Math 481/581 Mathematical Statistics I.

Notes 1

p.1

Random Variables.

A random variable X is a characteristic (= function on a sample space) that defines events (event = subset of sample space) by means of the statement $\{X \in A\}$, for some set A of real numbers. For example, if the sample space is
$\Omega = \{ttt, tth, tht, htt, thh, hth, hht, hhh\}$
corresponding to the experiment of tossing a coin 3 times, and if X is the number of heads minus the number of tails, then for $A = \{1,3\}$ we have

$$\{\omega \in \Omega : X(\omega) \in \{1,3\}\}$$
$$= \{tth, tht, htt, hhh\} \subseteq \Omega$$

Further we assume that a probability exists on Ω.
So $P\{X \in \{1,3\}\} = P\{tth, tht, htt, hhh\}$.
If we assume equal probabilities for each sample point $\omega \in \Omega$, then we obtain the probability $\frac{4}{8} = \frac{1}{2}$.

Note: We often refer to a random variable without reference to a sample space. All we require is the probabilities $P\{X \in A\}$, $A \subseteq \mathbb{R}$.

Discrete random variable

A discrete random variable is one that takes on only finitely or countably many values: there is a sequence $x_0, x_1, x_2, \ldots$ of real values so that $P(X = x_i) > 0$ for each i and

$$\sum_{i=0}^{\infty} P(X = x_i) = 1.$$

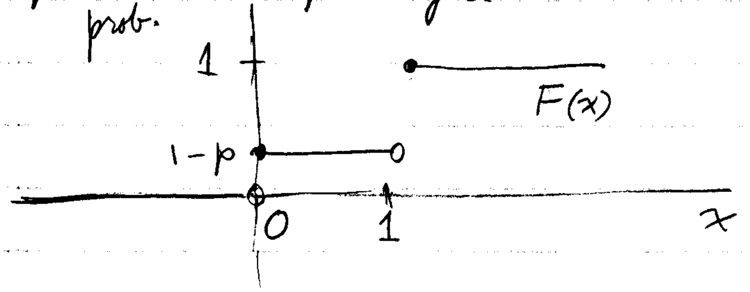
The simplest example is a Bernoulli random variable that takes values 0 and 1 only: $P(X=1) = p$ and $P(X=0) = 1-p$, for some $p \in [0,1]$.

The probability mass function or frequency function of a <u>discrete random variable</u> X is: $p(x) = P(X = x)$.
The cumulative distribution function (cdf) of <u>any</u> random variable is defined by:

$$F(x) := P(X \le x)$$
$$= P\{X \in (-\infty, x]\}, \quad x \in \mathbb{R}.$$

For the case of a <u>discrete random variable</u> this is simply:
$$F(x) = \sum_{x_i \le x} p(x_i)$$

<u>Example</u>. The cdf of a Bernoulli r.v. is shown:



Here $x_0 = 0$ and $x_1 = 1$
So,
$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ p(0) = 1-p, & \text{if } 0 \le x < 1, \\ p(0) + p(1) = 1, & \text{if } x \ge 1. \end{cases}$$

<u>Binomial random variable</u>: The number of successes in n independent trials with probability of success on each individual trial a constant p.
In this case we have a discrete random variable taking values $k \in \{0, 1, 2, \ldots, n\}$ with frequency function $P(X = k) = p(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, 2, \ldots, n$.
See Figure 2.3 of text.

<u>2.1.2 Example B</u>. A <u>majority decoder</u> determines the value of a binary digit that is transmitted n times in a row over a noisy communications channel. If n is odd then there is a strict majority of received values of the n transmissions. (over)

(p. 3) 2.1.2 Example B (continued) However the majority value may be incorrect. We calculate the probability of a truthful decoding of a single digit message for each of the cases $n=1,3,5$ of repeated transmissions. We assume that each transmission is subject to error in reception with prob. p, independently of all other transmissions.

Thus the total number of errors is binomially distributed with parameters n and p. We want the probability that fewer than one half of the n transmissions are received in error.

