

# 1 Discrete random variables

discretervs.tex draft, October 6, 2009

## 1.1 What is a discrete random variable?

A random variable is a discrete random-variable if it is a variable that can take only a *countable* number of values. Or to use the words of MGB, "A random variable  $X$  will be defined to be discrete if the range of  $X$  is countable."

I will use  $N$  to denote the countable number of values.  $X$  is the rv, and, depending on the context, use  $x$  or  $x_i$  to denote a specific value of  $X$ .

Not that it is not correct to say that the a discrete rv can take only a finite number of values. Consider for example the density function  $f(x) = \frac{1}{N}$ , where  $x = 1, 2, 3, \dots, N$ . This is a different density function for every value of  $N$ , no matter how large. This function can have mass at an infinite number of integers.

The probability density function for a discrete density function specifies which points have positive probability (mass) and what those probabilities are.

Note that the value of  $f(x)$  is the probability that  $X = x$ ; that is  $f(x) = \Pr[x]$ . This was not the case for continuous rvs: for a continuous rv,  $f(x)$  is not a probability.

If  $X$  is a discrete rv, we call  $f(x)$  a *discrete density function* or a *probability density function* or a *probability function*.

The probability density function is set of spikes (vertical lines) on the real line, a spike at each value of  $X$  that has positive probability. The height of the spike at  $X = x$  is the probability that  $X = x$ . The function takes the value of zero at all other points.

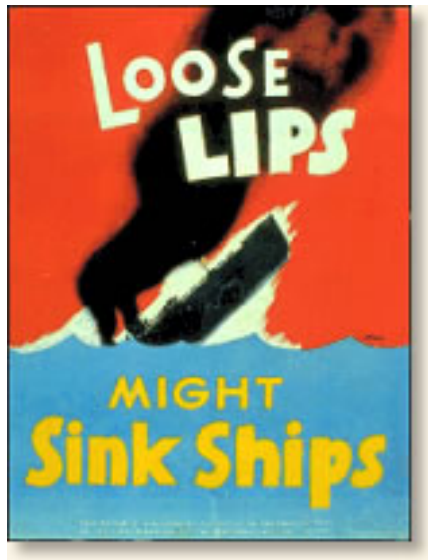
**insert example graph here**

$f(x_i) = \Pr[x_i] \geq 0$  for all possible values of  $X$ . So, it is required that

$$\sum_{i=1}^N \Pr[x_i] = 1$$

since the probability of something happening is one. Not that this implies that  $\Pr[x_i] \not\geq 1$ .

With discrete random variables, integration is "replaced" by summation. Of course, we can speak *loosely* and think of summization as a special type of integration, and then use the term *integration* to refer to both intergration and summation. One can always speak loosely. Of course,



With discrete distributions, there is a probability associated with each point in the real line (not like with continuous rvs) but most of these probabilities will be zero.

MGB formally defines a discrete density function as

**Definition 1** Any function  $f(\cdot)$  with domain the real line and counterdomain  $[0, 1]$  is defined to be a discrete density function if for some countable set  $x_1, x_2, \dots, x_n, \dots$ ,

$$f(x_j) > 0 \text{ for } j = 1, 2, \dots$$

$$f(x) = 0 \text{ for } x \neq x_j; j = 1, 2, \dots$$

$$\sum_{j=1}^N f(x_j) = 1, \text{ where the summation is over the points } x_1, x_2, \dots, x_n, \dots,$$

## 1.2 The expected value and variance of a discrete rv in terms of $f()$

$$\begin{aligned} E[X] &= \sum_{i=1}^N x_i \Pr[x_i] \\ &= \sum_{i=1}^N x_i f(x_i) \end{aligned}$$

and

$$\begin{aligned} \text{var}[X] &= \sum_{i=1}^N (x_i - E[x])^2 \Pr[x_i] \\ &= \sum_{i=1}^N (x_i - E[x])^2 f(x_i) \end{aligned}$$

Further note that

$$\begin{aligned} E[g(X)] &= \sum_{i=1}^N g(x_i) \Pr[x_i] \\ &= \sum_{i=1}^N g(x_i) f(x_i) \end{aligned}$$

These are the discrete analogs of corresponding expectation functions for continuous rv's

## 1.3 Some examples of discrete density functions

### 1.3.1 The Bernoulli distribution



This is Jakob Bernoulli, creator of the Bernoulli distribution - there was a shitload of famous Bernoulli mathematicians, and they did not all get along - a both functional and dysfunctional Swiss family.

