Statistics MATH414

Vivek Tewary

Discrete Random Variables

A “**discrete random variable**” is a function $X : S \to T$ where S is a sample space equipped with a probability P , and T is a countable (or finite) subset of the real numbers.

One can then define a **probability mass function** $f_X : T \to [0,1]$ on the range of X by $f_X(t) = P(X=t)$.

Given any event A in T, we can then define its probability to be

$$P(A) = \Sigma_{t \in A} f_X(X=t)$$

Common distributions

$X \sim \text{Uniform}(\{1, 2, \ldots, n\})$	$P(X = k) = 1/n$ for all $1 \leq k \leq n$
$X \sim \text{Bernoulli}(p)$	$P(X = 1) = p$ and $P(X=0) = 1-p$
$X \sim \text{Binomial}(n,p)$	$P(X = k) = C(n,k)\, p^k (1-p)^{n-k}$ for all $1 \leq k \leq n$
$X \sim \text{Geometric}(p)$	$P(X = k) = p(1-p)^{k-1}$ for all $k=1,2,3,\ldots$
$X \sim \text{Poisson}(\lambda)$	$P(X = k) = e^{-\lambda} \lambda^k / k!$ for all $k=0,1,2,3,\ldots$

Independence

Two random variables X and Y are independent if $(X \in A)$ and $(Y \in B)$ are independent for every event A in the range of X and every event B in the range of Y .

Two random variables X and Y are independent if $(X \in A)$ and $(Y \in B)$ are independent for every event A in the range of X and every event B in the range of Y .

Example:

It is possible to view the result of each die as a random variable in its own right, and then consider the possible results of the pair of random variables. Let X, Y ~ Uniform({1, 2, 3, 4, 5, 6}) and suppose X and Y are independent. If $x, y \in \{1, 2, 3, 4, 5, 6\}$ what is $P(X = x, Y = y)$?

By independence $P(X = x, Y = y) = P(X = x)P(Y = y) = 1/6 \cdot 1/6 = 1/36$.

For many problems it is useful to think about repeating a single experiment many times with the results of each repetition being independent from every other. Though the results are assumed to be independent, the experiment itself remains the same, so the random variables produced all have the same distribution.

The resulting sequence of random variables X_1 , X_2 , X_3 , . . . is referred to as “i.i.d.” (standing for “independent and identically distributed”).

Conditional, Joint, and Marginal Distributions

Let X be a random variable on a sample space S and let $A \subset S$ be an event such that $P(A) > 0$. Then the probability Q described by $Q(B) = P(X \in B \mid A)$ is called the **"conditional distribution"** of X given the event A.

If X and Y are discrete random variables, the **“joint distribution”** of X and Y is the probability Q on pairs of values in the ranges of X and Y defined by

$$Q((a, b)) = P(X = a, Y = b).$$

In many cases random variables are dependent in such a way that the distribution of one variable is known in terms of the values taken on by another.

Example

Let X ~ Uniform({1, 2}) and let Y be the number of heads in X tosses of a fair coin.

(Y | X = 1) ~ Bernoulli (½)

$$P(Y = 0 \mid X = 1) = \frac{1}{2}$$

$$P(Y = 1 \mid X = 1) = \frac{1}{2}$$

(Y | X = 2) ~ Binomial (2,½)

$$P(Y = 0 \mid X = 2) = \frac{1}{4}$$

$$P(Y = 1 \mid X = 2) = \frac{1}{2}$$

$$P(Y = 2 \mid X = 2) = \frac{1}{4}$$

Example

Let X ~ Uniform({1, 2}) and let Y be the number of heads in X tosses of a fair coin.

(Y | X = 1) ~ Bernoulli (½)

$$P(Y=0 \mid X=1) = \frac{1}{2}$$
$$P(Y=1 \mid X=1) = \frac{1}{2}$$

(Y | X = 2) ~ Binomial (2,½)

$$P(Y=0 \mid X=2) = \frac{1}{4}$$
$$P(Y=1 \mid X=2) = \frac{1}{2}$$
$$P(Y=2 \mid X=2) = \frac{1}{4}$$

	$X=1$	$X=2$
$Y=0$	$1/4$	$1/8$
$Y=1$	$1/4$	$1/4$
$Y=2$	0	$1/8$

Joint Distribution P(X=x and Y=y) computed using the formula

P(Y=y and X=x) = P(Y=y|X=x) P(X=x)

Just because X and Y are dependent on each other doesn't mean they need to be thought of as a pair. It still makes sense to talk about the distribution of X as a random variable in its own right while ignoring its dependence on the variable Y .

