ESE 520 Probability and Stochastic Processes

Lecture 4

" One-dimensional random variables and their probability distributions: concrete examples".

A. Discrete random variables/discrete probability distributions.

From the previous discussion we can summarize:

∘ Any discrete probability distribution on $\mathbb{R}$ is a set function which is concentrated on a countable number of points from $\mathbb{R}$. Since any probability distribution on $\mathbb{R}$ is completely described by its corresponding cdf, we can say that any discrete probability distribution has the cdf being a purely constant function with countably many jumps.

∘ If we switch to random variables, then we say that a random variable is a discrete one if its corresponding cdf is a purely constant function with countably many jumps. Alternatively, a random variable is said to be a discrete random variable if it has a countable number of possible values, say a set $D \subset \mathbb{R}$ so that the function $p(x) := P(X = x), x \in D$ has the property that $0 \leq p(x) \leq 1$ for any $x \in D$ and $\sum_{x \in D} p(x) = 1$.

Below are some examples of very important discrete probability distributions often arising in praxis and some (discrete) random variables having such probability distributions:

A1. Binomial probability distribution

Choose a finite set $D = \{0, 1, 2..., n\}$, where $n = 1, 2, ...$ arbitrary but fixed and a real number $0 < p < 1$. For any $x \in D$, define

$$p(x) := \binom{n}{x} p^x (1-p)^{n-x}, \tag{1}$$

where $\binom{n}{x} = \frac{n!}{(n-x)!x!}$, $x! = 1 \times 2 \times ... \times (x-1) \times x$ and $x = 0, 1, ...n$.

It is then not hard to verify (exercise!) that the function $p(x)$ defined in

(1) will satisfy the properties

$$0 \le p(x) \le 1 \text{ for any } x \in D$$

and

$$\sum_{x=0}^{n} p(x) = 1$$

so that the function $p(x), x = 0, 1, ..., n$ uniquely describes a probability distribution on $\mathbb{R}$ (defined as usually $P(A) = \sum_{x \in A} p(x)$ for any subset $A \subset \mathbb{R}$) which is concentrated on the points $x = 0, 1, 2, ..., n$. It is called the *binomial probability distribution.*

Definition 1. A (discrete) random variable X with values in $\{0, 1, ..., n\}$ and the probability distribution determined by the function $p(x)$ given in (1) is said to have a binomial probability distribution with parameters n and p.

Example 1. (”Bernoulli trial”) Consider an experiment of tossing a coin n times independently where in each toss the probability of occuring of ”heads” is p (so that the probability of ”tails” is $(1-p)$). Consider X=number of ”heads” occured. It is then easy to verify that

$$P(X = x) = p(x),$$

where $x = 0, 1, ..., n$ and $p(x)$ is as in (1). The random variable X has then the binomial probability distribution with parameters p and n. The parameter p is often interpreted as the probability of ”success” in a Bernoulli trial.

A2. Poisson distribution

Choose a countable set $D = \{0, 1, 2...\}$ and a fixed real number $\lambda > 0$. For any $x = 0, 1, ...,$ define

$$p(x) := e^{-\lambda} \frac{\lambda^x}{x!}. \tag{2}$$

Again, it can easily be seen that the function $p(x)$ defined in (2) will satisfy the properties

$$0 \le p(x) \le 1, x = 0, 1, ...$$

and

$$\sum_{x=0}^{\infty} p(x) = 1$$

so that the function $p(x), x = 0, 1, ...$ will define a probability distribution on $\mathbb{R}$ concentrated on the countable number of points $x = 0, 1, 2, ...$ It is called the *Poisson probability distribution with parameter* λ.

Definition 2. A (discrete) random variable X with values in the set $\{0, 1, ...\}$ and the probability distribution given in (2) is said to have a Poisson distribution with parameter $\lambda > 0$.

Typically, X is a good model for the number of times a particular event (usually, a rare event) occurs on a given time interval (usually a "unit" time interval as a month, a day, a year etc.) For example, X=number of earthquakes in CA in month of July in a particular year, X=number of phone calls received at an office in a given day, etc...

