14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 5

Konrad Menzel

February 19, 2009

We distinguish 2 different types of random variables: discrete and continuous.

1   Discrete Random Variables

Definition 1 *A random variable X has a* discrete *distribution if X can take on only a finite (or countably infinite) number of values $(x_1, x_2, \ldots)$.*

Definition 2 *If random variable X has a* discrete *distribution, the* probability density function *(p.d.f.) of X is defined as the function*

$$f_X(x) = P(X = x)$$

If $\{x_1, x_2, \ldots\}$ is the set of all possible values of X, then for any $x \notin \{x_1, x_2, \ldots\}$, $f_X(x) = 0$. Also

$$\sum_{i=1}^{\infty} f_X(x_i) = 1$$

The probability that $X \in A$ for $A \subset \mathbb{R}$ is

$$P(X \in A) = \sum_{x_i \in A} f_X(x_i)$$

Example 1 *If X is the number we rolled with a die, all integers $1, 2, \ldots, 6$ are equally likely. More generally, we can define the discrete uniform distribution over the numbers $x_1, x_2, \ldots, x_k$ by its p.d.f.*

$$f_X(x) = \begin{cases} \frac{1}{k} & \text{if } x \in \{x_1, x_2, \ldots, x_k\} \\ 0 & \text{otherwise} \end{cases}$$

This corresponds to the simple probabilities for an experiment with sample space $S = \{x_1, x_2, \ldots, x_k\}$.

Example 2 *Suppose we toss 5 fair coins independently from another and define a random variable X which is equal to the observed number of heads. Then by our counting rules, $n(S) = 2^5 = 32$, and $n(\text{"k heads"}) = \begin{pmatrix} 5 \\ k \end{pmatrix}$, using the rule on combinations. Therefore*

$$P(X = 0) = \begin{pmatrix} 5 \\ 0 \end{pmatrix} \frac{1}{32} = \frac{1}{32}, \qquad P(X = 1) = \begin{pmatrix} 5 \\ 1 \end{pmatrix} \frac{1}{32} = \frac{5}{32},$$

$$P(X = 2) = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \frac{1}{32} = \frac{10}{32}, \qquad P(X = 3) = \begin{pmatrix} 5 \\ 3 \end{pmatrix} \frac{1}{32} = \frac{10}{32},$$

$$P(X = 4) = \begin{pmatrix} 5 \\ 4 \end{pmatrix} \frac{1}{32} = \frac{5}{32}, \quad \text{and} \quad P(X = 5) = \begin{pmatrix} 5 \\ 5 \end{pmatrix} \frac{1}{32} = \frac{1}{32}.$$
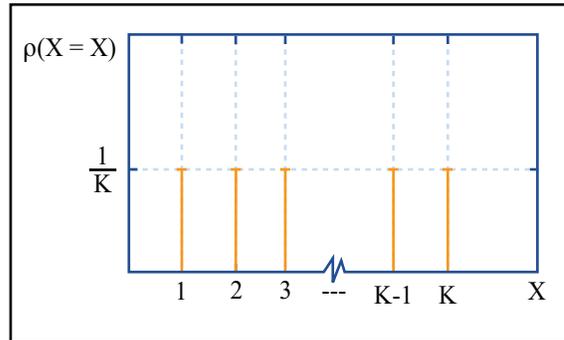
Note that these probabilities sum to 1.
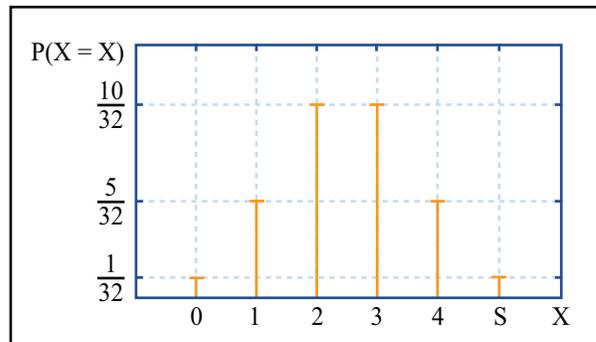
Note that in the die roll example, every single outcome corresponded to exactly one value of the random variable. In contrast for the five coin tosses there was a big difference in the number of outcomes corresponding to $X = 2$ compared to $X = 0$, say. So mapping outcomes into realizations of a random variable may lead to highly skewed distributions even though the underlying outcomes of the random experiment may all be equally likely.