$n=1$ $\quad P(0 \text{ errors}) = p(0) = \binom{1}{0} p^0 (1-p)^1 = \boxed{1-p}$

$n=3$ $\quad P(0 \text{ or } 1 \text{ error}) = p(0) + p(1)$

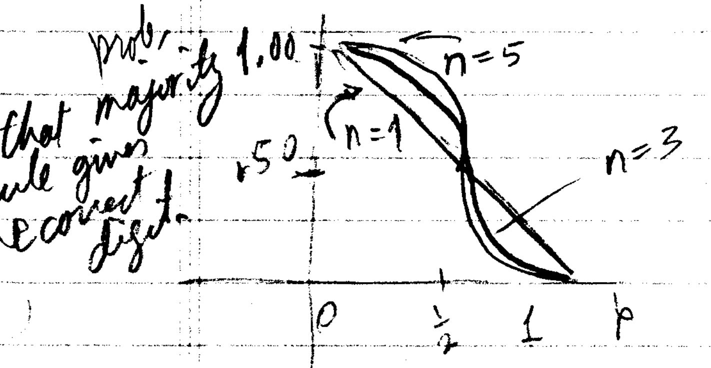
$$= \binom{3}{0} p^0 (1-p)^3 + \binom{3}{1} p^1 (1-p)^2$$

$$= (1-p)^3 + 3p(1-p)^2 = \boxed{(1-p)^2 [1+2p]}$$

$n=5$ $\quad P(0 \text{ or } 1 \text{ or } 2 \text{ errors}) = p(0) + p(1) + p(2)$

$$= \binom{5}{0} p^0 (1-p)^5 + \binom{5}{1} p^1 (1-p)^4 + \binom{5}{2} p^2 (1-p)^3$$

$$= (1-p)^5 + 5p(1-p)^4 + 10p^2(1-p)^3 = \boxed{(1-p)^3(1+3p+6p^2)}$$



If $p = 0.1$ then for

$n=1$: $1-p = .9000$

$n=3$: $(1-p)^2(1+2p) = .9720$

$n=5$: $(1-p)^3(1+3p+6p^2) = .99144$

Geometric Distribution.

Suppose a coin has prob. p of landing up heads. We toss the coin repeatedly until heads appears. In other words we perform independent Bernoulli trials with probability p of success on each trial until a success occurs. Let X be the number of trials up to and including the trial on which success occurs. Thus X takes values from the set $\{1, 2, 3, \ldots\}$. The probability of the event "$X = k$" for a given positive integer k can be calculated by noting that $X = k$ means exactly that there are precisely $k-1$ failures followed by 1 success. So $P(X = k) = P(F_1 \ldots F_{k-1}\, S_k) = (1-p)^{k-1} p^1$.

(where the subscript denotes the trial number)

Note the following "proof" of the geometric sum formula: $P(X < \infty) = 1 \Leftrightarrow$

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p = 1 \underset{l=k-1}{\Leftrightarrow} \sum_{l=0}^{\infty} (1-p)^l = \frac{1}{p}$$

Put $p = 1-r$ to obtain $\displaystyle\sum_{l=0}^{\infty} r^l = \frac{1}{1-r}$, $r \in (0,1)$

Otherwise stated the geometric sum formula gives that X is a proper (finite valued) random variable.

Negative Binomial Distribution

Suppose instead that X is the waiting time up to and including the time of the 2nd success. We must calculate the number of sequences of $k-2$ F's and 2 S's where the last element of the sequence is an S. There are $\binom{k-1}{1}$ such sequences. Thus $P(X = k) = \binom{k-1}{1} p^2 (1-p)^{k-2}$, $k \geq 2$.

p.5

Negative Binomial Distribution (continued).

If finally we let X denote the number of trials until the r^{th} success occurs we have $\binom{k-1}{r-1}$ ways to fill in $(r-1)$ S's in the first $k-1$ trials. Thus $P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$; $k = r, r+1, r+2, \ldots$

Hence $$\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{x-r} = 1$$

or (with $l=k-r$) $$\sum_{l=0}^{\infty} \binom{l+r-1}{r-1} (1-p)^l = \frac{1}{p^r} \iff \begin{matrix} s=r-1 \\ p=1-a \end{matrix}$$

$$\boxed{\sum_{l=0}^{\infty} \binom{l+s}{s} a^l = \frac{1}{(1-a)^{s+1}}}, \quad \forall s=0,1,2,\ldots \quad \forall a \in (0,1).$$

(generalization of the geometric sum formula).