Next, we consider some concrete problems starting first with one general example showing that any random variable generates on the Borel space a probability measure.

Example 2. Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. That is, X is a measurable map between spaces $(\Omega, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and P is a probability measure on $(\Omega, \mathcal{F})$. For any Borel set B, define

$$\mu(B) := P(w : X(w) \in (B)).$$

Show that μ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Solution: By definition, we have to show three properties:

i) $\mu(\emptyset) = P(w : X(w) \in \emptyset) = P(\emptyset) = 0$;

ii) $\mu(\mathbb{R}) = P(w : X(w) \in \mathbb{R}) = P(\Omega) = 1$;

iii) choose disjoint Borel sets $B_n, n = 1, 2, ...$ Then

$$\mu(\cup_{n=1}^{\infty} B_n) = P(w : X(w) \in \cup_{n=1}^{\infty} B_n) = P(X^{-1}(\cup_{n=1}^{\infty} B_n)) =$$

- we use here the properties of inverse images -

$$P(\cup_{n=1}^{\infty} X^{-1}(B_n)) = \sum_{n=1}^{\infty} P(X^{-1}(B_n) = \sum_{n=1}^{\infty} \mu(B_n)$$

since the sets $X^{-1}(B_n), n = 1, 2, ...$ will be pairwise disjoint sets.

Example 3. (problem 14 from chapter 2 from the book) We are given that all random variables $X_k, k = 1, 2, ..., m$ are independent and have a Poisson distribution with parameter $\lambda > 0$.

We are asked to find the probability $P(\cup_{k=1}^{m}\{X_k < 2\})$.
Using the independence and complement rule yields:

$$P(\cup_{k=1}^{m}\{X_k < 2\}) = 1 - P(\cap_{k=1}^{m}\{X_k \geq 2\}) =$$

$$1 - [P(X_1 \geq 2) \times ....P(X_m \geq 2)] =$$

$$1 - [(1 - P(X_1 \leq 1) \times ....(1 - P(X_m \leq 1)] =$$

$$1 - (1 - (e^{-\lambda} + \lambda e^{-\lambda}))^m.$$

Example 4. Assume that X has *a geometric probability distribution*. That is, X is a discrete random variable with values in the set $\{1, 2, ...\}$ so that

$$P(X = n) = (1 - p)p^{n-1}, n = 1, 2, ...$$

We calculate:

a)

$$P(X > n) = \sum_{k=n+1}^{\infty} (1 - p)p^{k-1} = \sum_{l=0}^{\infty} (1 - p)p^{l+n} =$$

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

where we used the geometric series fact

$$\sum_{l=0}^{\infty} p^l = \frac{1}{1 - p}.$$

b) Using the definition of conditional probability, we obtain

$$P(X > n + k|X > n) = \frac{P(X > n + k, X > n)}{P(X > n)} =$$

$$\frac{P(X > n+k)}{P(X > n)} = \frac{p^{n+k}}{p^n} = p^k$$

for all n and k.

The property in b) is called the *"property without the memory"* of the geometric distribution.

B. Continuous random variables/continuous probability distributions.

A probability distribution on $\mathbb{R}$ is said to be a continuous probability distribution (as noted before, a more correct name would be an absolute continuous probability distribution but since we don't consider here the third class of distributions - so-called singular distributions, we simply use the term continuous) if there is a real-valued function $f : \mathbb{R} \to \mathbb{R}$ satisfying the conditions

$$f(x) \geq 0, x \in \mathbb{R}$$

and

$$\int_{\mathbb{R}} f(x)dx = 1$$

so that the corresponding cdf has the form

$$F(x) = \int_{-\infty}^{x} f(y)dy,$$

where the both last integrals are the usual Riemann integrals.

If we extend that classification to random variables, then we call a random variable X to be a continuous random variable if its cdf $F_X(x)$ has the form

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy. \tag{3}$$

The relation (3) means that $F'(x) = f(x)$.