1.1 The Binomial Distribution

To generalize the preceding example, suppose we look at a sequence of n independent and identical trials, each of which can result in either a "success" or a "failure" (not necessarily with equal probability), and we are interested in the total number X of successes.

Example 3 *Suppose we sample 100 pieces from a batch of car parts at the producing plant for quality control. A piece passing the quality checks would constitute a "success", a piece falling short of one or more of the criteria would be a "failure". We are interested in the distribution of failures as a function of the total share of defective parts in the batch in order to infer from the sample whether we have good reason to believe that no more than, say, 1% of the pieces don't meet the standards.*

Suppose the probability of a success equals p, and the probability of a failure is therefore $q = 1 - p$. Since the trials are independent by assumption, the probability of any given sequence of x successes and $n - x$ failures *in fixed order* is

$$p^x(1 - p)^{n-x}$$

2

However, since we are only interested in the *number* X of successes, we have to take into account that there are $\begin{pmatrix} n \\ x \end{pmatrix}$ different sequences with x successes. Therefore, the probability of x successes is

$$P(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}$$

Definition 3 *A random variable X with probability density function*

$$f_X(x) = P(X = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} & \text{if } x \in \{1, 2, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$

is said to follow a binomial distribution *with parameters p and n, written as*

$$X \sim B(n, p)$$

You should notice that previously, we derived the probability distribution for every random experiment separately, writing down the number of possible outcomes, outcomes in each event etc. The binomial distribution serves as a model for a whole range of random experiments. For any given example which falls into this category, we can just look up the probabilities for a given set of parameters (n, p).

Example 4 *In order to make some money off your classmates, you obtained a bent 25 cent coin that comes up heads with a probability of $p_L = \frac{4}{5}$. Unfortunately, that coin got mixed up with your regular small coins, and only after you inserted 8 out of 9 quarters into the laundromat you notice your mistake. You hastily toss the coin 10 times, and it comes up heads for a total of 8 times. Would it be a good idea to continue to rip off your friends with the old coin trick or are you now stuck with a regular (fair) quarter with $p_F = \frac{1}{2}$? I.e. what is $P(A|B)$ for $A =$ "remaining coin is bent" and $B =$ "8 heads out of 10"? If the coin is fair,*

$$P(B|A^C) = \begin{pmatrix} 10 \\ 8 \end{pmatrix} p_F^8 (1-p_F)^{10-8} = \begin{pmatrix} 10 \\ 8 \end{pmatrix} \frac{1}{2^{10}}$$

If it is bent,

$$P(B|A) = \begin{pmatrix} 10 \\ 8 \end{pmatrix} p_L^8 (1-p_L)^{10-8} = \begin{pmatrix} 10 \\ 8 \end{pmatrix} \frac{4^8 \cdot 1^2}{5^{10}}$$

Now let's see what Bayes theorem has to say:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{\begin{pmatrix} 10 \\ 8 \end{pmatrix} \frac{4^8 \cdot 1^2}{5^{10}} \frac{1}{9}}{\begin{pmatrix} 10 \\ 8 \end{pmatrix} \frac{4^8 \cdot 1^2}{5^{10}} \frac{1}{9} + \begin{pmatrix} 10 \\ 8 \end{pmatrix} \frac{1}{2^{10}} \frac{8}{9}} = \frac{\frac{4^8}{5^{10}}}{\frac{4^8}{5^{10}} + \frac{8}{2^{10}}} \approx 46.21\%$$