	$X = 1$	$X = 2$	Sum
$Y = 0$	1/4	1/8	3/8
$Y = 1$	1/4	1/4	4/8
$Y = 2$	0	1/8	1/8
Sum	1/2	1/2	

Marginal Distribution P(X=x) computed using the formula

P(X=x) = sum over y in P(X=x , Y=y)

Distribution of $f(X)$ and $f(X_1, X_2, . . . , X_n)$

Distribution of $f(X)$ and $f(X_1, X_2, \ldots, X_n)$

$$P(f(X)=a) = P(X \in f^{-1}(\{a\}))$$

$X \sim \text{Uniform}(\{-2, -1, 0, 1, 2\})$

$f(x) = x^2$

Then

$P(f(X)=0) = P(X=0) = 1/5$

$P(f(X)=1) = P(X=1 \text{ OR } X=-1) = 2/5$

$P(f(X) = 4) = P(X=-2 \text{ OR } X=2) = 2/5$

$X, Y \sim \text{Bernoulli}(p)$

$Z = X+Y \sim \text{Binomial}(2,p)$

$X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$

$$Z = X+Y \sim \text{Poisson}(\lambda_1+\lambda_2)$$

Expected value

Let $X : S \to T$ be a discrete random variable (so T is countable). Then the expected value (or average) of X is written as $E[X]$ and is given by

$$E[X] = \Sigma_{t \in T}\, t \cdot P(X = t)$$

provided that the sum converges absolutely.

In this case we say that X has “finite expectation”. If the sum diverges to $\pm\infty$ we say the random variable has infinite expectation. If the sum diverges, but not to infinity, we say the expected value is undefined.

Suppose X is a random variable taking values in the range $T = \{2, 4, 8, 16, \dots\}$ such that $P(X = 2^n) = \frac{1}{2^n}$ for all integers $n \geq 1$.

This is the distribution of a random variable since

$$\sum_{n=1}^{\infty} P(X = 2^n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

But note that

$$\sum_{n=1}^{\infty} 2^n \cdot P(X = 2^n) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1$$

which diverges to infinity, so this random variable has an infinite expected value.

Suppose X is a random variable taking values in the range $T = \{-2, 4, -8, 16, \dots\}$ such that $P(X = (-2)^n) = \frac{1}{2^n}$ for all integers $n \geq 1$.

$$\sum_{n=1}^{\infty} (-2)^n \cdot P(X = 2^n) = \sum_{n=1}^{\infty} (-2)^n \frac{1}{2^n} = \sum_{n=1}^{\infty} (-1)^n.$$

Suppose that X and Y are discrete random variables, both with finite expected value and both defined on the same sample space S. If a and b are real numbers then

(1) $E[a] = a$;
(2) $E[aX] = aE[X]$;
(3) $E[X + Y] = E[X] + E[Y]$;
(4) $E[aX + bY] = aE[X] + bE[Y]$.
(5) If $X \geq 0$ then $E[X] \geq 0$.

Suppose that X and Y are discrete random variables, both with finite expected value and both defined on the same sample space S. If X and Y are independent, then

$$E[XY] = E[X]E[Y].$$

Common Properties

Let $X : S \to T$ be a discrete random variable and define a function $f : T \to U$. Then the expected value of f(X) may be computed as

$$E[f(X)] = \Sigma_{u \in U}\, u \cdot P(f(X) = u)$$

$$= \Sigma_{t \in T}\, f(t) \cdot P(X = t).$$

Variance and standard deviation

Let X be a random variable with finite expected value. Then the variance of the random variable is written as $Var[X]$ and is defined as

$$Var[X] = E[(X - E[X])^2]$$

The standard deviation of X is written as $SD[X]$ and is defined as

$$SD[X] = \sqrt{Var[X]}$$

Let $a \in R$ and let X be a random variable with finite variance (and thus, with finite expected value as well). Then,

(a) $Var[aX] = a^2 \cdot Var[X]$;
(b) $SD[aX] = |a| \cdot SD[X]$;
(c) $Var[X + a] = Var[X]$;
(d) $SD[X + a] = SD[X]$.