Consider a random variable  $x$  that can take one of two values: zero and one, such that  $P(1) = p$  and  $P(0) = (1 - p)$  where  $0 \leq p \leq 1$ . Is  $x$  a continuous or a discrete random variable? Discrete. The density function for  $x$  is

$$P[x] \equiv f_X(x) \equiv f_X(x; p) = \begin{cases} p^x(1-p)^{1-x} & \text{for } x = 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{cases}$$

or, written differently,

$$\begin{aligned} f(1) &= p \\ f(0) &= 1 - p \\ f(x) &= 0 \text{ if } x \text{ does not equal } 0 \text{ or } 1 \end{aligned}$$

Convince me that this is a legitimate density function. Since  $0 < p < 1$ ,  $p$  and  $(1 - p)$  are both positive, the function is never negative. In addition,

$$f_X(0) + f_X(1) = (1 - p) + p = 1$$

The Bernoulli is relevant whenever there are two alternatives, and the probabilities of the two alternatives are  $p$  and  $1 - p$ .

What is  $E[x]$ ?

$$E[x] = \sum_{x=0}^1 x f_X(x) = 0(1 - p) + 1(p) = p$$

What is  $var[x]$ ?

$$\begin{aligned}var[x] &= E[(x - E[x])^2] = \sum_{x=0}^1 (x - E[x])^2 f_X(x) \\&= \sum_{x=0}^1 (x - p)^2 f_X(x) = (0 - p)^2(1 - p) + (1 - p)^2 p \\&= p^2(1 - p) + (1 - 2p + p^2)p = p^2 - p^3 + p - 2p^2 + p^3 \\&= p(1 - p)\end{aligned}$$

Three examples of a random variable that would have a Bernoulli distribution:

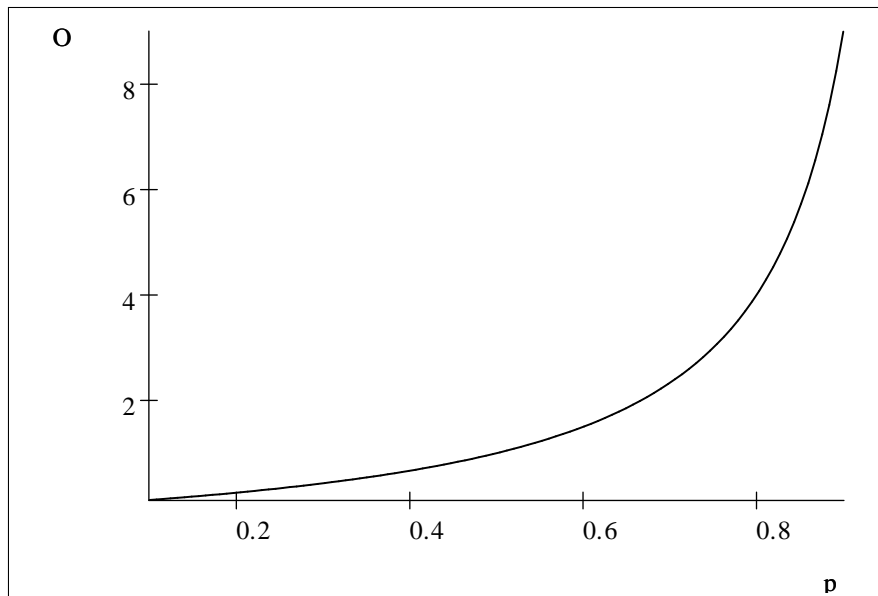
which side of a fair coin results from flipping a coin ( $p = .5$ ), whether a queen is drawn when a card is randomly drawn from a deck ( $p = \frac{4}{52}$ ).

The parameter  $p$  could be the probability that one votes for McCain. In which case  $1 - p$  is the probability that one votes for someone else (Obama, Ralph, ...). Imagine you cannot decide how to vote so make up your mind in the following fashion. You arrive at the polling station with an urn that contains 62 red balls and 38 blue balls, so  $p = .62$ . In the booth you randomly draw a ball from your urn and that determines your vote.