Therefore, it is more likely that the coin you are left with is in fact a regular quarter - which doesn't mean that in total, the probability of heads is still

$$P(H|B) = \frac{4}{5} \cdot 46.21\% + \frac{1}{2} \cdot 53.79\% \approx 63.86\%$$

However a better idea would of course be to do a few more tosses. If you can repeat the experiment arbitrarily often, you will eventually be able to distinguish the two types of coins with an arbitrary degree of certainty. As an aside, you can see this exercise as a very basic example for a hypothesis test.
Say you got another 8 heads out of another 10 trials (we'll call this event C), leaving the share of heads

at the same value as before. Then, doing the same steps as before, this time based on the first posterior $P(H|B)$ rather than the prior $P(B)$, the conditional probability would be

$$P(H|B \cap C) = \frac{\left(\begin{array}{c} 10 \\ 8 \end{array} \right) \frac{4^8 \cdot 1^2}{5^{10}} \cdot 46.21\%}{\left(\begin{array}{c} 10 \\ 8 \end{array} \right) \frac{4^8 \cdot 1^2}{5^{10}} \cdot 46.21\% + \left(\begin{array}{c} 10 \\ 8 \end{array} \right) \frac{1}{2^{10}} \cdot 53.79\%} = \frac{46.21\%}{46.21\% + 53.79\% \cdot \frac{5^{10}}{4^8 2^{10}}} \approx 85.51\%$$

Alternatively, if we just aggregate the two series of trials into 16 heads out of 20 trials, we get

$$P(A|B') = \frac{\left(\begin{array}{c} 20 \\ 16 \end{array} \right) \frac{4^{16}}{5^{20}}}{\left(\begin{array}{c} 20 \\ 16 \end{array} \right) \frac{4^{16}}{5^{20}} + \left(\begin{array}{c} 20 \\ 16 \end{array} \right) \frac{8}{2^{20}}} \approx 85.51\%$$

so that it doesn't matter whether we update simultaneously or sequentially. This is more generally a desirable property of Bayesian updating: the final result only depends on the overall information we are using, not the order in which we update.

2   Continuous Random Variables

Many types of data are results from measurements of some kind: weights, lengths, incomes etc., which can - at least conceptually - take any value in some interval (sometimes all) of the real numbers. In this case, the definition of a probability density function for discrete random variables is not practical, because (a) the number of possible outcomes is uncountable, so we can't just add up probabilities over single values, and (b) the probability of any *particular* value on the continuum typically has to be zero. This is why we have to deal with this type of random variables separately from the discrete case.

Definition 4 *A random variable X has a* continuous distribution *if X can take on any values in some interval - bounded or unbounded - of the real line.*

For discrete random variables, it was relatively straightforward to define a probability density function, since there was only a finite number of values. A continuous random variable can take more than countably many values, and therefore the derivation becomes a little bit more involved.

The idea goes as follows: we can "discretize" the distribution by putting the possible values the random variable can take into "bins", i.e. instead of looking at the probabilities $P(X = x)$, we'll look at probabilities for intervals, i.e. $P(x_1 \leq X \leq x_2)$. The graphical representation of this is a *histogram*: we fix a set of points $x_0 < x_1 < \ldots < x_n$ on the real line and calculate the probabilities for X falling into each "bin", i.e. interval between two subsequent points, so that $P(x_{i-1} \leq X \leq x_i)$. Then for values in the interval $[x_0, x_n]$, we can define a function

$$h_n(x) = \begin{cases} \frac{P(x_0 \leq X \leq x_1)}{x_1 - x_0} & \text{if } x \in [x_0, x_1) \\ \vdots \\ \frac{P(x_{i-1} \leq X \leq x_i)}{x_i - x_{i-1}} & \text{if } x \in [x_{i-1}, x_i) \\ \vdots \\ \frac{P(x_{n-1} \leq X \leq x_n)}{x_n - x_{n-1}} & \text{if } x \in [x_{n-1}, x_n) \end{cases}$$