Let X be a random variable for which $E[X]$ and $E[X^2]$ are both finite. Then

$$Var[X] = E[X^2] - (E[X])^2 .$$

If X and Y are *independent* random variables, both with finite expectation and finite variance, then

(a) $Var[X + Y] = Var[X] + Var[Y]$; and
(b) $SD[X + Y] = \sqrt{(SD[X])^2 + (SD[Y])^2}$.

Common Properties

Common distributions

Distribution	**Expected Value**	**Variance**
Bernoulli(p)	p	$p(1-p)$
Binomial(n,p)	np	$np(1-p)$
Geometric(p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	λ	λ
Uniform$(\{1,2,\ldots,n\})$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$

Let X be a discrete random variable with finite expected value and finite, non-zero variance. Then $Z = (X - E[X])/SD[X]$ has zero expectation and variance of 1.

Markov and Chebyshev Inequalities

(Markov’s Inequality) Let X be a discrete random variable which takes on only non-negative values and suppose that X has a finite expected value. Then for any $c > 0$,

$$P(X \geq c) \leq E[X]/c$$

(Chebyshev's Inequality) Let X be a discrete random variable with finite, non-zero variance. Then for any $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2$$

where μ is the expectation of X and σ is the standard deviation of X.

Conditional expectation and conditional variance

Let $X : S \to T$ be a discrete random variable and let $A \subset S$ be an event for which $P(A) > 0$. The "conditional expected value" of X given A is

$$E[X|A] = \sum_{t \in T} t \cdot P(X = t|A)$$

and the "conditional variance" of X given A is

$$Var[X|A] = E[(X - E[X|A])^2 |A].$$

A die is rolled. What are the expected value and variance of the result given that the roll was even?

Let X be the die roll. Then $X \sim \text{Uniform}(\{1,2,3,4,5,6\})$, but conditioned on the event A that the roll was even, this changes so that

$$P(X=1|A) = P(X=3|A) = P(X=5|A) = 0 \quad \text{while}$$

$$P(X=2|A) = P(X=4|A) = P(X=6|A) = \frac{1}{3}.$$

Therefore,

$$E[X|A] = 2(\frac{1}{3}) + 4(\frac{1}{3}) + 6(\frac{1}{3}) = 4.$$

Note that the (unconditioned) expected value of a die roll is $E[X] = 3.5$, so the knowledge of event A slightly increases the expected value of the die roll.

The conditional variance is

$$Var[X|A] = (2-4)^2(\frac{1}{3}) + (4-4)^2(\frac{1}{3}) + (6-4)^2(\frac{1}{3}) = \frac{8}{3}.$$

This result is slightly less than $\frac{35}{12}$, the (unconditional) variance of a die roll. This means that knowledge of event A sligthly decreased the typical spread of the die roll results.

Let $X : S \to T$ *be a discrete random variable and let* $\{B_i : i \geq 1\}$ *be a disjoint collection of events for which* $P(B_i) > 0$ *for all* i *and such that* $\bigcup_{i=1}^{\infty} B_i = S$. *Suppose* $P(B_i)$ *and* $E[X|B_i]$ *are known. Then* $E[X]$ *may be computed as*

$$E[X] = \sum_{i=1}^{\infty} E[X|B_i]P(B_i).$$

Covariance and correlation

Let X and Y be two discrete random variables on a sample space S. Then the "covariance of X and Y" is defined as

$$\mathrm{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$$

$$\mathrm{Cov}[X, X] = \mathrm{Var}[X].$$

Let X and Y be discrete random variables with finite mean for which $E[XY]$ is also finite. Then

$$\mathrm{Cov}[X, Y] = E[XY] - E[X]E[Y].$$

	$X = -1$	$X = 0$	$X = 1$
$Y = -1$	$1/15$	$2/15$	$2/15$
$Y = 0$	$2/15$	$1/15$	$2/15$
$Y = 1$	$2/15$	$2/15$	$1/15$

Find covariance

Let X, Y, and Z be discrete random variables, and let $a, b \in R$. Then,

(a) $Cov[X, Y] = Cov[Y, X]$;
(b) $Cov[X, aY + bZ] = a \cdot Cov[X, Y] + b \cdot Cov[X, Z]$;
(c) $Cov[aX + bY, Z] = a \cdot Cov[X, Z] + b \cdot Cov[Y, Z]$; and
(d) If X and Y are independent with a finite covariance, then $Cov[X, Y] = 0$.

