Section 3.7 Geometric Distributions

Tuesday, September 13, 2016 9:34 AM

Section 3.7 Geometri...

Appealing to the variance computing formula, we have

$$V(Y) = E(Y^2) - [E(Y)]^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

NOTE: WMS derive the binomial mean and variance using a different approach (not using the mgf). See pp 107-108. □

Example 3.15. Artichokes are a marine climate vegetable and thrive in the cooler coastal climates. Most will grow in a wide range of soils, but produce best on a deep, fertile, well-drained soil. Suppose that 15 artichoke seeds are planted in identical soils and temperatures, and let Y denote the number of seeds that germinate. If 60 percent of all seeds germinate (on average) and we assume a $b(15, 0.6)$ probability model for Y, the mean number of seeds that will germinate is

$$E(Y) = \mu = np = 15(0.6) = 9.$$

The variance of Y is

$$V(Y) = \sigma^2 = np(1-p) = 15(0.6)(0.4) = 3.6 \text{ (seeds)}^2.$$

The standard deviation of Y is $\sigma = \sqrt{3.6} \approx 1.9$ seeds. □

BERNOULLI DISTRIBUTION: In the $b(n, p)$ family, when $n = 1$, the binomial pmf reduces to

$$p_Y(y) = \begin{cases} p^y(1-p)^{1-y}, & y = 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

This is called the **Bernoulli distribution**. Shorthand notation is $Y \sim b(1, p)$ or $Y \sim \text{Bern}(p)$.

3.7 Geometric distribution

TERMINOLOGY: Envision an experiment where Bernoulli trials are observed. If Y denotes the trial on which the first success occurs, then Y is said to follow a **geometric distribution** with parameter p, where p is the probability of success on any one trial.

# of trials to observe the 1st success

$P(Y=1) = P$

$P(Y=2) = P($1st success on 2nd trial$)$

$= P($failure on 1st trial success on 2nd trial$)$

$= (1-P)P$

$\vdots$

GEOMETRIC PMF: The pmf for $Y \sim \text{geom}(p)$ is given by

$$p_Y(y) = \begin{cases} (1-p)^{y-1}p, & y = 1, 2, 3, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

is it valid?

RATIONALE: The form of this pmf makes intuitive sense; we first need $y - 1$ failures (each of which occurs with probability $1 - p$), and then a success on the yth trial (this occurs with probability p). By independence, we multiply

RATIONALE: The form of this pmf makes intuitive sense; we first need $y-1$ failures (each of which occurs with probability $1-p$), and then a success on the yth trial (this occurs with probability p). By independence, we multiply

$$\underbrace{(1-p)\times(1-p)\times\cdots\times(1-p)}_{y-1 \text{ failures}}\times p=(1-p)^{y-1}p.$$

$P(Y=y)=P(\text{1st till } (y-1)\text{th trials get failures; } y\text{th success})$

$=(1-p)^{y-1}p$

NOTE: Clearly $p_Y(y)>0$ for all y. Does $p_Y(y)$ sum to one? Note that

$$\sum_{y=1}^{\infty}(1-p)^{y-1}p=p\sum_{x=0}^{\infty}(1-p)^x=\frac{p}{1-(1-p)}=1.$$

$X=y-1$

$|1-p|<1$

$$\sum_{x=0}^{\infty} a r^x=\frac{a}{1-r} \text{ if } |r|<1$$

In the last step, we realized that $\sum_{x=0}^{\infty}(1-p)^x$ is an infinite geometric sum with common ratio $1-p$. $\square$

Example 3.16. Biology students are checking the eye color of fruit flies. For each fly, the probability of observing white eyes is $p=0.25$. What is the probability the first white-eyed fly will be observed among the first five flies that are checked?

Step 1:

SOLUTION: Let Y denote the number of flies needed to observe the first white-eyed fly. We can envision each fly as a Bernoulli trial (each fly either has white eyes or not). If we assume that the flies are independent, then a geometric model is appropriate; i.e.,

Step2: $Y\sim \text{geom}(p=0.25)$. We want to compute Step3: $P(Y\le 5)$. We use the pmf to compute

$$\begin{aligned}
P(Y=1)&=p_Y(1)=(1-0.25)^{1-1}(0.25) &=& \ 0.25\\
P(Y=2)&=p_Y(2)=(1-0.25)^{2-1}(0.25) &\approx& \ 0.19\\
P(Y=3)&=p_Y(3)=(1-0.25)^{3-1}(0.25) &\approx& \ 0.14\\
P(Y=4)&=p_Y(4)=(1-0.25)^{4-1}(0.25) &\approx& \ 0.11\\
P(Y=5)&=p_Y(5)=(1-0.25)^{5-1}(0.25) &\approx& \ 0.08.
\end{aligned}$$

Adding these probabilities, we get $P(Y\le 5)\approx 0.77$. The pmf for the geom$(p=0.25)$ model is depicted in Figure 3.3. $\square$