**An aside: the odds** When there are two alternatives,  $x = 1$  and  $x = 0$ , with

$\Pr[x = 1] = p$  (a Bernoulli) it is common to interpret  $x = 1$  as a success. In such common circumstances, statisticians, usually not economists, like to talk about the odds of a success rather than the probability of a success where the odds,  $O$ , are defined as

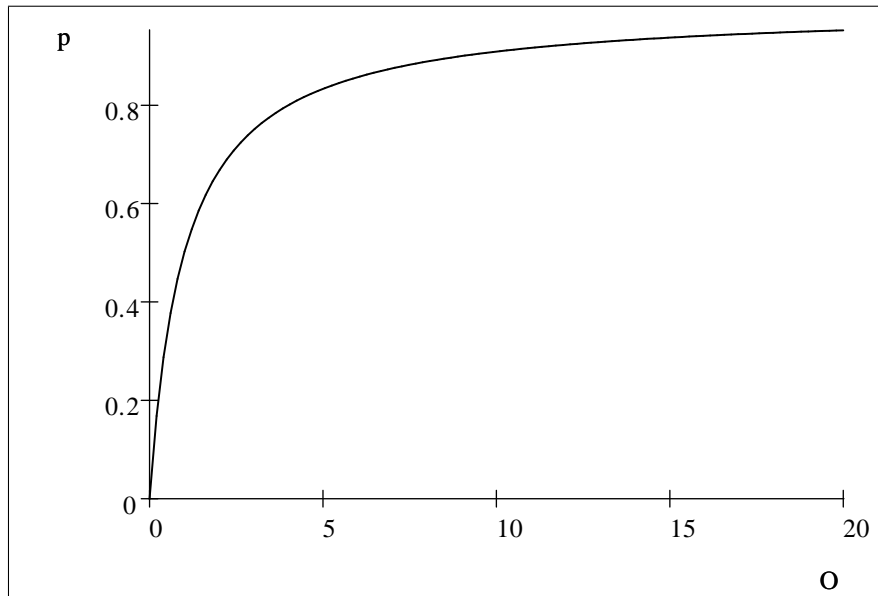
$$O = \frac{p}{1-p}$$



The odds,  $p/(1-p)$ , as a function of  $p$

$$\frac{.1}{1-.1} = 0.11111, \quad \frac{.5}{1-.5} = 1.0, \quad \frac{.6}{1-.6} = 1.5, \quad \frac{.7}{1-.7} = 2.3333$$

Note that  $0 \leq O$ , and  $O$  is undefined for  $p = 1$ . Solving  $O = \frac{p}{1-p}$  for  $p$ ,  
 $p = \frac{O}{(O+1)}$ .



$p$  as a function of the odds,  $O$

So, talking in terms of the odds and talking in terms of  $p$  are two different ways of saying the same thing ( $p \iff O$ ). Note the the odds of a sucess is 1 if  $p = .5$ , less than one if  $p < .5$  and greater than 1 if  $p > .5$ . Think about the odds of your horse winning the race (a sucess).

### 1.3.2 Add your own examples here

### 1.3.3 Assume the rv $X$ has a Poisson distribution

$X$  has a Poisson distribution if

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2, 3, \dots$$

where  $\lambda > 0$ .  $f_X(x) = 0$  if  $X$  is not a positive interger.

The Poisson is a discrete distribution that can take only integer values. It associates positive probability at each integer, so at an infinitie number of points.

It is often the distribution of choice, sometimes incorrectly, if one wants to *count* something; e.g. the number of times American's get married, or, to make a bad pun, the number of fish caught in a day of fishing. We might also use it to predict the number of kids you will have.

For the Poisson<sup>1</sup>

$$E[X] = \lambda = \text{var}[X]$$

can you show this?

Let's assume that  $\lambda = 2.2$  and determine the probability that you will get married  $x$  times.

Using my mathematical software, I got the probability for the first eight integers, starting with zero.

PoissonDen (0; 2.2)

0.110 8; that is, there is a 11% chance one will not get married

PoissonDen (1; 2.2)

0.243 77; that is, there is a 24% change one will marry once

PoissonDen (2; 2.2)

: 0.268 14; that is, there is a 26% chance one will marry twice

PoissonDen (3; 2.2)

: 0.196 64; a 19% chance one will marry thrice

---

<sup>1</sup>It is obviously restrictive to assume that the mean and variance are equal. This restriction can be relaxed by, for example, assuming a negative binomial distribution. See, for example, Green page xxx

PoissonDen (4; 2.2)

: 0.108 15; that is, there is a 10% chance one will marry four times

PoissonDen (5; 2.2)

:  $4.7587 \times 10^{-2}$ ; that is, there is a 4% chance one will marry five times

PoissonDen (6; 2.2)

:  $1.7448 \times 10^{-2}$ ; a 1% chance one will marry six times

PoissonDen (7; 2.2)

:  $5.4838 \times 10^{-3}$ ; a .5% chance that one will marry seven times.

