May 7 Lec—Probability Functions Cont., Expected Value, Variance, Bernoulli, Binomial

Properties of probability functions

Numbers 1 and 2 below assume that p_X is the probability function of a discrete random variable X.

1. $\forall x \in \mathbb{R}$ we have $0 \leq p_X(x) \leq 1$ (cuz it's the probability $P(\{X = x\})$, bleh)
2. $1 = \sum_{x \in \mathbb{R}} p_X(x)$ (Looks like summing over an uncountable set, but $p_x(x)$ is only nonzero at countably many points since X (and thus its target set) is discrete. Another way to write this sum is $\sum_{x \in X(S)} p_X(x)$.)
3. Let $p : \mathbb{R} \to [0, 1]$ be a function such that the set $\{x \in \mathbb{R} \mid p(x) \neq 0\}$ is discrete and $\sum_x p(x) = 1$. Then p is the probability function of some discrete random variable X.

Examples

1. Let $p : \mathbb{R} \to \mathbb{R}$ such that

$$p(x) = \begin{cases} 0 & x \notin \{1, 2, \ldots, N\} \\ C \cdot x & x \in \{1, 2, \ldots, N\} \end{cases}$$

   Find a constant C such that p is the probability function of some discrete random variable X.
   Note that $\{x \mid p(x) \neq 0\}$ is finite, therefore discrete.
   We must have $C > 0$. It is enough to have

$$\sum_{n=1}^{N} p_X(n) = 1$$

$$\implies \sum_{n=1}^{N} c \cdot n = 1$$

$$\implies c \cdot \sum_{n=1}^{N} n = 1$$

$$\implies C \cdot \frac{N(N+1)}{2} = 1$$

$$\implies C = \frac{2}{N(N+1)}$$

We saw this example last week with the die example.

2. Let $p : \mathbb{R} \to \mathbb{R}$ such that

$$p(x) = \begin{cases} 0 & x \notin \mathbb{N}^* = \{1, 2, \ldots\} \\ \frac{C}{x(x+1)}, & x \in \mathbb{N}^* \end{cases}$$

Find C such that p is a probability function.
Note that we already see that the set of values over which p is nonzero is discrete. We need

$$\sum_{n=1}^{\infty} \frac{C}{n(n+1)} = 1$$

$$C \cdot \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$$

This example is cooked up such that $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges. You can't set up this kind of question nilly-willy (since series like $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges). We observe (or rather, you should already know) that $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$ because it's a telescoping series. (Don't do this but I think you can also brute force it into a geometric sum.) So $C = 1$ is the solution.

The second example is a useful counterexample in many situations.

Expected Value

Definition: Expected value

Let X be a discrete random variable and p_X be the probability function of X. The **expected value** of X, denoted $\mathbb{E}(X)$, is the number (∞ is not a number, mind you) defined as $\mathbb{E}(X) = \sum_x x \cdot p_X(x)$.

(This term is also sometimes called the *mean* or the *average value* of X, for good reason—given a set of x_i's, the sum $\sum_{i=1}^n x_i \cdot \frac{1}{n}$ is the average of n of the x_i's.)

Digression: How does Minerva calculate the average of a course?

The method is absolutely nonsense: it converts each person's grade to a 4.0-scale, averages that, and converts it back to a letter grade. Averages (expected values) already have much less information than the probability function, let alone Minerva's "average."

Examples

1. Let $x \in \{0, 1, 2\}$ with probability function $p_X(0) = \frac{1}{2}, p_X(1) = \frac{1}{3}, p_X(2) = \frac{1}{6}$. We calculate $\mathbb{E}(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{6} = \frac{2}{3}$.
2. Consider the example we had a moment ago:

$$p(x) = \begin{cases} 0 & x \notin \mathbb{N}^* = \{1, 2, \ldots\} \\ \frac{1}{x(x+1)}, & x \in \mathbb{N}^* \end{cases}$$

We calculate

$$\mathbb{E}(X) = \sum_x x \cdot p(x)$$

$$= \sum_{n=1}^{\infty} n \cdot \frac{1}{n(n+1)}$$

$$= \sum_{n=1}^{\infty} \frac{1}{n+1}$$

which diverges. So $\mathbb{E}(X)$ is not defined (again, ∞ is not a number!!!)

