IPDA1005 Introduction to Probability and Data Analysis

Worksheet 6
Solution

1. A special case of the more general discrete uniform distribution of a random variable X is

$$p(x) = \begin{cases} \frac{1}{k} & x = 1, 2, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

The above expression applies, for example, to the number of dots appearing on the roll of a fair die. Sketch the pmf of the distribution above for $k = 6$, and identify the mean as the 'balance point' of the distribution. What is its value? Can you show that $E(X) = (k+1)/2$?

Solution: It's pretty straightforward to see that for the case of a fair die, a plot of the pmf will have vertical bars of height $1/6$ at $x = 1, 2, 3, 4, 5, 6$. The mean, or balance point, will be at $3\frac{1}{2}$. For $k \geq 2$,

$$\begin{aligned} E(X) &= \sum_{x=1}^{k} x \cdot p(x) \\ &= \sum_{x=1}^{k} \frac{x}{k} \\ &= \frac{1}{k} \sum_{x=1}^{k} x \\ &= \frac{1}{k}(1 + 2 + \ldots + k) \\ &= \frac{k(k+1)}{2k} = \frac{k+1}{2} \end{aligned}$$

2. Let X be a Bernoulli random variable with pmf

$$p(x; \alpha) = \begin{cases} 1 - \alpha & \text{if } x = 0 \\ \alpha & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Compute $E(X^2)$.

Solution: First of all, $E(X) = 0 \cdot (1-\alpha) + 1 \cdot \alpha = \alpha$, and similarly, $E(X^2) = 0^2 \cdot (1-\alpha) + 1^2 \cdot \alpha = \alpha$.

(b) Show that $\mathrm{Var}(X) = \alpha(1-\alpha)$.

Solution: Using the results above, $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \alpha - \alpha^2 = \alpha(1-\alpha)$.

(c) What is $E(X^n)$ for any positive power n?

Solution: From the calculation of $E(X^2)$ in part (a), you should be able to see that $E(X^n) = \alpha$ for any positive number n.

3. Suppose that $E(X) = 5$ and $E[X(X-1)] = 27.5$.

(a) Determine $E(X^2)$.

Solution: By the linearity of expectations,

$$E[X(X-1)] = E(X^2 - X) = E(X^2) - E(X) = 27.5$$

which implies that

$$E(X^2) = E[X(X-1)] + E(X) = 27.5 + 5 = 32.5.$$

(b) What is $\mathrm{Var}(X)$?

Solution: $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 32.5 - 5^2 = 7.5$.

(c) What is the general relationship among $E(X)$, $E[X(X-1)]$, and $\mathrm{Var}(X)$?

Solution: Substituting (a) into (b) yields $\mathrm{Var}(X) = E[X(X-1)] + E(X) - [E(X)]^2$.

4. The binomial random variable X has pmf

$$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Use the definition of $E(X)$ to show that $E(X) = np$. (**Hint:** Remember that any binomial distribution is a proper pmf so its sum over the set of values of the random variable is one.)

Solution:

Using the definition of $E(X)$,

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} x \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} np \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \end{aligned}$$

If we now let $y = x - 1$ and $m = n - 1$, this becomes

$$E(X) = np \sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{m-y} = np$$

since the last summation is the sum of all the values of a binomial distribution with parameters m and p, which is equal to one.

5. When circuit boards used in the manufacture of Blu-ray players are tested, the long-run percentage of defectives is 5%. Let X denote the number of defective boards in a random sample of size $n = 25$, so $X \sim \text{Bin}(25, 0.05)$.
(**NB**: You will need to use the binomial tables to answer parts (a)–(d) of this question.)

   The table provides the CDF $F(x)$ of binomial distributions for different values of n, the number of trials. We want the table for $n = 25$.

   (a) Determine $P(X = 2)$.

   Solution: In general, $P(X = a) = p(a) = F(a) - F(a-1)$, so $P(X = 2) = F(2) - F(1) = 0.873 - 0.642 = 0.231$.

   (b) Determine $P(X \geq 5)$

   Solution: Recall that the cdf $F(x) = P(X \leq x)$, so to use the table for calculating $P(X \geq 5)$, we calculate $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - F(4) = 1 - 0.993 = 0.007$.