The probability that one will marry twelve times is

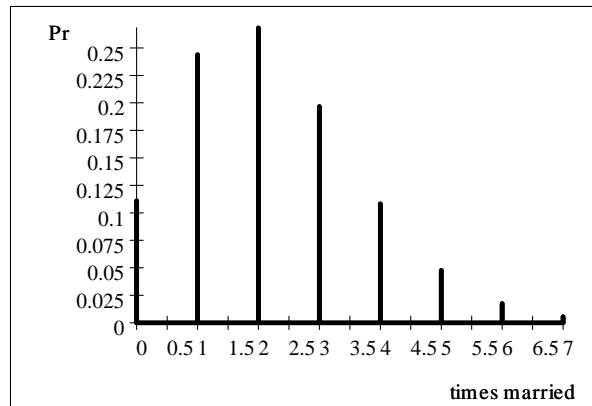
PoissonDen (12; 2.2)

:  $2.9736 \times 10^{-6}$ , which is not much.

Graphing this Poisson for 0 through 7  $\text{PoissonDen}(k; 2.2) = \frac{2.2^k e^{-2.2}}{k!}$

(0, 0, 0, 0.1108, 0, 0, 1, 0, 1, 0.24377, 1, 0, 2, 0, 2, 0.26814, 2, 0, 3, 0, 3, 0.19664, 3, 0, 4, 0, 4, 0.10815, 4, 0, 5, 0, 5,

$4.7587 \times 10^{-2}$ , 5, 0, 6, 0, 6,  $1.7448 \times 10^{-2}$ , 6, 0, 7, 0, 7,  $5.4838 \times 10^{-3}$ , 7, 0)



Poisson distribution with  $\lambda = 2.2$

the axis looks weird, but the plot is correct

Note that  $E[x] = 2.2 = \sigma^2$ .

A question we will address later is estimating  $\lambda$  using a sample. What would be your best estimate of  $\lambda$  if you had a sample of one dead guy who had been married three times?

Now let's make the problem a little more interesting. Assume that  $\lambda$  is a function of *age*; e.g.<sup>2</sup>

$$\lambda = \lambda_0 age_i$$

If, for example,  $\lambda_0 = .0488$  and you are 24,  $\lambda = .0488(24) = 1.1712$ .

In this case, we are talking about the probability of being married  $x$  times given your age.

Using my mathematical software, I got the probability for the first eight integers for a 24-year old, starting with zero marriages.

$$\text{PoissonDen}(0; 1.17)$$

: 0.310370; that is, there is a 31% chance one will not have been married

$$\text{PoissonDen}(1; 1.17)$$

: 0.36313; that is, there is a 36% change one will have been married once

$$\text{PoissonDen}(2; 1.17)$$

: 0.2124; that is, there is a 21% chance for twice

$$\text{PoissonDen}(3; 1.17)$$

:  $8.2848 \times 10^{-2}$ ; that is an 8% chance for one thrice.

In constrast if one is 80,  $\lambda = .0488(80) = 3.904$  and  $\text{PoissonDen}(3; 3.904) = 0.19994$ , a 20% chance of being married three times.

Obviously, one could make  $\lambda$  a function of a whole bunch of things.

---

<sup>2</sup>It is also common to assume a nonlinear function such as  $\lambda = \exp(\lambda_0 age)$ . See, for example, Green page xxx.

## 1.4 One can be very general in their specification of a discrete density function

For example, imagine a world with the discrete rv can only take the interger values 1 through 10. Then specify the density function as  $\Pr[j] = p_j$  where  $0 < p_j < 1$ ,  $j = 1, 2, \dots, 10$  and  $\sum_{j=1}^{10} p_j = 1$ . This probability function has nine parameters and imposes no restrictions on the functional form. Such a specification can be described as *non-parameteric*: given that the rv is restricted to 10 integer values, no restrictive assumption are imposed on  $f_X(X)$ .