Proposition

Let X be a discrete random variable, and let $F : \mathbb{R} \to \mathbb{R}$. Then $Y = F \circ X = F(X)$ is also a discrete random variable. Moreover,

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(F(X)) \\ &= \sum_x F(x) p_X(x)\end{aligned}$$

Note that if you don't use this then you need to compute $\sum_n n \cdot p_Y(n)$. The proof will be omitted here.

Example

Suppose we have a probability function $p_X(-1) = \frac{1}{4}, p_X(0) = \frac{5}{12}, p_X(1) = \frac{1}{3}$. Let $F : \mathbb{R} \to \mathbb{R}$, $F(x) = x^2$. The target set of $Y = F(X) = X^2$ is $\{0, 1\}$, with $p_Y(0) = \frac{5}{12}$ and $p_Y(1) = \frac{7}{12}$. Here you can see that $\mathbb{E}(Y) = \frac{7}{12}$, which is also what we get if we apply the formula above:

$$\mathbb{E}(Y) = (-1)^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{5}{12} + 1^2 \cdot \frac{1}{3} = \frac{7}{12}$$

Properties of the expected value

Let X be a discrete random variable.

1. If $F : \mathbb{R} \to \mathbb{R}$, $F(x) = C \; \forall x \in \mathbb{R}$, then $\mathbb{E}(F(X)) = C$ (duh).
2. (Linearity—addition) If F and G are two functions, then $\mathbb{E}(F(X) + G(X)) = \mathbb{E}(F(X)) + \mathbb{E}(G(X))$.
3. (Linearity—scalar multiplication) Let $\alpha \in \mathbb{R}$. Then $\mathbb{E}(\alpha \cdot F(X)) = \alpha \cdot \mathbb{E}(F(X))$.

Variance

Definition: Variance

If X is a discrete random variable and we let $\mu_X = \mathbb{E}(X)$, then the **variance** of X is defined as

$$V(X) = \mathbb{E}([X - \mu_X]^2)$$

Proposition: A better formula to compute the variance

$$V(X) = \mathbb{E}(X^2) - \mu_X^2$$

$$= \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Proof

$$V(X) = \mathbb{E}[(X - \mu_X)^2]$$

$$= \mathbb{E}(X^2 - 2\mu_X X + \mu_X^2)$$

By linearity,

$$= \mathbb{E}(X^2) - 2\mu_X \mathbb{E}(X) + \mathbb{E}(\mu_X^2)$$

Since μ_X is a constant, $\mathbb{E}(\mu_X^2) = \mu_X^2$. Then following above:

$$= \mathbb{E}(X^2) - 2\mu_X \mathbb{E}(X) + \mu_X^2$$

$$= \mathbb{E}(X^2) - 2\mu_X \cdot \mu_X + \mu_X \cdot \mu_X$$

$$= \mathbb{E}(X^2) - \mu_X^2$$

Example

Consider the example we had earlier: $p_X(-1) = \frac{1}{4}, p_X(0) = \frac{5}{12}, p_X(1) = \frac{1}{3}$.
We have $\mathbb{E}(X) = \frac{1}{12}, \mathbb{E}(X^2) = \frac{7}{12}$.
Calculating the variance:

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$$= \frac{7}{12} - \left(\frac{1}{12}\right)^2$$

$$= \frac{83}{144}$$

Scribe's aside

There are 3 types of mathematicians, those that can count and those that can't. I'm the third type.

Properties of variance

1. $V(X) \geq 0$
2. $V(X) = 0 \iff X$ is constant
3. $V(\alpha \cdot X) = \alpha^2 \cdot V(X)$ for any $\alpha \in \mathbb{R}$.

Bernoulli Random Variable

We have this kind of experiment:
Flip a coin (not necessarily fair) where $P(\{H\}) = p$, $0 < p < 1$ (so $P(\{T\}) = 1 - p$). Define a discrete random variable $X(T) = 0, X(H) = 1$. The probability function looks like this:

x	$p_X(x)$
0	$1-p$
1	p

Definition: Bernoulli distribution

In the above scenario, X is said to have a **Bernoulli** distribution with parameter p, denoted $X \sim \text{Bernoulli}(p)$ (or just $X \sim Ber(p)$).

Theorem: Expected value and variance of a Bernoulli distribution

If $X \sim Ber(p)$, then

1. $\mathbb{E}(X) = p$

2. $V(X) = p(1-p)$

Proof

1. $\mathbb{E}(X) = 0 \cdot (1-p) + 1 \cdot p = p$
2. $\mathbb{E}(X^2) = 0^2(1-p) + 1^2 p = p$, so
   $\implies V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p - p^2 = p(1-p)$.