Ex 3.16 $P(Y\le y)=F_Y(y)$

$=\sum_{i=1}^{y} P(Y=i)$

$=\sum_{i=1}^{y}(1-p)^{i-1}p$

$=p\sum_{i=1}^{y}(1-p)^{i-1}$

Let $k=i-1$

$=p\sum_{k=0}^{y-1}(1-p)^k$

$=p\,\frac{1-(1-p)^y}{1-(1-p)}$

$=1-(1-p)^y$

$\sum_{k=0}^{n} r^k = 1+r+r^2+r^3+\cdots+r^n$

$r\sum_{k=0}^{n} r^k = \quad r+r^2+\cdots+r^n+r^{n+1}$

$(1-r)\sum_{k=0}^{n} r^k = 1-r^{n+1}$

$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}$

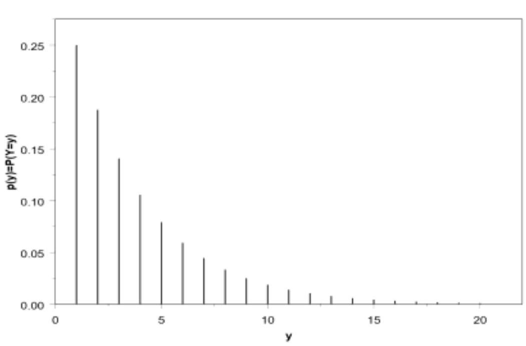

Figure 3.3: *Probability histogram for the number of flies needed to find the first white-eyed fly. This represents the geom$(p=0.25)$ model in Example 3.16.*

$$\boxed{\text{If } Y\sim \text{Geom}(p) \quad P(Y\le y)=1-(1-p)^y}$$

Back to Ex.3.16

$P(Y\le 5)=1-(1-.25)^5$

$=1-.75^5$

$P(3\le Y\le 100)=\sum_{k=3}^{100}(1-p)^{k-1}p$

$=P(Y\le 100)-P(Y<3)$

GEOMETRIC MGF: Suppose that $Y\sim\text{geom}(p)$. The mgf of Y is given by

$$m_Y(t)=\frac{pe^t}{1-qe^t},$$

where $q=1-p$, for $t<-\ln q$.

Proof. Exercise. $\square$ 1st HW Problem

MEAN AND VARIANCE: Differentiating the mgf, we get where $q = 1 - p$, for $t < -\ln q$.

Proof. Exercise. □ 1st HW Problem

MEAN AND VARIANCE: Differentiating the mgf, we get

$$\frac{d}{dt}m_Y(t) = \frac{d}{dt}\left(\frac{pe^t}{1-qe^t}\right) = \frac{pe^t(1-qe^t) - pe^t(-qe^t)}{(1-qe^t)^2}.$$

Thus,

$$E(Y) = \frac{d}{dt}m_Y(t)\bigg|_{t=0} = \frac{pe^0(1-qe^0) - pe^0(-qe^0)}{(1-qe^0)^2} = \frac{p(1-q) - p(-q)}{(1-q)^2} = \frac{1}{p}.$$

Similar (but lengthier) calculations show

$$E(Y^2) = \frac{d^2}{dt^2}m_Y(t)\bigg|_{t=0} = \frac{1+q}{p^2}.$$

$q = 1-p$

① Geom(p)
② pmf $P(Y \le y)$
③ Probability
④ $E(Y)$. $V(Y)$. MGF of Y

Finally,

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1+q}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2}. \square$$

NOTE: WMS derive the geometric mean and variance using a different approach (not using the mgf). See pp 116-117. □

Example 3.17. At an orchard in Maine, "20-lb" bags of apples are weighed. Suppose that four percent of the bags are underweight and that each bag weighed is independent. If Y denotes the the number of bags observed to find the first underweight bag, then $Y \sim \text{geom}(p = 0.04)$. The mean of Y is

# of trials to 1st success

$$E(Y) = \frac{1}{p} = \frac{1}{0.04} = 25 \text{ bags}.$$

The variance of Y is

$$V(Y) = \frac{q}{p^2} = \frac{0.96}{(0.04)^2} = 600 \text{ (bags)}^2. \square$$

3.8 Negative binomial distribution

NOTE: The negative binomial distribution can be motivated from two perspectives:

- as a generalization of the geometric
- as an "inverse" version of the binomial.

TERMINOLOGY: Imagine an experiment where Bernoulli trials are observed. If Y denotes the trial on which the rth success occurs, $r \ge 1$, then Y has a **negative binomial distribution** with waiting parameter r and probability of success p.

NEGATIVE BINOMIAL PMF: The pmf for $Y \sim \text{nib}(r, p)$ is given by

$$p_Y(y) = \begin{cases} \binom{y-1}{r-1}p^r(1-p)^{y-r}, & y = r, r+1, r+2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Of course, when $r = 1$, the nib(r, p) pmf reduces to the geom(p) pmf.

$= P(Y \le 100) - P(Y \le 2)$
$= P(Y \le 100) - P(Y \le 2)$
$= 1 - (1-p)^{100} - (1 - (1-p)^2)$
$= (1-p)^2 - (1-p)^{100}$

Find $E(Y)$. $V(Y)$

① $E(Y) = \sum_{y=1}^{\infty} y P_Y(y)$
$= \sum_{y=1}^{\infty} y (1-p)^{y-1} p$

$V(Y) = E(Y^2) - (E(Y))^2$

$\sum_{y=1}^{\infty} y^2 (1-p)^{y-1} p$

② mgf