   (c) Determine $P(1 \leq X \leq 4)$.

Solution: $P(1 \leq X \leq 4) = F(4) - F(0) = 0.993 - 0.277 = 0.715.$

(d) What is the probability that none of the 25 boards is defective?

Solution: $P(X = 0) = P(X \leq 0) = F(0) = 0.277.$

(e) Calculate the expected value and standard deviation of X.

Solution: $E(X) = np = 25(.05) = 1.25$ and $\text{Var}(X) = np(1-p) = 1.1875.$ $SD(X) = \sqrt{1.1875} = 1.09.$

6. Let X be a binomial random variable with fixed n and success probability p. As you saw above, its mean is np. Its variance is $\text{Var}(X) = np(1-p)$.

(a) Are there values of p $(0 \leq p \leq 1)$ for which $\text{Var}(X) = 0$? Explain why this is so.

Solution: $\text{Var}(X) = np(1-p) = 0$ when either $p = 0$ (where every trial is a failure so there is no variability in X) or $p = 1$ (where every trial is a success and again there is no variability in X).

(b) For what value of p is $\text{Var}(X)$ maximized?

Solution: You can either plot $\text{Var}(X)/n$ for different values of p to identify the maximum, or you can differentiate $\text{Var}(X)$ with respect to p, i.e, $\frac{d}{dp}[np(1-p)] = n[(1)(1-p) + p(-1)] = n[1-2p] = 0 \Longrightarrow p = 0.5$, which can be easily seen to correspond to a maximum value of $\text{Var}(X)$.

7. Derive the expression for the pmf of a negative binomial random variable X with parameters r (desired number of successes) and p (probability of a success) by using its relationship to a binomial pmf, i.e.,

$$nb(x; r, p) = \frac{r}{x+r} \cdot b(r; x+r, p)$$

Here the functions nb and b respectively represent the binomial and negative binomial PMF with the quantity before the semicolon representing the variable and the quantities after the semicolon representing the parameters.

Solution: If X has a negative binomial distribution, it is the number of failures before obtaining r successes when the probability of a success is p.

$$\begin{aligned}
\frac{r}{x+r} \cdot b(r; x+r, p) &= \frac{r}{x+r}\binom{x+r}{r} p^r (1-p)^{x+r-r} \\
&= \frac{r}{x+r}\frac{(x+r)!}{x!r!} p^r (1-p)^x \\
&= \frac{(x+r-1)!}{x!(r-1)!} p^r (1-p)^x \\
&= \binom{x+r-1}{r-1} p^r (1-p)^x \\
&= nb(x; r, p)
\end{aligned}$$

8. Suppose that $p = P(\text{female birth}) = 0.5$. A couple wishes to have exactly two female children in their family. Let Y be the number of children that the couple has until they have two girls.

   (a) Write out the pmf of Y.

   Solution: Here $r = 2$, and we are looking at the number of trials rather than the number of failures. So

   $$P(Y = y) = (y-1)(0.5^2)(0.5^{y-2})$$

   (b) What is the probability that the family has four children?

   Solution:
   $$P(Y = 4) = 3(0.5^2)(0.5^2) = 0.1875$$

   (c) What is the probability that the family has at most four children?

   Solution:

   $$P(Y \leq 4) = \sum_{y=2}^{4} P(Y = y) = 0.25 + 0.25 + 0.1875 = 0.6875$$

   (d) What is the probability that the family has x male children?

Solution: This is the number of failures before 2 successes (X) and hence has negative binomial distribution. So from the negative binomial PMF,

$$P(X = x) = \binom{x+1}{1} p^2(1-p)^x = (x+1)(.5)^{x+2}$$

(e) How many children would you expect this family to have? How many male children would you expect this family to have?

Solution: $E(Y) = r/p = 2/0.5 = 4$, which makes intuitive sense—if the probability that a baby is a girl is 0.5, then we would expect, on average, the family to have 4 children if two of them are girls.
$E(X) = rq/p = 2$ or, alternatively, If $E(X) = E(Y) - 2 = 2$.

(*Sources: All questions adapted from Devore and Berk (2012), except for Question 6.*)

Bibliography

1. Devore, J.L. and Berk, K.N. (2012) *Modern Mathematical Statistics with Applications.* Springer: New York.