The converse of the last statement is false.

The quantity $Cov[X,Y]/\sigma_X \sigma_Y$, where σ_X, σ_Y are standard deviations of X and Y respectively, is known as the "correlation" of X and Y and is often denoted as $\rho[X, Y]$.

Common Properties

Continuous Random Variables

Uncountable sample spaces and densities

If our sample space is uncountable, then we cannot give a positive probability to a singular outcome, since, in that case, additivity of disjoint events would force such a probability function to exceed 1, which we do not allow.

If our sample space is uncountable, then we cannot give a positive probability to a singular outcome, since, in that case, additivity of disjoint events would force such a probability function to exceed 1, which we do not allow.

Hence, we may only prescribe non-zero probabilities to other uncountable subsets of the sample space.

If our sample space is uncountable, then we cannot give a positive probability to a singular outcome, since, in that case, additivity of disjoint events would force such a probability function to exceed 1, which we do not allow.

Hence, we may only prescribe non-zero probabilities to other uncountable subsets of the sample space.

The sample space may be thought of as a subset of the Euclidean space R^n and therefore prescribing probabilities is like prescribing areas/volumes to subsets of R^n. It is impossible to prescribe "volume" to every subset of R^n under reasonable demands such as empty set should have zero volume, volume should not change under translations, volume is more for larger sets, etc.

A reasonable class of events to which a probability can be assigned is the class of **sigma-algebras** which is any collection of events of the sample space S such that

(1) $S \in F$

(2) If $A \in F$ then $A^c \in F$

(3) If $A_1, A_2, \ldots$ is a countable collection of sets in F then $\cup \{ A_n \mid n \in \mathbb{N} \} \in F$

Let S be a sample space and let F be a σ-algebra of S. A “probability” is a function $P : F \to [0, 1]$ such that

(1) $P(S) = 1$;

(2) If $E_1, E_2, \ldots$ are a countable collection of disjoint events in F, then

$$P(\cup E_j) = \Sigma P(E_j)$$

The triplet (S, F, P) is referred to as a probability space.

Let $f : R \to R$ is called a density function if f satisfies the following:

(i) $f(x) \geq 0$,
(ii) f is piecewise-continuous, and
(iii) $\int_R f(x)\, dx = 1$.

Given a density function, we can define a probability on R by $P(E) = \int_E f(x)\, dx$

Continuous Random Variables

Let (S, F, P) be a probability space and let $X : S \to R$ be a function. Then X is a random variable provided that whenever B is an event in R (i.e. a Borel set), $X^{-1}(B)$ is also an event in F.

Let (S, F, P) be a probability space. A random variable $X : S \to R$ is called a continuous random variable if there exists a density function $f_X : R \to R$ such that for any event A in R, $P(X \in A) = \int_A f_X(x)\,dx$. The function f_X is called the **probability density function** of X.

If X is a random variable then its cumulative distribution function $F : R \to [0, 1]$ is defined by $F(x) = P(X \leq x)$.

Common Distributions

X ~ Uniform(a, b)	f(x) = 1/(b-a) for x in (a,b), 0 o/w
X ~ Exp(λ)	$f(x) = \lambda e^{-\lambda x}$ for x > 0, 0 o/w
X ~ Normal(μ, σ^2)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Multiple continuous random variables

Consider the open rectangle in R^2 given by $R = (0, 1) \times (3, 5)$ and $|R| = 2$ denote its area.

Let (X, Y) have a joint density $f : R^2 \to R$ given by

$f(x,y) = ½$ for (x,y) in R and 0 elsewhere, then for any Borel set A in R,

$$P(A) = \iint_A f(x,y)\,dxdy \text{ defines a probability}$$

Thank you