4

Dividing by the length of the interval $[x_{i-1}, x_i)$ makes sure that the *area* under the graph for a given interval equals the probability of the random variable X taking a value in that interval, i.e. we can calculate

$$P(x_j \le X \le x_k) = \sum_{i=j+1}^{k} P(x_{i-1} \le X \le x_i) = \sum_{i=j+1}^{k} \left[\int_{x_{i-1}}^{x_i} h_n(t)dt \right] = \int_{x_j}^{x_k} h_n(t)dt \quad \text{for } x_j < x_k$$
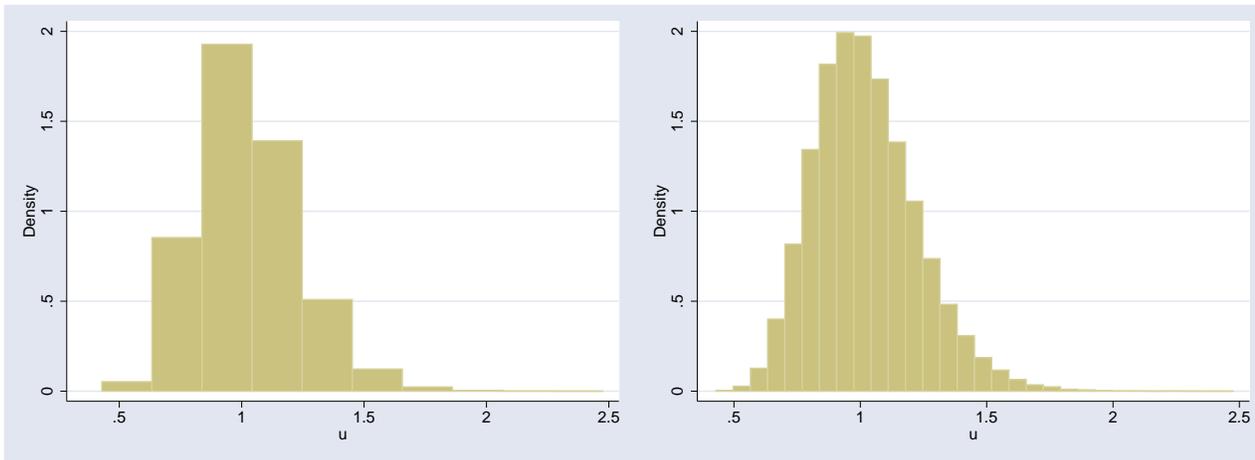
Figure 1: Histograms of the same Distribution for 10 and 30 Bins, respectively
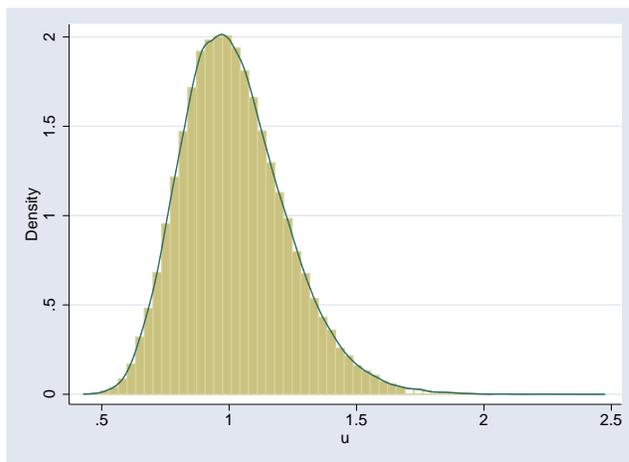
Figure 2: Histogram with 60 Bins and Continuous Density

This is not completely satisfying yet, since this only allows us to calculate the probability of X falling between any two of the grid points from $x_0 < x_1 < \ldots < x_n$, but not e.g. of a subinterval of, say, $[x_j, x_k]$. We can address this by making the grid of points $x_1, x_2, \ldots$ finer, and therefore the intervals narrower. If we shrink the distance between neighboring points x_{i-1}, x_i to an arbitrarily small "dx", we'll have a function $h_\infty(x)$ for which the probability that X lies between *any* two points a and b can be expressed

as the integral of $h_\infty(x)$ from a to b. This limit is called the probability density function for a continuous random variable:

Definition 5 *If random variable X has a* continuous *distribution, the* probability density function *(p.d.f.) of X is defined a positive function $f_X(x)$ such that for any interval $A \subset \mathbb{R}$*

$$P(X \in A) = \int_A f_X(t)dt$$

F(x)

$P(a \leq x \leq b)$

X

a   b
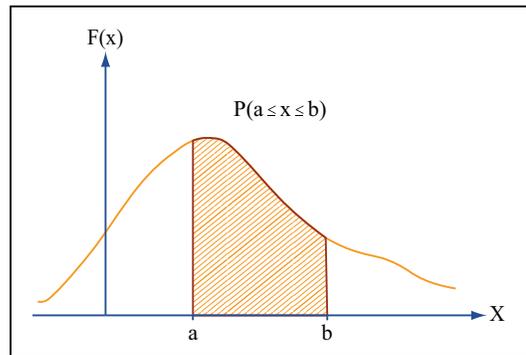
Image by MIT OpenCourseWare.

Figure 3: P.D.F. of a Continuous Random Variable

From the axioms for a probability function, we can see that any continuous p.d.f. must satisfy

$$f_X(x) \geq 0 \qquad \forall x \in \mathbb{R}$$

and

$$\int_{-\infty}^{\infty} f_X(x) = 1$$

Hence, if we want to know $P(X \in A)$ for $A = [a, b] \subset \mathbb{R}$, we can compute

$$P(X \in [a, b]) = P(a \leq X \leq b) = \int_a^b f(t)dt$$

Remark 1 *Note that for any $x \in \mathbb{R}$,*
$$P(X = x) = 0$$

if X has a continuous distribution.

This may seem a little counterintuitive in part also because we use continuous distributions to approximate things which are in fact discrete (e.g. income or time unemployed). So far we haven't seen any examples of continuous random variables in any of the probabilities we computed.

3   Examples

Suppose that a random variable is such that on some interval $[a, b]$ on the real axis, the probability of X belonging to some subinterval $[a', b']$ (where $a \leq a' \leq b' \leq b$) is proportional to the length of that subinterval.

Definition 6 *A random variable X is* uniformly *distributed on the interval $[a, b]$, $a < b$, if it has the probability density function*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

In symbols, we then write

$$X \sim U[a, b]$$

Figure 4: p.d.f for a Uniform Random Variable, $X \sim [a, b]$

For example, if $X \sim U[0, 10]$, then

$$P(3 < X < 4) = \int_3^4 f(t)dt = \int_3^4 \frac{1}{10}dt = \frac{1}{10}$$

What is $P(3 \leq X \leq 4)$? Since the probability that $P(X = 3) = 0 = P(X = 4)$, this is the same as $P(3 < X < 4)$.

Example 5 *Suppose X has p.d.f.*

$$f_X(x) = \begin{cases} ax^2 & \text{if } 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

What does a have to be? - since $P(X \in \mathbb{R}) = 1$, the density has to integrate to 1, so a must be such that

$$1 = \int_{-\infty}^{\infty} f_X(t)dt = \int_0^3 at^2 dt = \left[\frac{ax^3}{3}\right]_0^3 = \frac{27}{3}a - 0 = 9a$$

Therefore, $a = \frac{1}{9}$.
What is $P(1 < X < 2)$? - let's calculate the integral

$$P(1 < X < 2) = \int_1^2 \frac{t^2}{9}dt = \frac{2^3}{9 \cdot 3} - \frac{1^3}{9 \cdot 3} = \frac{7}{27}$$

What is $P(1 < X)$?

$$P(1 < X) = \int_1^{\infty} f_X(t)dt = \int_1^3 \frac{t^2}{9}dt = \frac{27 - 1}{27} = \frac{26}{27}$$