be visited once? Twice?
3. At least once?

# Example II: Planting trees

A certain type of tree has seedling randomly disperse in a large area, with the mean density of seedling being approximately five per squared yard. Suppose that forester randomly locates ten 1-square-yard regions in the area.

1. What is the probability that none of the regions will contain seedlings.

# Material Covered So Far

Review.