See 2.1.3 Example B for the frequency function of X when $r=2$ and $p=1/9$.

The Poisson Distribution.

We know that the mean and variance of the binomial distribution are respectively np and $np(1-p)$ (See Appendix A). Thus if $p \to 0$ and $n \to \infty$ such that $np \to \lambda$ it makes sense that the binomial distribution will have a limiting distribution with mean λ and variance $\lambda(1-0) = \lambda$, too. This is indeed the case, and the limiting distribution is of a random variable X taking values $k \in \{0,1,2,\ldots\}$ with frequency function $\boxed{p(k) = e^{-\lambda} \frac{\lambda^k}{k!}}, \; k=0,1,2,\ldots$

(See the bottom of p.42 to top of p. 44 in text.)

This is called the Poisson frequency function with parameter (or mean) λ.

(p.6) As shown in 2.1.5 Example A; let X be the number of occurrences of "double sixes" after rolling two fair six-sided dice 100 times. Since the probability p of a double-six on any one roll is $p = 1/36$, then X is distributed as a binomial variable with $n=100$ and frequency parameter $p = 1/36$. Then $\lambda = np = \frac{100}{36} = 2.78$. So we may approximate the frequency function of X by the Poisson frequency function $p(k) = e^{-2.78}(2.78)^k/k!$ The values of the exact binomial frequency function and the values of the Poisson frequency function are shown on p. 44.

Chap 2, Exercise 25 (w/correction) The probability of being dealt a five card hand that is a royal straight flush is

$$p = 4 \Big/ \binom{52}{5} = \frac{4}{2598960} = 1.539 * 10^{-6}.$$

number of hands: $n = 100 * 52 * 20 = 1.04 * 10^5$

Let $X :=$ # of occurrences of a royal straight flush

(a) $P(X=0) = \binom{104000}{0} p^0 (1-p)^{104000} = .8521$

Poisson approximation $\lambda = np = 1.04 * 10^5 * 1.539 * 10^{-6} = .160$. $P(X=0) \approx e^{-\lambda} = e^{-.160} = .8521$

(b) $P(X=2) = \binom{104000}{2} p^2 (1-p)^{103998} = .01091$

Poisson Approx. $e^{-.160} \frac{(.160)^2}{2!} = .01091$

A Poisson Process arises as a distribution of random events in a subset S, where S may be the line (time), plane, or higher dim. space.

The Poisson Process has the property that if $S_1, \ldots, S_n$ are disjoint subsets of S then the numbers of events $N_1, \ldots, N_n$ (an event is indicated by a point in S) in these sets are independently distributed as Poisson r.v.'s that have parameters $\lambda|S_1|, \ldots, \lambda|S_n|$ respectively. Here $|S_i|$ denotes the measure or volume of S_i and $N_1, \ldots, N_n$ are independent means $P(N_1 = k_1, \ldots, N_n = k_n) = P(N_1 = k_1) \cdots P(N_n = k_n)$ for all $k_1, \ldots, k_n$.

λ may be interpreted as the average number of events per unit volume. If telephone calls come into a switchboard as a Poisson process on the time line with parameter $\lambda = .5$ calls per minute then in any fixed 5 minute interval the number of calls is distributed as Poisson with parameter 5 (min.) $*$ .5 calls/min $= 2.5$ (calls).

An interesting statistical problem in the subject of <u>stochastic geometry</u> is whether a given seemingly random pattern of points in space is from a Poisson process or not. (Reference: "Stochastic Geometry and Its Applications, 2nd ed." (1995) by D. Stoyan, W. S. Kendall and J. Mecke.)

The Buffon's Needle problem (Chap 3 Exercise 5) is a prototype problem of Geometrical probability. The field of Stochastic Geometry embraces classes of problems that are generalizations of this type.