Exercise

For any discrete random variable X, and any constant $C \in \mathbb{R}$, we have $V(X+C) = V(X)$.

Example

Let X be a discrete random variable with target set (support) $\{a, b\}$ (where $a, b \in \mathbb{R}$) such that $p_X(a) = 1-p$ and $p_X(b) = p$. This is not a Bernoulli distribution, but it is the transform of a Bernoulli distribution. Indeed, let $Y \sim Ber(p)$, then $X = (b-a)Y + a$. Calculating, we get

$$\begin{aligned} \mathbb{E}(X) &= (b-a)\mathbb{E}(y) + a \\ &= (b-a)p + a \\ &= pb + (1-p)a \end{aligned}$$

and

$$\begin{aligned} V(X) &= V[(b-a)Y + a] \\ &= V[(b-a)Y] \\ &= (b-a)^2 \cdot p(1-p) \end{aligned}$$

Binomial Random Variable

We have an experiment that has two outcomes, S ("success") and F ("failure"). Let $P(\{S\}) = p$, $0 < p < 1$. Repeat the experiment N times, where N is a fixed positive integer, and assume that the different trials are independent. Let X be the number of "successes."
(Note that if $N = 1$ we get our good ol' friend the Bernoulli distribution.)
The support of X is $\{0, 1, \ldots, N\}$. Let us find $p_X(x)$:

- $p_X(0) = P(\{X = 0\}) = P(0 \text{ successes}) = P(\underbrace{FF \ldots F}_{N \text{ times}})$. By independence, $= [P(F)]^N = (1-p)^N$.
- $p_X(1) = P(\{X = 1\})$. There are N events $A_1, \ldots, A_N$ that correspond to this: $A_1 := SFF \ldots F$, $A_2 := FSF \ldots F$, ..., $A_N = FF \ldots FS$. The event $\{X = 1\}$ is the union of these N pairwise disjoint events $A_1, \ldots, A_N$. So $P(\{X = 1\}) = P(A_1) + \cdots + P(A_N) = Np(1-p)^{N-1} = C_1^N \cdot p^1 (1-p)^{N-1}$
- $p_X(2) = P(\{X = 2\})$. There are C_2^N events (ways that 2 successes can be put between $N-2$ failures) of probability $p^2(1-p)^{N-2}$ each. So $p_X(2) = C_2^N p^2 (1-p)^{N-2}$

Generalization: For any $k \in \{0, 1, \ldots, N\}$, we have

$$p_X(k) = C_k^N p^k (1-p)^{N-k}$$

Definition: Binomial distribution

A random variable with the above probability distribution is said to have a Binomial distrbution with parameter N, p. This is written as $X \sim \mathrm{Binomial}(N, p)$.

Scribe's aside

Another common notation that the professor didn't cover is $X \sim B(N, p)$. Because of the brevity of this notation, I will use this notation in my notes.

Proposition

Let $X \sim B(N, p)$.

1. $\mathbb{E}(X) = Np$
2. $V(X) = Np(1-p)$

Proof

1. $$\mathbb{E}(X) = \sum_{k=0}^{N} kp_X(k)$$

   Since $k = 0$ is just a zero term,

   $$= \sum_{k=1}^{N} kp_X(k)$$

   $$= \sum_{k=1}^{N} k \cdot C_k^N p^k (1-p)^{N-k}$$

   Let's try computing $k \cdot C_k^N$ because we really want k to disappear.

   $$k \cdot C_k^N = k \cdot \frac{N!}{k!(N-k)!}$$

   $$= \frac{N!}{(k-1)!(N-k)!}$$

   $$= N \cdot \frac{(N-1)!}{(k-1)!(N-k)!}$$

   $$= N \cdot C_{k-1}^{N-1}$$

   So going back to the original calculation:

   $$\mathbb{E}(X) = \sum_{k=1}^{N} N \cdot C_{k-1}^{N-1} p^k (1-p)^{N-k}$$

   Make a change of variable $l = k - 1$.

   $$= N \cdot \sum_{l=0}^{N-1} C_l^{N-1} p^{l+1} (1-p)^{N-1-l}$$

   $$= N \cdot p \cdot \sum_{l=0}^{N-1} C_l^{N-1} p^l (1-p)^{N-1-l}$$

By the binomial theorem,

$$= Np[p + (1 - p)]^{N-1}$$

Since $p + (1 - p) = 1$,

$$\mathbb{E}(X) = Np$$